Journal of Dynamics, Monitoring and Diagnostics, 2023, 2, 275-289
https://doi.org/10.37965/jdmd.2023.260

ISTP
RESEARCH ARTICLE

Rolling Bearing Fault Diagnosis Based On Convolutional Capsule Network

Guangjun Jiang,[1,2] Dezhi Li,[1,2] Ke Feng,[3] Yongbo Li,[4] Jinde Zheng,[5] Qing Ni,[6] and He Li[7]

[1]School of Mechanical Engineering, Inner Mongolia University of Technology, Hohhot, Inner Mongolia, China
[2]Inner Mongolia Key Laboratory of Advanced Manufacturing Technology, Hohhot, Inner Mongolia, China
[3]Department of Industrial Systems Engineering and Management, National University of Singapore, 117576, Singapore
[4]School of Aeronautics, Northwestern Polytechnical University, Xi'an 710068, China
[5]School of Mechanical Engineering, Anhui University of Technology, Ma'anshan 243002, China
[6]School of Mechanical and Mechatronic Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia
[7]Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

(Received 14 April 2023; Revised 28 May 2023; Accepted 05 June 2023; Published online 06 June 2023)

Abstract: Fault diagnosis technology has been widely applied and is an important part of ensuring the safe operation of mechanical equipment. In response to the problem of frequent faults in rolling bearings, this paper designs a rolling bearing fault diagnosis method based on convolutional capsule network (CCN). More specifically, the original vibration signal is converted into a two-dimensional time–frequency image using continuous wavelet transform (CWT), and the feature extraction is performed on the two-dimensional time–frequency image using the convolution layer at the front end of the network, and the extracted features are input into the capsule network. The capsule network converts the extracted features into vector neurons, and the dynamic routing algorithm is used to achieve feature transfer and output the results of fault diagnosis. Two different datasets are used to compare with other traditional deep learning models to verify the fault diagnosis capability of the method. The results show that the CCN has good diagnostic capability under different working conditions, even in the presence of noise and insufficient samples, compared to other models. This method contributes to the safe and reliable operation of mechanical equipment and is suitable for other rotating scenarios.

Keywords: continuous wavelet transform; convolutional capsule network; fault diagnosis; rolling bearings

I. INTRODUCTION

With the progress and development of intelligent manufacturing, mechanical equipment plays an important role in the trend of intelligent manufacturing, and rolling bearings are indispensable parts in most mechanical equipment [1]. Rolling bearings are more prone to damage due to their complex working environment of variable load and speed changes. The failure of rolling bearings not only affects the operation of equipment but also leads to serious safety issues and significant economic losses. Therefore, improving the accuracy and efficiency of fault diagnosis is very important.

Nowadays, the processing and analysis methods of fault signals have also been developed and widely used, from traditional time domain waveform analysis to new analysis methods such as wavelet analysis [2], Wigner-Ville technique [3], Hilbert demodulation [4], and other time–frequency analysis. The increasing number of mechatronic products in the market highlights the need for automated fault diagnosis methods. Traditional fault diagnosis methods rely heavily on empirical knowledge and require multiple indicators to reach a diagnosis, making it challenging to meet the demands of the rapidly growing market [5].

With the development of artificial intelligence, machine learning methods such as artificial neural networks [6], Bayesian classifiers [7], and support vector machines [8] began to emerge gradually, which are applied for extracting features to diagnose bearing failures. Even though bearing fault features are reliable indicators of machinery health, extracting these features usually requires complex mathematical techniques. Furthermore, the feature extraction methods used for different types of faults may vary. Therefore, manual extraction of fault features in diagnosis methods is highly dependent on the knowledge and expertise of experts [9]. As the number of bearing monitoring points and data volume grow, traditional diagnostic methods cannot meet the demands of big data analysis [10].

With the development of deep learning, which has achieved excellent results in various fields, it has also gained more and more attention in fault diagnosis [11]. In particular, the rise of convolutional neural networks (CNN) [12], a usual deep learning method, has gradually led to intelligent and automated fault diagnosis. Huang *et al.*

Corresponding authors: Ke Feng (e-mail: ke.feng@outlook.com.au), He Li (e-mail: he.li@centec.tecnico.ulisboa.pt).

[13] proposed an end-to-end process architecture and a rolling bearing fault diagnosis model based on convolutional neural attention module-CNN, capable of adaptively extracting fault features and free from the reliance on manual processing of complex signals. Guo *et al.* [14] constructed an improved convolutional generative adversarial network to improve the accuracy of bearing fault diagnosis under complex operating conditions with the help of the data generation capability of generative adversarial networks and the feature extraction capability of improved deep convolutional networks. Xie *et al.* [15] proposed a hybrid model based on CNN and individual classifiers to diagnose bearing faults. Pan *et al.* combined CNN with Long Short Term memory networks to achieve good results in bearing fault diagnosis [16]. Wang *et al.* proposed a multi-scale learning neural network that includes one-dimensional and two-dimensional CNN, significantly improving CNN [17]. Liu *et al.* used variational mode decomposition and CNN to perform fault diagnosis and degradation identification on planetary gears and achieved excellent results [18]. Gou *et al.* utilized continuous wavelet transform (CWT) and CNN models to achieve fault diagnosis of sensors in aircraft engine control systems [19]. Zhang *et al.* used three CNN networks with different activation functions to extract fault features of the original signal from different angles, then fused and transformed the fault features into 2D DCNN for fault diagnosis. The experimental results show that the features extracted by this method are more comprehensive and can significantly improve fault diagnosis accuracy [20]. Zhi *et al.* proposed an intelligent fault diagnosis method based on CNNs to solve the problem of imbalanced bearing data [21]. The above literature has applied CNN to various fields of engineering practice and achieved excellent results. However, the training of deep neural networks, such as CNN, relies on many samples, and the models suffer from overfitting factors when the training data is insufficient. Moreover, CNN is difficult to extract signal features contaminated by noise in noisy environments.

In traditional neural networks, each neuron is scalar in feature transmission and does not carry spatial position features, resulting in weak fault diagnosis capabilities. The capsule network was first proposed by Sabour *et al.* [22] in 2017, and its emergence solved this problem. Specifically, each neuron in the capsule network is a vector, not a scalar. This enables the capsule network to extract more comprehensive detailed features from the input data while reducing the loss of spatial feature information. To effectively address the problem of traditional deep learning models being unable to effectively extract spatial feature information when detecting fault signals, decreasing fault diagnosis ability. In this paper, we design a fault diagnosis method using convolutional capsule networks (CCN), which combines CNN with capsule networks capable of extracting more comprehensive features to perform complete extraction of fault features. The innovations and contributions of this paper are summarized as follows:

(1) This article designs a fault diagnosis method based on CCN, which avoids the drawbacks of manually extracting features and relying on expert experience. It can adaptively learn features and provide a foundation for implementing intelligent fault diagnosis.

(2) This article transforms the original vibration signal into a time-frequency domain signal after CWT processing, which can fully express the amplitude characteristics and frequency components of nonstationary signals, making the network more capable of learning features.

(3) This article improves the feature extraction layer of the capsule network to a combination of the convolutional layer and pooling layer, which can extract deeper features and reduce the number of parameters. However, the backend of CCN still uses the capsule network to vectorize and mine the spatial information of features.

(4) This article verifies the fault diagnosis ability of CCN in noisy environments and under insufficient samples, providing a foundation for solving the problem of difficulty in collecting fault data in engineering practice and the presence of noise pollution during the collection process.

The rest of this paper is organized as follows: Section II describes the theory, and the methodology is introduced in Section III; Section IV uses experimental datasets to verify the effectiveness of the developed methodology; conclusions are drawn in Section 5.

II. THEORY

A. CONVOLUTIONAL NEURAL NETWORKS

CNN was proposed by Lecun in 1998 [23] as one of the representative algorithms of deep learning. It is a class of feed-forward neural networks with a deep structure that includes convolutional computation. CNN has achieved proud results in several fields. In the field of fault diagnosis, CNN is also gaining more and more attention. Typically, CNN contains a convolutional layer, a pooling layer, and a fully connected layer, and their basic framework is shown in Fig. 1.

The role of the convolution layer is to extract information from the input image, which is called image features; these features are represented by each pixel in the image by combining or independently, for example, the texture features of the image and the color features. The image convolution operation is shown in Fig. 2.

The convolution expressions are

$$f(x) = \sum_{i,j}^{n} \theta_{i,j} * x_{i,j} + b \tag{1}$$

In Equation (1), $f(x)$ represents the output feature and $\theta_{i,j}$ represents the size of the convolution kernel elements in the ith row jth column, the $x_{i,j}$ represents the ith row jth column element size, * represents the convolution operation, and b is the bias.

The pooling layer is also one of the cores of CNN, and the pooling layer usually comes after the convolutional layer. The pooling operation can reduce a large number of training parameters while ensuring the invariance of the feature map space and can prevent the occurrence of overfitting. The feature map undergoes maximum and average pooling operations, as shown in Fig. 3.

