A New Fault Diagnosis Classifier for Rolling Bearing United Multi-Scale Permutation Entropy Optimize VMD and Cuckoo Search SVM

ZIJIAN GUO, MINGLIANG LIU, YUNXIA WANG, AND HUABIN QIN

1HLJ Province Key Laboratory of Senior-Education for Electronic Engineering, Heilongjiang University, Harbin 150080, China
2Key Laboratory of Information Fusion Estimation and Detection, Harbin 150080, China

Corresponding authors: Mingliang Liu (2002039@hlju.edu.cn) and Yunxia Wang (2001019@hlju.edu.cn)

This work was supported in part by the Basic Research Fund of Heilongjiang University, Heilongjiang, under Grant kjcX201806, and in part by the Graduate Innovation Research Fund Project under Grant YJSCX2019-061HLJU.

ABSTRACT Aiming at the influence of mixed noise of bearing vibration signal on the extraction of useful information, a fault diagnosis optimize classifier based on multi-scale permutation entropy (MPE) and cuckoo search algorithm (CS) is proposed. Firstly, the MPE threshold method is adopted to select the appropriate variational mode decomposition algorithm (VMD) parameters, and then the signal is reconstructed by adding neutral white noise, and the reconstructed signal is decomposed by MPE-OVMD algorithm to obtain the optimal IMF component. Finally, the cuckoo search algorithm is used to optimize the global optimal solution of the support vector machine, thereby achieving the classification model of support vector machine with the best parameters. The analysis results of motor signals show that the method can eliminate the phenomena of mode aliasing and signal over-decomposition. An analytical comparison of the CSSVM classifier is carried out with the performance of the learners such as recall rate, ROC curve, AUC. The contrast experiment shows that the classification model can avoid misrecognition of the fault sample as the normal condition and maximum the optimal maintenance time of the equipment under the premise of ensuring the accuracy. The classifier model of the cuckoo optimization algorithm has better fitting accuracy than others such as the Grid Search algorithm (GS), Particle Swarm Optimization (PSO), Genetic Algorithm search (GA), and the ensemble fault recognition rate is as high as 90%.

INDEX TERMS Variational mode decomposition, multi-scale permutation entropy, signal reconstruction, cuckoo algorithm, support vector machine.

I. INTRODUCTION Nowadays, with the development of the industrial field scale, DC motors are increasingly required to be increasingly larger scale, with better integration and higher speed. Their working conditions directly affect the operation and safety of the equipment. Reference [1] Fault diagnosis of DC motors are very important because the rolling bearing is the key component which is easy to be damaged in rotating machinery. The bearing signals represent non-stationary and non-linear, and the current decomposition methods mainly include empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), empirical wavelet transform (EWT) and so on. EMD is a new time-frequency analysis method first proposed by Huang. It can decompose the signal into finite intrinsic mode functions (IMF) and is suitable for processing non-linear and non-stationary signals [2], [3]. Yujing et al. [4] decomposed the fault signals of the rolling bearings by EMD to obtain the basic mode components, and then the Hilbert-Huang transform is used to obtain the envelope spectrum, thus extracting the fault information of rolling bearing. However, one of the most important shortcomings of EMD time-frequency analysis method is the problem of mode aliasing [5]. Yaguo et al. [6] proposed the Ensemble Empirical Mode Decomposition method, which can effectively solve the mode aliasing problem of empirical mode decomposition method. However, the white noise added by EEMD cannot be completely neutralized and is not complete.

In 2014, Dragomiretskiy et al. proposed an adaptive signal decomposition method of variational mode decomposition
(VMD) [7], which determines the central frequency and broadband of each mode by iteratively searching for optimal solution, and achieves the effective separation of signal components in frequency domain. Compared with other decomposition methods, it can accurately separate signals with a strong mathematical foundation and higher computing efficiency. In addition, it has the characteristics of wiener filtering, which can effectively remove noise. However, the modal number \( K \) needs to be set in advance, and the value of \( K \) can only be determined by experience. Therefore, the decomposition results are easily influenced by human factors, and there is a phenomenon of excessive decomposition or insufficient decomposition, that is, when the mode is too large, excessive decomposition occurs, and an abnormal white noise component is decomposed; and modal aliasing occurs, when the mode is too small. VMD algorithm is more sensitive to noise [8]–[10], that is, the decomposition results are susceptible to background noise. Especially in the environment with strong background noise, it is easier to generate false components which easily leads to misdiagnosis of subsequent fault identification [11]–[14]. For the adaptive determination method of modal number, Liying et al. [15] determined the decomposition level \( K \) in VMD algorithm by particle swarm optimization (PSO); Zhang et al. [16] optimized the parameters of VMD algorithm with the application of grasshopper optimization algorithm (GOA). In addition, other optimization algorithms such as ant colony algorithm [17], artificial fish swarm algorithm [18] are used to optimize the parameters of VMD algorithm. Compared with empirically determining the \( K \) value, these optimization algorithms can automatically determine the \( K \) value based on the original signal and have good adaptability. However, the drawbacks of these optimization algorithms are also obvious, such as large amount of calculation, high redundancy and low computational efficiency [19].

There are two main methods of traditional fault classification: artificial neural network (ANN) and machine learning method. References [20] and [21] first introduced two methods for fault diagnosis and achieved considerable results. In recent years, artificial intelligence methods such as support vector machine (SVM), artificial neural network (ANN) and Bayesian classifier have been introduced into rolling bearing fault diagnosis [22]. In reference [23] and [24], RBF neural network and BPNN are used to identify rolling bearings. However, since the early fault data of rotating machinery is not large enough, it is difficult to train an efficient neural network with a small amount of data. In reference [25], BP neural network is used to train and identify the cavitation state of centrifugal pump. Although this method has achieved certain effects, it requires hundreds of thousands of trainings, and the amount of data calculation is too large and the time is too long. Reference [26] utilizes particle optimization algorithm to optimize the SVM to classify rolling bearings. Although this method can get the classification results more accurately, PSO algorithm cannot be widely applied in fault diagnosis of rolling bearings because it is easy to fall into local minimum.

Inspired by the findings, this study proposes a combined model of MPE-OVMD algorithm and CS algorithm optimization SVM (CSSCM). Firstly, an optimization algorithm based on multi-scale permutation entropy is proposed, and the VMD algorithm is improved from the perspective of noise-assisted data analysis [27] to further enhance the signal-to-noise ratio of the signal. At the same time, to reduce the reconstruction error and completely neutralize the added white noise, two white noises with equal amplitude and opposite symbols are added to the original signal, and then decomposed by VMD algorithm respectively. Finally, the noise in the original signal is offset. Each layer of IMF obtained is integrated and averaged respectively, and then the signals are reconstructed based on the result of the integrated mean [28]. The reconstructed signals be decomposed by the MPE-OVMD. Finally, in the case of ideal model convergence speed, the CSSCM method is used to identify the four states of rolling bearing signal (normal, asymmetry type, load-shafting misalignment, rotor system misalignment-rubbing). Compared with the models PSOSVM, GSSVM and GASVM, it is found that the MPE-OVMD-CSSCM model has better overall performance in fault diagnosis of DC motor.

