ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

Multilevel Dimensionality Reduction Combined with IBFO-Optimized VMD-SVM for Bearing Fault Classification

Siwen Qi

Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao, Liaoning, China

Abstract: *To address the challenges of excessive fault features and low identification accuracy in motor bearings, this paper proposes a bearing fault identification method integrating multilevel dimensionality reduction with Improved Bitterling Fish Optimization (IBFO)-optimized Variational Mode Decomposition (VMD) and Support Vector Machine (SVM). Specifically, Tent chaotic mapping is introduced to optimize BFO's initial population and mitigate local optima entrapment, while convergence parameters are refined to accelerate convergence and enhance robustness, and Cauchy variation is incorporated to improve local search capability. For VMD requiring tuning of decomposition layer count K and penalty coefficient α, IBFO optimizes these parameters to enhance decomposition performance, with features from optimal components forming eigenvectors subjected to multilevel dimensionality reduction. Furthermore, IBFO optimizes penalty factor C and kernel parameter g of SVM to boost recognition accuracy. Experimental results demonstrate 7.29% accuracy improvement through algorithmic parameter optimization and an additional 9.369% gain via multilevel dimensionality reduction, with simulations confirming its efficacy in enhancing bearing fault identification accuracy.*

Keywords: *Bearings; Bitterling Fish Optimization; Multilevel Dimensionality Reduction; Fault Classification*

1. Introduction

Electric motors serve as driving units for equipment and play critical roles across various industries[1]. Under continuous operation in harsh environments, even minor faults may lead to severe consequences. Therefore, real-time monitoring of motor conditions has become a research priority. Timely fault detection/diagnosis and preventive maintenance can reduce economic losses caused by motor failures and mitigate their severe impacts on industrial production[2][3]. Rolling bearings undergo continuous rotation during operation and are prone to damage, accounting for 45%–55% of rotating machinery failures[4]. Thus, the operational condition of rolling bearings directly affects motor safety and reliability[5]. Since bearing vibration signals enable early-stage anomaly detection and capture comprehensive low-to-high-frequency information, vibration analysis is essential for rolling bearing fault diagnosis.

Early-stage faults typically exhibit weak vibration characteristics that are easily masked by noise. Such noise interferes with fault feature extraction, leading to misdiagnosis or missed detection. Extracting fault features while suppressing noise remains a key research challenge[6]. Researchers have developed vibration signal feature extraction methods with notable achievements: huang et al.[7] proposed Empirical Mode Decomposition (EMD), but this method suffers from end effects and mode mixing that compromise feature extraction. Wu Z.H. et al.[8] developed Ensemble Empirical Mode Decomposition (EEMD) by combining EMD with white noise statistical characteristics. While EEMD addresses some EMD limitations, the added noise increases computational load. Cai Xinyi et al.[9] introduced an improved composite interpolation envelope EMD method, but it lacks subsequent data processing and classification. Gilles J. et al.[10] proposed Empirical Wavelet Transform (EWT) to reduce computational complexity and accelerate decomposition. However, EWT may partition spectra suboptimally or excessively[11]. Dragomiretskiy et al.[12] established Variational Mode Decomposition (VMD) as a fully non-recursive model with significantly stronger robustness to sampling and noise. Nevertheless, VMD requires predefined parameters, and improper selection may cause over-decomposition or under-decomposition[13]. Shan Yuting et al.[14] optimized VMD

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

parameters using genetic algorithms but obtained suboptimal solutions due to weak local search capability despite strong global exploration. Liu Qiang et al.[15] optimized decomposition level K through energy ratio analysis but did not optimize the penalty factor α.

Additionally, feature processing and fault identification remain necessary after VMD decomposition. Lei Chunli et al.[4] applied Convolutional Neural Networks (CNN) without feature preprocessing. Yatsugi Kenichi et al.[16] diagnosed motor faults via Support Vector Machine (SVM) using stator current features, but overlapping features reduced accuracy. Liu Xinya et al.[17] employed Principal Component Analysis (PCA) for dimensionality reduction, but its linear nature cannot handle nonlinear data structures. Liu Yunhang et al.[18] developed a Center Modified Projection (CMP) method combined with Improved Grey Wolf Optimization (IGWO)-SVM for bearing fault classification, optimizing SVM parameters and applying CMP dimensionality reduction, though VMD parameters were not optimized. Lida Zareian et al.[19] demonstrated through benchmark testing that Bitterling Fish Optimization (BFO) outperforms Grey Wolf Optimization and Whale Optimization algorithms.

These studies confirm VMD's effectiveness for vibration signal decomposition/noise suppression and SVM's utility for bearing fault identification, yet both require parameter optimization. Moreover, nonlinear dimensionality reduction remains underexplored. Therefore, this paper proposes a bearing fault identification method combining multilevel dimensionality reduction with Improved BFO (IBFO)-optimized VMD-SVM. First, IBFO optimizes VMD decomposition level K and penalty parameter α to enhance decomposition performance. Second, vibration signals are decomposed using optimized VMD to extract effective components. Subsequently, multilevel dimensionality reduction is applied. Finally, IBFO optimizes SVM penalty factor C and kernel parameter g, with experimental validation. Results demonstrate the method's efficacy in improving rolling bearing fault diagnosis accuracy.

