

## Research Article

# Adaptive Fisher-Based Deep Convolutional Neural Network and Its Application to Recognition of Rolling Element Bearing Fault Patterns and Sizes

Peng Luo ,<sup>1,2</sup> Niaoqing Hu ,<sup>1,2</sup> Lun Zhang ,<sup>1,2</sup> Jian Shen ,<sup>1,2</sup> and Ling Chen ,<sup>1,2</sup>

<sup>1</sup>Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology, Changsha 410073, China

<sup>2</sup>College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

Correspondence should be addressed to Niaoqing Hu; hnq@nudt.edu.cn

Received 2 June 2020; Revised 5 August 2020; Accepted 10 September 2020; Published 25 September 2020

Academic Editor: John S. Sakellariou

Copyright © 2020 Peng Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep learning has the ability to mine complex relationships in fault diagnosis. Deep convolutional neural network (DCNN) with deep structures, instead of shallow ones, can be applied to mining useful information from the original vibration data. However, when the number of the training samples is small, the diagnosis accuracy will be affected. As an improvement of the DCNN, deep convolutional neural network based on the Fisher-criterion (FDCNN) can be used for the fault diagnosis of small samples. But the model parameters in the method are based on human labor or prior knowledge, which is bound to bring negative influence on the diagnosis accuracy. Therefore, a novel adaptive Fisher-based deep convolutional neural network (AFDCNN) method, which can optimize the model parameters adaptively, is proposed as an improvement of the FDCNN. Comparative verification test results show that AFDCNN has more outstanding performance.

## 1. Introduction

There are many benefits of the intelligent rotating machinery health monitoring and fault diagnosis; for example, it can reduce the dependence on the costly training and highly skilled operators and detect potential hazards before a catastrophic failure occurs [1, 2]. Meanwhile, it can also reduce the operation and maintenance costs of the complex engineering systems. Rolling bearings are widely used as critical moving parts of rotating machinery, its state of health matters [3, 4]. Therefore, the intelligent health condition monitoring and accurate fault diagnosis of the rolling element bearings are of great significance.

To meet the needs aforementioned, some fault diagnosis methods, such as BPNN and SVM, have been used for the machinery health monitoring [5–11]. While, with the larger scale, higher speed and much more complex of the rotating machinery, it is ideal for fault diagnosis method that can identify the health status of the diagnosis object accurately,

quickly, and intelligently. There are more stringent requirements that the fault diagnosis methods could be more intelligent indeed.

As a great progress of the diagnosis method, deep learning [12] has the ability to solve the problems that the traditional fault diagnosis methods have to extract features on the basis of prior knowledge and have limited capacity to mine the hidden relationships in the fault quantitative diagnosis. The deep convolutional neural network (DCNN) with deep structures can be established on the basis of the deep learning theory [13, 14]. It can mine distributed features from the original vibration data adaptively [15, 16]. Ever since deep learning theory has been used in the mechanical fault diagnosis, it has attracted a lot of attention [17]. Jun Lee and Kim [18] proposed a novel algorithm for localizing slab identification numbers (SINs) in factory scenes by using DCNN. Bai [19] used DCNN to extract features and achieved good diagnostic results. Guo et al. [20] proposed a novel hierarchical learning rate adaptive deep

convolution neural network based on an improved algorithm and applied it to the bearing fault diagnosis. Verstraete et al. [21] proposed a fault diagnosis method based on DCNN and time-frequency image analysis and achieved good results on two public datasets of rolling element bearing vibration signals. Zhuang et al. [22] proposed a novel deep learning method based on the DCNN and achieved ideal results as well. Zhang et al. [23] proposed a deep graph convolutional network on the basis of graph convolution operators, graph coarsening methods, and graph pooling operations; the experimental results demonstrate that the proposed method can be used to detect different kinds and severities of faults in roller bearings by learning from the constructed graphs. Wang et al. [24] proposed an enhanced intelligent diagnosis method based on multisensor data-fusion and DCNN, and the proposed method achieved higher prediction accuracy and more obvious visualization clustering effects.

The aforementioned applications show that the DCNN is a potential tool in dealing with fault diagnosis of rolling element bearing. While, as a diagnosis model based on the training samples, DCNN is influenced by the number of training samples as well [25]. Here comes the problem, the experimental vibration samples with labels cannot be always sufficient. In which, some of them are very difficult to obtain [26, 27]. The deep convolutional neural network based on the Fisher-criterion (FDCNN) is used for words recognition of small samples [28]. Aiming at the shortcoming of DCNN and learning from related methods in image recognition, this paper has adopted the Fisher classification criteria in the back propagation of model training. But the model parameters in [28] are set based on prior knowledge, which is bound to bring negative influence on the recognition accuracy. Therefore, a novel adaptive Fisher-based deep convolutional neural network (AFDCNN) method in which the model parameters can be optimized adaptively is proposed for the fault diagnosis of bearings in this paper. The advantages of the proposed method are stated again as follows:

- (1) The AFDCNN is able to extract fault features from the original data adaptively
- (2) The AFDCNN is able to establish the hidden relationship between the machinery health conditions and the signals measured adaptively
- (3) Based on limited samples, AFDCNN can achieve perfect performance compared to DCNN
- (4) The proposed method avoids dependence on expert experience to some extent

The architecture of the paper is organized as follows. First, a brief introduction to the traditional DCNN is given. Second, the DCNN model is improved based on the idea of Fisher-criterion. It can be more conducive to the classification characteristics direction. However, the model parameters in the method are based on prior knowledge, which is bound to bring negative influence on the diagnosis accuracy. Therefore, the FDCNN model can be improved by using the optimization algorithms for optimizing the parameter combination adaptively, and then, the AFDCNN is

proposed. Third, the collected bearing fault samples are mainly used for two purposes. One of them, the training sample set is used to build the model, and the other, the test sample set is used to verify the model. Furthermore, the contrast verification is expanded between the traditional methods and the AFDCNN method. Forth is the conclusion.

## 2. Brief Introduction to the DCNN

Essentially, a typical 10-layer DCNN model shown in Figure 1 has two parts [20]: the feature extractor and the Softmax classifier. The feature extractor has one inputting layer and three alternating convolutional layers (or C-layer), max-pooling layers (or P-layer), and two full connection layers (or FC-layer). The C-layer is used for feature extraction, and the P-layer is used for resampling. After several alternating C-layers and P-layers, the FC-layer is followed to compute the class scores. Then, the class scores are inputted into the Softmax classifier and the diagnosis results could be obtained.