II. VARIATIONAL MODE DECOMPOSITION

The essence of VMD algorithm is to find the optimal solution of variational model by iteration, and to determine the bandwidth and center frequency of each mode. Therefore, it can adaptively decompose signals in frequency domain and effectively separate each mode [29]. Its can be understood as decomposing the original signal into \( K \) modal functions \( u_k(t) \) and ensuring that the sum of the estimated bandwidth of each modal function is minimized. The above process can be expressed as follows:

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

where \( \{u_k\} \) represents the \( K \) sub-modes obtained by decomposition, and \( \{\omega_k\} \) represents the central frequency of each sub-mode. By introducing two parameters of secondary penalty factor \( \alpha \) and Lagrange multiplier \( \lambda(t) \), the expression is transformed from a non-constrained variational problem to a constrained variational problem:

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

In the formula, \( \lambda \) and \( \alpha \) represents Lagrange multiplier and penalty parameter. The Alternate Direction Method of Multipliers (ADMM) is adopted to solve the variational problem of VMD method. [30] The optimal solution of the above formula
is obtained by iteratively updating the sequences $u_{k}^{n+1}$, $\omega_{k}^{n+1}$ and $\hat{\lambda}_{k}^{n+1}$. The value of $u_{k}^{n+1}$ can be expressed as:

$$
\arg\min \left\{ \frac{1}{n!} \left\| u_{k}^{(t)} \right\|_{2}^{2} \right\}
$$

where $m$ denotes the embedding dimension, $\omega_{k}$ is equivalent to $\omega_{k}^{n+1}$ and $\sum_{i \neq k} u_{i}(t)$ is equivalent to $\sum_{i \neq k} u_{i}(t)^{n+1}$. By using Parseval’s theorem, the block formula can be changed to

$$
\hat{u}_{k}^{n+1}(\omega) = \left\{ \hat{f}(\omega) - \sum_{i \neq k} \hat{u}_{i}(\omega + \frac{\hat{\lambda}(\omega)}{2}) \right\} \cdot \frac{1}{1 + 2\alpha_{k}|\omega - \omega_{k}|^{2}}.
$$

The update result of the center frequency is as follows.

$$
\omega_{k}^{n+1} = \frac{\int_{0}^{\infty} \omega |\hat{u}_{k}(\omega)|^{2} \, d\omega}{\int_{0}^{\infty} |\hat{u}_{k}(\omega)|^{2} \, d\omega}
$$

where the result of $\omega_{k}^{n+1}(\omega)$ is equal to that of the current residual $\hat{f}(\omega) - \sum_{i \neq k} u_{i}(\omega)$, and $\omega_{k}^{n+1}$ represents the center of gravity of the power spectrum of the modal function. The specific steps of implementing the VMD algorithm are as follows:

1. The parameters $\{\hat{u}_{k}(\omega), \omega_{k}, \hat{\lambda}(\omega)\}$ are initialized.
2. Updating $u_{k}$, $\omega_{k}$ in frequency domain according to formulas (4) and (5).
3. Updating lambda, in which the expression is:

$$
\hat{\lambda}_{k}^{n+1}(\omega) = \hat{\lambda}_{k}(\omega) + \tau \left( \hat{f}(\omega) - \sum_{k} \hat{u}_{k}^{n+1}(\omega) \right)
$$

4. Stop iteration until $\sum_{k} \left\| \hat{u}_{k}^{n+1} - \hat{u}_{k}^{n} \right\|^{2} < \epsilon$.

### III. PERMUTATION ENTROPY

Permutation Entropy (PE), a method for measuring the complexity of time series was proposed by Bandt and Pompe [31]. It has the advantages of simple principle, high computational efficiency and good robustness and suit for non-linear data analysis [32], which can detect the catastrophic phenomena of complex dynamical systems.

The time series $\{s(n), n = 1, 2, \cdots N\}$ is spatially reconstructed to obtain the reconstructed vector.

$$
X_{j} = [s(j) \cdot s(j + \tau) \cdots s(j + (m - 1) \tau)]
$$

where $m$ denotes the embedding dimension, $\tau$ is the delay time, subscript $j = 1, 2, \cdots N - (m - 1) \tau$. The reconstructed vector $X_{j}$ can be used as the row vector of matrix $X$.

$$
j + (j_{1} - 1) \tau \leq j + (j_{2} - 1) \tau \leq \cdots \leq j + (j_{m} - 1) \tau
$$

wherein $j_{1}, j_{2}, \cdots, j_{m}$ represent the index number of the column. Therefore, for each row of matrix $X$ obtained by time series reconstruction $\{s(n), n = 1, 2, \cdots , N\}$, a set of symbolic sequences can be obtained.

$$
S(g) = (j_{1}, j_{2}, \cdots, j_{m})
$$

where $g = 1, 2, \cdots, 1$, and there is a common arrangement of the sequences of non-symbols in $1 \leq m!$. After calculation of the probability of occurrence of each symbol sequence $P_{1}, P_{2}, \cdots, P_{l}$, We get $\sum_{g=1}^{l} P_{g} = 1$.

### IV. MULTI-SCALE PERMUTATION ENTROPY

Being defined the entropy value of time series in different scales, multi-scale entropy (MPE) is obtained by coarsening process and reflects the complexity of time series. If the MPE of time series monotonically increases as the scale factor increases, the sequence contains more information on multiple scales [33], [34]. Firstly, the time series is multi-scale coarsened, and then the ranking entropy of different scale coarsened sequences is calculated. Set the length of one-dimensional time series $X = \{x_{i}, i = 1, 2, \cdots, N\}$ as $N$, the sequence can be obtained by coarsening:

$$
y_{j}^{(s)} = \frac{1}{s} \sum_{i = (j-1)s+1}^{j} x_{i}, j = 1, 2, \cdots, \left\lfloor \frac{N}{s} \right\rfloor
$$

where $s$ is the scale factor, $s = 1, 2, \cdots, \left\lfloor N/s \right\rfloor$ denotes the integer of $N/s$. When $s=1$, the coarse-grained sequence is the original sequence. Time sequence of coarse-grained sequence $y_{j}^{(s)}$:

$$
Y_{l}^{(s)} = \{y_{l}^{(s)}, y_{l+1}^{(s)}, \cdots, y_{l+(m-1)\tau}^{(s)}\}
$$

$$
\gamma_{l+i=(l_{1}-1)\tau}^{(s)} \leq \gamma_{l+i=(l_{2}-1)\tau}^{(s)} \leq \cdots \leq \gamma_{l+i=(l_{m}-1)\tau}^{(s)}
$$

Where $y_{l+i=(l-1)\tau}^{(s)} \leq y_{l+i=(l+1)\tau}^{(s)} \leq \cdots \leq y_{l+i=(m-1)\tau}^{(s)}$

For any coarse-grained sequence $y_{j}^{(s)}$, a set of symbolic sequences $S(r) = (l_{1}, l_{2}, \cdots, l_{m})$ can be obtained, where $r = 1, 2, \cdots, R$ and $R \leq m!$. The probability of occurrence of each symbol sequence is calculated. The ranking entropy $H_{p}(m)$ of different symbol sequences is defined as:

$$
H_{p}(m) = - \sum_{r=1}^{R} P_{r} \ln P_{r}
$$

When $P_{r} = 1/m!$, $H_{p}(m)$ reaches its maximum value $\ln (m!)$. For convenience, $H_{p}(m)$ is usually normalized:

$$
H_{p} = \frac{H_{p}(m)}{\ln (m!)}
$$

$H_{p}$ is the normalized ranking entropy. Obviously $0 \leq H_{p} \leq 1$, the smaller the value, the more regular the time series.

### V. OPTIMIZATION OF VMD BASED ON MULTI-SCALE PERMUTATION ENTROPY

In order to adaptively determine the number of decomposition layers, a multi-scale permutation entropy optimization (MPE-OVMD) VMD is proposed. The principle of the algorithm is to calculate the permutation entropy of each layer of
inherent mode function obtained by decomposing the original signal. Due to the randomness of the abnormal components, the permutation entropy of the abnormal components is much larger than that of the normal components. Therefore, after setting the threshold \( H_P \) of the permutation entropy, we can judge whether the permutation entropy of each IMF layer in the decomposition result is larger than the threshold \( H_P \), thereby determining whether there is an abnormal component in the decomposition result. To select the permutation entropy threshold \( H_P \), the value of \( H_P \) is set to 0.5 based on the linear relationship between the randomness of time series and its corresponding \( H_P \) value in reference [35]. The specific steps of the algorithm are as follows:

1. Set the initialization of \( K \) to 2, and the entropy threshold is setting 0.5.
2. The original signal is decomposed by VMD algorithm to obtain \( K \) modal components \( \text{IMF}_i(t) (i = 1 \sim K) \).
3. Calculate the multi-scale permutation entropy \( \text{MPE}_i (i = 1 \sim K) \) of each IMF in the decomposition result.
4. Judging whether \( \text{MPE}_i \) is greater than the threshold value of 0.5.

