

## Research Article

# An Integrated Method of Rolling Bearing Fault Diagnosis Based on Convolutional Neural Network Optimized by Sparrow Optimization Algorithm

Shuyuan Dong 

*School of Mathematics, Jilin University, Changchun, Jilin 130012, China*

Correspondence should be addressed to Shuyuan Dong; [dongsy0318@163.com](mailto:dongsy0318@163.com)

Received 18 March 2022; Revised 7 June 2022; Accepted 8 June 2022; Published 15 July 2022

Academic Editor: Punit Gupta

Copyright © 2022 Shuyuan Dong. 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.

Intending to solve the problems including poor self-adaptive ability and generalization ability of the traditional categorizing method under big data, a parameter-optimized Convolutional Neural Network (CNN) based on Sparrow Search Algorithm (SSA) is proposed in this research. Initially, the raw data regarding a series of bearing vibration signals are processed with Fast Fourier Transform (FFT) and Continuous Wavelet Transform (CWT) to attain groups of time-frequency maps. Then, Locally Linear Embedding (LLE) and linear normalization are introduced to make these maps proper for the input of CNN. Next, the pre-processed data sets are utilized as training and testing samples for CNN, and the accuracy rate of the testing is considered as the fitness of SSA, which is used to search for optimal parameter combinations for CNN by SSA. Meanwhile, the construction of the CNN is determined by experience and other previous researches. Finally, an NN-based defect diagnosis model for bearings will be constructed after the SAA has determined the appropriate parameters. The model's accuracy rate may reach 99.4 percent after repeated testing using samples, which is significantly superior to the classic fault detection approach and the fault diagnostic method based solely on shallow networks. This experimental result demonstrates that the suggested strategy may significantly increase the model's self-adaptive feature extraction capacity and accuracy rate, implying a higher performance in defect diagnosis in the presence of huge data.

## 1. Introduction

The scale of the modern manufacturing industrial system is becoming larger and larger, which will add uncertainty to the manufacturing process and makes it more complicated to control the condition of various production equipment. In order to ensure the safety and product quality of manufacturing, fault detection based on large data has become one of the hotspots in the field of process control in the era of intelligent manufacturing. For instance, Caggiano et al. [1] created a machine learning system for online fault recognition via automatic image processing to quickly identify material defects caused by process nonconformities in metal powder Selective Laser Melting (SLM). Liu et al. [2] methodically summarized MVCDFD-MTs in order to give academics and engineers with a theoretical foundation and roadmap for further research or development of MVCDFD-MTs based on

machined surface texture information. Xu and Yao [3] described how a laser line scanning sensor was integrated into a robot-based laser-aided additive manufacturing (LAAM) system to allow for part geometry measurement on machine.

Bearings are widely used in various rotating machines related to power supply, automobile, and military industries as core components of modern machinery. However, because these apparatuses are typically used in harsh environments such as high temperature, high moisture, and overload [4], bearing failure is unavoidable. The majority of shaft wear is difficult to detect. It will not be noticed unless the machine has a large jumping range, irregular noise, or an abnormal temperature. When these phenomena are discovered, the majority of the rolling shafts have already worn out, resulting in machine shutdown, which will almost certainly result in economic losses and accidents [5, 6]. Bearing condition monitoring and problem diagnostics can thus be used to

maintain the equipment, extend its service life, and avoid potentially dangerous mishaps. It will be easier to obtain some parameters such as vibration, noise, and temperature using auxiliary condition monitoring and then use appropriate fault diagnosis methods to diagnose faults. Existing defect detection algorithms fall into two broad categories: approaches based on signal analysis and methods based on machine learning [7]. Duan et al. [8] suggested a time-frequency Kurtosis Spectrum-based bearing fault detection technique to reliably extract the characteristic frequency of the rolling bearing damage, which uses Slice Wavelet Transform to decompose the time-frequency vibration signal to obtain the corresponding amplitude of each frequency component Kurtosis. The Kurtosis sequence is used to construct the time-frequency Kurtosis Spectrum of the signal. The corresponding frequency bands are determined according to the corresponding frequencies of the larger frequency peaks in the time-frequency Kurtosis Spectrum, and the time-frequency slices are selected in the time-frequency space. Then, the signal components are separated by reconstruction, and the envelope of the reconstructed signals is obtained by Envelope Demodulation. On this basis, the characteristic frequency of rolling bearing is determined by the equivalent power spectrum of the envelope signal. Li [9] developed a technique for diagnosing bearing faults based on a VMD-based bispectrum. VMD simplifies complicated nonstationary vibration signals by decomposing them into a series of Intrinsic Mode Functions (IMF). VMD detects IMF at the center frequency using the Alternating Direction Multiplier Method (ADMM). Bispectrum analysis can be used to determine the presence of phase coupling effects. Bispectrum is immune to Gaussian and non-Gaussian noise and is hence ideal for identifying rolling bearing local problems. When condition monitoring systems are used to collect real-time data from the device, massive data will be attained after the long-time operation promoting machinery health monitoring to become big data. These kinds of big data have the features such as large volume, diversity, and high velocity. The traditional methods based on signal analysis will have limitations on efficiency, accuracy, and capability to deal with these massive data. Therefore, how to extract fault features both accurately and efficiently has become a hotspot for many researchers [10, 11].

A machine learning-based approach that employs powerful machine learning tools such as Backpropagation Neural Networks (BPNN), Probabilistic Neural Networks (PNN), and Extreme Learning Machines (ELM) is attracting an increasing number of people to apply it to defect diagnosis. Deep learning is a hot topic among researchers in the machine learning and pattern recognition fields. It has had tremendous success in a wide range of fields, including speech recognition, computer vision, and natural language processing [11]. Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Deep Belief Networks (DBN) are critical components of big data solutions because they can extract useful knowledge from complex systems. Deep learning, as opposed to traditional learning methods, is thought to use a shallow structured learning architecture. It refers to machine learning approaches that use supervised or unsupervised schemes to automatically

learn hierarchical expressions in deep classification methods [12]. The application of CNN to image classification has recently achieved great success. CNN is now one of the most effective methods for detection and identification [13]. Liu et al. [14] proposed a method based on a time-frequency diagram obtained by Fourier Transform (FT) for each vibration signal frame and CNN. Time-frequency diagrams can then be constructed using the Short-Time Fourier Transform (STFT). As a result, the training signal's time-frequency diagram is utilized as the input of the convolution neural net to train the network. Rolling bearing faults may be identified after the time-frequency diagram of the test signal is entered into the network model. Consequently, the fault recognition rate is at least 97.64%. As some researchers have demonstrated already, network structure and training parameters, such as epochs, minibatch, and initial learning rate, have a considerable impact on test accuracy [15, 16]. In order to obtain a better recognition rate, this research uses Sparrow Search Algorithm (SSA) to optimize training parameters. SSA is a kind of intelligent optimization algorithm introduced by Xue [33] in 2020, which has more advantages, including aiming at more general problems and better preventing falling into local optimum compared with the traditional optimization algorithm. Wang et al. [17] established the possible danger to an aerial target depending on the commander's emotional state as determined by the SSA-BP model. In their research, they suggested a system for predicting possible threats while taking commander emotion into account (PTP-CE) by combining a Bidirectional LSTM (BiLSTM) network with a Backpropagation Neural Network (BP) optimized with the Sparrow Search Algorithm (SSA). The results show that the prediction accuracy of the SSA-BP is higher than that of the Genetic Algorithms-based Backpropagation Neural Network (GA-BP), the BP, and the General Regression Neural Network (GRNN), which indicates that the SSA performs more robustly while solving the global optimal solution issue. Experiments show that the method proposed in this research can effectively improve CNN's identification accuracy and solve the shortcomings of traditional methods under big data. This research employs a unique group optimization technique called SSA, inspired by the foraging and antipredatory behavior of sparrows, to maximize the CNN's parameter combination, demonstrating the method's application potential and workability. Moreover, it also established a practical and effective method for bearing diagnosis considering big data situation, which can obtain a higher test accuracy rate through a few iterations of SSA than other traditional or nonoptimized methods.

Involved theories are introduced in Section 2. Section 3 describes the construction of bearing fault diagnosis model based on parameter-optimized CNN. The case studied in this paper is discussed in Section 4. Section 5 makes a conclusion of the proposed method.

## 2. Theory

*2.1. Theory of Continuous Wavelet Transform (CWT).* The CWT is a kind of alternative method of transforming the original signal into a certain domain in order to analyze and

process [18, 19]. The foundation of CWT [20, 21] is a family of functions

$$\psi_{\sigma,\tau}(t) = \frac{1}{\sqrt{\sigma}} \psi\left(\frac{t-\tau}{\sigma}\right), \quad \sigma, \tau \in R, \quad (1)$$

where  $\psi$  is a fixed function, called ‘‘mother wavelet,’’ which is localized in both time and frequency. The function  $\psi_{\sigma,\tau}(t)$  is produced by dilating ( $\sigma$ -dilation) in the time domain and translating ( $\tau$ -translation) in the frequency domain to the mother wavelet. The mother wavelet utilized in this research is the Daubechies wavelet (dbN), which has no explicit expression (except for  $N=1$ , which is Harr wavelet), but the square mode of the transformation function  $h$  is clear [22].

The CWT is defined as the inner product of the wavelet family  $\psi_{\sigma,\tau}(t)$  and signal  $f(t)$ .