The fully connected layer acts as a classifier in a CNN, transforming all feature matrices into one-dimensional feature vectors. The fully connected layer is generally at

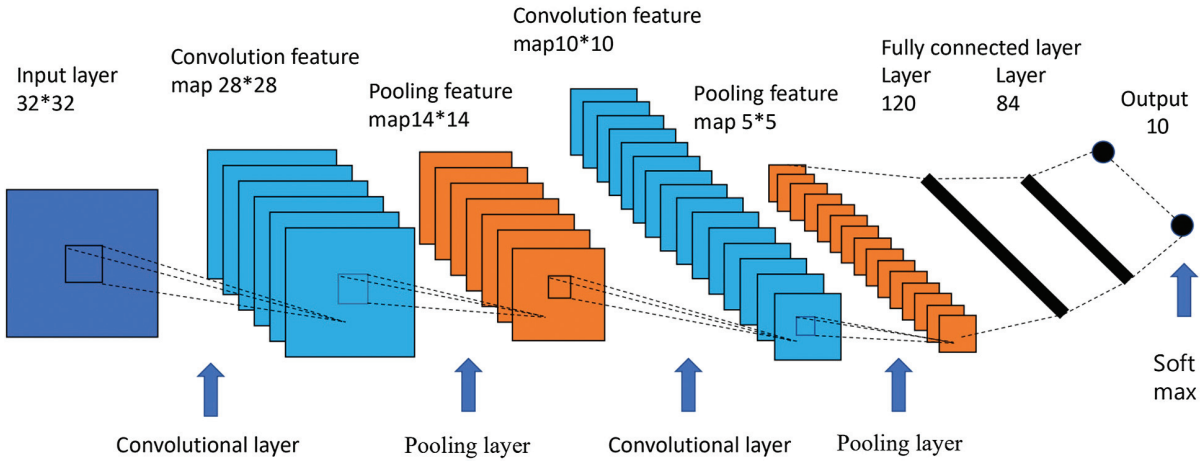

Fig. 1. Convolutional neural network framework diagram.

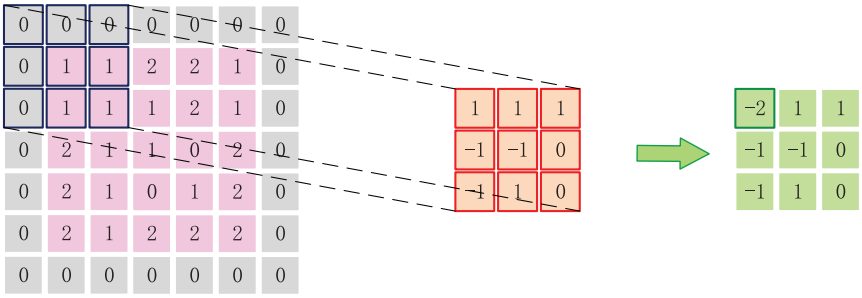

Fig. 2. Schematic diagram of the convolution operation.

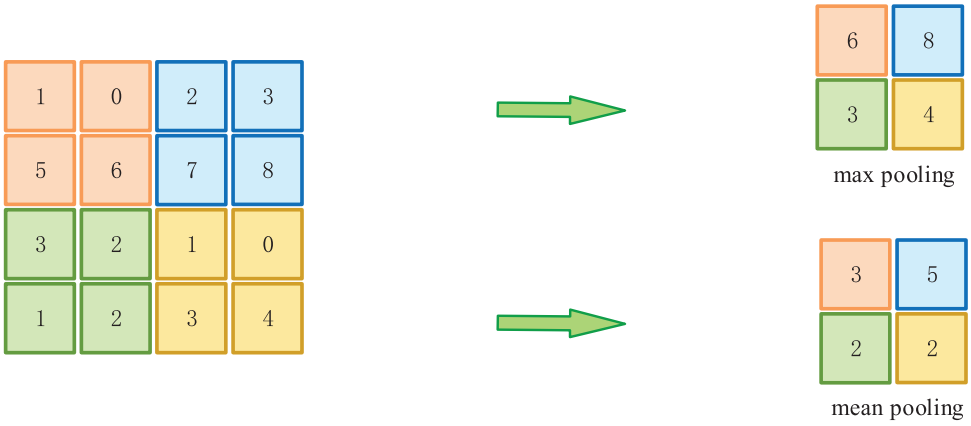

Fig. 3. Schematic diagram of maximum and average pooling.

the very end of the structure in a CNN and is responsible for the final output of the model.

B. CAPSULE NETWORKS

Capsule networks are used for feature delivery by vector neurons, which can well capture the position-relative relationship between features and avoid the loss of position feature information.

The computational process of the capsule network can be divided into three steps. In the first step, the prediction vector $u_{j|i}$ of the capsule network is the neuron u_i multiplied by the weight value w_{ij}, u_i is the ith neuron, w_{ij} is the weight matrix, and u_j is the jth vector generated by the prediction of the ith input feature. Then its formula can be expressed as

$$u_{j|i} = w_{ij}u_i \tag{2}$$

In the second step, the output vector s_j is obtained by multiplying the prediction vector $u_{j|i}$ by the coupling coefficient c_{ij}. The formula can be expressed by Equation (3) c_{ij} is the coupling coefficient between the ith vector in the main capsule layer and the jth vector in the digital capsule layer.

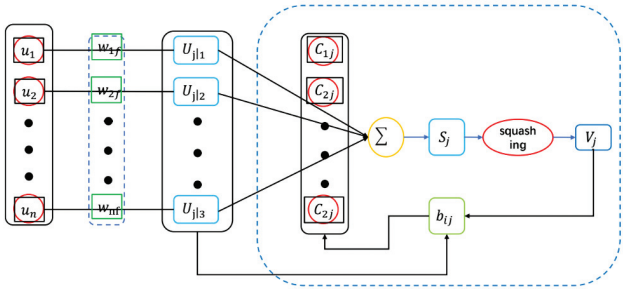

Fig. 4. Dynamic routing algorithm operation process diagram.

$$s_j = \sum_i c_{ij} u_{j|i} \tag{3}$$

In the third step, the output vector v_j is calculated by the nonlinear transformation of the total output vector s_j. s_j denotes the total output vector at the jth layer. The nonlinear transformation function is shown in Equation (4).

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \tag{4}$$

The coupling coefficient c_{ij} is obtained by a dynamic routing operation, the purpose of which is to allow the input neurons to be intelligently selected for transmission to the next layer of neurons according to the features they carry. It is calculated as in Equation (5) and Equation (6)

$$c_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}} \tag{5}$$

$$b_{ij} = b_{ij} + v_j u_{ji} \tag{6}$$

Figure 4 shows the operation process of the dynamic routing algorithm, in which the initialization of the paranoid coefficients b_{ij} is done using 0 pairs, and the coupling coefficient c_{ij} is calculated using Equation (5) to derive the output vector v_j. The value of the new paranoid coefficients b_{ij} is calculated using Equation (6) to calculate the value of the new c_{ij} and the value of s_j is further modified by the dynamic routing algorithm to change the value of the output vector v_j. The value of the output vector v_j is further modified by the dynamic routing algorithm.

C. CONTINUOUS WAVELET TRANSFORM

CWT is a signal processing method that is gaining popularity in the field of fault diagnosis [24]. The CWT can be implemented by the following operations [25]:

$$cwt(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \varphi^* \left(\frac{t-b}{a} \right) dt \tag{7}$$

In Equation (7), a is the scale parameter, b is the translation parameter, $x(t)$ is the original time-domain signal, φ is the wavelet function, and φ^* is the complex conjugate of φ.

The time–frequency map obtained by CWT can observe both time domain and frequency domain signals, so CWT plays an essential role in signal processing

III. PROPOSED METHODS

A. STRUCTURE OF THE MODEL

This paper proposes combining a CNN and a capsule network to form a CCN, which can effectively extract information from two-dimensional time-frequency maps. CCN, its network structure is shown in Fig. 5.

As shown in Fig. 5, the data need to be processed before being input to the CCN, and the original signal is first acquired for data processing to convert into a time–frequency image. The CCN mainly consists of two convolutional layers, a pooling layer, and a capsule layer, and the model has a simple structure with relatively few layers. The front end of the network uses the convolutional layer to extract feature information pooling layer to reduce the number of parameters. The back end uses the capsule network to convert the extracted features into vector neurons and the dynamic routing algorithm to achieve the transfer of features. After the second convolutional layer, CCN retains the pooling layer of the CNN. Here, the pooling operation is mainly used to reduce the number of parameters of the model and further improve the training speed of the model. Moreover, the pooling layer can also effectively prevent the occurrence of overfitting and improve the model's generalization ability. The primary capsule layer contains the convolution operation, while the convolution result is constructed as a capsule in vector form as the input of the digital capsule layer. The digital capsule layer accurately categorizes fault features by calculating the correlation between capsule layers through dynamic routing

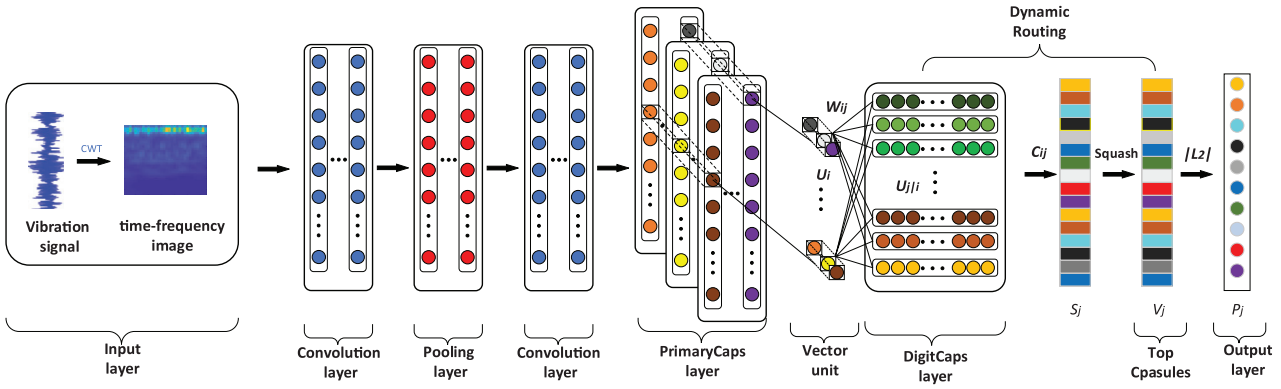

Fig. 5. CCN structure.

algorithms. The CCN front-end combines a convolutional layer and a pooling layer, which can reduce training parameters and extract fault characteristics at a deeper level. The backend of CCN uses a set of capsule structures to vectorize and mine the spatial information of features. In the process of feature transmission in CCN, neurons are vector neurons that can carry spatial position features. CCN can extract more effective spatial position features in different operating conditions. Therefore, it performs excellently in fault diagnosis under different operating conditions.

