Journal of Innovation and Development
ISSN: 2958-4752 | Vol. 10, No. 1, 2025

Rolling Bearing Fault Diagnosis Based on CBAM-WDCNN

Nan Yang[1, *], Zhigang Bing[2] and Di Zhao[3]

[1]School Automation and Electrical Engineering Tianjin University of Technology and Education, Tianjin, China
[2]School Automation and Electrical Engineering Tianjin University of Technology and Education, Tianjin, China
[3]Engineering Teaching Practice Training Center of Tiangong University, Tianjin, China
[*]Corresponding author: y18763619986@163.com

Abstract: Rolling bearing fault diagnosis methods based on convolutional neural networks (CNNs) have demonstrated excellent capabilities in processing complex nonlinear signals. Among them, the one-dimensional convolutional neural network greatly accelerates the signal processing process by directly processing the one-dimensional time-series signal and simplifying the input of the model. However, the traditional 1D convolutional neural network has problems such as poor feature recognition. In this paper, we propose a model based on the fusion of the attention mechanism CBAM and the deep convolutional neural network with a wide kernel in the first layer (WDCNN), which emphasizes the key features and suppresses the irrelevant noises, and significantly improves the model's recognition ability for fault features. The improved method is compared with other diagnostic algorithms to verify its higher accuracy and more stable training process.

Keywords: Bearing fault diagnosis; deep learning; convolutional neural network; convolutional neural attention mechanism.

1. Introduction

The health of rolling bearings is critical to the performance and safety of mechanical systems. As they operate at high speeds and high loads, bearings experience wear and failures that can lead to serious consequences if not addressed in a timely manner [1]. Therefore, accurate troubleshooting is critical to proper equipment operation and cost reduction.

Traditional rolling bearing fault diagnosis methods are mostly based on vibration signal analysis, which identifies different bearing fault modes by extracting the time-domain, frequency-domain and time-frequency-domain features of the signals [2]. For example, Min Lai's team [3] adopted a cyclic spectrum analysis technique based on information entropy to identify faults. For the same rolling bearing as well as gear fault diagnosis problem, Qu et al [4] implemented a composite approach combining dual-tree complex wavelet packet transform and multiple classifiers. While dealing with faults in induction motor rotors, Keskes et al [5] performed fault detection by using an integrated approach of wavelet packet techniques and support vector machines. Although these methods have played an important role in the past decades, they usually require complex signal preprocessing and feature engineering and strongly depend on the experience and knowledge of domain experts. In addition, traditional methods have limitations in dealing with nonlinear and nonsmooth signals, making it difficult to achieve accurate detection of early weak faults.

With the development of deep learning, CNNs have made progress in areas such as image and speech processing, and have been applied to rolling bearing fault diagnosis. Compared with signal analysis, CNN can automatically extract high-level features and improve the accuracy and robustness of fault classification. However, the standard CNN model is designed for two-dimensional images, which requires the conversion of one-dimensional vibration signals and may lose information.

Therefore, this paper proposes to design the first layer as a wide convolutional kernel on the basis of a one-dimensional convolutional neural network, which can help the network to better retain the key information of the input data, and help to improve the overall performance and generalization ability of the model; and add a convolutional neural attention module including channel attention and spatial attention (CBAM), which includes both channel attention and spatial attention, can focus on more important channel features, thus suppressing irrelevant or redundant information.

2. WDCNN Fault Diagnosis Model

2.1. 1D-CNN

Convolutional Neural Network (CNN), is a deep learning network specifically designed to process grid-like data such as images or time series. Its basic structure consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. Its advantage lies in its ability to perform effective feature extraction directly on the input signal in an end-to-end form without having to manually compute various features. The basic structure of CNN is shown in Figure. 1.

As can be seen from Figure.1, each convolutional layer contains multiple convolutional kernels that are used to perform convolutional operations on the input data to extract different features. The convolution kernels are usually learnable parameters that can be optimized according to the data features.

