Research Article

Blind Parameter Identification of MAR Model and Mutation Hybrid GWO-SCA Optimized SVM for Fault Diagnosis of Rotating Machinery

Wenlong Fu ^{1,2}, Jiawen Tan ^{1,2}, Xiaoyuan Zhang,³ Tie Chen,^{1,2} and Kai Wang^{1,2}

¹College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, 443002, China

²Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang, 443002, China

³College of Electrical Engineering, Henan University of Technology, Zhengzhou, 450001, China

Correspondence should be addressed to Wenlong Fu; ctgu_fuwenlong@126.com

Received 14 January 2019; Accepted 27 March 2019; Published 17 April 2019

Academic Editor: Marcin Mrugalski

Copyright © 2019 Wenlong Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As a crucial and widely used component in industrial fields with great complexity, the health condition of rotating machinery is directly related to production efficiency and safety. Consequently, recognizing and diagnosing rotating machine faults remain to be one of the main concerns in preventing failures of mechanical systems, which can enhance the reliability and efficiency of mechanical systems. In this paper, a novel approach based on blind parameter identification of MAR model and mutation hybrid GWO-SCA optimization is proposed to diagnose faults for rotating machinery. Signals collected from different types of faults were firstly split into sets of intrinsic mode functions (IMFs) by variational mode decomposition (VMD), the decomposing mode number K of which was preset with central frequency observation method. Then the multivariate autoregressive (MAR) model of all IMFs was established, whose order was determined by Schwartz Bayes Criterion (SBC), and all parameters of the model were identified blindly through QR decomposition, where key features were subsequently extracted via principal component analysis (PCA) to construct feature vectors of different fault types. Afterwards, a hybrid optimization algorithm combining mutation operator, grey wolf optimizer (GWO), and sine cosine algorithm (SCA), termed mutation hybrid GWO-SCA (MHGWOSCA), was proposed for parameter selection of support vector machine (SVM). The optimal SVM model was later employed to classify different fault samples. The engineering application and contrastive analysis indicate the availability and superiority of the proposed method.

1. Introduction

Rotating machinery is a crucial part in modern industrial manufacture, and its failure may result in serious safety accidents and economic losses. Therefore, one of the major topics to be investigated in preventing failures of mechanical systems is recognizing and diagnosing rotating machine faults [1, 2]. In various rotating machineries, rolling bearings are the most widely used components, which are greatly related to the remaining lifetime of machineries [3]. Unfortunately, rolling bearings are prone to failure to some extent because of complex operating conditions and structural characteristics [4]. Given this fact, effective and feasible fault diagnosis methods need to be developed for rolling bearings, to promote the reliability of rotating machinery [5].

Owing to rich information carried by vibration signals, most of fault diagnosis methods for rolling bearings rely on analyzing vibration signals [6]. However, vibration signal is commonly nonstationary in actual operation, which restricts the extraction effect for fault features, and thus affects accuracy of fault diagnosis. Addressing this issue, many time-frequency signal analysis methods are proposed to analyze nonstationary signal, including empirical mode decomposition (EMD) [7], ensemble empirical mode decomposition (EEMD) [8], empirical wavelet transform (EWT) [9] and variational mode decomposition (VMD) [10]. EMD has advantages in dealing with nonstationary signal and does not need to preset any basis function, but its performance is affected by end effects and mode mixing. To overcome these defects, EEMD as an improved version of EMD is proposed

by introducing noise-assisted analysis method, which is also increasing the computational cost and cannot completely neutralize the added noise. EWT is a self-adaptive signal decomposition method that inherits advantages of EMD and wavelet transform, but the selection of an appropriate wavelet filter bank remains to be a problem. Unlike the methods mentioned above, VMD is a quasi-orthogonal signal decomposition method and its main structure is to construct and solve a constrained variational problem. It can effectively realize the separation of each frequency component in signal, avoiding the problem of mode mixing or filter selection. Its advancement and effectiveness have been demonstrated in many previous studies [11, 12]. As the nonstationarity of signal weakened by above signal processing methods, fault features can be smoothly extracted. Due to the high efficiency and consideration of variable correlation, multivariate autoregressive (MAR) model [13] is introduced to extract features from vibration signal, which has been widely adopted for feature extraction and shows superior performance [14, 15]. Further, key components among the features extracted from MAR model are screened out via principal component analysis (PCA) to finally construct feature vectors [16].

The essence of fault diagnosis for rolling bearings is a pattern recognition problem; that is, the fault type is determined by fault features of samples. In engineering application, many alternative methods are available for addressing this issue. For instance, Bayesian decision has strong recognition ability by considering prior probability as well as class conditional probability, and its good accuracy requires assumption of a suitable prior model [17]. K-nearest neighbor (KNN) based on Euclidean or Manhattan distances is easy to realize and susceptible to distribution of samples [18]. Artificial neural network (ANN) has good ability in processing recognition problem with the support of a large number of samples [19]. Based on statistical learning and structural risk minimization theory, support vector machine (SVM) is an excellent machine learning method. SVM finds an optimal hyperplane that meets the classification requirements, where the hyperplane maximizes the interval between two classes, thus promoting classification performance. Currently, SVM combined with feature extraction method has been successfully employed in the field of pattern recognition [20, 21].

Although SVM has superior ability in pattern recognition, its performance is affected by parameters. In view of this, various optimization algorithms are developed and applied to search the best parameters, such as genetic algorithm (GA) [22], particle swarm optimization (PSO) [23], sine cosine algorithm (SCA) [24], and grey wolf optimizer (GWO) [25]. As novel optimization algorithms, SCA and GWO proposed by Mirjalili et al. have proved effective in many previous works [26–28]. To achieve better performance in convergence precision, a mutation hybrid GWO-SCA (MHGWOSCA) optimization method is proposed in this paper. To sum up, based on blind parameter identification of MAR model and mutation hybrid GWO-SCA optimized SVM, a fault diagnosis method for rotating machinery is presented in the research. Firstly, VMD was applied to decompose the vibration signal into a set of intrinsic mode functions (IMFs), and the decomposing mode number K

was preset with central frequency observation method. Then the MAR model of all IMFs was established, whose order was ascertained by Schwartz Bayes Criterion (SBC). And all parameters within the model were identified blindly through QR decomposition, where the key features were subsequently extracted via PCA to construct feature vectors of different fault samples. Afterwards, a hybrid optimization algorithm combining mutation operator, GWO, and SCA theory, called mutation hybrid GWO-SCA (MHGWOSCA), was proposed for parameter selection of SVM. Finally, the optimal SVM model was employed to classify different fault samples. The engineering application and contrastive analysis reveal the availability and superiority of the proposed method.

