

Research Article

Remaining Useful Life Prediction and Fault Diagnosis of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network

Shuang Zhou,¹ Maohua Xiao ,¹ Petr Bartos,² Martin Filip,² and Guosheng Geng¹

¹College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

²Faculty of Agriculture, University of South Bohemia, Studentska 1668, Czech Republic

Correspondence should be addressed to Maohua Xiao; xiaomaohua@njau.edu.cn

Received 16 July 2020; Revised 16 August 2020; Accepted 13 September 2020; Published 5 October 2020

Academic Editor: Changqing Shen

Copyright © 2020 Shuang Zhou 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.

Rolling bearings play a pivotal role in rotating machinery. The remaining useful life prediction and fault diagnosis of bearings are crucial to condition-based maintenance. However, traditional data-driven methods usually require manual extraction of features, which needs rich signal processing theory as support and is difficult to control the efficiency. In this study, a bearing remaining life prediction and fault diagnosis method based on short-time Fourier transform (STFT) and convolutional neural network (CNN) has been proposed. First, the STFT was adopted to construct time-frequency maps of the unprocessed original vibration signals that can ensure the true and effective recovery of the fault characteristics in vibration signals. Then, the training time-frequency maps were used as an input of the CNN to train the network model. Finally, the time-frequency maps of testing signals were inputted into the network model to complete the life prediction or fault identification of rolling bearings. The rolling bearing life-cycle datasets from the Intelligent Management System were applied to verify the proposed life prediction method, showing that its accuracy reaches 99.45%, and the prediction effect is good. Multiple sets of validation experiments were conducted to verify the proposed fault diagnosis method with the open datasets from Case Western Reserve University. Results show that the proposed method can effectively identify the fault classification and the accuracy can reach 95.83%. The comparison with the fault diagnosis classification effects of backpropagation (BP) neural network, particle swarm optimization-BP, and genetic algorithm-BP further proves its superiority. The proposed method in this paper is proved to have strong ability of adaptive feature extraction, life prediction, and fault identification.

1. Introduction

Electromechanical equipment is constantly developing in the direction of large-scale, precision, systemisation, and automation. The health of the main components must be monitored in real time to maintain the safe operation of the equipment [1–4]. Bearings are crucial parts in rotating machinery, and their health directly affects the performance of the entire equipment [5, 6]. A previous analysis showed that faults caused by vibration accounted for 70% of all mechanical faults. Faulty rolling bearings accounted for 30% of the faults caused by vibration [7]. Improving the accuracy of remaining life prediction and fault identification plays an important role when making correct

maintenance decisions on the basis of the operating status of equipment.

Traditional remaining useful life prediction methods require two basic steps, namely, establishing performance degradation indicators and studying the prediction model. One method for constructing performance degradation indicators is to extract a single statistical feature from the original signal and use it as a performance degradation indicator, such as the root mean square and kurtosis adopted by Li et al. [8], spectral entropy based on multiscale morphological decomposition adopted by Wang et al. [9], and binary multiscale entropy proposed by Li et al. [10]. These methods require substantial experience and professional knowledge [11]. A single statistical feature cannot ensure

subsequent prediction accuracy. Construction methods are used for performance degradation indicators that combine multiple indicators. Wang et al. [12] proposed a principal component analysis- (PCA-) based feature extraction method for rolling bearing fault classification. Liu et al. [13] introduced a new feature extraction method to obtain sensitive features through phase space reconstruction and joint feature matrix approximate diagonalisation. Wang et al. [14] reduced the dimension of extracted features through k-means PCA dimensionality reduction method and proposed a trend prediction method based on Pchip-EEMD-GM(1,1) to predict the remaining service life of rolling bearings. Part of the information is lost using data dimensionality reduction for feature fusion, and whether they are all correlated with the remaining life of the bearings remains uncertain. Various data-driven prediction methods for the remaining life of mechanical equipment, including the hidden Markov model adopted by Liu et al. [15], support vector machine (SVM) used by Shen et al. [16], and wavelet SVM used by Chen et al. [17], are proposed. The above methods all need to manually construct features, and the method used has a gap in model complexity and learning ability compared with the deep learning method represented by convolutional neural network (CNN). Thus, these methods can still be improved in terms of prediction accuracy.

Scholars have conducted numerous researches on bearing fault diagnosis. Liu et al. [18] proposed a bearing fault diagnosis method based on short-term matching method and SVM, which is superior to the traditional method in terms of weak shock signal oscillation feature extraction and early fault diagnosis. Attoui et al. [19] introduced a new time-frequency analysis method for identifying and classifying bearing faults and validated its effectiveness using public bearing data sets. Zheng et al. [20] proposed a bearing fault diagnosis method based on compound multiscale fuzzy entropy and SVM that can effectively identify different bearing fault types and fault severity. Yao et al. [21] combined the integration of empirical mode decomposition, permutation entropy, and improved SVM, thereby improving the accuracy of fault classification. Zheng et al. [22] presented a bearing fault diagnosis method based on generalised compound multiscale permutation entropy, PCA, and SVM and studied the influence of parameters on the calculation of generalised compound multiscale permutation entropy. Qi et al. [23] developed a bearing fault feature extraction method based on variational modal decomposition, local tangent space arrangement algorithm, and K-means classifier. The test results showed that the features obtained after dimensionality reduction are sensitive to faults. Yang et al. [24] used an extended ant lion optimiser to optimise the SVM model parameters, thereby improving the recognition rate of bearing failures at different test speeds. Jiang et al. [25] proposed a novel ICF-guided VMD method to accurately extract the weak damage features of rotating machines.

From the above analysis, traditional data-driven bearing life prediction or fault diagnosis methods usually need to understand certain signal processing techniques, manually

construct algorithms to extract and select features, and use classifiers for prediction and classification recognition. However, designing fault features that cover all information is difficult for mechanical big data with unclear and variable patterns and multifault information coupling. The features extracted and selected for a particular problem may be inapplicable to other problems. Therefore, adaptively extracting features using the model is necessary rather than manually extracting and selecting features.

In this study, short-time Fourier transform (STFT) and CNN are combined for bearing life prediction and fault diagnosis research. The rest of the paper is organized as follows. The relevant theoretical bases are briefly introduced in Section 2. A bearing remaining useful life prediction and fault diagnosis model using STFT and CNN is established in Section 3. In Section 4, the second datasets of the Intelligent Management System (IMS) bearing full-cycle life data are selected and converted into time-frequency maps and are inputted into the CNN for training and testing for life prediction. In Section 5, the proposed method is used to convert the time-frequency maps of the Case Western Reserve University (CWRU) bearing datasets, and the fault classification is completed by the CNN via training and testing. The proposed method is compared with the back-propagation NN (BPNN), particle swarm optimisation (PSO)-BP, and genetic algorithm (GA)-BP in terms of fault diagnosis classification effects. Finally, the conclusions are drawn in Section 6.

2. Fundamental Theory

2.1. STFT. STFT is widely used for studying nonstationary signals [26]. It can combine the time domain and frequency domain analyses of signals. Its basic idea is to divide the nonstationary signal into many small time intervals and consider it to be stationary in these time intervals; perform Fourier transform on these signals in different time intervals; synthesise the Fourier transform of each time interval. This operation avoids the loss of fault information, which can be regarded as a mapping from the time domain to the time-frequency domain. Therefore, STFT is a 2D function of time and frequency.

