

Research Article

Compound Fault Diagnosis for Gearbox Based Using of Euclidean Matrix Sample Entropy and One-Dimensional Convolutional Neural Network

Decai Zhang , Xueping Ren , and Hanyue Zuo 

School of Mechanical Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China

Correspondence should be addressed to Decai Zhang; 2397595377@qq.com

Received 3 December 2020; Revised 18 March 2021; Accepted 22 March 2021; Published 12 April 2021

Academic Editor: Giuseppe Ruta

Copyright © 2021 Decai Zhang 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.

Vibration signals of gearbox under different loads are sensitive to the existence of the fault and composite fault vibration signals are complex. Traditional fault diagnosis methods mostly rely on signal processing methods. It is difficult for signal processing methods to separate effective information from those fault signals. Therefore, traditional fault diagnosis methods are difficult to accurately identify those faults. In this paper, a one-dimensional convolutional neural network (1-D CNN) intelligent diagnosis method with improved SoftMax function is proposed. Local mean decomposition (LMD) decomposes the signals into different physical fictions (PF). PFs are input into the matrix sample entropy based on Euclidean distance (MESE), and the PFs which best reflect fault characteristics are selected. Finally, the PFs by MESE are used to train the CNN to identify the faults of parallel-shaft gearbox. Experiment shows that MESE can quickly and accurately select the PFs with the most significant fault features. 1-D CNN can get nearly 100% recognition rate with less time and the CNN of SoftMax improved can effectively eliminate LMD endpoint effect. This method can successfully identify single faults, combination faults, and faults under different loads of the gearbox. Compared with other methods, this method has the characteristics of high efficiency, accuracy, and strong anti-interference. Therefore, it can effectively solve the problem of complex fault signal decomposition of gearbox and can diagnose the gearbox fault under different load operation. It has great significance for gearbox fault diagnosis in actual production.

1. Introduction

Gearbox, as one of the most important parts of mechanical equipment, has a wide range of applications and a high utilization rate. Therefore, it will bring unimaginable safety problems and high maintenance costs when the fault occurs. So, the fault diagnosis of gearbox gear is very important. However, the structure of the gearbox is complex. The gears should be able to bear a heavy load, under complex and changeable operating conditions. The large gears, small gears, and bearings in the gearbox are prone to failure parts.

Gearbox diagnosis can be made from many aspects, such as vibration signal, sound signal, current signal, and oil-based signal. These signals can reflect the status of the gearbox. The analysis of vibration signal is the most commonly used and best effect. The traditional fault diagnosis methods are the application of signal processing experience

and manual feature extraction [1, 2]. The signal analysis method and feature extraction method are the main factors affecting the accuracy of fault diagnosis. Time-domain, frequency-domain, time-frequency domain analysis, and wavelet basis function [3], etc. are the most commonly used feature analysis methods in fault diagnosis [4, 5]. Rui et al. [6] designed a dynamic model of gear pairs with variable meshing stiffness to study the fault vibration characteristics of spur gear pair with local spalling defect. This method has great limitations due to the fact that gear spalling is likely to cause other faults and the fault is obvious in actual production. Brethee et al. [7] exploited a comprehensive dynamic model to analyze the effect of surface wear on the dynamic response of gears. It is based on the amplitude of grid vibration and the side band components increase significantly with the increase of wear degree. Sanz et al. [8] proposed a multistage algorithm for gear dynamic state

monitoring. The meshing dynamics of gears is monitored by means of the gear state information. But gear state information is changeable and complex and difficult to determine.

Due to the complexity of noise and gearbox system, the vibration signal collected from the gearbox through the sensor is usually mixed with a large number of noise interference signals. If the fault is complex, the signal mixed with noise will become more complex. So, some recursive mode decompositions are very suitable for gearbox fault diagnosis [9, 10]. Wu et al. [11] proposed a fault diagnosis method of planetary gear box based on two-dimensional variational decomposition (2-D VMD) and full vector spectrum technology and verified its correctness and effectiveness. Li et al. [12] proposed a VMD method that is complementary to the DDTFA method. Comparing this method with the current VMD and the overall empirical mode decomposition, it can be seen that VMD-DDTFA is very effective in fault diagnosis of gear crack and broken tooth measurement under variable working conditions. Xiao et al. [13] proposed a gear fault diagnosis method based on kurtosis criterion, variant modal decomposition (VMD), and self-organizing map (SOM) neural network. This method uses the VMD algorithm to decompose the gear vibration signal, extracts the kurtosis value of the IMF to form a feature vector and input SOM for diagnosis, and obtains a more ideal effect. But the SOM network is an unsupervised topology neural network, the final output size is difficult to determine, so it is not universal; Most of these methods use vibration signals or other signals through traditional wavelet, EMD [14], VMD, and other decomposition methods to decompose them into appropriate signal components [15, 16]. The type of fault is obtained by the parameters of the signal components. These methods can detect faults well, but they rely heavily on mechanical expertise and are inefficient due to the manual operation of actual diagnosis. Therefore, the development of traditional fault diagnosis in the direction of artificial intelligence is an inevitable trend. Therefore, the development of traditional fault diagnosis towards artificial intelligence is an inevitable trend.

With the development of artificial intelligence, machine learning is increasingly used in the field of fault diagnosis. The CNN has great advantages in feature classification. CNN [17] is currently mostly used for two-dimensional image recognition. Chen et al. [18] used two stacked CNNs to build a novel deep image saliency computing framework. The proposed framework highlights the objects of interest from complex background while preserving details. Sergey et al. [19] developed a method based on a partially observed guided policy search method and CNN, which was used to learn policies that map raw image observations directly to torques on the robot's motors. CNN can be used to directly process large data or multidimensional data samples, which is beneficial for more detailed local feature extraction and retaining the relative relationship of multidimensional data, leading to the acquisition of improved identification results. Deng et al. [20] proposed an improved quantum-inspired differential evolution, which uses the MSIQDE with global

optimization ability to optimize the parameters of the DBN and construct an optimal DBN model. Experimental results show that MSIQDE-DBN has higher classification accuracy. This is an improvement and perfection of the deep belief network. It shows that deep neural network is an algorithm suitable for fault classification. Chen et al. [21] proposed a convolutional neural network (CNN) for gearbox fault identification and classification, which is suitable for fault diagnosis of industrial reciprocating machinery. It also shows good performance in gearbox fault diagnosis, but the selection of parameters and hyperparameters is very difficult. John et al. [22] proposed a deep convolutional neural network (DCNN) based on hierarchical correlation propagation (LRP) method for gearbox fault diagnosis. It converts the vibration signals of time series data into time spectrum images by wavelet transform and then classifies them by DCNN. Yao et al. [23] input the time-domain and frequency-domain signals as the original signals to the end-to-end convolutional neural network (CNN) for gear fault diagnosis. This method is to analyze the sound signal and to identify the gear pitting fault problem. Li et al. [24] proposed a method that combines convolutional neural network (CNN) and gated recurrent unit (GRU) network and achieves an accuracy of more than 98% with fewer training samples. Li et al. [25] developed an improved deep neural network based on a domain-adaptive diagnosis model and used the original vibration signal for transfer learning. It used particle swarm optimization algorithm and L2 regularization algorithm to optimize the improved deep neural network. However, this method has higher requirements for data collection and classification. Chen et al. [26] proposed a gearbox fault diagnosis method based on feature learning of one-dimensional residual convolutional automatic encoder. This unsupervised learning method applies 1-D convolution automatic encoder for feature extraction and deconvolution for filter signal reconstruction. It performs well in signal denoising and feature extraction. But many times of convolution and deconvolution make the learning efficiency of network lower. The deeper neural network also makes the parameters more complex and less adaptable.

Through the survey and analysis of these literatures, most intelligent diagnostics rely heavily on traditional signal processing methods [27, 28], which makes intelligent diagnostics still limited by traditional methods. Many methods based on deep learning use two-dimensional CNN to analyze the image of the vibration signal. This makes the operating time and storage cost increase with the increase of data. As the neural network becomes deeper, multiple feature extraction and optimization such as feature image and convolution pooling will inevitably affect the accuracy of fault diagnosis, and the influence of irrelevant signals such as noise on the accuracy of diagnosis will also gradually expand.

Therefore, in order to solve these problems, local mean decomposition (LMD) and one-dimensional CNN (1-D CNN) are combined [29]. LMD has a better adaptive system and lower error rate compared with EMD, and the disadvantage of decomposition of finite vibration signal, endpoint effect, is greatly reduced. Compared with 2-D CNN, 1-D CNN directly extracts the vibration data, which reduces the

time and improves the accuracy. Compared with the traditional method, it is no longer limited by fault types and complexity, and different faults can be obtained by training different fault signals. The method of calculating the distance between the features of SE, which is more suitable for the correlation analysis of PF, is improved. After the analysis, the selected components will be more suitable for CNN training. The mode of data input is changed from 1-D to 2-D matrix, which is to improve the extraction and calculation speed of SE. The SoftMax activation function of CNN was improved and it will reduce the impact of LMD endpoint effect on CNN. LMD is used for adaptive local decomposition of fault signals. The CNN feature normalization ability and feature extraction ability enable this method identify single fault, faults under different loads, and composite faults. This method gets higher accuracy. The recognition accuracy of this method is nearly 98.8%.

The research in this paper is organized as follows: Section 2 introduces the basic principle of LMD and the improvement method of endpoint effect. Section 3 introduces the improved distance algorithm of SE and the method of fast solved by matrix. In Section 4, the principles of 1-D CNN and improved SoftMax activation function are introduced. In Section 5, experiments and results of gearbox fault diagnosis are provided to verify the effectiveness of the method, and the results are compared with other methods. Finally, the conclusion and some works in the future are given in Section 6.

2. LMD Principle Analysis and Endpoint Effect Compensation Method

LMD is to separate pure frequency modulation signals and envelope signals from the original signal. The PF component is the product of the frequency modulation signal and the envelope signal. The signal is repeatedly separated until the isolated PF component becomes monotone or reaches a given threshold. Each PF signal has its own unique physical meaning [30]. The specific decomposition process of any one-dimensional signal $x(t)$ is as follows:

Find all the local extreme points of $x(t)$ and call them a_i . l_i is the average value of a_i and a_{i+1} , which are every two adjacent extreme points. The algorithm is shown in the following equation:

$$l_i = \frac{a_i + a_{i+1}}{2}. \quad (1)$$

Connect all l_i and l_{i+1} with a straight line and then apply the lines smoothly with the moving average method. The local mean function is called $l_{11}(t)$.

Then, use the local extreme point a_i to get the envelope estimate b_i . Connect all b_i and b_{i+1} . The envelope estimation function, $b_{11}(t)$, is obtained by smoothing with the moving average method.

$$b_i = \frac{|a_i - a_{i+1}|}{2}. \quad (2)$$

The LMD decomposition process is a process of continuously separating high-frequency signals, but the actual signal length is limited [31]. This makes it necessary for LMD to infer the value near the endpoint when processing the final signal, which makes the local mean function and the estimation function generate severe errors.

A section of the original signal and the curve of the local mean function are compared to analyze the influence of the endpoint effect and its attenuation method. Figure 1 shows the envelope and local mean function of two actual vibration signals. The way to find the envelope of LMD is to find the maximum and minimum points of the signal, and the signal outside the endpoint is unknown, so this little piece of the envelope is going to be unknown. If it is an excellent situation, the endpoint value is exactly the extreme point, the local mean function and envelope function obtained will be accurate, and the endpoint effect is almost eliminated. The cosine function is used to verify the validity of the method. Figure 2 shows the component signals after the decomposition of two vibration signals.

Figure 3 shows the first PF component and the original vibration signal. After analyzing the results of the decomposition, it can be seen that the LMD component of the vibration signal whose end point passes through the origin is almost unaffected by the end point effect and can coincide with the original signal, while the other vibration signal has a large difference value after decomposition. The vibration amplitudes between the PF component and the residual component vary by 4 orders of magnitude, so the residual component can be ignored. It is proved that the end point of the original signal is the extreme point. The end effect can be mostly eliminated. So, reducing or moving the original signal to a qualified endpoint is a kind of effective means. The reduced dimension time is defined as t_0 and the original vibration signal function is $x(t)$, and substitute t_0 from the LMD decomposition formula. Since t_0 is very small, it hardly affects the values of b and l . It can be assumed that b and l do not change as a result of the signal moving. Separating the local mean function $l_{11}(t)$ from the shifted delta function $x(t - t_0)$, $m_{11}^0(t)$ which has been gained can be expressed as

$$m_{11}^0(t) = x(t - t_0) - l_{11}(t). \quad (3)$$

The vibration signal is similar to the local periodic function. According to the basic principle of periodic function, when the function moves, the conclusion is

$$x(t - t_0) = x(t) - x(t_0) + \tau, \quad (4)$$

where τ is the deviation of the function from the coordinate axis, which is a fixed value of a function. $m_{11}^0(t)$ divided by the envelope estimation function $b_{11}(t)$ demodulates to get $s_{11}^0(t)$:

$$s_{11}^0(t) = \frac{m_{11}^0(t)}{b_{11}(t)}. \quad (5)$$

Continue to iterate according to the LMD rule to the end. The process is as follows:

$$\left. \begin{aligned}
 m_{11}^0(t) &= x(t-t_0) - l_{11}(t) \\
 &= x(t) - l_{11}(t) - x(t_0) + \tau \\
 &= m_{11}(t) - x(t_0) + \tau, \\
 m_{12}^0(t) &= s_{11}^0(t) - l_{12}(t) \\
 &= m_{12}(t) - \frac{x(t_0) + \tau}{b_{11}(t)}, \\
 m_{13}^0(t) &= s_{12}(t) - \frac{x(t_0) + \tau}{b_{11}(t)b_{12}(t)} - l_{13}(t) \\
 &= m_{13}(t) - \frac{x(t_0) + \tau}{b_{11}(t)b_{12}(t)}, \\
 &\dots \\
 m_{1n}^0(t) &= m_{1n}(t) - \frac{x(t_0) + \tau}{\prod_{q=1}^{n-1} b_{1q}(t)},
 \end{aligned} \right\} \tag{6}$$

$$\left. \begin{aligned}
 s_{11}^0(t) &= \frac{m_{11}^0(t)}{b_{11}(t)} \\
 &= s_{11}(t) - \frac{x(t_0) + \tau}{b_{11}(t)}, \\
 s_{12}^0(t) &= \frac{m_{12}^0(t)}{b_{12}(t)} \\
 &= s_{12}(t) - \frac{x(t_0) + \tau}{b_{11}(t)b_{12}(t)}, \\
 s_{13}^0(t) &= \frac{m_{13}^0(t)}{b_{13}(t)} \\
 &= s_{13}(t) - \frac{x(t_0) + \tau}{b_{11}(t)b_{12}(t)b_{13}(t)}, \\
 &\dots \\
 s_{1n}^0(t) &= s_{1n}(t) - \frac{x(t_0) + \tau}{\prod_{q=1}^n b_{1q}(t)},
 \end{aligned} \right\}$$