The paper is organized as follows: Section 2 is dedicated to the basic knowledge of VMD, MAR, and SVM. Section 3 introduces the proposed fault diagnosis method based on blind parameter identification of MAR model and mutation hybrid GWO-SCA optimized SVM. Section 4 presents the engineering application, the result analysis of which demonstrates the availability of the proposed method. The conclusion is summarized in Section 5.

2. Fundamental Theories

2.1. Variational Mode Decomposition. Variational Mode Decomposition (VMD) is a new adaptive signal preprocessing method [29]. By setting a mode number K in advance, the given signal can be decomposed into K band-limited intrinsic mode functions (IMFs). The cyclic iterative method is used to obtain the optimal solution of constrained variational problem, through which the frequency center and bandwidth of each mode are determined. The constrained variation problem can be described as follows:

$$\begin{aligned} \min_{m_k, w_k} \quad & \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ \text{s.t.} \quad & \sum_{k=1}^K m_k(t) = f(t), \quad k = 1, 2, \dots, K \end{aligned} \quad (1)$$

where $m_k = \{m_1, m_2, \dots, m_k\}$ represents the set of K mode functions and $w_k = \{w_1, w_2, \dots, w_k\}$ represents the set of central frequencies, while ∂_t and $\delta(t)$ are the partial derivative of time t for the function and unit pulse function, respectively. $f(t)$ is the given real valued input signal.

For the above variational problem, quadratic penalty function term and Lagrange multiplier are employed to transform it into an unconstrained problem. Then problem (1) can be specified as follows:

$$\begin{aligned} L(m_k, w_k, \beta) &= \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ &+ \left\| f(t) - \sum_k m_k(t) \right\|_2^2 \\ &+ \left\langle \beta(t), f(t) - \sum_k m_k(t) \right\rangle \end{aligned} \quad (2)$$

where α and $\beta(t)$ represent penalty factor and Lagrange multiplier, respectively, and $f(t)$ is the given signal.

Then alternate direction method is utilized to obtain the saddle point of Lagrange multiplier, that is, solving (2) by optimizing m_k , w_k , and $\beta(t)$ alternately [30]. The optimization problems of m_k and w_k are formulated as (3) and (4), respectively.

$$m_k^{n+1}(t) = \min \left\{ \alpha \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-jw_k t} \right\|_2^2 + \left\| f(t) - \sum_i m_i(t) + \frac{\beta(t)}{2} \right\|_2^2 \right\} \quad (3)$$

$$w_k^{n+1} = \min \left\{ \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \quad (4)$$

Solving problems (3) and (4), the iterative equations are deduced as follows.

$$\hat{m}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{m}_i(w) + \hat{\beta}(w)/2}{1 + 2\alpha(w - w_k)^2} \quad (5)$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{m}_k(w)|^2 dw}{\int_0^\infty |\hat{m}_k(w)|^2 dw} \quad (6)$$

The Lagrange multipliers can be iterated with

$$\hat{\beta}^{n+1}(w) = \hat{\beta}^n(w) + \tau \left(f(w) - \sum_k \hat{m}_k^{n+1}(w) \right) \quad (7)$$

where τ is an updating parameter.

The main steps of VMD can be summarized as follows:

Step 1. Initialize $m_k^1, w_k^1, \beta^1, n=1$.

Step 2. Execute loop, $n=n+1$.

Step 3. Update m_k and w_k based on (5) and (6).

Step 4. Update β based on (7).

Step 5. If $\sum_k \|\hat{m}_k^{n+1} - \hat{m}_k^n\|_2^2 / \|\hat{m}_k^n\|_2^2 < \varepsilon$, loop end, else turn to Step 2 for next iteration.

2.2. Multivariate Autoregressive Model. Autoregressive (AR) model deduces the regressive variables with itself, namely, using a linear combination of white noise at time n and values of previous p times to describe the output of system at time n . The equation can be described as follows [31]:

$$x_n = \alpha_1 x_{n-1} + \alpha_2 x_{n-2} + \cdots + \alpha_p x_{n-p} + e_n \quad (8)$$

where α_i ($i = 1, 2, \dots, p$) is the autoregressive coefficient, e_n is the white noise with mean 0 and variance σ^2 , and p is the order of model AR(p).

When multivariate modeling is required, AR model must be analyzed and established multiple times. For this

reason, the model order determined by different variables may be inconsistent. Compared with AR model, multivariate autoregressive (MAR) model only models once and can reflect the internal relationship among different variables. Let x_n^k ($k = 1, 2, \dots, m$) be the value of the k -th variable at time n ; obviously it relates to not only the values of previous p times and the white noise at time n , but also the values of previous p times of other variables [13]. The equation of MAR model can be described as

$$\begin{aligned} x_n^1 &= \sum_{i=1}^p \alpha_{1i}^1 x_{n-i}^1 + \sum_{i=1}^p \alpha_{1i}^2 x_{n-i}^2 + \cdots + \sum_{i=1}^p \alpha_{1i}^m x_{n-i}^m + e_n^1 \\ x_n^2 &= \sum_{i=1}^p \alpha_{2i}^1 x_{n-i}^1 + \sum_{i=1}^p \alpha_{2i}^2 x_{n-i}^2 + \cdots + \sum_{i=1}^p \alpha_{2i}^m x_{n-i}^m + e_n^2 \\ &\vdots \\ x_n^m &= \sum_{i=1}^p \alpha_{mi}^1 x_{n-i}^1 + \sum_{i=1}^p \alpha_{mi}^2 x_{n-i}^2 + \cdots + \sum_{i=1}^p \alpha_{mi}^m x_{n-i}^m + e_n^m \end{aligned} \quad (9)$$

where p is the order of MAR model, α_{ki}^m is the autoregressive coefficient for the k -th variable when the m -th variable is delayed with step i , and e_n^m is the random error of the m -th variable.

2.3. Support Vector Machine. Support vector machine (SVM) is a machine learning model developed by Vapnik [32]. It can deduce the optimal solution between model complexity and learning ability with limited information, so as to obtain the best classification accuracy. SVM has unique advantages in solving problems of limited samples, nonlinear and high-dimensional data. For a given sample set $\{(x_i, y_i) \mid i = 1, 2, \dots, n\}$, the main idea of SVM is to map samples into a high-dimensional feature space and find a hyperplane to solve the corresponding nonlinear problem in low-dimensional space. The hyperplane function can be constructed as

$$w \cdot x + b = 0 \quad (10)$$

where w is weight vector and b is bias parameter, while $w \cdot x$ represents the inner product of w and x .

