Current Science
CS 5 (6), 5614-5633 (2025)
Received 2 November 2025 | **Accepted 20 November** 2025 | **Online Available 30 December** 2025
https://doi.org/10.52845/CS/2025-5-6-26

Current Science
Current Science

OPEN ACCESS
ISSN (O) 2795-8639

Original Article

A Fault Identification Method for Rotating Machinery Based on CEEMDAN-MDE-PCC-EWT and RCMDE

Liu Yan-Bo[1], Luo Le[2], Liu Yu[3]

College of Mechanical Engineering, Chongqing University of Technology, Chongqing, 400054 CHINA

***Corresponding Author: Liu Yu**

Abstract:
Addressing the challenge of severe noise interference in rotating machinery, which impedes effective fault feature extraction from nonlinear and non-stationary vibration signals, this paper proposes a novel fault identification method based on secondary decomposition and Refined Composite Multiscale Dispersion Entropy (RCMDE). First, the original vibration signal is decomposed using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to obtain Intrinsic Mode Functions (IMFs). Mixed evaluation indices, combining Multiscale Dispersion Entropy (MDE) and Pearson Correlation Coefficient (PCC), are then constructed to separate noise-dominant from signal-dominant IMFs. Subsequently, Empirical Wavelet Transform (EWT), with parameters optimized via the Grey Wolf Optimization (GWO) algorithm, is applied for signal filtering and reconstruction. Finally, RCMDE features are extracted from the reconstructed signal and input into a Support Vector Machine (SVM) for fault classification. Experimental results demonstrate the method's excellent performance in both simulation studies and tests on public bearing datasets. Notably, for the highly noisy planetary gear fault diagnosis task, the method significantly enhances the signal-to-noise ratio (SNR) by at least 3.14 dB and achieves a fault classification accuracy of 98.61%, outperforming benchmark methods. This research provides valuable insights for rotating machinery fault diagnosis.

Index Terms: Rotating Machinery, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Refined Composite Multiscale Dispersion Entropy (RCMDE), Empirical Wavelet Transform (EWT

1. Introduction

Rotating machinery is ubiquitous in industrial applications, where its safe operation is critical for ensuring equipment reliability and minimizing economic losses 1. This is particularly true for high-end equipment in technology-intensive sectors such as civil aviation and wind power, where reliability and efficiency are paramount 2. Consequently, deep mining of abnormal information from operational data has become essential for effective equipment fault diagnosis and ensuring the efficient performance of rotating machinery. In today's era of globalization and rapid technological advancement, researchers are increasingly focusing on fault prediction and health management strategies for these critical assets 34. Notably, nonlinear dynamics methods, particularly those employing multivariate multiscale entropy, have demonstrated significant effectiveness in quantifying fault characteristics using multi-sensor signals 5.

Traditional fault diagnosis methods primarily rely on classical signal processing and pattern recognition techniques. However, vibration signals from rotating machinery are often nonlinear, non-stationary, and contaminated by strong noise, making it difficult for conventional time-domain and frequency-domain analyses to extract effective fault features. Time-

frequency analysis offers distinct advantages in this context 6. This approach decomposes complex signals into multiple simpler components using time series decomposition. By subsequently combining the classification predictions from each decomposed layer, it enhances overall classification accuracy 7. To address specific challenges, researchers have developed various decomposition techniques. Zhou et al. 8 mitigated information redundancy by fusing features using Empirical Mode Decomposition (EMD) information entropy and employing Random Forest (RF) to rank feature significance. The Ensemble Empirical Mode Decomposition (EEMD) method was introduced to overcome EMD's mode aliasing issue by incorporating auxiliary noise. Huang et al. 9 proposed an EEMD-based strategy for diagnosing misfire faults in ten-cylinder diesel engines. They decomposed instantaneous crankshaft speed signals into Intrinsic Mode Functions (IMFs) and detected abnormal amplitude fluctuations within these IMFs. Addressing the impact of complex equipment structures and transmission paths on rolling bearing fault signals, Zhu et al. 10 developed a Multi-Signal Improved Empirical Fourier Decomposition (MS-IEFD) method for feature extraction. For challenges like weak signal features due to underwater acoustic noise, large intra-class variations, and high inter-class similarity in modulation recognition, Zhou et al. **Error! Reference source not found.** applied a denoising method combining Variational Mode Decomposition (VMD) and wavelet analysis, preserving effective information in low-correlation mode components. However, a critical limitation of these signal decomposition approaches is their strong dependence on predefined parameters. The rationality of parameter selection directly impacts the diagnostic accuracy and stability of the resulting models.

Although EEMD significantly mitigates the mode aliasing inherent in EMD, residual white noise often persists within the decomposed mode components, and issues like mode indeterminacy remain. To address these limitations, Torres et al. 12 proposed the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithm. Researchers have subsequently leveraged CEEMDAN in various applications. Wang et al. 13 developed a hybrid CEEMDAN-CNN-BiLSTM model for predicting mining area surface subsidence. This model excels at capturing spatiotemporal features and exhibits strong generalization, providing a more precise and reliable solution for short-term GNSS time series prediction. Addressing the challenge of extracting bearing fault frequency features in strong noise environments, Wu et al. 14 integrated CEEMDAN with Wavelet Threshold Denoising (WTD). This combined approach outperforms traditional single denoising methods in reducing random noise and isolating fault characteristic frequencies. For the problem of complex noise interference during on-site acoustic monitoring of high-voltage shunt reactors, Wang et al. 15 introduced a collaborative denoising method combining improved CEEMDAN with Singular Value Decomposition (SVD). The denoising effectiveness of this method surpasses that of Wiener filtering, wavelet packet denoising, VMD, and other single denoising algorithms.

Signal processing advancements have established se condary decomposition as an effective technique for multi-level, multi-scale signal decomposition to extr act multi-source information features, demonstrating significant application value in this domain **Error! Reference source not found.**. To address the challe nge of severe noise interference in rotating machiner y vibration signals—characterized by complex nonli nearity and non-stationarity, which impedes effective fault information extraction—this paper proposes a n ovel fault identification method based on secondary decomposition and Refined Compound Multi-scale Dispersion Entropy (RCMDE). Specifically, the met hod proceeds as follows: 1) The original vibration si gnal is decomposed using Complete Ensemble Empi rical Mode Decomposition with Adaptive Noise (CE EMDAN). 2) The resulting Intrinsic Mode Function

(IMF) components are clustered and screened utilizing Multiscale Dispersion Entropy (MDE). 3) A hybrid index based on MDE is constructed to classify the IMF components. 4) Empirical Wavelet Transform (EWT) filtering is applied to different component types, with its parameters optimized by the Grey Wolf Optimization (GWO) algorithm for effective noise removal. 5) RCMDE features are extracted from the reconstructed signal to form the feature set. 6) Fault classification and identification are performed using a Support Vector Machine (SVM) classifier.

I. Theory and Method

A. MDE

In recent years, complexity theory rooted in information entropy has been widely applied to distinguish noise from signal components after decomposition and to represent signal features. This approach reveals hidden patterns in time series while quantitatively characterizing their irregularity. Shang et al. 17 applied CEEMD to decompose original data, using Permutation Entropy (PE) thresholds for component separation. Compared to alternative methods, entropy-based nonlinear dynamic complexity metrics offer significant advantages: strong noise robustness, high classification accuracy, low data volume requirements, and independence from prior knowledge. Building on these concepts, Rostaghi and Azami (2016) introduced Dispersion Entropy (DE) – a novel nonlinear dynamic measure for irregularity quantification 18.

