IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 30, 2019, accepted January 10, 2020, date of publication January 14, 2020, date of current version January 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966582

Fault Diagnosis for Rolling Bearings Based on Composite Multiscale Fine-Sorted Dispersion Entropy and SVM With Hybrid Mutation SCA-HHO Algorithm Optimization

WENLONG FU, KAIXUAN SHAO, JIAWEN TAN, AND KAI WANG
College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China
Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang 443002, China

Corresponding author: Wenlong Fu (ctgu_fuwenlong@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51741907, in part by the Open Fund of Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station under Grant 2017KJX06, and in part by the Research Fund for Excellent Dissertation of China Three Gorges University under Grant 2019SSPY070.

ABSTRACT The health condition of rolling bearing possesses a significant impact on the safety and efficiency of rotating machinery. Accordingly, to diagnose the faults in rolling bearings effectively and accurately, a novel hybrid approach coupling variational mode decomposition (VMD), composite multiscale fine-sorted dispersion entropy (CMFSDE) and support vector machine (SVM) optimized by mutation sine cosine algorithm and Harris hawks optimization (MSCAHHO) is proposed in the paper. Firstly, VMD is employed to decompose raw vibration signals with various fault types into different sets of intrinsic mode functions (IMFs) to weaken the non-stationarity of signals, before which the parameter K of VMD is decided through central frequency observation method. Subsequently, CMFSDE is put forward in this paper to analyze the complexity of fault signals by fully considering the relationship between neighboring elements based on composite multiscale technique, with which the representative features of different fault samples are extracted to construct feature vectors. Later, an enhanced hybrid optimization approach called MSCAHHO is proposed by integrating sine cosine algorithm (SCA) and a periodic mutation strategy to improve Harris hawks optimization (HHO). Then, MSCAHHO is employed to optimize the parameters of SVM, after which the optimal SVM model is utilized for fault classification. Finally, the performance of the proposed methodology is evaluated with four validity indices through comparative experiments. The experimental results reveal that the proposed VMD-CMFSDE-MSCAHHO-SVM method achieves favorable diagnosis results comparing with other relevant methods.

INDEX TERMS Fault diagnosis, variational mode decomposition, composite multiscale fine-sorted dispersion entropy, support vector machine, hybrid mutation SCA-HHO.

I. INTRODUCTION

Rolling bearings, as the most important supporting component, are widely employed in rotating machinery systems, such as large generator sets, aero engines and advanced precision machine tools [1], [2]. Due to their low impact resistance, rolling bearings are relatively susceptible to fatigue, damage and various types of faults [3]. A small fault can affect the operating safety and reliability of the entire mechanical equipment, thus resulting in huge economic losses and even casualties to varying degrees [4], [5]. Therefore, effective and achievable methods are of critical significance for fault diagnosis of rolling bearing in industrial production.

Considering that the operating state of equipment can be reflected by vibration signals, vibration analysis based on vibrational signals has been generally accomplished for diagnosing faults of rolling bearings [6], [8]. Extracting fault features from the collected vibration signals is not only the key process of the vibrational analysis method, but also the premise for state recognition and fault diagnosis.

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

However, owing to the complexity and non-stationarity of the vibration signals, fault feature extraction cannot be always carried out effectively. For this purpose, several time-frequency approaches have been designed for processing non-stationary signals in previous studies, including wavelet transform (WT) [9], empirical mode decomposition (EMD) [10], ensemble empirical mode decomposition (EEMD) [11], and variational mode decomposition (VMD) [12]. Among the above methods, WT based on Fourier transform method has an outstanding performance of analyzing non-stationary signals, but different signals need to be analyzed by selecting different basis functions in WT, which brings applicability problem to signal processing. Unlike WT, EMD does not require pre-selected basis functions and is extensively utilized to process non-stationary signals into sets of time sequences adaptively. Nevertheless, EMD also exists some limitations, such as end effect and mode mixing. To address these limitations, based on the theory of EMD, an improved version called EEMD with the assistance of added gaussian white noise is proposed, but the results decomposed by EEMD may be influenced by the added noise: (1) the residual noise still exists in the reconstructed signal in application, which can easily inundate the fault-related information [13]; (2) different realizations of signal assisted with noise may produce different number of modes, making it difficult to calculate the means of these modes [13]. Different from the mentioned methods, VMD is a novel adaptive signal decomposition method by constructing and solving a constrained variational problem to achieve signal decomposition, thus avoiding the mode mixing in EMD, the noise effect in EEMD and the basis function selection in WT. Additionally, the ability and effectiveness of VMD in signal decomposition have been demonstrated in previous literature [14]–[16]. Therefore, VMD is adopted to preprocess the non-stationary vibration signals in this paper.

After decomposing the non-stationarity signals, the next procedure is to extract fault features effectively. Entropy, as an effective tool for estimating the irregularity and uncertainty of the signal, has been widely employed to extract features within non-stationary time series, such as sample entropy (SampEn) [17], permutation entropy (PE) [18], fuzzy entropy (FE) [19]. Nevertheless, these classical entropies remain some drawbacks and need much effort to be improved. For example, SampEn and FE trouble from computational cost; PE does not consider the relationship between signal amplitudes. In recent years, a new technique named dispersion entropy (DE) [20] is designed to measure the complexity and uncertainty of time series, which is time-saving and less affected by mutation signal as well as can consider the relationship between amplitudes, thus to overcome the drawbacks stated above to some extent. Unfortunately, DE remains two weaknesses. First, the relationship information between neighboring amplitudes is not fully considered, leading to that different vectors may be mapped to the same dispersion pattern, which would impact the assessment accuracy. Second, DE only analyzes time series under a single scale. As a result, a lot of valuable fault information hidden in other scales will be ignored. To tackle the first problem, a new dispersion entropy termed fine-sorted dispersion entropy (FSDE) [21] is proposed by adding an additional factor to distinguish the different sequences that are mapped to the same pattern. For overcoming the second weakness, the concept of composite multiscale dispersion entropy (CMDE) is emerged through composite multiscale coarse graining procedure to avoid the loss of much potentially useful information [22]. While FSDE and CMDE have made considerable improvements over DE, these methods only concern on solving a single weakness existing within DE. Therefore, in this paper, a promoted version of DE called composite multiscale fine-sorted dispersion entropy (CMFSDE) is put forward by incorporating the advantages of FSDE and CMDE for extracting representative features of time series. The contrastive experiments among FSDE, CMDE and CMFSDE are conducted later, indicating CMFSDE has better capability of feature extraction than FSDE and CMDE.

The essence of fault diagnosis is a classification issue. Various well-known approaches have been developed for the classification issue in engineering applications, containing Bayesian decision [23], k-nearest neighbor (KNN) [24], artificial neural network (ANN) [25], support vector machine (SVM) [26], etc. Among the above methods, Bayesian decision has a notable ability in recognition considering class conditional probability and prior probability. But the accuracy of Bayesian decision is reached through the assumption of an appropriate prior model. Based on Euclidean or Manhattan distances, KNN can be realized simply and is affected by the distribution of samples easily. ANN performs strong capacity in terms of pattern recognition with large data samples, yet it is time-consuming for adjusting the parameters in network structure. Based on statistical learning theory, SVM possesses the ability to solve small-sample, nonlinear and high-dimension classification problems with suitable kernel functions [27]. By mapping samples to high-dimensional space, the hyperplane of optimal classification is constructed in the space to satisfy the classification requirements. At present, combining with feature extraction method, SVM has been successfully and broadly applied for pattern recognition in fault diagnosis [28], [29]. Due to its remarkable and powerful property, SVM is utilized as the fault classifier in this study.

When deducing the fault pattern recognition, the performance of SVM is greatly influenced by the penalty factor C and kernel parameter g [21]. For a machine learning model, intelligent optimization algorithm can determine the suitable parameters of SVM to improve its performance [30]. Therefore, many optimization algorithms have been presented and employed to search the optimal parameters of SVM, such as particle swarm optimization (PSO) [31], grey wolf optimizer (GWO) [32] and sine cosine algorithm (SCA) [28]. Although the performance of the above algorithm has been verified in past publications, there may still be two defects within them, i.e., premature convergence and trapping into the

local optimum [33]. Hence, researchers from different areas have successfully developed several new modified and hybrid algorithms to alleviate defects of these standard algorithms that include PSO combined with SCA [34], quantum-behaved PSO [35], multi-objective hybrid GWO [36], integration of GWO, SCA and mutation operator [28], etc. Recently, Harris hawks optimization (HHO) [37] is proposed by Heidari *et al*. in 2019, which is a novel population-based optimization method with better behaviors than other swarm optimizers on benchmark functions and real constrained engineering problems. Nevertheless, similar to other optimization algorithms, the original HHO may still suffer from the troubles described above. On this account, an enhanced hybrid optimization method termed as mutation SCA-HHO (MSCAHHO) combining the respective advantages of periodic mutation strategy [38], SCA and HHO is proposed in this study, which has fast convergence speed and can approximate the best global optimum based on test of several benchmark functions, including unimodal, multi-modal and composite functions. In consideration of the favorable performance of the proposed MSCAHHO method, it is applied to optimize the parameters of SVM.

To sum up, a novel hybrid diagnosis model with the fusion of VMD, CMFSDE, MSCAHHO optimization strategy and SVM is put forward in this paper. Firstly, the vibration signals are decomposed by VMD into sets of intrinsic mode functions (IMFs), before which the decomposing mode number K is predefined by observing center frequencies of the IMFs. Later, the feature arrays of various fault categories are constructed by the improved novel dispersion entropy CMFSDE. Afterwards, the proposed MSCAHHO optimization method is introduced to search the optimal parameters of SVM. Subsequently, the SVM model optimized by MSCAHHO is applied to classify different fault samples. Lastly, the performance of the proposed model is verified with engineering application and contrastive experiment.

According to the above description, the innovative contributions of this research are made as follows:

(1) An improved version of DE, namely CMFSDE, is proposed for solving the shortages of DE by further considering the relationship information between neighboring amplitudes under multiscale. Subsequently, CMFSDE is applied to extract representative fault features from IMFs obtained by VMD.
(2) MSCAHHO optimization method, in which HHO is enhanced with SCA and a periodic mutation strategy, is originated for discovering the optimal parameters of SVM.
(3) A novel hybrid diagnosis model is systematically presented, whose superiority and effectiveness are verified quantitatively and objectively with four indices in contrastive experiments.

The organization of this article is arranged below: Section II aims at presenting the fundamental theories of VMD, dispersion entropy and SVM. Section III is devoting to describe the mathematical models of CMFSDE and MSCAHHO in detail. Subsequently, Section IV introduces an improved hybrid diagnosis model integrating VMD, CMFSDE, MSCAHHO optimization strategy and SVM. Next, the proposed methodology is validated with engineering application and comparative experiments in Section V. Finally, in Section VI, conclusions of the study are drawn.

II. BASIC THEORY

A. VARIATIONAL MODE DECOMPOSITION

VMD is a new method of adaptive and quasi-orthogonal signal preprocessing technique [39], through which a non-stationary signal can be decomposed into a sum of band-limited IMFs. By solving the optimal solution of constrained variational problem, the central frequency and band-limited of each mode can be decided [40], which can be construed as follows:

$$\min_{m_k,w_k}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*m_k(t)\right]e^{-jw_kt}\right\|_2^2\right\}$$
$$s.t.\ \sum_{k=1}^{K}m_k(t)=f(t),\quad k=1,2,\ldots,K \tag{1}$$

where K denotes the decomposition number of IMFs. m_k and w_k represent the time-domain signal and the central frequencies of the k-th IMF, respectively. While $f(t)$ is the input signal. ∂_t is the partial derivative function, and $\delta(t)$ is unit pulse function.