2. Theoretical Foundation

2.1 Variational Mode Decomposition

Variational Mode Decomposition (VMD) is an adaptive signal processing method that decomposes complex signals into several Intrinsic Mode Functions (IMFs). Each IMF represents a specific frequency component of the signal, enabling independent analysis of different frequency characteristics. The fundamental principles of VMD are as follows:

The objective of VMD is to decompose the input signal $f(t)$ into a sum of modal functions $u_k(t)$, each possessing a specific center frequency ω_k.

Each modal function $u_k(t)$ is obtained by solving the optimization problem in Equation (1):

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

α_t denotes the temporal gradient, $\delta(t)$ is the Dirac delta function, $*$ represents the convolution operation, and j is the imaginary unit.

The sum of all modal functions should approximate the original signal:

$$\sum_{k=1}^{K}u_k(t)=f(t) \tag{2}$$

K is the decomposition layer index, k is the current decomposition layer index, and f(t) is the input signal.

To solve the optimization problem in Equation (1), Lagrange multipliers and a penalty term are introduced to construct the augmented Lagrange function:

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha\sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e_2^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{3}$$

α is a penalty parameter and λ(t) is the Lagrange multiplier.

By solving the Lagrangian function in Equation (3), $u_k(t), \omega_k$, and are iteratively updated until convergence is achieved. The resulting modal functions $u_k(t)$ represent distinct frequency components of the signal, with each corresponding to a center frequency ω_k. This analysis confirms that the decomposition level K and penalty coefficient α constitute critical parameters in VMD.

2.2 Support Vector Machine

Support Vector Machine (SVM) is a supervised learning model for classification and regression analysis. In practical applications, it trains on collected bearing vibration data to find optimal decision boundaries separating different classes. The fundamental principles for linearly inseparable cases are as follows:

To handle linearly inseparable data, slack variables ξ_i and penalty factor are introduced, transforming the optimization problem into:

$$\min L(\omega) = \frac{1}{2}\omega^2 + C\sum_{i=1}^{n}\xi_i \tag{4}$$

$$y_i(\omega \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \tag{5}$$

The Lagrange function for Equation (4) is constructed as:

$$L(\omega,b,\xi,\alpha,\beta) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left[y_i(\omega \cdot x_i + b) - 1 + \xi_i\right] - \sum_{i=1}^{n}\beta_i\xi_i \tag{6}$$

According to the strong duality of convex optimization, it is transformed into a dual problem:

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) \tag{7}$$

$$\sum_{i=1}^{n}\alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq C \ \ \forall i \tag{8}$$

For nonlinear separable data, kernel function $K(x_i,x_j)$ replaces the dot product yielding the dual problem:

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j) \tag{9}$$

The Gaussian radial basis function (RBF) kernel is defined as:

$$K\left(x_i, x_j\right) = \exp\left(\frac{-\left\|x_i - x_j\right\|^2}{2g^2}\right) \tag{10}$$

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n}\alpha_i y_i K(x_i,x) + b\right) \tag{11}$$

These steps demonstrate SVM's capability to handle both linear and nonlinear classification problems. The core lies in determining appropriate penalty factor C and kernel parameter g, where optimal parameter selection enhances classification performance and recognition accuracy.

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

2.3 Bitterling Fish Optimization

Bitterling Fish Optimization (BFO) is a swarm intelligence optimization algorithm inspired by the behavior of bitterling fish, mimicking their collective intelligence in foraging, predator avoidance, and social interactions.

Initial population is created:

$$F_i^j = l + (u - l).r \tag{12}$$

The initial population is generated within a specified interval [*l*,*u*], where each dimension of the solution is determined by a random number r between 0 and 1.

Each fish (solution) is evaluated using an objective function f.

$$Fittness = f(F_i^j) \tag{13}$$

In BFO, each fish explores the problem space to locate better oyster positions. Each fish operates autonomously, roaming the environment in search of oysters. When a fish identifies a target oyster, it moves toward it without interference from other fish. This behavioral pattern enables each fish to independently select and pursue the most suitable oyster. The oyster occupation status is represented as follows:

$$F_i^{t+1} = \begin{cases} J.F_i^t + \left(F^+ - J.F_i^t\right).\delta & r \le P \\ J.F_i^t + \left(F^* - J.F_i^t\right).\delta & r > P \end{cases} \tag{14}$$

The i-th solution's current position at time t and new position at t+1, respectively, is the optimal solution and a randomly selected oyster from the population. δ and r are random numbers between 0 and 1, and J represents the step size or the rate at which fish move to evade or approach oysters.

The decrease in parameter J over time can be interpreted as a natural phenomenon where increased successful mating among male fish leads to a gradual reduction in their activity levels. This reduced activity may occur because they no longer need to search for oysters as frequently, having already found mates. Thus, as J decreases, the fish tend to focus more on local search. To reduce parameter J, the following formula is used:

$$J(t+1) = \left(J(1) - \frac{J(1).t}{Maxt}\right).\cos(t \times \cos^{-1}(U(t))) \tag{15}$$

J(1) is the initial step size and jump value for each fish; t is the current iteration count; Maxt is the maximum number of iterations; U(t) is a generated random sequence.

Over time, to enhance exploration around the optimal solution, parameter P is gradually decreased to meet the condition r > P in Equation (14), expressed as follows:

$$P = \left|1 - \frac{t}{\sqrt{1+t^2}}\right| + \frac{rand}{t^a} \tag{16}$$

Parameter a is a predefined exponent.

