

Article

# Mechanical Fault Diagnosis of a DC Motor Utilizing United Variational Mode Decomposition, SampEn, and Random Forest-SPRINT Algorithm Classifiers

Zijian Guo <sup>1,2</sup>, Mingliang Liu <sup>1,2,\*</sup>, Huabin Qin <sup>1</sup> and Bing Li <sup>1</sup>

<sup>1</sup> HLJ Province Key Lab of Senior-Education for Electronic Engineering, Heilongjiang University, Harbin 150080, China; gzj940823@163.com (Z.G.); 13181683786@163.com (H.Q.); libing\_entropy@163.com (B.L.)

<sup>2</sup> Key Laboratory of Information Fusion Estimation and Detection, Heilongjiang Province, Harbin 150080, China

\* Correspondence: 2002039@hlju.edu.cn; Tel.: +86-0451-8660-4401

Received: 9 April 2019; Accepted: 30 April 2019; Published: 6 May 2019



**Abstract:** Traditional fault diagnosis methods of DC (direct current) motors require establishing accurate mathematical models, effective state and parameter estimations, and appropriate statistical decision-making methods. However, these preconditions considerably limit traditional motor fault diagnosis methods. To address this issue, a new mechanical fault diagnosis method was proposed. Firstly, the vibration signals of motors were collected by the designed acquisition system. Subsequently, variational mode decomposition (VMD) was adopted to decompose the signal into a series of intrinsic mode functions and extract the characteristics of the vibration signals based on sample entropy. Finally, a united random forest improvement based on a SPRINT algorithm was employed to identify vibration signals of rotating machinery, and each branch tree was trained by applying different bootstrap sample sets. As the results reveal, the proposed fault diagnosis method is featured with good generalization performance, as the recognition rate of samples is more than 90%. Compared with the traditional neural network, data-heavy parameter optimization processes are avoided in this method. Therefore, the VMD-SampEn-RF-based method proposed in this paper performs well in fault diagnosis of DC motors, providing new ideas for future fault diagnoses of rotating machinery.

**Keywords:** variational mode decomposition; sample entropy; random forest; SPRINT algorithm

## 1. Introduction

Reliable operation of DC motors is crucial to factories, which has a huge impact not only on personal safety but also on efficiency of operations [1]. Therefore, it is of great significance to study the fault diagnosis of motors. As vibration signals generated by mechanical components often provide various dynamic information about the status of the mechanical system, many studies on DC motor fault diagnosis have focused on mechanical vibration signals [2,3], which have achieved successful application to a certain extent in the past decades.

The most commonly employed method of fault diagnosis combines signal processing techniques with the vibration signals of rotating machinery to extract fault features. Vibration signals obtained from rotating machinery are periodic and continuous. Thus, time-frequency analysis is the optimal signal process [4–6]. In recent years, various methods, such as wavelet decomposition (WD), wavelet packet decomposition (WPD), empirical mode decomposition (EMD), and local mean decomposition (LMD), have been extensively applied in the field of motor fault diagnosis. However, wavelet decomposition and wavelet packet decomposition are nonadaptive signal analysis methods, which need to select the wavelet basis function in advance [7,8]. The eigenmode function of EMD will be submerged

in the background of hostile noise, which makes it impossible to extract physical characteristic parameters [9,10]. As an improved method, ensemble empirical mode decomposition (EEMD) has effectively solved these problems. To provide a unified reference framework in the time-frequency domain, EEMD adds white noise of finite (rather than infinitesimal) amplitude in the original vibration signals [11], but the central section of the algorithm is still the EMD method, which still has endpoint effects and inaccurate calculation problems. EWT (empirical wavelet transform) and LMD are two adaptive signal decomposition methods that are based on a recursive mode [12–14]. However, LMD and EWT algorithms are essentially adaptive signal decomposition methods based on a recursive mode, with endpoint effects and modal aliasing problems [15,16]. In recent years, another important method for diagnosis-based signals is based on principal component analysis (PCA) of the data. Santos-Ruiz et al. utilized dynamic PCA and an extended Kalman filter (EKF) to describe, detect, and quantify fluid leaks. The paper firstly proposed a data-driven system based on PCA for describing, detecting, and quantifying fluid leaks in the experimental pipelines [17]. Compared with the two approaches of steady-state estimation and EKF, Santos-Ruiz et al. proposed a method for leak detection and isolation (LDI) in pipelines based on data fusion [18]. The two methods ensure normal operation of the pipelines by on-line monitoring of tubing flow through modeling. However, the working environment of the motor is changeable, making it difficult to model accurately. The two methods both need a lot of data, but this experiment was simulated to monitor the early fault of the motor and could not produce a lot of fault data. Thus, the two methods mentioned above are not suitable for early fault diagnosis of the motor. Variational mode decomposition (VMD) is a novel and adaptive signal decomposition method proposed by Dragomiretskiy and Zosso in 2014 [19]. As a nonrecursive variational mode signal decomposition method, it can decompose complex multicomponent signals into different modes. Compared with former methods, it has obvious advantages in signal decomposition accuracy and noise robustness. Meanwhile, mode aliasing can also be solved [20].

When the operational state of rotating machinery changes, its dynamic characteristics tend to show nonlinearity and complexity, and its vibration signals also present nonlinearity and nonstationarity [21,22]. Traditional nonlinear dynamic theories [23], such as fractal dimension [24,25] and Lyapunov exponent [26], have been employed in the field of fault diagnosis for many years. However, these methods require a large amount of data in parameter estimation, while early fault data of motors are generally scarce. Entropy is used to describe the overall statistical characteristics of the signal, which is strongly generalizable to the overall characteristics of the signal. Jaynes proved in 1957 that thermodynamic entropy was identical with information-theory entropy of the probability distribution [27]. Based on the research, Basaran and Yan presented a damage metric based on the second law of thermodynamics and statistical mechanics [28]; verification of the damage model has been performed by extensive comparisons with laboratory test data of low cycle fatigue. In 2010, Naderi postulated that the thermodynamic entropy of metals, undergoing repeated cyclic loads and reaching the point of fracture, is a constant that is independent of geometry, load, and frequency. Its generation has applications in determining the fatigue life of components that undergo cyclic bending, torsion, and tension–compression [29]. Furthermore, Sosnovskiy made an attempt to propose a generalized theory of evolution based on the concept of tribo-fatigue entropy. Meanwhile, a mechanothermodynamic function was constructed for the specific case of fatigue damage of materials as a result of temperature variation from 3 K to 80% of the melting temperature based on the analysis of 136 experimental results [30]. According to the probability of new patterns generated in signal time series, approximate entropy (ApEn) was proposed [31] by Pincus, which measured the signal from the aspect of measuring the complexity of the time series. ApEn of the vibration signal describes the complexity and irregularity of the signal. The ApEn variation of motor equipment under different operating conditions can characterize the operation status of the machinery during this period. As an improved algorithm based on ApEn, sample entropy (SampEn) can lower the deviation introduced by approximation entropy into self-matching entropy. Therefore, this paper employed the sampEn algorithm to extract feature signals [32].

The random forest (RF) algorithm was proposed by Breiman in 2001 [33]. RF is composed of multiple decision trees [34], which is similar to the bagging algorithm [35]. Based on the bootstrap method [36], it resamples and generates multiple training sets, and the final results are determined by the voting principle of multiple decision trees. Different from conventional algorithms, it is a method that randomly selects and splits attribute sets. Compared with LQV (learning vector quantization) [37], ELM (extreme learning machine) [38], and the decision tree, it has the advantages of a fast training speed, powerful approximation ability, and high simulation accuracy. However, when there are unrecognizable faults, such a fault diagnosis method based on traditional machine learning is still unable to help engineers find where the fault is, which reduces the efficiency of fault diagnosis [39]. To solve the above problems, a new fault diagnosis method based on the RF algorithm is proposed in this paper. Through improving the bagging method of decision trees, this method employs a conditional probability index to split unbiased nodes of the decision tree, and it synthesizes the classification results of the decision tree according to weighted voting principles.

As indicated above, the main contribution of this paper is to propose a novel mechanical fault diagnosis method for DC motors based on VMD, SampEn, and RF, during which the optimal RF classifier is established through the SPRINT algorithm. Firstly, the signal acquisition system was designed in LabVIEW 2014, and then the acquired signals were de-noised by the translation invariant of the wavelet. After extracting the physically meaningful modes by the VMD algorithm, the signal characteristics are calculated by the SampEn method. Finally, the improved RF classifier is employed to identify the labeled samples.

## 2. Variational Mode Decomposition (VMD)

### 2.1. A Brief Description of VMD

VMD extracts different vibration signals from different vibration statuses by continuously updating eigenmode functions and utilizing an alternating direction multiplier method to constantly update various intrinsic mode functions and central frequencies [20,21]. Subsequently, it decomposes the self-adaptive quasi-orthogonal signal  $f$  into a discrete quasi-orthogonal band-limited signal  $N_k$  with certain sparsity, and each mode revolves around the center frequency  $\omega_t$ . The estimated  $K$  central angular frequencies  $\omega_t$  are obtained by setting the finite broadband parameter ( $a$ ) and the central angular frequency initialization method. Furthermore, the modal functions  $U_k$  are obtained according to different central angular frequencies. Each modal function is a single component AM-FM function [40]:

The band-limited intrinsic mode function (BIMF) is defined as:

$$U_k = A_k(t) \cos[\phi(t)]. \quad (1)$$

In the formula, the phase function  $\phi(t)$  is a non-monotonic decreasing function,  $\phi'(t) > 0$ . The envelope  $A_k(t) > 0$ , and the change rate of  $A_k(t)$  as well as the instantaneous angular frequency  $A_k'(t) > 0$  is low, which is much slower than the phase function  $\phi(t)$ .

According to Carson's rule, the BIMF prior estimate [41] is:

$$BIMF_{AM-FM} = 2(f_{AM} + f_{PM} + \Delta f). \quad (2)$$

where  $\Delta f$  is maximum deviation of homeopathic frequency,  $f_{FM}$  refers to offset rate of instantaneous frequency, and  $f_{PM}$  means the highest frequency of envelope  $A_k(t)$ . In order to estimate the frequency bandwidth of IMF (Intrinsic Mode Function) components, the following steps must be taken:

1. For each modal function  $U_k$ , the marginal spectrum is obtained by a Hilbert transformation.
2. The exponent-modified method is utilized to move the spectrum of modal functions to their estimated central frequencies.
3. The bandwidth of each modal function is obtained by Gauss smoothing.

