



An Improved Fault Diagnosis Method of Rolling Bearings Based on Multi-Scale Attention CNN

Linfeng Deng · Yuanwen Zhang · Zhifeng Shi

Submitted: 17 November 2023 / in revised form: 18 April 2024 / Accepted: 6 May 2024 / Published online: 20 June 2024
© ASM International 2024

Abstract Rolling bearing fault diagnosis based on convolutional neural network is greatly effective for bearing maintenance, and it is of great significance for ensuring the safe operation of rotating machinery. However, the traditional convolutional neural network models only focus on the single-scale feature and ignore the multi-scale deep information, which results in low performance for conducting the complex fault diagnosis problem. Aiming at this problem, we propose an improved fault diagnosis method of rolling bearings based on multi-scale attention convolutional neural network. In the first stage, a one-dimensional convolutional neural network model integrating feature extraction and fault intelligence classification is established. Subsequently, a multi-scale structure with serial layer skipping connection is constructed to extract multi-scale features in different reception fields, and the “Concat” operator is used to fuse the extracted features of each layer with SE attention mechanism in order to obtain more important feature information. And then, the fault samples are input into the network model to realize the end-to-end fault diagnosis of rolling bearings result from the nonlinear fitting ability of deep learning. Extensive experiments on two well-known rolling bearing datasets validate that the proposed method not only achieves higher fault diagnosis accuracy on the fault data sets under constant load conditions, but also exhibits strong fault identification and transfer diagnosis ability on the fault data sets under variable load conditions.

Keywords Rolling bearing · Fault diagnosis · Convolutional neural network · Multi-scale attention mechanism

Introduction

Rolling bearings are critical components of rotating machinery, the health status of them having a significant impact on the performance, efficiency, and service life of mechanical equipment [1, 2]. Typically, bearings operate in harsh environments, being subjected to complex and varied working conditions to make them prone to anomalies. Once the faults occur, they may result in economic losses or even safety accidents [3]. Therefore, deep research on fault diagnosis of rolling bearings is of great significance for ensuring the safe and normal operation of mechanical equipment [4, 5].

In recent years, the research on rolling bearing fault diagnosis using deep learning method has become a mainstream in the field of fault diagnosis, and it is an extremely potential intelligent fault diagnosis method for rolling bearings [6–8]. Convolution neural network (CNN) is an important deep learning model, which has a very powerful feature extraction capability and can extract the inherent characteristics and useful information embedded in the original data. At present, CNN has achieved significant success in image recognition, natural language processing and other application fields of deep learning [9]. In the meantime, it has also been effectively applied in the field of machine fault diagnosis [10]. Wang et al. [11] proposed an adaptive deep convolutional neural network approach for fault diagnosis of rolling bearings, which automatically learns the essential features of faults from the

L. Deng (✉) · Y. Zhang · Z. Shi
School of Mechanical and Electrical Engineering, Lanzhou
University of Technology, Lanzhou 730050, China
e-mail: denglinfeng2002@163.com

input data layer by layer, eliminating the need for artificial feature extraction. Liu et al. [12] proposed an unsupervised domain adaptation method called deep feature alignment adaptive network (DFAAN), to address the issue of low fault diagnosis capability when there are distribution differences between the source and target data, which can enhance the adaptability of fault diagnosis models. Zhang et al. [13] designed the first-layer deep network model with wide convolution kernel, which can effectively resist noise and learn the intrinsic characteristics of faults, as well as automatically remove features that are not helpful to the diagnosis results. Han et al. [14] proposed a CNN combined support vector machines (CNN-SVM) model for bearing fault diagnosis to address the problem that it is difficult to meet the training requirements of complex models in the field of fault diagnosis under small sample data. Although these works mentioned above achieve the accurate diagnosis results for rolling bearing faults, usually these diagnosis models used the single-scale convolution kernels to extract fault features. In this situation, when rolling bearings operate under different load conditions, the corresponding relation between fault patterns and fault characteristics may be very complex. Under this circumstance, the single-scale convolutional neural network fails to capture the complete details of fault features, which makes part of the fault information lost and furthermore results in the reduction of the fault diagnosis accuracy.

The fault diagnosis method based on multi-scale feature fusion technology can extract and fuse fault features from different perspectives [15], which can achieve good results in fault information extraction. Due to the fact that multi-scale feature fusion technology can effectively eliminate the deficiencies of traditional single-scale analysis methods in comprehensively extracting bearing fault feature information in complex environments, it is able to provide a more complete and accurate diagnosis of faults. Wang Wei [16] used a multi-scale convolutional neural network as feature extractors to learn fault features and obtain the high-precision fault diagnosis results in experimental validation. Jiang et al. [17] proposed a multi-scale CNN structure that integrates features from different scales through multiple convolutional and pooling layers, which is able to better capture bearing fault features at different scales, so as to improve the accuracy and reliability of fault diagnosis. Seungmin et al. [18] proposed a multi-scale convolutional neural network (MSCNN) model, which learns more powerful feature information than traditional CNNs through multi-scale convolutional operations and reduces the number of parameters and training time. On the other hand, in order to enhance the fault recognition performance of the fault diagnosis models, different types of attention mechanisms are introduced into the multi-scale CNN models, and corresponding multi-scale CNN models

are proposed [19–21]. The convolution kernels in these multi-scale CNN models are all parallel connected to conduct the data. Because there is no particular hierarchical relationship among the features in the models, they may fail to capture the fault characteristics at different levels. Useful information of multiple frequency components and different time scales in the fault signal can be obtained through serially connecting the fault features in different layers via the serial skipping layer connection mode constructed by employing the convolution kernels. The multi-scale features of fault signals can be extracted more effectively, which is very beneficial to improve the diagnosis accuracy of rolling bearing faults.

From the above literature, it can be found that different convolutional network models show their own advantages and address the corresponding problems in the fault diagnosis application. It is really a feasible way for enhancing the effectiveness of bearing fault diagnosis results that useful parts of different models are integrated to construct a new fault diagnosis model which can reasonably utilize advantages and avoid weaknesses of the different models.

And therefore, here multi-scale attention mechanism and feature fusion module are introduced into a one-dimensional convolutional neural network, and we propose an intelligent identification method for rolling bearing fault diagnosis based on multi-scale attention convolutional neural network (MSACNN), which aims to address the issue that the single-scale convolutional kernels in convolutional networks struggle to wholly extract fault features in the process of bearing fault diagnosis. The main contributions of this paper are included as follows:

1. A multi-scale network with convolution kernel serial skipping layer connection is established to extract multi-scale rolling bearing fault characteristics to meet the adaptability of rolling bearing fault diagnosis model.
2. Introducing SE attention mechanism into the network can enhance the model's attention to key features and improve the fault identification accuracy of the fault diagnosis model.
3. Experimental verification and comparative analysis were implemented on two rolling bearing fault datasets, and the results show that the proposed method really outperforms other compared methods in fault diagnosis accuracy.

The rest of this article is organized as follows. Section “[Basic Theories](#)” explains the theoretical fundamentals related to the proposed method. Section “[Model construction and fault diagnosis process](#)” provides a detailed description about the fault diagnosis framework and the implementation process of the proposed method. The effectiveness of the proposed method is

demonstrated in Section “[Experimental validations](#)” through an experimental study with two rolling bearing fault cases. Finally, we draw conclusions in Section “[Conclusion](#)”.

Basic Theories

Convolutional Neural Network

Convolutional neural network (CNN) is a kind of multi-layer feedforward neural network, which extracts features of input data layer-by-layer by using convolutional layer and pooling layer alternately, and subsequently outputs the extracted features through the fully connected layer [22]. It should be noted that the one-dimensional convolutional neural network implements the convolution operation through the one-dimensional convolution kernel.

