



# An Improved Convolutional Neural Network Model by Multiwavelets for Rolling Bearing Fault Diagnosis

Gangxing Ren, Jing Yuan<sup>(✉)</sup>, Fengxian Su, Huiming Jiang, and Qian Zhao

School of Mechanical Engineering, University of Shanghai for Science and Technology, 516 Jun Gong Road, Shanghai 200093, China  
yuanjing@usst.edu.cn

**Abstract.** Convolutional neural network (CNN) is increasingly applied to data-driven fault diagnosis of mechanical equipment spare parts. However, CNN training network parameters need a large amount of fault data, and better training effect can be obtained by updating network parameters repeatedly. In this paper, an improved CNN by multiwavelets is introduced multiwavelets into convolution layer, the natural convolution attribute of multiwavelets is fused with convolution layer to fully release the two-channel feature extraction ability of multiwavelets transform. At the same time, we change the parameters of the multiwavelets convolution kernel to discuss the overall diagnostic performance of the network in the same dataset. Thus, different multiwavelets kernel parameters are customized according to different signal characteristics. The feasibility and effectiveness of improved CNN by multiwavelets for case Western Reserve University fault-bearing data are verified.

**Keyword:** Deeping learning · CNN · Multiwavelets · Signal processing · Fault identification

## 1 Introduction

Rotating machinery is an indispensable part of the equipment manufacturing industry, and rolling bearing is an important part of rotating machinery. Its small defects (such as bearing crack and bearing abrasion damage) may lead to catastrophic accidents of the whole mechanical structure. Therefore, the timely and accurate classification of rolling bearing fault types has attracted many scholars [1–3].

The development of data-driven mechanical fault diagnosis benefits from the rapid development of sensing technology, computing systems and information storage technology in recent years. These technologies provide technical guarantees for data acquisition, transmission, and storage in manufacturing systems [4, 5]. However, the abnormal response of rolling bearing caused by fault is usually irrelevant. Due to the complex nonlinear behaviors e.g., friction contact between components, radial clearance of bearing and small vibration, the original time-domain signal show not only transient phenomenon but also nonlinear dynamic effects. Noise and various uncertainties further

exacerbate this situation [6]. Traditional fault diagnosis methods consist of three main phases: Original signal acquisition, feature extraction and selection, fault classification and fault prediction. For example, Yuan et al. proposed high-fidelity noise-reconstructed empirical mode decomposition for mechanical multiple and weak fault extractions [7]. Qiao et al. used empirical mode decomposition, fuzzy feature extraction, and support vector machines to diagnose and verify the faults of steam turbine generator sets under three different working conditions [8]. Chen combined wavelet packet feature extraction technology and machine learning technology, an online monitoring model of logistic regression is proposed. The effectiveness of the scheme is verified by analyzing the tool wear vibration signal [9]. However, the traditional fault diagnosis methods listed above have their limitations. First, traditional methods rely on manual experience selection, and the quality of extracted features directly affects the performance of the final classification algorithm. Therefore, the stability of the diagnosis effect is not reliable. Second, In the current era of big data, a large amount of data needs to be processed in time and quickly. It takes time and effort to construct features manually. Even if the extracted features are used to train the model, the efficiency of the final model prediction or classification still needs to be improved.

Meanwhile, in order to meet the challenges of the big data era, people have done a lot of research on intelligent fault diagnosis methods in recent years. Deep learning models automatically find nonlinear features and realize classification by superimposing multiple network layers, which has achieved promising results in many tasks of artificial intelligence. Along with the development of the internet industry, lots of sensors are used in the operation detection system of mechanical equipment. The large explosion of data has made the traditional fault diagnosis methods difficult to meet the needs of the market, which also aroused people's interest in exploring the application of deep learning methods in fault diagnosis. For example: stacked Auto-Encoders (SAEs) [10], Deep Belief Networks (DBNs) [11], Recurrent Neural Networks (RNNS) [12], Generative Adversarial Networks (GAN) [13]. Due to deep learning's powerful function of nonlinear feature mapping and advantage of end-to-end learning advantages, the results of bearing diagnosis based on deep learning method have been significantly improved.

