

Article

ECNN: Intelligent Fault Diagnosis Method Using Efficient Convolutional Neural Network

Chao Zhang ^{1,*}, Qixuan Huang ^{1,†}, Chaoyi Zhang ^{2,†}, Ke Yang ^{3,†}, Liye Cheng ⁴ and Zhan Li ⁵

¹ Department of Integrated Technology and Control Engineering, School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China

² School of Civil Aviation, Northwestern Polytechnical University, Xi'an 710072, China

³ Beijing Aerospace Systems Engineering Research Institute, Beijing 100076, China

⁴ The Fifth Electronics Research Institute of Ministry of Industry and Information Technology, Guangzhou 510610, China

⁵ China Institute of Marine Technology & Economy, Beijing 100081, China

* Correspondence: caec_zc@nwpu.edu.cn

† These authors contributed equally to this work.

Abstract: With outstanding deep feature learning and nonlinear classification abilities, Convolutional Neural Networks (CNN) have been gradually applied to deal with various fault diagnosis tasks. Affected by variable working conditions and strong noises, the empirical datum always has different probability distributions, and then different data segments may have inconsistent contributions, so more attention should be assigned to the informative data segments. However, most of the CNN-based fault diagnosis methods still retain black-box characteristics, especially the lack of attention mechanisms and ignoring the special contributions of informative data segments. To address these problems, we propose a new intelligent fault diagnosis method comprised of an improved CNN model named Efficient Convolutional Neural Network (ECNN). The extensive view can cover the special characteristic periods, and the small view can locate the essential feature using Pyramidal Dilated Convolution (PDC). Consequently, the receptive field of the model can be greatly enlarged to capture the location information and excavate the remarkable informative data segments. Then, a novel residual network feature calibration and fusion (ResNet-FCF) block was designed, which uses local channel interactions and residual networks based on global channel interactions for weight-redistribution. Therefore, the corresponding channel weight is increased, which puts more attention on the information data segment. The ECNN model has achieved encouraging results in information extraction and feature channel allocation of the feature. Three experiments are used to test different diagnosis methods. The ECNN model achieves the highest average accuracy of fault diagnosis. The comparison results show that ECNN has strong domain adaptation ability, high stability, and superior diagnostic performance.

Keywords: intelligent fault diagnosis; efficient convolutional neural network; pyramidal dilated convolution; residual neural network; feature calibration and fusion



Citation: Zhang, C.; Huang, Q.; Zhang, C.; Yang, K.; Cheng, L.; Li, Z. ECNN: Intelligent Fault Diagnosis Method Using Efficient Convolutional Neural Network. *Actuators* **2022**, *11*, 275. <https://doi.org/10.3390/act11100275>

Academic Editors: Giorgio Olmi and Anyang Lu

Received: 20 August 2022

Accepted: 23 September 2022

Published: 26 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the era of Big Data, fault diagnosis has always played an important role in industrial production [1–3]. In practical environments, the equipment works with complex operating conditions and strong noise. Its core components (such as bearings, gears, and motors) occasionally fail, and some faults are challenging to locate using traditional methods, so intelligent fault diagnosis is essential to ensuring their safety and reliability [4–9].

Convolutional Neural Network (CNN) [10–15], one of the representative deep learning models, is increasingly applied to deal with fault diagnosis tasks relying on its outstanding advances in deep features learning and nonlinear classification. Zhang et al. [16] proposed a fault diagnosis method. The method uses multiple parallel convolutional layers to extract

rich and complementary fault features effectively and then transforms the one-dimension signal into a two-dimension signal by continuous wavelet transforms. Huo et al. [17] extended a convolutional neural network with transfer learning, which can adaptively process one-dimensional vibration signals into two-dimension matrices and reduce the data distribution distance between the source and the target domains. Wang et al. [18] proposed a new multi-sensor information fusion method constructing the time-domain signals into a rectangular two-dimension matrix and then used an improved 2D-CNN to realize signal classification. Wen et al. [19] employed LeNet-5 to develop a new CNN and convert signals into 2D images through a conversion method. The proposed method easily captures the features. S. et al. [20] applied a one-dimensional convolutional neural network to motor fault diagnosis and proposed adding feature classification and extraction to a single learning body. Zhao et al. [21] proposed a normalized CNN for rolling bearing diagnosis with different severity and fault directions under data imbalance and variable conditions. Abdeljaber et al. [22] presented a novel, fast and accurate structural damage detection system using 1D Convolutional Neural Networks (CNNs) that has an inherent adaptive design to fuse feature extraction and classification blocks into a single and compact learning body. Dibaj et al. [23] proposed a fault diagnosis approach based on variational mode decomposition and CNN for rotating machinery. Janssens et al. [24] proposed a DL model for condition monitoring using CNN and proved that the feature learning method was significantly better than the feature extraction method in fault diagnosis of rotating machines. A Convolutional Neural Network (CNN) machine learning algorithm is proposed to classify gearbox faults in [25], and the learning features of the CNN filters are visualized to understand the physical fault diagnosis phenomena. Cheng et al. [26] implied an intelligent fault diagnosis method for the rotation machine based on a new continuous wavelet transforming the local binary convolutional neural network. Therefore, in the field of fault diagnosis, scholars have utilized a lot of convolutional neural networks to improve the accuracy of diagnosis and finally achieved excellent results. Although these studies verify the efficiency of CNN in the fault diagnosis field, the two following problems remain.

In the course of fault diagnosis, firstly, converting vibration signals to spectrogram requires a quantity of computation. Secondly, spectrograms are resized as small images before the training model to decrease computation cost time; some signal nature will be lost during the compression process. Therefore, the proposed method is dedicated to one-dimension data [27–29]. In addition, the traditional 1D-CNN uses the pooling layer to obtain the receptive field. Information on the Avg/Max pooling region is insufficient to capture the importance of the pooling feature. Although stride convolution could learn from neighbor features, it fails to model the importance of down sampling procedures adaptively and limits shift-invariance because it focuses only on one fixed location within each sliding window and discards the rest [30].

Since the equipment is usually affected by load changes and environmental noises, some scholars use dilated convolution to replace the pooling layer in standard convolution. This process can retain the timing relation of original signals and obtain larger-scale feature information, which helps improve the feature learning ability of neural networks. In [31], Su et al. used a dilated convolution deep belief network dynamic multilayer perceptron for bearing fault recognition under alterable running states. Zhao et al. [32] proposed a novel transfer learning framework based on a Deep Multi-Scale Convolutional Neural Network (MSCNN), in which dilated convolution and global average pooling were introduced to realize intelligent fault diagnosis of rolling bearings. Han et al. [33] used dilated convolutions to construct a Novel Multi-scale Dilated Convolutional Neural Network (NMDCNN) to enrich the field of view coverage. Meanwhile, piecewise cross-entropy is used to balance the misclassification cost between healthy samples and faulty samples. Chu et al. [34] advanced a novel multi-scale convolution model based on Multi-Dilation Rates And Multi-Attention Mechanism (MDRMA-MSCM) for mechanical fault diagnosis. Wang et al. [35] proposed Cascade Convolutional Neural Network (C-CNN) for fault diagnosis, and a cascade structure was built to avoid information loss. In conclusion, dilated convolution

has made significant achievements in the field of fault diagnosis, and its advantages are as follows:

- (1) Dilated convolution replaces the pooling layer operation in feature processing. This process not only cannot reduce the receptive field of the network but also can leave the temporal relationship of the original signal completely, which is of great significance for mining the domain invariant characteristics of the signal.
- (2) The use of large convolution kernels in standard convolution will increase the computation amount. Compared with larger convolution kernels, dilated convolution has significantly less computation making the model more accurate in processing fault features.

However, dilated convolution still maintains the black-box feature, which assigns the same weight to different feature channels, and cannot adaptively adjust the corresponding weight according to the importance of the channel. In particular, the lack of attention mechanism ignores the exceptional contribution of the information data segment.

In recent years, the attention mechanism has been experimentally implied to solve this problem. It is inspired by the method that the human brain uses limited computation to obtain high-value information while processing information. Based on the idea, it has been applied in plenty of deep learning methods. Extracting the features of the original vibration signals, the attentional mechanism can dynamically enhance the weight of significant feature channels, thus improving diagnostic accuracy. Zhang et al. [36] proposed a fault diagnosis method, which is utilized to realize spatiotemporal feature fusion, where vibration signal fused features with attention weight. Li et al. [37] constructed a rolling bearing fault diagnosis model, which combines a Dual-stage Attention-based Recurrent Neural Network (DA-RNN) and Convolutional Block Attention Module (CBAM). Yang et al. [38] developed a method based on multilayer bidirectional gated recurrent units with an attention mechanism to access the interpretability of neural networks in fault diagnosis. Zhang et al. [39] proposed a Hybrid Attention improved Residual Network (HA-ResNet) based method to diagnose the fault of the wind turbines' gearbox. This method highlights the essential frequency bands of wavelet coefficients and the fault features of convolution channels. Cao et al. [40] built a deep domain-adaptive multi-task learning model Y-Net, which is exploited to enable domain-adaptive diagnosis of faults in planetary gearboxes. The Squeeze and Excitation Residual (SE-Res) modules are utilized to reduce the redundancy of the model and improve the separability of deep features.

Attention mechanisms can be used to solve the problem that dilated convolution cannot adaptively obtain the particular weight of informative data segments. At the same time, the dilated convolution is introduced into the attention mechanism so that the model can gather a larger receptive field and extract more features. Inspired by the dilated convolution and attention mechanism, an ECNN model was proposed in this study. The model employs a novel Pyramidal Dilated Convolution to extract more valuable features, and the Residual Network Feature Calibration and Fusion (ResNet-FCF) block is implied to assign different weights to different feature channels. This article contributes:

- (1) In this model, the pyramidal dilatated convolution is designed to extract the features of data segments, which dramatically improves the receiving field and captures more features.
- (2) The ResNet-FCF is designed as a novel attention network architecture. Based on the local interaction learning scheme, ResNet-FCF introduces a residual network with one-dimensional convolution for global cross-channel interaction. The module assigns precise weight to the informative data segments.
- (3) The model has a good diagnosis effect under three practical examples.

The rest of this paper is organized as follows. The basic theory of the convolutional neural networks and residual networks are briefly introduced in Section 2. Moreover, it introduces an ECNN model and a novel fault diagnosis method. In Section 3, the experiment results are analyzed and discussed. Ultimately, the conclusions are presented in Section 4.

2. Materials and Methods

2.1. Convolutional Neural Networks

The CNN model is used for fault diagnosis, which can not only automatically extract fault features but also process a large amount of real-time data generated by the equipment.

2.1.1. Convolutional Layer

The main module of CNN [41] is the convolutional layer, which uses filters to perform a series of convolutional operations on the input data and output the corresponding features. Its mathematical model can be expressed as

$$y_j^k = (X * W) = g\left(\sum W_{ij}^k \cdot X_i^{k-1} + b_j^k\right) \quad (1)$$

where X represents the vibration signal, $*$ represents the convolution operation, k represents the k -layer network, W represents the convolution kernel, and y_j^k is the k -th layer output. g stands for nonlinear activation function, and b_j^k represents the bias vector.

When $F(x)$ (the output of the residual block) is 0, the basic mapping function of the residual block becomes $H(x) = x$, which accomplishes the identity mapping of the input and output. At the same time, according to the rule of backpropagation, it can be known that the backpropagation gradient of the residual part is $(\partial H)/(\partial x) + 1$. The model is merely required to minimize the residual map for approximating the identity map, which aims to ensure that the backpropagation gradient of the bottom layer is non-zero. Furthermore, the increase in the smoothness of interaction takes complete advantage of deep neural networks. The structure of the residual network is shown in Figure 1.

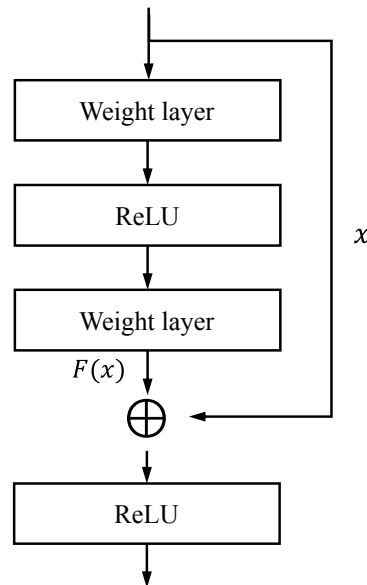


Figure 1. Residual neural network.