That is given by

$$F_w(\sigma, \tau) = \langle f(t), \psi_{\sigma,\tau}(t) \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{\sigma}} \bar{\psi}\left(\frac{t-\tau}{\sigma}\right) dt, \quad (2)$$

where  $\bar{\psi}$  is the complex conjugate of  $\psi$ ,  $F_w(\sigma, \tau)$  is the time-scale map, and  $f(t)$  satisfies the condition:

$$f^2 = \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty. \quad (3)$$

In order to make sure the existence of the inverse wavelet transformation, the mother wavelet  $\psi(t)$  should satisfy the admissibility condition:

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty, \quad (4)$$

where  $\hat{\psi}(\omega)$  is the Fourier Transform of  $\psi(t)$  and  $C_\psi$  is a constant for wavelet  $\psi$ .

The integrand in equation (6) has an integrable discontinuity at  $\omega = 0$ , which means that  $\int \psi(t) dt = 0$ . Then, the inverse wavelet transform can be attained by

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F_w(\sigma, \tau) \psi_{\sigma,\tau}(t) \frac{d\sigma d\tau}{\sigma^2}. \quad (5)$$

**2.2. Theory of Manifold Learning.** Manifold learning is a form of unsupervised learning. Compared to the standard linear dimensionality reduction approach, manifold learning successfully uncovers the underlying dimensionality of nonlinear high-dimensional data, hence facilitating dimensionality reduction and data analysis [23, 24]. It may be classified into linear and nonlinear manifold learning algorithms. Isomap [25], Laplacian Eigenmaps (LE) [26], and Locally Linear Embedding (LLE) [27] are all examples of nonlinear manifold learning algorithms. Linear manifold learning is a linear extension of nonlinear methods such as Principal Component Analysis (PCA) and Multidimensional Scaling (MDS). The LLE approach was used in this study.

The main idea of the LLE algorithm is to maintain the local order relationship between the data embedded in space and essential space. Our goal is to find low dimensional a dataset  $(y_1, y_2, \dots, y_N)$ , which exists in the space of  $\mathbb{R}^d$ ,

based on the given data set  $(x_1, x_2, \dots, x_N)$  in the high-dimensional space  $\mathbb{R}^D$  and  $d \ll D$ .

Firstly, a number of adjacent points  $x_{i1}, \dots, x_{in}$  of every point  $x_i$  in the data set should be found.  $x_{i1}, \dots, x_{in}$  can be the data points in the sphere that surrounds  $x_i$  with  $\varepsilon$  as the distance.

Every data point can be expressed by linear combination of its adjacent points  $x_{i1}, \dots, x_{in}$ :

$$x_i = \sum_{j=1}^n \omega_{ij} x_{ij}, \quad (6)$$

where  $\omega_i$  is a  $n \times 1$  column vector and  $\omega_{ij}$  is line  $j$  of  $\omega_i$ .

Then, it comes to minimizing the following loss function (9):

$$\arg \min_{\omega} \sum_{i=1}^N \left( \left\| x_i - \sum_{j=1}^k \omega_{ij} x_{ij} \right\|_2^2 \right). \quad (7)$$

The weight coefficient can be obtained by solving the above formula:

$$\omega = [\omega_1, \omega_2, \dots, \omega_N]. \quad (8)$$

Next, it is considered that, after reducing the original data from  $D$  dimension to  $d$  dimension  $x_i (D \times 1) \rightarrow y_i (d \times 1)$ , it can still be expressed by linear combination of its adjacent points  $y_{i1}, y_{i2}, \dots, y_{in}$ , and its combination coefficient remains unchanged. Once again, minimize the loss function:

$$\arg \min_Y \sum_{i=1}^N \left( y_i - \sum_{j=1}^k \omega_{ij} y_{ij} \right)^2. \quad (9)$$

Eventually, it is able to get the data in low dimensional space after dimensionality reduction

$$Y = [y_1, y_2, \dots, y_N]. \quad (10)$$

**2.2.1. Theory of Convolutional Neural Network (CNN).** CNNs, which are deep feed-forward artificial neural networks well-suited to extracting structural characteristics from two-dimensional signals, can help with both feature learning and time-frequency map recognition. For large-scale picture classification and recognition, it is also the most commonly employed deep learning model [28, 29]. Figure 1 depicts the general layout of the CNN.

The input of a convolutional layer is an image of size  $a * b$ . The convolutional layer consists of  $f$  filters of size  $r * s$ , and it has fewer dimensions than the input images. The output of that is a series of feature maps with the size of  $(a - r + 1) * (b - s + 1)$  with the step size 1. The filter is implemented by assigning weights to each pixel in the input image and calculating it as a weighted sum to extract some features contained in the image. Then, the weighted sum is added by the additive deviation, and the pixels in the convolution map are obtained through the nonlinear function. Lately, the usage of rectified linear units (ReLU) as a nonlinear activation function has gotten an amount of

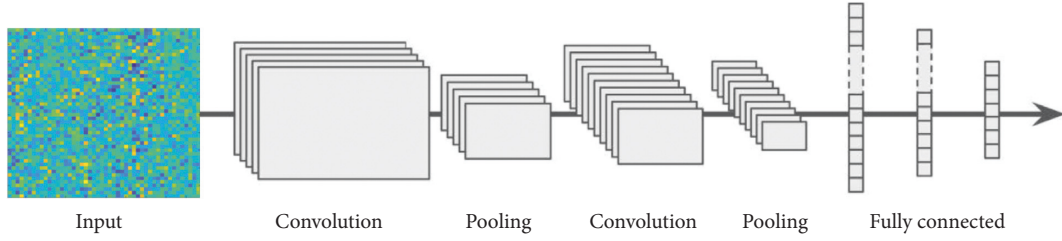


FIGURE 1: Basic structure of CNN.

popularity [30], which is utilized in the convolutional layer of this research, as shown as follows:

$$\phi(x) = \max(0, x). \quad (11)$$

The output  $y_j^l$  of a certain feature map  $j$  in the convolutional layer  $l$  is given as

$$y_j^l = \phi \left( b_j^l + \sum_{i \in Z_j^l} y_i^{l-1} \otimes f_{ij}^l \right), \quad (12)$$

where  $\phi$  is the nonlinear activation function;  $b_j^l$  is the scalar bias for the  $l$  th layer;  $Z_j^l$  is the selected feature map  $i$  in the  $l-1$  th layer, which is added up by the feature map  $j$  in the  $l$  th layer;  $\otimes$  indicates the convolutional operator that convolutes the activation  $y_i^{l-1}$  of the preceding layer;  $f_{ij}^l$  is the filter used to perform the convolutional operation.

Then, this is followed by a pooling layer. Each feature map is needed to be subjected to region-wise pooling to retain the main features while reducing parameters and computation to prevent overfitting. The pooling process used here is average pooling. That is, the average sampling value of  $2 \times 2$  is not overlapped in the output of the preceding layer. The output of the pooling layer will lead to a reduction of dimension. After reducing the dimension of the feature map  $d$  in a layer  $l$ , the output  $p_d^l$  is given as

$$p_d^l = X(d_j^l, N_j^l), \quad (13)$$

where  $X$  is the downsizing function downsized by a factor of  $N_j^l$  and  $d_j^l$  is a convoluted feature map which is to be downsized.

The last layer is a fully connected layer, and its output  $y_p$  is given by

$$y_p = \psi(b_p + Wf), \quad (14)$$

where  $b_p$  is the bias for the output layer;  $W$  is the weight matrix between the input and output layers of the fully connected layers;  $f$  is the feature maps of the fully connected input layer; and  $\psi$  is the softmax function [30]. The training method applied to this research is stochastic gradient descent (SGD). Every time a datum is read, the stochastic gradient descent algorithm will immediately calculate the cost function gradient to update the parameters. The gradient is calculated using the backpropagation method [31]. All filter weights and deviations are updated according to the objective function of each input sample until the best

representation of training samples is obtained. The cost function  $J$  employed here is cross entropy which is defined as

$$J(\theta) = \frac{1}{m} \left[ \sum_{i=1}^m (y^i \log h_\theta(x^i) + (1 - y^i) \log(1 - h_\theta(x^i))) \right], \quad (15)$$

where  $m$  is the number of training samples;  $h_\theta(x)$  is the value predicted by parameters  $\theta$  and  $x$ ;  $y$  is the value of the original training sample, which acts as the standard answer;  $i$  means the  $i$  th sample.

**2.2.2. Theory of Sparrow Search Algorithm (SSA).** The introduction of SSA in 2020 is primarily motivated by sparrow foraging and antipredator behavior. The technique is new and benefits from fast search and convergence. The fundamental requirements of the algorithm are as follows [32]. When foraging, sparrows are classified as either discoverers or followers. The discoverers are in charge of finding food within the population and establishing foraging areas and routes for the entire sparrow population, while the participants rely on the discoverers for food. Sparrows frequently use one of two foraging strategies: discoverers or participants. Individuals in the surveillance group recognize one another's behavior, and attackers compete for food resources with high intake partners to increase their predation rate. Furthermore, as the sparrow population becomes aware of the threat, antipredator behavior develops.

In the  $D$ -dimensional solution space, the location of each sparrow is  $X = (x_1, x_2, \dots, x_D)$ , and the fitness value is  $f_i = f(x_1, x_2, \dots, x_D)$ . There are  $N$  sparrows in the group. In each generation,  $P_N$  sparrows with the best position in the population are selected as discoverers, while the remaining  $N - P_N$  sparrows are selected as followers.

