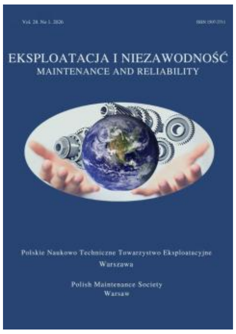


Eksploatacja i Niezawodnosc – Maintenance and Reliability
Volume 28 (2026), Issue 1
journal homepage: http://www.ein.org.pl

Article citation info:
Han F, Zhang X, Cao J, A novel approach of fault diagnosis for gearbox based on VMD optimized by GSWOA and improved RCMSE, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(1) http://doi.org/10.17531/ein/208672

A novel approach of fault diagnosis for gearbox based on VMD optimized by GSWOA and improved RCMSE

Indexed by:
Web of Science Group

Fei Han[a,*], Xiaodong Zhang[a,b], Jiahao Cao[a]

[a] School of Mechanical Engineering, Xinjiang University, China
[b] School of Mechanical Engineering, Xi'an Jiaotong University (XJTU), China

Highlights

- GSWOA is integrated to optimize VMD decomposition, significantly suppressing background noise and providing high-quality input for reliable feature extraction and fault diagnosis.
- An improved refined composite multiscale entropy (IRCMSE) method is proposed to accurately extract and distinguish diverse gear fault features, enhancing both diagnostic precision and stability.
- The proposed method demonstrates strong noise robustness and generalization across variable-speed and variable-load conditions, validated on both public and real-world datasets.

Abstract

Wind turbine gearboxes face alternating loads and complex noise over prolonged operation, leading to vibration signals affected by noise, nonlinearity, and non-stationarity. These challenges complicate fault diagnosis. This study proposes a novel approach, combining a modified whale optimization algorithm (GSWOA) to optimize variational mode decomposition (VMD) parameters and an improved refined composite multiscale sample entropy (IRCMSE) method. The approach enhances VMD decomposition, mitigating mode mixing and boundary artifacts. An advanced coarse-graining process generates IRCMSE-based features, improving sensitivity to weak fault signatures. These features are fed into a CNN-BiLSTM model, leveraging CNN for spatial feature extraction and BiLSTM for temporal behavior modeling, enabling precise fault classification. Experiments on the WFD-1000 wind turbine gearbox platform show superior performance in signal reconstruction, feature discrimination, and fault detection, highlighting the method's robustness under complex operational conditions.

Keywords

fault diagnosis, improved refined composite multiscale sample entropy, improved whale optimization algorithm, convolutional neural network, bidirectional long short-term memory network

1. Introduction

The gearbox, as a vital transmission unit in wind turbine systems, is essential for sustaining efficient power generation and ensuring long-term operational reliability. It transforms low-speed, high-torque mechanical energy into input suitable for electrical conversion, all while enduring complex and variable loading conditions, fluctuating ambient environments, and extended service durations[1]. Such operating scenarios render the gearbox vulnerable to a range of structural degradations, including gear wear, fatigue-induced cracking, and tooth breakage—failures that can induce abrupt shutdowns and result in significant economic impact[2].

Among various monitoring signals, vibration data are particularly advantageous in fault diagnosis due to their sensitivity to structural deviations and their ability to capture nuanced dynamic behavior. Compared with thermal or acoustic emission signals, vibration responses provide more comprehensive insight into meshing dynamics and internal structural alterations. However, the inherent complexity of gearbox architecture and the strong coupling among internal components lead to vibration signals that are nonlinear, non-

---

(*) Corresponding author.
E-mail addresses: F. Han (ORCID: 0009-0004-7640-9352) 13781759406@163.com, X. Zhang (ORCID: 0009-0008-5131-0523) xdzhang@xjtu.edu.cn, J. Cao (ORCID: 0009-0005-7729-5446) jhaocao@163.com

stationary, and frequently corrupted by harmonic interference and stochastic noise[3]. These characteristics impose substantial challenges on effective signal analysis and fault pattern recognition. As such, the pursuit of accurate and robust fault diagnosis techniques—capable of operating under real-world, variable working conditions—remains both a theoretical necessity and a pressing engineering concern[4].

Enhancing the precision of feature extraction and improving fault classification accuracy requires the application of effective preprocessing techniques to suppress background noise and highlight signal components related to mechanical degradation. Among conventional signal decomposition approaches, Empirical Mode Decomposition (EMD) has been widely employed in the analysis of nonlinear and non-stationary vibration data[5]. However, its tendency to induce mode mixing often compromises both the interpretability and consistency of the decomposed results. To address this issue, several improved variants—such as Ensemble EMD (EEMD), Complementary EEMD (CEEMD)[6], and CEEMD with Adaptive Noise (CEEMDAN)—have been proposed, introducing controlled noise and ensemble averaging to reduce mode aliasing[7]. Despite these enhancements, the core limitations of EMD remain unresolved due to their reliance on its intrinsic algorithmic structure.

An alternative strategy, Variational Mode Decomposition (VMD), formulated within a variational optimization framework, allows for the decomposition of signals into a set of finite-bandwidth intrinsic mode functions while inherently avoiding mode overlap and boundary distortions. This formulation not only improves mode separation but also offers superior resilience to noise interference, making it well-suited for vibration signals characterized by structural complexity and interference from multiple dynamic sources. Nevertheless, the diagnostic performance of VMD depends critically on the proper configuration of its primary parameters, including the number of decomposition modes and the penalty factor. Inadequate parameter settings may lead to suboptimal feature extraction and diminished fault identification reliability. Therefore, the development of an adaptive and computationally efficient strategy for parameter tuning remains a key challenge in maximizing the effectiveness of VMD-based diagnostic frameworks[8].

To address the issues of over- or under-decomposition caused by improper settings of the mode number K and the penalty factor αin Variational Mode Decomposition (VMD), existing studies primarily adopt two categories of approaches: empirical adjustment and intelligent optimization. Traditional strategies typically fix α and then adjust K incrementally based on subjective evaluation of the decomposition results. However, this trial-and-error process is highly experience-dependent, time-consuming, and unsuitable for complex signal environments. To achieve adaptive parameter selection, a Fault-Information-Guided VMD (FIVMD) method has been proposed[9], which constructs a statistical model using fault cyclic characteristics and the Ratio of Feature Component Amplitude (RFCA) to autonomously determine VMD parameters[10]. Although this method demonstrates good performance in weak feature extraction, it suffers from high computational complexity. With the advancement of swarm intelligence optimization algorithms, methods such as Particle Swarm Optimization (PSO)[11], Grey Wolf Optimizer (GWO)[12], Slime Mould Algorithm (SMA)[13], Harris Hawks Optimization (HHO)[14], and Whale Optimization Algorithm (WOA) have been widely applied to VMD parameter tuning[15]. Among them, GWO-based VMD has shown strong performance in suppressing mode mixing and improving fault recognition accuracy, while SMA-VMD and HHO-VMD have demonstrated excellent denoising and feature enhancement capabilities. Comprehensive analysis indicates that swarm intelligence algorithms significantly improve the scientific rigor and robustness of VMD parameter selection, thereby enhancing the effectiveness of vibration signal preprocessing under complex backgrounds. Based on these insights, this study employs VMD as the signal decomposition tool and introduces an advanced swarm intelligence optimization strategy to achieve adaptive parameter tuning, thereby improving the overall performance of subsequent feature extraction and fault diagnosis.

In order to enhance the discriminative power of extracted features, particularly when classifying closely related fault categories, the characterization of signal complexity plays a pivotal role. Entropy-based methods have been widely adopted in vibration signal analysis due to their ability to quantify dynamic irregularities. Conventional approaches such as Sample Entropy (SE), Approximate Entropy (APE)[16], and Energy Entropy (EE) have demonstrated effectiveness in extracting fault-related information[17]. However, these techniques are limited by their reliance on single-scale analysis, which restricts their ability to capture multi-scale behaviors inherent in mechanical systems[18].

To address this shortcoming, multiscale entropy (MSE) theory was introduced, allowing entropy measurements to extend across multiple time scales. Among its variants, Multiscale Sample Entropy has achieved notable results in practical diagnostics. Nonetheless, standard MSE is known to suffer from information degradation due to the coarse-graining process, which reduces sensitivity to localized fluctuations. To mitigate this issue, Refined Composite Multiscale Sample Entropy (RCMSE) was proposed[19], offering improved stability and feature preservation by refining the averaging mechanism used in coarse-graining. While RCMSE has shown improved performance in analyzing vibration signals, its effectiveness declines when the signal is heavily contaminated by noise or when fault characteristics are weak and difficult to isolate. In such cases, further enhancement of the entropy model becomes necessary to ensure reliable feature extraction and robust classification performance under challenging diagnostic conditions.

Once vibration signal features have been extracted and transformed into structured feature vectors, the subsequent classification stage becomes pivotal for effective fault identification. In this context, both traditional machine learning algorithms and modern deep learning frameworks have been widely explored in rotating machinery diagnostics. Among these, convolutional neural networks (CNNs) with residual mechanisms and bidirectional long short-term memory (BiLSTM) networks have demonstrated significant advantages in capturing spatial hierarchies and temporal dependencies, respectively—leading to enhanced recognition of complex fault signatures[20].

Despite their promising performance, the deployment of deep learning architectures in industrial scenarios often encounters practical constraints[21]. One notable limitation is the dependency on large-scale labeled datasets, which are frequently unavailable in real-world settings due to cost and time constraints. Moreover, conventional neural models such as probabilistic neural networks (PNN) and backpropagation neural networks (BPNN)[22], although effective under certain conditions, exhibit reduced generalization capability owing to their sensitivity to initial weights and intricate parameter configuration. The high computational burden and extended training time associated with deep models further restrict their use in applications requiring rapid response or real-time processing[23].

To address these challenges, a hybrid model integrating CNN and BiLSTM is proposed in this study. By leveraging the spatial abstraction capability of CNNs and the temporal learning strength of BiLSTM networks, the framework aims to improve diagnostic accuracy while maintaining robustness under complex and variable operating environments. This architecture is specifically tailored for fault detection in wind turbine gearboxes, with the goal of achieving reliable, efficient, and scalable diagnostic performance across diverse industrial conditions.