2.1.2. Pooling Layer

The pooling operation aims to reduce the dimension of the output of the previous layer. The essence is to effectively reduce the amount of calculation and retain important information, so the calculation resources and time are reduced in order to better preserve the data feature and make the model obtain a higher convergence speed. The pooling methods mainly include maximum pooling and average pooling. Maximum pooling p_{max} and average pooling p_{avg} refer to computing the maximum value and average pooling

in the pool window, respectively. The mathematical description of the two equations is shown as

$$p_{avg}^{l(i,j)} = \frac{1}{W} \sum_{t=(j-1)W+1}^{jW} a^{l(i,t)} \quad (2)$$

$$p_{max}^{l(i,j)} = \max_{(j-1)W+1 \leq i \leq jw} \{a^{l(i,t)}\} \quad (3)$$

where $a^{l(i,t)}$ is the activation value of the t neuron in frame i of layer l ; W is the width of the pool window.

2.1.3. Fully Connected Layer

Through a series of convolution and pooling, depth feature extraction is performed layer by layer, then the features are flattened and input into the fully connected layer, and the equation is

$$y_k = \sigma((w_f)_{sm}^T + b_f) \quad (4)$$

where w_f and b_f are the weight matrix and bias of the k -layer, respectively; y is the output of the fully connected layer. Usually, for obtaining the predicted output of the model, the softmax function is connected to realize the category classification after the fully connected layer, namely

$$f_k(y) = \exp(y_k) / \sum_{k=1}^C \exp(y_k) \quad (5)$$

where $f_k(y)$ is the predicted value of the softmax function for each category; C is the number of categories.

2.2. Residual Neural Network

As the deep neural network training process may lead to network degradation, He et al. [42] proposed a Residual Neural Network (ResNet), which builds a model by stacking residual blocks. The input of the residual block is x . After two convolutional layers, the basic mapping function $H(x)$ is acquired. The residual block introduces a skip layer connection to reconstruct the learning process of stacked network layers, so that the residual mapping function $F(x)$ of the network layer can fit $H(x) - x$, where the expression of $F(x)$ is as follows:

$$F(x) = W_2(ReLU(W_1(x))) \quad (6)$$

where W_1 and W_2 in represent the convolutional layer 1 and convolutional layer 2 in the residual block, and $ReLU$ represents the nonlinear activation function.

2.3. ECNN Model

As the standard CNN model still maintains the black-box feature, it not only cannot capture the features of different scales but also ignores the unique contribution of special data segments. In order to solve the problem, we start from three aspects: raw vibration signal using Pyramidal Dilated Convolution, ResNet-FCF block construction, and the general procedure of the proposed method.

Compared with CNN, this model adopts different architectural approaches in processing original signals, feature extraction and weight allocation of signal channels, as shown in Figure 2. The ECNN model uses dilated convolution to replace the pooling layer to avoid the loss of important features and has a larger receptive field under the same parameters with more useful features extracted. After the dilated convolution operation, the global average pooling (GAP) is used to obtain the descriptor of each feature channel. Then the local channel interaction and the global channel interaction based on the residual network are used for weight redistribution, strengthening the crucial features and suppressing the irrelevant features.

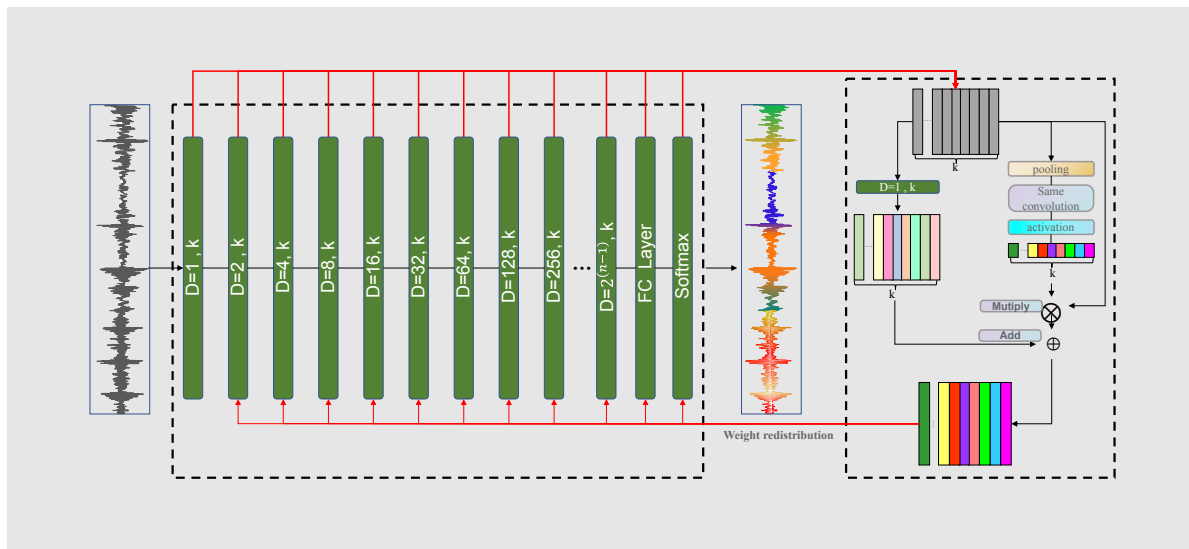


Figure 2. ECNN model of structures.

2.3.1. Pyramidal Dilated Convolution

Dilated convolution has huge advantages over standard convolution [43]. In this study, Pyramidal Dilated Convolution (PDC) is proposed, which has an adaptive receptive field with the same parameters. When a large amount of information is acquired, the parameters can be kept unchanged, resulting in a decrease in computing resources. Assuming that $x(a)$ represents the one-dimensional input, A shows the data feature size, the dilation rate is r , the 1D convolution kernel is $w(i)$, and $y(a)$ represents the output feature of data operation. The relationship equation between them is as follows:

$$y(a) = \sum_{i=1}^A x(a + r \times i) \times w(i). \tag{7}$$

The Pyramidal Dilated Convolution can not only expand the reception field of the original network but also compensate for the loss caused by under-sampling, as demonstrated in Figure 3. Therefore, it is feasible to replace standard convolution in neural networks with Pyramidal Dilated Convolution. Depending on the PDC with different dilation rates, the receptive field becomes enlarged, and in the meantime, features of data segments permit more access.

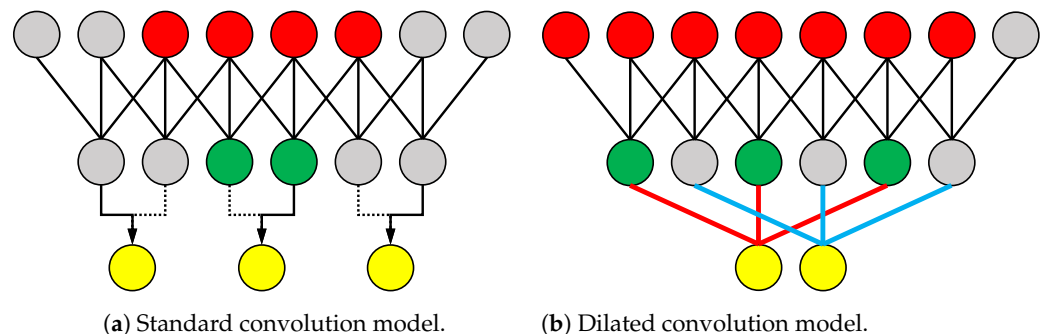


Figure 3. Comparison of two convolutional networks.

Under the condition that the size of the convolution kernel is the same, the receptive field of the Pyramidal Dilated Convolution in the i -th convolutional layer rises exponentially. The dilation rate d and the size of the convolution kernel jointly determine the receptive field size of the i -th Pyramidal Dilated Convolutional network. Equation (8) is the receptive field of dilation:

$$S_1 = kernel_size \tag{8}$$

$$S_i = S_{i-1} + (kernel_size_i - 1) * d_i \tag{9}$$

where S_i stands for the receptive field of the i -th layer, and $kernel_size_i$ and d_i represent the size of the convolution kernel and dilation rate of the i -th layer network, respectively.

2.3.2. ResNet-FCF Block

In recent years, it has been proved that channel attention exists with great potential in improving the performance of convolutional neural networks. In this study, we propose a novel ResNet-FCF block with the structure shown in Figure 4.

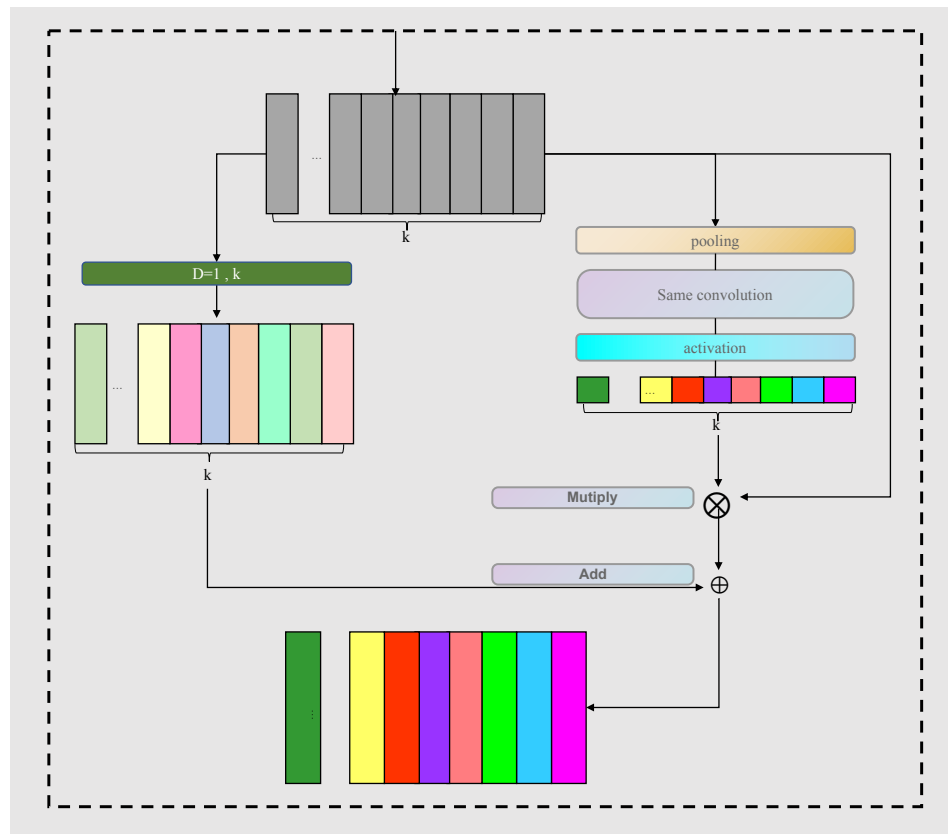


Figure 4. The construction of a ResNet-FCF block.

The specific working process of the module is initially to perform GAP [44] after convolution transformation, whose purpose is to obtain the global information of each channel. The statistic $z \in \mathbb{R}^C$ is generated by shrinking the dimension $H \times W$ of the input signal, and the p -th element of z can be expressed as:

$$z_p = F_{sq}(u_p) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_p(i, j) \tag{10}$$

where z_p can be interpreted as a collection of local features, the statistical information of these local features can express the entire signal and have a global receptive field. The statistic z does not require dimensionality reduction, and the attention weight of each channel can be obtained in the following way:

$$\rho = \sigma(\mathbf{W}_k z) \tag{11}$$

where ρ is the weight, and W_k contains $k \times C$ parameters, which is defined as:

$$\begin{bmatrix} w_1^1 & \cdots & w_1^k & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w_2^2 & \cdots & w_2^{k+1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w_C^{C-k+1} & \cdots & w_C^C \end{bmatrix} \quad (12)$$

σ represents the sigmoid [45] nonlinear activation function:

$$\sigma(x) = 1/(1+e^{-x}) \quad (13)$$

Equation (11) avoids complete independence between different channels and ensures efficiency and effectiveness while realizing partial cross-channel interaction. The weight of z_j is calculated only by considering the interaction between z_j and its k neighboring elements:

$$\rho_i = \sigma\left(\sum_{j=1}^k w_i^j y_i^j\right), y_i^j \in \Omega_i^k \quad (14)$$

where Ω_i^k represents the set of k channels adjacent to y_i . In order to reduce the complexity of the model, all channels can share the same parameters, namely:

$$\rho_i = \sigma\left(\sum_{j=1}^k w^j y_i^j\right), y_i^j \in \Omega_i^k \quad (15)$$

Equation (11) can be further simplified as a one-dimensional convolution operation:

$$\varphi = \sigma\left(\text{Conv}_1^k(z)\right) \quad (16)$$

where Conv_1^k represents the one-dimensional convolution, and k is the size of the corresponding convolution kernel. It is verified through experiments that the value of k is related to the number of channels C . The simplest mapping is a linear function.

