



# Rolling bearing fault diagnosis algorithm using overlapping group sparse-deep complex convolutional neural network

Fengping An · Jianrong Wang

Received: 8 September 2021 / Accepted: 18 February 2022 / Published online: 10 March 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

**Abstract** As the key component of a mechanical system, rolling bearings will cause paralysis of the entire mechanical system once they fail. In recent years, considering the high generalization ability and nonlinear modeling ability of deep learning, rolling bearing fault diagnosis methods based on deep learning have been developed, and good results have been achieved. However, because this kind of method is still in the initial development stage, its main problems are as follows. First, it is difficult to extract the composite fault signal feature of rolling bearings. Second, the existing deep learning rolling bearing fault diagnosis methods cannot address the problem of multi-scale information of rolling bearing signals well. Therefore, this paper first proposes the overlapping group sparse model. It constructs weight coefficients by analyzing the salient features of a signal. It uses convex optimization techniques to solve the sparse optimization model and applies the method to extract

features of rolling bearing composite faults. For the problem of extracting multi-scale feature information from rolling bearing composite fault signals, this paper proposes a new deep complex convolutional neural network model. This model fully considers the multi-scale information of rolling bearing signals. The complex information in this model not only has a rich representation ability but can also be used to extract more scale information. Finally, the classifier of this model is used to identify rolling bearing faults. This paper proposes a new rolling bearing fault diagnosis algorithm based on overlapping group sparse model-deep complex convolutional neural network. The experimental results show that the method proposed in this paper can not only effectively identify rolling bearing faults under constant operating conditions, but also accurately identify rolling bearing fault signals under changing operating conditions. Additionally, the classification accuracy of the method proposed in this paper is superior to that of traditional machine learning methods. The proposed method also has certain advantages over other deep learning methods.

---

F. An (✉)  
School of Physics and Electronic Electrical Engineering,  
Huaiyin Normal University, Huaian 223300, China  
e-mail: anfengping@163.com

F. An  
Beijing Institute of Technology, Beijing, China

J. Wang  
School of Mathematical Sciences, Shanxi University,  
Taiyuan 030006, SX, China  
e-mail: wangjian198010@163.com

**Keywords** Fault diagnosis · Rolling bearing · Overlapping group sparse · Deep complex convolutional neural network

## 1 Introduction

The timely identification of faults in the equipment operation process is of great significance to the safe operation of mechanical systems and can reduce or prevent major economic losses and catastrophic accidents [1, 2]. Rolling bearings are a key transmission component in the structure of an entire mechanical system; if they fail, they can cause an increase in the overall failure rate of the system, cause significant economic losses, and even cause serious safety accidents. Therefore, research on rolling bearing fault diagnosis methods has always been a key issue in the field of fault diagnosis. Rolling bearing fault diagnosis methods are generally divided into rolling bearing fault diagnosis methods based on feature analysis and rolling bearing fault diagnosis methods based on artificial intelligence [3–5].

Studies on rolling bearing fault diagnosis methods based on feature analysis mainly include the following. Bafroui et al. [6] used a discrete wavelet transform to study rotor fault feature extraction. Although it has good reconstruction characteristics and other advantages, this method has problems such as the loss of high-frequency fault characteristic information. Georgoulas et al. [7] used the empirical mode decomposition (EMD) method to obtain the original characteristics of vibration signals from a normal bearing and a faulty bearing, thereby realizing the abnormal recognition of a rolling bearing. Yu et al. [8] used ensemble EMD (EEMD) and singular value decomposition (SVD) to obtain useful fault features of rolling bearings and used the fault features for identification. Although the above methods have been widely used and promoted in the field of rolling bearing fault diagnosis, these methods have problems such as the artificial setting of feature information, weak adaptive ability, and poor robustness. Therefore, better rolling bearing fault diagnosis methods have been studied in the industry.

Fault diagnosis methods for rolling bearings based on artificial intelligence are divided into fault diagnosis methods for rolling bearings based on machine learning and deep learning. Fault diagnosis methods based on machine learning mainly include the method by Muruganatham et al. [9], who used SVD and feed forward back propagation neural networks (BPNNs) to diagnose different faults of rolling bearings. Ali et al. [10] used a BPNN as a classifier to diagnose the

running state of the rolling elements and inner and outer rings of a bearing. Li et al. [11] proposed a method for the fault diagnosis of rolling bearings based on a binary tree SVM model. Uddin et al. [12] proposed an enhanced k-nearest neighbors (KNN) classification algorithm using enhanced KNN to realize bearing fault diagnoses. The fault diagnosis methods based on machine learning have an improved diagnosis effect and adaptive ability compared with the fault diagnosis method based on feature analysis. Excellent machine learning methods have outstanding interpolation ability and rule processing between features, etc. Feature point methods mainly focus on accurate identification near feature points. However, this type of method has problems such as weak self-learning ability and weak robustness in the modeling process. In this context, deep learning [13, 14] has strong generalization and feature extraction abilities. Deep learning has been widely used in machine vision, image classification and natural language processing. Therefore, researchers have introduced deep learning to the field of fault diagnosis and have introduced fault diagnosis methods based on deep learning. Shao et al. [15] used a deep belief network (DBN) for the intelligent state monitoring of induction motors and used a DBN to automatically extract relevant features of vibration signals for state recognition. Shao et al. [16] used a DBN optimized by particle swarm optimization for rolling bearing fault diagnosis and introduced stochastic gradient descent to fine-tune the weights of the restricted Boltzmann machine (RBM) training process. Finally, the optimized DBN was used for fault diagnosis. Jiang et al. [17] proposed a multilayer deep learning convolutional neural network (CNN) for the fault diagnosis of rolling bearings. Wang et al. [18] proposed an adaptive CNN method and applied it to the fault diagnosis of rolling bearings. Islam et al. [19] used a deep learning model to monitor rolling bearing faults, used wavelet analysis to extract signal characteristics, and then used a deep learning model to classify faults. Zhou et al. [20] used a deep learning model to directly process vibration signals, and, combined with a regional adaptive method to diagnose faulty bearings, it can improve the model's diagnostic effect. Cabrera et al. [21] combined a deep CNN with a long- and short-term memory (LSTM) network model and used the combined model to estimate the bearing state. This approach achieved a better fault diagnosis effect.

In summary, deep learning models have been applied to the field of fault diagnoses of rolling bearing equipment. However, adaptive feature extraction and fault diagnoses based on deep learning have the following problems. First, it is difficult to extract the composite fault signal features of rolling bearings. Second, the existing methods cannot address the problem of multi-scale information of rolling bearing signals well. Therefore, this paper proposes applying the sparse characteristics of a structural group based on the overlapping group sparse model to represent a signal and then constructs the weight coefficients by analyzing the salient features of the signal. The existing convex optimization technology is used to solve the sparse optimization model. This method is applied to extract the features of weak composite faults of rolling bearings, and it can solve the problem of difficult feature extraction of composite fault signals of rolling bearings. In addition, to address the difficulty in extracting multi-scale information of rolling bearing signals with deep learning methods, this paper proposes a deep complex CNN. The model fully considers the multi-scale information characteristics of rolling bearing signals. The difference in scale characteristic information is used to distinguish fault category information. Plural information not only has a rich representation ability but also has the ability to promote memory and the retrieval of fault information. Experiments verify that the model has a good ability to identify rolling bearing faults and strong robustness.

