Hindawi
Journal of Sensors
Volume 2021, Article ID 2251530, 17 pages
https://doi.org/10.1155/2021/2251530

Research Article

Multisource Data Fusion Diagnosis Method of Rolling Bearings Based on Improved Multiscale CNN

Yulin Jin, Changzheng Chen, and Siyu Zhao

School of Mechanical Engineering, Shenyang University of Technology, Shenyang 110000, China

Correspondence should be addressed to Changzheng Chen; czchen@sut.edu.cn

Received 24 August 2021; Accepted 15 October 2021; Published 9 November 2021

Academic Editor: Ying-Ren Chien

Intelligent diagnosis applies deep learning algorithms to mechanical fault diagnosis, which can classify the fault forms of machines or parts efficiently. At present, the intelligent diagnosis of rolling bearings mostly adopts a single-sensor signal, and multisensor information can provide more comprehensive fault features for the deep learning model to improve the generalization ability. In order to apply multisensor information more effectively, this paper proposes a multiscale convolutional neural network model based on global average pooling. The diagnostic model introduces a multiscale convolution kernel in the feature extraction process, which improves the robustness of the model. Meanwhile, its parallel structure also makes up for the shortcomings of the multichannel input fusion method. In the multiscale fusion process, the global average pooling method is used to replace the way to reshape the feature maps into a one-dimensional feature vector in the traditional convolutional neural network, which effectively retains the spatial structure of the feature maps. The model proposed in this paper has been verified by the bearing fault data collected by the experimental platform. The experimental results show that the algorithm proposed in this paper can fuse multisensor data effectively. Compared with other data fusion algorithms, the multiscale convolutional neural network model based on global average pooling has shorter training epochs and better fault diagnosis results.

1. Introduction

A rolling bearing is an indispensable part of industry. In the application process, in order to improve the safety and reliability of machine operation, the state of rolling bearings must be evaluated regularly [1–4]. The traditional fault diagnosis method is to process and analyze the fault signal in the time and frequency domains by the signal processing method and rely on certain expert experience to achieve the purpose of fault diagnosis [5–7]. However, in the actual diagnosis process, this fault diagnosis method is more complicated. In recent years, with the development of machine learning, some machine learning algorithms, such as ANN [8], SVM [9–11], and KNN [12, 13], have been applied to fault diagnosis. These machine learning algorithms have been proved to solve classification problems. However, these algorithms need to extract input features manually. In the mechanical fault signal, especially in the early stage of faults or in the presence of noise interference, it is difficult to extract sensitive fault features effectively.

In recent years, deep learning based on the deep neural network has developed rapidly and has been successfully applied to computer vision [14, 15], natural language processing [16], speech processing [17], sleep-arousal detection [18], and other fields. Deep learning introduces the convolution kernel with variable weight into the deep network to extract features adaptively, which simplifies the complex work of feature extraction and avoids the inaccuracy of manual feature extraction. The advantage makes deep learning algorithms such as convolutional neural networks (CNN) [19] and deep belief networks (DBN) [20] widely used in the field of fault diagnosis. Fault diagnosis is expected to become more intelligent and more suitable for modern industrial trends. At present, signal preprocessing methods based on CNN can be roughly divided into fault diagnosis methods based on one-dimensional (1D) CNN and fault diagnosis methods based on two-dimensional (2D) CNN. The fault diagnosis method based on 1D CNN is to take the collected 1D time-domain signal as the input of the network directly. Ince et al. used a one-dimensional

convolutional neural network model with adaptive feature extraction for fault diagnosis [21]. Liu et al. constructed a multitask one-dimensional convolutional neural network to solve the problem of fault diagnosis [22]. Liu et al. combined a denoising convolutional autoencoder with a one-dimensional convolutional neural network for fault diagnosis in the noisy environment [23]. The fault diagnosis method based on 2D CNN is to transform the one-dimensional time-domain signal into a two-dimensional form and inputs it into the network. Transforming the one-dimensional time-domain signals into two-dimensional time-frequency diagrams is a common method [24–26]. Meanwhile, some researchers have also used ingenious ways to transform the time-domain signal in dimensionality [27–29]. In terms of model improvement, researchers have also adopted a variety of ways to improve the diagnostic ability of the model. Wang et al. improved the calculation speed and stability of the network by adding a batch normalization layer [30]. Jiang et al. constructed a 1D multiscale model for fault diagnosis [31]. Wang et al. constructed a joint attention module (JAM) and added it to the diagnosis model, which effectively improved the diagnosis performance of the model [32].

The above fault diagnosis problems are dealt with by the fault signal measured by a single sensor. In practice, unilateral information is often one-sided, and the information of a single sensor may not fully reflect the fault features due to the installation location, installation direction, and other factors. Using more comprehensive multisensor signals can further improve the diagnostic capability of the deep learning model. At present, researchers have adopted a variety of methods to diagnose faults with multisource data. Wang et al. constructed time-domain multisensor signals at different locations into a two-dimensional matrix and used improved two-dimensional CNN for diagnosis and classification [33]. Xu et al. proposed an integration model based on multisensor information fusion [34]. Shan et al. fused the extracted feature factors of multiple sensor signals and used CNN for fault classification [35]. Wang et al. converted multisensor data into color images and constructed an improved CNN based on LeNet-5 for classification [36]. Xia et al. input multisensor data into the CNN model through multiple channels to solve the composite diagnosis problem [37]. Gu et al. proposed a multisensor fault diagnosis model based on discrete wavelet transform (DWT) and long short-term memory (LSTM) [38]. Kou et al. fused the monitoring information of multisensors after normalization and classified it by CNN [39]. Li et al. proposed a fault diagnosis method based on multiscale permutation entropy (MPE) and multichannel fusion convolutional neural network (MCFCNN) [40]. Peng et al. extracted features from multisensor information through short-time Fourier transform (STFT) and input the fused features into the depth residual neural network to solve the problem of fault diagnosis [41].

The above method combines multisensor information to provide more comprehensive signal features for the deep learning model to improve the representation capability.

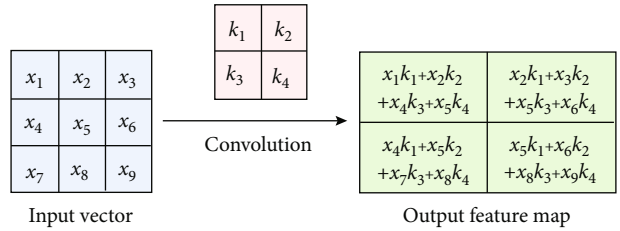


Figure 1: Schematic diagram of the convolution process.

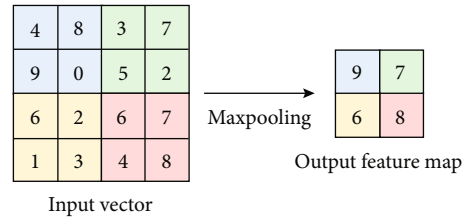


Figure 2: Schematic diagram of the maximum pooling process.

However, in the actual signal acquisition process, there is often noise interference, which will reduce the significance of fault features to a certain extent, weaken the feature differences between various fault categories, and make the training process of the deep learning model more difficult. When a certain sensor data is greatly affected by noise, adding it to the training sample will reduce the quality of the input sample. In this case, only the multisensor data is fused, but the model is not improved in a targeted manner, which will reduce the performance of the overall model.

In order to solve the problem mentioned above, this paper proposes a multiscale CNN model based on global average pooling (MSCNN-GAP). First, a data preprocessing method without expert experience is used to transform the one-dimensional time-domain signal into a two-dimensional form. Then, the processed data is input into the model. In this model, multiscale convolution kernels are introduced to extract sample features in diversity. Multifeature extraction makes the deep learning model more suitable for fault diagnosis with noise. The parallel feature extraction method in the multiscale structure is also better adapted to data fusion. In the fusion part of the network, in order to maintain the feature space of each branch channel, the global average pooling (GAP) layer is used to replace the traditional fully connected layer in the dimensionality reduction method, which further improves the performance of the network. Finally, the effectiveness of MSCNN-GAP is verified by data comparison.