Assuming a nonstationary signal $S(t)$, its STFT can be defined as

$$\text{STFT}_s(t, \omega) = \int_{-\infty}^{+\infty} S(\tau)h(\tau - t)e^{-j\omega\tau} d\tau, \quad (1)$$

where t is the time shift parameter, and $h(t)$ is the window function centred on t . The signal is truncated by the window function and divided into multiple segments. The intercepted signal can be expressed as

$$S_t = S(\tau)h(\tau - t), \quad (2)$$

where S_t is the signal at the fixed time t corresponding to the original signal, and $S(\tau)$ is the signal at τ corresponding to the execution time. A Fourier transform is performed on S_t to obtain the spectrum and is expressed as

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} S_t(\tau) d\tau = \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} S_t(\tau) h(\tau - t) d\tau. \quad (3)$$

The centre position of the window function can be changed by changing the size of translation parameter t to obtain the Fourier transform at different times. In each different time interval, a different frequency spectrum can be obtained, and the totality of these spectrums constitutes a time-frequency distribution.

2.2. CNN. CNN is a type of deep feedforward artificial NN. It is a deep learning network specially designed for extracting features of 2D signal structure. It is suitable for feature learning and recognition of time-frequency maps. The training of CNN is composed of the forward propagation of calculating the loss function through the convolution, pooling and fully connected layers, and the backward propagation of updating the network parameters layer by layer using the gradient descent algorithm [27–29].

2.2.1. Convolutional Layer. The convolution operation is where the convolution kernel slides on the picture matrix, the values at the corresponding position multiply, and accumulating the result to form a new picture matrix. The convolution kernel slides a feature picture to map it and complete feature extraction. After the convolution operation, the signal characteristics can be enhanced, and the noise interference can be reduced.

Assuming a picture P , the grey value at the point on the x^{th} row and y^{th} column is represented as $f(x, y)$, and the pixel size is $M \times N$. The convolution kernel can be expressed as $k(x, y)$, and its size is $a \times b$. The weight value on the convolution kernel indicates the contribution ability of the corresponding position point to the final convolution effect. The larger the weight, the greater the contribution ability. The matrix obtained by the convolution calculation of picture P and convolution kernel k is expressed as

$$C(s, t) = f(x, y) * k(x, y) = \sum_{x=1}^a \sum_{y=1}^b k(x, y) f(s+x-1, t+y-1), \quad (4)$$

where the ranges of values of s and t are $[1, M - a + 1]$ and $[1, N - b + 1]$, respectively.

The mathematical structure of the convolution layer is expressed as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j\right), \quad (5)$$

where l represents the current layer number, k_{ij} is the weight matrix of the convolution kernel, M_j represents the set of

input feature maps, and b_j is an offset term corresponding to each feature in the convolution layer.

2.2.2. Activation Function. After the convolution operation, the activation function will perform a nonlinear transformation on the logit value of each convolution output. The activation function aims to map the originally linear indivisible multidimensional features to another space, where the linear separability of the features will be enhanced. The activation functions commonly used in NNs are the “sigmoid” function, hyperbolic tangent function “tanh,” and “rectified linear unit” (ReLU), which are expressed as

$$a^{l(i,j)} = \text{Sigmoid}(y^{l(i,j)}) = \frac{1}{1 + e^{-y^{l(i,j)}}}, \quad (6)$$

$$a^{l(i,j)} = \text{Tanh}(y^{l(i,j)}) = \frac{e^{y^{l(i,j)}} - e^{-y^{l(i,j)}}}{e^{y^{l(i,j)}} + e^{-y^{l(i,j)}}}, \quad (7)$$

$$a^{l(i,j)} = f(y^{l(i,j)}) = \max\{0, y^{l(i,j)}\}, \quad (8)$$

where $a^{l(i,j)}$ is the activation value of the output $y^{l(i,j)}$ of the convolution layer.

2.2.3. Pooling Layer. The pooling operation can reduce the dimension of the feature map and ensure the invariance of the feature position. “Average pooling” and “max pooling” are commonly used. They output the mean and maximum values of neurons in the perceptual domain, respectively. Their mathematical descriptions are expressed as follows:

$$p^{l(i,j)} = \frac{1}{W} \sum_{t=(j-1)W+1}^{jW} a^{l(i,t)}, \quad (9)$$

$$p^{l(i,j)} = \max_{(j-1)W+1 \leq t \leq jW} \{a^{l(i,t)}\}, \quad (10)$$

where $a^{l(i,t)}$ is the activation value of the t^{th} neuron in the i^{th} frame of the l^{th} layer, and W is the width of the pooling area.

2.2.4. Fully Connected Layer. This layer is located before the output layer of the network. Its essence is a fully connected network, that is, each neuron in the network is connected to other neurons of different layers, and its parameters are the most in the entire network. Its mathematical model is

$$y = f(w \cdot x + b), \quad (11)$$

where y is the output matrix, x is the input matrix, w is the weight, b is the bias term, and f is the activation function.

2.2.5. BP and Updating of Parameters. The first-order partial derivatives of the loss function for each parameter in each

layer are calculated. The weights and offsets of the network are updated layer by layer using the gradient descent algorithm until the loss function converges or reaches the set number of iterations.

The least mean square error function is usually selected as the loss function:

$$E = \frac{1}{2} \sum_{k=1}^n (t_k^n - y_k^n)^2 = \frac{1}{2} \|t^n - y^n\|^2, \quad (12)$$

where E is the loss function, n is the number of samples, t is the actual output value, and y is the expected output value.

The gradient descent method is

$$\begin{cases} w' = w - \alpha \frac{\partial E}{\partial w}, \\ b' = b - \alpha \frac{\partial E}{\partial b}, \end{cases} \quad (13)$$

where w' and b' are the weight and offset after one iteration, w and b are the weight and offset of the previous time, E is the loss function, and α is the learning rate that determines the step size of the parameter updating process.

2.3. CNN Training Process. This process is shown in Figure 1, and its specific steps are summarized as follows:

- (1) Initialize the weight matrix and offset vector of the CNN.
- (2) The feature matrix is inputted from the CNN input layer, passes through the CONL, PL and FC layers, and outputs the predicted value through the CNN output layer.
- (3) Establish the error function between the true target value and the predicted value of the CNN output, that is, the loss function, and calculate the loss function value.
- (4) If the loss value is greater than the expected value, the loss value will be backpropagated to the CNN network to calculate the error value, the weight matrix and bias vector are continuously updated, and the training starts from the second step again. The training process is stopped when the loss value is less than or equal to the expected value.

3. Method and Steps

A method based on STFT and CNN is proposed to realise fault feature self-extraction and effectively complete bearing remaining life prediction and fault diagnosis for overcoming the limitations of bearing remaining useful life prediction and fault diagnosis based on “shallow learning” algorithm and 1D CNN. The specific process is shown in Figure 2.

The proposed method can be roughly divided into nine steps as follows:

- (1) Collecting rolling bearing vibration signals.

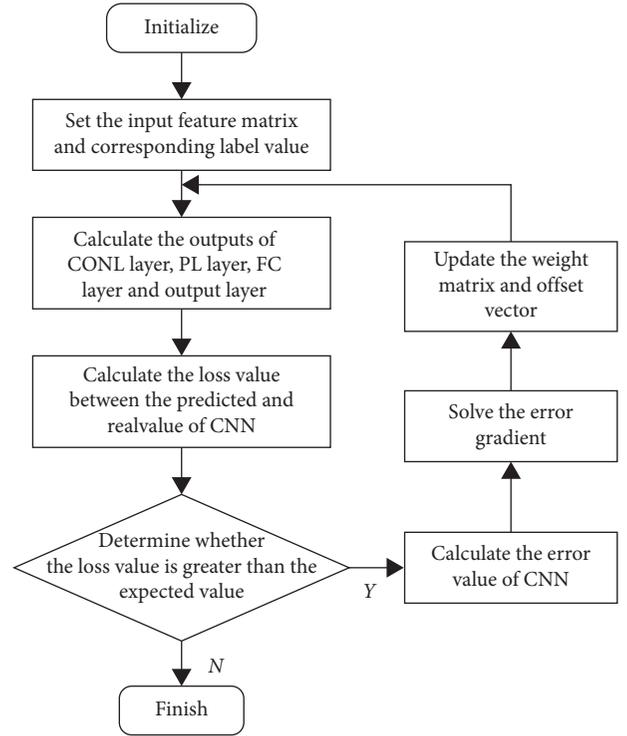


FIGURE 1: CNN training process flow chart.