The process of VMD can be considered as the construction and solution of a constrained variational problem, as described by Equation (3):

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left\{ \left[ \sigma(t) + \frac{j}{\pi t} \right] * u_k(t) \right\} e^{-j\omega_k t} \right\|_2^2 \right\}, \tag{3}$$

wherein  $u_k$  denotes the subsignals,  $\omega_k$  refers to the center frequencies of the submodes,  $\partial_t$  indicates the partial derivative of the function to find time  $t$ ,  $\sigma(t)$  means the unit pulse function,  $j$  indicates the imaginary unit, and  $*$  is the convolution. In order to render the problem unconstrained, a quadratic penalty term and Lagrangian multipliers are introduced, and the problem is renewed to Equation (4).

$$L(u_k, \omega_k, \lambda) = \alpha \sum_{k=1}^k \partial \left\| \left[ \left( \sigma(t) + \frac{j}{\pi t} \right) * u_k(t) e^{-j\omega_k t} \right] \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k \right\|_2^2 + \left( \lambda(t), f(t) \sum_{k=1}^K u_k(t) \right), \tag{4}$$

where  $\alpha$  is a penalty term (called equilibrium constraint parameter) and  $\lambda$  is a Lagrange multiplier. The saddle point is found by using the alternate direction method of multipliers (ADMM) [42].

The decomposition mode number  $K$  is determined artificially, and the frequency-domain expression of mode  $\hat{u}_k^1$ , the center frequency of each mode function  $\omega_k^1$ , and the largrangian multiplier are initialized. Meanwhile, the mode  $u_k$  and the center frequency  $\omega_k$  are updated separately by Equations (5) and (6):

$$\hat{u}_k^{n+1} = \frac{\hat{f} - \sum_{i < k} \hat{u}_i^{n+1} - \sum_{i > k} \hat{u}_i^n + \frac{\hat{\lambda}^n}{2}}{1 + 2\alpha(\omega - \omega_n^k)^2}; \tag{5}$$

$$\omega = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}|^2 d\omega}. \tag{6}$$

The modes  $u_k$  and the center frequencies  $\omega_k$  can be obtained through each updated equation. The largrangian multiplier is renewed by the following equation:

$$\hat{\lambda}^{n+1} = \hat{\lambda}^n + \tau \left( \hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right). \tag{7}$$

The iteration is updated until it converges to Equation (8).

$$\sum_k \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon \tag{8}$$

It can be concluded that this algorithm needs to set  $K$ ,  $\alpha$ ,  $\tau$ , and the tolerance convergence criterion in advance. Compared with the former two parameters,  $\tau$  and the tolerance convergence criterion barely have an impact on the decomposed signal. Hence, the default value was adopted in this paper. Prior experience was required to get  $K$ . If the parameter was set without ascertaining, it would be difficult to ensure the accuracy and efficiency of signal decomposition. Therefore, searching the best parameters is key to the method, which promotes the following research.

### 2.2. VMD Algorithm Simulation

The simulated signal was composed of three types of subsets, namely three different harmonics,  $x(t) = x_1 + x_2 + x_3$ , wherein  $x_1 = \cos(6 * \pi * t)$ ,  $x_2 = 1/9 \sin(52 * \pi * t)$ , and  $x_3 = 1/27 \cos(600 * \pi * t)$  [43]. The simulation signal  $x(t)$  is decomposed by VMD. The simulation signal and its decomposition diagram are shown in Figure 1.

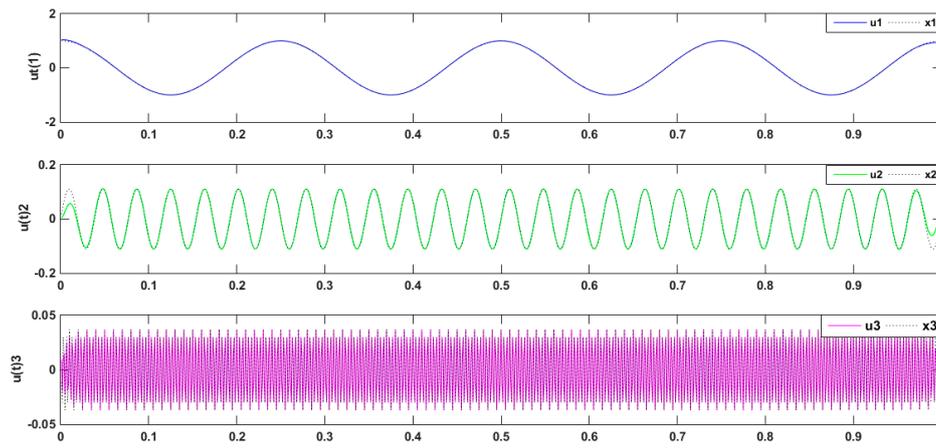


Figure 1. Comparison between simulated signal and modal components.

The three modal components (presupposition  $K = 3$ ) obtained from the decomposing simulation were signals  $u_1$ ,  $u_2$ , and  $u_3$ , which corresponded to the input signals  $x_1$ ,  $x_2$  and  $x_3$ , respectively. Both magnitude and frequency of the modal component and original signal matched well with each other, as shown Figure 1. Furthermore, the spectrum distribution of the input signal is shown in Figure 2, and the obtained spectrum distribution of each modal signal is shown in Figure 3. It can be observed that the three sums of modal frequency after decomposition can perfectly coincide with the frequency of the total decomposition signals in the frequency range. Compared with the original signal, results showed that the three decomposed modal frequencies can coincide with the original signals in the frequency range.

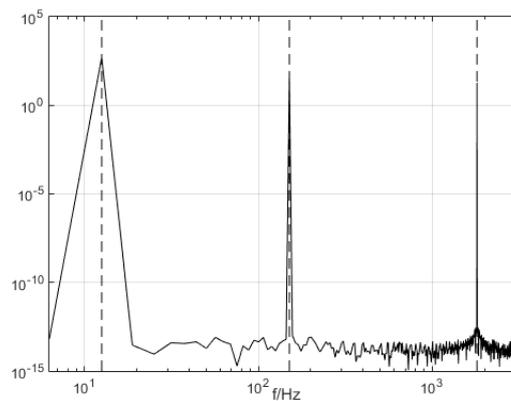


Figure 2. The frequency spectrum distribution of the input signal.

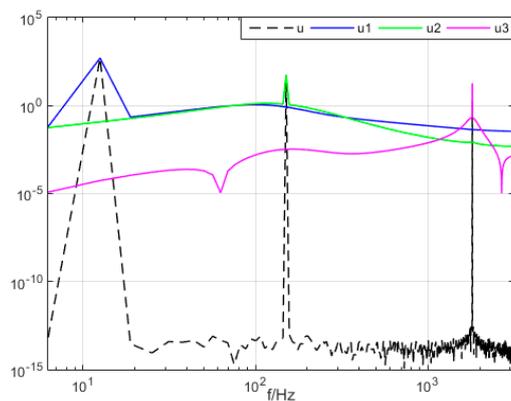


Figure 3. The mode signal spectrum with variational mode decomposition (VMD).

### 2.3. Study of the Parameter Decomposition Modes of the VMD

#### Simulation Experimental Results

Periodic signals were employed to verify the effect of the VMD modal value on signal decomposition. The mode number was set as  $K = 2, 3, 4,$  and  $5,$  and the penalty factor was set as  $a = 2000$ . As shown in Figure 4 below, when the mode number was set as  $K = 2,$  the decomposed signal frequencies of 4 and 26 Hz were superimposed, which caused modal mixing. Meanwhile, as shown in Figures 5 and 6 where  $u_1(t),$  and  $u_2(t)$  represent the cosines of frequencies 4 and 26 Hz, respectively, if the  $K$  value was higher, a new condition, called false mode, would appear. False mode  $u_3(t)$  was rendered when  $K = 4,$  while false modes of  $u_3(t), u_5(t)$  were rendered when  $K = 5.$  By tracing different  $K$  values, different central frequency curves can be obtained. As is shown in the Figure 7, when  $K = 3,$  the effects of the simulation signal will be better than the others, proving the right choice of parameters.

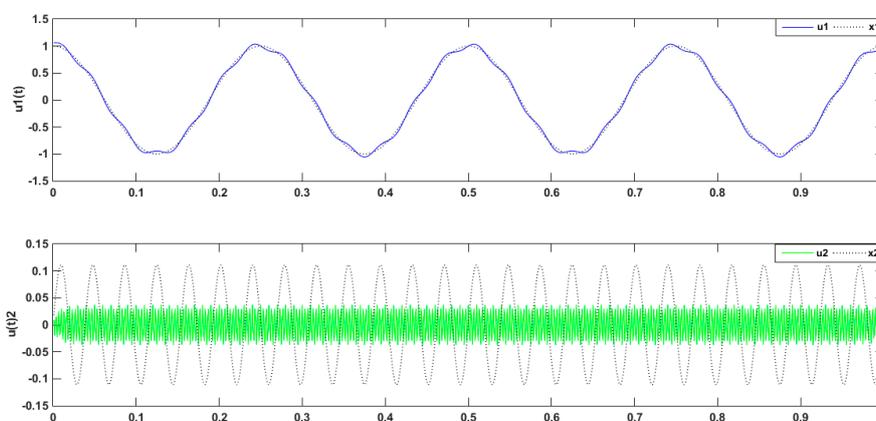


Figure 4. Decomposed signal by VMD ( $K = 2; a = 2000$ ).

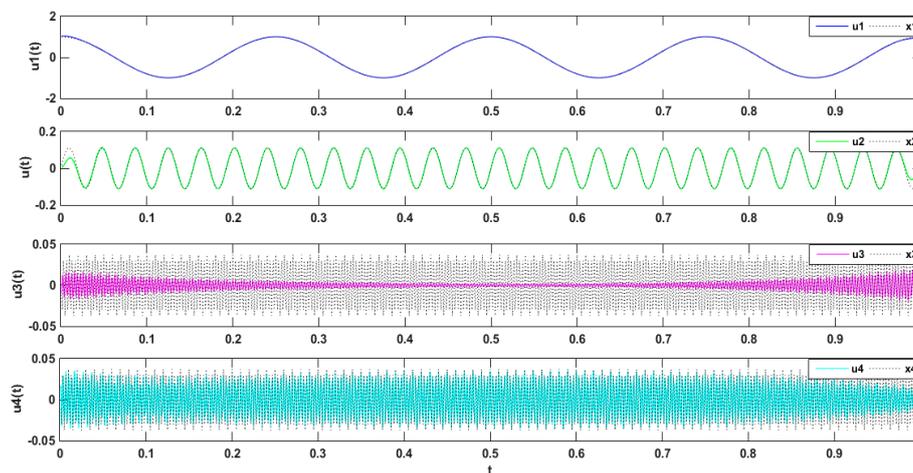


Figure 5. Decomposed signal by VMD ( $K = 4; a = 2000$ ).