The innovations and main contributions of this article are described as follows:

(1) For the multisensor fusion model, the influence of noise factors that may exist in the process of signal acquisition on the performance of the model is further considered

(2) In order to improve the adaptability of the model to noise, an MSCNN-GAP model with a parallel structure is established

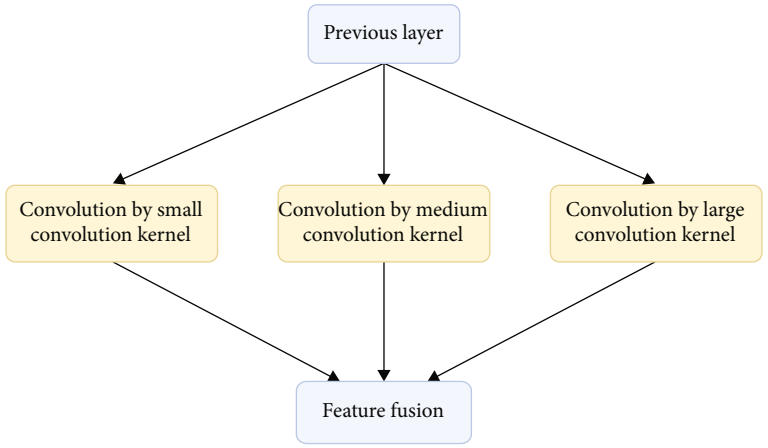


Figure 3: Schematic diagram of the multiscale CNN structure.

(3) In the feature fusion process, the GAP layer is used. The GAP layer reduces the number of network parameters and preserves the spatial structure of each feature during feature fusion to further improve the performance of the model

(4) A comprehensive experiment is designed and executed to verify the effectiveness and robustness of the proposed MSCNN-GAP comprehensively

This article is organized as follows. Section 2 introduces the basic theory of multiscale CNN and global average pooling. Section 3 introduces the diagnosis method of MSCNN-GAP. In Section 4, the experimental verification is carried out, and the experimental results are discussed. Finally, the conclusion is given in Section 5.

2. Basic Theory

2.1. Multiscale Convolutional Neural Network. CNN is a common network form in deep learning. It was first proposed by Lecun et al. and applied to classify handwritten digits [42, 43]. The basic CNN consists of three parts: convolution, activation, and pooling [44]. Convolution is the most basic and essential operation in CNN. The convolution operation adopts a fixed-size convolution kernel to slide on its input to extract the input features. The schematic diagram of the convolution operation is shown in Figure 1. The output result of the n-th feature map of the m-th layer can be expressed as

$$X_n^m = \sum_{k=1}^{K} x_k^{m-1} \cdot \omega_k + b_n^m, \tag{1}$$

where x_k^{m-1} represents the k-th output feature map of the previous layer, ω_k represents the weight of the convolution kernel corresponding to the k-th feature map, b_n^m represents the bias term, and K represents the total number of channels in the previous layer.

After the convolution operation, the activation function will process the convolution output nonlinearly. Adding a nonlinear activation layer can make the model nonlinear so that the network can be used to solve more complex nonlinear problems. The common nonlinear activation functions include the sigmoid function, tanh function, ReLU function [45], and PReLU function [46]. The output result after nonlinear activation can be expressed as

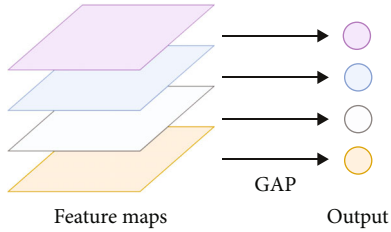


Figure 4: Schematic diagram of the GAP process.

$$z = h(X_n^m), \tag{2}$$

where $h(\cdot)$ represents the nonlinear activation function, X_n^m represents the output of the previous convolution layer, and z represents the output term.

Pooling is a way of downsampling. Adding a pooling layer after the activation function can reduce the number of parameters and extract the key features. The schematic diagram of the maximum pooling process is shown in Figure 2. Meanwhile, the pooling operation can also be used as a nonlinear operation to improve the representation capability of the model.

With the development of deep learning, CNN is used to deal with more complex problems. In the face of these problems, stacking deeper neural networks is a common solution. However, due to the excessive number of parameters of the deep neural network, the problem of gradient instability is prone to occur in the training process, which makes training more difficult. To solve this problem, the Google team first proposed a multiscale structure [47]. The multiscale structure uses multiple parallel CNN structures to replace the deep network and finally merges the feature extraction results of each branch. Multiscale CNN can also use diverse

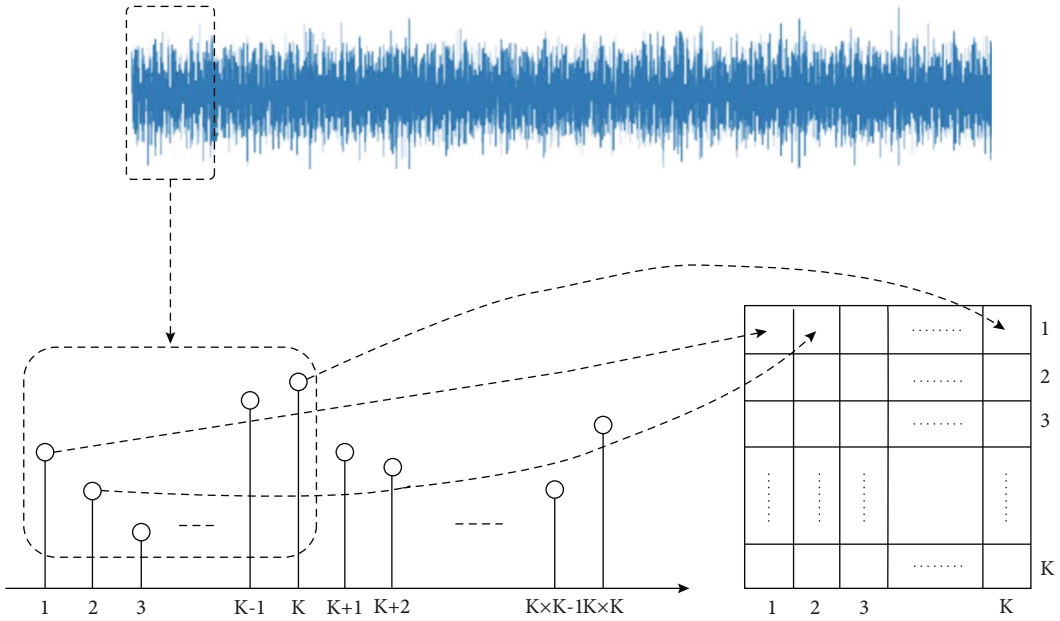


Figure 5: Data conversion process.

Table 1: Network parameters.

Layer	Input	Output	Batch normalization
C1	1 × 64 × 64	16 × 64 × 64	Yes
P1	16 × 64 × 64	16 × 32 × 32	—
C2	16 × 32 × 32	32 × 32 × 32	Yes
P2	32 × 32 × 32	32 × 16 × 16	—
C3	32 × 16 × 16	64 × 16 × 16	Yes
P3	64 × 16 × 16	64 × 8 × 8	—
C4	64 × 8 × 8	128 × 8 × 8	Yes
P4	128 × 8 × 8	128 × 4 × 4	—
GAP	128 × 4 × 4	128 × 1 × 1	—

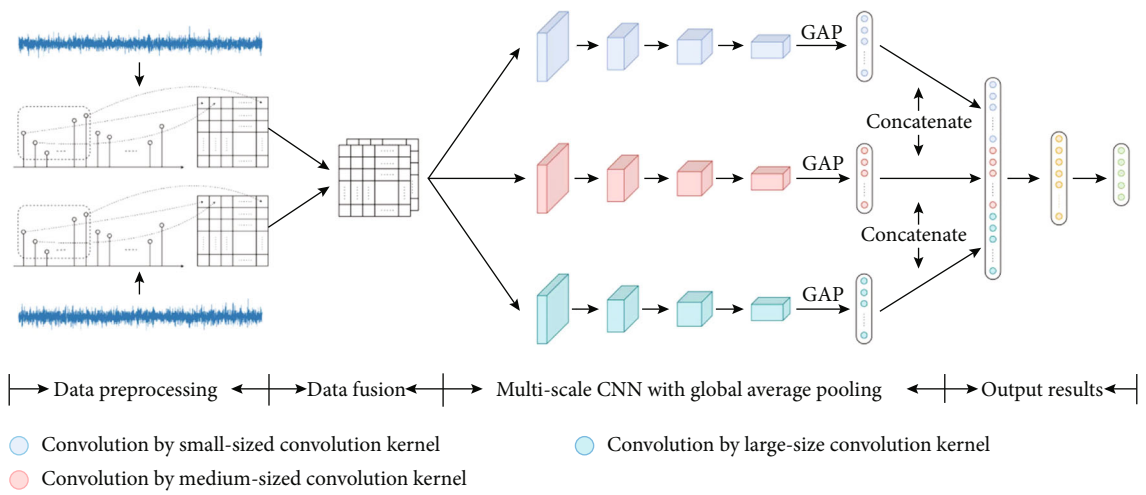


Figure 6: Fault diagnosis method based on MSCNN-GAP.

convolution kernels to extract input features at the same time, which further improves the performance of CNN. The schematic diagram of the multiscale structure is shown in Figure 3. For the multiscale structure of three branches, the fusion process can be expressed as

$$x = \text{Concatenate}(\sigma_1, \sigma_2, \sigma_3), \tag{3}$$

where σ_1, σ_2, and σ_3 represent the feature vectors extracted from each scale and Concatenate $(\cdot)$ represents the fusion process of one-dimensional vectors.