B. CONSTRUCTING THE MODEL LOSS FUNCTION

During the training of the model, the weight parameters of the model are updated by a backpropagation algorithm. And back propagation requires a loss function that can calculate the spacing of the model's output from the true value. The loss function is used in the training process, and the parameters of the model's weight values are updated continuously by back propagation. In this paper, the expression of the loss function used is

$$L_c = T_c \max\left(0, m^+ - p_c\right)^2 + \lambda\left(1 - T_c\right) max\left(0, p_c - m^-\right)^2 \quad (8)$$

where p_c is the final output of the model with ten probability values, and the subscript c is the cth digit of the output; T_c is the classification indicator function, assuming that the Kth digit of the output indicates the failure category K, the digit is responsible for predicting the probability of category K. Then when the input sample is of category K and $c = k$ when $T_c = 1$, otherwise $T_c = 0$, m^+ is the upper bound, here taken as a fixed value of 0.9, and when the probability value $p_c > 0.9$ when the loss function is set to 0, and m^- is the lower bound, where it takes a fixed value of 0.1, and the loss function is set to 0 when the probability value $p_c < 0.1$, λ is a proportionality factor and takes a value of 0.5.

IV. EXPERIMENTS AND ANALYSIS OF RESULTS

To verify the applicability and generalization of the methodology proposed in this paper. Therefore, this article selected the Case Western Reserve University Bearing Dataset [26] and the Patburn University Bearing Dataset [27] for simulation.

A. MODEL PARAMETER SETTINGS

Table I shows the model parameters of CCN, where the input time–frequency map size is 32*32. CCN uses two convolutional layers to extract features and one pooling layer for parameter reduction. The pooling layer is set with a small perceptual field, mainly to reduce the parameters without losing too many features. The two convolutional layers perform scale transformation, and feature extraction on the data, and the extracted data are sent to the capsule module for the initial capsule construction. The data of the digital capsule layer are finally output after squeezing, and the final output is ten probability values corresponding to ten types of faults.

B. INTRODUCTION TO THE COMPARISON METHODS

To evaluate the performance of the proposed method, this paper verifies the feature extraction capability of the CCN by comparing it with other deep learning models. The compared models are all traditional CNN, which are introduced as follows:

1. LeNet-5 was proposed by LeCun in 1998 to solve the handwritten digit recognition problem and is considered one of the seminal works in CNN [23]. This network was one of the first neural networks to be widely used for digital image recognition and one of the milestones in deep learning.
2. VGG16 is a classical CNN architecture [28]. VGG was developed to increase the depth of CNNs to improve the model performance.
3. ResNet18 is the model proposed by He *et al.* [29]. The innovation of the residual structure is to increase the depth of the CNN and to make the convergence of the CNN faster. It also allows the CNN to have significantly fewer parameters at deeper layers than previous deep CNN.

C. CASE STUDY 1

1) INTRODUCTION AND DATA ANALYSIS. As shown in Fig. 6, the test bench consists of a motor on the far left, a torque transducer in the middle, a force gauge on the right, and control electronics. The bearing under test is a motor support bearing, type SKF6205 bearing with deep groove balls.

The three operating conditions of the data set are presented in Table II. Data A is the data collected at a bearing speed of 1772 rpm and a load of 1 HP. Data B is collected when the bearing speed is 1750 rpm, and the load is 2HP. Data C is collected when the bearing speed is 1730 rpm, and the load is 3HP.

As shown in Table III, each data set contains nine fault types and one normal status type, for a total of 10 fault types. A damage degree of 0.007 inches indicates mild damage, 0.014 inches is moderate, and 0.021 inches is

Table I. Description of model parameters

Serial number	Module name	Layer name	Core size/step size	Output size
1	Convolution Module I	Convolutional layer 1	(3,3)/1	64*30*30
2	Convolution Module II	Convolutional layer 2	(3,3)/1	256*28*28
		Maximum pooling layer	(2,2)/1	256*14*14
3	Capsule Module	Main capsule layer	\	288*8
		Digital capsule layer	\	10*16
4	Output Module	Output layer	\	10

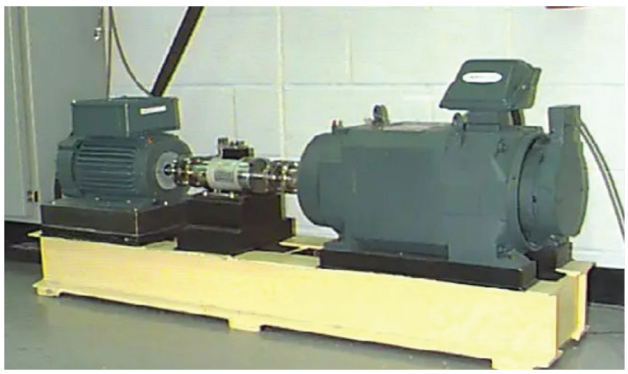

Fig. 6. Case Western Reserve University bearing data set collection system.

Table II. Introduction of various working conditions

Date	Rotational speed(rpm)	Load(HP)
A	1772	1
B	1750	2
C	1730	3

severe. Each sample is labeled using a ten-dimensional one-hot coding vector, where only one of the ten numbers in the vector has a value of 1, and the rest are 0. The location index with a value of 1 indicates a category.

In deep learning, due to the powerful fitting ability of artificial neural networks, too few samples in the training set can lead to overfitting of the model on the training set and a decrease in accuracy on the test set. To avoid overfitting, many samples are usually required as the training set. Therefore, this paper uses overlapping sampling to construct the dataset. The acquisition starts from the beginning of the original vibration signal, and each time 1024 data points are acquired, the acquisition is moved backward by 400 data points and continues. This is done until the complete original data is collected.

Table III. Introduction of the data set

Label	State	Degree of damage(inch)
0	Ball fault	0.007
1	Inner ring fault	0.007
2	Outer ring fault	0.007
3	Ball fault	0.014
4	Inner ring fault	0.014
5	Outer ring fault	0.014
6	Ball fault	0.021
7	Inner ring fault	0.021
8	Outer ring fault	0.021
9	Normal state	\

The original signal vibration waveform is plotted as shown in Fig. 7. After plotting the original vibration waveform, the segmented signal is subjected to a CWT CWT process. Cmor3-3 wavelet is selected as the CWT wavelet basis function. Figure 8 shows its corresponding time–frequency graph after compression processing. The size of the picture is 32*32. After pre-processing, 300 wavelets of each condition are obtained, of which 240 are selected as training samples and the remaining 60 as test samples.

This article develops the proposed model under the Python framework. The experimental equipment used was a desktop computer based on a 64-bit operating system under Windows 11, with a running memory of 16GB, an Intel (R) Core (TM) i5-10400 CPU, and an NVIDIA GeForce RTX3060Ti GPU. During the model's training, the optimization was performed using the Adam optimization algorithm with a batch setting of 64 and a learning rate of 0.001; the number of iterations of the model training was set to 50.

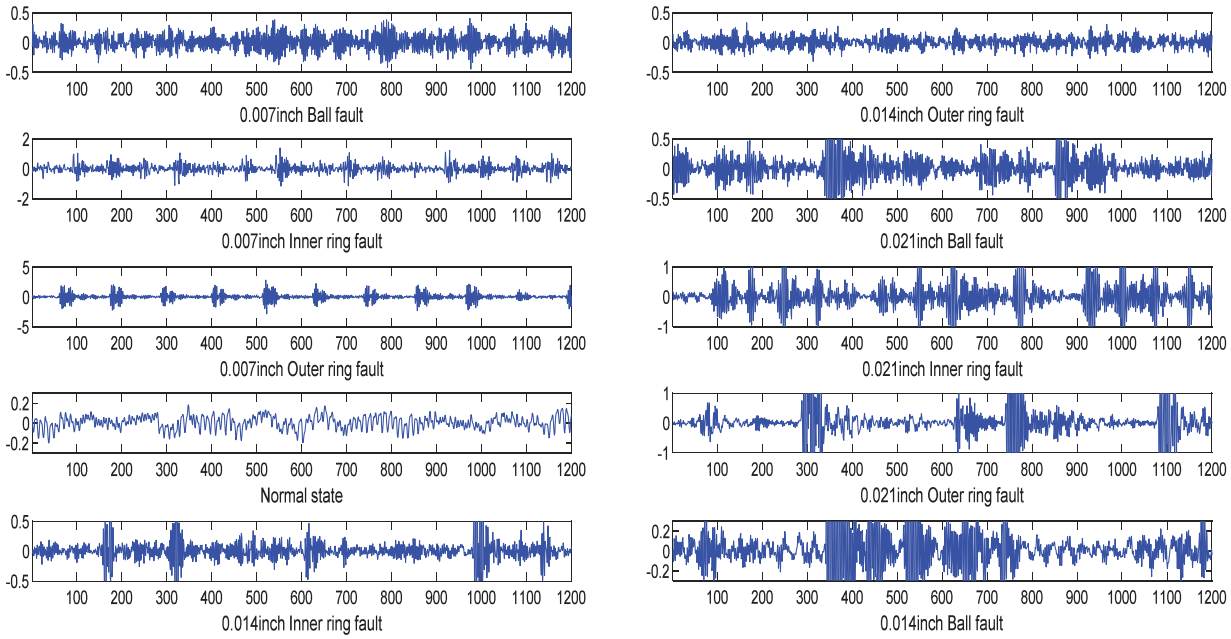

Fig. 7. Original vibration signal waveform.

