

Article

# Research on Multi-Domain Fault Diagnosis of Gearbox of Wind Turbine Based on Adaptive Variational Mode Decomposition and Extreme Learning Machine Algorithms

Hui Li <sup>1,\*</sup>, Bangji Fan <sup>1</sup>, Rong Jia <sup>1</sup>, Fang Zhai <sup>2</sup>, Liang Bai <sup>3</sup> and Xingqi Luo <sup>3</sup>

<sup>1</sup> School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China; eufanbangji@163.com (B.F.); Ja.rong@163.com (R.J.)

<sup>2</sup> School of Humanities and Foreign Languages, Xi'an University of Technology, Xi'an 710048, China; xautzf@126.com

<sup>3</sup> Institute of Water Resources and Hydro-Electric Engineering, Xi'an University of Technology, Xi'an 710048, China; xaut\_bl@126.com (L.B.); luoxq@xaut.edu.cn (X.L.)

\* Correspondence: lihui@mail.xaut.edu.cn; Tel.: +86-29-8885-2376

Received: 13 February 2020; Accepted: 9 March 2020; Published: 16 March 2020



**Abstract:** Since variational mode decomposition (VMD) was proposed, it has been widely used in condition monitoring and fault diagnosis of mechanical equipment. However, the parameters  $K$  and  $\alpha$  in the VMD algorithm need to be set before decomposition, which causes VMD to be unable to decompose adaptively and obtain the best result for signal decomposition. Therefore, this paper optimizes the VMD algorithm. On this basis, this paper also proposes a method of multi-domain feature extraction of signals and combines an extreme learning machine (ELM) to realize comprehensive and accurate fault diagnosis. First, VMD is optimized according to the improved grey wolf optimizer; second, the feature vectors of the time, frequency, and time-frequency domains are calculated, which are synthesized after dimensionality reduction; ultimately, the synthesized vectors are input into the ELM for training and classification. The experimental results show that the proposed method can decompose the signal adaptively, which produces the best decomposition parameters and results. Moreover, this method can extract the fault features of the signal more completely to realize accurate fault identification.

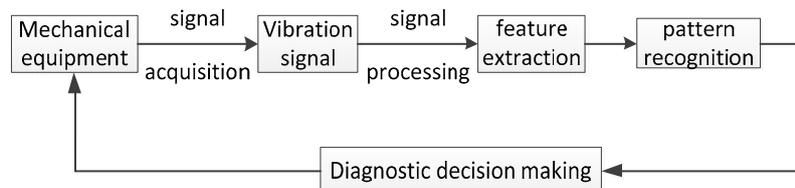
**Keywords:** VMD; grey wolf optimizer; principal components analysis (PCA); multi-domain fault diagnosis; ELM

## 1. Introduction

With the increasing depletion of traditional energies, wind energy, a clean energy, has been widely considered. Because of the instability and randomness of wind speeds in addition to the poor operating environment, the speed and load of a wind-based unit continuously change, which results in a high failure rate for the unit. It is difficult to find faults in the early stages, and once a fault truly occurs, the fault leads to very large economic losses. Therefore, it is urgent to study the early failure of the gearbox of a wind turbine and determine the cause of the failure because these characteristics are of great significance to ensuring the safe and stable operation of the wind turbine and avoiding catastrophic accidents.

The main contents of fault diagnosis generally include: Signal acquisition, feature extraction, pattern recognition, diagnostic decision making, and other basic links. The specific process is shown in Figure 1. The gearbox is an important part of the wind turbine, whose signals are non-linear and

non-stationary [1]. Therefore, it is necessary to use various signal processing methods to preprocess before feature extraction. Time-frequency methods such as the short-time Fourier transform (STFT), Wigner-Ville distribution (WVD), and discrete wavelet transform (DWT) have their own limitations, and it is difficult to achieve satisfactory results in the actual signal analysis [2–5]. EMD based signal processing methods are the mainstream at present. In reference [6], a fault feature extraction method based on local mean decomposition (LMD) and multi-scale entropy are proposed. Using LMD as preprocessing, the non-stationary vibration signal of rolling bearing is decomposed into several product functions, and its multi-scale entropy is taken as a feature vector. A feature extraction method based on energy entropy of the empirical mode decomposition (EMD) is proposed in reference [7]. In reference [8], the vibration signal is decomposed into a set of intrinsic mode functions (IMFs) by ensemble empirical mode decomposition (EEMD). The PE values of the first few IMFs are calculated as feature vectors, and the distance between clusters in the feature space is used to optimize the support vector machine (SVM) for fault types and severity classification of the degree. However, these methods only extract the time-frequency characteristic information of the vibration signal, and the extracted characteristic information is often not comprehensive enough. Reference [9] decomposes vibration signals into IMFs through EMD. 13 time-domain feature parameters and 16 frequency-domain feature parameters are extracted, and these parameters are input into a SVM for fault diagnosis. However, there are still some deficiencies in the research of multi-domain feature extraction. For example, traditional signal processing methods often have unsatisfactory results, and multi-domain feature vectors, while comprehensively extracting features, will also bring information redundancy and affect the difficulty of fault diagnosis.



**Figure 1.** Implementation process of fault diagnosis.

Empirical mode decomposition (EMD) is a kind of adaptive signal decomposition method that can adaptively decompose the signal into a series of intrinsic mode functions according to the local scale characteristics of the signal itself to reveal the internal characteristics of the signal [10–12]. In reference [13], a local mean decomposition (LMD) method is proposed that overcomes the problems of overpackaging and insufficient envelopes in EMD and has the advantages of fewer end effects and fewer iterations. However, LMD also faces the problem of modal mixing. For the sake of suppressing these problems, a method called ensemble empirical mode decomposition (EEMD) is proposed in reference [14]. However, a series of recursive decomposition methods, such as EMD, face the problem of modal mixing. The fundamental reason for these problems is the limitation of the recursive decomposition principle. Therefore, to fundamentally solve this problem, a new method is needed. VMD is a new method that is completely different from the recursive mode decomposition [15]. VMD takes the solution of the variational problem as its whole frame, has a solid theoretical basis, and can overcome the modal mixing. However, before using VMD for signal decomposition, the decomposition scale  $K$  and penalty factor  $\alpha$  need to be set in advance. For the determination of these two parameters, the central frequency observation method is used in reference [16], while references [17,18] determine the parameter value according to artificial experience. However, these methods do not fundamentally solve the problem of parameter determination. Recently, intelligent algorithms such as the particle swarm optimization (PSO) have been used to optimize VMD parameters with good results [19–21]. Pattern recognition is the second key in fault diagnosis. Compared with traditional pattern recognition algorithms such as back propagation (BP) neural network and SVM, extreme learning machine (ELM) has the advantages of fast training speed, high learning efficiency, and strong robustness [16,22,23].

Based on these researches, this paper applies an improved GWO algorithm [24,25] to VMD parameter optimization to achieve better adaptive VMD decomposition. In addition, combined with the ELM, this paper proposes a multi-domain fault diagnosis method and applies it to the fault diagnosis of the rolling bearing of the gearbox.

## 2. Relevant Theoretical Basis

### 2.1. Variational Mode Decomposition (VMD)

Unlike recursive modal decomposition methods such as EMD, VMD is a new method for signal decomposition and estimation with better time-frequency distribution. The VMD method can adaptively divide the frequency domain of the signal and effectively separate the components to obtain a series of intrinsic mode functions (IMFs) with sparse characteristics [26]. In essence, the VMD algorithm transforms the decomposition process of the signal into the solution process of the variational problem. Therefore, the algorithm can be divided into two parts, that is, the construction and solution of the variational problem.

#### (a) The construction of the variational problem

First, to obtain the single side spectrum of each IMF component  $u_k(t)$ , the Hilbert transform is applied to each component  $u_k(t)$ :

$$\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \quad (1)$$

where  $\delta(t)$  is a pulse function,  $j$  is the imaginary unit, and  $u_k(t)$  is the  $k$ -th IMF component decomposed.

Second, the frequency spectrum of each mode is modulated to the corresponding fundamental frequency band, that is, an estimated center frequency is added to the analytical signal of the IMF component, as follows:

$$\left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)\right] e^{-j\omega_k t} \quad (2)$$

Finally, the bandwidth of each IMF is estimated by solving the square L2 norm of the demodulation signal gradient. The variational problem can be expressed as follows:

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

$$s.t. \sum_{k=1}^K u_k = f$$

where  $\omega_k$  is the center frequency of the  $k$ -th IMF component  $u_k$  and  $f$  is the original input signal.

#### (b) The solution of the variational problem

First, the quadratic penalty factor  $\alpha$  and Lagrange multiplication operator  $\lambda(t)$  are introduced to transform the constrained variational problem into an unconstrained problem as follows:

$$L(\{u_k\}, \{\omega_k\}, \lambda)$$

$$= \alpha \sum_k \left\| \partial_t \left[ \left(\delta(t) + \frac{j}{\pi t}\right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|f(t) - \sum_k u_k(t)\|^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \quad (4)$$

Second, the alternating direction method of multipliers (ADMM) is used to solve the variational problem [27].