The location update formula of each generation of discoverers is as follows:

$$x_{i,d}^{t+1} = \begin{cases} x_{i,d}^t * \exp\left(\frac{-i}{\alpha * \text{iter}_{\max}}\right), & R_2 < ST, \\ x_{i,d}^t + Q, & R_2 \geq ST, \end{cases} \quad (16)$$

where  $x_{i,d}^t$  indicates represents the  $d$  th dimension position of the  $i$  th individual in the  $t$  th generation;  $\alpha$  is a uniform random number in  $(0, 1]$ ;  $Q$  is a standard normal distribution random number;  $R_2$  is a uniform random number in

$[0, 1]$ ;  $ST$  means the alarm value whose value range is  $[0.5, 1.0]$ .

Position update formula for followers is given as

$$x_{i,d}^{t+1} = \begin{cases} Q * \exp\left(\frac{xw_{i,d}^t - x_{i,d}^t}{t^2}\right), & i > \frac{n}{2}, \\ xb_{i,b}^t + \frac{1}{D} \sum_{d=1}^D \text{rand}\{-1, 1\} * (|xb_{i,d}^t - x_{i,d}^t|), & i \leq \frac{n}{2}, \end{cases} \quad (17)$$

where  $xw$  is the worst position in the current population and  $xb$  is the best position in the current population.

While sparrows are foraging, some of them will be vigilant. When danger comes near, they will give up the present food. The location update formula is as follows:

$$x_{i,d}^{t+1} = \begin{cases} xb_{i,d}^t + \beta * (x_{i,d}^t - xb_{i,d}^t), & f_i \neq f_g, \\ x_{i,d}^t + K * \left(\frac{x_{i,d}^t - xw_{i,d}^t}{|f_i - f_w| + \varepsilon}\right), & f_i = f_g, \end{cases} \quad (18)$$

where  $\beta$  indicates a random number conforming to the standard normal distribution;  $K$  means a uniform random number of  $[-1, 1]$ ;  $\varepsilon$  is a smaller number that prevents the denominator from being unique;  $f_w$  is the fitness value of the sparrow in the worst position. It can be seen from the formula (18) that if the sparrow is currently in the best position, it will escape to a location near itself, depending on the ratio of the worst food in the worst position to the food and its location. If the sparrow is not in the best position, it will escape to the nearest best location.

**2.2.3. Proposed Method.** The optimization steps of SSA for CNN parameters are as follows, which are shown in Figure 2.

- (1) Set parameters for SSA and initialize the population.
- (2) Use the CNN to train the test's accuracy rate as a fitness function, and the individual fitness value is evaluated considering the learning rate and batch size. Then, mark the optimal position.
- (3) Calculate whether the current iteration times have achieved the termination condition; if they have, terminate the loop and print the result; otherwise, continue.
- (4) Update the position in the current population and initialize the individuals beyond the upper and lower limits.

### 3. The Construction of Bearing Fault Diagnosis Model Based on Parameter-Optimized CNN

The proposed bearing diagnosis method, which is based on parameter-optimized CNN, combines traits derived from a large data monitoring device with the benefits of deep

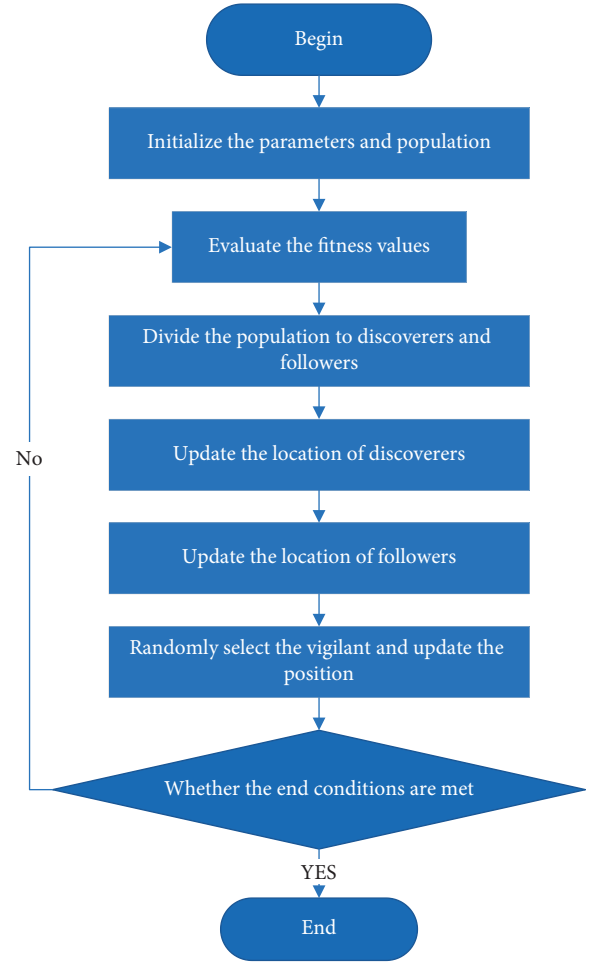


FIGURE 2: The flowchart of improved CNN based on SAA.

learning. It combines unsupervised and supervised learning to extract fault features and identify the running state of devices in large amounts of data in an adaptive manner. Nonetheless, this measure has a limited adaptive capacity for feature extraction and a shallow network's insufficient generalization capability for defect identification when compared to traditional approaches. Figure 3 depicts the entire procedure. The specific steps are as follows:

- (1) The bearing's vibration signal is preprocessed using FFT to get frequency domain data.
- (2) A two-dimensional Continuous Wavelet Transform is used to generate time-frequency maps of the data, taking both time and frequency domain information into account.
- (3) Reduce the amount of the data acquired using LLE and standardize it.
- (4) Minimize accuracy rate in CNN by searching for the appropriate mix of CNN parameters and SAA and then comparing the ideal architecture distribution of the network.
- (5) Input standard samples of the bearing's various states into the optimized CNN.

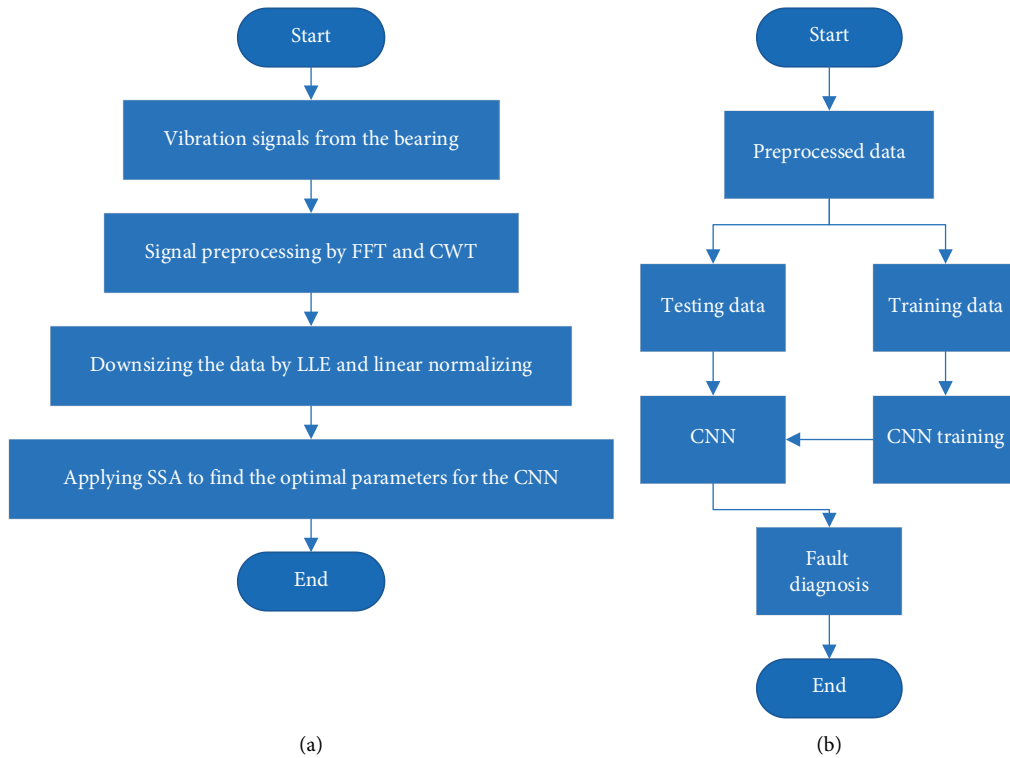


FIGURE 3: The flowchart of the proposed method. (a) Data preprocessing. (b) Bearing fault diagnosis.

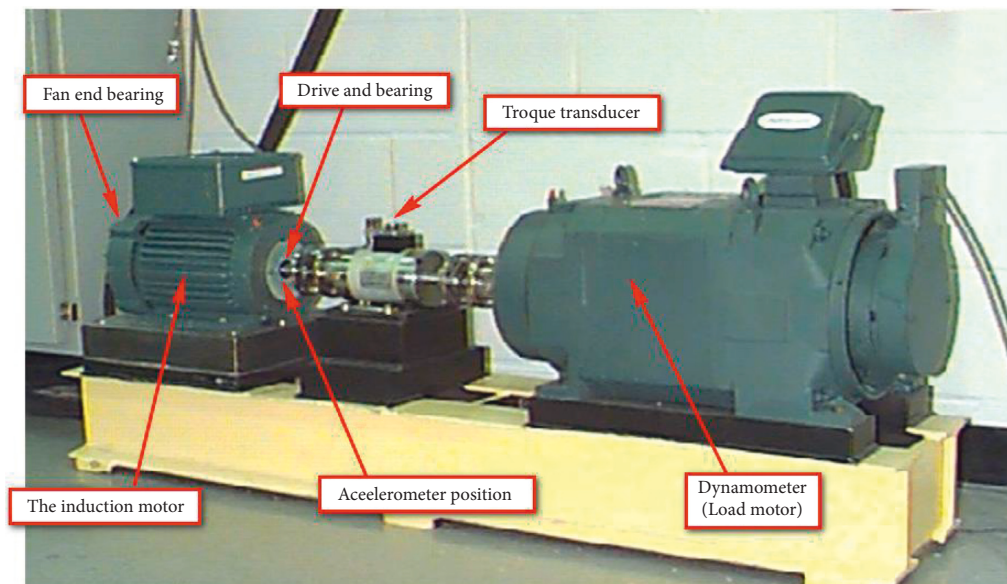


FIGURE 4: The bearing test station.

