Hindawi
Shock and Vibration
Volume 2019, Article ID 8213056, 20 pages
https://doi.org/10.1155/2019/8213056

Research Article

Fault Diagnosis of Variable Load Bearing Based on Quantum Chaotic Fruit Fly VMD and Variational RVM

Bo Xu,[1,2] Huipeng Li,[1,2] Fengxing Zhou,[1] Baokang Yan,[2] Yi Liu,[2,3] and Yajie Ma[1]

[1]*Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China*
[2]*School of Electronic Information, Huang Gang Normal University, Huanggang 438000, China*
[3]*School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China*

Correspondence should be addressed to Huipeng Li; yeshuip@163.com

Received 13 June 2018; Revised 11 November 2018; Accepted 29 November 2018; Published 3 January 2019

Academic Editor: Franck Poisson

Under normal circumstances, bearings generally run under variable loading conditions. Under such conditions, the vibration signals of the bearing malfunctions are often nonstationary signals, which are difficult to process effectively. In order to accurately and effectively diagnose the failure types and damage degree of bearings under variable load conditions, an intelligent diagnostic model based on the variational mode decomposition (VMD) of quantum chaotic fruit fly optimization algorithm (QCFOA) and a multiclassification variational relevance vector machine (VRVM) is proposed. First, the key parameters of the VMD are selected using the QCFOA. Secondly, the known bearing fault signal is decomposed by the optimized VMD, and the center frequency and marginal spectral entropy (MSE) of each natural modal component are extracted to construct two-dimensional MSE. Then, the probit model is used to replace the logistic model, and a simpler and more practical multiclassification model is constructed. The two-dimensional MSE of each intrinsic modal component is used as a learning sample for VRVM. Finally, the bearing fault data under 1 hp load are taken as training samples, and the bearing fault data under two loads of 0 hp and 3 hp are used as test samples to verify the effectiveness of the intelligent diagnosis model. The experimental results show that the intelligent fault diagnosis method proposed in this paper can accurately diagnose the type of fault and the degree of damage and has high robustness.

1. Introduction

Rolling bearings, which are the core components of rotating machinery, operate under harsh conditions and are easily damaged. It is necessary to monitor the fault. Once a fault occurs, it may cause huge economic losses and personal safety problems. In recent years, with the continuous update of knowledge in the field of digital signal processing and machine learning, intelligent fault diagnosis technology has become a major development trend [1]. Intelligent diagnosis is essentially a pattern recognition process, including two important steps of fault feature extraction and fault identification. In particular, how to effectively extract weak fault characteristics in the fault signal is key to fault diagnosis.

The vibration signal analysis is widely used in the fault diagnosis of bearing. In general, the vibration signals of the rolling bearing fault are mostly nonstationary signals, and the method suitable for processing nonstationary signals should be adopted. Empirical mode decomposition (EMD) [2] as a powerful tool of nonstationary signal processing has received extensive attention from researchers concerned with mechanical fault diagnosis. The method of bearing fault feature extraction based on EMD has been widely applied [3–5]. Inspired by the EMD method, Smith et al in [6] proposed another adaptive signal decomposition method named the "local mean decomposition (LMD)" in 2005, which has aroused the attention of a large number of scholars [7]. The EMD and LMD are excellent self-adaptive processing methods for nonstationary and nonlinear signals. However, they have unavoidable deficiencies, such as the end effect, overshoot, and mode mixing. Recently, many scholars proposed

optimization and improvement methods for such deficiencies. However, both of them were regarded as recursive mode decomposition methods, and inherent defects are difficult to solve fundamentally. Thus, a new adaptive signal processing method named the "variation mode decomposition (VMD)" was proposed by Dragomiretskiy and Zosso [8]. The method abandons the recursive decomposition mode and effectively alleviates or avoids a series of deficiencies in the EMD and LMD methods.

The excellent features presented by VMD have been used by scholars in the field of fault diagnosis [9]. However, an important feature of VMD is that the number of intrinsic modal components and their penalty parameters must be set in advance; once unsuitable parameter values are selected, the decomposition results will be seriously affected. Therefore, against the selection of key parameter values of VMD, some scholars introduced the particle swarm optimization algorithm to select the key parameters of VMD optimally and achieved certain results [10]. However, the traditional heuristic optimization algorithms, such as particle swarm optimization (PSO) [11], genetic algorithm (GA) [12], ant colony optimization (ACO) [13], and cuckoo algorithm (CA) [14], have some problems such as parameter dependence, computational complexity, convergence speed, and optimization accuracy, which restrict their practical applications. To solve this problem, this paper introduces the fruit fly optimization algorithm (FOA) [15] to optimize the key parameters of VMD. This algorithm is a new swarm intelligence algorithm based on the bionics principle of fruit fly foraging behavior. It is applied in many fields [16–18]. However, its convergence accuracy is very sensitive to the initial value. Once the initial value is not selected properly, the search is likely to fall into a local optimum, and the convergence accuracy is low. This paper combines the quantum logistic chaotic mapping [19] and FOA and propose a quantum chaotic fruit fly algorithm in the three-dimensional space search, which strengthens the ergodicity, avoids the search process falling into the local optimal value, and improves the search efficiency. Then, using the local minimum value of the MSE as the fitness function, the quantum chaotic fruit fly optimization algorithm (QCFOA) is used to search two key parameters of VMD and obtain the optimal combination value of the key parameters. Furthermore, the optimized VMD is used to process the known fault signals, and the effective intrinsic mode function (IMF) component and its MSE are obtained.

The traditional diagnosis method usually trains the fault diagnosis model under a certain load and has great limitations in diagnosing the state of equipment under a certain load. In practical applications, many mechanical devices work under variable load conditions. Both the load and the damage degree will affect the amplitude of the fault characteristic frequency of the bearing. Therefore, the single MSE cannot effectively characterize the damage degree of the fault in the variable load condition (the radial load is mainly discussed in this paper). Because the bearing is running under different load conditions, the normal contact load between the rolling body and the raceway will change, which leads to the change in the natural vibration frequency of the bearing. To address these challenges, the center frequency is introduced in this paper, and the one-dimensional MSE is extended into two-dimensional MSE as the learning sample of the variational correlation vector machine. Then, the method is used to identify the various types of faults and the damage degree of the bearing under variable load.

The technology of intelligent diagnosis for mechanical faults is changing fast. Artificial neural network (ANN) and support vector machine (SVM), as intelligent recognizers, have received the most extensive attention, and a multitude of recent research efforts have been made to explore the mechanical fault diagnosis. However, the ANN algorithm requires a large number of training samples and has some inherent problems, such as black box operation, low generalization ability, and overlearning [20–24]. Similarly, the SVM also has some unavoidable deficiencies. The method cannot get the probabilistic prediction and the uncertainty in prediction [25–29].

RVM is a new machine learning algorithm based on support vector machine (SVM) and Bayesian theory framework [30]. Compared with the SVM method, it can directly give the uncertainty of the result while giving the diagnosis results. The RVM training process needs fewer parameters, and its solution is more sparse [31]. So, the probability output of RVM accords with the actual mechanical fault diagnosis process and has high applied research value. However, when the standard RVM is very limited in the size of the data sample, the computational cost of the training is very high. For this problem, Bishop [32] proposed a method of computing and solving RVM by means of the variational method, named the "VRVM." In the case of very limited data samples, this method is better than the type II maximum edge likelihood estimation and can give the posterior distribution of parameters and hyperparameters. Compared with the standard RVM, the practicability and performance of the VRVM are better. In addition, VRVM is the same as the standard RVM, and its classification and regression are all mapped by logistic function. Therefore, when the regression problem is converted into a classification problem, the noise variable must be ignored, and the true value of the model cannot be accurately estimated. Therefore, the probit model is used instead of the logistic model in this paper, which makes the classification problem and the regression problem organically combined to avoid the approximate derivation of the logistic model from the continuous output to the discrete output mapping and to reduce the amount of computation.

Finally, the proposed method is used to diagnose the variable load fault data collected from the failure platform to verify the effectiveness and robustness of the method.

2. The Principle of VMD

In the VMD algorithm, the IMF is redefined as an AM-FM signal, which removes the loop iteration method used by the

EMD algorithm. Instead, the signal decomposition process is transferred to the variational structure. By constructing and solving the constrained variational problem and decomposing the original signal into a specified number of IMF components, the construction process of the corresponding variational problem is summarized as follows.

For each IMF component $u_k(t)$, the following analytic signal is obtained through the Hilbert transform:

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

For each analytic signal, a central frequency ω_k is estimated, and the frequency of each analytical signal is transformed to the baseband by shifting frequency:

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

The Gaussian smoothing index of the frequency shift signal is used to estimate the bandwidth of each IMF component, and then the corresponding constraint variational model is expressed as

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}}\left\{\sum\limits_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\}, \\ \text{s.t.}\sum\limits_k u_k=f, \end{cases} \tag{3}$$

where $\{u_k\}=\{u_1,u_2,\ldots,u_k\}$ represents the decomposition of the IMF component, k is the number, $\{\omega_k\}=\{\omega_1,\omega_2,\ldots,\omega_k\}$ represents the frequency center of each component, f is the input signal, and $e^{-j\omega_k t}$ is the estimated center frequency. In order to obtain the abovementioned constraint variation problem, the augmented Lagrange function is introduced as follows:

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda)=\alpha\sum\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega t}\right\|_2^2 \\ +\left\|f(t)-\sum_k u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_k u_k(t)\right\rangle, \end{aligned} \tag{4}$$

where α is the quadratic penalty parameter and λ is the Lagrange multiplier. The saddle point of the augmented Lagrange function is obtained by using the alternating direction multiplier algorithm, which is the optimal solution of equation (2) constraint variational model, and the original signal is decomposed into K narrowband IMF components. The solution procedure for the variational model is as follows (Algorithm 1).

It is known from the variational model solution process that the performance of VMD is closely related to the decomposition parameters, such as the total number of modalities K and the secondary penalty α.

The performance of VMD is very sensitive to the value of K. If the value of K is too small, the data will be undersegmented and some components will be contained in other modalities, and if the value of K is too large, problems such as modal copying will occur. If the value of α is too small, the bandwidth of the modal component will be too large; some components will be included in other modal components, or additional "noise" will be captured; if the value of α is too large, the bandwidth of the modal component will be too small and some components in the original signal will be lost.

Therefore, the design of an optimal VMD should be focused on how to obtain the optimal combination value of key parameters of the VMD.

3. VMD Optimization Based on Quantum Chaotic FOA

The VMD algorithm needs to set the number of IMF components in advance when processing signals. Different K values will result in different decomposition results. In addition, the penalty parameter α also has a great influence on the decomposition result. The smaller the α, the larger the bandwidth of each IMF component; conversely, the bandwidth of the component signal is smaller. Therefore, parameters K and α are two critical parameters that affect VMD performance. In practical application, the components of the bearing fault signal are very complex, and the selection of suitable parameters K and α is key to the effective extraction of bearing fault characteristics by the VMD algorithm. Meanwhile, α and K have interactive effects on the performance of VMD. Therefore, a swarm intelligence algorithm which can optimize parameters α and K is needed to avoid the contingency and blindness of manually setting parameters.

3.1. Quantum Chaotic Mapping. Quantum chaotic system has the natural properties of the classical chaotic system [33]. For the same classical chaotic system, different quantitative criteria can produce different quantum chaotic maps. In the work of Goggin [34], the classical logistic system is quantified by the recoil rotor model, and the corresponding quantum logistic mapping is obtained. The definition is as follows:

$$\begin{cases} x_{n+1}=\mu\left(x_n-\left|x_n\right|^2\right)-\mu y_n, \\ y_{n+1}=-y_n e^{-2\beta}+e^{-\beta}\mu\left[\left(2-x_n-\widetilde{x}_n\right)y_n-x_n\widetilde{z}_n-\widetilde{x}_n z_n\right], \\ z_{n+1}=-z_n e^{-2\beta}+e^{-\beta}\mu\left[2\left(1-\widetilde{x}_n\right)z_n-2x_n y_n-x_n\right], \end{cases} \tag{5}$$

where μ is a chaos control parameter, β is a dissipative parameter, x_n, y_n, and z_n are the state values of the system, and $\widetilde{x}_n$ and $\widetilde{z}_n$ are complex conjugation of x_n and z_n, respectively. If the initial value in the chaotic system is real, then the chaotic sequence generated by the system is real, and there are $\widetilde{x}_n=x_n$ and $\widetilde{z}_n=z_n$. In [35–37], it is proved that the pseudorandom sequence based on the quantum chaotic mapping not only has all the advantages of the traditional chaotic system but also has a weaker correlation and stronger ergodicity than the traditional chaotic system.