The main contributions of this paper are as follows: (1) This paper proposes an overlapping group sparse model, which can solve the problem of difficult extraction of composite fault signal features of rolling bearings. (2) This paper proposes a deep complex convolutional neural network, and gives the model design and process description. It can make full use of the multi-scale information of the fault signal and help to improve the fault classification effect. (3) Based on the steps (1)–(2), this paper proposes a rolling bearing fault diagnosis algorithm based on the overlapping group sparse model-deep complex convolutional neural network. The experimental part gives the specific experimental environment and parameter settings.

Section II describes the overlapping group sparse model. Section III introduces the deep complex CNN. Section IV establishes a rolling bearing fault diagnosis

algorithm based on overlapping group sparse model-deep complex CNN. Section V conducts an experimental analysis of the method proposed in this paper and compares it with other mainstream methods. Finally, summarize and analyze the content of the paper.

## 2 Overlapping group sparse model

### 2.1 Group of sparse models

To solve the inverse regularization problem,  $z$  is recovered from the rolling bearing signal  $y(y = z + w)$ . Assuming that  $z$  is nonsparse, a certain sparseness will appear in the transform base  $\Phi = [\Phi_1, \dots, \Phi_M]$ .  $z$  can be expressed as the following formula:

$$z = \Phi\theta \quad (1)$$

In the formula,  $\theta$  is the sparse representation coefficient. The process of solving  $\theta$  is a sparse approximation process. The sparse model can be expressed as:

$$\min_{\theta \in R^M} \frac{1}{2} \|y - \Phi\theta\|^2 + I(\theta) \quad (2)$$

In the formula,  $I(\theta)$  is the regularization penalty function that induces the sparse solution  $\theta$ . The choice of  $I(\theta)$  depends on the knowledge of the sparse structure of the solution  $\theta$ . If  $\theta$  is sparse, then the regularization function  $I(\theta)$  can choose the  $l_1$  norm. It can be expressed as:

$$I(\theta) = \lambda \|\theta\|_1 \quad (3)$$

In the formula,  $\lambda > 0$  is the penalty parameter. It is a parameter to adjust the degree of compression. When  $\lambda$  is greater, the degree of compression is greater. It will make more coefficients approach zero. Conversely, the smaller the  $\lambda$ , the smaller the degree of compression. It will cause more coefficients to be retained.

The above defines the  $l_1$  norm form as the Lasso model on the basis of minimizing the residual sum of squares. Because it can obtain sparse solutions of high-dimensional data, the Lasso model is widely used for feature selection of high-dimensional data. It adds the same penalty function to each variable. In other words, it compresses the coefficients of each variable to the same degree.

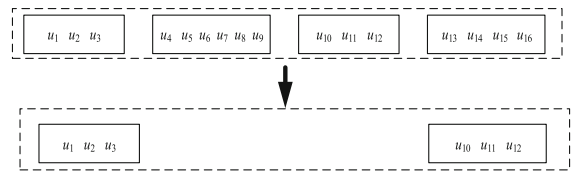
In some cases, there is a relationship among the coefficients. These related coefficients can be regarded as a whole. At this time, the Lasso model is not suitable for handling this relationship. It can be replaced by the group Lasso model. The group Lasso model is an expansion of the Lasso model and adds constraints to a set of coefficient vectors. It implements coefficient compression from the perspective of the group. If  $\theta$  is sparse, then  $I(\theta)$  is the group Lasso penalty function. It can be expressed as:

$$I(\theta) = \lambda \sum_{i=1}^I \|\theta_i\| \tag{4}$$

In the formula,  $i = 1, 2, \dots, I$ , all coefficients are divided into group I. With the change in  $\lambda$ , the sub-model vectors are either all 0 or not 0. The group Lasso model regards each group of coefficients as a “single” vector for selection under the condition that the coefficient vectors are divided into groups. If the coefficients in this group are not zero, then all coefficients in this group are selected. Conversely, if the set of coefficients are all zero, then the set of coefficients are all discarded. In the above formula,  $\lambda > 0$ . It is used to control the amount of contraction. The larger  $\lambda$  is, the more severe the compression, and its corresponding  $\theta_i$  is closer to zero. The complete group that has not contributed to the model is removed. Conversely, if  $\theta_i \neq 0$ , then all the coefficients in  $i$  are not zero. All the coefficients of this group are selected for the model to realize the selection from the perspective of the variable group. The selection effect of the group Lasso method is shown in Fig. 1. In Fig. 1,  $u_1, u_2, \dots, u_{12}$  represent components in the sparse group.

### 2.2 Overlapping group sparse model

The group sparse model does not consider that different sub-model vectors may contain the same certain variables. The combination of these variables is excluded by the group sparse model, which has limitations in practical applications. That is, in reality, there is overlap between groups and the variables contained in the groups. The aforementioned no overlapping group sparse model is no longer applicable in this situation. Therefore, this paper proposes an overlapping group sparse model, which allows variables between different groups to overlap, and then



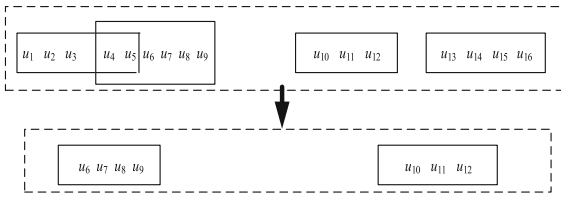
**Fig. 1** The sparse results of the group of Lasso (each rectangle in the figure is a group)

introduces the prior information of the “overlapping” structure into the model. The prior information here refers to the sparseness of the rolling bearing fault signal itself. Then, it constructs the weight coefficients by analyzing the salient features of the signal, and then uses the existing convex optimization technology to solve the sparse optimization model. This method is applied to the feature extraction of weak composite faults of rolling bearings. In other words, the prior information mainly refers to the sparseness and other characteristics contained in the rolling bearing signal. It can be integrated into subsequent fault diagnosis, and it can improve fault diagnosis modeling capabilities.

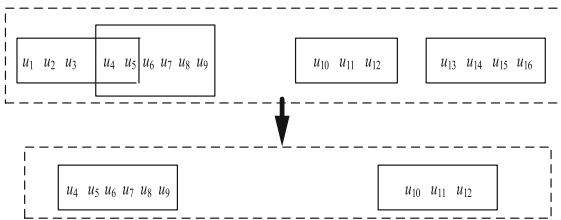
If  $\theta$  is a sparse group, the group overlaps with the group. Its regularization function  $I(\theta)$  can choose the overlapping group Lasso penalty function. It can be expressed as the following formula [22]:

$$I(\theta) = \lambda \sum_{j=1}^m \|\theta_{j,K}\|_2 \tag{5}$$

In the formula,  $\theta_{j,K} = [\theta(j), \dots, \theta(j + K - 1)] \in R^K$ , and it optimizes the experimental results by adjusting the values of parameters  $K$  and  $\lambda$ . Reference [23] noted that when groups have overlapping structures, the group lasso without overlapping structure in formula (5) is used to select variables within the group. Because group  $\{u_1, u_2, u_3, u_4, u_5\}$  is not selected, variables  $u_4$  and  $u_5$  are discarded when group  $\{u_1, u_2, u_3, u_4, u_5\}$  is discarded. Group  $\{u_4, u_5, u_6, u_7, u_8, u_9\}$  is selected, but the variables  $u_4$  and  $u_5$  are discarded. Therefore, its final selection effect will not contain variables  $u_4$  and  $u_5$ . The specific results are shown in Fig. 2. The reason for this situation is that the overlapping variables  $u_4$  and  $u_5$  between group  $\{u_1, u_2, u_3, u_4, u_5\}$  and group  $\{u_4, u_5, u_6, u_7, u_8, u_9\}$  are not considered. The variable selection effect of the overlapping group lasso is shown in Fig. 3. The group  $\{u_4, u_5, u_6, u_7, u_8, u_9\}$  is completely selected. Table 1