- (2) Transform the initial vibration signals into time-frequency graphs using STFT.
- (3) Process the time-frequency maps. Delete the non-featured parts of the time-frequency map (including blank spaces and coordinates) and compress the picture into squares of suitable size.
- (4) Set up the network and initialize the network parameters. In accordance with the samples and requirements, build a network model of appropriate depth and determine the network parameters (such as learning rate, number of iterations, and step size).
- (5) Network training and forward propagation. Input the sample into the network and obtain the error between the network output and the expected target through forward propagation.
- (6) Determine whether the network converges. If the network converges, go to step 8; otherwise, go to step 7.
- (7) BP and weight modification. Using the BP algorithm, the error obtained in step 5 is propagated back to each node layer by layer, and the weights are updated. Steps 5 to 7 are repeated until the network converges.
- (8) Determine whether the network meets the actual requirements in accordance with the accuracy of the test sample. If the requirements are met, go to step 9; otherwise, jump to step 4 and modify the network parameters.
- (9) The network is outputted for bearing remaining useful life prediction and fault diagnosis.

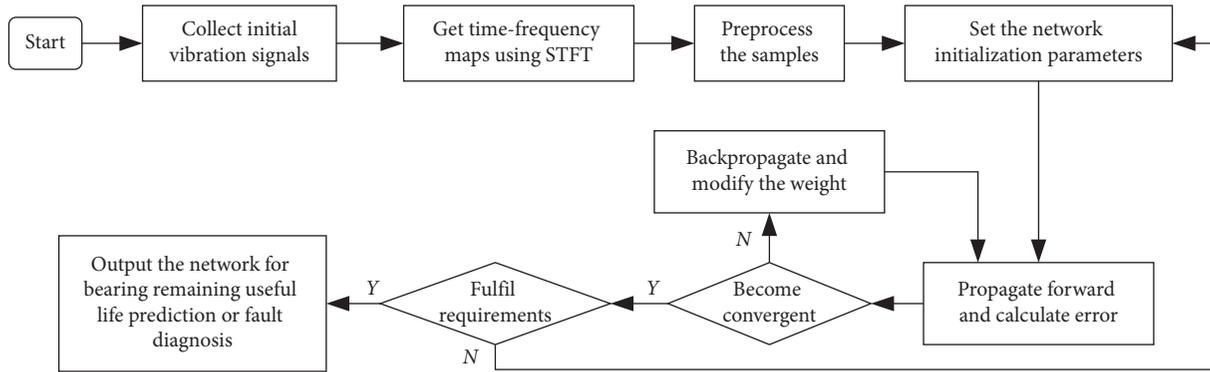


FIGURE 2: Bearing remaining useful life prediction and fault diagnosis based on STFT and CNN.

4. Bearing Remaining Useful Life Prediction Based on CNN

4.1. Experimental Data Description. The rolling bearing life cycle data sets for this experiment were provided by the Centre for IMS, University of Cincinnati [30]. The experimental data sets are collected from run-to-failure bearing test on the designed test platform, as shown in Figure 3.

Four bearings were installed on a shaft. The rotation speed was kept at 2,000 r/min using an AC motor coupled to the shaft via rub belts. A 6,000 lb radial load was added to the shaft and the bearings via a spring mechanism. A PCB 353B33 High Sensitivity Quartz ICP Accelerometer was installed on bearing housing. A vibration data sample of 20,480 points was collected every 10 min with a National Instruments DAQCard-6062E data acquisition card. The sampling rate was set to 20 kHz. The test was stopped when the accumulated debris that adhered to the magnetic plug exceeded a certain level, thereby causing an electrical switch to close [31].

Figure 4 displays the vibration signals of the tested bearing during its entire life cycle. The results show that the amplitudes of the vibration signals are stable in the early stage, whereas they significantly increase compared with the normal standard amplitude at the end of bearing life, indicating that the vibration signals contain the bearing performance degradation.

4.2. Construction of Time-Frequency Maps. The construction steps of time-frequency maps are as follows: transform the vibration signals into time-frequency maps using STFT, remove the coordinates, text, energy bars, and blank parts around the time-frequency maps, greyscale the time-frequency maps, and normalise the compression of the grid. The sample data collected in the experiment are converted into time-frequency maps using the above operation. To facilitate processing, each sample set forms a time-frequency map, and a total of 984 time-frequency maps are obtained, representing the life of the bearing at different stages from normal to failure. Some pictures are randomly selected, as shown in Figure 5. All the time-frequency maps are randomly classified into two groups, where 884 of them are chosen as training samples, and the remaining maps are

selected as test samples. They are taken as inputs for network training and testing, respectively.

4.3. Network Construction. The structure of the convolutional network layer directly affects the effect of the network model. Thus, the selection of network structure parameters is particularly important.

The derivative values of the sigmoid and tanh functions are close to zero when the absolute value of the input is large. This condition causes the error value not to propagate downward with the increase in the number of neural network layers when using the error BP to update the weight. Thus, the underlying network training is not thorough. This condition is also called the gradient dispersion phenomenon. The derivative value of the ReLU function is consistently one when the input value is greater than zero, which can overcome the gradient dispersion phenomenon well.

A “dropout” method is introduced in the fully connected layer. In particular, a certain neuron in the hidden layer stops working with a certain probability during the training process, thereby improving the generalization ability of the network and preventing overfitting.

In accordance with a large number of experiments, the network structure of CNN is composed of 1 input layer, 4 convolutional layers, 2 pooling layers, 1 fully connected layer, 4 ReLU activation layers, and 1 output layer. The size of the time-frequency map to be inputted via the input layer is 64×64 . The sizes of the convolution kernels of the four convolutional layers are all 5×5 , and the numbers of convolution kernels are 1, 16, 16, and 16, respectively. The two pooling layers are 2×2 in size. The ReLU is chosen as the activation function. The dropout and the learning rate are set to 30% and 0.01, respectively.

4.4. Experimental Result Analysis. The training and testing samples of time-frequency graphs characterising bearing life are inputted into the established CNN to perform full-cycle life prediction. Figure 6 shows an iterative graph of the accuracy and loss function during its training and testing. The accuracy rapidly rises when the number of iterations ranges from approximately 0–20. The accuracy reaches a relatively high value, tends to a stable value, and reaches 99.45% when the number of iterations reaches 50. On the

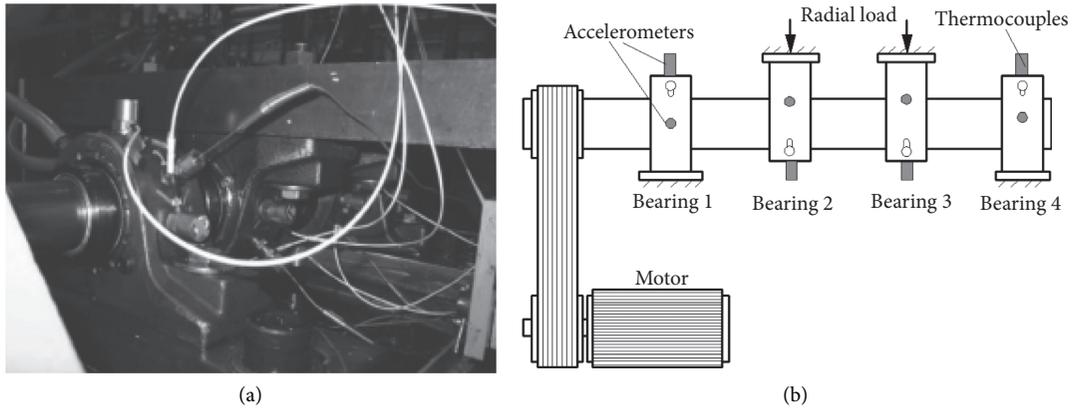


FIGURE 3: Bearing test platform. (a) Bearing test rig. (b) Sensor placement.