Initialize:$\{\hat{u}_k^1\}$, $\{\omega_k^1\}$, $\hat{\lambda}^1$, $n \longleftarrow 0$
repeat:
$n \longleftarrow n+1$
for $k = 1 : K$ **do**
Update $\hat{u}_k$ for all $\omega \geq 0$:
$\hat{u}_k^{n+1}(\omega) \longleftarrow \hat{f}(\omega) - \sum_{i<k}\hat{u}_i^{n+1}(\omega) - \sum_{i>k}\hat{u}_i^n(\omega) + (\hat{\lambda}^n(\omega)/2)/1 + 2\alpha(\omega - \omega_k^n)^2$
Update ω_k:
$\omega_k^{n+1} \longleftarrow \int_0^\infty \omega|\hat{u}_k^{n+1}(\omega)|^2 d\omega / \int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega$
end for
Dual ascent for all $\omega \geq 0$:
$\hat{\lambda}^{n+1}(\omega) \longleftarrow \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega))$
until convergence: $\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \varepsilon$

Algorithm 1: Complete optimization of VMD.

3.2. Quantum Chaotic FOA. The advantage of the FOA is that it is easy to understand, simple to search, and easy to implement. Therefore, it is widely used in parameter optimization problems [38, 39]. In this paper, a quantum logistic chaotic map is used to extend the search space of the FOA into a three-dimensional search space, and the location of the fruit fly group is initialized by the better characteristics of nonperiodicity, ergodicity, and class randomness and more sensitivity to system parameters and initial conditions than traditional chaotic systems. The method improves the diversity of population and strengthens the ergodicity of the search, avoiding the premature convergence of the search process to the local optimum and improving search efficiency. The sketch map of its three-dimensional search space foraging is shown in Figure 1.

The steps of the QCFOA are shown as follows:

Step 1. Initialize the fruit fly swarm location $(X_{\text{axis}}, Y_{\text{axis}}, Z_{\text{axis}}) = (X_0, Y_0, Z_0) = (X_0, Y_0, Z_0)$ with random function $\text{rand}(\cdot)$ randomly and parameters of the first iteration, including the maximum number of iterations T, number of fruit flies N, chaos control parameter μ, and dissipation parameter β.

Step 2. Update the position (X_i, Y_i, Z_i) of each fruit fly by using equation (5):

$$\begin{cases} X_i = X_{\text{axis}} + \text{rand}(1,1), \\ Y_i = Y_{\text{axis}} + \text{rand}(1,1), \\ Z_i = Z_{\text{axis}} + \text{rand}(1,1), \end{cases} \tag{6}$$

where $\text{rand}(\cdot) \in [0,1]$ represents the random variable of uniform distribution, $i = 1, 2, \ldots, N$. Substituting equation (5) into $\text{rand}(\cdot)$ of equation (6), we can get the following:

$$\begin{cases} X_i = X_{\text{axis}} + X_i', \\ Y_i = Y_{\text{axis}} + Y_i', \\ Z_i = Z_{\text{axis}} + Z_i', \end{cases} \tag{7}$$

where

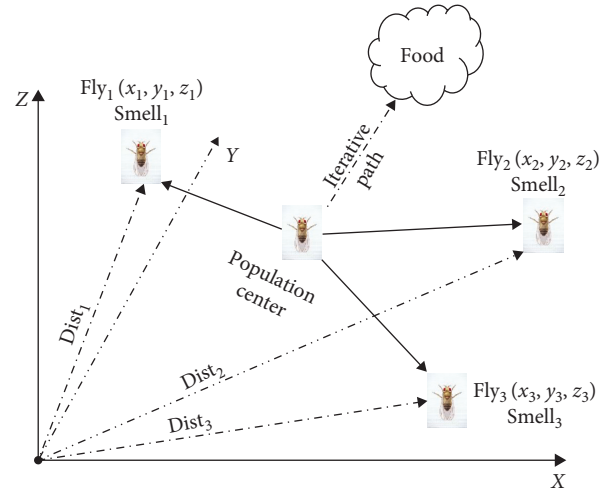


Figure 1: Three-dimensional space foraging map of the fruit fly swarm.

$$\begin{cases} X_i' = \mu\left(X_{\text{axis}} - \left|X_{\text{axis}}\right|^2\right) - \mu Y_{\text{axis}}, \\ Y_i' = -Y_{\text{axis}} e^{-2\beta} + e^{-\beta}\mu\left[(2 - 2X_{\text{axis}})Y_{\text{axis}} - 2X_{\text{axis}}Z_{\text{axis}}\right], \\ Z_i' = -Z_{\text{axis}} e^{-2\beta} + e^{-\beta}\mu\left[2(1 - X_{\text{axis}})Z_{\text{axis}\,n} - 2X_{\text{axis}}Y_{\text{axis}} - X_{\text{axis}}\right]. \end{cases} \tag{8}$$

Step 3. Each component from Step 2 is mapped to the value of odor concentration judgment via equation (7):

$$\begin{cases} \text{Distance}(i) = \sqrt{X_i^2 + Y_i^2 + Z_i^2}, \\ S_i = \dfrac{1}{\text{Distance}(i)}. \end{cases} \tag{9}$$

Step 4. Calculate odor concentration values based on fitness function:

$$\text{Smell}_i = \text{fitness function}(S_i). \tag{10}$$

Step 5. Select maximum odor concentration:

$$\text{best Smell Index} = \text{selection max}(\text{Smell}). \tag{11}$$

Step 6. Update maximum odor concentration:

$$\begin{cases} X_i = X(\text{best Smell Index}), \\ Y_i = Y(\text{best Smell Index}), \\ Z_i = Z(\text{best Smell Index}). \end{cases} \tag{12}$$

Step 7. Check the termination condition: this step compares the current maximum odor concentration to the previous maximum odor concentration. When the current maximum concentration is no longer superior to the previous one, or the current iteration is equal to the maximum number of iterations, the iteration process is terminated. Otherwise, execute Step 2.

3.3. VMD Model Based on Quantum Chaotic FOA

3.3.1. The Fitness Function. The decomposition capability of the VMD is heavily determined by the selected parameters α and K. For nonstationary and nonlinear signal processing, it is not feasible to search the optimal parameters of $[\alpha, K]$ artificially. In the study, the optimization algorithm of quantum chaotic fruit fly is used to search the optimal parameters of VMD. But, before the optimization, a fitness function needs to be determined.

Information entropy is a measure of uncertainty in information quality, which represents the average uncertainty of signals. The greater the entropy, the greater the uncertainty of the signal and the more complex the signal. This paper defines the VMD marginal spectral entropy H_p of the signal $x(t)$ based on the information entropy, which can represent the uncertainty of the signal in the frequency domain and measure the complexity of signal frequency.

Hilbert transformation for each IMF component $c_i(t)$ of VMD is as follows:

$$H[c_i(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{c_i(\tau)}{t-\tau} d\tau. \tag{13}$$

The analytic signal $s_i(t)$ is constructed as follows:

$$s_i(t) = c_i(t) + jH[c_i(t)] = a_i(t) e^{j\varphi_i(t)}, \tag{14}$$

and we can obtain the following:

$$x(t) = \text{Re} \sum_{i=1}^{n} a_i(t) e^{j\varphi_i(t)} = \text{Re} \sum_{i=1}^{n} a_i(t) e^{j2\pi \int f_i(t)dt}, \tag{15}$$

where n is the number of IMF components, $a_i(t)$ is the instantaneous amplitude, $\varphi_i(t)$ is the instantaneous phase, and $f_i(t)$ represents the instantaneous frequency.

The amplitude and frequency of the Hilbert transform are time-domain functions. The amplitude of the signal $x(t)$ can be expressed as a function of time and frequency in the three-dimensional space, and it is the Hilbert amplitude spectrum:

$$H(f,t) = \text{Re} \sum_{i=1}^{n} a_i(t) e^{j2\pi \int f_i(t)dt}. \tag{16}$$

The marginal spectrum of the signal $x(t)$ can be obtained by the time integral of $H(f,t)$:

$$h(f) = \int_{-\infty}^{+\infty} H(f,t)\, dt. \tag{17}$$

The variation rule of $x(t)$ amplitude with frequency is described by equation (15). The marginal spectrum entropy of the signal $x(t)$ is defined as follows:

$$\begin{cases} H_p = -\sum_{i=1}^{N} p_i \ln p_i, \\ p_i = \dfrac{h(i)}{\sum_{i=1}^{N} h(i)}, \end{cases} \tag{18}$$

where p_i is the probability of the corresponding amplitude of the ith frequency and $h(i)$ is the marginal spectrum of the ith IMF component. In order to facilitate analysis, the marginal spectrum entropy value is normalized: $H_E = H_p / \ln L$, and the value range is [0,1]. L determines the length of $h(i)$ sequence.

3.3.2. Proposed Improved Algorithm Framework. When the bearing is in fault, the vibration signal mainly appears as a periodic impact signal. Therefore, when using the VMD algorithm to deal with bearing fault signals, multiple IMF components will be obtained. According to the definition of marginal spectrum entropy in equation (18), if the IMF component contains more noise components, its corresponding MSE value is larger. Conversely, if the IMF component mainly contains the periodic impact component of bearing failure, the marginal spectrum entropy of VMD is very small. Therefore, the minimum value of the marginal spectrum entropy is expressed as a local minimum, and the corresponding IMF component is taken as the optimal component. The local minimum marginal spectral entropy (LMMSE) of the VMD is shown as follows:

$$\min_{L} H_E^{\text{IMF}} = \min - \sum_{i=1}^{L} p_i \ln p_i. \tag{19}$$

In this study, the LMMSE of the VMD is taken as the fitness value in the optimization process, and the global optimal IMF component is searched. The local minimum value of the marginal spectrum entropy is used as the final optimization target. The optimization process is the same as the FOA, and the process is carried out as follows:

Step 8. Initialize the fruit fly swarm parameters of the first iteration, including the fruit fly group number N, number of iterations T, and maximum iterations $T_{\max}$.

Step 9. The parameters in equation (3) are set as $\mu = 3.99$ and $\beta \geq 6$.

Step 10. Generate two large groups of U_α and U_K randomly by using equation (7), and the number of fruit flies in each group is N. The positions $U(X_i^\alpha, Y_i^\alpha, Z_i^\alpha)$ and $U(X_i^K, Y_i^K, Z_i^K)$ of each fruit fly are then continuously updated using equation (7).

Step 11. Calculate Distance(i, α) according to equation (5), and it is represented by $D(i, \alpha)$ and $D(i, K)$. Set $S(i, K) = 1/D(i, K)$ and $S(i, \alpha) = 1/D(i, \alpha)$, and S_i is represented by $S(i, \alpha)$ and $S(i, K)$. Get the odor concentration Smell_i by the fitness value, and obtain the best smell concentration value $\text{Smell}_{\text{best}}$ and the corresponding location $[S(i, \alpha), S(i, K)]$.

Step 12. Enter iterative optimization to repeat the implementation of Step 3, $T = T + 1$.When the current smell concentration is not superior to the previous iterative smell concentration any more or the iteration number reaches the maximum number of iterations $T_{\max}$, the circulation is terminated, and return optimal combination parameters $[\alpha_{\text{opt}}, K_{\text{opt}}]$. Otherwise, carry out Step 3.

3.4. Simulations and Analysis

3.4.1. Simulation Signal. In practice, rolling bearings generate periodic impact signals when pitting or cracking occurs. But in the early stage of failure, the impact signal is very weak and is easily drowned by noise, so it is difficult to find fault characteristic frequency in traditional signal analysis methods. This paper proposes a quantum chaotic fruit fly algorithm to search the optimal parameters of VMD, and the flow chart of algorithm is shown in Figure 2. In order to qualitatively and quantitatively analyze the validity and superiority of this method, the simulation signal of the early fault of the bearing under the simulated strong noise background is analyzed [30], and the expression of the simulation signal is as follows:

$$x(t) = x_1(t) + x_2(t) + x_n(t), \tag{20}$$

where $x_1(t)$ is an impact simulation signal with periodic pulse attenuation with a frequency of 12 Hz and maximum amplitude of 0.5 V, $x_2(t)$ is a cosine combined signal with a frequency of 35 Hz and 15 Hz, and $x_n(t)$ is a Gauss white noise signal. The time-domain waveform of the simulation signal is shown in Figure 3.

From the simulation signal of early fault of the bearing in Figure 3, it can be seen that the impact attenuation signal in the early stage of simulated bearing failure is almost submerged by low-frequency components and strong noise due to its small amplitude.

3.4.2. Parameter Optimization Analysis. In order to verify the effectiveness and superiority of the improved chaotic FOA parameter optimization proposed in this paper, the performance is compared with that of classical PSO, QPSO, FOA, and CFOA. In this experiment, the number of fruit flies is set as $N = 30$ and the maximum number of iterations

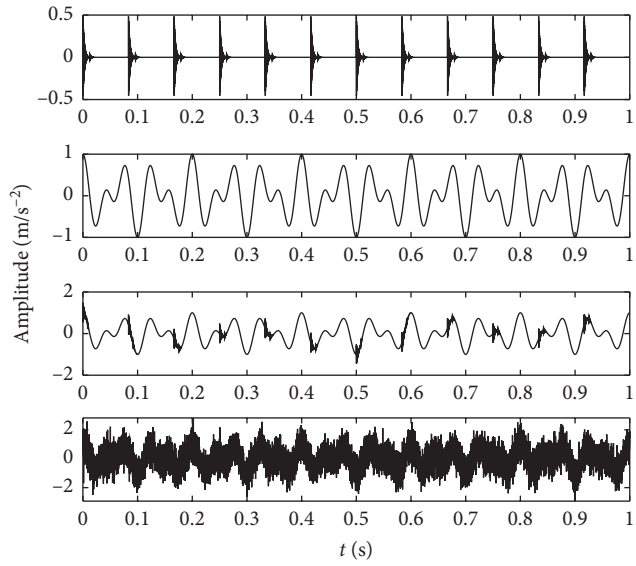


Figure 2: Simulation signal of early fault of the bearing.