**Fig. 2** The result of sparseness of group Lasso when the signal has overlapping group structure



**Fig. 3** The result of sparseness of the overlapping group Lasso when the signal has an overlapping group structure

**Table 1** Comparison of calculation efficiency of various structural sparse models

Model type	Structure sparsity	Algorithm complexity
Lasso	Unstructured sparsity	$O(NP\min\{N, P\})$
Group Lasso	Group sparsity	$O(P + \max\{d_1, \dots, d_{ G }\} \ln  G )$
Overlapping Group Lasso	Overlapping group sparsity	$O(P G )$

shows the structural sparsity characteristics and algorithm complexity of each structural sparsity model.  $N$  is the number of samples.  $P$  is the sample dimension.  $\max\{d_1, \dots, d_{|G|}\}$  is the maximum dimension of the model vector in the group.  $G$  is the potential of the group assembly.

### 3 Deep complex convolutional neural network model

#### 3.1 Basic principles of deep complex convolutional neural network model

The operation and characterization of common deep neural networks are all based on the real number domain, but complex number signals appear

increasingly frequently in practical applications. Theoretical analysis shows that complex numbers not only have richer representation abilities but also help to retrieve signal feature memories. However, there are relatively few studies on building deep CNN modules based on complex numbers. A deep complex CNN [24] used complex batch normalization, a complex weight initialization strategy and an end-to-end training scheme. It has been applied to music transcription tasks. Therefore, this section proposes a deep complex CNN for rolling bearing signals, which can fully consider the multi-scale information of rolling bearing signals. Different scale information can distinguish the fault category. Deep complex CNNs include complex convolution operations, complex pooling operations, complex activation functions, and complex classifier optimization. Different scale information can distinguish the fault category. Its specific content is as follows:

##### 3.1.1 Reconvolution operation

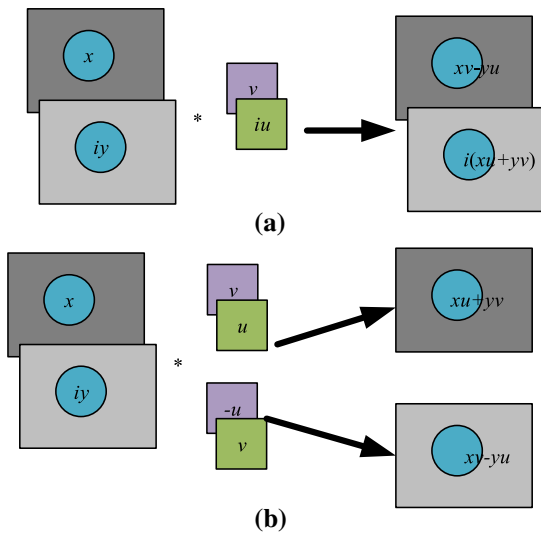
Let us first describe the complex number representation, specifically: a complex number  $v = a + ib$  has a real component  $a$  and an imaginary component  $b$ . For example, consider a typical real-valued convolution layer that has  $N$  feature maps; to represent these as complex numbers; we allocate the first  $N/2$  feature maps to represent the real components and the remaining  $N/2$  to represent the imaginary ones. The feature map here is evenly distributed to the real and imaginary parts of the complex number representation, and can be adjusted and optimized according to the actual situation of the data during actual operation.

In the complex convolution layer, the data of the input complex signal are  $v$ , the complex weight is  $w$ , and the offset information is  $c$ . Then the input convolution operation of the  $n$ th channel can be expressed by the following formula:

$$\begin{aligned} \Gamma &= v_h * w^n + c^n \\ &= (x_h + iy_h) * (v + iu) + (\alpha + i\beta) \\ &= (x_hv - y_hu + \alpha) + i(x_hu + y_hv + \beta) \end{aligned} \tag{6}$$

$$\text{Re}(\Gamma) = x_hv - y_hu + \alpha \tag{7}$$

$$\text{Im}(\Gamma) = x_hu + y_hv + \beta \tag{8}$$



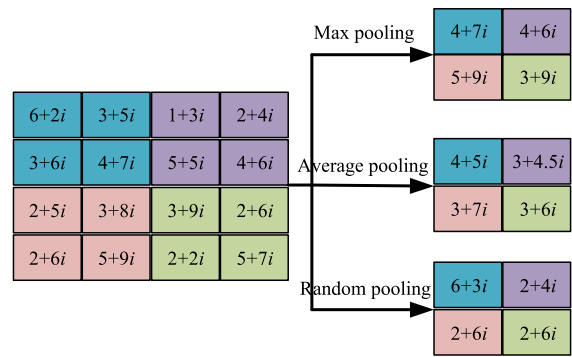
**Fig. 4** Equivalent complex and real convolution layer (**a** complex convolution, **b** real convolution)

In the formula,  $i = \sqrt{-1}$  is the imaginary unit.  $w^n = v + iu$  represents the complex connection weight of the CNN. It is the convolution kernel of the deep complex convolution network.  $c^n = +i$  represents the offset of all channels,  $u$  and  $v$  are real number matrices, and  $x$  and  $y$  are real number vectors. and are real numbers.  $\text{Re}(\Gamma)$  and  $\text{Im}(\Gamma)$  represent the real and imaginary parts of the complex number  $\Gamma$ , respectively. Figure 4 shows that complex convolution with a convolution kernel of  $v + iu$  is equivalent to a real network with two convolution kernels  $[v, u]$  and  $[-u, v]$ .

### 3.1.2 Repool operation

By performing convolution operation on a neighborhood in the image, it will get the neighborhood features of the image. After the characteristics of different locations are summarized, it is called pooling. Its main purpose is to reduce the amount of calculation by reducing the dimensionality of the input data. It has translation invariance. In the pooling layer, the input is broken down into blocks, and each block is replaced with a value. The complex pooling process with a core size of  $2 \times 2$  and a step size of 2 is shown in Fig. 5.