To transform the above constrained variational problem into unconstrained problem, quadratic penalty factor and Lagrange multiplier are utilized to modified the Equation (1), and then the augmented variational problem is expressed as:

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

where $\beta(t)$ is the Lagrange multiplier, α is the quadratic penalty factor.

Subsequently, following the dual decomposition and Lagrange theory, the alternate direction method of multipliers (ADMM) [41] is added to deal with this variational problem (2) by optimizing m_k, w_k and β alternately. The optimization issues of m_k and w_k are constructed by Equations (3) and (4), respectively.

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

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

Addressing Equations (3) and (4), the iterative equations are obtained in frequency domain as Equations (5) and (6):

$$m_k^{n+1}(w) = \frac{f(w) - \sum_{i \neq k} m_i(w) + \frac{\beta(w)}{2}}{1 + 2\alpha(w - w_k)^2} \tag{5}$$

$$w_k^{n+1} = \frac{\int_0^\infty w \left|m_k(w)\right|^2 dw}{\int_0^\infty \left|m_k(w)\right|^2 dw} \tag{6}$$

Meanwhile, the Lagrange multipliers can be updated as follows,

$$\begin{aligned} &L(m_k, w_k, \beta) \\ &= \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ &\quad + \left\| f(t) - \sum_k m_k(t) \right\|_2^2 + \left\langle \beta(t), f(t) - \sum_k m_k(t) \right\rangle \end{aligned} \tag{7}$$

where r presents an updating coefficient.

The calculation processes of VMD is the following:

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

Step 2: Start iteration, $n = n + 1$;

Step 3: Update m_k and w_k according to Equations (5) and (6);

Step 4: Update β using Equation (7);

Step 5: For a given solution accuracy ε, if $\sum_k \left\| m_k^{n+1} - m_k^n \right\|_2^2 < \varepsilon$, stop the iteration, else turn to Step 2 for the next iteration.

B. DISPERSION ENTROPY

DE is a recently proposed approach for evaluating the irregularity or complexity of time sequences [20]. For a given time sequence $\mathbf{x} = \{x_i, i = 1, 2, \ldots, N\}$ with length N, $\mathbf{x}$ is firstly mapped into by using the standard normal cumulative distribution function, that is:

$$y_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{\frac{-(t-\mu)^2}{2\sigma^2}} \mathrm{d}t \tag{8}$$

where μ and σ represent the expectation and variance of the normal distribution, respectively. $y_i \in (0, 1)$.

Subsequently, the phase space reconstruction matrix is generated by:

$$\mathbf{y}_j^m = \left[y_j, y_{j+\tau}, \ldots, y_{j+(m-1)d} \right] \tag{9}$$

where $j = 1, 2, \ldots, N - (m-1)\,d$, m and d represent the embedding dimension and time delay, respectively.

Future, $\mathbf{y}_j^m$ is mapped to the range $[1, 2, \cdots, c]$ by the following equations.

$$z_j^c = R\left(c \cdot y_j + 0.5\right) \tag{10}$$

$$\mathbf{z}_j^{m,c} = \left[z_j^c, z_{j+d}^c, \ldots, z_{j+(m-1)d}^c \right] \tag{11}$$

where c is the number of classes. z_j^c is the j-th member of the class series $\mathbf{z}_j^{m,c}$, while R denotes rounding. If $z_j^c = v_0$, $z_{j+d}^c = v_1$,..., $z_{j+(m-1)d}^c = v_{m-1}$, then the time series $\mathbf{z}_j^{m,c}$ is corresponded to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$.

Since $\pi_{v_0 v_1 \ldots v_{m-1}}$ is composed of m number and each number has c values, thus the number of possible dispersion patterns equal c^m. The corresponding rate of a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$ is calculated by

$$p = \frac{Number\left\{ j | j \leq N - (m-1)d, \pi_{v_0 v_1 \ldots v_{m-1}} \right\}}{N - (m-1)d} \tag{12}$$

where $Number\left\{ j | j \leq N - (m-1)d, \pi_{v_0 v_1 \ldots v_{m-1}} \right\}$ means the emergence number of each dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$ that is assigned to $\mathbf{z}_j^{m,c}$.

Finally, according to Shannon entropy, the DE value of the time series $\mathbf{x}$ is defined as

$$DE(\mathbf{x}, m, c, d) = -\sum_{\pi=1}^{c^m} p \cdot \ln(p) \tag{13}$$

The higher DE value demonstrates more irregularity of a time series, while the lower one means more regularity.

C. SUPPORT VECTOR MACHINE

SVM is designed for the classical two-classification issue and has unique advantages in solving small samples, nonlinear and high-dimension data classification problems [21]. The basic principle of SVM is to map samples to a high-dimensional feature space by nonlinear transformation and construct an optimal hyper-plane to solve the corresponding nonlinear problem in low-dimensional space. Assuming a sample set $\{(x_i, y_i) | i = 1, 2, \ldots, n\}$, x_i is the i-th input feature vector and y_i is the category label of x_i, the hyper-plane function is modeled by:

$$w \cdot x + b = 0 \tag{14}$$

where b and w are bias parameter and weight vector, respectively. Dot $\cdot$ presents the inner product.

For a binary classification issue, to correctly identify samples, all samples should be satisfied the constraint, that is:

$$w \cdot x_i + b \begin{cases} > 1 & for\ y_i = 1 \\ < -1 & for\ y_i = -1 \end{cases} \tag{15}$$

The maximum classification interval is $2/\left\|w\right\|^2$, which can be obtained by minimizing $\left\|w\right\|^2$. By bringing slack term ξ and penalty factor C into Equation (15), the linear indivisibility problem of SVM is converted into the following constrained optimization issue:

$$\begin{aligned} &\min f = \frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}^{n} \xi_i \\ &s.t.\ y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, n \end{aligned} \tag{16}$$

The computation of sample vectors in high-dimensional space is transformed into inner product by the kernel function. Since a kernel function determines the mapping of samples to feature space, the selection of kernel function is very meaningful for the classification of SVM. Due to the excellent property of radial basis function (RBF), it is always employed

as the kernel function of SVM in the utilization of pattern identification, describing by:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = \exp(-g||x_i - x_j||^2) \quad (17)$$

where g represents the kernel parameter.

To solve Equation (16), the Lagrange function is introduced to converted such quadratic programming problem to the corresponding dual problem, constructing by:

$$\max L = \sum_{i=1}^{n} \mu_i - \frac{1}{2} \sum_{i,j=1}^{n} \mu_i \mu_j y_i y_j K(x_i, x_j)$$
$$s.t. \sum_{i=1}^{n} \mu_i y_i = 0, \quad \mu_i \geq 0, \ i = 1, 2, \ldots, n \quad (18)$$

where μ_i is the Lagrange multiplier. Finally, by solving Equation (18), the linear classification decision function with RBF function is given by:

$$f(x) = sgn(\sum_{i=1}^{n} \mu_i K(x_i, x) + b) \quad (19)$$

III. THE PROPOSED METHOD FOR FAULT DIAGNOSIS OF ROLLING BEARINGS

A. COMPOSITE MULTISCALE FINE-SORTED DISPERSION ENTROPY

1) FINE-SORTED DISPERSION ENTROPY

DE is a powerful tool for measuring irregularity of time series, and can be adopted to process the fault signals of rolling bearings effectively [42]. However, according to the original definition of DE, the extent of the differences between samples are not fully considered. Consequently, the different sample vectors may be mapped into the same dispersion pattern, which makes DE unable to reflect fault information comprehensively. Figure 1 exhibits an example of different samples mapped into the same dispersion pattern [2 2 3 3], in which the left one presents the dispersion pattern, and the right one shows three different samples mapped into the same dispersion pattern.

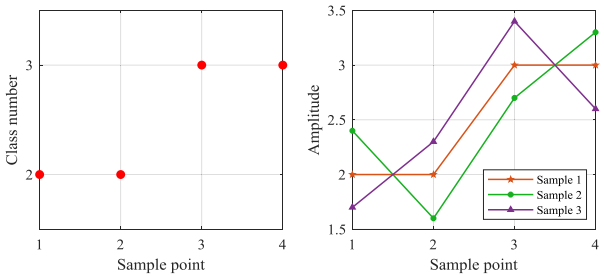


FIGURE 1. An example of different samples mapped into the same dispersion pattern.

Considering this account, a fine-sorted dispersion entropy (FSDE) [21] is proposed to improve the assessment accuracy of DE. In FSDE, factor f is introduced to measure the difference between elements in the vector $\mathbf{y}_j^m$ and then added in the class $\mathbf{z}_j^{m,c}$ as an additional element to distinguish different samples that are mapped into the same pattern. Thus, the relationship information between elements is taken into DE. The factor f is calculated by:

$$f = \left\lfloor \frac{max\left(|\mathbf{d}_y|\right)}{\rho \cdot std\left(|\mathbf{d}_x|\right)} \right\rfloor$$
$$\mathbf{d}_x = \{x_{i+1} - x_i | i = 1, 2, \ldots, n-1\}$$
$$\mathbf{d}_y = \{y_{i+1} - y_i | i = 1, 2, \ldots, m-1\} \quad (20)$$

where $\lfloor . \rfloor$ rounds the element to the nearest integer less than or equals to that element. *std* means the standard deviation. ρ is the adjusting coefficient.

To be specific, if $\rho > \frac{\max(|\mathbf{d}_y|)}{std(|\mathbf{d}_x|)}$, f has only one possible value, that is 0, which means that FSDE is the same as original DE. While, if $\rho \in \left(0, \frac{\max(|\mathbf{d}_y|)}{std(|\mathbf{d}_x|)}\right]$, f will have more possible value. The closer ρ tends to 0, the more dispersion patterns will be subdivided. For example, different time series [0.5326, 0.5249, 0.5158, 0.5072] and [0.5114, 0.5131, 0.5134, 0.5130] are mapped to the same class sequences based on DE with $\tau = 1, m = 4, \rho = 1, c = 6$, i.e., [4, 4, 4, 4] and [4, 4, 4, 4]. Meanwhile, adding a factor f as an additional element to the end of the class sequence, the new class sequences mapped by FSDE are [4,4,4,4,2] and [4,4,4,4,0], respectively. Obviously, the different time series originally mapped to the same pattern are mapped into different patterns. Figure 2 exhibits that f as an additional element is inserted to the end of the sequence $\mathbf{z}_j^{m,c}$ to further refine the sequence.

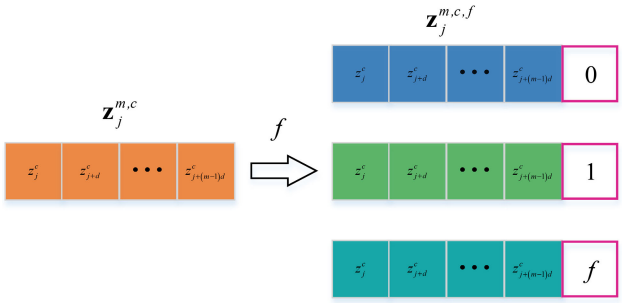


FIGURE 2. Different f values are added as an additional element at the end of the class sequence.