2.2. Global Average Pooling. Normally, CNN will reshape the feature maps into a set of one-dimensional feature vectors after the feature extraction and output the final classification results by using several fully connected layers. In the case of a large number of channels, there will be more parameters in the fully connected layer, and the model will be more complicated. In the reshaping process, the fully connected layer will also lose the spatial position information of the feature maps between the channels. GAP replaces the feature maps of each channel with a feature number. The GAP layer can reduce the amount of data in the model significantly, lower the risk of overfitting in the fully connected layer, and retain the spatial position information of each channel effectively. The global average pooling with c channels can be expressed as

$$x_c = \text{Avgpool}(y_c) = \frac{1}{P \times Q} \sum_{i=1}^{P} \sum_{j=1}^{Q} y_c(i, j), \tag{4}$$

where x is the output of GAP, y is the input of GAP, and P and Q are the width and height of the channel feature maps. The schematic diagram of GAP is shown in Figure 4.

3. Intelligent Diagnosis Method of MSCNN-GAP

3.1. Data Preprocessing. Data preprocessing is an essential part of the fault diagnosis process. An excellent preprocessing method can show the fault features better to improve the performance of the deep learning model. The most common preprocessing method is to transform time-domain signals into time-frequency graphs. However, this processing method is more complicated and requires a certain amount of expert experience. In this paper, one-dimensional time-domain data is transformed into two-dimensional matrices by dimension transformation [25]. The conversion diagram is shown in Figure 5. For a one-dimensional signal, $K \times K$ sampling points are taken as a sample, and each K sample point are taken as a row and arranged in columns in turn to form a two-dimensional matrix of $K \times K$. This preprocessing method does not require expert experience. Furthermore, it reduces the complexity of preprocessing.

3.2. Fault Diagnosis Method Based on MSCNN-GAP. In order to reflect the fault features more comprehensively, this paper uses multisensor data for fault diagnosis. The multisensor data will be input to the network model as multichannels. When the signal sample collection process is disturbed by noise factors, the fault features will be covered by the noise information, which will greatly impact the feature extraction process and reduce the diagnostic capability of the model. In order to weaken this influence and improve the robustness of the model, this paper constructs an MSCNN-GAP model. Compared with the single convolution kernel CNN, the multiscale CNN can select the convolution kernel size more flexibly to extract input features, thereby improving the noise tolerance of the model. Meanwhile, the multiscale CNN is a parallel structure. When facing training sample sets with noise, the multiscale structure can update the parameters of each branch more flexibly to better adapt to the multisource fusion dataset. In traditional CNN, the extracted feature maps will be reshaped into a set of one-dimensional feature vectors, and the final classification result will be output through the fully connected layer. This reshaping method will change the spatial structure of the feature maps. In the multiscale fusion process, maintaining the spatial structure of the feature maps of each branch is more helpful for training the network model. Therefore, the global average pooling is selected at the end of each branch instead of the traditional method of reshaping the

Figure 7: Experimental facility.

Table 2: Description of the bearing health condition.

	Condition description	Fault degree
NO	Normal condition	None
REF	Rolling element fault	Obvious
ORF	Outer ring fault	Obvious
IRF1	Inner ring faults	Obvious
IRF2	Inner ring faults	Serious

Table 3: Description of the fault diagnosis task.

Group	Vertical SNR	Horizontal SNR	Training sample	Test samples
A	0	0	500	500
B	-3	0	1000	500
C	-5	0	1000	500
D	0	-3	1000	500
E	0	-5	1000	500
F	-3	-3	1500	500
G	-5	-5	1500	500

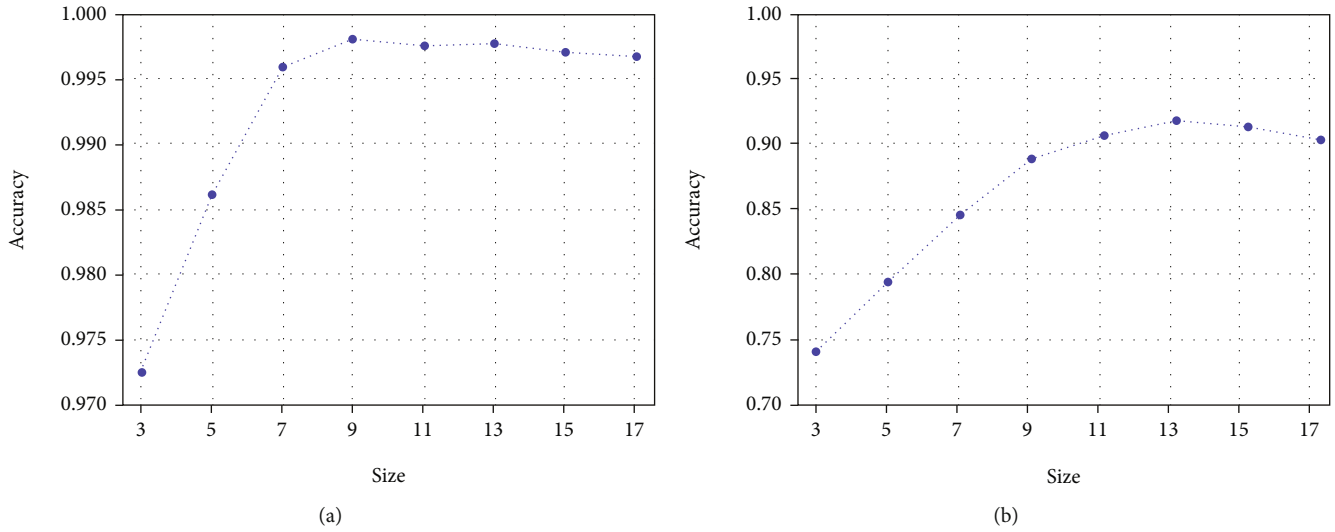


Figure 8: The influence of different sizes of convolution kernels on the model.

convolution layer into a one-dimensional vector. This makes the output results maintain the spatial structure of the feature maps in the fusion process and better adapt to the multiscale fusion process.

In order to contain as many signal features as possible, 4096 sampling points are selected as a group of signal samples. In the convolution process, three different sizes of convolution kernels are used to extract input features from the diversity. After each convolution and pooling calculation, nonlinear activation is performed by the ReLU function, and the batch normalization layer is added to improve the calculation speed. Finally, a hidden layer with 100 neurons is used to reduce the dimensionality of the fused one-dimensional vector, and the final classification result is output through the output layer. Table 1 shows the specific parameters of the convolution process of each branch. Figure 6 shows the structure diagram of the overall diagnosis model.

4. Experimental Verification

4.1. Experimental Dataset. In this paper, the fault data measured by the bearing fault experimental platform is used to verify the effectiveness of the algorithm. The experimental platform is shown in Figure 7. The platform is mainly composed of an AC motor, support bearing, base, rotating shaft, test bearing, and loading system. The model of the faulty bearing is SKFNU205. The fault types of the faulty bearing are shown in Table 2.

During the signal acquisition process, the loading system exerts a certain radial force on the faulty bearing. A horizontal sensor and a vertical sensor are used to collect signals in different directions, respectively. The sampling frequency of the signal is 16384 Hz, and the speed of the motor is 1487 r/min. In order to better verify the capability of the model to extract features, the collected fault signals are normalized in amplitude, which weakens the feature gap between the categories and further improves the difficulty of classification.

Table 4: Comparison of results of different algorithms.

Number	Algorithm	Accuracy (%)	Standard deviation
1	CNN (vertical)	99.08	0.858836
2	CNN (horizontal)	97.62	1.157411
3	MCCNN (combination)	98.6	0.758947
4	MSCNN (combination)	99.14	0.732393
5	MSCNN-GAP (combination)	99.93	0.149071

In order to better simulate the signal data in the case of noise, white noise with different signal-to-noise ratios (SNR) is added to the signals collected in different directions. The SNR can be defined as

$$\mathrm{SNR} = 10 \log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right), \tag{5}$$

where P_{signal} and P_{noise} are the power of the signal and noise.

Every 4096 sampling points of signal data are taken as a sample, and several groups of sample sets for training and testing are constructed by resampling. Among them, Group A has a relatively simple sample because there is no noise factor, and 500 training samples will be used to train the model. Group B, Group C, Group D, and Group E have certain noise factors, and 1000 training samples will be used. Group F and Group G are most affected by noise, and 1500 training samples are used for training. The specific parameters of several sample sets are shown in Table 3.