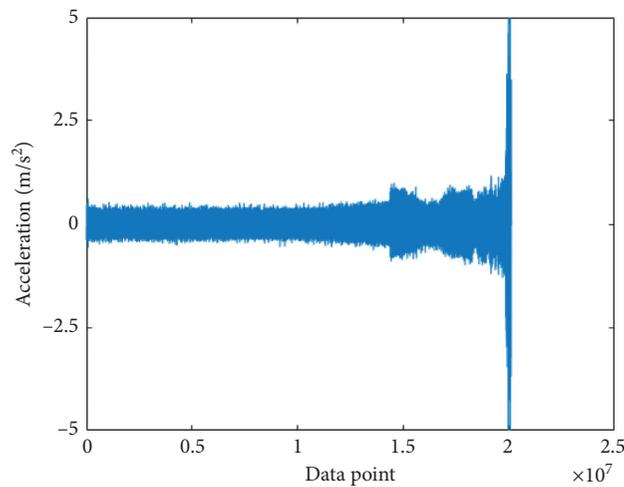


FIGURE 4: Run-to-failure vibration signals of bearing 1.

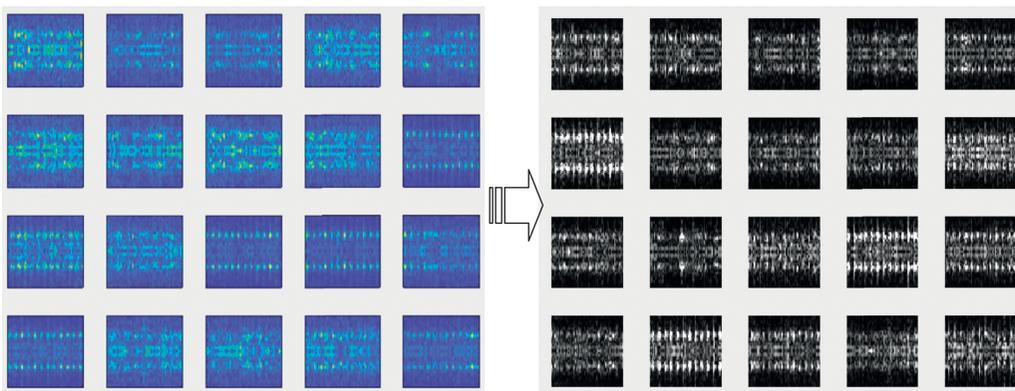


FIGURE 5: Partial time-frequency maps.

contrary, the loss function keeps decreasing and approaching zero, with a final value of 0.024. At this time, the network training has converged.

In the data set selected, the bearing experiences 9,840 min from normal operation to failure, which

represents the full cycle life of the bearing. For different sample sets, each sample corresponds to a lifetime value. In other words, the first sample set is measured when the bearing is only used for 10 min, and its remaining life is 9,830 min. In accordance with this rule, the last sample set

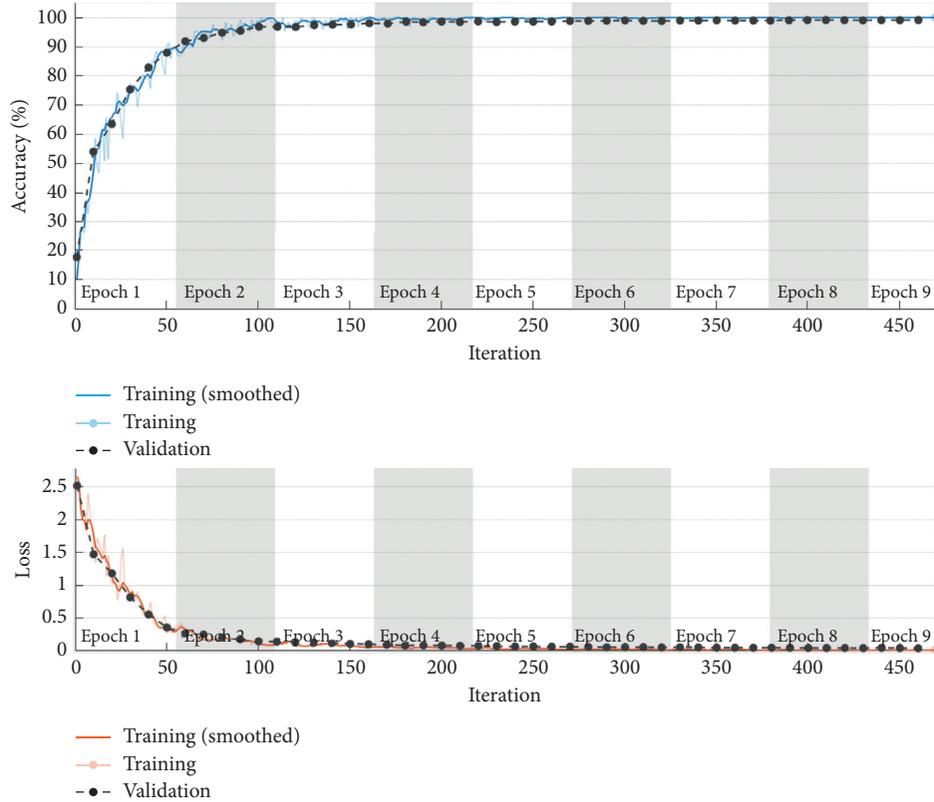


FIGURE 6: Iteration diagram of training and testing processes.

corresponds to a bearing life value of zero. The life prediction of all the sample sets was conducted, and the graphs relative to the true values were obtained, as shown in Figure 7.

To concretely show the comparison as previously mentioned, 100 testing samples are randomly selected, and their predicted and true remaining life value curves are drawn, as shown in Figure 8(a). Then, normalization is applied to process them, and the results are obtained, as shown in Figure 8(b).

The error between the real situation and the predicted situation during the testing phase can be obtained, as shown in Figure 9. The result is basically consistent with the final result obtained by continuous iteration during the previous training. From the error distribution diagram, the result is basically kept within a small error range, as shown in Figure 10. Therefore, the proposed bearing remaining life prediction method based on STFT and CNN has a good effect and can achieve a high estimation of bearing life.

5. Bearing Fault Diagnosis Based on CNN

5.1. Experimental Data Description. The vibration signal data comes from the open data set of CWRU [32]. The bearing test device is shown in Figure 11. In this experiment, the bearing at the motor drive end was used as the diagnostic object. Single-point damage was introduced on the inner ring, outer ring, and rolling elements of the test bearing to simulate three types of bearing faults. The damage

dimensions were 0.007, 0.014, and 0.021 inches, respectively. Under four different working conditions (different loads and speeds), the signals were collected with the acceleration sensor on the upper side of the motor drive end. The sampling frequency is 12 kHz.

The collected vibration signals are divided into many samples, and each sample contains 1,024 sampling points. Two hundred forty samples are found in the normal state at each motor speed. With regard to different fault types, damage sizes, and working conditions, 80 samples are found separately. Nine hundred sixty samples of the normal state and three fault types can be obtained. Table 1 shows the list of signal samples for different states of the bearing.