This paper fully preserves the integrity of the input data. The complex number pooling proposed in this



**Fig. 5** Schematic diagram of non-overlapping complex pooling

paper is a one-to-one correspondence between the real part and the imaginary part of the complex number. A complex number contains real and imaginary parts, and it cannot be compared. The complex random pooling method has the same rules as the random pooling method of the real number domain. Both methods calculate the probability of feature points in the neighborhood. The greater the probability is, the greater the probability of the feature point being selected. The maximum pooling of complex numbers can be achieved by calculating the modulus of the complex numbers. This outputs the complex number corresponding to the largest modulus. This approach can reduce the mean shift caused by the convolutional layer parameter error. More texture information will be retained. The formula for calculating the maximum pooling of the complex number  $z = x + yi$  is as follows:

$$h = \arg \max_{z \in patch} |z| \tag{9}$$

Complex average pooling averages the real and imaginary parts of feature points in the neighborhood. It can retain more background information.  $N_m$  is the nuclear size. The formula for calculating complex average pooling is:

$$h = \frac{1}{|N_m|} \sum_{j \in N_m} x_j + \frac{1}{|N_m|} \sum_{j \in N_m} y_j i. \tag{10}$$

### 3.1.3 Complex activation function

In the real number domain, most activation functions or transfer functions satisfy the characteristics of differentiability and boundedness. When the sigmoid

function of the real number domain is extended to the complex number domain  $f(z) = u(x,y) + iv(x,y)$ , the transfer characteristic of the complex number still satisfies the characteristics of boundedness and differentiability. Its complex activation function can be expressed as argmax, which is

$$f(z) = \frac{z}{d + \frac{1}{r}|z|} = \frac{x}{d + \frac{1}{r}\sqrt{x^2 + y^2}} + \frac{y}{d + \frac{1}{r}\sqrt{x^2 + y^2}}i \tag{11}$$

That is,

$$u(x,y) = \frac{x}{d + \frac{1}{r}\sqrt{x^2 + y^2}} \tag{12}$$

$$v(x,y) = \frac{y}{d + \frac{1}{r}\sqrt{x^2 + y^2}} \tag{13}$$

In the formula,  $c$  and  $r$  are positive real numbers. This nonlinear function is suitable for training a feedforward neural network and can be used for classification problems. The function is bounded, and the function is also bounded in the derivation. Its corresponding partial derivative is:

$$u_x = \begin{cases} \frac{r(dr|z| + y^2)}{|z|(dr + |z|)^2}, & \text{if } |z| \neq 0 \\ \frac{1}{d}, & \text{else} \end{cases} \tag{14}$$

$$u_y = \begin{cases} \frac{-rxy}{|z|(dr + |z|)^2}, & \text{if } |z| \neq 0 \\ 0, & \text{else} \end{cases} \tag{15}$$

$$v_x = \begin{cases} \frac{-rxy}{|z|(dr + |z|)^2}, & \text{if } |z| \neq 0 \\ 0, & \text{else} \end{cases} \tag{16}$$

$$v_y = \begin{cases} \frac{r(x^2 + dr|z|)}{|z|(dr + |z|)^2}, & \text{if } |z| \neq 0 \\ \frac{1}{d}, & \text{else} \end{cases} \tag{17}$$

### 3.1.4 Complex batch normalization

As the number of layers of the deep learning model increases, the output values of the last few layers are close to zero. Because the output value is the multiplication factor of the gradient, its back

propagation gradient is very small. In this case, it is difficult to update the parameters of the deep learning model. In order to avoid the network falling into the problem of local optimization, this section proposes to use complex batch normalization strategy to weaken the influence of model initialization [25]. The formula for complex batch normalization is as follows:

$$BN(h; \gamma, \beta) = \beta + \gamma \tilde{h} \tag{18}$$

$$h = \text{Re}(h) + \text{Im}(h) \tag{19}$$

$$\hat{h}_i = \frac{h_i - u_B(h)}{\sqrt{V}} \tag{20}$$

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \begin{pmatrix} \text{cov}(\text{Re}(h), (\text{Re}(h))) & \text{cov}(\text{Re}(h), \text{Im}(h)) \\ \text{cov}(\text{Im}(h), \text{Re}(h)) & \text{cov}(\text{Im}(h), \text{Im}(h)) \end{pmatrix} \tag{21}$$

In the formula,  $h$  represents the activation vector.  $h$  is the normalization of the mean value  $u_B = 0$ , the covariance  $K = 1$ , and the pseudocovariance matrix  $C = 0$ .  $V$  is the covariance matrix. and  $\gamma$  represent the displacement parameter and scaling parameter to be learned, respectively.  $\gamma$  is a positive semidefinite matrix, which can be expressed as:

$$\gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \tag{22}$$

### 3.1.5 Fully connected layer

In the entire deep complex convolutional network, the function of the complex fully connected layer is to map the distributed feature representation learned by the complex convolution layer, the complex maximum pooling layer and the complex activation function layer to the sample label space. Its purpose is to reduce the original rolling bearing signal input, and then use all the characteristic information to the maximum extent through the fully connected layer. The fully connected layer can be regarded as a convolutional layer with a  $1 \times 1$  convolution kernel, that is, it transforms the input data into a one-dimensional vector. It then points to the vector, and the specific formula is as follows:

$$\begin{aligned}
 o_i &= \sum_{j=1}^m u_j * w_{ij} + c_i \\
 &= \sum_{j=1}^m (x_i + jy_i) * (u_{ij} + jv_{ij}) + (\alpha_i + j\beta_i) \\
 &= \sum_{j=1}^m [(x_i u_{ij} - y_i v_{ij}) + (y_i u_{ij} + x_i v_{ij})j] + (\alpha_i + j\beta_i)
 \end{aligned}
 \tag{23}$$

In the formula,  $U = \{u_1, \dots, u_m\}$  is the input.  $O = \{o_1, \dots, o_m\}$  is the output.

### 3.1.6 Complex output layer

The output layer in the complex form is designed with reference to the superposition state in quantum mechanics. Because qubits in quantum computing can express two possible states, n qubits represent  $2^n$  states. Therefore, we can apply this idea to the model presented in this section. The rolling bearing signal enters the output layer after the convolution and pooling layer, and the probability corresponding to the amplitude of each rolling bearing signal is initialized to be equal and the cumulative sum is 1. The model performs a quantum observation, and each signal amplitude is mapped to a corresponding position that satisfies the probability distribution, and finally classified by probability. In addition, the function expressing the quantum state of the particle satisfies the normalization condition. The sum of the distribution probabilities of the particles is equal to 1. For the multi-classification problem of rolling bearing fault signals, the output value of the complex node corresponds to the probability of a category after being normalized by the softmax nonlinear function. It satisfies the constraint of probability distribution, that is,

$$\sum_{k=1}^K p(k) = 1$$

. The  $l_2$  norm of the complex number domain input  $z = x + yi$  ( $x, y \in R$ ) can be defined as A. For softmax regression, the probability of each class of the  $K$  class classifier corresponding to its input feature  $z$  is  $P(y = klz)$ . That is

$$h_\theta(x) = \begin{bmatrix} P(Y = 1|z; \theta_1) \\ \vdots \\ P(Y = k|z; \theta_k) \end{bmatrix} = \frac{1}{\sum_{h=1}^k \|e^{\theta_h^T Z}\|} \begin{bmatrix} \|e^{\theta_1^T Z}\| \\ \vdots \\ \|e^{\theta_k^T Z}\| \end{bmatrix}
 \tag{24}$$

In the formula,  $\theta_1, \dots, \theta_k \in Q^n$  are the parameters of the reconvolutional neural network model and  $e^{\theta_h^T Z} = e^{\theta_h^T x} + e^{\theta_h^T y}i$  is in the Softmax regression. For input data  $Z = \{z^{(1)}, \dots, z^{(m)}\}$ , its cost function is defined as follows:

$$I(\theta) = -\frac{1}{m} \left[ \sum_{g=1}^m \sum_{h=1}^k J\{Y^{(g)} = h\} \log \frac{\|e^{\theta_1^T Z^g}\|^2}{\|e^{\theta_h^T Z^g}\|^2} \right]
 \tag{25}$$

In the formula, when  $Y(g)$  is  $h$ , the index function  $J\{Yg = h\} = 1$ , otherwise it is 0.