**2.1. The Convolutional Layer (C-Layer).** The filter bank is described as follows:

$$W^l = \left( w_1^l, w_2^l, \dots, w_{d_h^l}^l \right), \quad (1)$$

in which  $w_k^l$  is a linear filter of the  $l$ -th layer, its size is  $m^l \times m^l$ , and  $k \in \{1, 2, \dots, d_h^l\}$ ,  $d_h^l$  is the number of different kernels or filters in the  $W^l$ . A matrix  $I_p^{l-1}$  with size  $\omega^{l-1} \times \omega^{l-1}$  is convolved with the filter  $w_k^l$ . The operation can be rewritten as

$$g_k^l = I_p^{l-1} * w_k^l. \quad (2)$$

The Softmax function is

$$\begin{aligned} x^l &= f(g_k^l) = \frac{1}{1 + \exp(-g_k^l)} \\ &= \frac{1}{1 + \exp(-I_p^{l-1} * w_k^l)}. \end{aligned} \quad (3)$$

**2.2. The Pooling Layer (P-Layer).** The P-layer is used for resampling. After the operation, the matrixes' size becomes

$$\text{size: } \frac{(\omega^{l-1} - m^l + 1)}{s} \times \frac{(\omega^{l-1} - m^l + 1)}{s}, \quad (4)$$

in which  $\omega^{l-1}$  is the size of inputting sample of the  $l$ -th layer and  $s$  is the down sampling size, for example, when the mean sampling method is used,  $s$  is 2.

**2.3. The Softmax Classifier.** The Softmax classifier can be described as follows:

$$p_{W^{(l)}}(y^l = i | x^l) = \frac{1}{1 + \exp(-W^{(l)T} x^l)}, \quad (5)$$

where  $p_{W^{(l)}}(\cdot)$  ( $i \in [0, 1]$ ) is an activation function and its parameter is  $W^{(l)}$ . The parameter  $W^{(l)}$  is learned by a

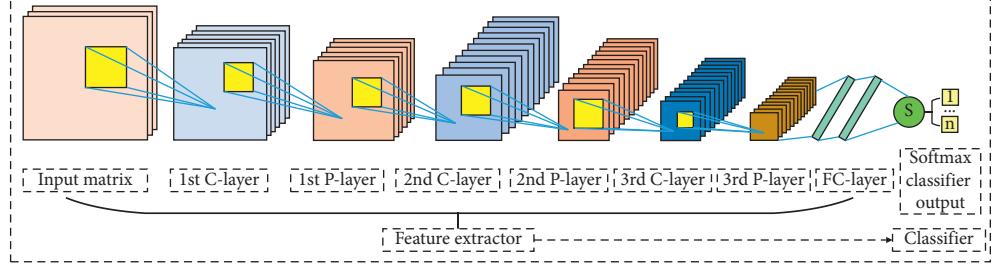


FIGURE 1: The traditional architecture of the DCNN.

training set, and  $x^l$  is the learned feature. The result of equation (5) is a label between 0 and 1. Furthermore, the predicted class  $i$  and score  $\hat{s}(i)$  can be described as

$$\begin{aligned} \hat{i} &= \left\{ i \left| \max_i p_{W^{(l)}}(y^l = i) \right. \right\}, \quad i \in [0, 1], \\ \hat{s}(i) &= p_{W^{(l)}}(y^l = \hat{i}), \quad \hat{i} \in [0, 1]. \end{aligned} \quad (6)$$

Compared to traditional fault diagnosis methods, DCNN has won widespread attention by relying on the advantages of adaptive feature extraction. The reconstruction errors between the inputs and outputs have been selected as the energy function in the method. The connection weights of the network will be optimized and adjusted through the forward and back propagation process. Then, the energy function can be minimized. The weights sharing principle has been used in the forward propagation process to reduce the complexity of the algorithm. The sample feature vector obtained will be adjusted by the weights and bias, and then the sample prediction labels can be obtained through an activation function. In order to obtain a better training model, the process of weight optimization will be one of the key factors.

### 3. Adaptive Fisher-Based Deep Convolutional Neural Network (AFDCNN)

**3.1. The Traditional Process of Model Training.** Assuming that  $m$  samples constituted the sample sets  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$  and they are  $n$  categories, respectively, the traditional energy function [26] can be represented as follows:

$$\begin{aligned} J(W, b) &= \frac{1}{m} \sum_{i=1}^m J(W, b; x^i, y^i) \\ &= \frac{1}{m} \sum_{i=1}^m \left( \frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right), \end{aligned} \quad (7)$$

where  $W$  is the weight value of each unit and  $b$  is the bias term and  $h_{W,b}(x^{(i)})$  is the output of the last neural network layer, namely, the fault-pattern index of the sample  $x^i$ . The target of the training network is to find the minimum value of the function  $J(W, b)$  by adjusting the  $W$  and  $b$ . Using the gradient descent method to optimize the objective function, the iterative formula can be represented as

$$\begin{aligned} W_{ij}^{(l)} &= W_{ij}^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b), \\ b_i^{(l)} &= b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b), \end{aligned} \quad (8)$$

where  $\alpha$  is the learning rate. Before using the back propagation algorithm, the first step is the forward propagation, and it has been used to calculate the output value  $h_{w,b}(x^{(i)})$  of the last layer of the network. Then, the error value between the  $h_{w,b}(x^{(i)})$  and the actual value can be calculated. The error can be represented as

$$\delta_i^{(nl)} = \frac{\partial J(\omega, b)}{\partial Z_i^{(nl)}} = \frac{\partial}{\partial Z_i^{(nl)}} \frac{1}{2} \|h_{w,b}(x^{(i)}) - y^{(i)}\|^2, \quad (9)$$

where  $nl$  is the order of the output layers,  $Z_i^{(l)}$  is the sum input of the  $l$  layer of the  $i$  unit,  $Z_i^{(nl)}$  is the sum input of the last layer of the  $i$  unit. The minimum error between the input tag value and prediction value has been used as the energy function in the back propagation process for the adjustment of the accurate weights.

**3.2. The Optimization Process of Model Training.** In the back propagation process of the DCNN, the adjustment of the weights can be more conducive to the classification characteristics direction based on the idea of Fisher-criterion. At the same time, the search space of the weights iteration is affected by the discriminant conditions, and it can be more conducive to the classification characteristics direction as well.