Convolutional neural network (CNN) is one of the representative deep learning algorithms, it contains convolution layer, pooling layer, and complete connection layer. With the help of these, CNN can directly train the original time-domain vibration signal and identify the fault features hidden in the signal, so as to diagnose and classify different types of faults, this effectively avoids subjective feature selection and human intervention. As such, fault diagnosis technology driven by original data set has become attractive. Weimer et al. adopt deep convolutional neural network, which overcomes the difficulty of redefining manual fault characteristics for each new situation in the production process and improves the automation and accuracy of monitoring [14]. Ince and Abdeljaber et al. use 1DCNN to detect motor faults and this method has higher accuracy than the model-based method [15, 16]. However, on one hand, the accuracy of CNN is extremely dependent on large-scale datasets, for the research of rolling bearing diagnosis, the fault data set is usually limited, and the cost of obtaining a large amount of training data is very high. On the other hand, the performance of the diagnosis method will decline when the training data set contains multiple fault features. The signal is caused by the coupled

composite fault characteristics in the dynamic signals of complex systems. Besides, when the data source distribution in the test set deviates from the training set data source in the target domain, the diagnostic performance of the network will be observed to decline. The signal differences are caused by different monitoring environments in which vibration signals are collected, e.g., different working loads, sensor positions, and rotation speeds. CNN network model is difficult to train the network parameters of multiple characteristic signals, but the actual working conditions are always changing, and the same fault type does not exist in a single form.

Therefore, this paper by introducing the combination of multiwavelets transform and CNN, the ability of multiwavelets containing multiple different frequency domain basis functions to extract multiple faults at the same time, multiwavelets also fully inherit the properties of single wavelets such as the property of orthogonality, symmetry, compact support [17]. Besides, the features extracted from the first layer of the convolutional neural network will affect the overall diagnostic performance of the network, the quality of feature extraction directly affects the accuracy of whole network fault identification and multiwavelets transformation is a natural convolution process based on multiple wavelet basis functions. Therefore, replacing the first convolution layer of convolution neural network with multiwavelets transform can not only give full play to the excellent attributes of multiwavelets transform but also have better compatibility.

Besides, the multiwavelets basic functions are similar to fault features could extract fault features in dynamic signals, this paper constructs customized multiwavelets layer parameters according to different input signals, customized multiwavelets construct the most matching multiwavelets basis function according to different signal characteristics. The improved CNN by multiwavelets reduces the network training parameters, reduces the problem of big data driving, and it improves the problem of poor fitting effect of classical CNN in the case of multiple fault classification.

The rest of the paper is organized as follows. Section 2 mainly expounds on the basic theory of multiwavelets and the basic structure of convolution neural networks. Section 3 describes the basic structure of the improved convolution neural network by replacing the multiwavelets. The simulation results are presented and compared with the existing scheme in Sect. 4. Finally, we conclude this paper.

## 2 Basic Theory of Method

### 2.1 Multiwavelets

#### Multiwavelets Multiscales Analysis

Multiwavelets refer to wavelets generated by two or more functions as scaling functions. The basic theory is to expand the multi-resolution analysis space generated by a single wavelet with multiscales functions [18].

The vector valued function  $F(x) = (f_1(x), f_2(x), \dots, f_r(x))^T$ , if there is  $f_j(x) \in L^2(\mathbb{R})$ ,  $j = 1, 2, \dots, r$ , it is recorded as  $F(x) \in L^2(\mathbb{R})^r$  ( $\frac{\pi}{2} - \theta$ ), if  $\Phi = (\phi_1, \dots, \phi_r)^T \in L^2(\mathbb{R})^r$  satisfy the two-scale relationship:

$$\Phi(t) = \sum_{k=0}^N H_k \Phi(2t - k) \quad (1)$$

where  $\{H_K\}, K = 0, 1, \dots, M$  is  $r \times r$  two-scales matrix sequence, which is  $r$ -order scale function. The  $r$  multiresolution analysis generated by  $\Phi(x)$  is defined as:

$$V_j = \text{clos}_{L^2(\mathbb{R})} \left\{ 2^{j/2} \varphi_i(2^j x - k) : 1 \leq i \leq r, k \in \mathbb{Z} \right\} \tag{2}$$

$V_j$  is subspace with a resolution of  $2^j$ .

If  $W_j$  is the complement subspace of  $V_j$  space in  $V_{j+1}$ , the vector function  $\Psi(x) = (\psi_1, \psi_2, \dots, \psi_r)^T \in L^2(\mathbb{R})^r$ , the expansion and translation components construct a Rise basis of  $W_j$  subspace

$$W_j = \text{clos}_{L^2(\mathbb{R})} \left\{ 2^{j/2} \psi_i(2^j x - k) : 1 \leq i \leq r, k \in \mathbb{Z} \right\} \tag{3}$$

There is a matrix sequence  $\{G_k\}_{k \in \mathbb{Z}}$ , make  $\Psi(x) = (\psi_1, \psi_2, \dots, \psi_r)^T$  satisfy the following two-scale relationship:

$$\Psi(x) = \sum_{k=0}^N G_k \Phi(2x - k) \tag{4}$$

where  $\{G_k\}, k = 0, 1, \dots, N$  is  $r \times r$  two-scales matrix sequence, which is  $r$ -order wavelet function. Studies on multiwavelets with multiplicity  $r > 2$  are rare. Hence,  $r = 2$  is studied in this paper.