However, the relationships characterized by linear functions are excessively limited. Exponential functions are often used to deal with this nonlinear mapping relationship. The expression is as follows:

$$C = 2^{(\gamma^k - b)} \quad (17)$$

the number of channels is C , the expression of k can be inversely deduced as

$$k = \left(\frac{\log_2 C}{\gamma} + \frac{b}{\gamma}\right) \quad (18)$$

in the equation: γ, b are adjustment parameters, which can be adjusted according to specific conditions.

$F_{sc}(\cdot)$ multiplies the input x by ρ and maps it to the output with the same dimension. The output is defined as follows:

$$F_{sc}(x, \rho) = Y \quad (19)$$

The process uses a local feature learning scheme to assign weights, then the skip part of the residual network is changed to convolution, which builds the global feature interaction. The output U of the residual block is defined as follows:

$$U = \text{Conv}_2^1(X) + Y \quad (20)$$

where Conv_2^1 represents the convolution whose kernel size is 1.

2.4. Intelligent Fault Diagnosis Process

In this study, a novel ECNN for intelligent fault diagnosis is developed. The specific framework of diagnosis is shown in Figure 5. The general process is summarized as follows:

- Step 1: Collect vibration signal data from equipment.
- Step 2: The raw signal is divided. Firstly, a certain proportion of the former is the training signal, and the rest is the test signal. Then, several samples are randomly selected from the corresponding signals to form a training set and a test set.
- Step 3: Construct an ECNN network model and select appropriate parameters.
- Step 4: Bring the training samples into the model for training. If the model does not converge, return to the previous step to redesign.
- Step 5: The performance of the proposed method is verified using test samples.

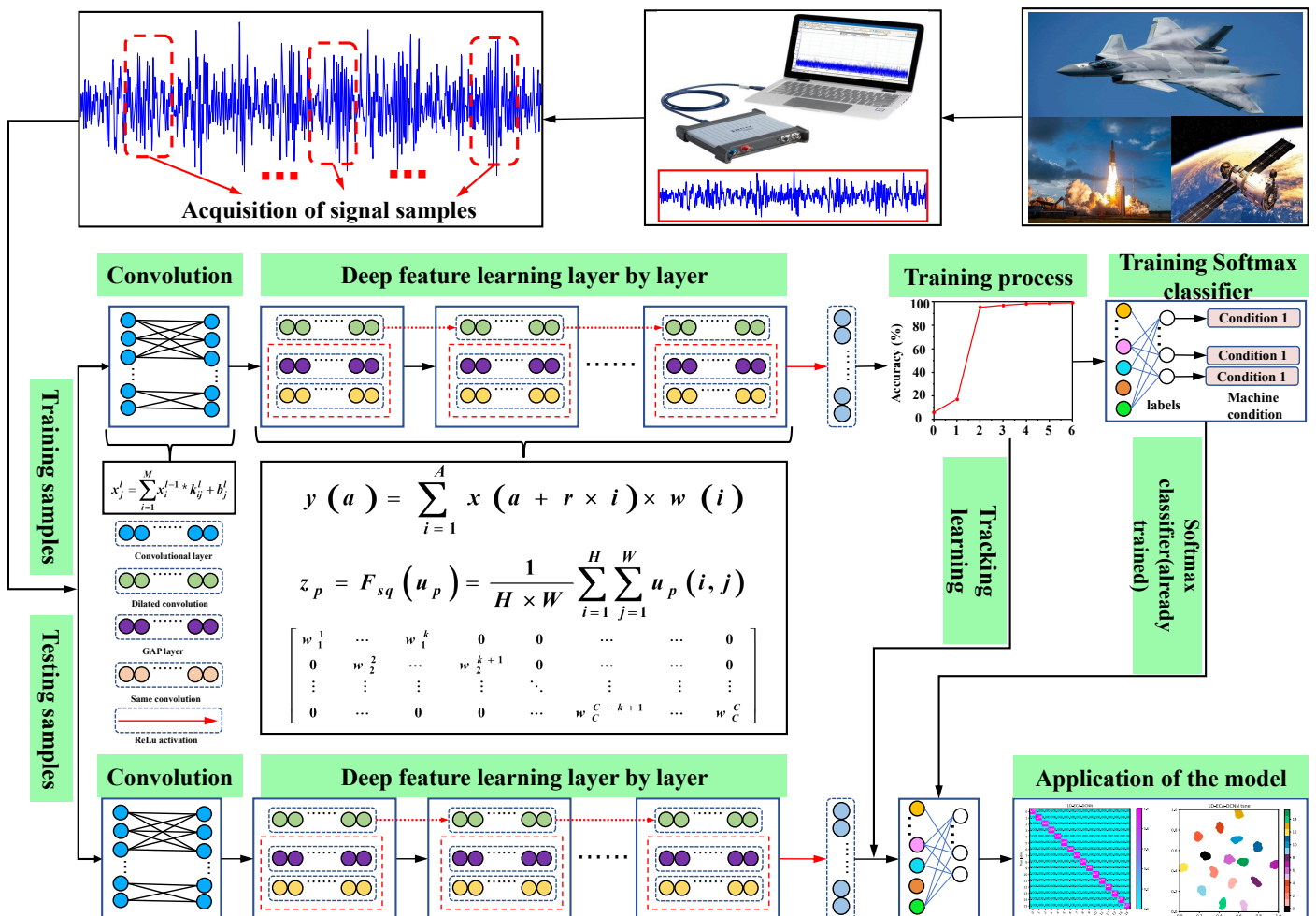


Figure 5. The structure of intelligent fault diagnosis.

3. Results

In this section, the Windows10 operating system is used. The Random Access Memory (RAM) and graphics process unit are 16GB and GeForce GTX 1060, respectively. The ECNN test model was built and trained under the Keras framework. Keras is an open-source artificial neural network library written in Python, which can be used as a high-level application interface for Tensorflow, Microsoft-CNTK and Theano to design, debug, evaluate, apply and visualize deep learning models. Three examples are given to verify that the ECNN fault diagnosis model can extract a considerable number of data features with equivalent calculations and assign much weight to essential features and little to irrelevant features. In the experiment, three different data sets are used to verify its effectiveness:

Case Western Reserve University (CWRU) rolling bearing, National Aeronautics and Space Administration (NASA) rolling bearing, and the Permanent Magnet Synchronous Motor (PMSM) of our lab, respectively.

3.1. The CWRU Rolling Bearing Example

The CWRU bearing data set is recognized as an authoritative fault diagnosis standard data set. In order to objectively compare the method with others, this paper selects the CWRU bearing data set for algorithm verification. In the experiment, the drive end data are selected. The bearing has three types of fault conditions: the inner ring, outer ring and rolling element. The depth and location of each type of fault are different, as shown in Table 1. There are four hundred samples for each condition. The load of this experiment is variable and can be divided into HP1, HP2 and HP3 load data. The experimental data contains 16 kinds of conditions. The signals are shown in Figure 6.

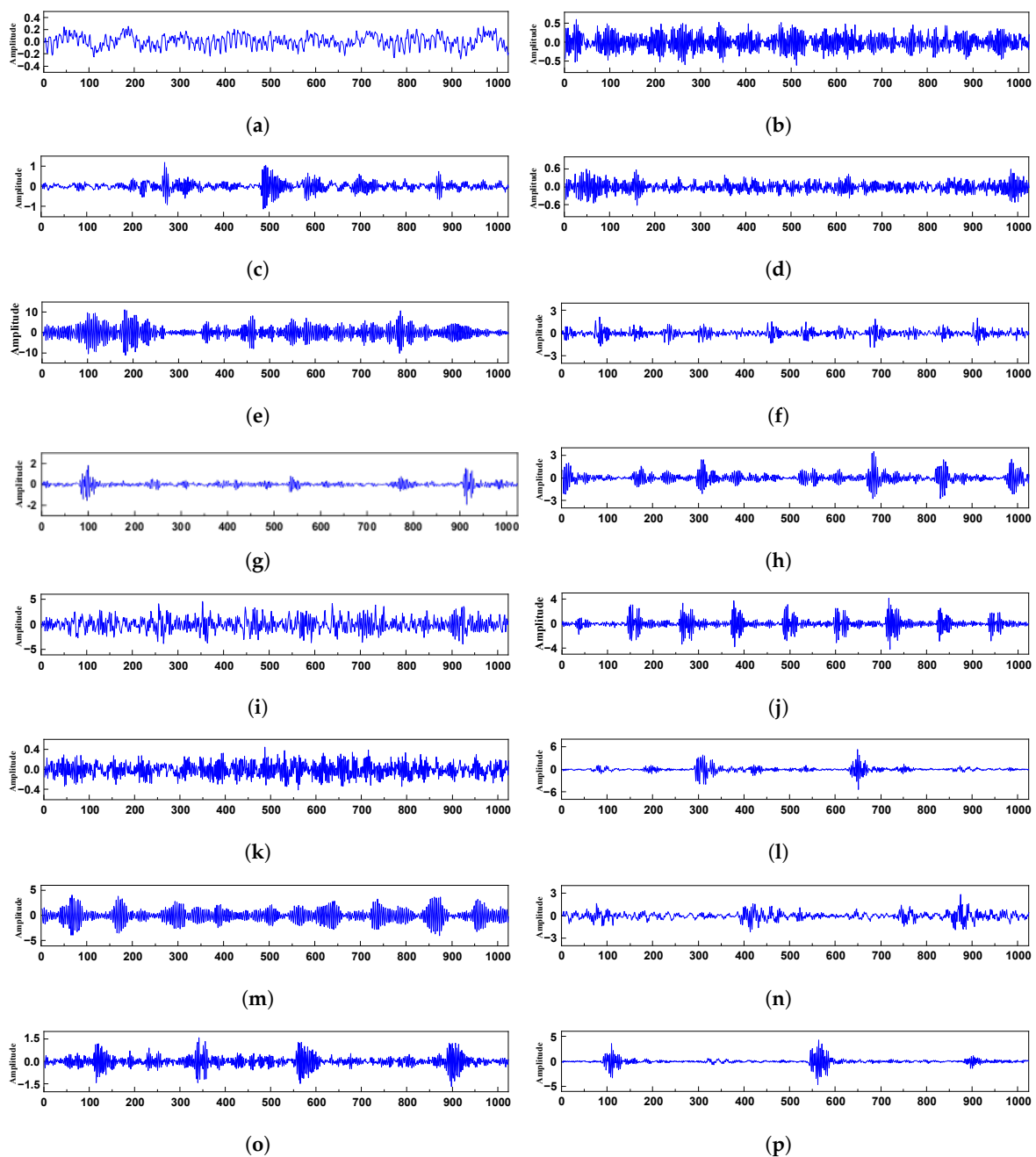


Figure 6. Raw waveforms of the bearing conditions.

3.1.1. Model Structure

In the experiment, six traditional methods and four convolutional neural networks were selected for the comparison. The conventional method extracts 11 time-domain features and 13 frequency-domain features from the samples and then classifies them using BPNN and SVM. More details about the eleven TD features and thirteen FD features can be seen in [46,47], respectively. CNN, dilated convolution and SE (Squeeze-and-Excitation) CNN are taken as the comparative experimental models of deep convolutional neural networks. The first layer of the ECNN model uses a one-dimensional way for data processing, and then different dilated rates are set for the other eight layers. This process can replace the pooling layer in standard CNN so more valuable features can be extracted while the receptive field increases. The designed ResNet-FCF block assigns different weights to feature channels after each layer. The model parameters of the deep learning network ECNN are shown in Table 2.

Table 1. Description of the bearing operation conditions.

Bearing Operation Conditions	Fault Diameter	Size of Training	Testing Samples	Label
Normal	-	300	100	0
Ball fault	0.007	300	100	1
Ball fault	0.014	300	100	2
Ball fault	0.021	300	100	3
Ball fault	0.028	300	100	4
Inner race fault	0.007	300	100	5
Inner race fault	0.014	300	100	6
Inner race fault	0.021	300	100	7
Inner race fault	0.028	300	100	8
Outer race fault 6 o'clock	0.007	300	100	9
Outer race fault 6 o'clock	0.014	300	100	10
Outer race fault 6 o'clock	0.021	300	100	11
Outer race fault 3 o'clock	0.007	300	100	12
Outer race fault 3 o'clock	0.021	300	100	13
Outer race fault 12 o'clock	0.007	300	100	14
Outer race fault 12 o'clock	0.021	300	100	15

Table 2. Parameters of the ECNN model.