In summary, wind turbine gearboxes operate under harsh conditions involving cyclic loads and alternating mechanical stress, where diverse fault types and persistent noise significantly hinder accurate diagnosis. These challenges complicate both feature extraction and fault identification, limiting the effectiveness of traditional methods. To address this, an integrated diagnostic framework is proposed, combining Variational Mode Decomposition (VMD) with parameter optimization via an improved Whale Optimization Algorithm (GSWOA), and enhanced feature representation through a refined Composite Multiscale Sample Entropy (IRCMSE) model. Experimental validation confirms the framework's superior capability in identifying gear transmission faults with high precision, offering a robust and scalable solution for condition monitoring under complex operational environments.

The main contributions of this study are as follows:

（1）This study introduces a signal preprocessing strategy based on the GSWOA-VMD-WAT framework, aiming to enhance diagnostic reliability under noisy conditions. The approach identifies and eliminates signal components exhibiting low correlation and minimal energy contribution relative to the original vibration signal. By selectively attenuating irrelevant background interference, the method improves the clarity of fault-related features and establishes a cleaner input foundation for subsequent feature extraction and classification stages.

（2）A refined version of the traditional RCMSE algorithm is introduced to improve the reliability of feature extraction from nonlinear and non-stationary signals. The enhanced approach effectively captures stable and informative characteristics within the vibration responses of wind turbine gearboxes, thereby providing a robust foundation for high-precision fault identification.

（3）An integrated fault diagnosis framework is constructed by combining GSWOA-VMD for signal decomposition, IRCMSE for feature extraction, and a CNN–BiLSTM architecture for classification. The proposed method is validated through a series of experiments and comparative evaluations, demonstrating its effectiveness and superior diagnostic performance.

The remainder of this paper is organized into five sections. Section 2 introduces the core theoretical foundations supporting this work, namely Variational Mode Decomposition (VMD), Improved Refined Composite Multiscale Sample Entropy (IRCMSE), the enhanced Whale Optimization Algorithm (GSWOA), and the combined CNN–BiLSTM model. Section 3 describes the construction of the proposed fault diagnosis framework designed for wind turbine gearboxes. Section 4 presents experimental validation and comparative analysis based on data obtained from the Wind Turbine Drivetrain Diagnostic Simulator (WFD-1000). Lastly, Section 5 offers a summary of the findings and outlines the conclusions drawn from the study.

2. Relevant theoretical foundations

2.1. VMD decomposition of vibration signals

Variational Mode Decomposition (VMD) is fundamentally based on solving a constrained variational problem[24], in which the original signal is decomposed into a predefined set of Intrinsic Mode Functions (IMFs) by minimizing an objective functional. The mathematical formulation of this optimization model is expressed as follows:

$$\begin{cases} \min_{\{\omega_m\},\ \{u_m\}} \left\{ \sum_m \left\| \partial_t \left[\left(\partial(t) + \frac{j}{\pi t} \right) u_m(t)^* \right] e^{-j\omega_m} \right\| \right\} \\ \text{s.t.} \sum u_m = f \end{cases} \tag{1}$$

Here, u_m denotes each Intrinsic Mode Function (IMF), and ω_m represents its corresponding center frequency. To convert the constrained variational problem into an unconstrained form, a balancing parameter α and a Lagrange multiplier λ are introduced:

$$\begin{cases} \zeta = \left\| f(t) - \sum_m u_m(t) \right\|_2^2 \\ \varphi = \alpha \sum_m \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right)^* u_m(t) \right] e^{-jt\omega_m} \right\|_2^2 + \zeta \\ \lambda(\{u_m\}, \{\omega_m\}, \lambda) = \varphi + \left\{ \lambda t, f(t) + \sum u_m(t) \right\} \end{cases} \tag{2}$$

The computational procedure of VMD is summarized as follows:

Step 1: Initialize the iteration by updating the counter $n = n + 1$

Step 2: Update u_m using the following equation:

$$u_m^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq m} u_m(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_m)^2} \tag{3}$$

Step 3: Update ω_m sing:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega^* |u_m(\omega)|^2 du}{\int_0^\infty |u_m(\omega)|^2 du} \tag{4}$$

Step 4: Update λ using the following equation:

$$\lambda^{n+1}(\omega) = \lambda^n(\omega) + \varepsilon \left| f(\omega) - \sum u_m^{n+1}(\omega) \right| \tag{5}$$

Step 5: If the convergence criterion is satisfied, terminate the iteration.

$$\sum_m \| u_m^{n+1} - u_m^n \|_2^2 < v \tag{6}$$

2.2. GSWOA-VMD Joint Wavelet Thresholding Noise Reduction

The performance of Variational Mode Decomposition (VMD) is highly sensitive to its parameters, particularly the number of modes K and the penalty factor α. To enable adaptive and efficient parameter selection, this paper proposes a Grey Wolf Strategy-enhanced Whale Optimization Algorithm (GSWOA), which performs global optimization of the parameter pair (K, α).

GSWOA integrates the spiral updating mechanism of the Whale Optimization Algorithm (WOA) with the α–β–δ leadership hierarchy from the Grey Wolf Optimizer (GWO). In each iteration, the three best-performing agents (α, β, and δ) guide the position updates of the remaining agents, thereby enhancing search diversity and convergence stability. Additionally, GSWOA introduces time-dependent control parameters A(t) and C(t), which dynamically adjust as the iteration progresses to regulate the search scope and speed at different stages:

$$A(t) = A_{min} - \left(\frac{A_{max} - A_{min}}{T}\right) \cdot t \tag{7}$$

$$C(t) = C_{min} + \left(\frac{C_{max} - C_{min}}{T}\right) \cdot t \tag{8}$$

To further mitigate the risk of premature convergence, a perturbation mechanism is embedded into the position updating process:

$$X_i^{t+1} = X_i^t + \varepsilon_c \cdot \sin\left(\frac{\pi t}{T}\right) \cdot \left(\frac{X_\alpha^t + X_\beta^t + X_\delta^t}{3} - X_i^t\right) \tag{9}$$

where X_i^t denotes the position of the *i*-th agent at iteration *t*, and ε_c is the perturbation coefficient (set to 0.01).

To evaluate the quality of VMD decomposition and guide the search process, a composite fitness function is defined as:

$$F(K, \alpha) = -\lambda_1 \cdot \frac{\sum_{i=1}^{K} \int_{f_m - \Delta f}^{f_m + nf_r + \Delta f} |S_i(f)|^2 \, df}{\sum_{i=1}^{K} \int \quad |S_i(f)|^2 df} + \lambda_2 \cdot \frac{1}{K(K-1)} \sum_{i \neq j} \rho\left(S_i, S_j\right) \tag{10}$$

where $S_i(f)$ represents the spectrum of the *i*-th Intrinsic Mode Function (IMF). The first term quantifies the energy concentration within the meshing frequency and its sidebands, while the second term reflects the degree of mode decoupling via the mean Pearson correlation among IMFs. The weighting coefficients are set as λ_1= 0.7 and λ_2=0.3, respectively. The search ranges are defined as K∈[3,8] and α∈[1000,5000], with a maximum number of iterations T=100. Early stopping is triggered when the fitness improvement remains below 1e-4 for 10 consecutive iterations.

The optimal parameter pair (K , α) obtained by GSWOA is then applied to the VMD process for signal decomposition. IMF components with significant energy and strong temporal correlation are retained for signal reconstruction. To further reduce high-frequency noise in the reconstructed signal, an adaptive wavelet thresholding method is employed. The complete denoising process is formulated as:

$$\widehat{w}_j(k) = \begin{cases} \operatorname{sgn}\left(w_j(k)\right) \cdot \left(\left|w_j(k)\right| - \lambda_j\right), & \text{if } \left|w_j(k)\right| > \lambda_j, \\ 0, & \text{otherwise} \\ \lambda_j = \frac{MAD}{0.6745} \cdot \sqrt{2\log N} \end{cases} \tag{11}$$

The fundamental workflow of GSWOA-VMD combined with wavelet-based adaptive thresholding denoising is illustrated in Figure 1[25]. Initially, vibration signals acquired from the experimental platform undergo preprocessing, including DC component removal and normalization, to improve signal consistency and computational efficiency. The key parameters of Variational Mode Decomposition, namely the number of decomposition modes K and the penalty factor α, are adaptively optimized using GSWOA, which iteratively searches for the configuration that yields the most effective decomposition structure. Once optimization is complete, VMD is applied to decompose the signal into a series of narrow-band components. Each resulting Intrinsic Mode Function (IMF) is then assessed in terms of its energy content and correlation with the original signal. Components meeting the selection criteria are retained and reconstructed to suppress noise, thereby enhancing the quality of time-domain signal representation.

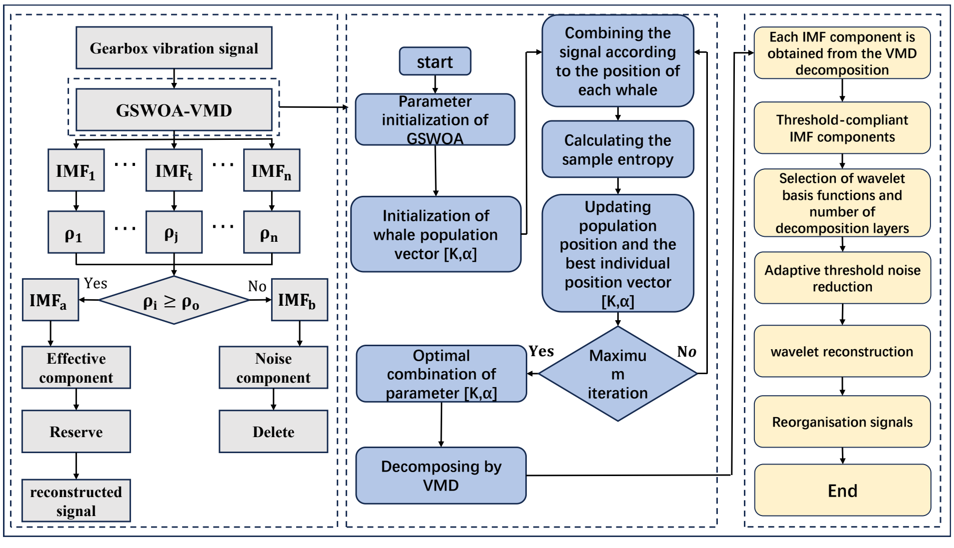


Fig.1. GSWOA-VMD Joint Wavelet Thresholding Noise Reduction Flowchart.