After developing a model for defect identification, test samples of the bearing in various states would be diagnosed.

#### 4. Case Study

**4.1. Experiment Setup.** The experimental data used in this study are rolling bearing vibration acceleration data from Case Western Reserve University's (CWRU) Bearing Data Center [33]. Figure 4 depicts the bearing test station. It is

equipped with an induction motor, fan end bearing, drive end bearing, torque transducer, and dynamometer (load motor). These indications originate from the accelerometer installed on the driving end of the induction motor's bearing box. The accelerometer is connected with the torque sensor and the dynamometer, as shown in Figure 5. The sampling frequency of the considered data is 48000 Hz collected from the drive end bearing (6205-2RS JEM SKF), and the samples are collected under four working conditions: (1) normal

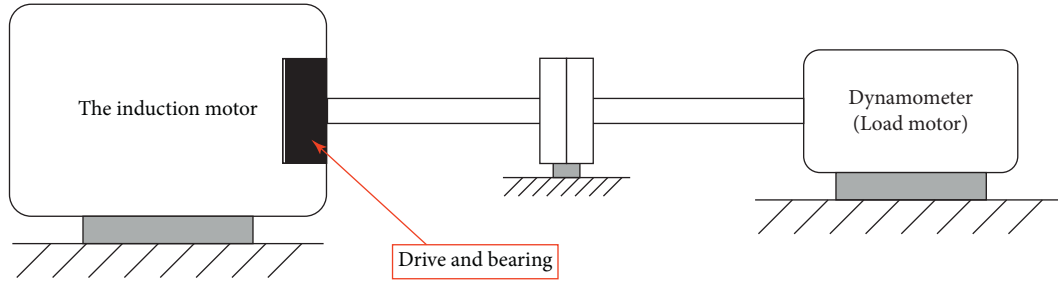


FIGURE 5: Sketch map of the structure.

TABLE 1: Parameters of the test station.

Motor load	Motor speed (r/min)	Sampling frequency (Hz)
3 PS	1730	48000

operating conditions; (2) inner race fault; (3) ball fault; (4) outer race fault. There are a total of 10 types of data, including one normal type and nine error types of different injures and fault diameters. Different components of the bearing are shown in Figure 5. The details of the data, including motor load and motor speed, are illustrated in Table 1. 2400 samples were collected from each category. Each sample contained 1000 sampling points, and a total of 24000 samples were collected. Nineteen thousand two hundred of the samples are selected for training, and the rest of the 4800 samples for testing (20% of the total). Table 2 illustrates the details of the sample sets from each condition and their corresponding status labels.

#### 4.2. Fault Diagnosis Using Parameter-Optimized CNN

**4.2.1. Data Preprocessing.** As previously stated, the vibration signal of the bearing is preprocessed. Figure 6 depicts the FFT spectrum under different operating conditions. Each signal is a superposition of multiple frequency domain components that can be dissected using frequency domain analysis. Each batch of data will be processed using FFT to obtain 1024 points in order to collect enough training samples and ensure that the model is properly trained. Because of the spectrum's symmetry, half of the data points are used as the eigenvector.

Then, the data will go through CWT to attain a series of time-frequency maps, and the diagrams of different states are shown in Figure 7. The size of each map is  $100 \times 1000$  and reshaped to  $1 \times 100000$ . Then, the processed samples of the same state are gathered to form a data group whose size is  $2400 \times 100000$ , and there are a total of ten data groups corresponding to 10 states.

**4.2.2. Data Dimension-Reduction by LLE.** The reprocessed data sets are linearly normalized to shorten training time and accelerate convergence to mitigate the effect of noise and aberrant samples on network training. As the CNN can only accept input in a specific format, the maps need to be downsized to an acceptable dimension through LLE (the number of neighbors is 15). Additionally, each group of

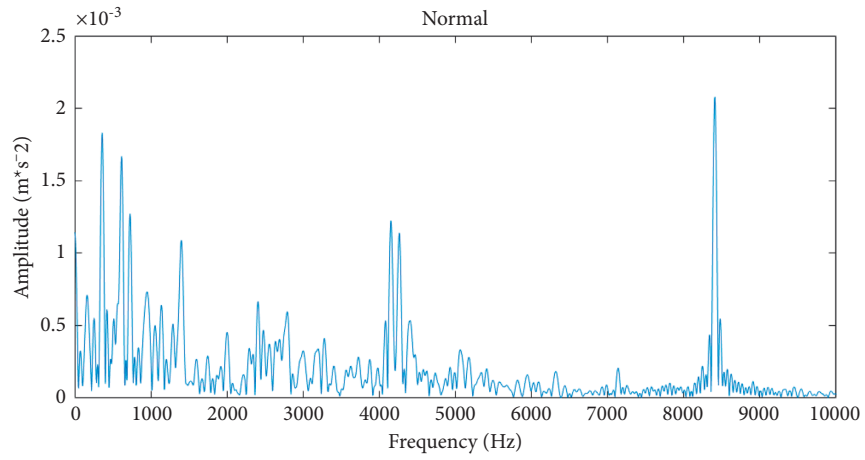
TABLE 2: Details of the sample sets.

Running state	Fault diameter	Training samples	Test samples	Status labels
Normal	//	19200	4800	1000000000
Inner race fault	0.1778 mm	19200	4800	0100000000
	0.3556 mm	19200	4800	0000100000
	0.5334 mm	19200	4800	0000000100
Ball fault	0.1778 mm	19200	4800	0010000000
	0.3556 mm	19200	4800	0000010000
	0.5334 mm	19200	4800	0000000010
Outer race fault	0.1778 mm	19200	4800	0001000000
	0.3556 mm	19200	4800	0000001000
	0.5334 mm	19200	4800	0000000001

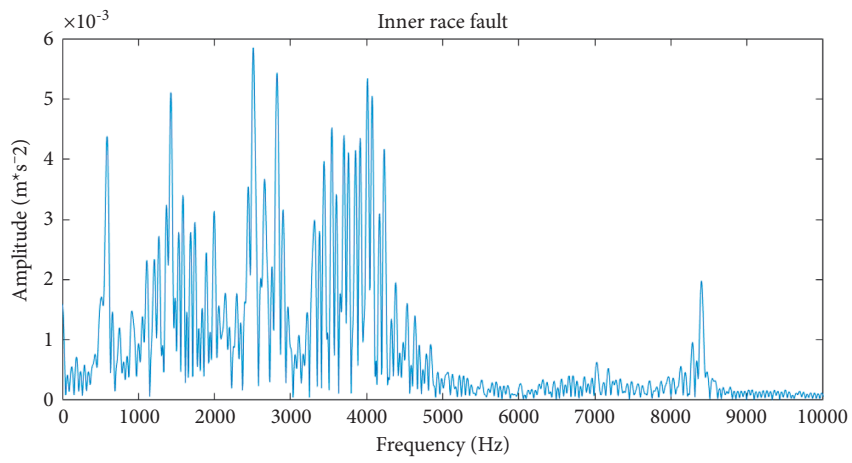
samples of different states is dimension-reduced separately to maintain the characteristics of each category of information. The dimension-reduced maps (sized  $48 \times 48$ ) are presented in Figure 8. For each data group (sized  $2400 \times 100000$ ), the dimension-reduced size is  $2400 \times 2304$ .

**4.2.3. Network Setup and Training.** Usually, CNN uses images as the input data. In this research, the dimension-reduced time-frequency maps derived from CWT and LLE are utilized as the input for the CNN with the size of  $48 \times 48$ . The input layer size has corresponded to the image's input which is  $48 \times 48$ . The network architecture is obtained from repeated experiments and references to previous kinds of literature. After inputting the images, the following is a convolutional layer with eight filters sized  $9 \times 9$  using ReLU as the activation function and a mean-pooling layer of size  $2 \times 2$ . A convolutional layer repeats this with 16 filters sized  $9 \times 9$  using ReLU as the activation function and a mean-pooling layer sized  $2 \times 2$ . The next section is fully connected layers which contain 360 neurons in the first layer and 60 neurons in the second one. ReLU is applied as the activation function for both of them. Then, the softmax function is used as the classification function. The training method of the network is SGD with ten epochs. The structure of the CNN is shown in Figure 9.