4.2. Selection of the Convolution Kernel. The size of the convolution kernel is one of the important hyperparameters in CNN. Convolution kernels of different sizes are adapted to different types of datasets. In order to maximize the

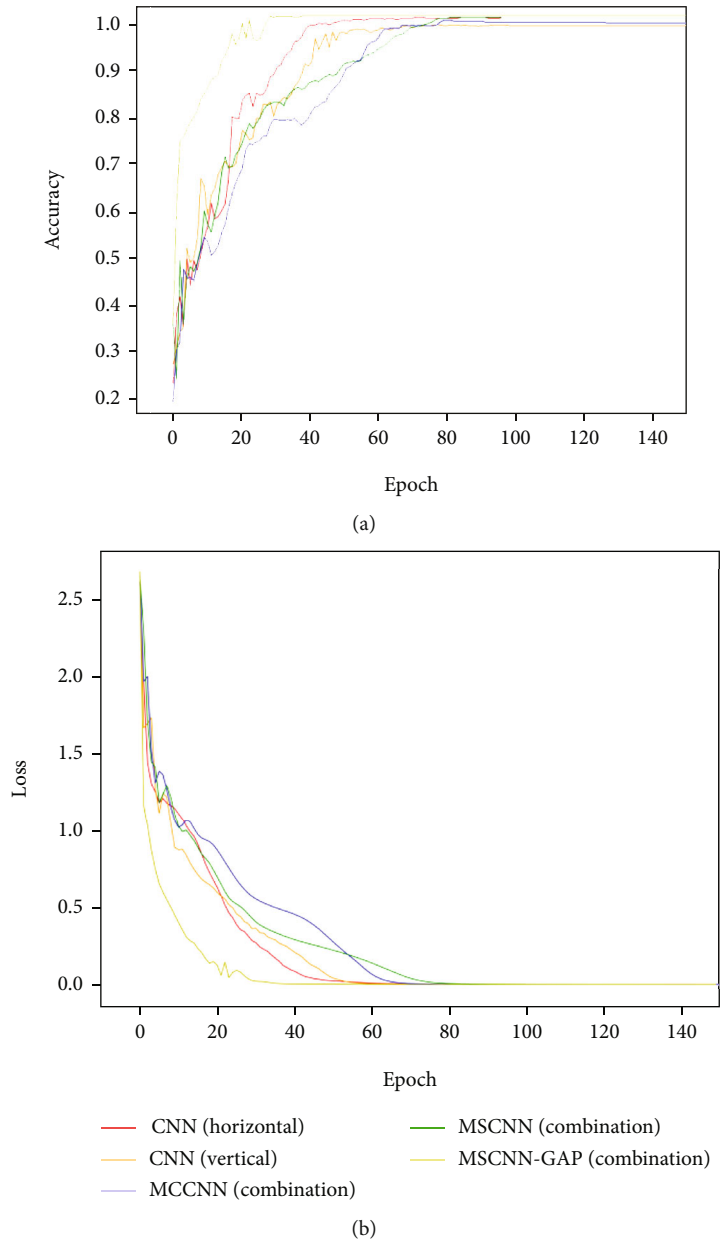


Figure 9: Model loss and classification accuracy for different algorithms.

diagnostic performance of CNN, this paper first tested the effect of different sizes of convolution kernels on the results. The signal data collected in the vertical direction is selected as the experimental sample. MSCNN-GAP with single-sensor information was chosen as the experimental algorithm. The experimental results are shown in Figure 8. The training set of Figure 8(a) uses 500 training samples. The training set in Figure 8(b) uses 1500 samples with SNR-5.

The results show that the smaller convolution kernel has a poor feature extraction effect due to the smaller receptive field for this dataset. To a certain extent, the performance of CNN is enhanced with the widening of the convolution kernel size. The convolution kernel with the size of 9 is the best when the sample has no noise interference. In the case of noise interference, the convolution kernel with the size of 13 has the best effect. After reaching the optimal value, with the gradual widening of the convolution kernel size, the diagnostic ability of CNN has a downward trend. The reason is that a large convolution kernel size is weak for local feature extraction. The comparison results of the two groups of data show that when the signal is disturbed by noise, the local features of the sample are weakened, and the larger convolution kernel that can better extract global features is more suitable for noisy datasets.

4.3. Comparative Analysis of Algorithms. In this section, we will verify the computational complexity of the proposed algorithm and its performance in multidata fusion. In order to better reflect the superiority of MSCNN-GAP, three other different algorithms are compared:

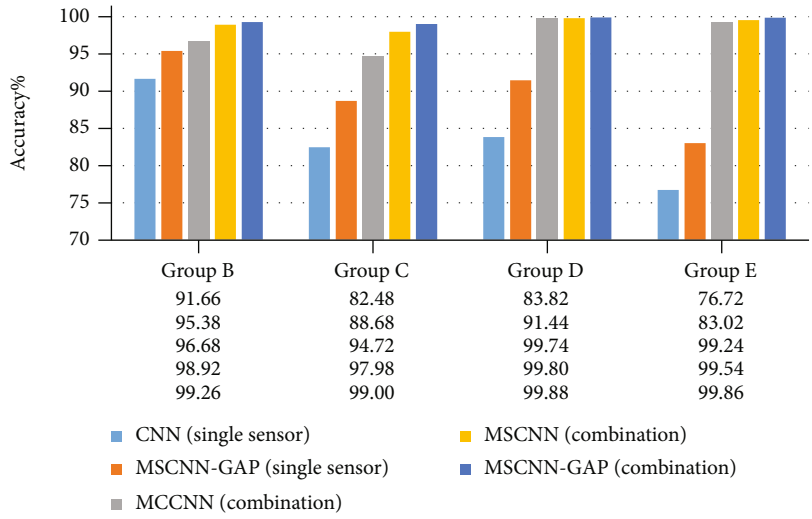


Figure 10: Comparison of results of different algorithms.

CNN (single sensor). This model uses a single-sensor signal as the input of CNN. The CNN model consists of four convolution modules and two fully connected layers. Each group of convolution modules is composed of a convolution layer, a pooling layer, and a normalization layer. At the end of convolution, the feature map is reshaped into a one-dimensional vector, and the dimension is reduced by a fully connected layer. Finally, the classification results of CNN are output [25].

Multichannel CNN (MCCNN). In order to apply multisensor information, multisensor data is input to MCCNN in the form of multichannel input. The specific structure of the MCCNN model is the same as the CNN model in algorithm 1 [34].

Multiscale CNN (MSCNN). Based on multichannel input in algorithm 2, in order to further improve the performance of CNN, this algorithm uses multiscale convolution kernel diversity to extract the features of the input. After the convolution operation of each branch, the feature map will be reshaped into a one-dimensional vector and reduced by the fully connected layer. Then, the one-dimensional vectors of each branch are merged and reduced again by the fully connected layer. Finally, the fully connected layer is used to output the final classification results [29].

MSCNN-GAP. The algorithm proposed in this paper uses multisensor data for fault diagnosis. Multisensor data will be input into the model as multiple channels. Based on CNN, a multiscale parallel structure is adopted to reduce the impact of noisy data on the model and improve the robustness of the model. After the convolution operation of each branch, the feature map of each branch will be reduced through the global average pooling layer, and the reduced feature vectors will be fused. GAP not only reduces the parameters of the model but also retains the spatial structure of the feature map before fusion. The fused feature vector will be reduced through a fully connected layer, and finally, the classification results will be output through a fully connected layer.

Computational complexity is a problem that must be considered in the deep learning model. The smaller the computational complexity, the shorter the computational time, and the higher the computational efficiency. This paper verifies the complexity of the proposed model from two aspects: time complexity and model parameters. For a convolutional neural network, its time complexity can be expressed as

$$\textbf{Time} \sim O\left(M^2 \cdot K^2 \cdot C_{\text{in}} \cdot C_{\text{out}}\right), \tag{6}$$

where M is the size of the output feature map, K is the size of the convolution kernel, C_{in} is the number of input channels of the convolution layer, and C_{out} is the number of output channels of the convolution layer [48].