2.3. Relying on IRCMSE to fully and adequately extract fault characteristics

2.3.1. Multi-scale sample entropy

Multiscale Sample Entropy (MSE) extends the single-scale analysis of Sample Entropy (SE) to multiple time scales, enabling the extraction of potentially valuable information embedded in complex signals. The basic computational procedure of MSE is as follows:

Step 1: Let $X = \{x_i, i = 1,2, \ldots, N\}$ denote the original time series. The corresponding coarse-grained sequence is constructed by applying the following transformation procedure.

$$\begin{cases} y_j^{\tau} = \dfrac{1}{\tau} \displaystyle\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq T \\ T = \left\lceil \dfrac{N}{\tau} \right\rceil \end{cases} \tag{12}$$

where $\tau = 1,2, \ldots$ is the scale factor, and $T = \frac{N}{\tau}$ represents the length of the coarse-grained time series. For a given scale factor τ, the corresponding coarse-grained time series is denoted as $y^{\tau} = \{y_1^{\tau}, y_2^{\tau}, \ldots, y_T^{\tau}\}$.

Step 2: For a given scale factor τ, reconstruct the coarse-grained time series y^{τ} into embedding vectors in dimensions m and m+1, respectively:

$$\begin{cases} Z_i^{(m)} = \{y_i^{\tau}, y_{i+\lambda}^{\tau}, \ldots, y_{i+(m-1)\lambda}^{\tau}\} \\ Z_i^{(m+1)} = \{y_i^{\tau}, y_{i+\lambda}^{\tau}, \ldots, y_{i+m\lambda}^{\tau}\} \\ 1 \leq i \leq T - m + 1 \end{cases} \tag{13}$$

where m is the embedding dimension, typically set to 1 or 2, and λ is the time delay, usually set to 1.

Step 3: For each i, compute the distance d between Z_i^{τ} and all other Z_i^{τ} in the mmm-dimensional space:

$$\begin{cases} d[z_i^{\tau}, z_j^{\tau}] = \max\{|z_i^{\tau} - z_j^{\tau}|\} \\ k = 0,1, \ldots, \tau \\ i, j = 1,2, \ldots, T - m + 1, \quad i \neq j \end{cases} \tag{14}$$

Step 4: The selection of the similarity tolerance parameter r plays a critical role in the stability of Multiscale Entropy (MSE) computation. Choosing a value that is either too small or too large can significantly distort the entropy estimation. To maintain consistency across varying signal magnitudes, r is typically defined as a fraction of the standard deviation of the original time series X, commonly ranging from 0.1 to 0.25 times SD(X). Given a coarse-grained sequence, the number of vector pairs $[z_i^{\tau}, z_j^{\tau}]$ for which the distance $\sum[z_i^{\tau}, z_j^{\tau}]$ is less than the threshold r is determined. This count is normalized by $T - m$ to yield the quantity $B_i^m(r)$. Aggregating over all template vectors produces the average similarity count $B^m(r)$, as defined by the following formulation:

$$\begin{cases} B_i^m(r) = \dfrac{1}{T - m} \sum d[z_i^{\tau}, z_j^{\tau}] < r \\ B^m(r) = \dfrac{1}{T - m + 1} \sum B_i^m(r) \\ i, j = 1,2, \ldots, T - m + 1, \quad i \neq j \end{cases} \tag{15}$$

Step 5: Based on the previous embedding dimension m, increase the dimension by one and repeat Steps 3 and 4 to obtain the value of $B^{m+1}(r)$.

Step 6: Finally, the value of Multiscale Entropy (MSE) is calculated using the obtained $B^m(r)$ and $B^{m+1}(r)$.

$$E_{MSE}(X, \tau, m, r) = \lim_{T \to \infty} \left\{ -\ln \frac{B^{m+1}(r)}{B^m(r)} \right\} \tag{16}$$

When T is a finite value, the Sample Entropy (SampEn) is calculated as follows:

$$E_{MSE}(X, \tau, m, r) = -\ln \frac{B^{m+1}(r)}{B^m(r)} \tag{17}$$

While Sample Entropy (SE) quantifies the complexity of a time series at a single resolution, Multiscale Entropy (MSE) extends this analysis across multiple temporal scales, allowing for a more comprehensive assessment of dynamic behavior. This multi-scale perspective addresses a key limitation of SE, which may fail to capture long-range dependencies or hidden features when confined to the original resolution. Despite these improvements, MSE is not without its shortcomings. During the coarse-graining stage, temporal correlations among data points are often neglected, potentially leading to the loss of critical information. Moreover, when applied to short-length signals with large scale factors, the method may yield undefined entropy values due to insufficient data points in the downsampled sequences.

2.3.2. Refined Composite Multiscale Entropy (RCMSE) Methodology

To overcome the limitations associated with traditional Multiscale Entropy (MSE), the Refined Composite Multiscale Entropy (RCMSE) approach has been introduced. This method enhances the conventional coarse-graining process by generating a sequence of refined sub-series under a given scale factor τ. Specifically, for each position in the original time series, a local segment of length τis extracted, and its mean is computed to form an element of the coarse-grained sequence. The starting index is incrementally advanced by one data point at a time until the end of the series is reached, resulting in a set of overlapping subsequences. This refined coarse-graining procedure improves feature preservation across scales and offers a more accurate representation of the signal's dynamic structure. The corresponding computational procedure is outlined below.

Step 1: Coarse-grain the original time series $X = \{x_i, i = 1,2,\dots,N\}$ at the scale factor τ. The refined coarse-grained time series is obtained by calculating the average of τ consecutive data points starting from each data point of the original series. The process is defined by the following equation:

$$\begin{cases} y_{j,k}^{\tau} = \dfrac{1}{\tau}\displaystyle\sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i \\ 1 \le j \le \dfrac{N}{\tau}, \quad 1 \le k \le \tau \end{cases} \tag{18}$$

Step 2: For all refined coarse-grained sequences y_j^{τ} obtained at the scale factor τ, calculate the Sample Entropy (SE) for each sequence. Subsequently, compute the average of these entropy values to obtain the RCMSE at scale τ. The calculation is expressed as follows:

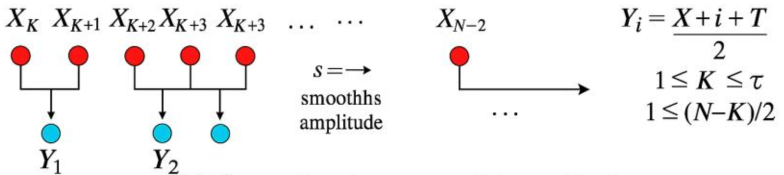
$$E_{\mathrm{RCMSE}}(x,\tau,m,r) = \frac{1}{\tau}\sum_{k=1}^{\tau}\left(-\ln\frac{\eta_{k,\tau}^{m+1}}{\eta_{k,\tau}^{m}}\right) \tag{19}$$

2.3.3. Improved Fine Composite Multiscale Sample Entropy

Refined Composite Multiscale Entropy (RCMSE) is a widely used technique for measuring time series complexity. In its standard formulation, the multiscale coarse-graining process typically employs arithmetic averaging across non-overlapping segments of the original signal. While this averaging mechanism can partially preserve structural information, it also inevitably suppresses high-energy transient components—such as impulsive responses—thereby degrading diagnostic sensitivity, especially in complex systems such as wind turbine gearboxes.

To enhance the representation of fault-related transient behaviors, this study introduces an Improved RCMSE (IRCMSE) algorithm that fundamentally reconstructs the coarse-graining phase. The method incorporates a cross-sampling strategy and an energy-driven selection mechanism, both of which are designed to retain amplitude-dominant dynamics and improve scale-wise adaptivity.

In the proposed approach, for a given scale factor τ, the original time series is no longer averaged within fixed-length windows. Instead, it is decomposed into τ interleaved subsequences, which together form a candidate pool for coarse-grained representation. Distinct from traditional schemes, this method evaluates the diagnostic potential of each subsequence using its root mean square (RMS) value as an energy-based criterion. The subsequence exhibiting the highest RMS is selected as the representative sequence for the current scale. Figure 2 illustrates the construction of the k-th refined segment when τ = 2. This selective strategy mitigates the dilution of informative content and significantly enhances sensitivity to high-frequency impulsive events.



(a) Conventional coarse-graining methods

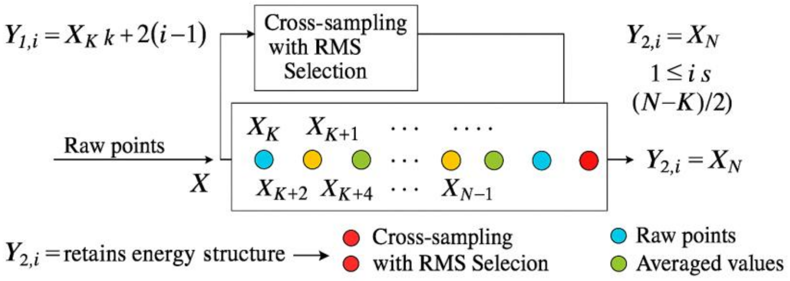


Fig. 2. Conventional and improved coarse-graining methods.

Moreover, to counteract the sample degradation and information loss typically observed in standard RCMSE at larger scale factors, the IRCMSE method further introduces a composite-scale enhancement mechanism. Specifically, at each scale τ, τ mutually independent subsequences are constructed from the original signal. These are then coarse-grained using scale factors ranging from 1 to τ, and the corresponding entropy values are computed in parallel. This composite representation effectively delays information decay and improves the completeness and robustness of feature characterization across multiscale domains.

In summary, IRCMSE achieves substantial improvements in response capability to impulsive and nonstationary signals through dual structural enhancements in both the coarse-graining procedure and multiscale decomposition framework. It thereby provides high-fidelity, high-resolution feature support for downstream fault pattern recognition tasks.

2.4. CNN–BiLSTM-Driven Framework for Intelligent Fault Diagnosis

Following the denoising and initial feature extraction stages, effective fault identification requires a model capable of interpreting the temporal and spatial complexities inherent in vibration signals. In this work, a composite deep neural network is constructed by integrating a Convolutional Neural Network (CNN) with a Bidirectional Long Short-Term Memory (BiLSTM) module. The input to the network is a two-dimensional feature matrix derived from the IRCMSE stage, with dimensions N×T, where N denotes the number of feature channels and T represents the temporal length. The CNN subnetwork is composed of two one-dimensional convolutional layers, employing 1×3 and 1×5 kernels with 16 and 32 output channels, respectively. Each layer uses ReLU activation and is followed by max-pooling to reduce dimensionality while preserving local discriminative structures such as sideband modulations and impulsive shock patterns. The resulting feature maps are then fed into a two-layer BiLSTM network, with each layer comprising 128 hidden units and operating in both temporal directions to capture contextual dependencies. Dropout layers with a rate of 0.5 are inserted between BiLSTM layers to prevent overfitting. A fully connected layer and a Softmax classifier are appended to produce multi-class fault predictions. The model is trained using the cross-entropy loss function and optimized via the Adam algorithm with an initial learning rate of 0.001, a batch size of 64, and a maximum of 200 epochs. Training and validation sets are partitioned in an 8:2 ratio, and early stopping is employed to monitor validation performance and avoid overfitting. By combining local spatial abstraction with robust temporal modeling, the proposed CNN-BiLSTM architecture offers high adaptability for complex fault diagnosis under nonstationary industrial conditions. The full model architecture is illustrated in Figure 3.