Figure 12 shows the vibration signal waveform diagram of the bearing in different states. Figures 12(a)–12(d) show the time domain waveforms of the bearing vibration signals under the circumstances of normal, inner ring failure, outer ring failure, and rolling element failure, respectively. Certain differences are found in the time domain waveform of different states, but completing the recognition of the signal state is impossible for nonprofessionals. These signals are only individual ideal signals. The signal waveforms of some states are extremely similar and difficult to distinguish. Therefore, solely relying on the time domain waveform of the signal for state recognition is unreliable.

5.2. Construction of Time-Frequency Maps. In this experiment, all the sample data in the above table are converted into time-frequency maps, and the sizes of the greyscale

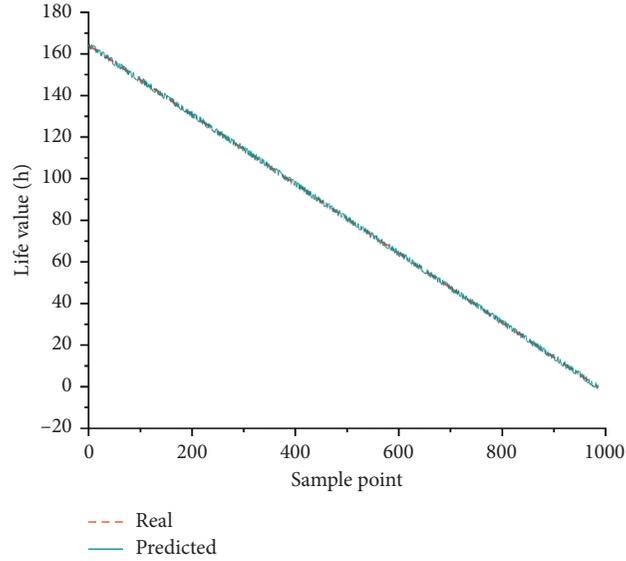


FIGURE 7: Comparison of the life values of all samples.

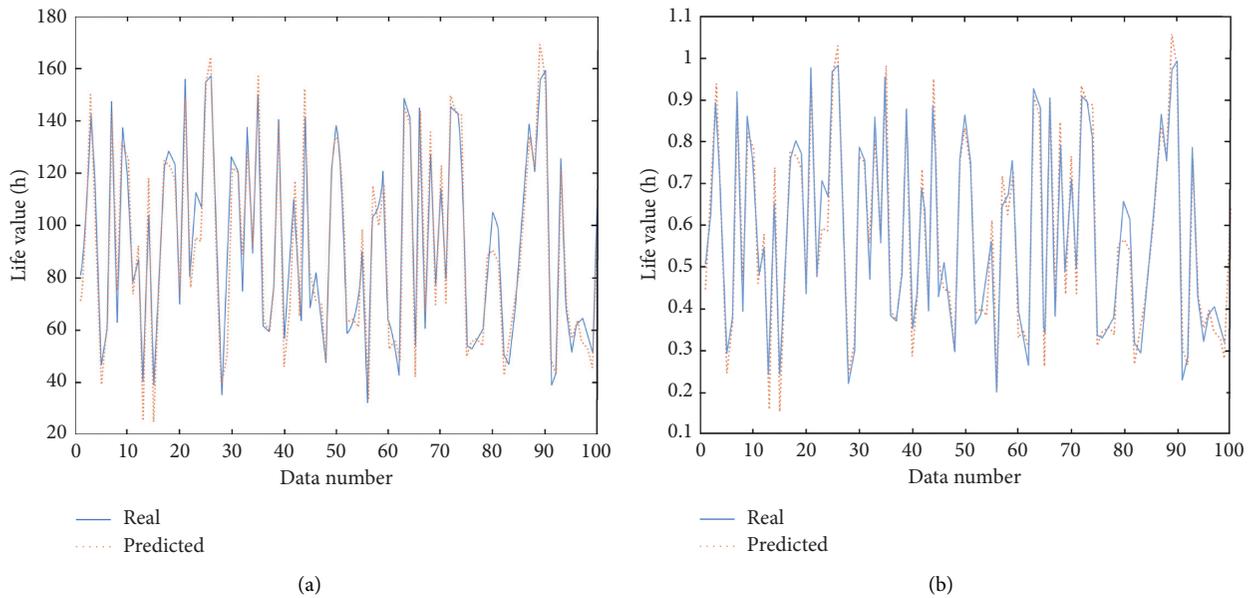


FIGURE 8: Comparison of the life values of testing samples. (a) Initial comparison. (b) Comparison after normalization.

time-frequency map are compressed to 36×36 . Figure 13 shows the final time-frequency maps randomly selected from normal signals and three types of fault samples. The time-frequency maps corresponding to the vibration signals in different states have evident differences, thereby providing the possibility for correctly completing the fault diagnosis and classification.

Nine hundred sixty samples of feature maps in each state of the bearing were obtained, where 768 of them were selected as training samples, and the remaining 192 as test samples. All the sample sets are configured with labels, as shown in Table 2.

5.3. Network Construction. The CNN model constructed in this experiment is mainly composed of input, convolutional, pooling, fully connected and output layers. Two convolution layers are found. The number of convolution kernels in each layer is 12, and the size of the convolution kernel is 5×5 . Two pooling layers are also found, and mean pooling is adopted. ReLU is used as the activation function. The learning rate is set to 0.06, and the dropout is set to 0.2.

A softmax classifier is used to classify the output after the fully connected layer. Softmax [33] is a generalization of logistic classifier that mainly solves the multiclassification problem. Assuming that the input sample in the training

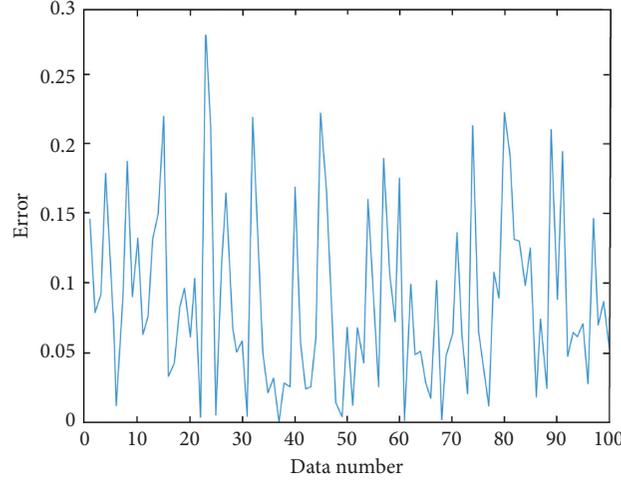


FIGURE 9: Predicted error curve.

TABLE 1: List of fault samples.

Damage size (inch)	Motor speed (r/min)	Normal signal	Inner ring failure	Outer ring failure	Rolling element failure
0	1797	240	—	—	—
	1772	240	—	—	—
	1750	240	—	—	—
	1730	240	—	—	—
0.007	1797	—	80	80	80
	1772	—	80	80	80
	1750	—	80	80	80
	1730	—	80	80	80
0.014	1797	—	80	80	80
	1772	—	80	80	80
	1750	—	80	80	80
	1730	—	80	80	80
0.021	1797	—	80	80	80
	1772	—	80	80	80
	1750	—	80	80	80
	1730	—	80	80	80
Total	—	960	960	960	960

data is x and the corresponding label is y , the probability that the sample is judged to be a certain category j is $p(y = j | x)$. Therefore, the output for a class K classifier will be a K -dimensional vector (the sum of the elements of the vector is 1), shown as follows:

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \dots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \dots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}, \quad (14)$$

where $\theta_1, \theta_2, \dots, \theta_k \in R^{n+1}$ are the model parameters, and $1/\sum_{j=1}^k e^{\theta_j^T x^{(i)}}$ is a normalization function that normalizes the probability distribution to make the sum of all probabilities be one.

