

Research on fault diagnosis of rolling bearings based on multi-method fusion

Shijun Yu¹, Changyou Guo², Haorui Liu³, Hengwei Zhu⁴

^{1, 2, 3, 4}School of Computer and Information, Dezhou University, Dezhou, 253000, China

³Key Laboratory of Intelligent Industrial Equipment Technology of Hebei Province, Hebei University of Engineering, Handan, 056038, China

¹Corresponding author

E-mail: ¹yushijun@dzu.edu.cn, ²49466087@qq.com, ³liuhaorui@dzu.edu.cn, ⁴1430059696@qq.com

Received 24 June 2025; accepted 11 November 2025; published online 13 December 2025

DOI <https://doi.org/10.21595/jve.2025.25159>



Copyright © 2025 Shijun Yu, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. To address the limitation that Variational Mode Decomposition (VMD) relies on empirical settings for the mode decomposition number K and penalty factor α , this paper proposed the RIME-VMD-KNN method for bearing fault diagnosis. Specifically, the RIME algorithm was used to intelligently optimize K and α of VMD, breaking the reliance on experience; Pearson Correlation Coefficient (PCC) was adopted to screen Intrinsic Mode Functions (IMFs) with high fault correlation for signal reconstruction, preserving key features; and the sample entropy of the reconstructed signal was input into KNN for fault identification. Experiments show that the optimization performance of RIME is superior to that of GA, GWO and AOA; the generalization ability is verified by supplementary tests on the XJTU-SY dataset; KNN is simpler and more efficient than SVM, proving the rationality of its selection; the confusion matrix and multiple random cross-validation confirm stability; and computing time and resource data are provided to verify the feasibility of embedded deployment. This method improves the reliability and real-time performance of diagnosis and has engineering value.

Keywords: multi-method fusion, variational mode decomposition, rime optimization algorithm, K-nearest neighbor algorithm.

1. Introduction

As a key part of rotating machinery, rolling bearings transmit rotational power and support rotor systems, with their status determining mechanical system stability and efficiency [1]. Yet their working environment is harsh. High temperature, dust, vibration in metallurgy and mining, plus frequent start-stop, load impact, oil pollution make them high-failure components [2-3]. Common faults include surface damage (pitting, spalling), cracks, and seal aging (worsening wear and vibration) [4-5]. These cause lower efficiency, higher energy use, even shutdowns and accidents [6]. Thus, timely, accurate bearing fault diagnosis is vital for equipment reliability and production efficiency.

Vibration analysis is common for bearing fault detection: acceleration sensors collect signals, analyzed via Fourier transform to identify issues like gear wear or bearing failures from spectral peaks [7-8]. AI is widely used too – artificial neural networks train on fault samples for accurate diagnosis [9], while KNN (non-parametric, efficient) classifies via similar historical vectors, handling noisy data well [10-11]. VMD and KNN are also researched [12-13]. Practical selection depends on bearing conditions and costs to ensure system stability.

As an advanced signal processing technology, VMD demonstrates significant advantages in the fault detection of hydraulic pumps. The VMD method was proposed by Dragomiretskiy and Zosso in 2014. It is capable of decomposing the original signal into multiple Intrinsic Mode Function (IMF) components with different center frequencies and bandwidths. This effectively avoids the end effect and mode mixing problems that exist in the traditional Empirical Mode Decomposition (EMD). It has a solid mathematical foundation and high computational efficiency [14]. In practical applications, many studies have optimized and expanded VMD. F. M. Zhou et

al. used the Particle Swarm Optimization (PSO) algorithm to optimize the parameters of VMD to obtain the best decomposition effect. By processing the vibration signals of the hydraulic pump, the IMF components containing rich fault information were screened out for reconstruction, thus obtaining the vibration signals after noise reduction and improving the accuracy of fault diagnosis [15]. Z. J. Guo et al. used VMD to decompose the vibration signals, and combined with the support vector machine optimized by MPE and the Cuckoo Search algorithm, they achieved effective fault diagnosis [16]. In addition, J. B. Zhang et al. further verified the advantages of VMD in signal decomposition by comparing VMD with other decomposition methods. Their research achievements provide strong support for the application of VMD in the field of mechanical fault diagnosis [17].

Wasim Zaman and others constructed a fault diagnosis method for centrifugal pumps based on a novel Sobel edge scale map and a convolutional neural network (CNN). They used the Sobel edge scale map to preprocess the vibration signals, effectively highlighting the fault features. Subsequently, the processed data was input into the CNN model. The powerful automatic feature extraction ability of the CNN enables the model to learn from a large number of preprocessed centrifugal pump fault sample data, and then accurately establish the mapping relationship between fault features and fault types, achieving efficient and accurate diagnosis of centrifugal pump faults [18]. However, this research method relies on a large number of high-quality samples, and the diagnostic accuracy will be affected when the samples are insufficient or unbalanced. The CNN-based fault analysis method proposed by V. Sinitsin et al. was based on a hybrid model. Vibration signals acquired by wireless acceleration sensors installed on a rotating shaft were converted into time-frequency images through Hilbert-Huang Transform (HHT) as the input of CNN. The features processed by CNN were combined with the signal power of resonance frequency to form a hybrid input, realizing the classification and localization of fault types [19]. The limitation of this method is that it directly inputs data into CNN for fault diagnosis, resulting in a complex model, long training time, and high demand for computing resources. The rolling bearing fault diagnosis method based on CEEMDAN and CNN-SVM proposed by L. Shi et al. decomposed and reconstructed signals through the CEEMDAN algorithm for noise reduction, converted the reconstructed signals into two-dimensional grayscale images as the input of CNN for feature extraction, and finally used SVM optimized by GWO for classification. This method utilizes CNN to automatically extract deep features, combines GWO-SVM to improve classification accuracy, and experiments show that the average diagnostic accuracy is high [20]. However, this method has limitations such as large computational complexity of CEEMDAN decomposition, requirement of a large amount of labeled data for CNN training, and increased algorithmic complexity due to the parameter optimization process of GWO.

The model proposed by P. X. Zhu et al. demonstrates excellent signal classification capabilities in industrial environments, especially for the identification of electromagnetic radiation signals in coal mine settings. Compared with traditional methods, RIME-SVM has significantly improved the classification accuracy, fully demonstrating its efficiency and precision in processing complex industrial signal data, and bringing important progress to the field of coal mine safety monitoring technology [21]. H. H. Song et al. proposed the RIME-SDAE model for cavitation fault diagnosis of centrifugal pumps. The model first performs SVD denoising on three-axis vibration signals and extracts multi-domain features, then uses the RIME algorithm to optimize the parameters of SDAE, and constructs a feature dataset to train the model. This approach can effectively solve the parameter selection problem of traditional models and improve the accuracy of fault identification [22]. Y. H. Shi et al. proposed an OLTC vibration signal denoising method based on the vibration signals of the converter transformer body, which uses the Rime optimization algorithm to optimize VMD-wavelet threshold, for monitoring and diagnosing the condition of on-load tap changers (OLTC) of converter transformers [23].

VMD has limitations in bearing fault diagnosis, as K and α rely on experience. Improper K worsens feature extraction, stability, efficiency and noise suppression. This paper proposes RIME-VMD-KNN: RIME optimizes K and α for VMD, overcoming manual dependence,

extracting fault info in noise. It selects optimal IMFs via PCC, uses sample entropy as features for KNN to diagnose bearing faults. Experiments show it boosts diagnosis reliability/efficiency.

2. Core algorithmic principles

2.1. RIME algorithm

The RIME (Frost Optimization) algorithm, proposed by scholar Huang in 2023, is bionically inspired by rime formation. Its two core mechanisms – soft frost diffusion (global search) and hard frost puncture (local development) – mirror rime growth, enabling adaptive solutions to complex optimization problems.