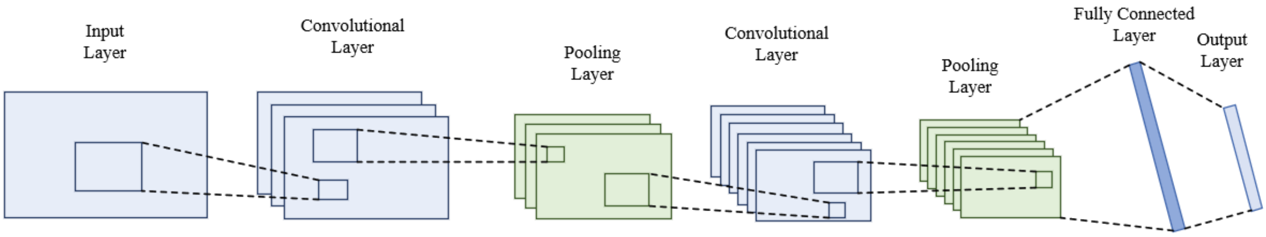

Figure 1. Basic structure of CNN.

The specific operations of the convolutional layer are as follows:

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

where M_j is the input feature vector; l is the layer l network; k is the convolution kernel; b is the network bias; x_j^l is the layer l output and x_i^{l-1} is the layer l input.

The pooling layer, on the other hand, reduces the feature map size and reduces the data dimensionality by downsampling, while retaining the most salient features. In pooling operations, there are usually three methods to choose from, including average pooling, maximum pooling, and minimum pooling. In practice, maximum pooling is the most widely used method. Namely:

$$P_i^{l+1}(j) = \max_{(j-1)W+1 \le t \le jW} \{q_i^l(t)\} \quad (2)$$

where: $q_i^l(t)$ denotes the value of the t neuron in the i feature vector of the l layer, $t \in [(j-1)W+1, jW]$; W is the width of the pooling region; $P_i^{l+1}(j)$ denotes the value corresponding to the neuron of the $l+1$ layer.

The fully connected layer then maps the feature vectors output from the pooling layer to the specified number of categories to complete the classification task.

In this paper, we use 1D-CNN, a one-dimensional convolutional neural network, which is a variant of CNN. It is mainly used to process one-dimensional sequence data, such as audio, text, etc. Local features are extracted from the input sequence by 1D convolutional operations, and more advanced features are gradually extracted by stacking multiple convolutional layers [6]. This structure can operate directly on the original one-dimensional data, retaining the original structure and timing information of the data, which can effectively avoid the problems of losing the timing relationship of the original data and affecting the feature extraction effect that may be caused by the process of converting images.

2.2. 1D-WDCNN

A traditional one-dimensional convolutional neural network is designed to use a larger convolutional kernel size for the first layer to cover more local features. Smaller convolutional kernels are used in the later layers to suppress overfitting while deepening the number of network layers. Compared to ordinary one-dimensional convolutional neural networks, the joint use of large and small convolutional layers allows the model to learn both macroscopic and microscopic features of the data [7], thus improving the model's ability to recognize complex patterns.

The WDCNN model is a neural network composed of four base blocks, which contain a convolutional layer, a bulk normalization layer (BN for short), an activation function layer, and a maximum pooling layer. The output of the previous block is used as the input of the next block, and the convolutional layer in module 1 uses a 15×15 large convolutional kernel, while the convolutional layers in the other modules are all 3×3 small convolutional kernels. After several experimental comparisons, the model parameters were determined as in Table 1.

Table 1. WDCNN model structural parameters.

Layer	Kernel Size	No. of Kernels	Output Shape
Input	——	——	1024
Conv	15	16	1010×16
Conv	3	32	1008×32
Max Pool	2	32	504×32
Conv	3	64	502×64
Conv	3	128	500×128
Adaptive Max Pool	——	128	4×128
Flatten	256	——	256
Flatten	64	——	64

3. Introduction of Convolutional Neural Attention Module

The Convolutional Neural Attention Module is able to adaptively focus on the important information in the input feature map and weight it. It centers on emphasizing the important information in the input feature map through a two-stage refinement process [8] - channel attention and spatial attention. The CBAM module is shown in Figure. 2.