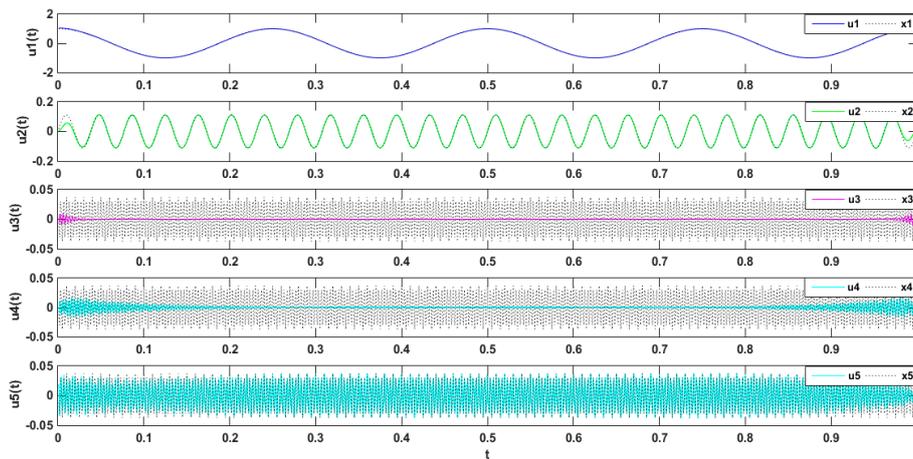


Figure 6. Decomposed signal by VMD ( $K = 5; a = 2000$ ).

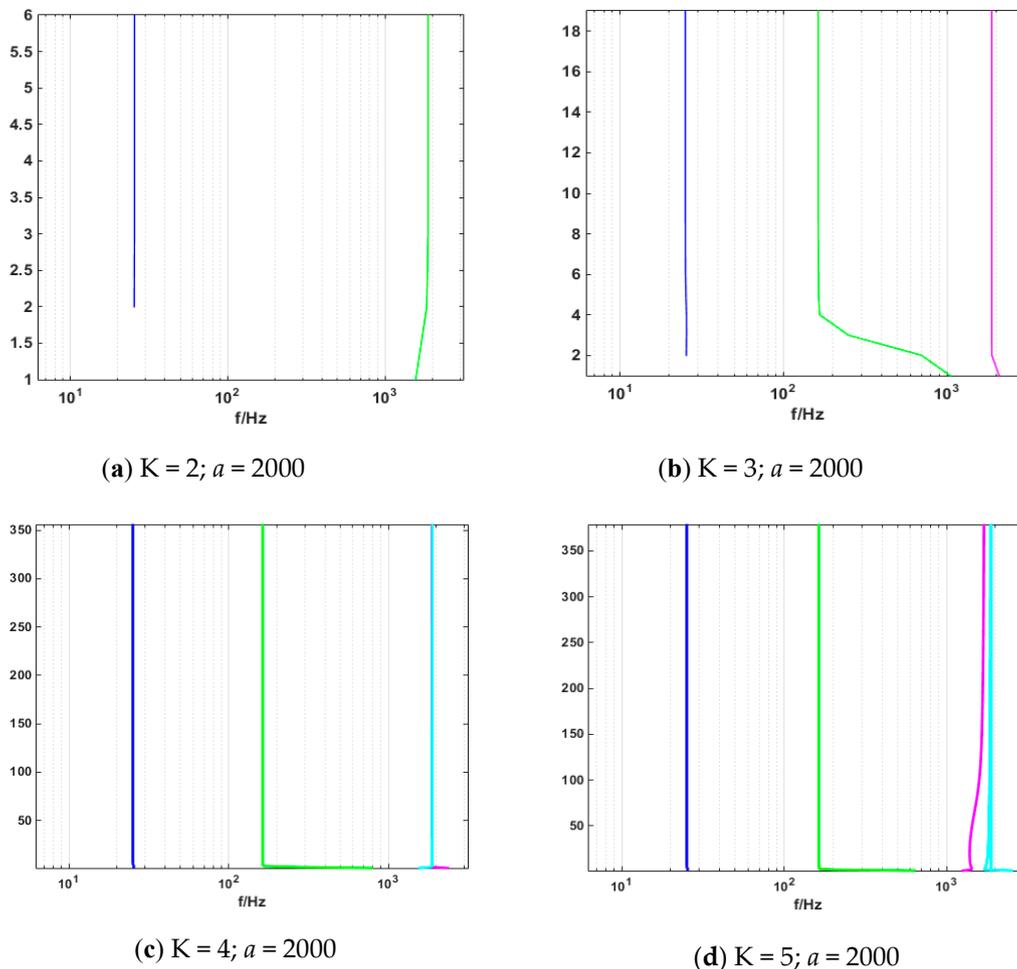


Figure 7. Change of center frequency.

The variation in central frequency of different  $K$  values is shown in Table 1. It can be observed that no sign of similar modal or modal mixing phenomenon appeared, when  $K = 3$ . But when  $K = 2$ , decomposition of periodic signals led to modal mixing. Moreover, false modes appeared in the process of decomposition as the value of  $K$  grew.

**Table 1.** Change of center frequencies in VMD under different mode values.

Mode Value	Center Frequency				
	U1	U2	U3	U4	U5
K = 2	25.76	1884.32			
K = 3	25.13	163.36	1884.96		
K = 4	25.13	163.36	1884.96	1884.96	
K = 5	25.13	163.36	1763.06	1874.27	1885.58

### 3. Sample Entropy

**For** a sequence of vectors of dimension  $m$ ,  $X_m(1)X_m(N - m - 1)$  **do**

1. The distance  $d[X_m(i), X_m(j)]$  between vector  $X_m(i)$  and  $X_m(j)$  is defined as the absolute value of the maximum difference between the two corresponding elements [44,45]

$$d[X_m(i), X_m(j)] = \max_{k=0, \dots, m-1} (|x(i+k) - x(j+k)|)$$

2. As for  $X_m(i)$ , the number of  $j$  ( $1 \leq j \leq N - m, j \neq i$ ) whose distance between  $X_m(i)$  and  $X_m(j)$  is less than or equal to  $r$  is counted and denoted as  $B_i$ . For  $1 \leq i \leq N - m$ , there is:

$$B_i^m = \frac{1}{N - m - 1} * B_i \tag{9}$$

3. Define  $B^{(m)}(r)$  as

$$B^{(m)}(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^{(m)}(r) \tag{10}$$

**For** the dimension raised to  $m + 1$ , the number of the distance between  $X_{m+1}(j)$  and  $X_{m+1}(i)$  ( $1 < j < N - m, j \neq i$ ) that is less than or equal to  $r$  is calculated and denoted as  $A_i$ .  $A_i^m(r)$  is defined as:

$$A_i^m(r) = \frac{1}{N - m - 1} A_i \tag{11}$$

Define  $A^m(r)$  as

$$A^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} A_i^m(r) \tag{12}$$

**End For**

$B^m(r)$  is the probability of two sequences to match  $m$  points under similar tolerance  $r$ , while  $A^m(r)$  is the probability of two sequences matching  $m + 1$  points [26].

**If**  $N \rightarrow \infty$ , sample entropy is defined

$$\text{SampEn}(m, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \left[ \frac{A^m(r)}{B^m(r)} \right] \right\} \tag{13}$$

**Else if**  $N$  is a finite value, it is seriated the following Equation:

$$\text{SampEn}(m, r, N) = -\ln \left[ \frac{A^m(r)}{B^m(r)} \right] \tag{14}$$

**End For**

It can be observed that the value of sample entropy was related to the values of  $m$  and  $r$ . Therefore, it was important to determine the values of the two parameters to calculate sample entropy.

Based on our research, when  $m = 1$  or  $2$ , or  $r = 0.25$  Std (Std is the standard deviation of initial data  $x(i), i = 1, 2 \rightarrow N$ ), the obtained sample entropy had the desired statistical properties. In this research  $m = 2$  and  $r = 0.2$  Std.

#### 4. Random Forest

##### 4.1. Bootstrap Resampling Algorithm

Bootstrap resampling is an integrated learning algorithm to sample the original data, which is an important component of the random forest algorithm. The algorithm requires extracting a sample from different samples  $\{x_1, x_2, \dots, x_n\}$  in set  $S$ . A new set  $S^*$  is formed, and the probability of excluding a sample  $x_i (i = 1, 2, \dots, n)$  in set  $S^*$  is [35]:

$$p = \left(1 - \frac{1}{n}\right)^n \tag{15}$$

When  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} p = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} \approx 0.368 \tag{16}$$

Although the total number of samples in the new set  $S^*$  is equal to the original set  $S$  (the total is  $n$ ), the new set  $S^*$  may contain duplicate samples. After removing the samples, it contains only about  $1 - 0.368 \times 100\% = 63.2\%$  samples in the original set  $S$ .

##### 4.2. Bagging Algorithm

A bootstrap algorithm is the earliest learning strategy, and the main idea is shown in Figure 8. The specific steps can be described as follows [36]:

1. Produce  $T$  training sets randomly through the bootstrap method of resampling;
2. Each training set is used to generate the corresponding decision tree;
3. As for the test set sample  $X$ , each decision tree is applied to test the corresponding categories;
4. Choose the category that has the most outcomes from the decision tree for testing sample  $X$  through voting.

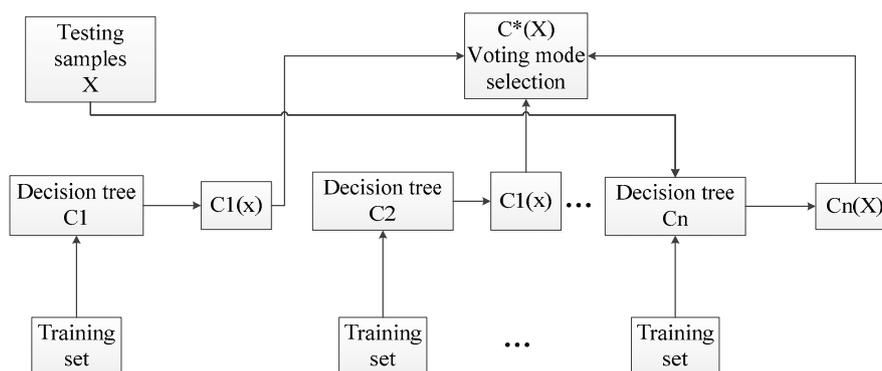
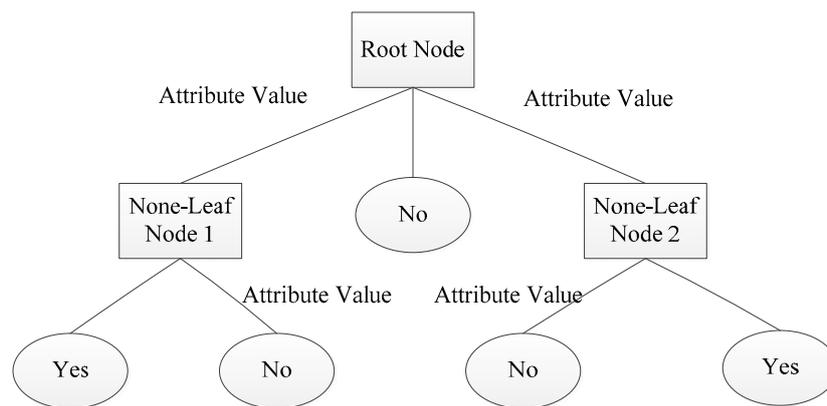


Figure 8. Bagging algorithm flowchart.