Convolution operation is a core part of CNN. In the convolutional layer, the input signal is convolved with the kernel, and then bias is added before being passed through an activation function to obtain the corresponding feature map. The convolution operation can be expressed as follows:

$$y^l = f\left(\sum_{i=1}^{c^{l-1}} w_{i,c}^l * x_i^{l-1} + b_i^l\right) \quad (\text{Eq 1})$$

where x_i^{l-1} denotes the output of the i -th channel of the layer $l-1$, c^{l-1} is the c -th channel of the layer $l-1$, y^l is the output of the l -th layer, $w_{i,c}^l$ is the weight matrix of the convolution kernel of the l -th layer, b_i^l is the bias term, $*$ is the symbol of the convolution operation, and $f(\cdot)$ is the activation function. Usually the activation function is rectified linear unit (ReLU), and its mathematical expression is defined as follows:

$$f(x) = \max(0, \log(1 + e^x)) \quad (\text{Eq 2})$$

In a neural network, a pooling layer is often connected after the convolutional layer. The main purpose of it is to reduce the parameters of the neural network and decrease the feature dimensions as well as prevent overfitting. The max pooling operation outputs the maximum value within the pooling layer in the perceptual domain of the input feature map. Its calculation formula is represented as follows:

$$p_i^{l+1} = \max_{(j-1)K+1 < t < jK} \{q_i^l(t)\} \quad (\text{Eq 3})$$

where $q_i^l(t)$ is the output value of the t -th neuron of the i -th channel in l -th layer, K is the size of the pooling kernel, j is the step size, and p_i^{l+1} is the output value of the i -th channel in layer $l + 1$

The fully connected layer first flattens the output features of the last convolutional layer or pooling layer into a one-dimensional vector, and then the vector is used to extract important feature information. After that, the outputs are connected to the Softmax classifier to complete the final classification task. The calculation formula for the fully connected layer can be represented as follows:

$$y^l = f((w^l)^T x^{l-1} + b^l) \quad (\text{Eq 4})$$

where x^{l-1} is the output value of layer $l-1$, w^l is the weight, y^l is the output of l -th layer, b^l is the bias term, and $f(\cdot)$ is the activation function.

Attention mechanism

Attention mechanism is essentially a distribution mechanism, with the core idea of highlighting certain important features of the object. Currently, it has been successfully applied to image processing, natural language processing and data prediction [23–25]. The Squeeze-and-excitation (SE) module, proposed by the Hu and his team in 2017 [26], is a novel network structure possessing attention mechanism, with the strong image classification ability. The core idea of SE is to recalibrate the global information of the feature maps into inter-channel correlations. The SE module mainly consists of two parts: the Squeeze operation and the Excitation operation. The structure of the SE module is shown in Fig. 1. Here, we have modified the SE attention mechanism to suit one-dimensional sequence data. It is noted that although the SE attention mechanism is widely used for two-dimensional image data, its core operation, recalibrating channel-wise feature responses, is inherently independent of data dimensionality.

In Fig. 1 X represents the original input data with H' , W' and C' denoting its height, width and the number of channels, respectively. The complete data conducting process of SE module can be described as follows. Firstly, the X undergoes the F_{tr} operation to perform convolution, resulting in the feature map U , where H , W and C represent its height, width and the number of channels, respectively. Secondly, the U is compressed using the F_{sq} operation into a $1 \times 1 \times C$ feature response value Z . Then, the Z is subjected to the F_{ex} operation to perform excitation, yielding a weight vector S with dimensions $1 \times 1 \times C$. Finally, the F_{scale} operation performs a multiplication operation to obtain \bar{X} .

The Squeeze operation adopts global average pooling to compress each $H \times W \times C$ -dimensional feature into a $1 \times 1 \times C$ feature response value Z with a global receptive field along the feature channel direction, and its calculation process can be expressed as follows:

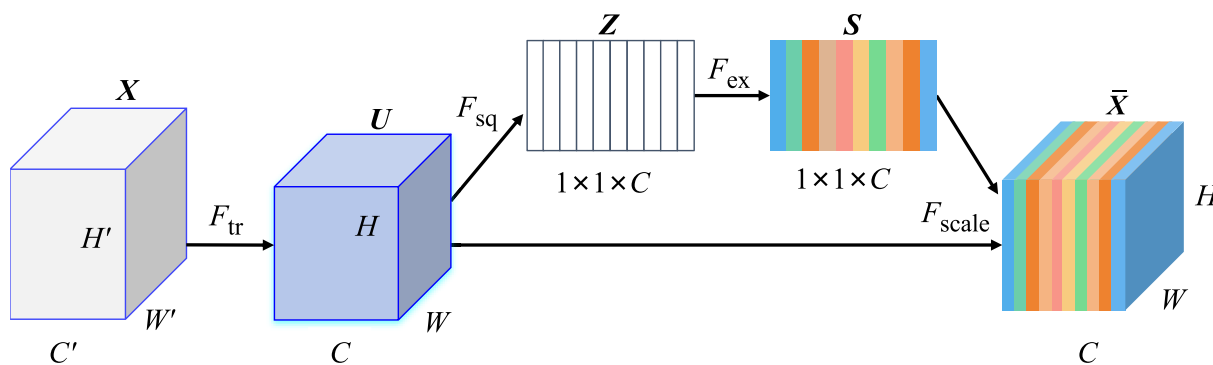


Fig. 1 Structure of the SE module

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (\text{Eq 5})$$

where z_c is the output after compression operation, F_{sq} is the extrusion operation, H is the height of the feature map, W is the width of the feature map, and $u_c(i, j)$ represents the output value of row i -th and column j -th in the c channel.

The Excitation operation utilizes two fully connected layers to constitute the gating mechanism. The first fully connected layer compresses C channels into C/r channels to reduce the computational cost; while, the second fully connected layer restores the number of channels to C . The operation formula can be expressed as follows:

$$s_c = F_{ex}(z_c, W) = \sigma(W_2 \delta(W_1 z_c)) \quad (\text{Eq 6})$$

where F_{ex} is the excitation operation, s_c is the weight obtained by F_{ex} , σ is the sigmoid function, δ is the ReLU activation function, $W_1 \in R^{C/r \times C}$ is the weight matrix of column C in row C/r , $W_2 \in R^{C \times C/r}$ is the weight matrix of column C/r in row C , and r is the scaling factor.

The function of Scale operation involves the channel-wise weights computed by the SE module respectively multiplying with the corresponding two-dimensional matrix of the original feature map, and then the final outputs can be obtained. Its calculation process is expressed as follows:

$$\bar{x}_c = F_{scale}(u_c, s_c) = s_c u_c \quad (\text{Eq 7})$$

where F_{scale} is the product operation and \bar{x} is the output result.

Model Construction and Fault Diagnosis Process

Architecture of Multi-Scale Feature Fusion Model

The convolutional neural network extracts feature through successive convolutional operations, the receptive field is one of its momentous concepts [27]. If the receptive field is too small, the network can only capture local features. And

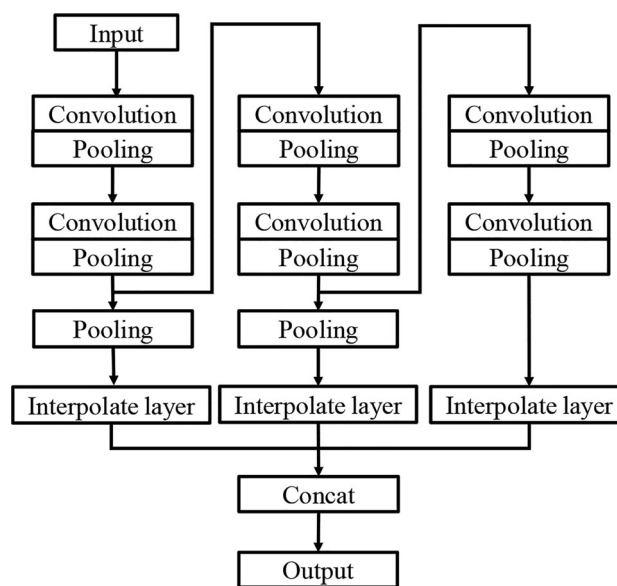


Fig. 2 Multi-scale structure of skip connection

meanwhile, if the receptive field is too large, although it may obtain a deeper understanding of global information, it also contains irrelevant information. In order to avert information redundancy and boost the effective receptive field, the signal with the different scales are sampled, which is conducive to capture the multi-scale features of the signal. There are two common types of multi-scale feature fusion, a parallel multi-branch network and a serial skip connection structure. Both of them carry on the feature extraction by using different receptive fields, and then feature fusion task will be continued.