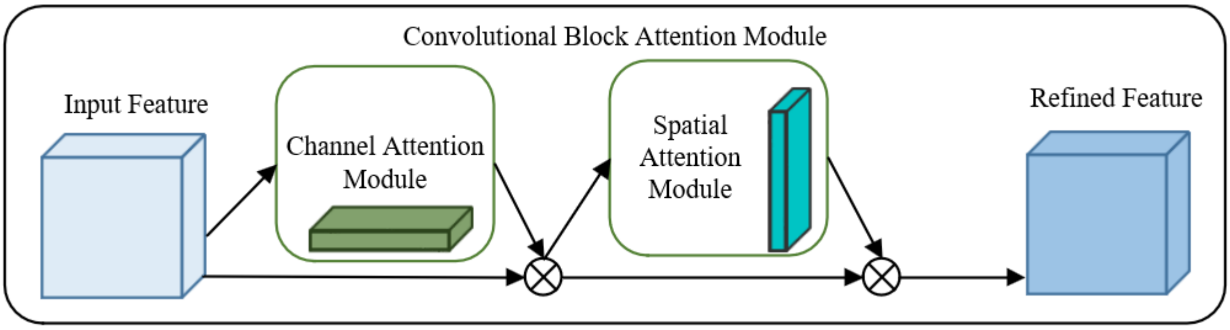

Figure 2. CBAM module.

The channel attention obtains the overall information of each channel through global pooling [9], and then calculates the channel weights through the fully connected layer and sigmoid function. These weights are used to adjust the channel strengths in the original feature map to enhance the useful channels and weaken the irrelevant channels, thus improving the network feature extraction. As shown in Figure 3.

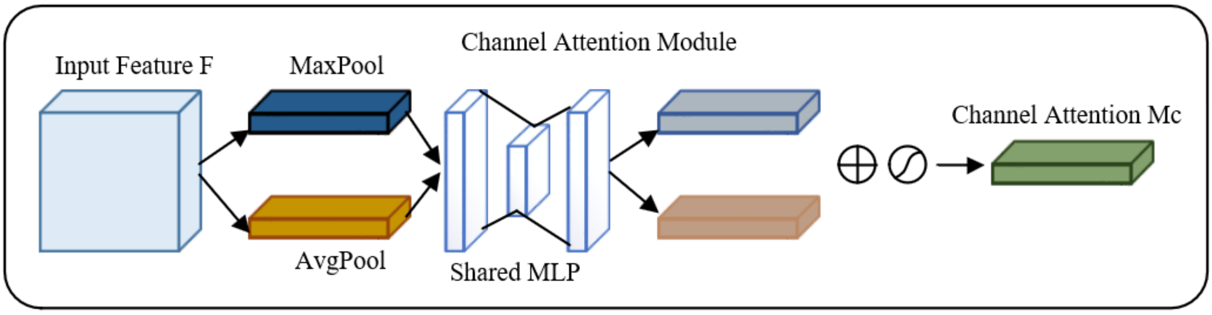

Figure 3. Schematic diagram of the channel attention module.

The spatial attention mechanism performs maximum and average pooling of the feature map to obtain spatial information after channel attention processing and obtains a weighted map by convolutional fusion and sigmoid [10]. This weighted map is multiplied with the original feature map to highlight important spatial regions and enhance target attention and detail capture. As shown in Figure 4.

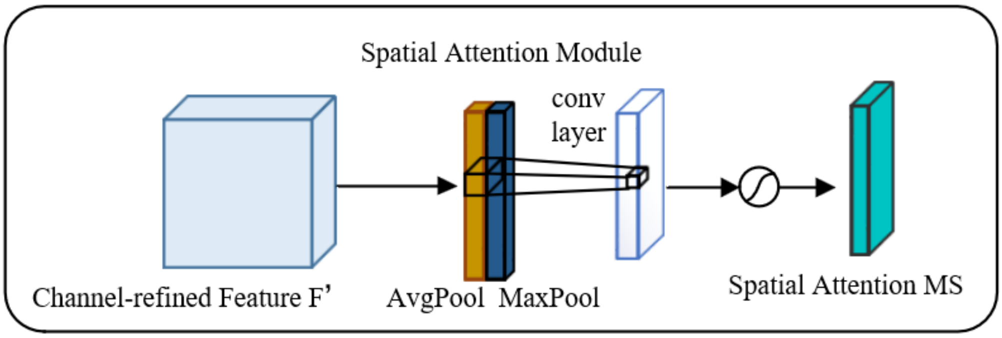

Figure 4. Schematic diagram of the spatial attention module.