Finally, the renewal formulas of the IMF component  $u(k)$  and its central frequency  $\omega(k)$  are obtained:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (5)$$

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

In fact, before the VMD decomposition of the signal, four parameters need to be set artificially: Decomposition scale  $K$ , penalty factor  $\alpha$ , noise tolerance  $\tau$ , and discrimination accuracy  $\varepsilon$ . It is found that the two parameters of noise tolerance and discrimination accuracy have little influence on the decomposition results of VMD, and the standard default value is usually adopted [19]. The  $K$  and  $\alpha$  are two important parameters of VMD decomposition [28]. In this paper, an improved GWO algorithm is selected to adaptively determine these two parameters to obtain the best combination of decomposed parameters. See Section 3 for details.

## 2.2. Fast Grey Wolf Optimizer (F-GWO)

The grey wolf optimizer (GWO) is inspired by the hunting behavior of wolves and is an optimization algorithm derived from the grey wolf's population mechanism and predatory behavior. The optimal solution can be obtained through continuous iteration. The basic process of this algorithm consists of the following three parts.

### (a) Rank Assignment

GWO imitates the hierarchy of grey wolves and divides them into four grades,  $(\alpha, \beta, \delta, \omega)$ . In the process of optimization,  $\omega$  is responsible for the path search, and  $(\alpha, \beta, \delta)$  are responsible for the guidance of the path search (optimization process). According to the rank assignment,  $\alpha$  obtains the optimal solution,  $\beta$  obtains the suboptimal solution, and  $\delta$  obtains the general solution.

### (b) Target Encircling

The  $\omega$  grey wolves find the best way to the target in all directions. To avoid falling into a local optimum, the  $\omega$  grey wolves need to traverse the whole path.

### (c) Target Attacking

At this stage,  $(\alpha, \beta, \delta)$  give instructions to guide  $\omega$  to move and gradually shorten the distance between wolves and prey. Then, the location is updated, and hunting is realized.

The specific steps and mathematical modeling of the GWO algorithm are as follows:

Step 1: Initialize the parameters, set the number of grey wolves  $\omega$ , number of iterations, etc.

Step 2: According to the position of each wolf after initialization, calculate the distance between each wolf and its prey, that is, the solution and fitness value. Then, the optimal solution, suboptimal solution and general solution are assigned to  $\alpha$ ,  $\beta$  and  $\delta$  according to the fitness value.

Step 3: According to the location of the prey and the distance between a grey wolf and the prey, change the search direction of the other grey wolves and update their location, that is, update the solution. This step is equivalent to the behavior of target encircling, and Formulas (7) and (8) are the corresponding mathematical modeling.

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \quad (7)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (8)$$

where  $\vec{D}$  represents the distance  $\vec{X}_p(t)$  and  $\vec{X}(t)$  are the position vectors of the prey and grey wolf, respectively.  $\vec{C} = 2\vec{r}_1$  and  $\vec{A} = 2\vec{a} \cdot \vec{r}_2 - \vec{a}$  are coefficient vectors,  $\vec{r}_1$  and  $\vec{r}_2$  are random vectors within the range of 0 to 1. As the number of iterations increases,  $\vec{a}$  decreases from 2 to 0.

Step 4: Keep the first three optimal positions for  $\alpha$ ,  $\beta$  and  $\delta$  until the end of the iteration. The other wolves  $\omega$  update their positions according to the positions of the first three ranks. In short,  $\alpha$ ,  $\beta$  and  $\delta$  estimate the position of the prey, while the other wolves follow the three wolves to update their

position around the prey and finally output the position of the wolf  $\alpha$ , that is, to obtain the best position. This step is equivalent to the final behavior of the target attack, and Formulas (9)–(11) are the corresponding mathematical modeling.

$$\begin{aligned}\vec{D}_\alpha &= \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right| \\ \vec{D}_\beta &= \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \\ \vec{D}_\delta &= \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right|\end{aligned}\quad (9)$$

$$\begin{aligned}\vec{X}_1 &= \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha) \\ \vec{X}_2 &= \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta) \\ \vec{X}_3 &= \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta)\end{aligned}\quad (10)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}\quad (11)$$

where  $\vec{D}_{\alpha,\beta,\delta}$  denotes the distances between the first three grey wolves and their prey,  $\vec{X}_{\alpha,\beta,\delta}$  denotes the positions of the first three grey wolves, and  $\vec{X}(t+1)$  denotes the positions of the other grey wolves updated by the positions of  $\alpha$ ,  $\beta$  and  $\delta$ .

Although the GWO algorithm is much better than the PSO and DE algorithms in performance, GWO has the disadvantages of slow convergence speed and weak global search ability. To solve this problem, the fast GWO algorithm [25], which improves the convergence factor and introduces the dynamic weight strategy, is used to optimize VMD parameters in this paper.

As the GWO algorithm is nonlinear in the convergence process, and the convergence factor  $\vec{a}$  decreases linearly from 2 to 0 with the number of iterations, a new nonlinear convergence method is introduced as follows:

$$a = 2 - 2\left(\frac{1}{e-1} \times (e^{\frac{t}{m}} - 1)\right)\quad (12)$$

To solve the problem of the GWO algorithm easily falling into a local optimum, a proportional weight based on the module value of the guiding position vector is introduced:

$$\omega_i = \frac{\vec{X}_i}{\vec{X}_1 + \vec{X}_2 + \vec{X}_3} \quad i = 1, 2, 3\quad (13)$$

where  $\omega_1$ ,  $\omega_2$  and  $\omega_3$  represent the learning rates of grey wolves  $\omega$  corresponding to wolves  $\alpha$ ,  $\beta$  and  $\delta$ , respectively.

Thereby, the global search ability of GWO is dynamically balanced. Finally, the update mode of the grey wolf positions is optimized as follows:

$$\vec{X}(t+1) = \frac{\omega_1 \cdot \vec{X}_1 + \omega_2 \cdot \vec{X}_2 + \omega_3 \cdot \vec{X}_3}{3}\quad (14)$$

### 3. Proposed Methods

#### 3.1. Adaptive VMD Algorithm

In this section, the F-GWO was introduced to optimize the parameters of VMD.

The fitness value is the core of the optimization algorithm. Because the value of entropy reflects the uncertainty, information entropy is an eminent index to judge signal sparseness [29]. In this paper,

the minimum average envelope entropy (MAEE) was selected as the fitness value of VMD optimization, which is expressed as follows:

$$\langle \widehat{K}, \widehat{\alpha} \rangle = \min_{(K,\alpha)} \left\{ \frac{1}{\widehat{K}} \sum_{i=1}^{\widehat{K}} H_{en}(i) \right\} \tag{15}$$

where  $\langle \widehat{K}, \widehat{\alpha} \rangle$  is the optimal combination of  $K$  and  $\alpha$ .  $H_{en}(i)$  is the envelope entropy of the  $i$ th IMF component. The calculation formulas are as follows in Formulas (16) and (17).

$$H_{en}(i) = - \sum_{n=1}^N p_i(n) \log_2(p_i(n)) \tag{16}$$

$$p_i(n) = b_i(n) / \sum_{n=1}^N b_i(n) \tag{17}$$

In the above formulas,  $b_i(n)$  is the envelope of the  $i$ th IMF component,  $N$  is the number of sampling points, and  $p_i$  is the normalized form of the envelope of the  $i$ th IMF component.

Based on the theoretical basis in the previous section and MAEE, the algorithm flow is shown in Figure 2.

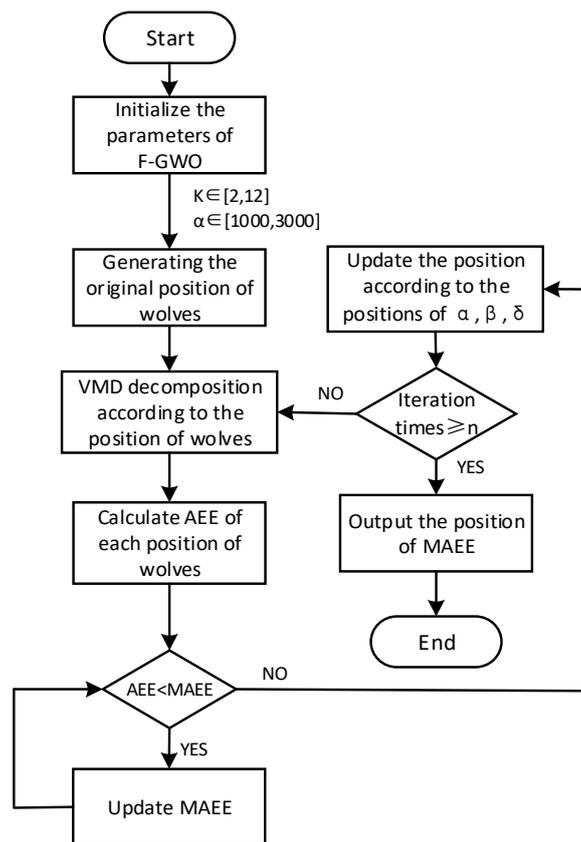


Figure 2. Flowchart of the adapted variational mode decomposition (VMD) method.

