

Article

A Rolling Bearing Fault Diagnosis Method Using Adaptive Decomposition and Impact Feature Enhancement Fusion

Siqi Peng¹ and Hanshan Li^{2,*} 
¹ School of Intelligent Science and Engineering, Xi'an Peihua University, Xi'an 710125, China; 003058@peihua.edu.cn

² School of Electronic and Information Engineering, Xi'an Technological University, Xi'an 710021, China

* Correspondence: lihanshan269@163.com

Abstract

To address the issues of parameter dependency on empirical settings, insufficient fault feature extraction capability, and limited classification accuracy in rolling bearing fault diagnosis using Variational Mode Decomposition (VMD), a novel fault diagnosis method based on adaptive signal decomposition and intelligent classification integration is proposed. The VMD parameters are adaptively optimized using the Subtraction-Average-Based Optimizer (SABO), and a kurtosis–correlation criterion is introduced to select a single fault-sensitive intrinsic mode function, from which time-domain features are extracted to construct fault feature vectors. The Moth-Flame Optimization Algorithm (MFOA) is employed to optimize the parameters of the Kernel Extreme Learning Machine (KELM) for fault state identification. From the perspective of methodological symmetry, the averaged population update of SABO is invariant to the ordering of search agents, VMD exhibits equivalence under permutation of mode labels, and KELM constructs the sample similarity matrix using a symmetric kernel function. These symmetry-related structures are integrated into the parameter optimization, modal decomposition, and fault classification stages of the proposed method. Experimental validation using the CWRU rolling bearing dataset demonstrates that the proposed method reaches a fault recognition accuracy of 96.73%, outperforming other comparative models and exhibiting superior diagnostic precision and robustness.

Keywords: rolling bearing fault diagnosis; adaptive signal decomposition; impact feature enhancement; variational mode decomposition; kernel extreme learning machine

1. Introduction

Rolling bearings are fundamental components of rotating machinery, and their operating condition directly affects the operational stability and safety of mechanical equipment. Under long-term high-speed rotation, heavy loads and complex working conditions, bearings are prone wear, impact and fatigue, resulting in problems such as inner ring faults, outer ring faults and rolling element faults. Once a bearing fails, it will not only cause increased vibration and decreased operational accuracy of the equipment, but in severe cases, it may even lead to the failure of the entire mechanical system. Therefore, accurately and quickly identifying the faults of rolling bearings has significant engineering application value [1].

The vibration signals of rolling bearings actually collected usually have obvious nonlinear, non-stationary and strong noise characteristics, and traditional time–frequency analysis methods are difficult to use for effectively extracting the fault impact features. In



Academic Editor: Yunqing Tang

Received: 11 June 2026

Revised: 24 July 2026

Accepted: 26 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

recent years, signal processing methods such as Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), Complete Ensemble Empirical Mode Decomposition (CEEMD), and Variational Mode Decomposition (VMD) have been widely applied in the field of bearing fault diagnosis [2]. Among them, VMD can effectively suppress the mode aliasing problem in the traditional EMD method by constructing a variational constraint model and has good anti-noise performance and frequency band separation ability. Therefore, it has received extensive attention in the analysis of mechanical fault signals [3]. However, the decomposition effect of VMD largely depends on the setting of the number of mode components K and the penalty factor α . If the parameters are not selected reasonably, it is easy to have insufficient decomposition or over-decomposition, which will affect the subsequent fault feature extraction and diagnosis results. Therefore, how to achieve the adaptive optimization of VMD parameters has become an important research direction at present. To address this issue, intelligent optimization methods such as Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), and Harris Hawk Optimization (HHO) have been used for VMD parameter optimization [4–6], but some algorithms are prone to falling into local optima during the later stage of the iteration process, resulting in a decrease in search accuracy. As a new intelligent optimization algorithm, the Subtractive Average-Based Optimizer (SABO) has the advantages of fast convergence speed, strong global search ability, and less likely to fall into local optima, showing good application potential in complex parameter optimization problems [7].

After completing the fault signal decomposition, how to select the effective components containing the main fault impact information from multiple IMF components is also very crucial. Traditional research mostly uses the minimum envelope entropy criterion for IMF selection, but envelope entropy mainly reflects the complexity of the signal and is not sensitive enough to the transient impact features in bearing faults. In contrast, kurtosis can more effectively reflect the impulse impact characteristics in vibration signals, and is therefore more suitable for the selection of fault feature components in rolling bearings [8]. In addition, Kernel Extreme Learning Machine (KELM) is widely used in mechanical fault classification and recognition due to its fast training speed and strong generalization ability [9], but its kernel parameters and regularization parameters also have a significant impact on classification performance, so it is necessary to further improve the model performance with the help of optimization algorithms. The Moth-Flame Optimization Algorithm (MFOA) has been proven to have excellent global optimization ability and convergence performance in parameter optimization problems [10].

Existing studies have separately improved VMD parameter selection, fault-sensitive component extraction, and classifier parameter optimization. However, the information transfer relationship between adaptive decomposition and subsequent fault classification has not been sufficiently emphasized. In particular, inappropriate VMD parameters affect the source and quality of fault features, whereas inappropriate KELM parameters affect the final classification boundary. Therefore, to address the empirical dependence of VMD parameters, the difficulty in extracting fault-induced impact features, and the sensitivity of classifier parameters, the main contributions of this study are summarized as follows:

- SABO is combined with VMD to adaptively optimize the mode number K and penalty factor α using kurtosis as the fitness function. A kurtosis–correlation criterion is then employed to select the fault-sensitive IMF. This process reflects the population-order invariance of SABO and the mode-permutation equivalence of VMD, introducing methodological symmetry into the proposed framework.
- Time-domain features are extracted from the optimal IMF, and MFOA is employed to optimize the KELM regularization coefficient C and kernel parameter σ . Accordingly,

a rolling bearing fault diagnosis method integrating adaptive decomposition, impact feature enhancement, and nonlinear classification is established.

- Comparative experiments are conducted on the CWRU rolling bearing dataset. Fixed data partitions, 20 repeated runs, comparison with seven diagnostic methods, confusion matrix analysis, and computational-time evaluation are used to assess the diagnostic performance and run-to-run variability of the proposed method.

The remainder of this paper is organized as follows: Section 2 reviews relevant research on rolling bearing fault diagnosis, signal decomposition, feature extraction, and intelligent classification; Section 3 presents in detail the proposed fault diagnosis framework for rolling bearings based on adaptive decomposition and impact feature enhancement fusion, including the overall framework, adaptive decomposition method, and impact feature enhancement strategy; Section 4 introduces the experimental dataset and validates the effectiveness of the proposed method from aspects such as signal preprocessing results and fault diagnosis performance; finally, Section 5 summarizes the entire work.

2. Related Work

In the field of rolling bearing fault diagnosis, vibration signals usually exhibit nonlinear, non-stationary, and noise-contaminated characteristics under varying load–speed operating conditions. Therefore, many researchers have focused on signal decomposition, feature extraction, and intelligent classification methods. In signal decomposition, Huang et al. [11] proposed empirical mode decomposition (EMD), which can adaptively decompose nonlinear and non-stationary signals into several intrinsic mode functions. Li et al. [12] proposed a VMD-FRFT-based rolling bearing fault diagnosis method to improve fault feature extraction in the early fault stage. Miao et al. [13] developed an improved VMD method for bearing fault diagnosis, which enhanced the ability to extract weak fault information. To solve the problem of VMD parameter selection, intelligent optimization algorithms have been introduced in recent studies. Fu et al. [14] proposed a kurtosis-based global VMD optimization method to extract weak bearing fault features under heavy noise. Wang et al. [15] proposed an SSA-optimized VMD and gray wolf optimization extreme learning machine model, which improved the diagnostic accuracy of rolling bearings. These studies indicate that optimization algorithms can improve VMD performance, but many existing fitness functions, such as envelope entropy and sample entropy, mainly measure signal complexity and may be insufficiently sensitive to transient impact features.

In terms of fault classification, the kernel extreme learning machine (KELM) has been widely used due to its fast training speed and strong nonlinear mapping ability. Li et al. [16] proposed a rolling bearing fault diagnosis method based on VMD and an improved KELM, achieving good classification performance. In addition, swarm intelligence algorithms have been used to optimize classifier parameters. Zhang et al. [17] employed a kurtosis-criterion-based VMD and a modulo square threshold to enhance bearing fault feature extraction under noisy conditions. However, the classification performance of KELM still depends on kernel parameters and regularization coefficients, and improper parameter settings may reduce model stability. More recently, several studies have further investigated the integration of adaptive signal processing and intelligent fault feature extraction. Koca et al. [18] combined VMD-based noise suppression with Transformer-based feature representation and rCCA-enhanced correlation modeling for bearing fault classification. Ren et al. [19] proposed a CEEMDAN-IAWT method to suppress noise interference and preserve useful fault information in rolling bearing vibration signals. Xia et al. [20] developed a hybrid diagnosis model combining optimized VMD with fuzzy dispersion entropy to improve nonlinear fault feature representation. Tian et al. [21] introduced an improved PSO-VMD method with multidimensional evaluation indices for bearing fault diagnosis in a crane

amplitude-changing mechanism. Liu et al. [22] combined SSA-VMD with MCKD to enhance periodic impulsive components and improve rolling bearing fault feature extraction. These studies demonstrate that adaptive decomposition, component evaluation, and impact enhancement can improve fault diagnosis performance, but most methods mainly focus on one processing stage and do not fully establish the information-transfer relationship between front-end signal decomposition and back-end classifier optimization.

In addition to methods combining signal decomposition with shallow classifiers, approaches that learn fault-sensitive representations directly from raw signals or time–frequency representations have received increasing attention. Mallat et al. [23] proposed the wavelet scattering transform, which constructs multiscale signal representations with translation stability through cascaded wavelet convolutions; however, its fixed filter structure limits its adaptability to different data distributions. Vincent et al. [24] developed stacked denoising autoencoders to learn robust nonlinear latent features from noise-contaminated data, although the reconstruction objective is not completely consistent with the fault classification objective. Hochreiter et al. [25] proposed the LSTM network, which can effectively capture long-term temporal dependencies in vibration sequences, but it generally requires more training data and computational resources and provides limited interpretability. Tasdemir et al. [26] compared statistical-feature-based machine learning methods with a raw-waveform CNN–Attention model and demonstrated that direct waveform learning can preserve dynamic information, although the model and training procedures are more complex. Yilmaz et al. [27] combined continuous wavelet transforms with CNN, BiGRU, a state-space model, and an autoencoder for UAV vibration anomaly detection, improving the joint learning of multiscale time–frequency and temporal features at the expense of increased computational complexity. Altunkaya et al. [28] integrated a semi-analytical model, a machine learning surrogate, and intelligent optimization, demonstrating the advantages of mechanism–data integration, although their study was not concerned with rotating-machinery fault diagnosis. Konar et al. [29] compared MLP, 1D-CNN, LSTM, and attention-based models and showed that different architectures have different capabilities for learning nonlinear and temporal features, while their performance remains sensitive to data size and parameter settings.