Subsequently, the position updates for fish in escape, failure to capture an oyster, and reproduction behaviors correspond to Equations (17) and (18) as follows:

$$F_i^{t+1} = \begin{cases} J.F_i^t + (F^* - J.M).\delta & r \le 0.5 \\ l + (u - l).\delta & r > 0.5 \end{cases} \tag{17}$$

$$F_i^{t+1} = F_i^t + R * rand(0,1) \tag{18}$$

R is the distribution radius around the oyster, which decreases over iterations, and M represents the average position of the fish swarm, calculated using Equation (19):

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

$$M = \frac{\sum_{i=1}^{n} F_i^t}{n} \tag{19}$$

Based on the fishing concept, the probability of losing a solution is given by:

$$d(F_i^t) = \frac{f(F_i^t)}{\sum_{i=1}^{n} f(F_i^t)} \tag{20}$$

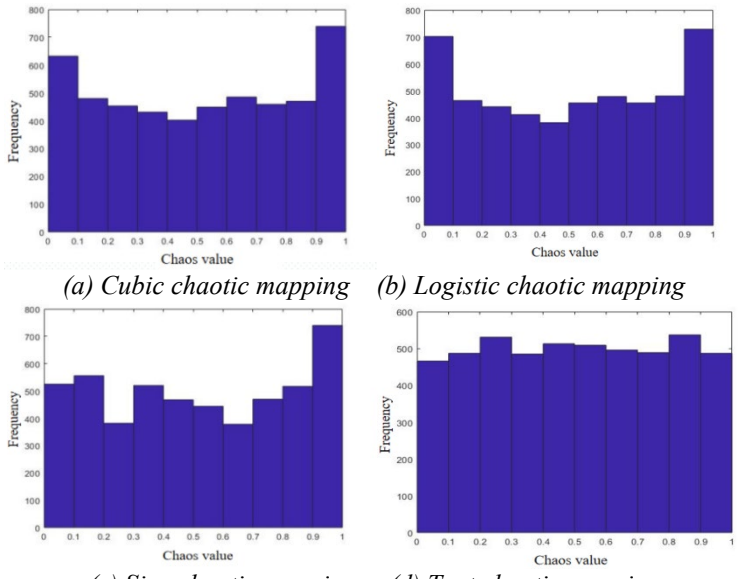
By simulating bitterling fish behavior and employing probabilistic decision-making and iterative updates, the BFO algorithm combines global and local search in the solution space to ultimately find the optimal solution. It underscores the importance of initial population setup, adaptive step size, and strategies to avoid local optima.

3. Parameter Optimization and Feature Dimensionality Reduction

3.1 Improved Bitterling Fish Optimization Algorithm

3.1.1 Initial Population Optimization

The original BFO algorithm randomly generates an initial population of bitterling fish, which then searches the problem space for suitable mating oysters. If the initial population distribution is uneven, it can easily lead to local optima and reduce global search capability. Since chaotic mapping performs better than pseudo-random number generation, this study employs chaotic mapping to optimize the initial population. Commonly used chaotic sequence models include Cubic, Logistic, Sine, and Tent mappings, as shown in Figure 1. Among these, the Tent mapping produces more uniformly distributed sequences compared to others. This uniform distribution enhances population diversity, achieves better optimization performance with less computational time, avoids local optima, and improves the algorithm's ability to find global optimal solutions.



(a) Cubic chaotic mapping *(b) Logistic chaotic mapping*

(c) Sine chaotic mapping *(d) Tent chaotic mapping*

Figure 1 Distribution of chaotic sequences

3.1.2 Convergence Factor Optimization

In the BFO algorithm, the convergence factor exhibits significant fluctuations in early iterations (Figure 2), which may destabilize the algorithm and affect convergence stability. In mid-to-late iterations, it decreases linearly at a slow rate, requiring more iterations to approach the optimal solution. To address this, this study optimizes the convergence factor by making it follow an exponential decay pattern.

In the early stage, the optimized convergence factor remains relatively stable, helping the algorithm find optimal solutions more steadily and reducing initial oscillations. As iterations progress, the

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

convergence factor gradually decreases and approaches zero more smoothly, maintaining smaller step sizes in later stages. This prevents premature convergence or entrapment in local optima, significantly enhancing the algorithm's local search capability. The optimized position update formula is given in Equation (21):

$$\omega = e^{-(3t/T)^2} \tag{21}$$

Among them, ω is the convergence factor value, t is the current iteration count, T is the total number of iterations. Equation (15) is replaced by Equation (21).

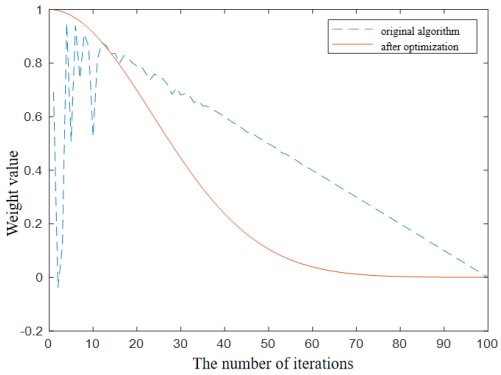


Figure 2 Adaptive weighting curves

3.1.3 Cauchy Mutation

Optimization problems aim to find the optimal solution for a given objective function. Many optimization algorithms tend to get trapped in local optima or converge prematurely. The Cauchy distribution, a heavy-tailed distribution, enables large jumps in the search space. By applying the Cauchy mutation operator to perturb the current best solution, new candidate solutions can be generated, mitigating the risk of local optima.