Thus, the time complexity of the fully connected layer and the global average pooling layer can be deduced as follows:

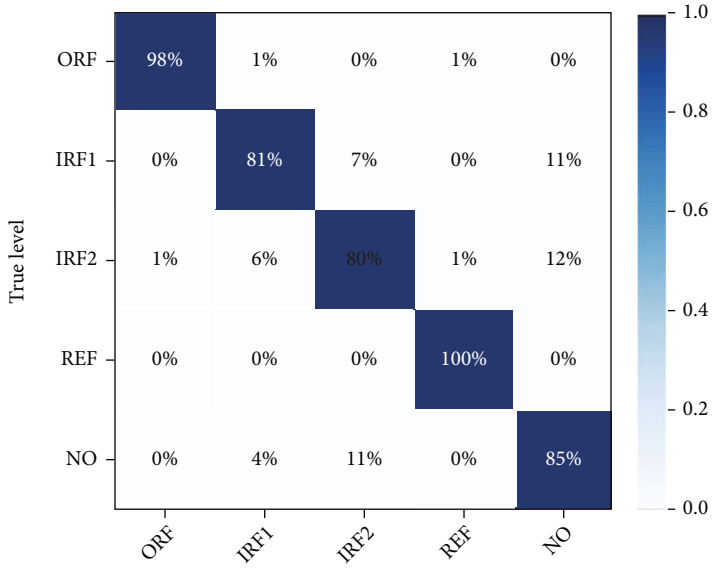
$$\begin{aligned} \textbf{Time} &\sim O\left(1^2 \cdot X^2 \cdot C_{\text{in}} \cdot C_{\text{out}}\right), \\ \textbf{Time} &\sim O\left(C_{\text{in}} \cdot C_{\text{out}}\right), \end{aligned} \tag{7}$$

where X is the number of neurons in the fully connected layer.

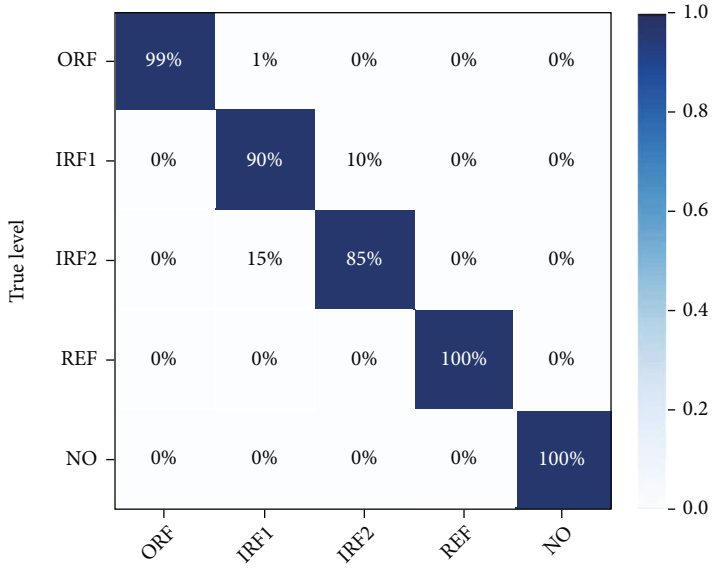
The algorithm proposed in this paper uses the global average pooling layer to replace the traditional fully connected layer at the end of the convolution operation. We input the network parameters of the fully connected layer and the global average pooling layer into the time complexity calculation formula. The time complexity involved is

$$128^2 \cdot 4^2 \cdot 4^2 \Rightarrow 128^2. \tag{8}$$

The former represents the time complexity of traditional multiscale CNN. The latter represents the time complexity of MSCNN-GAP. The result shows that the proposed algorithm is superior to the traditional multiscale CNN model in time complexity. From the point of view of the parameters

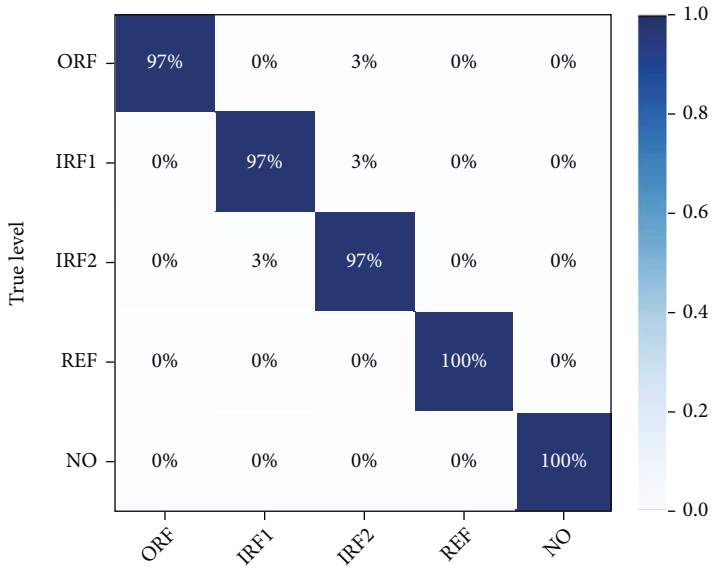


(a) MSCNN-GAP (single sensor)

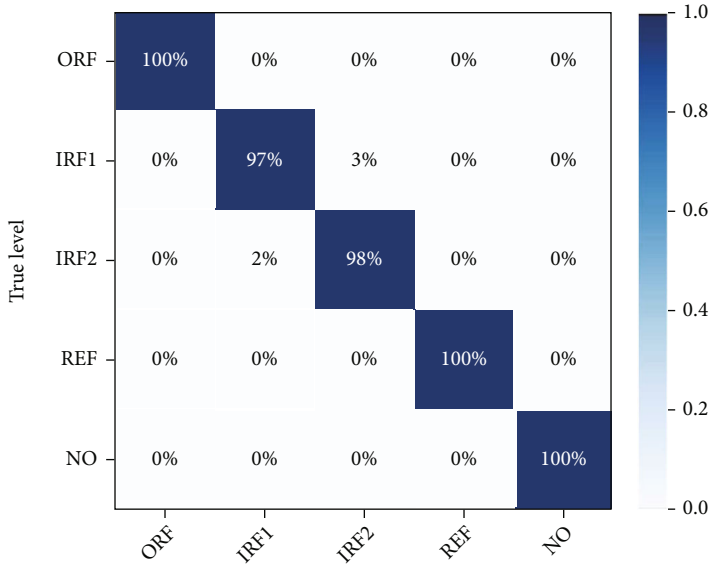


(b) MCCNN (combination)

Figure 11: Continued.



(c) MSCNN (combination)



(d) MSCNN-GAP (combination)

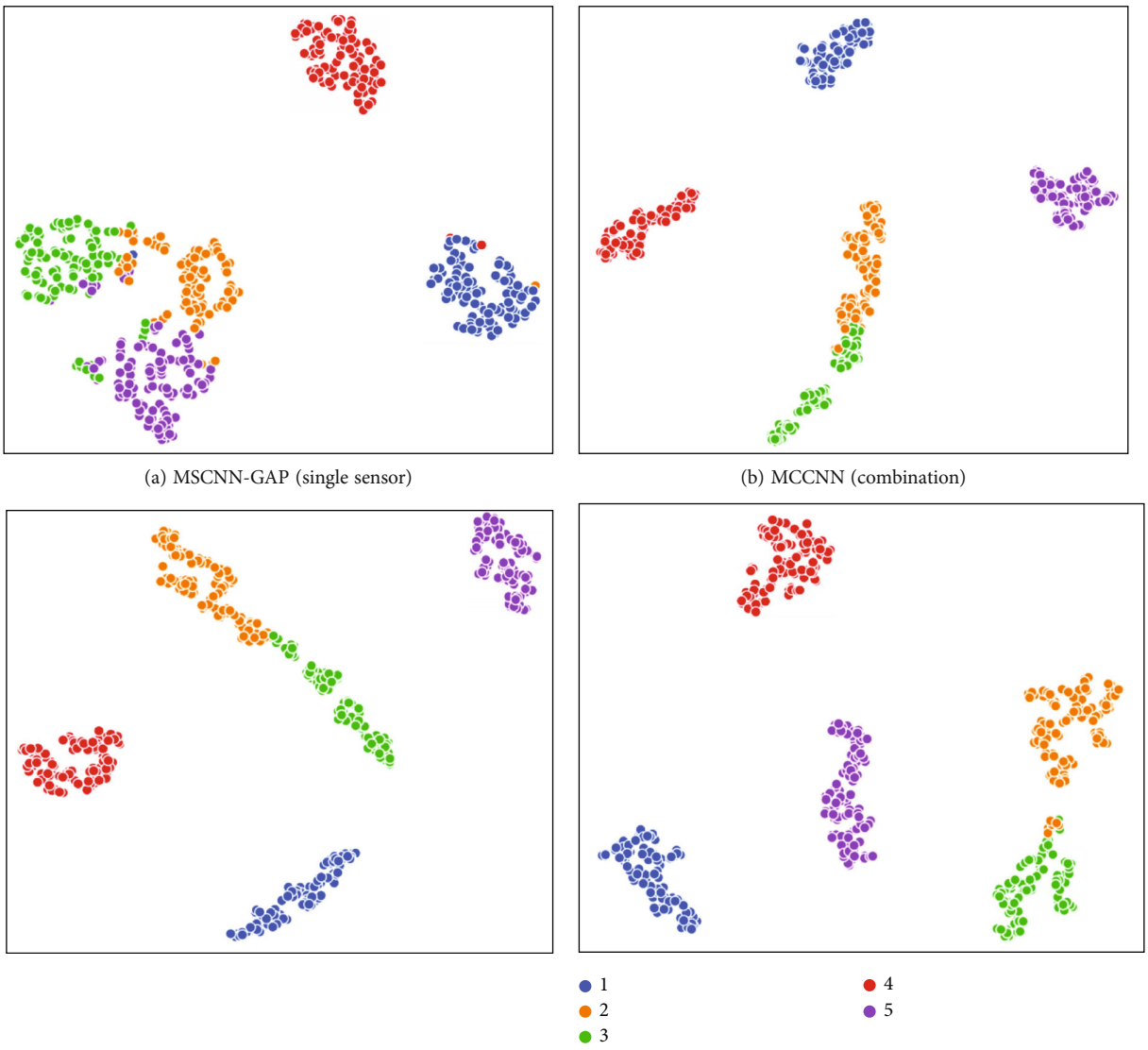
Figure 11: Confusion matrix results of each algorithm.

of the model, K variable weights are needed to calculate the final output for the fully connected layer containing K neurons. The global average pooling layer outputs the final result by obtaining the average value of its feature map and does not need other parameters to participate in the calculation. Therefore, the algorithm proposed in this paper has fewer parameters. In conclusion, the computational complexity of the proposed algorithm is less than that of the traditional multiscale CNN model.