Due to the characteristics that serial skip connection networks can scale the output feature maps of different convolutional layers to a uniform size so as to contain both global contextual information and local detailed information, the multi-scale structure of serial skip connection is selected in this study and shown in Fig. 2.

As depicted in Fig. 2, we utilize three multi-scale modules connected through two shortcuts to adapt varying

convolution kernel sizes and pooling layer numbers. This design considers the diversity of rolling bearing fault signals and the impact of variable load conditions. The multi-scale structure in neural networks improves adaptability and diagnostic accuracy in variable operating environments by utilizing different kernel sizes. This approach can capture a comprehensive range of fault features from noticeable patterns and subtle anomalies, because it employs serial skip-layer connections and feature concatenation to preserve and integrate both low-level and high-level information, resulting in a richer and more detailed feature set for fault detection. This new structure can extract and analyze features across multiple different scales, leading to accurate fault diagnosis results, particularly under the condition that fault characteristics frequently vary. Consequently, the multi-scale approach is able to outperform traditional single-scale neural networks, providing robust and precise diagnostics results for maintaining optimal operating efficiency in diverse and changing industrial settings.

Model Construction

To address the issues that deep convolutional network models are mainly only employed in pre-extracting feature process and one-dimensional convolutional neural networks with single-scale convolutional kernel are prone to lose part of information in the pooling layer, a multi-scale attention mechanism convolutional neural network model based on one-dimensional serial layer skipping connections is proposed. The structure of the proposed network is shown in Fig. 3.

As can be seen from Fig. 3, the model takes one-dimensional vibration signal as input, in the first layer, to improve the feature extraction and generalization capabilities of itself, with large convolution kernel utilized to increase the receptive field size of the convolution layer. To address the issue of partial information loss during feature extraction, a serial skip connection structure is employed to fully extract feature information from different scales. The Concatenation method is applied to fuse the features from each layer for enhancing the effective propagation of feature information through merging different channels. And meantime, the SE module is integrated into the network to make the network adaptively train and learn to possess the optimal performance. In the fully connected layer, the weight matrix is adopted to rearrange the extracted important features, and finally, the fault feature data are classified through the Softmax classifier.

Model Training

Once the MSACNN model is constructed, it is necessary to train the model to realize intelligent fault diagnosis. Since the output of the Softmax classifier is a probability value between 0 and 1, and the sum of all values is equal to 1, the calculation process of the Softmax classifier can be represented as follows:

$$\bar{Y} = \begin{bmatrix} p(y = 1|x; \theta) \\ p(y = 2|x; \theta) \\ \vdots \\ p(y = N|x; \theta) \end{bmatrix} = \frac{1}{\sum_{i=1}^N \exp(\theta_i x)} \begin{bmatrix} \exp(\theta_1 x) \\ \exp(\theta_2 x) \\ \vdots \\ \exp(\theta_i x) \end{bmatrix} \quad (\text{Eq 8})$$

where θ is the model parameter; x is the input value of Softmax; y indicates the actual category label of the fault category; \bar{Y} is the prediction probability vector of N categories output by Softmax; p is the calculation process of the probability of each class.

The obtained category probabilities are input into the cross-entropy loss function or error generation, with the loss function used to measure the consistency between the probability distribution of the output estimate of the model and the target value. The smaller loss function indicates a better fit of the model to the training samples and the calculation process is as follows:

$$L = - \sum_{i=1}^N y_i \log \bar{y}_i \quad (\text{Eq 9})$$

where y_i is the true classification result of fault, \bar{y}_i is the classification result output by the model, and L is the error loss value.

When the calculation of the loss function is finished, the parameters can be adaptively optimized through the error back propagation. To reduce training time and accelerate the training speed of the model, the Adam optimization algorithm is used to train the model. With the progress of the iterative training process, the weights of the network are updated repeatedly. Once the error loss decreases and tends to be stable, the network model simultaneously inclined to converge.

Fault Diagnosis Process

On the basis of the above theories and models, the intelligent fault diagnosis method of rolling bearing based on MSACNN is constructed, mainly consisting of three stages: data set preparation, model training and model testing. Figure 4 shows the specific flowchart of the method.

The specific implementation steps are described as follows:

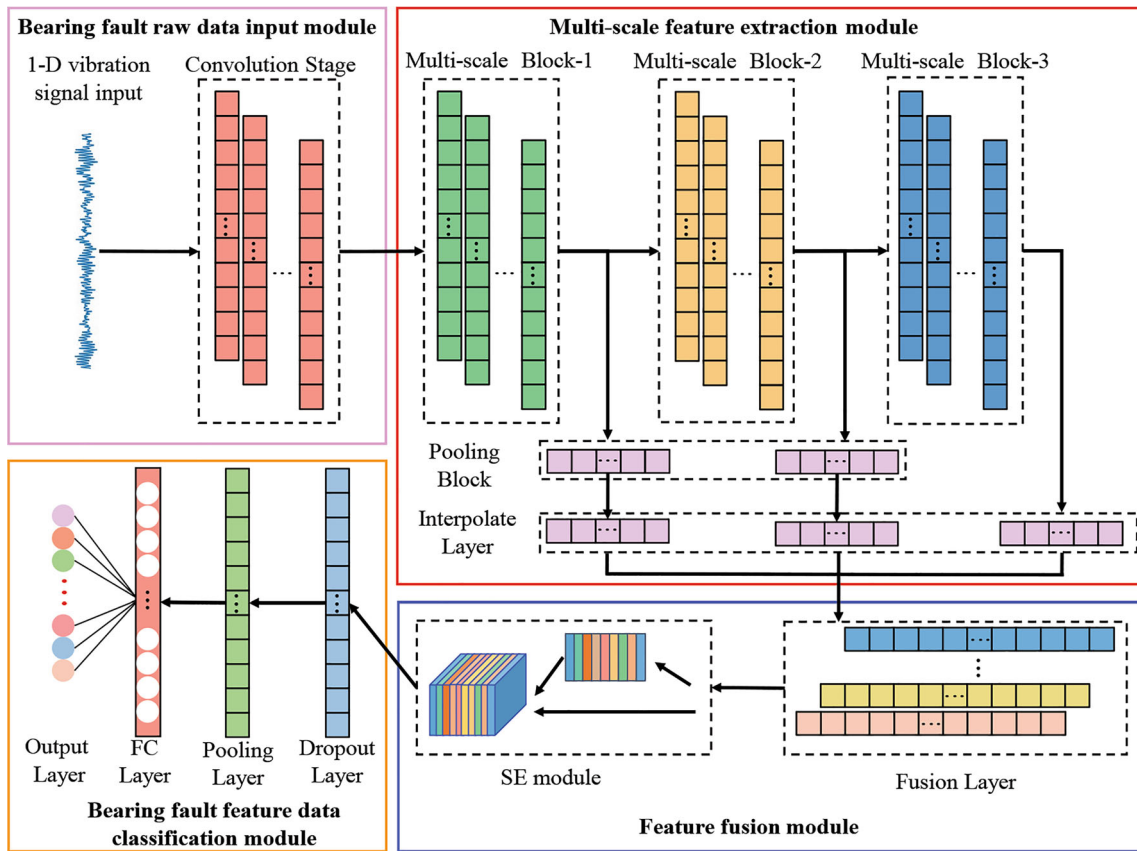


Fig. 3 Structure of the MSACNN model

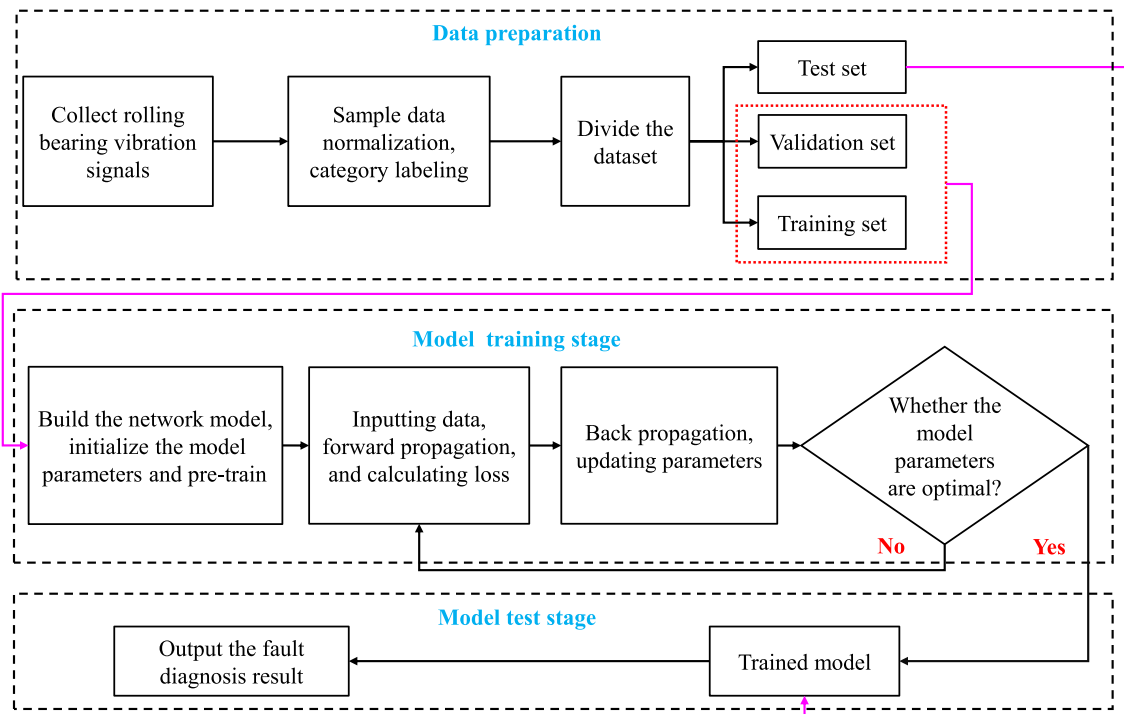


Fig. 4 Flowchart of the proposed fault diagnosis method

- (1) Collect the original vibration signals of rolling bearings with different faults during the practical operating process.
- (2) Pre-processing the collected sample data, including normalization, rearrangement and category labeling.
- (3) Divide the sample data into training set, validation set and test set.
- (4) Build the network model, initialize the model parameters, pre-train the model, and then save the model obtained after pre-training.
- (5) The training set is input into the pre-trained model, and subsequently, the forward propagation and loss function calculation are performed.
- (6) Use the validation set to check the change of the fault diagnosis accuracy of the trained model, backpropagate the loss function value, perform iterative calculation as well as update the model parameters.
- (7) Determine whether the value of the loss function tends to be stable. If yes, go to Step 8; otherwise, return to Step 5.
- (8) Input the test set into the trained model with the optimized parameters for fault intelligent diagnosis, and then output the fault classification results.

Experimental Validations

To verify the validity of the proposed model, two well-known rolling bearing data sets are adopted in this section for fault diagnosis experiments, concurrently, with the diagnosis results analyzed and discussed.

Case Study 1: CWRU Bearing Data Set

The experimental data set comes from the Case Western Reserve University (CWRU) in the United States, with it widely utilized for performance testing of fault diagnosis methods [28]. The experimental setup is shown in Fig. 5.

The test object is the drive end bearing of the experimental setup shown in Fig. 5, with a model number of SKF6205 and a sampling frequency of 12 kHz. The faults are artificially created by adopting electric discharge machining. The experiment is conducted under four different operating conditions with loads: 0, 1, 2, and 3HP, with each bearing running condition including normal (Normal), inner ring fault (IF), rolling body fault (BF) and outer ring fault (OF), and each fault containing 0.007 inch, 0.014 inch and 0.021 inch, three fault diameters, so 10 operating states can be formed. The constitution of the dataset is shown in Table 1.

Table 1 Constitution of the experiment datasets

Fault type	Load (HP)	Fault diameters (inch)	Label
Normal	0/1/2/3	...	Normal
BF	0/1/2/3	0.007	BF0.007
	0/1/2/3	0.014	BF0.014
	0/1/2/3	0.021	BF0.021
OF	0/1/2/3	0.007	OF0.007
	0/1/2/3	0.014	OF0.014
	0/1/2/3	0.021	OF0.021
IF	0/1/2/3	0.007	IF0.007
	0/1/2/3	0.014	IF0.014
	0/1/2/3	0.021	IF0.021

Fig. 5 Data acquisition system of bearing center of CWRU [28]

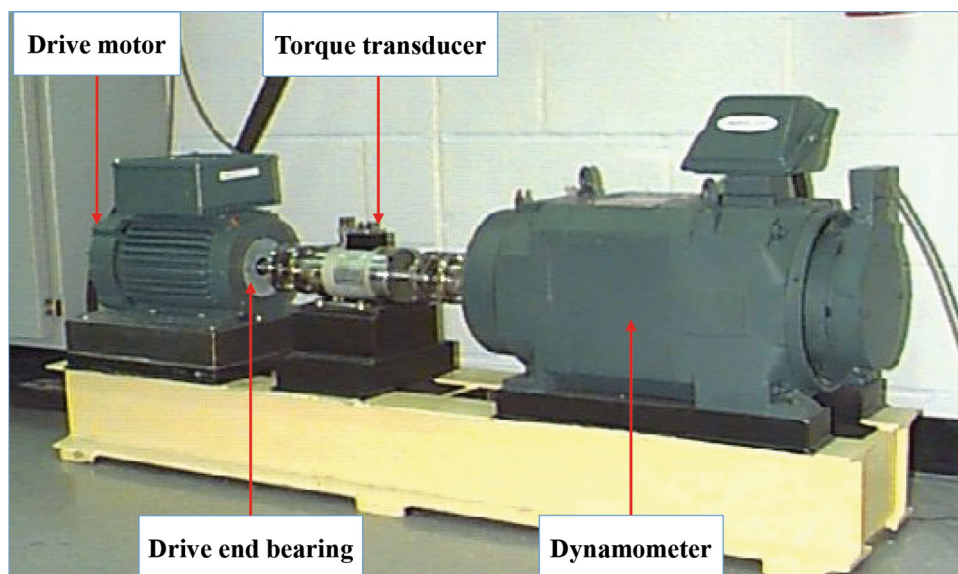
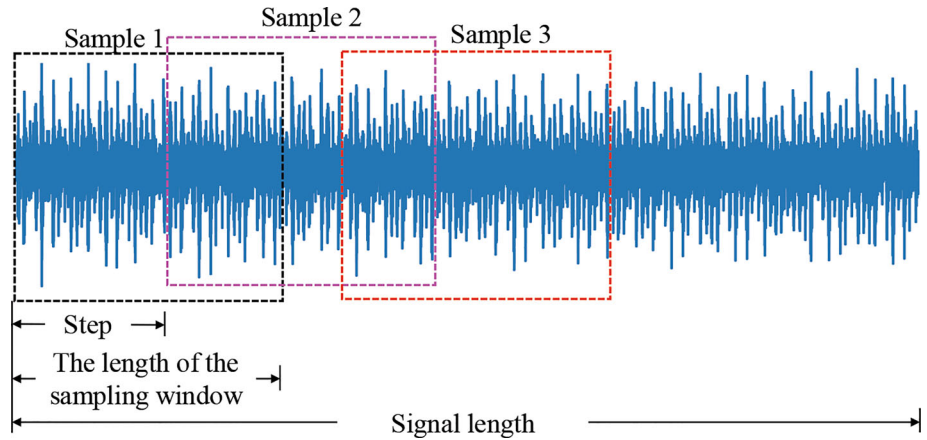


Fig. 6 Schematic diagram of the overlapping sampling method



Training a convolutional neural network model requires a large amount of data, enough training samples can enhance the generalization ability of the model. Since the sample size in the CWRU dataset is relatively small, the overlapping sampling method is adopted for data augmentation [29]. The process of overlapping sampling is illustrated in Fig. 6.

Through overlapping sampling, the expanded data set consists of 2100 training samples, 600 test samples and 300 verification samples, each of which contains 1024 data points and is normalized.

The validation experiments are conducted in a deep learning environment using the PyTorch framework. To avoid the randomness bias is caused by the results of a single experiment, each group of experiments is performed 10 times and the average value is taken.