Consider an initial time series $\{x_i, i = 1,2,\ldots,N\}$ representing a discrete signal of length *N*. The core principles of dispersion entropy algorithms operate as follows.

(1) Normalize the signal using the standard normal distribution function, yielding the mapped sequence $y = \{y_i, i = 1,2,\ldots,N\}$

$$y_i = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x_i} e^{\frac{-(t-\mu)^2}{2\sigma^2}}dt \# (1)$$

where μ and σ^2 denote the mean and variance of the original sequence *x*, respectively. By the properties of the standard normal cumulative distribution function, $y_i \in (0,1)$.

(2) Apply linear scaling to transform the sequence *y* into the range [1, 2, ..., c]

$$z_i^c = R(cy_i + 0.5)\# (2)$$

where *c* equals the number of categories, and $R(\cdot)$ denotes the rounding function.

(3) Perform phase space reconstruction on the initial sequence *x* to generate the matrix:

$$\begin{pmatrix} z_1 & z_1+d & \ldots & z_1+(m-1)d \\ z_2 & z_2+d & \ldots & z_2+(m-1)d \\ & & \ldots & \\ z_r & z_r+d & \ldots & z_r+(m-1)d \\ & & \ldots & \\ z_L & z_L+d & \ldots & z_L+(m-1)d \end{pmatrix} \# (3)$$

where *m* represents the embedding dimension, *d* the delay time, and $L = N - m + 1$ the row count of the matrix. Each row corresponds to an embedded vector $z_i^{m,c}$, namely.

$$z_i^{m,c} = \{z_i^c, z_{i+d}^c, \dots, z_{i+(m-1)d}^c\} \# (4)$$

where $i = 1,2,\dots,N-(m-1)d$.

(4) Calculate the dispersion pattern $\pi_{v0,v1,\dots,vm-1}(v = 1,2,\dots,c)$.If $z_i^c = v_0, z_{i+d}^c = v_1, \dots, z_{i+(m-1)d}^c = v_{m-1}$, then the dispersion pattern corresponding to $z_i^{m,c}$ is $\pi_{v0,v1,\dots,vm-1}$, and there are a total of c^m kinds of corresponding dispersion patterns.

(5) Calculate the probability $p(\pi_{v0,v1,\dots,vm-1})$ corresponding to each dispersion pattern, namely

$$p(\pi_{v0,v1,\dots,vm-1}) = \frac{C(\pi_{v0,v1,\dots,vm-1})}{N-(m-1)d} \# (5)$$

where $C(\pi_{v0,v1,\dots,vm-1})$ refers to the number of the dispersion patterns of type $\pi_{v0,v1,\dots,vm-1}$.

(6) The permutation entropy DE of the original signal x is defined as:

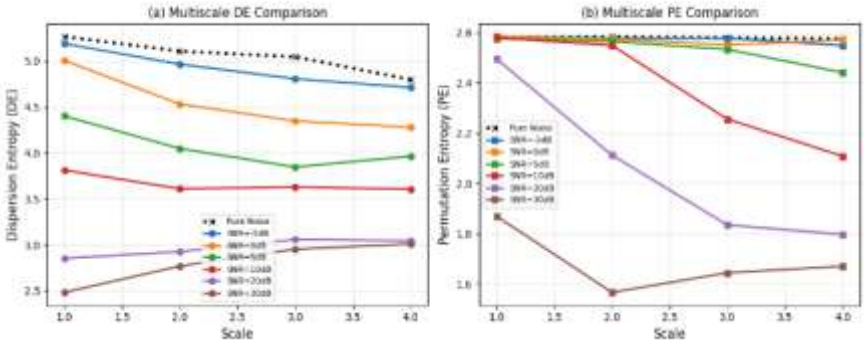
$$DE(X,m,c,d) = -\sum_{\pi=1}^{c^m} p(\pi_{v0,v1,\dots,vm-1}) \, ln[p(\pi_{v0,v1,\dots,v\,m-1})] \# \quad (6)$$

Dispersion Entropy (DE) retains the ordinal pattern characteristics of Permutation Entropy (PE) while additionally incorporating signal amplitude correlation information. This integration significantly enhances computational efficiency. DE also demonstrates robustness against minor fluctuations in adjacent signal amplitudes, ensuring stable category determination. Therefore, to more effectively capture the signal's complete information, this paper adopts Multi-scale Dispersion Entropy (MDE) as the criterion for evaluating IMF components.

Multi-scale Dispersion Entropy (MDE) quantifies time series complexity across multiple scales. The coarse-grained sequence derivation follows:

$$y_i(\tau) = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i \# (7)$$

Consider the original signal $x(t) = sin(20\pi t) + sin(40\pi\, t)$. Gaussian white noise of varying intensities was introduced to this base s ignal. Simulation results, as shown in Figure 1, s how the entropy curve of pure Gaussian noise as a black dashed line. Beyond 10 dB SNR, permutati on entropy (PE) converges with the reference cur ve, exhibiting negligible deviation. In contrast, dis persion entropy (DE) maintains superior discrimi native capability throughout the noise range.



(a) DE (b) PE

Figure 1 Multiscale entropy analysis comparison: Dispersion entropy and permutation entropy under different signal-to-noise ratios.

B. RCMDE

RCMDE constitutes an enhanced methodology for multiscale nonlinear feature extraction 19. This technique combines compound multiscale analysis with dispersion entropy to mitigate information loss inherent in traditional multiscale entropy's coarse-graining process, making it particularly suitable for characterizing non-stationary mechanical vibration signals. Consequently, RCMDE is selected for signal representation and is formally defined as:

$$RCMDE(X,m,c,d,\tau) = -\sum_{\pi=1}^{c^m} \overline{p}\left(\pi_{v0,v1,\ldots,vm-1}\right) ln\left[\overline{p}\left(\pi_{v0,v1,\ldots,vm-1}\right)\right] \# \ (8)$$

where $\overline{p}(\pi_{v0v1\ldots vm-1}) = \frac{1}{\tau}\sum_1^\tau p_k^\tau$represents the average probability of the dispersion pattern x_k^τ of the coarsened sequence.

The RCMDE algorithm flow is illustrated in Figure 2.