The performance of the algorithm will be verified below. According to the results of the previous section, we use a convolution kernel of size 9 as the feature extractor of CNN to extract the input features of noiseless samples. For noisy samples, we use a convolution kernel of size 13. For the multiscale model, the best convolution kernel size and its adjacent convolution kernel size are used for feature extraction.

In order to avoid contingency, each model is cross-validated ten times, and the average of the final results is obtained. The results are shown in Table 4.

The results show that the diagnostic ability based on the CNN algorithm can reach a high level in the absence of noise, which proves the feasibility of the CNN algorithm in fault diagnosis. In this experiment, because the



(a) MSCNN-GAP (single sensor)

(b) MCCNN (combination)

(c) MSCNN (combination)

(d) MSCNN-GAP (combination)

Figure 12: The t-SNE results of each algorithm.

surrounding noise interference is weak when the signal is collected, the fault features of each category are more obvious, and the results' gap between the algorithms is relatively small. Under the influence of different fault forms, fault locations, load directions, and other conditions, the significant degree of signal features collected by each direction sensor is different. In this experiment, since the load is applied in the vertical direction, compared with the horizontal direction, the signal features in the vertical direction will be more obvious. Therefore, the model trained by the signal collected by the vertical sensor has better results. In order to fuse multisensor data, only using multisensor information as multichannel input will make the result higher than the lower value trained by the single-sensor information and lower than the higher value. The reason is that for training samples with obvious features, the addition of samples with insufficient features causes a certain amount of noise interference and reduces the quality of the training set, which leads to the difficulty of feature extraction and reduces the performance of the model. In order to weaken the influence of this factor, a multiscale feature extraction method is introduced to effectively solve the problem of feature extraction difficulty and improve the accuracy of diagnosis. On this basis, the multiscale fusion method of global average pooling is introduced, which effectively retains the feature space structure while fusing so that the model achieves higher performance and more stable results.

Figure 9 shows the image of the accuracy and loss of each algorithm varying with the training epoch. The results show that CNN, MCCNN, and MSCNN all need more convergence epochs to achieve convergence. The convergence epoch of the MSCNN-GAP is greatly reduced. Meanwhile, the loss of the algorithm proposed in this paper decreases

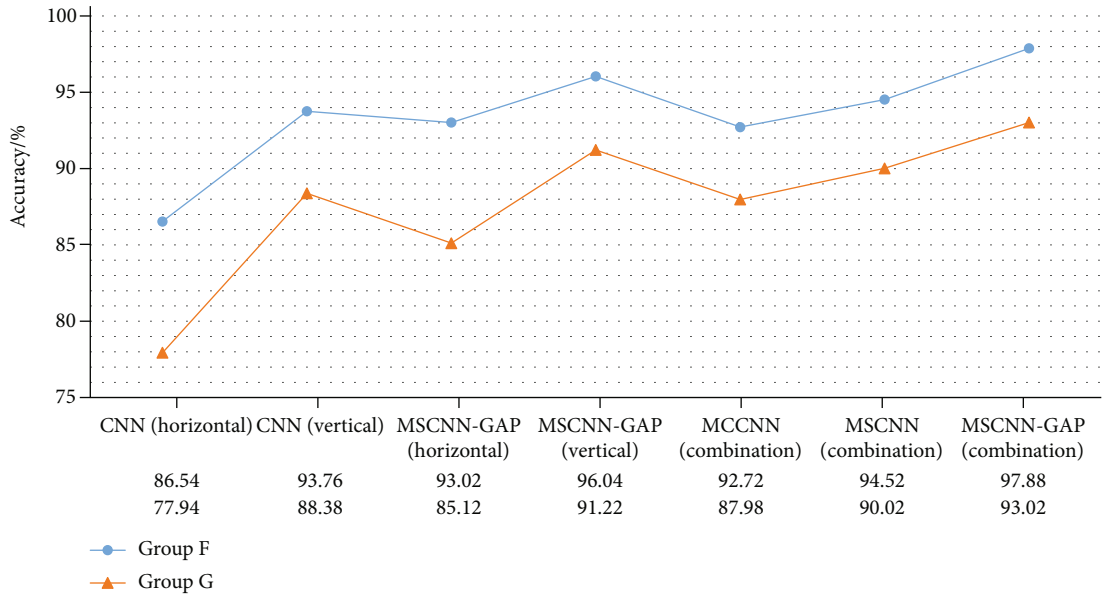


Figure 13: Comparison of diagnosis results of different algorithms.

faster, and the convergence process is more stable than other algorithms.

4.4. Comparative Analysis of Robustness. In the process of fault signal acquisition, it is often interfered with by the machine noise and surrounding noise. When there is much noise in the fault signal, the fault features will be covered by the noise information, which increases the difficulty of extracting effective features from the model. Noise will affect the quality of the dataset to a certain extent, thereby affecting the training process and the performance of the model. Therefore, the diagnostic model needs to have better robustness.

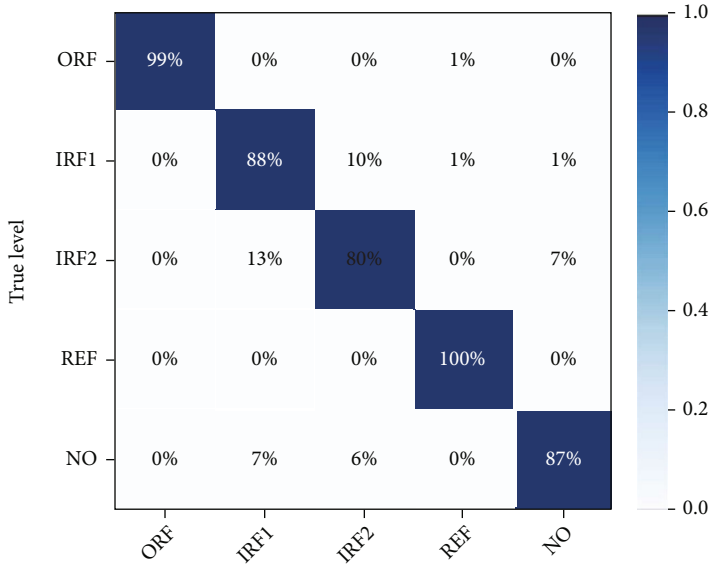
In this section, Groups B, C, D, and E will be used for comparative verification. The samples with noise factors are used as the training set for the algorithm using single-sensor data. Similarly, each algorithm is run ten times, and the average values are counted. The results are shown in Figure 10.

The results show that the diagnostic accuracy of the CNN model with single-sensor data is lower than 95% when the noise is weak. When the noise is serious, the diagnostic accuracy is lower than 85%. It also shows that the performance of the CNN model will be greatly affected by noise. The multiscale structure can extract more features and improve the diagnostic ability of the model. However, due to the interference of noise factors, the fault features are not obvious enough, and the diagnostic accuracy is relatively low. Combined with multisensor data, this problem can be alleviated effectively. In signal acquisition, due to the load in the vertical direction, the fault features of the signal collected in the vertical direction are more obvious. The feature extraction process will be more difficult when this group of signals is affected by the noise factors. This also makes the overall accuracy of Group D and Group E higher than Group B and Group C in the multisensor fusion algorithm. By adding multiscale information and using the global pooling method to save the feature space, the algorithm proposed in this paper better adapts to the samples with difficult feature extraction. It improves the diagnosis ability of the model.