### 3.2. Extraction of Multi-Domain Fault Feature

When an equipment fault occurs, its time and frequency domains often have corresponding feature changes. In this paper, the time-domain, frequency-domain and time-frequency-domain were synthesized to comprehensively extract fault features and realize the extraction of multi-domain fault features.

## (a) The character value of the time domain

According to the adaptive VMD (AVMD) proposed in Section 3.1, the original signal was decomposed to obtain the time-domain information of  $K$  IMF components, and the feature vector of the time domain, composed of  $K$  singular values, was obtained by singular value decomposition of the matrix formed by the components.

When decomposing different signals, the number of decomposed  $K$  was different, which led to different dimensions of  $K$  for the time-domain feature vector. To solve this problem, PCA was introduced, which can help analyze multi-dimensional data and reduce the dimension of the data. The PCA quoted in this paper has the following steps (with  $m$   $n$ -dimensional data):

Step 1: Make the original data into the matrix  $X$  ( $n * m$ );

Step 2: Zero average each row of  $X$ , that is, subtract the average of each row;

Step 3: Figure out the covariance matrix;

Step 4: Figure out the eigenvalues and eigenvectors of the covariance matrix;

Step 5: Arrange the eigenvectors into a matrix from top to bottom according to the corresponding eigenvalue size. Take the first  $k$  rows to form matrix  $P$ ;

Step 6: Then, reduce  $m$   $n$ -dimensional data to  $k$ -dimensional.

In this way, information redundancy can be reduced, and the problem of different dimension of  $K$  for the feature vector can be solved.

## (b) The character value of the frequency domain

The  $K$  IMF components are decomposed by the FFT to obtain the frequency-domain information of the signal. The next step is similar to the extraction of the time-domain features.

## (c) The character value of the time-frequency domain

The complexity of IMFs decomposed from vibration signals with different faults was different. The greater the complexity was, the greater the uncertainty and the greater the entropy were. In particular, for some specific faults in the gearbox, because fault information was usually concentrated in a sensitive frequency band, once a fault occurred, the complexity in the sensitive frequency band changed accordingly. In reference [30], a method, permutation entropy (PE), was proposed to detect the randomness and dynamic mutation of a time series. Compared with approximate entropy and sample entropy, PE had a faster calculation speed and stronger anti-interference ability. Based on this, the time-frequency characteristics of the signal can be reflected by PE. The specific implementation method was to decompose the vibration signal through the adaptive VMD algorithm. Each decomposed IMF component contained the characteristic information of different frequency bands from the original vibration signal, which can better reflect the local characteristics of the signal. Therefore, the PE of each IMF component can be calculated to form the feature vector in the time-frequency domain.

### 3.3. Method of Multi-Domain Fault Diagnosis

The framework of the method is shown in Figure 3.

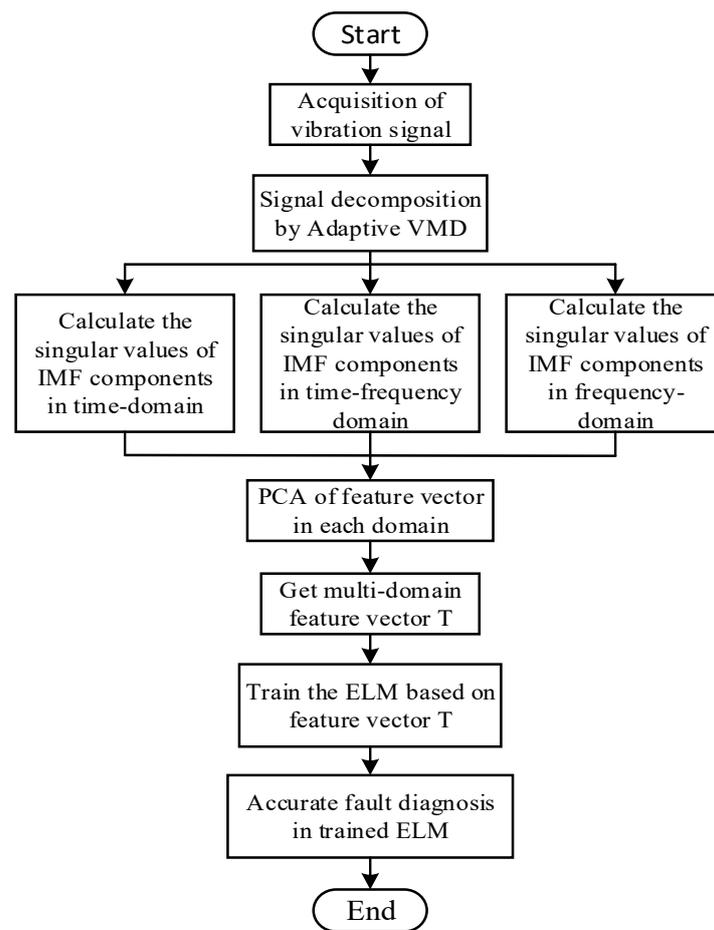


Figure 3. Flowchart of the multi-domain fault diagnosis.

#### 4. Simulation Verification

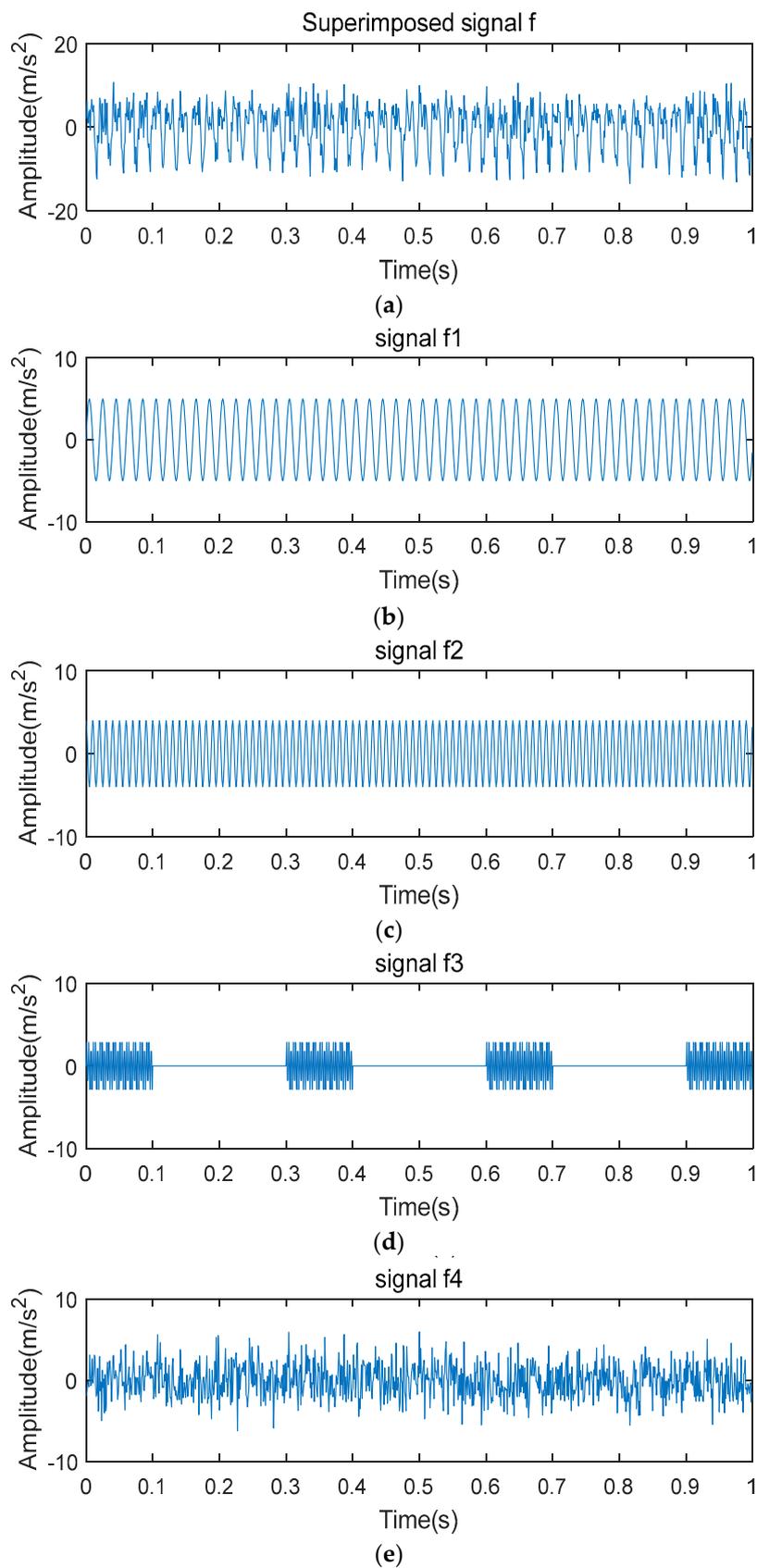
According to the research [31], the problem of modal mixing in a series of recursive decomposition methods, such as EMD, was usually caused by the interference of abnormal signals, such as noise signal and intermittent signal. Therefore, the simulated signal of Formulas (18) and (19) was established in MATLAB to verify the proposed FGWO-VMD method. The simulation time was 1 s, and the sampling rate was 1000 Hz.