The  $\theta$  of  $I(\theta)$  can be updated by the gradient descent method. The gradient expression formula is as follows:

$$\begin{aligned}
 \nabla_{\theta_{h(g)}} I(\theta) &= \frac{1}{m} \left[ J\{Y^{(g)} = h\} 2 \log \frac{(e^{\theta_{h(g)}^T Z^{(t)}})^2 Z_{q(t)}^2}{\|e^{\theta_h^T Z}\|^2} \right] \\
 &\quad - 2 \log \frac{(e^{\theta_{h(t)}^T Z^{(t)}})^2 Z_{q(t)}^g}{\sum_{l=1}^k \|e^{\theta_l^T Z}\|^2}
 \end{aligned}
 \tag{26}$$

In the formula,  $t = 1$  means calculating the real part.  $t = 2$  means that the imaginary part is calculated. The  $\theta$  vector update is defined as:

$$\theta^{new} = \theta^{old} - \alpha \nabla_{\theta_{h(g)}} I(\theta)
 \tag{27}$$

From the above analysis, it can be seen that the deep complex convolutional neural network model extends the real number domain model to the complex number domain model. Therefore, the overall framework of the model consists of several repeated reconvolution layers, complex pooling layers and complex activation functions.



### 3.2 Process description of deep complex convolutional neural network model

Weight optimization is a key part of the deep complex convolutional neural network model. In the weight update process, the gradient descent algorithm based on deep complex CNN is introduced in this section. The training process is shown in Algorithm 1 of Table 2.

**Table 2** Algorithm 1 process

---

$K$  represents the number of training layers,  $L$  represents the cost function,  $a_0$  is the minimum block input, clip() is a split weight operation, and batchnorm() normalizes complex batches;

Input: weight  $W_t$ , bias  $b_t$ , batch normalization parameter  $\theta_t$  and learning rate  $\gamma_t$ ;

Forward propagation

For  $k = 1$  to  $K$  do

$s_k \leftarrow a_{k-1}W_k$

$a_k \leftarrow \text{batchnorm}(s_k, \theta_k)$

Back propagation

Initialize activation gradient  $\partial L/\partial a_k$

For  $k = K$  to 2 do

Calculation  $\partial L/\partial a_{k-1}$

Parameter accumulation

$a \leftarrow \text{updateparameter}(a_k, \partial L/\partial a_k, \gamma_k)$  // Update rules: Stochastic Gradient Descent(SGD)

$\gamma_{k+1} \leftarrow \text{updatelearningrate}(\gamma_k, k)$

Output: update parameters  $W_{t+1}$  and  $b_{t+1}$ ;

---

**Table 3** Algorithm 2 (complex batch normalization) process

---

Input: activation vector  $h$ , small batch  $B = \{h_{1...m}\}$  and model parameters: and  $\gamma$ ;

Output:  $\{y_i = BN_{\beta, \gamma}(h_i)\}$

$u_B = E[\tilde{h}]$  // Average operation

$K = E[(\tilde{h} - u_B)(\tilde{h} - u_B)^*] = V_{11} + V_{12} + i(V_{21} - V_{12})$  // Variance operation

$C = E[(\tilde{h} - u_B)(\tilde{h} - u_B)] = V_{11} - V_{12} + i(V_{21} + V_{12})$  // Pseudocovariance operation

$\hat{h}_i = (h_i - u_B(h))/\sqrt{\bar{v}}$  // Standardization

$h_i = \beta + \gamma\hat{h}_i = BN(h_i; \beta, \gamma)$  // Scale and scale information

---

#### 3.2.1 Batch normalization to initialize complex weights

This paper uses the random initialization method to initialize the weights of the deep complex CNN. Its expression is as follows:

$$W = |W|e^{i\theta} = u + jv \tag{28}$$

In the formula,  $|W|$  and  $\theta$  represent the size and scale characteristics of the weight  $W$ , respectively.

The complex batch normalization operation decorrelates the real and imaginary parts of a unit. It can reduce the risk of overfitting. The specific normalization process is shown in Algorithm 2 of Table 3.

#### 3.2.2 Complex number back propagation operation

The purpose of complex back propagation training is to find suitable weights and biases through optimization. It enables the loss function  $L(W)$  to get the minimum value. Since the magnitude of the gradient value is very small, it is a gradient accumulation effect. After multiple accumulations, the gradient noise will be consumed by the gradient. For large data sets, the cost of evaluating the gradient of the entire data set is very expensive. The general method of its processing is the stochastic gradient descent method [26]. To deal with this problem, for deep complex convolutional neural networks, the real part and imaginary part of the weights are updated according to formula (7) and formula (8), respectively.

To improve the convergence speed of stochastic gradient descent, this paper adopts the batch norm conversion method. The average value of the weight parameters is calculated as shown in the following formula:

$$\bar{W}_t = \frac{1}{t} \sum_{j=1}^t w_j \tag{29}$$

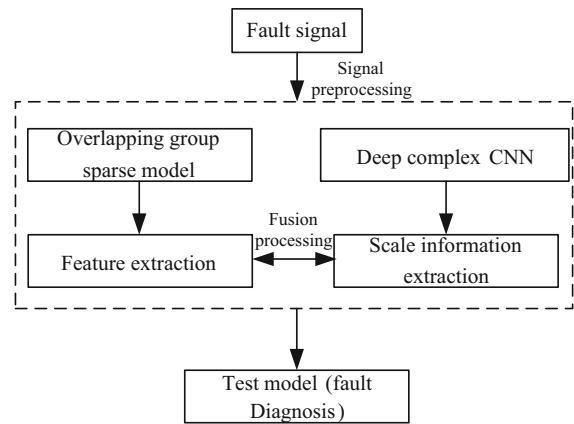
Then calculate the loss function based on the average weight  $\bar{W}_t$ . At the same time, standardized methods are used to standardize rolling bearing data. Its purpose is to reduce the difference in the order of magnitude of the weight parameters.

### 3.3 Rolling bearing fault diagnosis algorithm based on overlapping group sparse model-deep complex convolutional neural network

In this section, on the basis of Section II and Section III, this paper proposes a rolling bearing fault diagnosis algorithm based on overlapping group sparse model-deep complex CNN. The basic steps of the proposed fault diagnosis algorithm are as follows:

1. Data collection. The relevant rolling bearing data under different health conditions are obtained through public rolling bearing data sets.
2. Data preprocessing. The data obtained in step (1) are standardized.
3. Data feature extraction. The model proposed in Section II of this paper is used to extract features from rolling bearing data. The model can obtain richer and more complete feature information. It can better address the problem of the difficulty in extracting features of composite fault signals of rolling bearings.
4. Multi-scale information extraction and fault classification. The model proposed in Section III of this paper is used to extract multi-scale information from rolling bearing data. It can obtain different scale information of rolling bearing fault signals. Finally, the classifier of the model is used to classify and identify rolling bearing faults.
5. Actual test. The test samples are input into the rolling bearing fault diagnosis model trained in steps (1) to (4), and the model can give the rolling bearing fault category or result.