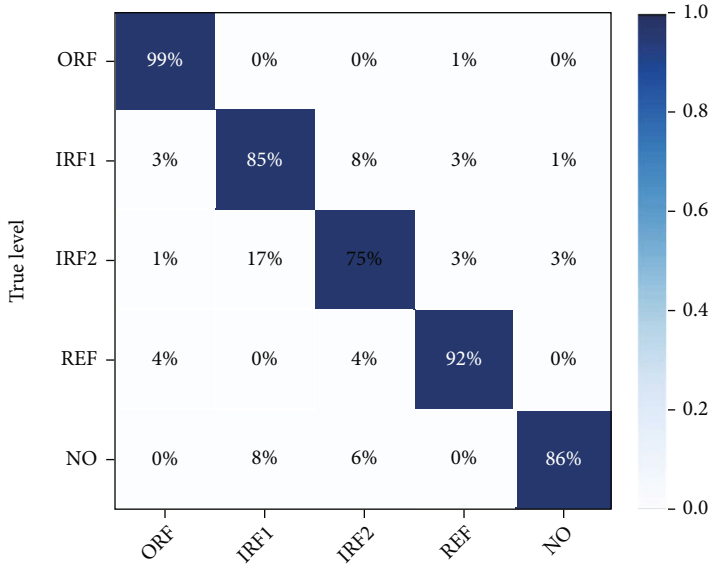
Figure 11 shows the confusion matrix of each algorithm. The results show that the two categories of inner ring faults are most likely to be confused due to the same fault form. After adding the noise, this category of NO also has a certain degree of misjudgment. The diagnostic capability of CNN has declined to a certain extent. Multisource data can improve the representation capability of the network. However, it is difficult to extract features between similar fault categories, and there is still a large degree of confusion between the two inner ring faults. The multiscale model can extract features better and reduce the confusion between the two categories. The results show that this paper further improves the performance of the model by the multiscale fusion method using global average pooling.

In order to intuitively show the effectiveness of the proposed algorithm, we use the t-SNE algorithm to reduce the dimensionality and visualize the result of feature extraction. Figure 12 shows the t-SNE visualization results of the four algorithms. Among them, 1-5 represent the five fault categories of ORF, IRF1, IRF2, REF, and NO. The results show that with the continuous optimization of the algorithm, the classification effect of each algorithm is gradually improved, which corresponds to the results of the confusion matrix. The proposed algorithm shows better clustering performance. There is more excellent separability between features of different categories. It further proves the effectiveness of the algorithm proposed in this paper.

In extreme cases, multisensor signals may have noise interference at the same time, which further improves the difficulty of feature extraction. In order to further verify

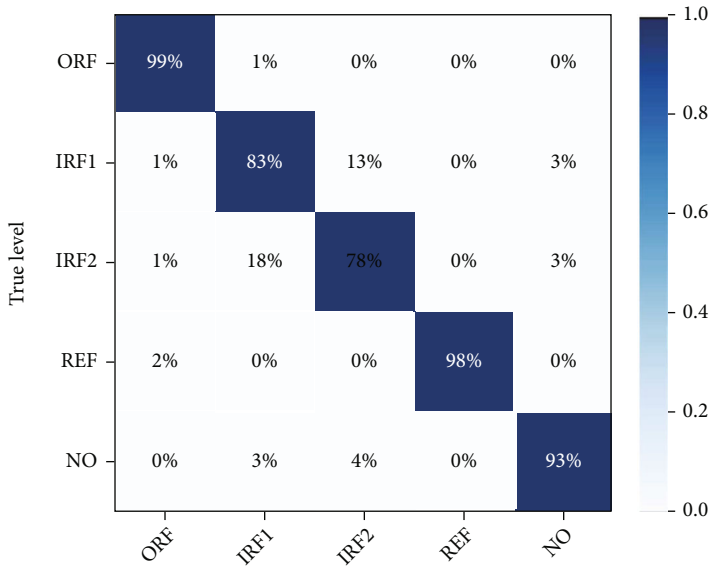


(a) MSCNN-GAP (vertical)

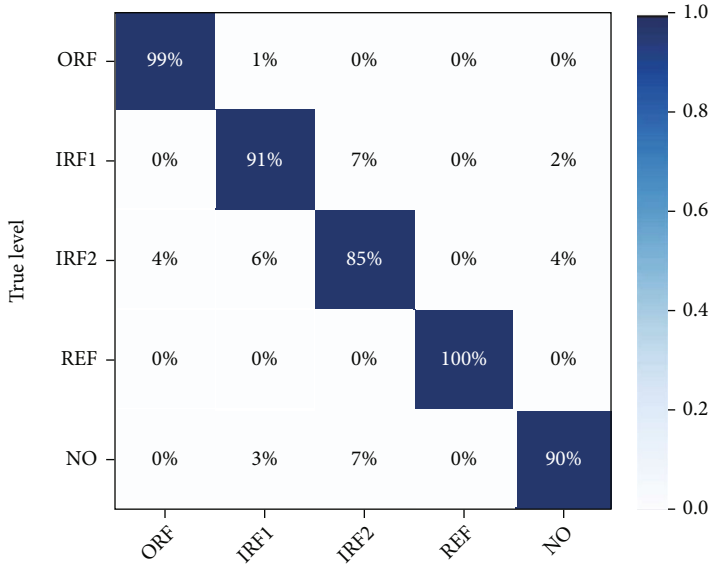


(b) MCCNN (combination)

Figure 14: Continued.



(c) MSCNN (combination)



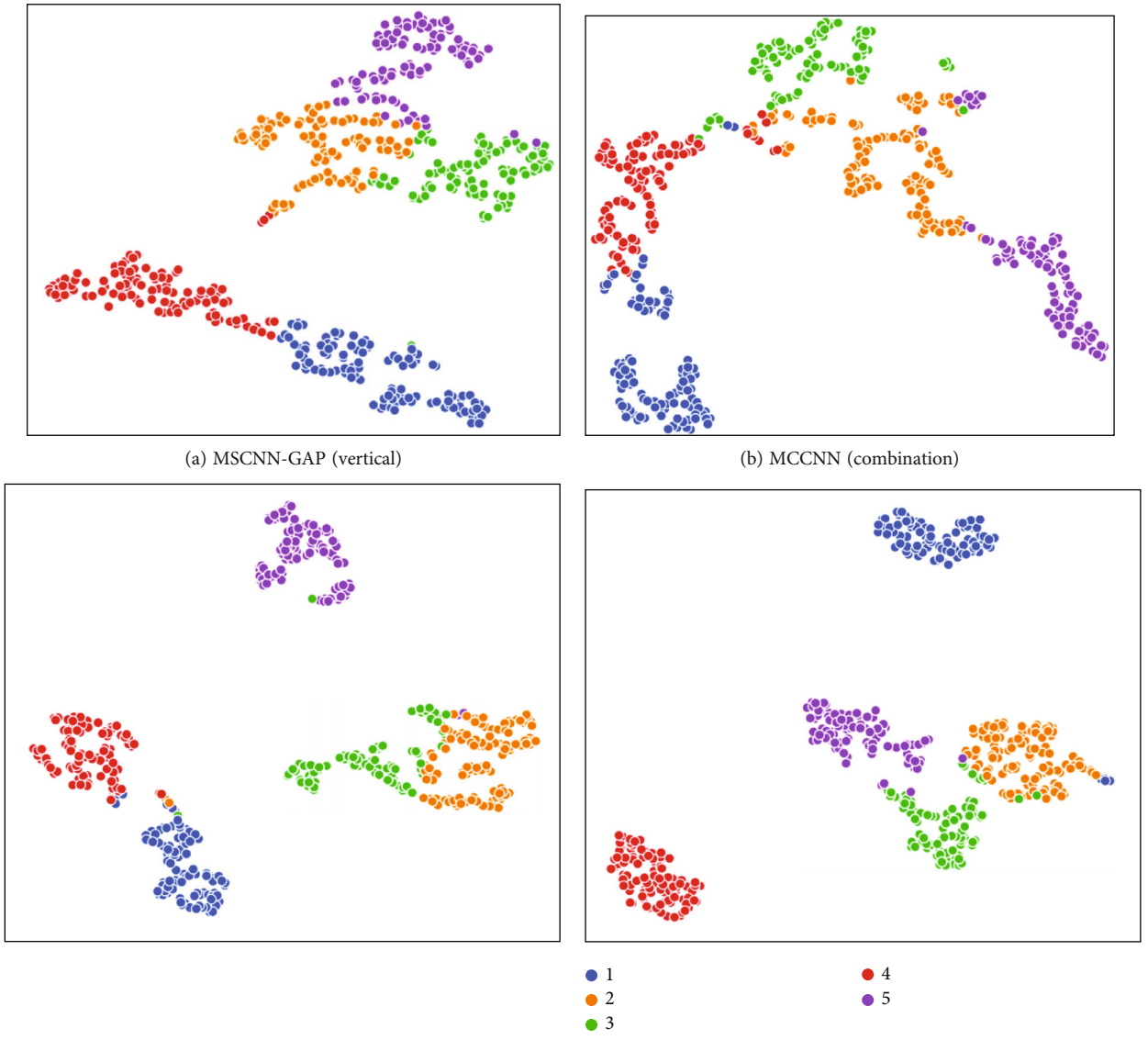
(d) MSCNN-GAP (combination)

Figure 14: Confusion matrix results of each algorithm.

the robustness of the algorithm, Group F and Group G are used to compare and verify the algorithms. Each algorithm is run ten times, and the average values are counted. The comparison results are shown in Figure 13.