It has four collaborative stages.

1) Population initialization stage. The solution space of the optimization problem is mapped to a “rime population”, where each individual (rime body) is composed of several particles (decision variables), denoted as $R = \{S_i\}$, where $S_i = \{x_{i,j}\}$. The population is initialized through random distribution or domain knowledge to ensure the diversity of initial solutions and lay a foundation for subsequent searches. The mathematical modeling is shown in Eq. (1):

$$R = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_i \end{pmatrix}. \quad (1)$$

2) Soft rime search mechanism (global exploration). Simulate the random diffusion of rime particles in a weak wind environment to achieve large-area coverage of the solution space. The position update formula is as follows:

$$R_{ij}^{new} = R_{best,j} + r_1 \cos \varphi \cdot \beta \cdot (h(Ub_{ij} - Lb_{ij}) + Lb_{ij}), \quad r_2 < E, \quad (2)$$

where: $R_{best,j}$ represents the j -th particle of the current optimal individual, guiding the search direction; the random number h controls the random number of the central distance between particles to adjust the search step size, and the value range of h is $(0, 1)$; β is the environmental factor (step function), simulating the influence of the external environment on diffusion, and the formula is as follows:

$$\beta = 1 - \left\lfloor \frac{\omega \cdot t}{T} \right\rfloor, \quad (3)$$

where: t is the current number of iterations; T is the maximum number of iterations; and ω adjust the number of step segments (default value: 5).

The adhesion coefficient E gradually decreases with the increase of the number of iterations, controlling the particle condensation probability. The formula is as follows:

$$E = \sqrt{\left(\frac{t}{T}\right)}. \quad (4)$$

3) Hard rime penetration mechanism (local development). Simulate the directional growth of rime under strong wind conditions to promote information exchange between ordinary individuals and the optimal solution. The position update formula is as follows:

$$R_{ij}^{new} = R_{best,j}, \quad r_3 < F^{normr}(S_i), \quad (5)$$

where: the random number determines whether to trigger the penetration operation, and its value range is $(-1, 1)$; $F^{normr}(S_i)$ is the normalized fitness value, which reflects the individual quality, the higher the value, the greater the exchange probability.

4) Forward greedy selection mechanism (iterative optimization). The update strategy is as follows:

$$F_{new}(S_i) \leq F_{old}(S_i). \quad (6)$$

2.2. Variational mode decomposition

As a signal decomposition and estimation method, when processing the original signal $f(t)$, the VMD algorithm decomposes it into K modal functions, and each modal function has its own central frequency. K is the number of modal components preset in advance. With such a mechanism, VMD can analyze the signal characteristics more accurately and provide a better data foundation for subsequent analysis and processing.

Decompose the faulty signal $f(t)$ into K modal functions u_k :

$$u_k(t) = A_k(t) \cos(\varphi_k(t)). \quad (7)$$

where: the phase $\varphi_k(t)$ is a non-decreasing function, $\varphi'_k(t) \geq 0$; the envelope is non-negative, $A_k(t) \geq 0$; the envelope $A_k(t)$ and the instantaneous frequency $\omega_k(t) = \varphi'_k(t)$ are slowly varying for the phase $\varphi_k(t)$.

The constrained variational model is constructed as follows:

$$\min \left\{ \sum_K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \quad (8)$$

$$\text{s. t. } \sum_K u_k(t) = f(t), \quad (9)$$

where: $\delta(t)$ is the Dirichlet function; $\{u_k\} = \{u_1, u_2, \dots, u_K\}$ is the K components; $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_K\}$ is the center frequency of each component.

The augmented Lagrangian function is introduced to further solve the optimization problem, and the corresponding results are obtained as follows:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) &= \alpha \sum_K \left\| \partial_t \left[\left(\delta_t + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_K u_k(t) \right\|_2^2 \\ &+ \langle \lambda(t), f(t) - \sum_K u_k(t) \rangle, \end{aligned} \quad (10)$$

where: α is the penalty factor; λ is the Lagrange multiplication operator.

The equations for the modal functions u_k , central frequencies ω_k , and Lagrange multipliers λ obtained using Parseval's theorem are shown as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}, \quad (11)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}, \quad (12)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[\hat{x}(\omega) - \sum_{k=1}^k \hat{u}_k^{n+1}(\omega) \right], \quad (13)$$

where: ω_k is the central frequency of the IMF.

3. Fault diagnosis and analysis

3.1. RIME optimizes VMD

RIME-optimized VMD is mainly used to enhance the parameter optimization effect of VMD in the field of fault diagnosis and other areas. First, it is necessary to clarify the key parameters to be optimized in VMD, which are mainly the decomposition layer number K and the penalty factor α . Determine a suitable fitness function. For example, permutation entropy can be used to measure the complexity of signals, helping to highlight fault characteristics, so as to quantify the optimization effect. Set parameters such as the population size and the maximum number of iterations T , and initialize the position of population individuals. The specific process is as follows:

- 1) Calculate the fitness value of each individual in the initial population, and find out the current optimal individual and its fitness value.
- 2) Within the number of iterations $t \leq T$, perform the following operations.
- 3) Calculate E using Eq. (4).
- 4) Generate a random number r_2 . If $r_2 < E$, update the individual position through the soft frost search strategy. This strategy is similar to simulating the random diffusion of frost, conducting global exploration in the solution space to avoid the algorithm from falling into local optimality.
- 5) Generate a random number r_3 . If $r_3 < F^{\text{normr}}(S_i)$, activate the hard frost penetration mechanism to achieve information exchange between individuals and enhance local development capabilities.
- 6) After updating the individual position, if the fitness value of the new individual $F_{\text{new}}(S_i) \leq F_{\text{old}}(S_i)$, update the current optimal solution using the forward greedy mechanism.
- 7) $t = t + 1$, continue the next round of iteration.

After the iteration ends, obtain the optimal individual position, i.e., the optimal parameter combination of $[K, \alpha]$. Use these parameters to decompose the original signal through VMD, yielding multiple Intrinsic Mode Function (IMF) components. Calculate the kurtosis values of each IMF component, and screen the IMF components based on the kurtosis magnitude for signal reconstruction to enhance the separability of fault features. Apply the reconstructed signal to fault diagnosis, such as combining machine learning classifiers (SVM, KNN model, etc.) for fault type identification and classification. The flow chart is shown in Fig. 1.

3.2. Diagnostic process

Bearing fault signals are nonlinear and non-stationary, and the fault analysis process is as follows: First, sensors are arranged in different directions of the rotor system to collect vibration data. Then, RIME-VMD is used to decompose the initial signal. The optimal IMF components rich in fault information are screened out according to the kurtosis value, and signal reconstruction is performed. The sample entropy of the reconstructed signal is calculated as a nonlinear fault feature. The reconstructed dataset is divided into a training set, a validation set, and a test set. The training set and validation set are used to train the KNN model, and the trained model is saved. Finally, the test set is used to evaluate the accuracy and generalization ability of the model to achieve rapid fault diagnosis of bearing faults. The process is shown in Fig. 2.