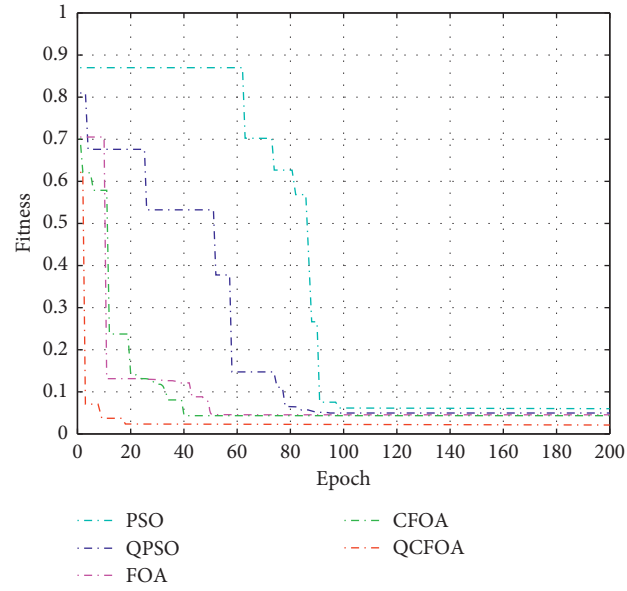


Figure 3: Search process diagram of three optimization algorithms.

is $T_{\max} = 200$. Other parameters are selected according to relevant literature to ensure the best results of each algorithm. The combination parameters $[\alpha_{\text{opt}}, K_{\text{opt}}]$ are optimized by the above method. Figure 4 and Table 1 show that different algorithms have different effects on the solution.

From Table 1, it can be observed that the QCFOA method achieves the minimum number of iterations and the best global optimal solution. The searched VMD parameter combination is [690, 4], and the reconstruction error is 1.02×10^{-5}. Figure 4 demonstrates the iterative process using PSO, QPSO, FOA, CFOA, and QCFOA. As shown in Figure 4, the convergence speed of the QCFOA algorithm is the fastest and the reconstruction error of the signal is the smallest.

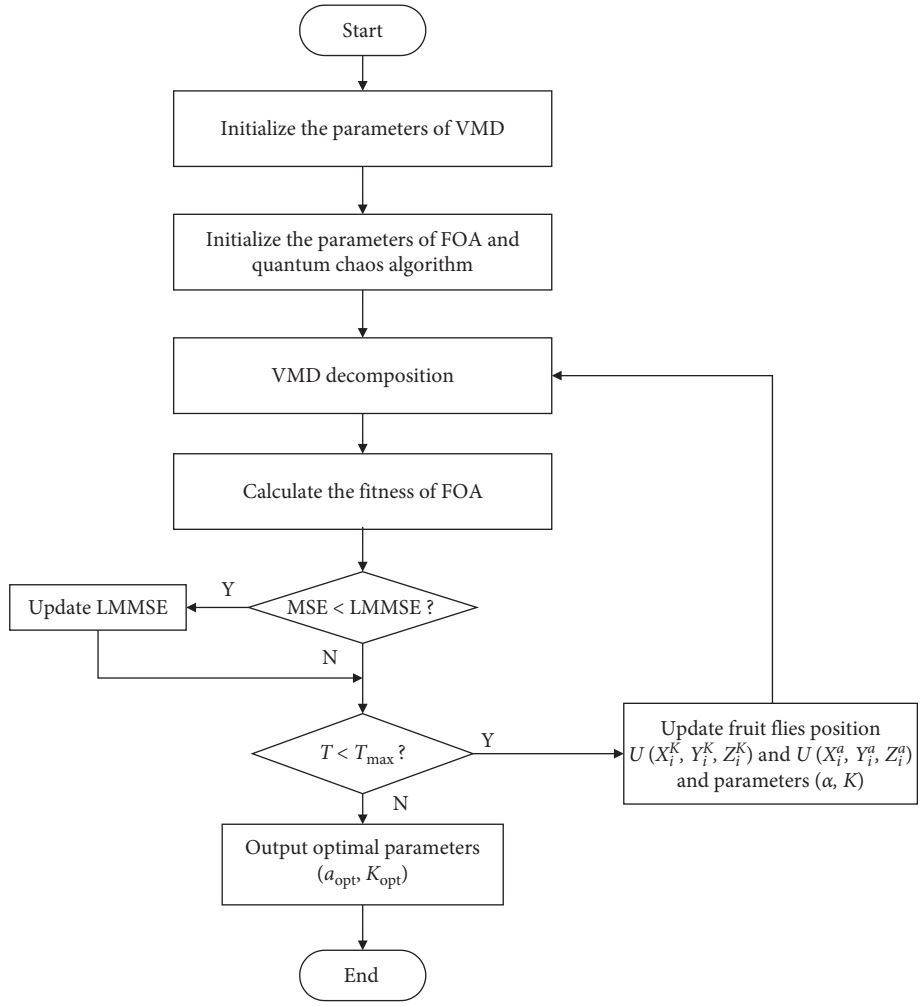


Figure 4: Flow chart of the optimal VMD algorithm.

Table 1: Computation efficiency comparison among PSO, QPSO, FOA, CFOA, and QCFOA methods.

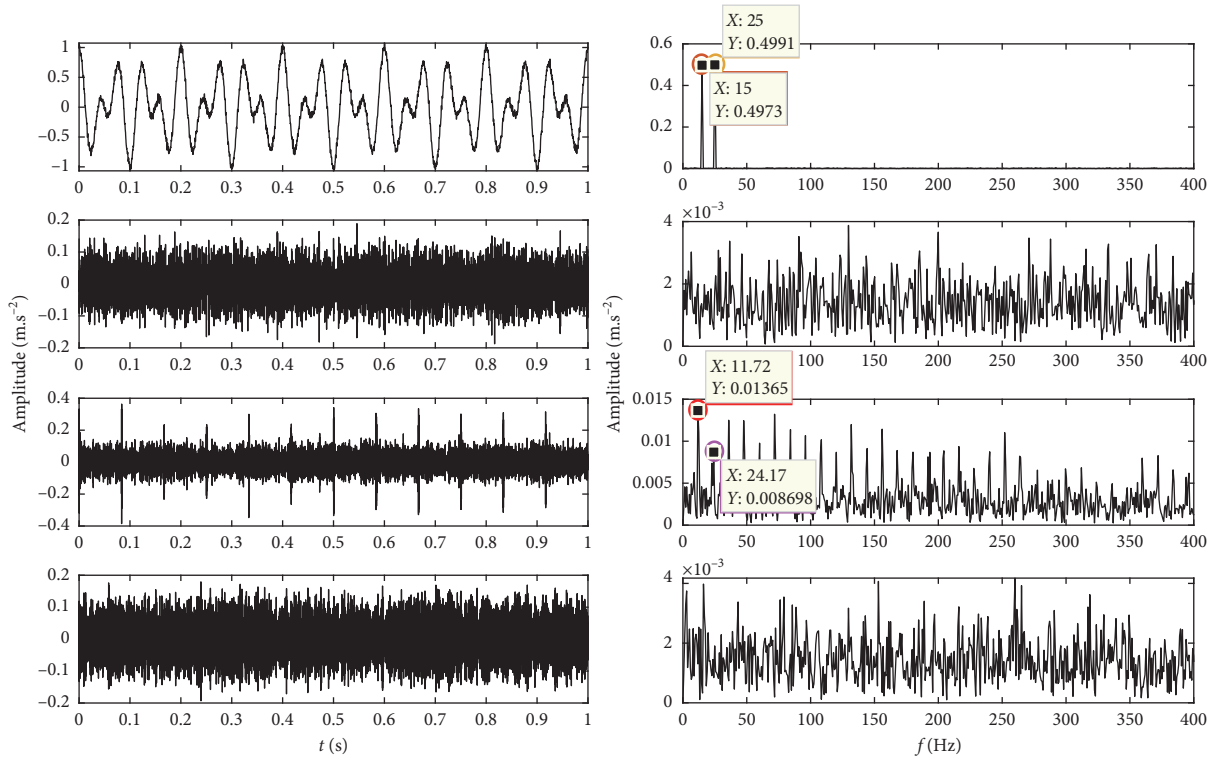
Algorithms	Global optimal solution	Iterative number	$[\alpha, K]$	Reconstruction error
PSO	0.0602	110	[1207, 7]	3.01×10^{-4}
QPSO	0.0501	94	[1150, 6]	1.47×10^{-4}
FOA	0.0454	75	[1040, 6]	1.15×10^{-4}
CFOA	0.0443	40	[970, 5]	6.17×10^{-5}
QCFOA	0.0236	18	[690, 4]	1.02×10^{-5}

In the search process, the LMMSE changes with the evolution of the population, and the corresponding VMD optimal parameter combination is shown in Table 1.

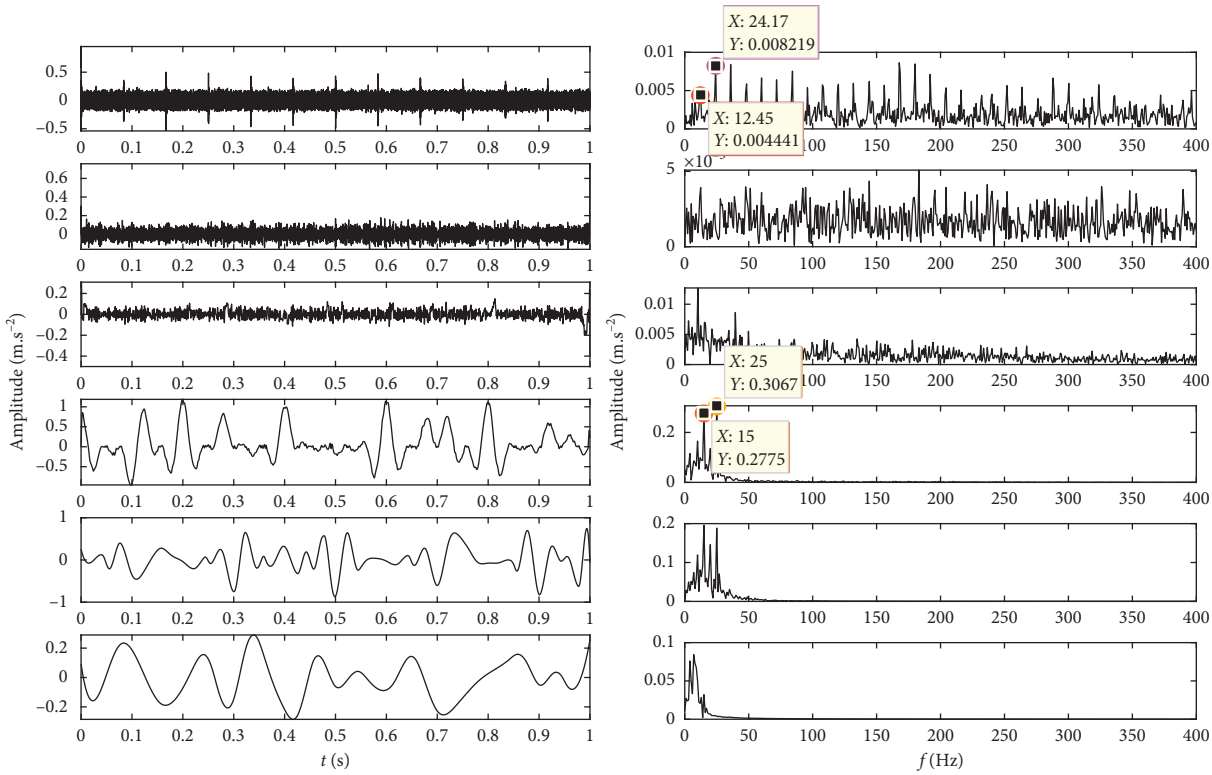
From Table 1, it can be observed that the QCFOA method achieves the minimum number of iterations and the best global optimal solution. The searched parameter combination of VMD $[\alpha, K]$ is [690, 4], its reconstruction error is 1.02×10^{-5}, and it has the least number of iterations. Figure 4 demonstrates the search process by using PSO, QPSO, FOA, CFOA, and QCFOA. The experimental results indicate that the convergence speed of the QCFOA is faster than others. Therefore, the experimental results indicate that the proposed method is more effective and superior than the mentioned optimization algorithms.