$$f(t) = f_1(t) + f_2(t) + f_3(t) + f_4(t) \quad (18)$$

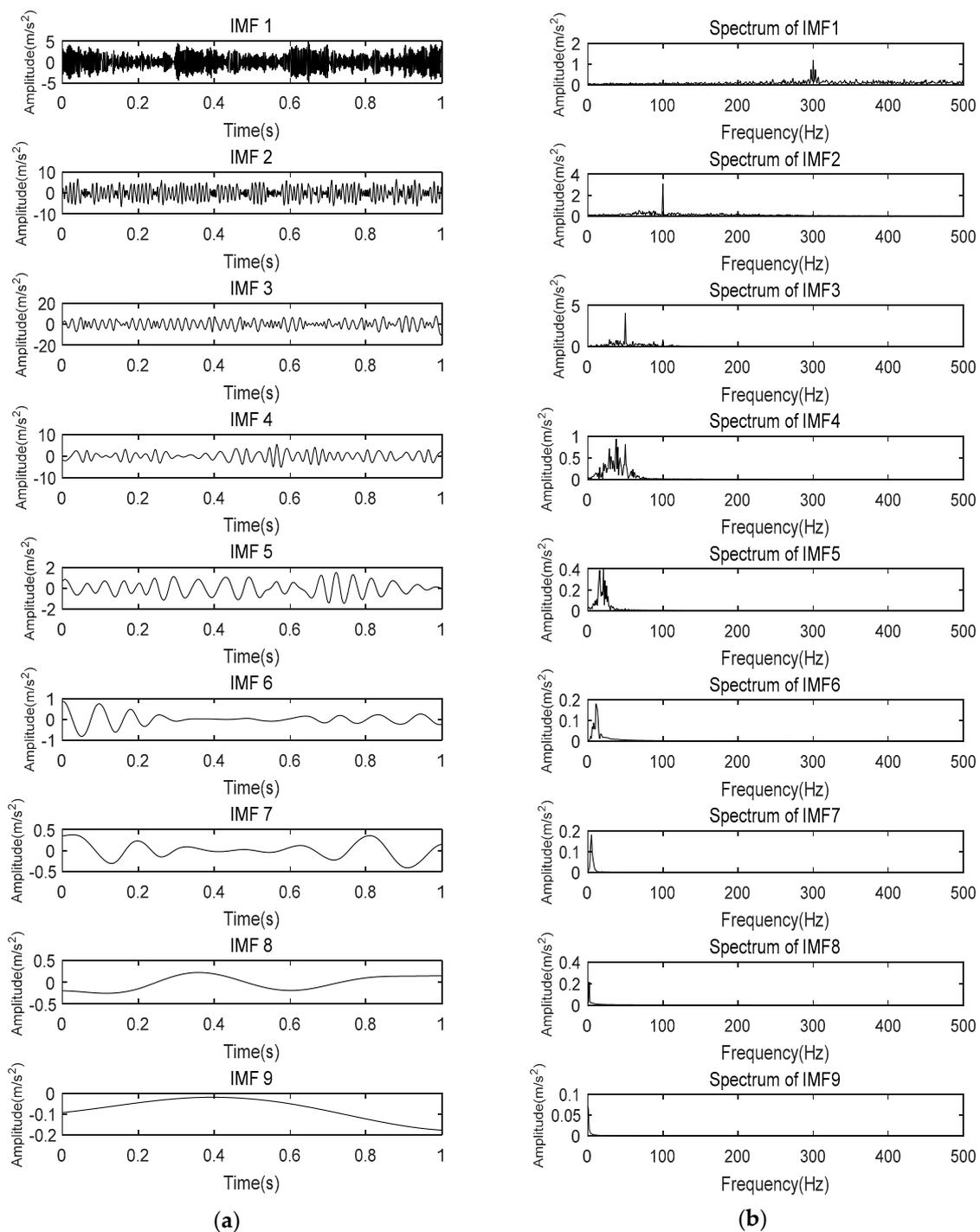
$$\begin{aligned} f_1(t) &= 5 \cdot \sin \cdot (2\pi \cdot 50 \cdot t) \\ f_2(t) &= 4 \cdot \cos \cdot (2\pi \cdot 100 \cdot t) \\ f_3(t) &= \begin{cases} 3 \cdot \sin \cdot (2\pi \cdot 300 \cdot t) & t \in [[0, 0.1], [0.3, 0.4], [0.6, 0.7], [0.9, 1]] \\ 0 & t \in [(0.1, 0.3), (0.4, 0.6), (0.7, 0.9)] \end{cases} \end{aligned} \quad (19)$$

It can be seen that  $f_1(t)$  and  $f_2(t)$  were sine cosine signals with amplitudes of 5, 4, and frequencies of 50 Hz and 100 Hz, respectively;  $f_3(t)$  is high frequency intermittent signal;  $f_4(t)$  is white noise with a mean value of zero and variance of 4. The four signals were combined as superimposed signals  $f(t)$ . Figure 4 shows the simulation signal diagram.

Firstly, EMD, LMD and EEMD are used to decompose the signal  $f(t)$ . The decomposition results and their spectra are shown in Figures 5–7, respectively.



**Figure 4.** Time-domain of simulation signal: (a) signal  $f(t)$ ; (b) signal  $f_1(t)$ ; (c) signal  $f_2(t)$ ; (d) signal  $f_3(t)$ ; (e) signal  $f_4(t)$ .

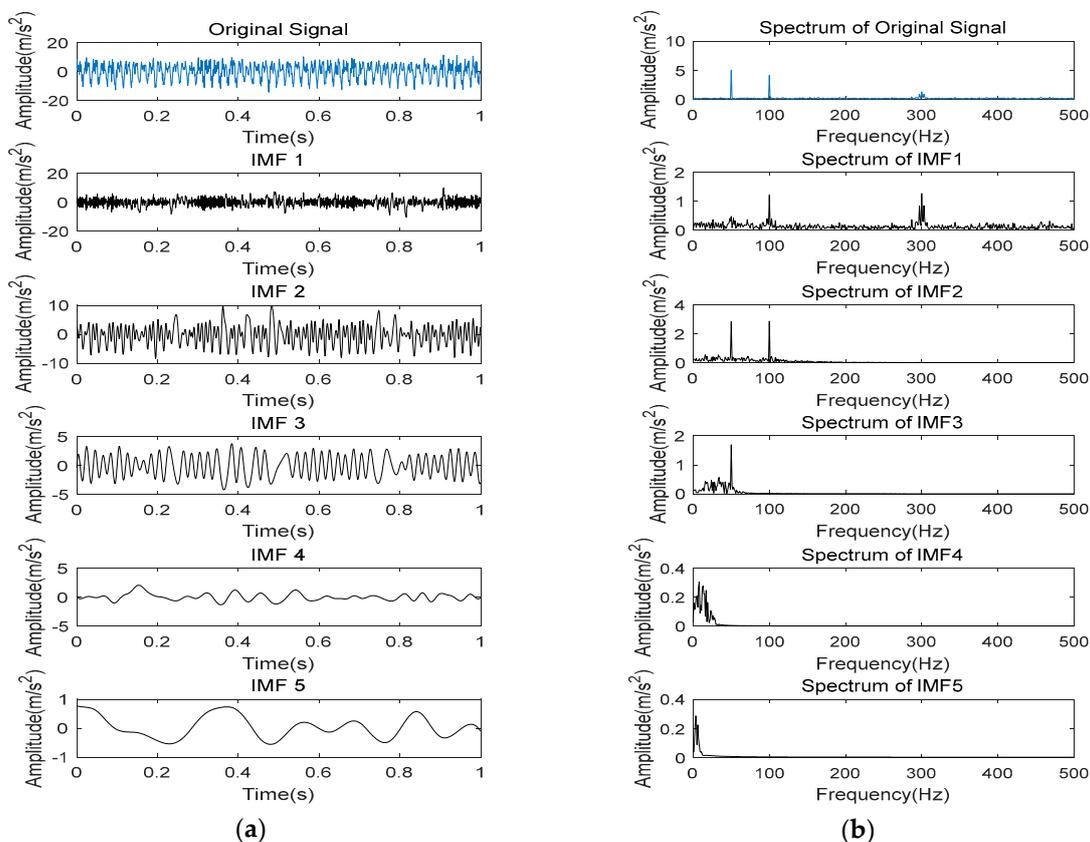


**Figure 5.** The decomposition results of empirical mode decomposition (EMD): (a) Time-domain; (b) Frequency-domain.

It can be seen from Figures 5 and 6 that EMD decomposed the signal into 9 components, while the actual signal was composed of 3 components and noise, and there was serious modal mixing. LMD decomposed the signal into 5 components, and IMF1 and IMF2 contained 2 different center frequencies. In Figure 7, EEMD decomposed the signal into 6 signals and improved the problem of modal mixing. However, by observing the spectrum of IMF4, it was found that EEMD still did not eliminate the mode mixing. Moreover, the 3 methods all over decomposed the signal.