4. Research on Fault Diagnosis Method Based on CBAM-WDCNN

4.1. Model structure

The rolling bearing fault diagnosis process based on CBAM-WDCNN is shown in Figure 5. The original vibration signals of rolling bearings are preprocessed and divided by data. The training set data is input into the CBAM-WDCNN model for training. The ideal training effect is achieved by tuning the training model, and finally the validation set data is input to output the final diagnosis results.

In this paper, CBAM-WDCNN model improves the WDCNN network in the previous paper. CBAM-WDCNN includes input layer, convolutional layer, activation layer, BN layer, CBAM, pooling layer and fully connected layer. The CBAM module is added behind the BN layer in the first and third layers as shown in Figure. 6. The convolutional layer extracts features from the input data; ReLU alleviates the problem of vanishing gradients and speeds up the convergence speed when the network is trained; the BN layer, i.e., the batch normalization layer, accelerates the training of the network, reduces the problem of vanishing/exploding gradients, and improves the model's generalization ability; the CBAM module helps the convolutional neural network to be more focused on the useful features [11] while suppressing the unwanted information, thus improving the model's feature extraction ability and generalization performance.

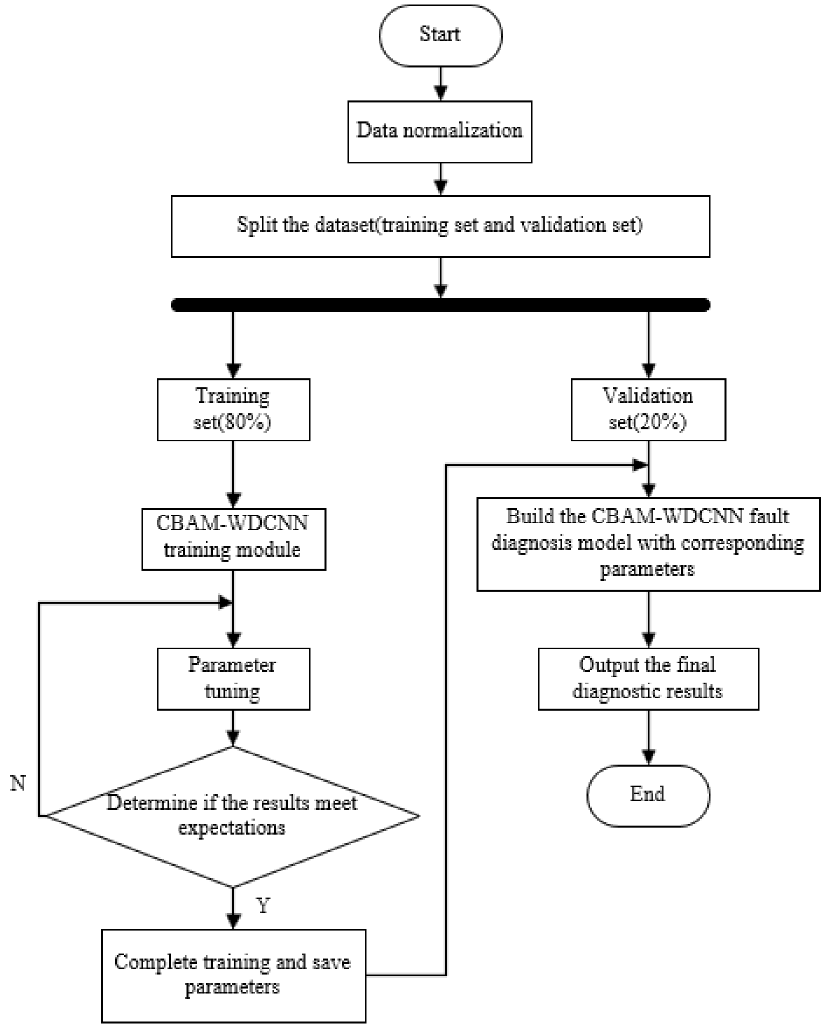

Figure 5. CBAM-WDCNN based rolling bearing fault diagnosis process.