More recently, researchers have continued to investigate the coordinated integration of adaptive signal decomposition and intelligent feature learning. Li et al. [30] introduced the Newton–Raphson-Based Optimization algorithm to optimize the key parameters of VMD and developed a WDCNN–Informer dual-branch architecture to fuse local impulsive features with global temporal dependencies. This method improves the completeness of fault representation and the diagnostic capability under noisy conditions; however, the multi-branch deep network increases model complexity and computational cost, while its internal decision process remains difficult to interpret. Liu et al. [31] proposed a Levy–Cauchy Optimized Sparrow Search Algorithm to adaptively determine the VMD mode number and penalty factor and subsequently employed a CNN–BiLSTM model to extract local patterns and long-term temporal information from the decomposed signals. The method improves the adaptability of VMD parameter selection and the robustness of fault classification, but its performance remains dependent on the selected fitness function, network architecture, training-data scale, and hyperparameter configuration. These recent studies show that rolling bearing fault diagnosis is developing toward adaptive decomposition, multilevel feature fusion, and intelligent classification. Nevertheless, several challenges remain, including the stable optimization of decomposition parameters under varying load–speed and noise conditions, the reliable selection of fault-sensitive modes, the effective transfer of information between front-end signal decomposition and back-end classification, and balancing diagnostic accuracy, robustness, interpretability, and computational cost.

In summary, existing bearing fault diagnosis methods can be broadly divided into decomposition-based interpretable methods and automatic feature learning methods. Decomposition-based methods provide explicit signal-processing stages and physically meaningful fault indicators, but they are sensitive to decomposition and classifier parameters. Learned-feature approaches can automatically extract multiscale, nonlinear, or temporal representations, but their performance is often affected by model architecture, training data, and computational cost. To address these issues, this study establishes a unified framework that combines SABO-VMD adaptive decomposition, kurtosis–correlation-based fault-sensitive IMF selection, nine-dimensional time-domain feature construction, and MFOA-KELM classification, thereby incorporating these symmetry-related structures into parameter optimization, modal decomposition, and fault classification.

3. Methods

3.1. Overall Framework of the Proposed Method

To improve the extraction of fault-related information from nonlinear and non-stationary vibration signals and enhance fault identification performance, a rolling bearing fault diagnosis method based on adaptive decomposition and impact feature enhancement fusion is proposed. The proposed framework integrates adaptive signal decomposition, impact feature enhancement, feature fusion, and intelligent fault classification into a unified diagnostic scheme. The population update of SABO is invariant to the ordering of search agents, VMD exhibits equivalence under permutation of mode labels, and KELM constructs the sample similarity matrix using a symmetric kernel function. These symmetry-related properties provide structural consistency among parameter optimization, modal decomposition, and fault classification. By combining parameter-adaptive decomposition, fault-sensitive feature enhancement, and symmetry-related algorithmic structures, the proposed method can effectively improve the representation capability of fault features and the accuracy of rolling bearing fault diagnosis. The overall framework of the proposed method is shown in Figure 1.

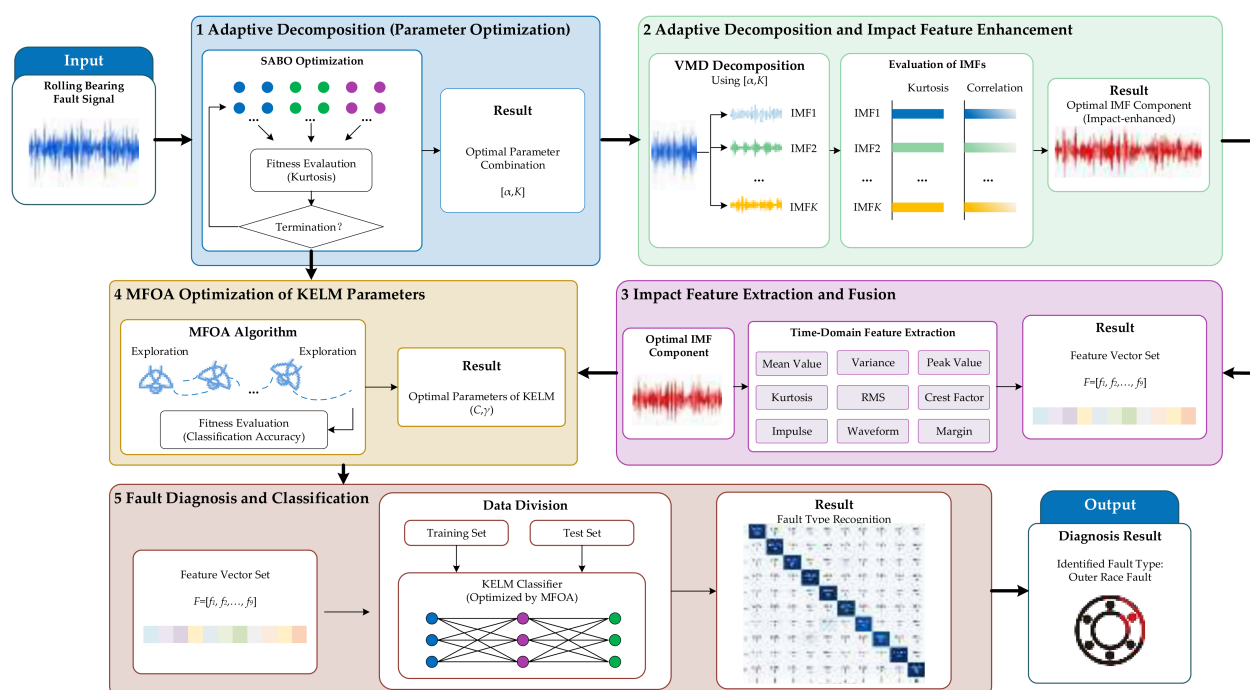


Figure 1. Overall Framework of the Proposed Rolling Bearing Fault Diagnosis Method.

The overall framework of the proposed rolling bearing fault diagnosis method based on adaptive decomposition and impact feature enhancement fusion is illustrated in Figure 1, and the detailed procedure can be summarized as follows.

(1) Different types of rolling bearing fault signals are first input into the adaptive decomposition framework. Taking kurtosis as the fitness evaluation index, the SABO algorithm is employed to optimize the key parameters of VMD, including the penalty factor and the number of decomposition modes, thereby obtaining the optimal parameter combination $[\alpha, K]$.

(2) The optimized parameter combination is then introduced into VMD to adaptively decompose the original fault signal into multiple IMF components. The kurtosis values and correlation characteristics of each IMF component are subsequently evaluated, and the component containing the most significant impulsive fault information is selected to achieve impact feature enhancement.

(3) Based on the selected optimal IMF component, representative time-domain features are extracted, including the mean value, variance, peak value, kurtosis, root mean square, crest factor, impulse factor, waveform factor, and margin factor. These features are fused to construct the rolling bearing fault feature vector set.

(4) To further improve fault identification performance, the MFOA is employed to optimize the kernel parameter and regularization coefficient of the KELM classifier. Through iterative search, the optimal classifier parameters are obtained, enhancing both classification accuracy and generalization capability.

(5) The constructed fault feature vector set is divided into mutually independent training, validation, and testing subsets. The training subset is used to train candidate KELM classifiers, whereas the validation subset is used to calculate the MFOA fitness value and determine the optimal KELM parameters. After parameter optimization, the training and validation subsets are combined to retrain the final KELM classifier. The independent testing subset is used only for final fault state identification and performance evaluation.

3.2. Adaptive Decomposition Method

Rolling bearing vibration signals usually exhibit strong nonlinearity, non-stationarity, and noise interference under varying load–speed operating conditions. Directly extracting fault features from raw signals often leads to information loss and reduced diagnostic accuracy. Therefore, an adaptive decomposition method is adopted to separate fault-related components from complex vibration signals and enhance the representation of fault characteristics.

In the proposed framework, Variational Mode Decomposition (VMD) is employed to decompose the original vibration signal into several intrinsic mode functions (IMFs) with different frequency characteristics. Since the decomposition performance of VMD is highly dependent on the penalty factor and the number of decomposition modes, the Subtraction-Average-Based Optimizer (SABO) is introduced to adaptively optimize these key parameters. By combining SABO with VMD, the optimal decomposition results can be obtained while effectively reducing mode mixing and over-decomposition. The overall adaptive decomposition process is illustrated in Figure 1.

To better describe the proposed adaptive decomposition framework, the principles of VMD [3] and SABO [7] are introduced in the following subsections.

First, the original signal $s(t)$ is analytically decomposed into K finite-bandwidth modal functions through the Hilbert transform to obtain the single-sideband spectrum $u_k(t)$. An exponential term containing the center frequency $w_k(t)$ is introduced to modulate $u_k(t)$, thereby shifting the frequency spectrum of $u_k(t)$ to the baseband.

Second, the bandwidth of $u_k(t)$ is estimated by calculating the gradient L2-norm of the modulated signal, and the variational model with constraints is constructed by Equation (1).

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \partial_k \left\| \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \\ \text{s.t. } \sum_k u_k = s(t) \end{cases} \quad (1)$$

where $\{u_k\}$ and $\{\omega_k\}$ represent the modal component set and the center frequency set, respectively; ∂_t is the partial derivative operator; and “*” denotes the convolution operator. $\delta(t)$ is the unit impulse function.

Third, by introducing the Lagrange multiplier $\lambda(t)$ and the quadratic penalty term a , the augmented Lagrange expression can be obtained by Equation (2).

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & a \sum_k \partial_k \left\| \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \\ & \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (2)$$

Fourth, the alternating direction multiplier method is employed to iteratively update the modal component set and the center frequency set until the optimal solution of the variational model is obtained, thereby realizing the decomposition of the original signal $s(t)$.

After the VMD is completed, in order to extract the most impactful information components, this paper uses the kurtosis–correlation coefficient criterion [32] to evaluate each IMF component. Kurtosis can reflect the strength of the impact components in the signal, and it can be defined by Equation (3).

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{s(i) - \mu}{\sigma} \right)^4, \quad (3)$$

where μ is the mean value, $s(i)$ is the vibration amplitude corresponding to the discrete point of the time-domain waveform sequence, N is the number of discrete sequence points, and σ is the standard deviation.

According to the kurtosis formula, when the standard deviation σ of the signal decreases, that is, when the fluctuation or dispersion degree of the signal decreases, the kurtosis value correspondingly increases. When the kurtosis value approaches 3, it indicates that the distribution of the vibration signal is close to a normal distribution, which means that the operating unit is in a normal operating condition without faults. Once impact or abnormal signals appear in the signal, the overall distribution of the vibration signal deviates from the normal distribution, thereby causing the kurtosis value to increase significantly.