Subsequently, a new class sequence $\mathbf{z}_j^{m,c,f}$ is constructed by the above procedures. $\pi^f_{v_0 v_1 \ldots v_{m-1}}$ is a dispersion pattern corresponding to a $\mathbf{z}_j^{m,c,f}$. Thus, the corresponding frequency of a new dispersion pattern can be measured by:

$$p^f = \frac{Number\left\{j | j \leq N - (m-1)d, \pi^f_{v_0 v_1 \ldots v_{m-1}}\right\}}{N - (m-1)d} \quad (21)$$

At last, like the calculation of DE, FSDE is computed as:

$$FSDE(\mathbf{x}, m, c, d, f) = -\sum_{\pi=1}^{c^m} p^f \cdot \ln\left(p^f\right) \quad (22)$$

2) COMPOSITE MULTISCALE DISPERSION ENTROPY

DE is a single-scale entropy method, which inevitably will cause many useful and important information hidden in multiple scales to be neglected. Therefore, the application of DE might be limited by analyzing non-linear and non-stationary fault signals of rolling bearings with one single scale. To overcome this shortage of DE, the composite multiscale dispersion entropy (CMDE) is proposed by employing composite multiscale coarse-grained procedure [22], in which multiple coarse-grained time series at the same scale are generated to capture more feature information of signals. Frist of all, assume an input time series with length N: $\mathbf{u} = \{u_i, i = 1, 2, \ldots, N\}$, $\mathbf{u}$ is divided into the composite multiscale coarse graining sequence $\mathbf{x}_k^\tau = \left\{x_{k,j}^\tau\right\}$ defined as:

$$x_{k,j}^\tau = \frac{1}{\tau}\sum_{i=(j-1)\tau+k}^{j\tau+k-1} u_i, \quad 1 \le j \le \left\lfloor \frac{N}{\tau} \right\rfloor = p, \ 1 \le k \le \tau \tag{23}$$

where $x_{k,j}^\tau$ is the k-th coarse grained time series with a scale factor τ, $\lfloor . \rfloor$ means the nearest integer less than or equal to that element. p is the length of $x_{k,j}^\tau$, j represents the j-th point of k-th coarse grained time series $x_{k,j}^\tau$. Thus, in CMDE, for a given scale factor τ, τ different time series are generated by dividing the original signals into some segments with a length based on the different starting points. Figure 3 illustrates the coarse-grained procedure of CMDE under the scale $\tau = 2$. As seen from Figure 3 two different time series are obtained for a scale factor $\tau = 2$. There is a similar process for other scale factors.

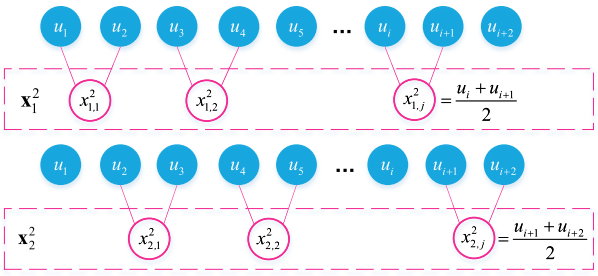


FIGURE 3. The coarse-grained procedure of CMDE under the scale $\tau = 2$.

Then, the DEs of each coarse-grained sequence are calculated at a scale factor τ based on Equation (13).

Finally, CMDE is expressed by averaging all τ DEs in the scale τ, such that,

$$CMDE(\mathbf{u}, \tau, m, c, d) = \frac{1}{\tau}\sum_{k=1}^{\tau} DE\left(\mathbf{x}_k^\tau, \tau, m, c, d\right) \tag{24}$$

According to the calculation procedure of CMDE, compared with DE, more information among elements is taken into full consideration by introducing the composite multiscale coarse-grained procedure in processing the original signals. In addition, averaging DEs can effectively weaken the influence of starting point position on the DE value.

3) COMPOSITE MULTISCALE FINE-SORTED DISPERSION ENTROPY

FSDE and CMDE only focus on adequately overcoming one problem of DE. Thus, the enhanced dispersion entropy CMFSDE is proposed by assembling the advantages of FSDE and CMDE to address the problems of DE comprehensively. Primarily, the coarse graining procedure of time series is the same as the procedure in CMDE. Subsequently, the FSDEs of each coarse-grained sequence $\mathbf{x}_k^\tau$ are calculated at a scale factor τ based on Equation (22). Finally, CMFSDE is defined by averaging all τ FSDEs in the scale τ, that is,

$$CMFSDE(\mathbf{u}, \tau, m, c, d, f) = \frac{1}{\tau}\sum_{k=1}^{\tau} FSDE\left(\mathbf{x}_k^\tau, \tau, m, c, d, f\right) \tag{25}$$

The flowchart of the proposed CMFSDE method is presented in Figure 4, and its main steps are described in detail as follows:

Step 1: Set paraments $\tau_{\max}, m, c, d, \rho$, where $\tau_{\max}$ is the largest scale factor;

Step 2: Generate coarse graining time series based on Equation (23);

Step 3: Calculate the factor f as an additional element in each class sequence based on Equation (20);

Step 4: Calculate FSDEs of all time series at the same scale based on Equation (22);

Step 5: Calculate CMFSDE based on Equation (25);

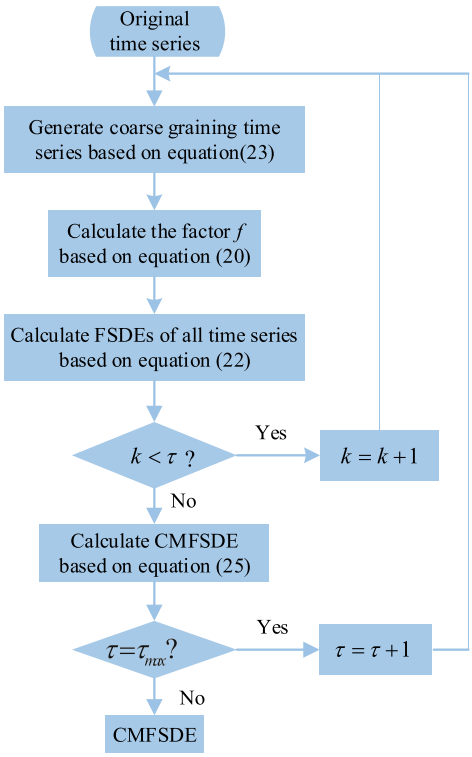


FIGURE 4. The flowchart of CMFSDE.

Step 6: If $\tau = \tau_{\max}$, stop the procedure, else turn to Step 2 and $\tau = \tau + 1$.

A contrastive experiment among CMDE, refined composite multiscale dispersion entropy (RCMDE) [43] and CMFSDE was conducted. To obtain more objective comparison results, 100 groups of white gaussian noise (WGN) are generated to measure these three entropy values. Each group noise signal has data length of 1000 points. To make a fair comparison, the parameters of three entropies are the same: $m = 2$, $d = 1$, $c = 6$, and $\tau_{\max} = 20$, as set in [43]. Subsequently, the mean values and standard deviations (SD) of the generated 100 groups are measured by three entropies. The coefficient of variation (CV) [43] defined as the SD divided by the mean is adopted to compare the results. As done in [43], the results of WGN are investigated at scale factor 10 as a trade-off between short and long scales. Figure 5 shows the waveform of WGN. Table 1 records the CV values and computation time of three entropies at scale factor 10 for WGN. From Table 1, CVs of RCMDE and CMFSDE are significantly less than that of CMDE. Moreover, the smallest CV for WGN is reached by CMFSDE. Besides, the computation time of three entropies is close by measuring 100 groups of WGN, among which the computation time of CMFSDE is the second. Overall, CMFSDE is more suitable for detecting dynamic changes of the complex signal comparing with CMDE and RCMDE.

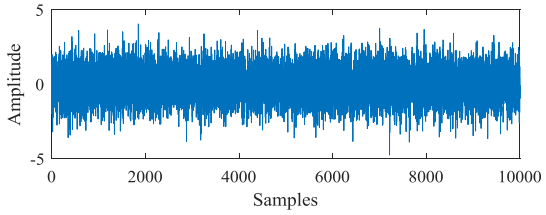


FIGURE 5. The waveform of WGN.

TABLE 1. CV values of three entropies at scale factor 10 for WGN as well as computation time.

	CMDE	RCMDE	CMFSDE
CV	0.0883	0.0374	0.0268
Time (s)	89.8	84.7	88.4

B. MUTATION SCA-HHO OPTIMIZATION

1) SINE COSINE ALGORITHM

A new stochastic optimization algorithm called SCA is proposed by Mirjalili [44] using simple sine and cosine functions as operators to tackle optimization issues. A simple and elaborate process of exploration and exploitation enables SCA to discover optimal results quickly in the search region. The updating strategy of the search agents is described as follows:

$$Z_i^{l+1} = \begin{cases} Z_i^l + r_1 \times \sin(r_2) \times |r_3 P_i^l - Z_i^l|, & r_4 < 0.5 \\ Z_i^l + r_1 \times \cos(r_2) \times |r_3 P_i^l - Z_i^l|, & r_4 \geq 0.5 \end{cases} \quad (26)$$

where Z_i^l and P_i^l are the current position and the best position of i-th search agent at l-th iteration, respectively. $r_1 = a - l(a/l_{\max})$ is a parameter for determining the region of search agent i at next iteration, where a, l and $l_{\max}$ are a constant, the current number of iterations and the maximum number of iterations, respectively. The random parameter r_2 inside $[0, 2\pi]$ defines the distance that the current solution should towards or away from the target optimal solution; The parameter r_3 randomly assigns a weight in [0, 2] to the target optimal solution, enhancing ($r_3 > 1$) or weakening ($r_3 < 1$) the effect of the current optimal solution; Finally, the parameter r_4 selected randomly from [0,1] is to fairly switch the units of the sine and cosine in Equation (26).

2) HARRIS HAWKS OPTIMIZATION

HHO proposed by Heidari *et al.* [37] in 2019 is a novel population-based and nature-inspired optimization method to solve various optimization problems. HHO algorithm emulates the hunting behaviors of Harris hawks in nature and contains the exploration phase and exploitation phase.

For the exploration phase, the locations of Harris' hawks are randomly updated as follows:

$$X^{l+1} = \begin{cases} X_{rand}^l - r_1 \left| X_{rand}^l - 2r_2 X^l \right| \\ \quad q \geq 0.5 \\ \left(X_{rabbit}^l - X_m^l \right) - r_3 \left(LB + r_4 (UB - LB) \right) \\ \quad q < 0.5 \end{cases} \quad (27)$$

where X^{l+1} is the position order of hawks in the next iteration l, X^l is the position of rabbit, X_{rabbit}^l is the current position vector of hawks, r_1, r_2, r_3, r_4, and q are assigned randomly in (0, 1) updating in each iteration, LB and UB are the upper and lower bounds of variables, respectively, X_{rand}^l is a randomly selected hawk from the current population, and $X_m^l = \frac{1}{n} \sum_{i=1}^{n} X_i^l$ is the average position of the current population of hawks. X_i^l represents the location of i-th hawk in iteration l, and n is the number of hawks.

For the exploitation phase, there are four stages: soft besiege, hard besiege, soft besiege with progressive rapid dives and hard besiege with progressive rapid dives.

In the soft besiege, the behavior of hawk is modeled by:

$$\begin{aligned} X^{l+1} &= \Delta X^l - E \left| J X_{rabbit}^l - X^l \right| \\ \Delta X^l &= X_{rabbit}^l - X^l \end{aligned} \quad (28)$$

where $E = 2E_0(1 - l/l_{\max})$ is used to switch from exploration to exploitation. $l_{\max}$ is the maximum iteration number, and E_0 is the initial value of E. ΔX^l is the difference between the position of the rabbit and the current location in iteration l, a random number r_5 is in (0,1), and $J = 2(1 - r_5)$ changes randomly about the jump strength of the rabbit in each iteration.

In the hard besiege, the update of current positions can be given by:

$$X^{l+1} = X^l_{rabbit} - E|\Delta X^l| \tag{29}$$

In the soft besiege with progressive rapid dives, the positions of hawks are performed by:

$$X^{l+1} = \begin{cases} Y & if\ F(Y) < F(X^l) \\ Z & if\ F(Z) < F(X^l) \end{cases} \tag{30}$$

where $Y = X^l_{rabbit} - E\left|JX^l_{rabbit} - X^l\right|$, $Z = Y + S \times LF\ (D)$, D shows the dimension of problem, S denotes a $1 \times$ D random matrix. LF means the levy flight function, calculating by:

$$LF(x) = 0.01 \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}}, \quad \sigma = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}}\right)^{\frac{1}{\beta}}, \quad \beta = 1.5 \tag{31}$$

where random numbers u and v are assigned in (0,1).