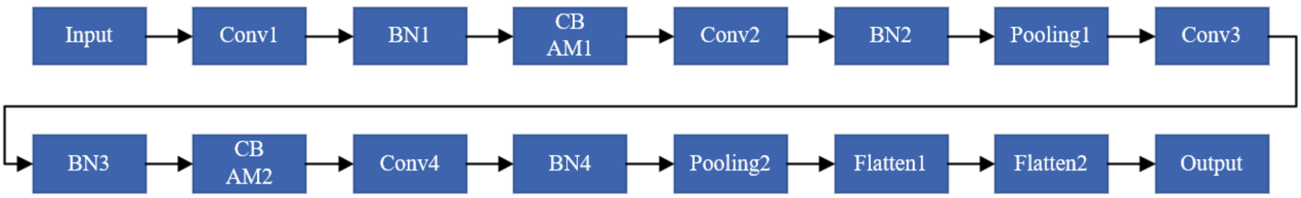

Figure 6. CBAM-WDCNN rolling bearing fault diagnosis model.

4.2. Model Training

PyTorch was used as a deep learning framework in the experiments. In order to train the neural network, the parameters need to be set initially. An initial learning rate of 0.001 was selected, and the Adam optimizer was used to adjust the network parameters to automatically adjust the learning rate during the training process. epoch size was set to 100 times, indicating that the entire dataset will be iterated 100 times, and batch_size was set to 64, which means that 64 samples were selected each time for training.

5. Experimental Results and Analysis

5.1. Experimental data set

In order to validate the practical effectiveness of the constructed CBAM-WDCNN model in bearing fault diagnosis, the widely known Case Western Reserve University (CWRU) bearing dataset was used for rigorous testing. This dataset is widely used as a benchmark for evaluating the performance of bearing fault diagnosis models because it covers a wide range of bearing fault types and different degrees of fault data.

In the dataset used, the drive end bearing is an SKF 6205 with a sampling frequency of 12 kHz and a table speed of 1797 r/min. Bearing failures include inner ring failures, outer ring failures, and rolling element failures, each of which corresponds to three different levels of damage.

In this paper, three sets of inner ring failure data with different damage degrees, three sets of outer ring failure data with different damage degrees, three sets of rolling body failure data with different damage degrees and one set of normal data are selected, and all the above data are divided into training and validation sets according to the ratio of 80% and 20%. The detailed division of the data set is shown in Table 2.

Table 2. Description of the bearing data set.

Fault Location	Fault Diameter(inch)	Fault Size(mm)
Normal	0	0
Ball	0.007	0.1778
Ball	0.014	0.3556
Ball	0.021	0.5334
Inner race	0.007	0.1778
Inner race	0.014	0.3556
Inner race	0.021	0.5334
Outer race	0.007	0.1778
Outer race	0.014	0.3556
Outer race	0.021	0.5334

5.2. Experimental results

Figure 7 shows the comparison of the accuracy curves of the optimized and improved CBAM-WDCNN rolling bearing fault diagnosis model and the WDCNN rolling bearing fault diagnosis model. From the figure, it can be seen that the accuracy of the CBAM-WDCNN model is around 99% after 20 iterations, and the accuracy of the WDCNN model is around 95%.

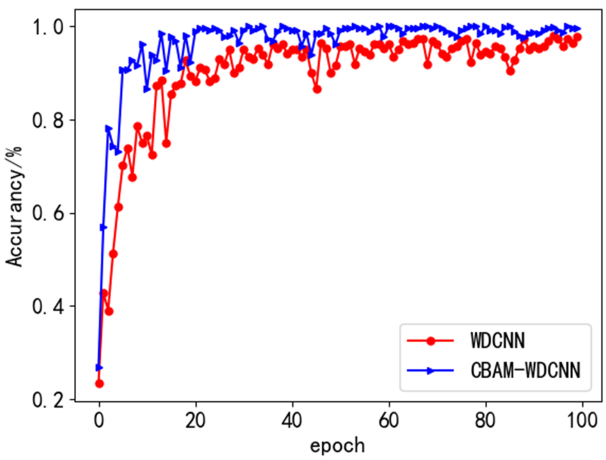

Figure 7. Variation curve of recognition accuracy.