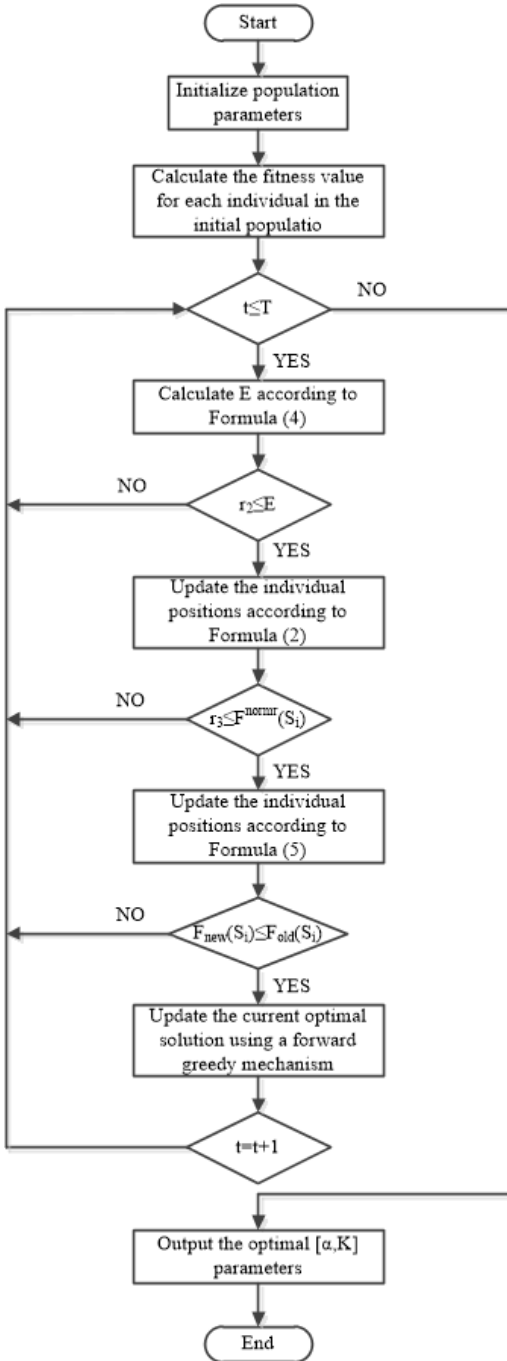


Fig. 1. Flow chart of RIME optimized VMD

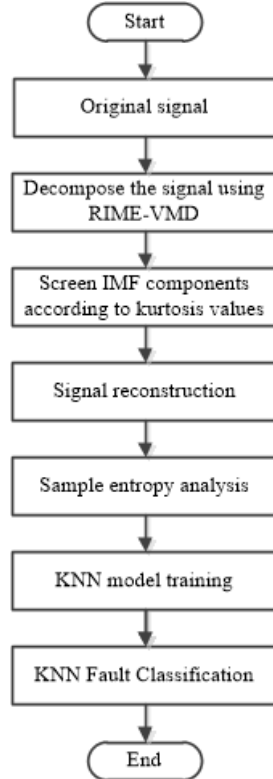


Fig. 2. Diagnostic flow chart

3.3. Fault diagnosis analysis

1) Data collection. The data in this study was sourced from the Electrical Engineering Laboratory of Case Western Reserve University. The SKF6205 deep groove roller bearing was selected as the research object. This bearing was equipped with 9 rollers, and its fault characteristics were simulated by a preset damage with a diameter of 0.1778 mm. The experiment

was driven by a 2-horsepower motor, operating under a stable rotational speed of 1750 r/min. The bearing vibration signals were captured in real time at a high-frequency sampling rate of 12 kHz, forming a time-series sample of 2048 data points for each collection. The study covered four typical states of the bearing: normal operation, inner ring fault, ball fault, and outer ring fault. Through systematic signal collection, a multi-dimensional fault feature database was constructed, providing accurate and reliable data support for subsequent model training and performance verification.

2) Analysis of original time series signals. The time-domain and frequency-domain representations of signals in four different states are shown in Fig. 3 and Fig. 4.

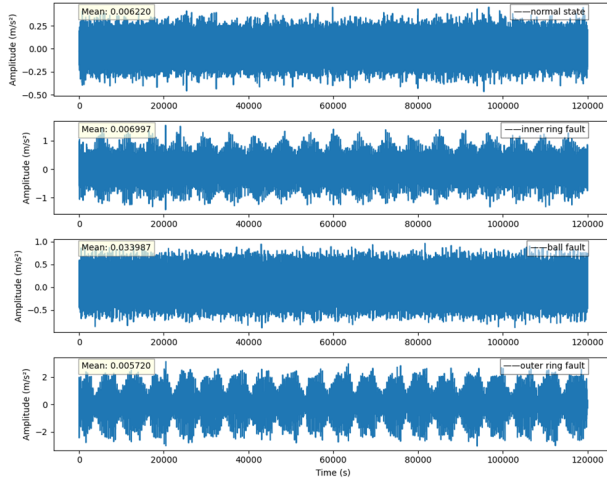


Fig. 3. Original time series of the four states

The figure shows the original time-series signals of rolling bearings in four states (normal, inner ring fault, ball fault, and outer ring fault). In terms of time-series signals, although each state has differences in fluctuation characteristics, the boundaries between different states are blurred and difficult to distinguish intuitively due to the complexity of the signals and noise interference. In the frequency-domain diagrams, the frequency distributions of different fault states overlap, making it difficult to accurately determine the characteristic frequencies. Since it is difficult to effectively distinguish each state based solely on these two types of diagrams, feature extraction methods based on intelligent algorithms, such as optimized decomposition algorithms like RIME-VMD, need to be adopted to deeply excavate the hidden features of the signals, improve the accuracy of judging the operating status and fault types of rolling bearings, and provide more reliable support for fault diagnosis and equipment maintenance.

3) Decomposition and reconstruction of signals. In the field of signal decomposition, the RIME-VMD, GA-VMD, GWO-VMD and AOA-VMD methods were tested and compared, showing different performance due to differences in the setting of their parameters $[K, \alpha]$. The decomposition mode number K controls the fineness of signal deconstruction: a too small value is prone to missing key features, while a too large value causes over-decomposition. The penalty factor α coordinates decomposition accuracy and smoothness, and improper values will lead to signal distortion or detail loss. To accurately explore the influence of parameters, the search interval of K was set to $[2, 10]$, and the range of α was set to $[100, 2500]$. After T iterations of trial and error, each algorithm screened out the optimal parameter strategy. Among them, the RIME algorithm, relying on a unique optimization mechanism, finally determined $(K, \alpha) = (9, 400)$ as the best parameter combination, which can effectively capture the multi-scale features of the signal under this configuration. The GA-VMD method set the parameter combination as $(K, \alpha) = (9, 1600)$ through experience and experiments, balancing the integrity of signal

decomposition and detail retention. The GWO-VMD algorithm found the parameter solution of $(K, \alpha) = (8, 2000)$, achieving a balance between signal smoothness and accuracy with a larger penalty factor α . The AOA-VMD algorithm, leveraging its swarm intelligence optimization characteristics, identified $(K, \alpha) = (8, 1900)$ as the suitable parameter combination, and achieved a balance between decomposition accuracy and detail capture. These parameter settings provide a specific reference for the comparison of signal decomposition effects of the three methods. Taking the inner race fault as an example, a data sample of 500 points was selected, and the signal decomposition is shown in Fig. 5.