$J_1$  is the similarity measure function in the class, and it is defined as the sum of all the samples with the category average distance.  $J_2$  is the similarity measure function between the classes, and it is defined as the sum of the average distance classes of all samples.

$$\begin{aligned} J_1 &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \|h_{W,b}(x^{(i,j)}) - M^{(i)}\|^2, \\ J_2 &= \frac{1}{2} \sum_{i=1}^m \sum_{j=i+1}^n \|M^{(i)} - M^{(j)}\|^2, \end{aligned} \quad (10)$$

where  $M^{(i)}$  is the mean value of the  $i$  category samples, and it can be represented as

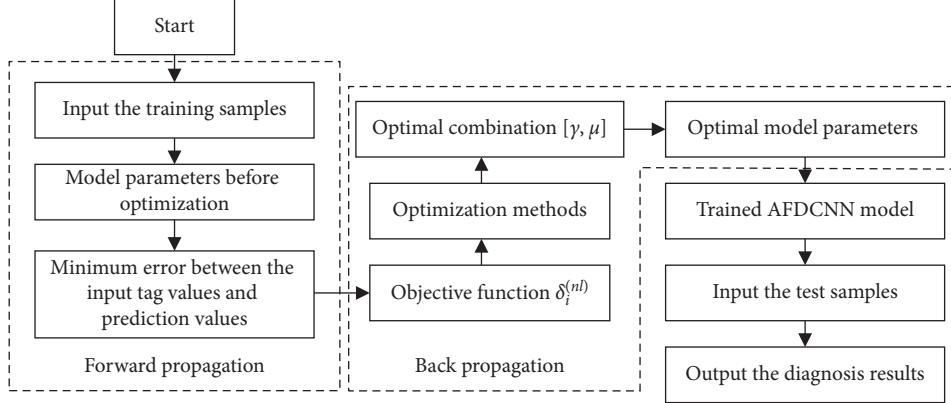


FIGURE 2: The flow chart of AFDCNN.

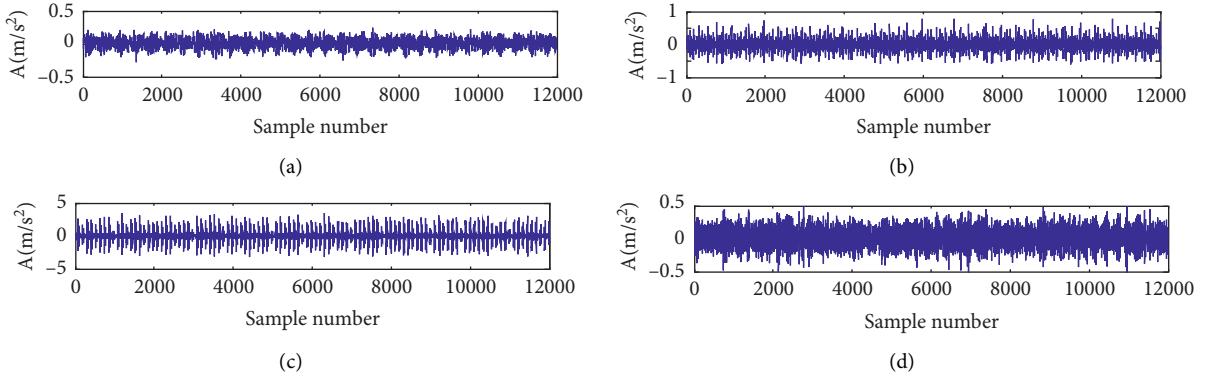


FIGURE 3: Data samples in different bearing health conditions: (a) Normal, (b) OF, (c) IF, and (d) RF.

$$M^{(i)} = \frac{\sum_{j=1}^n h_{W,b}(x^{i,j})}{n}. \quad (11)$$

When  $J_1$  is used as an energy function in the gradient algorithm, after each iteration, the prediction category will be closer to the actual one, and when  $J_2$  is used, the distance between the different categories will be bigger. In order to make the features learned by each DCNN layer more conducive to the diagnosis, the model as follows is used.

$$J = R + \gamma J_1 - \mu J_2, \quad (12)$$

in which  $R$  is the energy function of the DCNN and  $J$  is the overall energy function. The parameter combination  $[\gamma, \mu]$  is depending on the expert experience, which is bound to bring negative influence on the training model.

### 3.3. The Improved Optimization Process of Model Training.

In order to avoid the influence of human factors on model training and obtain the parameters adaptively, several optimization algorithms have been adopted and compared.

Before the optimization process, the objective functions can be derived as follows.

For the function  $J_1$ , the calculation formula of the output layer of residual error for each unit can be represented as

$$\begin{aligned}
\delta_i^{(nl)} &= \frac{\partial J_1}{\partial Z_i^{(nl)}} = \frac{\partial}{\partial Z_i^{(nl)}} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \|h_{W,b}(x^{(i,j)}) - y^{(i)}\|^2 \\
&= \sum_{i=1}^m \sum_{j=1}^n (h_{W,b}(x^{(i,j)}) - M^{(i)}) \cdot ((a_i^{(nl)})' - (M^{(i)})') \\
&= \sum_{i=1}^m \sum_{j=1}^n (a_i^{(nl)} - M^{(i)}) \cdot (a_i^{(nl)} \cdot (1 - a_i^{(nl)})) - \left( nM^{(i)} \cdot \left( \frac{1}{n} - M^{(i)} \right) \right) \\
&= \sum_{i=1}^m \sum_{j=1}^n \left( (a_i^{(nl)} - M^{(i)}) \cdot (a_i^{(nl)} \cdot (1 - a_i^{(nl)})) - \frac{a_i^{(nl)} \cdot (1 - a_i^{(nl)})}{n} \right) \\
&= \sum_{i=1}^m \sum_{j=1}^n (a_i^{(nl)} - M^{(i)}) \cdot (a_i^{(nl)} \cdot (1 - a_i^{(nl)})) \left( 1 - \frac{1}{n} \right).
\end{aligned} \quad (13)$$

For the function  $J_2$ , the calculation formula of the output layer of residual error for each unit can be represented as

TABLE 1: Parameters of rolling bearings.