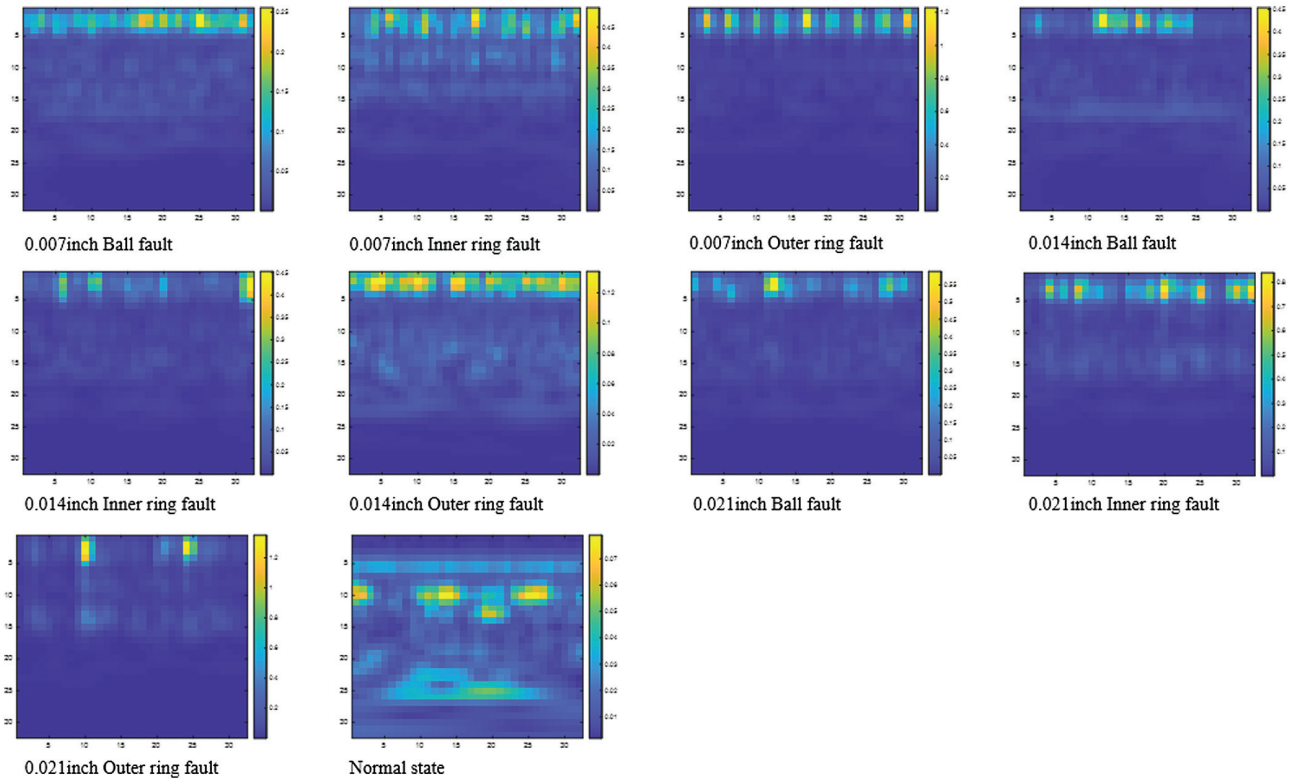

Fig. 8. Time-frequency diagram of the continuous wavelet transform.

2) RESULTS ANALYSIS. This section comprehensively analyzes and discusses the CCN model designed in this paper. Specifically, firstly, we conduct experiments on the model under different working conditions and prove that the model has good generalization and robustness. Secondly, we conduct experiments on the model under different signal-to-noise ratios to verify the noise immunity of the designed model. Finally, we conduct experiments with insufficient training data to verify the powerful feature extraction capability of the CCN designed in this paper.

• (1) FAULT DIAGNOSIS UNDER DIFFERENT WORKING CONDITIONS. In practical application scenarios, rolling bearings typically operate at different speeds and loads. [30–32]. Therefore, it is of great practical engineering importance to evaluate the fault diagnosis capability of the model under different working conditions [33–36]. To verify the fault diagnosis performance of the model under different working conditions, data from three different working conditions are selected for testing. The training and testing sets selected for this experiment are under the same working conditions and divided into three different working conditions for discussion.

The accuracy curves of multiple deep learning models at a load of 1hp (data A) are shown in Fig. 9. From Fig. 9, it can be seen that: in terms of the diagnostic capability of the network, the CCN can reach a smooth convergence state quickly, and the diagnostic accuracy is much higher compared to other models, indicating that the proposed method has more robust feature extraction capability compared to other deep learning network models.

In this paper, the processed time–frequency map is used as input to verify the effectiveness of the proposed method by diagnosing its fault classes. To reduce the influence of random factors and verify the stability of the proposed method, the method and other deep learning network models are repeated five times under three different operating conditions. Meanwhile, to quantitatively compare the diagnostic accuracy of the four diagnostic methods, the diagnostic accuracy of each test and the average diagnostic accuracy are listed as shown in Table IV.

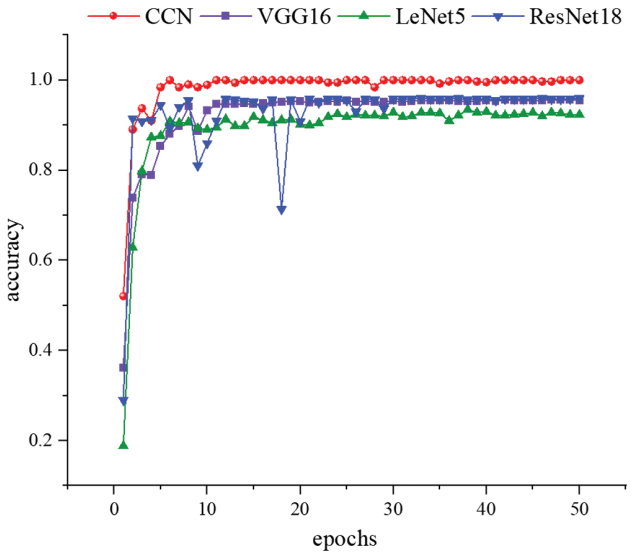

Fig. 9. Model accuracy curve.

As shown in Table IV, the average test accuracy of the CCN was 100% under the three different working conditions. Compared with VGG16, ResNet18, and LeNet-5, it improved by 1.53%, 1.83%, and 7.81%, respectively. The experimental results show that the proposed model has a

Table IV. Recognition accuracy of different algorithms

Methods	Data A Accuracy /%	Data B Accuracy /%	Data C Accuracy /%	Average accuracy rate /%
CCN	100	100	100	100
VGG16	98.00	98.78	98.63	98.47
ResNet18	97.64	98.24	98.64	98.17
LeNet-5	92.13	92.24	92.19	92.19

more stable diagnostic performance and has a good classification effect for multiple types of faults in different working conditions.

To better verify the advantages of the proposed method, the extracted features are reduced to two dimensions using the T-SNE method and plotted. The results of data A are shown in Fig. 10, and the dots of different colors indicate different faults. Among them, the classification effect of the LeNet-5 network is relatively poor, and the classification effect of CNN and ResNet18 is relatively good, but the distance between the features is close. The classification ability and the distance between capsule network features are better than the other models.

Figure 11 shows the confusion matrix obtained by different models using date A. leNet-5, VGG16, and ResNet18 all have incorrect samples, while the CCN has no incorrect samples. This result indicates that the CCN has the best classification performance.

• (2) FAULT DIAGNOSIS UNDER NOISE CONDITIONS. Due to the fact that rolling bearings are usually located in complex environments, they are inevitably affected by noise interference during actual equipment operation. Therefore, it is of great practical engineering importance to evaluate the noise immunity performance of the model in a noisy environment. To verify the fault diagnosis performance of the model in a noisy environment, add additive Gaussian white noise with different signal to noise ratio (SNR) to the test data set. The SNR is an important index to evaluate the amount of noise contained in the signal. Its calculation formula is

$$SNR = 10\log\frac{P_{signal}}{P_{noise}} \tag{9}$$

where P_{signal} is the original vibration signal power, P_{noise} is the noise signal power, and SNR is the signal-to-noise ratio.

In the noise resistance experiments, Gaussian white noise with different signal-to-noise ratios is chosen to be added to the data set A for the experiments. The diagnosis results of different algorithms in different noise environments are shown in Fig. 12.