The termination condition for the iteration is expressed as

$$\lim_{n \rightarrow \infty} b_{1n}(t) = 1, \quad (7)$$

$$\prod_{q=1}^n b_{1q}(t) = b_1(t).$$

Multiply $b_1(t)$ and the frequency modulation signal $s_{1n}^0(t)$ to get the first PF component which is reduced to the end effect:

$$PF_1^0(t) = b_1(t) \left[s_{1n}(t) - \frac{x(t_0) + \tau}{b_1(t)} \right] = PF_1(t) - x(t_0) + \tau. \quad (8)$$

Therefore,

$$c_1^0(t) = x(t) - PF_1(t) + x(t_0) + \tau = c_1(t) + x(t_0) + \tau. \quad (9)$$

Repeat the above algorithm:

$$PF_2^0(t) = b_2(t) \times \left[s_{2n}(t) - \frac{x(t_0) + \tau}{b_2(t)} \right]$$

$$= PF_2(t) - x(t_0) + \tau,$$

$$PF_3^0(t) = b_3(t) \times \left[s_{3n}(t) - \frac{x(t_0) + \tau}{b_3(t)} \right]$$

$$= PF_3(t) - x(t_0) + \tau \frac{PF_3^0(t)}{x(t_0)} = \frac{PF_3(t)}{x(t_0)} + \frac{\tau}{x(t_0)} - 1, \quad (10)$$

where $x(t_0)$ and $PF(t)$ are very different in order of magnitude, and $x(t_0)$ can be processed by taking an extreme value, and $(\tau/x(t_0))$ is the ratio of the vibration offset to the endpoint offset. The endpoint effect of LMD can be reduced by a small shift value.

3. Matrix Sample Entropy Based on Euclidean Distance

SE measures the complexity of the sequence by measuring the probability of generating a new pattern in the signal. The greater the probability of the new pattern, the greater the complexity of the sequence will be. The PF decomposed by LMD is actually a high-frequency signal with physical significance. Since the sample entropy has higher self-adaptability and consistency, sample entropy is easier to make sure whether the PF component's regularity is suitable for CNN [32]. The basic principle and improvement of sample entropy are as follows:

There is a time series, $\{x(n)\} = x(1), x(2), \dots, x(N)$, which is composed by N data. A set of vector sequences of dimension m , $X_m(1), X_m(2), \dots, X_m(N-m+1)$,

composed by sequence number, $X_m(i)$, in this sequence is $\{x(i), x(i+1), \dots, x(i+m-1)\}$. These vectors represent m consecutive x values starting from the i -th point. The PF components of length m are arranged and calculated according to the following matrix. Each X is a vector sequence of PF components:

$$X = \begin{bmatrix} x(1) & \cdots & x(m+1) \\ \vdots & \ddots & \vdots \\ x(N-m) & \cdots & x(N) \end{bmatrix}, \quad (11)$$

where $d[X_m(i), X_m(j)]$ is defined as the distance between the vector $X_m(i)$ and $X_m(j)$ is the absolute value of the maximum difference in the corresponding elements. The feature space x is an n -dimensional vector space R^n . $x_i, x_j \in \chi$, $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})^T$, $x_j = (x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(n)})^T$, and the distance of x_i and x_j is defined as

$$L_p(x_i, x_j) = \left(\sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}|^p \right)^{(1/p)}. \quad (12)$$

The Euclidean distance when p is equal to 2:

$$L_2(x_i, x_j) = \left(\sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}|^2 \right)^{(1/2)}. \quad (13)$$

If p is equal to infinity, each distance is the maximum, which is the original distance of SE:

$$L_\infty(x_i, x_j) = \max |x_i^{(l)} - x_j^{(l)}|. \quad (14)$$

Because the fault data are all displacement, velocity, and acceleration signals in the same coordinates, this method converts the one-dimensional distance characteristic into a high-dimensional distance. The algorithm makes the features of various directions equally important and reduces the difference value of each feature point of the same component. The method can normalize the feature of each direction and better identify the PF component which can best represent the gearbox fault:

$$d[X_m(i), X_m(j)] = \max \sqrt{|x(i+m)|^2 - |x(j+m)|^2}. \quad (15)$$

Calculate the matrix distance of x and x_i for each i :

$$XX_i = \begin{bmatrix} |x(i+1) - x(i)| & \cdots & |x(i+1+m) - x(i+m)| \\ \vdots & \ddots & \vdots \\ |x(N-m) - x(i)| & \cdots & |x(N) - x(i+m)| \end{bmatrix}. \quad (16)$$

The number of j ($1 \leq j \leq N-m$, $j \neq i$) is the $X_m(i)$ whose distance between $X_m(i)$ and $X_m(j)$ should be less than or equal to r . The number is called B_i . If $1 \leq i \leq N-m$ is right, the definition of $B_i^m(r)$ is expressed as

$$B_i^m(r) = \frac{1}{N-m-1} B_i. \quad (17)$$

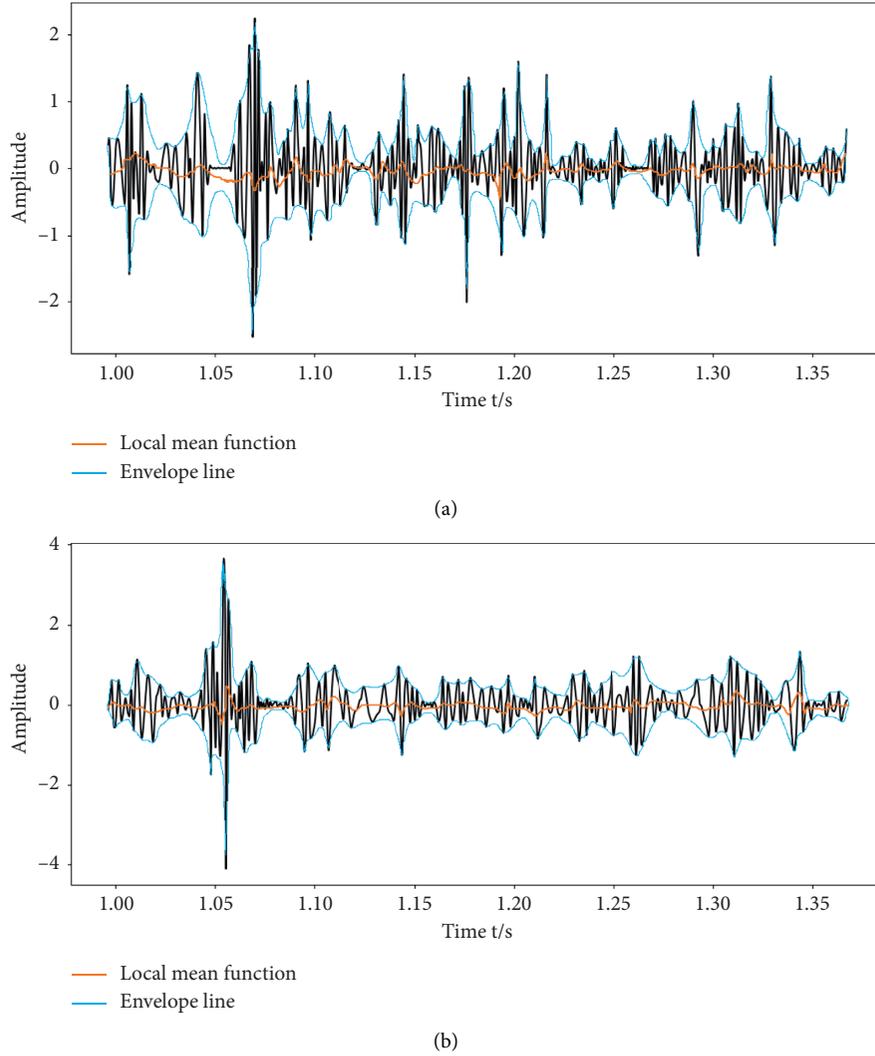


FIGURE 1: Vibratory signals with two different endpoints.

Increase the dimension of the matrix from m to $m + 1$ and then calculate the value of the matrix which is less than the threshold r . It is the distance, which is between $X_{m+1}(i)$ and $X_{m+1}(j)$ is less than or equal to r .

$$B_i^m = \text{Num} \left[\max_{1 \leq j \leq m} (XX_i(p, j)) < r \right], \quad (18)$$

$$B_{i+1}^m = \text{Num} \left[\max_{1 \leq j \leq m+1} (XX_i(p, j)) < r \right].$$

The definition of $B^{(m)}(r)$ is expressed as

$$B^{(m)}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r). \quad (19)$$

When $1 \leq p \leq N - m - i$, it becomes

$$B^m = \sum_{i=1}^{N-m} B_i^m, \quad (20)$$

$$B^{m+1} = \sum_{i=1}^{N-m} B_i^{m+1}.$$

where $B^{(m)}(r)$ is the probability of two sequences matching m points under similar tolerance r , and $B^{m+1}(r)$ is the probability that two sequences match $m + 1$. So, the sample entropy is defined as the following formula:

$$\text{SampEn}(m, r) = \lim_{s \rightarrow \infty} \left\{ -\ln \left[\frac{B^m(r)}{B^{m+1}(r)} \right] \right\}. \quad (21)$$

The value of sample entropy is related to the value of m and r . For mechanical vibration signals, Euclidean distance

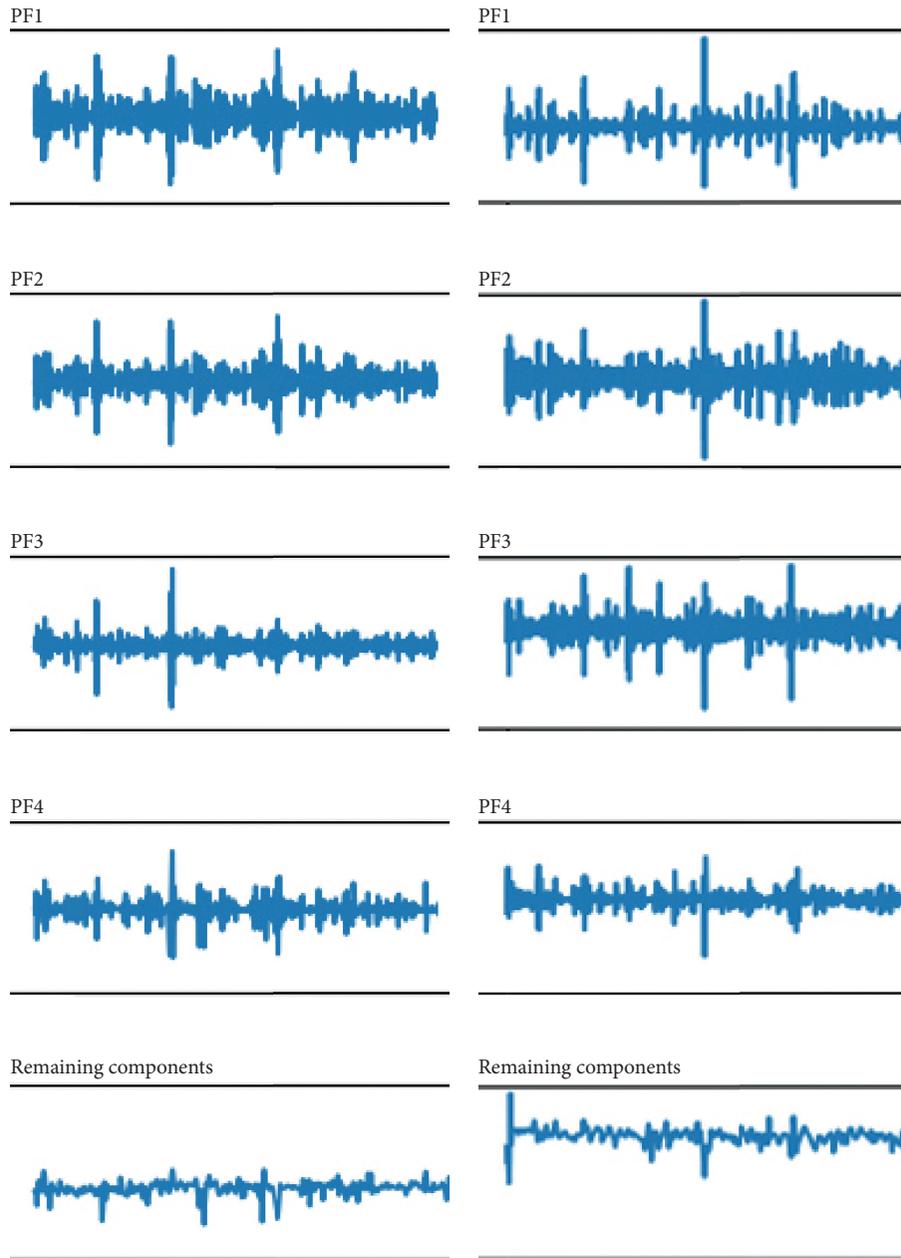


FIGURE 2: Decomposition components of two vibration signals.

greatly reduces the difference between different dimensions of the same fault, and the matrix algorithm has greatly improved the rapid entropy value of large quantities of vibration data.

4. 1-D CNN and the Improvement of SoftMax

4.1. The Principle of One-Dimensional CNN. Both 1-D CNN and 2-D CNN have the same characteristics and the same processing methods. Their key difference is the data and the sliding method of the filter [33]. The working mode of 1-D CNN is studied through 2-D CNN. The upper figure of Figure 4 shows the working mode of a convolutional layer of 1-D CNN. Each row represents one element, and the

convolution kernel convolves with the same width as the convolutional layer with a step length of 1, and the convolution is carried out 8 times. The below figure of Figure 4 shows the convolution method of a two-dimensional CNN. The convolution kernel is a 2×2 freely defined convolution kernel, and the convolution kernel is a three-layer convolution kernel with the characteristics of RGB, and the step size is 1. However, the convolution is carried out reciprocating convolution along a certain direction until the convolution kernel goes through the whole image.