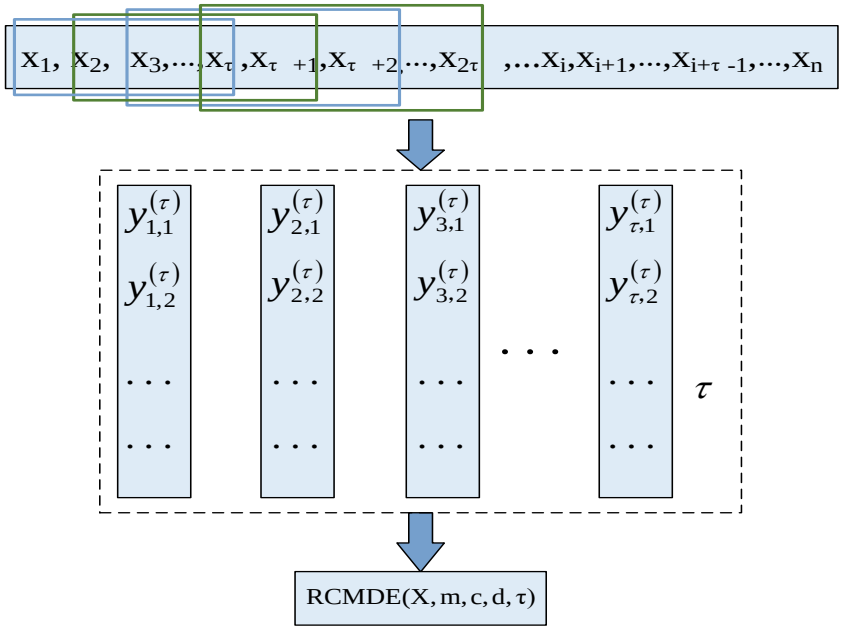


Figure 2. Flowchart of the RCMDE algorithm.

C. CEEMDAN

Let the initial time series sequence be defined as $x[n]$, Suppose the *k*-th mode component obtained from the CEEMDA*N* decomposition is denoted as $I\widetilde{M}F_k$, $E_j(\cdot)$ is the j-th mode component obtained after EMD decomposition,$\omega_j$is a Gaussian white noise signal that follows a standard normal distribution ,ε_k is the signal-to-noise ratio,$i = 1,2,\ldots,I$ represents the number of times white noise is added. Then, the algorithm flow of CEEMDAN is as follows.

(1) Add Gaussian white noise to the initial sequence $x[n]$ to obtain the new signal $x[n]+\varepsilon_k\omega^i[n]$.Perform EMD decomposition on the new signal to obtain its first-order mode $IMF_1^i[n]$ as:

$$IMF_1^i[n] = E_1\big(x[n] + \varepsilon_0\omega^i[n]\big)\#(9)$$

The overall averaging of the I generated modal components leads to obtaining the first modal component of the CEEMDAN decomposition as:

$$I\widetilde{M}F_1[n] = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i[n]\ \#\ (10)$$

(2) The residual after eliminating the first modal component is computed as:

$$r_1[n] = x[n] - I\widetilde{M}F_1^i[n]\#\ (11)$$

(3) White noise is incorporated into $r_1[n]$ to obtain the new signal$r_1[n] + \varepsilon_1 E_1(\omega^i[n])$. Subsequently, EMD decomposition is conducted to acquire its first-order mode component$IMF_2^i[n]$. Thus, the second mode

component of the CEEMDAN decomposition can be derived as:

$$I\widetilde{M}F_2[n] = \sum_{i=1}^{1} E_1\left(r_1[n] + \varepsilon_1 E_1\left(\omega^i[n]\right)\right) \#\ (12)$$

(4) With respect to $k = 2,3,...,K$, the residual after the removal of the kth modal component is computed as:

$$r_k[n] = r_{k-1}[n] - I\widetilde{M}F_k[n] \#(13)$$

the CEEMDAN decomposition can be derived.

(5) By way of analogy, white noise is incorporated into $r_k[n]$ to yield a new signal $r_k[n]+\varepsilon_k\omega^i[n]$. Subsequently, EMD decomposition is carried out on B to acquire its first-order mode component $IMF_{k+1}^{i}[n]$. Thus, the (k + 1) th mode component of

(6) Return to Step (5) and proceed to calculate the next mode until the residual is no longer decomposable, that is, until the termination condition of EMD decomposition is fulfilled. The residual obtained eventually is:

$$R[n] = x[n] - \sum_{k=1}^{K} I\widetilde{M}\,F_k \#(14)$$

where K represents the total number of modes. Therefrom, the reconstructed signal of the original signal sequence can be derived as:

$$x[n] = \sum_{k=1}^{K} I\widetilde{M}\,F_k + R[n] \#\ (15)$$

D. Construction of Hybrid Evaluation Indicators

The IMF components generated by CEEMDAN decomposition are arranged in natural descending order of frequency. By selecting and superimposing IMFs from specific frequency bands during reconstruction, three adaptive filtering effects can be achieved, namely low-pass (retaining low-frequency IMFs), high-pass (retaining high-frequency IMFs) and band-pass (combining targeted IMFs). This data-driven approach requires no preset parameters while better preserving non-stationary signals' time-frequency characteristics. However, certain IMF components - particularly high-frequency ones - often contain both signal and noise elements. Directly discarding these risks losing valid signals along with noise19. Unlike single-scale entropy analysis, Multiscale Dispersion Entropy (MDE) quantifies time series complexity across multiple scale factors, enabling assessment of IMF components' effective information content. IMFs with extremely low entropy values are typically noise-dominated, whereas those with moderate entropy generally contain significant signal components.

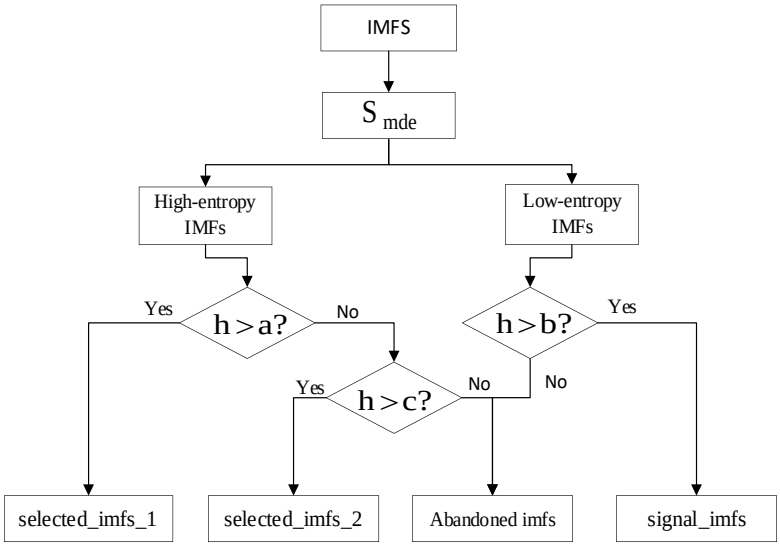


Figure 3. Mixed evaluation index.

In this paper, a hybrid evaluation index based on multi-scale dispersion entropy and Pearson correlation coefficient is constructed to serve as the clustering criterion for IMF components. As illustrated in Figure 3. Initially, the multi-scale dispersion entropy S_{mde} of the IMFs decomposed by CEEMDAN is computed under the scale factor τ, and the IMFs are clustered into high-entropy components and low-entropy components using the K-shape approach. Subsequently, the Pearson correlation coefficient h between the IMFs and the original signal is calculated. The greater the absolute value of h, the stronger the correlation. Subsequently, based on the Pearson correlation coefficient h, the IMFs are further partitioned to mitigate the potential misjudgment induced by a single indicator. Specifically, the portions of the high-entropy components featuring a Pearson correlation coefficient greater than the threshold a are regarded as the primary useful signals with relatively severe "contamination" (designated as selected_imfs_1); the portions of the low-entropy components with a Pearson correlation coefficient exceeding the threshold b are considered as "clean" useful signals (designated as signal_imfs) and retained; the portions of the high-entropy components with a Pearson correlation coefficient ranging between the threshold c and the threshold a are regarded as potential useful signals (designated as selected_imfs_2); the remaining IMFs are eliminated.