If the judgment is correct, it indicates that the signal is excessively decomposed, resulting in abnormal components, and the optimal mode number for stopping the cycle is \( K-1 \), otherwise, it means that for modes overlap, and then return to step, repeat the above steps to continue the VMD decomposition of the original signal. In order to improve the signal-to-noise ratio (SNR), a signal reconstruction method based on neutral and Gaussian white noise is proposed. According to reference 37, CEMD adds positive and negative Gaussian white noise pairs to the undecomposed signal to reduce the signal reconstruction errors and completely neutralize the added white noise. According to formula 4 and formula 5, if Gaussian white noise is simply added to the undecomposed signal, the reconstruction error will be increased. Based on the discussion, this paper proposes a signal decomposition method based on neutral and gaussian white noise, which neutralizes the idea of adding positive and negative Gaussian noise pairs: the white noise added in each cycle is neutralized by two positive and negative white noise pairs with the same amplitudes and opposite symbols. After adding the auxiliary white noise, it is decomposed by VMD algorithm. After several cycles, the IMFs of each layer obtained from each cycle are separately integrated and averaged, and the signals are reconstructed based on the results of the integrated average. The specific steps are as follows:

1. Parameter initialization, the value of \( K \) is determined the number of cycles \( N \), the amplitude of white noise are set.
2. The original signal \( x(t) \) is added with a pair of Gaussian white \( \text{noise}_1, \text{noise}_2 \), with opposite symbols, mean value 0 and constant standard deviation of amplitude value, and two signals to be decomposed \( x_1(t), x_2(t) \) are obtained:

\[
x_1(t) = x(t) + \text{noise}_1
\]
\[
x_2(t) = x(t) - \text{noise}_2
\]

3. Two sets of IMFs are obtained by VMD decomposition of \( x_1(t), x_2(t) \) as follows

\[
\begin{aligned}
\text{imf}_{1,i,j}(t) & \quad (i = 1 \sim N, j = 1 \sim K) \\
\text{imf}_{2,i,j}(t) & \quad (i = 1 \sim N, j = 1 \sim K)
\end{aligned}
\]  \( (16) \)

where \( \text{imf}_{1,i,j}(t) \) denotes the jth IMF component after the first decomposition of signal \( x_1(t) \), \( \text{imf}_{1,i,j}(t) \) denotes the jth IMF component after the first decomposition of signal \( x_2(t) \).

4. Repeat steps 2, 3, the final \( 2N*K \) IMFs are integrated averagely:

\[
\text{imf}_j(t) = \frac{1}{2N} \sum_{i=1}^{N} \left[ \text{imf}_{1,i,j}(t) + \text{imf}_{2,i,j}(t) \right] \quad (j = 1 \sim K)
\]  \( (17) \)

where \( \text{imf}_j(t) \) represents the integrated mean of IMF components in layer J of all decomposition results.

5. The signal is reconstructed and the reconstructed signal \( x_0(t) \) is obtained, and then decomposed by VMD to get the final IMF component:

\[
x_0(t) = \sum_{j=1}^{K} \text{imf}_j(t) \quad (j = 1 \sim K)
\]  \( (18) \)

The process is shown in figure 1:

![Figure 1. Flow-process diagram of MPE-OVMD method.](image)

The simulated signal consists of three types of subsets, namely three different harmonics, \( x(t) = x_1 + x_2 + x_3 \). Wherein, \( x_1 = \cos(6 * \pi * t) \), \( x_2 = 1/9 \sin(52 * \pi * t) \), \( x_3 = 1/27 \cos(600 * \pi * t) \). The simulation signal \( x(t) \) is decomposed by VMD. The simulation signal and its decomposition diagram are shown in Figure 2. Figure 2 b is IMF component by MPE-OVMD. By comparison, the MPE-OVMD can be eliminated the beginning and end of the signals \( u_2 \) and \( u_3 \) are relatively large error.

The three modal components (presupposition \( K=3 \)) obtained by the decomposing simulation are signals \( u_1, u_2 \) and \( u_3 \), which correspond to the input signals \( x_1, x_2 \) and \( x_3 \),
respectively. The amplitude and frequency of the modal component are well matched to the original signal, as shown in Figure 1. The spectrum distribution of the input signal is shown in Figure 3 and the obtained spectrum distribution of each modal signal is shown in Figure 4. Compared with the original signal, the results show that the three modal frequencies are consistent with the original signals in the frequency range.

Table 1 shows the variation of the central frequency for different K values. It can be indicated that when the K value in the table is 3, there is no similar modal and modal mixing phenomenon at the central frequency. When K=2, modal mixing occurs in the decomposition of periodic signals. When K=4 or 5, the number of modes in Table 1 is found to be similar, and as the value of k increases, a pseudo modality occurs during the decomposition.

| Mode Number | Center Frequency |
|-------------|------------------|
| U1          | 25.76            |
| U2          | 1884.32          |
| U3          | 1884.96          |
| U4          | 1884.96          |
| U5          | 1885.58          |

Then, it is verified whether the decomposition layers of simulation signal can be accurately determined by MPE method. By adding positive and negative Gauss function to the simulation signal, a new simulation signal is constructed, and the simulation signal is decomposed layer by layer by VMD algorithm to obtain multi-scale permutation entropy values of each layer, the end threshold is set to 0.5. Table 2 indicates that when K=4, a multi-scale permutation entropy value is greater than the threshold value of 0.5. Therefore, the optimum mode number for determining the signal is 3, which is consistent with the result obtained by the observing center frequency method. It is proved that this method is theoretically feasible.

| Mode Number | Multi-scale Permutation Entropy |
|-------------|---------------------------------|
| K=2         | 0.1085  0.2256  0.000  0.000  0.000 |
| K=3         | 0.1071  0.2315  0.2328  0.000  0.000 |
| K=4         | 0.1074  0.2308  0.2394  0.5305  0.000 |
| K=5         | 0.1074  0.2318  0.2402  0.4886  0.4946 |
VI. MULTI-CLASS SVM AND PARAMETER DESIGN
Support vector machine (SVM) is a machine learning algorithm based on statistical learning theory, which is an approximate implementation of structural risk minimization and can solve small sample and nonlinear problems well. When dealing with multi-classification problems, multi-classifiers can be constructed by combining multiple binary classifiers.

A. TWO CLASSES OF LINEAR SEPARABILITY
The black and white dots in figure 5 represent two types of samples, with the solid line representing the optimal classification surface and the two dashed lines passing through the points closest to the optimal classification of the two types of samples, parallel to the optimal classification surface. The interval between the two dashed lines is called the classification interval [37]. The optimal classification line not only separates the samples without errors, but also maximizes the classification interval [38].

Assuming linear separable samples \((x_i, y_i)\), where \(i = 1, 2, 3 \ldots, n\), \(x_i \in R^n\), \(y_i \in \{-1, +1\}\) is a class label. The general form of linear discriminant function in n-dimensional space is \(g(x) = \omega \cdot x + b\). The classification surface equation is \(\omega \cdot x + b = 0\). In order to make the classification line correctly classify all samples, it should satisfy:

\[
y_i \left[ (\omega \cdot x_i + b) \right] \geq 1 \quad i = 1, 2, \ldots, n
\]

The sample points established the above formula are the points through which two dashed lines pass, which is called support vector. Under the constraint condition 13, the minimum of \(||w||^2\) is obtained. Therefore, Lagrange functions can be defined:

\[
\mathcal{L}(\omega, b, \alpha) = \frac{1}{2} (\omega \cdot \omega) - \sum_{i=1}^{n} \alpha_i [y_i (\omega \cdot x_i) + b] - 1
\]

(20)

Lagrange coefficient \(\alpha_i > 0\), the original problem can be simplified to the following dual problem under constraint conditions

\[
\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \alpha_i \geq 0 \quad i = 1, 2, \ldots, n
\]

(21)

Solving Maximum

\[
Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i x_i \cdot x_j
\]

(22)

If \(\alpha_i^*\) is the optimal solution:

\[
\omega^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i
\]

(23)

\[
f(x) = \text{sgn} (\omega^* \cdot x + b^*) = \text{sgn} \left( \sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^* \right)
\]

(24)

sgn is a symbolic function, \(b^*\) is the threshold for classification. \(x\) is the sample to be classified, and \(x_i\) is the training sample.