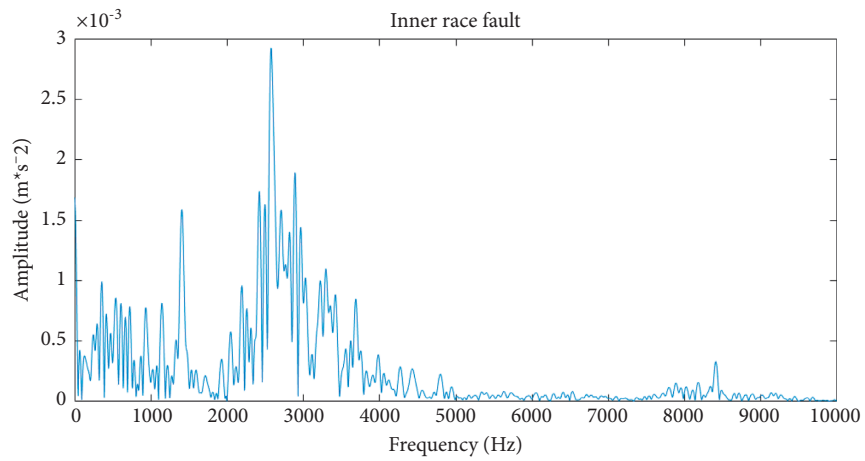
**4.2.4. Determination for Optimal Parameter of CNN with SSA.** The optimal parameters of learning rate and batch size in CNN are searched by SSA with a search range of respectively  $[0, 1]$  and  $[0, 100]$ . The detailed setting for SSA is presented in Table 3. It is also quite significant to decide the fitness of the chosen parameters during the SSA, and the



(a)



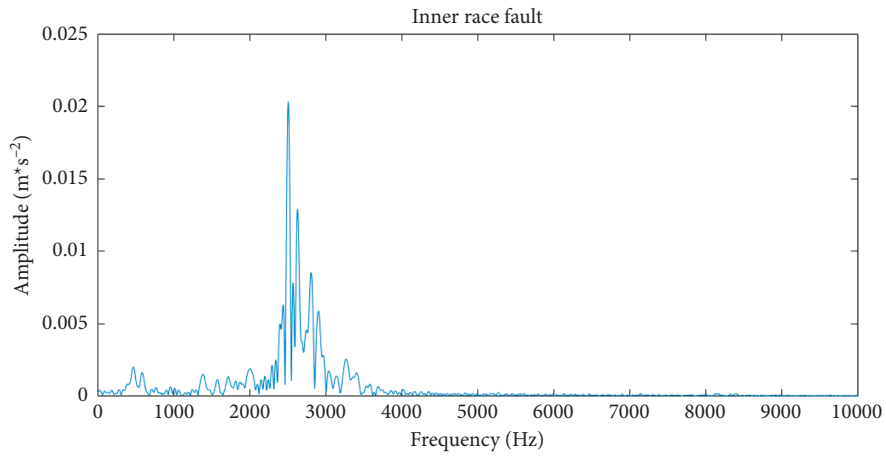
(b)



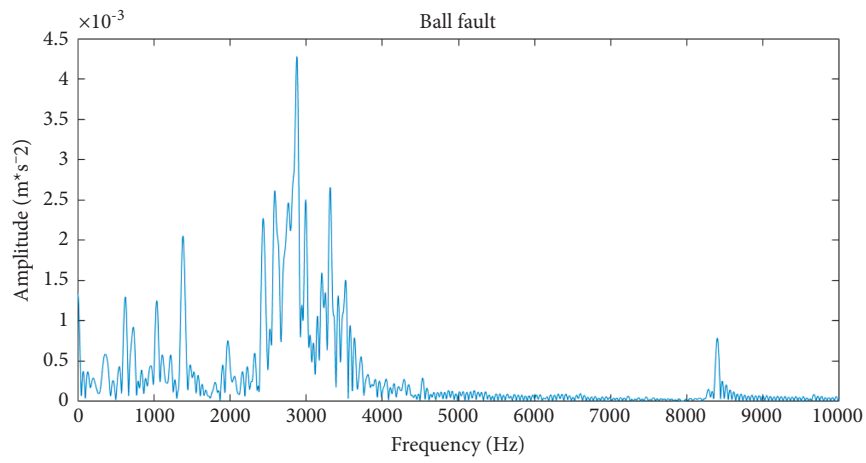
(c)

FIGURE 6: Continued.

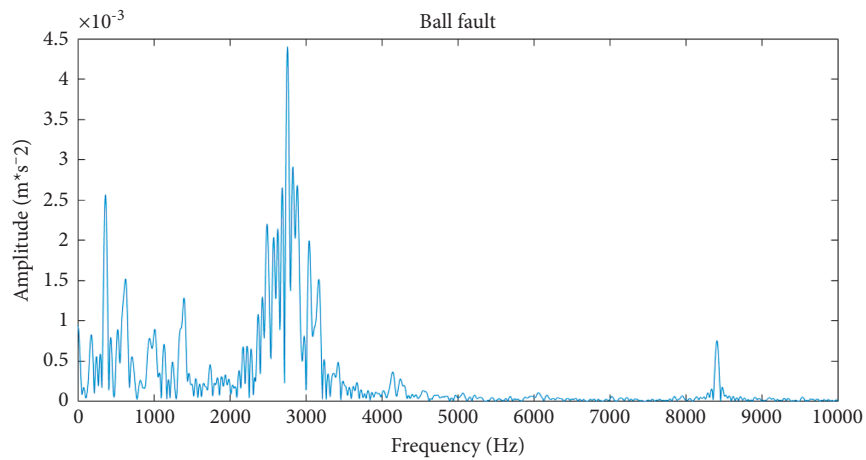




(d)

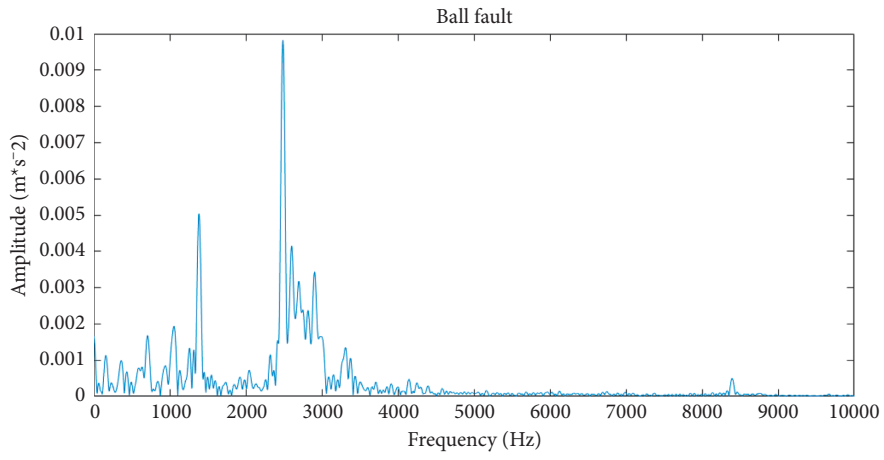


(e)

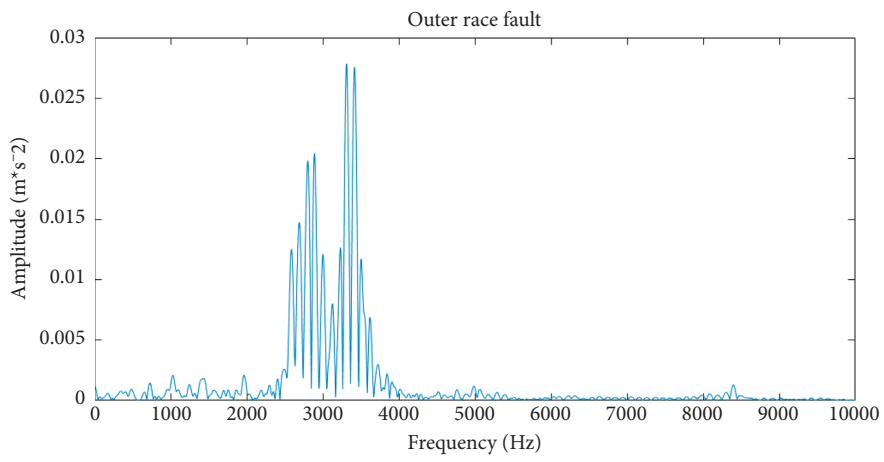


(f)

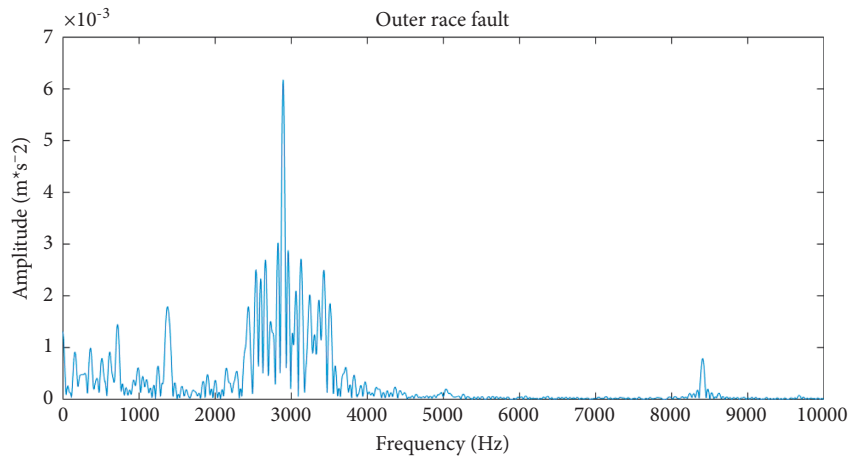
FIGURE 6: Continued.



(g)



(h)



(i)

FIGURE 6: Continued.