Because kurtosis is sensitive to abnormal impacts, the correlation coefficient R is introduced as an index to quantify the degree of correlation between the IMF component and the original signal. It can be calculated by Equation (4).

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(s(i) - \bar{s})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (s(i) - \bar{s})^2}}, \quad (4)$$

where $\bar{x}$ and $\bar{s}$ are the expectations of x and s , respectively. The closer the absolute value of R is to 1, the stronger the correlation between the variables. After VMD, the correlation

coefficient is first used to evaluate the degree of information preservation between each IMF and the original vibration signal. Kurtosis is then employed to characterize the impulsiveness of each IMF. By jointly considering the two indicators, the single IMF component with the best comprehensive performance is selected as the fault-sensitive component for subsequent feature extraction and fault classification. The selected IMF index is determined using the training data and is then fixed for the validation and testing sets to prevent testing information from influencing feature selection.

This paper uses SABO to optimize the VMD parameters. SABO is a mathematical behavior-based optimization algorithm that updates the spatial positions of group members using the subtractive average value. It has a fast convergence speed and strong optimization ability. The specific optimization process is as follows:

Firstly, each optimization problem has a search space, which is a subset of the dimensional space, and the dimension d is equal to the number of decision variables of the given problem. Based on its position in the search space, the algorithm search agent determines the values of the decision variables. Therefore, each search agent contains information about the decision variables. The population is randomly initialized according to Equations (5) and (6).

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}_{N \times n} = \begin{bmatrix} x_{1,1} & \dots & x_{1,d} & \dots & x_{1,n} \\ x_{2,1} & \dots & x_{2,d} & \dots & x_{2,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,d} & \dots & x_{N,n} \end{bmatrix}_{N \times n}, \quad (5)$$

$$x_{i,d} = l_{b,d} + r_{i,d}(u_{b,d} - l_{b,d}), \quad (6)$$

In Equation (5), X is the SABO population matrix, X_i is the i -th search agent (population member), and $x_{b,d}$ is the dimensional decision variable in the search space. N is the number of search agents, m is the number of decision variables, and r is a random number within $[0, 1]$. $u_{b,d}$ and $l_{b,d}$ represent the upper and lower bounds of the decision variables, respectively.

The objective function of each search agent represents a candidate solution of the decision variables, and the objective function is expressed by Equation (7).

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(X_1) \\ F(X_2) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1}, \quad (7)$$

In Equation (7), F is the vector of objective function values, and F_i is the evaluation value of the objective function based on the i -th search agent.

The SABO algorithm can effectively avoid falling into local optimal solutions, which is defined by Equation (8).

$$A - vB = \text{sign}[F(A) - F(B)](A - v * B), \quad (8)$$

In Equation (8), $\vec{v}$ is a vector with dimension m , and “ $*$ ” represents the Hadamard product of two vectors. $F(A)$ and $F(B)$ are the objective function values of A and B , respectively.

In the algorithm, the updated position X_i^{new} of each search agent X_i is obtained by averaging the v -subtraction operation. The update formula is expressed by Equation (9).

$$X_i^{new} = X_i + \tau_i \cdot \frac{1}{N} \sum_{j=1}^N (X_j - X_i), i = 1, 2, \dots, N, \quad (9)$$

For the update of Equation (9), the updated candidate position of X_i can be obtained by Equation (10). when the condition is satisfied, one update is completed; otherwise, the result is discarded, and one iteration is completed.

$$X_i = \begin{cases} X_i^{new}, & F_i^{new} < F_i \\ X_i, & \text{else} \end{cases}, \quad (10)$$

The decomposition performance of VMD is jointly affected by the mode number K and penalty factor α . Inappropriate parameter settings may lead to under-decomposition, over-decomposition, or mode mixing, thereby affecting the extraction of impulsive characteristics from rolling bearing fault signals. To reduce the dependence of VMD parameters on empirical selection, the Subtraction-Average-Based Optimizer (SABO) is employed to jointly optimize K and α . The kurtosis of the decomposed IMF components is used as the fitness criterion so that the optimized decomposition can emphasize the transient impacts generated by localized bearing faults. The SABO-based VMD parameter optimization procedure is summarized in Algorithm 1.

Algorithm 1: SABO-Based Optimization of VMD Parameters

Input:

Training vibration signal $s(t)$
 Search ranges $K \in [3, 10]$ and $\alpha \in [100, 1500]$
 Population size $N = 20$
 Maximum number of iterations $T = 20$

Output:

Optimal VMD parameters (K^*, α^*)
 IMF set $\{u_1, u_2, \dots, u_{K^*}\}$

- 1: Initialize the SABO population $X_i = [K_i, \alpha_i], i = 1, 2, \dots, N$
 - 2: for $t = 1$ to T do
 - 3: for $i = 1$ to N do
 - 4: Round K_i to the nearest integer
 - 5: Apply boundary constraints to K_i and α_i
 - 6: Decompose $s(t)$ using VMD with $[K_i, \alpha_i]$
 - 7: Calculate the kurtosis of each IMF component
 - 8: Evaluate the kurtosis-based fitness of X_i
 - 9: end for
 - 10: Record the current optimal solution
 - 11: Update the population using the SABO subtraction-average strategy
 - 12: end for
 - 13: Output the optimal parameter combination (K^*, α^*)
 - 14: Decompose $s(t)$ using VMD with (K^*, α^*)
 - 15: Obtain the optimal IMF set $\{u_1, u_2, \dots, u_{K^*}\}$
 - 16: return (K^*, α^*) and $\{u_1, u_2, \dots, u_{K^*}\}$
-

The specific steps for optimizing the VMD parameters using SABO are as follows:

(1) Obtain the original vibration signals from the rolling bearing training set, initialize the SABO population, and determine the search ranges of the VMD parameters, namely $K \in [3, 10]$ and $\alpha \in [100, 1500]$.

(2) Each search agent represents a candidate VMD parameter combination $[K, \alpha]$. The different candidate parameter combinations are successively substituted into VMD to decompose the rolling bearing vibration signals and obtain the corresponding IMF components.

(3) Calculate the kurtosis of each IMF under every candidate parameter combination and determine the fitness of the candidate solution according to the kurtosis evaluation. A larger kurtosis indicates that the corresponding decomposition result contains more prominent transient impact characteristics.

(4) Update the search agents according to the subtraction-average position-update mechanism of SABO. The updated value of K is rounded to the nearest integer, and boundary constraints are applied when K or α exceeds the predefined search range.

(5) After each iteration, update and retain the current optimal fitness value and its corresponding parameter combination. The VMD, fitness evaluation, and population update processes are repeated until the maximum number of iterations is reached.

(6) Output the optimal parameter combination $[K^*, \alpha^*]$, and substitute it into VMD to decompose the rolling bearing fault signal into K^* IMF components. The kurtosis and correlation coefficient of each IMF are then calculated, and a single fault-sensitive IMF is selected for nine-dimensional time-domain feature extraction and subsequent fault classification.

In this study, the SABO population size and maximum number of iterations are both set to 20 to balance parameter-search capability and computational cost. The search ranges of the VMD mode number K and penalty factor α are set to 3–10 and 100–1500, respectively. Since K is a discrete parameter, it is rounded during each iteration, whereas α is continuously updated according to the SABO position-update results. The final optimal parameter combination is consistently applied to the training, validation, and testing signals to ensure that all data subsets follow the same signal-decomposition procedure.

3.3. Impact Feature Enhancement Method

In this paper, the Moth-Flame Optimization Algorithm (MFOA) is employed to jointly optimize the regularization coefficient C and kernel parameter σ of KELM [9]. MFOA [10] simulates the spiral flight behavior of moths around flames, in which the position of each moth represents a candidate KELM parameter combination $[C_i, \sigma_i]$, whereas the flame represents a high-quality solution obtained during the current optimization process. Through iterative fitness evaluation and position updating, MFOA searches for the optimal parameter combination and thereby improves the fault classification performance of KELM.

To prevent the testing set from participating in model selection, the training set is used to train KELM under each candidate parameter combination, whereas the validation set is used to evaluate the corresponding classification performance. The validation classification error is employed as the MFOA fitness value. The independent testing set is not involved in fitness calculation or parameter optimization and is used only for the final evaluation of the optimized classifier.

1. Initialize the moth population. The moth population can be represented using the matrix of Formula (11).

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1d} \\ m_{21} & m_{22} & \cdots & m_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nd} \end{bmatrix}, \quad (11)$$

In the Formula (11), n represents the population size of moths; d represents the dimension of the input data [10];

2. Conduct fitness evaluation for the moth population. The fitness value vector can be expressed by Formula (12).

$$O_M = \begin{bmatrix} O_{M1} \\ O_{M2} \\ \vdots \\ O_{Mn} \end{bmatrix}, \quad (12)$$

3. Repeat the following process until the stopping criterion is met:

(1) Adaptively update the number of flames n . When the iteration count is 1, the number of moths is equal to the number of flames;

(2) Sort the fitness of the moth population and select the n moths with the better fitness as the flames;

(3) Update the search parameters of the moths;

(4) Update the positions of the moths based on the corresponding flames and the flight parameters of each moth. The moth position matrix can be expressed by Formula (13).

$$F = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1d} \\ F_{21} & F_{22} & \cdots & F_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \cdots & F_{nd} \end{bmatrix}, \quad (13)$$

The equation for the mechanism of updating the position of each moth relative to the flame can be expressed by Formula (14).

$$M_i = S(M_i, F_j), \quad (14)$$

where M_i represents the i -th moth, F_j represents the j -th flame, and S denotes the spiral function, it can be expressed by Formula (15).

$$S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j, \quad (15)$$

where D_i represents the distance between the i -th moth and the j -th flame; b is a constant defining the logarithmic spiral shape; and the path coefficient t is a random number within $[-1, 1]$. The expression of D_i is given by Formula (16).

$$D_i = |F_j - M_i|, \quad (16)$$

4. Output the obtained optimal solution (flame).

Kernel extreme learning machine is a machine learning algorithm that combines the kernel method with the extreme learning machine. Extreme learning machine is a fast single-hidden-layer feedforward neural network, characterized by randomly generating the weights and biases of the hidden layer and then using the least-squares method to solve the output weights. The kernel extreme learning machine introduces the idea of kernel function on this basis, thereby improving the generalization ability of the algorithm and its ability to handle nonlinear problems. The kernel extreme learning machine can enhance the prediction performance of the model while retaining the advantages of extreme learning machine. This paper uses the optimized kernel extreme learning machine as a fault diagnosis model to diagnose and predict faults.

In the extreme learning machine (ELM), the weights w and biases b of the hidden neurons are randomly generated. The connection weight β from the input layer to the

output layer is solved by the least-squares method. The activation function of the hidden layer is denoted as $g(\cdot)$, and the linear transformation from the input layer to the hidden layer is expressed by Formula (17).

$$h_i = g(w_i \cdot x + b_i), \quad (17)$$

where w_i is the weight vector from the input layer to the i -th hidden neuron, and b_i is the bias of the i -th hidden neuron.