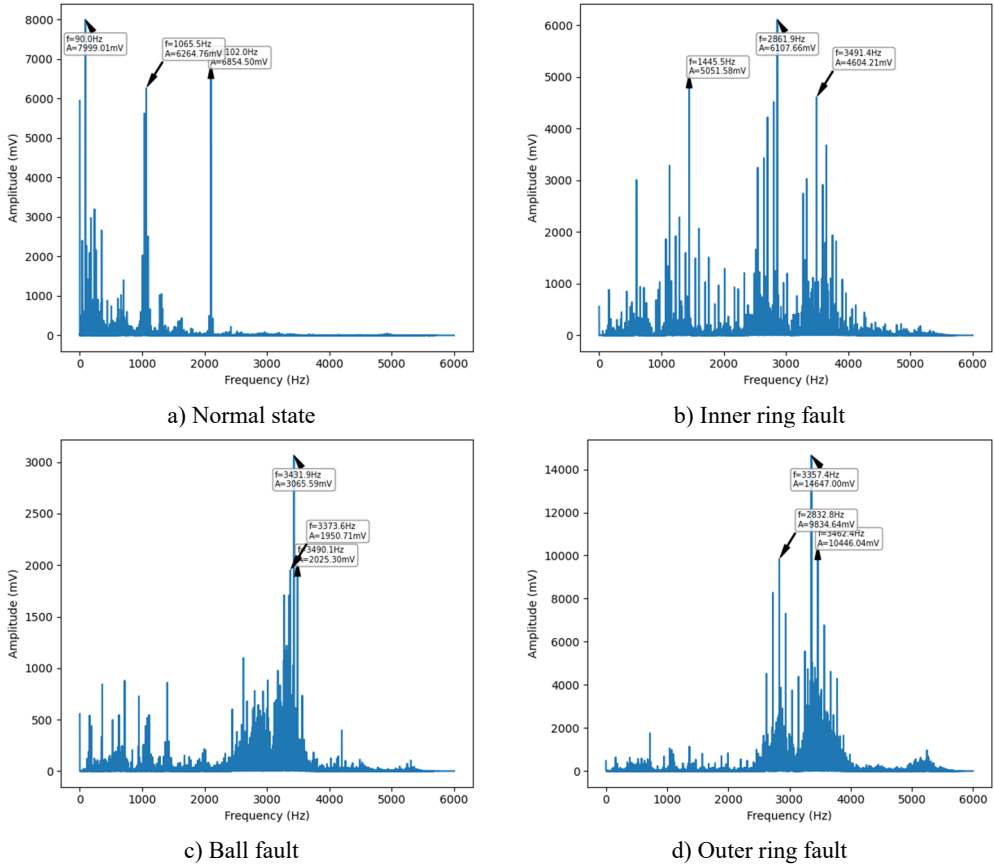


Fig. 4. Frequency-domain display of four states

As shown in Fig. 5, each method decomposes the original signal into multiple sub-signals. These sub-signals exhibit very similar time-domain and frequency-domain characteristics, showing no obvious differences and containing identical or highly overlapping information, which leads to redundancy. Processing a large number of sub-signals increases the computational burden and reduces efficiency. Among them, over-decomposition occurs in the other three methods except RIME-VMD. Over-decomposition may mistake noise for important signal components, thereby amplifying the impact of noise. Therefore, further processing and analysis of the decomposed sub-signals are required to remove redundant information and noise and extract useful features. To further compare the advantages and disadvantages of the three algorithms, an iterative graph analysis can be performed, as shown in Fig. 6.

As shown in Fig. 6, in the initial several iterations, the median values of all algorithms show a significant downward trend, indicating that they are all rapidly converging to the optimal solution.

The RIME-VMD algorithm descends the fastest in the first few iterations, demonstrating the fastest convergence speed. After approximately the 9th iteration, the median values of all algorithms tend to stabilize with minimal variation, suggesting that each algorithm has approached or reached a local optimal solution. At the end of 18 iterations, the RIME-VMD algorithm achieves the lowest median value. Overall, the RIME-VMD algorithm performs best in both convergence speed and final results.

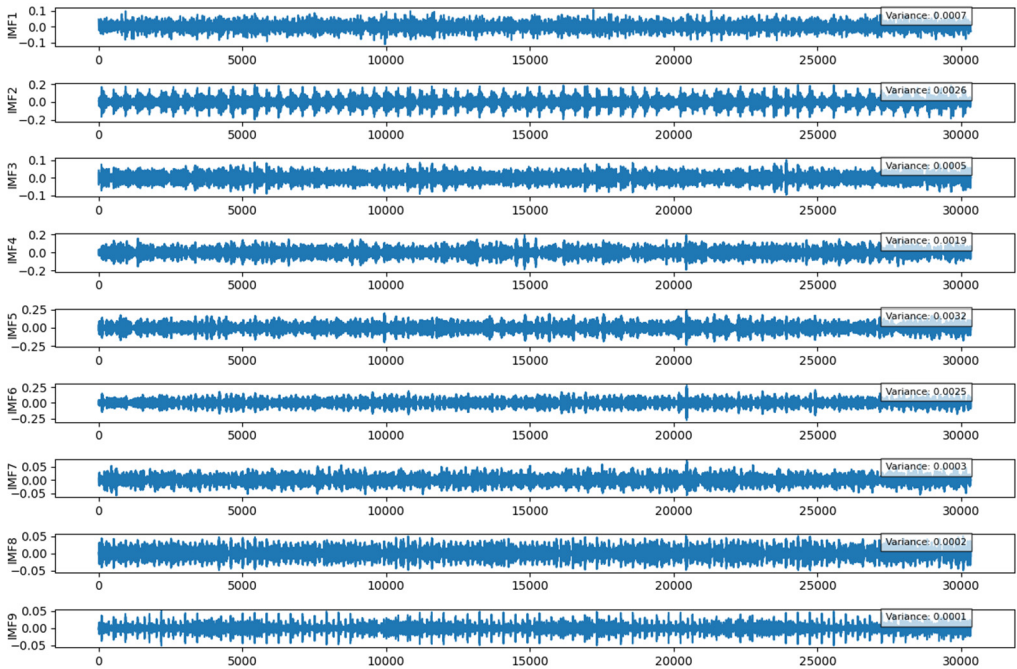


Fig. 5. Signal decomposition

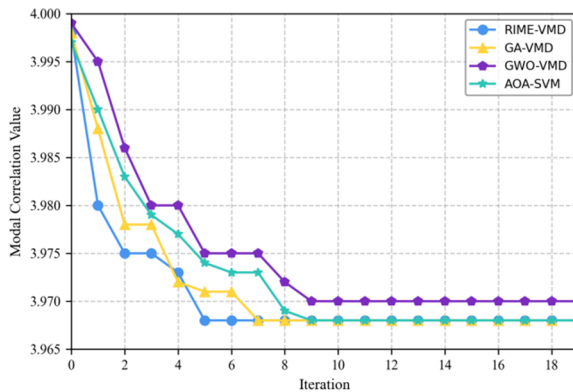


Fig. 6. Iterative diagram for algorithm comparison

The Pearson Correlation Coefficient (PCC) is a statistic used to measure the linear correlation degree between two continuous variables, with its value range between -1 and 1 . The method combining the Pearson Correlation Coefficient (PCC) with the IMF components has demonstrated significant advantages. The PCC value between each IMF component and the fault state is calculated to quantify their linear correlation. The closer the PCC value is to 1 or -1 , the stronger the correlation between the IMF component and the fault state, thereby screening out the key

components reflecting fault characteristics. This method can not only accurately locate the fault source but also effectively separate the noise and fault features in the signal, improving the accuracy of diagnosis. The relationship between the magnitude of the PCC value and the linear relationship between the two variables is shown in Table 1.

Table 1. Linear relationship between PCC values and two variables

PCC value range	Correlation degree	Explanation
1.0	Perfect positive correlation	The two variables have a perfect linear relationship: as one variable increases, the other variable increases accordingly
0.7-0.9	Strong positive correlation	There is a strong positive linear relationship between the two variables
0.4-0.6	Moderate positive correlation	There is a moderate positive linear relationship between the two variables
0.1-0.3	Weak positive correlation	There is a weak positive linear relationship between the two variables
0.0	No correlation	There is no linear relationship between the two variables.
-0.1- -0.3	Weak negative correlation	There is a weak negative linear relationship between the two variables
-0.4 - -0.6	Moderate negative correlation	There is a moderate negative linear relationship between the two variables
-0.7 - -0.9	Strong negative correlation	There is a strong negative linear relationship between the two variables
-1.0	Perfect negative correlation	The two variables have a perfect linear relationship: as one variable increases, the other variable decreases accordingly.

