

Research Article

Fault Diagnosis of Planetary Gearbox Based on Motor Current Signal Analysis

Ziyuan Jiang, Qinkai Han , and Xueping Xu

The State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Qinkai Han; hanqinkai@hotmail.com

Received 7 July 2020; Revised 7 August 2020; Accepted 7 September 2020; Published 14 September 2020

Academic Editor: Changqing Shen

Copyright © 2020 Ziyuan Jiang 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.

Planetary gearbox is one of the most widely used core parts in heavy machinery. Once it breaks down, it can lead to serious accidents and economic loss. Induction motor current signal analysis (MCSA) is a noninvasive method that uses current to detect faults. Currently, most MCSA-based fault diagnosis studies focus on the parallel shaft gearbox. However, there is a paucity of studies on the planetary gearbox. The effect of various signal processing methods on motor current and the performance of different machine learning models are rarely compared. Therefore, fault diagnosis of planetary gearbox based MCSA is conducted in this study. First, the effects of various faults on motor currents are studied. Specifically, the characteristic frequencies of a fault in sun/planet/ring gears and supporting bearings of the planetary gearbox are derived. Then, a signal preprocessing method, namely, singular spectrum analysis (SSA), is proposed to remove the supply frequency component in the current signal. Subsequently, four classical machine learning models, including the support vector machine (SVM), decision tree (DT), random forest (RF), and AdaBoost, are used for fault classifications based on the features extracted via principal component analysis (PCA). The convolutional neural network (CNN), which can automatically extract features, is also adopted. The dynamic experiment of the planetary gearbox with seven types of faults, including tooth chipping in sun/planet/ring gears, inner race spall in planet bearing, inner/outer races, and ball spalls in input support bearing, is conducted. Raw current signal in the time domain, reconstructed signal by SSA, and the current spectra in the frequency domain are used as the inputs of various models. The classification results show that the PCA-SVM is the best model for learned data while CNN is the best model for unlearned data on average. Furthermore, SSA mainly increases the accuracy of CNN in the time domain and exhibits a positive effect on unlearned data in the time domain. The classification accuracy increases significantly after transforming the time domain current data to the frequency domain.

1. Introduction

A planetary gearbox exhibits the characteristics of a compact structure with a large transmission ratio and high transmission efficiency [1]. The mechanical power system is composed of a planetary gearbox and the induction motor is widely used in heavy-duty equipment, such as helicopters, wind turbines, and large cranes. Due to heavy loads and the undesirable working environment, the key components in a planetary gearbox, such as the planet gear, sun gear, ring gear, and supported bearing, are prone to contact fatigue, wear, and even fracture, which may lead to accidents. Once there is a fault in the planetary gearbox, the system is shut down for maintenance. This leads to a reduction in

production efficiency and economic loss. Therefore, it is necessary to conduct fault diagnosis of planetary gearboxes.

A fault in rotating machinery equipment affects the motor current via torque transmission [2]; thus motor current signal analysis (MCSA) was proposed for condition monitoring. This is a noninvasive method that utilizes the current signal to detect faults. Given that the induction motor usually has its own current monitoring system, there is no need to install additional sensors. However, acquisition of other signals, such as vibration, temperature, and acoustics [3–5], requires installation of additional intrusive sensors, which is costly and may damage the original structure [6]. Fault diagnosis of rotating machinery based on MCSA has received

significant attention in extant studies. Most studies were focused on the parallel shaft gearboxes and motor bearings.

Yilmaz and Ayaz [7] used the sum of the power spectral densities for certain frequency intervals of motor current and vibration signal, as input features, and added temperature as another feature. Subsequently, they used adaptive neuro-fuzzy inference systems (ANFIS) to classify the fault of motor bearing. Lessmeier et al. [8] conducted experiments on motor bearings with inner and outer race faults and classified motor current signals detected in the experiment using a variety of classical machine learning models. In addition to the abovementioned method of signal analysis, Han et al. [9] proposed an intelligent triboelectric bearing, which exhibits the ability of self-powering and self-sensing wherein the output current changes according to the bearing fault. Stator current modeling for defective rolling bearings based on magnetic equivalent circuits was also conducted [10]. As for fault diagnosis of a gearbox, Mohanty and Kar [11] studied the induction motor current signal of a three-stage gearbox under fault condition via amplitude demodulation and frequency demodulation. Jiang et al. [12] analyzed vibration data measured from a transmission gearbox with gear fault and discussed the impact of initial center frequencies on variational mode decomposition. When compared with the parallel shaft gearbox, the periodic situation of the planetary gearbox is more complex. First, there are multiple planet gears that mesh with the sun gear. This can lead to signal superposition [13]. Second, the planet gear not only rotates but also revolves with the carrier. Third, the original property of the periodic signal is varied with many modulation signals and a complex transmission path. Thus, fault diagnosis in planetary gearbox becomes more difficult than those of parallel shaft gearbox and motor bearing due to these characteristics. Furthermore, there is a paucity of studies on this topic. For planetary gearbox under different fault conditions, time-frequency space vector modulus analysis [14] and magnetic equivalent model [15] of stator current have been established. Furthermore, Zhang et al. [16] analyzed the motor current under the condition of ring gear tooth fracture via a resonance residual method, which proves that the frequency spectrum near the natural frequency of the transmission system contains rich information.

In terms of research methods, demodulation, wavelet transform [17], empirical mode decomposition [18], Hilbert–Huang transform [19], extracting and analyzing nonlinearities [20], and other methods are commonly adopted for extracting fault features from signals. With the enhancement of computing performance, a data-driven machine learning model has been developed [21, 22]. Increasingly, more researchers are beginning to use signal analysis and statistical methods as the basis of feature extraction and combine the traditional machine learning model for intelligent fault diagnosis. Lei et al. [23] conducted filtering and empirical mode decomposition for vibration signals of the motor under different motor bearing faults in time and frequency domains. Furthermore, Lei selected features and classified them with the genetic algorithm and

ANFIS. Peng et al. [24] designed two support vector machine (SVM) models to classify the vibration and current signals of the gearbox in wind turbines according to the extracted features and used D-S information fusion to determine the fault category. Cheng et al. [25] proposed a self-encoder, which reduces the dimension of input signal from different fault experiments and then adopted a multiclass SVM model to classify the output of the self-encoder.

To extract representative features from signals, the methods of intelligent fault diagnosis mentioned above are time-consuming. To a great extent, the methods depend on the prior knowledge of signal processing technology and rotor dynamics. However, unsupervised learning and deep learning represented by neural networks reduce the requirement of prior knowledge because of their ability to automatically extract features and lead to direct diagnoses of the original signal or simply processed signal. Shen et al. [26] proposed a method based on stacked contractive autoencoder for feature extraction from amplitude spectrums. Jiang et al. [27] used the deep belief network to fuse the vibration signal and current signal to classify the fault of a two-stage gearbox. Hoang and Kang [28] adopted a convolutional neural network to obtain the features from two phases of motor currents and conducted the fault diagnosis based on the information fusion of perceptron, support vector machine, and k -nearest neighbor. Li et al. [29] proposed a deep random forest model, which integrates different signals as input, to diagnose a two-stage parallel gearbox. Lei et al. [30] proposed a two-stage bearing fault diagnosis method. In the first stage, unsupervised sparse filtering was used to extract features from raw signals, and the softmax function was adopted for classification in the second stage.