The one-dimensional convolutional neural network model is established according to the design criteria in Section “Model construction”, with the main hyperparameters initialized, such as the number of convolution kernels, size and step size, and optimized by pre-training. The experimental process follows the principle of single variable, the final model parameters are shown in Table 2.

In the process of model training, the size of Batch size affects the performance of the model. A too small Batch size may make it difficult for the model to converge; while, a too large Batch size may reduce the generalization ability of the model. Therefore, to select an appropriate Batch size, experiments are conducted with Batch sizes set to 16, 32, 64 and 128, respectively. The results are shown in Fig. 7.

As can be seen from Fig. 7, when the Epoch is less than 9, the accuracy of the model with a Batch size of 16 is higher than the other three cases. However, when the Epoch is greater than 9, the model with a Batch size of 64 is superior to the other three cases. Moreover, the diagnostic accuracy of the model is highest when the Batch size is 64. Taking into account the influence of Batch size on

Table 2 Structure parameters of the network model

Layer type		Size × Sep	Output size	Padding
Input	1024
Convolution Stage	Conv1	32 × 8	64 × 128	Yes
	Pool1	2 × 2	64 × 64	No
	BN
Multi-scale Block-1	Conv2-1	7 × 1	64 × 64	Yes
	Pool2-1	2 × 2	64 × 32	No
	Conv2-2	7 × 1	256 × 32	Yes
	Pool2-2	2 × 2	256 × 16	No
Multi-scale Block-2	Conv3-1	5 × 1	64 × 16	Yes
	Pool3-1	2 × 2	64 × 8	No
	Conv3-2	5 × 1	256 × 8	Yes
	Pool3-2	2 × 2	256 × 4	No
Multi-scale Block-3	Conv4-1	3 × 1	64 × 4	Yes
	Pool4-1	2 × 2	64 × 2	No
	Conv4-2	3 × 1	256 × 2	Yes
	Pool4-2	2 × 2	256 × 1	Yes
Pooling block	Pool2-3	2 × 2	256 × 8	No
	Pool3-3	2 × 2	256 × 2	No
Interpolate Layer	...	256 × 8	No	...
Fusion layer	...	768 × 8
SE-Block
Pooling Layer	8 × 1	768 × 1	No	...
Dropout Layer	...	64 × 768
FC Layer	...	64 × 10
Softmax Layer	...	64 × 10

model training speed and accuracy, the Batch size should be set to 64.

In addition, during the training of the diagnostic model, cross-entropy is utilized as the loss function and Adam is adopted as the optimizer [21]. In summary, the basic parameters of the MSACNN model are shown in Table 3.

In order to verify the fault diagnosis ability of the MSACNN model under constant load condition, the

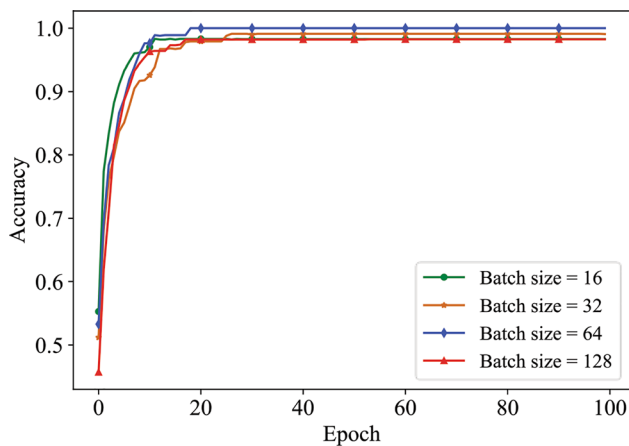


Fig. 7 Experimental results with different Batch size

Table 3 Parameter settings of MSACNN model

Learning rate	Epoch	Batch size	Dropout	Optimizer
0.002	100	64	0.5	Adam

training set samples with loads of 0, 1, 2 and 3HP are employed to train the model, and the test set samples under the same load are used for fault diagnosis test. The training process of the MSACNN model under different load conditions is shown in Fig. 8.

It can be found from Fig. 8 that when Epoch is around 25, the fault recognition accuracy of the MSACNN model under the four different loads in the constant load condition converges and can reach 100%.

To demonstrate that the MSACNN model possesses good diagnostic performance, the other four models are selected for comparison experiments. Among them, the Inception model is a traditional convolutional neural network model with parallel convolutional kernels, the Inception-SE is the Inception model with the addition of SE attention module, and two deep learning models proposed in other two literatures: WDCNN [13] and MSCNN [16]. The final experimental results are shown in Fig. 9.

Figure 9a, b, c and d respectively represents the training process and diagnosis results of five different models under four operating conditions with loads of 0 HP, 1 HP, 2 HP and 3 HP. It can be seen from Fig. 9 that under the four working conditions, the Inception-SE model can preferentially converge Inception model and the fault diagnosis accuracy is also higher than that of Inception model, which indicates that the Inception-SE model has a stronger fault diagnosis capability than Inception model. Therefore, it is demonstrated that introducing the SE attention module can improve the fault diagnosis performance of the model. Moreover, under all four working conditions, the proposed MSACNN model is the first to converge, and its fault

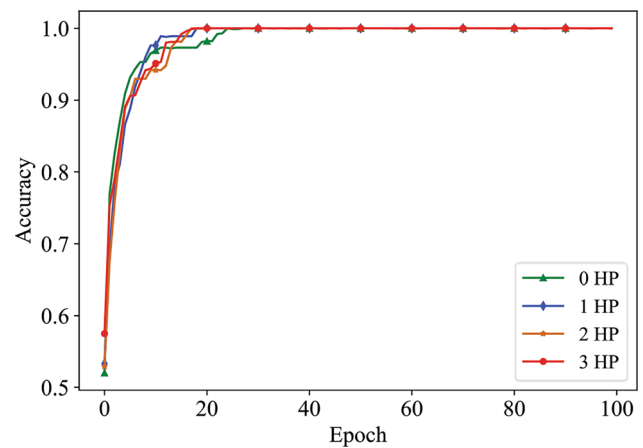


Fig. 8 Training process of the MSACNN model under different load conditions

recognition accuracy is higher than the other four models. That is because the MSACNN model enhances the ability of fault feature extraction by introducing serial skip-layers connection structure and attention mechanism, thereby improving the fault diagnosis performance.

The experimental results mentioned above demonstrate that the proposed MSACNN model takes on high fault recognition rate under constant working conditions, with a recognition accuracy of 100%, which confirms that the MSACNN model exhibits excellent fault data classification capabilities under four constant working conditions and is an effective model for identifying faults in rolling bearings under such conditions.

For the sake of visually demonstrating the fault classification performance of the MSACNN model, the classification results are displayed using a confusion matrix and shown in Fig. 10.

The confusion matrix shows the classification of the MSACNN model for each class of fault samples, where the horizontal axis represents the predicted sample labels and the vertical axis represents the true sample labels; numbers on the main diagonal indicate the ratio of predicted sample labels consistent with the true sample labels. As can be seen from Fig. 10, the samples used for the model test are completely correct, indicating the excellent fault classification performance of the model.

In practical applications, rolling bearings operate under different load conditions, which can lead to change in their fault vibration frequencies, the fault features also vary accordingly. Consequently, it is essential to verify the fault diagnosis capability of the MSACNN model under varying load conditions.