For N samples, the hidden layer output can be represented by Formula (18).

$$\mathbf{H} = [h(x_1), h(x_2), \dots, h(x_N)]^T, \quad (18)$$

The objective of ELM is to obtain the output weight β by minimizing the output error. KELM introduces a kernel function $K(x_i, x_j)$ on the basis of ELM to directly calculate the similarity between samples. The kernel matrix can be defined by Formula (19).

$$K(x_i, x_j) = h(x_i) \cdot h(x_j), \quad (19)$$

Therefore, the kernel matrix K can be gained by Formula (20).

$$\mathbf{K} = \mathbf{H}\mathbf{H}^T, \quad (20)$$

After obtaining the kernel matrix K , the output weight β can be calculated by Formula (21).

$$\beta = \left(\mathbf{K} + \frac{\mathbf{I}}{C} \right)^{-1} \mathbf{Y}, \quad (21)$$

where I denotes the identity matrix, and C is the regularization coefficient used to balance model complexity and generalization ability.

For a new input sample x_{new} , the prediction result is obtained by calculating the kernel similarity between the new sample and the training samples, it can be gained by Formula (22).

$$\hat{y}_{\text{new}} = K(x_{\text{new}}, X_{\text{train}})\beta, \quad (22)$$

The setting of the penalty coefficient C and the kernel function σ directly affects the diagnostic results of KELM. Reasonable setting of the value of C will reduce the randomness of the model, and σ will influence the scope of action of the kernel function. MFOA is introduced to optimize the regularization coefficient C and kernel parameter σ of KELM, aiming to obtain a KELM model with the optimal parameter configuration and improve the accuracy of fault diagnosis. The optimization process is shown in Figure 1. MFOA is introduced to optimize the regularization coefficient C and kernel parameter σ of KELM.

The regularization coefficient C is used to balance the fitting error and model complexity, whereas the kernel parameter σ determines the effective mapping range of the radial basis kernel function. Therefore, MFOA is introduced to jointly optimize these two parameters. The optimal parameter combination obtained by MFOA is assigned to KELM to improve its classification performance for different rolling bearing fault states.

The fault diagnosis of rolling bearings based on the optimized KELM is divided into 5 steps:

- (1) Determine the structure of the KELM classifier and initialize the MFOA parameters, including the moth population, maximum number of iterations, and the search ranges of the regularization coefficient C and kernel parameter σ .

- (2) Initialize the moth population. Each moth represents a candidate KELM parameter combination $[C_i, \sigma_i]$. For each candidate solution, the training feature set is used to train the corresponding KELM classifier.
- (3) Evaluate the classification performance of each candidate KELM model using the validation feature set. The validation classification error is used as the fitness value. The moths are then sorted according to their fitness values, and the better moth positions are selected as flames.
- (4) Calculate the distance between each moth and its corresponding flame, and update the moth positions according to the logarithmic spiral search mechanism. Repeat the fitness evaluation and position update until the maximum number of iterations is reached, and output the optimal regularization coefficient C^* and kernel parameter σ^* .
- (5) After the optimal parameters are obtained, combine the training and validation feature sets to retrain the final KELM classifier. The independent testing feature set is then used for final fault classification and performance evaluation, and the corresponding rolling bearing fault category is output.

3.4. Adaptive Decomposition and Impact Feature Enhancement Fusion Method

The proposed adaptive decomposition and shock feature enhancement fusion method in this paper consists of four modules: SABO-VMD adaptive decomposition, kurtosis–correlation IMF filtering, nine-dimensional time-domain feature fusion, and MFOA-KELM classification. SABO optimizes VMD parameters by leveraging the subtraction average relationship among search individuals, and its overall update result is not affected by the order of individual arrangement; VMD is equivalent under modal number permutation; and KELM constructs the sample similarity matrix using a symmetric kernel function. These symmetric correlation structures correspond to the damage of the approximate periodic response of healthy bearings caused by localized faults, thereby integrating adaptive decomposition, shock feature enhancement, and intelligent classification into the same diagnostic framework. To better illustrate the implementation process of the proposed adaptive decomposition and impact feature enhancement fusion method, the complete fault diagnosis procedure is shown in Algorithm 2 and Figure 2.

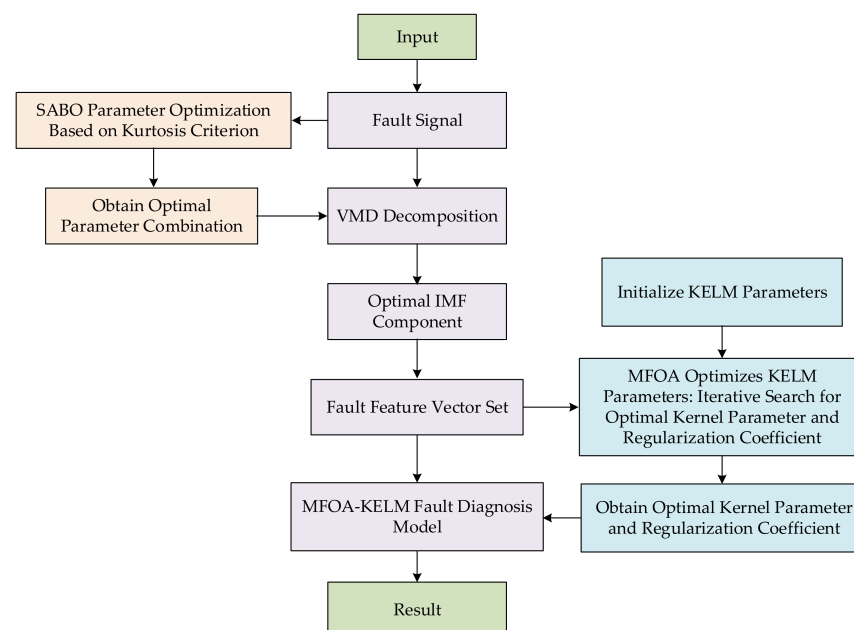


Figure 2. Specific process of the fault diagnosis method.

Algorithm 2: Adaptive Decomposition and Impact Feature Enhancement Fusion Method

Input:

Training signals S_{tr} and labels Y_{tr}
 Validation signals S_{val} and labels Y_{val}
 Testing signals S_{te} and labels Y_{te}
 VMD parameter ranges $K \in [K_{\min}, K_{\max}]$ and $\alpha \in [\alpha_{\min}, \alpha_{\max}]$
 SABO population size N_s and maximum iterations T_s
 KELM parameter ranges $C \in [C_{\min}, C_{\max}]$ and $\sigma \in [\sigma_{\min}, \sigma_{\max}]$
 MFOA population size N_m and maximum iterations T_m

Output:

Predicted fault labels $\hat{Y}_{te}$
 Optimal parameters $(K^*, \alpha^*, C^*, \sigma^*)$

- 1: Initialize the SABO population
 $X_i = [K_i, \alpha_i], i = 1, 2, \dots, N_s$
- 2: for $t = 1$ to T_s do
- 3: for $i = 1$ to N_s do
- 4: Round K_i to the nearest integer and apply boundary constraints
- 5: Decompose S_{tr} using VMD with candidate parameters (K_i, α_i)
- 6: Obtain IMF set $U_i = \{u_{i,1}(t), u_{i,1}(t), \dots, u_{i,K_i}(t)\}$
- 7: Calculate the kurtosis-based VMD fitness J_i
- 8: end for
- 9: Record the current optimal parameter combination
- 10: Update X_i using the SABO subtraction-average strategy
- 11: end for
- 12: Obtain the optimal VMD parameters (K^*, α^*)
- 13: Decompose S_{tr} , S_{val} , and S_{te} using VMD with (K^*, α^*)
- 14: Obtain IMF sets U_{tr} , U_{val} , and U_{te}
- 15: for $k = 1$ to K^* do
- 16: Calculate the kurtosis K_{uk} of the k -th training IMF
- 17: Calculate the correlation coefficient R_k of the k -th IMF and the original training signal
- 18: end for
- 19: Select the single fault-sensitive IMF index K^*
 according to the kurtosis–correlation criterion
- 20: Fix K^* for the training, validation, and testing sets
- 21: Obtain u_{tr,K^*} , u_{val,K^*} and u_{te,K^*}
- 22: Extract nine time-domain features from each selected IMF:
 $f = [\text{Mean, Variance, Peak, Kurtosis, RMS, Crest Factor, Impulse Factor, Waveform Factor, Margin Factor}]$
- 23: Construct the feature sets F_{tr} , F_{val} , and F_{te}
- 24: Initialize the moth population
 $M_i = [C_i, \sigma_i], i = 1, 2, 3, \dots, N_m$
- 25: for $t = 1$ to T_m do
- 26: for $i = 1$ to N_m do
- 27: Train a candidate KELM using F_{tr} , Y_{tr} , C_i , and σ_i
- 28: Calculate the validation accuracy $Acc_{val,i}$ using F_{val}
- 29: Calculate the MFOA fitness: $F_i = 1 - Acc_{val,i}$
- 30: end for

Algorithm 2: *Cont.*

```

31:   Sort the moths according to  $F_i$  and determine the flame positions
32:   Update  $M_i$  using the logarithmic spiral search mechanism
33:   Record the current optimal KELM parameters
34: end for
35: Obtain the optimal KELM parameters ( $C^*$ ,  $\sigma^*$ )
36: Combine  $F_{tr}$  with  $F_{val}$  and combine  $Y_{tr}$  with  $Y_{val}$ 
37: Retrain the final KELM using ( $C^*$ ,  $\sigma^*$ )
38: Predict the testing labels  $\hat{Y}_{te}$  using  $F_{te}$ 
39: Evaluate the final fault diagnosis performance using  $\hat{Y}_{te}$  and  $Y_{te}$ 
40: return  $\hat{Y}_{te}$  and ( $K^*$ ,  $\alpha^*$ ,  $C^*$ ,  $\sigma^*$ )

```

From the Algorithm 2 and Figure 2, it is not difficult to find that the adaptive decomposition and impact feature enhancement fusion method involves three steps. First, SABO jointly optimizes the VMD mode number K and penalty factor α using kurtosis as the fitness criterion, and the fault signal is decomposed using the obtained optimal parameters. Second, a single fault-sensitive IMF is then selected according to kurtosis and correlation, from which nine time-domain features are extracted to construct the fused feature vector. Final, MFOA optimizes the KELM regularization coefficient C and kernel parameter σ using the validation classification error as the fitness value, and the optimized KELM is employed for final fault identification. The fault-sensitive IMF and fused feature vector connect front-end representation optimization with back-end classification optimization, forming a complete diagnostic chain of adaptive decomposition, impact feature enhancement, feature fusion, and fault identification.