By the dilations of Eq. (1) and Eq. (4), the following recursive relationship between the coefficients  $(c_{1,j,k}, c_{2,j,k})^T$  can be obtained.

$$\begin{pmatrix} c_{1,j-1,k} \\ c_{2,j-1,k} \end{pmatrix} = \sqrt{2} \sum_{n=0}^K H_n \begin{pmatrix} c_{1,j,2k+n} \\ c_{2,j,2k+n} \end{pmatrix}, \quad j, k \in \mathbb{Z} \tag{5}$$

$$\begin{pmatrix} d_{1,j-1,k} \\ d_{2,j-1,k} \end{pmatrix} = \sqrt{2} \sum_{n=0}^K G_n \begin{pmatrix} c_{1,j,2k+n} \\ c_{2,j,2k+n} \end{pmatrix}, \quad j, k \in \mathbb{Z} \tag{6}$$

**Hermite Spline Multiwavelets**

Hermite spline multiwavelets is the orthogonal multiwavelets constructed by using third-order spline function, which have 2-order continuous differentiability. The multiscales support interval  $[0, 2]$ , and they have 4-order approximation order. The support interval of the multwavelets function is  $[0, 3]$  and they have 2-order approximation order. The two scale relationship is shown in Eq. (7) and Eq. (8).

$$\Phi(x) = \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \end{bmatrix} = H_0 \Phi(2x) + H_1 \Phi(2x - 1) + H_2 \Phi(2x - 2) \tag{7}$$

where  $H_0 = \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{1}{8} & -\frac{1}{8} \end{bmatrix}, H_1 = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{8} \end{bmatrix}, H_2 = \begin{bmatrix} \frac{1}{2} & -\frac{3}{4} \\ \frac{1}{8} & -\frac{1}{8} \end{bmatrix}.$

$$\Psi(x) = \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = G_0 \Psi(2x) + G_1 \Psi(2x - 1) + G_2 \Psi(2x - 2)$$

$$+ G_3\Psi(2x - 3) + G_4\Psi(2x - 4) \tag{8}$$

where

$$G_0 = \begin{bmatrix} \frac{67}{240} & \frac{7}{240} \\ -\frac{95}{972} & -\frac{1}{162} \end{bmatrix}, G_1 = \begin{bmatrix} -1 & \frac{187}{60} \\ \frac{89}{243} & \frac{91}{81} \end{bmatrix}, G_2 = \begin{bmatrix} \frac{173}{120} & 0 \\ 0 & \frac{26}{9} \end{bmatrix}, G_3 = \begin{bmatrix} -1 & \frac{187}{60} \\ -\frac{89}{243} & \frac{91}{81} \end{bmatrix},$$

$$G_4 = \begin{bmatrix} \frac{67}{240} & \frac{7}{240} \\ -\frac{95}{972} & -\frac{1}{162} \end{bmatrix}$$

## 2.2 CNN

### Convolutional Layer

Convolution layer extracts data features by convoluting the convolution layer parameters with the input data. Usually, a convolution layer has multiple convolution kernels. Because the same convolution kernel shares parameters in the process of convolution, one convolution kernel learns a class of features, which is called mapping graph. To calculate the output  $y^i$ , first the inputs  $x^1, x^2, \dots, x^d$  are convoluted with convolutional kernels  $W^{i,d}$ . Then, add all the convolutional operation results, the sum of convolution results is added to scalar offset value  $b^i$ . The output of the convolutional layer  $Z^i$  can be obtained as.

$$Z^i = \mathbf{W}^i \otimes \mathbf{X} + b^i = \sum_1^d W^{i,d} \otimes x^d + b^i \tag{9}$$

where  $\otimes$  represents the convolutional operation and  $W^i \in \mathbb{R}^{m \times n \times d}$  is the convolution kernel. Based on the nonlinear activation function, the output feature map  $y^i$  can be represented as.

$$y^i = g(Z^i) \tag{10}$$

where  $g(\cdot)$  represents the nonlinear activation function. In this research, convolution layer is adopt the rectified linear unit (ReLU) function as the activation function.

### Polling Layer

The pooled layer performs subsampling by checking the input data and extracts features while reducing the dimension of the data. Pooling includes maximum pooling and average pooling, among which maximum pooling has the best effect, which can be described as the following.

$$p^{1(i,j)} = \max_{(j-1)w < t < jw} \{a^{1(i,t)}\} j = 1, 2, \dots, q \tag{11}$$

where  $p^{1(i,j)}$  denotes  $t$ -th neuron of the  $i$ -th feature map in layer 1,  $w$  represent width od convolutional kernel,  $j$  represent  $j$ -th pooling kernel.