3.4.3. Comparison and Analysis of EEMD, LMD, and VMD Methods. In order to validate the effectiveness and superiority of the VMD method based on the QCFOA, EEMD, LMD, and VMD methods are used, respectively, to process the simulation signals in Figure 3, and the results are shown in Figure 5.

From Figure 5, it can be seen that the EEMD and LMD methods can effectively extract the characteristic frequencies of the low-frequency cosine components in the simulation signals including 25 Hz and 15 Hz but cannot extract the characteristic frequency of the weak impact signal of 12 Hz, whereas there is modal mixing which is an unavoidable deficiency in the decomposition of EEMD and LMD. The low-frequency cosine components of 25 Hz and 15 Hz



(a)



(b)

Figure 5: Continued.

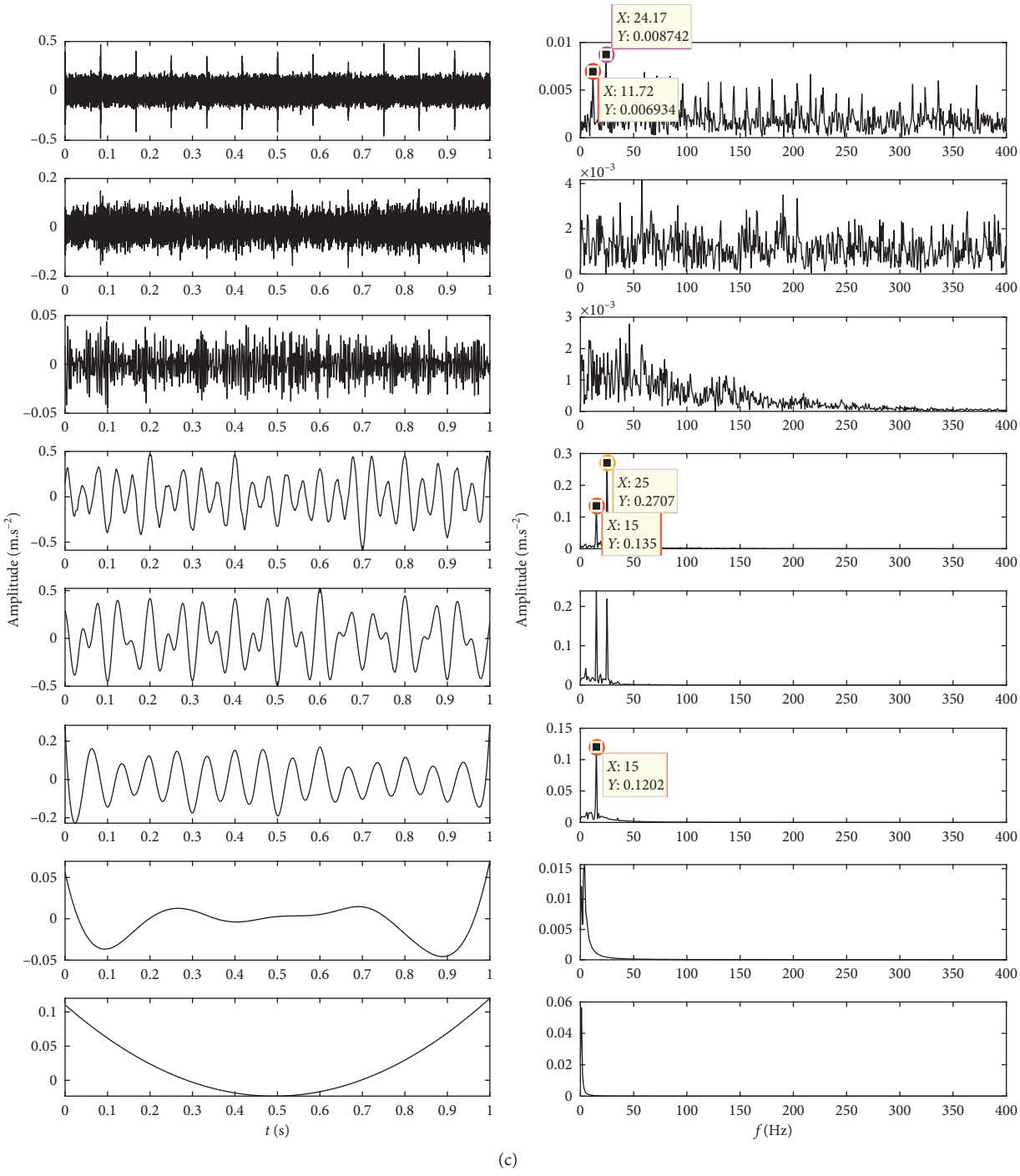


Figure 5: The characteristic components and corresponding marginal spectrum of the simulation signal based on (a) VMD, (b) LMD, and (c) EEMD methods.

appear in different components. In addition, the frequency amplitude of the extracted 12 Hz impact signal is also relatively weak. The VMD method based on the QCFOA proposed in this paper not only can effectively extract the characteristic frequency of the low-frequency cosine components of 25 Hz and 15 Hz but also can effectively extract the characteristic frequency of the weak impact signal of 12 Hz and its corresponding doubling frequency. The number of signal components decomposed by the improved VMD method is also significantly less than that obtained by EEMD and LMD. The experimental results show the better effectiveness and superiority of the proposed method.

4. Kernel Parameter Self-Optimization Variational Relevance Vector Machine

4.1. Variational Relevance Vector Machine. Many problems in machine learning belong to the category of supervised

learning, giving an input vector $X = \{x_n\}_{n=1}^N$ and the corresponding target output $T = \{t_n\}_{n=1}^N$. For regression problems, they can be arbitrary values. For classification problems, they are class labels. For regression, t_n can be any value, and for classification, t_n is the category label.

For the standard RVM regression model, the formula is defined as follows:

$$\begin{aligned} t_n &= y(x_n; w) = \left[\sum_{n=1}^{N} w_n K(x, x_n) + w_0\right] + \varepsilon_n \\ &= w^T \phi(x_n) + \varepsilon_n, \end{aligned} \tag{21}$$

where $\phi(x_n) = [1, K(x_n, x_1), \ldots, K(x_n, x_N)]^T$, $\{w_n\}$ is the weight parameter of the model, $K(x, x_n)$ is the kernel function, and an RBF kernel function is selected in this section: $K(x, x_n) = \exp(\|x - x_n\|^2/h^2)$, $h > 0$. ε_n represents the noise, which obeys normal distribution $N(0, \sigma^2)$.

If the input vectors $X = \{x_n\}_{n=1}^N$ are independent and identically distributed, then the likelihood function of the corresponding target output $T = \{t_n\}_{n=1}^N$ is defined as

$$P\left(T \mid X, w, \sigma^2\right) = \prod_{n=1}^{N} P\left(t_n \mid x_n, w, \sigma^2\right), \tag{22}$$

where the parameter w is Gaussian prior distribution and $\sigma^2$is noise variance:

$$P(w \mid \alpha) = \prod_{n=1}^{N} N\left(w_n \mid 0, \alpha_n^{-1}\right), \tag{23}$$

where $\alpha = \{\alpha_n\}$ is a hyperparameter vector, and each weight value w_n is independently assigned a parameter α_n. In order to make parameter learning more flexible, ultra-prior distribution is defined for α and noise variance σ^2, respectively. The appropriate prior distribution is the gamma distribution:

$$\begin{aligned} P(\alpha) &= \prod_{n=0}^{N} \text{Gamma}\left(\alpha_n \mid a_n, b_n\right), \\ P\left(\sigma^2\right) &= \text{Gamma}\left(\sigma^2 \mid c, d\right), \\ \text{Gamma}(\alpha \mid a, b) &= \Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha}, \end{aligned} \tag{24}$$

where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}\, dt$ is the gamma function, and it is usually defined that the hyperparameter is a very small value such as $a_n = b_n = c = d = 10^{-4}$; such a hyperparameter before does not provide information for posterior learning, so the posterior depends entirely on the data.

When the model is established, the posterior distribution of w, α, and σ^2 can be obtained by the variable Bayesian method [40]. In the process of iterative solution, most of α_n tends to infinity and the corresponding w_n is zero, realizing the sparse model.

RVM classification and regression essentially use the same framework model, except that the conditional distribution of the target value is changed. For the binary classification, the logistic function $\sigma(y) = (1/1 + e^{-y})$ is used in the continuous latent variable $y(x_n; w)$, and assuming that $P(T \mid X)$ is the Bernoulli distribution, the likelihood function is as follows [41]:

$$P(T \mid X, w) = \prod_{n=1}^{N} \sigma\left(y(\mathrm{x}_n; \mathrm{w})\right)^{t_n} \left[1 - \sigma\left(y(\mathrm{x}_n; w)\right)\right]^{1-t_n}. \tag{25}$$

It is important to note that, in the classification problem without considering the noise variable ε_n, it is very difficult to directly use the variational method to solve the above model. In [28], a lower bound is introduced by using the inequality:

$$\begin{aligned} &\sigma\{y(x_n; w)\}^{t_n} \left[1 - \sigma\{y(x_n; w)\}\right]^{1-t_n} \\ &\quad = \sigma(z_n) \geq \sigma(\xi_n) \exp\left(\frac{z_n - \xi_n}{2} - \lambda(\xi_n)\left(z_n^2 - \xi_n^2\right)\right), \end{aligned} \tag{26}$$

where $\lambda(\xi_n) = (1/4\xi_n) \times \tanh(\xi_n/2)$, $z_n = (2t_n - 1) y(x_n; w)$, ε_n is a variational parameter, and when ε_n=z_n, equation (26) is established. Finally, the variational method is used to solve the lower boundary.

The VRVM method is the posterior distribution of RVM model parameters and superparameters by variational Bayesian (VB) function. In the VB method [42] based on the RVM model, the observed variables are $X = \{x_n\}_{n=1}^N$ and $T = \{t_n\}_{n=1}^N$, the hidden variables are $\{y_n\}_{n=1}^N$, and the parameter is $\left\{w_{pm}, \alpha_p, \sigma^2\right\}_{p=1,m=1}^{P \quad M}$; therefore, the logarithmic edge likelihood function can be written as

$$\begin{aligned} \ln P(T \mid X) &= \ln \iint\!\!\iint P\left(T \mid X, y, w, a, \sigma^2\right) dy\, dw\, da\, d\sigma^2 \\ &= \ln \iint\!\!\iint Q\left(y, w, \alpha, \sigma^2\right) \frac{P\left(T \mid X, y, w, a, \sigma^2\right)}{Q\left(y, w, a, \sigma^2\right)}\, dy\, dw\, da\, d\sigma^2 \\ &\geq \iint\!\!\iint Q\left(y, w, a, \sigma^2\right) \ln \frac{P\left(T \mid X, y, w, a, \sigma^2\right)}{Q\left(y, w, a, \sigma^2\right)}\, dy\, dw\, da\, d\sigma^2, \end{aligned} \tag{27}$$

where $Q(y, w, a, \sigma^2)$ is the joint probability distribution function between the hidden variable and parameter. The variational Bayesian algorithm assumes that y, w, a, and σ^2 are independent of each other; therefore, the joint probability distribution of four variables can be written approximately as

$$Q\left(y, w, a, \sigma^2\right) \approx Q(y) Q(w) Q(a) Q\left(\sigma^2\right). \tag{28}$$

Following the assumption of equation (28), equation (21) can be rewritten as follows:

$$\begin{aligned} \ln P(T \mid X) \geq \iint\!\!\iint & Q(y) Q(w) Q(a) Q\left(\sigma^2\right), \\ & \ln \frac{P\left(T \mid X, w, a, \sigma^2\right)}{Q(y) Q(w) Q(\alpha) Q(\sigma^2)}\, dy\, dw\, d\alpha\, d\sigma^2 \\ & = F\left(Q(y), Q(w), Q(a), Q\left(\sigma^2\right)\right). \end{aligned} \tag{29}$$

It can be seen from equation (29) that the log-likelihood function has a lower bound and that the real value can be

approximated by maximizing the lower bound $F(Q(y), Q(w), Q(a), Q(\sigma^2))$. Then, the lower bound of the log-likelihood is solved using EM (expectation-maximization), and the posterior distribution of the hidden variable and all parameters is obtained. The lower bound expression is as follows:

$$\begin{aligned} F&\left(Q(y), Q(w), Q(a), Q\left(\sigma^2\right)\right) \\ &= \iint\iint Q(y)Q(w)Q(a)Q\left(\sigma^2\right)\Big\{\ln P(t \mid y) \\ &\quad + \ln P\left(y \mid Xw, \sigma^{-2}\right) + \ln P\left(w \mid 0, \alpha^{-1}\right) + \ln P(a \mid a, b) \\ &\quad + \ln P\left(\sigma^2 \mid c, d\right) - \ln Q(y) - \ln Q(w) - \ln Q(a) \\ &\quad - \ln Q\left(\sigma^2\right)\Big\}\, dy\, dw\, da\, d\sigma^2. \end{aligned} \tag{30}$$

Because the variable distributions of $Q(y)$, $Q(w)$, $Q(a)$, and $Q(\sigma^2)$ are all conjugate prior distributions, they have the same distribution form as their posterior distribution, and the following is their probability distribution:

$$Q(y) = \prod_{n=1}^{N} N^{t_n}(y_n; \widetilde{m}_n, \widetilde{\sigma}_n), \tag{31}$$

$$Q(w) = N(w; \widetilde{\mu}_w, \widetilde{\Sigma}_w), \tag{32}$$

$$Q\left(\sigma^2\right) = \text{Gamma}\left(\sigma^2; \widetilde{c}, \widetilde{d}\right), \tag{33}$$

$$Q(a)\left\{\prod_{p=1}^{P} \text{Gamma}\left(\alpha_p; \widetilde{a}_p, \widetilde{b}_p\right)\right\}^{M}, \tag{34}$$

where $N^{t_n}(\cdot)$ represents truncated normal distribution, and the direction of truncation $N^{+}(\cdot)$ or $N^{-}(\cdot)$ is determined by t_n.