4. Experiments

4.1. Dataset Description

The experimental data used in this study were obtained from the Case Western Reserve University (CWRU) rolling bearing dataset. The 12 kHz drive-end vibration signals collected from the SKF6205 bearing were selected. Four load–speed operating conditions were considered: 0 hp at approximately 1797 r/min, 1 hp at approximately 1772 r/min, 2 hp at approximately 1750 r/min, and 3 hp at approximately 1730 r/min. Under each operating condition, the data included the normal condition, inner-race fault, outer-race fault, and rolling element fault. The inner-race, outer-race, and rolling element faults each contained three fault diameters of 0.1778 mm, 0.3556 mm, and 0.5334 mm. Therefore, ten bearing-health categories were considered under each load–speed condition, including one normal category and nine fault categories.

To ensure the completeness of the data construction procedure and prevent adjacent or overlapping signal segments from appearing in different data subsets, the original continuous vibration signals were first divided chronologically into mutually exclusive training, validation, and testing intervals. A guard interval containing no fewer than 1024 sampling points was retained between the training and validation intervals and between the validation and testing intervals. The samples within the guard intervals were not used for data construction. Sliding-window segmentation was then performed separately within the three data intervals. Each signal sample contained 1024 consecutive sampling points, corresponding to approximately 0.0853 s at a sampling frequency of 12 kHz. The sliding-window stride was set to 512 points, resulting in a 50% overlap between adjacent windows. Window overlap occurred only within the same data subset. No window crossed the boundary of the training, validation, testing, or guard intervals, and no adjacent or overlapping windows were shared among the three data subsets.

For each fault category, 72 training samples, 18 validation samples, and 30 testing samples were generated, resulting in 120 samples per category. A total of 1200 samples were constructed for the ten categories, including 720 training samples, 180 validation samples, and 300 testing samples. The dataset partitioning was completed before signal decomposition, IMF evaluation, feature extraction, and classifier optimization and remained fixed throughout all experiments. The training set was used for model fitting, the validation set was used for parameter selection, and the independent testing set was used only for final performance evaluation. All comparison methods employed exactly the same original signal intervals, window length, window stride, sample sizes, subset partitions, and class labels to ensure a fair comparison.

In order to further analyze the differences in vibration characteristics of rolling bearings under different operating conditions, this paper conducted time-domain and frequency-domain analyses of the vibration signals under normal and faulty conditions. The corresponding results are shown in Figures 3 and 4. Compared with the vibration signals under normal conditions, the vibration amplitude of the faulty signals significantly increases, and the waveform periodic characteristics also undergo significant changes. As can be seen from Figures 3 and 4, there are obvious differences in the time-domain and frequency-domain characteristics between the normal equipment signals and the faulty equipment signals. In the time-domain waveform, the normal signals are generally stable overall, with relatively small vibration amplitudes and a relatively stable periodic change pattern, and no obvious impact vibration phenomenon is observed. In contrast, the vibration amplitude of the faulty signals significantly increases, the waveform fluctuation is more intense, and there are also strong random impact components, indicating abnormal mechanical vibrations during equipment operation. In the frequency-domain characteristics, the frequency energy of the normal signals is mainly concentrated in the low-frequency region, and the spectral distribution is relatively simple; while the faulty signals have obvious frequency peaks near the fundamental frequency, and there are also more frequency components and wide-band energy aggregation phenomena in the medium- and high-frequency regions, indicating that the vibration signals contain strong impact modulation characteristics.

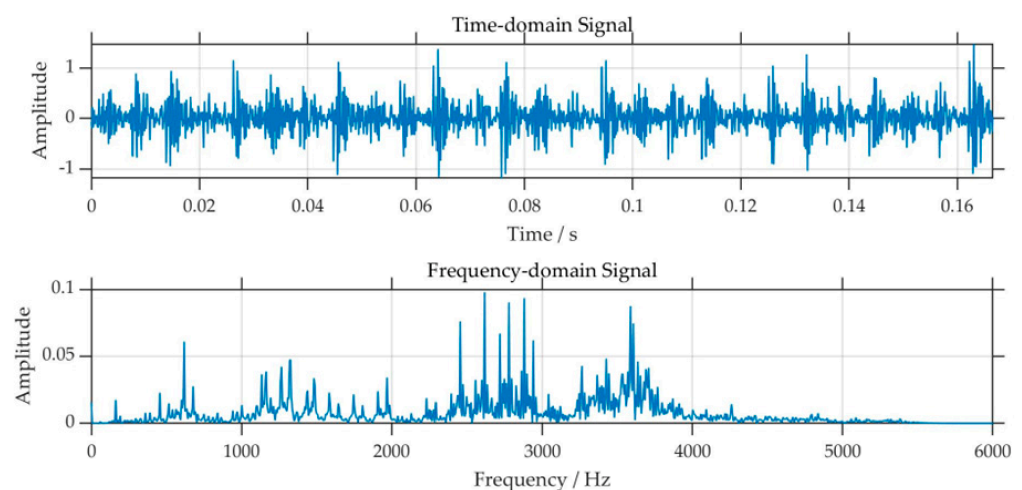


Figure 3. Time- and Frequency-Domain Representations of an IR007 Inner-Race Fault Signal at 0 HP and 1797 r/min.

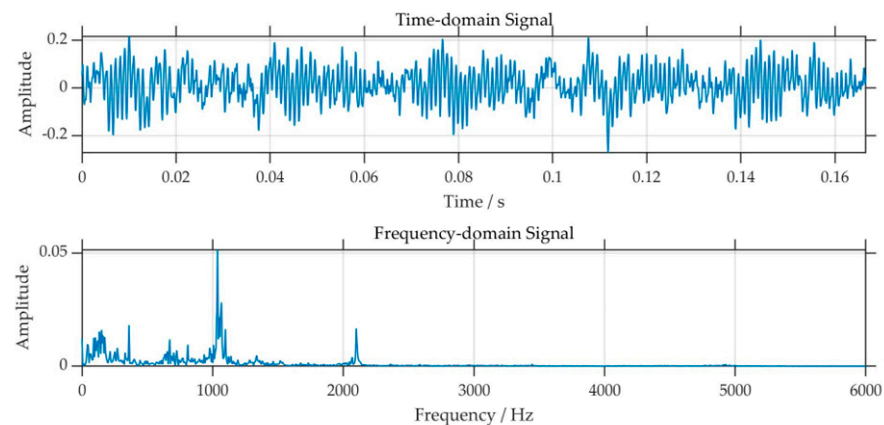


Figure 4. Time- and Frequency-Domain Representations of a Normal Bearing Signal.

4.2. Signal Preprocessing Results Based on Adaptive Decomposition

The decomposition effect of VMD is jointly influenced by the penalty factor and the number of modal components. Inappropriate parameter settings can easily lead to modal aliasing or over-decomposition, thereby affecting the effect of fault feature extraction. To improve the adaptive ability of VMD parameter selection, this paper uses SABO to optimize the VMD parameters $[\alpha, K]$. Specifically, the search range for the number of modal components K is set to $[3, 10]$, the search range for the penalty factor α is set to $[100, 1500]$, the population size is set to 20, and the maximum number of iterations is set to 20. Parameter optimization is carried out. Figure 5 presents the SABO-based VMD parameter optimization process for the IR007 inner-race fault signal under the 0 HP and 1797 r/min operating condition. As shown in Figure 5, the kurtosis-based fitness value converged to 7.71693 after four iterations. The optimal VMD parameters obtained were a mode number of $K = 4$ and a penalty factor of $\alpha = 725$.

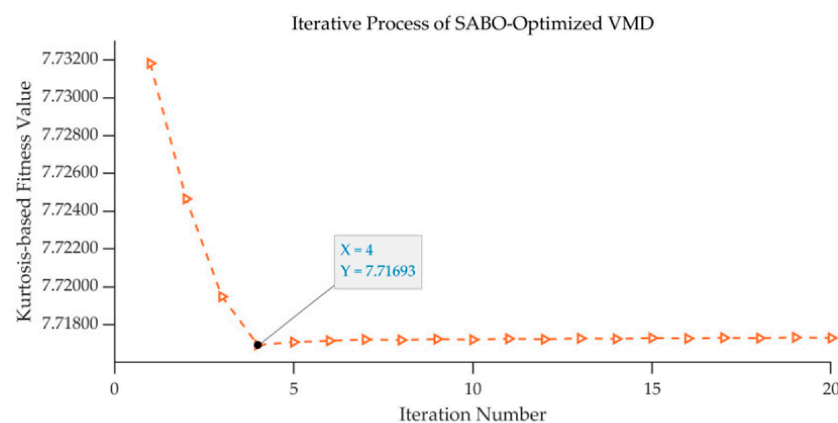


Figure 5. SABO-VMD Optimization Curve for IR007 at 0 HP and 1797 r/min.

To analyze the influence of different operating conditions on the VMD performance, the 0.007-inch inner-race fault (IR007) signals were selected as the research objects. The vibration signals under four load–speed conditions, namely 0 HP–1797 r/min, 1 HP–1772 r/min, 2 HP–1750 r/min, and 3 HP–1730 r/min, were decomposed using VMD. The corresponding frequency-domain decomposition results for $K = 3$, $K = 4$ and $K = 5$ are shown in Figures 6, 7 and 8 respectively.