Network Layer	Size/Stride	Kernels	Dilate Factor	Output Shape	Receptive Field	Padding
Input	—	—	—	1024 × 1	1	—
Conv 1	3 × 1/1 × 1	64	1	1024 × 64	3	Yes
ResNet-FCF	-	-	-	1024 × 64	3	-
Conv 2	3 × 1/1 × 1	64	2	1020 × 64	7	No
ResNet-FCF	-	-	-	1020 × 64	7	-
Conv 3	3 × 1/1 × 1	64	4	1012 × 64	15	No
ResNet-FCF	-	-	-	1012 × 64	15	-
Conv 4	3 × 1/1 × 1	64	8	996 × 64	31	No
ResNet-FCF	-	-	-	996 × 64	31	-
Conv 5	3 × 1/1 × 1	64	16	964 × 64	63	No
ResNet-FCF	-	-	-	964 × 64	63	-
Conv 6	3 × 1/1 × 1	64	32	900 × 64	127	No
ResNet-FCF	-	-	-	900 × 64	127	-
Conv 7	3 × 1/1 × 1	64	64	772 × 64	255	No
ResNet-FCF	-	-	-	772 × 64	255	-
Conv 8	3 × 1/1 × 1	64	516 × 64	511	No	-
ResNet-FCF	-	-	-	516 × 64	511	-
Conv 9	3 × 1/1 × 1	64	256	4 × 64	1023	No
ResNet-FCF	-	-	-	4 × 64	1023	-
Flatten	—	—	—	128	—	—
FC	128	—	—	128	—	—
SoftMax	16	—	—	16	—	—

3.1.2. Diagnosis Results of Different Methods

In order to obtain more accurate results, the experiment was divided into ten groups. The results show that the performance of BPNN and SVM depends on feature extraction to a large extent. When the selected input becomes sensitive features, the diagnostic results are further improved to 79.63% and 87.14%, respectively, but this is expensive and time-consuming. The accuracy of the proposed method is 99.12% higher than that of the nine methods, as shown in Table 3 and Figure 7.

Table 3. Parameters of the ECNN model.

Network Layer	Size/Stride	Kernels	Dilate Factor	Output Shape	Receptive Field	Padding
Input	—	—	—	1024 × 1	1	—
Conv 1	3 × 1/1 × 1	64	1	1024 × 64	3	Yes
ResNet-FCF	-	-	-	1024 × 64	3	-
Conv 2	3 × 1/1 × 1	64	2	1020 × 64	7	No
ResNet-FCF	-	-	-	1020 × 64	7	-
Conv 3	3 × 1/1 × 1	64	4	1012 × 64	15	No
ResNet-FCF	-	-	-	1012 × 64	15	-
Conv 4	3 × 1/1 × 1	64	8	996 × 64	31	No
ResNet-FCF	-	-	-	996 × 64	31	-
Conv 5	3 × 1/1 × 1	64	16	964 × 64	63	No
ResNet-FCF	-	-	-	964 × 64	63	-
Conv 6	3 × 1/1 × 1	64	32	900 × 64	127	No
ResNet-FCF	-	-	-	900 × 64	127	-
Conv 7	3 × 1/1 × 1	64	64	772 × 64	255	No
ResNet-FCF	-	-	-	772 × 64	255	-
Conv 8	3 × 1/1 × 1	64	516 × 64	511	No	-
ResNet-FCF	-	-	-	516 × 64	511	-
Conv 9	3 × 1/1 × 1	64	256	4 × 64	1023	No
ResNet-FCF	-	-	-	4 × 64	1023	-
Flatten	—	—	—	128	—	—
FC	128	—	—	128	—	—
SoftMax	16	—	—	16	—	—

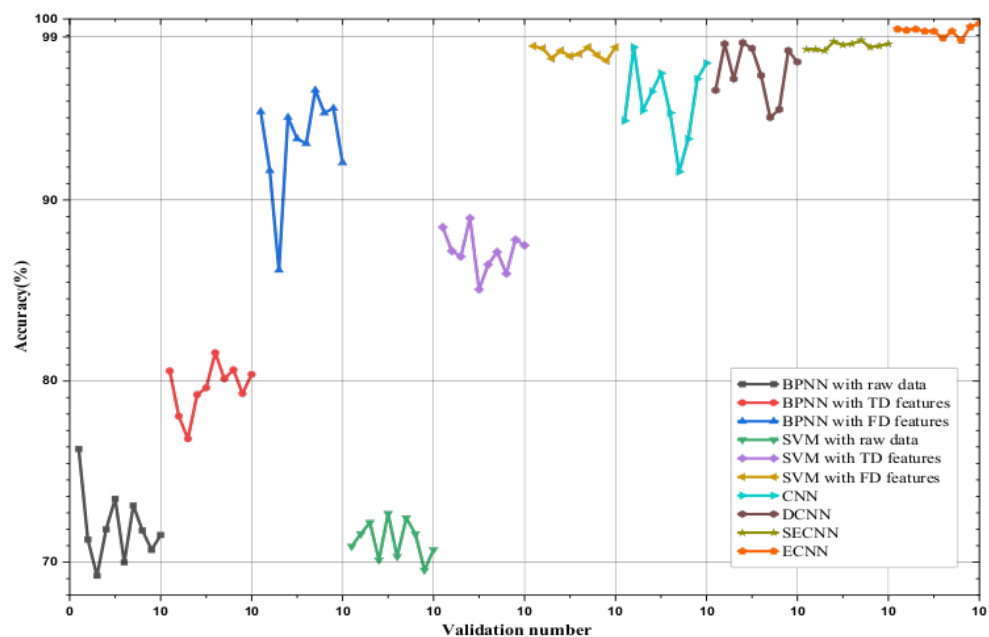


Figure 7. Diagnosis results of the ten trials using different methods.

3.1.3. Visualization of the Fault Diagnosis Process

ECNN is set for fault diagnosis, in which nine layers exist. Each layer initially uses different dilation rates for feature extraction and then utilizes the ResNet-FCF block to assign appropriate weights for the extracted features effectively. Thereby the informative data segments will receive more attention and their corresponding channel weights increase. This paper takes advantage of the t-SNE dimensionality reduction technique to visualize the features in each layer. The diagram of the structure is shown in Figure 8.

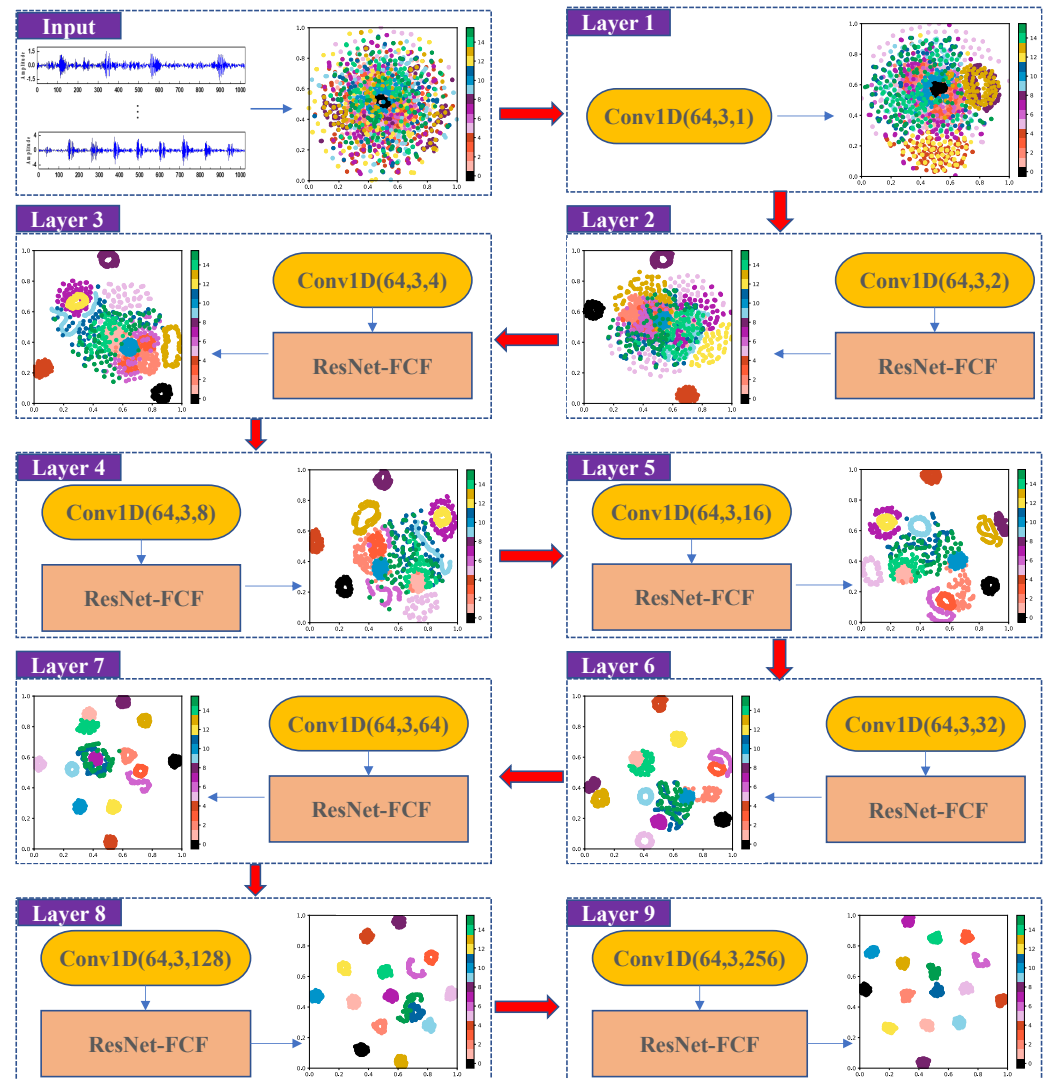


Figure 8. The proposed method fault classification results.

3.1.4. Weight Comparison of ECNN and SECNN

To explore the role of ResNet-FCF in every convolutional layer of ECNN, we visualize the weights generated by each convolutional layer and then compare the effects of ResNet-FCF and SE blocks at each layer in detail. As shown in Figure 9, the results indicate that when different types of samples are input, the weight generated by the fault diagnosis model is more related to the category information. The combination of local cross-channel and global cross-channel interactions in the ResNet-FCF block significantly improves the ability of channel attention.

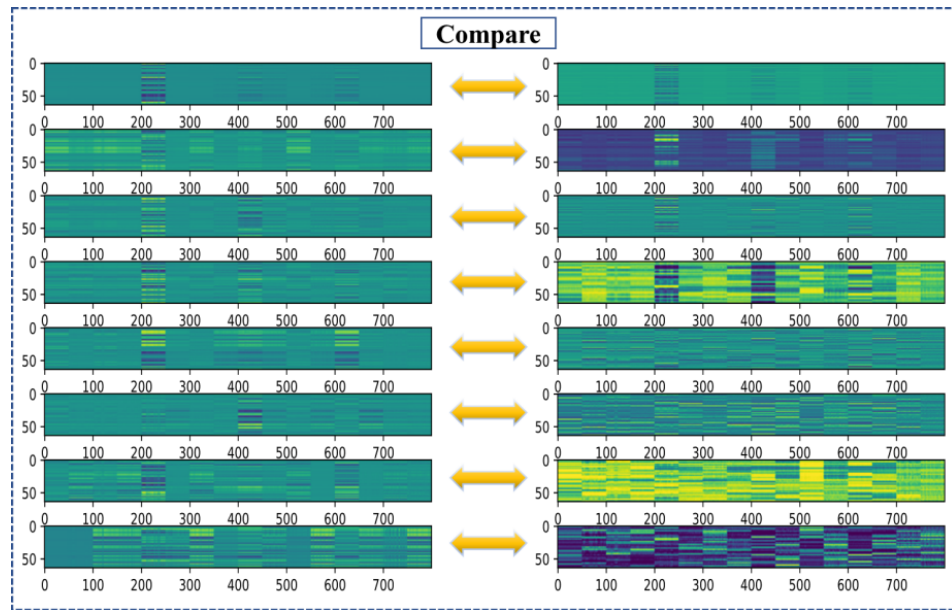


Figure 9. Visual comparison of weights for ResNET-FCF and SE blocks.

3.1.5. The Influence of Hyperparameters

The structural parameters of the model are determined by experiments. Two key parameters k (convolution kernel size) and C (number of channels) are selected through cross-validation, as shown in Figure 10. The results of the 3D hyperparametric graph show that the accuracy of the model is the highest when the size of the convolution kernel is sixty-four, and the number of channels is three.