This structure ensures that the network can learn as many features as possible and prevents overfitting.

5.4. Experimental Result Analysis. The training and test samples' time-frequency maps are inputted into the constructed CNN. For the training and testing processes, the error iteration processes of the network are shown in Figure 14. The error value of the NN remains stable at a small value after iteration, indicating that the network has finished training and converged.

When the program is completed, the final classification results of the four data sets listed above can be obtained, as shown in Table 3. The average accuracy of this experiment reaches 95.83%, showing that the fault diagnosis method based on time-frequency maps and CNN can identify and classify bearing faults and maintain a high accuracy rate. The feature that the CNN can independently extract the features of the time-frequency map, avoids artificially calculating the feature values, and immensely reduces the recognition error.

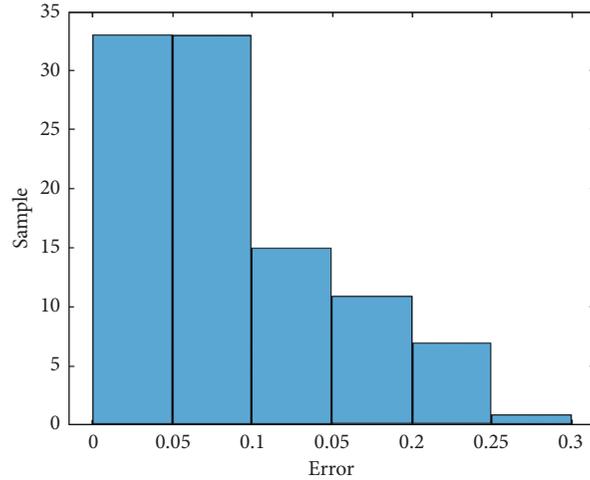


FIGURE 10: Test error distribution.

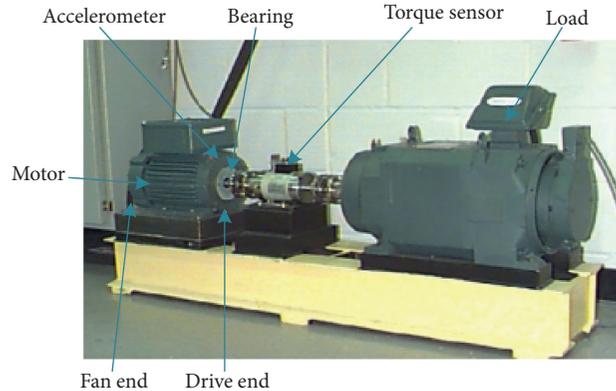


FIGURE 11: Schematic of bearing test device.

TABLE 2: Different fault samples and label configuration.

Classification	Training sample set	Testing sample set	Data set label
Normal signal	768	192	1
Inner ring failure	768	192	2
Outer ring failure	768	192	3
Rolling element failure	768	192	4
Total	3072	768	—

For different fault diameters but the same type of faults, they can still be classified into the same category although differences are found, reflecting the generalization ability of the proposed method. For different types of faults that have the same point in the feature maps, they can also be correctly distinguished, reflecting the powerful fault recognition ability of the proposed method.

The proposed method is compared with the fault diagnosis effects of several other methods, including traditional BPNN and BPNN optimized by PSO and GA, to prove its effectiveness. BPNN is a multilayer feedforward neural network trained according to error backpropagation algorithm. PSO can use the shared information of individuals to

make the group movement produce an evolution process from disorder to order in the problem solving space, thereby obtaining the optimal solution. GA is a calculation model that can search for the optimal solution by simulating natural selection in the biological evolution and biological evolution in the genetic mechanism [34]. BPNN, PSO-BP, and GA-BP all have excellent performance in fault classification. To avoid the accidental results of the experiment, multiple experiments are conducted on them separately to increase the reliability of the results.

The analysis results of the several methods to diagnose open-source rolling bearing fault data are shown in Table 4 and Figure 15. Compared with the traditional BPNN, the

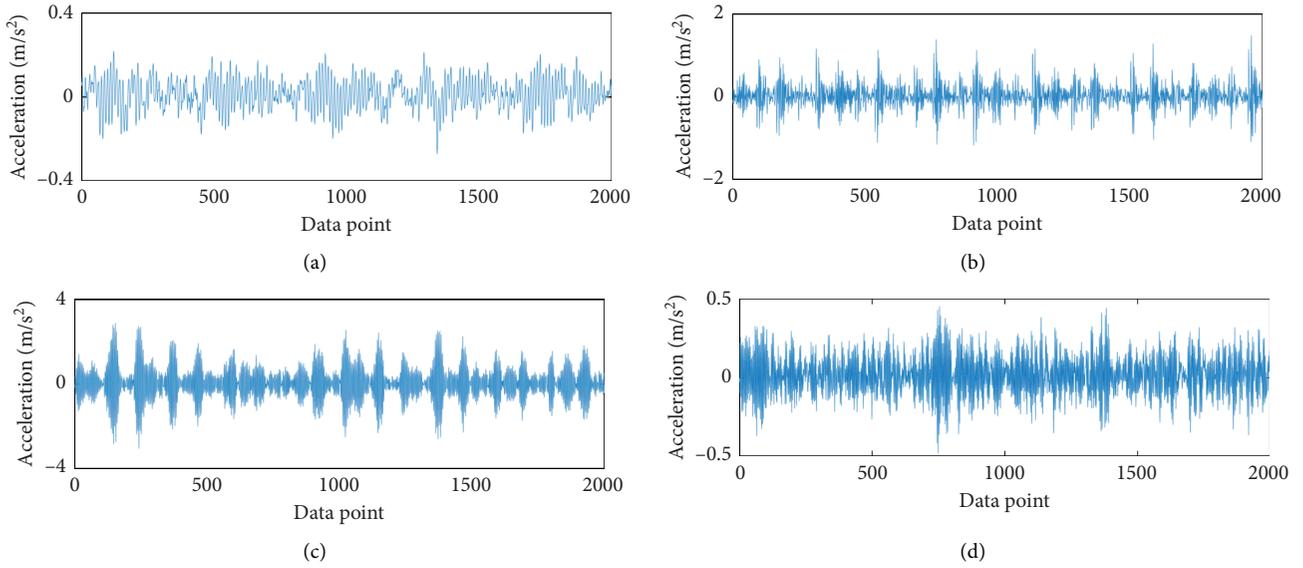


FIGURE 12: Time domain waveforms of bearing vibration signals in different states: (a) normal signal, (b) inner ring failure, (c) outer ring failure, and (d) rolling element failure.

TABLE 3: Experimental results.

Label	Normal signal	Inner ring failure	Outer ring failure	Rolling element failure	Total	Accuracy (%)	Average accuracy (%)
1	935	4	11	10	960	97.39	
2	3	908	32	17	960	94.58	95.83
3	7	16	932	5	960	97.08	
4	6	30	19	905	960	94.27	

TABLE 4: Diagnostic accuracy of multiple tests.