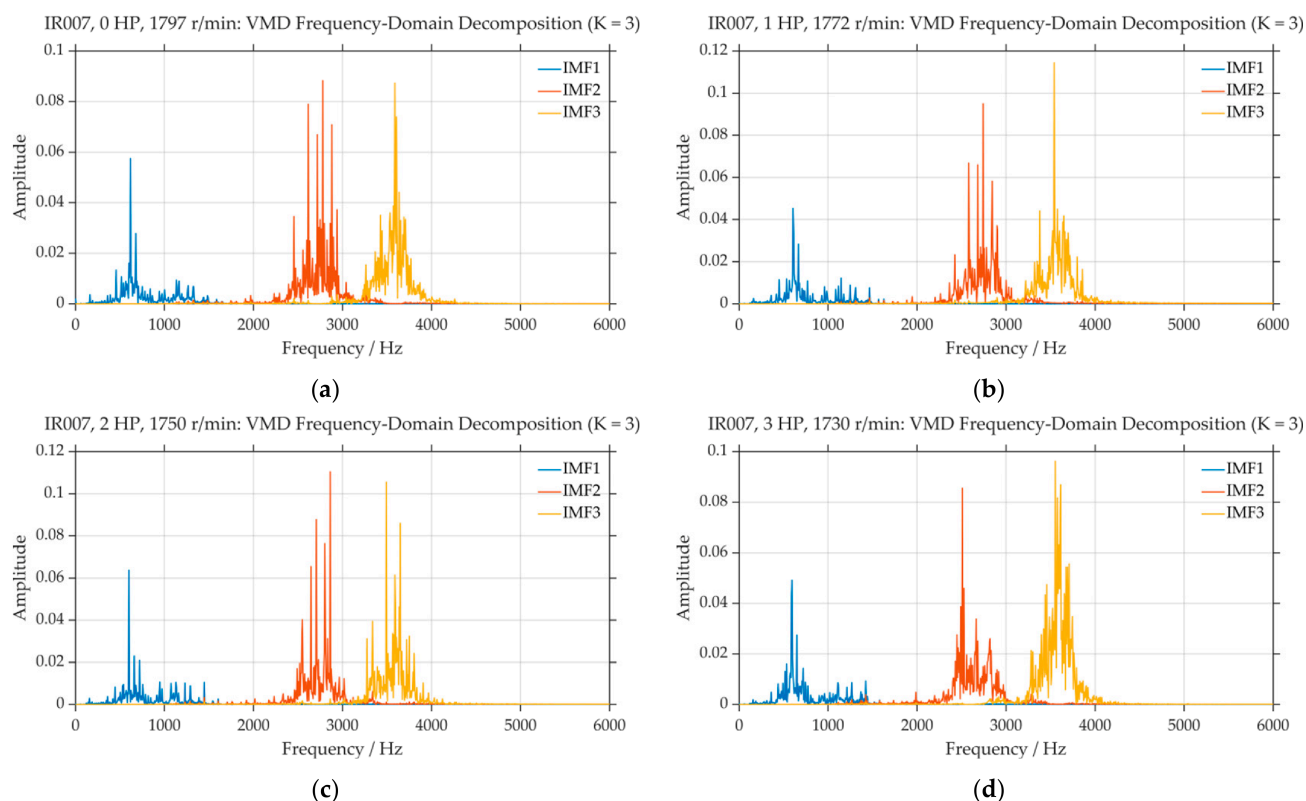


Figure 6. VMD frequency-domain decomposition results with $K = 3$ under four load-speed operating conditions: (a) 0 HP and 1797 r/min; (b) 1 HP and 1772 r/min; (c) 2 HP and 1750 r/min; and (d) 3 HP and 1730 r/min.

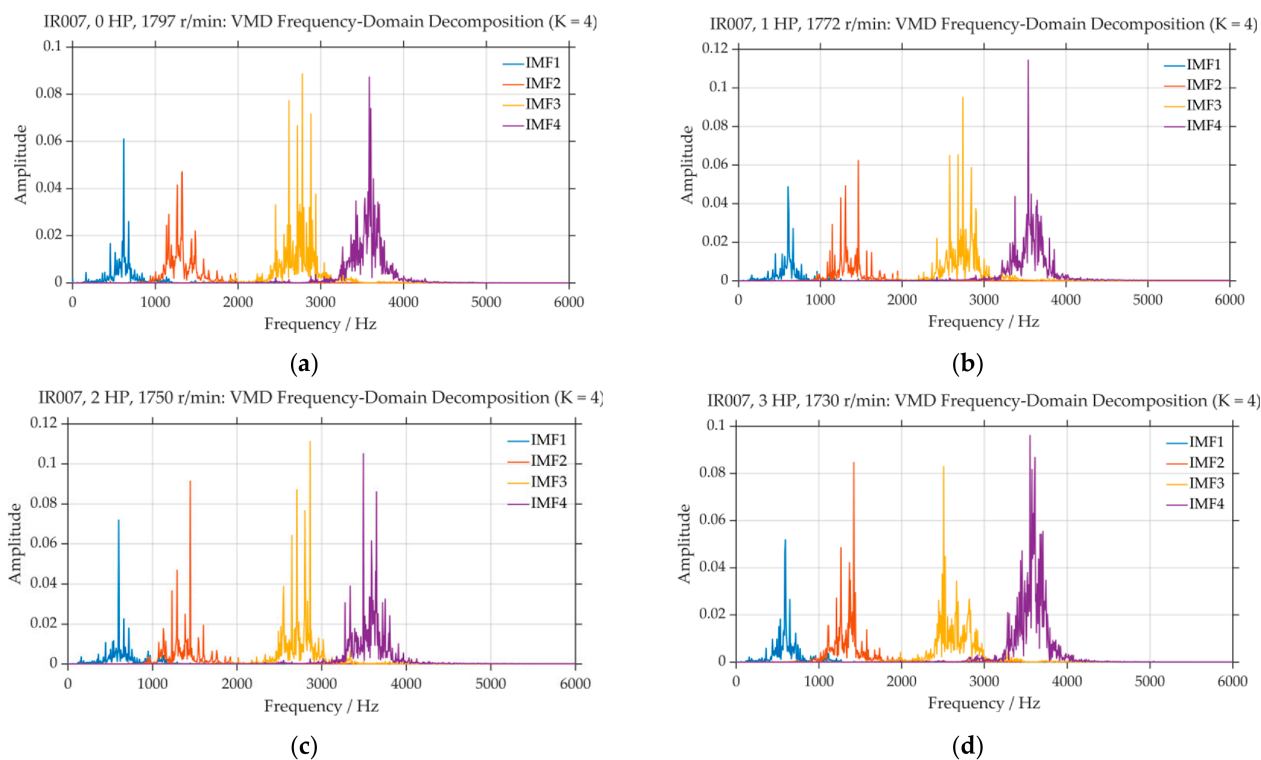


Figure 7. VMD frequency-domain decomposition results with $K = 4$ under four load-speed operating conditions: (a) 0 HP and 1797 r/min; (b) 1 HP and 1772 r/min; (c) 2 HP and 1750 r/min; and (d) 3 HP and 1730 r/min.

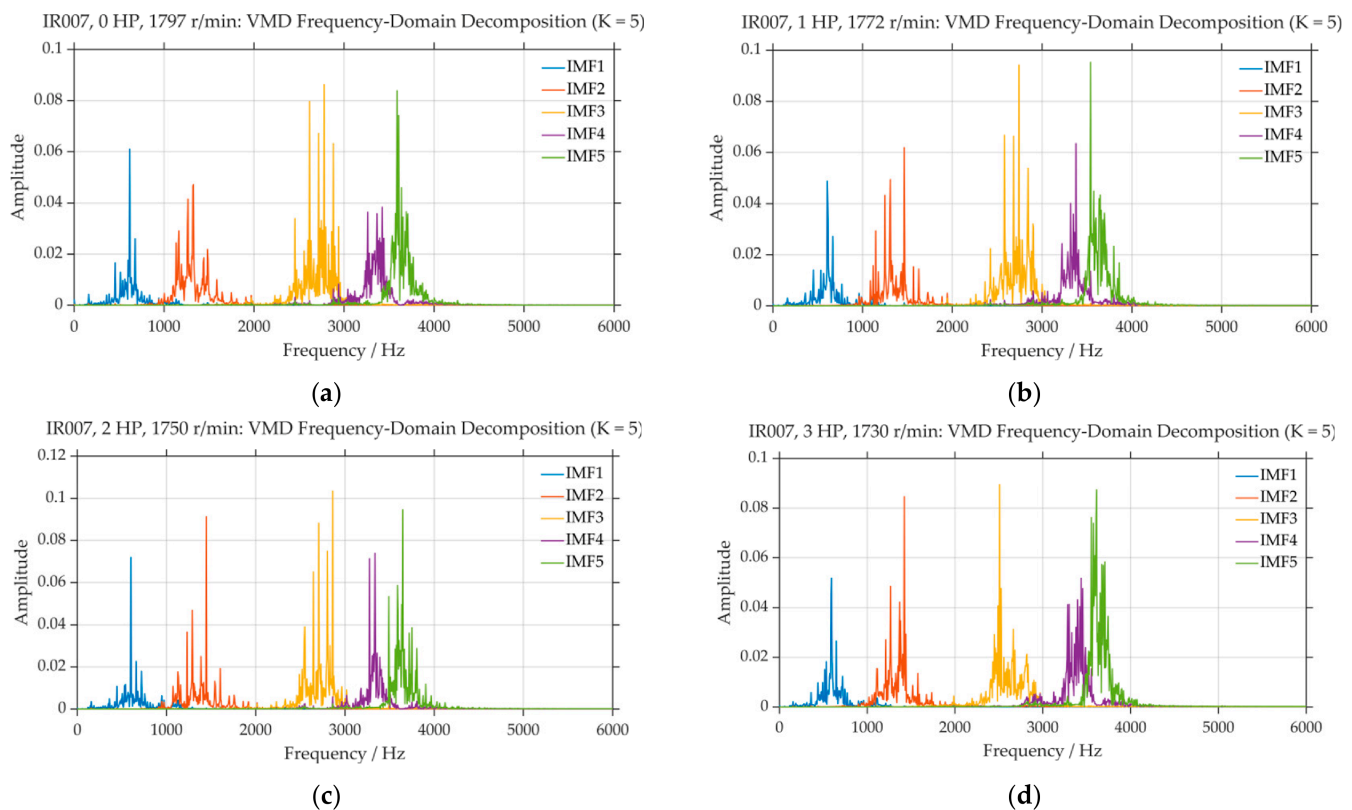


Figure 8. VMD frequency-domain decomposition results with $K = 5$ under four load–speed operating conditions: (a) 0 HP and 1797 r/min; (b) 1 HP and 1772 r/min; (c) 2 HP and 1750 r/min; and (d) 3 HP and 1730 r/min.

From the VMD frequency-domain decomposition results of the IR007 fault signals under four different load–speed operating conditions and with different numbers of modes, as shown in Figures 6–8, it can be seen that when $K = 3$, the frequency band division of each IMF component is relatively broad, and there is a significant overlap of some frequency components, indicating that the VMD is insufficient at this time and that the fault features are still coupled to some extent with other frequency components; when $K = 5$, although the frequency separation ability is further enhanced, the frequency bands of some IMF components are too close, resulting in the dispersion of fault frequency energy, indicating that there is a certain over-decomposition problem at this time. In contrast, when $K = 4$, the frequency distribution of each IMF component is clearer, and the different frequency bands have better independence. It can not only effectively separate the low-frequency vibration components but also relatively completely retain the medium- and high-frequency impact features, and the modal aliasing phenomenon is significantly reduced.

Based on the analysis of the SABO optimization results, it can be seen that when $K = 4$, the IMF4 component has the largest kurtosis value, indicating that this component has the most obvious impact characteristics and can more effectively represent the periodic impact information generated by the rolling bearing fault. From the spectral graph, it can also be observed that IMF4 has a relatively obvious energy concentration and characteristic peak in the high-frequency region, suggesting that this component contains more abundant fault feature information. Therefore, considering the modal separation effect, the band independence, and the kurtosis feature analysis, after VMD using the optimized parameters, IMF4 was identified as the fault-sensitive component according to the joint kurtosis–correlation evaluation, and uses it for subsequent fault feature extraction and classification diagnosis.