E. Noise Reduction Model

Integrating the aforementioned theoretical framework, Figure 4 presents the noise reduction model developed in this study. The denoising procedure comprises the following steps:

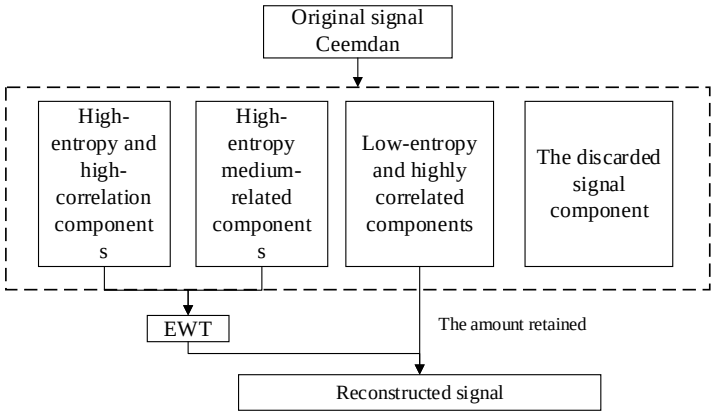


Figure 4. Signal processing module.

(1) Initial Signal Processing: A fixed-length data sample is selected from the original vibration signal and preprocessed. The signal is then decomposed using CEEMDAN to obtain a series of IMF components and a residual component.

(2) Component Classification by MDE: MDE is employed as the clustering criterion to categorize the IMF components into two primary classes: noise-dominant components and signal-dominant components.

(3) Secondary Screening & Classification: Utilizing the constructed hybrid evaluation index, a secondary screening is performed on both the noise-dominant and signal-dominant components. This process classifies the signal components into four distinct categories: 1) high-entropy, mainly useful signals (selected_imfs_1); 2) high-entropy, potentially useful signals (selected_imfs_2); 3) low-entropy useful signals (signal_imfs); and 4) discarded components.

(4) Optimized EWT Filtering for selected_imfs_1: EWT is applied to decompose the selected_imfs_1 signal. The resulting components are ranked based on their correlation, and the n components with the lowest correlation are discarded. Concurrently, the Grey Wolf Optimization (GWO) algorithm is utilized to optimize the key EWT parameters: the number of decomposed components (N) and the number of components to discard (n).

(5) Optimized EWT Filtering for selected_imfs_2: Similarly, EWT decomposition is performed on the selected_imfs_2 signal. The GWO algorithm is again employed to optimize the EWT parameters N and n for this specific signal category.

(6) Signal Reconstruction: The processed selected_imfs_1 and selected_imfs_2 signals, along with the retained signal_imfs components, are reconstructed to generate the final denoised signal.

F. Fault Identification Model

Based on the above theoretical framework, a fault diagnosis model was proposed. The workflow first denoises the raw vibration signals collected by the sensors. After signal reconstruction, RCMDE features are extracted and processed by a support vector machine (SVM) classifier to achieve fault classification and identification.MODEL VALIDATION

A. Noise Reduction Model

To validate the noise reduction capability of the signal processing module, simulation analysis was conducted.

Let the original signal be composed of a 10Hz fundamental wave (amplitude 1.2) and its 30Hz (amplitude 0.4) and 50Hz (amplitude 0.3) harmonic components, that is:

$$x(\mathrm{t}) = 0.4\,sin(\,2\pi \cdot 30t) + 0.3\,cos(\,2\pi \cdot 50t)$$

The fundamental component simulates the main vibration or power frequency component of the system, and the third and fifth harmonics reflect

the characteristic harmonics generated by typical nonlinear systems, such as mechanical transmissions or power electronics. The synthesized signal exhibits a periodic non-sinusoidal waveform with a total harmonic distortion rate (THD) of about 41.7%, which conforms to the harmonic spectrum characteristics of typical nonlinear loads in the IEEE 519-2014 standard.

Gaussian noise with a signal-to-noise ratio SNR=10 is added to the signal to simulate the actual working conditions, and the initial parameters are set to the scale factors τ=5, c=6, *m*=3, *d*=1. In terms of evaluation indexes, this paper uses three quantitative indexes with clear physical significance, namely signal-to-noise ratio (SNR), root mean square error (RMSE) and coefficient of determination (R^2), which are evaluated from three dimensions: noise suppression ability, signal reconstruction accuracy and waveform matching degree, respectively. The definitions of each evaluation index are as follows:

$$SNR = 10\,lg\frac{P_{signal}}{P_{noise}} \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x(i)-x_d(i))^2}{N}} \tag{17}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i-\hat{x}_i)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \tag{18}$$

Among them, SNR represents the ratio of signal power to noise power (in logarithmic scale), where a larger value indicates less noise pollution. RMSE is calculated as the square root of the mean squared point-by-point differences between the denoised signal and the original (noise-free) signal, and a smaller value implies higher denoising accuracy. $R^2$is utilized to quantify the explanatory capacity of the denoised signal for the original signal, within the range of (-∞, 1], and the closer to 1, the better the denoising outcome.

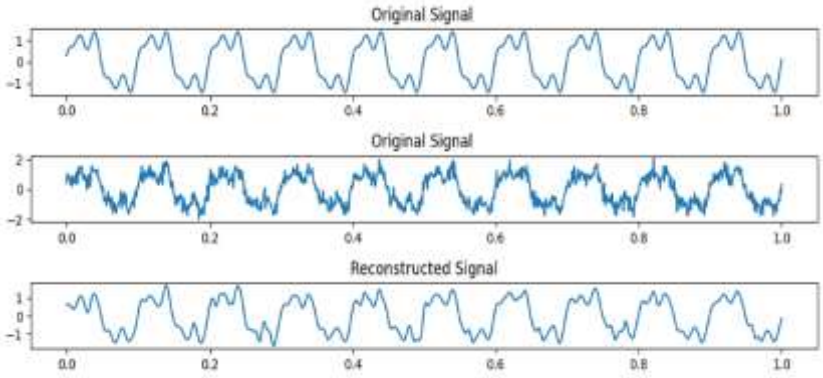


Figure 5. Waveform diagrams of the original simulation signal, noise addition signal and noise reduction signal.

The original signal, the original signal after addin g noise, and the signal denoised by the CEEMDA N-MPE-PCC-EWT method proposed in this pape r are shown in Figure 5. The various evaluation in dicators obtained through simulation analysis are presented in Table 1. It can be observed that the d

enoised signal demonstrates a remarkable noise suppression effect, maintaining a high degree of consistency with the original signal in the time-domain waveform, especially accurately restoring the signal's mutation points and characteristic peaks. While effectively eliminating the interference of background noise, it preferably retains the key characteristics of the original signal. Simultaneously, the signal-to-noise ratio (SNR) of the denoised signal has been significantly enhanced.

Table 1 The noise reduction result from the simulation signal.

Indicator	Value
SNR	15.96
RMSE	0.146
R^2	0.975

B. Fault Identification Model

The bearing fault dataset of Case Western Reserve University (CWRU) was employed to validate the fault identification model constructed in this paper, with a particular emphasis on the feature extraction capability based on RCMDE.