This study integrates the Cauchy mutation strategy into the BFO algorithm. The mathematical expression is as follows:

$$x_{newbest} = x_{best} + x_{best} \times \text{Cauchy}(0,1) \tag{22}$$

$x_{newbest}$ is the updated optimal solution, x_{best} is the current best solution, Cauchy(0,1) is the Cauchy factor.

3.2 IBFO-Optimized VMD and SVM

When applying VMD for fault signal decomposition, the decomposition results are significantly influenced by the number of modes K. An inappropriate K value may lead to unsatisfactory diagnostic outcomes. Additionally, the penalty factor α affects the smoothness of data and the sparsity of IMFs, necessitating proper adjustment of this parameter to achieve better decomposition performance. In SVM, the penalty factor C serves as a regularization term that balances fitting training data and preventing overfitting. Meanwhile, the kernel parameter g influences data partitioning and distribution. Therefore, selecting appropriate C and g parameters is crucial for improving SVM classification performance.

To optimize VMD parameters (K and α) and SVM parameters (C and g), this study employs the Improved Bitterling Fish Optimization (IBFO) algorithm. The IBFO-based parameter optimization can effectively enhance the overall performance of fault signal decomposition and classification, thereby improving the accuracy and reliability of fault diagnosis. The specific implementation steps are as follows:

Step 1: Initialize IBFO parameters, including parameter dimensions, value ranges, population size, and maximum iteration count.

Step 2: Generate initial values for K, α, C, and g using chaotic mapping.

Step 3: Perform VMD decomposition on the signal and conduct SVM training with cross-validation.

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

Step 4: Calculate the fitness value of decomposed components using cross-validation accuracy as the fitness metric. Update the optimal parameters if a better solution is found.

Step 5: Terminate the process if reaching maximum iterations; otherwise, return to Step 2.

3.3 Multi-Level Dimensionality Reduction

As VMD-extracted optimal components often contain high-dimensional features that increase data complexity, a two-stage dimensionality reduction approach is adopted. First, Principal Component Analysis (PCA) is applied for initial dimensionality reduction by projecting data onto directions with maximum variance, preserving essential information while removing redundancy and noise. Subsequently, Uniform Manifold Approximation and Projection (UMAP) is employed to further reduce dimensions to 2D for visualization, capturing complex nonlinear relationships and detailed features [20].

Stage 1: Preliminary Dimensionality Reduction via PCA.

The data samples are centered ($\sum_i x_i=0$). A new orthogonal coordinate system is established, where each ω_i is an orthonormal basis vector ($\|\omega_i\|_2=1$, $\omega_i^T\omega_j=0$). By discarding less significant coordinates and reducing dimensions to $d'<d$, the projection of sample x_i in the low-dimensional space becomes $z_i=(z_{i1};z_{i2};\cdots;z_{id'})$, where $z_{ij}=\omega_j^T x_i$ represents the j-th coordinate of x_i. Reconstructing from yields $\hat{x}_i=\sum_{j=1}^{d'} z_{ij}\,\omega_j=Wz_i$, where W is the orthonormal basis matrix. For the entire training set, the objective is to minimize the distance between all samples and the hyperplane, as expressed in Equation (23).

$$\sum_{i=1}^{m}\| \hat{x}_i - x_i \|_2^2 \tag{23}$$

By rearranging Equation (23), we obtain:

$$\sum_{i=1}^{m}\| \hat{x}_i - x_i \|_2^2 = -\mathrm{tr}\left(\mathbf{W}^{\mathrm{T}}\left(\sum_{i=1}^{m}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\right)\mathbf{W}\right)+\sum_{i=1}^{m}x_i^T x_i \tag{24}$$

$\sum_i^m x_i x_i^T$ is the covariance matrix of the dataset, and $\sum_i^m x_i^T x_i$ is a constant. Minimizing Equation (25) is equivalent to:

$$argmin - \mathrm{tr}\left(\mathbf{W}^{\mathrm{T}}\left(\sum_{i=1}^{m}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\right)\mathbf{W}\right) \tag{25}$$

$$\mathrm{s.t.}\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I} \tag{26}$$

Using the Lagrange function and taking the derivative with respect to W yields:

$$\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{W}=\lambda\mathbf{W} \tag{27}$$

Thus, we only need to perform eigenvalue decomposition on the covariance matrix. The obtained eigenvalues are sorted as $\lambda_1\geq\lambda_2\geq\cdots\lambda_d$. Then, by setting a reconstruction threshold β, we select the minimum value that satisfies Equation (28):

$$\frac{\sum_{i=1}^{d'}\lambda_i}{\sum_{i=1}^{d}\lambda_i}\geqslant t \tag{28}$$

The corresponding eigenvectors form $W=(\omega_1,\omega_2,\cdots\omega_{d'})$, which constitutes the solution of principal component analysis.

Stage 2: Final Dimensionality Reduction via UMAP.