The kurtosis values and correlation coefficients of each IMF component calculated are shown in Table 1. The correlation coefficient reflects the similarity between the IMF

component and the original signal. The larger the correlation coefficient, the more complete the original fault information retained by the IMF component. As can be seen from Table 1, the kurtosis values and correlation coefficients of different IMF components show significant differences. Among them, the correlation coefficient of IMF4 reaches 0.7816, which is the maximum among all IMF components, indicating that this component has a high similarity with the original fault signal and can retain the original vibration information more completely. Further analysis of the kurtosis values reveals that the kurtosis value of IMF4 reaches 4.0826, also the maximum among all IMF components, indicating that the shock characteristics in this component are the most obvious and can better reflect the transient shock information generated by the rolling bearing fault. Based on the above analysis results, it can be concluded that when $K = 4$, the VMD can effectively highlight the fault shock characteristics while ensuring modal independence. Therefore, this paper finally selects IMF4 as the best IMF component and conducts subsequent feature extraction and fault classification analysis on it.

Table 1. Kurtosis Values and Correlation Coefficients.

Modal Component	IMF1	IMF2	IMF3	IMF4
Kurtosis Value	3.1247	3.2861	3.5174	4.0826
Correlation Coefficient	0.3142	0.4875	0.6528	0.7816

4.3. Impact Feature Enhancement and Intelligent Fault Identification Results

After the original signal preprocessing based on SABO-VMD was completed, this paper used the MFOA optimization algorithm to optimize the penalty coefficient C and the kernel function σ of KELM. The number of optimization iterations was set to 50, the population size was 10, and the upper and lower bounds were $[0, 1000]$. The best IMF component was selected and 9 time-domain features were extracted to form a feature vector. For each candidate parameter combination, the training feature set was used to train KELM, whereas the validation classification error was used as the MFOA fitness value. The independent testing set was used only for final performance evaluation. The fitness curve of MFOA optimizing KELM is shown in Figure 9.

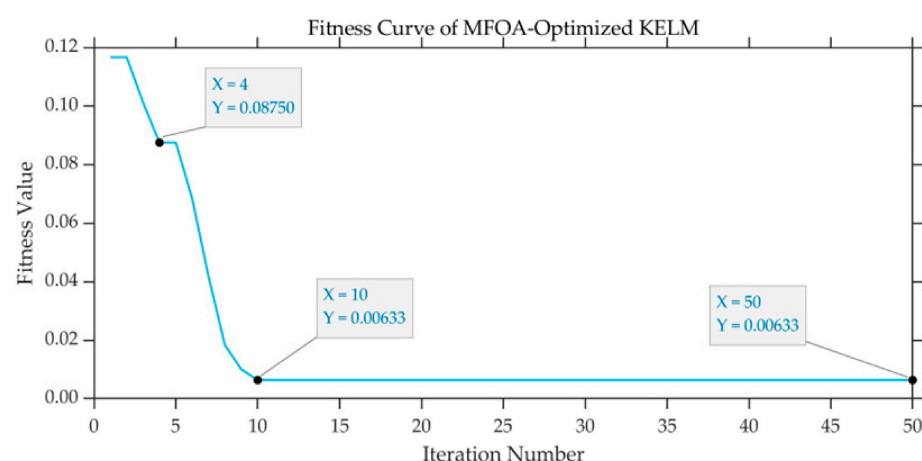


Figure 9. MFOA-KELM Fitness Curve.

As shown in Figure 9, after 4 iterations, the algorithm fell into a local optimal solution state, and the fitness value at this time was 0.08750; at the 10th iteration, it broke out of the local optimal solution and obtained the best solution found during the optimization process, and the fitness value at this time was 0.00633.

To further evaluate the average diagnostic performance and run-to-run variability of different rolling bearing fault diagnosis models, seven methods, including KELM, VMD-KELM, SABO-VMD-KELM, WS-KELM, SAE-KELM, LSTM, and the proposed SABO-VMD-MFOA-KELM, were evaluated over 20 independent runs. All methods were tested using the same training, validation, and testing datasets and the same random seeds to ensure a fair comparison. The classification accuracy distributions obtained from the 20 independent runs are presented in Figure 10, while the average classification accuracy, corresponding standard deviation, and computational time of each method are summarized in Table 2.

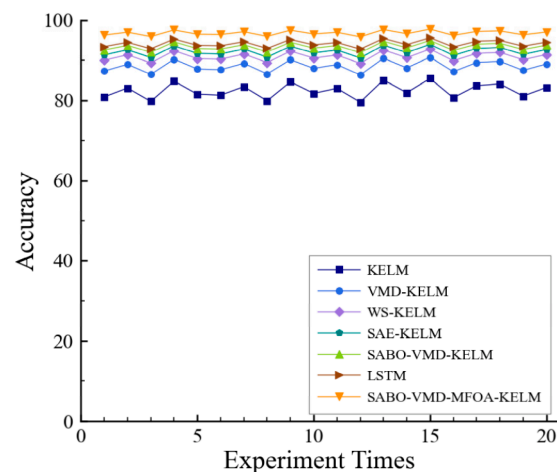


Figure 10. Classification Accuracy of Different Fault Diagnosis Methods over 20 Independent Runs.

Table 2. Performance comparison of different rolling bearing fault diagnosis models over 20 independent runs.

Model	Accuracy (%)	Time (s)
KELM	82.37 ± 1.84	1212.43 ± 18.67
VMD-KELM	88.41 ± 1.39	1418.73 ± 21.46
SABO-VMD-KELM	93.29 ± 0.96	1525.81 ± 24.37
WS-KELM	90.87 ± 1.12	1476.29 ± 20.83
SAE-KELM	92.14 ± 1.07	1589.62 ± 26.41
LSTM	94.08 ± 0.89	1664.27 ± 28.93
SABO-VMD-MFOA-KELM	96.73 ± 0.61	1731.47 ± 31.26

As shown in Figure 10, the classification accuracies of all methods exhibit different degrees of fluctuation over the 20 independent runs. Since all methods used the same fixed training, validation, and testing sets, these fluctuations mainly resulted from the stochastic initialization and parameter optimization processes rather than random data partitioning. Among the compared methods, KELM obtained the lowest average classification accuracy of $82.37\% \pm 1.84\%$, indicating that directly applying KELM to the original fault features was insufficient to fully distinguish different bearing fault states. After introducing VMD for signal decomposition, the average accuracy of VMD-KELM increased to $88.41\% \pm 1.39\%$, demonstrating that signal decomposition improved the representation of fault-related information. The WS-KELM and SAE-KELM methods achieved average accuracies of $90.87\% \pm 1.12\%$ and $92.14\% \pm 1.07\%$, respectively, showing that multiscale feature representation and nonlinear latent-feature learning can further enhance fault classification performance. By employing SABO to optimize the VMD parameters, SABO-VMD-KELM achieved an average accuracy of $93.29\% \pm 0.96\%$, indicating that adaptive parameter optimization improved the quality of VMD. LSTM achieved an average accuracy

of $94.08\% \pm 0.89\%$, reflecting its ability to learn temporal fault characteristics directly from vibration sequences.

The proposed SABO-VMD-MFOA-KELM method achieved the highest average classification accuracy of $96.73\% \pm 0.61\%$. Compared with KELM, VMD-KELM, SABO-VMD-KELM, WS-KELM, SAE-KELM, and LSTM, the average diagnostic accuracy was improved by 14.36, 8.32, 3.44, 5.86, 4.59, and 2.65 percentage points, respectively. Moreover, the proposed method obtained the smallest standard deviation among all compared methods, indicating lower performance fluctuation over the 20 independent runs. As summarized in Table 2, the average computational time of the proposed method was 1731.47 ± 31.26 s, which was higher than that of the comparison methods because SABO and MFOA were employed to optimize the VMD and KELM parameters, respectively. Nevertheless, the improvement in diagnostic accuracy and the reduction in performance fluctuation demonstrate that the proposed method achieves an effective balance between classification performance and computational cost. Therefore, the integration of SABO-VMD adaptive decomposition, fault-sensitive IMF selection, time-domain feature fusion, and MFOA-KELM classification can effectively improve the fault identification performance of rolling bearings under the investigated operating conditions.

However, the overall accuracy alone cannot fully reflect the classification capability of different models for individual fault categories. Therefore, the confusion matrices of the four models are further analyzed to evaluate the recognition accuracy and misclassification relationships of each fault type. The corresponding confusion matrix results are shown in Figure 11.

From the confusion matrix results, it can be seen that using only the KELM model has obvious misclassification, and the identification confusion between some fault categories is relatively serious. After introducing VMD for signal decomposition, the classification accuracy of the VMD-KELM model improved to a certain extent, indicating that VMD can effectively extract fault feature information. Further optimization of the VMD parameters using SABO has significantly improved the classification performance of the SABO-VMD-KELM model, and most fault categories achieved relatively high identification accuracy, with the misclassification phenomenon significantly reduced. The SABO-VMD-MFOA-KELM model proposed in this paper has achieved the highest classification accuracy, with only a very small number of samples misjudged, indicating that optimizing VMD parameters using SABO and the KELM parameters using MFOA, it can more effectively extract the impact characteristics of rolling bearing faults and improve the model's classification ability, thereby verifying the effectiveness and superiority of the proposed method for rolling bearing fault diagnosis.

Overall, the proposed method constructs a unified fault diagnosis procedure by integrating SABO-based parameter optimization, VMD-based adaptive decomposition, kurtosis–correlation-based fault-sensitive IMF selection, and MFOA–KELM classification. Its main advantage is that VMD can separate fault signals with overlapping frequency bands and similar characteristics, while the joint kurtosis–correlation criterion emphasizes fault components with strong impulsiveness and relatively complete information preservation. The subsequent fusion of nine time-domain features and MFOA-based classifier optimization further improves the representation of stochastic fault information under complex operating conditions, reduces the misclassification of similar fault categories, and provides an effective approach for rolling bearing condition monitoring and fault detection. However, because the proposed method integrates parameter optimization, signal decomposition, feature selection, and classifier optimization, its computational time is higher than that of some comparison methods, although it achieves higher fault recognition accuracy and lower observed run-to-run variability under the current experimental setting.