B. LINEARLY NONSEPARABLE CASE
In the case of linear inequalities, some training samples do not satisfy equation 23. Adding a slack term in 24:

\[
y_i \left[ (\omega \cdot x_i + b) \right] \geq 1 - \varepsilon_i \quad (i = 1, 2, \ldots, n)
\]

(25)

Calculating the minimum of following formula:

\[
\phi(\omega, \epsilon) = \frac{1}{2} (\omega \cdot \omega) + C \sum_{i=1}^{n} \epsilon_i
\]

(26)

Here C is the penalty coefficient, which controls the degree of penalty for fault samples. \(\alpha_i\) is expressed:

\[
0 \leq \alpha_i \leq C, \quad (i = 1, 2, \ldots, n)
\]

(27)

The input space is transformed into a high-dimensional space by a non-linear transformation, and then the optimal linear classification surface is obtained in the new space. The non-linear transformation is implemented by an appropriate inner product function. Radial Basis Function:

\[
K(x_i \cdot x) = \exp \left( -\frac{||x_i - x||^2}{2g^2} \right)
\]

(28)

Now, the corresponding discriminant function becomes:

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{n} \alpha_i^* y_i K(x_i \cdot x) + b^* \right)
\]

(29)

The above formula is SVM, which converts \((x_i \cdot x)\) into \(K(x_i \cdot x)\) in 28, \(g\) is the kernel function parameter, \(x_i\) is the training sample, and \(x\) is the test sample.

C. CUCKOO ALGORITHM
Cuckoo Search(CS) [42] is a meta-heuristic algorithm which is based on the parasitic reproduction strategy of cuckoo population itself. It not only combines the unique levy flight mode of birds and Drosophila to search, but also increases the information exchange between groups, and speeds up the convergence. Moreover, it has fewer parameters and is easy
to implement. In order to simulate the nest-seeking behavior of cuckoos, the CS algorithm sets three rules as follows [43]:

1. Cuckoos lay only one egg at each time, which represents a solution to the problem, and the eggs are randomly placed in the nest for hatching.

2. Among these nests, some of which contain high-quality eggs, is a good solution to the problem, and these nests will be reserved for the next generation.

3. The probability that the owner of a nest finds that the egg is an outsider is set to be \( P_a \), where \( P_a \in [0, 1] \). \( x_i(0) \) denotes the location of the first nest in the T generation and \( L(\lambda) \) denotes the random search path, then the formula for updating the path and location of cuckoo’s nest is as follows:

\[
x_i(t+1) = x_i(t) + \delta \odot L(\lambda) \quad i = 1, 2, \ldots, n
\]  

(30)

In the formula, \( \delta \) represents the step size control while \( \odot \) denotes point-to-point multiplication. After location is updated, a number \( r \) of [0, 1] is randomly generated. If \( r > P_a \), \( x_i(t+1) \) will be changed randomly, and vice versa. Finally, a group of nest locations \( x_i(t+1) \) with better test values is retained, which is still recorded as \( x_i(t+1) \).

The step size generated by levy flight is random and lacking self-adaptability, which cannot guarantee fast convergence. Therefore, the step size is adaptively and dynamically adjusted according to the search results of different stages, and the adaptive step-size adjustment strategy for the optimal nest location is set as follows:

\[
step_i = step_{\text{min}} + (step_{\text{max}} - step_{\text{min}}) \cdot d_i
\]  

(31)

\[
d_i = \frac{\|n_i - n_{\text{max}}\|}{d_{\text{max}}}
\]  

(32)

In the formula, \( n_i \) denotes the position of the first nest, \( n_{\text{max}} \) denotes the optimal position of the nest at the moment, and \( d_{\text{max}} \) denotes the maximum distance from the optimal position to other nests. When the position of the first nest is closer to the optimal position, the step size becomes smaller, the farther away from the optimal position, the larger the step size. In this way, based on the results of previous iterations, the current moving step size can be dynamically updated with better adaptability.

CS optimize SVM Parameters steps:

1. The range of parameters \( c \) and \( g \) of SVM, the minimum step size of CS algorithm \( \text{min}_{\text{step}} \), maximum step size of CS algorithm \( \text{max}_{\text{step}} \), and the number of iterations \( N \) are determined according to experience.

2. Set the initial probability parameter \( P_a \) to 0.25, and randomly generate the position \( x_i(0) = [x_1(0), x_2(0), x_3(0), \ldots, x_n(0)]^T \) of nests. Each nest corresponds to a set of parameters \( (c, g) \).

3. Calculating the fitting degree of training set for each nest position, and find the best nest position \( x_b(0) \) and the best fitting degree \( F_{\text{max}} \). The mean square error (MSE) of SVM output and the expected output are calculated as the fitness, and the formula is as follows:

\[
E_{\text{ms}} = \frac{1}{X} \sum_{i=1}^{X} (\mu_i^* - \mu_{i1})^2
\]  

(33)

Expected output is \( \mu_i^* \) and the actual output is \( \mu_{i1} \). The initial global optimal position is selected and retained to the next generation.

4. Retaining the position \( x_i(0) \) of the best nest of the previous generation, calculating the Levy flight step according to formula (30) and formula (31), and updating the position of other nests by Levy flight to get a new set of nest positions, and calculating their fitness \( F \).

5. According to the fitness \( F \), the position of the new nest is compared with that of the previous generation of nest \( p_{i-1} \). A group of new nest positions \( p_s = [x_1(t), x_2(t), x_3(t), \ldots, x_m(t)]^T \) are obtained by replacing the worse ones with the better ones.

6. By comparing the random number \( R \) with \( P_a \), the nests with less probability of being found in \( p_s \) are retained, and the nests with higher probability are updated to calculate the fit of the new nests, which is compared with that of the nests in \( p_s \). A new set of better nests are obtained by replacing the bad ones with the better ones. Location \( P_x(t+1) \) in step (6) and determining whether the fitness \( F \) meets the requirement. Stop searching if it meets the requirement, and output the global best fitness \( F_{\text{max}} \) and its corresponding optimal nest \( x_b(t) \). If it does not meet the requirement, return to step (4) to continue optimizing.

7. The parameters of SVM are set according to the optimal parameters \( c \) and \( g \) corresponding to the optimal nest location \( x_b(t) \). Its specific flow chart is as follows in figure 6:

**FIGURE 6. CS algorithm optimize support vector machine parameter.**

**VII. EXPERIMENTAL ANALYSIS**

The data acquisition system model which was used for mechanical vibration is INV1612. Among its many strengths are accurate measurement, high sampling rate, and powerful anti-jamming performance. The type of DC motor...
was a ZHS-5 multi-function rotor experiment bench, and SLM-6000 sensor signal conditioner acquired data has six test channels and an ADC resolution of 16 bits. These advantages were suitable for continuous signal acquisition in this experiment. Besides the rotor experiment bench and signal conditioner, an LC0159 piezoelectric acceleration sensor, a DC stabilized voltage source, and an AFT-0931 signal conditioner were applied to set up the vibration signal acquisition system. The frequency range of the LC0159 piezoelectric acceleration sensor was 112,000Hz, and the range and sensitivity were 500g and 10mV/g, respectively, while the frequency response and gain of the AFT-0931 signal conditioner were 0.5–45kHz and 5 kHz, respectively.