**Fully Connected and Output Layers**

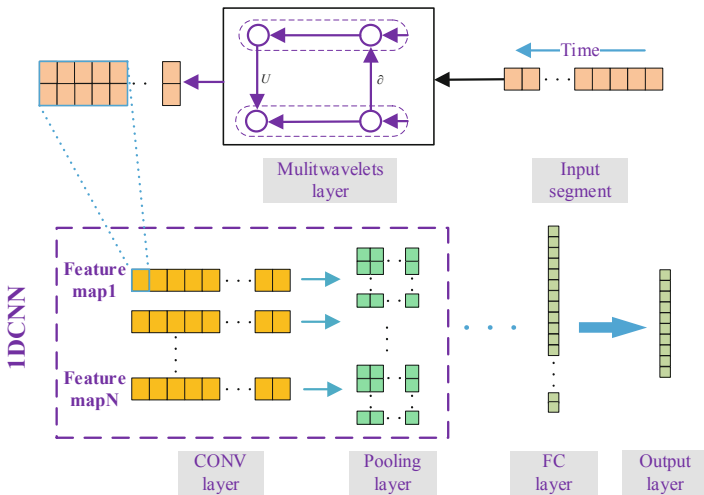
The signal features extracted from the upper network are input to the first full connection layer for one-dimensional sequence expansion. Each  $\lambda_c$  matrix by the finally output layer will be input into a softmax function  $\varphi(\cdot)$ , which is defined by

$$\varphi(\lambda_c) = \frac{e^{\lambda_c}}{\sum_{c=1}^C e^{\lambda_c}}, c = 1, \dots, C \tag{12}$$

The  $\varphi = [\varphi(v_1), \dots, \varphi(v_C)]$  represent a C-dimensional probability vector. Represents the probability distribution under C kinds of test conditions. The output value of the softmax function represents the probability distribution of the input signal in each tag.

**3 The Presented Network Structure**

Aiming at bearing fault, this paper improves the network framework based on CNN is illustrated in Fig. 1. Original time-domain signals from sensor are directly extracted features by multiwavelets layer of improved CNN, the feature signal extracted by multiwavelets is transmitted to deep network composed of one-dimensional convolutional and pooling layers. Since the characteristic signal after multiwavelets transform is a two-dimensional signal, the convolution kernel of the next convolution layer is also a two-dimensional matrix, and the characteristic signal after passing through the first convolution layer will be the one-dimensional characteristic signal, the dimension of convolution kernel in other convolution layers is one-dimensional convolution kernel. Finally, a fully connected (FC) layer and a multi category output layer composed of softmax function are used as the bottom architecture of the network.



**Fig. 1.** The structure of the improved CNN by multiwavelets.

### 3.1 Multiwavelets Layer

The specific implementation steps of the multiwavelets layers are as follows.

Split sublayer: the signal  $f$  is divided into two new signal according to the odd and even bits of the data sequence, even sequence samples are  $p$  and odd sequence samples are  $q$ .

$$p(x) = f(2x) \quad (13)$$

$$q(x) = f(2x + 1)x \in z \quad (14)$$

Predict sublayer: we use the optimizer matrix  $\partial$  to convolute even samples  $q$  to predict odd samples  $p$ . The error between the predicted value and real value is  $\Delta$  that defined as detail coefficients.

$$\Delta = p - \partial * q \quad (15)$$

where  $\partial$  is the matrix vector of prediction operator, the symbol  $*$  represent matrix vector convolution operation.

Update sublayer: the detail coefficients obtained by the predictor are transferred into the update sublayer, it convoluted with updater  $U$  composed of the parameter matrix to perform the operation and adding the result to  $p$ . The update sequence  $p^U$  represent the vector of approximation coefficients.

$$p^U = p + U * \Delta \quad (16)$$

where  $U$  represents the vector of two dimensionals matrix as updater.

The custom multiwavelet layer is based on the evolution of adaptive multiwavelet theory, and also integrates the biorthogonal perfect reconstruction multi filter bank to process the input signal. The transformation of parameter  $k$  control the matrix low-pass and highpass filters  $\{H, G, \tilde{H}, \tilde{G}\}$  as well as predict sublayer operation and update sublayer operation  $\{\partial, U\}$ .

$$\mathbf{G}(k) = k(\mathbf{I} - \partial(k^2)/k) \quad (17)$$

$$\mathbf{H}(k) = \mathbf{I} + \mathbf{U}(k^2)\mathbf{G}(k) \quad (18)$$

$$\tilde{\mathbf{H}}(k) = \mathbf{I} + \partial(k^2)/k \quad (19)$$

$$\tilde{\mathbf{G}}(k) = (\mathbf{I} - \mathbf{H}(k)k\mathbf{U}(k^2))/k \quad (20)$$

In order to ensure the linear phase of predictor filters, it require symmetry, the operator  $\partial$  of predict sublayer is subjected to the symmetry condition.