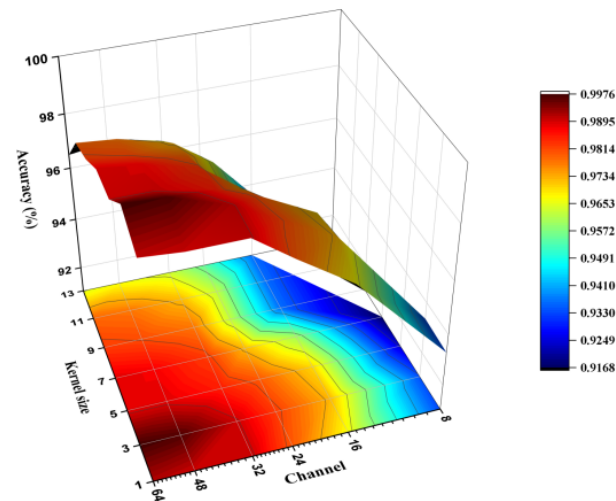


Figure 10. The effect of C and k on accuracy.

3.1.6. The Influence of Different Segmentation Ratios

In the experiment, with the increased number of training samples, the accuracy and experiment time of each network structure also increase. As seen in Table 4, under the conditions of different sample segmentation ratios, the ECNN fault diagnosis model has achieved the greatest results.

3.1.7. The Influence of Different SNR

To further research the performance of the model, various levels of white noise were added to the traditional learning model and the deep learning model, respectively. The results show that the traditional methods are less competitive compared with CNN in terms

of noise-resistance. The model behaves brilliantly in terms of robustness compared with other deep learning networks. Table 5 shows the accuracy of each model under noise conditions. Figure 11 shows the noise-resistance curve.

Table 4. Accuracy of different segmentation ratios.

Methods		BP-r	BP-T	BP-F	SVM-r	SVM-T	SVM-F	CNN	DCNN	SECNN	ECNN
350/50	Correct samples	12816	11410	15037	9787	13795	15717	15782	15640	15818	15818
	Accuracy	82.96%	75.67%	95.33%	71.88%	88.56%	98.36%	98.17%	99.02%	98.98%	99.34%
	Accuracy std	1.05%	2.81%	0.36%	0.55%	1.11%	0.27%	0.38%	0.32%	0.70%	0.40%
	Time	6.95	6.29	7.29	12.37	2.56	1.86	30.83	59.97	44.69	71.68
325/75	Correct samples	12816	11410	15037	9787	13795	15717	15782	15640	15818	15818
	Accuracy	80.72%	74.58%	94.61%	71.25%	87.58%	98.33%	97.75%	98.87%	98.79%	99.27%
	Accuracy std	0.84%	3.50%	0.34%	1.47%	1.00%	0.31%	1.21%	0.60%	0.58%	0.42%
	Time	6.31	6.21	6.93	12.17	2.5	1.81	28.38	58.69	41.7	70.87
300/100	Correct samples	12816	11410	15037	9787	13795	15717	15782	15640	15818	15818
	Accuracy	79.63%	71.91%	93.17%	71.20%	87.14%	98.15%	95.47%	97.07%	98.72%	99.12%
	Accuracy std	1.38%	1.99%	2.84%	1.05%	1.17%	0.29%	2.08%	1.50%	0.41%	0.37%
	Time	6.1	6.14	6.8	12.09	2.4	1.78	28.05	52.56	38.36	62.89
275/125	Correct samples	12816	11410	15037	9787	13795	15717	15782	15640	15818	15818
	Accuracy	78.48%	69.56%	92.98%	71.07%	86.12%	98.13%	95.40%	97.01%	96.95%	98.79%
	Accuracy std	1.27%	2.00%	1.66%	0.56%	1.28%	0.17%	2.64%	0.49%	2.83%	0.33%
	Time	5.89	5.63	6.65	11.5	2.38	1.76	25.19	52.14	37.78	62.46
250/150	Correct samples	12816	11410	15037	9787	13795	15717	15782	15640	15818	15818
	Accuracy	77.85%	69.02%	92.42%	70.60%	86.08%	98.10%	95.02%	96.82%	96.72%	98.58%
	Accuracy std	0.63%	2.70%	3.16%	1.00%	2.08%	0.19%	5.68%	2.07%	1.03%	1.02%
	Time	5.74	5.59	6.32	11.26	2.31	1.74	24.16	45.59	34.93	54.35
225/175	Correct samples	12816	11410	15037	9787	13795	15717	15782	15640	15818	15818
	Accuracy	72.34%	67.82%	89.83%	70.37%	86.08%	98.07%	93.61%	96.66%	96.56%	97.87%
	Accuracy std	1.50%	0.57%	3.48%	0.50%	1.57%	0.23%	3.70%	0.58%	2.25%	0.41%
	Time	5.62	5.48	6.27	10.81	2.19	1.71	22.89	44.74	33.34	54.22
200/100	Correct samples	12816	11410	15037	9787	13795	15717	15782	15640	15818	15818
	Accuracy	71.35%	66.72%	88.43%	69.07%	85.37%	98.04%	93.48%	96.24%	95.67%	97.23%
	Accuracy std	0.94%	3.74%	3.29%	1.59%	0.31%	0.38%	6.77%	1.33%	5.45%	1.82%
	Time	5.37	5.1	6.14	1.18	2.16	1.68	19.57	38.64	32.81	45.94

In this table, the accuracy reflects the average accuracy; the time represents the average time and units in seconds (s).

Table 5. The accuracy of each method under different SNRs.

Method	SNR(dB)							
	3	2	1	0	-1	-2	-3	-4
BPNN with raw data	64.13%	60.81%	57.25%	54.25%	45.63%	43.81%	34.43%	30.31%
BPNN with TD features	77.69%	77.63%	76.37%	76.00%	75.88%	74.63%	74.18%	74.06%
BPNN with FD features	90.94%	90.81%	87.88%	83.63%	82.81%	79.00%	73.87%	72.68%
SVM with raw data	69.63%	68.25%	66.56%	66.25%	64.88%	64.06%	62.93%	62.68%
SVM with TD features	84.75%	84.69%	84.25%	84.19%	84.00%	83.31%	81.18%	80.87%
SVM with FD features	96.44%	95.56%	94.38%	92.44%	92.00%	90.12%	85.31%	83.43%
CNN	96.75%	95.38%	93.56%	87.93%	85.25%	83.06%	79.25%	77.87%
DCNN	97.87%	97.75%	97.37%	96.81%	95.69%	92.31%	89.37%	80.87%
SECNN	97.69%	97.31%	97.06%	96.75%	95.50%	92.06%	88.58%	78.25%
ECNN	98.56%	98.06%	98.00%	97.75%	96.74%	94.44%	92.12%	87.43%

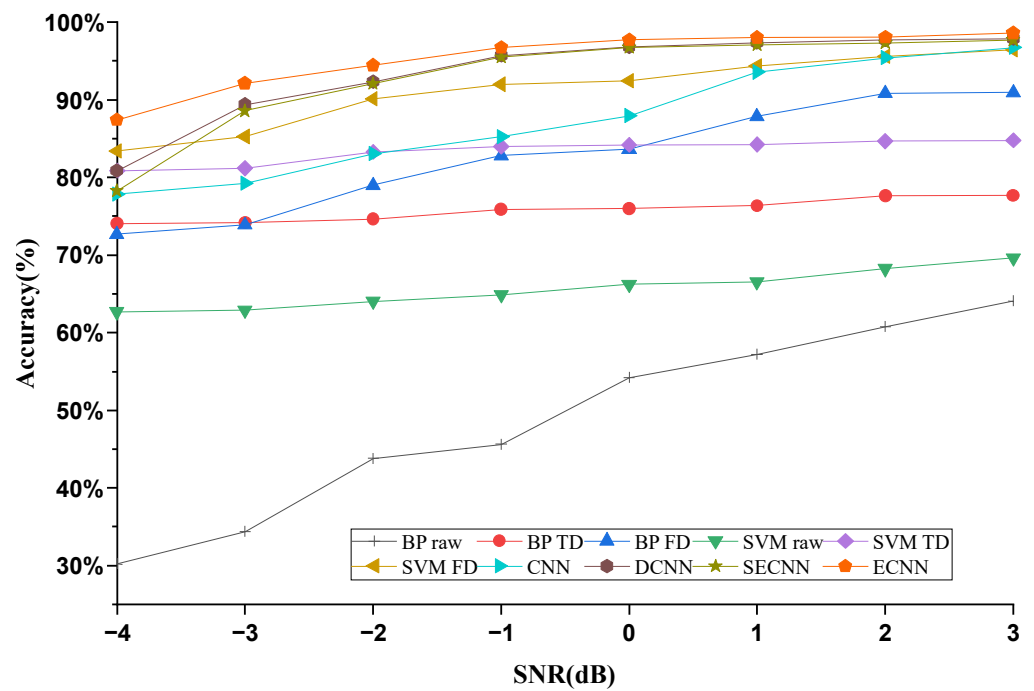


Figure 11. Accuracy curve of each method under different SNRs.

3.1.8. The Influence of Variable Load

In order to analyze the generalization ability of these models, every method is required to be trained under different loads, and another load will be used as a test set. The experiment results are shown in Figure 12 and Table 6.

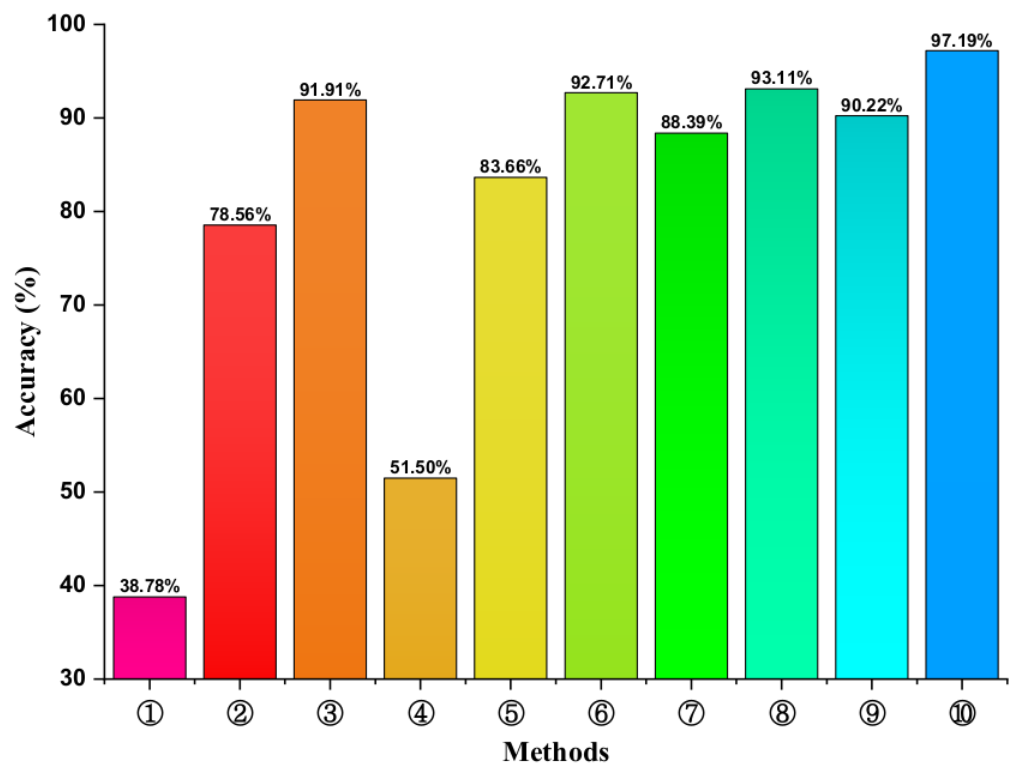


Figure 12. Average accuracy under different loads.

Table 6. Variable load accuracy rate of each method.