From the above review, it is confirmed that the existing research on MCSA-based fault diagnosis of planetary gearboxes is relatively less, and the performance of different feature extraction methods and classification models should be examined. Therefore, fault diagnosis of planetary gearbox based MCSA is conducted in this study. All the techniques adopted in this paper are common, and one of the aims of this study is to aid engineers and researchers in selecting an appropriate strategy for practical planetary gearbox fault detection via discussing preprocessing, feature extraction, and classification models. After the effects of various fault on characteristic frequencies of motor current are studied, a signal preprocessing method, singular spectrum analysis (SSA), is proposed to remove the supply frequency component and obtain a reconstructed signal. Subsequently, a convolutional neural network (CNN) and four classical machine learning models are selected to determine the type of fault. In this case, principal component analysis (PCA) is performed to extract features as the inputs of various machine learning models. Thus, dynamic experiments on planetary gearbox with seven types of faults, including tooth chipping in sun/planet/ring gears, inner race spall in planet bearing, inner/outer races, and ball spalls in input support bearing, are conducted. Finally, a few conclusions are summarized at the end.

2. Influence of Fault on Motor Current

A typical planetary gearbox structure is shown in Figure 1, which consists of sun/planet/ring gears and supporting bearings. When a fault occurs, such as tooth chipping in gears or spall in bearings, it causes the load to oscillate.

Therefore, the load torque under fault condition can be expressed as the sum of a fixed torque and periodic oscillation torque due to fault. The periodic oscillation torque can be expanded via Fourier expansion. In the case of neglecting the higher harmonics,

$$M(t) = M_0 + M_c \cos(\omega_c t), \quad (1)$$

where M_c denotes the amplitude of oscillation torque, $\omega_c = 2\pi f_c$, f_c denotes the characteristic frequency due to fault. Given that the oscillating torque is generated via rotation, f_c is generally considered a multiple of the input shaft rotation frequency f_r , that is, $f_c = N f_r$. N is greater than 0, as determined by the structure.

In the planetary gearbox, the fault characteristic frequency of sun gear, planet gear, and ring gear can be expressed as follows [31]:

$$f_c = \frac{f_m}{Z^*} = f_r \frac{(Z_{\text{ring}}/(1 + (Z_{\text{ring}}/Z_{\text{sun}})))}{Z^*}, \quad (2)$$

where f_m denotes the meshing frequency, Z_{ring} denotes the number of ring gear teeth, Z_{sun} denotes the number of sun gear teeth, and Z^* denotes the number of teeth corresponding to the faulty part.

In addition to gears, rolling element bearings are also used to support the rotary components of the planetary gearbox. By considering the rolling bearing that supports the sun gear, as an example, the fault characteristic frequencies of the outer and inner race and the ball are shown, respectively, as follows:

$$\begin{aligned} f_{c_outer} &= f_r \frac{n}{2} \left(1 - \frac{d}{D}\right), \\ f_{c_inner} &= f_r \frac{n}{2} \left(1 + \frac{d}{D}\right), \\ f_{c_ball} &= f_r \frac{D}{d} \left(1 - \left(\frac{d}{D}\right)^2\right), \end{aligned} \quad (3)$$

where n denotes the number of balls, d denotes the diameter of the ball, and D denotes the bearing pitch diameter.

Under normal working conditions, the load torque of the motor is in balance with the output torque of the motor. As long as the fault appears, the resultant torque that acts on the mechanical system is $-M_c \cos(\omega_c t)$. The total moment is equal to the moment of inertia times the angular acceleration, and it can be expressed as follows:

$$-M_c \cos(\omega_c t) = J\alpha. \quad (4)$$

The rotor phase θ_r can be obtained by integrating t twice on both sides of the equation as follows:

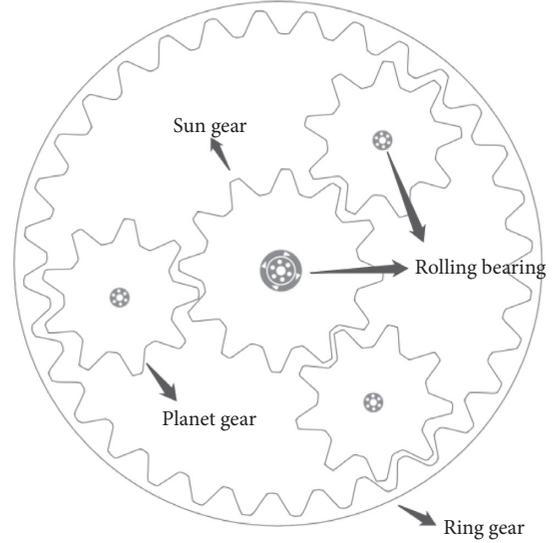


FIGURE 1: Schematic diagram of the planetary gearbox.

$$-\theta_r(t) = \int \omega_r(t) dt = \frac{M_c}{J\omega_c^2} \cos(\omega_c t) + \omega_{r0} t, \quad (5)$$

where ω_{r0} denotes the constant component of the rotor speed.

By considering the slip between the synchronous speed of the stator ω_s and the speed of the rotor ω_{r0} , using the transformation of the reference system [2], the magnetomotive force of the rotor can be obtained as follows:

$$F_r(\theta, t) = F_r \cos\left(\theta p - \omega_s t - \frac{M_c}{J\omega_c^2} \cos(\omega_c t) p\right). \quad (6)$$

The magnetomotive force of the stator is not affected by the fault. Hence, it can be directly obtained as follows:

$$F_s(\theta, t) = F_s \cos(\theta p - \omega_s t - \varphi_s), \quad (7)$$

where φ_s denotes the phase difference of initial magnetomotive force between the stator and rotor.

The airgap flux density $B(\theta, t)$ is determined by the sum of the magnetomotive forces of the stator and rotor and airgap permeance. According to the magnetic flux in the winding, the magnetic induction of electromotive force can be obtained via differentiation as follows:

$$\begin{aligned} E(t) &= -\omega_s \phi_s \sin(\omega_s t + \varphi_s) - \omega_s \phi_r \sin\left(\omega_s t + \frac{M_c}{J\omega_c^2} \cos(\omega_c t) p\right) \\ &\quad + \frac{M_c}{J\omega_c^2} \omega_c \phi_r \sin(\omega_c t) \sin\left(\omega_s t + \frac{M_c}{J\omega_c^2} \cos(\omega_c t) p\right). \end{aligned} \quad (8)$$

In practice, $(M_c/J\omega_c^2)$ is much less than 1. Hence, the last item can be ignored. Subsequently, the stator current can be expressed as follows:

$$I(t) = I_s \sin(\omega_s t + \varphi_s) + I_r \sin\left(\omega_s t + \frac{M_c}{J\omega_c^2} \cos(\omega_c t) p\right). \quad (9)$$

It is shown that the stator current is composed of two parts after the occurrence of fault. The first part is from the stator and the second part is from the oscillation of the rotor. Thus, the fault frequency can be expressed as follows:

$$f = f_s \pm f_c, \quad (10)$$

where f_s denotes the supply frequency of the stator, and its relation with ω_s is $\omega_s = (2\pi f_s/p)$, where p is the number of poles.