Next, the VMD method optimized in this paper was used to decompose the signal. Since signal  $f(t)$  is composed of 3 signals and white noise, the decomposition parameter  $K$  of VMD should

be 3. Then, the original GWO algorithm and FGWO algorithm were used to optimize the VMD decomposition parameters.



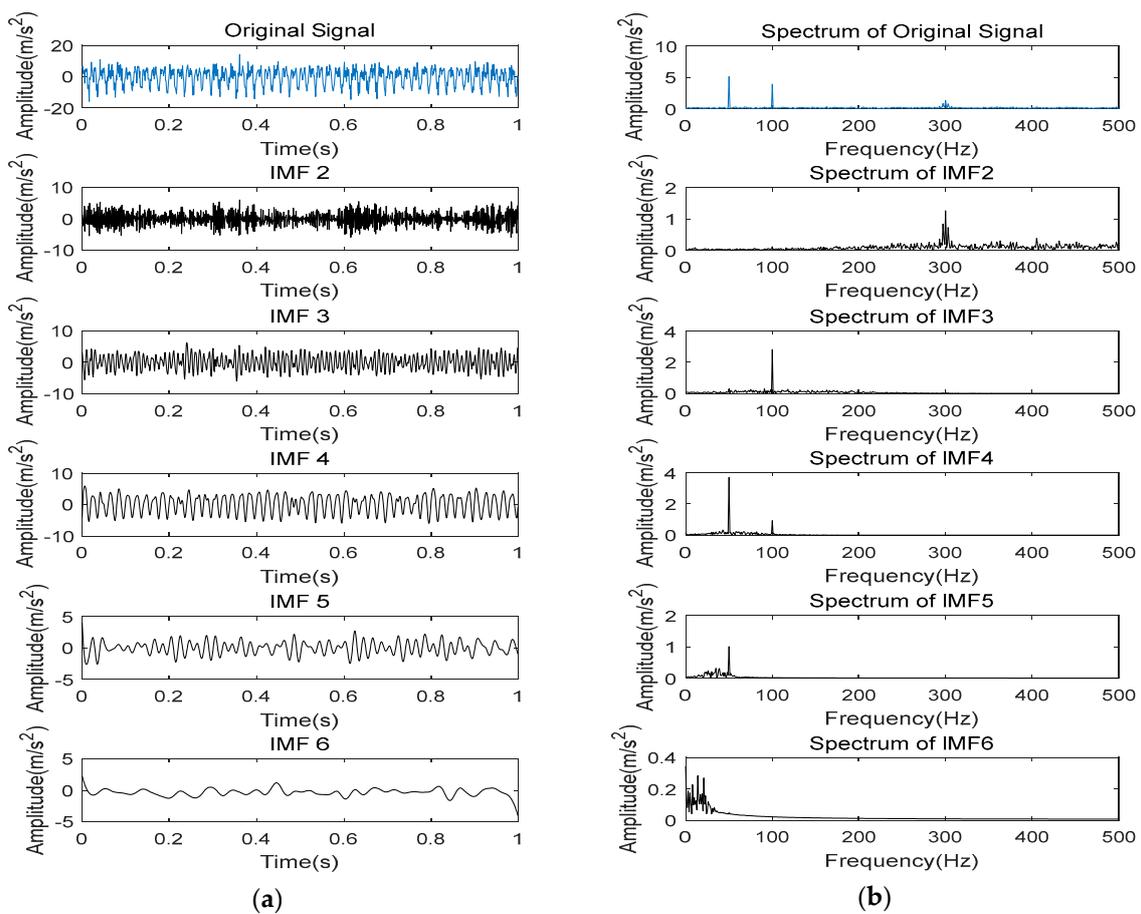
**Figure 6.** The decomposition results of local mean decomposition (LMD): (a) Time-domain; (b) Frequency-domain.

The number of grey wolves in the groups of GWO and FGWO was set at 100, and the maximum number of iterations was set at 10. The decomposition parameter of VMD was the position vector of the grey wolves. Figure 8 shows the convergence curve of the 2 optimization algorithms.

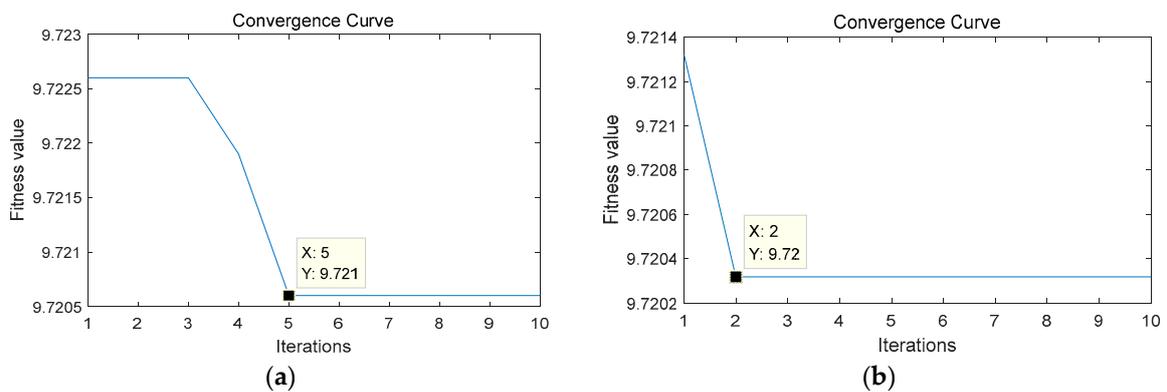
It can be seen from the convergence curve in Figure 8 that the fitness value of GWO reached the minimum value of 9.721 in the 5th iteration, while that of FGWO converged to the minimum value of 9.72 in the 2nd iteration. Therefore, both FGWO and GWO can converge to the global minimum, and FGWO does thus in less time. The position vector corresponding to MAEE was (3, 1126), which was used to decompose the simulation signal  $f(t)$ . The IMF components and their spectra are shown in Figure 9.

It was easy to see that under the optimal parameters obtained from FGWO, VMD decomposed the simulation signal into 3 IMF components with 50 Hz, 100 Hz, and 300 Hz as the center frequency, that was, the simulation signal  $f(t)$  was decomposed into  $f_1(t)$ ,  $f_2(t)$ , and  $f_3(t)$ , effectively. The problem of modal mixing was eliminated. Moreover, VMD eliminated some of the interference from the white noise  $f_4(t)$ .

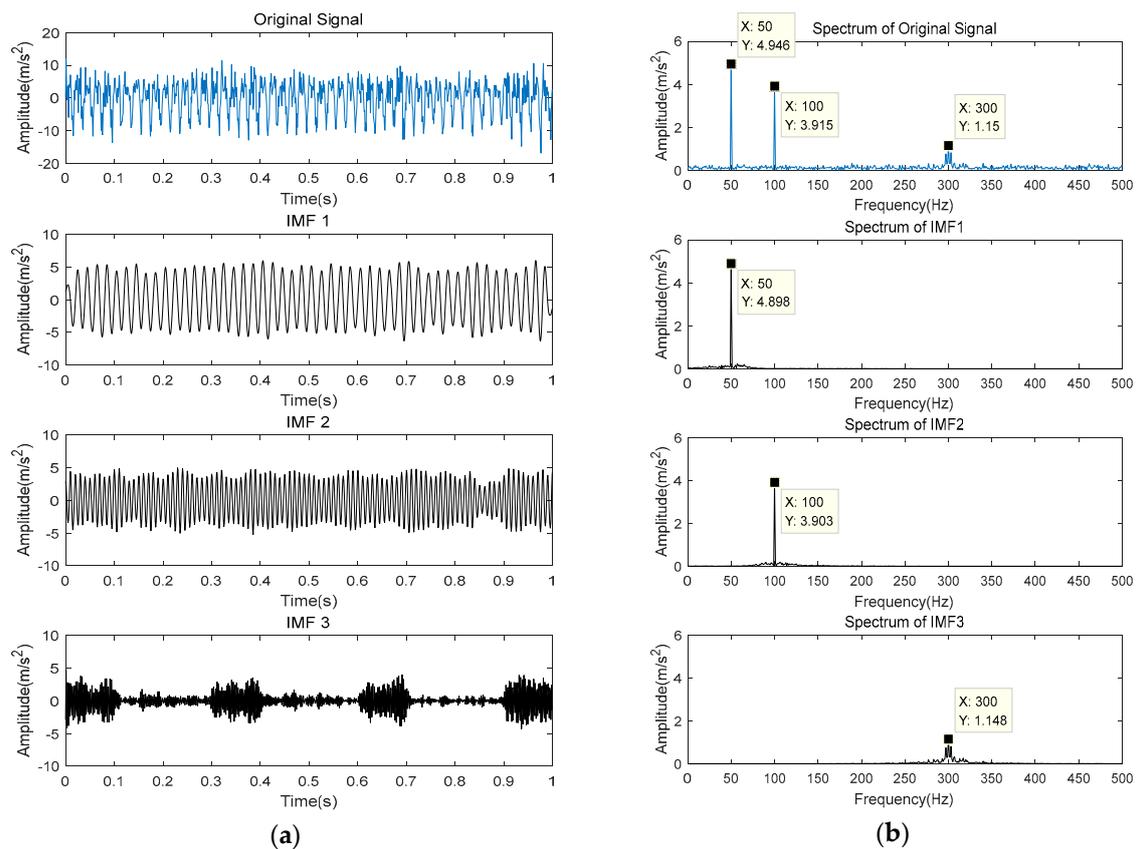
It can be seen that VMD had great advantages in signal processing. In this paper, FGWO was introduced to optimize VMD, thus that VMD can adaptively determine the optimal decomposition parameters, get the better result of decomposition.