Methods	BP-r	BP-T	BP-F	SVM-r	SVM-T	SVM-F	CNN	DCNN	SECNN	ECNN	
1→2	1	45.77%	78.30%	96.58%	50.73%	85.20%	94.53%	96.80%	98.42%	99.64%	99.23%
	2	42.95%	76.77%	96.69%	48.85%	86.24%	94.93%	96.12%	98.91%	98.83%	99.98%
	3	44.03%	77.38%	96.83%	49.29%	86.02%	94.86%	93.32%	98.83%	96.95%	99.97%
	4	43.39%	79.28%	96.67%	51.25%	85.23%	94.64%	94.85%	99.94%	97.17%	99.98%
	5	42.47%	80.92%	97.68%	50.57%	84.94%	94.66%	93.45%	99.65%	94.74%	99.81%
	Avg Std	43.72% 1.28%	78.53% 1.64%	96.89% 0.45%	50.14% 1.02%	85.52% 0.57%	94.72% 0.16%	94.91% 1.55%	99.15% 0.62%	97.47% 1.89%	99.79% 0.33%
2→1	1	38.83%	79.70%	94.97%	56.83%	83.51%	94.39%	95.00%	97.35%	92.77%	97.90%
	2	39.80%	79.94%	94.59%	54.23%	84.01%	94.43%	90.03%	98.60%	94.59%	98.15%
	3	38.51%	78.41%	94.16%	51.10%	84.39%	94.32%	95.85%	98.44%	94.85%	98.69%
	4	38.53%	82.01%	94.68%	52.41%	85.39%	94.37%	92.03%	97.38%	94.51%	98.56%
	5	39.43%	80.26%	95.06%	53.72%	80.98%	94.18%	92.90%	96.94%	93.97%	97.95%
	Avg Std	39.02% 0.57%	80.06% 1.29%	94.69% 0.35%	53.66% 2.15%	83.65% 1.65%	94.34% 0.10%	93.16% 2.33%	97.74% 0.73%	94.14% 0.83%	98.25% 0.36%
1→3	1	38.79%	80.07%	85.59%	55.02%	83.85%	91.23%	87.61%	90.83%	91.35%	98.94%
	2	40.56%	78.30%	88.83%	52.90%	83.44%	91.09%	92.39%	95.94%	90.17%	98.97%
	3	40.20%	77.68%	85.34%	57.02%	82.73%	90.33%	83.51%	98.33%	95.06%	98.04%
	4	37.86%	83.54%	83.08%	55.59%	83.00%	92.29%	89.66%	97.53%	88.27%	99.29%
	5	40.03%	80.76%	87.72%	54.14%	87.12%	91.12%	92.77%	97.06%	90.99%	99.15%
	Avg Std	39.49% 1.13%	80.07% 2.31%	86.11% 2.24%	54.93% 1.55%	84.03% 1.78%	91.21% 0.70%	89.19% 3.81%	95.94% 2.98%	91.17% 2.48%	98.88% 0.49%
3→1	1	34.26%	77.78%	91.22%	51.86%	83.49%	88.35%	82.34%	84.30%	81.14%	91.83%
	2	34.83%	76.33%	87.23%	50.11%	81.99%	88.59%	82.34%	83.68%	87.12%	91.70%
	3	34.09%	70.04%	90.26%	49.89%	81.46%	91.18%	80.57%	85.81%	80.22%	90.93%
	4	33.46%	77.97%	92.82%	48.01%	83.41%	92.96%	80.74%	87.09%	79.99%	92.19%
	5	33.84%	77.35%	87.58%	49.83%	82.95%	89.17%	81.64%	86.74%	85.56%	91.97%
	Avg Std	34.10% 0.51%	75.89% 3.33%	89.82% 2.39%	49.94% 1.37%	82.66% 0.90%	90.05% 1.97%	81.53% 0.85%	85.52% 1.49%	82.81% 3.30%	91.72% 0.48%
2→3	1	39.27%	79.50%	89.14%	58.38%	85.37%	94.92%	85.15%	90.42%	90.61%	98.93%
	2	41.24%	80.24%	89.41%	55.21%	85.73%	94.72%	89.98%	93.07%	91.10%	99.62%
	3	41.37%	77.46%	87.18%	53.96%	85.42%	94.34%	86.66%	97.63%	92.00%	99.86%
	4	41.05%	82.84%	87.59%	55.54%	84.66%	94.21%	86.27%	91.68%	93.23%	98.86%
	5	41.93%	82.73%	90.56%	52.73%	85.43%	95.10%	90.29%	87.17%	89.76%	98.55%
	Avg Std	40.97% 1.01%	80.56% 2.28%	88.78% 1.38%	55.16% 2.11%	85.32% 0.40%	94.66% 0.38%	87.67% 2.32%	91.99% 3.83%	91.34% 1.33%	99.16% 0.55%
3→2	1	35.31%	81.12%	94.08%	47.52%	81.98%	91.42%	82.99%	87.26%	81.12%	95.47%
	2	35.05%	74.39%	94.13%	45.11%	81.14%	91.09%	83.62%	86.66%	85.70%	95.36%
	3	35.42%	71.81%	96.18%	43.88%	81.19%	90.69%	84.34%	87.47%	81.27%	95.11%
	4	35.45%	77.67%	96.28%	44.65%	80.22%	92.14%	84.50%	87.36%	86.92%	95.33%
	5	35.67%	76.22%	95.11%	44.73%	79.40%	91.02%	83.87%	92.68%	87.11%	95.48%
	Avg Std	35.38% 0.22%	76.24% 3.50%	95.15% 1.06%	45.18% 1.39%	80.79% 0.99%	91.27% 0.55%	83.86% 0.61%	88.29% 2.47%	84.42% 3.00%	95.35% 0.15%

3.2. The NASA Rolling Bearing Example

In this section, the NASA dataset is further used to demonstrate the superiority of the proposed method in feature extraction and contributions from different data segments. From the beginning to the end of the data collection, there are four cases in the data set: normal, inner ring failure, outer ring failure and rolling element failure. The test time of normal, inner ring and rolling elements are all thirty-five days, with the exception of thirty days for the test of the outer ring. The data set extracted in this experiment was collected after frequent failures, and the data of the last five days were taken as samples shown in Figure 13 and Table 7.

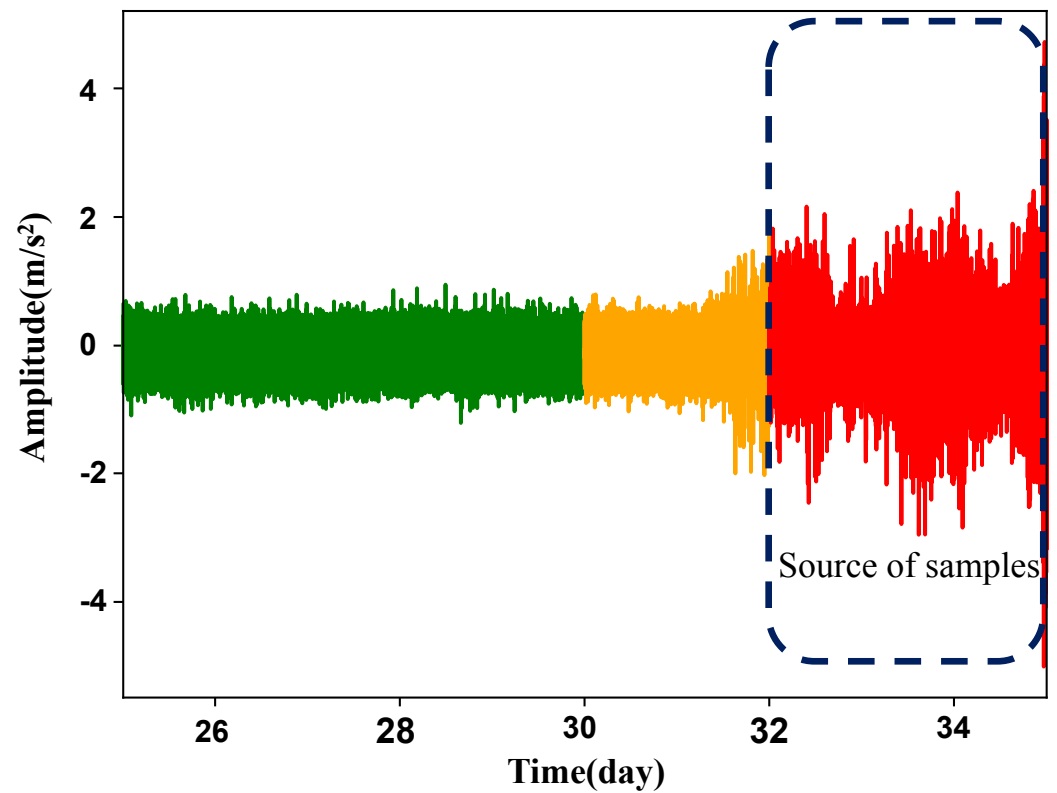


Figure 13. Collection of NASA datasets.

Table 7. Data description.

Conditions	Days	Train Samples	Test Samples	Dataset	Bearing	Label
Normal	25–35	300	100	No.1	No.1	0
Inner face	25–35	300	100	No.1	No.3	1
Outer face	20–30	300	100	No.3	No.3	2
Roller element	25–35	300	100	No.1	No.4	3

This experiment obtained 300 training samples and 100 testing samples from each condition, and the length of each sample was set to 1024. Ten experiments were conducted for each one of the four deep learning networks. The experiment results are shown in Figure 14 and Table 8. Compared with other deep learning methods, the proposed method has obvious advantages and enables to accurately identify different types of faults.

Table 8. Diagnosis results of the ten trials using different methods.

Methods	Total Sample	Correct Sample	Average Testing Accuracy
CNN	4000	3815	95.38% ± 0.54%
DCNN	4000	3924	98.10% ± 0.42%
SECNN	4000	3913	97.83% ± 0.50%
ECNN	4000	3982	99.55% ± 0.33%

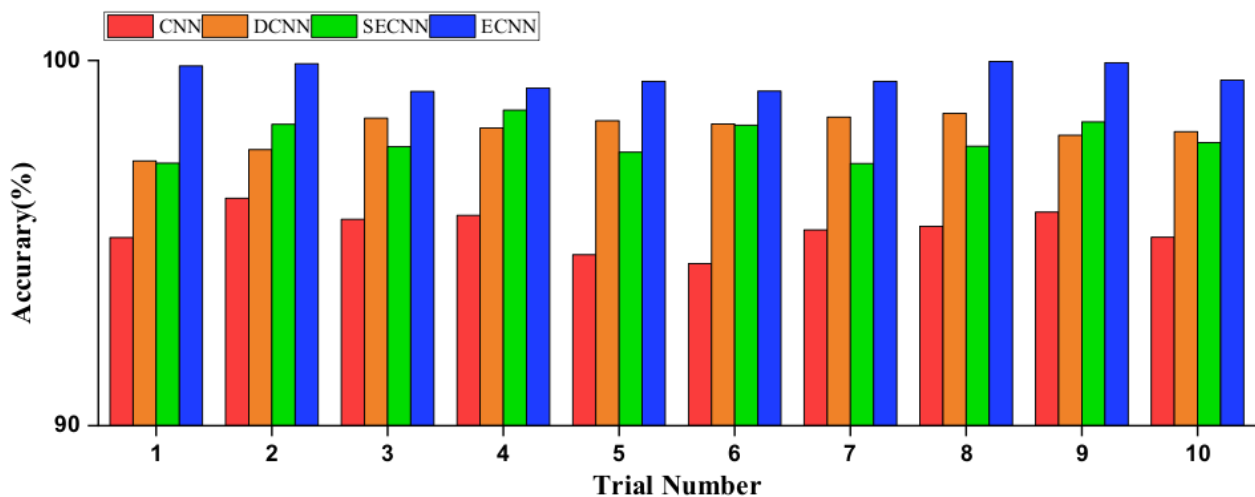


Figure 14. Diagnosis results of the ten trials using different methods.

The following conclusions can be drawn: different changes in load have different influences on signals. Compared with other deep learning methods, the ECNN fault diagnosis model has a high average accuracy rate that reaches 97.19% under variable working conditions. The results show that the model has a strong generalization ability and domain adaptability.

3.3. The Example of PMSM

Further study on the electrical systems of Permanent Magnet Synchronous Motors (PMSM) will be carried out to verify the transferability and extensibility of the ECNN-based method.

The experimental platform comprises the following parts: upper computer, electromechanical actuator and its controller, sensor, data acquisition system, loading system and power supply system. The structure of the experimental platform is shown in Figure 15.

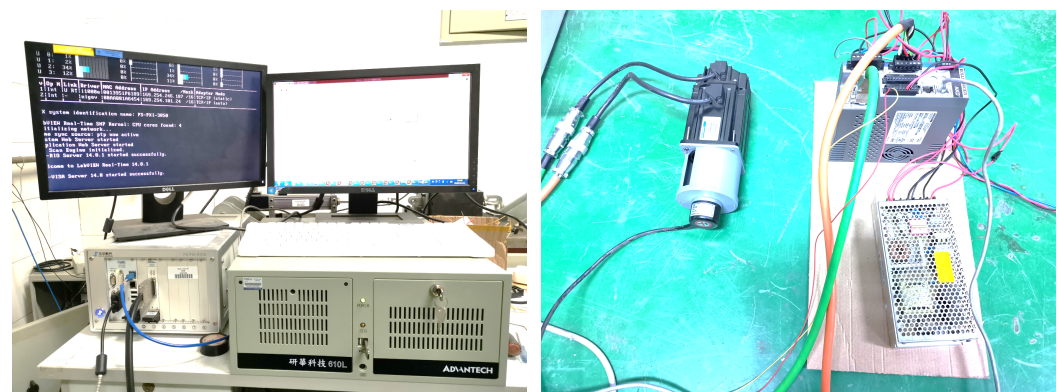


Figure 15. PMSM experimental setup.