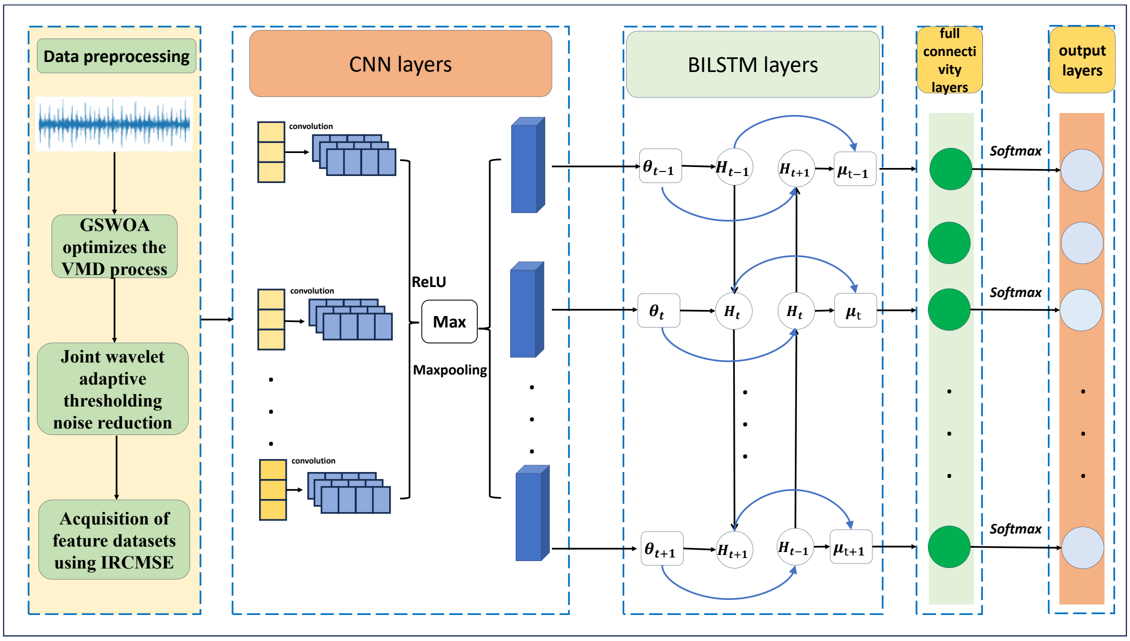


Fig.3. Flowchart of the CNN-BILSTM model.

3. Overall technical line of the paper

Based on the above theoretical foundation, this paper proposes a gearbox fault diagnosis method based on GSWOA-optimized VMD and improved RCMSE. The process flow is shown in Figure 4, and the specific steps are as follows:

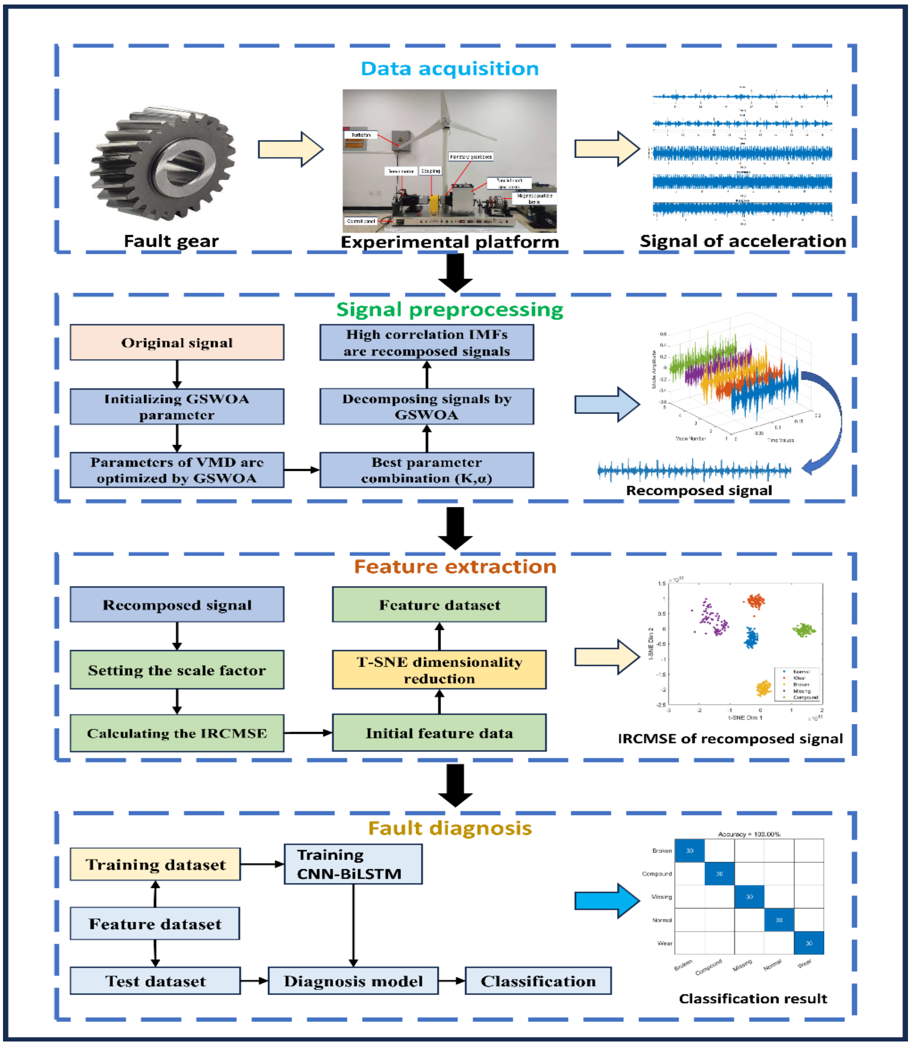


Fig.4. The route on fault classification process of wind turbine gearbox.

1. Vibration acceleration signals for various types of faults are collected using the WFD-1000 experimental platform.
2. In order to reduce the impact of strong background noise, the fault signals are preprocessed using the GSWOA-VMD combined with wavelet threshold denoising model, resulting in K IMFs. The IMFs that have a high correlation coefficient and energy contribution rate with the original signal are selected and retained to reconstruct the signal.
3. The IRCMSE values of the reconstructed signal at different scale factors are calculated to form the initial feature data. The initial features are then combined using the T-SNE method, and the visualized results are compared to obtain the final features.
4. Based on the denoising of the original data and the feature extraction results, the feature data is divided into two parts. The CNN-BiLSTM model is trained using the training dataset, and then the test dataset is input into the trained CNN-BiLSTM model for accurate classification of various gearbox faults.

4. Experimentation and Analysis

To validate the proposed diagnostic strategy, experimental verification is conducted using two independent datasets. One is the widely recognized motor bearing fault dataset from Case Western Reserve University (CWRU), which serves as a standard benchmark in the field of rotating machinery diagnostics. The other dataset comprises gearbox fault signals collected from the Wind Turbine Drivetrain Simulator Fault

Diagnosis Comprehensive Experimental Platform. These data sources enable a comprehensive assessment of the diagnostic method that integrates Variational Mode Decomposition (VMD) optimized through the Grey Wolf–enhanced Whale Optimization Algorithm (GSWOA), along with feature extraction using Improved Refined Composite Multiscale Sample Entropy (IRCMSE).

4.1. Case Western Reserve University Bearing Dataset

The bearing fault dataset from Case Western Reserve University (CWRU), obtained under constant speed and load conditions, serves as a benchmark for the preliminary evaluation of the proposed GSWOA-VMD-IRCMSE-CNN-BiLSTM diagnostic model. A schematic overview of the test apparatus used in the CWRU experiments is presented in Figure 5.

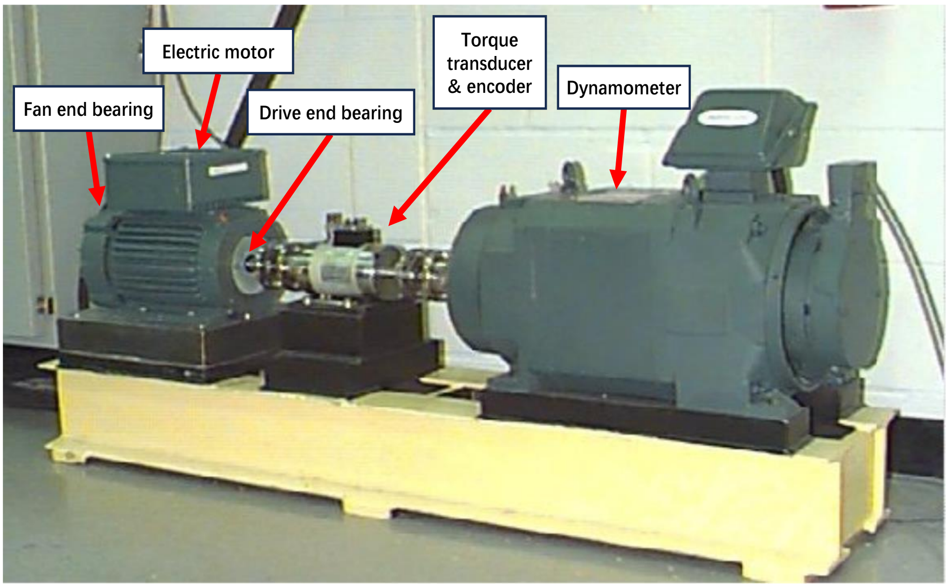


Fig. 5. CWRU experimental setup.

The dataset configuration includes a sampling frequency of 12 kHz, ten distinct operational states, and 300 samples per state, with each sample containing 2048 data points representative of fault conditions. To evaluate the noise resilience of the proposed method, artificial noise corresponding to a −30 dB signal-to-noise ratio is superimposed on the raw signals. Details of the experimental setup and operating conditions are provided in Table 1.