##### 4.3. Decision Tree

Decision tree learning is an inductive learning method based on specific examples. Generally, the recursive method of "from top to bottom, divide and rule" is employed to divide the search space into several nonoverlapping subsets. The schematic diagram of the decision tree is shown in Figure 9. Each non-leaf node represents the input attribute of the training set data, and leaf node represents the value of the target category attribute. "yes" and "no" represent the positive examples and counterexamples in the instance set [34].



**Figure 9.** Schematic diagram of a decision tree.

The purpose of decision tree learning is to generate trees with strong generalization abilities. There are many categories of decision trees, such as ID3, C4.5, and SPRINT. Although C4.5 overcomes the shortcomings of ID3 [46], it still has problems such as complex branches, large scales, and low efficiencies of the decision tree.

The SPRINT algorithm is a scalable, parallel, and inductive decision tree, which operates fast and allows multiple processors to create a collaborative decision tree model. Compared with the above two algorithms, it can better build a decision tree in parallel, making it suitable for larger data sets [33]. The SPRINT algorithm is based on the Gini index [47]:

$$\text{gini}(T) = 1 - \sum_{j=1}^n p_j^2, \quad (17)$$

where  $p_j$  is the frequency of occurrence of class  $j$ . If set  $T$  is divided into two parts,  $T_1$  and  $T_2$ , corresponding to  $M_1$  and  $M_2$  respectively, then the Gini index of this division is:

$$\text{gini}_{\text{division}} = \frac{m_1}{m} \text{gini}(T_1) + \frac{m_2}{m} \text{gini}(T_2). \quad (18)$$

The minimum Gini value is selected as the standard of segmentation. If it is a numerical range or continuous field, the possible segmented points are the midpoint of the two values. As for a discrete field, they are all subsets of attribute values.

#### 4.4. Random Forest Algorithm

The traditional machine learning model, like a neural network, is accurate in its prediction. However, the learning speed of a neural network is slow, and simple problems need to be trained hundreds of thousands of times, which will easily fall into local minimums. Compared with a neural network, random forest can effectively balance the error when there is imbalance in classification with high-speed network training. Based on the above statements, a RF approach based on SPRINT was proposed in this paper.

Similar to the bagging algorithm, this method was based on a bootstrap method to resample and generate multiple training sets. Unlike the bagging algorithm, RF chooses split attribute sets randomly when reconstructing a decision tree. The detailed random forest algorithm flow is shown in the Figure 10 (Assuming the number of attributes of the sample is  $M$ ,  $m$  is an integer greater than zero and less than  $M$ ):

1. Samples are collected by the bootstrap method, and training sets  $S_1, S_2, S_3 \rightarrow S_T$  will be generated randomly;
2. Each training set is employed to generate the corresponding decision tree  $C_1, C_2, C_3 \rightarrow C_T$ ;
3. Every tree grows intact without pruning;

4. As for the test set sample  $X$ , it gets the corresponding category by testing each decision number;
5. The category with the most output from  $T$  decision tree is taken as the category of test set sample  $X$  by voting.

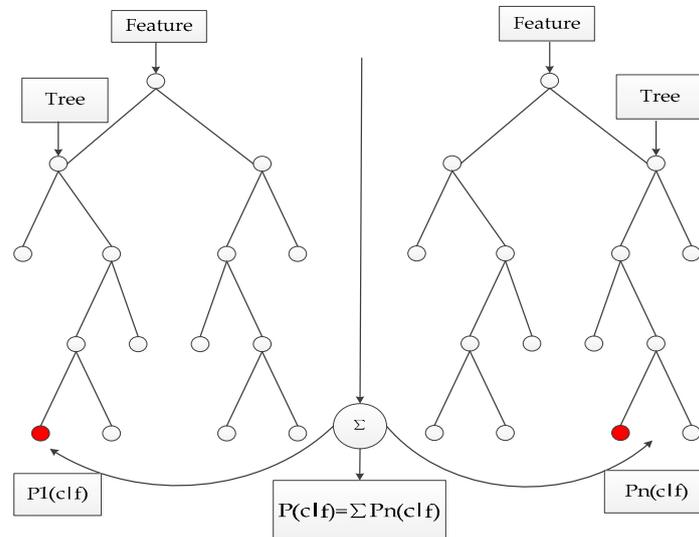


Figure 10. Schematic diagram of random forest.

4.5. Random Forest-United SPRINT Algorithm

Given the classifiers  $h_1(x), h_2(x), \dots, h_k(x)$  and random vector  $Y, X$ , the edge function is defined as:

$$mg(X, Y) = av_k I(h_k(x) = Y) - max_{j \neq Y} av_k I(h_k(x) = j). \tag{19}$$

In the formula,  $I(\ )$  is an indicative function,  $av_k(\rightarrow)$  is the average value, and, thus, the generalization error of the classifier is defined as:

$$PE^* = P_{x,y}(mg(x, y))$$

The above conclusions are extended to the random forest  $h_k(X) = H(X, \theta_k)$ . If the number of trees in the forest is large, the following theorems can be obtained by the law of large numbers and the structure of trees:

**Theorem 1.** With the increase of the number of trees, all random vectors  $\theta \rightarrow PE^*$  trends

$$P_{x,y}(p_\theta(h(x, \theta) = y) - max_{j \neq Y} p_\theta(h(x, \theta) = j) < 0). \tag{20}$$

It has been proven in literature [10] that over-fitting does not occur with the increase of numbers. In the meantime, the generalization error will converge to an upper bound.

Random forest edge function:

$$mv(X, Y) = P_\vartheta(h(X, \theta) = Y) - max_{j \neq Y} P_\vartheta(h(X, \theta) = j). \tag{21}$$

Classification strength of classifier  $\{h(X, \theta)\}$ :

$$S = E_{x,y}mr(x, y). \tag{22}$$

If  $S \geq 0$ , according to Chebyshev inequality:

$$PE^* = \text{var}(mr) / S^2. \quad (23)$$

$\text{Var}(mr)$  is required to have the following form:

$$\begin{cases} \text{var}(mr) = \bar{\rho}(E_{\theta}sd(\theta))^2 \\ \text{var}(mr) < \bar{\rho}(E_{\theta}\text{var}(\theta)) \end{cases}. \quad (24)$$

In the formula,  $\bar{\rho}$  is the mean value of the correlation coefficient, showing the correlation degree among trees in a random forest.

$$\bar{\rho} = E_{\theta, \theta'}(\rho(\theta, \theta')sd(\theta)sd(\theta')) / E_{\theta, \theta'}(sd(\theta)sd(\theta')), \quad (25)$$

and

$$\begin{cases} E_{\theta}\text{var}(\theta) \leq E_{\theta}(E_{x,y}mg(\theta, x, y))^2 - S^2 \\ E_{\theta}\text{Var}(\theta) \leq 1 - S^2 \end{cases}. \quad (26)$$

The following conclusions are received from Formulas (23), (25), and (26).

**Theorem 2.** *The upper bound of generalization error for random forest is defined as*

$$PE^* \leq \bar{\rho}(1 - S^2) / S^2. \quad (27)$$

$S$  is the classification strength of trees. The upper bound of the generalization error is determined by the classification accuracy of each decision tree and the degree of correlation between trees. It increases with the degree of correlation between trees, and it is also proportional to the classification intensity of each tree.

RF employs a classification and decision tree (SPRINT) to grow a single classification tree, which is different from the traditional SPRINT tree. The specific steps of random forest generation are as follows:

1. The random forest utilizes the bootstrap resampling method to extract 63.2% of the samples from the original training sample set to generate a subsample set, where each subsample corresponds to a classification tree. At the same time, the original samples that are not sampled are called out-of-bag data (OOB). OOB data are used to evaluate the classification accuracy of classifiers.
2. Each sub-sample set becomes a single classification tree. At each tree node,  $m$  feature vectors are picked out of  $M$  feature vectors, according to the empirical formula  $m = \text{int} \sqrt{M}$ . According to the principle of minimum nonpurity of nodes, feature  $\alpha$  is chosen as the classification attribute of the node.
3. According to the feature  $\alpha$ , the node is divided into two branches, and then the best feature from the remaining features is searched. The classification tree can grow sufficiently and the impurity of each node can be minimized through the above method.
4. In the classification phase, the classification label is a combination of the results of all classification trees. RF is based on the voting principle.

$$C_p = \text{argmax}_c \left( \frac{1}{N} \sum_{i=1}^N I \left( \frac{n_{h_i, c}}{n_{h_i}} \right) \right)$$

In the formula,  $N$  is the number of decision trees in the forest,  $n_{h_i, c}$  is the classification result of class  $C$  for tree  $h_i$ , and  $n_{h_i}$  is the number of leaf nodes of the tree.

## 5. Experimental Results and Analysis

An INV-1612 data acquisition system from the China Orient Institute of Noise & Vibration (COINV) Company was used for mechanical vibration signal acquisition. It had the following advantages: accurate measurement, high sampling rate, and a large number of ports. In this paper, the type of DC motor was a ZHS-5 multi-function rotor experiment bench. Therefore, before we collected its vibration signals, we had to consider the frequency range of the vibration signals and the sampling rate of devices. The SLM-6000 sensor signal conditioner acquired data with six test channels and an ADC resolution of 16 bits, which can extend external trigger circuits. These advantages were suitable for continuous signal acquisition in this experiment.

Besides the rotor experiment bench and signal conditioner, an LC0159 piezoelectric acceleration sensor (LANCE TECHNOLOGIES INC., Hebei, China), a DC stabilized voltage source, and an AFT-0931 signal conditioner (Electronic Technology Co., Ltd., Shijiazhuang, Hebei, China) were applied to build the vibration signal acquisition system for the DC motor. The frequency range of the LC0159 piezoelectric acceleration sensor was 1–12,000 Hz, and the range and sensitivity were 500 g and 10 mV/g, respectively, while the frequency response and gain of the AFT-0931 signal conditioner were 0.5–45 and 5 kHz, respectively.