The diagnostic performance of LeNet-5 in the noisy environment is significantly lower than the other three

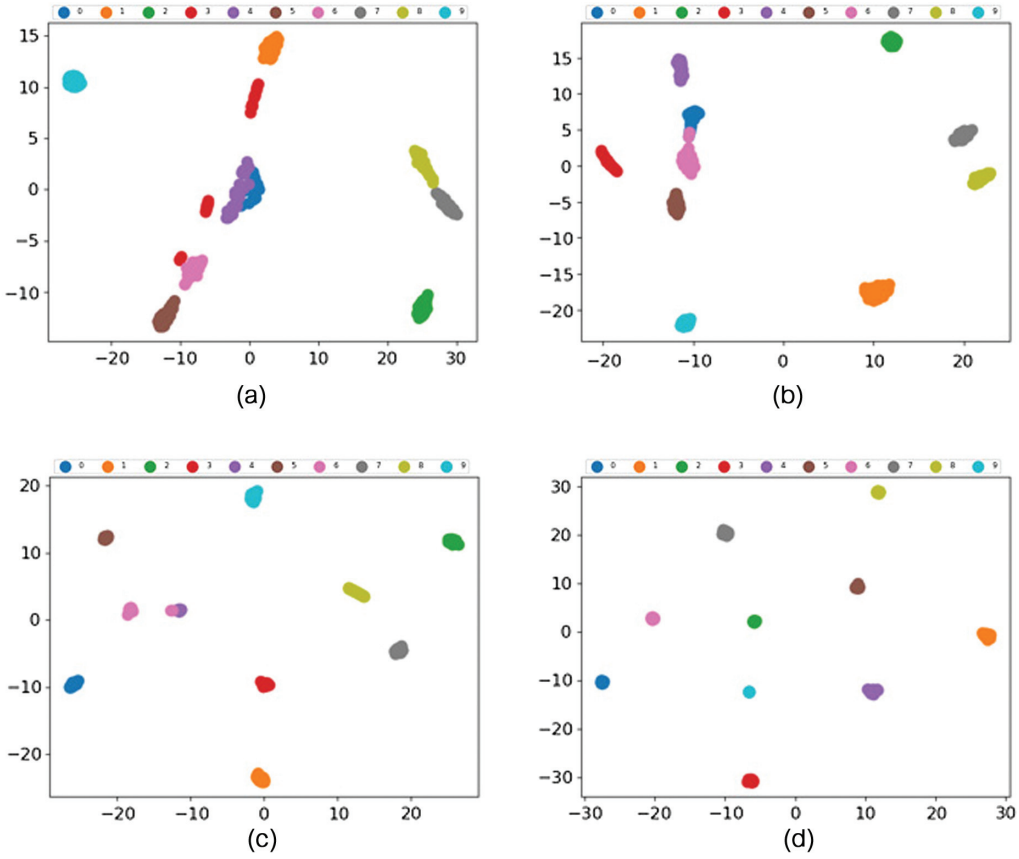

Fig. 10. Visualization results of data A classification features. features of (a) LeNet-5, (b) ResNet18, (c) VGG16, (d) CCN.

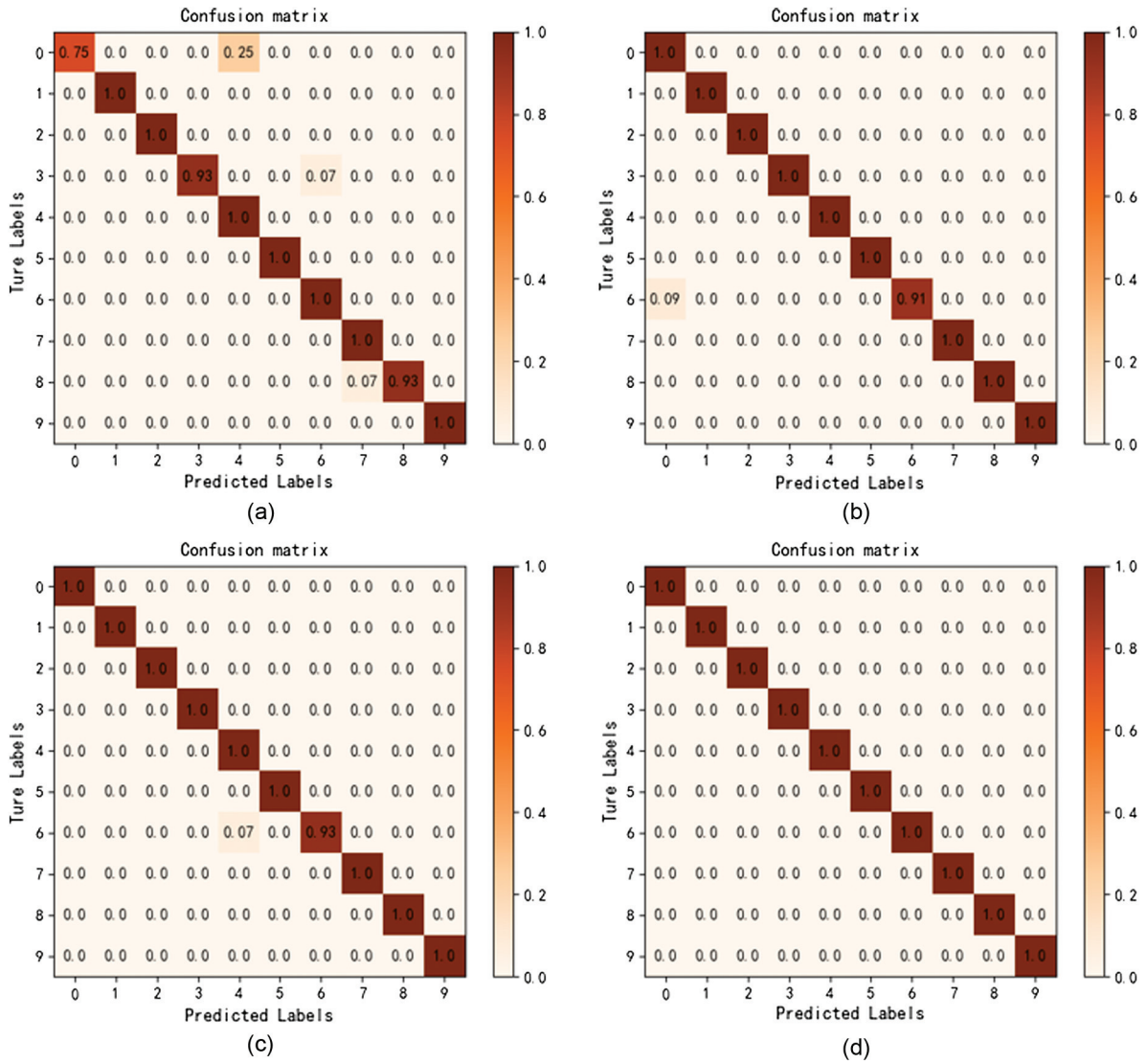

Fig. 11. Confusion matrix plots obtained for different models using test set A (a) LeNet-5, (b) ResNet18, (c) VGG16, (d) CCN.

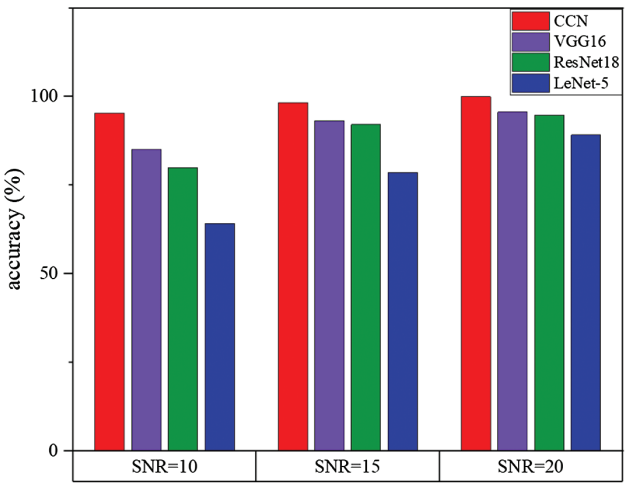

Fig. 12. Comparison of recognition accuracy under different noisy environments.

networks, as shown in Fig. 13. In the environment with SNR = 10, the recognition accuracy of VGG16, ResNet18, and LeNet-5 all start to show a serious decline. At SNR = 10, the accuracy rates of VGG16, ResNet18 and LeNet-5 were only 85.21%, 79.96% and 64.28%, respectively. Compared with the other network models, CCN only replaces the fully connected layer with the capsule layer, but the noise immunity has been improved significantly. This also shows that using vector neurons can extract more detailed information, enabling it to maintain a high recognition rate even in noise-polluted signals.