In order to reduce the risk of overfitting, the convolutional layer usually has fewer parameters than the fully connected layer. Figure 5 is a 1-D CNN model training simulation graph. The original data are obtained by three

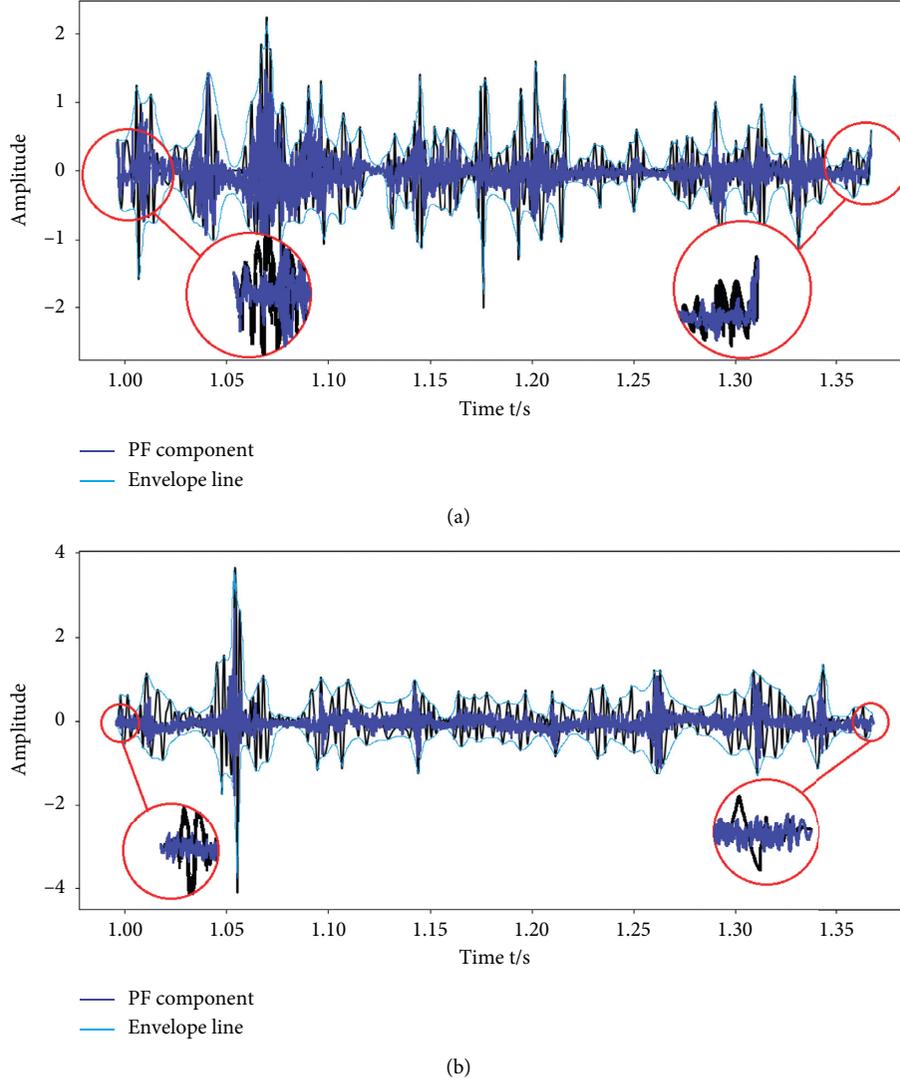


FIGURE 3: The first PF component and the original vibration signal.

convolution layers, three pooling layers, and three full connection layers to obtain specific eigenvalues.

4.2. Weight Compensation SoftMax Algorithm. If the original data are processed directly by vibration migration, it not only needs to process each group of data but also the offset point of each group of data cannot be determined. In order to make the method more general, SoftMax is improved. Adding an adaptive offset correction to the function theoretically cannot completely eliminate the impact of the end effect, but it will no longer affect the recognition of the neural network.

SoftMax is a numerical processing of output unit that deals with multiple classification problems. The basic definition of the SoftMax function is as follows:

$$S_i = \frac{e^{V_i}}{\sum_i^C e^{V_i}}. \quad (22)$$

Because of the existence of exponential operation, the value after operation is often overlapped, so the element is degraded and extracted by taking a value and V is the output value:

$$\alpha = \max(V),$$

$$S_i = \frac{e^{V_i - \alpha}}{\sum_i^C e^{V_i - \alpha}}. \quad (23)$$

For the linear classifier output, the input x is multiplied by the weight matrix, and the linear mixing function corresponding to the correct category is set as S_{y_i} which is output for corresponding SoftMax. So, to log on S_{y_i} operation does not affect the characteristics of the function. After a series of manipulations, the degradation results have become a simple SoftMax loss function. The function is expressed as

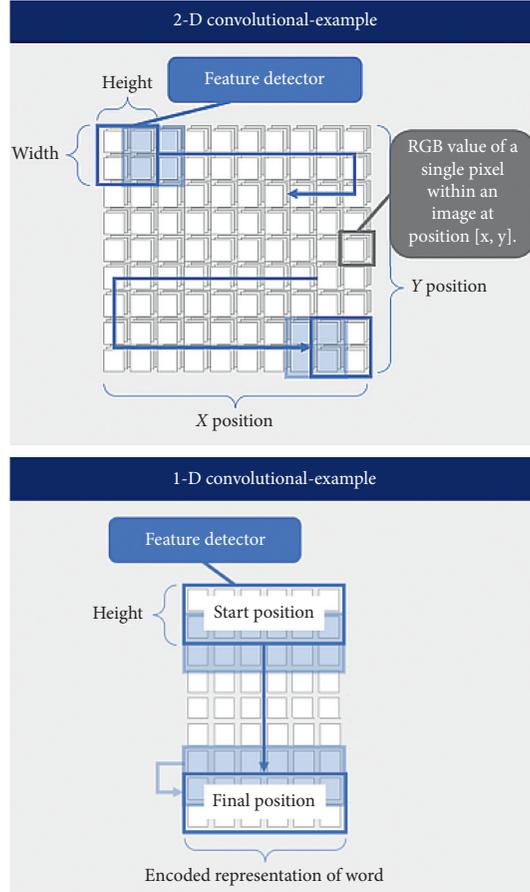


FIGURE 4: Working mode of 1-D CNN and 2-D CNN.

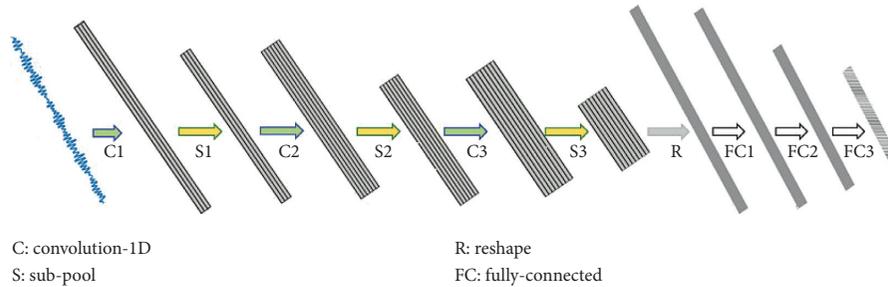


FIGURE 5: A 1-D CNN model for hybrid fault diagnosis.

$$L_i = -\log \frac{e^{S_{y_i}}}{\sum_{j=1}^C e^{s_j}} = -\left(S_{y_i} - \log \sum_{j=1}^C e^{s_j} \right) = -S_{y_i} + \log \sum_{j=1}^C e^{s_j}, \quad (24)$$

where S_{y_i} is usually the output of the fully connected layer in CNN. Put the weight of the LMD endpoint compensation, $(1/W_j^T)x_i + b_j$, into equation. In the formula, b_j is (τ/W_j^T) . So, the equation can be expressed as

$$L_i = -\log \frac{e^{(1/W_{y_i}^T)x_i + b_j}}{\sum_{j=1}^C e^{(1/W_{y_i}^T)x_i + b_j}} = -\log \left(\frac{e^{w_{y_i}x_i \tan(\theta_{y_i,i}) + b_{y_i}}}{\sum_j e^{w_jx_i \tan(\theta_{j,i}) + b_j}} \right). \quad (25)$$

In the iterative process, the centralization weight is $w_j = (1/\tau)$. It changes the distance calculation method by metric learning which converts $(1/W_{y_i}^T)x_i + b_j$ to $\tan(\theta_{y_i,i}) + b_{y_i}$. It is used to reduce class-inner-distance and

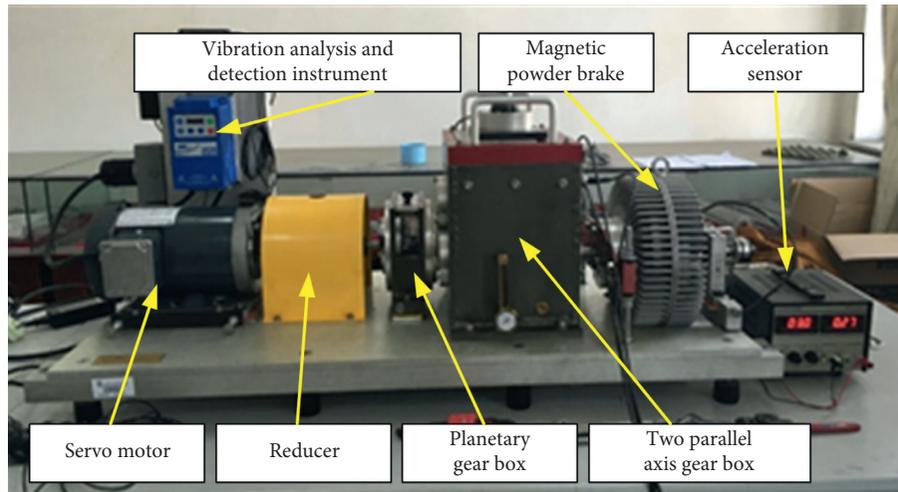


FIGURE 6: Fault diagnosis testbed of rotating machinery.

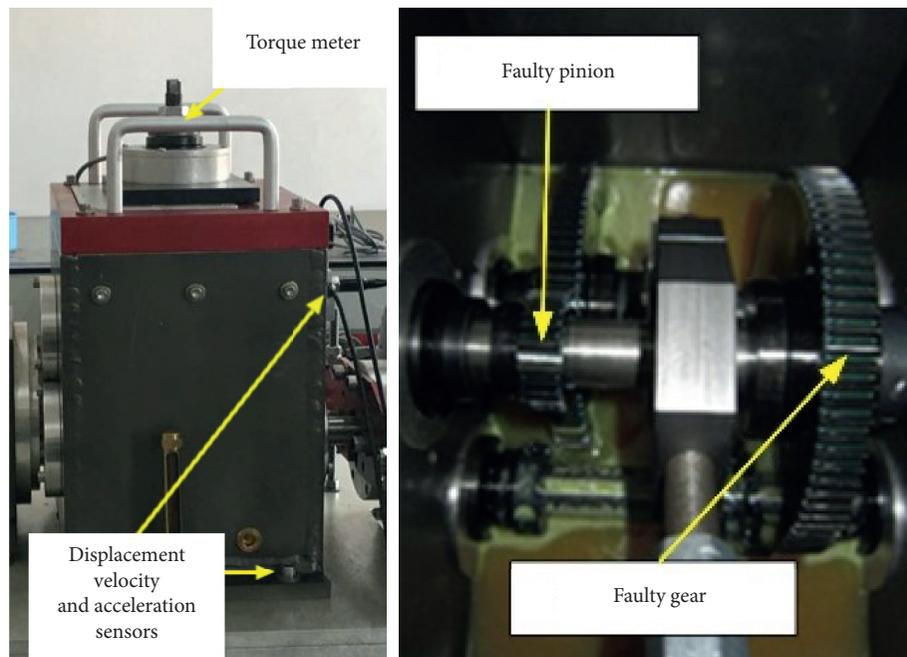


FIGURE 7: The position of sensors and internal structure of gearbox.

increase class-out-distance. It is achievable that the centering weight compensates for the end point offset in the distance. The range that can be preset is [1, 5] when a model network is trained. Debug the parameters until CNN is optimal.

5. Experiment and Verification

5.1. Data Collection. The experimental data collection is carried out using the experimental rig as shown in Figure 6. Figure 7 shows the sensor distribution and gearbox internal structure. The synthetic experimental system is composed of a first-stage planetary gearbox and a second-stage parallel-shaft spur gearbox. In order to accurately diagnose the gear

faults of parallel-shaft gearboxes, planetary gearboxes are equipped with intact gears without sampling. Experiment is performed with large gear and small gear in the parallel-shaft gearbox which is also an input shaft. There is servo motor, magnetic powder brake and controller, 4 magnetic accelerometers, and displacement sensors in the driving equipment. The sensors are placed in the axial and radial directions of the gear housing on the input and output sides to pick up the radial and axial vibration signals of the bearing housing. The large and small gears of the first input shaft are replaced. The large gear used in the experiment is S45 C gear with a modulus of 2 teeth and 75 teeth, and the pinion has a tooth number of 55, which is lubricated by immersion type.

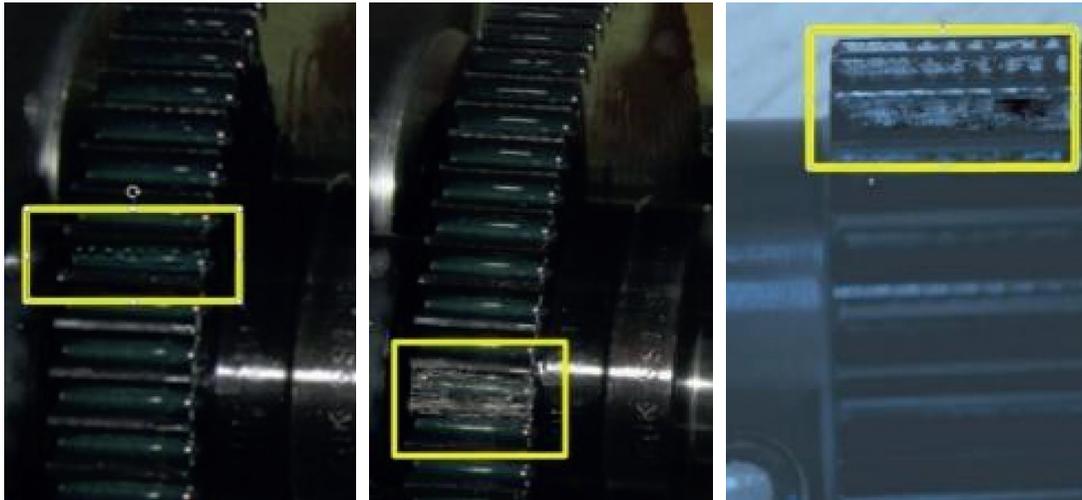


FIGURE 8: Fault pattern of the gears which are pitting, tooth fracture, and wear.