3. Preprocessing Technique: Singular Spectrum Analysis (SSA)

In the current signal, the supply frequency of the current is the dominant component, while the fault characteristic frequency components are usually weak. Hence, it is necessary to separate the different components of the current signal via preprocessing methods.

Singular spectrum analysis (SSA) is often used to deal with the nonlinear time series data [32]. By decomposing and reconstructing the time series, the sequences that represent different components can be extracted, which is helpful in separating weak fault feature information.

The first step of SSA for reconstruction is to transform time series T into vector series with a multidimensional delay. The method involves sliding a window with length L on T and intercepting the time series in the window to form the trace matrix X .

It is assumed that a time series T of length n is expressed as $T = [t_1, t_2, \dots, t_n]$. The elements on each inverse diagonal of $L * L$ part on the left of the trace matrix X are equal. Specifically, X is also termed as the Hankel matrix and can be expressed as follows:

$$X = [X_1, X_2, \dots, X_n] = \begin{bmatrix} t_1 & t_2 & \dots & t_K \\ t_2 & t_3 & \dots & t_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ t_L & t_{L+1} & \dots & t_n \end{bmatrix}, \quad (11)$$

where $K = n - L + 1$, $2 \leq L \leq (n/2)$.

The second step of SSA for reconstruction involves performing a singular value decomposition (SVD) on the trace matrix X :

$$X = U \sum V^T = \sum_{i=1}^r \sigma_i U_i V_i^T = \sum_{i=1}^r X^{(i)}, \quad (12)$$

where U denotes the unit orthogonal matrix of size $L * L$, U_i denotes the i th column vector; Σ denotes the diagonal matrix of size $L * K$, and the elements on the diagonal are the eigenvalues σ_i of the trace matrix X that are arranged from large to small; V^T denotes a unit orthogonal matrix of size $K * K$, V_i^T is its i th column vector; r denotes the rank of X , where $X^{(i)}$ is the trace matrix corresponding to the i th component of the signal.

Given that the elements on each inverse diagonal of $L * L$ part on the left side of $X^{(i)}$ are not equal, the diagonal average method can be used to change the elements on each

inverse diagonal of $X^{(i)}$ into the average value of the elements, and thus we obtain a new $\tilde{X}^{(i)}$. Furthermore, each element in $\tilde{X}^{(i)}$ can be calculated as follows:

$$\tilde{x}_{m,n} = \begin{cases} \frac{1}{s-1} \sum_{l=1}^s x_{l,s-l}, & 2 \leq s \leq L+1, \\ \frac{1}{L} \sum_{l=1}^{L+1} x_{l,s-l}, & L+2 \leq s \leq K+1, \\ \frac{1}{L+K-s+1} \sum_{l=s-K}^L x_{l,s-l}, & K+2 \leq s \leq K+L, \end{cases} \quad (13)$$

where $s = m + n$.

After the trace matrix of each component is obtained, different types of signal $\tilde{T}^{(i)}$ can be reconstructed as follows:

$$T = \sum_{i=1}^r \tilde{T}^{(i)}, \quad (14)$$

where $\tilde{T}^{(i)}$ can represent the trend, period, noise, or any other compositions of the original signal.

4. Feature Extraction and Classification Model

4.1. Principal Component Analysis (PCA). Principal component analysis (PCA) is a classical dimension reduction method in data mining. When a group of data has more than one variable, the problem becomes complicated. Typically, there is a certain correlation between these variables, and the information reflected by the two (or more) variables may be redundant. In this case, PCA uses one variable to reflect multiple variables with high similarity and thereby reduces the number of variables.

In signal processing, it is considered that the principal component has a larger variance and the noise exhibits a smaller variance. A set of zero-mean signal samples X can be expressed as $X = [X_1, X_2, \dots, X_n]$. The length of each sample X_i is the characteristic dimension m , which is equivalent to the length of the signal. To project the m -dimensional variables to the new p -dimensional space, a group of linearly independent basis in the original m -dimensional space should be constructed to ensure that the correlation between different dimensions is zero and the variance of each sample is maximum. This group of basis can be expressed as $U = [U_1, U_2, \dots, U_m]$. The projection variance maximization can be realized by the following constrained optimization problem:

$$\max \frac{1}{n} \sum_{i=1}^n (U_i^T X_i)^2 = \frac{1}{n} U_i^T X_i X_i^T U_i, \quad (15)$$

where $X_i X_i^T$ denotes the actual covariance matrix V :

$$V = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{bmatrix} = \frac{1}{n} XX^T. \quad (16)$$

Furthermore, $\sigma_{ij} = \text{Cov}(X_i, X_j)$, $i, j = 1, 2, \dots, m$.

With the Lagrange multiplier method, the optimization problem can be transformed into matrix eigenvalue decomposition problem. The eigenvalues of the covariance matrix V from large to small can be listed as $\{\lambda_i\}$, $i = 1, 2, \dots, m$. By preserving the first p orders of eigenvectors, the eigenmatrix is constructed as $\tilde{U} = [U_1, U_2, \dots, U_p]$.

Given that the covariance matrix is a real symmetric matrix, the unit eigenvectors are orthogonal to each other. Multiplying this eigenmatrix \tilde{U} with X , we can get the projection of these sample features in p -dimension. The percentage of the components on a different basis of vector space in the overall information can be calculated by the following equation. Hence, during the experiment, it can be judged whether the extracted features can represent the original data:

$$\sqrt{\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^m \lambda_i}}. \quad (17)$$

4.2. Classical Machine Learning Model. Four classical machine learning models, including support vector machine (SVM), decision tree (DT), random forest (RF), and AdaBoost, are adopted in this study. They are realized with Scikit-learn, an open-source machine learning library. The structure and principle of these models are briefly introduced as follows:

The SVM has been widely used in classification and regression problems due to its good robustness and strong generalization ability for unknown data. The core idea of SVM is to find a hyperplane, which can divide the samples linearly in space based on the training set samples. The samples that are closest to the hyperplane are called support vectors. The optimization direction of the model involves maximizing the distance between the support vector and hyperplane. When it is difficult to divide the low dimensional space linearly, SVM can use the kernel function for mapping the data to the high dimensional space to find the hyperplane. The main parameters of SVM in programming are kernel function, kernel coefficient gamma, and regularization parameter C. Usually, the kernel function is Gauss kernel function. As for gamma, if it is too large, the variance of Gaussian distribution is too small. Hence, the model can only operate on the support vector sample, which may lead to overfitting. Conversely, if gamma is too small, then the Gaussian distribution will be too smooth, resulting in the underfitting of training results. As for C, if it is too large, then the model increases the punishment in case of classification error, which leads to an overfit model.

The structure of DT is similar to that of a tree. Each internal node represents a judgment on the attribute, each branch represents an output of the judgment result, and, finally, each leaf node represents a classification result. The priority among different attributes is determined by the contribution of entropy. The main parameters of the decision tree are the criteria function, maximum depth, minimum number of samples for node subdivision, and minimum number of samples that are required at a leaf node. Furthermore, criteria function is used for calculating the entropy. Other parameters are used to determine the structure of the tree.