**Figure 7.** The decomposition results of ensemble empirical mode decomposition (EEMD): (a) Time-domain; (b) Frequency-domain.



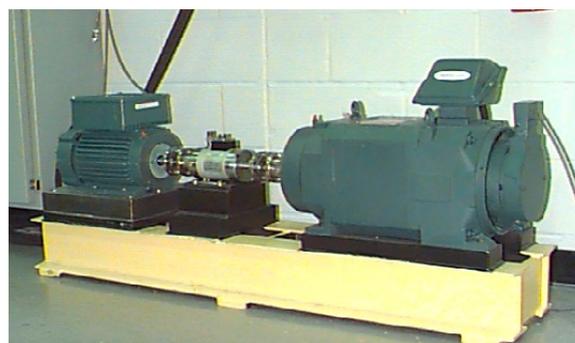
**Figure 8.** Convergence curve of optimization algorithm. (a) grey wolf optimizer; (b) fast grey wolf optimizer.



**Figure 9.** Intrinsic mode factors (IMF) components of optimal parameters of VMD: (a) Time-domain; (b) Frequency-domain.

## 5. Experimental Analysis

In this section, the proposed method was applied to process and analyze real data from the laboratory, of which the source was the bearing data center of Case Western Reserve University, as shown in Figure 10.



**Figure 10.** Test platform.

Table 1 shows the MATLAB data information collected under the normal bearing. Vibration data were collected through accelerometers, which were attached to the housing with magnetic bases. In addition, signals were collected through a 16 channel DAT recorder, and were post-processed in a MATLAB environment.

The data processed and analyzed in this section were the bearing data from the drive end. The following 3 cases of fault diagnosis are studied:

- (a) Diagnosis of different faults. The 4 kinds of data used were normal data and fault data of the inner race, ball, and outer race when the rotating speed was 1750 rpm and the load was 2 hp;
- (b) Diagnosis of the same fault under different working conditions. The 4 kinds of data used were the fault data of the inner race with the same fault degree under 4 speeds and loads;
- (c) Diagnosis of different fault degrees. The 4 kinds of data used were the fault data of the inner race under the same working conditions and 4 fault degrees. The fault diameters (fault degrees) were 0.01778 cm, 0.03556 cm, 0.05334 cm, and 0.07112 cm.

**Table 1.** Gathering information on normal bearing.

Data Name	Motor Load (HP)	Motor Speed (rpm)	Sampling Frequency (HZ)
97.mat	0	1797	12,000
98.mat	1	1772	
99.mat	2	1750	
100.mat	3	1730	

The specific fault data information is shown in Table 2. Each type of data was divided into 50 groups, with 2048 sampling points in each group. The first 30 groups were used for training, and the last 20 groups were used for testing.

**Table 2.** Specific information of fault diagnosis.

	TYPE	Motor Load (HP)	Motor Speed (rpm)	Fault Diameter (cm)
A	Normal	2	1750	0.01778
	Inner Race			
	Ball			
	Outer Race			
B	Inner Race	0	1797	0.01778
		1	1772	
		2	1750	
		3	1730	
C	Inner Race	2	1750	0.01778
				0.03556
				0.05334
				0.07112

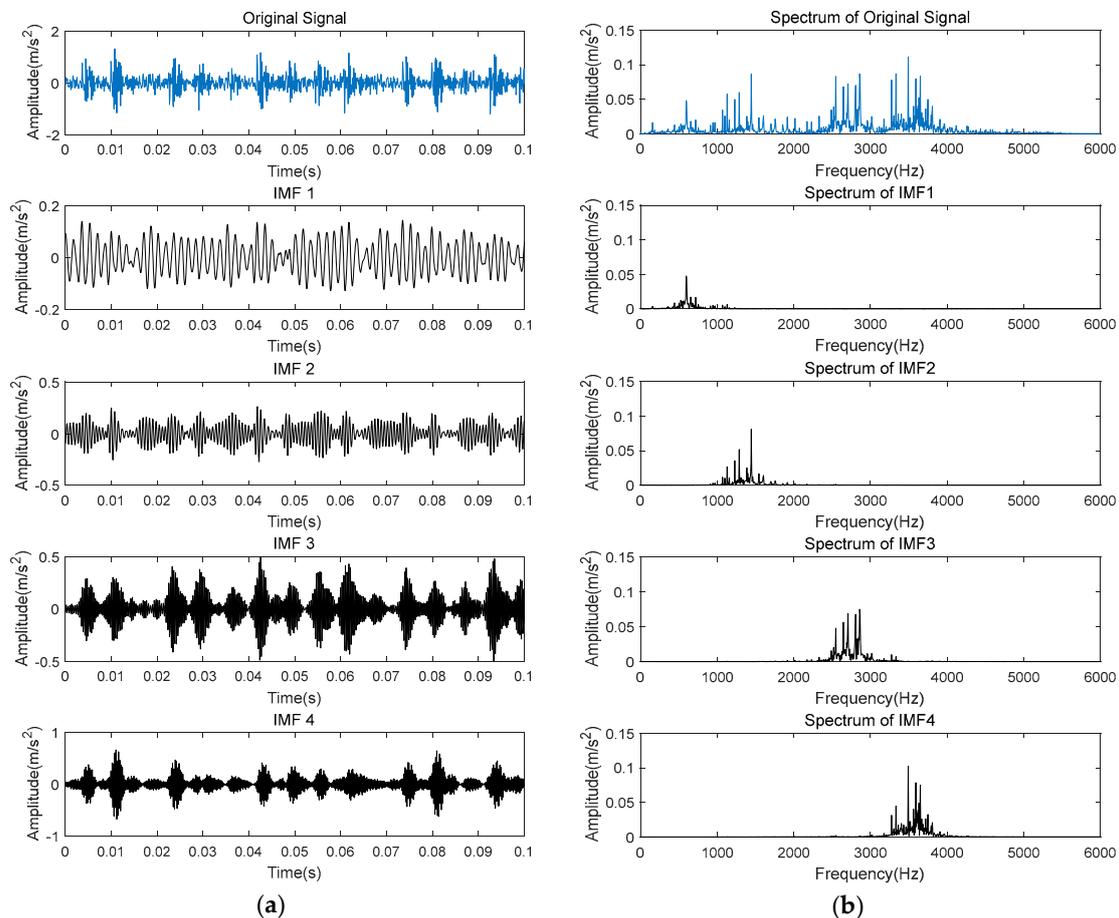
### 5.1. Diagnosis of Different Faults

The data used in this subsection are shown in A of Table 2.

First, FGWO-VMD decomposition and FFT were performed on the signal to obtain the time and frequency domain information of  $K$  IMF components. Taking the inner race fault as an example, the best decomposition parameter combination of VMD optimized by FGWO was (4, 2000), and the time-domain and frequency-domain information is shown in Figure 11. VMD decomposed the signal into  $K$  simple and easy-to-analyze IMF components and eliminated some noise interference. Each IMF component can jointly represent the characteristics of the original signal.