Method	Accuracy in different testing times (%)					Average accuracy (%)
	1	2	3	4	5	
BPNN	83.32	80.58	82.62	83.30	83.25	82.61
PSO-BP	83.35	85.67	87.32	86.54	86.70	85.92
GA-BP	86.53	88.24	87.34	87.50	85.42	87.01
STFT-CNN	95.83	97.30	94.78	97.65	96.25	96.36

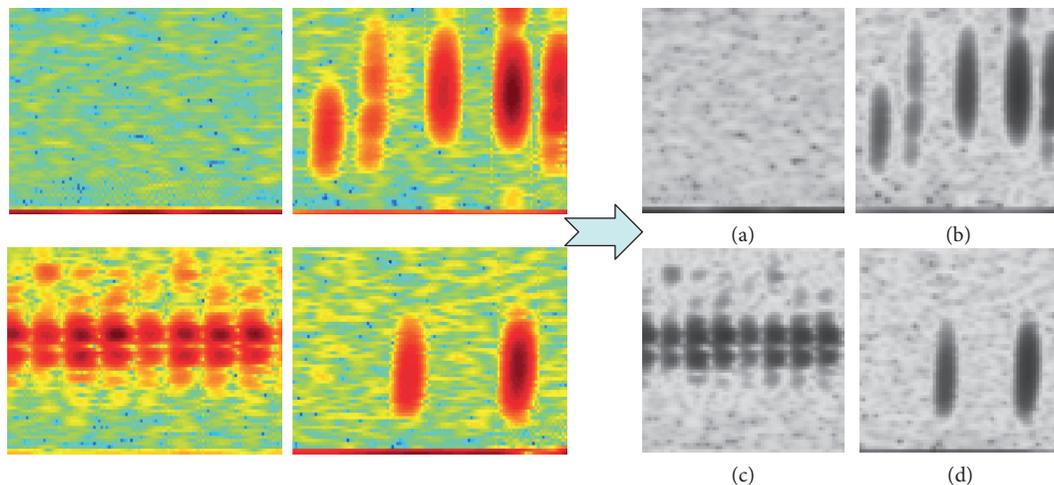


FIGURE 13: Time-frequency maps at different states: (a) normal signal, (b) inner ring failure, (c) outer ring failure, and (d) rolling element failure.

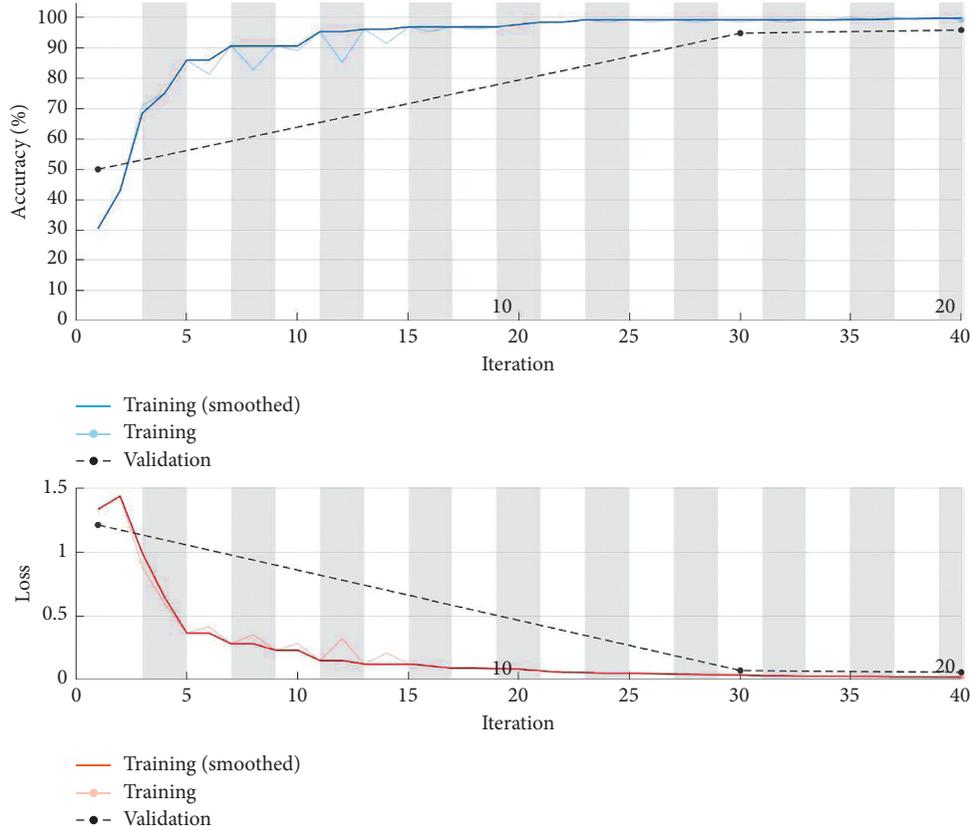


FIGURE 14: Training and testing processes.

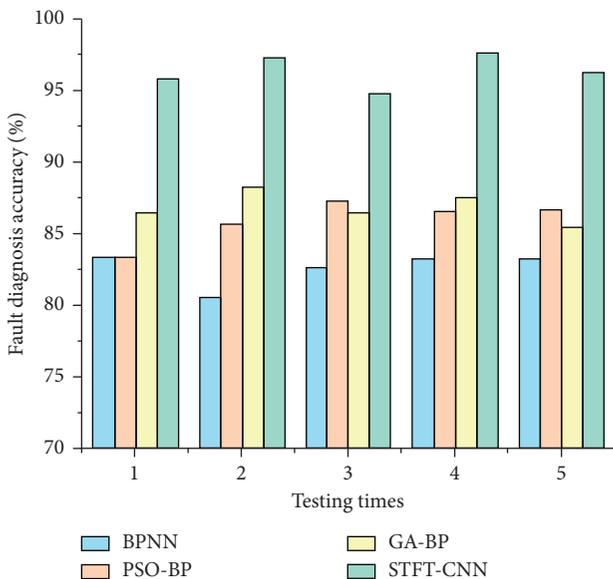


FIGURE 15: Comparison graph of the accuracy of multiple fault diagnosis tests.

BPNN optimized by PSO and GA immensely improves the diagnostic accuracy. However, STFT and CNN have the highest accuracy for fault diagnosis. This finding is mainly because the addition of algorithms optimizes the input eigenvalues and immensely improves the convergence ability

of the BPNN. However, the input values for such methods need to be calculated manually, and the calculation efficiency is low. The feature value calculation will take correspondingly considerable time when using a large sample. The STFT combined with CNN’s fault diagnosis method can directly reconstruct the 2D picture of the original vibration signal without manual calculation and has high diagnostic efficiency.

6. Conclusions

In this study, a bearing remaining useful life prediction and fault diagnosis method based on STFT and CNN was proposed. The bearing life prediction research was conducted using the rolling bearing life cycle datasets from the IMS. Then, the open source bearing data from CWRU were used for fault diagnosis, and its experimental results were compared with several fault diagnosis methods based on traditional BPNN. The following conclusions are obtained:

- (1) The rolling bearing remaining useful life prediction method based on STFT and CNN can effectively predict the life of the bearing. It can estimate the life of the bearing at any time in the entire cycle from normal operation to failure. Its accuracy can reach approximately 99.45%.
- (2) The rolling bearing fault diagnosis method based on STFT and CNN can better identify the different

failure modes, and its diagnosis accuracy is significantly higher than other methods based on traditional NN.

- (3) The proposed method does not need to manually calculate the eigenvalues, thereby effectively improving the prediction and diagnosis efficiency. Improving the structural and training parameters of the CNN can effectively improve the accuracy of life prediction and the correct recognition rate and stability of faults.

For future research, experimental verification will be carried out on rolling bearing data other than the database. In addition, improved CNN structure and more other techniques for improving the prediction and diagnosis accuracy will be investigated.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Agricultural Science and Technology Independent Innovation Fund of Jiangsu Province (CX(19)3081), the Fundamental Research Funds for the Central Universities (KYGD202005), and the Key Research and Development Program of Jiangsu Province (BE2018127).