The basic framework diagram of the rolling bearing fault diagnosis algorithm proposed in this paper is shown in Fig. 6.



**Fig. 6** The basic framework of the fault algorithm proposed in this paper

**Table 4** Rolling bearing parameter information table

Bearing type	Bearing pitch diameter (mm)	Rolling body diameter (mm)	Number of rolling elements	Contact angle (°)
SKF6205	39.04	7.94	11	0

## 4 Experimental verification

### 4.1 Data set description

This experiment uses bearing data published by Case Western Reserve University in the US for verification. The bearing failure test bench mainly includes induction motors, torque sensors, and dynamometers. This section takes SKF6205 deep groove ball bearings as the research object. It collects vibration signals under four different working conditions: normal state, inner ring failure, outer ring failure, and rolling element failure. The specific bearing parameters are shown in Table 4. The rolling elements and outer ring collect vibration signals of four failure levels. The outer ring collects vibration signals of three failure levels. All vibration signals are collected under 0, 1, 2, and 3 hp motor loads. The sampling frequency is 12 kHz. Norm means no fault. G1, G2, G3, and G4 indicate rolling element failures of 0.007, 0.014, 0.021, and 0.028 inches, respectively. IR1, IR2, IR3, and IR4 indicate

0.007, 0.014, 0.021, and 0.028 inch inner ring failure, respectively. OR1, OR2, and OR3 indicate 0.007, 0.014, and 0.021 inch outer ring failure, respectively. The experiment separately studies the recognition performance of the method proposed in this paper for fault data under constant working conditions and changing working conditions. (Data Sources: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>).

For constant conditions, this experiment divides the data into two groups to verify the method proposed in this article. It contains normal bearing, inner ring failure data, outer ring failure data, and rolling element failure data. The specific grouping information is shown in Table 5. The first 204,800 points of each type of experimental signal under constant working conditions are divided into 200 sets of data samples. The length of each data sample is 1024 points. Then 150 sets of sample data are randomly selected as training samples, and the remaining 50 sets are used as test samples for state recognition. A total of  $4 \times 150 = 600$  training samples and  $4 \times 50 = 200$  test samples are formed in each group of experiments. Each set of experiments is equivalent to a four-category problem.

Under changing conditions, the data are divided into two groups, and the experimental verification is conducted. Both experiments use Norm, IR1, OR1, and B1 as training samples under zero loads. Then, Norm, IR1, OR1, and B7 under 1, 2, and 3 loads are

**Table 5** Grouping information table under constant working conditions

Group category	Drehzahl (r/min)	Normal	Innerer ring	Outer ring	Rolling element
Group 1	1750	Norm	IR1	OR1	G1
Group 2	1750	Norm	IR2	OR2	G2

**Table 6** Grouping information table under changing conditions

Experiment group	Sample type	Drehzahl (r/min)	Normal	Innerer ring	Outer ring	Rolling element
Group 1	Training samples	1750	Norm	IR1	OR1	G1
	Test samples	1772	Norm	IR1	OR1	G1
Group 2	Training samples	1797	Norm	IR1	OR1	G1
	Test samples	1772	Norm	IR1	OR1	G1

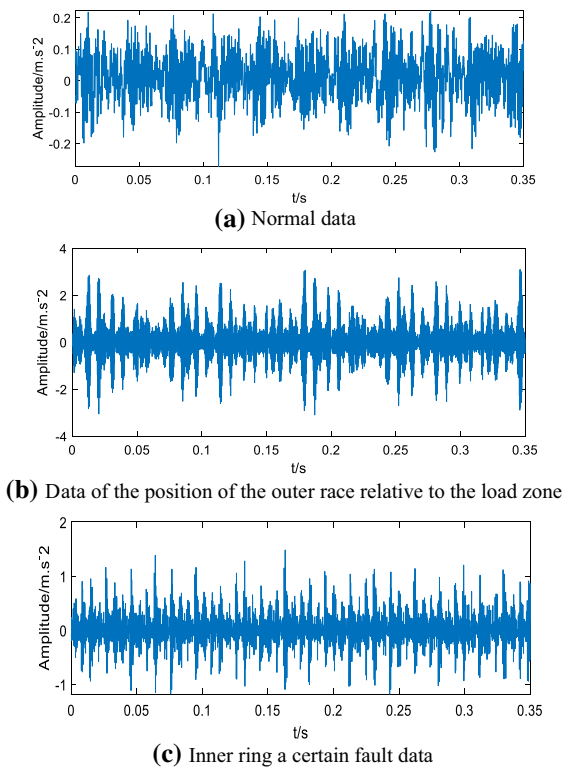
taken as test samples for state recognition. The two sets of experimental information are shown in Table 6. The first 204,800 points of each type of experimental signal under the four loads are divided into 200 sets of data samples. The length of each data sample is 1,024 points. A total of 600 data samples under zero loads are used as training samples. Two hundred data samples under 1, 2, and 3 loads are used as test samples for state recognition. Each set of experiments is a four-category problem.

The time-domain waveform of a certain data sample in the above four working states is shown in Fig. 7. Figure 7 shows that the time-domain waveform of the bearing vibration signal has nonlinear and nonstationary characteristics. For this signal, this article first uses the model proposed in the second part to characterize the fault. Then, the model proposed in the second part of the deep model is used to extract the dimension information of rolling bearings. Finally, the classifier of the deep complex convolutional neural network model is used to identify the type and degree of bearing fault.

#### 4.2 Experimental environment and parameter settings

This experiment deploys the proposed fault diagnosis algorithm on the Pytorch framework. This experiment is based on 2 GeForce RTX 3090GPU (24G memory) for accelerated training.

In deep learning, the optimization algorithm mainly affects the strategy of parameter update when the network is back propagated. Practice has proved that the adaptive algorithm has strong applicability and better convergence in network training. In this method, Adam [27] is selected to optimize the network parameters. Adam method can calculate the adaptive learning rate of each parameter. Adam method



**Fig. 7** Sample diagram of test data set

converges faster than other adaptive learning rate algorithms. In addition, it can solve problems such as large fluctuations in the loss function.

The optimized initial learning rate is set to 0.01, and the number of optimization iterations is set to 200. The learning rate is adjusted using the Step method. The specific setting is to adjust the learning rate to one-tenth of the original when the number of training times is 130 and 230. In the training process, the weight attenuation strategy is used to regularize the model parameters to prevent over fitting, and the attenuation factor is set to  $5 \times 10^{-5}$ .