RF and AdaBoost are ensemble models that are the integration of traditional machine learning algorithms. Their core idea involves combining multiple weak classifiers to obtain a strong classifier. When compared with a single classifier, it reduces the possibility of overfitting the model. In this study, the base classifier for both types of ensemble models is the decision tree.

RF establishes multiple weak classifiers that are independent and equal. When a new sample is input, the voting method is used to select its category. Each classifier is trained to obtain the combined model.

AdaBoost uses an adaptive method to iteratively learn each weak classifier when a new sample is input. Each iteration increases the sample weight of the last error classification, reduces the sample weight of the correct classification, combines multiple classifiers into a strong classifier linearly, and provides a larger weight to the weak classifier with a lower error rate.

4.3. One-Dimensional Convolutional Neural Network (CNN). The neural network differs from the classical machine learning model because it represents the learning method that can automatically extract and select features from the data.

For a given sample x and its corresponding label y , the neural network can use the error backpropagation algorithm to continuously approximate the real mapping function between them. Convolutional neural network is a type of feedforward neural network, which consists of convolutional layers and fully connected layers. An example of a one-dimensional convolutional neural network is shown in Figure 2.

In the convolutional layer, features are extracted from the input data:

$$y = f(x * k + b), \quad (18)$$

where x denotes the input data, $(*)$ denotes convolution operation, k denotes the kernel, b denotes bias, and f denotes the activation function.

The convolution kernel performs the convolution operation by window sliding on the input signal. Each convolutional layer can use multiple convolution kernels to extract different features from the current input. The purpose of the activation function is to introduce nonlinearity into the output of a neuron. If the activation function is not used, then all the neurons are linear combined. The

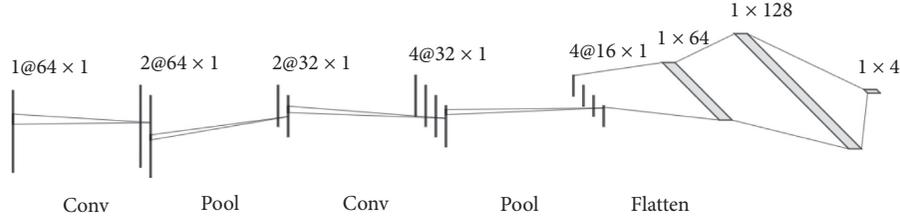


FIGURE 2: One-dimensional convolutional neural network.

commonly used activation functions are rectified linear unit (ReLU) and sigmoid:

$$\begin{aligned} \text{ReLU}(x) &= \max(0, x), \\ \text{sigmoid}(x) &= \frac{1}{1 + e^{-x}}. \end{aligned} \quad (19)$$

Subsequently, the feature obtained via convolution enters the pooling layer. The pooling operation is also realized by sliding a certain length of window on feature maps. When the window slides, the average or maximum value in each window is calculated to form a new feature map. The essence of pooling is actually a resampling operation, which reduces the amount of data.

Suppose that a signal sample input into the rollout layer is $x^{(0)} \in R^{1 \times m}$. By using n convolution kernels of size $1 \times k^{(0)}$, it can obtain n feature maps. After pooling, the input of the next layer becomes $x^{(1)} \in R^{n \times m'}$. Meanwhile, the size of each convolution kernel in the next layer should be $n \times k^{(1)}$. Normally, $k^{(i)}$ in each layer is selected from the commonly used values in CNN research and n is determined by the complexity of the classification task, which is also known as the number of channels.

By alternately stacking multiple convolution operations and pooling layers, the overall structure of convolutional layers can be obtained. The output feature map of the last convolutional layer is directly flattened into a one-dimensional vector and input into the fully connected layer. For the fully connected layer, the operation process of each layer can be expressed as follows:

$$\begin{aligned} z^{(l)} &= W^{(l)} \cdot x^{(l-1)} + b^{(l)}, \\ x^{(l)} &= f_l(z^{(l)}), \end{aligned} \quad (20)$$

where $W^{(l)}$ denotes the weight matrix of layer l , $b^{(l)}$ denotes the bias of layer l , $x^{(l-1)}$ denotes the input of layer l , $x^{(l)}$ denotes the output of layer l , which is also the input of the next layer, and f_l is the activation function of layer l .

The length of the last output vector from the fully connected layer is equal to the number of categories N . Then, softmax is used to convert each element of the output vector to a value on an interval (0, 1), and the sum is 1. Therefore, the conversion result can be regarded as the probability of each category. Finally, the result is the category that corresponds to the maximum probability. The softmax equation is as follows:

$$p_i = \frac{\exp(x_i)}{\sum_{j=1}^N \exp(x_j)}. \quad (21)$$

With respect to the training process, the cross-entropy loss function is used to measure the error between the predicted and real fault type because this is a classification problem. The cross-entropy loss function is as follows:

$$\text{CrossEntropy_Loss} = -\sum_{t=0}^T Y_t \times \log(\hat{Y}_t), \quad (22)$$

where Y_t denotes the real fault category label and \hat{Y}_t denotes the fault category label predicted by the network.

5. Experiment

5.1. Test Bench and Dataset. The test rig, as shown in Figure 3, consists of an induction motor (rated power: 3 kw, rated speed: 1480 rpm), a 2K-H planetary gearbox, and a magnetic power break. A torque/speed transducer is placed between the motor and gearbox via mechanical couplings. The torque load is applied via the magnetic power break. The planetary gearbox has three planet gears, and the tooth numbers are 16 (sun gear), 63 (planet gear), and 143 (ring gear). The input rolling bearing, which is used to support the rotation of sun gear, is a deep-groove rolling bearing (6208). The type of planet bearing is a cylindrical roller bearing (NJ204).

Seven types of faults have been seeded in the test gearbox. They are chipping teeth in sun/planet/ring gears (SG, PG, RG), inner race spall in planet bearing (PG-IR), inner/outer races, and ball spalls in input support bearing (IR, OR, B). The pictures of fault components are shown in Figure 3. For comparison, there is a control group with no fault.

A current sensor was utilized to acquire the phase A current signal of the motor. There are five types of torque loads in the system that run under each type of fault, and the experiments are conducted four times under each operating condition (see Table 1). Each experiment collects 30 s data, and the sampling frequency is 25.6 kHz.