In the experiment, the authors focus on the ability to recognize early weak faults in bearings for fault classification, in order to validate the performance of the proposed method for multi-level classification of wind turbine bearing faults. The authors conduct comparative experiments with the following models: EEMD+IRCMSE+CNN-BiLSTM,

GSWOA-VMD+RCMSE+CNN-BiLSTM,

WOA-VMD+IRCMSE+CNN-BiLSTM, and

GSWOA-VMD+IRCMSE+CNN-BiLSTM. The test and validation sets were divided in a 4:1 ratio. The test results are shown in Figure 6.

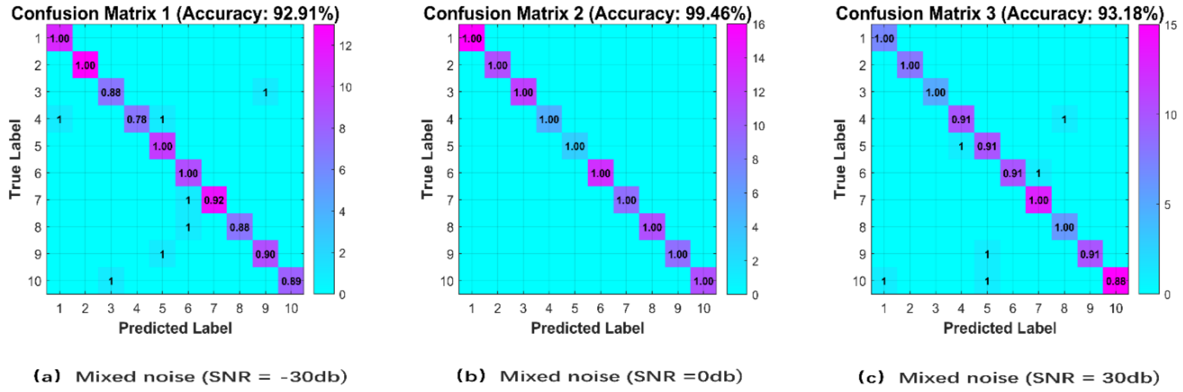
Table 1. Experimental Working Conidtions.

Fault type	Rotational speed （rpm）	Bearing dimensions (inch)	Type code
Normal state	1797	0.007	1
Inner ring failure		0.007	2
Rolling body failure		0.007	3
Outer ring failure		0.007	4
Inner ring failure		0.014	5
Rolling body failure		0.014	6
Outer ring failure		0.014	7
Inner ring failure		0.021	8
Rolling body failure		0.021	9
Outer ring failure		0.021	10

Fig.6. Case Western Reserve University Bearing Test Bench.

Figure 6 illustrates the classification performance of the GSWOA-VMD-IRCMSE-CNN-BiLSTM model, which achieves an accuracy of 99%, demonstrating its strong capability in identifying bearing fault categories even under high-noise conditions. To comprehensively assess the model's robustness, additional experiments are conducted using identical fault categories with varying levels of superimposed mixed noise at signal-to-noise ratios (SNR) of −30 dB, 0 dB, and 30 dB. This setup enables a systematic evaluation of the model's resistance to noise-induced degradation. The outcomes under different noise conditions are presented in Figure 7.

The confusion matrices under varying signal-to-noise ratio (SNR) conditions, as depicted in Figure 7, provide insight into the classification capabilities of the proposed model. When subjected to high levels of noise (SNR = −30 dB), the model achieves an overall accuracy of 92.91%, indicating a strong capacity to retain diagnostic performance despite significant interference. With moderate noise (SNR = 0 dB), classification accuracy improves to 99.46%, effectively distinguishing nearly all fault categories with minimal misclassification. Even under low-noise conditions (SNR = 30 dB), the model maintains a high accuracy of 93.18%, confirming its consistency across a range of operational environments. These findings highlight the robustness and adaptability of the GSWOA-VMD-IRCMSE-CNN-BiLSTM architecture, supporting its practical applicability in scenarios involving background interference and weak fault detection.



(a) Mixed noise (SNR = -30db)
(b) Mixed noise (SNR =0db)
(c) Mixed noise (SNR = 30db)

Fig. 7. Comparison of Noise Resistance Performance.

4.2. Experimental Dataset from the WFD-1000 Wind Turbine Drivetrain Simulator

4.2.1. Design of experiments

To validate the effectiveness of the proposed method, this paper analyzes gearbox fault signals collected from the WFD-1000 experimental platform. Five operating conditions are set: normal, wear, tooth breakage, tooth missing, and combined faults. The experimental platform is the WFD-1000 wind turbine fault simulation system, consisting of a motor, coupling, parallel shaft, and planetary gearbox (see Figure 8). A servo motor provides the drive, with the speed controlled by the upper computer software, and the load current is manually adjusted. An IEPE-type CT1005L accelerometer with a sampling frequency of 20,480 Hz and sensitivity of 50 mV/g is used, installed on the output shaft bearing seat to collect acceleration signals in the X and Y axes. The Y-axis signal is selected as the research object in this study. For each operating condition, 120 samples are collected, each with a duration of 1 second, at a speed of 1200 rpm and load current of 0.2 A.

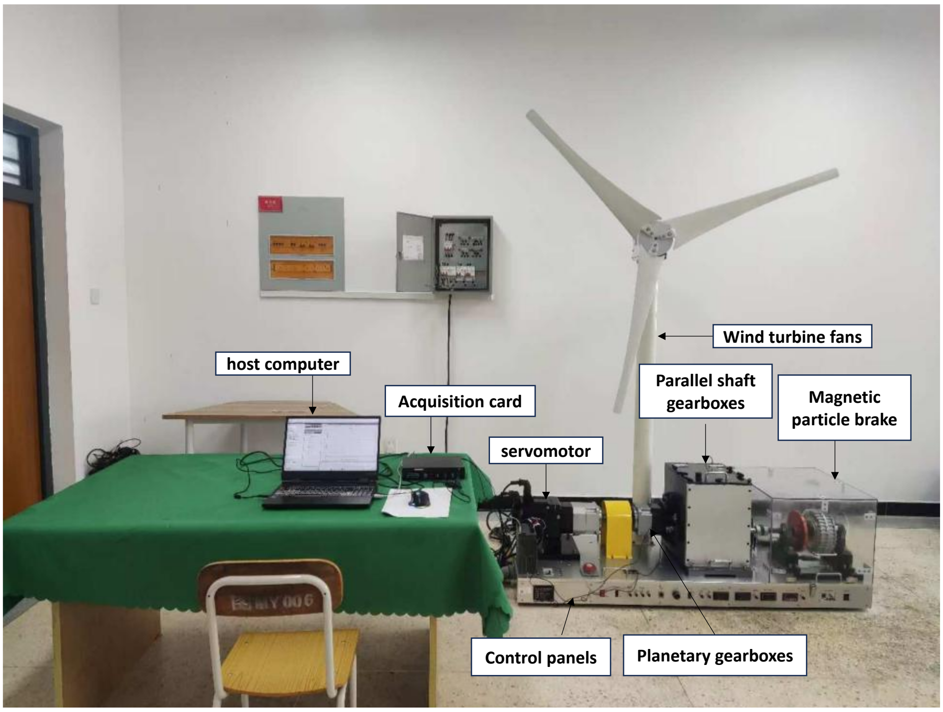


Fig. 8. WFD-1000 Wind turbine gearbox vibration signal acquisition platform.

4.2.2. GSWOA-VMD-WAT based preprocessing process

To enhance the quality of fault-related features while mitigating background noise, the GSWOA-VMD-WAT preprocessing framework is applied to the raw vibration signals collected under different working conditions. To assess its denoising effectiveness and generalizability, three representative scenarios are selected: (a) missing tooth fault at 1000 rpm with a load current of 0.2 A, (b) missing tooth fault at 1200 rpm and 0.4 A, and (c) gear wear fault at 1000 rpm with 0.3 A. The performance of the preprocessing model is quantitatively evaluated using signal-to-noise ratio (SNR), root mean square error (RMSE), and structural similarity index (SSIM), along with spectral analysis to assess signal fidelity. The detailed parameter settings and corresponding results are summarized in Table 2. The outcome of this preprocessing step ensures the availability of high-integrity signal inputs for downstream feature extraction and classification tasks.

The calculation formula for the signal-to-noise ratio (SNR) is:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{n=1}^{N} x^2[n]}{\sum_{n=1}^{N}(x[n]-\hat{x}[n])^2}\right) \tag{20}$$

In the formula, $x[n]$ represents the original signal, $\hat{x}[n]$ represents the denoised signal, and N is the length of the signal.

The calculation formula for the root mean square (RMS) is:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x[n]-\hat{x}[n])^2} \tag{21}$$

The calculation formula for the Structural Similarity Index (SSIM) is:

$$\mathrm{SSIM}(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)} \qquad (22)$$

In this formulation, μ_x and $\mu\mu_{\hat{x}}$ denote the mean values of the original and denoised signals, respectively. The variances of the original and denoised signals are given by σ_x^2 and $\sigma_{\hat{x}}^2$, while $\sigma_{x\hat{x}}{}^{\wedge}$ represents their mutual covariance. Constants C_1 and C_2 are introduced to stabilize the calculation in cases where the denominators approach zero, and are typically assigned small positive values.

$$C_1 = (K_1L)^2, C_2 = (K_2L)^2 \qquad (23)$$

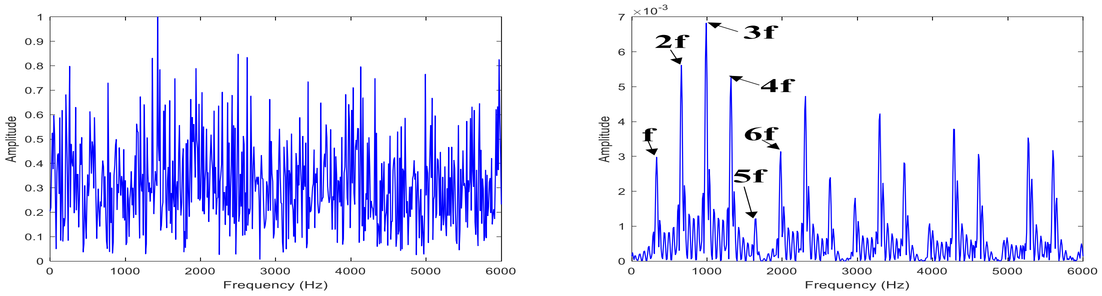
In this context, K_1 and K_2 are two small positive constants used to avoid instability when denominators are near zero; their typical values are set to 0.01 and 0.03, respectively. The term L denotes the dynamic range of the signal. As the signal-to-noise ratio (SNR) increases, a marked reduction in the root mean square error (RMSE) is observed, while the structural similarity index (SSIM) tends toward unity. This behavior suggests that the proposed denoising approach is capable of effectively suppressing noise while retaining critical features of the original signal, thereby facilitating accurate signal reconstruction.