Comparison of the loss function curves of the model before and after optimization is shown in Figure. 8. From the figure, it can be seen that after 20 iterations, the Loss value of the CBAM-WDCNN model is reduced to 0.04, and the model reaches full convergence, and the Loss value of the WDCNN is reduced to 0.1.

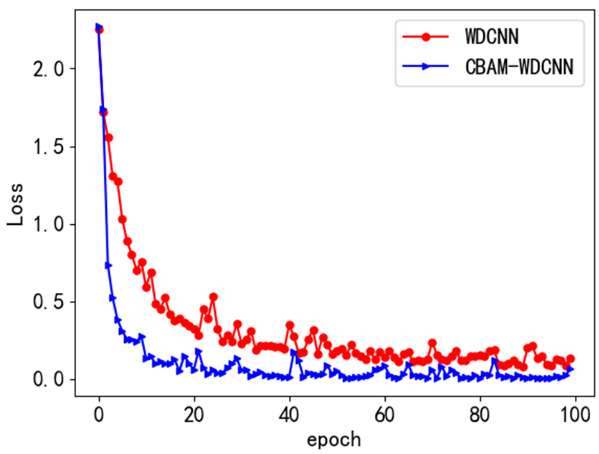

Figure 8. Loss function variation curve.

5.3. Comparison experiments with other algorithms

In order to further verify the effectiveness of the improved CBAM-WDCNN rolling bearing fault diagnosis model, it is experimentally compared with the three models of LeNet, ResNet and AleNet. All the experiments are carried out in the same environment and the same dataset, and the experimental results are shown in Figures. 9 and 10.

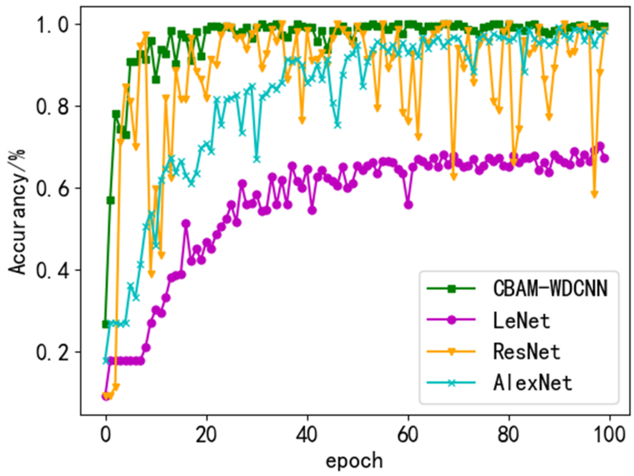

Figure 9. Comparison of model accuracy.

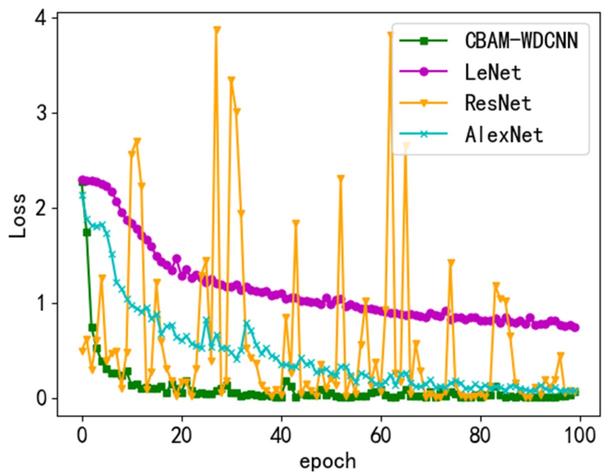

Figure 10. Comparison of loss functions.