In this study, 20 s data from two files under torque loads of 0 N·m, 5.25 N·m, and 10.5 N·m are selected to build the model. They are split by a window of size 0.1 s. Thus, a total of 9600 samples are acquired. The 9600 samples were divided into a training set and test set 1 by 70% and 30%. Specifically, 3200 samples (10 s data from two files under a torque load of

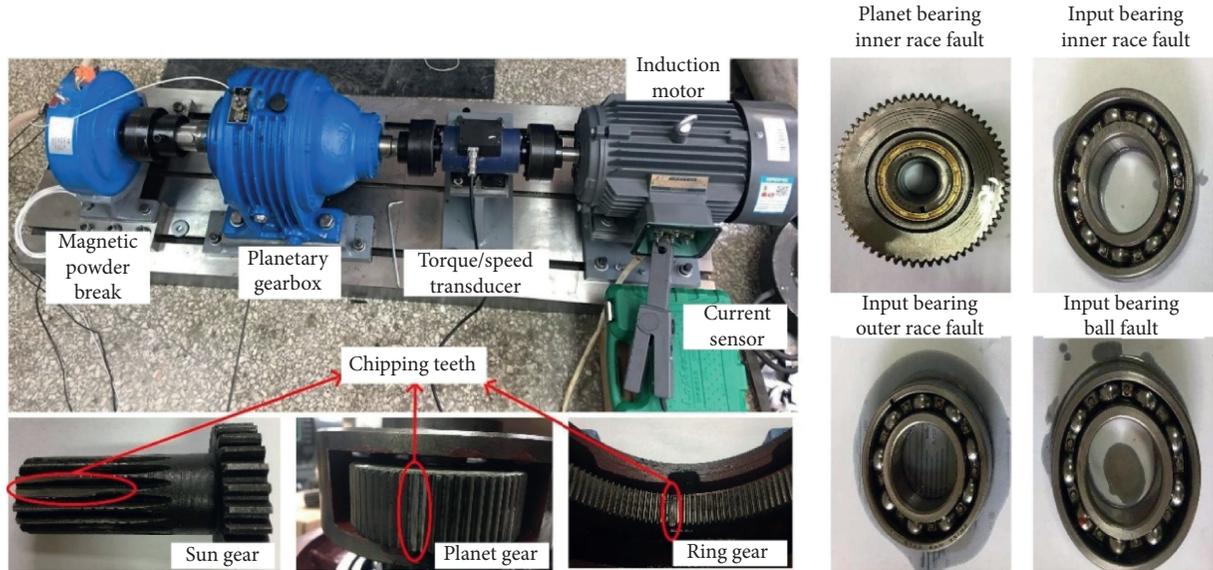


FIGURE 3: Test bench and seven different types of faults.

TABLE 1: Five operating conditions.

Torque load (N·m)	0	2.6	5.25	7.8	10.5
File number	101, 102, 103, 104	201, 202, 203, 204	301, 302, 303, 304	401, 402, 403, 404	501, 502, 503, 504

2.6 N·m and 7.8 N·m) as unlearned condition were used to evaluate the performance of the model to inspect its robustness. This set of 3200 samples is termed as test set 2.

5.2. Signal Analysis. The current signals of seven different types of fault and the normal condition are analyzed. As shown in Figure 4, the supply frequency at 50 Hz and its frequency multiplication are the most evident features. No other fault characteristic frequency is observed in the amplitude spectrum. There is no clear visual difference between eight amplitude spectrums.

From others' research [33], there are sidebands or many other characteristic frequencies in the amplitude spectrum, which may not be observed. To eliminate the strong effect of the supply frequency and expose the signal that may contain the fault information (besides the SSA we proposed), wavelet packet decomposition (WPD), ensemble empirical mode decomposition (EEMD), and independent component analysis (ICA) are considered for removing the power frequency. However, EEMD is not suitable for massive signal processing, and the experiment result indicates that it cannot isolate the complete supply frequency component for certain samples. As for ICA, if it is used for dimension reduction; then its function is the same as PCA in this case. If it is used for blind signal separation, then there should be multiple sensors that can acquire multiple signals. Therefore, these two methods are not adopted in this study.

After processing the current signal with SSA, by considering the inner race fault bearing as an example (as shown in Figure 5), the reconstructed signal by SSA can

be divided into two categories, namely, the signal with the same periodicity as the supply frequency and residual signal.

Therefore, the periodic components of the original signal can be removed, and the signal can be reconstructed with the residual signal. The periodic component is compared with the residual and original signals, as shown in Figure 6.

Besides SSA, an approach based on WPD is also considered for comparison with SSA. After removing the nodes, which represent frequencies near the supply frequency in the binary decomposition tree, signals can be reconstructed.

Figure 7 presents the amplitude spectra of the inner race fault signal in Figure 4 after the usage of SSA and WPD. Evidently, the SSA exhibits a better ability to eliminate power supply frequency components.

5.3. Experimental Results. Before the experiment, signals are labeled according to the types of fault. Given the fact that traditional machine learning models require feature engineering, PCA functions as a feature extractor. Their combination, PCA-SVM, PCA-DT, PCA-RF, and PCA-AdaBoost, with CNN is used in this classification task. Signal processing methods can be divided into four groups according to whether they are preprocessed by SSA and whether they are transformed into the frequency domain via fast Fourier transform (FFT). Given the fact that SSA involves large-scale matrix operation and resampling of the original data can lead to information loss, the window size used should not be extremely large. In this case, it is 640.

For the experiment, we implement the Python-based programming. First, PCA compresses the raw signal to 20