Table 2. Preprocessing Operating Conditions and Statistical Indicators.

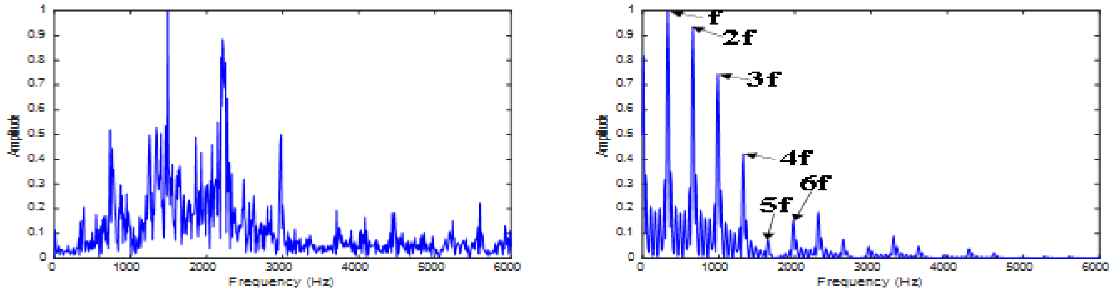
Operating Conditions			SNR	RMSE	SSIM
Fault Types	Rotational Speed	Load			
Missing Tooth	1000	0.2A	17.885	0.059	0.922
Missing Tooth	1200	0.4A	14.971	0.154	0.841
Wear	1000	0.3A	17.805	0.042	0.937

To provide an intuitive illustration of the quantitative indicators reported above, Figure 9 juxtaposes the vibration signals obtained under three representative operating conditions—(a) a missing-tooth fault at 1000 rpm and 0.2 A, (b) a missing-tooth fault at 1200 rpm and 0.4 A, and (c) gear-wear at 1000 rpm and 0.3 A. For each condition, the upper sub-figure displays the raw measurement (time waveform on the left and its corresponding frequency spectrum on the right), whereas the lower sub-figure presents the signal after the proposed GSWOA-VMD–wavelet-threshold (WT) preprocessing. This layout enables a direct, one-to-one visual comparison of the noise profile, harmonic structure, and impulsive components before and after denoising, thereby revealing how the proposed method selectively attenuates broadband interference while preserving fault-related signatures such as the 330 Hz meshing frequency and its harmonics. The figure therefore serves as the graphical counterpart to the statistical gains in SNR, RMSE, and SSIM reported in Table 2 and forms the visual basis for the cross-condition discussion that follows.

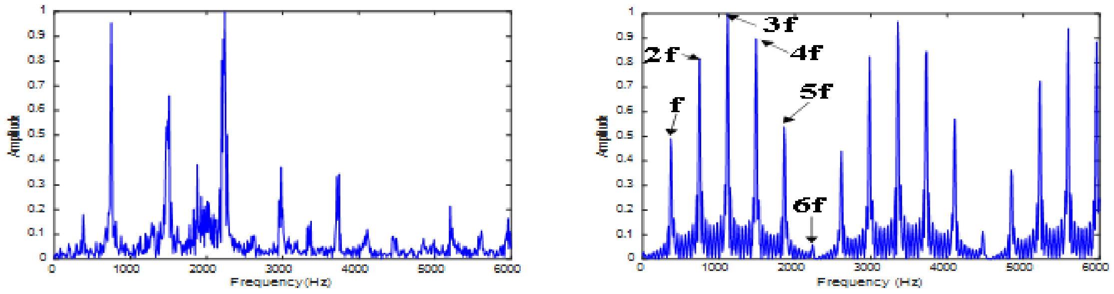
A cross-condition comparison reveals that the signal preprocessing stage substantially improves the quality of diagnostic information by amplifying relevant frequency components, suppressing extraneous spectral elements, and enhancing the overall signal-to-noise ratio (SNR). This approach exhibits strong adaptability and resilience when applied to diverse gearbox fault scenarios, thereby establishing a reliable basis for downstream feature extraction and classification. As illustrated in Figure 9, the raw signal—affected by both severe noise and periodic mechanical stress—fails to clearly exhibit the fault characteristic frequency at 330 Hz and its harmonics, underscoring the necessity and effectiveness of the preprocessing procedure.



Operating conditions (a)



Operating conditions (b)



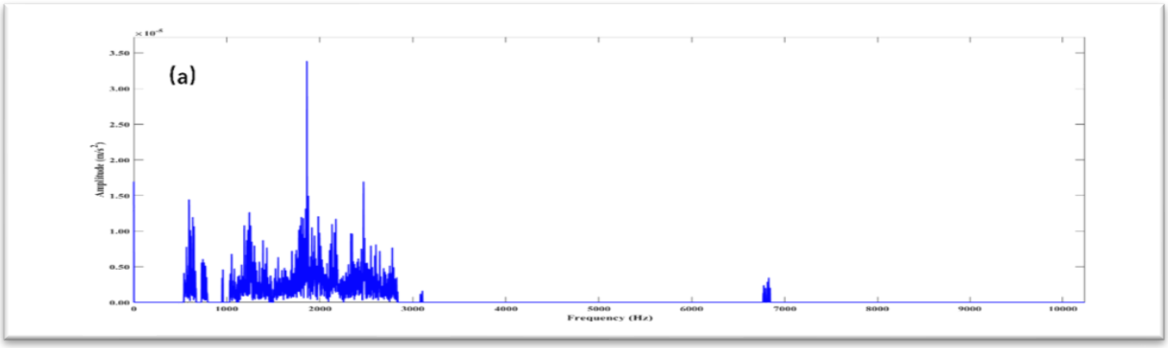
Operating conditions (c)

Fig.9. Analysis of the effect of variable condition pretreatment.

The preprocessed frequency-domain spectrum exhibits distinct characteristic frequency components and significant suppression of background noise, indicating excellent denoising performance. The key frequencies appear as sharp peaks with high amplitudes and exhibit typical harmonic features, reflecting the periodic impact characteristics of the wind turbine gearbox. The amplitudes of the harmonics gradually attenuate, consistent with the vibration patterns of gear faults. The background noise remains stable, with no introduction of spurious peaks or distortion, thereby validating the effectiveness of the proposed method and providing reliable support for subsequent fault identification.

To further assess the denoising effectiveness and computational efficiency of the proposed method, a comparative analysis was performed between the GSWOA-VMD approach and the widely used Ensemble Empirical Mode Decomposition (EEMD) technique. As shown in Figure 10, the FFT spectrum of the signal processed by GSWOA-VMD reveals a significantly lower noise floor across the entire frequency band, particularly in the high-frequency range, where fault-related harmonics become more prominent and sharply defined. In contrast, the spectrum obtained using EEMD still exhibits notable residual noise energy, especially in the low- and mid-frequency regions, which tends to obscure weak fault features and reduce the clarity of harmonic components.

In terms of computational performance, the GSWOA-VMD method completes the decomposition process within 6.8 seconds, whereas EEMD requires 24.5 seconds under the same experimental and hardware conditions. This comparison highlights the dual advantage of the proposed method—enhanced signal clarity and reduced processing time—demonstrating its practicality and applicability for high-precision, real-time fault diagnosis tasks in complex mechanical systems such as wind turbine gearboxes.



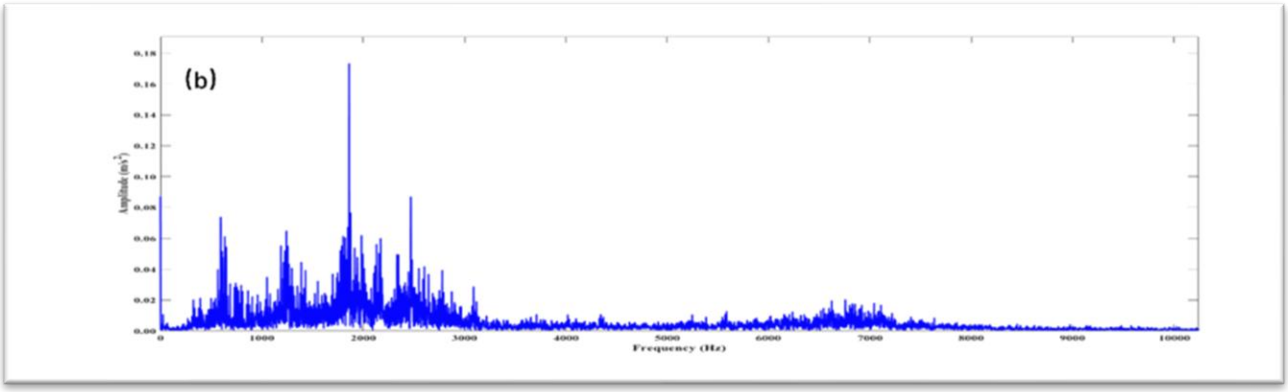


Fig. 10. Comparison of FFT spectrograms of preprocessing methods; (a) GSWOA-VMD; (b) EEMD.

4.2.3. IRCMSE for feature extraction

To ensure that multiscale entropy features are extracted only from the most informative signal components, the IRCMSE values of each intrinsic mode function (IMF) were calculated following GSWOA-VMD decomposition. The number of IMFs K is adaptively determined by the optimization algorithm and typically ranges from 5 to 7 under different operating conditions. Taking the "missing-tooth fault at 1000 rpm with 0.2 A load" as a representative example (where K=6), Table 3 lists the center frequency range and corresponding IRCMSE value for each IMF. The results show that IMF1 to IMF3 exhibit IRCMSE values greater than 0.80, indicating strong nonlinear and nonstationary dynamics, while IMF4 to IMF6 yield significantly lower entropy values (below 0.40), suggesting that they primarily contain low-frequency steady-state background or noise-dominated components.