References

- [1] Y. Lei, F. Jia, X. Zhou, and J. Lin, "A deep learning-based method for machinery health monitoring with big data," *Journal of Mechanical Engineering*, vol. 51, no. 21, pp. 49–56, 2015.
- [2] X. Xu, L. Zhang, Y. Jiang, and N. Chen, "Active control on path following and lateral stability for truck-trailer combinations," *Arabian Journal for Science and Engineering*, vol. 44, no. 2, pp. 1365–1377, 2019.
- [3] W. H. Wei, Y. L. Li, T. M. Xue et al., "Tool wear during high-speed milling of wood-plastic composites," *BioResources*, vol. 14, no. 4, pp. 8678–8688, 2019.
- [4] M. H. Xiao, X. J. Shen, Y. Ma et al., "Prediction of surface roughness and optimization of cutting parameters of stainless steel turning based on RSM," *Mathematical Problems in Engineering*, vol. 2018, Article ID 9051084, 15 pages, 2018.
- [5] Y. J. Wang, S. D. Wang, S. Q. Tang, Q. Y. Wang, and V. I. Mikulovich, "Prediction method of remaining useful life of rolling bearings based on improved gforest," *Proceedings of the CSEE*, vol. 40, no. 15, pp. 5032–5043, 2020.
- [6] Y. Wei and S. Liu, "Numerical analysis of the dynamic behavior of a rotor-bearing-brush seal system with bristle interference," *Journal of Mechanical Science and Technology*, vol. 33, no. 8, pp. 3895–3903, 2019.
- [7] Z. J. He and Z. Y. Huang, *Selection of Mechanical Fault Diagnosis*, Xi'an Jiaotong University Press, Xian, China, 1991.
- [8] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7762–7773, 2015.
- [9] B. Wang, H. R. Li, and B. H. Xu, "Motor bearing forecast feature extracting and degradation status identification based on multi-scale morphological decomposition spectral entropy," *Journal of Vibration and Shock*, vol. 32, no. 22, pp. 124–128+139, 2013.
- [10] H. R. Li, H. Yu, Z. K. Tian, and B. C. Li, "Degradation trend prediction of rolling bearings based on two-element multi-scale entropy," *China Mechanical Engineering*, vol. 28, no. 20, pp. 2420–2433, 2017.
- [11] C. Q. Shen, Y. M. Qi, J. Wang, G. G. Cai, and Z. K. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [12] F. T. Wang, J. Sun, D. W. Yan, S. H. Zhang, L. M. Cui, and Y. Xu, "A feature extraction method for fault classification of rolling bearing based on PCA," *Journal of Physics: Conference Series*, vol. 628, 2015.
- [13] Y. B. Liu, B. He, F. Liu, S. L. Lu, Y. L. Zhao, and J. W. Zhao, "Remaining useful life prediction of rolling bearings using PSR, JADE, and extreme learning machine," *Mathematical Problems in Engineering*, vol. 2016, Article ID 8623530, 13 pages, 2016.
- [14] F. T. Wang, X. F. Liu, C. X. Liu, H. K. Li, and Q. K. Han, "Remaining useful life prediction method of rolling bearings based on Pchip-EEMD-GM(1, 1) model," *Shock and Vibration*, vol. 2018, p. 10, Article ID 3013684, 2018.
- [15] Z. LIU, Q. Li, X. Liu, and C. Mu, "A hybrid LSSVR/HMM-Based prognostic approach," *Sensors*, vol. 13, no. 5, pp. 5542–5560, 2013.
- [16] Z. Shen, X. F. Chen, Z. J. He, C. Sun, X. L. Zhang, and Z. W. Liu, "Remaining life predictions of rolling bearing based on relative features and multivariable support vector machine," *Journal of Mechanical Engineering*, vol. 49, no. 2, pp. 183–189, 2013.
- [17] F. F. Chen, Y. Yang, B. J. Chen, and C. P. Chen, "Degradation trend prediction of rolling bearings based on fuzzy information granulation and wavelet support vector machine," *China Mechanical Engineering*, vol. 27, no. 12, pp. 1655–1661, 2016.
- [18] R. Liu, B. Yang, X. Zhang, S. Wang, and X. Chen, "Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 75, pp. 345–370, 2016.
- [19] I. Attoui, N. Fergani, N. Boutasseta, B. Oudjani, and A. Deliou, "A new time-frequency method for identification and classification of ball bearing faults," *Journal of Sound and Vibration*, vol. 397, pp. 241–265, 2017.
- [20] 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.
- [21] D. C. Yao, J. W. Yang, X. Q. Cheng, and X. Wang, "Railway rolling bearing fault diagnosis based on multi-scale IMF permutation entropy and SA-SVM classifier," *Journal of Mechanical Engineering*, vol. 54, no. 9, pp. 168–176, 2018.
- [22] J. D. Zheng, T. Liu, R. Meng, and Q. Y. Liu, "Generalized composite multiscale permutation entropy and PCA based

- fault diagnosis of rolling bearings,” *Journal of Vibration and Shock*, vol. 37, no. 20, pp. 61–66, 2018.
- [23] X. L. Qi, X. D. Ye, J. L. Cai, J. D. Zheng, Z. W. Pan, and X. Q. Zhang, “Fault feature extraction method of rolling bearings based on VMD and manifold learning,” *Journal of Vibration and Shock*, vol. 37, no. 23, pp. 133–140, 2018.
- [24] D. L. Yang, J. J. Miao, F. Y. Zhang, J. Tao, G. B. Wang, and Y. P. Shen, “Bearing fault diagnosis using a support vector machine optimized by an improved ant lion optimizer,” *Shock and Vibration*, vol. 2019, Article ID 9303676, 20 pages, 2019.
- [25] X. Jiang, C. Shen, J. Shi, and Z. Zhu, “Initial center frequency-guided VMD for fault diagnosis of rotating machines,” *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.
- [26] X. Zhao, J. Wu, Y. Zhang, Y. Shi, and L. Wang, “Fault diagnosis of motor in frequency domain signal by stacked denoising auto-encoder,” *Computers, Materials & Continua*, vol. 57, no. 2, pp. 223–242, 2018.
- [27] X. Huang, R. X. Chen, and Y. Huang, “Convolution neural networks for mechanical equipment fault diagnosis: the application and challenge,” *Manufacturing Technology & Machine Tool*, no. 1, pp. 96–100, 2019.
- [28] J. D. Zhang, Y. S. Zou, J. L. Deng, and X. L. Zhang, “Bearing remaining life prediction based on full convolutional layer neural networks,” *China Mechanical Engineering*, vol. 30, no. 18, pp. 2231–2235, 2019.
- [29] H. Pei, C. H. Hu, X. S. Si, J. X. Zhang, Z. N. Pang, and P. Zhang, “Review of machine learning based remaining useful life prediction methods for equipment,” *Journal of Mechanical Engineering*, vol. 55, no. 8, pp. 1–13, 2019.
- [30] H. Qiu, J. Lee, J. Lin, and G. Yu, “Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics,” *Journal of Sound Vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.
- [31] H. Qiu, J. Lee, J. Lin, and G. Yu, “Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics,” *Journal of Sound and Vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.
- [32] K. Loparo, *Bearing Vibration Data Set*, Case Western Reserve University, Cleveland, Ohio, 2003.
- [33] H. B. Wang, Y. X. Chen, and Y. Q. Li, “Face recognition method based on principal component analysis and Softmax regression model,” *Journal of Hefei University of Technology*, vol. 38, no. 6, pp. 759–763, 2015.
- [34] X. Xu, Y. Jiang, and H. Pueh Lee, “Multi-objective optimal design of sandwich panels using a genetic algorithm,” *Engineering Optimization*, vol. 49, no. 10, pp. 1665–1684, 2017.