In the Hard besiege with progressive rapid dives, the updating positions are operated by Equation (30), but Y and Z are calculated using new equations. $Y = X^l_{rabbit} - E\left|JX^l_{rabbit} - X^l_m\right|$, X^l_m is introduced in exploration phase, $Z = Y + S \times LF\ (D)$.

3) MUTATION SCA-HHO OPTIMIZATION METHOD

Although SCA and HHO have their own excellences, some deficiencies still require to be promoted, involving premature convergence and trapping into local optima, etc. Therefore, the enhanced algorithm called mutation SCA-HHO (MSCAHHO) is proposed with the integration of mutation operator, SCA and HHO. Also, it is expected to improve the convergence behavior and the quality of solutions. Additionally, the application of periodic mutation strategy can result in a more sufficient search with jumping greatly and periodically in the search region, trying to escape from local optimum [38]. Thus, more various solutions may be produced.

The hierarchical form of the proposed MSCAHHO is illustrated in Figure 6, where the top layer is performed with HHO individuals and the individuals are updated by SCA in the bottom layer. The top layer contains M HHO search agents corresponding to M groups in the bottom layer and each bottom group includes N SCA individuals. The new

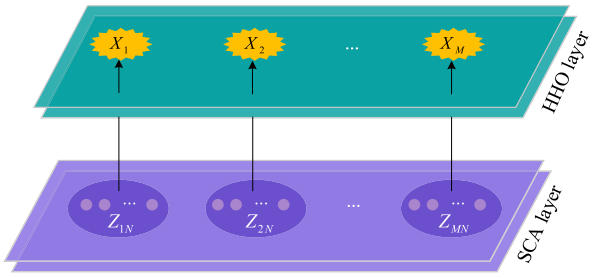


FIGURE 6. The hierarchical form of the proposed MSCAHHO.

positions are updated firstly by executing the SCA in the bottom layer. Subsequently, the optimal solution searched by each bottom layer group is kept by each corresponding agent in the top layer. Then, based on the obtained optimal solutions, the positions of HHO individuals are iterated in the top layer. Thus, the bottom layer individuals are updated by:

$$Z^{l+1}_{ij} = \begin{cases} Z^l_{ij} + r_1 \times \sin(r_2) \times \left|r_3 X^l_i - Z^l_{ij}\right|, & r_4 < 0.5 \\ Z^l_{ij} + r_1 \times \cos(r_2) \times \left|r_3 X^l_i - Z^l_{ij}\right|, & r_4 \geq 0.5 \end{cases} \tag{32}$$

where Z^l_{ij} describes the position of j-th individual in the bottom layer corresponding to the i-th search agent in the top layer. X^l_i denotes the position of i-th search agent in the top layer. l is the number of current iterations.

To enrich diversity of individuals and run away from local optima, mutation operator [38] is introduced to periodically change the location update strategy in Equation (32). Thus, the following updating rules contained the mutation operator are constructed in the bottom layer.

$$Z^{l+1}_{ij} = \begin{cases} \begin{cases} Z^l_{ij} + r_1 \times \sin(r_2) \times \left|r_3 X^l_i - Z^l_{ij}\right|, \\ \quad r_4 < 0.5 \\ Z^l_{ij} + r_1 \times \cos(r_2) \times \left|r_3 X^l_i - Z^l_{ij}\right|, \\ \quad r_4 \geq 0.5 \end{cases} & l \neq nT \\ Z^l_{ij} \times [1 + A \times (0.5 - rand)] & l = nT \end{cases}, \quad n = 1, 2, \ldots \tag{33}$$

$$X^{l+1} = X^l \times [1 + A \times (0.5 - rand)], \quad l = nT,\ n = 1, 2, \ldots \tag{34}$$

where T ($T < l_{\max}$) and A are the mutation periodicity and mutation amplitude, respectively. *rand* means the real random number in accordance with uniform distribution U (0,1) [28]. X^l means the set of positions of all top groups. It can be seen that the position of Z^l_{ij} and X^l will be mutated periodically as the updating process, which may enable the algorithm to jump out of the local optima.

It is a common approach for evaluating stochastic optimization algorithms by adopting benchmark functions with various property, which can ensure that the obtained results of the algorithm are not accidental [45]. Following this fact, several benchmark functions are introduced to verify the validity of the proposed MSCAHHO method, including unimodal, multimodal and fixed-dimension multimodal benchmark functions [37], [46], which are listed in Table 2. In addition, MSCAHHO is compared with the standard SCA, HHO and PSO. The parameter settings of these algorithms are the same to make a fair comparison in the experimental test. The number of search agents are set to 30 with 150 iterations.

TABLE 2. The description of benchmark function.

No.	Name	Function	Dim	Range	$F_{\min}$
Unimodal benchmark functions					
1	Sphere	$F=\sum_{i=1}^{n}x_i^2$	30	[-100, 100]	0
2	Schwefel	$F=\sum_{i=1}^{n}\left(\sum_{j-1}^{i}x_j\right)^2$	30	[-100, 100]	0
3	Rosenbrock	$F=\sum_{i=1}^{n-1}\left[100\left(x_{i+1}-x_i^2\right)^2+\left(x_i-1\right)^2\right]$	30	[-30, 30]	0
Multimodal benchmark functions					
4	Rastrigin	$F=\sum_{i=1}^{n}\left[x_i^2-10cos\left(2\pi x_i\right)+10\right]$	30	[-5.12, 5.12]	0
5	Ackley	$F=-20exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right)-exp\left(\frac{1}{n}\sum_{i=1}^{n}cos\left(2\pi x_i\right)\right)+20+e$	30	[-30, -32]	0
6	Griewank	$F=\frac{1}{4000}\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	30	[-600, 600]	0
Fixed-dimension multimodal benchmark functions					
7	Shekel's Foxholes	$F=\left(\frac{1}{500}+\sum_{j=1}^{25}\frac{1}{j+\sum_{i=2}^{2}\left(x_i-a_{ij}\right)^6}\right)^{-1}$	2	[-65, 65]	1
8	Kowalik	$F=\sum_{i=1}^{11}\left[a_i-\frac{x_1\left(b_i^2+b_ix_2\right)}{b_i^2+b_ix_i+x_4}\right]^2$	4	[-5, 5]	0.0003
9	Goldstein-Price	$F=\left[1+\left(x_1+x_2+1\right)^2\left(19-14x_1+3x_1^2-14x_2+6x_2^2\right)\right]\times\left[30+\left(2x_1-3x_2\right)^2\times\left(18-32x_1+12x_1^2+48x_1x_2+27x_2^2\right)\right]$	2	[-2, 2]	3

Moreover, all tests are performed 10 times independently on each benchmark function. The average convergence curves of four algorithms are plotted in Figure 7, where the experimental results appear that the newly modified MSCAHHO makes more outstanding performance as comparison to others on different testing classical functions with respect to convergence speed and searching optimal solutions.

The following gives main steps of the proposed MSCAHHO optimization approach:

Step 1: Set mutation parameters and initialize M searching agents in the top layer. $M\cdot N$ individuals in the bottom layer randomly in the given range of variables;

Step 2: Calculate the fitness value of each search agent;

Step 3: Execute iteration, $l=l+1$;

Step 4: For the bottom layer, update Z_{ij}^{l} and X according to Equations (33) and (34);

Step 5: Update X_i based on the best solution searched by the corresponding group in the bottom layer;

Step 6: For the top layer, update in HHO;

Step 7: If $l\neq l_{\max}$, turn to Step 3;

Step 8: Return the position X_{rabbit} as the optimal solution.

IV. FAULT DIAGNOSIS BASED ON CMFSDE AND SVM OPTIMIZED BY MSCAHHO

In this research, a novel fault diagnosis approach is proposed based on CMFSDE as the feature extractor and MSCAHHO optimized SVM as the fault classifier to diagnose faults in rolling bearings effectively, which can be divided into the following main steps:

Step 1: Set the modal decomposition number K with central frequency observation technique;

Step 2: Decompose the raw vibration signals into sums of IMFs by VMD;

Step 3: Calculate the CMFSDE value of each IMF;

Step 4: Construct feature vectors of different fault samples with CMFSDE values;

Step 5: Optimize the parameters C and g of SVM by the proposed MSCAHHO;

Step 6: Train the SVM model with optimal parameters C and g;

Step 7: Apply the optimal SVM model to classify various fault samples.

From the above, Figure 8 shows the detailed scheme of the proposed fault diagnosis model. As seen from Figure 8,

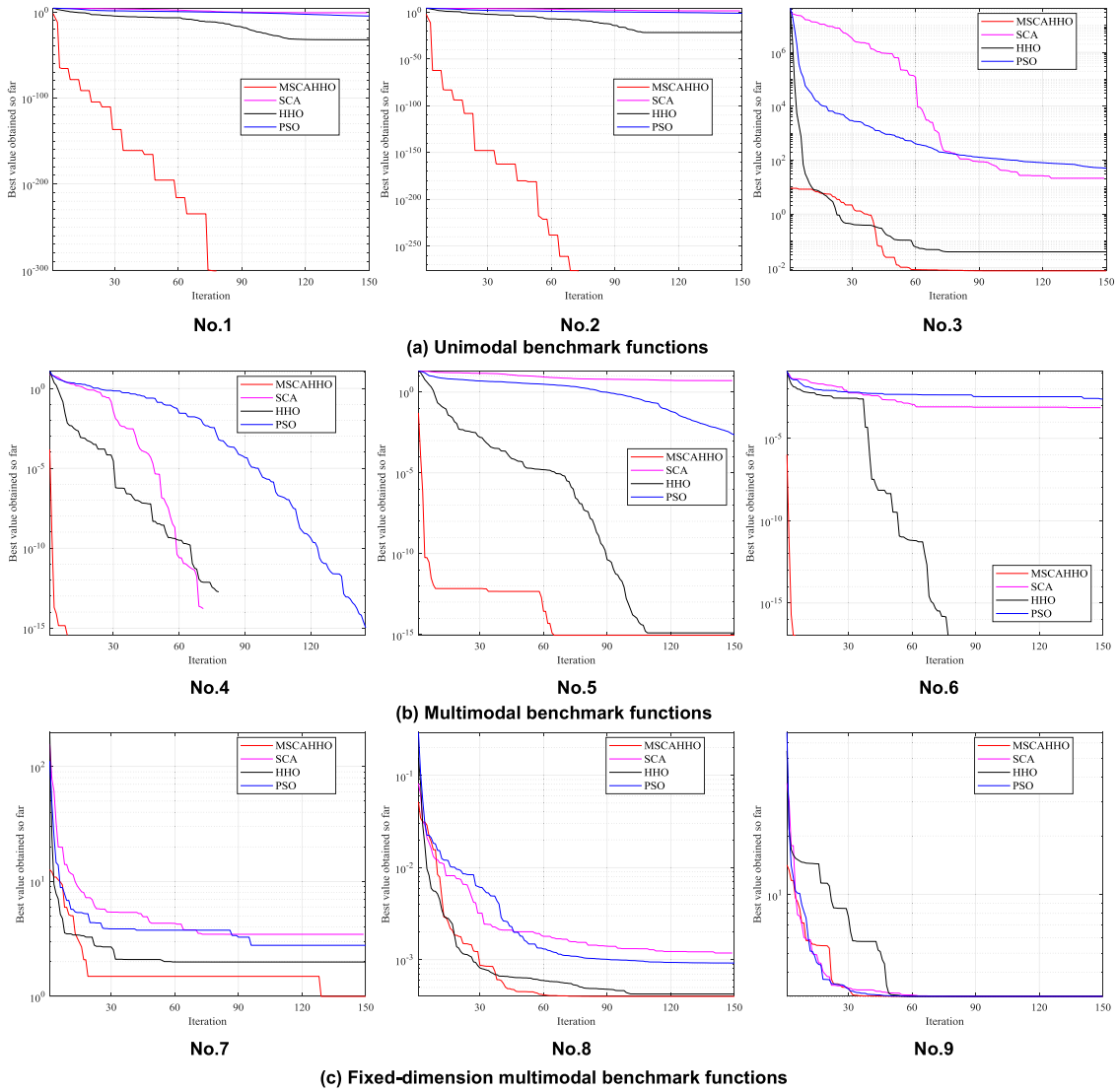


FIGURE 7. Convergence curves of PSO (blue), SCA (magenta), HHO (black) and proposed MSCAHHO (red) on various benchmark functions after 10 runs.

the scheme of the proposed approach comprises four main parts: signal processing, feature extraction, parameter optimization and fault classification. At first, the original vibration fault signals are decomposed into groups of IMFs by VMD in signal processing part. Then, CMFSDE is utilized as the feature extractor to extract representative features. Later, the parameters C and g of SVM are optimized by the proposed MSCAHHO optimization approach. At last, the fault feature vectors constructed with extracted representative features are imported into the SVM employed for fault classification.