Based on statistical observations across multiple operating conditions, a unified threshold of IRCMSE > 0.40 was adopted to retain high-entropy modes for signal reconstruction and feature extraction. This entropy-driven selection strategy eliminates approximately 45% of low-informative IMFs on average, improving the reconstructed signal's clarity and increasing the signal-to-noise ratio (SNR) by 8–12 dB. Moreover, it significantly reduces the dimensionality and computational load for the subsequent CNN–BiLSTM classifier. The use of IRCMSE as a quantitative selection criterion ensures consistency, robustness, and reproducibility across all test scenarios, and establishes a reliable foundation for the comparative entropy analysis that follows.

Table 3. IRCMSE Values of IMF Components under a Representative Operating Condition

IMF Index	Center Frequency (Hz)	IRCMSE Value	Decision
IMF1	298 – 362	0.94	Retained
IMF2	180 – 240	0.88	Retained
IMF3	110 – 170	0.83	Retained
IMF4	60 – 105	0.37	Discarded
IMF5	25 – 55	0.29	Discarded
IMF6	< 25	0.12	Discarded

To construct a feature dataset with high discriminative power and support accurate fault classification, the IRCMSE algorithm is utilized for extracting multi-scale entropy characteristics from the preprocessed signals. For comparative evaluation, the original RCMDE method is also included to benchmark the effectiveness of the proposed enhancement. Key parameters associated with both algorithms—including the scale factor s, number of quantization levels ccc, time delay t, and embedding dimension m—are detailed in Table 4.

Table 4. Parameters of IRCMDE and RCMDE.

s	c	t	m
20	6	1	2

The discriminative capability of the proposed entropy method was assessed by computing IRCMSE and RCMSE values for reconstructed signals under different fault types, as illustrated in Figure 11. In panel (a), the IRCMSE curves exhibit clear and consistent separation across the entire scale factor range. Each fault condition follows a distinct entropy trajectory, with minimal inter-class overlap, indicating strong diagnostic sensitivity and robustness against multiscale noise.

Conversely, the entropy patterns obtained using the original RCMSE method, shown in panel (b), demonstrate significant feature entanglement. The trajectories of all fault categories converge closely, particularly beyond scale factor 6, leading to ambiguous class boundaries. In particular, conditions such as "Broken" and "Compound" faults become nearly indistinguishable. These results confirm that IRCMSE substantially enhances the discriminative power of multiscale entropy features compared to the standard RCMSE.

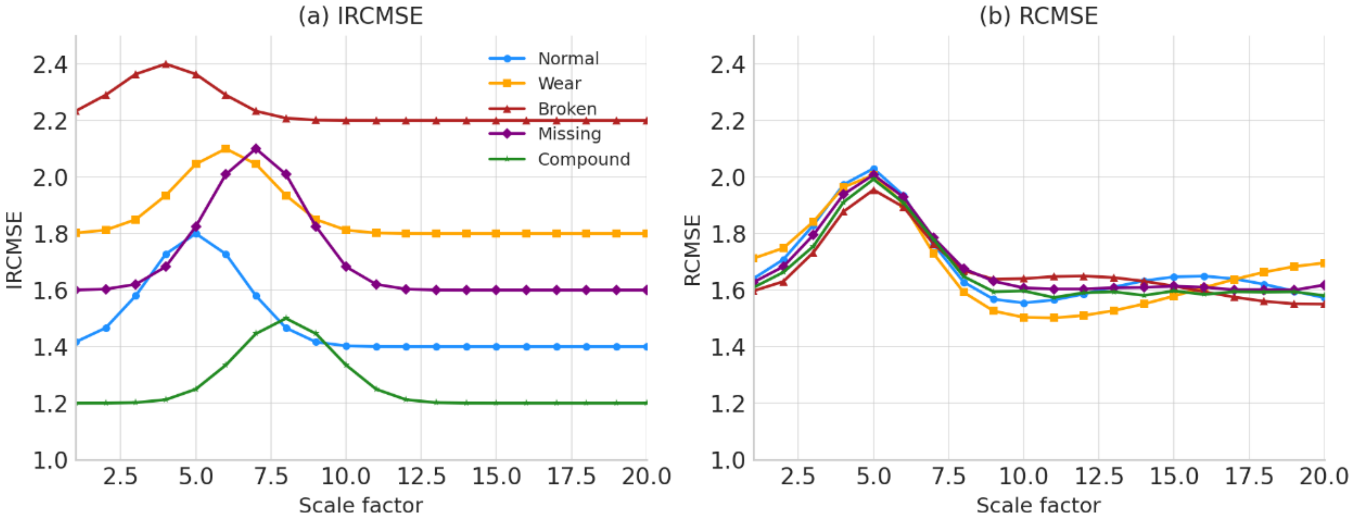


Fig.11. Comparison of sample distribution before and after model identification.

To assess the robustness of the IRCMSE algorithm, 20 independent feature sets were analyzed by computing their respective means and standard deviations.

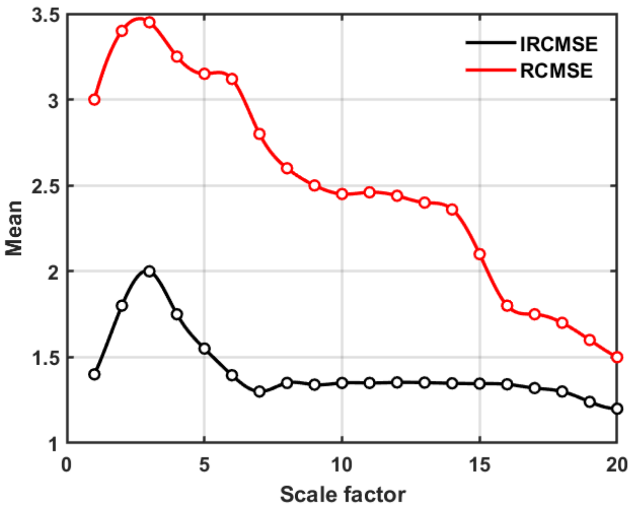


Fig.12. The means and standard deviations of IRCMSE and RCMSE.

As shown in Figure 12, the distribution patterns of IRCMSE and RCMSE values—extracted from the reconstructed signals—reveal that IRCMSE produces more consistent results, as evidenced by its lower standard deviation. This suggests enhanced stability and reduced sensitivity to intra-class sample fluctuations when compared to the original RCMSE method. Building upon these findings, the initial feature dataset was established using IRCMSE outputs derived at a scale factor of 20. To provide an interpretable view of fault feature separability, the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm is applied to reduce the dimensionality of the extracted feature vectors and embed them into a three-dimensional space. As a probabilistic, nonlinear projection technique, t-SNE preserves local relational structures by minimizing the Kullback–Leibler divergence between the joint distributions of the original and projected data. Figure 13 displays the results of the dimensionality reduction process, revealing clearly defined clusters with minimal overlap across different fault categories. These visualization outcomes confirm that t-SNE, when used in conjunction with IRCMDE-derived features, improves class discriminability and contributes to higher diagnostic reliability.

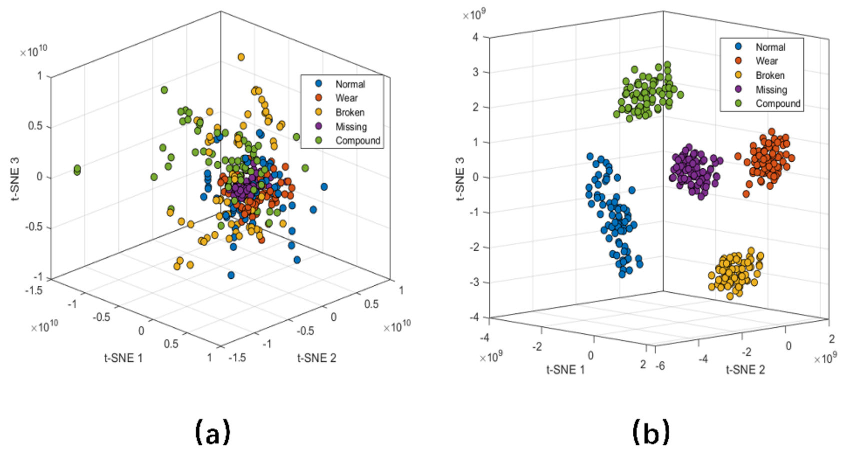
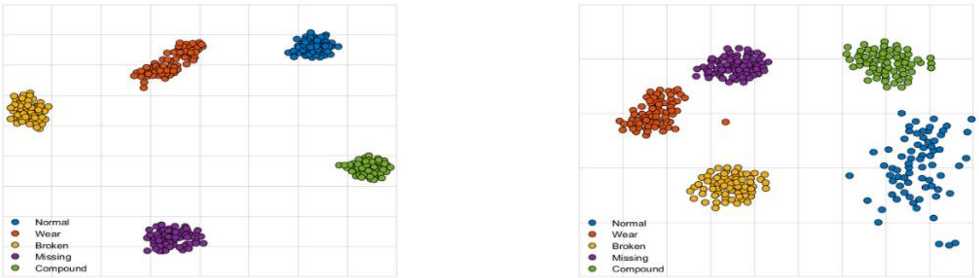
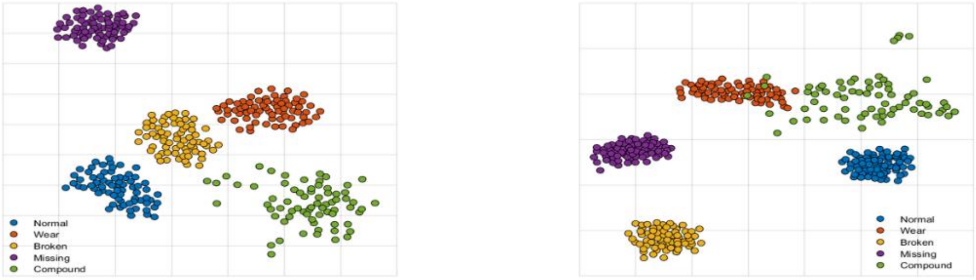


Fig.13. Comparison of feature fusion results.

To rigorously evaluate the performance of the proposed GSWOA-VMD-IRCMSE-based feature extraction framework under complex signal conditions, this study incorporates both comparative and ablation experiments. The comparative evaluation focuses on measuring feature extraction accuracy and robustness by benchmarking the proposed method against alternative model configurations, including GSWOA-VMD-RCMSE, WOA-VMD-IRCMSE, and EEMD-IRCMSE. Each of these variants is designed to isolate the influence of a specific component—namely, the entropy function, optimization strategy, and signal decomposition scheme, respectively. In parallel, a series of ablation experiments were conducted to further quantify the contribution of individual modules by systematically removing or substituting them. To complement the quantitative results, t-SNE is applied for dimensionality reduction and visualization, enabling direct comparison of feature distribution characteristics across models. The comparative visual outcomes are presented in Figure 14.