In order to verify the effectiveness of the proposed MPE-OVMD and CS-SVM based rolling bearings fault diagnosis methods, this paper adopts INV-1612 rolling bearing experimental data to study the fault diagnosis. The test drive end bearing is the 6205-2RSJEMSKF deep groove ball bearing. The EDM technology is used to make a single point of damage to the bearing, and the acceleration signal is collected by an acceleration sensor mounted on the end of the end drive bearing at a sampling frequency of 12 kHz. The experimental equipment is shown in the figure 7, 8. Fault location and vibration sensor location have been marked in the figure 7, the vibration sensor can be placed at any source position.

When the motor speed is 1750 r/min, vibration signals of the bearing four typical states (normal, fault I: rotor asymmetry, fault II: load-shafting misalignment, fault III: rotor system misalignment-rubbing coupling fault) are analyzed. Each state takes 50 sets of data with a length of 2048. We add positive and negative white noise pairs to the signal to be decomposed. Compared with EEMD simply adding noise pairs, the method of adding positive and negative Gauss white noise pairs to signals can reduce reconstruction error and promote the mutual neutralization of white noise. As shown in figure 9, this figure is the original to-be-decomposed signal without reconstruction. It can be seen from the figure that the sampling signal is too sparse in the acquisition process of the equipment, which seriously affects the decomposition of the signal. Figure 10 shows the sampled signal optimized by the MPE algorithm. In contrast, the MPE optimization algorithm can densify the sparse signal and eliminate the
noise in the acquisition process, which proves that the optimization algorithm can achieve better results in the signal denoising direction.

As shown in Figure 10, the four type reconstructed signal is modally determined by MPE-OVMD algorithm. The optimal modulus of VMD algorithm is determined based on whether the multi-scale permutation entropy with reference to different K-value modes is higher than the threshold value of 0.5.

As shown in Table 3, this experiment takes the multi-scale permutation entropy table obtained by fault-signal decomposition as an example. When $K=7$, the seventh mode MPE value has exceeded the preset threshold of 0.5. The other three faults are simultaneously decomposed by MPE-VMD algorithm, and the all MPE values are larger than the preset threshold when $K=7$. Therefore, the modal number K value of the VMD algorithm in this paper is 6.

In Figure 11, the reconstructed signal is decomposed into six-time domain waveforms of signal decomposition by MPE-OVMD algorithm. The Shannon entropy of all modes of each signal is calculated, and the modal with the maximum information entropy value is found as the optimum mode, which is the mode with the most fault information contained in the optimal decomposition layer of VMD algorithm. Its envelope entropy and amplitude spectrum contain abundant fault information. By analyzing its envelope entropy and amplitude spectrum, abundant fault information of four states in time domain waveforms can be obtained, which further proves the effectiveness of the proposed MPE-OVMD algorithm.

This paper compares this algorithm with the EMD algorithm. The modal components obtained by decomposing EMD algorithm in the figure 12. The algorithm divides the sequence into several IMF components without deviating from the time domain. EMD is not based on physics (principle), instead, these modes provide many signals in these data. However, the modulus of the traditional EMD algorithm is uncertain, and the phenomenon of modal aliasing is prone to occur during signal decomposition. Although EMD methods (such as EEMD and CEEMD) have been improved in recent years, the modal aliasing phenomenon has been somewhat mitigated, but it cannot be eliminated. As shown in Figure 11, each IMF component decomposed by the MPE-OVMD algorithm, that is, the instantaneous frequencies of each analytical signal, has an actual physical meaning. The algorithm can eliminate modal aliasing. In Figure 12, the four type reconstructed signal will have several IMF components through EMD algorithm. It can be seen from the figure that the fourth or fifth component EMD algorithm will generate residual components, which indicates that the algorithm will generate a large misunderstanding in signal decomposition, so that the decomposed signal cannot fully reflect the vibration information carried by the original signal.

As the same time, according to the characteristics of envelope entropy and amplitude spectrum, the envelope

| MODE VALUE | MULTI-SCALE PERMUTATION ENTROPY |
|------------|-------------------------------|
| K=2        | 0.2731 0.4357 0 0 0 0 0 0    |
| K=3        | 0.2679 0.3935 0.4349 0 0 0 0 0 |
| K=4        | 0.2624 0.3917 0.4321 0.4428 0 0 0 0 |
| K=5        | 0.2317 0.3702 0.4267 0.4526 0.4601 0 0 0 |
| K=6        | 0.2279 0.3648 0.4246 0.4335 0.4508 0.4523 0 0 |
| K=7        | 0.2229 0.3680 0.4221 0.4343 0.4320 0.4457 0.5014 0 |

FIGURE 10. Reconstructing Time Domain Waveform of Signal.
entropy and amplitude spectrum of the optimal modes of MPE-OVMD and EMD are given. When a signal is decomposed into several modes, if the IMF component contains more noise and the periodic impact characteristics related to the faults are not obvious, the sparsity of the component signal is diminished and the envelope spectrum will show multiple peaks. If the IMF component contains a lot of fault signature information, a regular impulse will appear in the waveform, and the signal will show a strong sparse characteristic, and the envelope spectrum will show a single peak. Taking frequency as independent variable and the amplitude of each frequency component as dependent variable, such a frequency function is called amplitude spectrum, which characterizes the distribution of signal amplitude with frequency.

Based on the above discussion, the information entropy values of all IMF components processed by VMD algorithm and EMD algorithm under this parameter are calculated, and the one with the smallest information entropy values is called the local minimum value, which is represented by $\min_{1 \leq p \leq L} E_p^{IMF}$. The IMF components corresponding to the global minimum entropy value are the best components which contain abundant fault feature information.

In this paper, the optimal modal corresponding components of four signals are found by this method, and their envelope entropy spectrum and amplitude spectrum are extracted for analysis. Figure 13 is an envelope spectrum and amplitude spectrum of the optimal component of IMF in MPE-OVMD algorithm. The amplitude spectrum shows the amplitude of the instantaneous frequency of the best component and its amplitude is expressed in the form of single peak value, which shows that the characteristic frequency of the vibration signal has been completely decomposed, and the envelope entropy spectrum shows a single peak value, and there is no multi-peak phenomenon. The envelope spectrum also proves that the vibration signal of this mode is completely extracted, and the envelope entropy value completely presents a single peak state. The characteristic frequency and the double spectrum line are very prominent, which indicates that the characteristic frequency has been completely extracted.

In order to verify the advantages of this method, the EMD algorithm is used to process the measured signals, and an envelope demodulation operation is performed. The best envelope spectrum is compared to the method in this paper. In figure 12, it is a result of performing EMD processing on the measured signals, and a total of decomposition components are obtained. After comparison, it is found that the characteristic frequency related component only appears in the envelope spectrum of the C1 component, and the demodulation effect is the best. The result is shown in figure 14. Compared with the decomposition method proposed in this paper, the fault characteristic frequency and the spectral line amplitude at double frequency are not prominent, and with serious background noise and excessive interference spectrum lines. The analysis effect is not good.

A. VMD DECOMPOSITION RESULTS
In order to verify that the proposed method is superior to common VMD algorithm, the above four faults are decomposed by VMD and MPE-OVMD respectively. Table 4 represents a central frequency observation method for determining the number of layers by VMD algorithm. The method observes
the central frequency of the IMF component of VMD algorithm, and finally selects the best mode. When K=2, the difference in center frequency is too large, indicating that mode aliasing has occurred in the signal. When K = 5, the center frequency of the fourth IMF is similar to that of the fifth IMF, indicating that the VMD algorithm has been over-decomposed. Based on the above discussion, the VMD algorithm takes the mode number K=4. The figure 15 is the waveform of VMD decomposition. It can be seen from decomposition waveform that there is no mode aliasing in VMD algorithm, but is K=4 the best decomposition layer of the measured signal?