Items	Health		Pattern 1			Pattern 2			Pattern 3		
	Fault location	None	OF	OF	OF	IF	IF	IF	RF	RF	RF
Fault size (mm)	0	0.18	0.36	0.54	0.18	0.36	0.54	0.18	0.36	0.54	0.54
Testing samples no.	10	10	10	10	10	10	10	10	10	10	10
Training samples no.	10	10	10	10	10	10	10	10	10	10	10
Target value	1		2			3			4		

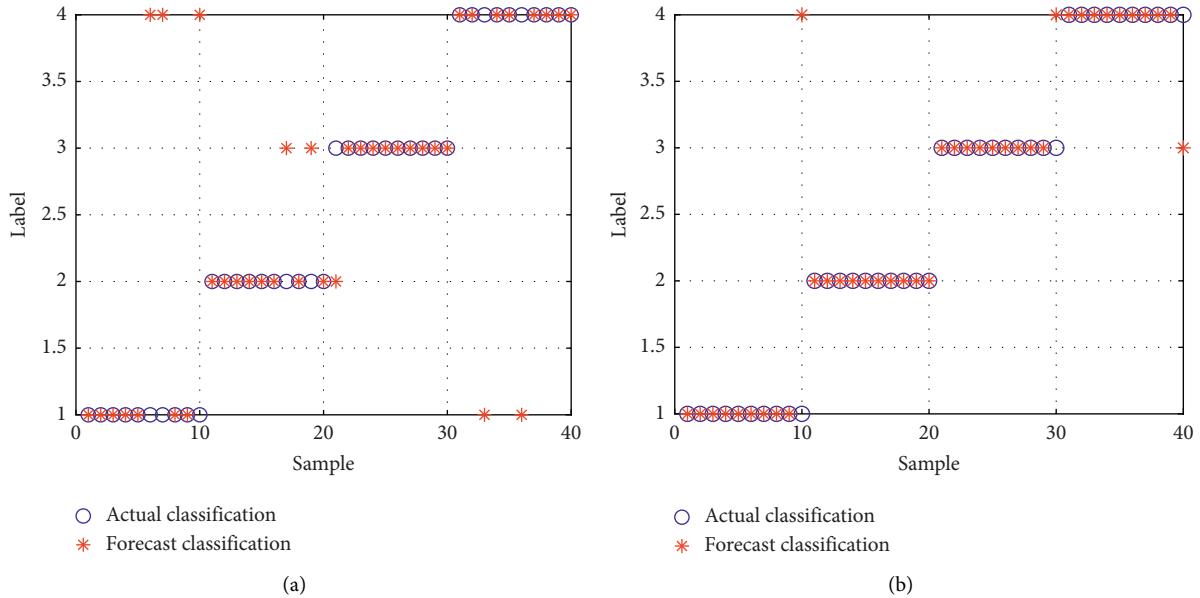


FIGURE 4: (a) Diagnostic classification of DCNN (diagnosis accuracy 82.5%). (b) Diagnostic classification of FDCNN (diagnosis accuracy 92.5%).

$$\begin{aligned}
\delta_i^{(nl)} &= \frac{\partial J_2}{\partial Z_i^{(nl)}} = \frac{\partial}{\partial Z_i^{(nl)}} \frac{1}{2} \sum_{i=1}^m \sum_{j=i+1}^n \|M^{(i)} - M^{(j)}\|^2 \\
&= \left( M^{(i)} - M^{(j)} \right) \cdot \left[ \left( M^{(i)} \right)' - \left( M^{(j)} \right)' \right] \\
&= \left( M^{(i)} - M^{(j)} \right) \cdot \left[ nM^{(i)} \cdot \left( \frac{1}{n} - M^{(i)} \right) \right. \\
&\quad \left. - nM^{(j)} \cdot \left( \frac{1}{n} - M^{(j)} \right) \right].
\end{aligned} \tag{14}$$

The particle swarm optimization (PSO) and stochastic gradient decent (SGD) are adopted for optimizing the parameter combination  $[\gamma, \mu]$ , adaptively and respectively.

In the model, all the weights can be obtained from the BP algorithm after the last layer residual error is minimized. According to the different working conditions of the same object being diagnosed, the optimal parameter combination  $[\gamma, \mu]$  obtained by the use of the optimization algorithm should satisfy the condition that the AFDCNN diagnosis model can be quick and accurate.

In this paper, the rolling element bearing is used as the object being diagnosed. Assume that the object being diagnosed has  $m$  kinds of faults, the  $i^{th}$  ( $i = 1, 2, \dots, m$ ) category has  $n_i$  samples and the sampling frequency is  $f$ .

The proposed method includes three convolutional layers, three max-pooling layers, and two full connection layers. The flow chart for AFDCNN is shown in Figure 2.

#### 4. Experimental Comparison

The bearing data are provided by the Case Western Reserve University (CWRU) [29]. The main components of the experimental apparatus were a 2-hp motor, a torque transducer, and a dynamometer. The motor shaft was supported by 6205-2RS JEM SKF bearings. The data were collected with the sampling frequency of 12 kHz, and the sampling time was 1 s. Figure 3 shows the time domain samples of four kinds of health conditions that are normal (N), outer race fault (OF), inner race fault (IF), and roller fault (RF). Table 1 shows the sample division of the dataset obtained. The configuration of the computer is Intel(R) Core(TM) i7-7400 CPU 16G RAM.

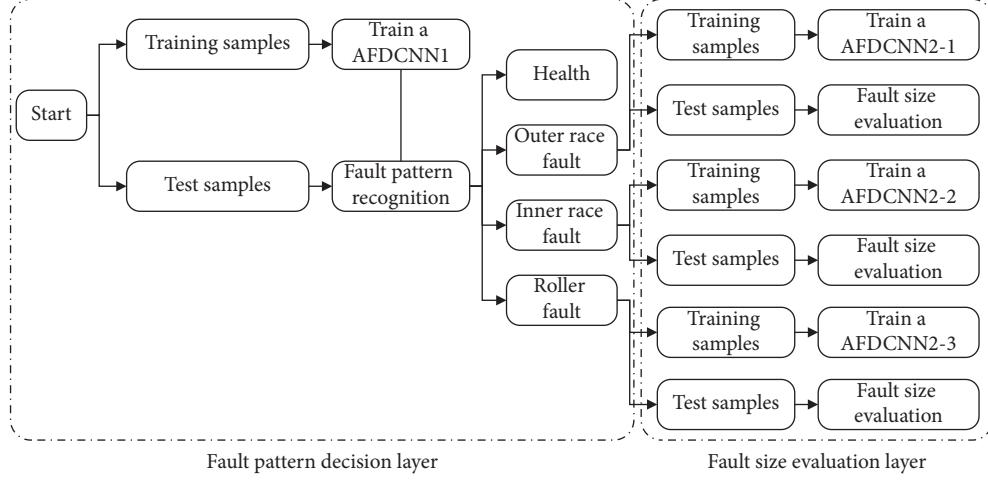


FIGURE 5: Hierarchical framework of AFDCNN.