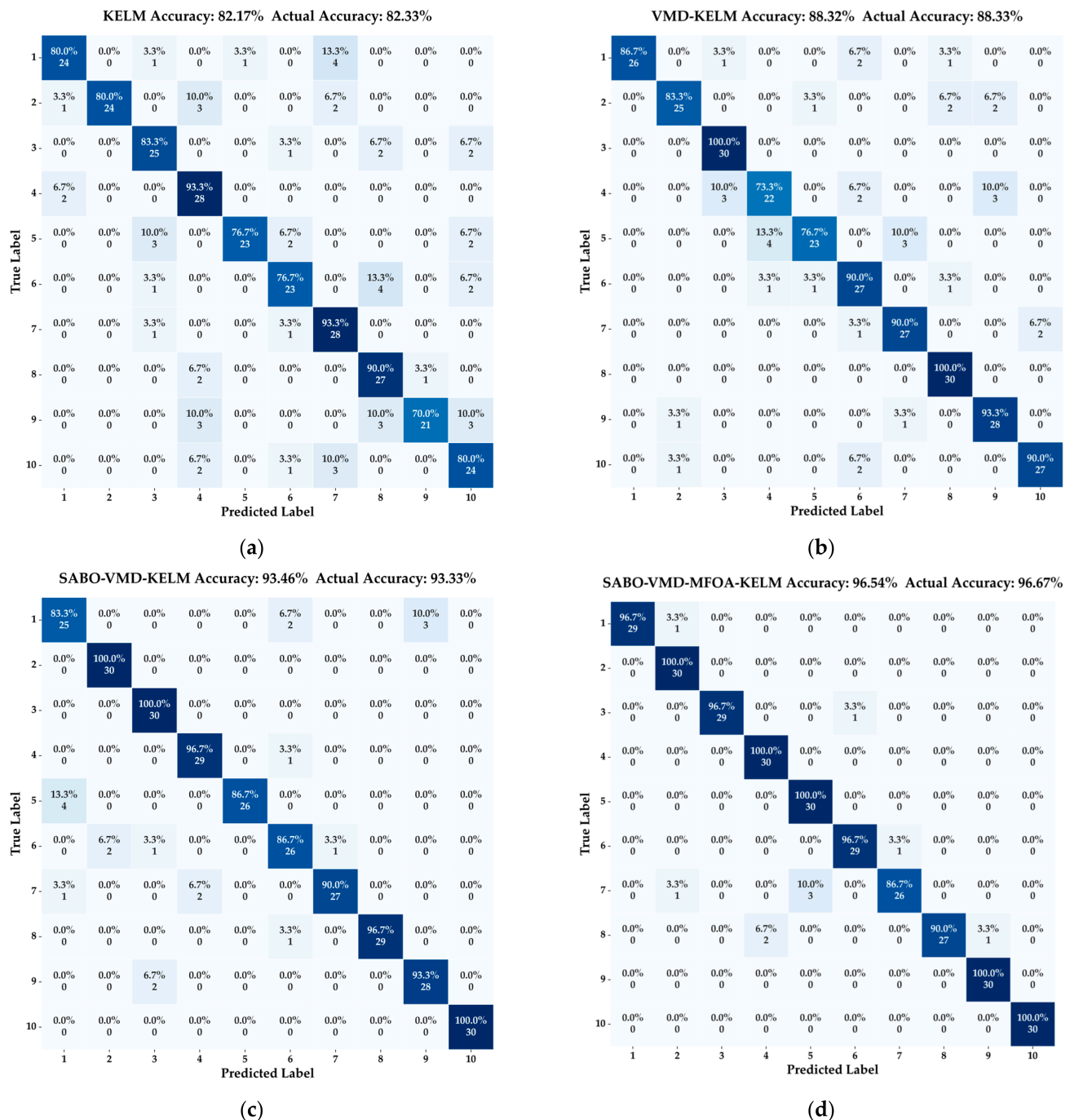


Figure 11. Confusion matrix results of different rolling bearing fault diagnosis models: (a) KELM; (b) VMD-KELM; (c) SABO-VMD-KELM; (d) SABO-VMD-MFOA-KELM.

5. Conclusions

This paper proposes a rolling bearing fault diagnosis method based on adaptive signal decomposition and intelligent classification fusion. The method was experimentally validated using the CWRU rolling bearing dataset, and we obtained the conclusion as follows:

(1) SABO was employed to optimize the mode number and penalty factor of VMD, reducing the dependence on empirical parameter selection and improving the decomposition of fault-related vibration components.

(2) Kurtosis and correlation coefficients were jointly used to select a single fault-sensitive IMF component. Nine time-domain features were then extracted to construct the fault feature vector for subsequent classification.

(3) MFOA was used to optimize the regularization coefficient and kernel parameter of KELM. Over 20 independent runs, the proposed SABO-VMD-MFOA-KELM method achieved the highest classification accuracy of $96.73\% \pm 0.61\%$, outperforming KELM, VMD-KELM, SABO-VMD-KELM, WS-KELM, SAE-KELM, and LSTM. Although its computational time increased to 1731.47 ± 31.26 s, the proposed method obtained the highest accuracy and the smallest standard deviation, demonstrating its effectiveness and relatively stable diagnostic performance under the investigated operating conditions.

Author Contributions: Conceptualization, methodology, software and validation, investigation, writing original draft, writing—review and editing, S.P.; methodology, software, investigation, supervision, writing—original draft, writing—review and editing, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shaanxi Provincial Department of Education (Grant No. 25JK0593).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Randall, R.B.; Antoni, J. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [\[CrossRef\]](#)
2. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [\[CrossRef\]](#)
3. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
4. He, C.; Li, C.; Chen, X.; Liu, Y. Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm-optimized extreme learning machine. *Measurement* **2021**, *173*, 108636. [\[CrossRef\]](#)
5. Zhou, J.; Xiao, M.; Niu, Y.; Ji, G. Rolling bearing fault diagnosis based on WGWOA-VMD-SVM. *Sensors* **2022**, *22*, 6281. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [\[CrossRef\]](#)
7. Trojovský, P.; Dehghani, M. Subtraction-average-based optimizer: A new swarm-inspired metaheuristic algorithm for solving optimization problems. *Biomimetics* **2023**, *8*, 149. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Dibaj, A.; Hassannejad, R.; Ettefagh, M.M.; Ehghaghi, M.B. Incipient fault diagnosis of bearings based on parameter-optimized VMD and envelope spectrum weighted kurtosis index with a new sensitivity assessment threshold. *ISA Trans.* **2021**, *114*, 413–433. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2012**, *42*, 513–529. [\[CrossRef\]](#)
10. Mirjalili, S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowl.-Based Syst.* **2015**, *89*, 228–249. [\[CrossRef\]](#)
11. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [\[CrossRef\]](#)
12. Li, X.; Yang, Y.; Pan, H.; Cheng, J.; Cheng, J. Fault diagnosis for rolling bearing based on VMD-FRFT. *Measurement* **2020**, *155*, 107554. [\[CrossRef\]](#)
13. Miao, Y.; Zhao, M.; Lin, J. Application of an improved variational mode decomposition method for bearing fault diagnosis. *ISA Trans.* **2019**, *92*, 257–276.
14. Fu, L.; Zhang, J.; Wang, H.; Li, X. An improved bearing fault diagnosis method based on global variational mode decomposition optimization. *J. Mech. Sci. Technol.* **2023**, *37*, 1783–1796. [\[CrossRef\]](#)
15. Wang, X.; Li, Y.; Zhang, Y.; Wang, H. Fault diagnosis method of rolling bearing based on SSA-VMD and GWO-ELM. *Sci. Rep.* **2024**, *14*, 29624. [\[CrossRef\]](#)

16. Li, K.; Su, L.; Wu, J.; Wang, H.; Chen, P. A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine. *Appl. Sci.* **2017**, *7*, 1004. [\[CrossRef\]](#)
17. Zhang, X.; Luan, Z.; Liu, X. Fault diagnosis of rolling bearing based on kurtosis criterion VMD and modulo square threshold. *J. Eng.* **2019**, *2019*, 8685–8690. [\[CrossRef\]](#)
18. Koca, T.; Er, M.B.; Çıtlak, A. Transformer-Based Bearing Fault Classification with VMD-Based Noise Suppression and rCCA-Enhanced Correlation Modeling. *Machines* **2026**, *14*, 507. [\[CrossRef\]](#)
19. Ren, H.; Wei, C.; Tan, Z.; Luo, L.; Ding, X. Denoising of rolling bearing vibration signals based on CEEMDAN-IAWT method. *J. Vib. Shock* **2023**, *42*, 199–206.
20. Xia, X.; Wang, X.; Chen, W. A hybrid fault diagnosis model for rolling bearing with optimized VMD and fuzzy dispersion entropy. *Int. J. Rotating Mach.* **2025**, *2025*, 7990867. [\[CrossRef\]](#)
21. Tian, Z.; Ke, Z.; Liu, H.; Zhao, B.; Zhu, X.; Tang, Y.; Zhu, C. Bearing fault diagnosis method based on an improved particle swarm optimization–variational mode decomposition (PSO–VMD) multidimensional index and its application in a crane amplitude-changing mechanism. *Mech. Sci.* **2025**, *16*, 431–443. [\[CrossRef\]](#)
22. Liu, Z.; Li, S.; Wang, R.; Jia, X. Research on Fault Feature Extraction Method of Rolling Bearing Based on SSA–VMD–MCKD. *Electronics* **2022**, *11*, 3404. [\[CrossRef\]](#)
23. Mallat, S. Group invariant scattering. *Commun. Pure Appl. Math.* **2012**, *65*, 1331–1398. [\[CrossRef\]](#)
24. Vincent, P.; Laroche, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408. [\[CrossRef\]](#)
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Taşdemir, A.T.; Özkut, E.C.; Yalçın Özkut, G.; Gül, F. Linking tea aroma chemistry to quality grades via a single MOS gas sensor: Classical machine learning vs. deep learning. *Sensors* **2026**, *26*, 3877. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Yılmaz, A.; Özkut, E.C.; Gül, F. Signal intelligence: Vibration-driven deep learning for anomaly detection of rotary-wing UAVs. *Drones* **2026**, *10*, 321. [\[CrossRef\]](#)
28. Altunkaya, A.N.; Özkut, E.C.; Avci, M. Analytical-to-AI pipeline: Modeling and optimization of entropy generation in pulsating non-Newtonian heat flow. *Comput. Math. Appl.* **2026**, *205*, 195–211. [\[CrossRef\]](#)
29. Konar, M.; Ayaz, H.A.; Özkut, E.C.; Türkmen, A. Aircraft takeoff speed prediction with deep learning: A comparative study of MLP, 1D-CNN, LSTM and attention-based architectures on Boeing 737-300 data. *Aeronaut. J.* **2026**, *1–18*. [\[CrossRef\]](#)
30. Li, Z.; Sun, J.; Chen, D.; Shi, P.; Su, X. Fault diagnosis of rolling bearings based on VMD-NR modal decomposition and multi-feature fusion. *Electronics* **2026**, *15*, 950. [\[CrossRef\]](#)
31. Liu, F.; Yue, X. A fault diagnosis method for rolling bearings based on enhanced sparrow search algorithm-optimized VMD and CNN-BiLSTM. *Sensors* **2026**, *26*, 3239. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Zou, Y.; Zhang, X.; Zhao, W.; Liu, T. Rolling bearing fault diagnosis based on optimized VMD combining signal features and improved CNN. *World Electr. Veh. J.* **2024**, *15*, 544. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.