The CWRU bearing fault dataset encompasses vibration signals collected under diverse working conditions for multiple typical bearing fault types, such as pitting, cracks, and broken teeth, among others. In this paper, ten categories of fault signals with serial numbers (225, 258, 260, 133, 105, 170, 236, 210, 146, 118) are chosen as the experimental dataset. Each type of fault incorporates the collected signals from three position sensors. The sample size is set at 1200, with an overlap ratio of 50%. The quantity of samples for each fault type is 200. 80% of the total samples are randomly selected as the training set, and the remaining 20% serve as the validation set.

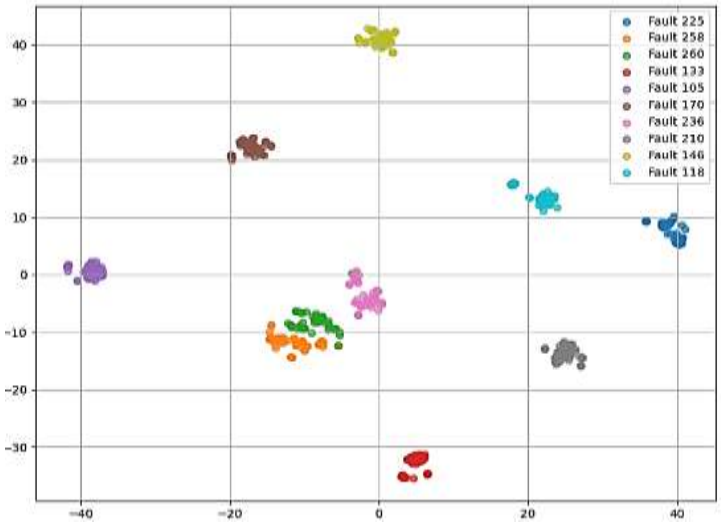


Figure 6 Classification result

The experimental results demonstrate that the fault diagnosis model constructed in this paper has achieved a 100% classification accuracy on the test set. This outstanding performance was visually verified through the t-SNE dimensionality reduction visualization, as shown in Figure 6. It can be observed that samples of different fault categories have formed well-defined and independent clusters in the feature space, and the intra-class aggregation degree is

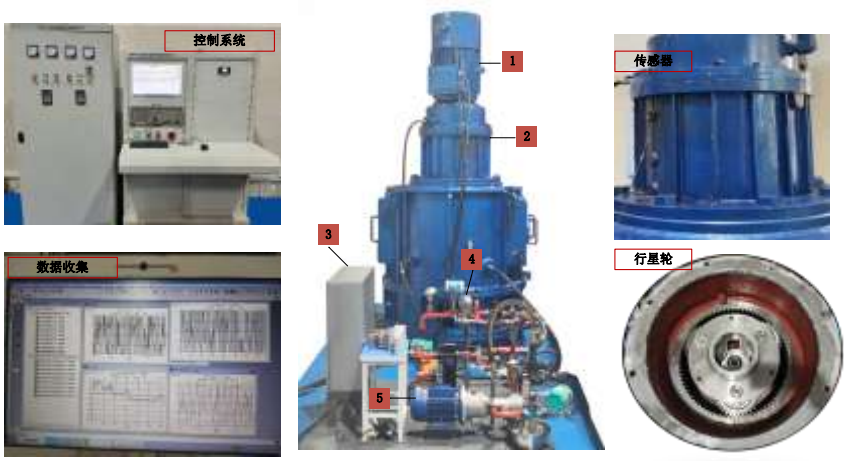
significantly higher than the inter-class distance, revealing excellent feature separability. It is worth noting that the model still maintains perfect classification performance for the three types of faults with different fault locations but similar failure mechanisms with serial numbers 133, 260 and 236.

II. Application

To further validate the effectiveness of the proposed model and method, a planetary gear fault classification and identification experiment was conducted on a planetary gearbox.

A. Planetary Gear Failure Test

Based on the actual structure and size of the nuclear power planetary gearbox, a scaled-down planetary gearbox test bench was constructed, as shown in Figure 7. This bench simulates the operational state of the nuclear power circulating water pump planetary gearbox system. The test bench comprises a control system, transmission device, power system, loading system, and lubrication system. The transmission device transfers power from the vertical motor to the horizontal load motor, establishing a functional linkage between them. The lubrication system provides lubrication to the transmission device and performs filtration and cooling of the lubricating oil. The control system regulates the input rotational speed and load. During testing, a BK accelerometer served as the sensor, connected to a Pulse data acquisition system via a BK Type 3053-B-120 acquisition card. In accordance with test requirements, the transmission path of the planetary gear set was analyzed. To ensure high-quality vibration signal acquisition, the sensor was mounted on the housing wall nearest to the planetary gear.



1 - Input motor, 2 - Planetary gearbox, 3 - Distribution pump, 4 - Oil pump of lubrication system, 5 - Motor of lubrication system

Figure 7. Gear experimental platform.

The parameters of the test gears are detailed in Table 2. The test specimens and their counterparts were manufactured in accordance with specifications JB/T10186-2007 and JB/T10817-2007. Specifically, the sun gear and planet gears, fabricated from 20CrMnMo steel, underwent surface carburizing and quenching to enhance wear resistance and bending capacity. Based on heat treatment recommendations for the nuclear power gearbox sun and planet gears from our collaborating unit, the achieved parameters were: effective case hardening depth of 0.6-0.8 mm, core hardness of HRC 32-42, tooth flank hardness of HRC 58-62, gear accuracy grade 6, addendum coefficient h*=1, and dedendum coefficient C*=1.25. Meanwhile, the inner gear ring, requiring high fatigue resistance, impact resi

stance, and excellent low-temperature impact toughness, was manufactured from 42CrMo steel; its surface treatment replicated the process used by British Allen Gears Company for the inner ring of a specific nuclear power station reducer. Following quenching and tempering, this component exhibited: hardness of 262-302 HBS, nitriding layer depth of 0.4 - 0.6 mm, surface hardness of ≥ 600 HV, gear accuracy grade 7, addendum coefficient h*= 1, and dedendum coefficient C*=1.25.

Table 2 Gear parameters

Name	Number of teeth	Modulus /mm	Tooth width/mm	Pressure angle/°
Sun gear	19	4.5	56	20
Planetary gear	33			
Internal gear annulus	86			

Table 3 Eight types of gear faults

No.	Planetary gear	Sun gear
1	Minor spalling	Normal
2	crack	Normal
3	Normal	Minor spalling
4	Normal	crack
5	Normal	Severe flaking
6	Normal	Normal
7	Slight pitting corrosion	Minor spalling
8	severely worn	Normal

Figure 8 Test site.

This paper focuses on experimental investigations of faults in planetary gears and the sun gear. The eight distinct fault types are detailed in

Table **3**

Test conditions were as follows: (1) Faults were introduced on specific teeth of either the planetary gears or the sun gear using electrical discharge machining (EDM). (2) Lubrication was provided via an oil bath system. The lubrication system's oil pump ensured lubrication, filtration, and cooling of t

he transmission oil to maintain normal gear operation. (3) The input rotational speed was 750 r/min, achieved with corresponding voltage and frequency settings of 380 V and 50 Hz, respectively. (4) Vibration signals were acquired by sensors mounted externally on the nuclear power gearbox housing. Data was collected for 1 minute at a sampling frequency of 25.6 kHz. The test setup is depicted in Figure 8. Ultimately, vibration datasets representing the eight distinct states of the planetary gear set were obtained and labeled 1 through 8.