To further verify the optimal modal number of the measured signals, the envelope entropy and amplitude spectrum analysis of the four states of the VMD algorithm are carried out in the figure 16, and compared with the MPE-OVMD algorithm proposed in. Compared with the envelope entropy and amplitude spectrum of MPE-OVMD, the envelope entropy of VMD algorithm presents a single peak, but its fourth graph shows more than one peak, indicating that modal aliasing may have occurred in the algorithm. To further prove this conjecture, this paper gives the fault characteristic frequencies of the two methods (taking asymmetry type as an example). Compared to other IMF components, the fault characteristic frequencies of the fourth IMF component have a larger fault frequency range, which shows that the fault frequency can be continually decomposed. It can also continue to decompose, indicating that when K=4, mode aliasing has already occurred in the original signal decomposition.

Figure 17 b is the fault signature frequencies for the six IMF components of MPE-OVMD (asymmetry type). It can
be seen from the figure that the six IMFs of the original signal of inner-loop fault can completely extract the fault frequencies, and the frequency amplitude is higher than that of VMD algorithm. Compared with the incomplete extraction of feature frequency in the VMD algorithm (figure 17a),
the proposed algorithm can extract all fault information to completely decompose the fault signal and enable the information contained in each mode to be the best information of the fault signal. The effectiveness of the proposed method in this paper is verified again. Appeal comparison method proves that the traditional VMD algorithm is flawed, and
the proposed algorithm can perfectly separate the fault signal without modal aliasing. This method is more time-efficient and applicable in practical measurement.

At the same time, the orthogonal performance index is given. When the IO value is smaller, it indicates that the mode number is changed to the optimal mode number. As shown in Figure 6, when \( k = 6 \), the IO value is the minimum, which indicates that the traditional central frequency method can not determine the modal number well, and the VMD algorithm based on parameter optimization can accurately determine the modal number.

**B. VIBRATION SIGNAL RECOGNITION OF ROLLING BEARING**

In this paper, IMF modal features are extracted and organized into three common bearing fault conditions: inner ring fault, balls fault, outer ring fault (the load area is centered at 6:00)

---

**TABLE 6. Fault types and data classification of data sets.**

| Fault Type | Normal | Asymmetry Type | Load-shafting Misalignment | Rotor System Misalignment-rubbing |
|------------|--------|----------------|---------------------------|----------------------------------|
| Training Set | 30     | 30             | 30                        | 30                               |
| Test Set   | 20     | 20             | 20                        | 20                               |
| Total Data Set | 50     | 50             | 50                        | 50                               |

Thirty sets of data for each state sample are extracted from 200 sets of bearing data samples as the training sets, and the remaining 20 sets of data are used as the test sets. The training set and test set are partitioned as Table 7.

According to the classification principle of SVM classifier, when RBF kernel function is used to solve the optimal decision surface, the selection of parameter penalty coefficient parameter \( C \) and kernel function parameter \( g \) has a great influence on the performance of SVM classifier. Therefore, it is necessary to study the parameter optimization of SVM. At present, little research has been done on the optimization of SVM parameters in the field of fault diagnosis of high voltage circuit breakers. In reference [44], Particle Swarm Optimization (PSO) algorithm is used to optimize single-class support vector machines and is applied to fault diagnosis of rotary bearings as well. In reference [45], the combination of grid search and cross-validation is used to parameterize the support vector machines, and it is applied in the mechanical fault diagnosis of high voltage circuit breakers. The limitations of local extremum in PSO parameterization method are pointed out. Based on the above discussion, this paper optimizes the parameters of SVM by various methods, and observes the optimization process of GS, PSO, GA and CS algorithm respectively. The cuckoo algorithm has better fitness than the PSO, GA and GS algorithms, and the convergence speed of cuckoo algorithm is quite fast. The cuckoo algorithm converges to the optimal nest in the 15th generation, which has obvious advantages over PSO, GA, and GS.

For ease of comparison, particle swarm optimization (PSO), genetic algorithm (GA) and grid search algorithm (GS) are used to diagnose transformer faults with the same training set and test set samples. Figure 18, 19 and 20 represent the optimization process of GS, GA and PSO algorithms, respectively. They show the comparison of the results of the corresponding transformer fault samples. As can be seen from figures 18a and 18b, the accuracies of training set and testing set of GS algorithm are 90% and 80%, respectively.

Figure 19a and 19b show that the fitness curve tends to be flat in the 80th generation and eventually tends to be straight, but the average fitness is constantly updated, and the average fitness result is much lower than its optimal fitness. The accuracies of training set and test set are 94.17% and 80%, respectively:
Figure 20a and 20b represent that the fitness curve tends to flatten and converge in the 20th generation, but the average fitness of SVM optimization of PSO is far lower than the optimal fitness. The parameter optimization results of the algorithm need to be discussed. The accuracies of training set and test set are 90.833 % and 86.25%, respectively, which is the worst performance of the four classifiers.

In this paper, set the parameters c of SVM and g of RBF kernel function, and set the total number of nests to n=25. The probability PA of finding exotic birds is 0.25 and the maximum number of iterations is 100. Figure 21 shows the fitness curve of parameter optimization of CSSVM model.
TABLE 7. Comparison of actual classification and predict classification of testing samples.

| Parameters Setting (C, g) | Normal | Asymmetry Type | Load-shafting Misalignment | Rotor System Misalignment -rubbing | Test Samples |
|---------------------------|--------|----------------|---------------------------|-----------------------------------|--------------|
| GSSVM                     | C=11.313, g=16 | 95% | 85% | 80% | 90% | 87.5% |
| GASVM                     | C=11.639, g=31 | 100% | 85% | 75% | 90% | 87.5% |
| PSOSVM                    | C=2.658, g=72 | 100% | 75% | 80% | 90% | 86.25% |
| CSSVM                     | C=79.65, g=2.45 | 95% | 90% | 80% | 95% | 90% |
| GWSVM                     | C=38.75, g=40 | 90% | 90% | 80% | 85% | 85% |
| ELM                       | Activation Function (Sig); | 90% | 90% | 85% | 80% | 86.25% |
| CNN                       | Convolution layer(2) | 95% | 90% | 85% | 85% | 88.75% |
|                           | Convolution kernel(3*3) |                |                |                |              |
|                           | Activation Function (Sigmoid) |                |                |                |              |

The optimal parameters are C=79.64, g=2.4586. As can be seen from Figure 21, the fitness curve converges faster in the first seven cycles, then gradually becomes flat, and finally tends to be straight, thus realizing parameter optimization. Fig 21b shows that in the 80 sets of test samples (20 sets of faults per set), three sets of samples diagnose one sample incorrectly. The accuracies of the training set and test set are 99% and 90%, respectively. The conclusion shows that the method can be effectively applied to fault diagnosis of rotating machinery bearings.

Figure 22 shows the cuckoo algorithm search process. Set the final traversal interval to [0,25], and set the traversal step to 5.

It can be seen from Fig. 20 that the parameter search process under the same parameters shows the trend of discounted fluctuations in the interval, and the best parameters are obtained in the fifteenth and eleventh generations.

TABLE 8. Confusion matrix of binary classification model.

| True situation | Positive examples | Negative examples |
|----------------|-------------------|-------------------|
| Forecast Result | TP (True Positive) | FN (False Negative) |
| Positive examples |                   |                   |
| Negative examples | FP (False positive) | TN (True Negative) |

Table 8 lists the parameters of the decision tree and shows the final recognition results of these three classifiers and the recognition rates of different parameters. Optimal parameters for each method are shown in this table. Due to the similar fault modes, the recognition rates of fault I and fault II were not much different, indicating that the vibration signals of the fault III state greatly differed from those of the other three states. The optimized random forest classifier was compared with CSSVM, GASVM, and PSOSVM classifiers. The results revealed that the recognition rate of the improved SVM classifier was higher than those of other methods by the same It can be seen from figure 20 that the parameter search process under the same parameters shows the trend of discounted fluctuations in the interval, and the best parameters are obtained in the fifteenth and eleventh generations.