TABLE 1: Sampling situations.

Characteristic	Value
Sample frequency (Hz)	2000 * 2.56
Sample time (s)	10.4
Power (W)	750
Speed (rpm)	880, 1470
Load (A)	0, 0.05, 0.1, 0.2
Number of sensors	8
Sensibility (mV/g)	150

TABLE 2: The description of each fault condition.

No.	Description
1	Normal
2	Main gear teeth breakage about 50%
3	Main gear with pitting on tooth depth 0.05 mm, width 0.5 mm, and length 0.05 mm
4	Pinion with chafing on tooth about 50%
5	Main gear teeth breakage about 50% and pinion with chafing on tooth about 50%
6	Main gear with pitting on tooth depth 0.05 mm, width 0.5 mm, and large 0.05 mm and pinion with chafing on tooth about 50%

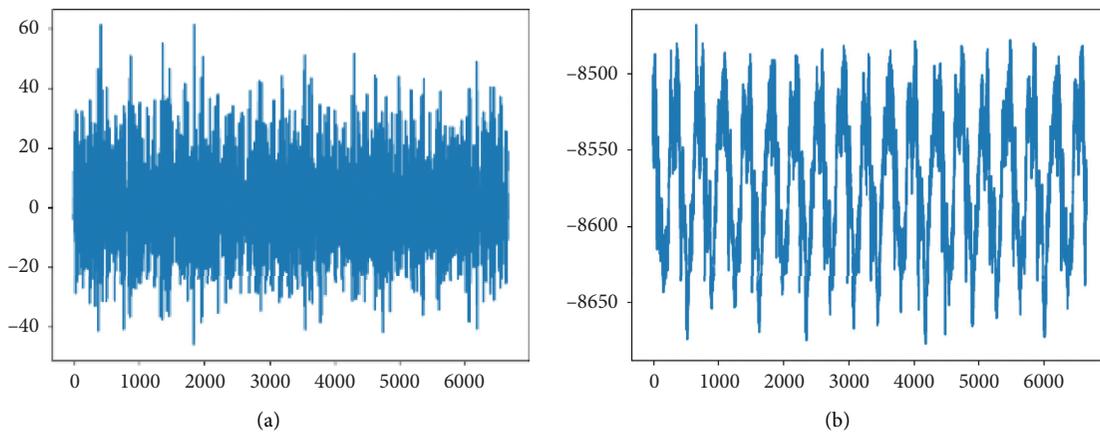


FIGURE 9: Continued.

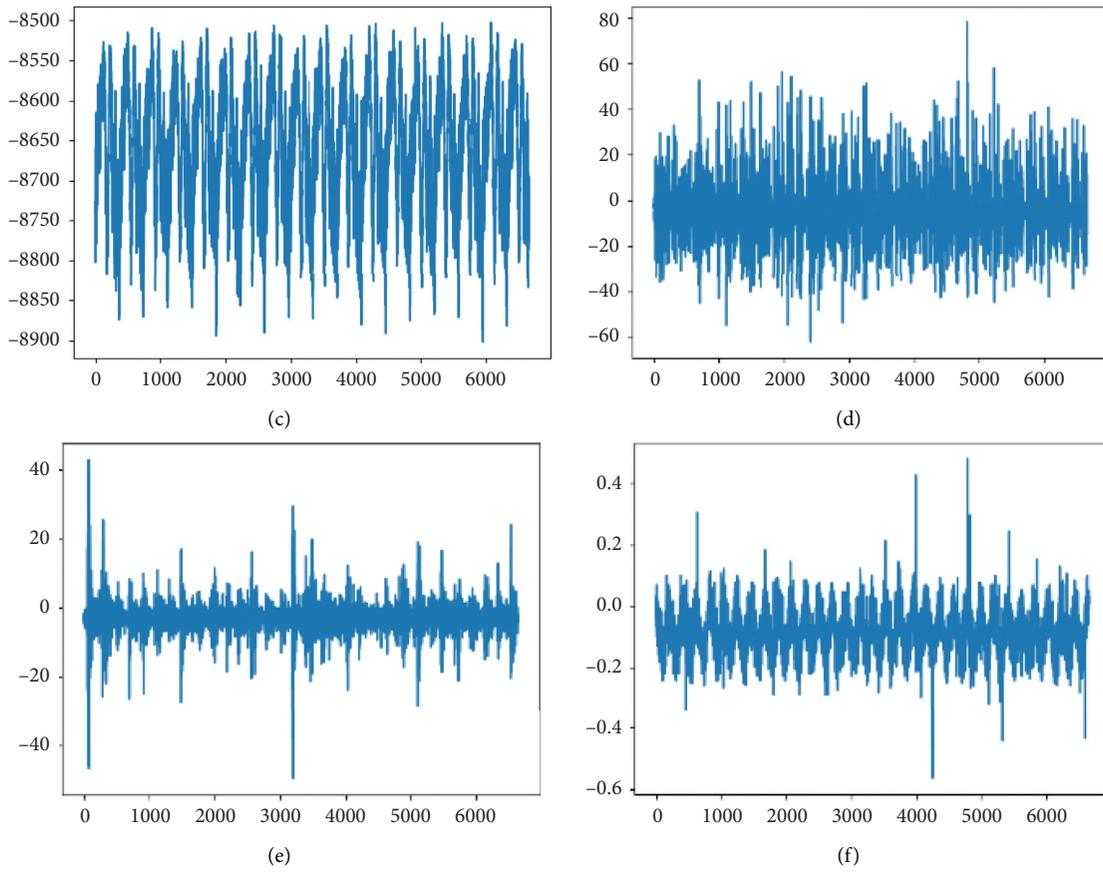


FIGURE 9: Time-domain waveforms of six different faults.

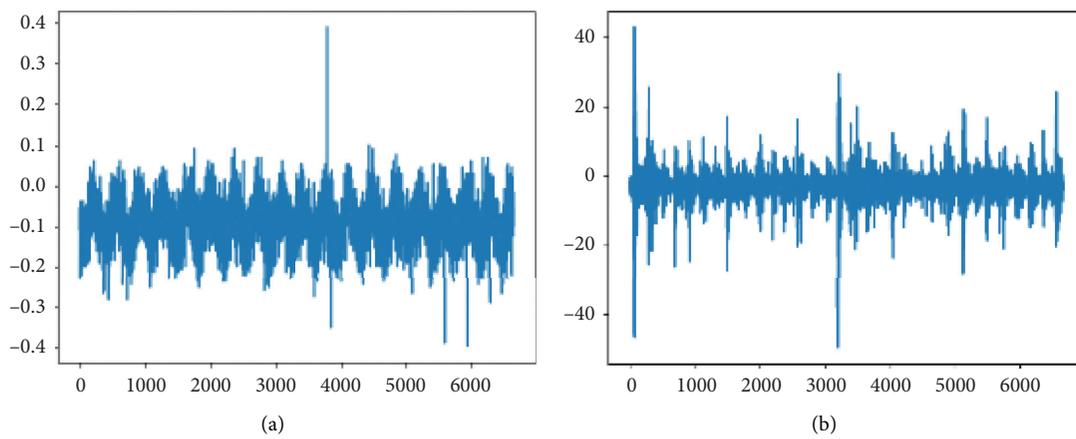


FIGURE 10: Continued.

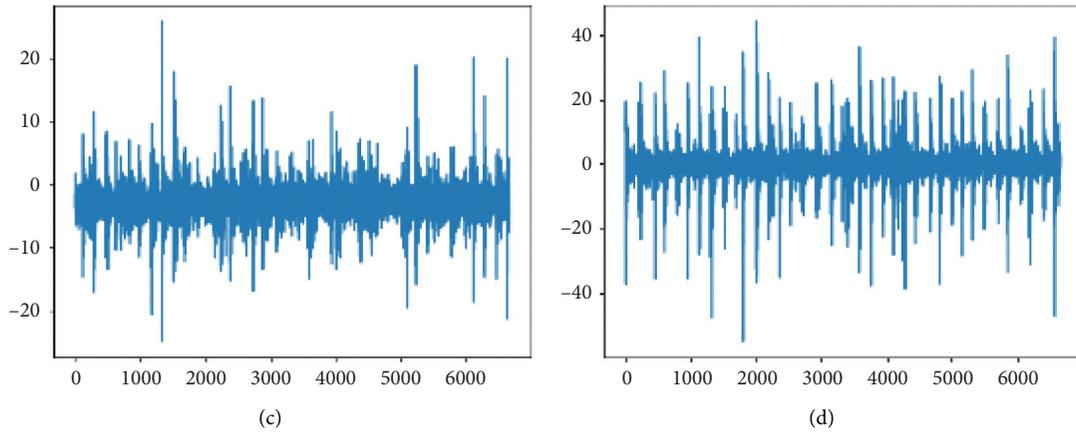


FIGURE 10: Time-domain waveforms of large gear pitting pinion wear.

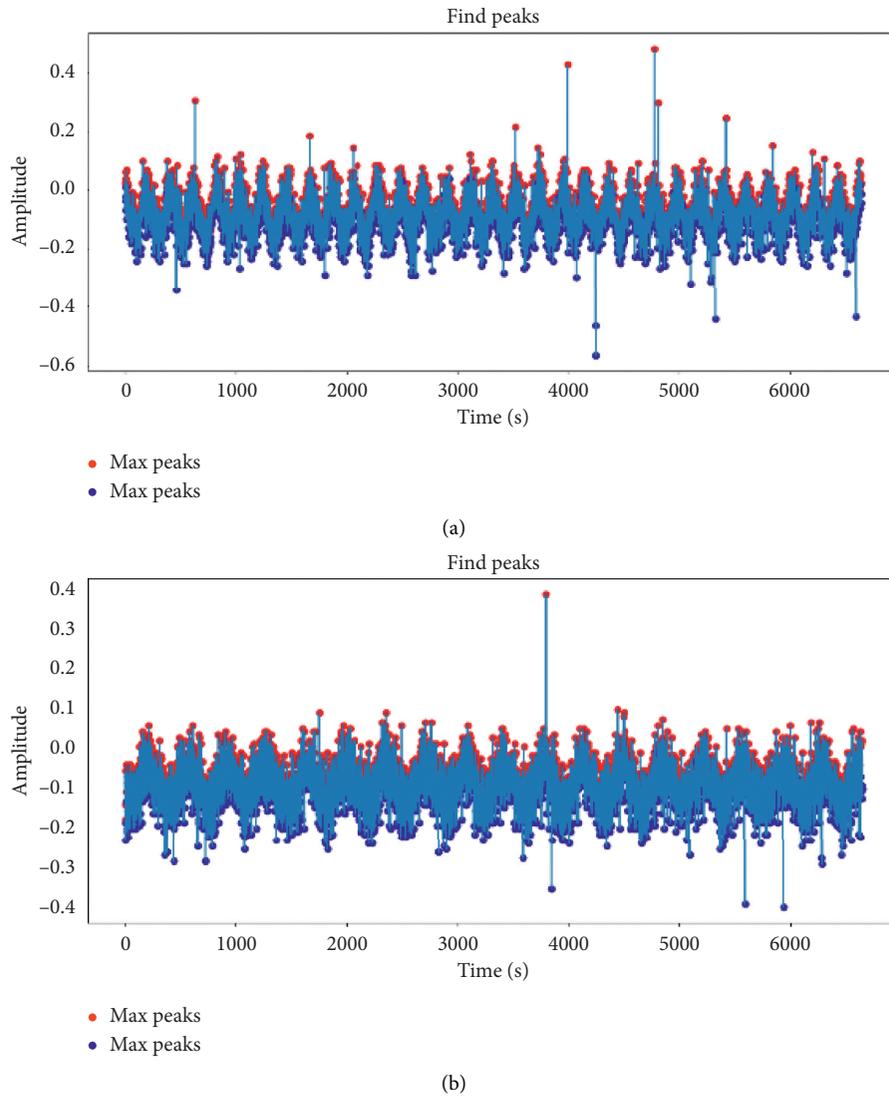


FIGURE 11: Vibration signal and envelope curve.