V. ENGINEERING APPLICATION

A. DATA COLLECTION

In the paper, the experimental vibration signals are collected from Bearings Data Center of Case Western Reserve University [47] due to that these data are well-known for their diversities and efficacities in validating new assessment techniques. The establishment procedure of original fault dataset is described as follows. To begin with, the collection of vibration signals was from the drive end under the sample frequency of 12000Hz and the rated load of 2 hp at the rotation speed of 1750 rpm in fault database; Furthermore, to fully investigate the fault characteristics, fault conditions were set to nine types, namely inner race fault, ball fault, and outer race fault with defect sizes of 0.007, 0.014, and 0.021 inches, respectively. In other words, each fault location has three different defect sizes; Subsequently, to demonstrate the effectiveness of the proposed method, the vibration signal of each fault type was split into 59 non-overlapped samples, as well as each sample has equal length data segment contained 2048 sample points. Thus, the dataset was established

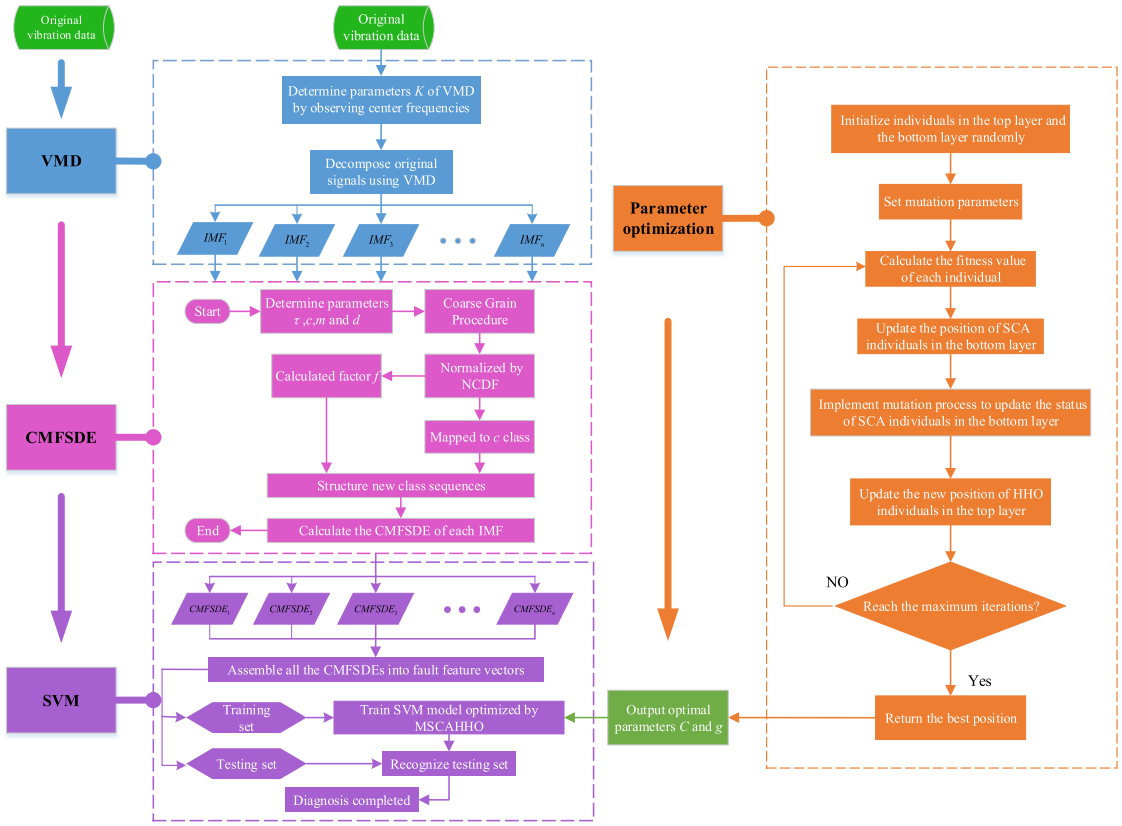


FIGURE 8. The detailed scheme of the proposed fault diagnosis model.

TABLE 3. The description of experimental data under 2 hp load.

Fault position	Severity (inches)	Number of total samples	Number of training samples	Number of testing samples	Label
Inner race	0.007	59	30	29	L1
Inner race	0.014	59	30	29	L2
Inner race	0.021	59	30	29	L3
Ball	0.007	59	30	29	L4
Ball	0.014	59	30	29	L5
Ball	0.021	59	30	29	L6
Outer race	0.007	59	30	29	L7
Outer race	0.014	59	30	29	L8
Outer race	0.021	59	30	29	L9

with a total of 59×9 samples. The detailed experimental fault data adopted in this research are listed in Table 3. The raw vibration signal waveforms of various fault types under 2 hp load are shown in Figure 9, presenting that the waveforms of various fault types are non-stationary and quite different.

B. ENGINEERING APPLICATION IN FAULT DIAGNOSIS OF ROLLING BEARING

1) COMPARATIVE EXPERIMENT AND EVALUATION INDICES

To verify the availability of the developed fault diagnosis method based on VMD-CMFSDE-MSCAHHO-SVM, comparative experiments were performed in the feature extraction and parameter optimization stages. In the feature extraction stage, FSDE and CMDE were compared with CMFSDE; While, HHO and SCA were added as comparison to MSCAHHO in the parameter optimization part. The parameter settings of contrastive methods were set as the same in the comparative experiments to make a fair comparison.

Four commonly measured evaluation metrics were introduced to assess the classification performance of different methods statistically [21], [48], including Adjusted Rand Index (ARI), Normalized Mutual Information (NMI),

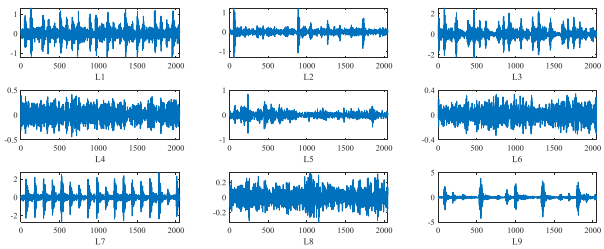


FIGURE 9. Time-domain waveforms of vibration signals of various fault types under 2 hp load.

TABLE 4. Four evaluation metrics for assessment in the experiment.

Abbreviation	Expression
ARI	$ARI = \dfrac{n_{11} + n_{00}}{C_n^2}$
NMI	$NMI = \dfrac{\sum_{\varphi\in\Phi}\sum_{\omega\in\Omega} p(\varphi,\omega)log\left(p(\varphi,\omega)/\,p(\varphi)p(\omega)\right)}{\sqrt{\left(\sum_{\varphi\in\Phi} p(\varphi)log\left(p(\varphi)\right)\right)\left(\sum_{\omega\in\Omega} p(\omega)log\left(p(\omega)\right)\right)}}$
F	$F = 2\dfrac{\left(TP/\left(TP+FP\right)\right)\left(TP/\left(TP+FN\right)\right)}{TP/\left(TP+FP\right)+TP/\left(TP+FN\right)}$
ACC	$ACC = \dfrac{TP+TN}{TP+FN+FP+FN}$

F-measure (F), and Accuracy (ACC), which were noted in Table 4. The range of ARI is $[-1, 1]$, while the other indices are inside $[0, 1]$. The value closer to 1, the better classification performance of a method.

The following notations are adopted: Φ and Ω present the sets of given actual label classified result, respectively; n_{11} is the number of sample pairs with the same label in both Φ and Ω; 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; $p(\varphi, \omega)$ denotes the joint probability function of Φ and Ω; $p(\varphi)$ and $p(\omega)$ are the probability functions of Φ and Ω, respectively; *TP*, *FP*, *FN*, *TN* mean true positive, false positive, false negative, true negative according to actual label and classification result, respectively.

2) SIGNAL PROCESSING

In the signal processing stage of the proposed method, the sample of each fault type was decomposed to a group of IMFs by VMD. The number of IMFs plays a significant impact on results of fault diagnosis. If K is set too large, center frequencies of neighboring IMFs may be relatively similar, something that might lead to mode mixing. On the contrary, if K is taken too small, the input signal may be not decomposed effectively, resulting in the neglect of more valuable information. Accordingly, the number K of IMFs is determined with central frequency observation method firstly.

To determine the appropriate K, the ball fault with a defect size of 0.007 inches (L4) was employed as an example in the experiment. The normalized center frequencies of IMFs with different K are displayed in Table 5 while the distribution of central frequency during the iterative calculation of VMD with various K values are plotted in Figure 10.

As seen from Table 5, similar normalized center frequencies appear when K equals 5 or larger, meaning that the phenomenon of excessive decomposing happens. Similarly, there arise two curves of normalized central frequencies whose distance interval is relatively close when K is 5 or 6 in Figure 10. According to the above analysis, K is recommended to 4 in this paper as highlighted in bold in Table 5.

Subsequently, the original fault signals were decomposed into four IMFs by VMD. The time-domain waveforms of the IMFs decomposed from nine kinds of fault sequences are presented in Figure 11, in which the IMFs of each fault category have their own fluctuation characteristic and can be employed to identify faults to a certain extent.

3) FEATURE EXTRACTION

After obtaining the IMFs through signal decomposition, CMFSDE of each component was calculated to construct the feature vector of sample. As an entropy-based method, the parameters of CMFSDE should be determined beforehand. Here, five parameters need to be pre-set, including maximum scale factor $\tau_{\max}$, embedding dimension m, time delay d, adjusting coefficient ρ and the number c of classes. Referring to previous literature [21], [43], the parameters of CMFSDE chosen in the experiment are listed in Table 6.

For all the first three samples of different fault types (L1–L9), CMFSDE and CMDE of the IMF1 component under different scale factors are recorded in Table 7, in which both CMFSDE and CMDE of the IMF1 component from the same fault type under the same scale are similar. For instance, the CMFSDE values of the classification label L1 are similar when $\tau = 1$. In addition, for different fault types, the entropy values of each IMF1 are distinct at the same scale factor.

TABLE 5. Normalized central frequencies with different decomposition number *K*.