The correlation coefficients of each IMF component are calculated using PCC, and the calculation results are shown in Table 2.

Table 2. Comparison of IMF correlation coefficients

IMF index	RIME-VMD	GA-VMD	GWO-VMD	AOA-VMD
1	0.35	0.33	0.35	0.34
2	0.35	0.32	0.33	0.33
3	0.52	0.5	0.55	0.51
4	0.36	0.35	0.36	0.33
5	0.38	0.35	0.36	0.36
6	0.37	0.35	0.35	0.36
7	0.36	0.34	0.36	0.35
8	0.35	0.34	0.35	0.33
9	0.35	0.32	0.35	0.35

In all states, the PCC coefficients of IMF3 and IMF5 are generally high, indicating that these IMF components contain more fault characteristic information. For each IMF combination, signal reconstruction was performed, followed by classifying signals under different states using the same classification model. Table 3 presents the classification accuracy corresponding to different IMF combinations, from which it could be observed that the combination of IMF3 and IMF5 was more reasonable.

To evaluate the stability of PCC selection under different noise levels, noises of varying intensities (specifically low, medium, and high noise levels) were added to the original signal by setting different noise variances. Table 4 presents the performance of PCC coefficients for each IMF component under different noise levels. It can be observed from the table that the PCC coefficients of IMF3 and IMF5 consistently maintain relatively high values across different scenarios (low, medium, and high noise). This indicates that selecting IMF3 and IMF5 via PCC for signal reconstruction exhibits good stability when confronted with different noise interferences.

Table 3. Classification accuracy corresponding to the same IMF combination

IMF combination	Classification accuracy for normal state (%)	Classification accuracy for inner ring fault state (%)	Classification accuracy for ball fault state (%)	Classification accuracy for outer ring fault state (%)	Average classification accuracy (%)	Performance grade
IMF3+IMF5	94.6	92.8	91.5	93.2	93.0	Excellent
IMF1+IMF4	81.4	77.8	76.3	78.5	78.5	General
IMF2+IMF4	83.7	80.2	79.1	81.5	81.1	Good
IMF2+IMF5	81.9	84.5	83.2	80.6	82.6	Good

Table 4. Performance of PCC coefficients for each IMF component under different noise levels

IMF component / noise level	Low noise	Medium noise	High noise	Stability evaluation
IMF1	0.35	0.20	0.15	The coefficient decreases significantly as noise intensity increases, with large fluctuations and poor stability.
IMF2	0.32	0.25	0.18	The coefficient decreases significantly as noise intensity increases, with large fluctuations and poor stability.
IMF3	0.52	0.50	0.48	The coefficient remains consistently high, with small fluctuations and good stability.
IMF4	0.36	0.28	0.22	The coefficient decreases significantly as noise intensity increases, with large fluctuations and poor stability.
IMF5	0.38	0.37	0.35	The coefficient remains consistently high, with small fluctuations and good stability.

These two components were selected for signal reconstruction to effectively extract and retain key fault characteristic information in the signal, while removing noise and other irrelevant signal components. The reconstructed signal is shown in Fig. 7.

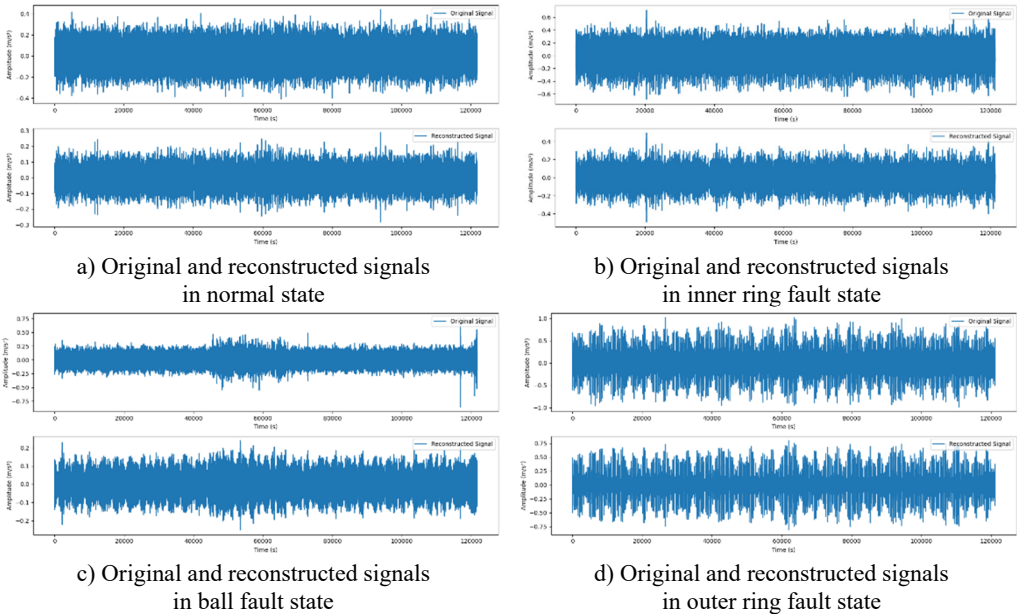


Fig. 7. Reconstructed signals

4) Fault Diagnosis Analysis. Fig. 8 shows the visualization results of multi-scale sample entropy using RIME-VMD, GA-VMD, GWO-VMD and AOA-VMD methods, all of which use t-SNE dimensionality reduction technology to project multi-dimensional data onto a

two-dimensional plane. In Fig. 8(b), although the signal reconstructed by the GA-VMD method can distinguish different states to a certain extent, some categories overlap, leading to unclear classification. The same applies to GWO-VMD in Fig. 8(c) and AOA-VMD in Fig. 8(d). In contrast, the signal reconstructed by the RIME-VMD method in Fig. 8(a) shows more compact distribution and higher separation of points in each category, with significantly reduced overlapping. It can be seen that the RIME-VMD method can more clearly distinguish different states when reconstructing signals, reducing classification uncertainty, which is superior to the GA-VMD, GWO-VMD and AOA-VMD methods. This is of great significance for fields such as fault diagnosis and condition monitoring, as it helps to more accurately identify the status of equipment and potential problems.