Define the metric space d as $X\times X\in R\geq 0$. Given the input hyperparameter k, for each ω_i, use the nearest neighbor descent algorithm to compute the k-nearest neighbor matrix for the set $\{\omega_{i1},\omega_{i2}\cdots,\omega_{ik}\}$. The distance is determined to the first nearest neighbor sample and the normalization factor σ_i:

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

$$\rho_i = \min\{d(\omega_i, \omega_{i_j}) \mid 1 \le j \le k, d(\omega_i, \omega_{i_j}) > 0\} \tag{29}$$

$$\sum_{j=1}^{k} \exp\left(\frac{-max\left(0, d\left(\omega_i, \omega_{i_j}\right) - \rho_i\right)}{\sigma_i}\right) = \log_2(k) \tag{30}$$

In the high-dimensional space, the relationship between the initial point and other points can be represented by Equation (31):

$$W(\omega_i, \omega_{i_j}) = \exp\left(\frac{-\max(0, d(\omega_i, \omega_{i_j}) - \rho_i)}{\sigma_i}\right) \tag{31}$$

$$W_{ij} = W_{i|j} + W_{j|i} - W_{i|j} W_{j|i} \tag{32}$$

Among them, W_{ij} represents the overall relationship, $W_{i|j}$ describes the relationship between initial point and other points j, and $W_{j|i}$ describes the relationship between initial point j and other points i. Equation (32) combines these into a unified topological representation while ensuring symmetry.

In the low-dimensional space, the relationships between points are described by Equation (33):

$$V_{ij} = \left[1 + a\left(y_i - y_j\right)^{2b}\right]^{-1} \tag{33}$$

Among them, hyperparameters a and b can adjust the clustering behavior of the mapped low-dimensional data.

$$V_{ij}^{*} = \begin{cases} 1 & |y_i - y_j| \le \min\_dist \\ e^{-(y_i - y_j) - \min\_dist} & |y_i - y_j| > \min\_dist \end{cases} \tag{34}$$

In practical applications, Equation (34) is curve-fitted based on the set hyperparameter min_dist.

The low-dimensional embedding is optimized by minimizing the difference between the high-dimensional and low-dimensional graphs. This optimization problem can be formulated using the cross-entropy loss function:

$$\begin{aligned} &CE(\omega, y) = \\ &\sum_{i \ne j} W_{ij}(\omega) \log_2 W_{ij}(\omega) \\ &-W_{ij}(\omega) \log_2 V_{ij}(y) \\ &+(1 - W_{ij}(\omega)) \log_2 (1 - W_{ij}(\omega)) \\ &-(1 - W_{ij}(\omega)) \log_2 (1 - V_{ij}(y)) \end{aligned} \tag{35}$$

UMAP employs stochastic gradient descent to minimize the aforementioned loss function, adjusting the coordinates and in the low-dimensional space to make the structure of the low-dimensional graph as close as possible to that of the high-dimensional graph.

3.4 Diagnostic Procedure

First, the collected vibration signals are processed and decomposed into several modal components using VMD. Next, the optimal components are extracted from each modal component through envelope entropy, and their feature values are extracted and subjected to dimensionality reduction. Finally, IBFO is used to optimize SVM parameters to train the fault diagnosis model, achieving identification and diagnosis of bearing faults. The diagnostic flowchart is shown in Figure 3.

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

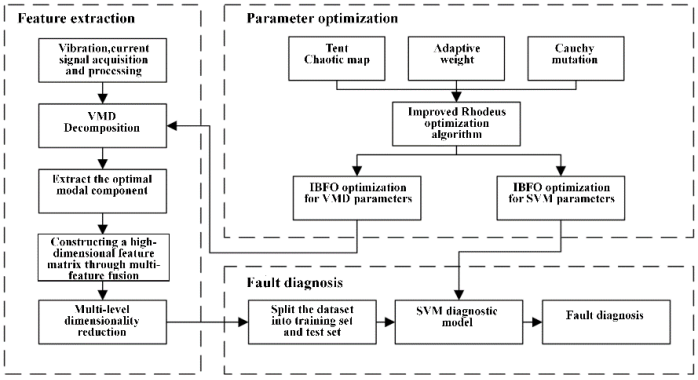
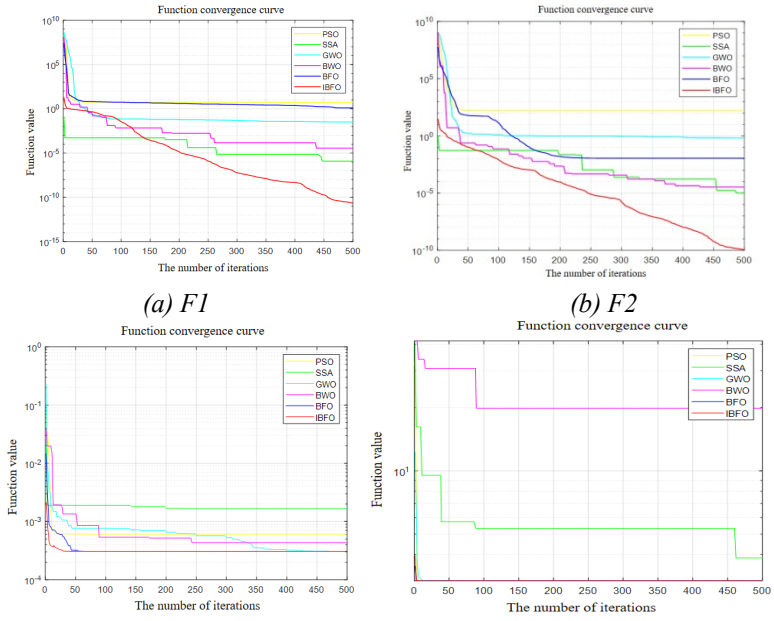