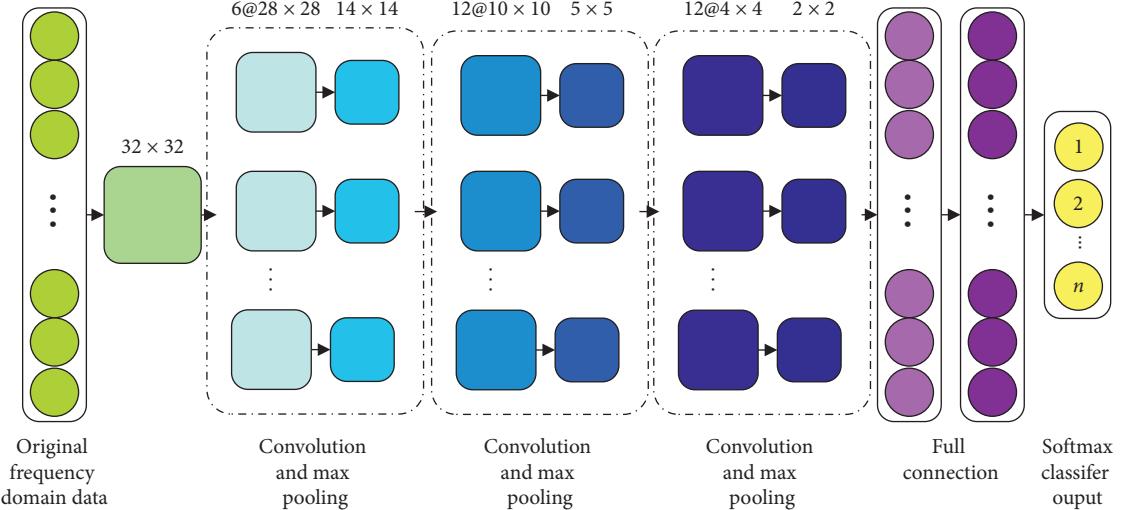


FIGURE 6: Architectural hierarchy of AFDCNN.

TABLE 2: Parameters of AFDCNN model during training.

Layer part parameter	The first layer		The second layer	
	AFDCNN1	AFDCNN2-1	AFDCNN2-2	AFDCNN2-3
No. of layers	10	10	10	10
Learning rate	Adaptive	Adaptive	Adaptive	Adaptive
No. of epochs	600	200	200	200
Batch size	10	10	10	10
No. of kernels	[6, 12, 12]	[6, 12, 12]	[6, 12, 12]	[6, 12, 12]

4.1. *Description of the Data.* Description of the data are provided in Figure 3 and Table 1.

4.2. *Comparison with DCNN, FDCNN, PSO-AFDCNN, and SGD-AFDCNN Method.* The convolutional neural network structure of the DCNN and FDCNN are 6C-2S-12C-2S-12C-2S; it means that the model includes three convolutional

layers and three pooling layers; the size of the convolution kernel is  $5 \times 5$ . Based on the experience, the model parameters combination  $[\gamma, \mu] = [0.02, 0.01]$ . The fault recognition results are shown in Figures 4 and 5.

The convolutional neural network structure of the AFDCNN is the same to the DCNN and FDCNN. The flow chart of Figure 5 described the hierarchical framework of the proposed method. The flow chart of Figure 6 described the

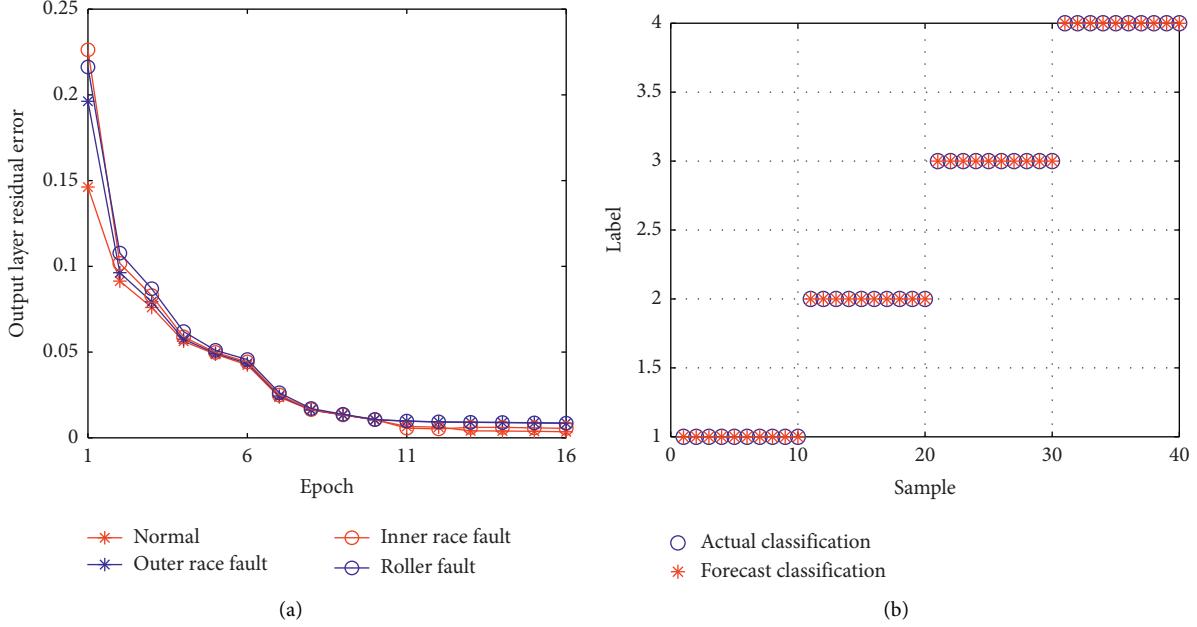


FIGURE 7: The output layer residuals corresponding to the evolution algebra and diagnostic classification of PSO-AFDCNN (diagnosis accuracy 100%).