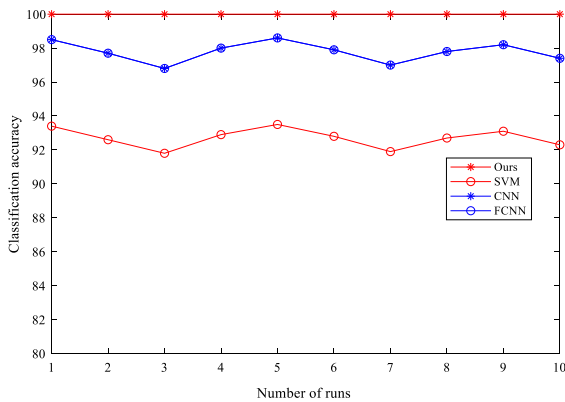
**Table 7** Fault diagnosis results obtained after running different methods 10 times (Group 1)

Method type	Diagnostic accuracy obtained by different methods (%)		
	Max	Min	Average
SVM	93.5	91.8	92.7
CNN	98.6	96.8	97.8
Deep CNN	100	98.2	99.0
Ours	100	100	100

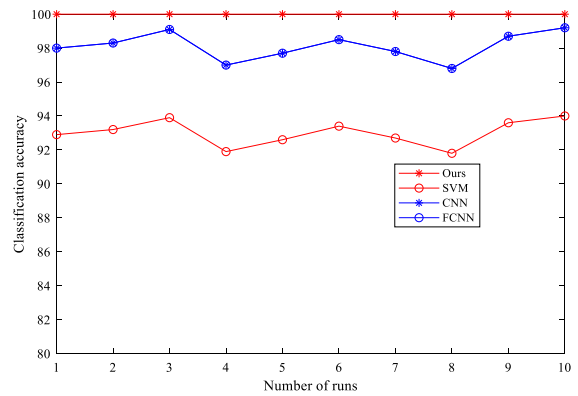
### 4.3 Diagnosis results and comparative analysis under stable conditions

First, the data of group 1 are used under constant working conditions for verification. First, the model proposed in the second part of this article is used to extract the characteristic information of the fault signal on the same timescale. Then, the model proposed in the third part of this article is used to extract different timescale feature information and fault classification information of the signal. Then, 600 data samples are randomly selected as the training set and the remaining 200 are used as the test set. Furthermore, to verify the effectiveness and advantages of the method proposed in this paper, the traditional SVM [28], the CNN [29], and the deep CNN [30] are used to perform fault analysis on the group 1 data. To avoid the contingency of the fault diagnosis results, the above four methods are all run 10 times. The specific results are shown in Table 7 and Fig. 8. Similarly, using the method in this paper, the traditional SVM [28], the CNN [29], and the deep CNN [30], experiments are performed on the group 2 data. The experimental results are shown in Table 8 and Figs. 9, 10.

From Tables 7, 8 and Figs. 8, 9, we can see that for group 1 and group 2, the classification accuracy of the SVM method is approximately 93%. However, the CNN method achieves 98% classification accuracy. Furthermore, the classification accuracy of the deep CNN is 1% higher than that of the CNN because this method is an optimized deep learning model. This shows that the optimized deep learning method has a certain effect on improving the accuracy of rolling bearing fault diagnosis. The method in this paper has the highest classification accuracy among all methods, and its accuracy is 100%. This shows that the method proposed in this paper not only greatly improves the classification accuracy of the traditional



**Fig. 8** The fault classification results obtained after running different methods 10 times (Group 1)

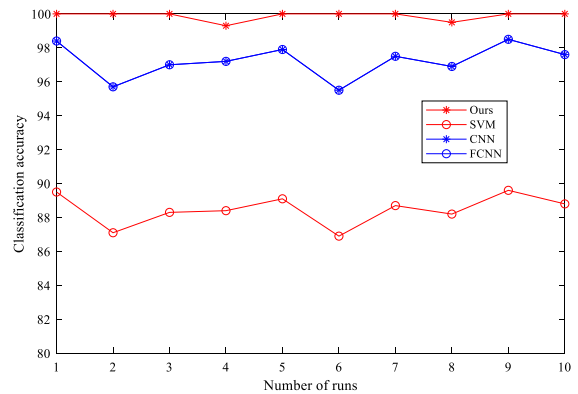


**Fig. 9** The fault classification results obtained after running different methods 10 times (Group 2)

SVM. It also improves the classification accuracy of deep learning methods, such as the CNN and deep CNN methods. This shows that the method proposed in this paper is highly adaptable to rolling bearing signals. It can better extract the fault feature information and multi-scale information of rolling bearing signals. This verifies that the rolling bearing fault diagnosis algorithm proposed in this paper has good stability and robustness. This is mainly because the method proposed in this paper is based on the characteristics of rolling bearing signals. It not only solves the problem of difficult extraction of rolling bearing signal features but also better extracts the multi-scale information of rolling bearing signals.

#### 4.4 Diagnosis results and comparative analysis under changing working conditions

To further verify the effectiveness of the fault diagnosis method proposed in this paper, the experimental process is similar to (2) experiment. This section uses the method of this paper, the traditional SVM [28], the CNN [29], and the deep CNN [30] to conduct experiments on the data samples of group 1



**Fig. 10** The fault classification results obtained after running different methods 10 times (Group 3)

and group 2 under changing conditions. To avoid the contingency of the fault diagnosis results, the above four methods are all run 10 times. The specific results are shown in Table 9 and Figs. 9, 10.

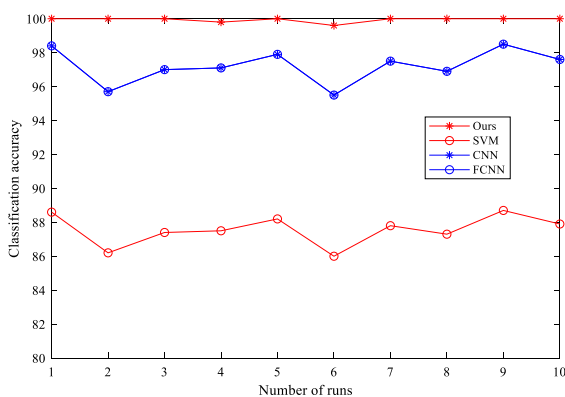
It can be seen in Table 8 and Figs. 10, 11 that for the group 1 and group 2 data, the classification accuracy of the SVM method is only 87%, which is approximately 6% lower than the classification accuracy under

**Table 8** Fault diagnosis results obtained after running different methods 10 times (Group 2)

Method type	Diagnostic accuracy obtained by different methods (%)		
	Max	Min	Max
SVM	93.9	91.8	93.0
CNN	99.2	96.8	98.1
Deep CNN	100	98.3	99.1
Ours	100	100	100

**Table 9** Fault diagnosis results obtained after running 10 times of different methods

Group type	Method type	Diagnostic accuracy obtained by different methods (%)		
		Max	Min	Average
Group 1	SVM	89.6	86.9	88.5
	CNN	98.5	95.5	97.2
	Deep CNN	100	96.7	98.3
	Ours	100	99.3	99.9
Group 2	SVM	88.7	86.0	87.6
	CNN	98.5	95.5	97.2
	Deep CNN	100	96.6	98.3
	Ours	100	99.6	99.9

**Fig. 11** The fault classification results obtained after running different methods 10 times (Group 2)

constant working conditions. This shows that the machine learning method that is similar to SVM has a poor fault classification effect under changing working conditions. The classification accuracy of the CNN method is only reduced by approximately 0.9% compared with the classification accuracy under constant working conditions. The classification accuracy of the deep CNN method is only reduced by approximately 0.8% compared with the classification accuracy under constant working conditions. Its reduction is smaller than that of the CNN method. For both the CNN and deep CNN models, the accuracy of fault classification under changing conditions is still very high. Their accuracies can reach more than 97%. This shows that deep learning methods can adapt to data changes very well. This is due to the great generalization ability of deep learning. The method in this paper still has the highest classification accuracy of all methods, and its accuracy is 99.9%. This shows

that the classification accuracy of the method proposed in this paper is lower than the classification accuracy of deep learning methods, such as the CNN and deep CNN methods, under changing conditions. This objectively verifies that the method proposed in this paper is more adaptable to the signal characteristics of rolling bearings than other deep learning methods. This proves that the proposed method not only has good classification accuracy but also has better adaptability and robustness. The main reasons why the method in this paper has such a good classification effect are as follows. First, the method in this paper is proposed to identify the signal characteristics of rolling bearings. Second, this paper solves the problem of single fault signal feature extraction and the problem of multi-scale information extraction of compound fault signals.