To better reflect the fault diagnosis performance of the CCN in the noisy environment, Fig. 13 shows the classification results of each fault for different models. As can be seen from the figure, the accuracy of each classification of LeNet-5 in the noisy environment is significantly lower than the other three networks. At SNR = 20, the classification ability of VGG16, ResNet18, and CCN showed relatively stable performance. However, in the SNR = 10 environment, the classification accuracies of VGG16 and ResNet18 all start to show a severe decline, and CCN can still accurately classify seven types of faults. On the other hand, it proves that CCN

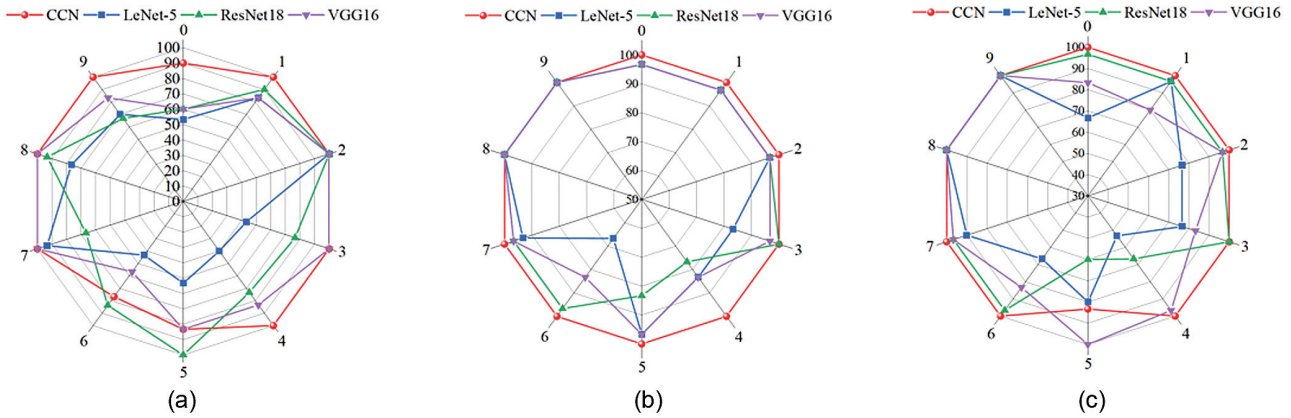

Fig. 13. Classification accuracy per class of faults for each model with different noise: (a) SNR = 10, (b) SNR = 15, (C) SNR = 20.

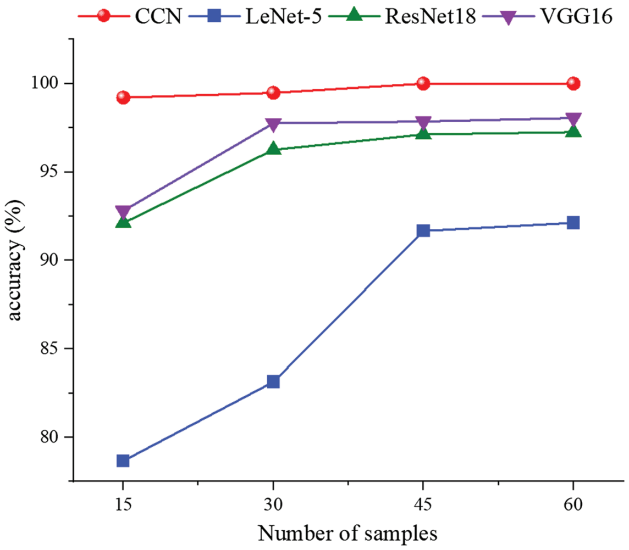

Fig. 14. Fault diagnosis accuracy of each model with different samples.

has a more vital feature extraction ability and better noise immunity than other deep learning models.

• (3) FAULT DIAGNOSIS UNDER INSUFFICIENT SAMPLES. In practical engineering, due to the difficulties in collecting rolling bearing data, we are unable to obtain sufficient fault data. Therefore, verifying the model's performance in fault diagnosis with insufficient samples has important practical significance. Therefore, different numbers of the training set samples in date A were selected for testing. The numbers of training samples in each category in the experiment are 15, 30, 45, and 60, respectively, and the number of samples in each test set is 150.

As shown in Fig. 14, the diagnostic performance of LeNet-5 in the under-sample condition is significantly lower than the other three networks. When the training samples are 30 to 60. the accuracy rates of VGG16, ResNet18 and CCN are relatively stable. However, when the training sample is 15, the fault diagnosis accuracies of VGG16, ResNet18, and LeNet-5 are 92.81%, 92.12%, and 78.67%, respectively. While CCN still maintains 99.21% recognition accuracy, which is 6.4%, 7.09%, and 20.54% higher compared to VGG16, ResNet18, and LeNet-5, respectively. The results show that CCN has stronger feature stunning performance, enabling it to maintain high fault diagnosis ability even under insufficient samples.

D. CASE STUDY 2

1) INTRODUCTION AND DATA ANALYSIS. The data for the Paderborn University bearing data set Alpha were collected from the modular test stand shown in Fig. 15, where (1) is the motor, (2) is the torque measurement shaft, (3) is the rolling bearing test module, (4) is the flywheel, and (5) is the load motor.

The various working conditions of the constructed dataset are shown in Table V. Data D is the data collected when the bearing speed is 1500 rpm, the torque is 0.1 Nm, and the radial load is 1000 N. Data E is the data collected when the bearing speed is 1500 rpm, the torque is 0.7 Nm, and the radial load is 400 N. Data F is the data collected when the bearing speed is 1500 rpm, the torque is 0.7 Nm, and the radial load is 1000 N.

As shown in Table VI, each data set contains five types of faults and one normal state type, for a total of 6 fault types. N represents the normal state, IF represents the inner ring fault, and OF represents the outer ring fault, where a damage degree of 1 indicates mild damage, 2 is moderate

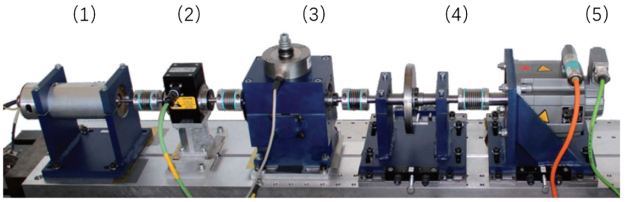

Fig. 15. Experiment rig of the Paderborn dataset.

Table V. Introduction of various working conditions

Date	Rotational speed (rpm)	Torque (N.m)	Radial load (N)
D	1500	0.1	1000
E	1500	0.7	400
F	1500	0.7	1000

Table VI. Introduction of the data set

Label	State	Bearing	Degree of damage
0	N	K004	\
1	IF	K121	1
2	IF	K118	2
3	IF	K116	3
4	OF	KA04	1
5	OF	KA16	2

damage, and 3 is severe damage. Each sample is labeled using a six-dimensional one-hot coding vector, where only one of the six numbers in the vector has a value of 1, and the rest are 0. The location index with a value of 1 indicates a category.

The overlapping sampling method is used to construct the data set, starting from the beginning of the original vibration signal, acquiring 1024 data points each time, and then moving backward 500 data points to continue acquiring after the acquisition is completed until 300 samples of the original data are collected. The rest of the experimental parameters are consistent with the data set of Case 1.

2) RESULTS ANALYSIS. To verify the generalization and robustness of the model, we also analyzed the fault diagnosis capability of the method under different operating conditions, noisy conditions, and insufficient samples in Case 2 and compared it with other deep learning models.

• (1) FAULT DIAGNOSIS UNDER DIFFERENT WORKING CONDITIONS. The accuracy of various models for the fault diagnosis experiments under different operating conditions is shown in Table VII.

The accuracy of various models under data set D is shown in Fig. 16.

To better verify the advantages of the model, we used data D for testing, and the extracted features were reduced to two dimensions and visualized using the T-SNE method, and the visualization results are shown in Fig. 17.

Figure 18 shows the confusion matrix obtained for the different models using the test set D.

• (2) FAULT DIAGNOSIS UNDER NOISE CONDITIONS. Since Case 2 has six fault types, the difficulty of fault diagnosis is lower compared to Case 1, with ten fault types, so a smaller signal-to-noise ratio is chosen. The accuracy of various models under different noise conditions is shown in Fig. 19.

• (3) FAULT DIAGNOSIS UNDER INSUFFICIENT SAMPLES. The accuracy of various models under different samples is shown in Fig. 20.

Analysis of experimental results: It can be observed from the above experimental results that the conclusions we got in case 1 have been proved successful. Firstly, in experiments under different working conditions, the accuracy rate of the method designed by us reaches 100%, which proves that the method has good generalization and robustness. Secondly, in the experiments under the condition of noise and the condition of insufficient training samples, the method designed in this paper performs better than other traditional deep learning models. It is again proved that the method designed in this paper has a specific antinoise ability and strong feature extraction ability in the case of insufficient samples

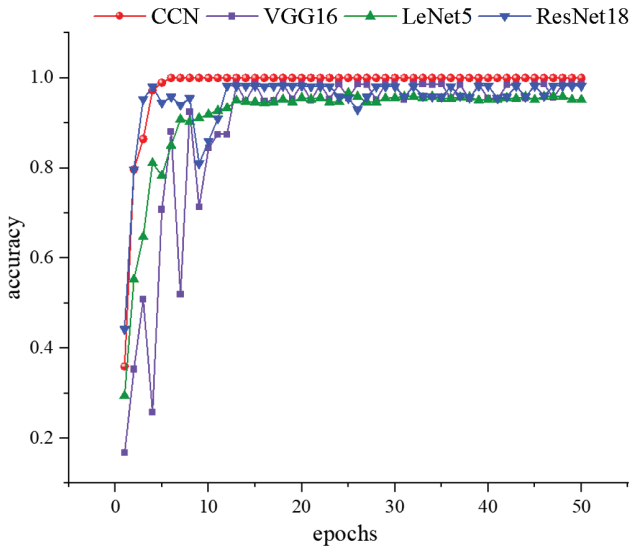

Fig. 16. Model accuracy curve.