Decomposition Number (K)	Normalized central frequency							
2	0.2741	0.0720						
3	0.2806	0.2346	0.1104					
4	**0.2806**	**0.2343**	**0.1120**	**0.0384**				
5	0.2854	0.2709	0.2305	0.1173	0.0527			
6	0.3516	0.2851	0.2706	0.2304	0.1168	0.0521		
7	0.3527	0.2864	0.2742	0.2466	0.2212	0.1163	0.0527	
8	0.3513	0.2904	0.2819	0.2684	0.2295	0.1154	0.0579	0.0154

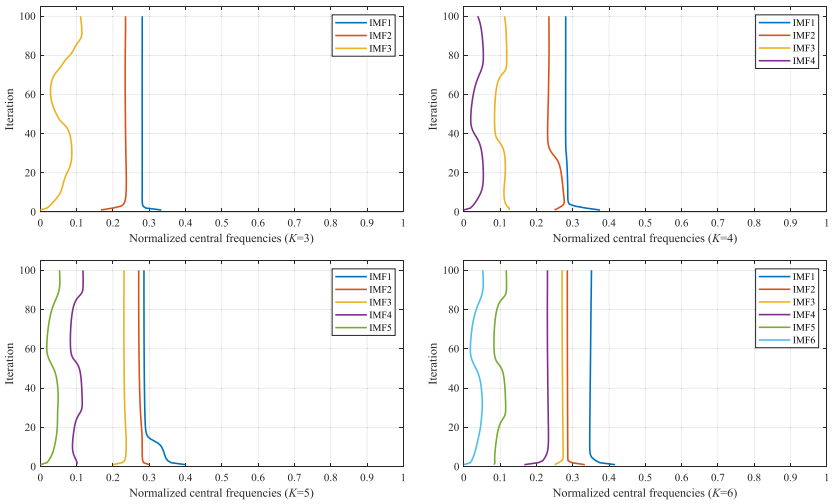


FIGURE 10. The distribution of central frequency with various K values for each iteration.

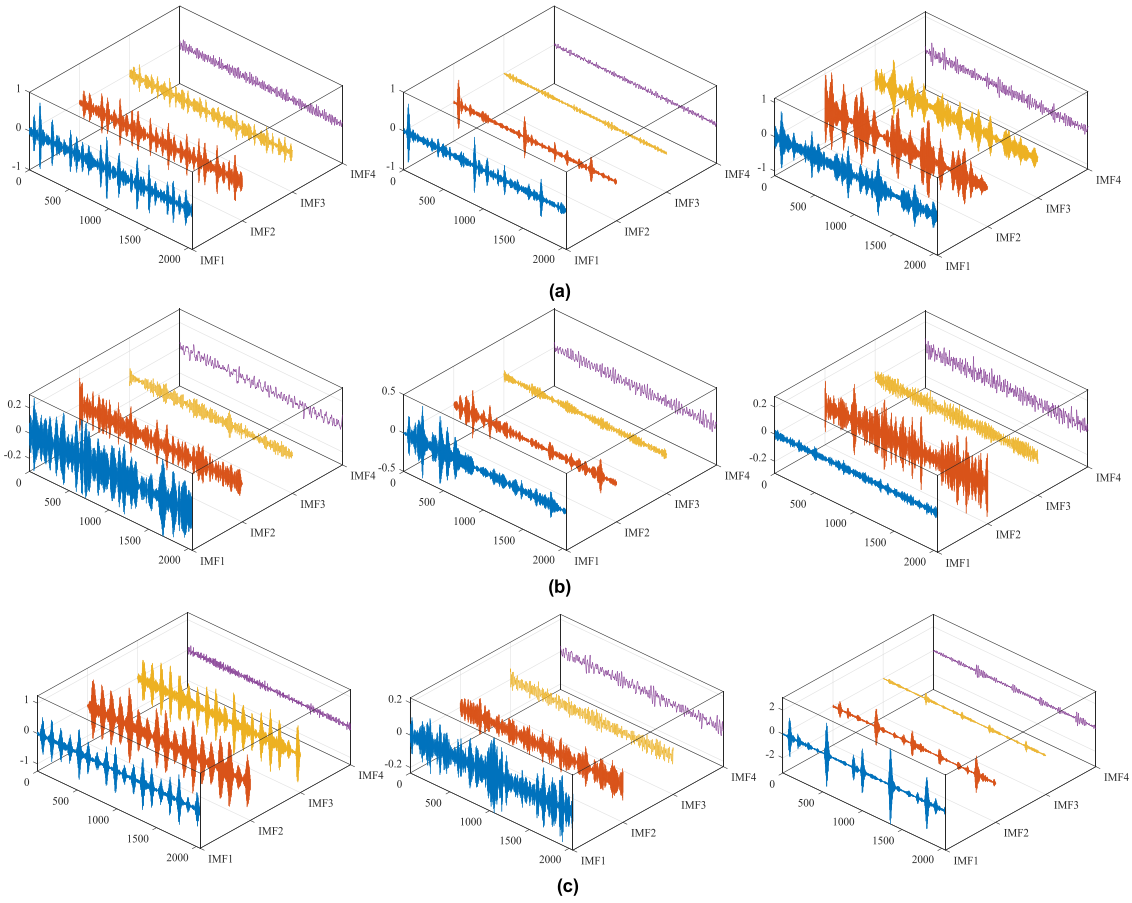


FIGURE 11. The decomposition results of all fault signals by VMD: (a) inner race fault (L1, L2, L3), (b) ball fault (L4, L5, L6), (c) outer race fault (L7, L8, L9).

Moreover, the CMFSDE value of every IMF1 is higher than the corresponding CMDE value, which means that the proposed CMFSDE is expected to represent the complexity of signals. Thus, the representative fault feature vectors constructed with CMFSDE values are introduced for the fault classification application later.

TABLE 6. Parameter setting of CMFSDE.

Parameter	τ_{max}	m	d	c	ρ
Value	3	3	1	6	2.9

4) PARAMETER OPTIMIZATION

To confirm the validity of the proposed approach, 59 feature vectors of fault types were divided randomly into two groups: 30 vectors selected for training and the remaining 29 ones for testing. After that, an important issue is the suitable parameter selection of SVM to boost its classification performance. Hence, the proposed MSCAHHO method is employed to search the optimal values of the penalty factor C and the kernel parameter g. The optimization experiment was carried out with 30 searching agents and 100 iterations. The searching region of parameter C and g were both in $[2^{-10}, 2^{10}]$. The mutation amplitude and mutation period were set 1 and 5 [28], respectively. The five-fold cross-validation was adopted in the experiment. Subsequently, the SVM model was trained with C and g optimized by the proposed MSCAHHO method, thus to achieve the fault classification. Moreover, to demonstrate the effectiveness of this method, the experiment was run 10 times with different random training samples repeatedly. Later, the average fitness value was calculated in each iteration. The optimal parameter values of C and g were recorded corresponding to the highest training accuracy.

The distribution of convergence curves of MSCAHHO method in 10 times is represented with the shaded area in Figure 12, where the averaged fitness values are marked with the solid black line. As seen from Figure 12, the average fitness value keeps rising and changes in stages. Then it tends to be stable as the iteration progresses, which means the method approaches the global optimum solution. In addition, the utilization of the mutation operator brings in more effective search by mutating periodically, which tries to escape from the local optima without impacting the overall convergence tendency. To verify the superiority the proposed MSCAHHO, the convergence curves of SCA and HHO as well as MSCAHHO are plotted in Figure 13 simultaneously. Through comparative analysis of the curves, it can be concluded the proposed MSCAHHO is superior to SCA and HHO.

5) FAULT CLASSIFICATION AND COMPARATIVE ANALYSIS

Based on the obtained optimal parameters C and g, the optimal SVM model was trained and then applied to identify testing samples. Each diagnosis result and computation cost were averaged on 10 independence runs with different random training samples on a laptop with Core i5 2.3GHz / 8GB RAM. Evaluation metrics including ARI, NMI, F and ACC were adopted to assess the performance of different models. The higher value of the four metrics means better performance.

TABLE 7. The CMFSDEs and CMDEs of IMF1 of three same samples of all fault signals (L1–L9) under different scale factors.

Classification Label	Sample Number	CMDE IMF1			CMFSDE IMF1		
		τ=1	τ=2	τ=3	τ=1	τ=2	τ=3
L1	1	0.4110	0.3188	0.3472	0.4463	0.3679	0.4086
	2	0.4210	0.3162	0.3434	0.4496	0.3619	0.4031
	3	0.4149	0.3158	0.3418	0.4468	0.3641	0.4078
L2	1	0.3568	0.2931	0.3394	0.3802	0.3292	0.3967
	2	0.3601	0.2990	0.3447	0.3868	0.3338	0.3943
	3	0.3543	0.2924	0.3384	0.3837	0.3279	0.3923
L3	1	0.3962	0.2692	0.3380	0.4300	0.2978	0.4040
	2	0.3702	0.2627	0.3404	0.4008	0.2905	0.4013
	3	0.3744	0.2480	0.3399	0.4031	0.2946	0.4013
L4	1	0.3153	0.2688	0.3407	0.4039	0.3327	0.4163
	2	0.3150	0.2698	0.3386	0.4074	0.3351	0.4164
	3	0.3117	0.2666	0.3385	0.4002	0.3304	0.4168
L5	1	0.3063	0.2350	0.3427	0.3765	0.2860	0.4109
	2	0.3044	0.2272	0.3330	0.3716	0.2815	0.4008
	3	0.3313	0.2463	0.3401	0.3640	0.2748	0.3891
L6	1	0.2882	0.3684	0.3372	0.3512	0.4362	0.4217
	2	0.3100	0.2582	0.3360	0.3872	0.3137	0.4151
	3	0.2849	0.3627	0.3367	0.3527	0.4377	0.4214
L7	1	0.4407	0.3351	0.3348	0.4561	0.3801	0.3951
	2	0.4440	0.3207	0.3312	0.4631	0.3731	0.3891
	3	0.4379	0.3324	0.3293	0.4543	0.3742	0.3912
L8	1	0.3241	0.2894	0.3337	0.4012	0.3472	0.4104
	2	0.3325	0.3139	0.3499	0.4165	0.3801	0.4172
	3	0.3244	0.2878	0.3365	0.4011	0.3462	0.4071
L9	1	0.4691	0.3782	0.3537	0.4735	0.3818	0.3994
	2	0.4655	0.3742	0.3477	0.4722	0.3778	0.4015
	3	0.4711	0.3703	0.3447	0.4791	0.3758	0.3996

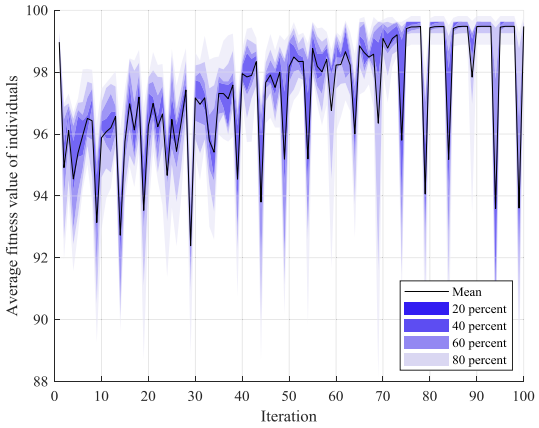


FIGURE 12. The distribution of convergence curves of MSCAHHO method in 10 times.

To illustrate the effectiveness of the proposed method, nine relevant methods were selected for comparison. Fault diagnosis results of the methods in terms of the four evaluation indices are exhibited in Table 8 and Figure 14, from which several conclusions can be found below:

(1) Comparing the results of all methods, the proposed VMD-CMFSDE-MSCAHHO-SVM method is superior to other methods with the highest values of four metrics about 0.9897, 0.9896, 0.9954 and 0.9954, which are highlighted in bold. Meanwhile, the evaluation deviations of the proposed method are also within a small range.