The results show that under the influence of noise factors, the diagnostic accuracy of each algorithm is affected to a certain extent. Among the algorithms that use single-sensor data, MSCNN-GAP has a better classification effect than traditional CNN. In terms of the data fusion algorithm, the result of data fusion using multichannel input is slightly lower than that of the model using the vertical signal. The data fusion algorithm proposed in this paper can obtain better results than MSCNN-GAP using single-sensor data. Therefore, MSCNN-GAP can apply multisource data more efficiently. The reason is that the multiscale model is a parallel structure, and the parameters of multiple branches can be adjusted to each other. Compared with the traditional CNN algorithm, MSCNN-GAP is more flexible. Therefore, the MSCNN-GAP has better robustness and can better adapt to the fault diagnosis of multisource data under noise factors.

Similarly, Figures 14 and 15 show the confusion matrix and t-SNE visualization results of each algorithm. The results show that under the influence of noise



(a) MSCNN-GAP (vertical)

(b) MCCNN (combination)

(c) MSCNN (combination)

(d) MSCNN-GAP (combination)

Figure 15: The t-SNE results of each algorithm.

factors, the result of the data fusion algorithm with only multichannel input is confusing relatively. Meanwhile, the visualization shows that the boundaries between categories are also blurred. The MSCNN-GAP model can distinguish the confusing categories better, and the boundaries between the categories are more obvious than other algorithms.

5. Conclusion

In order to use multisensor data more efficiently, this paper proposes an MSCNN-GAP model. MSCNN-GAP uses multiscale convolution kernels to extract more diverse features, which alleviates the problem of poor robustness of the CNN model under noise factors effectively. The parallel structure of the multiscale model is better adapted to the data fusion process. In the fusion process, instead of reshaping the feature maps into a one-dimensional vector, the global average pooling method is adopted, which effectively retains the spatial structure and position of the feature maps in the fusion process. The proposed model is verified by the bearing fault data collected from the experimental platform. Experimental results show that, compared with other data fusion methods, the algorithm proposed in this paper makes more effective use of multisensor information and obtains a higher diagnostic accuracy and a shorter convergence period. The proposed algorithm has stronger robustness and better classification results when the signal samples are affected by noise factors.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under grant 51675350.

References

[1] Z. Liu and L. Zhang, "A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings," *Measurement*, vol. 149, no. 107002, 2020.

[2] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, no. 106587, 2020.

[3] M. Cerrada, R. V. Sánchez, C. Li et al., "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 2018.

[4] H. Zhao, H. Liu, Y. Jin, X. Dang, and W. Deng, "Feature extraction for data-driven remaining useful life prediction of rolling bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 3511910, 2021.

[5] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 108–126, 2013.

[6] Z. Li, Y. Jiang, Q. Guo, C. Hu, and Z. Peng, "Multi-dimensional variational mode decomposition for bearing-crack detection in wind turbines with large driving-speed variations," *Renewable Energy*, vol. 116, no. B, pp. 55–73, 2018.

[7] H. Cao, F. Fan, K. Zhou, and Z. He, "Wheel-bearing fault diagnosis of trains using empirical wavelet transform," *Measurement*, vol. 82, pp. 439–449, 2016.

[8] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.

[9] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[10] K. Zhu, X. Song, and D. Xue, "A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm," *Measurement*, vol. 47, pp. 669–675, 2014.

[11] M. Kang, J. Kim, J. Kim, A. C. C. Tan, E. Y. Kim, and B. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Transactions on Power Electronics*, vol. 30, pp. 2786–2797, 2015.

[12] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1793–1803, 2016.

[13] D. He, R. Li, and J. Zhu, "Plastic bearing fault diagnosis based on a two-step data mining approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3429–3440, 2013.

[14] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: a review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[15] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature Methods*, vol. 16, no. 12, pp. 1233–1246, 2019.

[16] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[17] D. Wang and J. Chen, "Supervised speech separation based on deep learning: an overview," *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 26, no. 10, pp. 1702–1726, 2018.

[18] Y. Chien, C. Wu, and H. Tsao, "Automatic sleep-arousal detection with single-lead EEG using stacking ensemble learning," *Sensors*, vol. 21, no. 18, p. 6049, 2021.

[19] J. Jiao, M. Zhao, J. Lin, and K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, vol. 417, pp. 36–63, 2020.

[20] S. Xing, Y. Lei, S. Wang, and F. Jia, "Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2617–2625, 2021.

[21] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, 2016.

[22] Z. Liu, H. Wang, J. Liu, Y. Qin, and D. Peng, "Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 3501711, 2021.

[23] X. Liu, Q. Zhou, J. Zhao, H. Shen, and X. Xiong, "Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network," *Sensors*, vol. 19, no. 9724, 2019.

[24] Y. Xu, Z. Li, S. Wang, W. Li, T. Sarkodie-Gyan, and S. Feng, "A hybrid deep-learning model for fault diagnosis of rolling bearings," *Measurement*, vol. 169, no. 108502, 2021.

[25] Q. Zhou, Y. Li, Y. Tian, and L. Jiang, "A novel method based on non-linear auto-regression neural network and convolutional neural network for imbalanced fault diagnosis of rotating machinery," *Measurement*, vol. 161, no. 107880, 2020.

[26] P. Liang, C. Deng, J. Wu, and Z. Yang, "Intelligent fault diagnosis of rotating machinery via wavelet transform, generative adversarial nets and convolutional neural network," *Measurement*, vol. 159, no. 107768, 2020.

[27] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2018.

[28] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mechanical Systems and Signal Processing*, vol. 140, no. 106683, 2020.

[29] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, "Dislocated time series convolutional neural architecture: an intelligent fault diagnosis approach for electric machine," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1310–1320, 2017.

[30] J. Wang, S. Li, B. Han et al., "Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis," *Measurement Science and Technology*, vol. 30, no. 1, p. 015106, 2019.

[31] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196–3207, 2019.

[32] H. Wang, Z. Liu, D. Peng, and Y. Qin, "Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 5735–5745, 2020.

[33] J. Wang, D. Wang, S. Wang, W. Li, and K. Song, "Fault diagnosis of bearings based on multi-sensor information fusion and 2D convolutional neural network," *IEEE Access*, vol. 9, pp. 23717–23725, 2021.

[34] X. Xu, Z. Tao, W. Ming, Q. An, and M. Chen, "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," *Measurement*, vol. 165, no. 108086, 2020.

[35] P. Shan, H. Lv, L. Yu, H. Ge, Y. Li, and L. Gu, "A multisensor data fusion method for ball screw fault diagnosis based on convolutional neural network with selected channels," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 7896–7905, 2020.

[36] H. Wang, S. Li, L. Song, L. Cui, and P. Wang, "An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2648–2657, 2020.

[37] M. Xia, Z. Mao, R. Zhang, B. Jiang, and M. Wei, "A new compound fault diagnosis method for gearbox based on convolutional neural network," in *Proceedings of 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS'20)*, pp. 1077–1083, Liuzhou, China, 2020.

[38] K. Gu, Y. Zhang, X. Liu, H. Li, and M. Ren, "DWT-LSTM-based fault diagnosis of rolling bearings with multi-sensors," *Electronics*, vol. 10, no. 207617, 2021.

[39] L. Kou, Y. Qin, X. Zhao, and X. Chen, "A multi-dimension end-to-end CNN model for rotating devices fault diagnosis on high-speed train bogie," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 2513–2524, 2020.

[40] H. Li, J. Huang, X. Yang, J. Luo, L. Zhang, and Y. Pang, "Fault diagnosis for rotating machinery using multiscale permutation entropy and convolutional neural networks," *Entropy*, vol. 22, no. 8518, 2020.

[41] B. Peng, H. Xia, X. Lv et al., "An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network," *Applied Intelligence*, pp. 1–15, 2021.

[42] Y. Lecun, B. Boser, J. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

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

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[45] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *International Conference on International Conference on Machine Learning*, Haifa, Israel, 2010.

[46] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," in *IEEE International Conference on Computer Vision*, pp. 1026–1034, New York, 2015.

[47] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, New York, 2015.

[48] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5353–5360, Boston, America, 2015.