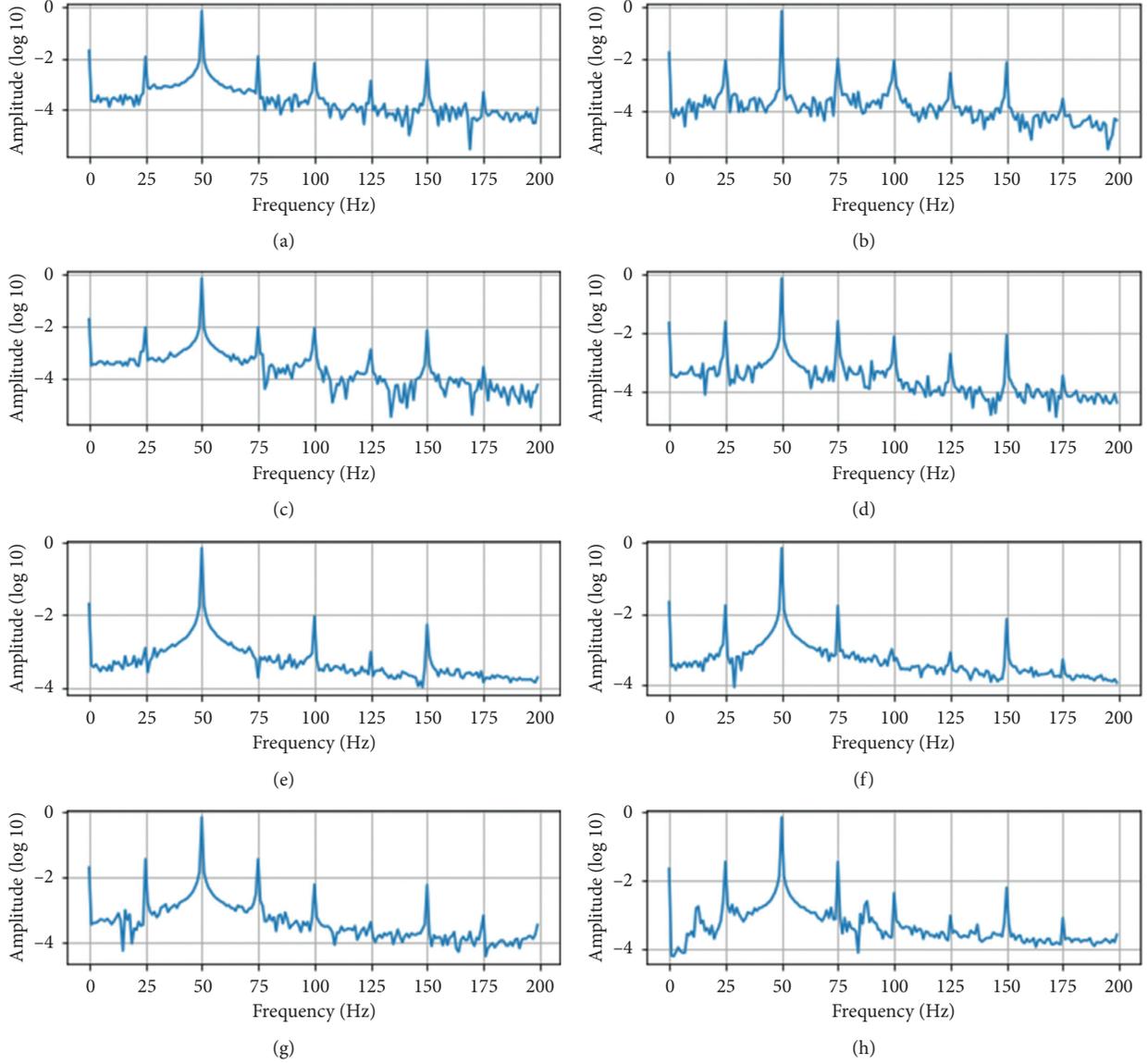


FIGURE 4: Amplitude spectrum analysis of different types of faults. (a) Bearing inner race fault, (b) bearing outer race fault, (c) bearing ball fault, (d) planet gear fault, (e) planet bearing inner race fault, (f) sun gear fault. (g) ring gear fault, and (h) no fault.

dimensions in advance. By using equation (17) to calculate the proportion of information in the 20 dimensions, the result exceeds 95%. Thus, these 20-dimension data represent the original data. With respect to SVM, the commonly used gamma and C are obtained by selecting the best combination according to the result. In this study, the parameter combination that enhances the result is as follows: gamma is the default and C takes 100. For DT, Gini impurity is selected as the criteria function, the maximum depth is set to 30, the minimum number of samples for node subdivision is 10, and the minimum number of samples required to be at a leaf node is set to 5. The parameters of the base classifier of RF and AdaBoost are the same as the DT above.

However, these machine learning models, which should extract features in advance, are not suitable for the original long sequence signal. Thus, CNN is adopted.

By considering the computational burden, the specific structure designed is shown in Table 2. The kernel size and number of channels are popular choices in CNN study. There are 4 convolutional layers and a fully connected layer. In the case of the CNN with time domain signal input, the adaptive moment estimation (Adam) optimizer is selected, and its learning rate can adapt to the training process. In the case of the CNN with frequency domain signal input, the stochastic gradient descent (SGD) optimizer is selected, and the learning rate α is set to 0.0015. Different optimizers are experimentally determined. They, respectively, reach higher training accuracies in different domains. It has been proven that Adam and SGD exhibit different performances in different tasks. Adam usually converges faster than SGD. However, it may not converge in some cases and can miss

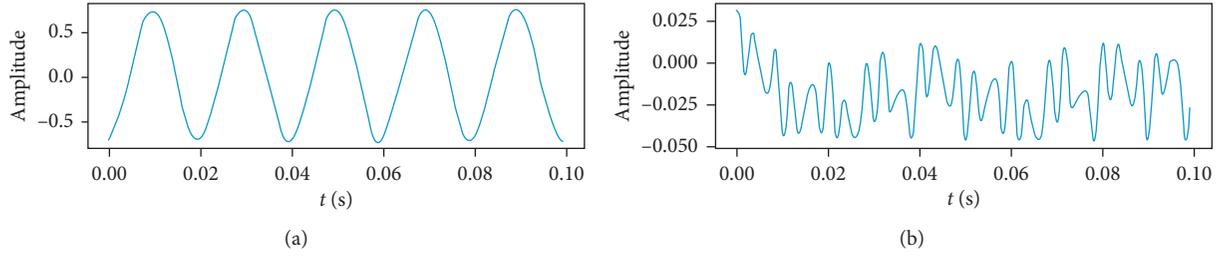


FIGURE 5: Different reconstructed signals from the components. (a) Supply frequency signal and (b) residual signal.

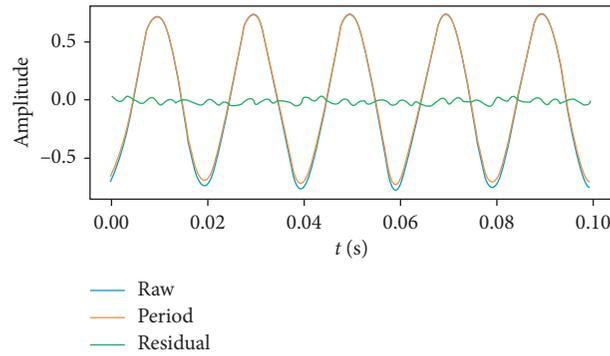


FIGURE 6: Comparison of the periodic signal with the residual signal and the original signal.

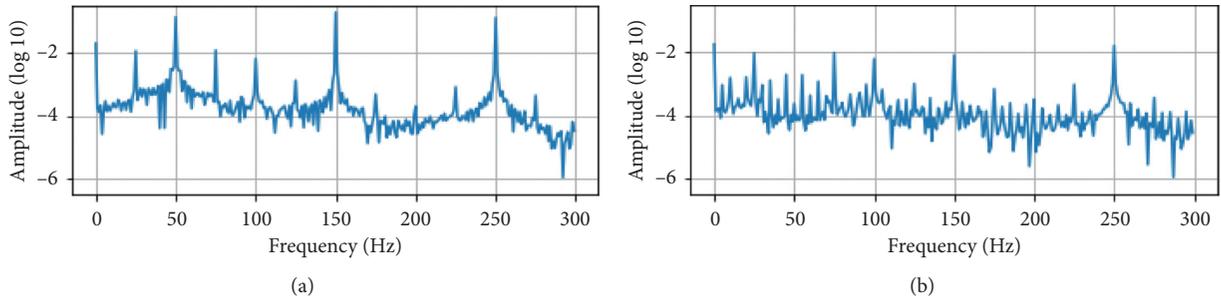


FIGURE 7: Comparison of amplitude spectrum after the usage of WPD and SSA. (a) WPD result. (b) SSA result.

the optimum solution [34]. In another study [28], a similar strategy using different optimizers is also adopted.

After the training, the current signals of test set 1 and test set 2 are inputted into each model to obtain the classification results. The accuracy of test set 1 is shown in Table 3.

The comparison of different models reveals that PCA-SVM exhibits the best performance in the time domain, whether to use preprocessing or not. It is also the best model in the frequency domain for reconstructed signals. It exhibits a maximum accuracy of 99.83% and an average accuracy of 93.92%. Furthermore, PCA-AdaBoost exhibits the second-highest average accuracy of 92.57%. It is the best model for signals, which are not preprocessed in the frequency domain. The average accuracy of CNN is slightly inferior to that of PCA-AdaBoost. These three models are very stable. PCA-RF is ranked after these three models. The worst model

among them is PCA-DT. Its maximum accuracy is 94.13%. However, its average accuracy is lower than 80%.

The preprocessing method, SSA, seems to have a weak impact on the result of test set 1. It only improves the performance of CNN in the time domain and that of PCA-SVM slightly in the frequency domain. The accuracy of PCA-DT and that of two ensemble models in two domains decrease, which suggests that components of the reconstructed signal may contain certain misleading features learned by the three models. The performance of WPD is even worse than that of SSA, which suggests that SSA may be a better signal reconstruction solution.