At the same time, ELM and CNN network parameters is set. It is much better to refer to multiple small convolution kernels for superposition than to use one large convolution kernel alone. Under the condition of constant connectivity, the number of parameters and calculation complexity are greatly reduced. Therefore, the CNN convolution kernel is set as 3*3. Compared with the diagnosis results of ELM and CNN, the diagnosis results of the two mainstream neural networks are not higher than CSSVM. In this case, due to the lack of data in the early DC motor operation, it directly leads to the lack of data in the neural network training, and can not
accurately identify faults. Although, ELM, CNN and other neural networks are efficient, they need a lot of data support and are not suitable for this DC motor bearing diagnosis.

VIII. PERFORMANCE METRICS OF LESRNERS

In the field of machine learning, the indicators commonly used to measure the performance of learners are accuracy, precision, recall, score, PR curve, ROC curve, AUC, etc. These indicators are applicable to classification problems:

1. Accuracy: Accuracy measures the ratio of the number of samples correctly identified by the classifier to the total number of samples in the test set.

2. Precision and recall: For binary classification problems, the category concerned is usually called positive examples, while the other category is called negative examples. Therefore, the classification results of the classifier are as follows:

   TP: The classifier identifies the positive samples as the number of positive examples (True Positive).
   FN: Classifier identifies positive samples as the number of negative examples (False Negative).
   FP: The classifier identifies negative samples as the number of positive examples (False Positive).
   TN: The classifier identifies positive samples as the number of positive examples (True Negative).

   Based on the above classification results, the confusion matrix is defined as Table 9:

   The precision rate measures the proportion of true positive in all predicted positive results, which can be defined as:
   \[ P = \frac{TP}{TP + FP} \]

   The recall rate measures the proportion of true positive examples that can be identified by the classifier in a real case. It can be defined as:
   \[ R = \frac{TP}{TP + FN} \]

   PR curve: A curve with abscissa as the recall rate and ordinate as the curve of the accuracy rate.
   ROC curve: Receiver Operating Characteristic curve, the area under ROC curve area is called AUC (Area Under ROC Curve), and it is better when the classifier’s AUC value is close to 1.

   The features extracted from the feature extraction scheme of MPE-OVMD-CSSVM are analyzed, and the “one-to-remain” multi-classification SVM model is used for learning. The results of random search optimization are as follows:

   \[ C_{\text{best}} = 79.64, \ g_{\text{best}} = 2.45 \]

   According to the optimal SVM classification model obtained by learning, the test set samples are identified, and four types of signals are taken as the positive samples in turn. The remaining signal samples are negative samples, and the ROC curve is shown in Figure 22. The vertical axis TPR of the ROC curve is the recall rate, while the horizontal axis FPR represents the proportion of the positive samples of the classifier that are divided into negative examples.

   Figure 23(b) indicates that Fault III has been eliminated, and the other three kinds of faults can be accurately classified. Figure 21(a) is an overall ROC curve of four states, and Figure 23(b) is the ROC curve of four states, respectively. Figure 23(b) indicates that Fault III has been eliminated, and the other three kinds of faults can be accurately classified.

   Table 10 shows the confusion matrix corresponding to figure 23. It can be seen that the test set consists of 80 samples, and each mechanical state contains 20 samples.
Ordinary samples are misidentified as Fault I type sample, Fault I type signals are identified as normal, and 10 samples of the other two fault types are correctly identified. The final accuracy of the entire test set is $\text{Auc} = 90\%$. The experimental results verify the effectiveness of the cuckoo search optimization and MPE-OVMD feature extraction scheme.

This paper argues that the performance of classifiers should not only focus on accuracy, but also on the diagnosis of mechanical fault. It is often hoped that fault types can be discovered in time to avoid delaying the optimal time of fault diagnosis. Therefore, the recall rate of fault sample identification results is required to be high. According to the confusion matrix shown in Table 10, the performance indicators of the classifier shown in Table 11 are obtained.

There are 60 samples of all faults, 53 of which are identifiable. If all fault samples are taken as positive samples, the recall rate is 90%, and the accuracy rate of the classifier is 90%. The recall rate and the accuracy rate of the classifier are both 90%, indicating that good recognition results are obtained.

### IX. CONCLUSION

In this paper, the vibration signal acquisition platform is designed by virtual instrument technology, and the feature extraction and recognition methods of the mechanical vibration signal of the INV1618 rotary motor bearing are deeply studies. The research mainly involves three parts: signal denoising pretreatment, signal feature extraction and signal recognition. The main conclusions are as follows:

1. In this paper, the positive and negative Gauss white noise signal reconstruction and addition method based on signal denoising preprocessing is adopted. Signal reconstruction denoising method is a new processing method of adding Gauss white noise to the original signal. The simulation results of signal denoising show that the reasonable addition of the Gauss white noise with opposite symbols and the same amplitude can improve the signal-to-noise ratio and reduce the mean square deviation, maintain the smoothness and retain the mutation information of the original signal.

2. In view of the shortcomings of EMD, EEMD and VMD methods, a method based on multi-scale permutation VMD algorithm for decomposing the measured signals is proposed. Through comparative research, the decomposition results of the simulated signal show that the EMD and EEMD methods have severe mode aliasing phenomenon, and the decomposed mode function does not describe the real-time frequency characteristics of the signal well. Therefore, it is not suitable for the study of mechanical fault diagnosis of circuit breakers. The traditional VMD algorithm needs to determine the mode number $K$ by observing the center frequency, but this method will cause errors to some extent and easily lead to modal aliasing. The signal decomposition method of multi-scale sample entropy optimization VMD algorithm proposed in this paper has high time-frequency resolution and can extract the narrowband intrinsic mode functions with practical physical significance. Transform the decomposed signal into a Hilbert transform. The time-frequency analysis results of the measured signals show that the frequency range of the mechanical vibration signals of the rolling bearing enclosing in this paper is mainly distributed below 6 KHz.

3. In view of the small scale of the vibration samples of rotary motors bearing, the uneven distribution of sample fractions and the low recall rate in the case of fault status are the problems. SVM can solve the problem of uneven distribution of samples, and can learn the decision surface which is closely supported. It is good for detecting abnormal samples well and increasing the recall rate of abnormal samples. Therefore, a classification model of CSSVM classifier is proposed, and the signal classification is realized by using sklearn machine learning package. Compared with the grid search, particle swarm optimization and genetic algorithm, the support vector machine classification model based on cuckoo search optimization has faster convergence speed. The recognition result of the classifier is: taking all fault samples as an example, the accuracy rate is 95% while the recall rate is 96.67%. Expect misunderstanding between the outer ring fault states, the vibration signals of other states can be correctly identified, which reduces the probability of identifying weak faults as normal and misses the optimal maintenance opportunities of the equipment.

### REFERENCES

[1] G. Wan and Y.-B. Shaft, “Alignment induced vibration in high capacity steam-turbine generator set,” Poweren Gineering, vol. 24, no. 6, pp. 768–774, 2004.
[2] J. Cheng, D. Yu, J. Tang, and Y. Yang, “Local rub-impact fault diagnosis of the rotor systems based on EMD,” *Mechanism Mach. Theory*, vol. 44, no. 4, pp. 784–791, Apr. 2009.

[3] B. Yang and C. S. Suh, “Interpretation of crack-induced rotor non-linear response using instantaneous frequency,” *Mech. Syst. Signal Process.*, vol. 18, no. 3, pp. 491–513, May 2004.

[4] W. Yujing, J. Yicheng, K. Shouqiang, Y. Guangxue, and C. Yanna, “Diagnosis method of fault location and performance degradation degree of rolling bearing based on optimal ensemble EMD,” *Chin. J. Sci. Instrum.*, vol. 34, no. 8, pp. 1834–1840, Aug. 2013.