Figure 3 Diagnostic process

4. Experimental Analysis

4.1 Algorithm Comparison Test

This paper selects six algorithms for comparison experiments: Grey Wolf Optimization (GWO), Particle Swarm Optimization (PSO), Sparrow Search Algorithm (SSA), Beluga Whale Optimization (BWO), BFO, and IBFO. The population size is set to 50, and the maximum number of iterations is 500. Multimodal (F1, F2) and fixed-dimensional multimodal (F3, F4) test functions are selected to evaluate the global search performance and local search performance of the algorithms. Their expressions are shown in Table 1, and the test results are shown in Figure 4.



(a) F1 *(b) F2*

(c) F3 *(d) F4*

Figure 4 Test results

As can be seen from Figure 4, in F1 and F2, SSA has faster convergence speed with optimization accuracy of 10^(-5), while other algorithms have relatively slower convergence speed. However, IBFO's optimization accuracy improves as the number of iterations increases. In F3 and F4, IBFO achieves the fastest convergence speed when reaching the optimum compared to other algorithms, with higher convergence accuracy. By balancing global and local search, the algorithm not only improves convergence speed but also enhances convergence accuracy, proving the effectiveness of the improvement strategy and providing a solid foundation for using IBFO for parameter optimization.

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

4.2 Dimensionality Reduction Performance Analysis

Table 1: Test Functions

Name	Function	Dimension	Optimum
F1	$f_1(x)=\frac{\pi}{n}\{10\sin(\pi y_1)+\sum_{i=1}^{n-1}(y_i-1)^2[1+10\sin^2(\pi y_{i+1})]$ $+(y_n-1)^2+\sum_{i=1}^{n}u(x_i,10,100,4)$	30	0
F2	$f_2(x)=0.1\{\sin^2(3\pi x_1)+\sum_{i=1}^{n}(x_i-1)^2[1+\sin^2(3\pi x_i+1)]$ $+(x_n-1)^2[1+sin^2(2\pi x_n)]\}+\sum_{i=1}^{n}u(x_i,5,100,4)$	30	0
F3	$f_3(x)=\sum_{i=1}^{11}[a_i-\frac{x_1(b_i^2+b_ix_2)}{b_i^2+b_ix_3+x_4}]^2$	4	0
F4	$f_4(x)=[1+(x_1+x_2+1)^2*(19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2)]$ $*[30+(2x_1-3x_2)^2*(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2)]$	2	3

This experiment utilized the publicly available Case Western Reserve University (CWRU) dataset, comprising four groups of bearing drive-end vibration signals. The experimental data specifications are presented in Table 2. For each fault type, 120 samples were prepared, with each sample containing 1,000 sampling points. For every sample, features listed in Table 3 were extracted through IBFO-optimized VMD parameters, subsequently forming a feature matrix.

Table 2: Case Western Reserve University Rolling Bearing Test Data Sheet

Label	Motor speed	Fault diameter	Fault category	Data
1	1797		Normal	97.mat
2	1797	0.007	Inner ring	103.mat
3	1797	0.007	Rolling elements	118.mat
4	1797	0.007	Outer ring	130.mat

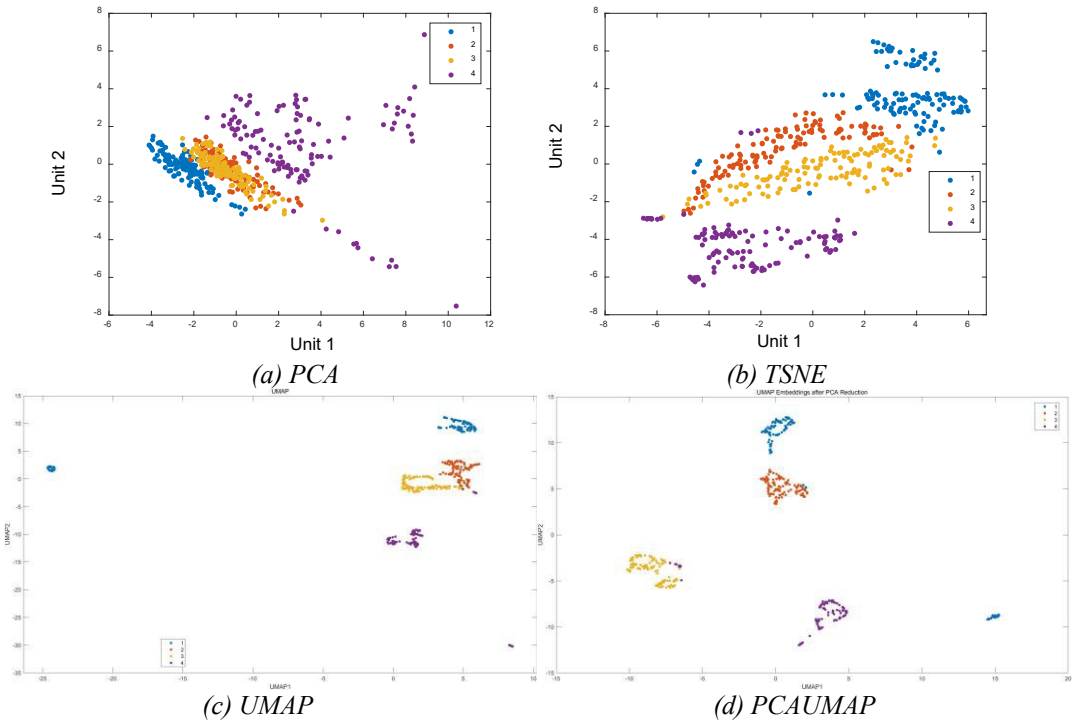
Table 3: Characterization parameter table