3.3.1. The Faults of Motor

In order to guarantee the safety of the test and prevent irreversible damage to the motor, mainly the inter-turn short circuit fault of the motor is simulated in this study, as shown in Figure 16. The inter-turn short circuit fault is mainly reflected in the change in the winding resistance value. Therefore, the three-phase winding of the motor is externally connected, in which the resistance of the two-phase windings has the same value (the other phase is not a series of resistance or a different resistance value of the resistance) to simulate the three-phase winding asymmetry. At the same time, the three-phase current and speed signals of the motor under different fault depths are collected.

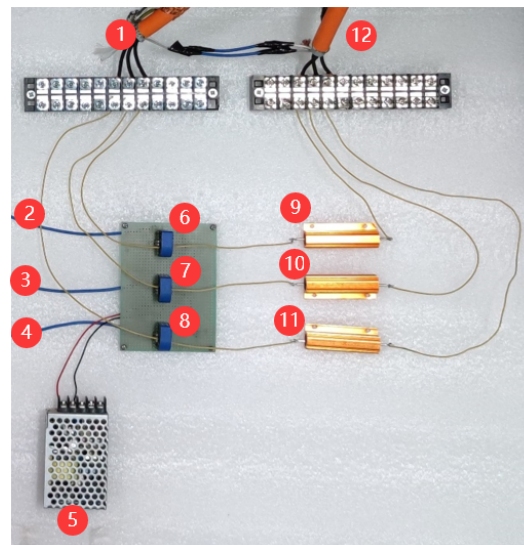


Figure 16. Failure testing of PMSM.

As shown in Figure 16, number 1 represent the controller of the output bus, numbers 2, 3, and 4 represent the three current sensor output signal, number 5 on behalf of the 5V DC power supply, numbers 6, 7 and 8 are used to measure the a, b, and c three-phase current of the current sensor respectively, number 9, 10 and 11 are concatenated in an a, b, and c three-phase resistor respectively, and number 12 represents the input bus at the end of the motor end.

3.3.2. Influence of Short Circuit on Three-Phase Current of Motor

The 1%, 2%, 4%, 8% and 16% inter-turn short circuit faults were injected for a PMSM stator winding. (a), (b), (c), (d), (e) and (f) are the three-phase current curves of the motor under different working conditions. The red line is the a-phase current curve, the purple line is the B-phase current curve, and the blue line is the C-phase current curve. As shown in Figure 17.

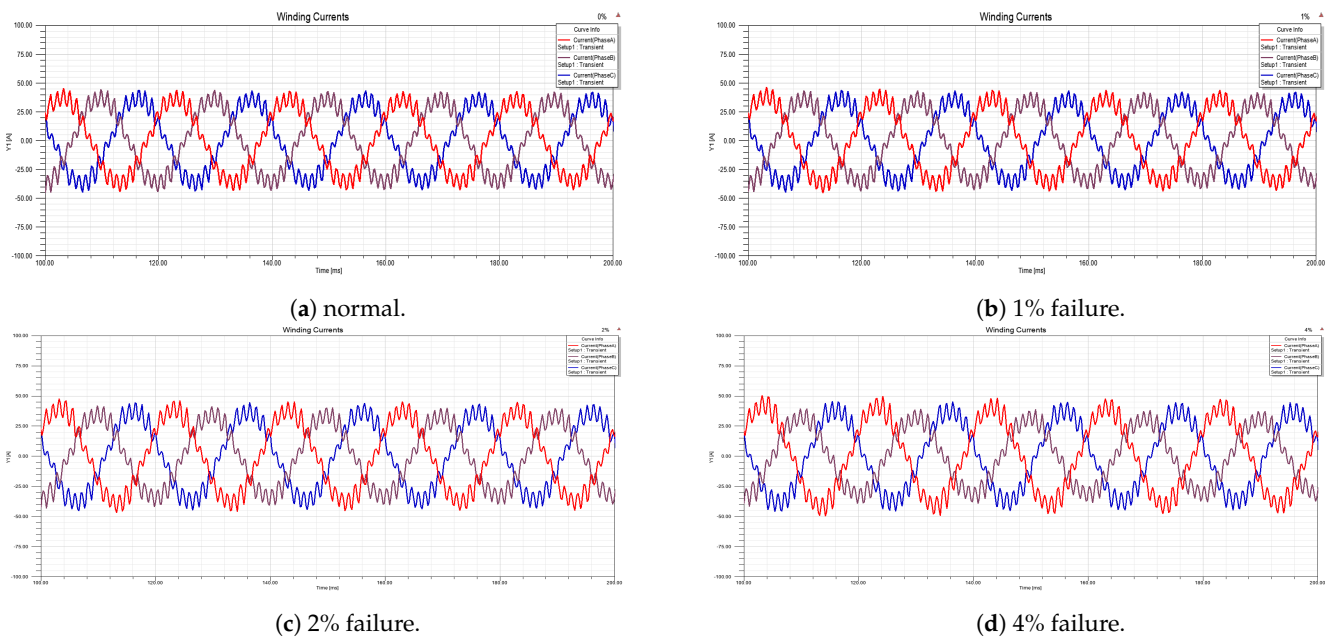
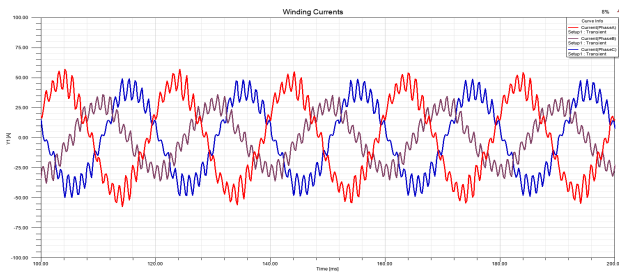
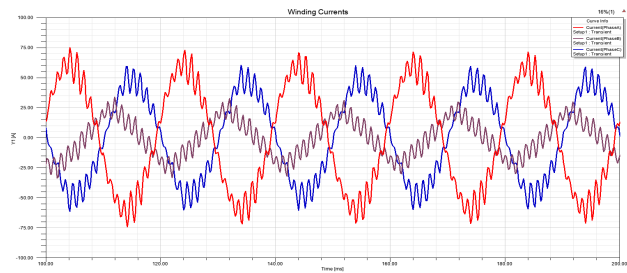


Figure 17. Cont.



(e) 8% failure.



(f) 16% failure.

Figure 17. Three phase current curve of the motor under different conditions.

3.3.3. Verification of Different Deep Learning Algorithms

Four deep learning models were trained ten times to obtain average accuracy. As shown in Figure 18. The results show that ECNN can improve the diagnosis ability of small faults and has a good recognition ability for inter-turn short circuit faults with the same fault type and different fault depths. In the experiment, the accuracy of CNN is 93.38%, while the accuracy of optimized DCNN is improved to 96.61%. The recognition accuracy of the algorithm combining SECNN is 97.25%, while the accuracy of the ECNN network is the highest, reaching 99.87%. Deep network ECNN has great advantages in feature extraction and state recognition. The confusion matrix of the four deep learning models is shown in Figure 19.

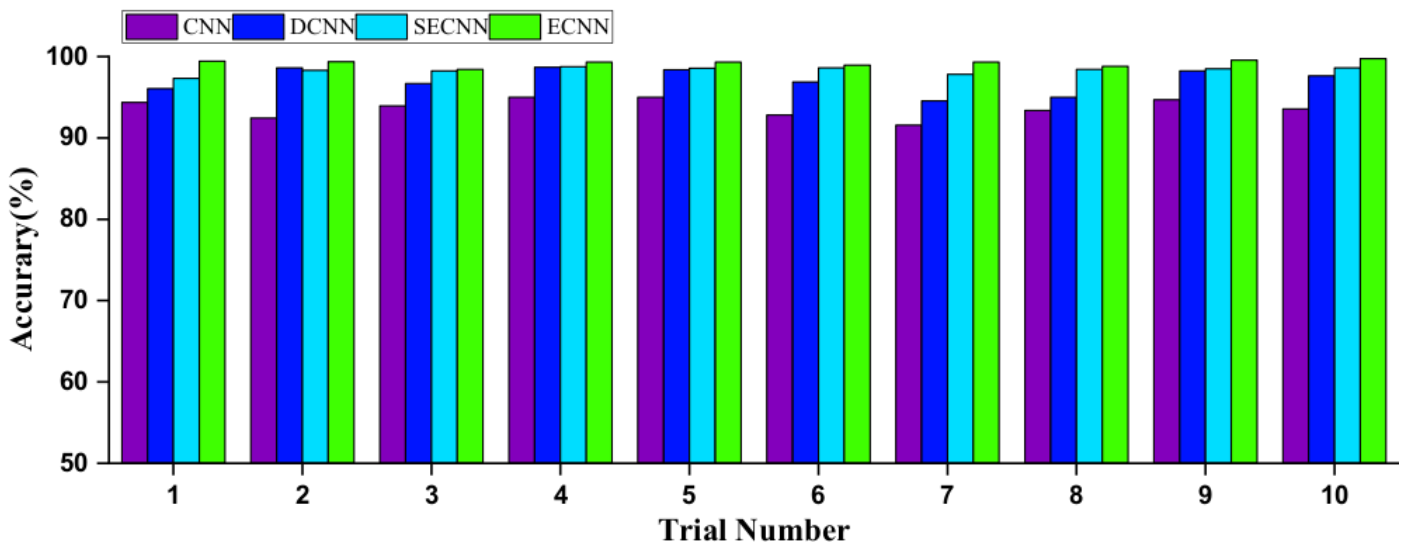


Figure 18. Diagnosis results of the four trials using different methods.

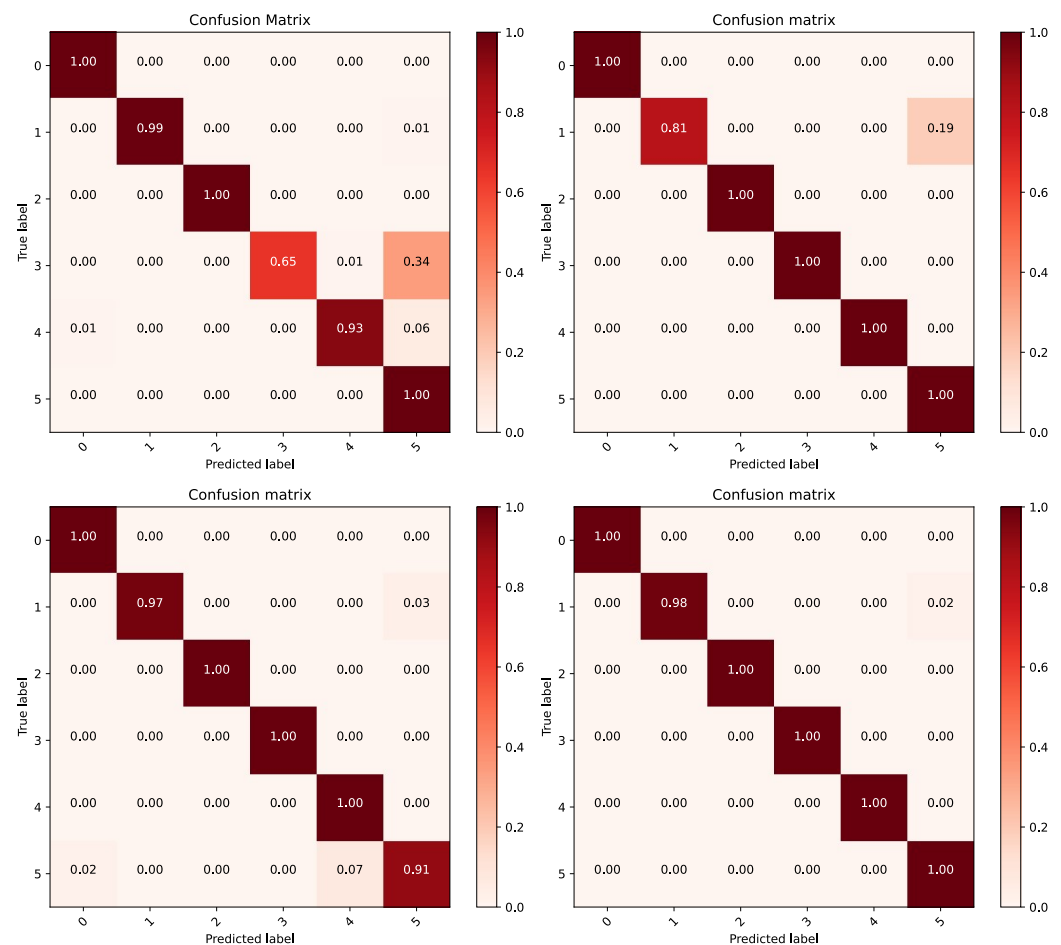


Figure 19. The confusion matrix of the four deep learning models.