[5] Z. H. Wu and N. E. Huang, “Ensemble empirical mode decomposition: A noise-assisted data analysis method,” *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[6] L. Yaguo, “Machinery fault diagnosis based on improved Hilbert-Huang transform,” *J. Mech. Eng.*, vol. 47, no. 5, pp. 71–77, May 2011.

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

[8] X. Zhao, S. Zhang, and L. J. Zhishen, “Application of new denoising method based on VMD in fault feature extraction,” *J. Vib. Meas. Diagnosis*, vol. 38, no. 1, pp. 1–12, 2018, doi: 10.16450/j.cnki.issn.1004-6801.2018.01.002.

[9] A. Xu, J. Yan, and X. Wu, “Rotating machinery fault feature extraction based on VMD and MP algorithm,” *Foreign Electron. Meas. Technol.*, vol. 36, pp. 11–17, 2017, doi: 10.16450/j.cnki.issn.1004-6801.2018.02.

[10] S. Mohanty, K. K. Gupta, and K. S. Raju, “Hurst based vibro-acoustic feature extraction of bearing using EMD and VMD,” *Measurement*, vol. 117, pp. 200–220, Mar. 2018.

[11] L. Chen, C. Xiao, X. Li, Z. Wang, and S. Huo, “A seismic fault recognition method based on ant colony optimization,” *J. Appl. Geophys.*, vol. 152, pp. 1–8, May 2016.

[12] Z. Wang, L. Jia, and Y. Qin, “Adaptive diagnosis for rotating machineries using information geometric kernel ELM based on VMD-SVD,” *Entropy*, vol. 20, no. 1, pp. 73–91, 2018.

[13] Z. Wang, J. Wang, Z. Zhao, and R. Wang, “A novel method for multi-fault feature extraction of a gearbox under strong background noise,” *Entropy*, vol. 20, no. 1, pp. 1–10, Dec. 2017.

[14] Z. Li, J. Chen, Y. Zi, and J. Pan, “Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive,” *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, Feb. 2017.

[15] J. Liying, L. Xiaodong, W. Jin, and C. Jianguo, “Feature analysis method of fault location and performance degradation degree of rolling bearing based on optimal ensemble EMD,” *Chin. J. Sci. Instrum.*, vol. 34, no. 8, pp. 1834–1840, Aug. 2013.

[16] J. D. Zheng, J. S. Cheng, and Y. Yang, “Modified EEMD algorithm and its applications,” *J. Vib. Shock*, vol. 32, no. 21, pp. 21–26, 2013.

[17] Z. H. Wu and N. E. Huang, “Ensemble empirical mode decomposition: A noise-assisted data analysis method,” *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[18] L. Yaguo, “Machinery fault diagnosis based on improved Hilbert-Huang transform,” *J. Mech. Eng.*, vol. 47, no. 5, pp. 71–77, May 2011.

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

[20] X. Zhao, S. Zhang, and L. J. Zhishen, “Application of new denoising method based on VMD in fault feature extraction,” *J. Vib. Meas. Diagnosis*, vol. 38, no. 1, pp. 1–12, 2018, doi: 10.16450/j.cnki.issn.1004-6801.2018.01.002.

[21] A. Xu, J. Yan, and X. Wu, “Rotating machinery fault feature extraction based on VMD and MP algorithm,” *Foreign Electron. Meas. Technol.*, vol. 36, pp. 11–17, 2017, doi: 10.16450/j.cnki.issn.1004-6801.2018.02.

[22] S. Mohanty, K. K. Gupta, and K. S. Raju, “Hurst based vibro-acoustic feature extraction of bearing using EMD and VMD,” *Measurement*, vol. 117, pp. 200–220, Mar. 2018.

[23] L. Chen, C. Xiao, X. Li, Z. Wang, and S. Huo, “A seismic fault recognition method based on ant colony optimization,” *J. Appl. Geophys.*, vol. 152, pp. 1–8, May 2016.

[24] Z. Wang, L. Jia, and Y. Qin, “Adaptive diagnosis for rotating machineries using information geometric kernel ELM based on VMD-SVD,” *Entropy*, vol. 20, no. 1, pp. 73–91, 2018.

[25] Z. Wang, J. Wang, Z. Zhao, and R. Wang, “A novel method for multi-fault feature extraction of a gearbox under strong background noise,” *Entropy*, vol. 20, no. 1, pp. 1–10, Dec. 2017.

[26] Z. Li, J. Chen, Y. Zi, and J. Pan, “Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive,” *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, Feb. 2017.

[27] J. Liying, L. Xiaodong, W. Jin, C. Jianguo, and Y. Mingyue, “Feature parameters extraction method of gear based on PSO-VMD,” *Des. Res.*, vol. 2017, no. 11, pp. 65–71, Nov. 2017.

[28] X. Zhang, Q. Miao, and H. Zhang, “A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery,” *Mech. Syst. Signal Process.*, vol. 108, pp. 58–72, Aug. 2018.

[29] Y. Zhou, F. He, and Y. Qiu, “Dynamic strategy based parallel ant colony optimization on GPUs for TSPs,” *Sci. China Inf. Sci.*, vol. 60, no. 6, Jun. 2017.

[30] W. Shen, X. Guo, C. Wu, and D. Wu, “Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm,” *Knowl.-Based Syst.*, vol. 24, no. 3, pp. 378–385, Apr. 2011.

[31] N. Huang, H. Chen, G. Cai, L. Fang, and Y. Wang, “Mechanical fault diagnosis of high voltage circuit breakers based on variational mode decomposition and multi-layer classifier,” *Sensors*, vol. 16, no. 11, p. 1887, Nov. 2016.

[32] J. Chen and R. J. Patton, *Robust Model-based Fault Diagnosis for Dynamic Systems*. New York, NY, USA: Springer, 1999.

[33] P. K. Kankar, S. C. Sharma, and S. P. Harsha, “Fault diagnosis of ball bearings using machine learning methods,” *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1876–1886, Mar. 2011.

[34] W. He, M. Kong, and H. Zhao, “Research on meteorological prediction with Bayesian classifier,” *Comput. Eng. Des.*, vol. 28, no. 15, pp. 3783–3782, Jan. 2007.

[35] K. G. Narendra, V. K. Sood, K. Khorasani, and R. Patel, “Application of a radial basis function (RBF) neural network for fault diagnosis in a HVDC system,” *IEEE Trans. Power Syst.*, vol. 13, no. 1, pp. 177–183, Feb. 1998.

[36] J. Tao, Y. Liu, D. Yang, and G. Bin, “Rolling bearing fault diagnosis based on bacterial foraging algorithm and deep belief network,” *J. Vib. Shock*, vol. 36, no. 23, pp. 68–74.1, 2017.

[37] H. Guo, C. Yuliang, W. Xiaochuan, M. Tingfeng, and S. Yongsheng, “Characteristic analysis of cavitation pressure fluctuation in centrifugal pump,” *J. Wuhan Univ. Technol.*, vol. 41, no. 4, pp. 549–553, Apr. 2017.

Z. Guo et al.: New Fault Diagnosis Classifier for Rolling Bearing United MPE
MINGLIANG LIU was born in Zhaoyuan, Heilongjiang, China, in 1980. He received the Ph.D. degree in forestry engineering automation from Northeast Forestry University, in 2017. He is currently a Full Professor and a master’s Tutor with the School of Electrical Engineering, Heilongjiang University. His research interests include intelligent detection, fault diagnosis, and signal processing.

YUNXIA WANG was born in Shandong, China, in 1978. She received the M.S. degree in signal and information processing from Heilongjiang University, in 2008. She is currently with the College of Electronic Engineering, Heilongjiang University, as a Lecturer. Her research interests include intelligent detection, fault diagnosis, signal processing, and so on.

HUABIN QIN graduated from the Qilu University of Technology. He is currently pursuing the degree with Heilongjiang University. His research interests include deep learning, fault diagnosis, machinery health management, and multiobjective optimization algorithm.

***