According to the theory of the EM algorithm, the posterior distribution of the variables is actual expectation of the logarithm of the complete likelihood function with respect to other variables (indicated by $\langle\cdot\rangle$), and then the terms related to the variable are extracted. The posterior distributions of y, w, a, and σ^2 are, respectively, solved as follows:

Step 13. Extract the term related to variable y in equation (30):

$$\begin{aligned} \widetilde{F}(Q(y)) = \int Q(y)\Big\{&\ln P(t \mid y) \\ &+ \iint Q(w)Q\left(\sigma^2\right)\ln P\left(y \mid Xw, \sigma^{-2}\right) dw\, d\sigma^2 - \ln Q(y)\Big\}\, dy. \end{aligned} \tag{35}$$

Set the derivative of $Q(y)$ in equation (35) to be 0; hence,

$$\begin{aligned} &\ln P(t \mid y) + \iint Q(w)Q\left(\sigma^2\right)\ln P\left(y \mid Xw, \sigma^{-2}\right) dw\, d\sigma^2 \\ &\quad - \ln Q(y) - 1 = 0. \end{aligned} \tag{36}$$

Solve equation (8) and get the following equation:

$$\begin{aligned} \widetilde{m}_n &= x_n\langle w\rangle, \\ \widetilde{\sigma}_n &= \left\langle \sigma^2 \right\rangle^{-1} \end{aligned} \tag{37}$$

Step 14. Extract the term related to variable w in equation (30):

$$\begin{aligned} \widetilde{F}(Q(w)) = \int Q(w)\Big\{&\iint Q(y)Q\left(\sigma^2\right)\ln P\left(y \mid Xw, \sigma^{-2}\right) dy\, d\sigma^2 \\ &+ \int Q(a)\ln P\left(w \mid 0, \alpha^{-1}\right) da - \ln Q(w)\Big\}\, dw. \end{aligned} \tag{38}$$

Set the derivative of $Q(w)$ in equation (38) to be 0. Hence,

$$\begin{aligned} \widetilde{\Sigma}_w &= \left(diag(\langle a\rangle) + \left\langle \sigma^2 \right\rangle \sum_{n=1}^{N} x_n^T x_n \right)^{-1}, \\ \widetilde{\mu}_w &= \left\langle \sigma^2 \right\rangle \widetilde{\Sigma}_w \sum_{n=1}^{N} x_n^T \langle y_n\rangle. \end{aligned} \tag{39}$$

Step 15. Extract the term related to variable σ^2 in equation (29):

$$\begin{aligned} \widetilde{F}\left(Q\sigma^2\right) = \int Q\left(\sigma^2\right)\Big\{&\iint Q(y)Q(w)\ln P\left(y \mid Xw, \sigma^{-2}\right) dy\, dw \\ &+ \int Q(a)\ln P\left(\sigma^2 \mid c, d\right) - \ln Q\left(\sigma^2\right)\Big\} d\sigma^2. \end{aligned} \tag{40}$$

Set the derivative of $Q(\sigma^2)$ in equation (40) to be 0. Hence,

$$\begin{cases} \widetilde{c} = c + 0.5, \\ d = d + \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \frac{1}{2}\left(\left\langle y_n^2 \right\rangle - 2\langle y_n\rangle\left\langle w^T \right\rangle x_n^T + x_n\left\langle ww^T \right\rangle x_n^T\right). \end{cases} \tag{41}$$

Step 16. Extract the term related to variable a in equation (29):

$$\begin{aligned} \widetilde{F}(Q(a)) = \int Q(a)\Big\{ &\int Q(w)\ \ln P\left(w \mid 0, a^{-1}\right) dw \\ &+ \ln P(a \mid a, b) - \ln Q(a) gBi\}\, da. \end{aligned} \tag{42}$$

Set the derivative of $Q(a)$ in equation (42) to be 0. Hence,

$$\begin{aligned} \widetilde{\alpha}_p &= \alpha_p + \frac{1}{2}, \\ \widetilde{b}_p &= b_p + \frac{1}{M}\sum_{n=1}^{N} \frac{1}{2}\left\langle w_{pm}^2 \right\rangle. \end{aligned} \tag{43}$$

Through the process above, the parameter iteration formulas in each variable's posterior distribution function are obtained, but they are expressed by the expectations of

other variables. Therefore, the expectation of these parameters needs to be found [42].

For equation (32), $t_n (t_n \in \{-1, +1\})$ determines the truncation direction of the truncation normal distribution; hence,

$$\begin{aligned}\langle y_n \rangle &= x_n \tilde{\mu}_w + t_n \frac{1}{\langle \sigma^2 \rangle^{1/2}} \frac{\text{normpdf}\left((x_n \tilde{\mu}_w) / \langle \sigma^2 \rangle^{-1/2} \right)}{\text{normpdf}\left(t_n (x_n \tilde{\mu}_w) / \langle \sigma^2 \rangle^{-1/2} \right)}, \\ \left\langle y_n^2 \right\rangle &= \frac{1}{\langle \sigma^2 \rangle} + (x_n \tilde{\mu}_w)^2 + t_n (x_n \tilde{\mu}_w) \left\langle \sigma^2 \right\rangle^{-1/2} \\ &\quad \cdot \frac{\text{normpdf}\left((x_n \tilde{\mu}_w) / \langle \sigma^2 \rangle^{1/2} \right)}{\text{normpdf}\left(t_n (x_n \tilde{\mu}_w) / \langle \sigma^2 \rangle^{-1/2} \right)},\end{aligned} \tag{44}$$

where normpdf () is a normal probability density function Moreover,

$$\begin{aligned}\langle w \rangle &= \tilde{\mu}_w, \\ \left\langle w w^T \right\rangle &= \tilde{\Sigma}_w + \tilde{\mu}_w \tilde{\mu}_w^T, \\ \left\langle \alpha_p \right\rangle &= \frac{\tilde{a}_p}{\tilde{b}_p}, \\ \left\langle \ln \alpha_p \right\rangle &= \Psi\left(\tilde{a}_p \right) - \ln \tilde{b}_p, \\ \left\langle \sigma^2 \right\rangle &= \frac{\tilde{c}}{\tilde{d}}, \\ \left\langle \sigma^2 \right\rangle &= \Psi(\tilde{c}) - \ln \tilde{d}.\end{aligned} \tag{45}$$

The Ψ function is defined as follows:

$$\Psi(a) = \frac{d \ln \Gamma(a)}{da}. \tag{46}$$

After the model is trained, for a test sample x_*, its probability of prediction can be calculated as follows:

$$\begin{aligned}P\left(t_* = 1 \mid x_* \right) &= \int P\left(t_* = 1 \mid y_* \right) P\left(y_* \mid x_*, \langle w \rangle, \left\langle \sigma^2 \right\rangle \right) dy_* \\ &= \int_0^\infty N\left(y_*; x_* \langle w \rangle, \left\langle \sigma^2 \right\rangle^{-1} \right) dy_* \\ &= \text{normcdf}\left(\frac{x_* \tilde{\mu}_w}{\langle \sigma^2 \rangle^{-1/2}} \right).\end{aligned} \tag{47}$$

where normcdf () is a normal cumulative integral distribution function. According to the prediction probability, the test sample can be classified and identified. When $P(t_* = 1 \mid \mathrm{x}_*) \geq 0.5$, the test sample is judged as a classification; otherwise, it is judged as another class.

4.2. Introduction of Probit Model. The binary logistic model of standard RVM is

$$P\left(t_n = 1 \mid w^T \phi(x_n) \right) = \frac{1}{1 + \exp\left(-w^T \phi(x_n) \right)}. \tag{48}$$

The logistic function is a mapping from continuous variables to binary output t_n. Logistic mapping function is easy to understand, but it is not a standard probability function; there are many difficulties in the process of reasoning. In addition, the traditional RVM classification model introduces a lower bound to the likelihood function, which is an approximate derivation. Therefore, the real value of the model cannot be estimated accurately. To solve this problem, a mapping method from continuous quantity to discrete quantity through the probit model is defined as follows:

$$P\left(t_n, y_n \mid w^T \phi(x_n), \sigma^2 \right) = P\left(t_n \mid y_n \right) N\left(y_n; w^T \phi(x_n), \sigma^{-2} \right), \tag{49}$$

where the hidden variable $\{y_n\}_{n=1}^N$ is a continuous random variable hidden behind t_n. Based on the requirement of the model, the target value $\{t_n\}_{n=1}^N$ is assumed to be −1 or +1. The probability relationship is as follows:

$$P\left(t_n \mid y_n \right) = I\left(t_n = \text{sign}\left(y_n \right) \right) = \begin{cases} 1, & t_n = \text{sign}\left(y_n \right), \\ 0, & \text{otherwise}, \end{cases} \tag{50}$$

where $I(\cdot)$ is an indicator function. By integrating y_n, it can be found that

$$\begin{aligned}&P\left(t_n, y_n \mid w^T \phi(x_n), \sigma^2 \right), \\ &= \int P\left(t_n = 1, y_n \mid w^T \phi(x_n), \sigma^2 \right) dy_n, \\ &= \int_0^\infty N\left(y_n; w^T \phi(x_n), \sigma^{-2} \right) dy_n, \\ &= \text{normcdf}\left(\frac{w^T \phi(x_n)}{\sigma^{-1}} \right),\end{aligned} \tag{51}$$

where normcdf $(\cdot)$ is a normal cumulative distribution function.

The advantage of the probit model is to transform the problem of binary-classification output into a regression problem by introducing hidden variables. It makes the models of classification and regression completely equivalent, and the noise variable must be ignored by using the logistic model. It can be seen form Figure 6 that the probit model approximates the logistic model very well. Therefore, the reasoning algorithm based on the probit model can be flexibly applied to the classification model directly. In addition, the use of the probit model can also easily extend the binary classification to multiple classifications [43]. Compared with the multivariate logistic model [44], the multivariate probit model can also avoid complex approximate calculations, has more simple and practical characteristics, and can be well approximated to the logistic model.

5. Analysis of Bearing Fault Diagnosis

5.1. Experimental Scheme. In order to further verify the effectiveness of the proposed method for bearing fault

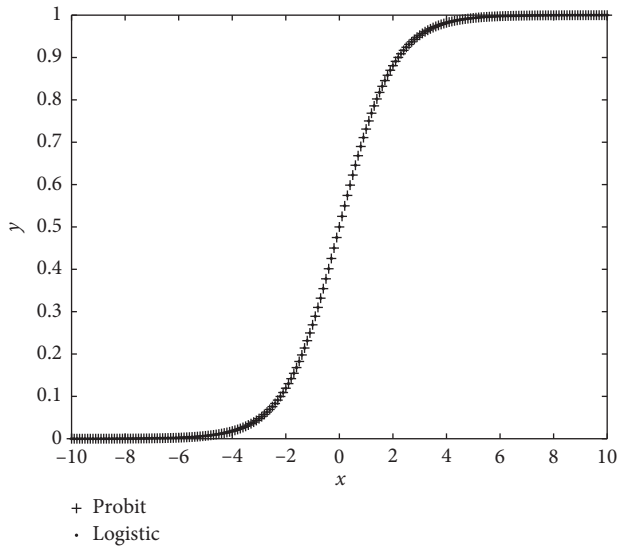


Figure 6: Comparison between the probit model and logistic model ($\sigma^2 = 1$).

diagnosis under unknown variable load conditions, the experiments are conducted to collect vibration signals from bearings operating under variable load conditions. The experimental platform is shown in Figure 7; it uses bearing type 6203-2RS JEM SKF deep groove rolling bearing. The bearing inner diameter is 25 mm, outer diameter is 52 mm, and thickness is 15 mm; the rolling diameter is 8.18 mm, and the pitch diameter of the bearing is 44.2 mm; the sampling frequency is 12 kHz, and the sampling data length is 12000. The fault with different etch diameters (simulating varying degrees of damage to the inner ring, the outer ring, and the rolling body) is processed by electric spark. Bearing loads include 3 types: load 0 (0 N, 1797 r/min, simulated no load), load 1 (800 N, 1772 r/min, simulated light load), and load 2 (1600 N, 1750 r/m, simulated heavy load). The pitting diameter of the bearing is 0.1778 mm (it simulates minor fault), 0.3556 mm (it simulates medium fault), and 0.5334 mm (it simulates serious fault), which is used to simulate three different damage degrees of the bearing.