As can be seen from Figures 9 and 10, with the increase in the number of training rounds, the accuracies of the four models on the whole show an upward trend, the loss rate of the ResNet model has a larger oscillation, and the loss rates of the other three models drop to a relatively stable interval of oscillation. The accuracy of LeNet stays at about 62%, and the loss value drops to about 0.9.The accuracy of ResNet is around 98%, but with too large ups and downs. ResNet accuracy is around 98% but the fluctuation is too large, and the convergence effect is not good. aleNet is more stable after convergence, the accuracy is around 95% with small oscillations, and the loss value is stable at around 0.1. In contrast, the improved CBAM-WDCNN model in this study has relatively smooth curves and smaller fluctuations than the other models, which is obviously more stable.

6. Conclusion

For the traditional convolutional neural network model in the processing of image data is often time-consuming and complex, easy to ignore the local key features of the original vibration signal, and does not fully take into account the focus of attention between the different channels and spatially important areas and other issues, this paper selects the U.S. Case Western Reserve University bearing data to analyze the fault diagnosis model of rolling bearing, and proposes to design the first layer of the original one-dimensional convolutional neural network. We propose to design the first layer of the original one-dimensional convolutional neural network with a wide kernel and add the attention mechanism CBAM module, and propose a rolling bearing fault diagnosis method based on the CBAM-WDCNN model.

Through the joint use of the larger convolutional kernel in the first layer and the smaller convolutional kernel in the later layers the model can learn both macro and micro features of the data, the CBAM module makes full use of the fault information contained in the vibration signals, and it effectively suppresses the high-frequency noise in the original input signals. By comparing and analyzing the diagnosis results with other different algorithms, it is shown that the network model has higher fault diagnosis accuracy, lower loss value and more stable training process.

References

[1] Li X, Ma Z, Yuan Z, et al. (2024). A review on convolutional neural network in rolling bearing fault diagnosis[J]. Measurement Science and Technology, 35(7). DOI:10.1088/1361-6501/AD356E.

[2] Sha M, Liu L. (2015). A review of bearing fault diagnosis techniques based on vibration signals[J]. Bearing, (09):59-63. https://link.cnki.net/doi/10.19533/j.issn1000-3762.2015.09.017

[3] Li M, Yang J, Wang X. (2015). Information entropy based cyclic spectrum analysis method and its application in rolling bearing fault diagnosis[J]. Journal of Vibration Engineering, 28(01):164-174. https://link.cnki.net/doi/10.16385/j.cnki.issn.1004-4523.2015.01.021

[4] Qu J, Zhang Z, Gong T. (2015). A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion[J]. Neurocomputing, 171837-853. DOI:10.1016/j.neucom.2015.07.020.

[5] Keskes H ,Braham A ,Lachiri Z. (2013). Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM[J]. Electric Power Systems Research, 97151-157. DOI:10.1016/j.epsr.2012.12.013.

[6] Qu J, Yu L, Yuan T, et al. (2018). Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network[J]. Journal of Instrumentation, 39(07):134-143. DOI:10.19650/j.cnki.cjsi.J1803286.

[7] Zhang W. (2018). Research on bearing fault diagnosis algorithm based on convolutional neural network [D]. Harbin Institute of Technology.

[8] Yao Q, Bie S, Yu J, et al. (2022). A bearing fault diagnosis method combining improved Inception V2 module and CBAM[J]. Journal of Vibration Engineering, 35(04):949-957. https://link.cnki.net/doi/10.16385/j.cnki.issn.1004-4523.2022.04.019

[9] Li J, Ma Y, Hu X, et al. (2023). Wind turbine bearing fault diagnosis based on CBAM-InceptionV2-dual-stream CNN[J]. Intelligent Power, 51(06):28-33.

[10] Xu S, Yuan R, Lv Y, et al. (2023). A novel fault diagnosis approach of rolling bearing using intrinsic feature extraction and CBAM-enhanced InceptionNet[J]. Measurement Science and Technology, 34(10). DOI:10.1088/1361-6501/ACE19C.

[11] Du X,Gong B,Yu P,et al. (2022). Fault diagnosis of analog circuits based on CBAM-CNN[J]. Control and Decision Making, 37(10):2609-2618. DOI:10.13195/j.kzyjc.2021.1111.