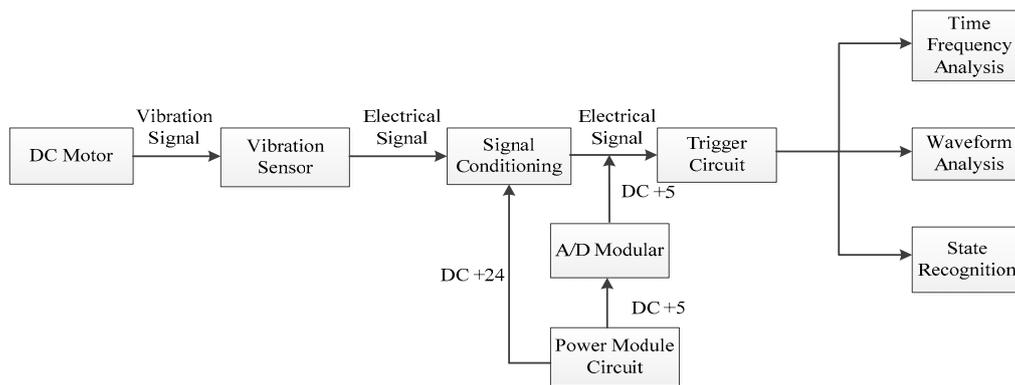
In this paper, vibration signals were collected by the INV-1612 fault data acquisition system, as shown in Figures 11 and 12. Firstly, the vibration sensor LC0159 was employed to transform the vibration signals into analog electrical signals. Secondly, the analog electrical signals were inputted into the SLM-6000 sensor signal conditioner, and the filtering and amplification processes of the analog electrical signals were then completed in the signal conditioner. In addition, the signal conditioner could prevent shock signals with big amplitudes from damaging the host computer. Finally, the analog electrical signal was inputted into the SLM-6000 for AD conversion. The design framework of the entire signal acquisition system block diagram is shown in Figure 13.



Figure 11. ZHS-5 multi-function rotor experiment bench.

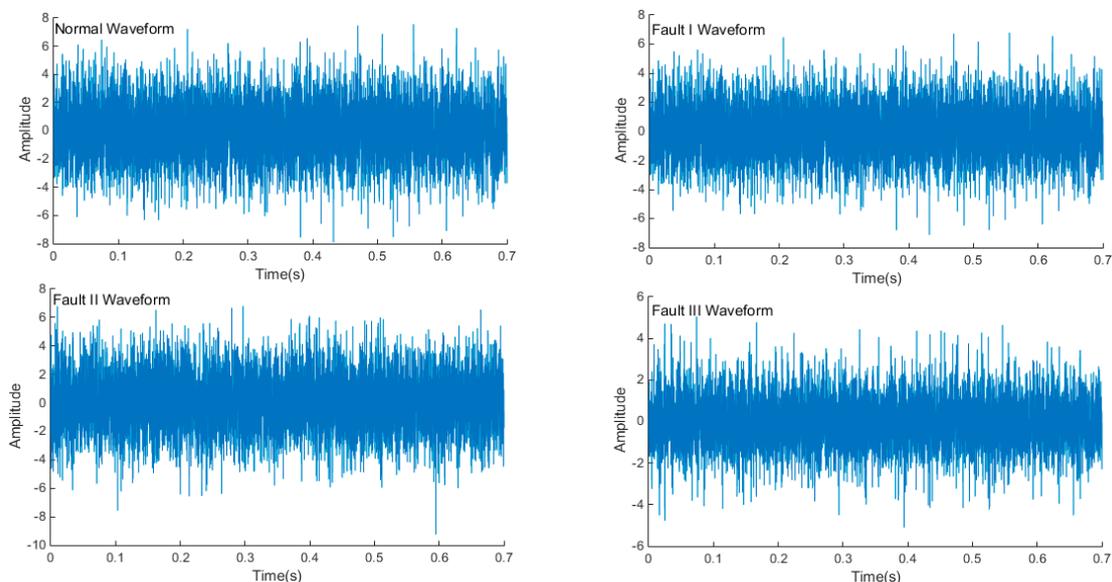


Figure 12. Model SLM-6000 sensor signal conditioner.



**Figure 13.** Design framework of the entire signal acquisition system.

In this paper, 50 samples were collected for each mechanical state by the acquisition system, and there was a total of 200 samples. Figure 14 shows the time domain waveform of the normal signal and three types of fault signals (rotor asymmetry, load-shafting misalignment, and rotor system misalignment-rubbing coupling faults). The sampling frequency was 10 Ks/s, and the sampling time was 0.7 s. The three kinds of faults were more different from normal conditions. These differences were mainly caused by the impact or friction of each mechanical component of the operating mechanism—when the mechanical fault occurred, the impact or friction process changed to a certain extent; therefore, the instantaneous frequency, instantaneous amplitude, and occurrence time of each vibration event must be different.



**Figure 14.** Time domain waveform of each mechanical state.

As mentioned previously, after signal acquisition work was completed, signals needed to be de-noised. Utilizing a variable time window and frequency window to analyze the signal, a wavelet transform was very suitable for vibration signal de-noising of rotating machinery based on a nonstationary wavelet threshold function. Meanwhile, the smoothness of the de-noised signal was considered. A wavelet transform method was applied in signal de-noising in this paper. It was an effective method to extract the vibration signal through the wavelet threshold de-noising method. Daubechies-9 wavelet was employed after experimental comparison. The de-noised signals of four types of original signals are shown in Figure 15.

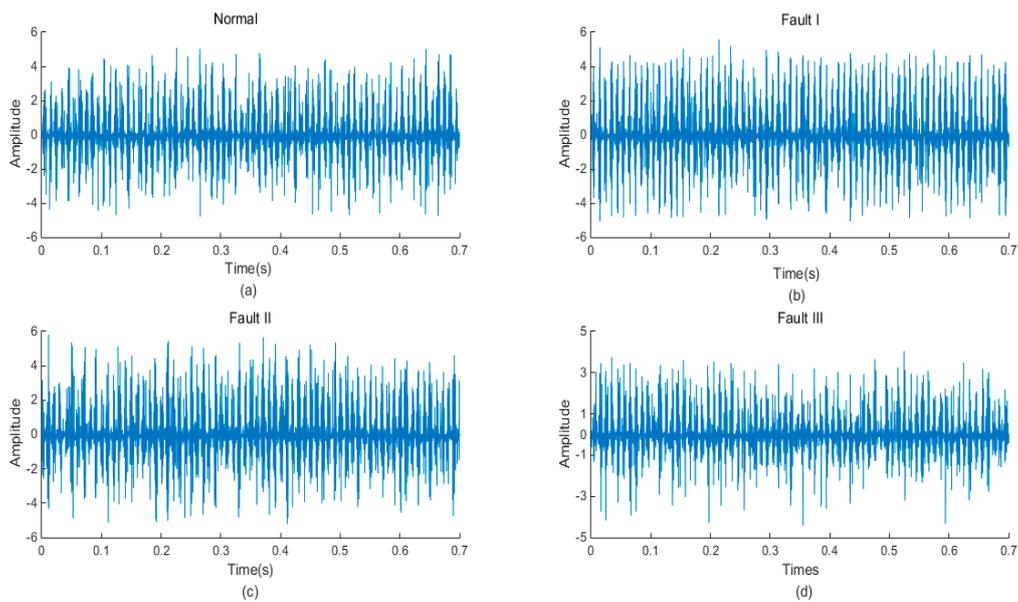


Figure 15. The de-noised signal normal: (a); Fault I: (b); Fault II: (c); Fault III: (d).

5.1. Signal Decomposition and Feature Extraction

The measured signals were more complicated and different from simulation signals. Here, seven components have been decomposed from measured signals. As shown in Figure 16, each mode component was a narrow band AM-FM signal. The frequency distribution of the vibration signals changed because of the mechanical component.

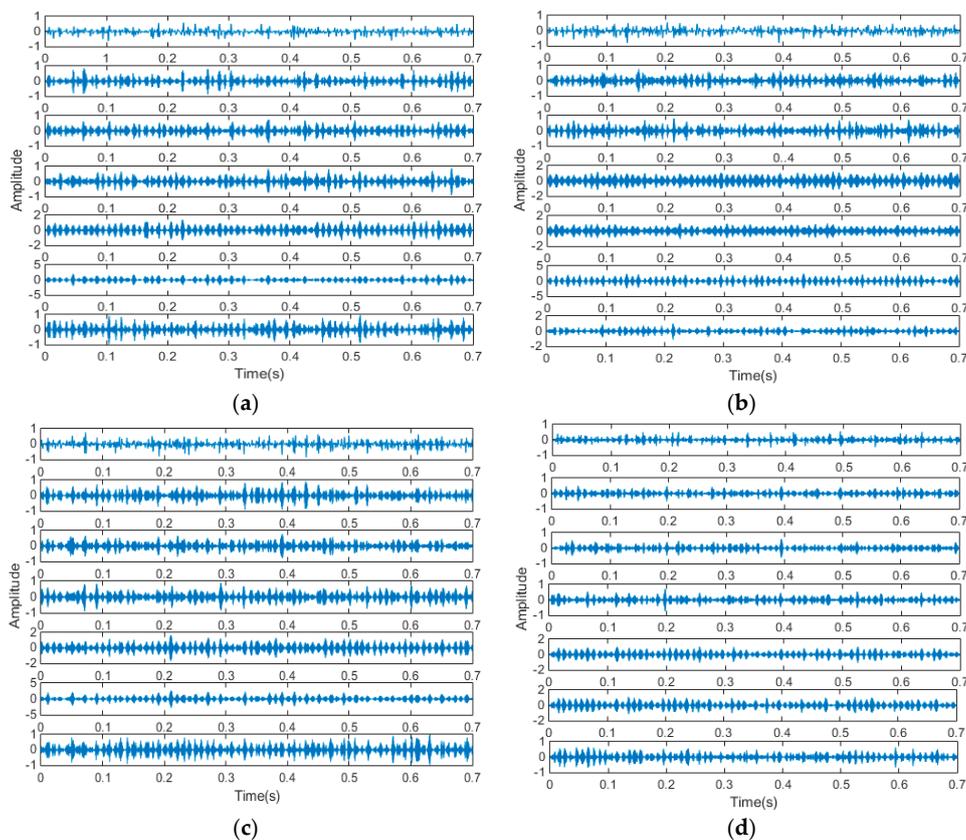


Figure 16. Vibration Components of the four types of signals obtained by the EWT method. (a) Normal type; (b) Fault I; (c) Fault II; and (d) Fault III.

The key was to determine the modal value  $K$ . When the signal was over-decomposed, the central frequency of the mode was identical. In the meantime, the correlation coefficients of each mode with different  $K$  values were compared with the original signals to verify the practicability of the method (center frequency). Based on the formula, the center frequency of the fault was calculated, as is shown in Table 2. It was proved that the center frequencies were greatly different when  $K = 3$  and 4. When  $K = 8$ , a similar number of modalities appeared, thus, the best modal value was seven [38].