5.2. Fault Feature Extraction. To further verify the effectiveness of the VMD method based on the quantum chaotic FOA in the early bearing fault feature extraction, this paper divides the fault feature extraction into two cases: the fault characteristics of the minor faults under different loads and the fault characteristics of different damage faults under the same load.

5.2.1. Fault Feature Extraction of Weak Damage Degree under Different Loads. Under three different loads, the measured signals of small bearing failure (pitting diameter 0.1778 mm) and small features of the rolling body fault are used as experimental signals. The fault signals are shown in Figure 8. The parameter $[\alpha, K]$of the VMD method optimized by the quantum chaotic FOA is used to search the optimal value, and the search results are shown in Table 2.

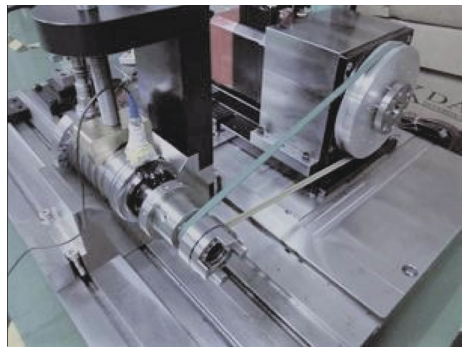


Figure 7: Experimental platform.

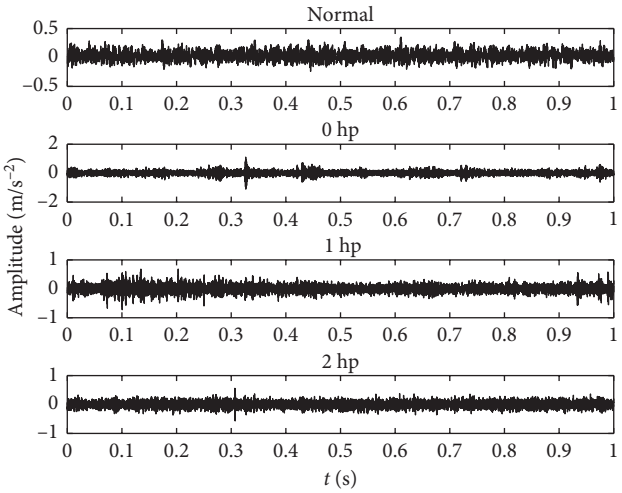


Figure 8: Rolling body minor fault signals under three different loads.

The optimal IMF component and its corresponding marginal spectrum and center frequency for the small faults of rolling bodies under three different loads are shown in Figures 9(a)–9(c), respectively. f_r represents the rotor

Table 2: Optimal search results.

Load	0 hp	1 hp	2 hp
The calculated value of the fault frequency (Hz)	119.39	118.01	116.61
Fault measured value (Hz)	119.4	117.9	116.5
[α, K]	[2000, 6]	[1600, 6]	[1400, 8]
Optimal IMF	IMF4	IMF4	IMF6
Normalized center frequency	0.4947	0.4224	0.3644
LMMSE	0.3228	0.3231	0.3275

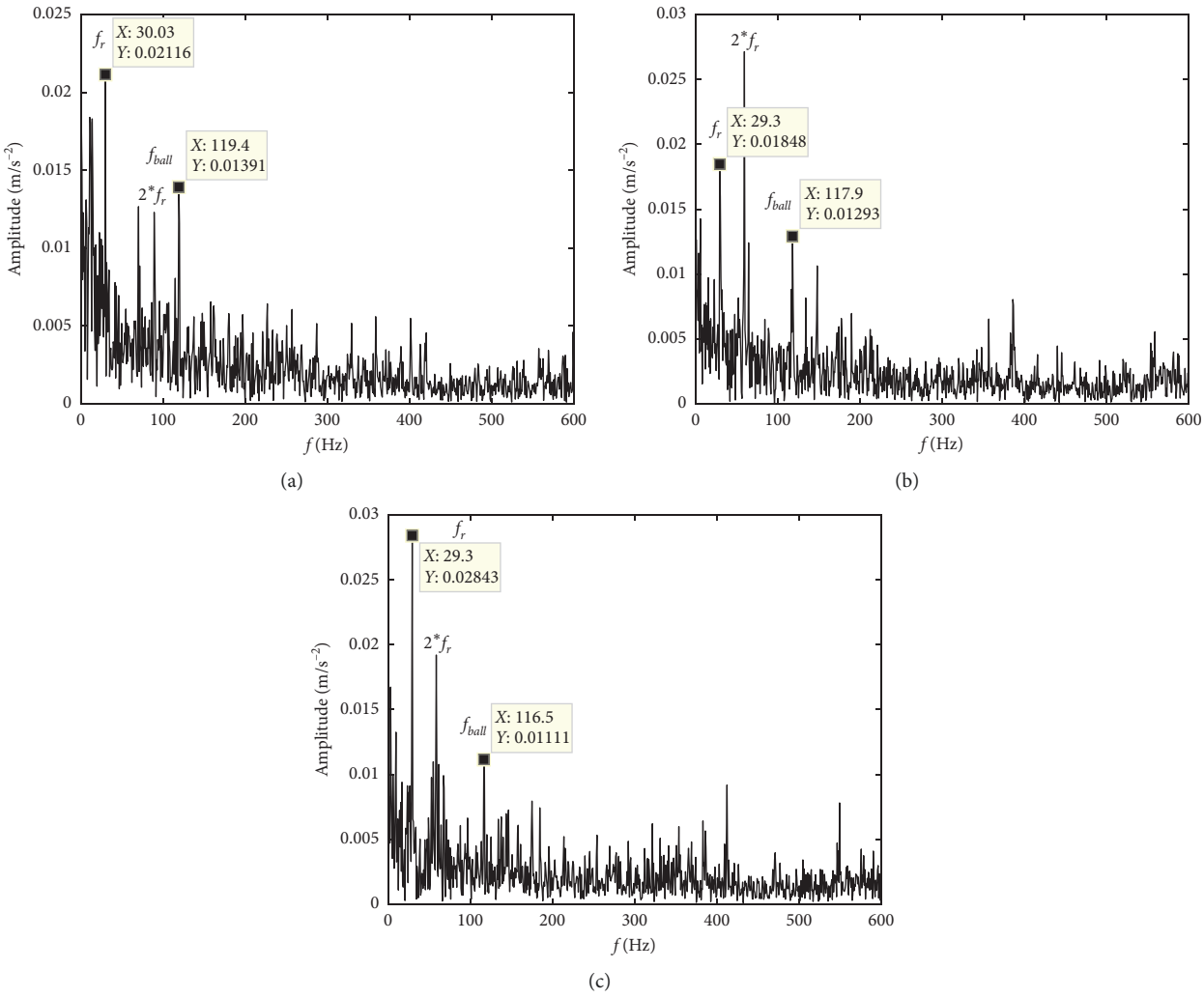


Figure 9: The best IMF component and its corresponding marginal spectrum and center frequency. Marginal spectrum of IMF4 components under load: (a) 0 hp; (b) 1 hp; (c) 2 hp.

frequency, and f_{ball} represents the fault characteristic frequency of the rolling element. It can be seen from Figure 9 that the weak fault characteristic frequencies can be effectively extracted under the three loads, the frequency amplitude decreases with the increase of the load, and the bearing fault feature frequency amplitude under load 2 hp is minimum. In addition, it can be seen from Table 2 that the center frequency of the slight fault of the rolling body is also different under different loads. To further validate the effectiveness of VMD based on the QCFOA, the EEMD and LMD methods are used to deal with the rolling body fault signal with the pitting diameter of 0.1778 mm under load 2 hp, respectively. It can be seen from Figure 9 that the characteristic frequency of the bearing fault obtained by the EEMD method is very weak and almost drowned by other frequency components. The LMD method has improved relative to the EEMD method, but the fault characteristic frequency is still very weak. Compared with the VMD method in Figure 10, the characteristic frequency of the bearing fault is superior than that of EEMD and LMD. Therefore, the above experimental results indicate that the VMD method using the quantum chaotic FOA can

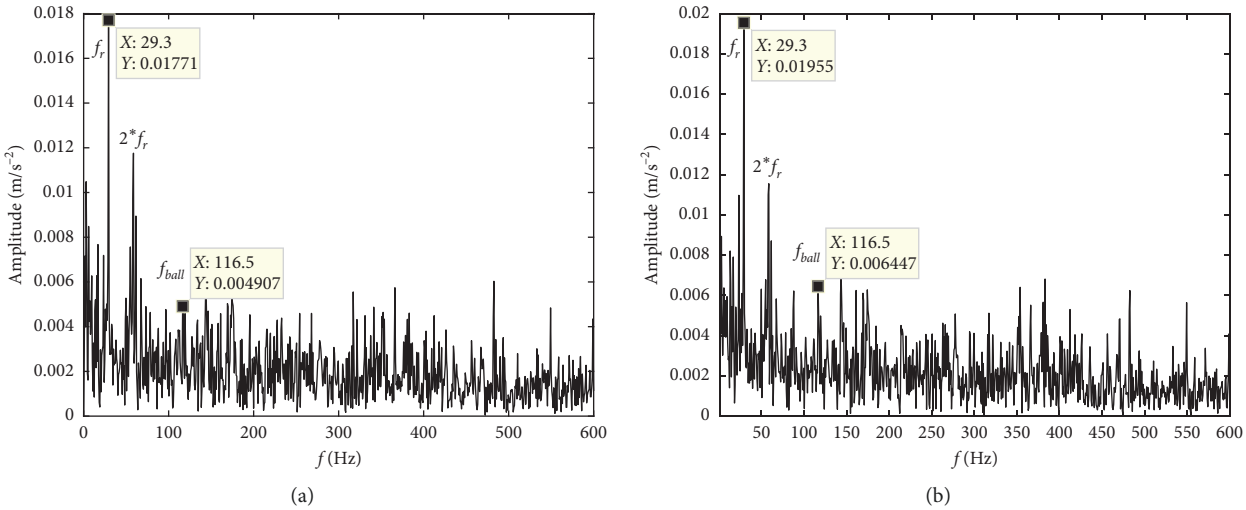


Figure 10: Marginal spectrum of minor rolling body fault under heavy load conditions. (a) EEMD decomposition. (b) LMD decomposition.

accurately extract the characteristic frequency of the weak rolling body fault under heavy loads, and the validity of the proposed method is also verified.

5.2.2. Fault Feature Extraction of Different Defects under the Same Load. The experiment is conducted on a bearing with different degrees of fault, and the collected vibration signals shown in Figure 9 are of a bearing with different defects.

It can be seen from Figure 11 that the frequency of fault features with different damage degrees under the same load can be effectively extracted, the amplitude of the frequency increases with the increase of the degree of rolling damage, and the frequency amplitude of the minor damage fault features is the least. The center frequencies of rolling body fault with different damage degrees used for the experiment are given in Table 3, and the changes are minor.

5.3. Selection of Eigenvectors. The marginal spectrum can accurately reflect the distribution of the actual frequency components of the signal and the degree of uncertainty of the signal spectrum. The smaller the marginal spectral entropy of the signal is, the greater the concentration of the energy spectrum of the signal is and the more concentrated the marginal spectrum of the signal is. The analysis of the marginal spectrum energy can effectively reflect the working state of the bearing. From Table 2, it is seen that there is a small difference in the LMMSE of the same fault type and the fault level at different loads, but the difference in their center frequency values is large. It can be seen from Table 3 that the LMMSE under the same load is significantly different for the same fault type and different fault degrees, but the difference between their central frequencies is small. The experimental data show that the magnitude of load and the degree of fault have a certain effect on the marginal spectrum entropy of failure. Therefore, the marginal spectrum entropy of a single IMF component is difficult to accurately characterize the degree of fault under variable load conditions. For the above reasons, the center

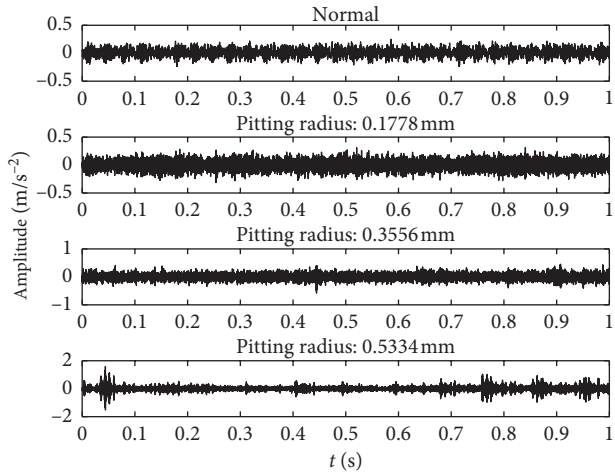


Figure 11: Vibration signals under different pitting radius.

Table 3: Experimental data under different fault degrees.