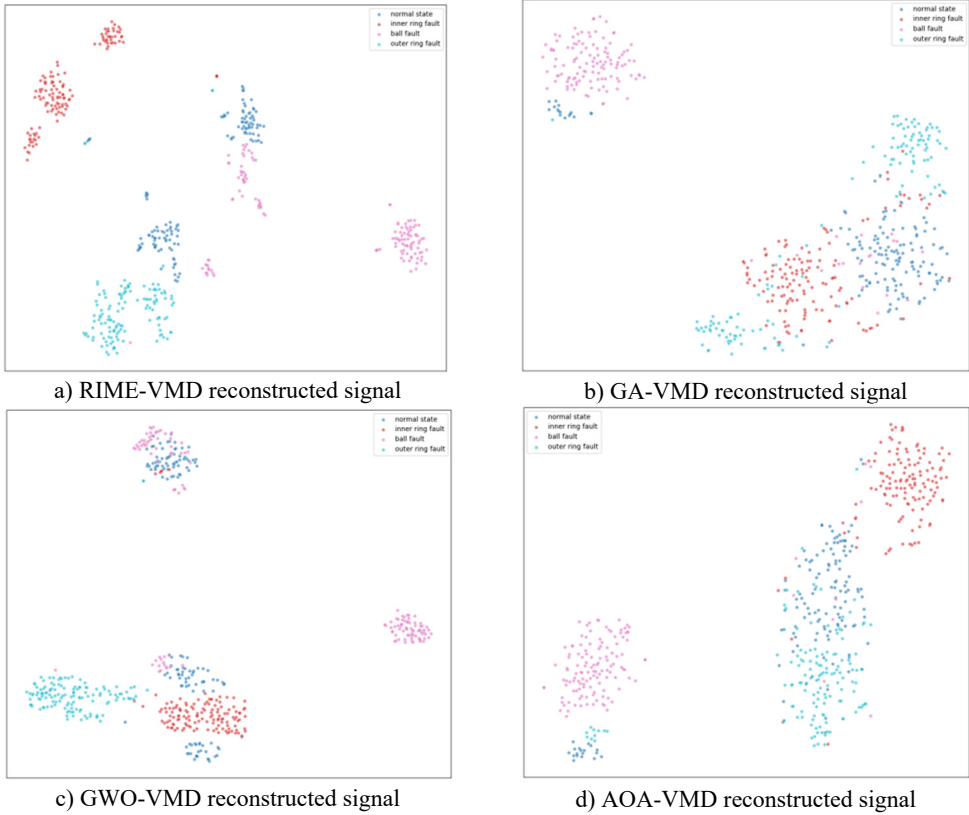


Fig. 8. Visualization of multi-scale sample entropy

To further measure the accuracy of several methods, 100 time windows were selected for each modal component of the four states to extract multi-scale sample entropy, resulting in 100×20 training samples. Finally, 400×20 samples of the four states were labeled and used as inputs to the KNN classifier. The results are shown in Fig. 9.

It can be seen from Fig. 9 that the GA-VMD, GWO-VMD and AOA-VMD methods can predict fault categories well for most categories, but there are certain deviations between the predicted values and the true values for individual categories. In contrast, the identification results of RIME-VMD show higher overall accuracy and stability. Although there are also some deviations in the results of RIME-VMD, these deviations are smaller than those of the other three methods, and the distribution is more concentrated. Especially for some key categories, the predicted values of RIME-VMD almost completely coincide with the true values, demonstrating its superior performance in fault identification.

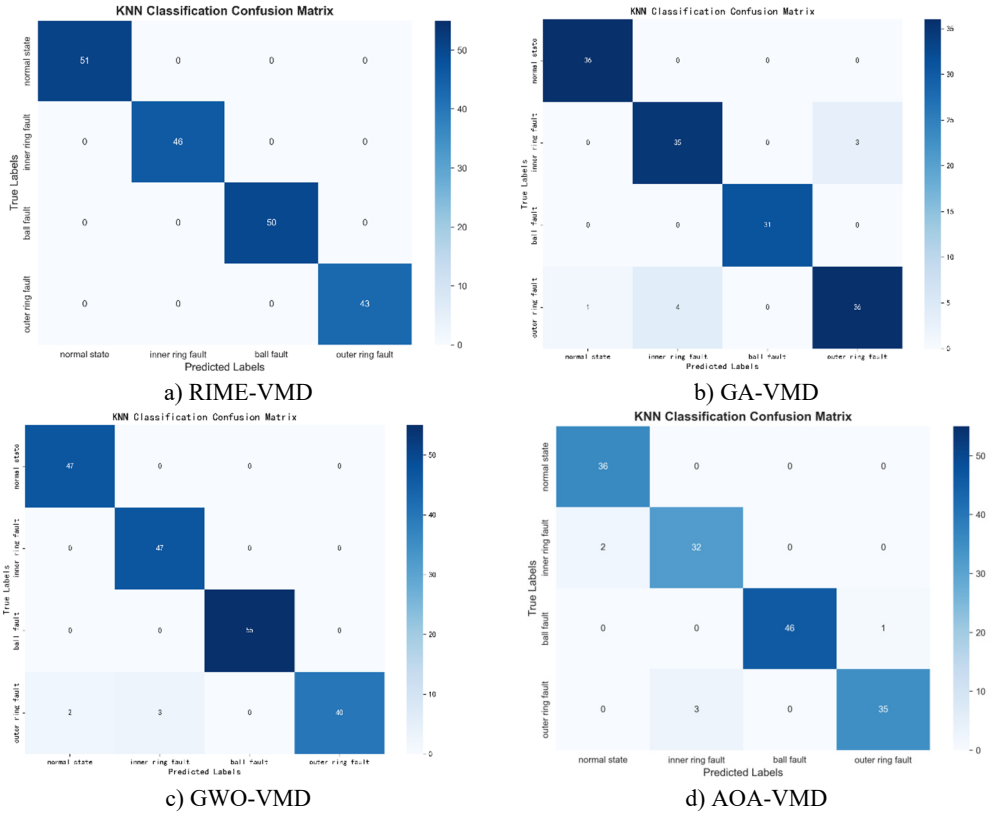


Fig. 9. KNN classification confusion matrix

To evaluate the stability and significance of the RIME-VMD method, validation was conducted using a “multiple random partitioning + cross-validation” approach, with the following steps:

(1) Data partitioning: Divide all data into five equal parts.

(2) Cyclic Evaluation:

– Use the first sample as the test set and the remaining four samples as the training set, then calculate the metric once.

– Use the second sample as the test set and the remaining four samples as the training set, then recalculate the metrics.

– Repeat the above steps until all 5 portions have served as the test set, yielding 5 sets of metrics.

(3) Statistical Results: The calculated averages are summarized in tabular form. See Table 5 for details.

In summary, we can conclude that the RIME-VMD method is superior to the GA-VMD, GWO-VMD and AOA-VMD methods in fault identification. Especially when dealing with complex and changeable fault data, RIME-VMD can provide more accurate and reliable prediction results. This finding is of great significance for practical engineering applications and helps to improve the efficiency of equipment fault diagnosis and maintenance.

To further verify the method proposed in this paper, a Data Combination test was conducted. The bearing load was set to 2 horsepower, and the damage diameters were 0.1778 mm, 0.3556 mm, 0.5334 mm, and 0.7112 mm respectively; tests and analyses were carried out for these four operating states. Each Intrinsic Mode Function (IMF) under each state was constructed into a 100×20 training sample. Finally, a total of 400×20 samples from the four states were labeled

and input into SVM and KNN classifiers respectively. As can be seen from Table 6, the test results show that there were a small number of sample misclassifications when using SVM, while the accuracy rate of the KNN comprehensive judgment method reached 100 %.

Table 5. Performance comparison

Evaluation dimension	Performance indicator (%)	RIME-VMD	GA-VMD	GWO-VMD	AOA-VMD	Criteria for judging the merits of indicators
Basic classification ability	Accuracy	98.6	92.3	94.8	93.5	The higher, the better (reflecting the overall correctness of prediction)
Precision of positive class recognition	Precision	97.9	89.5	91.2	90.7	The higher, the better (to reduce the misjudgment of normal cases as faults)
Completeness of positive class coverage	Recall	98.2	90.1	93.5	92.1	The higher, the better (to reduce the missed judgment of faults)
Ability to balance precision and coverage	F1-Score	98.0	89.8	92.3	91.4	The higher, the better (integrating Precision and Recall)

Table 6. Comparison of classification results

		Multi-scale MSE		Single scale SE	
		KNN	SVM	KNN	SVM
VMD	IMF1	100 %	99 %	100 %	99 %
	IMF2	100 %	100 %		
	IMF3	100 %	100 %		
	IMF4	100 %	99 %		
	IMF5	100 %	100 %		

3.4. Verification of real-time performance

To verify the real-time performance of the proposed method, experiments were designed from three aspects: processing delay, computational efficiency, and adaptability to embedded/edge computing. The experimental platforms and parameter settings are as follows:

Experimental Platforms: General-purpose computing platform: Intel i7-12700H CPU, 16GB DDR4 memory, Windows 11 operating system; Edge computing platform: Raspberry Pi 4B (4-core ARM Cortex-A72 CPU, 4GB LPDDR4 memory).