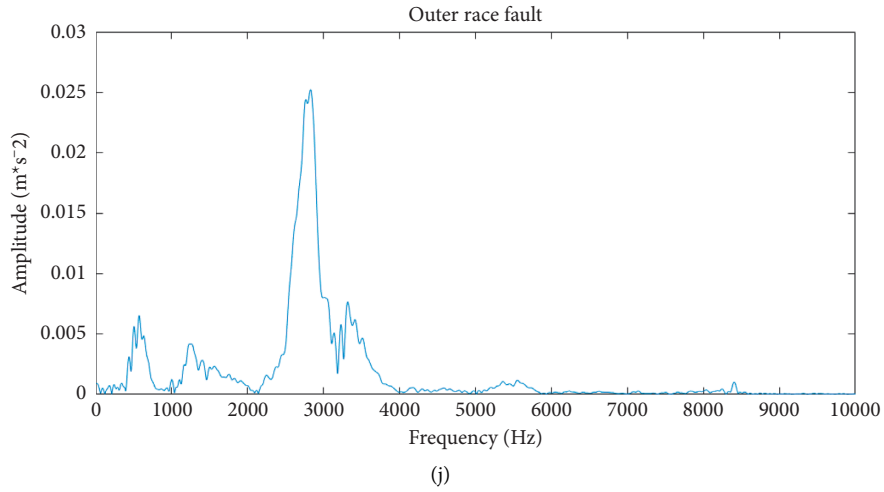


FIGURE 6: FFT spectrum of different running states: (a) normal, (b)–(d) inner race fault of different fault diameters, (e)–(g) ball race fault with different fault diameters, and (h)–(j) outer race fault of different fault diameters (from 0.1778 mm to 0.5334 mm).

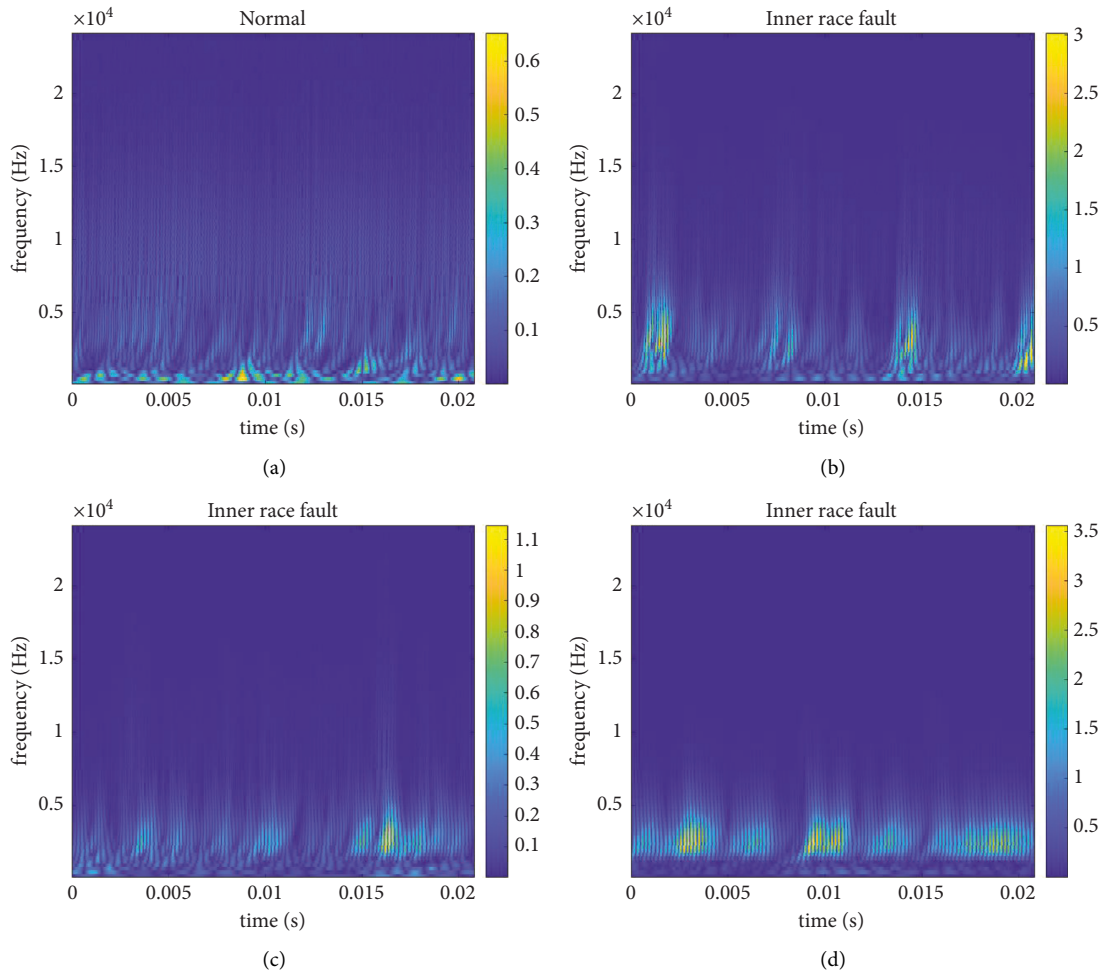


FIGURE 7: Continued.

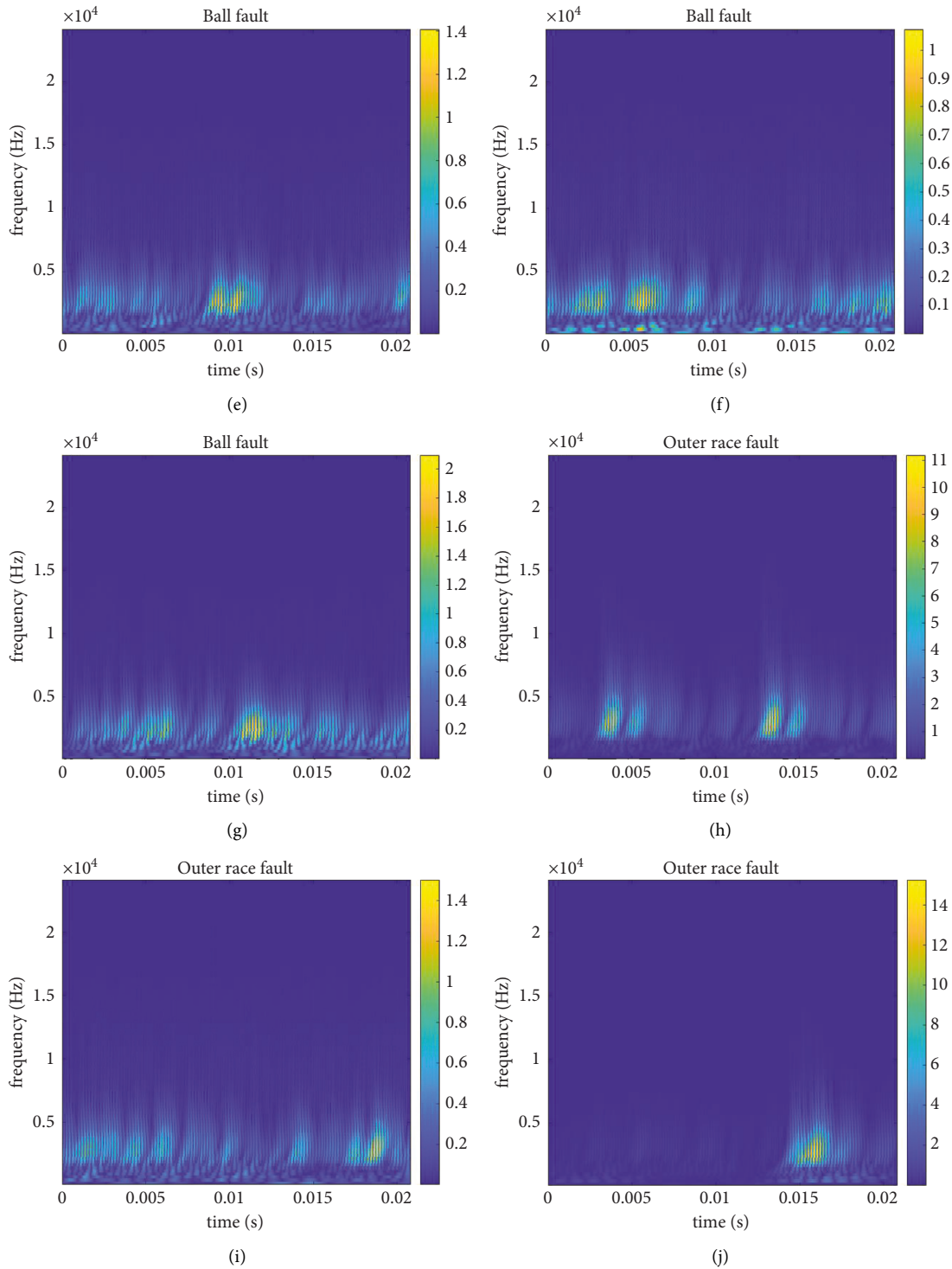


FIGURE 7: Time-frequency maps of different running states: (a) normal, (b)–(d) inner race fault of different fault diameters, (e)–(g) ball race fault with different fault diameters, and (h)–(j) outer race fault of different fault diameters (from 0.1778 mm to 0.5334 mm).

parameter determination criterion of this research is the accuracy rate of the testing, which is shown as follows:

$$\text{accuracy rate} = \frac{\text{the number of the correctly classified samples}}{\text{the total number of samples}}. \quad (19)$$