Taking binary classification issue as an example, to classify samples correctly with sufficient classification interval, all samples are required to meet the following constraint.

$$w \cdot x_i + b \begin{cases} > 1 & \text{for } y_i = 1 \\ < -1 & \text{for } y_i = -1 \end{cases} \quad (11)$$

In the sequel, classification interval can be calculated as $2/\|w\|^2$, and maximizing the interval is equivalent to minimizing $\|w\|^2$. To solve the linear indivisibility problem, slack term ξ and penalty factor C are brought into (11), and hence construction of the optimal hyperplane is transformed

into a constrained optimization problem.

$$\begin{aligned} \min \quad & f = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i (w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, n \end{aligned} \quad (12)$$

When mapping samples to high-dimensional space, different inner product kernel functions will form different algorithms. Among them, radial basis function is widely employed in application of pattern recognition, which can be described as

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = \exp(-g \|x_i - x_j\|^2) \quad (13)$$

where g is kernel parameter.

The solution of constrained optimization problem (12) is determined by saddle point of Lagrange function, and this quadratic programming problem is transformed into the corresponding dual problem; that is,

$$\begin{aligned} \max \quad & L = \sum_{i=1}^n \mu_i - \frac{1}{2} \sum_{i,j=1}^n \mu_i \mu_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad & \sum_{i=1}^n \mu_i y_i = 0, \quad \mu_i \geq 0, \quad i = 1, 2, \dots, n \end{aligned} \quad (14)$$

where μ_i is Lagrange multiplier.

With the optimal value of μ_i obtained from above dual problem, the optimal classification discriminant function can be ascertained as follows.

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \mu_i K(x_i, x) + b \right) \quad (15)$$

3. Fault Diagnosis by Blind Parameter Identification of MAR Model and Mutation Hybrid GWO-SCA Optimized SVM

3.1. Blind Parameter Identification of MAR Model. Multivariate autoregressive (MAR) model is the multivariate version of autoregressive (AR) model. By synchronous autoregressive modeling and parameter identification of multiple variables, it possesses advantages of high efficiency and consideration of variable correlation. Parameter identification is a critical part of MAR analysis and appropriate parameters will contribute to a better fitting effect, during which process the order selection of MAR model not only relates to the extraction of essential features implicated in the sequence, but also has a great impact on recognition efficiency. In this research, Schwartz Bayes Criterion (SBC) is adopted to estimate the order of MAR model. The SBC function is as follows [33]:

$$S(p) = \frac{l_p}{m} - \left(1 - \frac{n_p}{N}\right) \ln N \quad (16)$$

where $n_p = mp + 1$, $l_p = \ln \det d_p(R_{22}^T R_{22})$, and R_{22} is the fourth part of the upper triangular matrix obtained by QR

decomposition of K matrix of the model (see [33] for more details).

The specific method of order selection is that, within the given range of model order, $S(p)$ will gradually decrease with the order increasing. If $S(p)$ takes the minimum value or there is no significant decreasing in $S(p)$ as the order increases, the corresponding p value is the model order.

Considering the inner relation among different variable sequences, there will be a mass of parameters in MAR model. Generally, minimum variance criterion is used to estimate model parameters directly [34], but it would be computationally expensive for MAR model. To achieve better computational efficiency, a fast calculation method based on QR decomposition is applied to identify the parameters (see [33] for more details). Then all the identified parameters are combined to construct the feature vector of the multivariate sequences.

3.2. Mutation GWO-SCA Optimization

3.2.1. Sine Cosine Algorithm. Sine cosine algorithm (SCA) is a newly swarm optimization algorithm, which is simple in structure and easy to realize. Using only sine and cosine functions, SCA achieves good optimization effect. Exploration and exploitation are the two key phases when SCA processes optimization problem [24]. With an elaborate mechanism of exploration and exploitation, SCA is able to search for feasible solutions quickly in the searching space. Let $Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{id})^T$ be the position of i -th ($i = 1, 2, \dots, m$) individual, whereupon each solution of the optimization problem corresponds to a position of corresponding individual in the searching space, where d is the dimension of individuals. The core updating equation of SCA is as follows [24]:

$$Z_i^{l+1} = \begin{cases} Z_i^l + r_1 \times \sin(r_2) \times |r_3 P_i^l - Z_i^l|, & r_4 < 0.5 \\ Z_i^l + r_1 \times \cos(r_2) \times |r_3 P_i^l - Z_i^l|, & r_4 \geq 0.5 \end{cases} \quad (17)$$

where $P_i = (P_{i1}, P_{i2}, \dots, P_{id})^T$ is the best position of individual i . The parameter $r_1 = a - l(a/l_{\max})$ determines the region of individual i at next iteration, where $l_{\max}$, l , and a are the maximum number of iterations, the current number of iterations, and a constant, respectively. The parameter r_2 is a random number in the scope of $[0, 2\pi]$, which defines the distance that next movement should be towards. To stochastically emphasize ($r_3 > 1$) or deemphasize ($r_3 < 1$) the effect of the best position of individual, a random weight r_3 within $[0, 2]$ is brought in the equation. Finally, the parameter r_4 is a random number in the range of $[0, 1]$ to switch fairly between sine and cosine components.

3.2.2. Grey Wolf Optimizer. Grey wolf optimizer is a swarm intelligence algorithm based on strict organization system of grey wolves and its sophisticated cooperative hunting method. GWO algorithm simulates the hierarchy and hunting behavior of grey wolves in the wild. Wolves are divided into four subgroups, i.e., α , β , δ , and ω . The first three subgroups are in turn the three wolves with the best fitness values, which guide other wolves (ω) to search for the target.