(2) In terms of feature extraction, as comparison to the results obtained by VMD-FSDE-MSCAHHO-SVM (No.3), VMD-CMDE-MSCAHHO-SVM (No.6) and VMD-CMFSDE-MSCAHHO-SVM (No.9), the average classification accuracies (ACC) of FSDE-based, CMDE-based and CMFSDE-based methods are 0.9816, 0.9927 and 0.9954, respectively. The results show that all versions of DE can extract fault features effectively with high recognition accuracy. Among these methods, CMFSDE-based method achieves the highest accuracy with 0.9954, indicating that CMFSDE is more suitable for feature extraction than

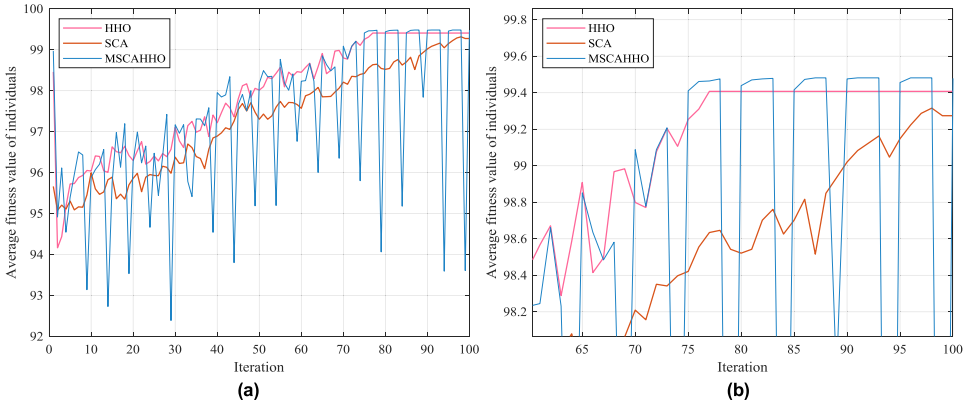


FIGURE 13. MSCAHHO compared with SCA and HHO: (a) overall tendency, (b) local detail.

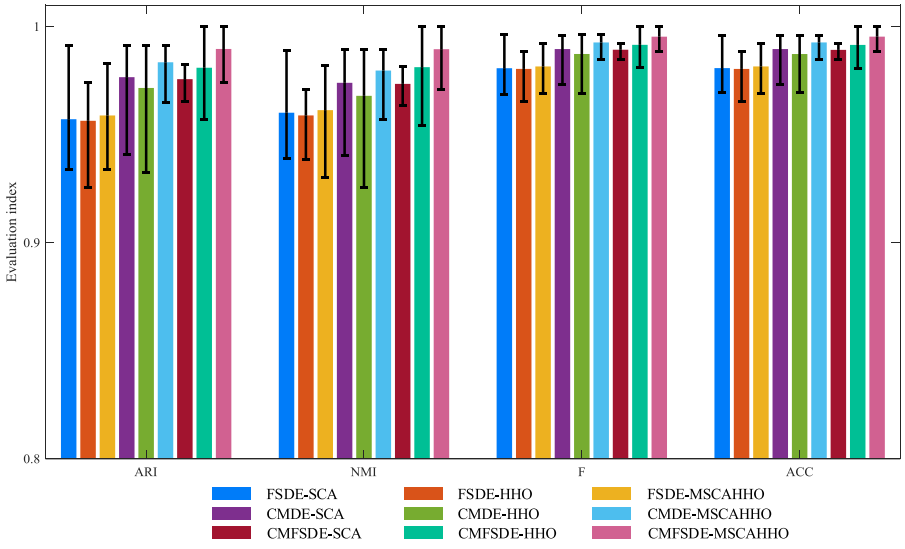


FIGURE 14. Comparison of evaluation results of different methods.

TABLE 8. Fault diagnosis results of different methods.

No.	Method	Best C	Best g	Evaluation result				Time (s)
				ARI	NMI	F	ACC	
1	VMD-FSDE-SCA-SVM	145.54	65.66	0.9572 [-0.0233, 0.0341]	0.9602 [-0.0210, 0.0286]	0.9808 [-0.0121, 0.0154]	0.9808 [-0.0115, 0.0153]	425.7
2	VMD-FSDE-HHO-SVM	11.80	765.43	0.9565 [-0.0307, 0.0176]	0.9589 [-0.0204, 0.0121]	0.9805 [-0.0490, 0.0080]	0.9805 [-0.0149, 0.0080]	415.5
3	VMD-FSDE-MSCAHHO-SVM	34.42	114.31	0.9589 [-0.0250, 0.0239]	0.9614 [-0.0312, 0.0208]	0.9816 [-0.0126, 0.0107]	0.9816 [-0.0123 0.0107]	420.4
4	VMD-CMDE-SCA-SVM	464.59	3.33	0.9766 [-0.0360, 0.0146]	0.9740 [-0.0336, 0.0156]	0.9897 [-0.0165, 0.0065]	0.9897 [-0.0165, 0.0065]	414.2
5	VMD-CMDE-HHO-SVM	687.52	163.49	0.9717 [-0.0393, 0.0196]	0.9680 [-0.0423, 0.0216]	0.9874 [-0.0181, 0.0088]	0.9874 [-0.0180, 0.0088]	428.9
6	VMD-CMDE-MSCAHHO-SVM	70.06	12.85	0.9835 [-0.0185, 0.0077]	0.9797 [-0.0228, 0.0099]	0.9927 [-0.0080, 0.0035]	0.9927 [-0.0080, 0.0034]	421.4
7	VMD-CMFSDE-SCA-SVM	390.23	2.97	0.9757 [-0.0102, 0.0069]	0.9735 [-0.0098, 0.0079]	0.9893 [-0.0046, 0.0031]	0.9893 [-0.0460, 0.0031]	421.8
8	VMD-CMFSDE-HHO-SVM	388.83	61.60	0.9810 [-0.0240, 0.0190]	0.9813 [-0.0269, 0.0187]	0.9916 [-0.0107, 0.0084]	0.9916 [-0.0107, 0.0084]	433.2
9	**VMD-CMFSDE-MSCAHHO-SVM**	**503.07**	**15.95**	**0.9897 [-0.0156, 0.0103]**	**0.9896 [-0.0185, 0.0104]**	**0.9954 [-0.0070, 0.0046]**	**0.9954 [-0.0069, 0.0046]**	420.2

TABLE 9. Parameters of PSO and BPNN.

Method	Parameter	Value
PSO	Learning factor c_1	1.2
	Learning factor c_2	1.2
	Inertia factor w	0.5
BPNN	The largest training number	500
	Number of hidden layers	20
	Learning rate	0.1
	Training error	0.001

FSDE and CMDE. Besides, the proposed CMFSDE-based method still has the highest values of the other three indices, i.e., 0.9897, 0.9896 and 0.9954. Similarly, it can be concluded from other CMFSDE-based methods, such as VMD-FSDE-SCA-SVM (No.1), VMD-CMDE-SCA-SVM (No.4) and VMD-CMFSDE-SCA-SVM (No.7). In brief, these comparative results demonstrate the availability of the proposed CMFSDE to extract fault features.

(3) Taking CMDE-based methods (No.4, No.5, No.6) as examples for validity analysis, four evaluation values of CMDE-MSCAHHO are the highest (0.9835, 0.9797, 0.9937, 0.9927) compared with those of CMDE-SCA (0.9766, 0.9740, 0.9797, 0.9897) and CMDE-HHO (0.9717, 0.9680, 0.9874, 0.9874), revealing that the proposed MSCAHHO is superior to SCA and HHO in terms of optimizing the parameters of SVM. Additionally, MSCAHHO performs better evaluation results than SCA and HHO both in FSDE-based and CMFSDE-based methods, which also testifies the superiority of MSCAHHO.

(4) As seen from Table 8, the computation time of each model is close. To be specific, the computation time of the proposed model is the third, while the accuracy of the proposed is the highest (0.9954). This verifies the efficiency and capacity of the proposed model considering the accuracy and computation cost in fault diagnosis of rolling bearings.

Based on the above results and analyses, the proposed VMD-CMFSDE-MSCAHHO-SVM model (No. 9) can achieve the highest evaluation results and competitive stability in comparison with other methods.

6) COMPARISION AMONG DIFFERENT CLASSIFIERS

Among previous literature, various classification algorithms have been researched and applied in different fields, among which support vector data description (SVDD) [49] and back-propagation neural network (BPNN) [5] are two typical ones. In this section, a comparative experiment among BPNN, SVDD and SVM was conducted. In the comparative experiment, the parameters of BPNN are set the same as in [5], and the parameters of SVM and SVDD are selected by a classical optimization algorithm PSO. The parameters of PSO and BPNN are listed in Table 9. Thus, three comparative models are: VMD-CMFSDE-BPNN, VMD-CMFSDE-PSO-SVDD, VMD-CMFSDE-PSO-SVM. For each model, the final diagnosis results and computation time are the average of 10 trials on the same computer. The diagnosis results and computation time of different classifiers are exhibited in Table 10. Figure 15 shows the performance errors of different classifiers.

TABLE 10. Comparison results among different classifiers.

Method	Time (s)	Evaluation result			
		ARI	NMI	F	ACC
VMD-CMFSDE-BPNN	362.0	0.9558 [-0.0397, 0.0268]	0.9632 [-0.0304, 0.0256]	0.9800 [-0.0180, 0.0123]	0.9801 [-0.0184, 0.0123]
VMD-CMFSDE-PSO-SVDD	423.6	0.9432 [-0.0499, 0.0481]	0.9423 [-0.0432, 0.0465]	0.9744 [-0.0237, 0.0162]	0.9743 [-0.0241, 0.0218]
VMD-CMFSDE-PSO-SVM	423.0	0.9625 [-0.0494, 0.0375]	0.9630 [-0.0504, 0.0370]	0.9838 [-0.0215, 0.0162]	0.9835 [-0.0218, 0.0165]
VMD-CMFSDE-MSCAHHO-SVM	420.2	**0.9897 [-0.0156, 0.0103]**	**0.9896 [-0.0185, 0.0104]**	**0.9954 [-0.0070, 0.0046]**	**0.9954 [-0.0069, 0.0046]**

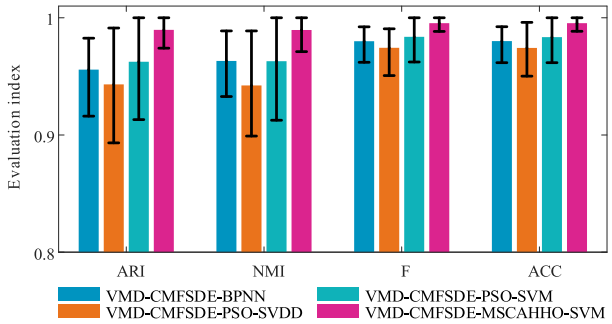


FIGURE 15. Performance errors of different classifiers.

As shown in Table 10, the average time of BPNN-based model is the lowest, but its diagnosis accuracy is about 0.9801, which is lower than those of SVM-based methods. The average time of PSO-SVM-based model is very close to that of PSO-SVDD-based model, but the diagnosis performance of PSO-SVM-based model (0.9625, 0.9630, 0.9838 and 0.9835) is much better than that of PSO-SVDD-based model (0.9432, 0.9423, 0.9744 and 0.9743). These results verify the advantage of SVM over BPNN and SVDD in fault diagnosis of rolling bearings. Compared with other three models, the proposed model combining MSCAHHO-SVM can achieve the best diagnosis performance (0.9897, 0.9896, 0.9954 and 0.9954), and its computation time is the second. Moreover, as seen from Figure 15, the performance error of the proposed model is the smallest. The results illustrate that MSCAHHO outperforms PSO for improving the performance of SVM. Overall, considering the diagnosis performance as well as computation time, SVM optimized by optimization algorithm is more suitable for fault diagnosis of rolling bearings.

VI. CONCLUSION

To analyze the complexity of vibration signal and promote the accuracy of fault diagnosis in rolling bearing, a novel hybrid fault diagnosis approach is proposed based on CMFSDE as the feature extractor and MSCAHHO optimized SVM as the fault classifier in the paper.