Experimental Parameters: The bearing fault dataset from the “data combination” was adopted, which has a load of 2 horsepower and a damage diameter range of 0.1778-0.7112 mm; Sample size: 400×20.

Each platform was run independently for 10 experiments, and the average value and standard deviation were calculated for final results.

1) Analysis of processing delay. Table 7 presents the comparison of processing delays across different platforms.

Table 7. Comparison of processing latency

Platform type	Method	Sample size	Average latency (ms)	Standard deviation (ms)
General computing platform	RIME-VMD	400×20	18.3	1.5
General computing platform	GA-VMD	400×20	45.8	3.5
General computing platform	GWO-VMD	400×20	38.2	2.8
General computing platform	AOA-VMD	400×20	28.5	2.1
Edge computing platform	RIME-VMD	400×20	69.5	3.2
Edge computing platform	GA-VMD	400×20	153.7	6.8
Edge computing platform	GWO-VMD	400×20	126.3	5.6
Edge computing platform	AOA-VMD	400×20	93.6	4.3

As can be seen from the table, on the general-purpose computing platform, the average delay of RIME-VMD in processing 400×20 samples is 18.3 ms with a standard deviation of 1.5 ms, both of which are lower than the corresponding values of the other three methods. This indicates that the proposed method has better real-time performance on general-purpose devices.

On the edge computing platform, the average delay of RIME-VMD is 69.5 ms, which is not only lower than the requirement threshold of “delay < 100 ms” for industrial real-time diagnosis but also lower than that of the other three methods. This proves that the proposed method still has real-time processing capability on edge devices.

2) Analysis of Computational Efficiency. The computational efficiency is analyzed from two aspects: “Samples Processed Per Second (SPS)” and “Resource Utilization”, with specific details shown in Table 8.

Table 8. Analysis of computational efficiency

Method	Sample set specification	Samples per second (SPS)	Average CPU utilization (%)	Peak memory usage (MB)
RIME-VMD	100×20	5.7	38	46
GA-VMD	100×20	3.1	63	89
GWO-VMD	100×20	3.6	56	76
AOA-VMD	100×20	4.1	48	62

Efficiency: On the Raspberry Pi 4B, the SPS (Samples Processed Per Second) of the proposed method is 5.7 (i.e., processing 5.7 sets of 100×20 samples per second), which is 1.8 times that of GA-VMD. This indicates that the proposed method has higher throughput and can cope with scenarios with higher-frequency data input.

Resource Occupancy: During processing, the proposed method has an average CPU utilization rate of 38 % and a peak memory occupancy of 46 MB. In contrast, GA-VMD has a CPU utilization rate of 63 % and a memory occupancy of 89 MB. This proves that the proposed method has lower hardware resource requirements and is more suitable for resource-constrained environments.

3) Feasibility Conclusion for Embedded/Edge Computing. Based on the above data, RIME-VMD can operate stably on edge devices, with the following key performance indicators meeting application requirements: The processing delay satisfies industrial real-time demands (i.e., < 100 ms); Its resource occupancy is lower than that of the other three methods, requiring no high-performance hardware support and thus being adaptable to low-cost embedded devices.

Therefore, RIME-VMD is feasible for embedded/edge computing scenarios and can be applied to on-site real-time fault diagnosis.

3.5. Verification of generalization ability

To comprehensively evaluate the generalization ability of the model, on the basis of verification using the CWRU dataset, the XJTU-SY dataset (Xi'an Jiaotong University - Shaoxing University of Arts and Sciences Bearing Dataset) was supplemented. This dataset covers “full-life cycle faults” (ranging from normal conditions to early-stage, mid-stage, and late-stage faults) and includes operating conditions with variable rotational speeds (1000-3000 r/min) and variable loads (0-3000 N). In contrast to the fixed operating conditions of the CWRU dataset, it is used to test the model's generalization ability under varying operating conditions.

After the simulation model is trained in the benchmark scenario, it is migrated to a similar but distinct laboratory scenario to verify its generalization ability across different operating conditions. The normal/single-fault samples from the CWRU dataset are used as the training set (divided according to the original proportion, with 70 % allocated for training), while the full-life cycle fault samples from the XJTU-SY dataset serve as the test set (100 % used for testing, with no training data included). The accuracy of the original study, which was based on training with the CWRU dataset and subsequent testing with the same dataset (CWRU training → CWRU testing), is retained as the benchmark. Additionally, the accuracies of GA-VMD, GWO-VMD,

and AOA-VMD under the same cross-dataset testing are newly incorporated. The comparison of cross-dataset testing accuracies is presented in Table 9.

Table 9. Comparison of cross-dataset testing accuracy

Model	Training set	CWRU test set (Baseline)	XJTU-SY test set (variable operating conditions)
RIME-VMD	CWRU	99.2 %	85.3 %
RIME-VMD	CWRU+XJTU	99.5 %	91.7 %
GA-VMD	CWRU+XJTU	98.8 %	82.5 %
GWO-VMD	CWRU+XJTU	98.5 %	82.1 %
AOA-VMD	CWRU+XJTU	97.6 %	81.3 %

As can be seen from the table, when trained solely on the CWRU dataset, the model achieves an accuracy of only 85.3 % on the variable operating condition dataset. This is because the faults in the CWRU dataset occur under fixed operating conditions, while those in the XJTU-SY dataset occur under variable operating conditions – this discrepancy results in the model having weak ability to identify faults under variable operating conditions. After training with multi-source data, the accuracy on the variable operating condition dataset increases to 91.7 %. This indicates that by combining data from the CWRU and XJTU-SY datasets, the model learns more generalized fault features and reduces its reliance on the distribution of a single dataset. It can also be observed from the table that the RIME-VMD model has advantages over other models: its accuracy in composite fault testing is higher than that of other models. The reason lies in the attention module of this model, which can focus on the multi-fault features in composite faults, whereas other models are easily interfered with by redundant features. This thus verifies the generalization ability of RIME-VMD under varying operating conditions.

4. Conclusions

To tackle the issues of subjective parameter selection and insufficient decomposition accuracy in complex noise for traditional VMD in rolling bearing fault diagnosis, this paper proposed a method integrating the RIME optimization algorithm, VMD, and KNN. The RIME algorithm optimizes VMD's core parameters (K and α) globally, breaking the reliance on experience to decompose vibration signals accurately under complex noise. PCC then screens key IMF components for signal reconstruction, preserving core fault features. Finally, sample entropy of the reconstructed signal is input as a feature vector into the KNN classifier for fault identification. Multi-dimensional experiments show RIME outperforms GA, GWO, and AOA in parameter optimization with faster convergence and higher accuracy. The method has good generalization on both test bench and XJTU-SY datasets, adapting to different equipment and conditions. KNN is simpler and more efficient than SVM, meeting engineering needs. Multiple cross-validations and confusion matrices confirm its stability, while computation time and resource data verify deployment feasibility. Future research will explore its universality in multi-condition/multi-type bearing faults and combine edge computing to enhance online processing, advancing intelligent fault diagnosis.

Acknowledgements

The authors have not disclosed any funding.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Shijun Yu: conceptualization, software, writing-original draft preparation. Changyou Guo: project administration, software. Haorui Liu: methodology, validation. Hengwei Zhu: investigation, supervision.