During the optimization process, the wolves update the positions of α , β , δ , and ω . The updating equations of grey wolves can be described as [25]

$$\begin{aligned} \text{Distance updating: } D_\alpha &= |C_1 \cdot X_\alpha(l) - X(l)| \\ D_\beta &= |C_2 \cdot X_\beta(l) - X(l)| \end{aligned} \quad (18)$$

$$\begin{aligned} \text{Position updating: } X_1(l+1) &= X_\alpha(l) - A_1 \cdot D_\alpha \\ X_2(l+1) &= X_\beta(l) - A_2 \cdot D_\beta \\ X_3(l+1) &= X_\delta(l) - A_3 \cdot D_\delta \end{aligned} \quad (19)$$

$$X(l+1) = \frac{(X_1 + X_2 + X_3)}{3} \quad (20)$$

where D_α , D_β , and D_δ are distances between the position of individual α , β , and δ and an individual, respectively, while C_1 , C_2 , and C_3 are random numbers in $[0, 2]$. $A_1 = 2a \cdot r - a$, $A_2 = 2a \cdot r - a$, and $A_3 = 2a \cdot r - a$, where $a = 2 - 2(l/l_{\max})$ is an adaptive parameter which will decrease linearly to 0 with iteration and r is a real random number in $[0, 1]$. X_α , X_β , and X_δ represent the current position of individual α , β , and δ , respectively. X indicates the final position of an individual which would be in a random place within a circle defined by individual α , β , and δ . Finally, l is the current number of iterations.

3.2.3. Mutation Hybrid GWO-SCA Optimization. In the updating strategy of GWO, the position of α is of the most importance, which guides the search behavior of other individuals. Besides in SCA, sine and cosine functions in (17) possess outstanding ability of balancing exploration and exploitation process. Hence the combination of capabilities between GWO and SCA is expected to improve the convergence performance, which makes the hybrid algorithm, grey wolf optimizer-sine cosine algorithm (GWO-SCA), powerful in both exploring different regions and exploiting promising regions. The part of individual α in (18) is improved with the updating strategy of SCA [35]:

$$\begin{aligned} D_\alpha(l+1) &= \begin{cases} r_1 \times \sin(r_2) \times |C_1 X_\alpha(l) - X(l)|, & r_4 < 0.5 \\ r_1 \times \cos(r_2) \times |C_1 X_\alpha(l) - X(l)|, & r_4 \geq 0.5 \end{cases} \end{aligned} \quad (21)$$

where values of parameter r_1 , r_2 , and r_4 are the same as in SCA, and the meaning of D_α , C_1 , X_α , and X are the same as in GWO, while l is the current time of iterations. Except for the subsequent mutation operation, the rest of the processing procedure of GWO remains unchanged.

Analysis of GWO shows that diversity of individuals is gradually losing with iteration and it is likely to result in the local optimum in the later stage of iteration, which will affect the optimization effect. To solve this problem, mutation operators are introduced to enrich diversity of individuals [36]. The new equation containing the mutation operators is applied to enhance (19):

$$\begin{aligned} X_1(l+1) &= \begin{cases} X_\alpha(l) - A_1 D_\alpha(l), & l \neq nT \\ X_1(l) [1 + G(0.5 - \text{unif})], & l = nT \end{cases} \\ X_2(l+1) &= \begin{cases} X_\beta(l) - A_2 D_\beta(l), & l \neq nT \\ X_2(l) [1 + G(0.5 - \text{unif})], & l = nT \end{cases} \\ X_3(l+1) &= \begin{cases} X_\delta(l) - A_3 D_\delta(l), & l \neq nT \\ X_3(l) [1 + G(0.5 - \text{unif})], & l = nT \end{cases} \end{aligned} \quad (22)$$

$n = 1, 2, \dots$

where G and T ($T < l_{\max}$) are the given mutation amplitude and mutation period, respectively, while unif represents a random number that conforms to the uniform distribution $U(0, 1)$. It can be observed that, during the updating process, the position of individual α , β , and δ will mutate periodically, which gives them a certain ability to jump out of local optimum even in the later stage of iterations.

The main steps of proposed mutation hybrid GWO-SCA (MHGWOSCA) are shown below:

Step 1. Initialize the population randomly in the given range of variables and set mutation parameters.

Step 2. Calculate the fitness value of each individual.

Step 3. Choose and save the best three individuals as α , β , and δ .

Step 4. Update other individuals (ω) according to (21) and (22).

Step 5. Update parameters a , A , and C .

Step 6. If the maximum number of iterations is not met, go to Step 2.

Step 7. Output the position of individual α as the optimal solution.

3.3. Diagnosis Method Based on Blind Parameter Identification of MAR Model and Mutation Hybrid GWO-SCA Optimization. In this section, to promote the fault diagnosis accuracy, a fault diagnosis method based on blind parameter identification of MAR model and mutation hybrid GWO-SCA optimized SVM is proposed. The specific steps are detailed as follows:

Step 1. Collect the vibration signals.

Step 2. Determine the mode number K of VMD by central frequency observation method.

Step 3. Decompose the samples into sets of IMFs with VMD.

Step 4. Establish the MAR model of all IMFs and apply SBC to determine the model order; then MAR parameter vectors of different fault samples can be identified by QR decomposition.

TABLE 1: Description of the experimental data.

Position of fault	Defect size (inches)	Label of classes	Number of samples
Inner race	0.007	L1	59
Inner race	0.014	L2	59
Inner race	0.021	L3	59
Ball	0.007	L4	59
Ball	0.014	L5	59
Ball	0.021	L6	59
Outer race	0.007	L7	59
Outer race	0.014	L8	59
Outer race	0.021	L9	59

TABLE 2: The confusion matrix.

		Actual label	
		Positive	Negative
Classification result	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

Step 5. Extract principal features from parameter vectors based on PCA; thus the feature vectors of different fault types are obtained.

Step 6. Optimize the parameters C and g for SVM with the proposed MHGWOSCA optimization method.

Step 7. Train the SVM model with optimal parameters C and g ; then the optimal SVM model is obtained.

Step 8. Apply the optimal SVM model to classify different fault samples and evaluate the performance of the model.

The flowchart of the proposed fault diagnosis model is shown in Figure 1.

4. Engineering Application

4.1. Data Collection. Vibration signals with different fault locations and sizes gathered from Bearings Data Center of Case Western Reserve University [37] were employed as the experiment data in this research. The experiment stand was mainly composed of a motor, an accelerometer, a torque sensor/encoder, and a dynamometer. The bearing was a deep groove ball bearing with model SKF6205-2RS. Accelerometers were applied to collect vibration signal; meanwhile, the torque sensor/encoder was for the measurement of power and speed. By using electro-discharge machining, the experiment device simulated three fault states of the rolling bearing: inner race fault, ball element fault, and outer race fault, and the depth of faults was 0.011 inches. In the experiment, vibration signals collected from the drive end (DE) were taken as the research objects, where the sample frequency was 12kHz and the rotation speed was 1750 rpm under the rated load of 2 hp. To fully verify the validity of the proposed fault diagnosis method, 9 types of samples were used in this paper, namely, inner race fault, ball fault, and outer race fault with diameters of 0.007, 0.014, and 0.021 inches (i.e., each of the three types of faults has three defect sizes). Further, vibration signal of each fault type contained

59 samples and each sample contained 2048 sample points. Image of the experiment device is shown in Figure 2, and the experimental data [5] are listed in Table 1.