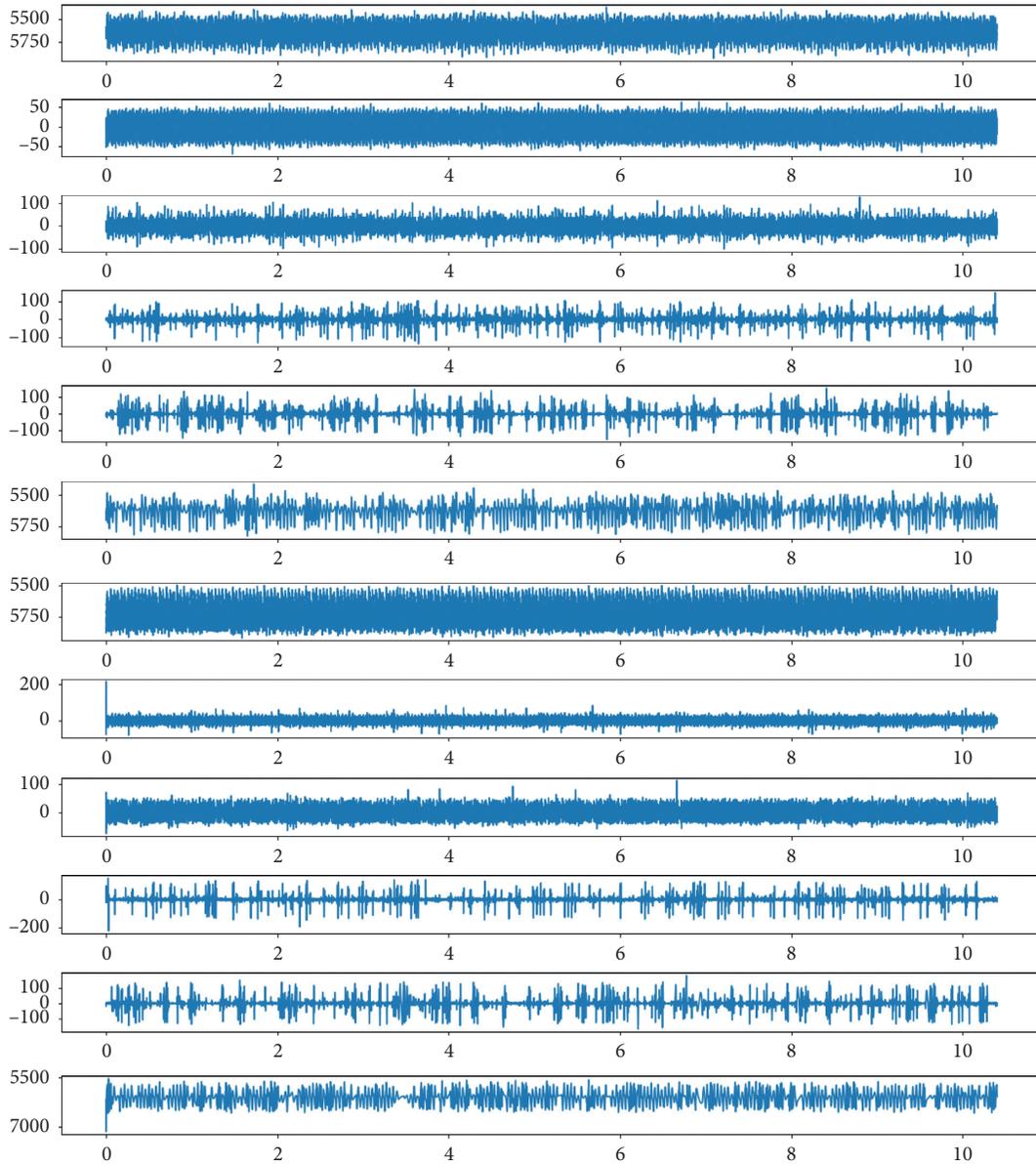


FIGURE 12: PF component and spectrum diagram of LMD. Two faults are large gear broken teeth and small gear wear and large gear pitting and small gear wear.

There are six different kinds of failure data, including gear broken teeth, gear pitting, pinion wear, gear broken teeth and pinion wear, big gear pitting and small gear wear, and normal conditions. The sampling frequency is 5120 Hz. Compound fault signals are collected by the system. Because the damage of one gear is likely to cause the damage of other parts, considering this situation, complex multigear faults are added into fault classification. Each dataset includes 6656 vibration signals collected by 9 sensors, 2 speeds, and 3 loads under each working condition, which is collected 8 times in each state interval. There are 2,160 sets of data, in which 1296 groups were used as training samples and other 864 groups were used as test samples. The gears' fault patterns are given in Figure 8. Table 1 is the specific experimental sampling situation.

Table 2 shows 6 specific fault descriptions of the large and small gears of the input shaft of the gearbox and the load and speed during sampling. There are 21,600 groups of data which were collected by 8 sensors. 60% of the data were used for CNN training and 40% of the samples were used for testing.

5.2. Local Mean Signal Decomposition. LMD is used to decompose the vibration signal into time domain, and they are shown in Figure 9. The figure shows the time-domain waveform of the normal state, the broken teeth of the big gear, the pitting erosion of the big gear, the wear of the big gear, and the pitting erosion of the big gear. They are all under the 0.2a load with the measured speed of 1475 rpm.

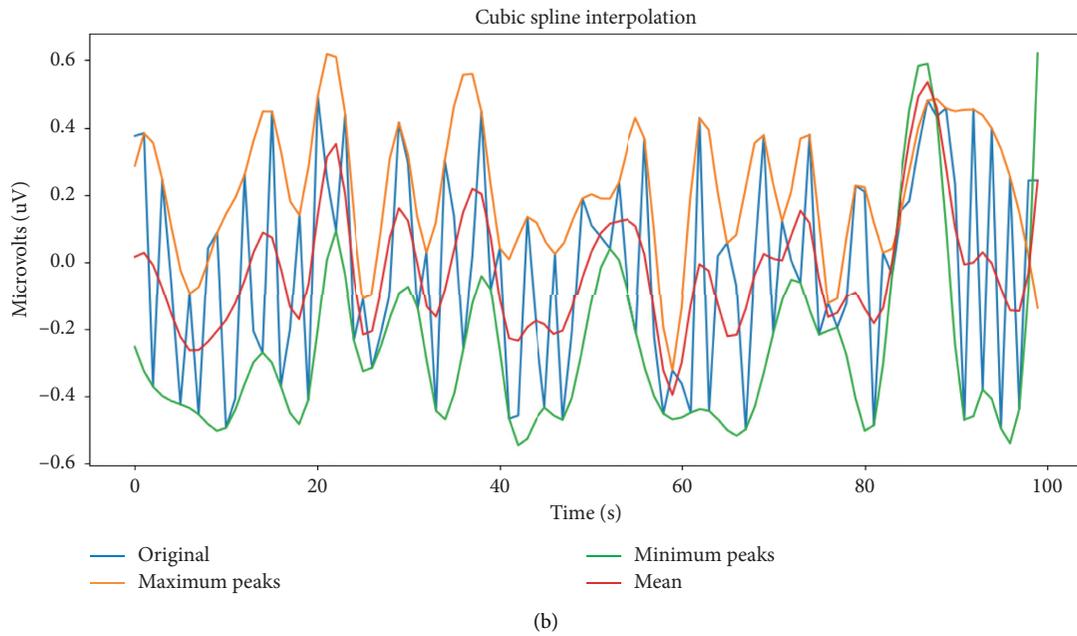
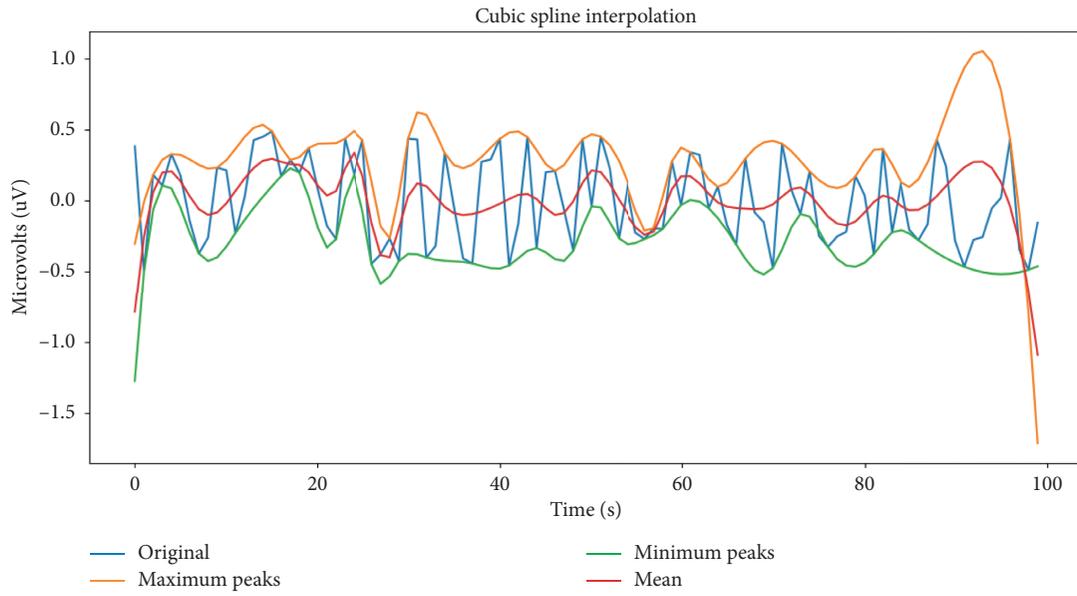


FIGURE 13: Continued.

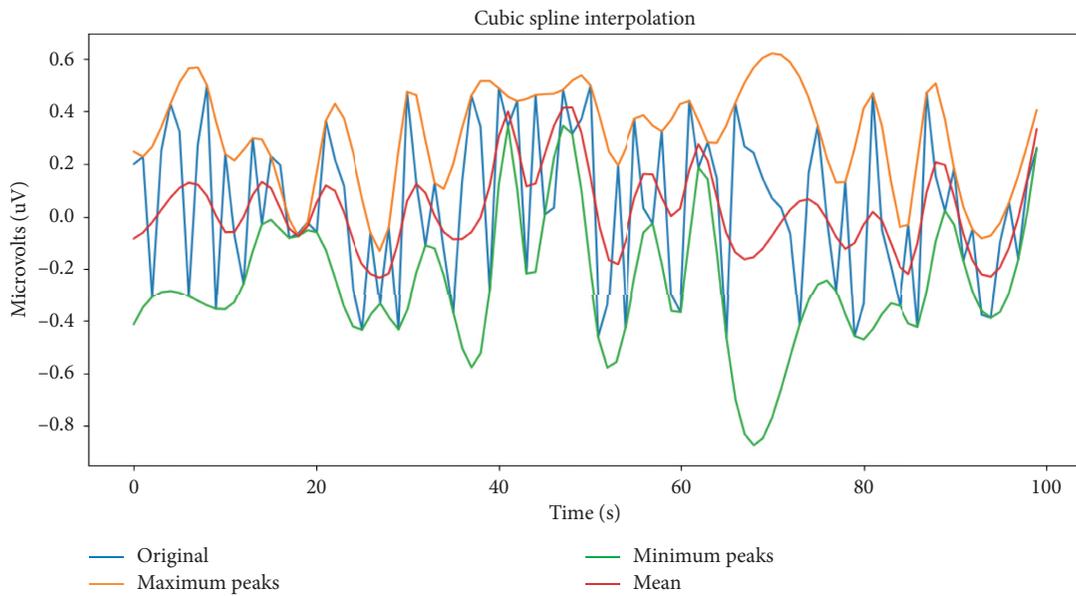
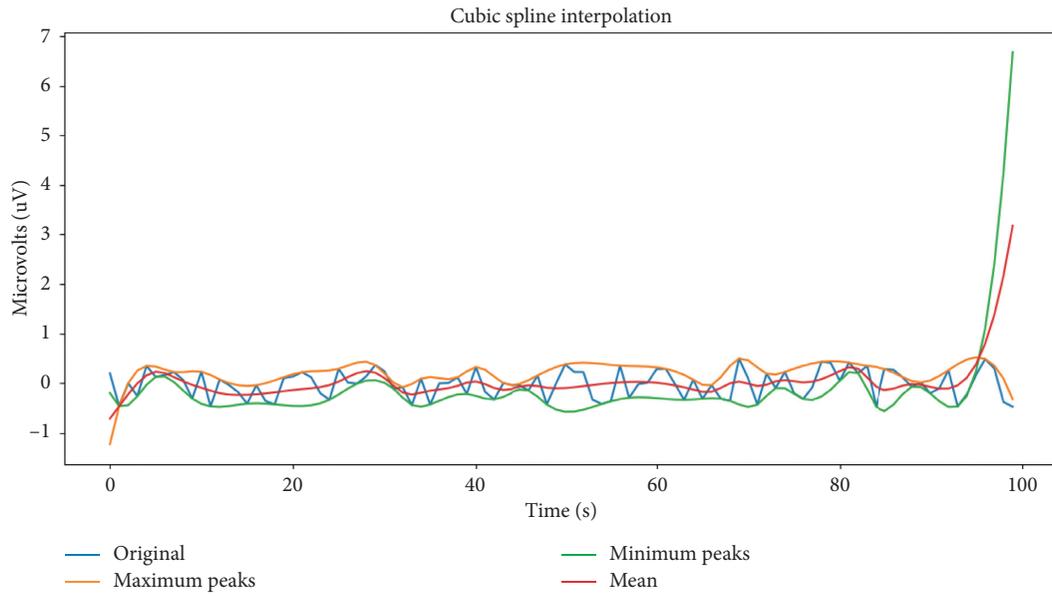
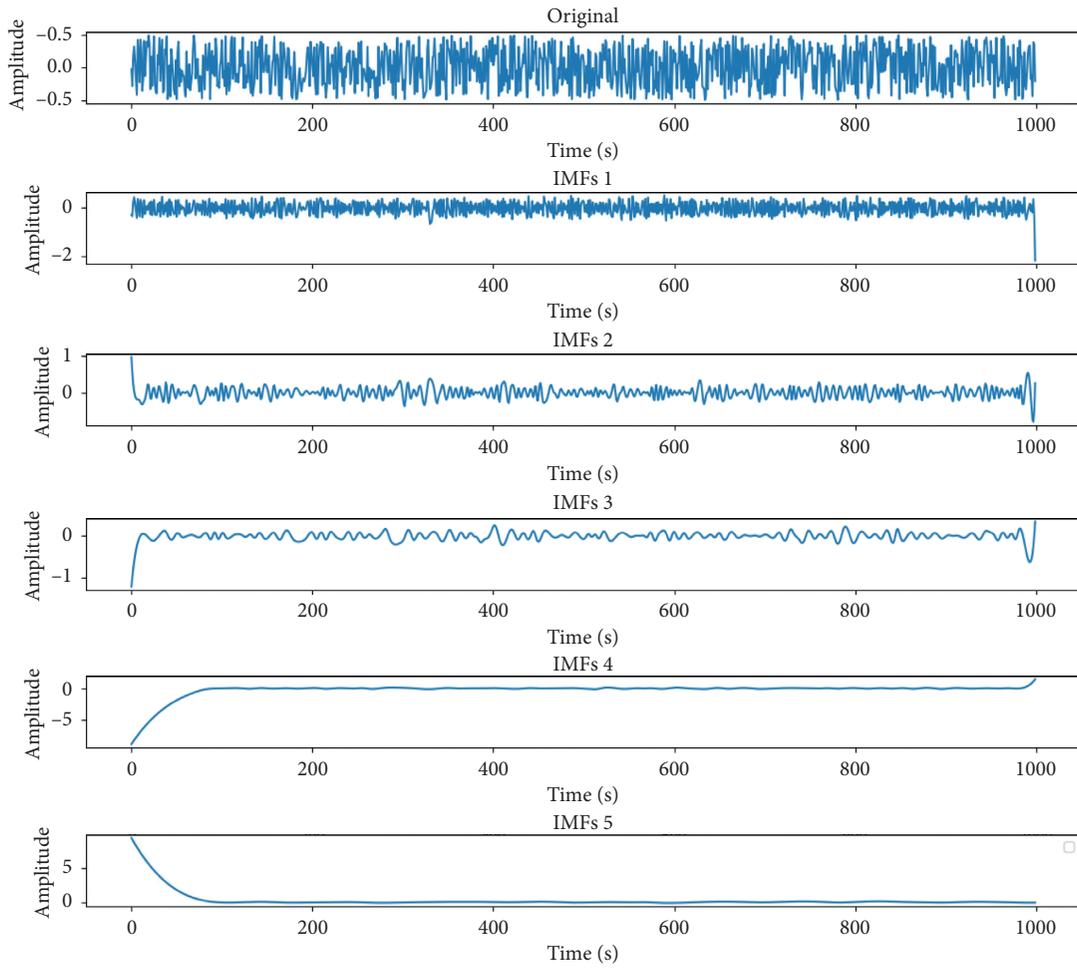


FIGURE 13: Continued.