After the signals are transformed into the frequency domain via FFT, the average classification accuracy increases. As for CNN and PCA-SVM, their performance in the time domain becomes extremely good. Thus, accuracy is easily affected, resulting in no obvious improvements or a

TABLE 2: CNN structure.

Input signal data (2560 × 1)	
Convolutional layer 1	Input channel: 1, output channel: 4, kernel size: 3 ReLU Max pooling, filter size: 5 Output size (512 × 4)
Convolutional layer 2	Input channel: 4, output channel: 8, kernel size: 3 ReLU Max pooling, filter size: 4 Output size (128 × 8)
Convolutional layer 3	Input channel: 8, output channel: 16, kernel size: 3 ReLU Max pooling, filter size: 4 Output size (32 × 16)
Convolutional layer 4	Input channel: 16, output channel: 32, kernel size: 3 ReLU Max pooling, filter size: 4 Output size (8 × 32)
Flatten	Feature size (256 × 1)
Fully connected layer	Input size: 256, output size: 2560 ReLU Input size: 2560, output size: 8
Final output (classified by softmax)	

TABLE 3: Fault diagnosis accuracy of test set 1.

Methods	Time domain			Frequency domain			Average (%)
	None (%)	SSA (%)	WPD (%)	None (%)	SSA (%)	WPD (%)	
CNN	96.88	99.13	77.78	98.37	98.16	78.12	91.41
PCA-SVM	99.83	99.20	83.33	99.20	99.34	82.64	93.92
PCA-DT	80.73	67.08	53.12	94.13	89.24	64.58	74.81
PCA-RF	95.62	91.49	69.79	98.75	97.47	77.78	88.48
PCA-AdaBoost	99.41	96.94	73.61	99.76	99.24	86.46	92.57
Average	94.49	90.77	71.53	98.04	96.69	77.92	

slight decrease in accuracy. For the other three models, with data not preprocessed, the score of PCA-DT increase by 13.40%, that of PCA-RF by 3.13%, and that of PCA-AdaBoost by 0.35%. As for the reconstructed signal of SSA, the accuracy of PCA-DT increases by 22.16%, that of PCA-RF by 5.98%, and that of PCA-AdaBoost by 2.30%. Furthermore, WPD also shows the same improving trend.

In general, PCA-SVM is the best model in the test set 1. The deep learning model may not be better than the classical machine learning model. SSA mainly works on CNN in the time domain but does not significantly improve the performance of other models. Thus, the preprocessing method should be designed carefully in case of adverse effects. Transforming the time domain data into the frequency domain can improve diagnosis accuracy.

Figure 8 presents the confusion matrix of the raw signals classification result by CNN. It is shown that the main error of the classification is actually concentrated on the bearing. The accuracy of identifying ball spalls in input supporting bearing is the lowest. It is mostly confused with the outer race fault followed by the sun gear fault. Inner race fault is likely to be identified as outer race fault or ball fault. Outer

race fault and sun gear fault are easily confused with the planet gear fault. Other types of faults are recognized with high accuracy. These confusions are probably because the rolling bearing supports the sun gear, and the distance between them is small, which leads to similar signal transmission paths.

For test set 2, the accuracy of different methods is shown in Table 4. In the time domain, CNN is the best model for both preprocessed and raw signals. As for the frequency domain, PCA-AdaBoost and PCA-SVM exhibit the best performance, which is the same as that for test set 1. On average, CNN exhibits the highest accuracy. Models ranked behind CNN are PCA-SVM, PCA-AdaBoost, PCA-RF, and PCA-DT. The results are similar to those of test set 1.

With respect to test set 2, the reconstruction of the original current signal by SSA plays a positive role, which is different from that for test set 1. The accuracy of all the models in the time domain increased significantly after SSA, especially that of CNN, which suggests that SSA increases the robustness in the time domain. Furthermore, it still outperforms WPD. In test set 2, the component that contains weak fault characteristics may be exposed via the

		Predicted fault							
		IR	OR	B	PG	IR(PG)	SG	RG	N
Real fault	IR	92.2%	2.2%	2.2%	0	0.8%	2.0%	0	0.6%
	OR	0	96.4%	0	2.2%	0.6%	0.3%	0	0.6%
	B	0.8%	7.1%	89.5%	0	0	2.0%	0	0.6%
	PG	0	0	0	100.0%	0	0	0	0
	IR(PG)	0	0	0	0	100.0%	0	0	0
	SG	0	0.8%	0	2.0%	0	96.5%	0	0.6%
	RG	0	0	0	0	0	0	100%	0
	N	0	0.3%	0	0	0.3%	0	0	99.4%

FIGURE 8: Confusion matrix of classification result with the original signal and CNN. IR represents bearing inner race fault; OR represents bearing outer race fault; B represents bearing ball fault; PG represents planet gear fault; IR (PG) represents that bearing in the gearbox has an inner race fault; SG represents sun gear fault; RG represents ring gear fault; and N represents no fault.

TABLE 4: Fault diagnosis accuracy of test set 2.

Methods	Time domain			Frequency domain			Average (%)
	None (%)	SSA (%)	WPD (%)	None (%)	SSA (%)	WPD (%)	
CNN	63.53	81.22	36.88	76.41	74.88	61.88	65.80
PCA-SVM	57.09	67.56	48.12	80.91	81.16	65.00	66.64
PCA-DT	32.09	44.50	28.75	71.03	69.09	45.62	48.51
PCA-RF	44.72	64.22	41.25	78.37	78.53	63.44	61.76
PCA-AdaBoost	47.53	67.53	45.94	81.34	79.91	60.31	63.76
Average	48.99	65.01	40.19	77.61	76.71	59.25	

decomposition method of SSA rather than WPD. However, the positive effect in the frequency domain is not significant.

The advantage of transforming to frequency domain is more evident, whether the preprocessing method is used or not. The performance in the frequency domain is better than that in the time domain. Furthermore, even the accuracy of SVM increases.

Our paper also uses vibration signals for comparison. As far as CNN is concerned, the accuracy of model classification for vibration signals is 99.0%. Although the best classification accuracy of the current signal in test set 1 exceeds that of the vibration signal, the training time consumed by the former is much longer than that of the latter. In the experiment, only five epochs were required, and the classification accuracy of the vibration signal can reach 99.0% while the current signal requires more than

100 epochs. Additionally, the model of vibration signal can easily attain an accuracy of 95% in test set 2, while the current signal can only attain a maximum accuracy of 81.34%. This proves that the model trained via vibration signal exhibits higher robustness. Therefore, the method based on the vibration signal is still better than the method based on the current signal.

6. Conclusions

In this paper, deep learning and different machine learning models are used for fault diagnosis of bearings and planetary gearboxes. The current signal of the motor is processed via different methods, and the results are compared.

From the perspective of models, PCA-SVM is the best model for learned data while CNN is the best model for

unlearned data in the time domain. In the frequency domain, PCA-AdaBoost and PCA-SVM exhibit the best performance in both test sets. PCA-DT is the least recommended model. From the perspective of preprocessing, SSA mainly increases the accuracy of CNN in the time domain and exhibits a positive effect on unlearned data in the time domain. However, in other cases, the accuracy decreases due to SSA. This suggests that an effective preprocessing method should be designed for the target model in the future. Finally, by transforming the time domain data to the frequency domain data, the accuracy increases significantly, which in turn shows that fault features are more likely to be exposed in the frequency domain.