4.2. Evaluation Metrics. In this application, adjusted Rand index (ARI), normalized mutual information (NMI), F-measure (F), and accuracy (ACC) as four widely used evaluation measures are employed to evaluate the diagnosis results [38], which reflect the matching degree between classification results and real sample distribution information.

Let Φ be the given actual label set and Ω be the classified result set. ARI considers the number of samples with the same labels and different labels. The scope of ARI is $[-1, 1]$ and the higher value indicates that more samples are labeled correctly. It can be expressed as

$$ARI = \frac{n_{11} + n_{00}}{C_n^2} \quad (23)$$

where n_{11} is the number of sample pairs with the same label in both Φ and Ω , while n_{00} is the number of sample pairs with different labels in Φ and Ω . C_n^2 is all possible combinations of sample pairs.

NMI is one of the common external evaluation metrics that can measure the degree of agreement between two data distributions [39]. The value of NMI lies between 0 and 1, and the higher value indicates a better classification performance. The calculation equation is as follows:

$$NMI = \frac{\sum_{\varphi \in \Phi} \sum_{\omega \in \Omega} p(\varphi, \omega) \log(p(\varphi, \omega) / p(\varphi) p(\omega))}{\sqrt{(\sum_{\varphi \in \Phi} p(\varphi) \log(p(\varphi))) (\sum_{\omega \in \Omega} p(\omega) \log(p(\omega)))}} \quad (24)$$

where $p(\varphi, \omega)$ is the joint probability function of Φ and Ω , and $p(\varphi)$ and $p(\omega)$ are the probability distribution functions of Φ and Ω , respectively.

F and ACC are derivations from the confusion matrix whose structure is reflected in Table 2 [40, 41]. F comprehensively considers the proportion of true positive result in

FIGURE 1: The flowchart of the proposed fault diagnosis method.

FIGURE 2: Experiment stand and its measurement sensors in bearing data center.

FIGURE 3: Time domain waveforms of signals with different faults: (a) inner race 0.007 inches, (b) inner race 0.014 inches, (c) inner race 0.021 inches, (d) ball 0.007 inches, (e) ball 0.014 inches, (f) ball 0.021 inches, (g) outer race 0.007 inches, (h) outer race 0.014 inches, (i) outer race 0.021 inches.

positive actual label and in positive classified result. ACC represents the proportion of all correctly classified results in total results. The scope of these two metrics is both $[0, 1]$, and the higher value indicates a better classification performance. F and ACC can be defined as (25) and (26), respectively.

$$F = 2 \frac{(TP / (TP + FP)) (TP / (TP + FN))}{TP / (TP + FP) + TP / (TP + FN)} \quad (25)$$

$$ACC = \frac{TP + TN}{TP + FN + FP + FN} \quad (26)$$

FIGURE 4: The VMD decomposition results of 0.007 inches fault signals: (a) inner race, (b) ball, (c) outer race.

4.3. Application to Fault Diagnosis of Rolling Bearing. To demonstrate the superiority of the proposed VMD-MAR-MHGWOCSA-SVM method, the experiment was carried out in a comparative form at each stage, as follows: EMD was applied for comparison in the signal processing stage; fuzzy entropy (FE) [42] and permutation entropy (PE) [5] were used in the feature extraction stage for comparison; GWO, SCA, and HGWOCSA were employed for comparison in the parameter optimization stage. The parameters of the contrastive experiments were set in the same way as the proposed method.

In the proposed method, the first step is to decompose signal of each fault type into a set of IMFs. The mode number K of VMD needs to be determined in advance with central frequency observation method. If K is too large, mode mixing would occur because central frequencies of adjacent IMFs are too close. In contrast, it is difficult to weaken the nonstationarity effectively if K is too small. Central

frequencies under different K value are listed in Table 3, where K was ascertained by the sample of inner race fault with diameter of 0.007 inches (L1). As illustrated in Table 3, when K was set to 5, the first two central frequencies were relatively close, which meant excessive decomposition had occurred; hence parameter K was set to 4. The original signal waveforms of different fault types are shown in Figure 3 and decomposition results of 0.007 inches faults (L1, L4, L7) are illustrated in Figure 4. It can be observed that waveforms from 9 types of faults are quite different and IMFs of one fault type are also discriminative to a certain extent.

When decomposed components of a sample were obtained, a multivariate autoregressive (MAR) model was established for all IMFs. The values of SBC value $S(p)$ are shown in Figure 5, where the order of MAR model was determined according to the sample of fault type L1. The figure shows that as the model order p increases, $S(p)$ decreases gradually, but the value of $S(p)$ is no longer significantly

TABLE 3: Normalized central frequencies with different K values.

Number of modes	Normalized central frequencies									
2	0.2256	0.0907								
3	0.2978	0.2275	0.0999							
4	0.2471	0.1738	0.0860	0.0127						
5	0.3066	0.2834	0.2271	0.1148	0.0525					
6	0.3072	0.2854	0.2364	0.2142	0.1139	0.0519				
7	0.3170	0.2981	0.2757	0.2270	0.1242	0.0982	0.0493			
8	0.3178	0.2990	0.2777	0.2363	0.2147	0.1231	0.0975	0.0492		
9	0.4022	0.3133	0.2970	0.2768	0.2363	0.2146	0.1230	0.0975	0.0492

FIGURE 5: Selection of MAR model order p .

decreased when p increases to 10. Hence, the MAR model order p is selected as 10 in the experiment. After establishing the MAR model, all parameters were identified based on the QR decomposition and thus the fault feature vectors were formed. Further, principal features were extracted from all features by using principal component analysis (PCA), thus achieving a balance between the computational cost and the integrity of feature information. The principal component (PC) distribution of the feature vectors obtained by blind parameter identification of MAR model is illustrated in Figure 6 and Table 4, while the 3D projection of feature vectors for different faults is shown in Figure 7.