$$\partial(0) = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ c & -\frac{1}{4} \end{pmatrix}, \quad \partial(-1) = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} \\ -c & -\frac{1}{4} \end{pmatrix} \quad (21)$$

Next, operator  $U$  of update sublayer closely related to  $\partial$  can be calculated as

$$U = \{U(0), U(1)\}, \quad U(0) = \partial(-1)/2 = \begin{pmatrix} \frac{1}{4} & -\frac{1}{8} \\ -\frac{c}{2} & -\frac{1}{8} \end{pmatrix}, \quad U(1) = \partial(0)/2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{8} \\ \frac{c}{2} & -\frac{1}{8} \end{pmatrix} \quad (22)$$

Equation (22) show that the parameter  $c$  affects not only the matrix vector of predictor  $\partial$  but also the matrix vector updater  $U$ . In other words, free parameter  $c$  will change the multiscales and multiwavelets functions due to Eq. (21).

### 3.2 Multiwavelets Layer Parameters

The customized multiwavelets layer relies on the basic function to change its kernel function with different input signals, to match different fault features more accurately. Table 1 shows the important parameters of improved CNN by multiwavelets, here each data sample of the input network contains 1024 sampling points. The shallow convolution layer is replaced by a multiwavelets layer with a kernel size of 10, and the number of output channels is set to 1, which greatly reduces the shallow training parameters of the network and improves the overall training speed and convergence speed of the network. The fault features are extracted by multiwavelets layer and transferred into convolution layers, the fault feature information is further mined. Two convolutional layers are set, Conv1D represent one vector between convolution kernel and the input parameters. The next adaptive maximum pooling layer adopt to the kernel number 16. Then pass feature signal to the full connection layer. The final dimension of output data is  $m$  which is consistent with the type of input label.

Figure 2 shows the data flow framework structure of improved CNN by multiwavelets, described as follows.

Step1: In the data preprocessing stage, the collected time-domain signals containing various fault types are divided into training set and test set according to the ratio of 6:4.

Step2: In the training stage, two different fault timing signals are packaged into the same two-dimensional matrix, the improved CNN is initialized.

Step3: According to different fault signal types, set different parameter  $c$  constraint matrices  $\partial$  and  $U$ , therefore, the kernel parameters of the multiwavelet layer are determined.

Step4: The data is processed by the split sublayer, predict sublayer, and update sublayer to extract fault features.

Step5: The fault features extracted from the shallow layer are introduced into two one-dimensional convolutional layers and adaptive maximum pooling layer, the fault features characteristics of the input signals will be fully mined.

Step6: The results are calculated by the softmax function get the probability of each feature label. At the same time, the gradient of each layer in the network model is calculated by using the back propagation algorithm to continuously modify the network parameters and improve the training accuracy.

Step7: The loss rate is calculated according to the cross entropy loss function to judge whether the network accuracy has completed the training.

Step8: The well-trained network model is applied to the testing data.

Step9: Identify the fault type label for the input signal.



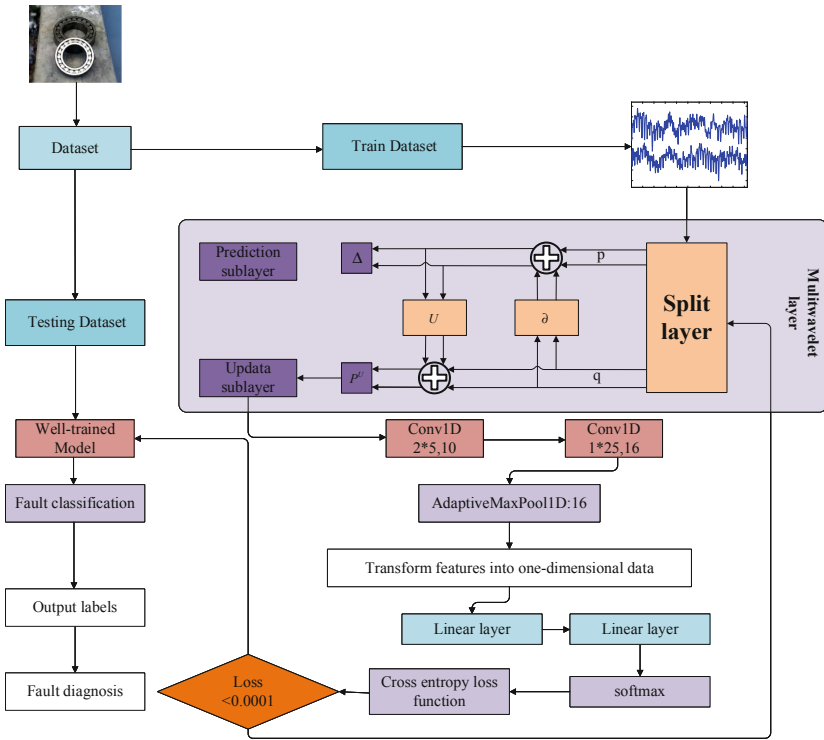


Fig. 2. The flowchart of improved CNN by multiwavelets.