A higher accuracy rate means a better training effect and parameters fitness. Through observing the accuracy rate, the training result of the model can be judged. While increasing the number of iterations improves the outcome of the defect diagnosis, it also significantly increases the computation time. Considering the SSA's effect and the time necessary,

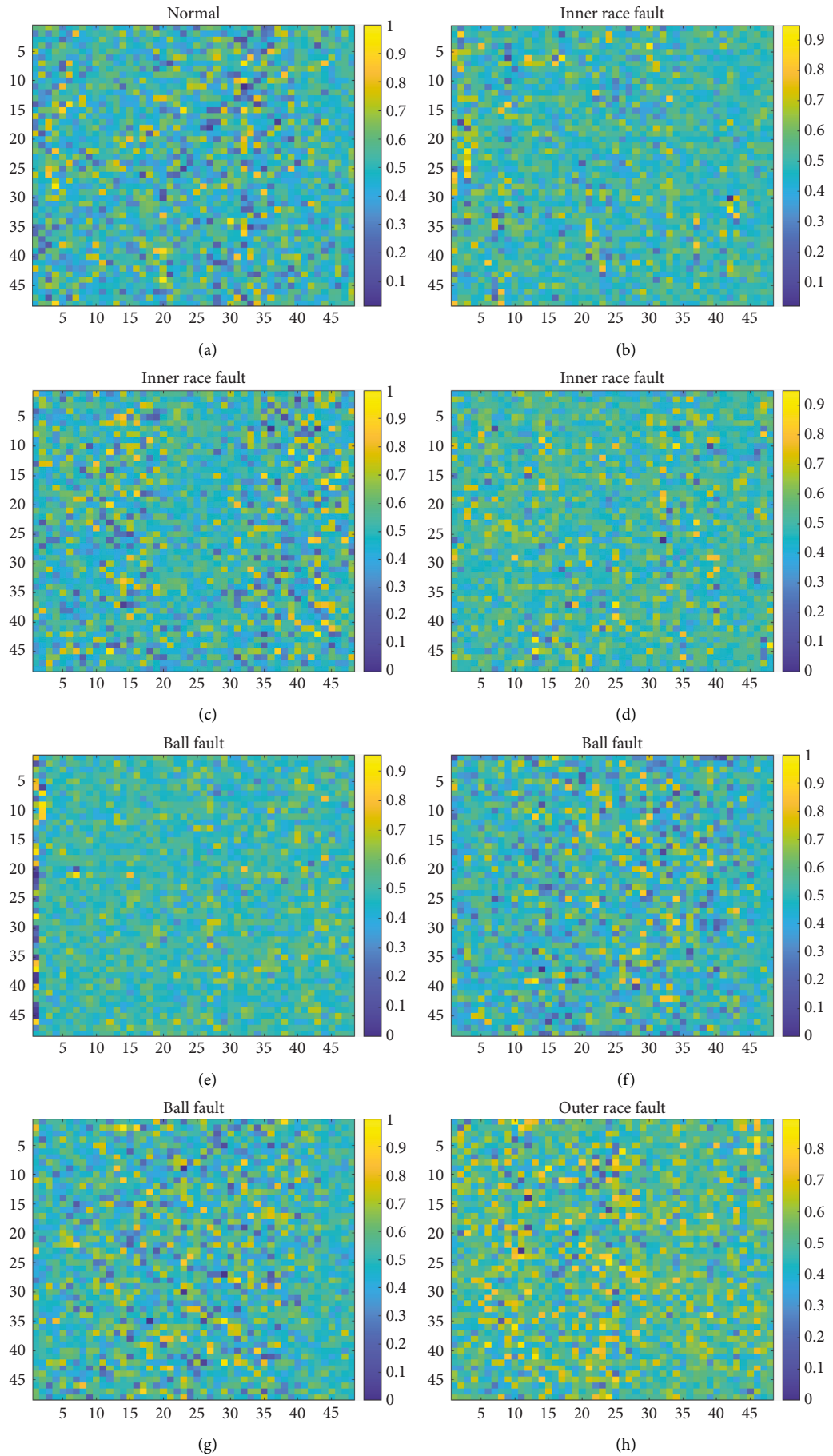


FIGURE 8: Continued.

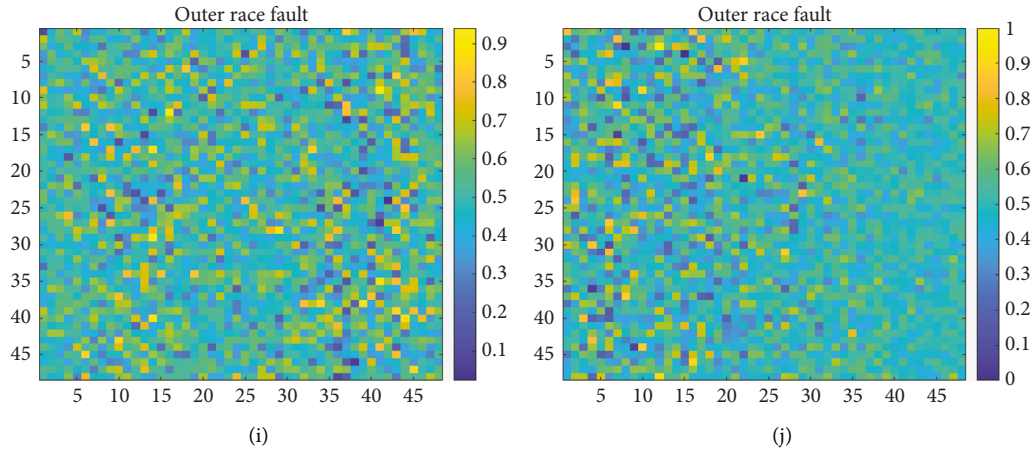


FIGURE 8: Dimension-reduced maps of different running state: (a) normal, (b–d) inner race fault of different fault diameters, (e–g) ball race fault with different fault diameters, and (h–j) outer race fault of different fault diameters (from 0.1778 mm to 0.5334 mm).

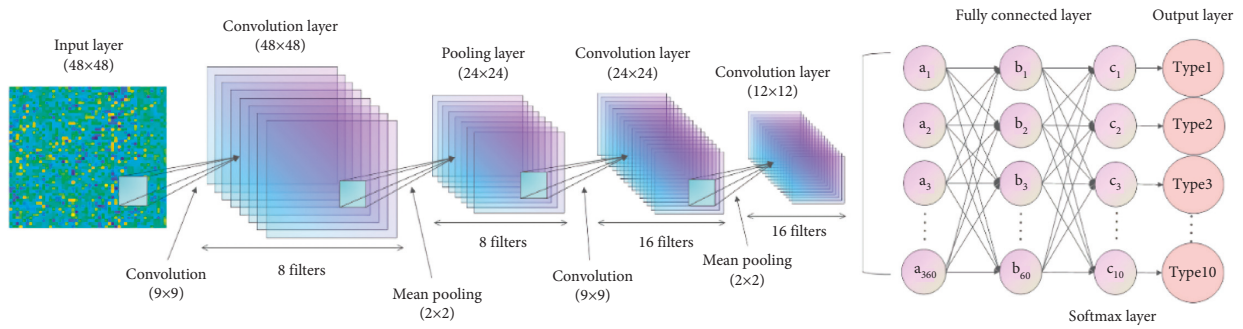


FIGURE 9: Structure of the CNN.

TABLE 3: Settings of the SSA.

Number of search agents	Maximum number of iterations	Variable dimension	Range of learning rate	Range of batch number
5	10	2	[0, 0.01]	[1, 100]

TABLE 4: Optimized parameters of CNN.

Learning rate	Batch size	Momentum value	Epoch
0.0073	80	0.95	10

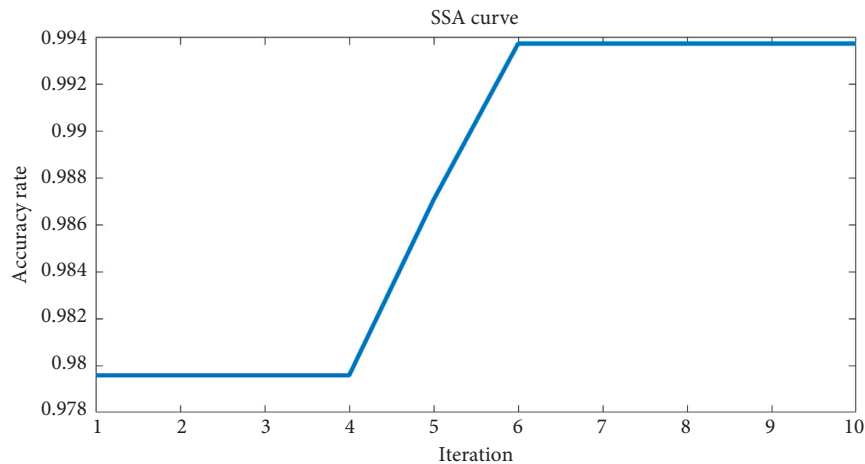


FIGURE 10: Variation of the accuracy rate.

TABLE 5: Optimization process of SAA from iteration 3 to 7.