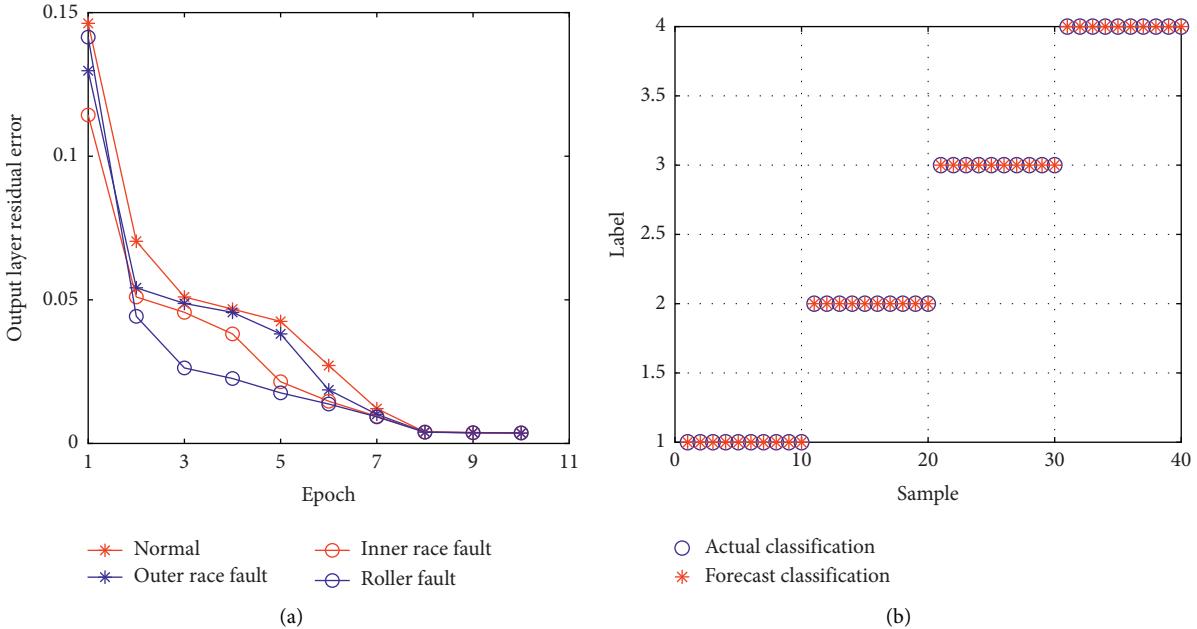


FIGURE 8: The output layer residuals corresponding to the evolution algebra and diagnostic classification of SGD-AFDCNN (diagnosis accuracy 100%).

architectural hierarchy of the AFDCNN. Table 2 stated some of the parameters of the AFDCNN model during training. By using the PSO and SGD, the optimal parameters combination  $[\gamma, \mu]$  is obtained, respectively. As shown in Figures 7(a) and 8(a), it can find that the minimum stable value of the  $\delta_i^{nl}$  appeared in the eleventh generation and eighth generation, respectively. The optimized combinations obtained are  $[\gamma, \mu] = [0.03, 0.021]$  and  $[\gamma, \mu] = [0.03, 0.021]$ ,

respectively. The fault recognition results are shown in Figures 7(b) and 8(b).

Table 3 stated the models adopted in this paper and the diagnosis results. From the diagnosis results of the different models, it can find that both the PSO-AFDCNN and SGD-AFDCNN models have the superior ability on the recognition rate, and because of the difference of optimization speed, the SGD-AFDCNN has shown better performance.

TABLE 3: The models adopted in this paper and the diagnosis results (No. 1 and No. 2 represent no. of testing samples and no. of correct identification, respectively).

Models	Model parameters	No. 1	No. 2	Calculation time (s)	Recognition rate (%)
DCNN	Prior knowledge	40	33	33.1260	82.5
FDCNN	Prior knowledge	40	37	63.1432	92.5
PSO-AFDCNN	Adaptively	40	40	123.4221	100
GSD-AFDCNN	Adaptively	40	40	96.2492	100