Number	Features	Number	Features
1	Min	6	Kurtosis
2	Max	7	Root mean square
3	Average	8	Crest factor
4	Variance	9	Pulse factor
5	Peak-to-peak	10	Margin factor

The high-dimensional feature matrix underwent dimensionality reduction using PCA, t-SNE, UMAP, and PCAUMAP respectively. For t-SNE, the algorithm was configured with exact computation mode and Mahalanobis distance metric. Both UMAP and PCAUMAP were set with a target embedding closeness of 0.1 and Mahalanobis distance metric. All four methods generated 2-dimensional visualizations, as shown in Figure 5.

As shown in Figure 5(a), PCA exhibits severe aliasing between categories 2 and 3 with poor separation. Category 1 shows compact clustering while category 4 appears scattered. Figure 5(b) demonstrates that t-SNE achieves better separation than PCA, though minor aliasing persists between categories 2 and 3, with clusters being relatively dispersed. In Figure 5(c), UMAP yields superior clustering for categories 1 and 4, but category 2 and 3 clusters show partial aliasing with indistinct boundaries. Figure 5(d) reveals that PCAUMAP delivers optimal clustering results: only two instances of category 1 are misclassified into category 2, and six instances of category 4 appear in category 3. Other categories show significant inter-cluster distances with clear separation, indicating effective preservation of global information and extraction of high-dimensional nonlinear features while maintaining excellent clustering performance.

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403



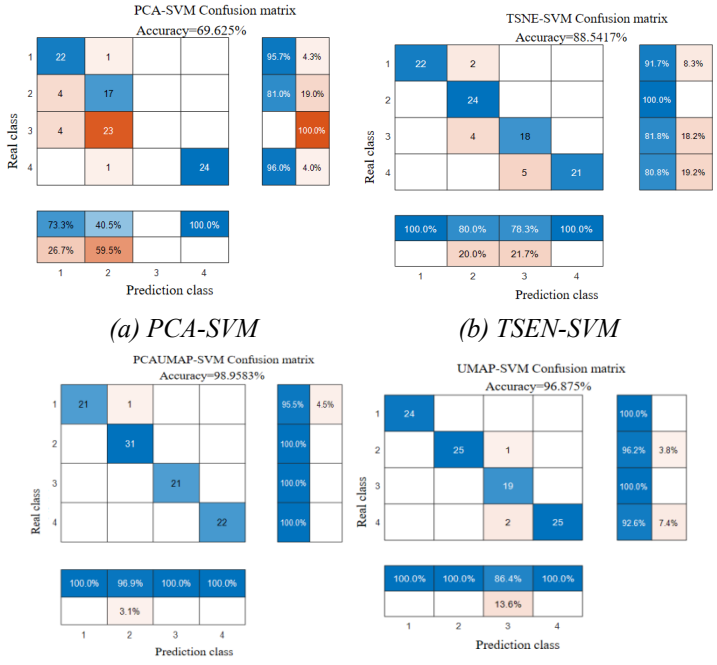
(a) PCA *(b) TSNE*

(c) UMAP *(d) PCAUMAP*

Figure 5 Results of dimensionality reduction

4.3 Dimensionality Reduction Comparison Experiment

Experimental data from Table 2 were processed through IBFO-optimized VMD parameters to extract Table 3 features, forming a feature matrix. After dimensionality reduction, samples were randomly split (80% training, 20% testing) for SVM model training and fault classification. Classification results are shown in Figure 6.



(a) PCA-SVM *(b) TSEN-SVM*

(c) UMAP-SAM *(d) PCAUMAP-SVM*

Figure 6 Classification results of confusion matrix

Figure 6(d) indicates that PCAUMAP-based classification achieves 98.9583% accuracy, with only one instance of true category 1 misclassified as category 2. This represents a 2.0833% absolute

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

improvement over UMAP and 33.3333% over PCA, confirming that PCA-to-UMAP sequential reduction better preserves high-dimensional information and enhances inter-class discriminability.

4.4 Fault Recognition Accuracy Comparison

To validate the proposed diagnostic model, Table 4 compares accuracy across methods with/without parameter optimization and dimensionality reduction:

Table 4: Fault Recognition Accuracy of Different Methods

Number	Method	Accuracy%
1	VMD-SVM	73.9583%
2	BFO-VMD-SVM	82.2917%
3	IBFO-VMD-SVM	89.5833%
4	IBFO-PCAUMAP-VMD-SVM	98.9523%

As evidenced by Table 4, Group 1 exhibits the lowest accuracy (73.9583%), indicating that using VMD and SVM alone yields limited feature extraction and classification performance. Group 2 achieves improved accuracy (82.2917%) through BFO-optimized VMD and SVM parameters, confirming that optimized VMD more effectively extracts discriminative features and enhances SVM performance. Group 3 demonstrates IBFO's superior optimization capability compared to conventional BFO. Group 4 shows a 9.369% absolute accuracy gain over Group 3, proving that integrating multilevel dimensionality reduction with optimization algorithms significantly enhances classifier performance and achieves high recognition accuracy.