Iteration	Batch size	Learning rate (keeping two significant digits)
3	160	0.48
	160	1.7
	160	0.0096
	160	0.0096
	4	0
4	100	0
	160	0
	160	0.0091
	4	0.0096
5	4	0.78
	4	0.27
	80	0.0073
	160	0.0096
	160	0.0099
6	4	0.18
	80	0.49
	160	0.23
	64	0.0066
	80	0.0073
7	100	0.036
	80	0.014
	80	0.0075
	40	0
	80	0.0070
	4	0

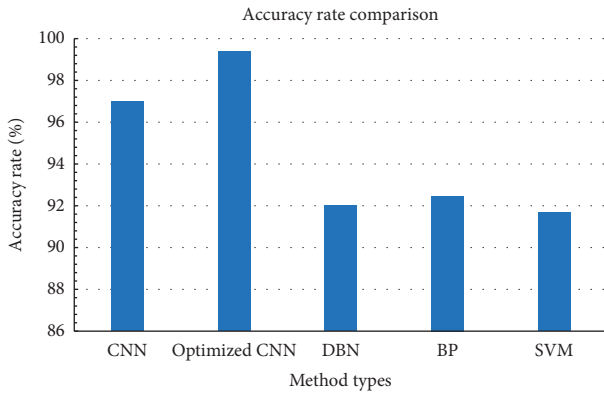


FIGURE 11: The accuracy rate comparison of different methods.

the number of iterations is set at 10. Table 4 summarizes the parameters determined for CNN. Figure 10 illustrates the variance in accuracy rate of the best optimization coefficient of each iteration as a function of the number of rounds. Table 5 shows some details about the variance of optimization coefficient during this process.

**4.2.5. Comparative Analysis of the Results with Other Methods.** To demonstrate the suggested method’s benefit in defect identification, the accuracy rate of the improved CNN is compared to the CNN established only by practice (trained ten times). The suggested technique has an accuracy rate of 99.4 percent, indicating that this model has the benefit of high accuracy, efficiency, and stability for identifying

bearing faults under a variety of conditions. The accuracy rate of the CNN using experience-based parameters is 97 percent, somewhat lower than that of the optimized one. Then, using the same experimental data provided from Case Western Reserve University’s Bearing Data Center, the accuracy of DBN is 92 percent [34], BP is 92.5 percent [35], and SVM is 91.7 percent [36]. Compared to shallow network models, deep network models are more suited for adaptive defect identification in huge data sets and complicated environments. In comparison to the deep network model, standard fault detection approaches suffer from limitations in terms of adaptive fault feature extraction, monitoring, and diagnostic accuracy, as well as generalization performance. Figure 11 illustrates the comparison figure.

### 5. Conclusion

The combination of parameters including epoch, batch size, and learning rate for CNN has a pronounced influence on training. To obtain the optimal combination of the aforementioned parameters, SSA using the accuracy rate of testing as the fitness is proposed for optimizing the parameters of CNN. Then, the improved method is utilized for the fault diagnosis, and its effectiveness is well demonstrated. Eventually, the detailed conclusions can be drawn as follows:

- (1) The excellent correctness of the parameter-optimized CNN demonstrates that the approach can adaptively extract defect information from the bearing vibration signal spectrum, reducing the need for a wide variety of signal processing techniques and diagnostic experience. It offers significant benefits in terms of defect diagnostic capability and generalization performance.
- (2) This research develops a unique integrated defect diagnostic model based on FFT, CWT, LLE, and optimized CNN that outperforms shallow layer networks and classic character recognition and pattern recognition techniques. It has practical utility in a world of huge data.

### Data Availability

All data, models, and code generated or used during the study appear in the submitted paper.

### Conflicts of Interest

The author declares no conflicts of interest.

### References

- [1] A. Caggiano, J. Zhang, V. Alfieri, F. Caiazzo, R. Gao, and R. Teti, “Machine learning-based image processing for on-line defect recognition in additive manufacturing,” *CIRP Annals*, vol. 68, no. 1, pp. 451–454, 2019.
- [2] Y. Liu, L. Guo, H. Gao, Z. You, Y. Ye, and B. Zhang, “Machine vision based condition monitoring and fault diagnosis of machine tools using information from machined surface texture: a review,” *Mechanical Systems and Signal Processing*, vol. 164, Article ID 108068, 2022.

- [3] P. Xu and Yao, "In-process adaptive dimension correction strategy for laser aided additive manufacturing using laser line scanning," *Journal of Materials Processing Technology*, vol. 303, Article ID 117544, 2022.
- [4] S. Zhang, B. Habetler, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics - a review," *2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, pp. 257–263, 2019.
- [5] M. El Hachemi Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 984–993, 2000.
- [6] X. Jin, F. Cheng, Y. Peng, W. Qiao, and L. Qu, "Drivetrain gearbox fault diagnosis: vibration- and current-based approaches," *IEEE Industry Applications Magazine*, vol. 24, no. 6, pp. 56–66, 2018.
- [7] Y. Lei, L. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: a review," *Measurement*, vol. 48, no. 1, pp. 292–305, 2014.
- [8] C. Duan, P. Gao, X. Xu, and Q. Gao, "A damage diagnosis method for rolling bearings based on time-frequency kurtosis spectrum," *Journal of Mechanical Engineering*, vol. 51, no. 11, pp. 78–83, 2015.
- [9] H. Li, *VMD Based Bi-spectrum for Bearing Fault Diagnosis*, 2021.
- [10] Y. G. Lei, F. Jia, and X. Zhou, "A deep learning-based method for machinery health monitoring with big data," *Journal of Mechanical Engineering*, vol. 51, no. 21, pp. 55–62, 2015.
- [11] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146–157, 2018.
- [12] X. W. Chen and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE Access*, vol. 2, no. 2, pp. 514–525, 2014.
- [13] *ImageNet Classification with Deep Convolutional Neural Networks*, Curran Associates Inc, 2012.
- [14] B. Liu, X. B. Shu, and Ou Qiao, "Fault diagnosis of rolling bearing based on time frequency diagram and CNN," *Journal of Nanchang Aeronautical University (NATURAL SCIENCE EDITION)*, no. 2, pp. 86–91, 2018.
- [15] Li Xu, Li Ge, and Yu Liang, "Fault diagnosis of rolling bearing based on wavelet packet and CNN," *Journal of Sichuan University of Science and Engineering (self Science Edition)*, vol. 31, no. 3, 2018.
- [16] S. Zhang, S. Zhang, B. Habetler, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics-A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.
- [17] X. Wang, J. Liu, T. Hou, and P. Chao, "The SSA-BP-based potential threat prediction for arialtarget considering commander emotion," *Defence Technology*, no. 3, 2021.
- [18] S. Soltani, "On the use of the wavelet decomposition for time series prediction[J]," *Neurocomputing*, vol. 48, no. 1-4, pp. 267–277, 2002.
- [19] S. J. Yao, Y. H. Song, L. Z. Zhang, and X. Y. Cheng, "Wavelet transform and neural networks for short-term electrical load forecasting," *Energy Conversion and Management*, vol. 41, no. 18, pp. 1975–1988, 2000.
- [20] S. K. Lee, "An acoustic decay measurement based on time-frequency analysis using wavelet transform," *Journal of Sound and Vibration*, vol. 252, no. 1, pp. 141–153, 2002.
- [21] S. Sinha, P. S. Routh, and P. D. Anno, "Spectral decomposition of seismic data with continuous wavelet transform," *Geophysics*, vol. 70, no. 6, 2008.
- [22] Y. Wang and J. Chu, "Based on Daubechies wavelet analysis, information extraction of Misfire Fault Diagnosis of automotive electronic control engine," *forest engineering*, vol. 30, no. 2, pp. 138–142, 2014.
- [23] R. Xu, F. Jiang, and H. Yao, "Summary of manifold learning," *Journal of Intelligent Systems*, vol. 1, no. 1, pp. 44–51, 2006.
- [24] H. Zhou, W. Wu, and Q. Teng, "Algorithm research in manifold learning," *computer application research*, vol. 24, no. 7, pp. 214–217, 2007.
- [25] J. B. Tenenbaum and D. Silva, "A global geometric framework for nonlinear dimensionality reduction," vol. 290, no. 5500, Science, 2000.
- [26] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems*, vol. 14, no. 6, pp. 585–591, 2001.
- [27] T. R. Sam and K. S. Lawrence, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [28] B. Sravyapranati, D. Suma, and C. Manjulatha, *Large-scale video classification with convolutional neural networks*, 2020.
- [29] E. Maggiori, Y. Tarabalka, and G. Charpiat, "Fully convolutional neural networks for remote sensing image classification," in *Proceedings of the Geoscience & Remote Sensing Symposium*, IEEE, Beijing, China, July 2016.
- [30] Y. Lecun, K. Kavukcuoglu, and E. Culurciello, "Large-scale FPGA-based convolutional networks," *Machine Learning on Very Large Data Sets*, pp. 1–26, Cambridge University Press, Cambridge, UK, 2011.
- [31] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [32] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.
- [33] K. Loparo, *Bearing Vibration Data set*, Cleve land: Case Western Reserve University, Cleveland, OH, USA, 2003.
- [34] G. Qing-sheng, L. Xiao-bo, and L. Shan, "fault diagnosis of rolling bearing based on DBN with dynamic addition algorithm," *Journal of nanchang hangkong university(Natural science edition)*, vol. 35, no. 2, pp. 92–99, 2021.
- [35] S. Jie and W. Xing, "BP neural network based bearing fault diagnosis with differential evolution & EEMD denoise," in *Proceedings of the 2017 9th International Conference on Modelling, Identification and Control (ICMIC)*, IEEE, Kunming, China, 10-12 July 2017.
- [36] Y. Bu, J. Wu, and J. Ma, "The Rolling Bearing Fault Diagnosis Based on LMD and LS-SVM," in *Proceedings of the Control & Decision Conference*, IEEE, Changsha, China, May 2014.