The model is trained utilizing the training set data under 1HP, 2HP and 3HP loads respectively, and then the test set data under the other two load conditions are simultaneously diagnosed to verify the fault diagnosis capability of the

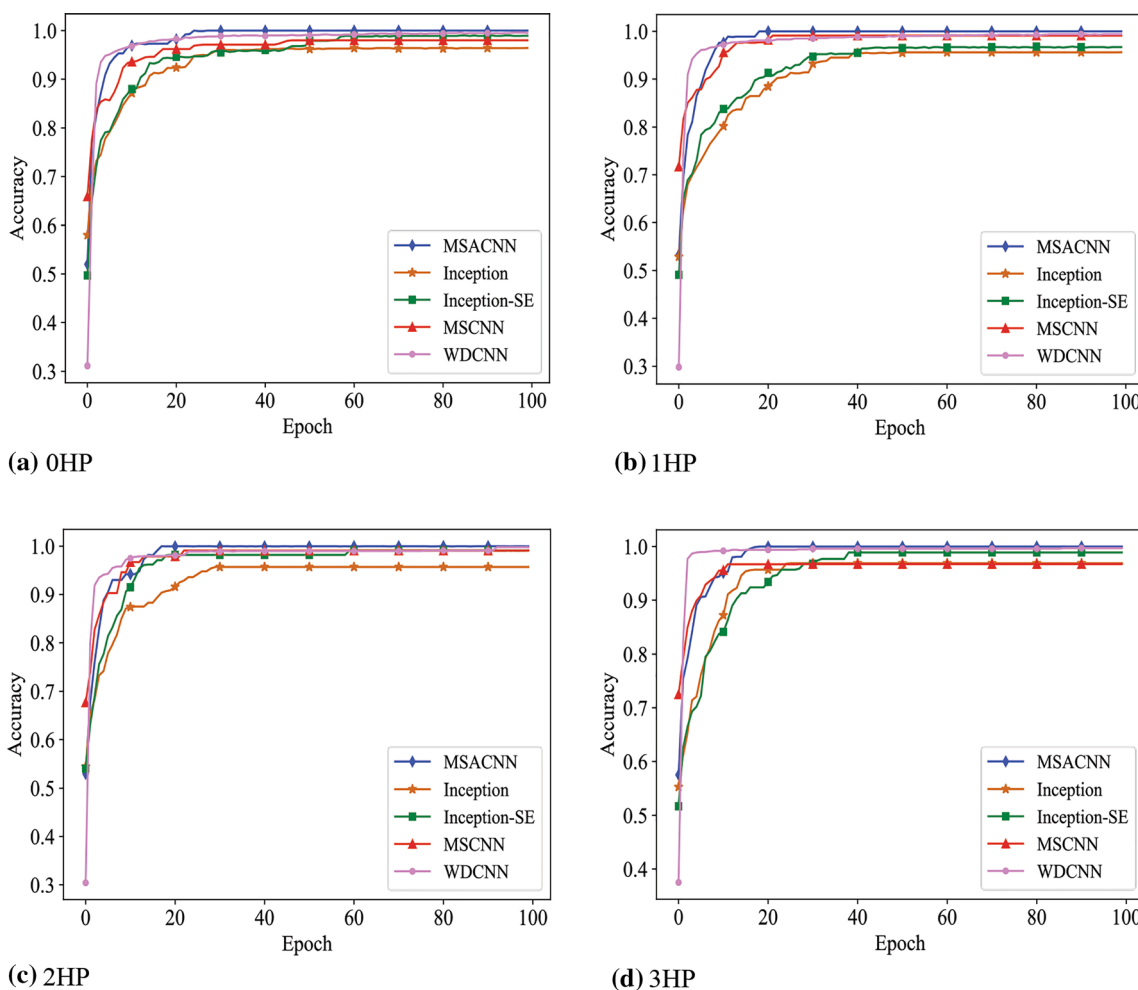


Fig. 9 Accuracy variation of training process of different models at four loads

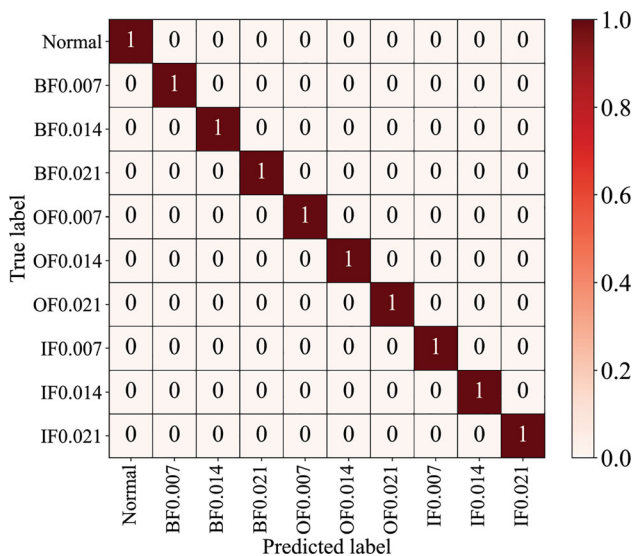


Fig. 10 Confusion matrix diagram of diagnosis result of test set

MSACNN model under variable load conditions. Additionally, compared with the experimental diagnosis results of the basic CNN model, the artificial feature selection and SVM (AFS + SVM) model [30] and the WDCNN model [13] on the CWRU data set, the results are presented in Fig. 11.

From Fig. 11, it can be observed that the traditional AFS + SVM method has a lower accuracy compared to the other three deep learning-based intelligent diagnosis methods, mainly because the weak adaptability of manual fault feature extraction and the insufficient nonlinear expression ability of SVM lead to its low fault recognition rate under different loads. Although the basic CNN model possesses strong nonlinear expression capability, its fault recognition rate is not high, and it has poor generalization performance. The WDCNN model utilizes a large convolutional kernel of size 116×1 to capture short-term features and extract more comprehensive fault feature information, leading to a fault recognition rate of up to 90%. The MSACNN model, with multi-scale module and

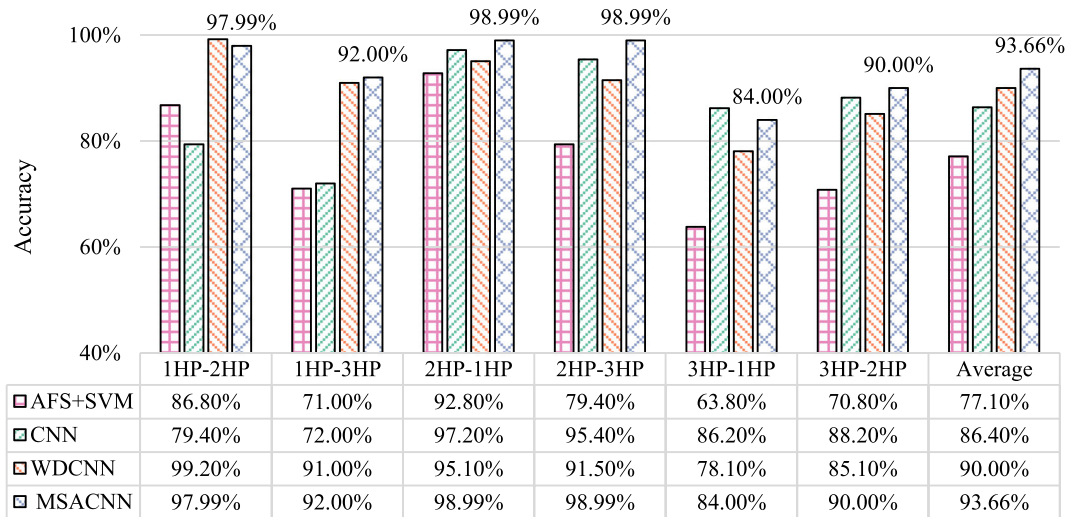


Fig. 11 Variable load experiment results of different models

SE module, captures crucial fault feature information and enables more accurate decision-making, with the fault transfer diagnostic accuracy reaching as high as 98.99%. However, in 1HP-2HP, the accuracy of WDCNN is higher than that of the MSACNN, which implies WDCNN model, featuring wider and deeper convolutional layers, is able to better adapt to the specific characteristics associated with the 1HP-2HP power settings. In other situations, the classification accuracies of MSACNN are significantly higher. This outcome aligns with the expectation, as increasing divergences between the target and source domains heighten the identification difficulty, leading to low classification accuracies. In summary, the MSACNN model possesses a good generalization performance under variable load conditions, maintaining the average recognition rate of over 90%.

In order to further verify the fault diagnosis capability of the MSACNN model under variable load conditions, the experimental results of the four aforementioned models are plotted as a box plot for fault diagnosis under changing conditions and shown in Fig. 12.