## 5 Conclusion

Aiming at the difficulty of extracting features and scale information of composite fault signals of existing rolling bearings, this paper proposes a new overlapping group sparse model, which can effectively extract the single feature extraction problem of the composite fault signal of rolling bearings. Additionally, to better extract the scale information of the composite fault signals of rolling bearings, this paper proposes a new deep complex convolutional neural model. The plural form in the model not only has richer characterization ability but also helps to extract different time scale information of rolling bearing composite fault signals. It also helps to remember the fault information. This paper proposes a method for

composite fault diagnosis of rolling bearings, that is, a fault diagnosis algorithm for rolling bearings based on an overlapping group sparse model-deep complex CNN.

The related bearing data experiments show that the method proposed in this paper can accurately identify all faults under constant working conditions. It not only greatly improves the recognition accuracy compared with the SVM method but also improves the recognition accuracy to a certain extent compared with other deep learning methods. This directly verifies the effectiveness of the fault diagnosis method proposed in this article. Under changing conditions, the fault classification accuracy obtained by the method proposed in this paper is the highest. The classification accuracies of the method proposed in this paper, SVM, and other deep learning methods under changing working conditions, which are 0.8%, 6%, and 2%, respectively, are lower than the classification accuracies under constant working conditions. It can be seen that the accuracy of the method proposed in this article decreases the least. This further verifies that the method proposed in this paper can adapt to different working conditions for the diagnosis and identification of rolling bearing fault signals.

**Acknowledgements** This paper is supported by National Natural Science Foundation of China (No. 61701188), Natural Science Foundation of Jiangsu Province (No. BK20201479), China Postdoctoral Science Foundation funded project (No. 2019M650512), and Future Network Scientific Research Fund Project (No. FNSRFP-2021-YB-43).

#### Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Data Availability Statement** The data used to support the findings of this study are included within the paper. (Data Sources: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>).

#### References

- Manieri, F., Stadler, K.: Morales-Espejel G E", The origins of white etching cracks and their significance to rolling bearing failures". *Int. J. Fatigue* **120**, 107–133 (2019)
- Li, X., Zhang, W., Ding, Q.: Multi-layer domain adaptation method for rolling bearing fault diagnosis. *Signal Process.* **157**, 180–197 (2019)
- Huang, W., Cheng, J., Yang, Y.: Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection. *Mech. Syst. Signal Process.* **114**, 165–188 (2019)
- Chen, B., Shen, B., Chen, F.: Fault diagnosis method based on integration of RSSD and wavelet transform to rolling bearing. *Measurement* **131**, 400–411 (2019)
- An, F.P.: Rolling bearing fault diagnosis algorithm based on FMCNN-sparse representation. *IEEE Access* **7**, 102249–102263 (2019)
- Bafroui, H.H., Ohadi, A.: Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions. *Neurocomputing* **133**, 437–445 (2014)
- Georgoulas, G., Loutas, T., Stylios, C.D.: Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. *Mech. Syst. Signal Process.* **41**(1–2), 510–525 (2013)
- Yu, K., Lin, T.R., Tan, J.W.: A bearing fault diagnosis technique based on singular values of EEMD spatial condition matrix and Gath-Geva clustering. *Appl. Acoust.* **121**, 33–45 (2017)
- Muruganatham, B., Sanjith, M.A., Krishnakumar, B.: Roller element bearing fault diagnosis using singular spectrum analysis. *Mech. Syst. Signal Process.* **35**(1–2), 150–166 (2013)
- Ali, J.B., Fnaiech, N., Saidi, L.: Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **89**, 16–27 (2015)
- Li, Y., Xu, M., Wang, R.: A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **360**, 277–299 (2016)
- Uddin, S., Islam, M., Khan, S.A.: Distance and density similarity based enhanced-nn classifier for improving fault diagnosis performance of bearings. *Shock. Vib.* **2016**, 1–11 (2016)
- Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2006)
- Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006)
- Shao, S., Sun, W., Wang, P.: Learning features from vibration signals for induction motor fault diagnosis. In *IEEE International Symposium on Flexible Automation (ISFA)*, pp. 71–76, (2016)
- Shao, H., Jiang, H., Zhang, X.: Rolling bearing fault diagnosis using an optimization deep belief network. *Meas. Sci. Technol.* **26**(11), 1–17 (2015)
- Jiang, H., Wang, F., Shao, H.: Rolling bearing fault identification using multilayer deep learning convolutional neural network. *J. Vibroeng.* **19**(1), 138–149 (2017)
- Fuan, W., Hongkai, J., Haidong, S.: An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Meas. Sci. Technol.* **28**(9), 95–104 (2017)
- Islam, M.M.M., Kim, J.M.: Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network. *Comput. Ind.* **106**, 142–153 (2019)

20. Zhou, F., Yang, S., Fujita, H.: Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl.-Based Syst.* **187**, 104–118 (2020)
21. Cabrera, D., Guamán, A., Zhang, S.: Bayesian approach and time series dimensionality reduction to LSTM-based model-building for fault diagnosis of a reciprocating compressor. *Neurocomputing* **380**, 51–66 (2020)
22. Yuan, L., Liu, J., Ye, J.: Efficient methods for overlapping group lasso. *Adv. Neural Inf. Process. Syst.* pp. 352–360 (2011)
23. Obozinski, G., Jacob, L., Vert, J.P.: Group lasso with overlaps: the latent group lasso approach. *arXiv preprint arXiv:1110.0413* (2011)
24. Trabelsi, C., Bilaniuk, O., Zhang, Y.: Deep complex networks. *arXiv preprint arXiv:170509792* (2017)
25. Cooijmans, T., Ballas, N., Laurent, C.: Recurrent batch normalization. *arXiv preprint arXiv:1603.09025* (2016)
26. Wang, L., Yang, Y., Min, R.: Accelerating deep neural network training with inconsistent stochastic gradient descent. *Neural Netw.* **93**, 219–229 (2017)
27. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. “*arXiv preprint arXiv:1412.6980* (2014)
28. Zhao, H., Gao, Y., Liu, H.: Fault diagnosis of wind turbine bearing based on stochastic subspace identification and multi-kernel support vector machine. *J. Modern Power Syst. Clean Energy* **7**(2), 350–356 (2019)
29. Huang, W., Cheng, J., Yang, Y.: An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. *Neurocomputing* **359**, 77–92 (2019)
30. Xu, Z., Li, C., Yang, Y.: Fault diagnosis of rolling bearing of wind turbines based on the variational mode decomposition and deep convolutional neural networks. *Appl. Soft Comput.* pp. 1–15 (2020)

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