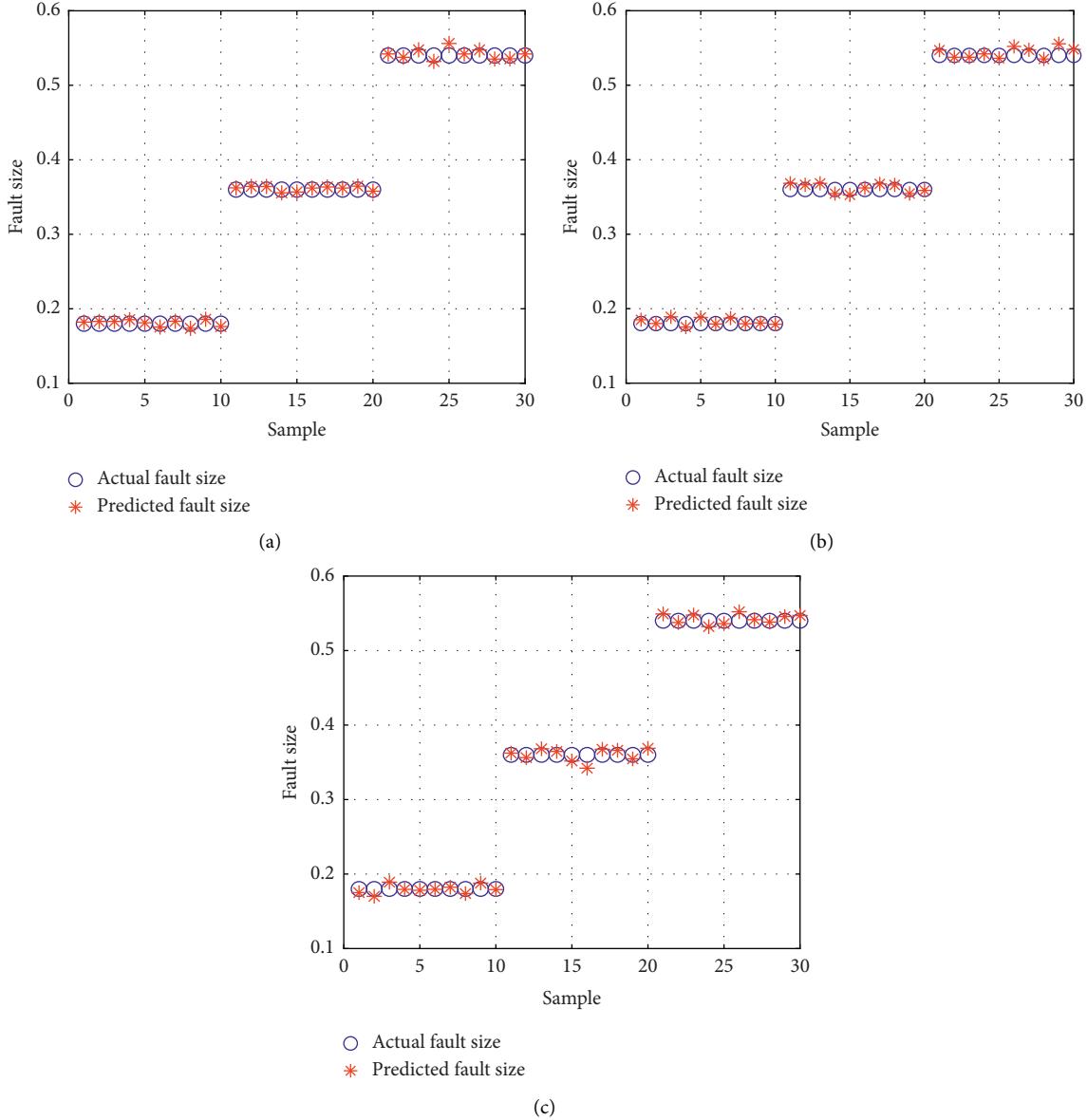


FIGURE 9: PSO-AFDCNN predictions of the (a) OF size, (b) IF size, and (c) RF size for testing samples.

Furthermore, the comparisons of the bearing fault quantitative diagnosis between the PSO-AFDCNN model and SGD-AFDCNN model are adopted, and the diagnosis results are shown as Figures 9 and 10.

To further analyze the evaluation performance of the two methods, a statistical indicator is used to quantify the accuracy of the second layer of the proposed system. The

accumulation error, which denotes the maximum deviation from actual fault size, is defined as follows:

$$\text{Accumulation error} = \sum(|f_{\text{actual}} - f_{\text{predicted}}|). \quad (15)$$

The formula above is used to calculate the maximum error achieved using the PSO-AFDCNN and SGD-

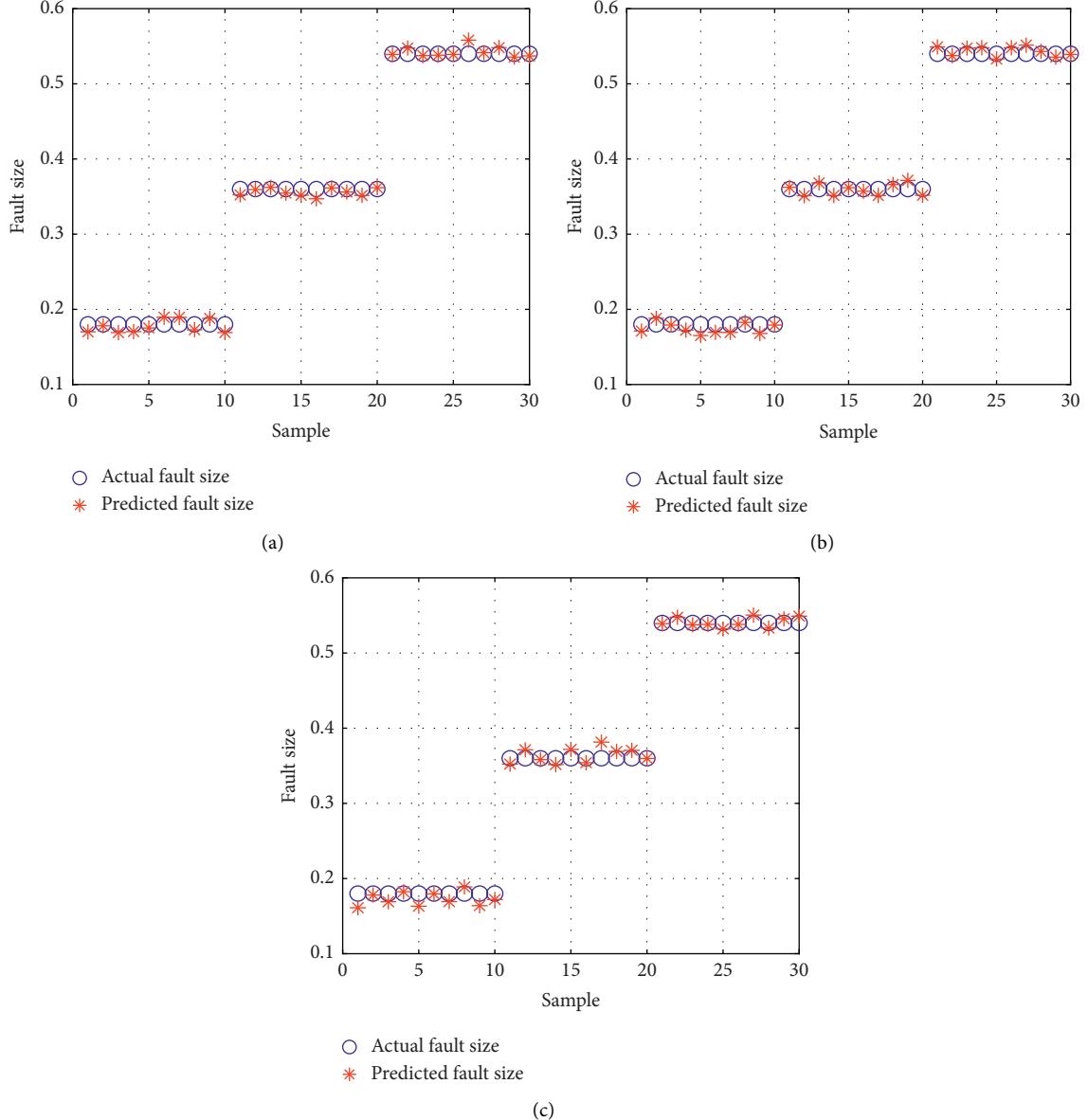


FIGURE 10: SGD-AFDCNN predictions of the (a) OF size, (b) IF size, and (c) RF size for testing samples.