Damage degree	Minor	Moderate	Severe
Fault frequency calculated value (Hz)	116.61	116.61	116.61
Measured value (Hz)	116.5	116.5	116.5
[α, K]	[1800, 6]	[1600, 6]	[1400, 8]
Optimal IMF	IMF4	IMF4	IMF6
Normalization center frequency	0.4586	0.4624	0.4667
LMMSE	0.4875	0.3994	0.3211

frequency is combined with the basis of one-dimensional MSE to extend it into two-dimensional MSE in this paper, and its central frequency can be obtained through equation (2).

In order to verify the effectiveness and robustness of the proposed method under variable load conditions, the experimental data under 1 hp load are used as a training sample, with 0 hp and 2 hp representing the unknown load, and the experimental data are used as a test set. The length of each data sample is 4096 points. The detailed data description is shown in Table 4.

Table 4: Experimental data under variable load conditions.

Training samples (1 hp)	Test samples (0 hp + 2 hp)	Pitting radius (mm)	Bearing state	Fault identifier
40	40	0	Normal	0
80	80	0.1778	Outer race	1
80	80	0.3556	Outer race	2
80	80	0.5334	Outer race	3
80	80	0.1778	Rolling element	4
80	80	0.3556	Rolling element	5
80	80	0.5334	Rolling element	6
80	80	0.1778	Inner race	7
80	80	0.3556	Inner race	8
80	80	0.5334	Inner race	9

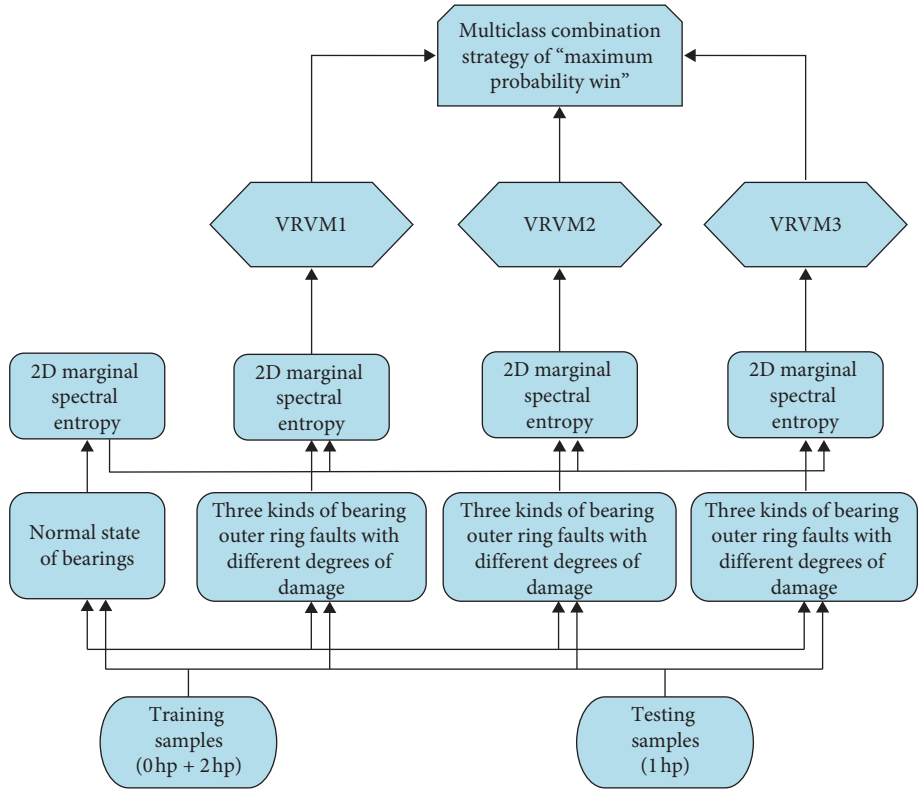


Figure 12: Variable load fault diagnosis process based on VRVM.

5.4. Fault Diagnosis

5.4.1. The Process of Intelligent Diagnosis. First, the sample data are decomposed by VMD, and the corresponding IMF components are obtained. The two-dimensional MSE synthesized by marginal spectrum entropy and the central frequency corresponding to each IMF component is taken as the eigenvector. Then, the most commonly used RBF kernel function is selected as the kernel function of VRVM. The parameters of the kernel function are optimized through 10-fold cross-validation. The detailed diagnosis process is shown in Figure 12.

In Figure 12, three multiclassifiers of VRVM are used. Although the probit model is proposed in this paper, it can easily extend the binary classification to the multiple classification and avoid complex approximate calculation. However, as the number of classes increases, the Hessian matrix in the process of constructing the model will also increase, resulting in an increase in computational complexity. Therefore, in this paper, three VRVM multiclassifiers are constructed and a multiclassification intelligent diagnosis model is constructed based on the combination strategy of "maximum probability win".

5.4.2. Diagnostic Results. In order to verify the validity and robustness of the proposed method in intelligent fault diagnosis under variable load conditions, the fault data of different damage degrees under the same load and the fault data with slight damage under different radial loads are tested for the diagnosis model in this section. The three different levels of damage are as follows: (1) minor damage, (2) moderate damage, and (3) severe damage. The three different radial loads are 0 hp, 1 hp, and 2 hp, respectively. 0 hp represents the radial load force of 0 N, 1 hp represents the radial load force of 400 N, and 2 hp represents the radial load force of 800 N.

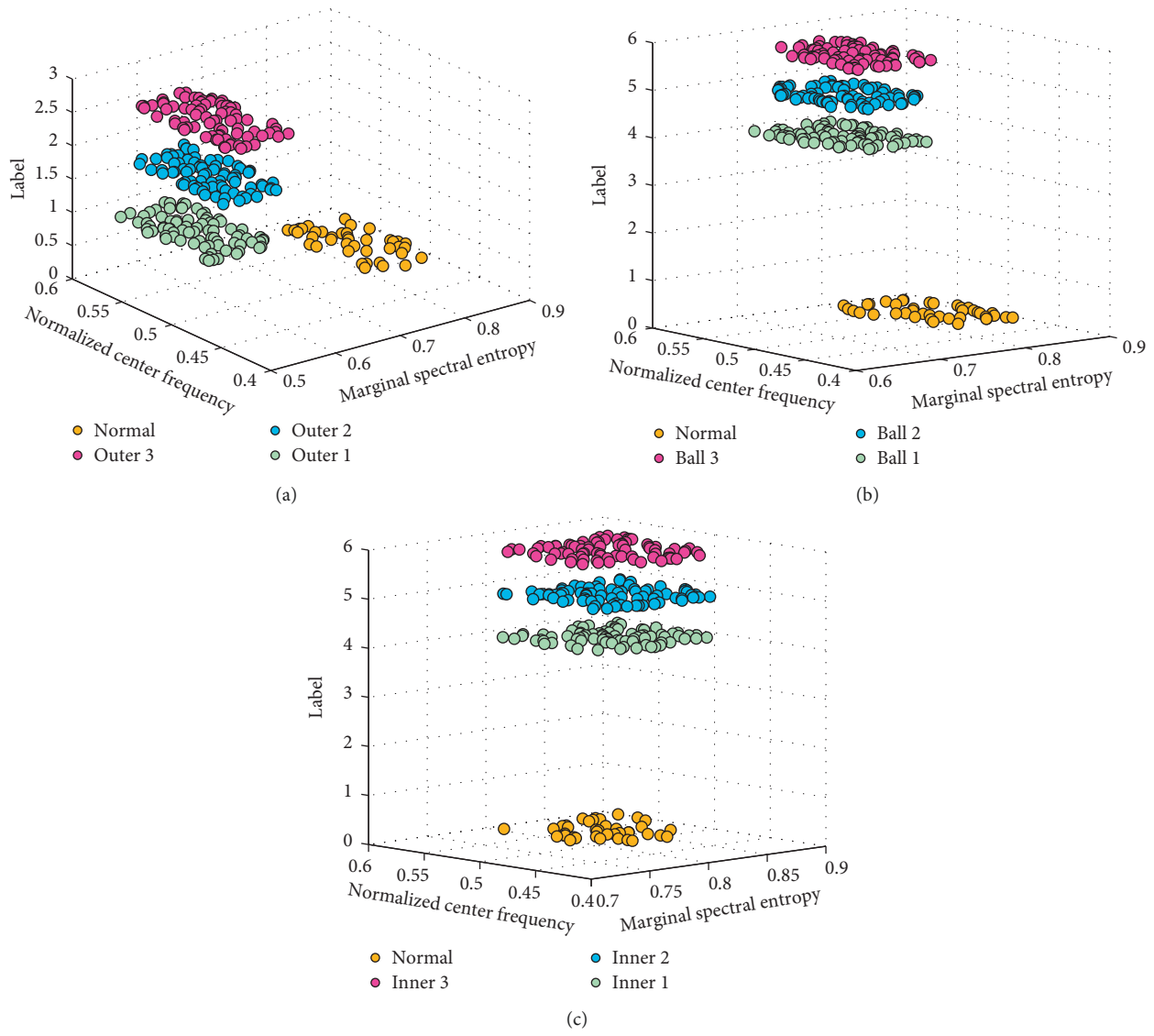


Figure 13: Intelligent diagnosis of different bearing faults under variable load. Diagnosis results of three different damage degrees of the bearing outer race under 0 hp load (a), bearing rolling balls under 1 hp load (b), and bearing inner race under 2 hp load (c).

According to the proposed bearing fault intelligent diagnosis procedure, the diagnosis results of bearing outer race damage with different degrees under 0 hp load are shown in Figure 13(a). It can clearly distinguish the running data of bearing outer race under normal and three different damage degree conditions. Similarly, the diagnosis results of bearing rolling balls under 1 ph load and bearing inner race under 2 hp load are shown in Figures 13(b)–13(c), respectively. Under different loading conditions, the fault diagnosis results of the bearing inner ring, rolling body, and outer ring with slight damage are shown in Figure 14.

According to the proposed procedure shown in Figures 13 and 14, fault types and degrees can be effectively identified under variable loads with the intelligent diagnosis method. The fault recognition rate is 100%, and the fault degree recognition rate also reaches 90%; the overall recognition rate is 95%, which proves the validity of the method. Meanwhile, VRVM is compared with other classification methods in terms of classification accuracy and running time. The results are shown in Table 5. The experiment results show that the VRVM method has a good performance in classification accuracy, but the running time is longer. In practice, equipment fault samples are often scarce, and it is even more difficult to get data of different fault types under the same load. It is a practical significance to identify fault types and fault levels under unknown loads by using fault data under known loads.

6. Conclusions

In this paper, the optimal variational mode decomposition (VMD) based on the quantum chaotic fruit fly optimization algorithm and variational relevance vector machine (VRVM) are combined as a hybrid method to diagnose the bearings' fault under variable load conditions.

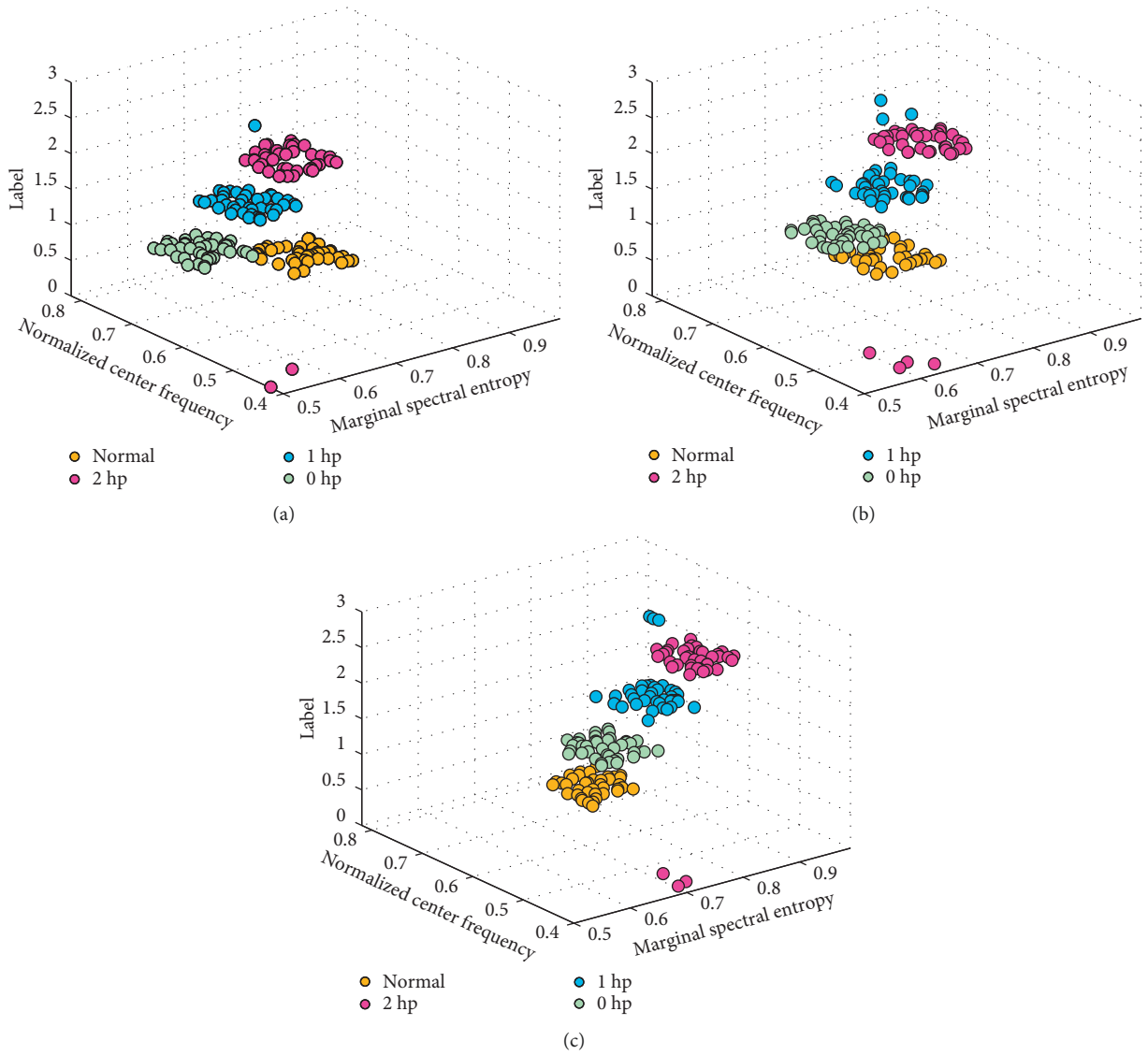


Figure 14: Intelligent diagnosis of bearing faults with minor damage under three different load conditions. (a) Bearing outer race fault. (b) Bearing rolling ball fault. (c) Bearing inner race fault.