Table 1. The parameters of improve CNN model.

	Name	Kernel size	Channel	Data size
	Data size			$1 \times 1024$
	Preprocessing			$2 \times 512$
L1	Multiwavelets layer	$2 \times 5$	1	$2 \times 1 \times 256$
L2	Conv1D	$2 \times 5$	10	$1 \times 10 \times 52$
L3	Conv1D	$1 \times 25$	16	$1 \times 16 \times 227$
L4	AdaptiveMaxPool1D		16	$1 \times 16 \times 16$
L5	Linear layer		1	$1 \times 1 \times 100$
L6	Softmax layer		1	m

## 4 Experiment Verification

### 4.1 Experiment Description

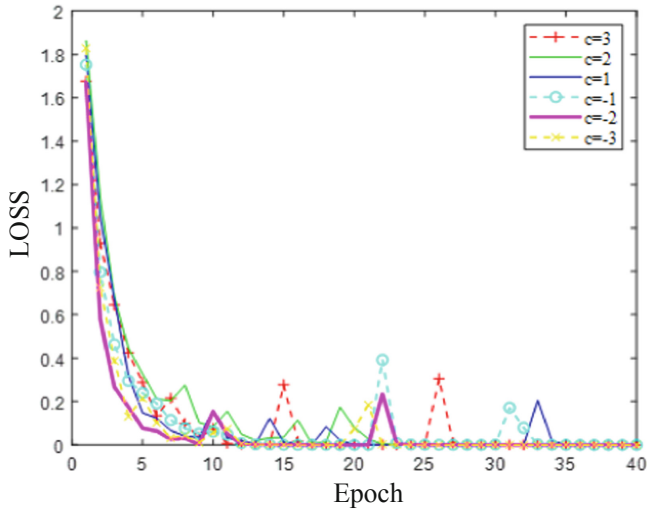
In this section, the improved CNN by multiwavelets model is applied to the laboratory bearing dataset of Case Western Reserve University (CWRU) for verification [19]. The vibration signals of 12000 samples/second are collected through two accelerometers (sensors) installed at the drive end of an electric motor in the test rig under four different conditions. Table 2 shows the detailed working condition data of ten faults. The damage diameter of three basic faults is 0.007 in., 0.014 in., and 0.021 in.. The signal sampling frequency is 12 kHz. The collected signal of each fault state is divided into 100 training samples and each sample contains 1024 points. The original signal is allocated to training samples and testing samples by a ratio of 3:2.

**Table 2.** Rolling bearing operation state

	0.007	0.014	0.021	Motor speed (rpm)
Normal	–	–	–	1797
Ball	✓	✓	✓	1797
Inner race	✓	✓	✓	1797
Outer race	✓	✓	✓	1797

### 4.2 Parameter Optimization of Multiwavelets Layer

In the process of training parameters, the network will initialize parameters, but the value of random initialization parameters is uncertain, so it is difficult to fully show the feature extraction ability of the multiwavelets layer, Fig. 3 shows the loss function decreases during the iterative training of network parameters, which compares the network training speed and fitting effect of parameters under different values, it can be observed from the figure that when the  $c = -2$ , the network training speed and the stability of parameter fitting reach the best, Table 3 shows fault classification average accuracy of multiwavelets layer under different parameter values, and when multiwavelets parameter  $c = -2$ , the network model full play to performance. Therefore, in the case of parameter  $c = -2$ , the network model can achieve the best speed, stability, and diagnostic accuracy of multi-fault classification.



**Fig. 3.** Changes of cross-entropy loss in different parameters.

**Table 3.** The accuracy of the dataset.

Dataset	$C = -3$	$C = -2$	$C = -1$	$C = 1$	$C = 2$	$C = 3$
Accuracy (%)	99.67	99.83	99.63	99.87	99.67	99.33

### 4.3 The Performance of Bearing Fault Diagnosis

Multilayer Perceptron (MLP), and 1DCNN are taken for comparison with the presented solution, improve CNN model  $c = -2$  as the parameter of multiwavelet. MLP consists of four layers: input layer, output layer, and two hidden layers, 1DCNN has the same convolution layer parameters as the improve CNN, in which the convolution kernel size of the shallow network is 10. The two groups of experimental schemes are fully in line with the principle of controlling a single variable.