(e)

FIGURE 13: Continued.

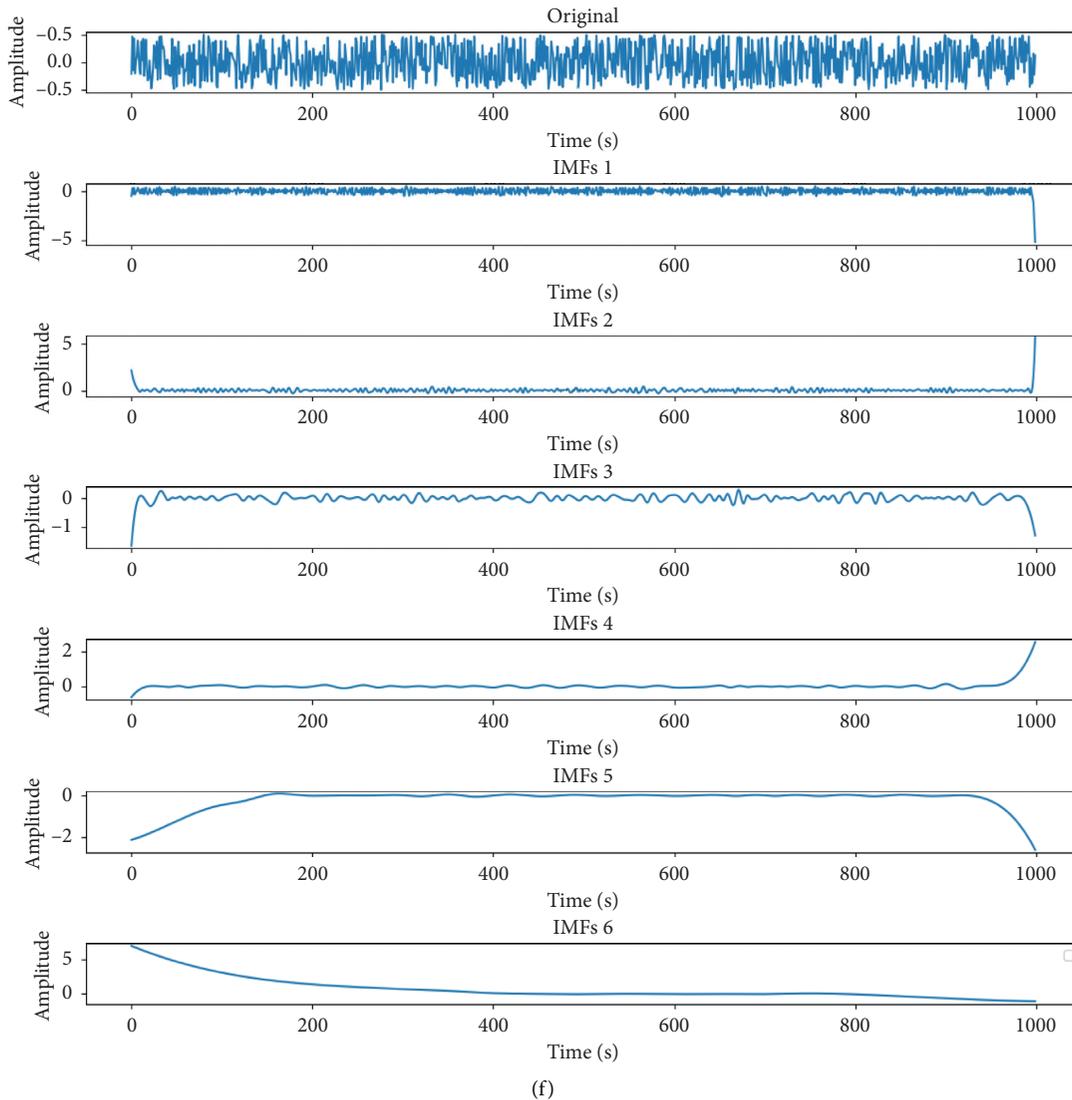


FIGURE 13: Cubic spline interpolation and INFs of EMD. The two faults are large gear broken teeth and small gear wear and large gear pitting and small gear wear.

Comparing these time-domain signals, the vibration interval is changing constantly. The waveforms of the last two compound faults are very similar, both of which have small amplitude and large random vibrations. Their average amplitude is also very similar, but the entropy of the large gear pitting and the pinion wear is larger than that of the large gear broken tooth and pinion wear. So, it is impossible to distinguish different compound faults only by time-domain and frequency-domain decomposition.

Time-domain signal decomposition is performed for different loads and speeds of the same fault. Four waveforms in Figure 10 are the time-domain waveforms of large gear pitting pinion wear under load of 0.05 A, 0.1 A, and 0.2 A at 880 rpm. By comparing the time-domain waveforms under different loads, it can be noticed that the increase of the load under compound fault will result in the increase of amplitude jumping frequency and larger amplitude vibration. However, the average amplitude tends to decrease. It can be

indicated that the load has an amplifying effect on the fault amplitude of the gear, but it will generate more classes for CNN training.

5.3. Comparison of LMD and EMD. There are many similarities between LMD and EMD as signal decomposition. Both of them have end-effects [34, 35]. In order to compare the advantages and disadvantages of the two, the fault signal is decomposed by EMD and LMD in turn. Figures 11–13 show the waveform and envelope diagrams generated after decomposition.

Both methods use small frequency ratio modal aliasing for amplitude correction. The vibration signals are decomposed at 1447 rpm and the same speed without load. LMD decomposes signals into PF0, PF1, PF2, PF3, and the remaining components. High-frequency signals can better represent the characteristics of the signal. After

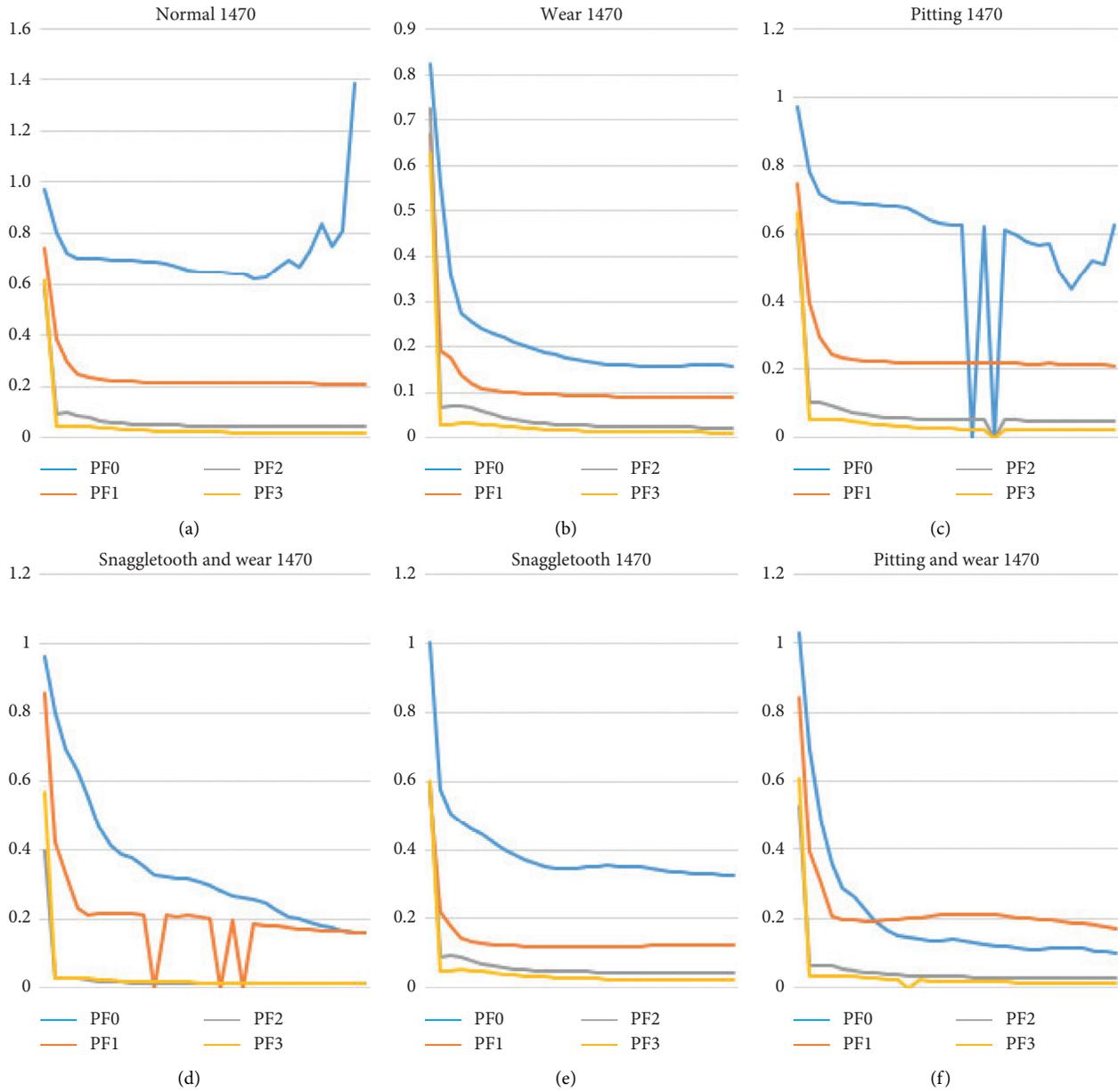


FIGURE 14: The curve of MESE.

TABLE 3: Two entropies of compound fault sample.

No. Component	Matrix sample entropy based on Euclidean distance		Original sample entropy	
	PF2	PF3	PF2	PF3
0.7	0.39566858	0.56896635	0.52580551	0.60952587
	0.02913883	0.02946744	0.06547757	0.03529882
0.3	1.4943305	1.8380982	1.9083757	1.9643323
	0.3207245	0.1855201	0.55637974	0.22477861

TABLE 4: Parameters of CNN based on LMD.

Layer (tyPF)	Output (ShaPF)	Param
ReshaPF_1 (ReshaPF)	(None, 6656, 1)	0
Conv1_1 (Conv1D)	(None, 3325, 16)	144
Conv1_2 (Conv1D)	(None, 1663, 16)	2064
Conv1_3 (Conv1D)	(None, 832, 16)	2064
Max_pooling1d_1n (Maxpooling1)	(None, 208, 16)	0
Conv1_4 (Conv1D)	(None, 104, 64)	4160
Conv1_5 (Conv1D)	(None, 52, 64)	16448
Max_pooling1d_2 (Maxpooling1)	(None, 26, 64)	0
Conv1_6 (Conv1D)	(None, 13, 256)	65792
Conv1_7 (Conv1D)	(None, 7, 256)	262400
Max_pooling1d_3 (Maxpooling1)	(None, 3, 256)	0
Conv1_8 (Conv1D)	(None, 3, 512)	262656
Conv1_9 (Conv1D)	(None, 3, 512)	524800
Max_pooling1d_4 (Maxpooling1)	(None, 1, 512)	0
Global_average_pooling1d_1	(None, 512)	0
Dropout_1 (dropout)	(None, 512)	0
Dense_1 (dense)	(None, 6)	3078

Total params: 1143606. Trainable params: 1143606. Nontrainable params: 0.