4. Conclusions

This paper presents a new ECNN model for fault diagnosis. Firstly, a novel pyramidal dilated convolution method is proposed. It not only expands the receptive field of the model but also plays an important role in feature extraction. Secondly, aiming at the problem that all feature channels of fault information after pyramid dilated convolution are treated equally, a novel ResNet-FCF block is proposed. The ResNet-FCF block can effectively assign appropriate weights to each channel through local cross-channel interaction and global cross-channel interaction learning. The three examples' results show the proposed method's effectiveness and superiority. However, there is still room for improvement in the parameter tuning of this model, which can be further optimized on the number of convolution kernels on the convolution layer. Meanwhile, the recognition rate from the "3" domain to the "1" domain needs to be improved. In the future, other algorithms to improve the adaptive ability of the domain can be added to further improve the fault identification ability under complicated conditions.

Author Contributions: Conceptualization, Q.H.; methodology, Q.H.; software, Q.H.; validation, C.Z. (Chao Zhang), L.C. and Z.L.; formal analysis, C.Z. (Chao Zhang); investigation, Q.H.; data curation, C.Z. (Chaoyi Zhang) and K.Y.; writing—original draft preparation, Q.H.; writing—review and editing, C.Z. (Chaoyi Zhang) and K.Y.; visualization, Q.H.; supervision, C.Z. (Chao Zhang), L.C. and Z.L.; project administration, C.Z. (Chao Zhang); funding acquisition, C.Z. (Chao Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Projects of China (JSZL2020203B004, JSZL202160113001 and JCKY2021607B018).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, W.; Chen, Z.; He, G. A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery. *IEEE Trans. Ind. Inform.* **2020**, *17*, 1753–1762. [[CrossRef](#)]
2. Yan, X.; Liu, Y.; Xu, Y.; Jia, M. Multichannel fault diagnosis of wind turbine driving system using multivariate singular spectrum decomposition and improved Kolmogorov complexity. *Renew. Energy* **2021**, *170*, 724–748. [[CrossRef](#)]
3. Valizadeh Yaghmourali, Y.; Ahmadi, N.; Abbaspour-sani, E. A thermal-calorimetric gas flow meter with improved isolating feature. *Microsyst. Technol.* **2017**, *23*, 1927–1936. [[CrossRef](#)]
4. Shashank, R.; Rai, A.A.; Rao, R.B. Intelligent fault diagnosis of Rolling-element bearing using 1-D Deep Convolutional Neural Networks. *Int. J. COMADEM* **2021**, *24*, 11–18.
5. Eren, L.; Ince, T.; Kiranyaz, S. A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *J. Signal Process. Syst.* **2019**, *91*, 179–189. [[CrossRef](#)]
6. Ahmed, Q.; Khan, F.; Ahmed, S. Improving safety and availability of complex systems using a risk-based failure assessment approach. *J. Loss Prev. Process. Ind.* **2014**, *32*, 218–229. [[CrossRef](#)]
7. Khorram, A.; Khalooei, M.; Rezghi, M. End-to-end CNN+ LSTM deep learning approach for bearing fault diagnosis. *Appl. Intell.* **2021**, *51*, 736–751. [[CrossRef](#)]
8. Vyas, N.S.; Satishkumar, D. Artificial neural network design for fault identification in a rotor-bearing system. *Mech. Mach. Theory* **2001**, *36*, 157–175. [[CrossRef](#)]
9. Arunthavanathan, R.; Khan, F.; Ahmed, S.; Imtiaz, S.; Rusli, R. Fault detection and diagnosis in process system using artificial intelligence-based cognitive technique. *Comput. Chem. Eng.* **2020**, *134*, 106697. [[CrossRef](#)]
10. Zhao, B.; Zhang, X.; Li, H.; Yang, Z. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. *Knowl.-Based Syst.* **2020**, *199*, 105971. [[CrossRef](#)]
11. Miao, M.; Sun, Y.; Yu, J. Deep sparse representation network for feature learning of vibration signals and its application in gearbox fault diagnosis. *Knowl.-Based Syst.* **2022**, *240*, 108116. [[CrossRef](#)]
12. Su, K.; Liu, J.; Xiong, H. Hierarchical diagnosis of bearing faults using branch convolutional neural network considering noise interference and variable working conditions. *Knowl.-Based Syst.* **2021**, *230*, 107386. [[CrossRef](#)]
13. Wang, X.; Mao, D.; Li, X. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. *Measurement* **2021**, *173*, 108518. [[CrossRef](#)]
14. He, C.; Ge, D.; Yang, M.; Yong, N.; Wang, J.; Yu, J. A data-driven adaptive fault diagnosis methodology for nuclear power systems based on NSGAI-CNN. *Ann. Nucl. Energy* **2021**, *159*, 108326. [[CrossRef](#)]
15. Long, Y.; Zhou, W.; Luo, Y. A fault diagnosis method based on one-dimensional data enhancement and convolutional neural network. *Measurement* **2021**, *180*, 109532. [[CrossRef](#)]
16. Zhang, K.; Wang, J.; Shi, H.; Zhang, X.; Tang, Y. A fault diagnosis method based on improved convolutional neural network for bearings under variable working conditions. *Measurement* **2021**, *182*, 109749. [[CrossRef](#)]
17. Huo, C.; Jiang, Q.; Shen, Y.; Qian, C.; Zhang, Q. New transfer learning fault diagnosis method of rolling bearing based on ADC-CNN and LATL under variable conditions. *Measurement* **2022**, *188*, 110587. [[CrossRef](#)]
18. Wang, J.; Wang, D.; Wang, S.; Li, W.; Song, K. Fault diagnosis of bearings based on multi-sensor information fusion and 2D convolutional neural network. *IEEE Access* **2021**, *9*, 23717–23725. [[CrossRef](#)]
19. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [[CrossRef](#)]
20. Yang, S.; Sun, X.; Chen, D. Bearing fault diagnosis of two-dimensional improved Att-CNN2D neural network based on Attention mechanism. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS), Dalian, China, 20–22 March 2020; pp. 81–85.
21. Zhao, D.; Zhang, H.; Liu, S.; Wei, Y.; Xiao, S. Deep rational attention network with threshold strategy embedded for mechanical fault diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–15. [[CrossRef](#)]
22. Abdeljaber, O.; Avci, O.; Kiranyaz, S.; Gabbouj, M.; Inman, D.J. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J. Sound Vib.* **2017**, *388*, 154–170. [[CrossRef](#)]
23. Dibaj, A.; Etefagh, M.M.; Hassannejad, R.; Ehghaghi, M.B. A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults. *Expert Syst. Appl.* **2021**, *167*, 114094. [[CrossRef](#)]
24. Janssens, O.; Slavkovikj, V.; Vervisch, B.; Stockman, K.; Locufier, M.; Verstockt, S.; Van de Walle, R.; Van Hoecke, S. Convolutional neural network based fault detection for rotating machinery. *J. Sound Vib.* **2016**, *377*, 331–345. [[CrossRef](#)]
25. Senanayaka, J.S.L.; Van Khang, H.; Robbersmvr, K.G. CNN based Gearbox Fault Diagnosis and Interpretation of Learning Features. In Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; pp. 1–6.
26. Cheng, Y.; Lin, M.; Wu, J.; Zhu, H.; Shao, X. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network. *Knowl.-Based Syst.* **2021**, *216*, 106796. [[CrossRef](#)]

27. Liu, P.Y.; Chen, C.C.; Liong, S.T.; Tsai, M.H.; Hsieh, P.C.; Wang, K.C. Intelligent Fault Diagnosis Based on Multi-Resolution and One-Dimension Convolutional Neural Networks. In Proceedings of the 2021 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh City, Vietnam, 26–28 August 2021; pp. 319–322.
28. Gong, W.; Wang, Y.; Zhang, M.; Mihankhah, E.; Chen, H.; Wang, D. A Fast Anomaly Diagnosis Approach Based on Modified CNN and Multisensor Data Fusion. *IEEE Trans. Ind. Electron.* **2022**, *69*, 13636–13646. [[CrossRef](#)]
29. Song, X.; Cong, Y.; Song, Y.; Chen, Y.; Liang, P. A bearing fault diagnosis model based on CNN with wide convolution kernels. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *13*, 4041–4056. [[CrossRef](#)]
30. Li, X.; Wu, Y.; Fu, Y.; Tang, C.; Zhang, L. Neighbour Feature Attention-Based Pooling. *Neurocomputing* **2022**, *501*, 285–293. [[CrossRef](#)]
31. Su, H.; Yang, X.; Xiang, L.; Hu, A.; Xu, Y. A novel method based on deep transfer unsupervised learning network for bearing fault diagnosis under variable working condition of unequal quantity. *Knowl.-Based Syst.* **2022**, *242*, 108381. [[CrossRef](#)]
32. Zhao, B.; Zhang, X.; Zhan, Z.; Pang, S. Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling bearings under variable working conditions and domains. *Neurocomputing* **2020**, *407*, 24–38. [[CrossRef](#)]
33. Han, S.; Shao, H.; Zhao, R. Gearbox Fault Diagnosis using Novel Dilated CNN and Piecewise Loss Function under Unbalanced Data. In Proceedings of the 2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing), Nanjing, China, 15–17 October 2021; pp. 1–5.
34. Chu, C.; Ge, Y.; Qian, Q.; Hua, B.; Guo, J. A novel multi-scale convolution model based on multi-dilation rates and multi-attention mechanism for mechanical fault diagnosis. *Digit. Signal Process.* **2022**, *122*, 103355. [[CrossRef](#)]
35. Wang, F.; Liu, R.; Hu, Q.; Chen, X. Cascade convolutional neural network with progressive optimization for motor fault diagnosis under nonstationary conditions. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2511–2521. [[CrossRef](#)]
36. Zhang, X.; He, C.; Lu, Y.; Chen, B.; Zhu, L.; Zhang, L. Fault diagnosis for small samples based on attention mechanism. *Measurement* **2022**, *187*, 110242. [[CrossRef](#)]
37. Li, J.; Liu, Y.; Li, Q. Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method. *Measurement* **2022**, *189*, 110500. [[CrossRef](#)]
38. Yang, Z.B.; Zhang, J.P.; Zhao, Z.B.; Zhai, Z.; Chen, X.F. Interpreting network knowledge with attention mechanism for bearing fault diagnosis. *Appl. Soft Comput.* **2020**, *97*, 106829. [[CrossRef](#)]
39. Zhang, K.; Tang, B.; Deng, L.; Liu, X. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement* **2021**, *179*, 109491. [[CrossRef](#)]
40. Cao, X.; Chen, B.; Zeng, N. A deep domain adaption model with multi-task networks for planetary gearbox fault diagnosis. *Neurocomputing* **2020**, *409*, 173–190. [[CrossRef](#)]
41. Zhang, C.; Feng, J.; Hu, C.; Liu, Z.; Cheng, L.; Zhou, Y. An Intelligent Fault Diagnosis Method of Rolling Bearing Under Variable Working Loads Using 1-D Stacked Dilated Convolutional Neural Network. *IEEE Access* **2020**, *8*, 63027–63042. [[CrossRef](#)]
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
43. He, J.; Wu, P.; Tong, Y.; Zhang, X.; Lei, M.; Gao, J. Bearing Fault Diagnosis via Improved One-Dimensional Multi-Scale Dilated CNN. *Sensors* **2021**, *21*, 7319. [[CrossRef](#)]
44. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
45. Liu, Y.; Yang, Q.; An, D.; Nai, Y.; Zhang, Z. An improved fault diagnosis method based on deep wavelet neural network. In Proceedings of the 2018 Chinese Control And Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 1048–1053.
46. Lei, Y.; He, Z.; Zi, Y. A new approach to intelligent fault diagnosis of rotating machinery. *Expert Syst. Appl.* **2008**, *35*, 1593–1600. [[CrossRef](#)]
47. Lei, Y.; He, Z.; Zi, Y. EEMD method and WNN for fault diagnosis of locomotive roller bearings. *Expert Syst. Appl.* **2011**, *38*, 7334–7341. [[CrossRef](#)]