During optimization stage, the feature vectors of different fault types were divided randomly into two parts with 30 and 29 vectors, respectively, and the 30 ones were used for training while the 29 ones for testing. The proposed MHGWOSCA approach was applied to optimize the penalty factor C and the kernel parameter g , where individual number and iteration time were set to 30 and 100, respectively, while searching ranges of C and g were both $[2^{-10}, 2^{10}]$. Considering the maximum number of iterations and the number of optimization parameters, mutation amplitude was set to 0.5 and mutation period was set to 5. During optimization process, fitness values were calculated by fivefold cross validation for the training samples. Hence, the optimization effect of the given C and g can be measured by average accuracy of cross

FIGURE 6: Principal component distribution of feature vectors.

validation, and then the optimal C and g were obtained. The convergence procedure of MHGWOSCA is depicted in Figure 8, from which it can be observed that, with the progress of iteration, the average fitness value of individuals rises rapidly at first and then tends to be stable; that is, the individuals are close to the global optimal solution. In addition, due to the effect of the mutation operators, the best three individuals α , β , and δ change greatly and periodically, trying to jump out of the local optimum without affecting the overall convergence trend. The shaded part shows the distribution of the convergence curve in 10 runs. The convergence curves of different optimization methods are compared in Figure 9, while the average fitness values for different optimization methods in the last 20 iterations are listed in Table 5. The convergence data are the average of 10 runs. The figure shows that the convergence rate of SCA is significantly lower than other methods; GWO and HGWOSCA both have good convergence trend; MHGWOSCA first converges to a high level in the initial stage and shows the best convergence effect at later stage of iterations among all methods.

With the optimal parameters C and g , the SVM model optimized by MHGWOSCA was trained and applied to recognize testing samples. For an in-depth verification of the proposed method, diagnosis results were averaged over 10 runs and training samples were chosen at random in every run. Furthermore, deviation of each result was calculated as a reference for result evaluation. The optimal C and g are presented corresponding to the best accuracy among 10 runs.

TABLE 4: Principal components of feature vectors.

Fault label	Principal components												
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
L1	-0.465	0.498	-0.055	-0.516	0.095	0.008	-0.101	0.334	0.151	0.100	-0.040	0.078	-0.097
	-0.459	0.510	-0.076	-0.515	0.116	0.003	-0.082	0.297	0.137	0.083	-0.032	0.076	-0.095
L2	-0.362	-0.252	0.257	-0.034	0.268	-0.861	-0.699	-0.181	-0.131	-0.702	0.053	0.071	-0.185
	0.396	-0.547	-0.194	-0.347	0.161	-0.432	-0.415	0.160	0.059	-0.438	0.510	-0.070	-0.393
L3	0.883	0.374	1.127	0.456	0.168	0.145	-0.120	0.229	-0.036	0.061	0.054	-0.031	0.020
	0.903	0.393	1.113	0.462	0.140	0.162	-0.158	0.221	-0.054	0.065	0.081	-0.054	0.030
L4	-0.724	-1.063	0.770	0.037	-0.593	0.167	0.025	-0.271	-0.099	0.045	0.053	0.124	-0.244
	0.277	-1.649	0.483	-0.168	0.614	1.125	0.160	-0.182	0.285	-0.320	0.017	0.495	0.519
L5	-0.387	0.075	-0.232	-0.161	0.232	0.566	0.057	-0.033	-0.077	-0.069	0.124	0.013	0.072
	-0.334	-0.002	-0.292	-0.107	0.242	0.600	0.082	0.003	-0.120	-0.102	0.127	0.039	0.082
L6	0.350	-0.174	0.128	-0.874	-0.099	-0.441	0.413	0.533	-0.149	0.306	-0.065	0.204	0.002
	-1.335	0.256	-0.407	0.647	-0.262	0.114	0.339	0.128	-0.068	0.374	-0.047	-0.265	-0.284
L7	1.847	0.174	0.321	0.410	0.250	-0.255	-0.026	-0.182	-0.008	0.198	-0.283	0.083	-0.011
	1.995	0.217	0.045	0.276	0.170	-0.364	-0.070	-0.297	-0.048	0.130	-0.086	0.061	0.005
L8	-1.110	0.912	0.024	1.937	-1.912	-0.797	0.348	0.720	-0.036	-0.714	0.657	0.399	0.280
	-1.337	0.116	-0.700	0.553	0.894	-0.020	0.247	-0.459	-0.077	0.289	-0.036	0.308	0.335
L9	-0.845	0.547	-0.570	0.382	-0.046	0.345	0.295	0.237	0.015	-0.127	-0.100	-0.218	0.005
	0.753	0.129	-0.638	0.636	0.267	0.380	0.208	-0.340	-0.274	-0.473	-0.328	-0.078	-0.162

FIGURE 7: 3D projection of feature vectors for different faults.

FIGURE 8: Convergence procedure of proposed optimization approach: (a) overall trend, (b) partial enlarged detail.

Fault diagnosis results for different methods are shown in Table 6 and Figure 10. It can be seen from Table 6 that the evaluation values of the proposed VMD-MAR-MHGWOSCA-SVM method are the highest among all four metrics, i.e., 0.9573, 0.9554, 0.9808, and 0.9808, and the deviations of the evaluation values are also at a small level. Specifically, the comparison between EMD-MAR-MHGWOSCA-SVM and

the proposed method indicates the effectiveness of VMD; the comparison of VMD-FE-HGWOSCA-SVM, VMD-PE-HGWOSCA-SVM, and the proposed method shows MAR model possess a better feature extraction ability; similarly, it can be concluded from the contrastive experiment among VMD-MAR-GWO-SVM, VMD-MAR-HGWOSCA-SVM, VMD-MAR-SCA-SVM, and the proposed method that

FIGURE 9: Comparison of different optimization methods: (a) overall trend, (b) partial enlarged detail.

FIGURE 10: Comparison of evaluation values for different methods.

MHGWOSCA improves the diagnosis performance of SVM. Figure 11 is the boxplots of evaluation values, illustrating the performance of different diagnosis methods. As shown in the figure, the proposed method achieves the highest evaluation values and possesses outstanding stability among all methods.