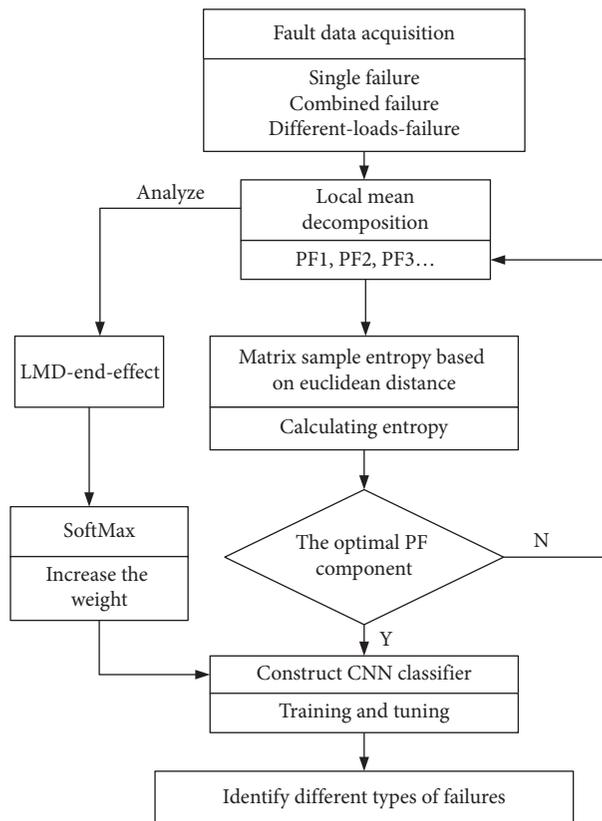


FIGURE 15: The flow chart of final program.

decomposition, it is easy to see that PF0 and PF1 have a large difference in value; PF2, PF3, and the remaining components also show different change disciplines. EMD separates the trends of different feature scales in the signal layer by layer and it generates a series of signal components IMF1 to

IMF6 with different feature scales. Comparing the IMF of the two compound faults, the endpoint effect of EMD is very serious [36]. Due to the end effect, the decomposition of the same signal can completely be useless. Because false end components will gradually interfere with the entire signal

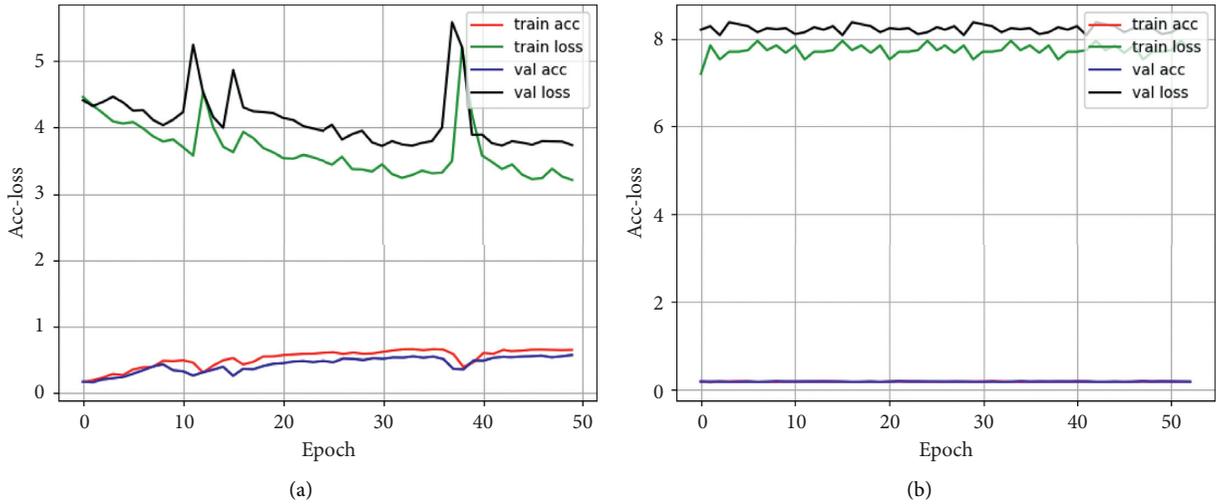


FIGURE 16: CNN trained by original failure data and CNN trained by all LMD-PFs.

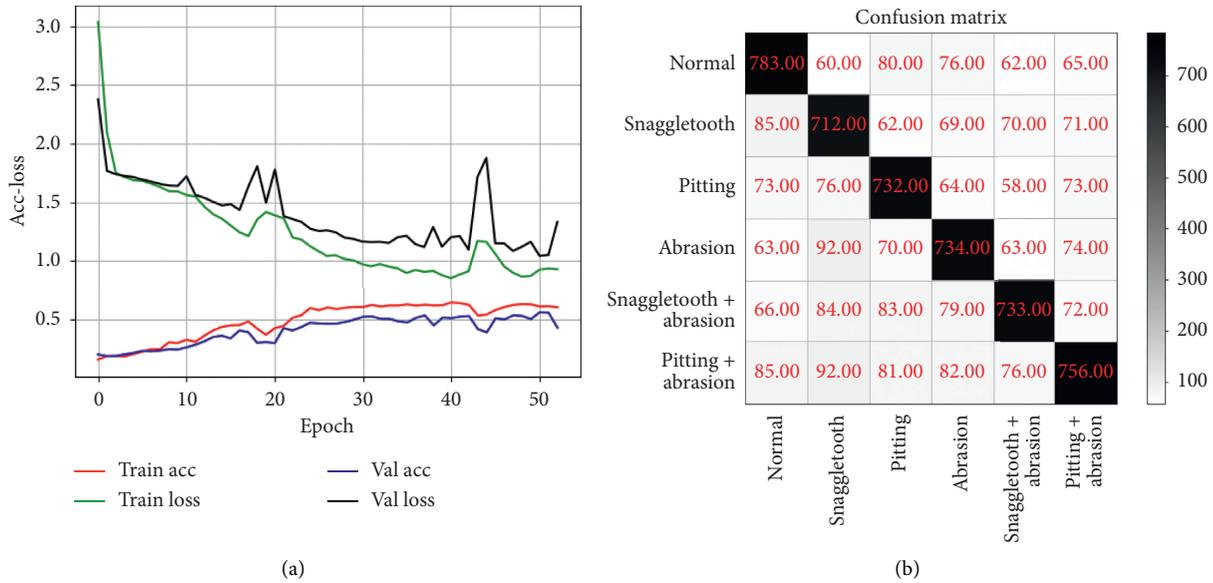


FIGURE 17: The 1-D CNN curve of accuracy rate and loss rate trained by IMF2 of EMD and confusion matrix.

sequence from the end point, EMD is suitable to deal with faulty signals. The fault diagnosis with EMD has great disadvantages.

5.4. *Selecting the Most Useful PFs by MESE.* The original signal standard deviation is set to 0.7. The entropy template vector length is set to 5×5 , and $m = 2$ for sample entropy extraction. Figure 14 shows results of extraction for each state.

The sample entropy curve of PF0, which is supposed to fall steadily, shows jump and reverse growth. Compared with PF2 and PF3, the signals are very similar without obvious pulsation. Therefore, sample entropy with standard deviation of 0.3 is collected for further comparison and the

original sample entropy of the same data is collected for comparison. Table 3 lists the SE and MESE values of PF2 and PF3 with standard deviations of 0.3 and 0.7.

LMD is a method to separate signals from high frequency to low frequency successively, so PF2 should have higher sample entropy than PF3. Entropy of PF3 is higher than that of PF2, but it is lower than that of a second-order sample. This is caused by the endpoint effect of LMD in the vicinity of PF2 so that the entire PF3 data may become false. After analysis, this set of data PF2 is more suitable for CNN training than PF3. After comparison, it can be seen that the PF value obtained by MESE has a greater difference than the value obtained by the original one. In other words, MESE can better increase the difference between each entropy value than the original SE. The larger component of

characteristic entropy is selected and used in CNN training. The CNN which trained by MESE has higher accuracy. Formula (26) ensures the MESE of a PF component should be monotonically decreasing. Formula (27) ensures the MESE of high-order PF component should be higher than the low-order:

$$\prod_{m=1}^{50} \left[\frac{\text{MESE}(m, r) - \text{MESE}(m+1, r)}{\text{MESE}(m, r)} \right] > 0, \quad (26)$$

$$\prod_{n=0}^4 \prod_{m=1}^{50} [PF_n \times (\text{MESE}_m) - PF_{n+1} \times (\text{MESE}_m)] > 0. \quad (27)$$

5.5. Construction of One-Dimensional CN. The weight-optimized SoftMax is used as the activation function of the final classification to train the one-dimensional CNN under different working conditions to classify complex fault types. In the training process, the Adam algorithm is used to optimize the model, and the generator which can greatly save the time to read the data is used to read the data.

The model is built using Keras in a Python environment. For the first time, a filter with a height of 10 is defined. In order to enable the CNN to learn a single feature, 100 filters are defined in the first layer. The output of the first layer is $71 * 100$ matrices, and each column of the output matrix contains a weight of the filter.

Maximum pooling layer of size 4 is selected in the pooling layer, which makes the pooled matrix only 1/4 of the original matrix [37], and it reduces the complexity of the next one and the probability of data overfitting. Continuous convoluting and pooling make the state and fault characteristics of multiple sensors of vibration data appear as much as possible.

An average pooling layer is added to further reduce the probability of overfitting during training after multiple convolution and pooling. The average pooling layer is the average value of the two weights, and the output size is $512 * 1$. Each feature detector has only one weight through this layer.

Due to the diversity of sensor sampling and the change of load, the ReLu function is used for activation in the first three, ranging from 0.1–0.6 loss rate for training and the last activation function is changed to sigmoid. It improves the generalization ability of the network. During training, dropout can be used to reduce the loss rate. The dropout layer can randomly assign neurons to 0. Because of too many convolutions and multiple sensors in different ways, the weight is assigned 0.8, which means 80% of the neurons will have zero weight. The network will not respond sensitively to small changes in data. This layer does not change the size of the output matrix. A value of 1.0 was assigned to the τ of improved SoftMax function and it forms a classification layer. It is determined that $\tau = 2.6$ is the most appropriate after a series of training. In order to increase or decrease the number of test periods and improve the accuracy, the strategy of extracting stop method is adopted to stop in time when verifying errors of training. Table 4 shows the specific situation of CNN training.

5 kinds of data are input into one-dimensional convolutional neural network for training and testing, and parameters are adjusted repeatedly to adjust dropout to 1.0 and learning rate to 0.05, and SoftMax is selected as the activation function. The arithmetic mean value of each category index is used as the evaluation standard. The equation is expressed as

$$F(i) = \frac{1}{n} \sum_i^n Fl_i = \frac{1}{n} \sum_i^n \frac{2 \times P_i \times R_i}{P_i + R_i}, \quad (28)$$

where P_i is the precision corresponding to the i -th category and R_i is the recall corresponding to the i -th category.

The flow chart of final program is expressed as Figure 15.

5.6. Comparison of Fault Recognition with LMD-MESE-CNN and Other Methods. The raw data are input into CNN for fault diagnosis; the result is shown in Figure 16(a). The accuracy is only 0.71. A preliminary conclusion can be drawn: the model that directly trains the CNN has a low recognition rate and the net cannot be used for fault diagnosis. All PF components are directly input into CNN for fault diagnosis; the result is shown in Figure 16(b). After LMD decomposition, the faulty data are obviously distributed in each PF component. If all PF components are input into CNN for training, the accuracy rate of CNN drops to only 10%. Therefore, neither data input method is suitable for fault diagnosis.

The analysis curve and the obfuscation matrix are shown in the left of Figures 17–19. The corresponding obfuscation matrices [38, 39], whose schematic is in Figure 20, are shown in the right of those figures. The accuracy of EMD-CNN training models keeps increasing, but there are some random declines and the accuracy cannot be achieved appropriately. When the selected PF signals are input into CNN for training, the net obtains high accuracy and low loss value. However, feature curve does not increase linearly as the number of training data increases. It can also be seen from the confusion matrix that most conditions can be separated. But a small number of confusion features does not. Because some false data are generated by LMD endpoint effect, errors are produced in the training model. The location of these errors is difficult to estimate, so it is essential to find a way to reduce the loss caused by the endpoint effect [40]. The last one shows that those PFs, selected by MESE, are input into CNN with weight-SoftMax and it is found that the curve jump is disappeared. The results show that one-dimensional CNN with selected PFs can effectively reduce the impact of endpoint effect. This net can accurately identify combined failure. Table 5 shows all the accuracy and loss of these methods.

5.7. Comparing the Proposed Method with Other Methods. Compared with the traditional 1-D CNN, the signal analysis method based on LMD and MESE proposed in this paper can not only reduce the network parameters and speed up the training speed but also obtain good results for the diagnosis of mixed faults. The network structure proposed by

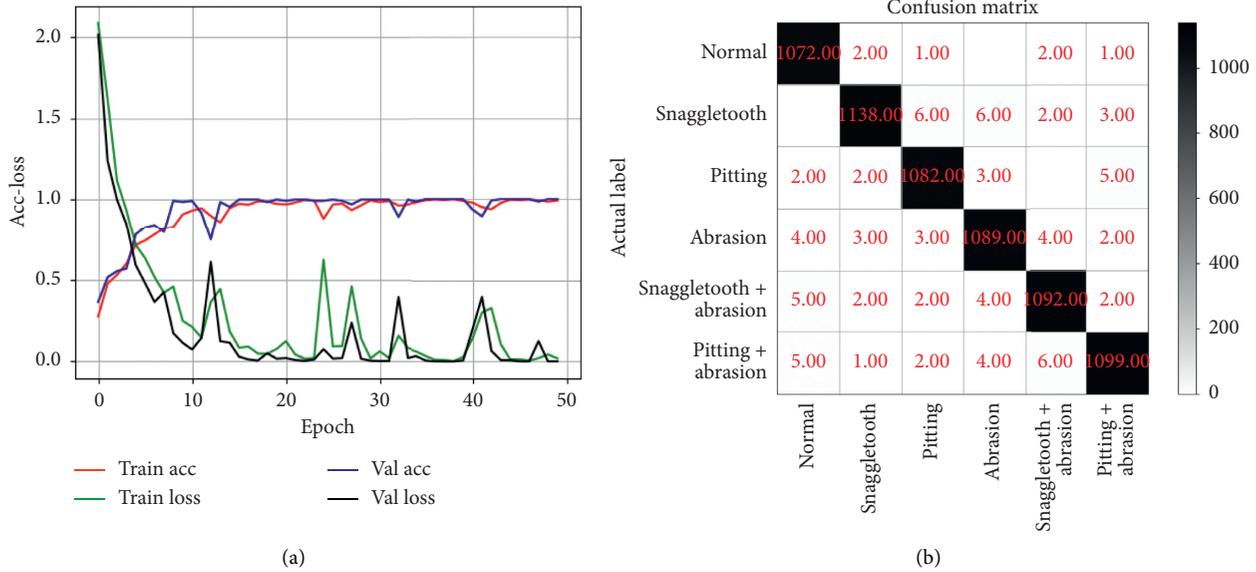


FIGURE 18: The 1-D CNN curve of accuracy rate and loss rate trained by PF of LMD and confusion matrix.