Figure 4 is shown the CNN and improved CNN by multiwavelets decreased rapidly in the first five iterations and converged in the seventh iteration, the MLP is difficult fully converge in the multi-fault classification experiment, and the Cross entropy loss function fluctuates greatly, this phenomenon proved that MLP model has obvious shortcoming in nonlinear fitting problem. CNN network shows a good training effect in multi-fault nonlinear fitting fault. However, CNN fluctuated greatly in the 14th, 37th and 38th iterations show that the training parameters are unstable in the process of multiple training. The improved CNN by multiwavelets model not only inherits the excellent nonlinear fitting ability of the CNN model in the face of multi-fault problems but also has the spatial-temporal features extracted by the multiwavelets, which makes the model maintain better network stability and reliability in multiple iterations compared with the classical CNN model.

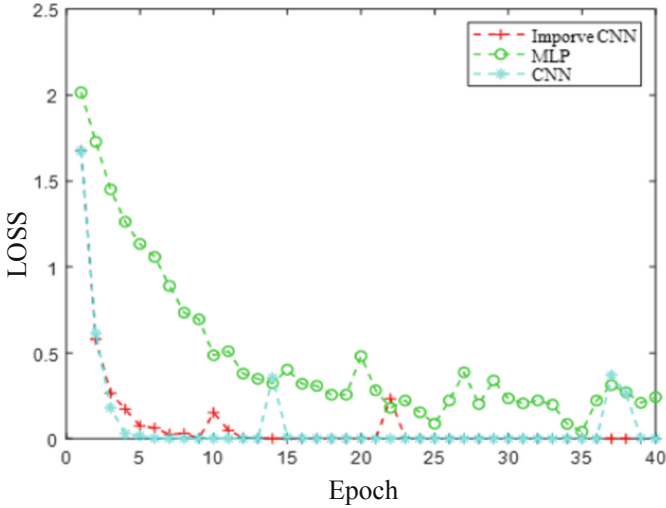


Fig. 4. Variation of loss value in different experimental schemes.

Table 4. Classification results in case.

Method	Max accuracy	Min accuracy	Mean accuracy
IMPROVE CNN	100.0%	99.3%	99.8%
MLP	59.0%	53.1%	54.0%
1DCNN	99.8%	97.4%	97.5%

The accuracy of these methods on common Dataset is listed in Table 4. After training these three schemes five times, this paper brings the training parameters of each time into the test set for verification, and the average accuracy shows that improved CNN has a stronger ability to extract fault features. Fig. 5 visualizes the verification results of 400 sets of test sets by three methods, Figure (a) shows the confusion matrix thermodynamic diagram of the improved CNN by multiwavelets. Figure(b) and Figure(c) show the confusion matrix thermodynamic diagram of CNN and MLP respectively. These figures indicate that the improved CNN has higher accuracy in fault classification.

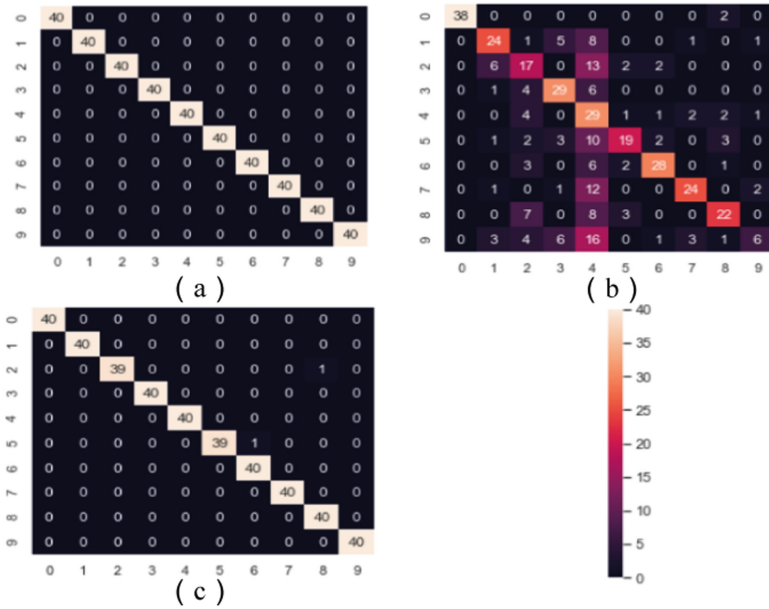


Fig. 5. Sample classification in validation set.

## 5 Conclusion

CNN method relies on training a large number of the same fault data and updating network parameters through backpropagation to establish the recognition ability of the same type of fault. However, under the actual working conditions, the types of bearing faults are constantly changing, and it is difficult to collect sufficient data for the same type of faults. When small datasets are employed, the first layer of CNN is difficult to effectively extract deep feature parameters and it influences the performances of the entire network. Therefore, an improved CNN by multiwavelets is proposed in this paper. The first layer of the model is the multiwavelets transform in signal processing, multiwavelets layer completely inherits the advantages of the fast extraction of features signals from multiple wavelets base of multiwavelets. Original signals in the time domain are directly input into the improved CNN, multiwavelets layer is used as a multi-channel filter to extract multi fault features at the same time, subsequently, the features extracted by the multiwavelets layer are fused as inputs to the next layer. By enhancing the ability of network shallow feature extraction, accurately extract fault features and reduce network parameters. By incorporating multiwavelets feature extraction, this proposed CNN method can accurately diagnose faults with smaller data set samples in practical bearing diagnosis, and by customizing multiwavelets layer parameters, the influence of shallow layer parameters on the overall network performance is discussed. The validity of this method is fully verified by publicly experimental datasets and comparison with the traditional methods and classical methods.