5. Conclusions

For rolling bearing fault diagnosis, a new method based on blind parameter identification of MAR model and mutation

hybrid GWO-SCA optimized SVM is proposed in this study. Due to the nonstationarity of original signal, signals collected from different types of faults were firstly split into a set of IMFs by VMD, the decomposing mode number K of which was determined with central frequency observation method. Then all the IMFs were employed to establish a MAR model, whose order was determined by SBC. Later, all parameters of the MAR model were identified blindly through QR decomposition and key features of parameter vectors were extracted via PCA, thus forming the fault feature vectors. Finally, a hybrid optimization method combining mutation

FIGURE 11: Boxplots of evaluation values for different methods; the x-axis tick labels correspond to 1: VMD-FE-HGWOSCA-SVM, 2: VMD-PE-HGWOSCA-SVM, 3: EMD-MAR-GWO-SVM, 4: EMD-MAR-MHGWOSCA-SVM, 5: VMD-MAR-GWO-SVM, 6: VMD-MAR-HGWOSCA-SVM, 7: VMD-MAR-SCA-SVM, and 8: VMD-MAR-MHGWOSCA-SVM.

TABLE 5: The average fitness value of individuals during iterations.

Iteration number	Optimization methods			
	MHGWOSCA	HGWOSCA	GWO	SCA
81	88.12	97.25	97.15	95.95
82	97.32	97.25	97.17	96.06
83	97.32	97.25	97.17	95.87
84	97.32	97.25	97.17	96.03
85	97.33	97.26	97.17	96.13
86	89.70	97.25	97.17	96.16
87	97.32	97.25	97.17	96.04
88	97.33	97.25	97.18	96.02
89	97.32	97.25	97.18	96.08
90	97.32	97.25	97.17	96.09
91	88.87	97.26	97.19	96.25
92	97.32	97.25	97.18	96.33
93	97.32	97.25	97.18	96.34
94	97.32	97.25	97.19	96.29
95	97.32	97.26	97.19	96.43
96	88.99	97.26	97.19	96.37
97	97.32	97.26	97.19	96.40
98	97.32	97.26	97.18	96.43
99	97.32	97.26	97.19	96.46
100	97.32	97.26	97.19	96.46

TABLE 6: Fault diagnosis results with different methods.

Processing methods	C	g	Adjusted Rand index (ARI)	Result evaluation		
				Normalized mutual information (NMI)	F-measure (F)	Accuracy (ACC)
VMD-FE-HGWOSCA-SVM	184.0764	2.2262	0.9001 [-0.034, 0.034]	0.9046 [-0.032, 0.034]	0.9519 [-0.019, 0.017]	0.9525 [-0.018, 0.017]
VMD-PE-HGWOSCA-SVM	2.5155	414.0703	0.9216 [-0.037, 0.053]	0.9221 [-0.043, 0.048]	0.9635 [-0.017, 0.025]	0.9636 [-0.017, 0.025]
EMD-MAR-GWO-SVM	172.6222	0.0010	0.9284 [-0.031, 0.038]	0.9244 [-0.033, 0.039]	0.9662 [-0.017, 0.018]	0.9667 [-0.016, 0.018]
EMD-MAR-MHGWOSCA-SVM	254.3134	0.0339	0.9296 [-0.072, 0.053]	0.9236 [-0.074, 0.055]	0.9661 [-0.041, 0.026]	0.9667 [-0.039, 0.026]
VMD-MAR-GWO-SVM	515.8658	0.1028	0.9325 [-0.048, 0.033]	0.9311 [-0.054, 0.039]	0.9691 [-0.024, 0.016]	0.9693 [-0.023, 0.015]
VMD-MAR-HGWOSCA-SVM	324.7595	0.0925	0.9380 [-0.039, 0.037]	0.9366 [-0.053, 0.037]	0.9719 [-0.018, 0.017]	0.9720 [-0.018, 0.016]
VMD-MAR-SCA-SVM	212.6435	0.3115	0.9431 [-0.028, 0.032]	0.9401 [-0.041, 0.032]	0.9741 [-0.012, 0.014]	0.9743 [-0.013, 0.014]
VMD-MAR-MHGWOSCA-SVM	261.4197	0.1909	0.9573 [-0.034, 0.025]	0.9554 [-0.037, 0.023]	0.9808 [-0.015, 0.012]	0.9808 [-0.015, 0.011]

operator, GWO, and SCA theory, termed mutation hybrid GWO-SCA (MHGWOSCA), was proposed to optimize SVM model which was subsequently applied to realize the pattern recognition for different fault samples. The contrastive experiments in Section 4 were implemented among the proposed VMD-MAR-MHGWOSCA-SVM method and other seven relevant methods, where nine types of faults with different locations and sizes were successfully diagnosed. The analysis of results indicates that the proposed method exhibits the best performance in all classification metrics, showing its outstanding precision and stability among all methods.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 51741907 and 51409095), the Fundamental Research Project for Application Supported by Yichang Science and Technology Bureau (no. A17-302-a12), the Open Fund of Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station (no. 2017KJX06), and the Research Fund for Excellent Dissertation of China Three Gorges University (no. 2019SSPY070).