In Fig. 12, the upper quartile represents the data point that is at the 75% percentage when the data are sorted in ascending order, and the lower quartile represents the data point at the 25% percentage, with the distance between the upper quartile and the lower quartile in the box plot of fault diagnosis reflecting the degree of data fluctuation to some extent. It can be seen from Fig. 12 that the distance between the upper quartile and the lower quartile in the box plot of fault diagnosis accuracy for the MSACNN model is relatively small, which illustrates that the proposed method exhibits good stability. All these results in Figs. 8, 9, 10, 11, 12 demonstrate that the constructed model has higher

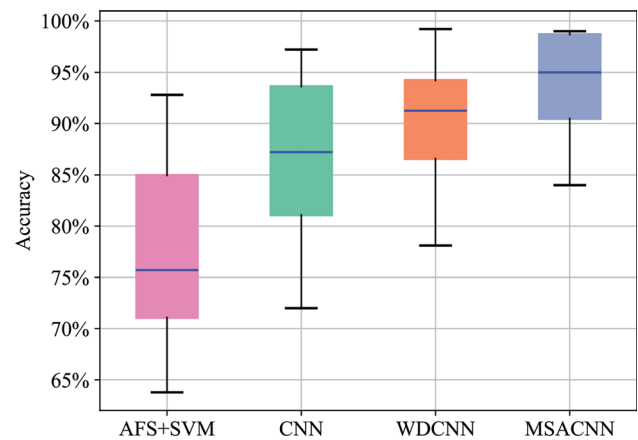


Fig. 12 Box plot of fault diagnosis results of different models

accuracy in bearing fault diagnosis even though the bearing operates under varying rotating speed and load conditions.

Case Study 2: PU Bearing Data Set

To further verify the effectiveness of the proposed model, the rolling bearing data set of the University of Paderborn (PU) in Germany is also employed to implement the fault classification and identification [31].

This dataset was constructed by utilizing accelerometers to collect vibration signals from the bearing seat, with a sampling frequency of 64 kHz. By changing the speed of the drive system to regulate the radial force on the bearing and the load torque to the drive system, the fault data under four different working conditions are easily obtained. In this case study, the data under condition 0 (i.e., rotating speed of 1500 r/min, load torque of 0.7 N·m, and radial force of 1000 N) was selected.

The bearing faults in this dataset are artificially induced by using three different methods: electric discharge machining (EDM), electric etching (ee), and drilling (dr). The dataset includes two types of faults and one normal type. The outer ring (OR) fault forms include electric etching single damage level 1, electric etching single damage level 2, drilling single damage level 1, drilling single damage level 2, and electric discharge machining single damage level 1. And the inner ring (IR) fault forms include electric etching single damage level 1, electric etching single damage level 2, and electric discharge machining single damage level 1. So, there are nine different operating states including normal. The whole dataset consists of 3780 training samples, 1080 test samples and 540 validation samples, and the length of each sample is 1024. The constitution of the dataset is shown in Table 4.

The model structure and super-parameters used in this experiment are the same as those described in case study 1. Similarly, to avoid randomness caused by the experiment only implemented one time, each experiment is implemented 10 times, and the average accuracy is taken as the experimental result. The experimental results of the proposed method are compared with those of the WDCNN and MSCNN methods, and the corresponding experimental results are shown in Fig. 13.

From Fig. 13, it can be clearly seen that the proposed method outperforms the other two models in terms of fault diagnosis results, achieving an accuracy of 99.54% and 98.11% for training set and test set, respectively. That is because the MSACNN method utilizes a multi-scale structure with serial skip connections and attention mechanism which are greatly profitable to extract more comprehensive fault features from the raw data. And consequently, the proposed method is capable of improving the accuracy of fault diagnosis for rolling bearings especially under different operating conditions.

To intuitively illustrate the data classification results obtained by the MSACNN method in this experiment, the

fault diagnosis results are displayed via the confusion matrix and shown in Fig. 14.

It can be seen from Fig. 14 that the MSACNN method performs better than the other two methods in terms of classifying each type of faults. In the confusion matrix of the MSACNN method, the highest misclassification rate is that a few OR01(EDM) samples are incorrectly classified as Normal samples, with a false classification rate of 3%. Compared with the other two methods, the MSACNN method has the fewest misclassification results of each operating state. In other words, the misjudgment rate in the confusion matrix of the MSACNN method is obviously smaller than that of the MSCNN method and WDCNN method, which also demonstrates that the MSACNN method is indeed superior to the other two methods in feature extraction and fault classification of rolling bearings.

Conclusion

To address the issue of insufficient feature extraction caused by single-scale convolution kernel of traditional CNN in rolling bearing fault diagnosis, from the perspective of feature information fusion, multi-scale feature extraction modules with different receptive fields and SE modules that can obtain important feature information are integrated into a one-dimensional convolutional neural network, thereby directly extracting important feature information of bearing faults from the raw vibration data. This method utilizes the nonlinear fitting ability of deep learning to automatically implement fault feature extraction and fault data classification, and then obtain ultimate fault classification results, the whole process of which can achieve intelligent “end-to-end” fault diagnosis. The experimental results show that the proposed method can extract sensitive features of faults by using a strategy of increasing network width and depth through the use of multi-scale convolutional layers. The proposed method not

Table 4 Constitution of PU dataset in this case study

Fault type	Damage type	Damage level	Label
Normal	Normal
OR	Electric etching	1	OR01(ee)
	Electric etching	2	OR02(ee)
	Drilling	1	OR01(dr)
	Drilling	2	OR02(dr)
	Electric discharge machining	1	OR01(EDM)
IR	Electric etching	1	IR01(ee)
	Electric etching	2	IR02(ee)
	Electric discharge machining	1	IR01(EDM)

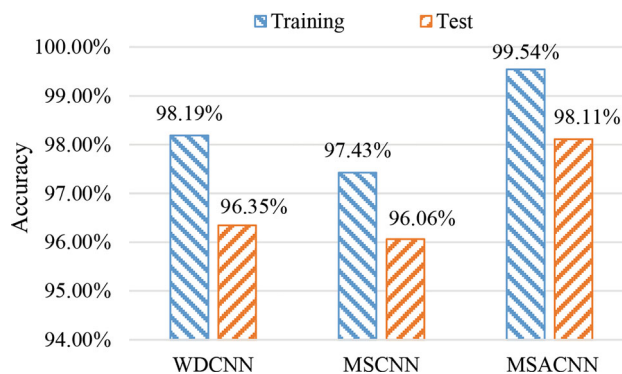
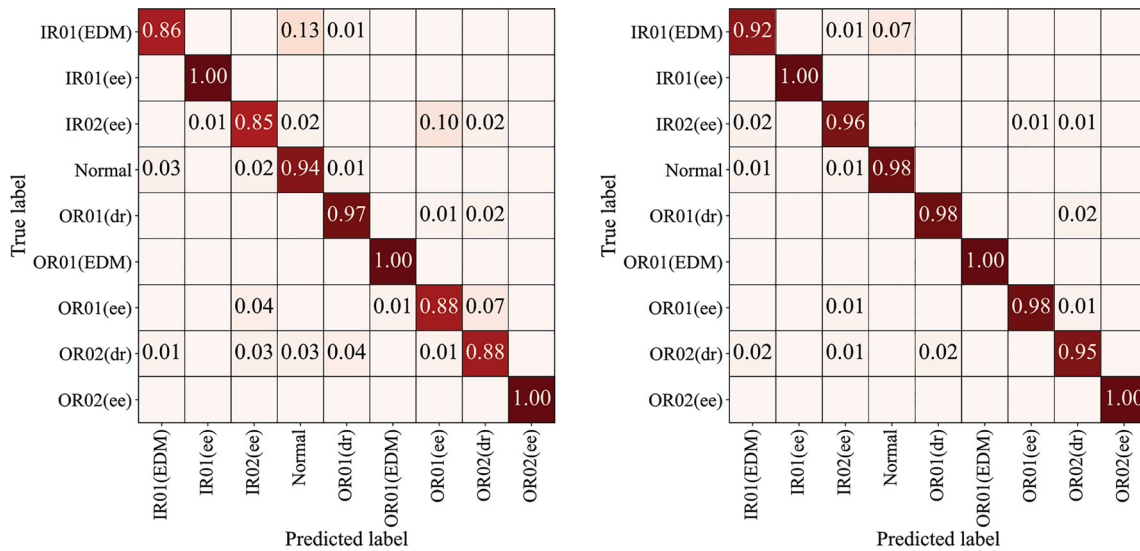
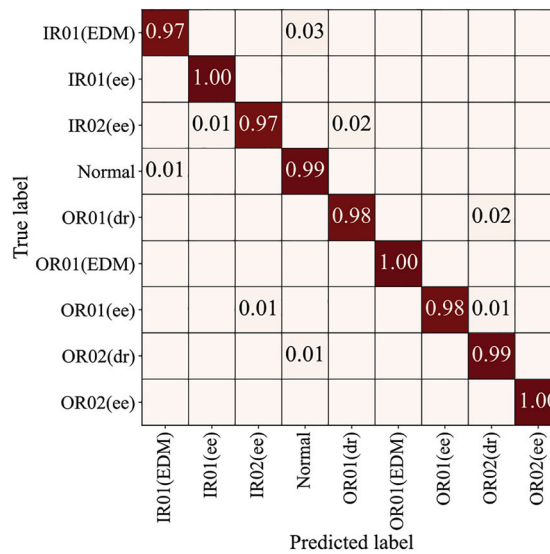