B. Performance Comparison of Noise Reduction

Figure 9 displays the time-domain waveforms of vibration signals corresponding to the eight gear fault types. As evident from the waveforms, the experimental planetary gearbox exhibits significant noise, making it challenging to visually distinguish the fault types.

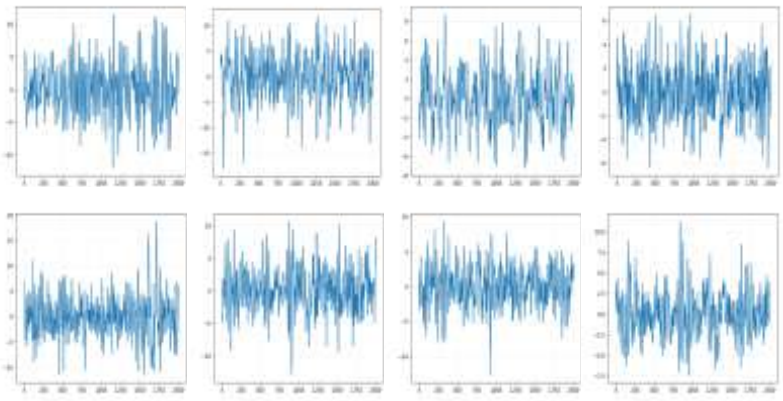


Figure 9 The time-domain waveforms of the eight fault types

Using fault type 1 as an example, a randomly selected signal segment was decomposed using the noise reduction method CEEMADN proposed in this paper. The decomposition results are shown in Figure 10.

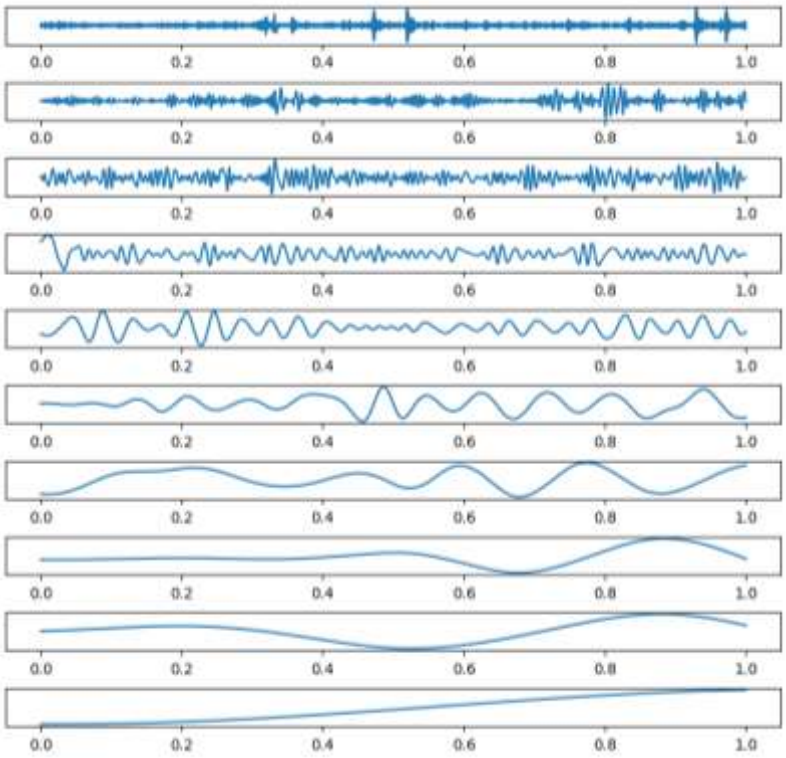


Figure 10 The IMFs after CEEMDAN decomposition (Fault 1)

The signal was decomposed into nine IMF components. Multi-scale entropy (MSE) calculations were performed for IMF1-IMF9, followed by cluster analysis using the K-shape method. This analysis categorized IMF1-IMF4 as high-entropy components (average MSE: 4.0586) and IMF5-IMF9 as low-entropy components (average MSE: 2.2375). These results clearly show that after CEEMDAN decomposition, the IMF entropy values exhibit a significant spread and can be distinctly grouped into high- and low-entropy categories. Consequently, the high-entropy category is initially considered to represent noise-dominated components.

Based on the previously constructed hybrid evaluation index and extensive experimentation, thresholds a, b, and c were established as 0.5, 0.15 and 0.3, respectively. The entropy values and Pearson correlation coefficients for each IMF component are listed in Table 4. Consequently, for Fault 1, the IMF classifications were determined as follows: IMF4 (low-entropy & high-correlation, signal_imfs); IMF1 & IMF2 (high-entropy & high-correlation, selected_imfs_1); IMF3 (high-entropy & medium-correlation, selected_imfs_2). The remaining IMF components were discarded.

Table 4 The multi-scale entropy and correlation coefficient of each imf (Fault 1).

IMF	Multiscale entropy value (scale factor = 5)	Pearson correlation coefficient
1	4.4268, 4.5496, 4.7169, 4.4230, 4.5886	0.1166
2	3.6309, 4.2070, 4.4326, 4.4909, 4.3696	0.5548
3	3.3008, 3.8309, 4.2219, 4.4864, 4.6385	0.5761
4	2.6969, 3.1572, 3.4562, 3.7172, 3.8303	0.4766
5	2.3216, 2.6602, 2.9183, 3.1229, 3.2791	0.4147
6	2.1059, 2.3277, 2.5099, 2.6629, 2.8021	0.1301
7	1.9032, 2.0110, 2.1054, 2.1919, 2.2716	0.0865
8	1.8517, 1.9218, 1.9852, 2.0433, 2.0984	-0.0100
9	1.7324, 1.7525, 1.7700, 1.7862, 1.8023	0.0121

Empirical Wavelet Transform (EWT) was applied to decompose the high-entropy and high-correlation component IMFs1 and the high-entropy and medium-correlation component IMFs2. We applied the Grey Wolf Optimization (GWO) algorithm to optimize the parameters N and n for each component set. The optimized parameters were N=6, n=1 for IMFs1 and N=10, n=4 for IMFs2.