References

- [1] S. Riaz, H. Elahi, K. Javaid, and T. Shahzad, "Vibration feature extraction and analysis for fault diagnosis of rotating machinery—a literature survey," *Asia Pacific Journal of Multidisciplinary Research*, vol. 5, no. 51, pp. 103–110, 2017.
- [2] D. Zhong, C. Yi, H. Xiao, H. Zhang, and A. Wu, "A novel fault diagnosis method for rolling bearing based on improved sparse regularization via convex optimization," *Complexity*, vol. 2018, Article ID 2169364, 10 pages, 2018.
- [3] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 60, pp. 252–272, 2015.
- [4] C. P. Mbooi and K. Hameyer, "Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection," *IEEE Transactions on Industry Applications*, vol. 52, no. 5, pp. 3861–3868, 2016.
- [5] W. Fu, J. Tan, C. Li, Z. Zou, Q. Li, and T. Chen, "A hybrid fault diagnosis approach for rotating machinery with the fusion of entropy-based feature extraction and SVM optimized by a chaos quantum sine cosine algorithm," *Entropy*, vol. 20, no. 9, p. 626, 2018.
- [6] R. Yuan, Y. Lv, and G. Song, "Fault diagnosis of rolling bearing based on a novel adaptive high-order local projection denoising method," *Complexity*, vol. 2018, Article ID 3049318, 15 pages, 2018.
- [7] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.
- [8] M. Luo, C. Li, X. Zhang, R. Li, and X. An, "Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings," *ISA Transactions*, vol. 65, no. 8, pp. 556–566, 2016.
- [9] X. Jiang, L. Wu, and M. Ge, "A novel faults diagnosis method for rolling element bearings based on ewt and ambiguity correlation classifiers," *Entropy*, vol. 19, no. 6, p. 231, 2017.
- [10] H. Liu, J. Jing, and J. Ma, "Fault diagnosis of electromechanical actuator based on VMD multifractal detrended fluctuation analysis and PNN," *Complexity*, vol. 2018, Article ID 9154682, 11 pages, 2018.
- [11] W. Fu, K. Wang, J. Zhou, Y. Xu, J. Tan, and T. Chen, "A hybrid approach for multi-step wind speed forecasting based on multi-scale dominant ingredient chaotic analysis, KELM and synchronous optimization strategy," *Sustainability*, vol. 11, no. 6, p. 1804, 2019.
- [12] W. Fu, K. Wang, C. Li, X. Li, Y. Li, and H. Zhong, "Vibration trend measurement for a hydropower generator based on optimal variational mode decomposition and an LSSVM improved with chaotic sine cosine algorithm optimization," *Measurement Science and Technology*, vol. 30, no. 1, p. 015012, 2019.
- [13] H. Hytti, R. Takalo, and H. Ihalainen, "Tutorial on multivariate autoregressive modelling," *Journal of Clinical Monitoring and Computing*, vol. 20, no. 2, pp. 101–108, 2006.
- [14] S. Ganapathy, "Multivariate autoregressive spectrogram modeling for noisy speech recognition," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1373–1377, 2017.
- [15] Y. Shekofteh and F. Almasganj, "Autoregressive modeling of speech trajectory transformed to the reconstructed phase space for ASR purposes," *Digital Signal Processing*, vol. 23, no. 6, pp. 1923–1932, 2013.
- [16] C. Liu, D. Zhou, Z. Wang, D. Yang, and G. Song, "Damage detection of refractory based on principle component analysis and gaussian mixture model," *Complexity*, vol. 2018, Article ID 7356189, 9 pages, 2018.
- [17] B. Cai, L. Huang, and M. Xie, "Bayesian networks in fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2227–2240, 2017.
- [18] D. Wang, "K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: revisited," *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 201–208, 2016.
- [19] N. Raj, G. Jagadanand, and S. George, "Fault detection and diagnosis in asymmetric multilevel inverter using artificial neural network," *International Journal of Electronics*, vol. 105, no. 4, pp. 559–571, 2017.
- [20] H. Chen, J. Wang, J. Li, and B. Tang, "A texture-based rolling bearing fault diagnosis scheme using adaptive optimal kernel time frequency representation and uniform local binary patterns," *Measurement Science and Technology*, vol. 28, no. 3, Article ID 035903, 2017.
- [21] P. Shi, K. Liang, D. Han, and Y. Zhang, "A novel intelligent fault diagnosis method of rotating machinery based on deep learning and PSO-SVM," *Journal of Vibroengineering*, vol. 19, no. 8, pp. 5932–5946, 2017.
- [22] C. Zhang, T. Peng, C. Li, W. Fu, X. Xia, and X. Xue, "Multi-objective optimization of a fractional-order PID controller for

- pumped turbine governing system using an improved NSGA-III algorithm under multiworking conditions,” *Complexity*, vol. 2019, Article ID 5826873, 18 pages, 2019.
- [23] M. Jiang, L. Jiang, D. Jiang, F. Li, and H. Song, “A sensor dynamic measurement error prediction model based on NAPS0-SVM,” *Sensors*, vol. 18, no. 1, p. 233, 2018.
- [24] S. Mirjalili, “SCA: a sine cosine algorithm for solving optimization problems,” *Knowledge-Based Systems*, vol. 96, pp. 120–133, 2016.
- [25] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” *Knowledge-Based Systems*, vol. 69, no. 2014, pp. 120–133, 2013.
- [26] S. Eswaramoorthy, N. Sivakumaran, and S. Sekaran, “Grey Wolf optimization based parameter selection for support vector machines,” *COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 35, no. 5, pp. 1513–1523, 2016.
- [27] M. Abd Elaziz, D. Oliva, and S. Xiong, “An improved opposition-based sine cosine algorithm for global optimization,” *Expert Systems with Applications*, vol. 90, pp. 484–500, 2017.
- [28] W. Fu, K. Wang, C. Li, and J. Tan, “Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM,” *Energy Conversion and Management*, vol. 187, pp. 356–377, 2019.
- [29] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [30] R. H. Chan, M. Tao, and X. Yuan, “Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 680–697, 2013.
- [31] C.-C. Wang, Y. Kang, P.-C. Shen, Y.-P. Chang, and Y.-L. Chung, “Applications of fault diagnosis in rotating machinery by using time series analysis with neural network,” *Expert Systems with Applications*, vol. 37, no. 2, pp. 1696–1702, 2010.
- [32] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
- [33] A. Neumaier and T. Schneider, “Estimation of parameters and eigenmodes of multivariate autoregressive models,” *ACM Transactions on Mathematical Software*, vol. 27, no. 1, pp. 27–57, 2001.
- [34] W. Fu, J. Tan, and Q. Li, “Feature extraction for fault diagnosis utilizing blind parameter identification of MAR model,” in *Proceedings of the International Conference on Sensing, Diagnostics, Prognostics, and Control*, Xi’an, China, 2018.
- [35] N. Singh and S. Singh, “A novel hybrid GWO-SCA approach for optimization problems,” *Engineering Science and Technology, an International Journal*, 2017.
- [36] Y. V. Pehlivanoglu, “A new particle swarm optimization method enhanced with a periodic mutation strategy and neural networks,” *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 3, pp. 436–452, 2013.
- [37] Bearing Data Center of the Case Western Reserve University, <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.
- [38] A. Fahad, N. Alshatri, Z. Tari et al., “A survey of clustering algorithms for big data: taxonomy and empirical analysis,” *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 3, pp. 267–279, 2014.
- [39] M. Yuwono, Y. Guo, J. Wall et al., “Unsupervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in heating ventilation and air conditioning systems,” *Applied Soft Computing*, vol. 34, pp. 402–425, 2015.
- [40] S. Shah, B. Shah, A. Amin et al., “Compromised user credentials detection in a digital enterprise using behavioral analytics,” *Future Generation Computer Systems*, vol. 93, pp. 407–417, 2019.
- [41] O. Maruatona, P. Vamplew, R. Dazeley, and P. A. Watters, “Evaluating accuracy in prudence analysis for cyber security,” in *Proceedings of the International Conference on Neural Information Processing*, pp. 407–417, 2017.
- [42] H. Zhao, M. Sun, W. Deng, and X. Yang, “A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing,” *Entropy*, vol. 19, no. 1, p. 14, 2016.