In Figure 11, the time-domain information of IMFs was decomposed into singular values, and the vector  $T1$  composed of  $K$  singular values was obtained. PCA was performed on  $T1$ , as shown in Figure 12. It can be seen from Figure 12 that the contribution rate of the first feature value of  $T1$  after PCA was very high, thus it was selected as the feature value  $t_1$  of the time-domain information to realize the dimensionality reduction of the time-domain feature vector. Similar to the extraction method of the time-domain feature values, the vector  $T2$  composed of  $K$  singular values and the  $T3$  composed of  $K$  values of PE were obtained, and PCA was carried out for them, as shown in Figures 13 and 14. It was easy to see from Figure 13 that the first feature value of  $T2$  after PCA was selected as

the feature value  $t_2$  of the frequency-domain information to complete the dimensionality reduction of the frequency-domain feature vector. It can be seen from Figure 14 that after the feature vector of the time-frequency domain was processed by PCA, the contribution rate of the first 2 values reached 80%. Therefore, these values were selected as the feature values  $t_3$  and  $t_4$  to represent the characteristics of the signal in the time-frequency domain.



**Figure 11.** The intrinsic mode functions (IMF) information of inner race fault: (a) Time-domain; (b) Frequency-domain.

Now, the multi-domain feature extraction and dimensionality reduction of the signal were complete, and next, the time-domain, frequency-domain, and time-frequency-domain feature values were synthesized to form the feature vector  $T = [t_1, t_2, t_3, t_4]$ , which can fully represent the characteristics of the signal.

Finally, according to this multi-domain feature extraction method, the feature vectors of 50 groups of signals of each fault were extracted, and a total of 200 feature vectors were obtained. A total of 120 feature vectors were randomly selected to train the ELM, and then 80 feature vectors were randomly selected as the signals to be tested, which are input into the ELM after training. Then, the feature vectors can be classified to realize the diagnosis of different faults. The results of the diagnosis are shown in Figure 15.

In Figure 15, 1 represents normal, 2 represents the inner race fault, 3 represents the ball fault, and 4 represents the outer race fault. The correct rate of the ELM training was 95%. When fault diagnosis was carried out in the trained ELM, the correct rate was as high as 98.75%. Only one error occurred in 80 test samples, that was, an inner race fault was misjudged as a ball fault.

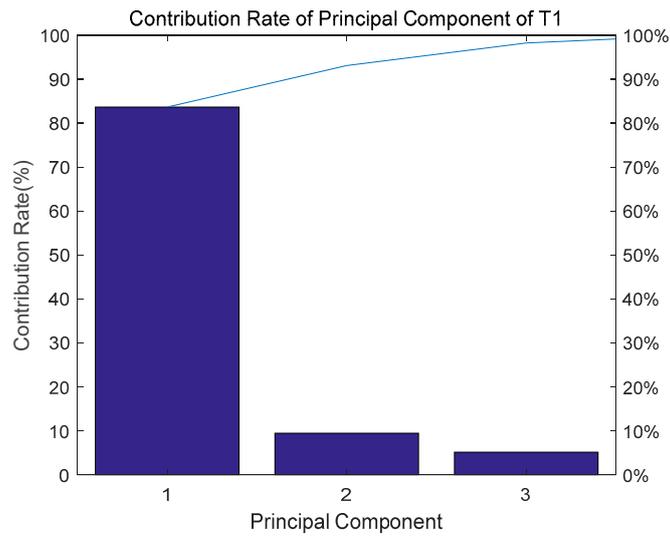


Figure 12. Principal component analysis (PCA) of the feature vector in time-domain.

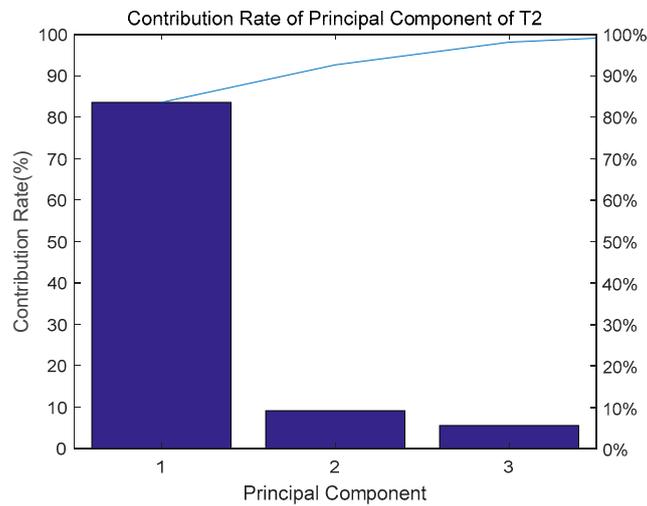


Figure 13. Principal component analysis (PCA) of the feature vector in frequency-domain.

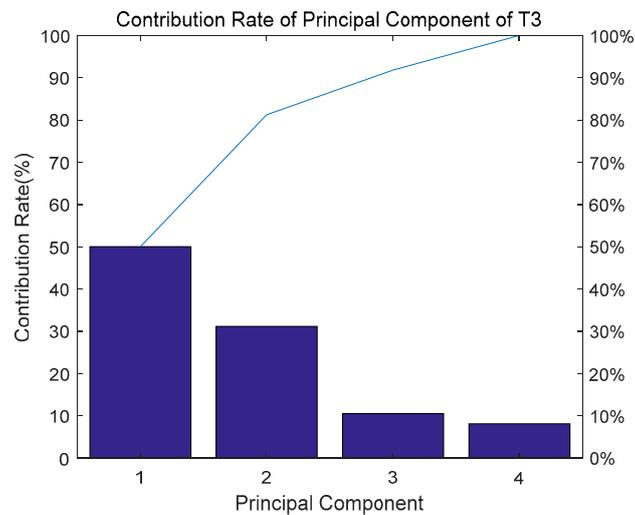
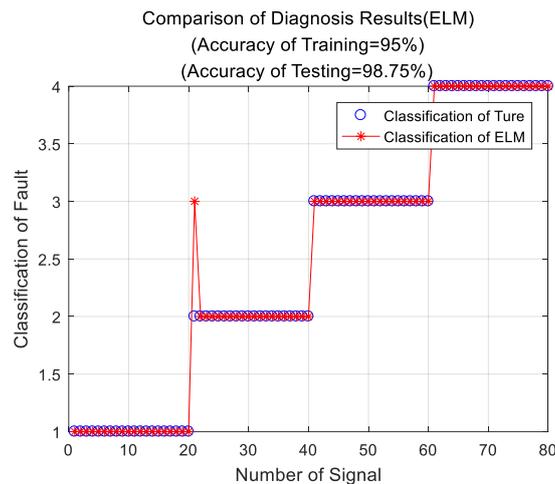


Figure 14. Principal component analysis (PCA) of the feature vector in time-frequency-domain.

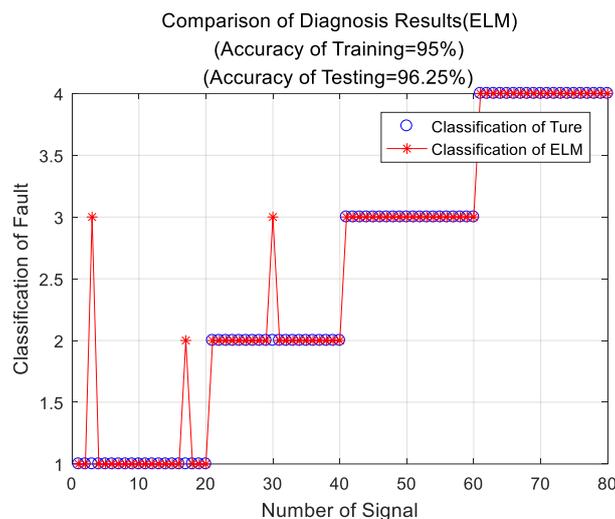


**Figure 15.** Diagnosis results of case A.

### 5.2. Diagnosis under Different Working Conditions

The data used in this subsection are shown in B of Table 2.

For the method in Section 5.1 and under the same fault of the inner race and different working conditions, the results of diagnosis are shown in Figure 16.



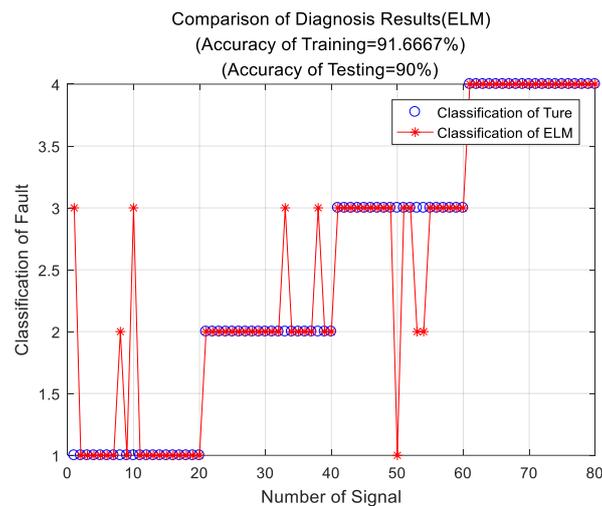
**Figure 16.** Diagnosis results of case B.