**Table 2.** Change of center frequencies in VMD under different  $K$  values (Fault III).

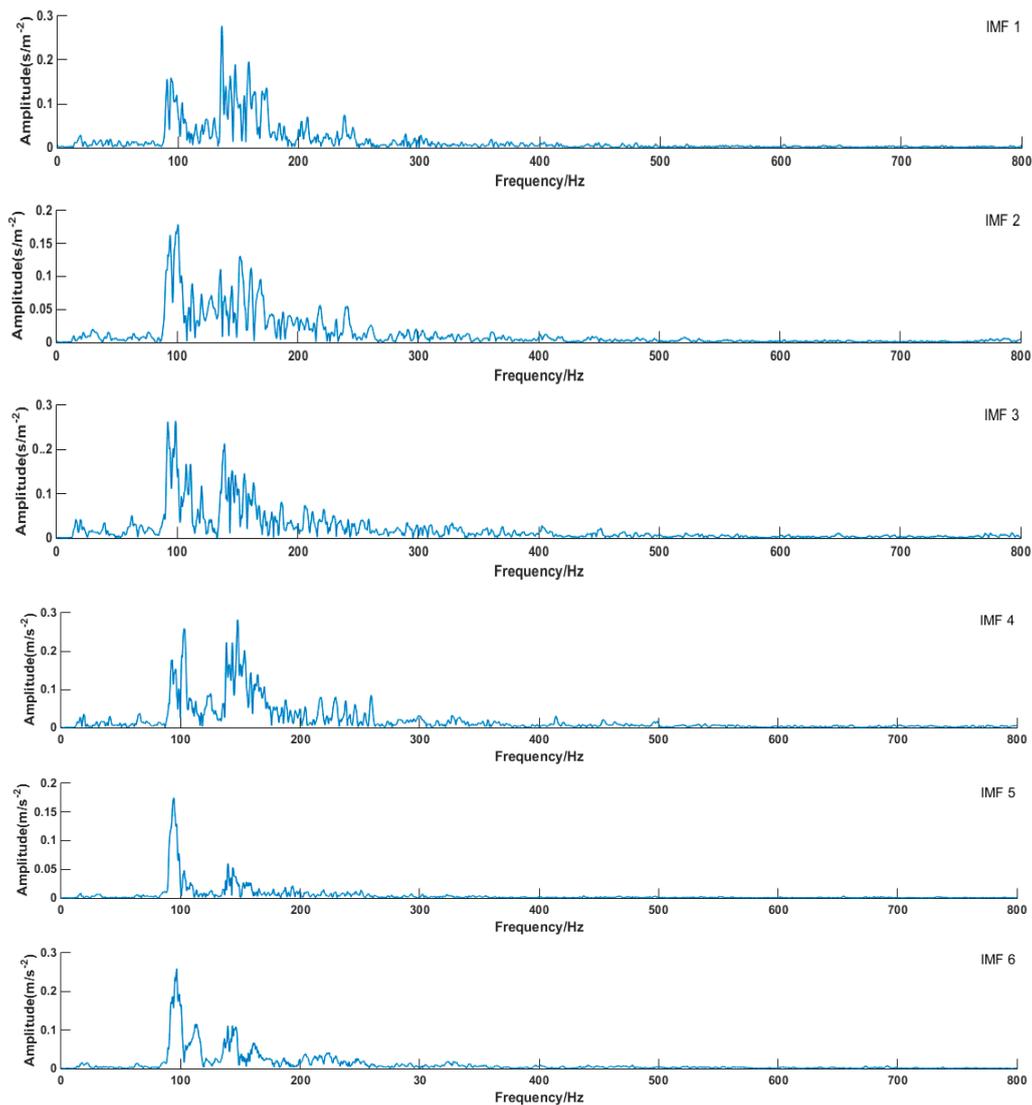
Model Value	Center Frequency							
	U1	U2	U3	U4	U5	U6	U7	U8
$K = 3$	95.86	456.25	554.80					
$K = 4$	89.69	200.74	462.75	554.80				
$K = 5$	87.17	198.56	372.27	464.10	554.96			
$K = 6$	83.56	117.50	283.25	384.21	463.85	645.28		
$K = 7$	83.56	117.49	283.21	382.64	464.32	553.70	645.28	
$K = 8$	21.67	98.64	183.31	285.25	382.27	464.32	553.70	554.88

For further proof, the correlation coefficients between each mode and the original signal were obtained by formulas. In Table 3, the correlation coefficient increased at first and then decreased. However, when  $K = 8$ , it shows that over-decomposition occurred. When  $K = 3, 4,$  and 5, modal mixing appeared. The optimum mode number was obtained when  $K = 7$ , which further proved the accuracy of the center frequency method.

**Table 3.** Change of Correlation Coefficients in VMD under different  $K$  values (Fault III).

Model Value	Correlation Coefficient							
	$\rho_1$	$\rho_2$	$\rho_3$	$\rho_4$	$\rho_5$	$\rho_6$	$\rho_7$	$\rho_8$
$K = 3$	0.3366	0.5512	0.5930					
$K = 4$	0.3029	0.3213	0.5485	0.5916				
$K = 5$	0.2945	0.3055	0.3417	0.5324	0.5876			
$K = 6$	0.2928	0.3029	0.3355	0.5258	0.5669	0.3570		
$K = 7$	0.2824	0.2825	0.2844	0.3191	0.5202	0.5650	0.3553	
$K = 8$	0.1996	0.2681	0.2720	0.2808	0.3176	0.5467	0.5176	0.3548

Modal mixing does not easily happen in the VMD algorithm. The first six envelop spectrums were compared in this paper. Figure 17 schematically illustrated the envelope spectrums of all modes that were decomposed by VMD (take Fault I for example). It can be seen from the diagram that the envelope spectrum of each component of decomposed signal can clearly reflect fault information. The envelope spectrum of decomposed signal components by EMD is shown in Figure 18. The phenomenon of modal mixing appeared in mode 4. When the number of modes was six, the envelope spectrum could not accurately reflect fault information, which proved that the decomposition signal effect of the VMD algorithm was better than that of EMD.



**Figure 17.** Envelope spectrum of each mode (VMD).

Envelope entropy processed the envelope signals obtained after demodulation operations into a probability distribution sequence, and its calculated entropy reflected the sparse characteristics of the original signals. If the IMF component contained more fault characteristic information, the envelope diagram would be relatively dense. As shown in Figure 17, compared with EMD, the envelope spectrum generated by VMD decomposition was very dense, which included much of the fault information of the physical system. The IMF 1 envelope spectrum indicated it was an optimum component by EMD decomposition, as shown in Figure 18. However, the IMF 1 envelope spectrum was too messy, and it was hard to extract valid fault information.

Meanwhile, the envelope spectrums of EWT, LMD, and VMD were compared. LMD and EWT were new nonstationary signal processing methods. They could adaptively decompose the signal according to the characteristics of the signal itself. They were superior to the EMD method in terms of restraining the endpoint effect, reducing iteration times, and preserving signal information integrity. However, LMD and EWT had lots of shortcomings. When the number of smoothing times was large, hysteresis occurred, and the step size could not optimally be determined in smoothing. Finally, compared with other methods, the VMD algorithm could extract all the information of the fault signal accurately in the field of signal decomposition, which ensured that modal mixing could be thoroughly solved.

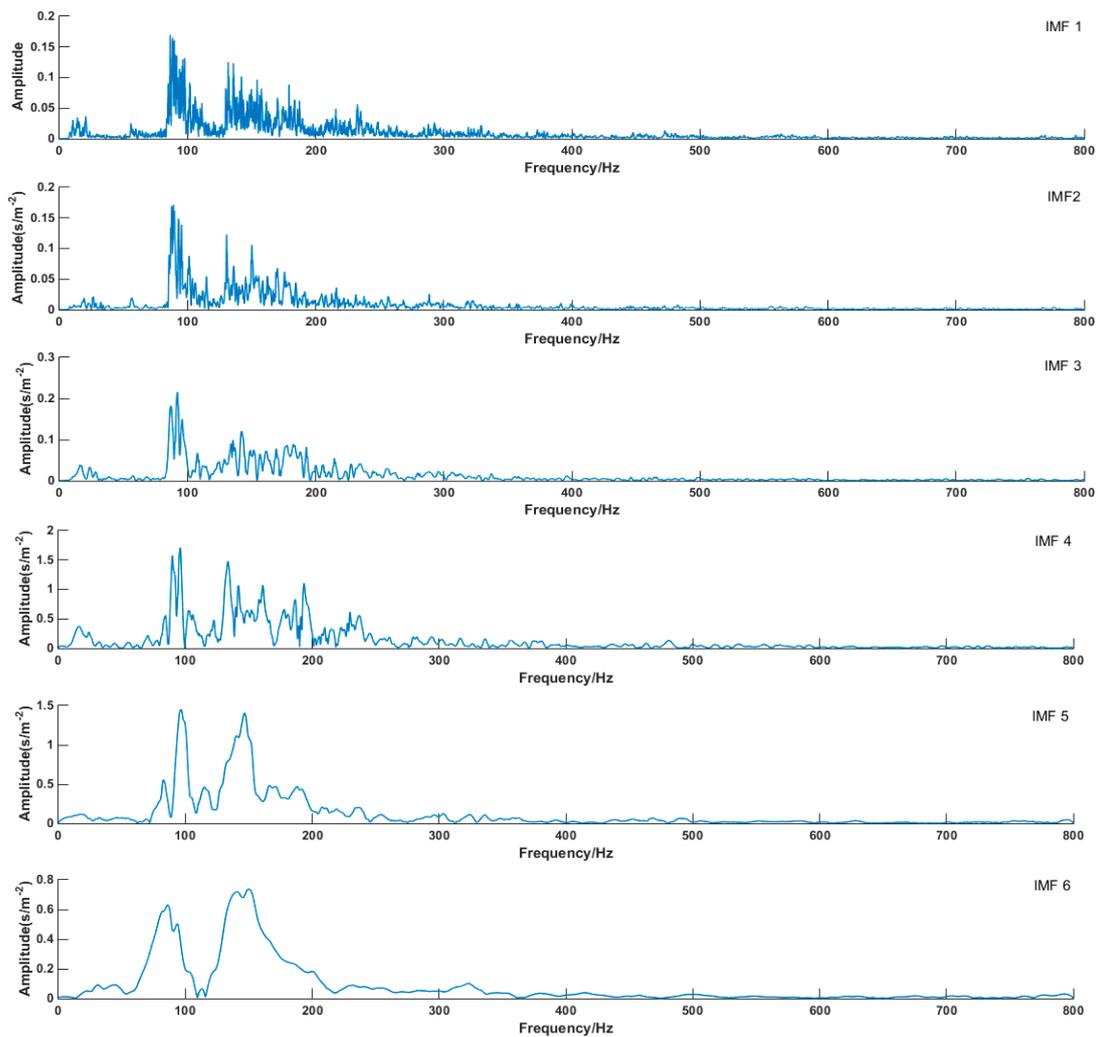


Figure 18. Envelope spectrum of each mode (EMD).

5.2. Feature Extraction

When  $M$  was one, characteristics of the random forest algorithm could not be shown, for the random forest algorithm required selecting  $m$  eigenvectors from  $M$  eigenvectors. In addition, in order to prove that random forest could show better performance under relatively simple eigenvector conditions, the time domain characteristics of motor vibration signals were extracted in this paper, and the feature vectors of each signal were extracted by using sample entropy, which were denoted as  $C = [C_1, C_2, C_3, C_4, C_5, C_6, C_7]$ . The entropy value of samples under different fault conditions is shown in Table 4.