Based on the CNN shallow layer replacement multiwavelets model, this paper mainly finds the multiwavelets layer parameters most suitable for fault characteristics through

parameter optimization. In the future research direction, the multiwavelets layer parameters can be adaptively matched with fault characteristics to improve the effect of network diagnosis efficiency, and the underlying logic of CNN network training parameters can be discussed based on the physical meaning of multiwavelets layer.

**Acknowledgements.** This research is sponsored by the National Natural Science Foundations of China (No. 51975377 and 52005335), Shanghai Sailing Program (No. 21YF1430600).

## References

1. Attoui, I., Oudjani, B., Boutasseta, N., et al.: Novel predictive features using a wrapper model for rolling bearing fault diagnosis based on vibration signal analysis. *Int. J. Adv. Manuf. Technol.* **106**(7), 3409–3435 (2020)
2. Wang, H., Chen, J., Zhou, Y., et al.: Early fault diagnosis of rolling bearing based on noise-assisted signal feature enhancement and stochastic resonance for intelligent manufacturing. *Int. J. Adv. Manuf. Technol.* **107**(3), 1017–1023 (2020)
3. Xu, Q., Lu, S., Jia, W., et al.: Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning. *J. Intell. Manuf.* **31**(6), 1467–1481 (2020)
4. Lund, D., MacGillivray, C., Turner, V., et al.: Worldwide and regional internet of things (IoT) 2014–2020 forecast: a virtuous circle of proven value and demand. *Int. Data Corp. (IDC), Tech. Rep.* **1**(9) (2014)
5. Chen, Z., Li, W.: Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.* **66**(7), 1693–1702 (2017)
6. Sousa, R., Antunes, J., Filipe, C., et al.: Robust cepstral-based features for anomaly detection in ball bearings. *Int. J. Adv. Manuf. Technol.* **103**(5–8), 2377–2390 (2019)
7. Yuan, J., Xu, C., Zhao, Q., Jiang, H., Weng, Y.: High-fidelity noise-reconstructed empirical mode decomposition for mechanical multiple and weak fault extractions. *ISA Trans.* (2022). <https://doi.org/10.1016/j.isatra.2022.02.017>
8. Hu, Q., He, Z.J., Zi, Y., et al.: Intelligent fault diagnosis in power plant using empirical mode decomposition, fuzzy feature extraction, and support vector machines. *Key Eng. Mater.* **293**, 373–382 (2005)
9. Chen, B., Chen, X., Li, B., et al.: Reliability estimation for cutting tools based on a logistic regression model using vibration signals. *Mech. Syst. Signal Process.* **25**(7), 2526–2537 (2011)
10. Haidong, S., Hongkai, J., Xingqiu, L., et al.: Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. *Knowl. Based Syst.* **140**, 1–14 (2018)
11. Shao, H., Jiang, H., Zhang, H., et al.: Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network. *IEEE Trans. Ind. Electron.* **65**(3), 2727–2736 (2017)
12. Pan, H., He, X., Tang, S., et al.: An improved bearing fault diagnosis method using one-dimensional CNN and LSTM. *Strojnicki Vestnik/J. Mech. Eng.*, 64 (2018)
13. Liu, H., Zhou, J., Xu, Y., et al.: Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks. *Neurocomputing* **315**, 412–424 (2018)
14. Weimer, D., Scholz-Reiter, B., Shpitalni, M.: Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Ann.* **65**(1), 417–420 (2016)

15. Ince, T., Kiranyaz, S., Eren, L., et al.: Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Ind. Electron.* **63**(11), 7067–7075 (2016)
16. Abdeljaber, O., Avci, O., Kiranyaz, S., et al.: Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J. Sound Vib.* **388**, 154–170 (2017)
17. Sun, H., He, Z., Zi, Y., et al.: Multiwavelet transform and its applications in mechanical fault diagnosis—a review. *Mech. Syst. Signal Process.* **43**(1–2), 1–24 (2014)
18. Keinert, F.: *Wavelets and Multiwavelets*. CRC Press, Boca Raton (2003)
19. Bearing Data Center, Case Western Reserve University, Cleveland, OH, USA (2004). <http://csegroups.case.edu/bearingdatacenter/home>