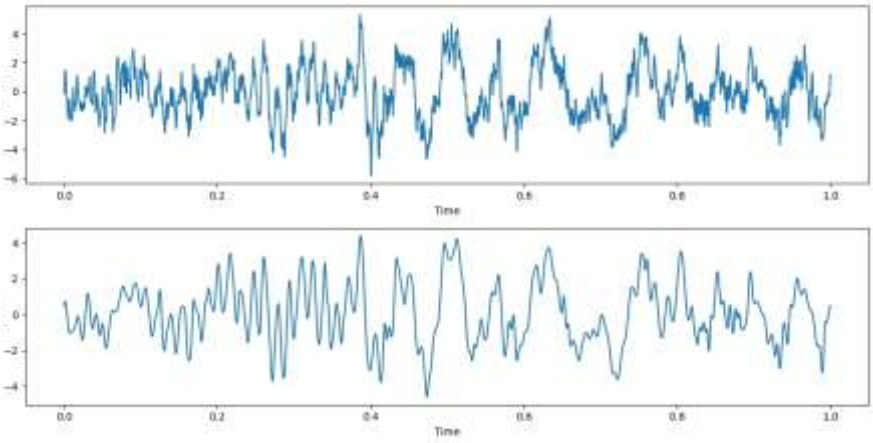


FIGURE 11. Comparison of signal waveforms before and after noise reduction

FIGURE 11 compares the signal waveforms before and after noise reduction. The denoised signal e

xhibits significant attenuation of transient disturba nces ("glitches") in the time-domain waveform w hile effectively preserving the fundamental charac teristics of the original signal. Furthermore, the ti me-domain profile appears smoother, indicating r educed interference from random fluctuations. Th ese observations demonstrate that the proposed no ise reduction method effectively eliminates extran eous high-frequency noise while preserving signal integrity.

Figure 12 compares the frequency spectra of the original and denoised signals, revealing the denoising effect primarily through: (1) significant suppression of high-frequency noise, evidenced by substantial attenuation of spectral energy in that range, indicating effective filtering of most high-frequency noise components; (2) moderate attenuation of low-frequency background noise, while ensuring the fundamental frequency components remain largely unaffected; (3) excellent preservation of characteristic frequency peaks, with amplitudes maintained near original levels, confirming retention of key signal features; and (4) enhanced spectral clarity characterized by sharper peak contours and a significantly reduced noise floor between peaks. This improved energy concentration within the main frequency components effectively enhances the signal-to-noise ratio (SNR).

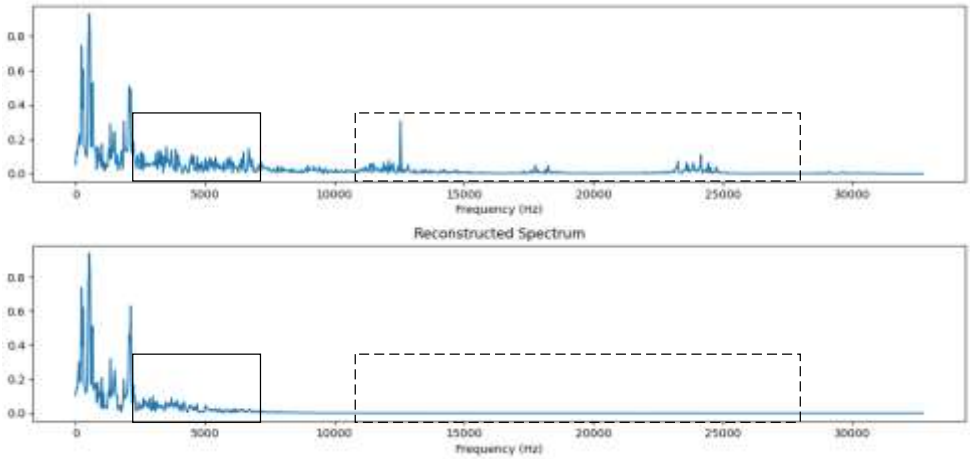


Figure 12. Comparison of signal frequency domain graphs before and after noise reduction.

The results demonstrate that the proposed noise re duction method achieves excellent performance in the planetary gear fault experiment. It effectively preserves the intrinsic characteristics of the signa l while substantially improving signal quality. Fre quency-domain analysis confirms the method's str ong adaptability across different frequency bands , capable of effectively suppressing broadband noi se while preserving critical characteristic frequenc y components. This balanced noise reduction outc ome is crucial for subsequent signal analysis and processing.

Furthermore, the proposed noise reduction metho d was compared against four representative appro aches: (1) the CEEMDAN-PE method (denoising based on permutation entropy), (2) traditional wa velet soft-threshold denoising, (3) wavelet hard-th reshold denoising, and (4) the CEEMDAN-MDE-EWT method—which employs EWT to filter and reconstruct a selected proportion of components , specifically designed to validate the efficacy of se gmented denoising. To ensure a fair comparison u nder identical conditions, the parameters of all me thods were rigorously optimized, with detailed set tings provided in Table 5. Specifically, various co mbinations of wavelet basis functions and decom position levels were evaluated for the wavelet thre sholding methods, while the IMF selection strateg

y was carefully optimized for the CEEMDAN-based methods (CEEMDAN-PE and CEEMDAN-MDE-EWT).

Table 5 Settings of various model parameters

Denoising Approaches	Parameter
CEEMDAN-PE	Embedding dimension m=3, delay time t=1, removing the number of Imfs n=2
Wavelet soft and hard threshold denoising	General threshold
CEEMDAN-MDE-EWT	m=3, t=1, category number c=6, scale factor$\tau = 5$
CEEMDAN-MDE-PCC-EWT	categories c=6, scale factor + mixed evaluation indicators$\tau = 5$

Table 6 compares the evaluation metrics of the various noise reduction methods. The results clearly demonstrate the superior performance of the proposed method across three key aspects: noise suppression capability, signal reconstruction accuracy, and waveform similarity. Specifically, the proposed method achieved an SNR increase of at least 3.14 dB, maintained RMSE below 0.84, and achieved an R^2 value of 0.94. These metrics collectively confirm the method's exceptional ability to suppress noise while preserving signal fidelity.

Table 6 Performance metrics comparison of denoising methods

Denoising Approaches	Performance metrics		
	SNR	RMSE	R^2
CEEMDAN-PE	6.4207	1.3939	0.7722
Wavelet soft and hard threshold denoising	6.1701 /8.1428	1.4352 /1.0193	0.7584 /0.8782
CEEMDAN-MDE-EWT	6.1075	1.4456	0.7549
CEEMDAN-MDE-PCC-EWT	11.2876	0.8383	0.9409

C. Performance Comparison of Fault Classification Methods

For the denoised vibration data collected under the eight planetary gear fault states, the dataset was segmented into samples of 1000 points with a 25% overlap, yielding 30 valid samples per fault type. Each sample contained synchronous measurements from the five sensors, providing a multi-dimensional basis for feature extraction. The samples were randomly partitioned, with 70% (168 samples) allocated to the training set and the remaining 30% (72 samples) to the validation set.

Planetary gear fault diagnosis was performed using the previously constructed model. The RCMDE features served as the core input to the SVM classifier. Key parameters included: embedding dimension m = 3, delay time τ = 1, number of categories c = 6, and scale factors = 5. Multi-scale analysis yielded 25-dimensional feature vectors for each fault type.

Employing the SVM classifier for all methods, the proposed fault classification and identification approach was compared against alternative methods. The identification accuracy rates are compared in Table 7. The results demonstrate the significant advantage of the proposed method, which utilizes secondary decomposition and RCMDE. Its performance surpasses that of other methods, achieving at least a 6.94% higher classification accuracy. On the validation set (72 samples), this model produced only one misclassification, resulting in an

error rate of 1.39%. These findings validate the effectiveness and reliability of the proposed fault classification method.