E. IMPACT OF ITERATION TIMES ON DYNAMIC ROUTING ALGORITHMS

The dynamic routing algorithm is the core algorithm of capsule networks, used to calculate the similarity weight coefficients of capsules and update similarity. The dynamic routing algorithm is equivalent to doing a fully connected mapping, where each path requires a fully connected mapping of all dimensions of the upper and lower capsules, resulting in a vast number of parameters. Too many iterations of dynamic routing algorithms can lead to excessive training parameters, while too few iterations can lead to incomplete mapping and insufficient diagnostic ability. Therefore, evaluating the number of iterations of dynamic routing algorithms for stopping CCN is of great significance. This section selects data with SNR = 10 in date A for validation, and the results are shown in Table VIII.

The first column in Table VIII shows the number of iterations of the dynamic routing algorithm, the second

Table VII. Recognition accuracy of different algorithms

Methods	Data D Accuracy /%	Data E Accuracy /%	Data F Accuracy /%	Average accuracy rate /%
CCN	100	100	100	100
VGG16	98.61	99.13	98.72	98.82
ResNet18	98.24	98.57	98.26	98.36
LeNet-5	95.17	95.69	95.12	95.32

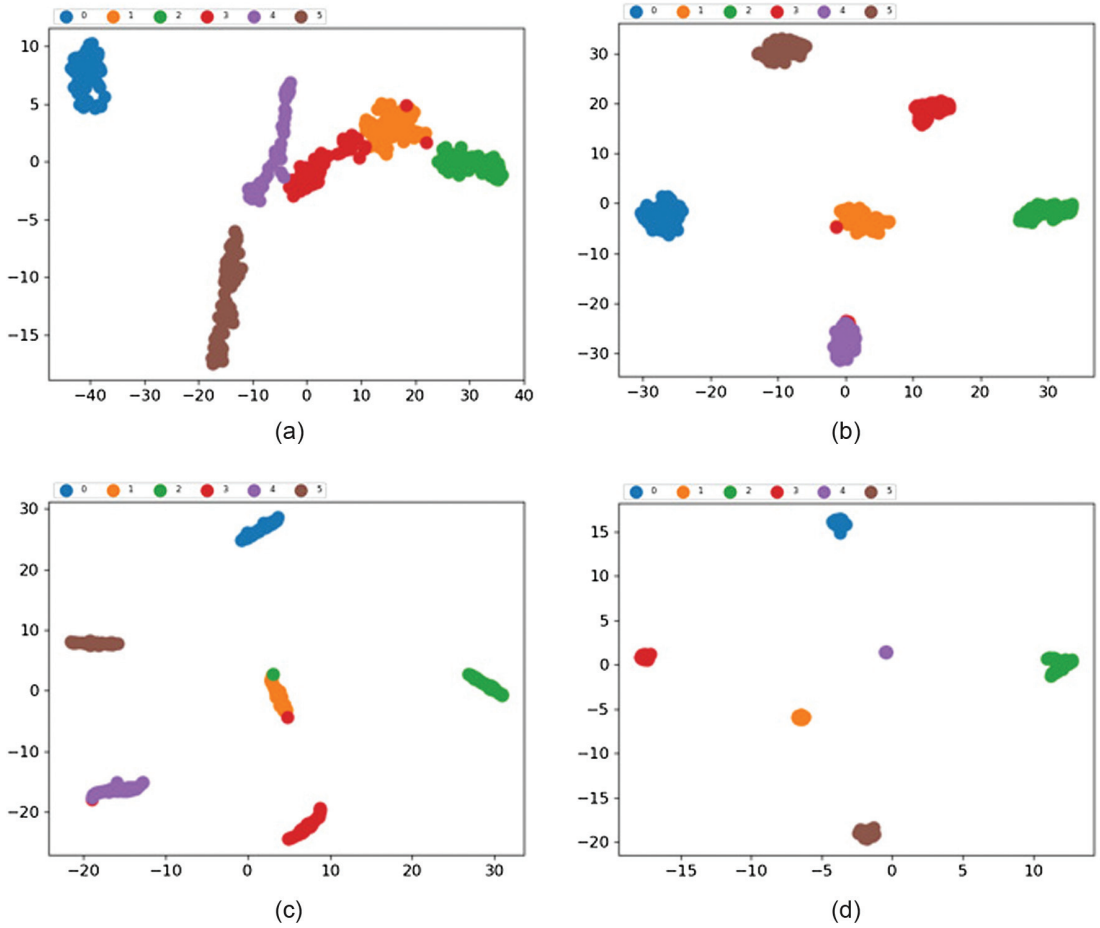

Fig. 17. Visualization results of data D classification features. features of (a) LeNet-5, (b) ResNet18, (c) VGG16, (d) CCN.

column shows the accuracy of diagnosis, and the third column shows the execution time of each training round. The table shows that the accuracy is highest when the number of iterations is 3, and the execution time is also in an intermediate position. Therefore, the dynamic routing algorithm selected in this article has an iteration number of 3.

V. CONCLUSION

This article designs a rolling bearing fault diagnosis method based on CCN, and the conclusions are summarized as follows:

(1) Compared to other traditional machine learning models, the method proposed in this paper can adaptively extract fault features, avoiding the drawbacks of manual feature extraction and relying on expert experience, and providing a foundation for implementing intelligent fault diagnosis.

(2) This article transforms the original vibration signal into a time–frequency domain signal after CWT processing as input to the network model. The network can learn features more fully to fully express the amplitude characteristics and frequency components of nonstationary signals.

(3) The CCN front-end combines a convolutional layer and pooling layer, which can reduce training parameters and extract fault characteristics at a deeper level. The backend of CCN uses a set of capsule structures to vectorize and mine the spatial information of features. In the process of feature transmission in CCN, neurons are vector neurons that can carry spatial position features. CCN can extract more effective spatial position features in different operating conditions. Therefore, it performs excellently in fault diagnosis under different operating conditions.

(4) To address the difficulties in collecting fault data in engineering practice, such as the lack of training samples and the presence of noise, experiments were conducted in small samples and noisy environments. The experimental results show that the accuracy of the proposed method in this paper still maintains a high level even in the presence of insufficient training samples and noise.

(5) To verify the applicability and generalization of the method proposed in this article, data from six operating conditions from two datasets were used for validation. The results indicate that the fault diagnosis ability is stable in various working conditions, and the diagnostic accuracy is significantly higher than other comparison methods.

Although capsule networks have good diagnostic performance under noisy conditions and insufficient samples,

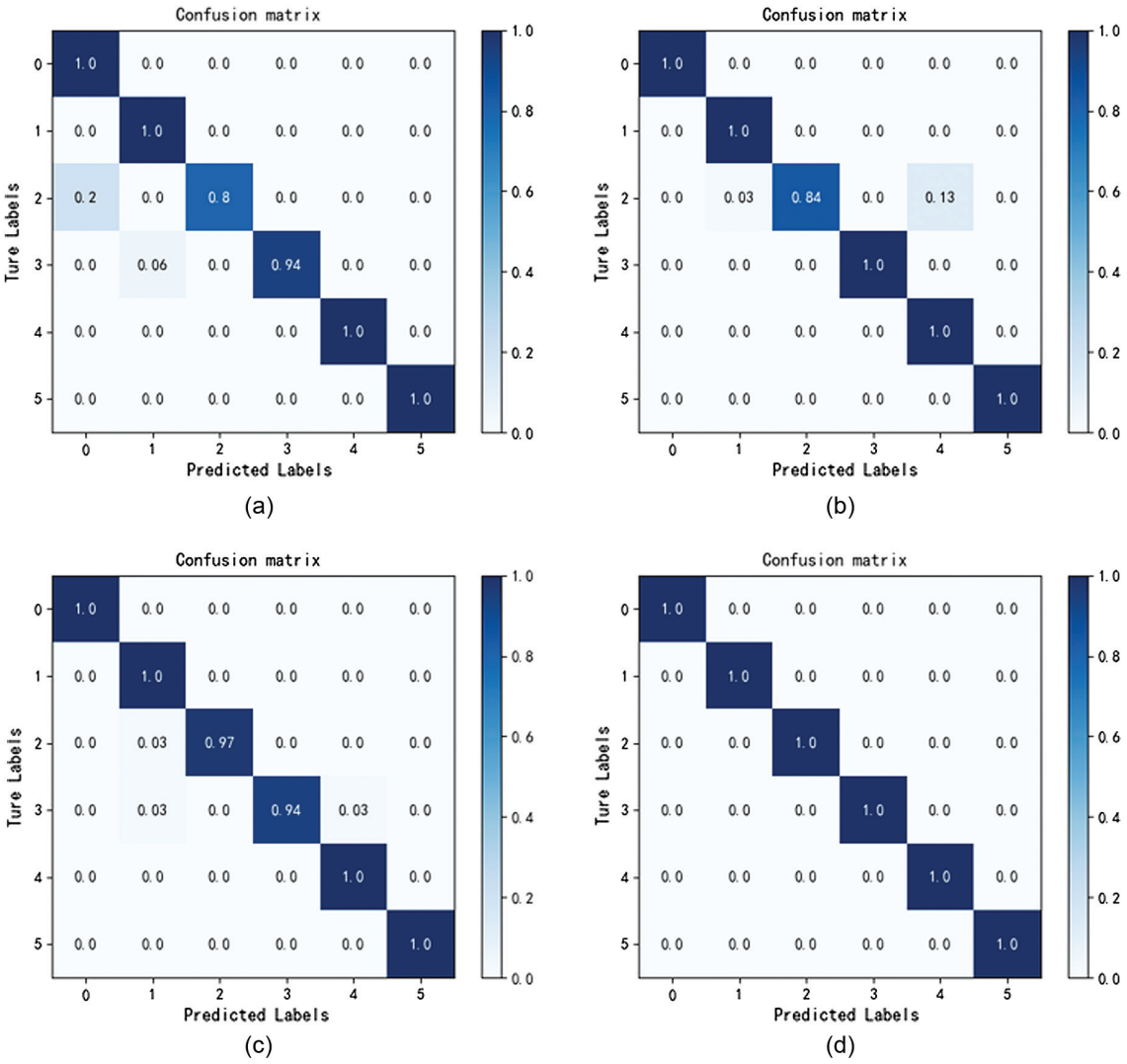

Fig. 18. Confusion matrix plots obtained for different models using test set D (a) LeNet-5, (b) ResNet18, (c) VGG16, (d) CCN.

they still have the problem of difficult training due to large training parameters. In future research and learning processes, the capsule network model will be improved to address the drawback of high equipment requirements due to the large parameters of the capsule network model.