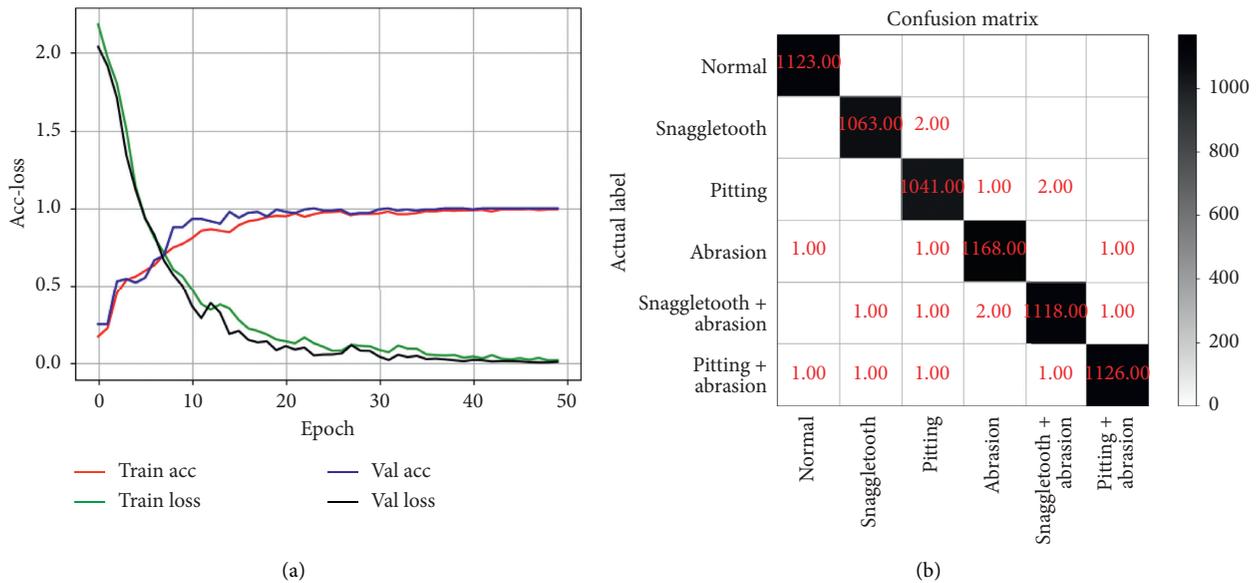


FIGURE 19: The optimization of CNN curve of accuracy rate and loss rate trained by PF of LMD and confusion matrix.

this method is compared with other methods proposed in accuracy and running time. As shown in Table 6, the test accuracy of the DCNN method based on LRP is 99.90%, which is very close to the proposed method. It can be seen from Table 6 that the deep learning method has a higher accuracy for the compound fault diagnosis of the gearbox. The method proposed in this paper is slightly better than other deep learning methods on the testing set.

The number of training parameters and training time of the four best methods are shown in Table 6. It can be seen from Figure 21 that the training parameters of 1-D CNN fault diagnosis method based on LMD and MESE proposed in this paper are about 10% less than that of the

traditional 1-D CNN. The training parameters of this method are much lower than those of LRP-DCNN and GRU-CNN method. Although LPR-CNN has a high accuracy rate, the training parameters and time are much higher than the proposed method. The training time of the traditional CNN method is shorter, but the accuracy rate is much lower than the proposed method. The method proposed in this paper can reduce the number of training parameters and reduce the training time based on excellent diagnostic performance.

In the future, transfer learning to reduce the time of data tabulation and classification will be applied to achieve better universality and higher efficiency.

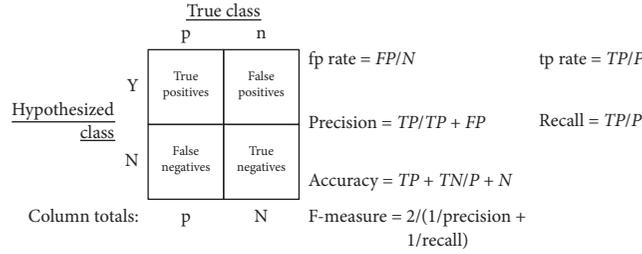


FIGURE 20: The structure and principle of the confusion matrix.

TABLE 5: The accuracy and loss rate of 1-D CNN.

Method	Fault detection			
	Train-acc	Val-acc	Train-loss	Val-loss
CNN	72.68	66.37	3.21	3.76
LMD-CNN	36.67	35.21	7.68	8.24
EMD-CNN	61.37	46.32	0.91	1.39
LMD-PF2-CNN	99.21	99.07	0.07	0.11
LMD-SoftMax CNN	100.00	99.97	0.03	0.06

TABLE 6: The accuracy comparison between the proposed method in this paper and the other methods.

Accuracy	Training set	Testing set
Proposed method	100.00	99.97
Traditional CNN	98.40	96.80
LRP-DCNN	100.00	99.90
GRU-CNN	99.81	99.75
DTL	97.50	93.64
1-D CAE	98.80	97.50

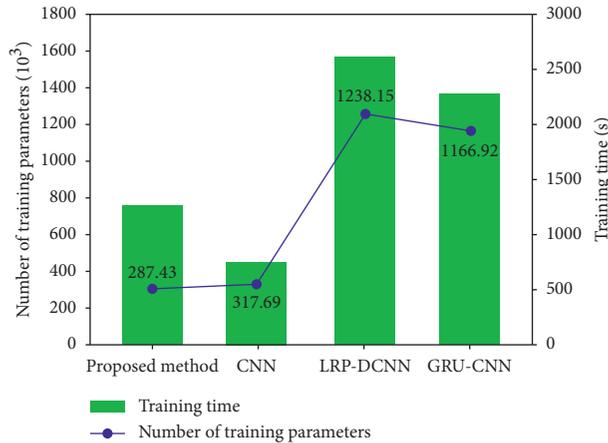


FIGURE 21: The number of training parameters and the training time of the four methods.

6. Conclusions

For vibration signals extracted from a working gearbox contain complex noise, nonlinearity, and nonstationarity, a one-dimensional convolutional neural network algorithm (1-D CNN) with weighted activation function based on the advantage of CNN depth to extract signal features is designed. Then, an improved matrix sample entropy based on Euclidean distance (MESE) is proposed to improve the

selection accuracy of components, speed up the classification rate, and avoid the identification error caused by too large difference of PF components. The LMD signal decomposition method is used to decompose the fault signals. The end effect of LMD is solved, and the appropriate PF component is obtained. The gearbox fault data with 24 different loads were used to verify the effectiveness of the proposed method. From the experiment results, the recognition speed of PF component by MESE is several times faster than SE, and PF

component with greater difference can be obtained. For the end effect of LMD, the improved 1-D CNN algorithm applies the method of weight compensation, which can eliminate the effect from the signal starting point and obtain the correct PF component. For the fault signal of different loads, four times convolutions and poolings and parameters adjustment of neural network can effectively extract the fault feature. The use of 1-D CNN greatly accelerates the speed and accuracy of convolutions and poolings. It avoids the problem of overfitting and too much parameter and reduces the running time of the algorithm. Therefore, this method can solve the end effect of LMD, accelerate the screening of component differences and classification, and quickly diagnose the complex faults and faults with different loads. This study provides a good choice for the gearbox fault diagnosis with nonideal conditions. The fault diagnosis based on vibration signals is not only for single faults but also for more complex and uncertain gearbox faults.

Due to the deepening of the number of CNN layers, more data are needed for multiple types of fault recognition, which makes the method take more time to train new models for new faults. It is important to make the trained one-dimensional CNN model have better versatility and mobility in the future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request. Data were collected from the parallel-shaft gearbox in the Fault Diagnosis Laboratory of Inner Mongolia University of Science and Technology. The data are precious, so the tutor hopes that the data and code will not be uploaded to the database. If it is necessary, the readers can contact the corresponding author via email: 2397595377@qq.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] L. Cui, C. Kang, H. Wang, and P. Chen, "Application of composite dictionary multi-atom matching in gear fault diagnosis," *Sensors*, vol. 11, no. 6, pp. 5981–6002, 2011.
- [2] Z. Feng and M. Liang, "Complex signal analysis for planetary gearbox fault diagnosis via shift invariant dictionary learning," *Measurement*, vol. 90, pp. 382–395, 2016.
- [3] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, vol. 100, 2021.
- [4] D. Yu, M. Wang, and X. Cheng, "A method for the compound fault diagnosis of gearboxes based on morphological component analysis," *Measurement*, vol. 91, pp. 519–531, 2016.
- [5] Y. Zhang, W. Lu, and F. Chu, "Planet gear fault localization for wind turbine gearbox using acoustic emission signals," *Renewable Energy*, vol. 109, pp. 449–460, 2017.
- [6] M. Rui, Y. Chen, and Q. Cao, "Research on dynamics and fault mechanism of spur gear pair with spalling defect," *Journal of Sound and Vibration*, vol. 331, no. 9, pp. 2097–2109, 2012.
- [7] K. F. Brethee, D. Zhen, F. Gu, and A. D. Ball, "Helical gear wear monitoring: modelling and experimental validation," *Mechanism and Machine Theory*, vol. 117, pp. 210–229, 2017.
- [8] S. Javier, P. Ricardo, and C. Huerta, "Gear dynamics monitoring using discrete wavelet transformation and multi-layer perceptron neural networks," *Soft Computing*, vol. 12, no. 9, pp. 2867–2878, 2012.
- [9] S. Guoji, S. McLaughlin, X. Yongcheng, and P. White, "Theoretical and experimental analysis of bispectrum of vibration signals for fault diagnosis of gears," *Mechanical Systems and Signal Processing*, vol. 43, no. 1, pp. 76–89, 2014.
- [10] L. Bai, Z. Han, Y. Li, and S. Ning, "A hybrid de-noising algorithm for the gear transmission system based on CEEM-DAN-PE-TFPPF," *Entropy*, vol. 20, no. 5, p. 361, 2018.
- [11] Z. Wu, Q. Zhang, L. Cheng, and S. Tan, "A new method of two-stage planetary gearbox fault detection based on multi-sensor information fusion," *Applied Sciences*, vol. 9, no. 24, p. 5443, 2019.
- [12] F. Li, R. Li, L. Tian, L. Chen, and J. Liu, "Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions," *Mechanical Systems and Signal Processing*, vol. 116, pp. 462–479, 2019.
- [13] D. Xiao, J. Ding, X. Li, and L. Huang, "gear fault diagnosis based on kurtosis criterion VMD and SOM neural network," *Applied Sciences*, vol. 9, no. 24, p. 5424, 2019.
- [14] R. Ricci and P. Pennacchi, "Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions," *Mechanical Systems and Signal Processing*, vol. 25, no. 3, pp. 821–838, 2011.
- [15] Y. Li, M. Xu, Y. Wei, and W. Huang, "An improvement EMD method based on the optimized rational hermite interpolation approach and its application to gear fault diagnosis," *Measurement*, vol. 63, pp. 330–345, 2015.
- [16] M. Hafida, C. Fakher, and A. Felkaoui, "Detection of gear faults in variable rotating speed using variational mode decomposition (VMD)," *Mechanics and Industry*, vol. 17, no. 2, p. 207, 2016.
- [17] S. Hoochang, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [18] T. Chen, L. Lin, L. Liu, X. Luo, and X. Li, "DISC: deep image saliency computing via progressive representation learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1135–1149, 2015.
- [19] L. Sergey, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1334–1373, 2016.
- [20] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7319–7327, 2020.
- [21] Z. Chen, C. Li, and R.-V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock and Vibration*, vol. 2015, Article ID 390134, 10 pages, 2015.
- [22] G. John, P. Wang, C. Sun, and R. X. Gao, "Explainable convolutional neural network for gearbox fault diagnosis," *Procedia CIRP*, vol. 80, pp. 476–481, 2019.
- [23] Y. Yao, H. Wang, S. Li et al., "End-to-end convolutional neural network model for gear fault diagnosis based on sound signals," *Applied Sciences*, vol. 8, no. 9, p. 1584, 2018.

- [24] X. Li, J. Li, Y. Qu, and D. He, "Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals," *Applied Sciences*, vol. 9, no. 4, p. 768, 2019.
- [25] J. Li, X. Li, D. He, and Y. Qu, "A domain adaptation model for early gear pitting fault diagnosis based on deep transfer learning network," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 234, no. 1, pp. 168–182, 2020.
- [26] S. Chen, J. Yu, and S. Wang, "One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes," *Journal of Process Control*, vol. 87, pp. 54–67, 2020.
- [27] S. Rezaian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the 2014 Computer Vision and Pattern Recognition Conference*, pp. 512–519, Columbus, OH, USA, June 2014.
- [28] O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, no. 1, pp. 331–345, 2016.
- [29] S. Ning, Z. Han, Z. Wang, and X. Wu, "Application of sample entropy based LMD-TFPP de-noising algorithm for the gear transmission system," *Entropy*, vol. 18, no. 11, p. 414, 2016.
- [30] H. Li, "Fault diagnosis of gear wear based on local mean decomposition," *Advanced Materials Research*, vol. 459, pp. 289–302, 2012.
- [31] C. Zhang, Z. Li, C. Hu, S. Chen, J. Wang, and X. Zhang, "An optimized ensemble local mean decomposition method for fault detection of mechanical components," *Measurement Science and Technology*, vol. 28, 2017.
- [32] N. Zhang, A. Lin, H. Ma, P. Shang, and P. Yang, "Weighted multivariate composite multiscale sample entropy analysis for the complexity of nonlinear times series," *Physica A: Statistical Mechanics and Its Applications*, vol. 508, pp. 595–607, 2018.
- [33] X. Liu, Q. Zhou, J. Zhao, H. Shen, and X. Xiong, "Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network," *Sensors*, vol. 19, no. 4, p. 972, 2019.
- [34] W. Huang, J. Zeng, Z. Wang, and J. Liang, "Partial noise assisted multivariate EMD: an improved noise assisted method for multivariate signals decomposition," *Biomedical Signal Processing and Control*, vol. 36, pp. 205–220, 2017.
- [35] Z. Liu, X. Chen, Z. He, and Z. Shen, "LMD method and multi-class RWSVM of fault diagnosis for rotating machinery using condition monitoring information," *Sensors*, vol. 13, no. 7, pp. 8679–8694, 2013.
- [36] T. Ma, Z. Liu, and Z. Zhou, "Inhibition of envelope correction method to EMD end effect," *Applied Mechanics and Materials*, vol. 347-350, pp. 3940–3944, 2013.
- [37] S. Kim and J.-H. Choi, "Convolutional neural network for gear fault diagnosis based on signal segmentation approach," *Structural Health Monitoring*, vol. 18, no. 5-6, pp. 1401–1415, 2019.
- [38] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," *Communications in Statistics-Theory and Methods*, vol. 49, no. 9, pp. 2080–2093, 2020.
- [39] L. Amalia, A. Carrasco, A. Martín et al., "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 2019.
- [40] F. W. Wang, Y. Q. Zhou, and S. J. Zhou, "Study on methods for improving LMD end effect," *Applied Mechanics and Materials*, vol. 851, pp. 559–566, 2016.