Table 7 Comparison of diagnostic model performance

Fault diagnosis methods	Accuracy %
MDE-SVM	73.61
CMDE-SVM	77.78
RCMDE-SVM	90.28
CEEMDAN-PE-RCMDE-SVM	65.28
Wavelet soft and hard threshold denoising-RCMDE-SVM	76.38/ 91.67
CEEMDAN-MDE-EWT-RCMDE-SVM	86.11
CEEMDAN-MDE-CC-EWT-RCMDE-SVM	98.61

III. Conclusions

This paper addresses the challenge of severe noise contamination in rotating machinery and the difficulty in extracting effective fault information from complex, nonlinear, and non-stationary vibration signals. We propose a fault diagnosis method based on secondary decomposition and RCMDE.

(1) The proposed CEEMDAN-MDE-PCC-EWT noise reduction method leverages the distinct strengths of its constituent techniques: CEEMDAN's ability to construct adaptive noise models enhances decomposition robustness and stability; MDE effectively captures the complexity and nonlinearity of time series across multiple scales; and EWT excels at mitigating mode aliasing issues. By integrating CEEMDAN decomposition, MDE-based clustering and screening, hybrid evaluation indices, and EWT-based denoising, this method achieves significant noise suppression while preserving signal fidelity, demonstrating superior denoising performance.

(2) The proposed fault classification and identification method, leveraging secondary decomposition and RCMDE, capitalizes on two key strengths: 1) secondary decomposition's enhanced capability to capture temporal dependencies in sequential data, improving prediction accuracy; and 2) RCMDE's exceptional performance in multi-scale analysis. By integrating these techniques, the method achieves high-precision fault classification and identification.

(3) The proposed method demonstrates excellent performance in both simulation studies and tests on public bearing datasets. Notably, it also achieves strong denoising and diagnostic results on the high-noise planetary gear fault dataset, significantly increasing the SNR by at least 3.14 dB and achieving a fault classification accuracy of 98.61%. This performance surpasses the benchmark methods by a significant margin.

The findings of this study offer valuable insights for rotating machinery fault diagnosis. However, verification experiments revealed two primary limitations: 1) The model's numerous parameters necessitate extensive experimentation for tuning and threshold determination. 2) High computational complexity during training results in prolonged calculation times. Future work should focus on developing adaptive parameter optimization algorithms and model acceleration strategies.

Funding: This work was supported in part by Chongqing Natural Science Foundation Innovation Development Joint Fund under Grant CSTB2022NSCQ-LZX0024, National Natural Science Foundation of China under Grant 52075062, Chongqing University of Technology Graduate Innovation Project under Grant gzlcx20253004.

References

1. L. Yan, D. Zhao, and L. Cui.(2025). "Single-trend component extraction for fault diagnosis of rotating machinery under time-varying speed co

nditions," *Measurement*, vol. 251, p. 117302.
2. S. Cui, X. Liu, X. Li, et al.(2025). "Remaining useful life prediction and uncertainty quantification method for high-end rotating machinery," *China Mech. Eng*. [Online]. Available: http://kns.cnki.net/kcms/detail/42.1294.TH.20250328.1352.002.html
3. P. Gao, J. Wang, Z. Shi, et al.(2024)."Long-term temporal attention neural network with adaptive stage division for remaining useful life prediction of rolling bearings," *Reliab. Eng. Syst. Saf*., vol. 251.
4. R. Lin, Y. Yu, H. Wang, et al.(2022)."Remaining useful life prediction in prognostics using multi-scale sequence and long short-term memory network," *J. Comput. Sci.*, vol. 57, p. 101508.
5. F. Chen, Z. Zhao, X. Hu, D. Liu, X. Yin, and J. Yang.(2025). "A nonlinear dynamics method using multi-sensor signal fusion for fault diagnosis of rotating machinery," *Adv. Eng. Inform.*, vol. 65, pt. B, p. 103190.
6. T. Wang and B. Zhang.(2019). "Application of improved empirical wavelet transform in fault feature extraction of rolling bearings," *Railw. Locomot. Car*, vol. 39, no. 5, pp. 53–58.
7. L. Xian, K. He, C. Wang, et al.(2020). "Factor analysis of financial time series using EEMD-ICA based approach," *Sustain. Futures*, vol. 2, pp. 100003–100030.
8. Y. Zhou, S. Wu, and W. Xing.(2025)."Gearbox fault diagnosis based on feature fusion," *Noise Vib. Control*, vol. 45, no. 2, pp. 63–69.
9. Y. Huang, Z. Li, J. Wang, et al.(2025). "Misfire fault diagnosis based on EEMD decomposition of diesel engine crankshaft instantaneous speed signal," *J. Beijing Inst. Technol.* [Online]. Available: https://doi.org/10.15918/j.tbit1001-0645.2024.165
10. D. Zhu, W. He, and Q. Zhu.(2024)."Fault feature extraction method based on multi-signal and improved empirical Fourier decomposition," *J. Mech. Electr. Eng.*, vol. 41, no. 10, pp. 1853–1864.
11. F. Zhou, S. Wei, and G. Qiao.(2025). "Underwater acoustic signal modulation recognition method combining VMD-wavelet denoising and bilinear ResNet with coordinate attention mechanism," *J. Harbin Eng. Univ*. [Online]. Available: http://kns.cnki.net/kcms/detail/23.1390.u.20250305.1733.002.html
12. M. E. Torres et al.(2011). "A complete ensemble empirical mode decomposition with adaptive noise," in Proc. IEEE ICASSP, pp. 4144–4147.
13. K. Wang, X. Xiao, Y. Yu, et al.(2024). "CEEMDAN-CNN-BiLSTM hybrid model for surface subsidence prediction in mining areas," *J. Navig. Position.*, vol. 12, no. 5, pp. 156–163, doi: 10.16547/j.cnki.10-1096.20240519.
14. Y. Wu, J. Long, Y. Wei, et al.(2025). "Bearing vibration signal denoising method based on CEEMDAN-WTD-DBO," *Mod. Electron. Tech.*, vol. 48, no. 6, pp. 91–98, doi: 10.16652/j.issn.1004-373x.2025.06.015.
15. G. Wang, J. He, Y. Min, et al.(2025). "Collaborative denoising method for sound signals of high-voltage shunt reactor based on CEEMDAN-SVD with improved cut-off frequency," *Elect. Mach. Control*. [Online]. Available: http://kns.cnki.net/kcms/detail/23.1408.TM.20250303.1741.002.html
16. D. Wang.(2025)."Fault diagnosis method for abnormal vibration of pump motor bearing based on secondary decomposition," *Tech. Superv. Water Resour*., no. 2, pp. 282–286.
17. L. Shang, D. Jia, D. An, et al.(2025). "Ultra-short-term load forecasting model based on CEEMD feature combination," *J. Guangxi Normal Univ. (Nat. Sci. Ed.)*. [Online]. Available: https://doi.org/10.16088/j.issn.1001-6600.2024090603
18. M. Rostaghi and H. Azami.(2016)."Dispersion entropy: A measure for time-series analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614.
19. H. Azami, M. Rostaghi, and J. Escudero.(2017). "Refined composite multiscale dispersion entropy and its application to biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2872–2879, Dec.

20. Fan, C. Li, S. Yang, et al.(2023). "Application o f CEEMDAN combined with wavelet threshold algorithm in suppressing scattering clutter in un derwater lidar," *Acta Phys. Sin.*, vol. 72, no. 22, pp. 106–113.