Table 5: Performance comparison of different classification methods.

Algorithms	Classification accuracy (%)		Run times (s)
	Testing precision	Training precision	
LSSVM	71.0	73.1	27.1
FSVM	80.6	82.1	24.1
KFCM-FSVM	84.3	87.8	36.2
RVM	82.4	84.7	34.8
M-RVM	83.2	88.2	38.7
V-RVM	89.1	95.0	31.6

The results of the experiment and application demonstrate the superiority of the proposed method. The conclusions are summarized as follows:

(1) The VMD method is a new adaptive signal processing method. In the process of bearing fault signal processing, its performance is influenced by the two parameters such as the number of its own components and the penalty factor. Therefore, using the quantum chaotic fruit fly optimization algorithm to filter the two key parameters of the optimal value can guarantee the effectiveness and reliability of VMD performance.

(2) FOA is a new swarm intelligence optimization algorithm. Its principle is simple and easy to implement and has strong local search capability; however, its global search is weak. If the initial value is not set properly, it will easily fall into local minimum, thus losing population diversity and premature convergence. Compared with the traditional chaotic system, the quantum chaotic system has the characteristics of better aperiodicity, ergodicity, and class randomness and more sensitivity to system parameters and initial conditions. Therefore, the location of the fruit fly population is initialized by the proposed quantum chaotic system, which can improve the diversity of

the population, strengthen the ergodicity of the search, avoid the search process falling into the local optimal value prematurely, and improve the search efficiency.

(3) The posterior distribution of all parameters and superparameters is obtained by using RVM. Then, the probit model is used to replace the logistic model in the original variational RVM classification so that the classification and regression are organically combined. The approximate deduction of the logistic model from continuous output to discrete output mapping is avoided so that the reasoning algorithm of the RVM regression model can be directly applied to the classification model.

(4) Usually, the fault diagnosis model is trained by the traditional diagnosis method under a specific load, and it has a great limitation. In practice, most mechanical equipment work under variable load conditions. A single marginal spectrum entropy cannot effectively characterize the degree of the fault under variable load conditions. Therefore, by introducing the central frequency of VMD, the one-dimensional marginal spectrum entropy is extended to two-dimensional marginal spectrum entropy as the learning sample of VRVM.

(5) To develop an intelligent fault diagnosis model which is a systematic problem, it includes experimental condition design, feature extraction, feature selection, and model training, and each step will affect the validity of the final model. The method proposed in this paper provides an effective diagnostic strategy for multiclass faults and fault degree diagnosis under variable load conditions. However, identification of the accuracy and running time of bearing damage recognition needs to be improved in this paper, which is also a problem that needs further study in the later stage.

Data Availability

All data included in this study are available upon request by contacting the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61174106 and 51707079), Research Project of Hubei Provincial Department of Education, China (No. B2016006), and Hubei Provincial Natural Science Foundation of China (No. 2016CFB463).

References

[1] M. Kang, J. Kim, L. M. Wills, and J.-M. Kim, "Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7749–7761, 2015.

[2] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, and Q. Zheng, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings Mathematical Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[3] Y. P. Cai, G. H. Xu, A. H. Li, and T. Wang, "An improved EMD algorithm based on local integral average constraint cubic spline interpolation by adding dense points and its application in rotary machine fault diagnosis," *Journal of Vibration and Shock*, vol. 35, no. 19, pp. 81–87, 2016.

[4] A. Humeau-Heurtier, M. A. Colominas, G. Schlotthauer, M. Etienne, L. Martin, and P. Abraham, "Bidimensional unconstrained optimization approach to EMD: an algorithm revealing skin perfusion alterations in pseudoxanthoma elasticum patients," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 233–239, 2017.

[5] T. Guo and Z. Deng, "An improved EMD method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing," *Applied Acoustics*, vol. 127, pp. 46–62, 2017.

[6] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.

[7] Y. Li, M. Xu, R. Wang, and W. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy," *Journal of Sound and Vibration*, vol. 360, pp. 277–299, 2016.

[8] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[9] K. Li, L. Su, and J. J. Wu, "A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine," *Applied Sciences-Basel*, vol. 7, no. 10, p. 1004, 2017.

[10] Y. L. Jiang and F. Y. Zhang, "Clock synchronization algorithm based on particle swarm optimization with natural selection," *Journal of Southwest Jiaotong University*, vol. 52, no. 3, pp. 593–599, 2017.

[11] C. C. Yi, Y. Lv, and Z. Dang, "A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition," *Shock and Vibration*, vol. 2016, Article ID 9372691, 10 pages, 2016.

[12] H. L. Zhang, W. Luo, X. M. Liu, and Y. He, "Measurement of soil organic matter with near infrared spectroscopy combined with genetic algorithm and successive projection algorithm," *Spectroscopy and Spectral Analysis*, vol. 37, no. 2, pp. 584–587, 2017.

[13] M. F. Zeng, S. Y. Chen, W. Q. Zhang, and C. H. Nie, "Generating covering arrays using ant colony optimization: exploration and mining," *Journal of Software*, vol. 27, no. 4, pp. 855–878, 2017.

[14] S. Kang, M. Kim, and J. Chae, "A closed loop based facility layout design using a cuckoo search algorithm," *Expert Systems With Applications*, vol. 93, pp. 322–335, 2018.

[15] H. Dai, G. Zhao, J. Lu, and S. Dai, "Comment and improvement on "A new fruit fly optimization algorithm: taking the financial distress model as an example"," *Knowledge-Based Systems*, vol. 59, pp. 159-160, 2014.

[16] W.-T. Pan, "A new fruit fly optimization algorithm: taking the financial distress model as an example," *Knowledge-Based Systems*, vol. 26, no. 2, pp. 69–74, 2012.

[17] F. Ye, X. Y. Lou, and L. F. Sun, "Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data," *PLoS One*, vol. 12, no. 12, Article ID e0188746, 2017.

[18] X. H. Liu, Y. Shi, and J. Xu, "Parameters Tuning approach for proportion integration differentiation controller of magnetorheological fluids brake based on improved fruit fly optimization algorithm," *Symmetry*, vol. 9, no. 7, p. 109, 2017.

[19] A. Akhshani, A. Akhavan, S.-C. Lim, and Z. Hassan, "An image encryption scheme based on quantum logistic map," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 12, pp. 4653–4661, 2012.

[20] P. Jayaswal, S. Verma, and A. Wadhwani, "Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis," *Journal of Vibration and Control*, vol. 17, no. 8, pp. 1131–1148, 2010.

[21] A. Moosavian, H. Ahmadi, A. Tabatabaeefar, and M. Khazaee, "Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing," *Shock and Vibration*, vol. 20, no. 2, pp. 263–272, 2013.

[22] K. C. Ashwani and K. P. Raj, "bearing fault classification based on wavelet transform and artificial neural network," *IETE Journal of Research*, vol. 59, no. 3, pp. 219–225, 2013.

[23] Z. Zhang, "Automatic fault prediction of wind turbine main bearing based on SCADA data and artificial neural network," *Open Journal of Applied Sciences*, vol. 08, no. 6, pp. 211–225, 2018.

[24] B. Pramod and B. T. Lina, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 980–987, 2015.

[25] S. R. Luo, J. S. Cheng, and H. L. Ao, "Application of LCD-SVD technique and CRO-SVM method to fault diagnosis for roller bearing," *Shock and Vibration*, vol. 2015, Article ID 847802, 8 pages, 2015.

[26] H. L. Ao, J. S. Cheng, K. L. Li, and T. K. Truong, "A roller bearing fault diagnosis method based on LCD energy entropy and ACROA-SVM," *Shock and Vibration*, vol. 2014, Article ID 825825, 12 pages, 2014.

[27] S.-w. Fei, "fault diagnosis of bearing based on wavelet packet transform-phase space reconstruction-singular value decomposition and SVM classifier," *Arabian Journal for Science and Engineering*, vol. 42, no. 5, pp. 1967–1975, 2017.

[28] Y. Li, W. Zhang, Q. Xiong, D. Luo, G. Mei, and T. Zhang, "A rolling bearing fault diagnosis strategy based on improved multiscale permutation entropy and least squares SVM," *Journal of Mechanical Science and Technology*, vol. 31, no. 6, pp. 2711–2722, 2017.

[29] X. Liu, L. Bo, and H. Luo, "Bearing faults diagnostics based on hybrid LS-SVM and EMD method," *Measurement*, vol. 59, pp. 145–166, 2015.

[30] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[31] W. L. Zhao, Z. L. Wang, J. Ma, and L. F. Feng, "Fault diagnosis of a hydraulic pump based on the CEEMD-STFT time-frequency entropy method and multiclass SVM classifier," *Shock and Vibration*, vol. 2016, Article ID 2609856, 8 pages, 2016.

[32] C. M. Bishop and M. E. Tipping, "Variational relevance vector machines," *Uncertainty in Artificial Intelligence proceedings*, vol. 28, no. 3, pp. 46–53, 2000.

[33] A. A. A. El-Latif, L. Li, N. Wang, Q. Han, and X. Niu, "A new approach to chaotic image encryption based on quantum chaotic system, exploiting color spaces," *Signal Processing*, vol. 93, no. 11, pp. 2986–3000, 2013.

[34] M. E. Goggin, B. Sundaram, and P. W. Milonni, "Quantum logistic map," *Physical Review A*, vol. 41, no. 10, pp. 5705–5708, 1990.

[35] A. Akhshani, A. Akhavan, A. Mobaraki, S.-C. Lim, and Z. Hassan, "Pseudo random number generator based on quantum chaotic map," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 1, pp. 101–111, 2014.

[36] M. E. Goggin, B. Sundaram, and P. W. Milonni, "Quantum logistic map," *Physical Review A*, vol. 41, no. 10, pp. 5705–5708, 1990.

[37] A. Tajima, A. Tanaka, W. Maeda, S. Takahashi, and A. Tomita, "Practical quantum cryptosystem for metro area applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 13, no. 4, pp. 1031–1038, 2007.

[38] D. X. Niu, H. C. Wang, H. Y. Chen, and Y. Liang, "The general regression neural network based on the fruit fly optimization algorithm and the data inconsistency rate for transmission line icing prediction," *Energies*, vol. 10, no. 12, p. 2066, 2017.

[39] M. Yang, N. B. Liu, and W. Liu, "Image 1D OMP sparse decomposition with modified fruit-fly optimization algorithm," *Cluster Computing*, vol. 20, no. 4, pp. 1–8, 2017.

[40] M. Girolami and S. Rogers, "Variational Bayesian multinomial probit regression with Gaussian process priors," *Neural Computation*, vol. 18, no. 8, pp. 1790–1817, 2006.

[41] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[42] C. M. Bishop and M. E. Tipping, "Variational relevance vector machines," in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 46–53, San Francisco, CA, USA, June-July 2000.

[43] X. Zhou, X. Wang, and E. R. Dougherty, "Multi-class cancer classification using multinomial probit regression with bayesian gene selection," *IEEE Proceedings-Systems Biology*, vol. 153, no. 2, p. 70, 2006.

[44] D. Y. Zhou, "Radar target recognition based on kernel projection vector space using high-resolution range profile," in *Proceedings of 3th International Conference on Intelligent System Design and Engineering Applications*, pp. 1077–1080, Hong Kong, China, January 2013.