TABLE 4: Comparison of fault size evaluation performance of PSO-AFDCNN and SGD-AFDCNN.

Fault patterns Methods	OF		IF		RF	
	PSO-AFDCNN	SGD-AFDCNN	PSO-AFDCNN	SGD-AFDCNN	PSO-AFDCNN	SGD-AFDCNN
Calculation time (s)	90.1391	56.5645	81.3253	64.6385	87.5374	65.2342
Accumulation error (mm)	0.1286	0.1834	0.1606	0.2062	0.1758	0.2378

AFDCNN methods, and the results are listed in Table 4 for comparison.

The conclusion that both PSO-AFDCNN and SGD-AFDCNN have perfect diagnosis ability on the bearing experiment data can be obtained from the compassion results. Furthermore, the SGD-AFDCNN showed faster diagnosis speed, while the PSO-AFDCNN obtained better diagnosis accuracy. The superiority of the proposed

hierarchical AFDCNN model is confirmed by the experimental comparison results collectively.

## 5. Conclusion

In this paper, a novel DCNN model, which can be called as AFDCNN, is proposed, and the contrast verification

between DCNN, FDCNN, PSO-AFDCNN, and SGD-AFDCNN has been done on the bearing dataset.

The advantages of the AFDCNN are stated as follows:

- (1) It is able to extract fault features from the original data adaptively
- (2) It is able to establish the hidden relationship between the machinery health conditions and the signals measured adaptively
- (3) Both SGD-AFDCNN and PSO-AFDCNN have perfect performance on the bearing fault-pattern recognition, and SGD-AFDCNN showed better calculation ability than PSO-AFDCNN, while, in the process of quantitative diagnosis, POS-AFDCNN obtained better diagnosis accuracy
- (4) The proposed method avoids dependence on expert experience to some extent

The results of the experiments demonstrated that the proposed AFDCNN model has superior ability compared to other methods, such as DCNN and FDCNN. The AFDCNN model achieved a high degree of fault diagnosis accuracy and offered an automatic feature extraction method which could be a practical and convenient method for the bearing fault diagnosis.

## Data Availability

The data are available in the website: <https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant nos. 51975576 and 51475463) and Defense Industrial Technology Development Program (Grant no. WDZC20195500305).

## References

- [1] L. Wang and Y. Shao, "Fault mode analysis and detection for gear tooth crack during its propagating process based on dynamic simulation method," *Engineering Failure Analysis*, vol. 71, pp. 166–178, 2017.
- [2] Z. Wei, Y. Wang, S. He, and J. Bao, "A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection," *Knowledge-Based Systems*, vol. 116, pp. 1–12, 2017.
- [3] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 61, pp. 252–272, 2015.
- [4] Y. Yang, P. Luo, L. Gan, and J. Cheng, "SADBN and its application in fault classification and identification of rolling bearings," *Vibration and Shock*, vol. 38, pp. 11–16, 2019.
- [5] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 73, pp. 303–315, 2016.
- [6] R. Jegadeeshwaran and V. Sugumaran, "Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines," *Mechanical Systems and Signal Processing*, vol. 53, pp. 436–446, 2015.
- [7] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, vol. 5, no. 18, pp. 1077–1095, 2004.
- [8] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mechanical Systems and Signal Processing*, vol. 85, pp. 746–759, 2017.
- [9] C. Shen, D. Wang, Y. Liu, F. Kong, and P. W. Tse, "Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines," *Smart Structures and Systems*, vol. 13, 2014.
- [10] K. C. Gryllias, I. A. Antoniadis, and A. Antoniadis, "A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 2, pp. 326–344, 2012.
- [11] X. Li, A. n. Zheng, X. Zhang, C. Li, and L. Zhang, "Rolling element bearing fault detection using support vector machine with improved ant colony optimization," *Measurement*, vol. 46, no. 8, pp. 2726–2734, 2013.
- [12] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [13] M. Gan, C. Wang, and C. A. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 73, pp. 92–104, 2016.
- [14] Q. Li, Z. Jin, C. Wang, and D. D. Zeng, "Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems," *Knowledge-Based Systems*, vol. 107, pp. 289–300, 2016.
- [15] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Advanced Engineering Informatics*, vol. 32, pp. 139–151, 2017.
- [16] O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2015.
- [17] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016.
- [18] S. Jun Lee and S. W. Kim, "Localization of the slab information in factory scenes using deep convolutional neural networks," *Expert Systems with Applications*, vol. 77, pp. 34–43, 2017.
- [19] S. Bai, "Growing random forest on deep convolutional neural networks for scene categorization," *Expert Systems with Applications*, vol. 71, pp. 279–287, 2017.
- [20] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [21] D. Verstraete, A. Ferrada, E. López Drogue, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using

- time-frequency image analysis of rolling element bearings,” *Shock and Vibration*, vol. 2017, Article ID 5067651, 2017.
- [22] Z. Zhuang, H. Lv, J. Xu, Z. Huang, and W. Qin, “A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions,” *Applied Sciences*, vol. 9, no. 9, p. 1823, 2019.
  - [23] D. Zhang, E. Stewart, M. Entezami, C. Roberts, and D. Yu, “Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network,” *Measurement*, vol. 156, Article ID 107585, 2020.
  - [24] H. Wang, Li Shi, 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. 1, p. 99, 2019.
  - [25] P. Luo, N. Hu, G. Shen, L. Zhang, and Z. Cheng, “DCNN with explicable training guide and its application to fault diagnosis of the planetary gearboxes,” *IEEE Access*, vol. 8, p. 99, 2020.
  - [26] C. Huang, X. Huang, Y. Fang et al., “Sample imbalance disease classification model based on association rule feature selection,” *Pattern Recognition Letters*, vol. 133, pp. 280–286, 2020.
  - [27] N. Huang, X. Yang, G. Cai, X. Song, Q. Chen, and W. Zhao, “Fault depth confrontation diagnosis for main bearings of fans using unbalanced small sample data,” *Chinese Journal of Electrical Engineering*, vol. 40, no. 2, pp. 563–574, 2020.
  - [28] Y. Sun, G. Qi, and Y. Hu, “Deep convolutional neural network recognition algorithm based on Fisher-criterion,” *Journal of Beijing University of Technology*, vol. 6, pp. 835–841, 2015.
  - [29] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the case Western reserve university data: a benchmark study,” *Mechanical Systems and Signal Processing*, vol. 65, pp. 100–131, 2015.