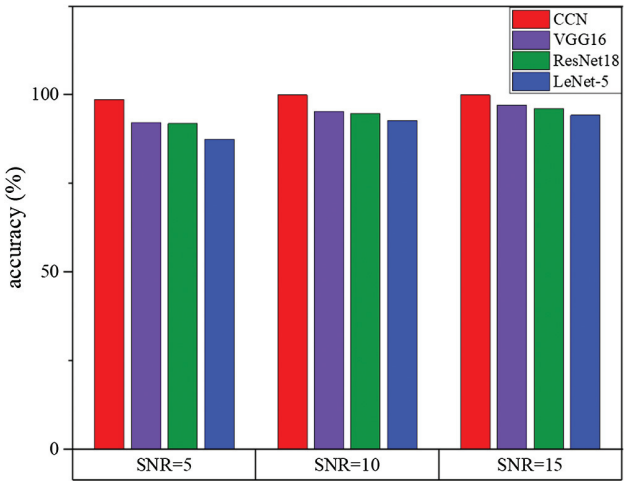

Fig. 19. Comparison of recognition accuracy under different noisy environments.

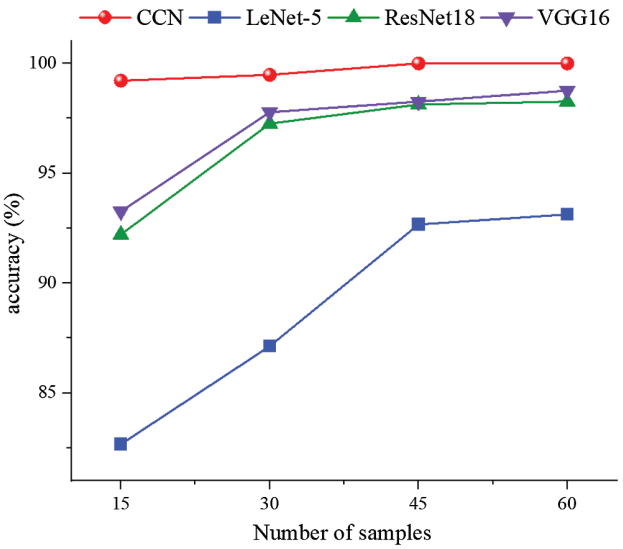

Fig. 20. Fault diagnosis accuracy of each model with different samples.

Table VIII. Impact of iteration times of dynamic routing algorithms

Number of times	Accuracy (%)	Execution time
1	88.54	2.26s
2	92.18	4.17s
3	95.31	6.09s
4	93.75	8.01s
5	92.18	9.96s

Acknowledgments

This work was partially supported by the Science and Technology Planning Project of Inner Mongolia of China under contract number 2021GG0346.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

References

[1] Q. Wang et al., "A domain adaptation method for bearing fault diagnosis using multiple incomplete source data," *J. Intell. Manuf.*, pp. 1–15, 2023.

[2] X. Lu, J Ma, and C. Ying, "Small-sample automotive planetary gearbox fault diagnosis based on continuous wavelet transform and model-free element learning," *Mech. Trans.*, vol. 46, no. 09, pp. 159–164, 2022.

[3] S. K. Gundewar and P V. Kane, "Bearing fault diagnosis using time segmented Fourier synchrosqueezed transform images and convolution neural network," *Measurement*, vol. 203, p. 111855, 2022.

[4] T. Tian and K. Ding, "Hilbert transform and its application in fault diagnosis," *Vib. Shock*, vol. 15, no. 02, pp. 24–27, 1996.

[5] Y. Sun and G. L. Peng, "Improved capsule network for rolling bearing fault diagnosis method," *J. Harbin Inst. Technol.*, vol. 53, no. 01, pp. 23–28, 2021.

[6] K. Feng et al., "Digital twin-driven intelligent assessment of gear surface degradation," *Mech. Syst. Signal Process.*, vol. 186, p. 109896, 2023.

[7] F. Sağlam, E. Yıldırım, and M. A. Cengiz, "Clustered Bayesian classification for within-class separation," *Exp. Syst. Appl.*, vol. 208, p. 118152, 2022.

[8] J. Xu, X. Zhang, Y. Li, "New developments in support vector machines," *Control Decis. Making*, vol. 19, no. 5, pp. 481–484, 2004.

[9] K. Feng et al., "Use of cyclostationary properties of vibration signals to identify gear wear mechanisms and track wear evolution," *Mech. Syst. Signal Process.*, vol. 150, p. 107258, 2021.

[10] K. Xu et al., "A bearing fault diagnosis method without fault data in new working condition combined dynamic model with deep learning," *Adv. Eng. Inf.*, vol. 54, p. 101795, 2022.

[11] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, 2018.

[12] Y. Liu et al., "Fault diagnosis of rolling bearings based on attention module and 1D-CNN," *J. Solar Energy*, vol. 43, no. 03, pp. 462–468, 2022.

[13] Huang C and J. Zhang, "A CBAM-CNN based fault diagnosis method for rolling bearings," *Modern Manuf. Eng.*, no. 11, pp. 137–143, 2022.

[14] W. Guo and X. Xing, "An intelligent diagnosis method for few-sample bearings based on improved convolutional generative adversarial networks," *China Mech. Eng.*, vol. 33, no. 19, pp. 2347–2355, 2022.

[15] W. Xie et al., "Evaluation of different bearing fault classifiers in utilizing CNN feature extraction ability," *Sensors*, vol. 22, no. 9, p. 3314, 2022.

[16] H. Pan et al., "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," *Strojniski Vestnik/ J. Mech. Eng.*, vol. 64, pp. 443–452, 2018.

[17] D. Wang et al., "Application of multiscale learning neural network based on CNN in bearing fault diagnosis[J]," *J. Signal Process. Syst.*, vol. 91, pp. 1205–1217, 2019.

[18] C. Liu et al., "Planetary gears feature extraction and fault diagnosis method based on VMD and CNN," *Sensors*, vol. 18, no. 5, p. 1523, 2018.

[19] L. Gou et al., "Aeroengine control system sensor fault diagnosis based on CWT and CNN," *Math. Prob. Eng.*, vol. 2020, 2020.

[20] X. Zhang et al., "Multi-model ensemble deep learning method for intelligent fault diagnosis with high-dimensional samples," *Front. Mech. Eng.*, vol. 16, no. 2, pp. 340–352, 2021.

[21] Z. Xing et al., "Intelligent fault diagnosis of rolling bearing based on novel CNN model considering data imbalance," *Appl. Intell.*, vol. 52, pp. 1–13, 2022.

[22] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules.2017," *arXiv preprint arXiv:1710. 09829*, 2017.

[23] Y. Lecun and L. Bottou, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[24] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Process.*, vol. 96, pp. 1–15, 2014.

[25] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Process. Mag.*, vol. 8, no. 4, pp. 14–38, 1991.

[26] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case Western Reserve University data: a benchmark study," *Mech. Syst. Signal Process.*, vol. 64, pp. 100–131, 2015.

[27] C. Lessmeier et al., *KAt-DataCenter, Chair of Design and Drive Technology*. Paderborn, Germany: Paderborn University, 2019.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[29] K. He et al., "Deep residual learning for image recognition," *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, pp. 770–778, 2016.

[30] K. Feng et al., "A novel order spectrum-based Vold-Kalman filter bandwidth selection scheme for fault diagnosis of gearbox in offshore wind turbines," *Ocean Eng.*, vol. 15, no. 266, p. 112920, 2022.

[31] K. Feng et al., "A novel similarity-based status characterization methodology for gear surface wear propagation monitoring," *Tribol. Int.*, vol. 174, p. 107765, 2022.

[32] Y. Xu et al., "CFCNN: a novel convolutional fusion framework for collaborative fault identification of rotating machinery," *Inf. Fusion*, vol. 95, pp. 1–6, 2023.

[33] Q. Ni et al., "A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis," *Mech. Syst. Signal Process.*, vol. 164, p. 108216, 2022.

[34] K. Feng et al., "A review of vibration-based gear wear monitoring and prediction techniques," *Mech. Syst. Signal Process.*, vol. 182, p. 109605, 2023.

[35] K Feng et al., "A novel vibration indicator to monitor gear natural fatigue pitting propagation," *Struct. Health Monit.*, p. 14759217221142622, 2023.

[36] Y. Zhang et al., "Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing," *Reliab. Eng. Syst. Saf.*, vol. 234, p. 109186, 2023.