Table 4. The Entropy Value of Samples Under Different Fault Conditions.

Motor Types	Label	SampEn I	SampEn II	SampEn III	SampEn IV	SampEn V	SampEn VI	SampEn VII	Desired Output
Normal	1	0.3241	0.4299	0.3757	0.2647	0.1455	0.2659	0.4664	1
		0.3091	0.4368	0.3818	0.2506	0.1338	0.2599	0.4445	1
		0.2914	0.4567	0.3811	0.2742	0.1630	0.2888	0.4837	1
Fault I	2	0.2239	0.5127	0.2319	0.1434	0.1818	-0.5116	0.3566	2
		0.2854	0.2787	0.2934	0.2787	0.2018	0.8924	0.4550	2
		0.3838	0.4191	0.4041	0.3033	0.1318	-0.8236	0.6149	2

Table 4. Cont.

Motor Types	Label	SampEn I	SampEn II	SampEn III	SampEn IV	SampEn V	SampEn VI	SampEn VII	Desired Output
Fault II	3	0.3084	0.3857	0.5754	0.2425	0.7845	0.3938	0.6215	3
		0.0684	0.5457	0.5754	0.2625	1.5645	0.4307	0.2915	3
		0.2384	0.5557	0.5130	0.2225	1.4085	0.4553	0.5515	3
Fault III	4	0.2749	0.6835	0.1855	0.1931	0.1958	0.3765	0.5568	4
		0.4348	0.3811	0.1955	0.5119	0.2458	0.3519	0.4968	4
		0.4840	0.5323	0.1755	0.2435	0.3758	0.3273	0.4568	4

Firstly, sample entropy was calculated by Formulas (9)–(14) and regarded as the feature vector. At the same time, the corresponding four states were labeled  $D = [1,2,3,4]$  for the expected input of the classifier. Table 4 shows that the eigenvectors of each state were similar, and the eigenvectors of different states were distinct. Subsequently, according to the process of classifier design, 200 feature vectors were randomly divided into testing sets and training sets. For each type, 30 feature vectors were randomly selected as training samples, and 20 other feature vectors were used as testing samples. In addition, 10-fold cross-validation was employed in the paper, and 120 training samples were equally divided into two data sets again, one for training, where cross validation was performed 10 times. Besides, the loop traversal interval was  $[0.1, 2]$ , and traversing step length TL was 0.1. Feature vectors were inputted into the optimal classifier (RF with “tree = 500”) for testing. Finally, the classifier was evaluated by the testing set.

### 5.3. Fault Diagnosis of the DC Motor

There are many classifiers (decision tree) in random forest. The parameters of each classifier are different, and the training sample sets of each classifier are also different. Therefore, different classifiers were generated to make the diagnosis results of each classifier diverse. Finally, final classification results were generated by voting. It effectively improved the accuracy of diagnosis.

From the figure below, the recognition rates of four conditions were demonstrated. Blue circles represented correct recognition samples, and red stars represented wrong recognition samples. The result shows that, in the 80 samples set in the testing set, the normal state was not misidentified, and there was one sample recognition error in the other three faults. The vibration signal recognition rates were: Normal 100%, Fault I 95%, Fault II 90%, and Fault III 90%. This paper also discussed the effect of random forest generalization. In Figure 19, the abscissa represented all decision trees (500 trees) of random forest and the training set samples; the ordinate represented the test set samples. The abscissa and longitudinal coordinates satisfied  $X + Y = 500$ .

If the error classification sample was close to the center of the graph (the intersection of line  $y = x$  and  $X + Y = 500$ ,  $P(250, 250)$ ), and the training set was the same as the decision tree of the test set in the implementation of the algorithm, the error recognition of the sample was acceptable (Figure 19b), verifying good performance of sample generalization.

Conversely, if the misclassified samples deviated from the center, as shown in Figure 19c,d, there was a gap between the decision tree of the training set and the test set in the whole algorithm process, and the samples would be misclassified. This was unreasonable and errors were not allowed, revealing that the generalization performance of the classifier was slightly poor.

Random forest contains different decision trees, which also exerts a certain impact on their generalization performance. Many experiments were conducted in this paper for detailed discussion and research. In order to reduce the impact of randomness, 100 random forest models were established after the decision tree was determined, and then the average of the accuracy was taken as the classification accuracy under the current decision tree. As shown in Figure 20, it was ideal to select

400–500 decision trees in random forest in the case of fault identification of rotating machinery, taking the speed of the decision tree and modeling in random forest into consideration.

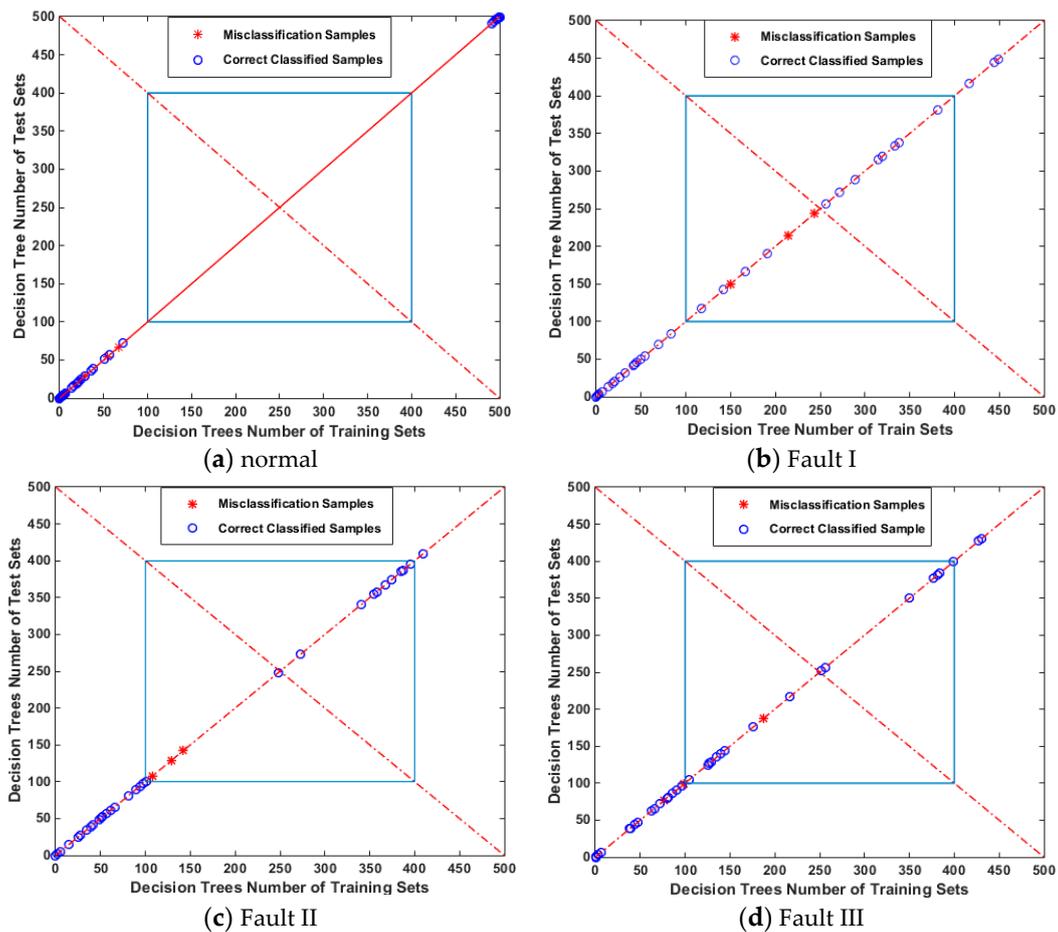


Figure 19. Performance Analysis of Forests Classifier.

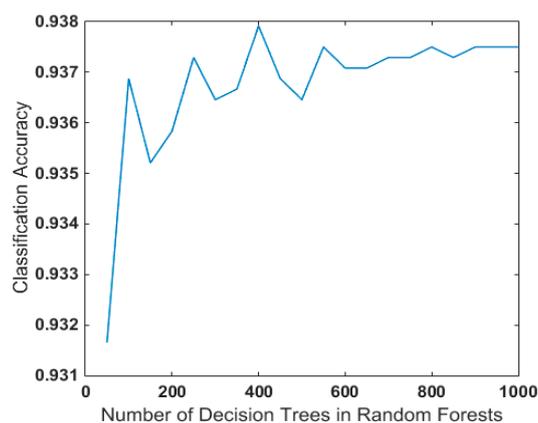


Figure 20. Influence of Decision Tree Number on Performance in Random Forest.

LQV, ELM, and decision trees are widely employed classifiers in rotating machinery fault diagnosis, and they all work perfectly to a certain extent. In order to prove the ability to differentiate normal and fault types of these classification methods, the LQV, ELM, decision tree, and random forest classifiers were compared in this paper, as shown in Table 5.

**Table 5.** Final Recognition Results.

Classifier	Experimental Parameters	Recognition Rate				
		Normal	Fault I	Fault II	Fault III	All Test Samples
LQV	Hidden Layer = 5;	75%	75%	80%	80%	77.5%
	Hidden Layer = 20;	85%	90%	90%	90%	88.75%
	Hidden Layer = 40	80%	85%	85%	85%	83.75%
ELM	Activation Function (Sin);	80%	80%	85%	85%	82.5%
	Activation Function (Sig);	85%	90%	90%	90%	88.75%
	Activation Function (Hardlim)	80%	85%	85%	85%	84.5%
Decision Tree	Decision Tree (100);	85%	80%	80%	85%	83.75%
	Decision Tree (500)	95%	90%	85%	85%	88.75%
	Decision Tree (800)	90%	85%	80%	80%	83.5%
Random Forest	Decision Tree (100);	100%	90%	90%	90%	92.5%
	Decision Tree (500)	100%	95%	90%	90%	93.75%
	Decision Tree (800)	95%	90%	90%	90%	91.25%

The value of the hidden layer neurons played a crucial role in the prediction results of the neural network. As for the ELM classifier, the value of the hidden layer neurons was the number of samples in the training set, and the activation function of hidden layer neurons was ‘sig’. Regarding the LQV classifier, we followed the literature [3] to design the LQV into a three-layer neural network, which included an input layer, a competition layer, and a linear output layer, and the number of hidden layer neurons was 20.

The parameters of the decision tree are listed in the above Table 5, which also shows the final recognition results of these three classifiers and recognition rates for different parameters. Optimal parameters for each method can be found in this table. Because of the similar fault patterns, the recognition rates of fault II and fault III showed little difference, indicating that the vibration signals of the fault I condition were very different from the vibration signals of the other three conditions. The optimized random forest classifier was compared with LQV, ELM, and decision tree classifiers. The result revealed that with the same feature extraction method, the decision tree method was better than other two neural networks, and the recognition rate of the improved RF classifier was higher than other methods.