As shown in Figure 16, for the same fault under different working conditions, the method proposed in this paper was applied to feature extraction and classification, and only 3 misjudgments appeared in 80 test samples. The effect was still considerable, and the diagnostic accuracy was as high as 96.25%. This result showed that the decomposition effect of adaptive VMD was great. Combined with the subsequent work of the multi-domain feature extraction, all the original signal features were extracted comprehensively, thus increasing the accuracy of diagnosis.

### 5.3. Diagnosis of Different Fault Degrees

The data used in this subsection are shown in C of Table 2.

The steps in Section 5.1 were followed to obtain the diagnosis results of different fault degrees, as shown in Figure 17. It can be seen from Figure 16, that although the classification effect was not as good as those for cases A and B, when diagnosing the same fault of different degrees, the accuracy rate of diagnosis was 90%.



**Figure 17.** Diagnosis results of case C.

In order to further illustrate the effectiveness of the proposed method, EEMD was combined with 4 pattern recognition methods: SVM, genetic algorithm back propagation (GA-BP) neural network, ELM, and deep convolutional neural networks (Deep-CNN) to obtain accuracy rate of diagnosis in 3 cases. The results are shown in Table 3.

**Table 3.** The accuracies of five fault diagnosis methods.

Method	Case A	Case B	Case C
EEMD-SVM	0.984	0.962	0.804
EEMD-GABP	0.961	0.938	0.721
EEMD-ELM	0.973	0.956	0.791
EEMD-DeepCNN	0.978	0.955	0.821
AVMD-ELM	0.988	0.963	0.900

It can be seen from Table 3 that in case A and B, because of the multi-domain feature extraction method proposed in this paper, the accuracies of 5 fault diagnosis methods were high. Therefore, the improvement of AVMD-ELM was not obvious. In case C, affected by signal processing methods, the accuracies of the 4 fault diagnosis methods were low. Due to the improvement of signal processing methods, the fault accuracy of AVMD-ELM was far higher than those of the other 4 methods.

## 6. Conclusions

There are two major contributions to fault diagnosis in this paper. In the view of the signal processing method before feature extraction, this paper optimizes the VMD algorithm; in the view of feature extraction, this paper proposes a multi-domain fault diagnosis method. Through these two improvements, the accuracy of fault diagnosis is improved. The conclusions are as follows:

- It can be seen from the experiment that VMD can eliminate the problem of modal mixing. Through the optimization of this paper, VMD can also adaptively determine the optimal decomposition parameters to obtain better decomposition effect;
- In view of the problem of signal feature extraction, PCA is carried out thus that information redundancy is eliminated, the feature vectors representing each domain become more concise, and the features of each domain are more prominent thus that the accuracy of the subsequent classification is greatly enhanced;
- In view of the problem of fault diagnosis accuracy, this paper starts from the fundamental problem that affects the accuracy of classification, synthesizes the extracted three-domain features to form

a multi-domain feature vector that can comprehensively represent the fault characteristics of signals, and greatly improves the accuracy of fault diagnosis.

Finally, there are still some improvements that can be made in this method. Because of the instability of wind speed and the poor operating environment of the fan, the fan sometimes runs in a condition of variable speed. Therefore, on the basis of the research work in this paper, the technology of order tracking can be introduced to realize the automatic diagnosis of gearbox fault under variable conditions, broadening the application range of this method.

**Author Contributions:** Conceptualization, H.L. and B.F.; methodology, R.J.; validation, H.L., B.F., and R.J.; formal analysis, L.B.; investigation, X.L.; resources, X.L.; writing—original draft preparation, B.F.; writing—review and editing, H.L.; writing—translation and sentence checking, F.Z.; visualization, L.B. and F.Z.; supervision, R.J.; project administration, H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (grant number 51779206) and Scientific Research Program funded by the Shanxi Provincial Education Department (Program No.17JK0570).

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

## References

- Feng, Z.; Qin, S.; Liang, M. Time–frequency analysis based on Vold-Kalman filter and higher order energy separation for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. *Renew. Energy* **2016**, *85*, 45–56. [[CrossRef](#)]
- Strangas, E.; Aviyente, S.; Zaidi, S. Time–Frequency Analysis for Efficient Fault Diagnosis and Failure Prognosis for Interior Permanent-Magnet AC Motors. *IEEE Trans. Ind. Electron.* **2008**, *55*, 4191–4199. [[CrossRef](#)]
- Tang, B.; Liu, W.; Song, T. Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-ville distribution. *Renew. Energy* **2010**, *35*, 2862–2866. [[CrossRef](#)]
- Li, H.; Li, H.; Huang, G.; Ou, J. Early Fault Diagnosis for Planetary Gearbox Based on Adaptive Parameter Optimized VMD and Singular Kurtosis Difference Spectrum. *IEEE Access* **2019**, *7*, 31501–31516.
- Chen, B.; Zhang, Z.; Sun, C.; Li, B.; Zi, Y.; He, Z. Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors. *Mech. Syst. Signal Process.* **2012**, *33*, 275–298. [[CrossRef](#)]
- Gao, Q.; Liu, W.; Tang, B.; Li, G. A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine. *Renew. Energy* **2018**, *116*, 169–175. [[CrossRef](#)]
- Ben Ali, J.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **2015**, *89*, 16–27. [[CrossRef](#)]
- Zhang, X.; Liang, Y.; Zhou, J.; Zang, Y. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement* **2015**, *69*, 164–179. [[CrossRef](#)]
- Shen, Z.; Chen, X.; Zhang, X.; He, Z. A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM. *Measurement* **2012**, *45*, 30–40. [[CrossRef](#)]
- Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
- Lei, Y.; Lin, J.; He, Z.; Zuo, M.J.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [[CrossRef](#)]
- Junsheng, C.; Dejie, Y.; Yü, Y. The application of energy operator demodulation approach based on EMD in machinery fault diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 668–677. [[CrossRef](#)]
- Smith, J.S. The local mean decomposition and its application to EEG perception data. *J. R. Soc. Interface* **2005**, *2*, 443–454. [[CrossRef](#)] [[PubMed](#)]
- Wu, Z.; Huang, N.E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]

15. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [[CrossRef](#)]
16. Li, Y.; Cheng, G.; Liu, C.; Chen, X. Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks. *Measurement* **2018**, *130*, 94–104. [[CrossRef](#)]
17. Li, Z.; Jiang, Y.; Guo, Q.; Hu, C.; Peng, Z. Multi-dimensional variational mode decomposition for bearing-crack detection in wind turbines with large driving-speed variations. *Renew. Energy* **2018**, *116*, 55–73. [[CrossRef](#)]
18. Paternina, M.R.A.; Tripathy, R.K.; Zamora-Mendez, A.; Dotta, D. Identification of electromechanical oscillatory modes based on variational mode decomposition. *Electr. Power Syst. Res.* **2019**, *167*, 71–85. [[CrossRef](#)]
19. Zhang, X.; Miao, Q.; Zhang, H.; Wang, L. A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery. *Mech. Syst. Signal Process.* **2018**, *108*, 58–72. [[CrossRef](#)]
20. Yan, X.; Jia, M. Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings. *Mech. Syst. Signal Process.* **2019**, *122*, 56–86. [[CrossRef](#)]
21. Yang, K.; Wang, G.; Dong, Y.; Zhang, Q.; Sang, L. Early chatter identification based on an optimized variational mode decomposition. *Mech. Syst. Signal Process.* **2019**, *115*, 238–254. [[CrossRef](#)]
22. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2005; Volume 2, pp. 985–990.
23. Tian, Y.; Ma, J.; Lu, C.; Wang, Z. Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mech. Mach. Theory* **2015**, *90*, 175–186. [[CrossRef](#)]
24. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [[CrossRef](#)]
25. Guo, Z.; Liu, R.; Gong, C.; Zhao, L. Research on the improvement of Gray Wolf algorithm. *Appl. Res. Comput.* **2017**, *34*, 3603–3606.
26. Li, Y.; Li, G.; Wei, Y.; Liu, B.; Liang, X. Health condition identification of planetary gearboxes based on variational mode decomposition and generalized composite multi-scale symbolic dynamic entropy. *ISA Trans.* **2018**, *81*, 329–341. [[CrossRef](#)]
27. Li, J.; Tang, W.; Wang, J.; Zhang, X. Multilevel thresholding selection based on variational mode decomposition for image segmentation. *Signal Process.* **2018**, *147*, 80–91. [[CrossRef](#)]
28. Yao, J.; Xiang, Y.; Qian, S.; Wang, S.; Wu, S. Noise source identification of diesel engine based on variational mode decomposition and robust independent component analysis. *Appl. Acoust.* **2017**, *116*, 184–194. [[CrossRef](#)]
29. Randall, R.; Antoni, J. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [[CrossRef](#)]
30. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)]
31. Huang, N.E.; Shen, Z.; Long, S.R. A new view of nonlinear water waves: The Hilbert Spectrum. *Annu. Rev. Fluid Mech.* **1999**, *31*, 417–457. [[CrossRef](#)]