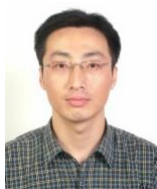
Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] L. Yuan, D. Lian, X. Kang, Y. Chen, and K. Zhai, "Rolling bearing fault diagnosis based on convolutional neural network and support vector machine," *IEEE Access*, Vol. 8, pp. 137395–137406, Jan. 2020, <https://doi.org/10.1109/access.2020.3012053>
- [2] J. Yan, J. Kan, and H. Luo, "Rolling bearing fault diagnosis based on Markov transition field and residual network," *Sensors*, Vol. 22, No. 10, p. 3936, May 2022, <https://doi.org/10.3390/s22103936>
- [3] J. Zhou, M. Xiao, Y. Niu, and G. Ji, "Rolling bearing fault diagnosis based on WGWAO-VMD-SVM," *Sensors*, Vol. 22, No. 16, p. 6281, Aug. 2022, <https://doi.org/10.3390/s22166281>
- [4] Y. Wang, S. Zhang, R. Cao, D. Xu, and Y. Fan, "A rolling bearing fault diagnosis method based on the WOA-VMD and the GAT," *Entropy*, Vol. 25, No. 6, p. 889, Jun. 2023, <https://doi.org/10.3390/e25060889>
- [5] Z. Wang, L. Yao, G. Chen, and J. Ding, "Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals," *ISA Transactions*, Vol. 114, pp. 470–484, Aug. 2021, <https://doi.org/10.1016/j.isatra.2020.12.054>
- [6] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, Vol. 66, No. 8, pp. 1926–1935, Aug. 2017, <https://doi.org/10.1109/tim.2017.2674738>
- [7] P. Casoli, M. Pastori, F. Scolari, and M. Rundo, "A vibration signal-based method for fault identification and classification in hydraulic axial piston pumps," *Energies*, Vol. 12, No. 5, p. 953, Mar. 2019, <https://doi.org/10.3390/en12050953>
- [8] K. Zhang, Y. Xu, Z. Liao, L. Song, and P. Chen, "A novel fast entrogram and its applications in rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, Vol. 154, p. 107582, Jun. 2021, <https://doi.org/10.1016/j.ymssp.2020.107582>
- [9] S. Tang, Y. Zhu, and S. Yuan, "A novel adaptive convolutional neural network for fault diagnosis of hydraulic piston pump with acoustic images," *Advanced Engineering Informatics*, Vol. 52, p. 101554, Apr. 2022, <https://doi.org/10.1016/j.aei.2022.101554>
- [10] Q. Lu, X. Shen, X. Wang, M. Li, J. Li, and M. Zhang, "Fault diagnosis of rolling bearing based on improved VMD and KNN," *Mathematical Problems in Engineering*, Vol. 2021, pp. 1–11, Oct. 2021, <https://doi.org/10.1155/2021/2530315>
- [11] Q. Zhenya and Z. Xueliang, "Rolling bearing fault diagnosis based on CS-optimized multiscale dispersion entropy and ML-KNN," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, Vol. 44, No. 9, Aug. 2022, <https://doi.org/10.1007/s40430-022-03643-3>
- [12] W. Chen, M. Yu, and M. Fang, "Research on identification and localization of rotor-stator rubbing faults based on AF-VMD-KNN," *Journal of Vibration Engineering and Technologies*, Vol. 9, No. 8, pp. 2213–2228, Aug. 2021, <https://doi.org/10.1007/s42417-021-00357-z>
- [13] G. Liu, Y. Ma, and N. Wang, "Rolling bearing fault diagnosis based on SABO-VMD and WMH-KNN," *Sensors*, Vol. 24, No. 15, p. 5003, Aug. 2024, <https://doi.org/10.3390/s24155003>
- [14] M. Ye, X. Yan, and M. Jia, "Rolling bearing fault diagnosis based on VMD-MPE and PSO-SVM," *Entropy*, Vol. 23, No. 6, p. 762, Jun. 2021, <https://doi.org/10.3390/e23060762>
- [15] F. Zhou, X. Yang, J. Shen, and W. Liu, "Fault diagnosis of hydraulic pumps using PSO-VMD and refined composite multiscale fluctuation dispersion entropy," *Shock and Vibration*, Vol. 2020, No. 1, pp. 1–13, Aug. 2020, <https://doi.org/10.1155/2020/8840676>
- [16] Z. Guo, M. Liu, Y. Wang, and H. Qin, "A new fault diagnosis classifier for rolling bearing united multi-scale permutation entropy optimize VMD and Cuckoo Search SVM," *IEEE Access*, Vol. 8, pp. 153610–153629, Jan. 2020, <https://doi.org/10.1109/access.2020.3018320>

- [17] J. Zhang, Y. Zhao, X. Li, and M. Liu, "Bearing fault diagnosis with kernel sparse representation classification based on adaptive local iterative filtering-enhanced multiscale entropy features," *Mathematical Problems in Engineering*, Vol. 2019, No. 1, Jun. 2019, <https://doi.org/10.1155/2019/7905674>
- [18] W. Zaman, Z. Ahmad, M. F. Siddique, N. Ullah, and J.-M. Kim, "centrifugal pump fault diagnosis based on a novel SobelEdge scalogram and CNN," *Sensors*, Vol. 23, No. 11, p. 5255, Jun. 2023, <https://doi.org/10.3390/s23115255>
- [19] V. Sinitsin, O. Ibryaeva, V. Sakovskaya, and V. Eremeeva, "Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model," *Mechanical Systems and Signal Processing*, Vol. 180, No. 15, p. 109454, Nov. 2022, <https://doi.org/10.1016/j.ymssp.2022.109454>
- [20] L. Shi, W. Liu, D. You, and S. Yang, "Rolling bearing fault diagnosis based on CEEMDAN and CNN-SVM," *Applied Sciences*, Vol. 14, No. 13, p. 5847, Jul. 2024, <https://doi.org/10.3390/app14135847>
- [21] P. Zhu, S. Huang, L. Liu, and Z. Zhang, "RIME-SVM for electromagnetic radiation signal identification in coal mining environment," in *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Vol. 88, pp. 1342–1347, Jul. 2024, <https://doi.org/10.1109/icpics62053.2024.10797089>
- [22] H. Song, H. Sun, and N. Chen, "Cavitation fault diagnosis of centrifugal pump based on RIME-SDAE," in *Vibroengineering Procedia*, Vol. 54, pp. 46–52, Apr. 2024, <https://doi.org/10.21595/vp.2024.24039>
- [23] Y. Shi, Y. Yang, Y. Ruan, T. Zhang, M. Lin, and Z. Luo, "Optimization of VMD wavelet threshold based on rime optimization algorithm for denoising vibration signals of converter on load tap changer," in *2024 IEEE 2nd International Conference on Power Science and Technology (ICPST)*, pp. 648–653, May 2024, <https://doi.org/10.1109/icpst61417.2024.10602301>



Shijun Yu received Master of Engineering degree in Harbin University of Science and Technology, Harbin, China, in 2008. Now he works at Dezhou University. His current research interests include engineering vibration and application, deep learning and machine learning, etc.



Changyou Guo received Ph.D. degree in University of Science and Technology Beijing, Beijing, China, in 2016. Now he works at Dezhou University. His current research interests include engineering vibration and application, deep learning and machine learning, etc.



Haorui Liu received Ph.D. degree in China University of Mining and Technology, Beijing, China, in 2010. Now he works at Dezhou University. His current research interests include engineering vibration and application, signal processing, etc.



Hengwei Zhu received Master of Engineering degree in Xihua University, Chengdu, China, in 2014. Now he works at Dezhou University. His current research interests include engineering vibration and application, deep learning and machine learning, etc.