## 6. Conclusions

This paper presents a new method for mechanical fault diagnosis of DC motors. Firstly, the original vibration signals were collected by the designed acquisition system, which was designed by Labview2014. Secondly, VMD was employed to decompose the vibration signals into a series of physically meaningful modes. Thirdly, the SampEn method was used for feature extraction. Finally, the improved random forest method was adopted to identify four mechanical conditions, where the parameter “tree” of RF was optimized by a loop traversal method. The simulation and practical test results demonstrate the following advantages of the new method:

1. Compared with EMD, LMD, and EWT methods, the VMD method can better decompose the signals into a series of physically meaningful modes, and it also can solve the mode aliasing well.
2. The proposed SampEn method is more superior in feature extraction. The feature curves of different types of signals obtained by the SampEn method are more dispersed than approximate entropy method, which is good for classification.
3. The parameter “tree” of RF needs to be determined, which usually depends on human choice. Combined with SPRINT, the loop traversal method can effectively find the best parameter of the decision tree.

- According to the final recognition rates, the VMD-SampEn feature extraction method and the optimal RF classifier are more suitable for mechanical fault diagnosis of DC motors. The proposed RF-SPRINT method provides a new idea for mechanical feature extraction of DC motors.

**Author Contributions:** Z.G. wrote and revised the paper; M.L. and H.Q. checked the paper and provided important guidance for paper; Z.G. conceived the idea and realized it by program; M.L. and B.L. provided important guidance for experiment. All authors have read and approved the final manuscript.

**Funding:** This work is supported by Basic research fund of Heilongjiang University in Heilongjiang Province under grant (No. kjcx201806).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Liu, Z.; Guo, W.; Hu, J.; Ma, W. A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and Twin SVM. *ISA Trans.* **2016**, *66*, 249. [[CrossRef](#)] [[PubMed](#)]
- Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
- Manana, M.; Arroyo, A.; Renedo, C.J.; Perez, S.; Delgado, F. Field winding fault diagnosis in DC motors during manufacturing using thermal monitoring. *Appl. Therm. Eng.* **2011**, *31*, 978–983. [[CrossRef](#)]
- Zhong, T.; Li, Y.; Wu, N.; Nie, P.; Yang, B. A study on the stationarity and Gaussianity of the background noise in land-seismic prospecting. *Geophysics* **2015**, *80*, 67–82. [[CrossRef](#)]
- Yin, J.; Wang, W.; Man, Z.; Khoo, S. Statistical modeling of gear vibration signals and its application to detecting and diagnosing gear faults. *Inf. Sci.* **2014**, *259*, 295–303. [[CrossRef](#)]
- Cui, L.; Huang, J.; Zhang, F. Quantitative and Localization Diagnosis of a Defective Ball Bearing Based on Vertical-Horizontal Synchronization Signal Analysis. *IEEE Trans. Ind. Electron.* **2017**, *64*, 8695–8706. [[CrossRef](#)]
- Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [[CrossRef](#)]
- Wang, D.; Miao, Q.; Fan, X.; Fan, X.; Huang, H.Z. Rolling element bearing fault detection using an improved combination of Hilbert and wavelet transforms. *J. Mech. Sci. Technol.* **2009**, *23*, 3292–3301. [[CrossRef](#)]
- Daubechies, I. The Wavelet Transform. Time-Frequency Localisation and Signal Analysis. *J. Renew. Sustain. Energy* **1990**, *36*, 961–1005. [[CrossRef](#)]
- Liu, S.; Zhao, H.; Hui, W.; Rui, M. Application of improved EMD algorithm for the fault diagnosis of reciprocating pump valves with spring failure. In Proceedings of the 9th International Symposium on Signal Processing and Its Application, Sharjah, UAE, 12–15 February 2007.
- Hu, X.; Peng, S.; Hwang, W.L. EMD Revisited: A New Understanding of the Envelope and Resolving the Mode-Mixing Problem in AM-FM Signals. *IEEE Trans. Signal Process.* **2012**, *60*, 1075–1086. [[CrossRef](#)]
- Lei, Y.; He, Z.; Zi, Y. Application of the EEMD method to rotor fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2009**, *23*, 1327–1338. [[CrossRef](#)]
- Bin, G.F.; Gao, J.J.; Li, X.J.; Dhillon, B.S. Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network. *Mech. Syst. Signal Process.* **2012**, *27*, 696–711. [[CrossRef](#)]
- Ling, X.; Li, Y. Application of Empirical Wavelet Transform in Fault Diagnosis of Rotary Mechanisms. *J. Chin. Soc. Power Eng.* **2015**, *12*, 975–981.
- Dragomiretskiy, K.; Zosso, D. Two-Dimensional Variational Mode Decomposition. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*; Tai, X.C., Bae, E., Chan, T.F., Lysaker, M., Eds.; Springer: Cham, Switzerland, 2015.
- Gilles, J. Empirical Wavelet Transform. *IEEE Trans. Signal Process.* **2013**, *61*, 3999–4010. [[CrossRef](#)]
- Santos-Ruiz, I.; López-Estrada, F.R.; Puig, V.; Pérez-Pérez, E.J.; Mina-Antonio, J.D.; Valencia-Palomo, G. Diagnosis of Fluid Leaks in Pipelines Using Dynamic PCA. *IFAC* **2018**, *51*, 373–380. [[CrossRef](#)]
- Santos-Ruiz, I.; Bermúdez, J.R.; López-Estrada, F.R.; Puig, V.; Torres, L.; Delgado-Aguíñaga, J.A. Online leak diagnosis in pipelines using an EKF-based and steady-state mixed approach. *Control Eng. Pract.* **2018**, *81*, 55–64. [[CrossRef](#)]

19. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
20. Mohanty, S.; Gupta, K.K.; Raju, K.S. Comparative study between VMD and EMD in bearing fault diagnosis. In Proceedings of the 9th International Conference on Industrial and Information Systems (ICIIS), Gwalior, India, 15–17 December 2014. [[CrossRef](#)]
21. Luo, J. Wavelet Fractal Technology and Its Application to Aeroengine Fault Diagnosis. *J. Projectiles Rockets Missiles Guidance* **2006**. [[CrossRef](#)]
22. Lin, J.; Qu, L.S. Feature Extraction Based on Morlet Wavelet And Its Application For Mechanical Fault Diagnosis. *J. Sound Vib.* **2000**, *234*, 135–148. [[CrossRef](#)]
23. Rosenstein, M.T. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D* **1993**, *65*, 117–134. [[CrossRef](#)]
24. Cheng, J.; Yu, D.; Yang, Y. Fault Diagnosis for Rotor System Based on EMD and Fractal Dimension. *China Mech. Eng.* **2005**, *6*, 1088–1091.
25. Goldberger, A.L. Non-linear dynamics for clinicians: Chaos theory, fractals, and complexity at the bedside. *Lancet* **1996**, *347*, 1312–1314. [[CrossRef](#)]
26. Li, Y.; Shao, W.; Pu, Y.; Wang, H.; Kui-Long, Z.; Wei-Guo, Z.; Yu-Long, T. Application of Lyapunov Exponents to Fault Diagnosis of Rolling Bearing. *Noise Vib. Control* **2007**, *5*, 104–105.
27. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
28. Basaran, C.; Yan, Y. A Thermodynamic Framework for Damage Mechanics of Solder Joints. *J. Electron. Packag.* **1998**, *120*, 379–384. [[CrossRef](#)]
29. Naderi, M.; Amiri, M.; Khonsari, M.M. On the thermodynamic entropy of fatigue fracture. *Proc. R. Soc. Lond.* **2010**, *466*, 423–438. [[CrossRef](#)]
30. Leonid, S.; Sergei, S. Mechano-thermodynamic Entropy and Analysis of Damage State of Complex Systems. *Entropy* **2016**, *18*, 268. [[CrossRef](#)]
31. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*. [[CrossRef](#)]
32. Lake, D.E.; Richman, J.S.; Griffin, M.P.; Moorman, J.R. Sample entropy analysis of neonatal heart rate variability. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2002**, *283*, R789–R797. [[CrossRef](#)] [[PubMed](#)]
33. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In *Ensemble Machine Learning*; Zhang, C., Ma, Y.Q., Eds.; Springer: New York, NY, USA, 2012; pp. 157–175.
34. Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **1985**, *39*, 783–791. [[CrossRef](#)]
35. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
36. Quinlan, J.R. Induction on decision tree. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
37. Tripoliti, E.E.; Fotiadis, D.I.; Manis, G. Dynamic construction of Random Forests: Evaluation using biomedical engineering problems. In Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine, Corfu, Greece, 3–5 November 2010.
38. Sun, J.; Wang, C.; Sun, J.; Wang, L. Analog circuit fault diagnosis based on PCA and LVQ. *J. Circuits Syst.* **2013**, *8*. [[CrossRef](#)]
39. He, X.; Wang, H.; Lu, J.; Jiang, W. Analog circuit fault diagnosis method based on preferred wavelet packet and ELM. *Chin. J. Sci. Instrum.* **2013**, *34*, 2614–2619.
40. Yan, X.; Jia, M.; Xiang, L. Compound fault diagnosis of rotating machinery based on OVMD and a 1.5-dimension envelope spectrum. *Meas. Sci. Technol.* **2016**, *27*, 075002. [[CrossRef](#)]
41. Hestenes, M.R. Multiplier and gradient methods. *J. Optim. Theory Appl.* **1969**, *4*, 303–320. [[CrossRef](#)]
42. Peng, Z.K.; Tse, P.W.; Chu, F.L. A comparison study of improved Hilbert-Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mech. Syst. Signal Process.* **2005**, *19*, 974–988. [[CrossRef](#)]
43. Wang, H.; Chen, J.; Dong, G. Fault Diagnosis Method for Rolling Bearing's Weak Fault Based on Minimum Entropy Deconvolution and Sparse Decomposition. *J. Mech. Eng.* **2013**, *49*, 88–94. [[CrossRef](#)]
44. Pincus, S.M. Assessing serial irregularity and its implications for health. *Ann. N. Y. Acad. Sci.* **2010**, *954*, 245–267. [[CrossRef](#)]
45. Alcaraz, R.; Rieta, J.J. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms. *Biomed. Signal Process. Control* **2010**, *5*, 1–14. [[CrossRef](#)]

46. Sathyadevan, S.; Nair, R.R. Comparative Analysis of Decision Tree Algorithms: ID3, C4.5 and Random Forest. In *Computational Intelligence in Data Mining*; Jain, L., Behera, H., Mandal, J., Mohapatra, D., Eds.; Springer: New Delhi, India, 2015; pp. 549–562.
47. Luo, Z.K.; Sun, H.Y.; Wang, D. An Improved SPRINT Algorithm. *Adv. Mater. Res.* **2012**, 532–533, 1685–1690. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).