These conclusions can aid in the selection of methods for future studies related to the fault diagnosis of planetary gearbox based on MCSA.

Data Availability

The motor current 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

The research work described in the paper was supported by the National Science Foundation of China under Grant no. 11872222 and the State Key Laboratory of Tribology under Grant no. SKLT2019B09.

References

- [1] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis-part II: signals and signal processing methods," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6546–6557, 2015.
- [2] M. Blodt, M. Chabert, J. Regnier, and J. Faucher, "Mechanical load fault detection in induction motors by stator current time-frequency analysis," *IEEE Transactions on Industry Applications*, vol. 42, no. 6, pp. 1454–1463, 2006.
- [3] M. Lebold, K. McClintic, R. Campbell, C. Byington, and K. Maynard, "Review of vibration analysis methods for gearbox diagnostics and prognostics," in *Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology*, pp. 623–634, Virginia Beach, VA, USA, May 2000.
- [4] R. R. Schoen, T. G. Habetler, F. Kamran, and R. G. Bartfield, "Motor bearing damage detection using stator current monitoring," *IEEE Transactions on Industry Applications*, vol. 31, no. 6, pp. 1274–1279, 1995.
- [5] J. Hou, W. Jiang, and W. Lu, "Application of a near-field acoustic holography-based diagnosis technique in gearbox fault diagnosis," *Journal of Vibration and Control*, vol. 19, no. 1, pp. 3–13, 2013.
- [6] X. Jin, F. Cheng, Y. Peng, W. Qiao, and L. Qu, "A comparative study on vibration- and current-based approaches for drivetrain gearbox fault diagnosis," in *Proceedings of the IEEE Industry Applications Society Annual Meeting*, pp. 1–8, Portland, OR, USA, October 2016.
- [7] M. S. Yilmaz and E. Ayaz, "Adaptive neuro-fuzzy inference system for bearing fault detection in induction motors using temperature, current, vibration data," in *Proceedings of the IEEE EUROCON*, pp. 1140–1145, St-Petersburg, Russia, May 2009.
- [8] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification," in *Proceedings of the European Conference of the Prognostics and Health Management Society*, Bilbao, Spain, July 2016.
- [9] Q. Han, Z. Ding, Z. Qin, T. Wang, X. Xu, and F. Chu, "A triboelectric rolling ball bearing with self-powering and self-sensing capabilities," *Nano Energy*, vol. 67, Article ID 104277, 2020.
- [10] Q. Han, Z. Ding, X. Xu, T. Wang, and F. Chu, "Stator current model for detecting rolling bearing faults in induction motors using magnetic equivalent circuits," *Mechanical Systems and Signal Processing*, vol. 131, no. 15, pp. 554–575, 2019.
- [11] A. R. Mohanty and C. Kar, "Fault detection in a multistage gearbox by demodulation of motor current waveform," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 4, pp. 1285–1297, 2006.
- [12] 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, no. 24, pp. 36–55, 2018.
- [13] Y. Lei, Z. He, J. Lin, D. Han, and D. Kong, "Research advances of fault diagnosis technique for planetary gearboxes," *Journal of Mechanical Engineering*, vol. 47, no. 19, pp. 59–67, 2011.
- [14] X. Chen and Z. Feng, "Time-frequency space vector modulus analysis of motor current for planetary gearbox fault diagnosis under variable speed conditions," *Mechanical Systems and Signal Processing*, vol. 121, pp. 636–654, 2019.
- [15] Q. Han, T. Wang, Z. Ding, X. Xu, and F. Chu, "Magnetic equivalent modeling of stator currents for localized fault detection of planetary gearboxes coupled to electric motors," *IEEE Transactions on Industrial Electronics*, 2020.
- [16] J. Zhang, J. S. Dhupia, and C. J. Gajanayake, "Stator current analysis from electrical machines using resonance residual technique to detect faults in planetary gearboxes," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 9, pp. 5709–5721, 2015.
- [17] C. Kar and A. R. Mohanty, "Monitoring gear vibrations through motor current signature analysis and wavelet transform," *Mechanical Systems and Signal Processing*, vol. 20, no. 1, pp. 158–187, 2006.
- [18] J. Wang, G. Zhou, X. Zhao, and S. Liu, "Gearbox fault diagnosis and prediction based on empirical mode decomposition scheme," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 1072–1075, Hong Kong, China, August 2007.
- [19] R. Yan and R. X. Gao, "Hilbert-Huang transform-based vibration signal analysis for machine health monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 6, pp. 2320–2329, 2006.
- [20] Y. Yongfeng, W. Qinyu, W. Yanlin, W. Qin, and K. Lu, "Dynamic characteristics of cracked uncertain hollow-shaft," *Mechanical Systems and Signal Processing*, vol. 124, no. 1, pp. 36–48, 2019.

- [21] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.
- [22] Y. Yoo and J.-G. Baek, "A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network," *Applied Sciences*, vol. 8, no. 7, p. 1102, 2018.
- [23] Y. Lei, Z. He, Y. Zi, and Q. Hu, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2280–2294, 2007.
- [24] Y. Peng, W. Qiao, L. Qu, and J. Wang, "Gearbox fault diagnosis using vibration and current information fusion," in *Proceedings of the IEEE Energy Conversion Congress and Exposition*, pp. 1–6, Milwaukee, WI, USA, September 2016.
- [25] F. Cheng, J. Wang, L. Qu, and W. Qiao, "Rotor-Current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier," *IEEE Transactions on Industry Applications*, vol. 54, no. 2, pp. 1062–1071, 2018.
- [26] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. 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.
- [27] G. Jiang, J. Zhao, C. Jia, Q. He, P. Xie, and Z. Meng, "Intelligent fault diagnosis of gearbox based on vibration and current signals: a multimodal deep learning approach," in *Proceedings of the Prognostics and System Health Management Conference*, pp. 1–6, Qingdao, China, October 2019.
- [28] D. T. Hoang and H. J. Kang, "A motor current signal-based bearing fault diagnosis using deep learning and information fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3325–3333, 2020.
- [29] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems and Signal Processing*, vol. 76-77, pp. 283–293, 2016.
- [30] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
- [31] Z. Feng and M. J. Zuo, "Vibration signal models for fault diagnosis of planetary gearboxes," *Journal of Sound and Vibration*, vol. 331, no. 22, pp. 4919–4939, 2012.
- [32] S. Krishnannair, "Fault detection of Tennessee eastman process using kernel dissimilarity scale based singular spectrum analysis," in *Proceedings of the 13th IFAC Workshop on Adaptive and Learning Control Systems*, pp. 204–209, Winchester, UK, December 2019.
- [33] E. Artigao, A. Honrubia-Escribano, and E. Gómez-Lázaro, "In-service wind turbine DFIG diagnosis using current signature analysis," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2262–2271, 2019.
- [34] S. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in *Proceedings of the International Conference on Learning Representations*, Vancouver, Canada, May 2018.