5. Conclusions

This paper proposes a bearing fault identification method integrating PCAUMAP with IBFO-optimized VMD-SVM parameters to address low modal recognition accuracy and global feature loss. Key conclusions: (1) IBFO incorporating Tent mapping, convergence factor optimization, and Cauchy mutation effectively optimizes VMD/SVM parameters. This enhances modal extraction and increases recognition accuracy by 7.29%. (2) For high-dimensional features, PCA retains maximum-variance components to reduce noise, while UMAP preserves local neighborhoods and global structures. Sequential PCA-to-UMAP reduction achieves efficient dimensionality reduction and improves accuracy by 9.369%. (3) The integrated PCA-UMAP-IBFO-VMD-SVM methodology processes high-dimensional vibration signals effectively, achieving 98.95% accuracy and enhanced robustness in fault diagnosis.

References

[1] ZHANG Hui, GEO Baojun, HAN Bin, et al. Fault diagnosis method of motor bearing based on GAF-Caps Net[J]. Journal of Electrotechnology,2023,38(10):2675-2685.
[2] XIA Zhiling, HU Kaibo, LIU Xinyue, et al. Fault diagnosis of asynchronous motor rotor broken bars based on variable modal decomposition[J]. Journal of Electrotechnology,2023,38(08):2048-2059.
[3] K. An et al. Edge Solution for Real-Time Motor Fault Diagnosis Based on Efficient Convolutional Neural Network[J]. IEEE Transactions on Instrumentation and Measurement, 2023,72:1-12.
[4] SONG Xiangjin, SUN Wenju, LIU Guohai, et al. Deep subdomain adaptive network electric machine rolling bearing fault diagnosis across operating conditions[J]. Journal of Electrotechnology, 2024, 39(01):182-193.
[5] TIAN Jing, WANG Yingjie, WANG Zhi, et al. Rolling bearing fault diagnosis method based on EEMD and airspace correlation noise reduction[J]. Journal of Instrumentation, 2018, 39(7):144-151.
[6] Shen Zhe. Vibration noise characteristics and optimization design of underwater vehicle stern structure[D]. Harbin: Harbin Engineering University, 2018.
[7] HUANG N E, SHEN Z,LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for non-linear and nonstationary time series analysis[J]. Proceedings of the Royal Society London, 1998,454:903-995.
[8] WU Z H, HUANG N E. Ensemble empirical mode composition: A noise-assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1):1-41.
[9] X.I. Cai, J. Ma, X. Li. Improved composite interpolation envelope empirical modal decomposition for rolling bearing fault feature extraction[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(01): 191-203.

ISSN 2616-5767 Vol.8, Issue 4: 18-30, DOI: 10.25236/AJETS.2025.080403

[10] GILLES J. Empirical wavelet transform[J]. IEEE Transactions on Signal Processing, 2013, 61(16): 3999-4010.
[11] QIAO Zhicheng, LIU Yongqiang, LIAO Yingying. Application of improved empirical wavelet transform and minimum entropy deconvolution in railroad bearing fault diagnosis[J]. Vibration and Shock, 2021, 40(02): 81-90, 118.
[12] Dragomiretskiy Konstantin, Zosso Dominique. Variational Mode Decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62.
[13] MIAO Y H, ZHAO M, LIN J. Identification of mechanical compound fault based on the improved parameter adaptive variational mode decomposition [J]. ISA Transactions, 2019, 84: 82-95.
[14] SAN Yuting, LIU Tao, CHU Wei, et al. Application of genetic algorithm optimized variational modal decomposition in bearing fault feature extraction[J]. Noise and Vibration Control, 2024, 44(01): 148-153, 204.
[15] LIU Qiang, ZHAO Rongzhen, YANG Zeben. Research on rolling bearing fault identification method by K-VMD fusion of envelope entropy and SVM[J]. Noise and Vibration Control, 2022, 42(03): 92-97, 121.
[16] Yatsugi Kenichi, Pandarakone Shrinathan-Esakimuthu, Mizuno Yukio. Common Diagnosis Approach to Three-Class Induction Motor Faults Using Stator Current Feature and Support Vector Machine[J]. IEEE Access, 2023, 11: 24945-24952.
[17] LIU Xinya, MA Chao, HUANG Min, et al. Research on bearing fault diagnosis method based on angle resampling and PCA-XGBoost under variable rotational speed[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(03): 45-54.
[18] LIU Yunhang, SONG Yubo, ZHU Dapeng. Center-corrected projection combined with IGWO-SVM for rolling bearing fault classification[J]. Vibration and Shock, 2023, 42(24): 267-275.
[19] Lida Zareian, Javad Rahebi, Mohammad Javad Shayegan. Bitterling fish optimization (BFO) algorithm[J]. Multimedia Tools and Applications, 2024, 83: 75893-75926.
[20] Mcinnes L, Healy J, Saul N, et al. UMAP: Uniform Manifold Approximation and Projection for Dimension Re-duction[J].The Journal of Open Source Software,2018,3(29):1-63.