Fig. 13 Fault diagnosis results of different models



(a) WDCNN

(b) MSCNN



(c) MSACNN

Fig. 14 Confusion matrix graphs of diagnosis results of different models

only has high fault recognition accuracy under fixed operating conditions, but also achieves an average fault migration diagnosis accuracy of 93.66% under variable operating conditions, which indicates that the method has strong generalization capability. It can be reasonably deduced that the strategy of multi-scale feature information fusion described in this study actually possesses extraordinary potential in vibration signal processing and can provide a new solution for rolling bearing fault diagnosis under complex operating conditions.

In the actual working conditions, rolling bearings often face a more complex and changeable operating environment, so in the future work, the influences of varying

rotating speed conditions, noise interference and other factors on the fault diagnosis performance of rolling bearings will be considered, and the application of multi-scale neural network in engineering practices also will be investigated deeply.

Acknowledgments The authors would like to sincerely thank the Case Western Reserve University Bearing Data Center and the University of Paderborn Bearing Data Center for supplying fault bearing datasets, and the anonymous reviewers for their constructive suggestions and comments on this paper.

Funding This work was supported by the National Natural Science Foundation of China (Grant No. 62241308) and the Key Program of

Natural Science Foundation of Gansu Province (Grant No. 23JRRA761).

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. M. Yu, C. Yang, L. Liu, J. Su, An approach to recognize combined faults of rolling bearing by combing discrete Wavelet Transform and Generalized S Transform. *J. Fail. Anal. Prev.* **23**(1), 258–270 (2023)
2. J. Zhang, Y. Sun, L. Guo, H. Gao, X. Hong, H. Song, A new bearing fault diagnosis method based on modified convolutional neural networks. *Chin. J. Aeronaut.* **33**(2), 439–447 (2020)
3. G. An, Q. Tong, Y. Zhang, R. Liu, W. Li, J. Cao, Y. Lin, An improved variational mode decomposition and its application on fault feature extraction of rolling element bearing. *Energies.* **14**(4), 1079 (2021)
4. K. Wang, Y. Cheng, H. Zheng, L. Gao, Z. Feng, Based on stochastic resonance to enhance micro-fault signal features. *J. Fail. Anal. Prev.* **23**(3), 1203–1215 (2023)
5. E. Song, F. Gao, C. Yao, Y. Ke, Research on rolling bearing fault diagnosis method based on improved LMD and CMWPE. *J. Fail. Anal. Prev.* **21**(5), 1714–1728 (2021)
6. B. Qin, Q. Luo, Z. Li, C. Zhang, H. Wang, W. Liu, Data screening based on correlation energy fluctuation coefficient and deep learning for fault diagnosis of rolling bearings. *Energies.* **15**(7), 2707–2728 (2022)
7. B. Zhang, Rolling bearing fault detection system and experiment based on deep learning. *Comput. Intell. Neurosci.* **2022**, 8913859 (2022)
8. Z. Jin, Y. Sun, Research on bearing variable condition fault diagnosis based on RDADNN. *J. Fail. Anal. Prev.* **23**(4), 1663–1674 (2023)
9. E.H. Dhekra, E. Ines, K.N. Mai, H. Atef, A conic radon-based convolutional neural network for image recognition. *Int. J. Intell. Syst. Appl.* **15**(1), 1–12 (2023)
10. Y. Lecun, Y. Bengio, G. Hinton, Deep learning. *Nature.* **521**(7553), 436–444 (2015)
11. F. Wang, H. Jiang, H. Shao, W. Duan, S. Wu, An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Meas. Sci. Technol.* **28**(9), 095005 (2017)
12. S. Liu, H. Jiang, Y. Wang, K. Zhu, C. Liu, A deep feature alignment adaptation network for rolling bearing intelligent fault diagnosis. *Adv. Eng. Inform.* **52**, 101598–101610 (2022)
13. W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors.* **17**(2), 425–446 (2017)
14. T. Han, L. Zhang, Z. Yin, A. Tan, Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. *Measurement.* **177**, 109022 (2021)
15. X. Li, W. Zhang, Q. Ding, Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliab. Eng. Syst. Saf.* **182**, 208–218 (2019)
16. W. Wang, Study on motor fault diagnosis based on multi-scale convolutional neural network. *China Univ. Min. Technol.* **2020**, 38–50 (2020)
17. G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Trans. Industr. Electron.* **66**(4), 3196–3207 (2019)
18. H. Seungmin, O. Seokju, J. Jongpil, Bearing fault diagnosis based on multiscale convolutional neural network using data augmentation. *J. Sens.* **2021**, 6699637–6699651 (2021)
19. H. Wang, Z. Liu, D. Peng, Y. Qin, Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis. *IEEE Trans. Industr. Inf.* **16**(9), 5735–5745 (2019)
20. W. Zhang, D. Chen, Y. Xiao, H. Yin, Semi-supervised contrast learning based on multiscale attention and multitarget contrast learning for bearing fault diagnosis. *IEEE Trans. Industr. Inf.* **19**(10), 10056–10068 (2023)
21. L. Jia, T.W. Chow, Y. Wang, Y. Yuan, Multiscale residual attention convolutional neural network for bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **71**, 3519413 (2022)
22. K. Zhao, F. Jia, H. Shao, Unbalanced fault diagnosis of rolling bearings using transfer adaptive boosting with squeeze-and-excitation attention convolutional neural network. *Meas. Sci. Technol.* **34**(4), 044006 (2023)
23. J. Bai, B. Ding, Z. Xiao, L. Jiao, H. Chen, A.C. Regan, Hyperspectral image classification based on deep attention graph convolutional network. *IEEE Trans. Geosci. Remote Sens.* **60**, 5504316 (2022)
24. A. Galassi, M. Lippi, P. Torroni, Attention in natural language processing. *IEEE Transact. Neural Netw. Learn. Syst.* **32**(10), 4291–4308 (2021)
25. M. Cheng, J. Lin, S. Lu, S. Dong, X. Dong, Seismic data reconstruction based on multiscale attention deep learning. *IEEE Trans. Geosci. Remote Sens.* **61**, 5914718 (2023)
26. J. Hu, S. Li, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(1), 2011–2023 (2020)
27. T. Kobayashi, Disentangled convolution for optimizing receptive field. *Pattern Recogn. Lett.* **169**, 67–74 (2023)
28. Case School of Engineering, Case Western Reserve University Bearing Data Center Seeded Fault Test Data. <https://engineering.case.edu/bearingdatacenter>
29. Z. Feng, S. Wang, M. Yu, A fault diagnosis for rolling bearing based on multilevel denoising method and improved deep residual network. *Digital Signal Process.* **140**, 104106–104124 (2023)
30. J. Deng, Y. Zou, X. Zhang, Y. Jiang, L. Zhang, Application of an improved CNN in bearing fault diagnosis. *Modern Manuf. Eng.* **4**, 142–147122 (2020)
31. C. Lessmeier, J.K. Kimotho, D. Zimmer, W. Sextro (2016) *Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification*. In Proceedings of the European Conference of the Prognostics and Health Management Society. Bilbao, Spain, 5–8 July 2016, pp 1–17.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.