Within the proposed approach, the non-stationary original vibration signals were firstly decomposed into several IMFs by VMD whose decomposition number K is deduced with central frequency observation method. Afterwards, to overcome some problems existing in FSDE and CMDE, CMFSDE was put forward to construct the representative feature vectors from different fault samples. Theoretically, the proposed CMFSDE coupled the advantages of FSDE and composite multiscale coarse-grained process. Compared with FSDE, CMDE and RCMDE, CMFSDE is more suitable in feature extraction of fault signals. Subsequently, the improved optimization method named MSCAHHO, in which HHO was enhanced with SCA and mutation operator, outperforms standard SCA and HHO in terms of convergence speed and local optima avoidance by testing with the benchmark functions. Thereafter, the optimized SVM model, which employed MSCAHHO for optimization of the penalty factor C and kernel parameter g, was adopted to recognize and classify different fault samples. Lastly, the proposed VMD-CMFSDE-MSCAHHO-SVM method was compared to several relevant methods in terms of four evaluation indices including ARI, NMI, F and ACC in the contrastive experiments. The experiment results confirm that CMFSDE-based method performs the highest level of accuracy in classification comparing with FSDE-based and CMDE-based methods, and MSCAHHO is superior to the standard PSO, SCA and HHO in optimizing the parameters of SVM. Furthermore, the proposed method not only achieves much better classification performance, but also attains distinguished precision and stability as comparison to others. These results demonstrate that the proposed novel hybrid methodology could be effectively employed to the area of fault diagnosis of rolling bearings.

REFERENCES

[1] W. Gong, H. Chen, Z. Zhang, M. Zhang, R. Wang, C. Guan, and Q. Wang, "A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion," *Sensors*, vol. 19, no. 7, p. 1693, Apr. 2019.

[2] J. W. Tan, W. L. Fu, K. Wang, X. M. Xue, W. B. Hu, and Y. H. Shan, "Fault diagnosis for rolling bearing based on semi-supervised clustering and support vector data description with adaptive parameter optimization and improved decision strategy," *Appl. Sci.*, vol. 9, no. 8, p. 1676, Apr. 2019.

[3] Z.-X. Hu, Y. Wang, M.-F. Ge, and J. Liu, "Data-driven fault diagnosis method based on compressed sensing and improved multiscale network," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3216–3225, Apr. 2020.

[4] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.

[5] X. Yan and M. Jia, "Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection," *Knowl.-Based Syst.*, vol. 163, pp. 450–471, Jan. 2019.

[6] J. Ben 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," *Appl. Acoust.*, vol. 89, pp. 16–27, Mar. 2015.

[7] Z. Li, Y. Tao, A. Abu-Siada, M. A. S. Masoum, Z. Li, Y. Xu, and X. Zhao, "A new vibration testing platform for electronic current transformers," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 3, pp. 704–712, Mar. 2019.

[8] J. Zheng, Z. Dong, H. Pan, Q. Ni, T. Liu, and J. Zhang, "Composite multi-scale weighted permutation entropy and extreme learning machine based intelligent fault diagnosis for rolling bearing," *Measurement*, vol. 143, pp. 69–80, Sep. 2019.

[9] S. Wan and X. Zhang, "Teager energy entropy ratio of wavelet packet transform and its application in bearing fault diagnosis," *Entropy*, vol. 20, no. 5, p. 388, May 2018.

[10] W. Fu, K. Wang, J. Tan, and K. Zhang, "A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 205, Feb. 2020, Art. no. 112461.

[11] Q. Fu, B. Jing, P. He, S. Si, and Y. Wang, "Fault feature selection and diagnosis of rolling bearings based on EEMD and optimized Elman_AdaBoost algorithm," *IEEE Sensors J.*, vol. 18, no. 12, pp. 5024–5034, Jun. 2018.

[12] 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 Convers. Manage.*, vol. 187, pp. 356–377, May 2019.

[13] R. Abdelkader, A. Kaddour, A. Bendiabdellah, and Z. Derouiche, "Rolling bearing fault diagnosis based on an improved denoising method using the complete ensemble empirical mode decomposition and the optimized thresholding operation," *IEEE Sensors J.*, vol. 18, no. 17, pp. 7166–7172, Sep. 2018.

[14] W. Fu, K. Wang, J. Tan, and K. Shao, "Vibration tendency prediction approach for hydropower generator fused with multi-scale dominant ingredient chaotic analysis, adaptive mutation grey wolf optimizer and KELM," *Complexity*, vol. 2020, no. 2020, 2020, Art. no. 4516132.

[15] M. Zhang, Z. Jiang, and K. Feng, "Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump," *Mech. Syst. Signal Process.*, vol. 93, pp. 460–493, Sep. 2017.

[16] 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," *Meas. Sci. Technol.*, vol. 30, no. 1, Jan. 2019, Art. no. 015012.

[17] H. Shang, K. Lo, and F. Li, "Partial discharge feature extraction based on ensemble empirical mode decomposition and sample entropy," *Entropy*, vol. 19, no. 9, p. 439, Aug. 2017.

[18] W. Zhang and J. Zhou, "Fault diagnosis for rolling element bearings based on feature space reconstruction and multiscale permutation entropy," *Entropy*, vol. 21, no. 5, p. 519, May 2019.

[19] J. Zheng, J. Cheng, and Y. Yang, "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy," *Mechanism Mach. Theory*, vol. 70, pp. 441–453, Dec. 2013.

[20] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time-series analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614, May 2016.

[21] W. Fu, J. Tan, Y. Xu, K. Wang, and T. Chen, "Fault diagnosis for rolling bearings based on fine-sorted dispersion entropy and SVM optimized with mutation SCA-PSO," *Entropy*, vol. 21, no. 4, p. 404, Apr. 2019.

[22] J. Zheng, L. I. Congzhi, and H. Pan, "Application of composite multi-scale dispersion entropy in rolling bearing fault diagnosis," (in Chinese), *Noise Vibrat. Control*, vol. 38, no. S2, pp. 653–656, 2018.

[23] B. Cai, L. Huang, and M. Xie, "Bayesian networks in fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2227–2240, Oct. 2017.

[24] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral Kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.

[25] Y. Yu, D. Yu, and J. Cheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," *J. Sound Vibrat.*, vol. 294, nos. 1–2, pp. 269–277, Jun. 2006.

[26] S. Zhou, S. Qian, W. Chang, Y. Xiao, and Y. Cheng, "A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier," *Sensors*, vol. 18, no. 6, p. 1934, Jun. 2018.

[27] W. Fu, K. Wang, C. Zhang, and J. Tan, "A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine," *Trans. Inst. Meas. Control*, vol. 41, no. 15, pp. 4436–4449, Nov. 2019.

[28] W. Fu, J. Tan, X. Zhang, T. Chen, and K. Wang, "Blind parameter identification of mar model and mutation hybrid GWO-SCA optimized SVM for fault diagnosis of rotating machinery," *Complexity*, vol. 2019, pp. 1–17, Apr. 2019.

[29] H. Zhou, T. Shi, G. Liao, J. Xuan, J. Duan, L. Su, Z. He, and W. Lai, "Weighted Kernel entropy component analysis for fault diagnosis of rolling bearings," *Sensors*, vol. 17, no. 3, p. 625, Mar. 2017.

[30] S. Zhu, X. Yuan, Z. Xu, X. Luo, and H. Zhang, "Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast," *Energy Convers. Manage.*, vol. 198, Oct. 2019, Art. no. 111772.

[31] Z. Liu, H. Cao, X. Chen, Z. He, and Z. Shen, "Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings," *Neurocomputing*, vol. 99, pp. 399–410, Jan. 2013.

[32] Z. Dong, J. Zheng, S. Huang, H. Pan, and Q. Liu, "Time-shift multi-scale weighted permutation entropy and GWO-SVM based fault diagnosis approach for rolling bearing," *Entropy*, vol. 21, no. 6, p. 621, Jun. 2019.

[33] Z. Zhao, J. Yang, W. Yang, J. Hu, and M. Chen, "A coordinated optimization framework for flexible operation of pumped storage hydropower system: Nonlinear modeling, strategy optimization and decision making," *Energy Convers. Manage.*, vol. 194, pp. 75–93, Aug. 2019.

[34] M. Issa, A. E. Hassanien, D. Oliva, A. Helmi, I. Ziedan, and A. Alzohairy, "ASCA-PSO: Adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise local sequence alignment," *Expert Syst. Appl.*, vol. 99, pp. 56–70, Jun. 2018.

[35] W.-J. Niu, Z.-K. Feng, C.-T. Cheng, and J.-Z. Zhou, "Forecasting daily runoff by extreme learning machine based on quantum-behaved particle swarm optimization," *J. Hydrol. Eng.*, vol. 23, no. 3, Mar. 2018, Art. no. 04018002.

[36] C. Li, W. Wang, and D. Chen, "Multi-objective complementary scheduling of hydro-thermal-RE power system via a multi-objective hybrid grey wolf optimizer," *Energy*, vol. 171, pp. 241–255, Mar. 2019.

[37] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Gener. Comput. Syst.*, vol. 97, pp. 849–872, Aug. 2019.

[38] Y. V. Pehlivanoglu, "A new particle swarm optimization method enhanced with a periodic mutation strategy and neural networks," *IEEE Trans. Evol. Comput.*, vol. 17, no. 3, pp. 436–452, Jun. 2013.

[39] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[40] 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, Mar. 2019.

[41] R. H. Chan, M. Tao, and X. Yuan, "Constrained total variation Deblurring models and fast algorithms based on alternating direction method of multipliers," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 680–697, Jan. 2013.

[42] M. Rostaghi, M. R. Ashory, and H. Azami, "Application of dispersion entropy to status characterization of rotary machines," *J. Sound Vibrat.*, vol. 438, pp. 291–308, Jan. 2019.

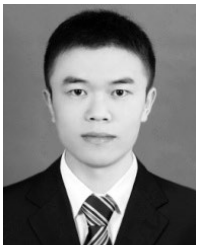
[43] H. Azami, M. Rostaghi, D. Abásolo, and J. Escudero, "Refined composite multiscale dispersion entropy and its application to biomedical signals," *IEEE Trans. Bio-Med. Eng.*, vol. 64, no. 12, pp. 2872–2879, Dec. 2017.

[44] S. Mirjalili, "SCA: A sine cosine algorithm for solving optimization problems," *Knowl.-Based Syst.*, vol. 96, pp. 120–133, Mar. 2016.

[45] S. N. Chegini, A. Bagheri, and F. Najafi, "PSOSCALF: A new hybrid PSO based on sine cosine algorithm and levy flight for solving optimization problems," *Appl. Soft Comput.*, vol. 73, pp. 697–726, Dec. 2018.

[46] N. Singh and S. Singh, "A novel hybrid GWO-SCA approach for optimization problems," *Eng. Sci. Technol., Int. J.*, vol. 20, no. 6, pp. 1586–1601, Dec. 2017.

[47] *Bearing Data Center of the Case Western Reserve University*. Accessed: Dec. 28, 2018. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file
[48] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Y. Zomaya, S. Foufou, and A. Bouras, "A survey of clustering algorithms for big data: Taxonomy and empirical analysis," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 267–279, Sep. 2014.
[49] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.



WENLONG FU received the B.S. and Ph.D. degrees in hydraulic and hydropower engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2011 and 2016, respectively. He is currently an Associate Professor with the College of Electrical Engineering and New Energy, China Three Gorges University. His research interests include machine learning, signal processing, and fault diagnosis.

KAIXUAN SHAO is currently pursuing the master's degree in electrical engineering with China Three Gorges University (CTGU). His research interests include machine learning, signal processing, and fault diagnosis.

JIAWEN TAN is currently pursuing the master's degree in electrical engineering with China Three Gorges University (CTGU). His research interests include machine learning, signal processing, and fault diagnosis.

KAI WANG is currently pursuing the master's degree in electrical engineering with China Three Gorges University (CTGU). His research interests include machine learning and signal processing.