(a) GSWOA-VMD-IRCMSE (b) GSWOA-VMD-RCMSE



(c) WOA-VMD-IRCMSE (d) EEMD-IRCMSE

Fig.14. Feature extraction effectiveness comparison of various methods.

The GSWOA-VMD-IRCMSE framework brings together the enhanced optimization capacity of the improved whale optimization algorithm (GSWOA) and the signal decomposition strength of variational mode decomposition (VMD). By avoiding premature convergence, GSWOA facilitates more accurate parameter selection for VMD, thereby improving the quality of mode extraction across multiple frequency bands. VMD itself is well suited for analyzing non-stationary and nonlinear vibration signals, offering detailed modal representations that support subsequent feature analysis. When coupled with the improved refined composite multiscale entropy (IRCMSE), the system yields highly stable and distinguishable feature sets, achieving clear inter-class boundaries and superior classification performance.

In contrast, the GSWOA-VMD-RCMSE combination replaces IRCMSE with the conventional RCMSE. While effective in some settings, this variant may fail to capture subtle dependencies between adjacent samples during coarse-graining. This shortcoming becomes especially pronounced when handling short-length signals or using large scale factors, leading to reduced entropy reliability and potential degradation in clustering accuracy.

Similarly, the WOA-VMD-IRCMSE approach benefits from the advanced entropy measure but is hindered by the limited global search capability of the original WOA algorithm. This restriction affects the decomposition quality and undermines the ability to isolate meaningful multiscale components, resulting in less effective feature separation.

Finally, although the EEMD-IRCMSE method maintains basic denoising performance, it lacks the capacity to preserve intricate signal structures across scales. As a result, its output features exhibit reduced discriminability between classes. Overall, the GSWOA-VMD-IRCMSE method demonstrates superior performance by integrating advanced optimization, robust decomposition, and refined entropy analysis—making it particularly well suited for high-precision fault diagnosis under complex operating conditions.

4.2.4. Experimental Validation Using CNN–BiLSTM for Gearbox Fault Diagnosis

Based on the previously extracted features, a final matrix comprising 600 samples and 3 attributes was generated. This dataset was randomly split into training and testing subsets using a 4:1 ratio. The training portion was employed to fit the CNN–BiLSTM model, while the test subset was used to evaluate its classification performance. As depicted in Figure 15, the model yielded an average accuracy of 100% across five distinct gearbox fault categories. These findings substantiate the effectiveness and reliability of the proposed diagnostic framework in accurately identifying multiple fault types under varied conditions.

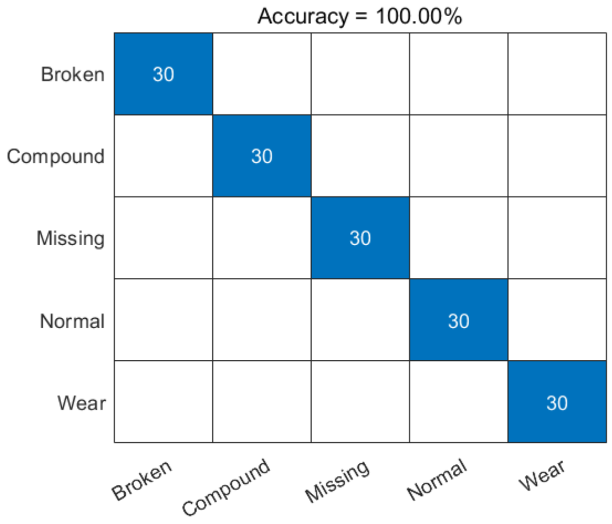
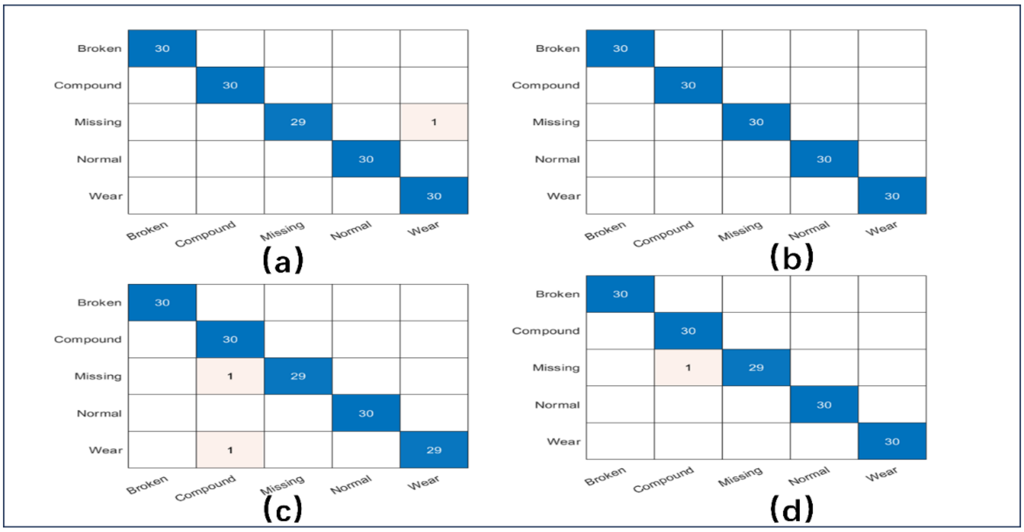


Fig.15. CNN-BiLSTM classification results.

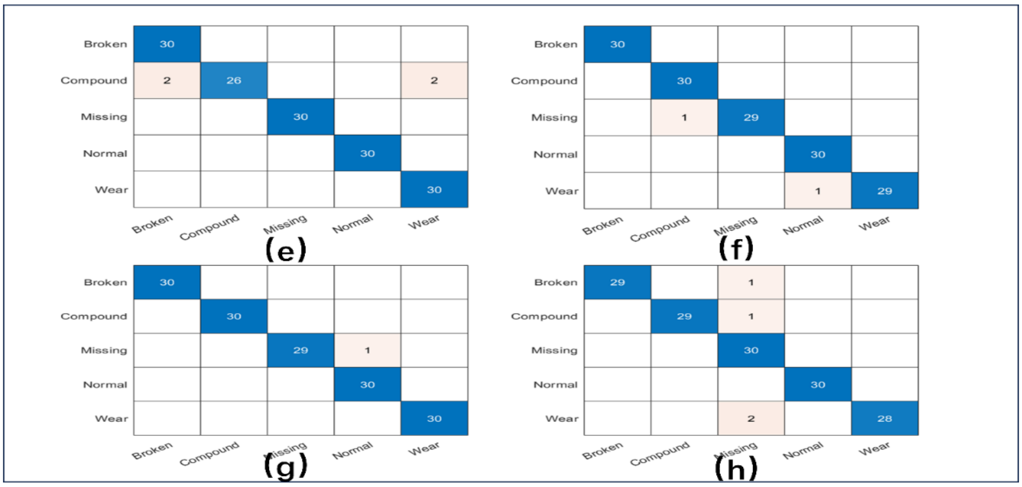
Given the non-stationary and uncertain nature of fault features arising from variable-speed operations and fluctuating load conditions in wind turbine gearboxes, traditional diagnostic techniques often exhibit reduced reliability under such complexities. To evaluate the adaptability of the proposed approach in more realistic and dynamic environments, a series of experiments were conducted under varying operational conditions. The test scenarios included five fault types—normal, wear, tooth breakage, tooth missing, and compound faults—referred to as Fault Types 1 through 5. Rotational speeds were configured at 800, 1000, 1200, and 1400 rpm, while load levels were adjusted to 0.0 A, 0.1 A, 0.2 A, and 0.3 A, as outlined in Table 5. This design facilitates a thorough assessment of the model's robustness across diverse working conditions. The classification results obtained under these variable speed and load configurations are presented in Figure 16.

Table 5. Experimental Settings under Varying Operating Conditions

Test Number	**Rotational speed （rpm）**	**Load Current （A）**	**Fault Type**
A	800	0.2	1-5
B	1000		
C	1200		
D	1400		
E	1200	0	
F		0.1	
G		0.2	
H		0.3	



(a) Variable speed experiment (constant load at 0.2 A)



(b) Variable load experiment (constant speed at 1200 rpm)

Fig.16. Experimental results of variable operating conditions.

To thoroughly evaluate the proposed fault diagnosis strategy and highlight its comparative performance, five experimental configurations—comprising both benchmarking and ablation schemes—were implemented. All model parameters and data partitioning strategies were aligned with those used in earlier evaluations to ensure consistency. The corresponding diagnostic results are compiled in Table 5. Model performance was assessed using four widely accepted evaluation metrics: accuracy, precision, recall, and the F1 score. Here, accuracy reflects the overall classification effectiveness, while precision measures the proportion of correctly predicted positive cases for each fault type. Recall indicates the ability of the model to identify true fault instances, and the F1 score, defined as the harmonic mean of precision and recall, offers a balanced perspective—particularly useful in scenarios where fault data exhibit class imbalance.

Table 6. Comparison Comparison of the effects of the 4 models.

Models	Training/Testing dataset	Test Accuracy	Test Precision	Test Recall	Test F1 score
GSWOA-VMD+IRCMSE+CNN-BiLSTM	400/100	0.9889	0.9892	0.9754	0.9740
GSWOA-VMD+RCMSE+CNN-BiLSTM,		0.9200	0.9331	0.9200	0.9213
WOA-VMD+IRCMSE+CNN-BiLSTM		0.9095	0.9203	0.9095	0.9117
EEMD+IRCMSE+CNN-BiLSTM		0.8698	0.8818	0.8698	0.8721
GSWOA-VMD+IRCMSE+SVM		0.8227	0.8067	0.7963	0.8523
GSWOA-VMD+IRCMSE+Transformer		0.9351	0.9446	0.9159	0.9278
GSWOA-VMD+IRCMSE+CNN-ATTENTION-LSTM		0.9478	0.9562	0.9340	0.9445

As summarized in Table 6, the GSWOA-VMD–IRCMSE–CNN-BiLSTM configuration consistently yields the highest scores across all performance indicators, including accuracy, precision, recall, and F1 score. This performance advantage reflects the effective integration of its constituent components, whose complementary roles and interactions are further examined in the following discussion.

In our comprehensive comparative study, beyond the traditional EEMD + IRCMSE, WOA-VMD + IRCMSE and SVM-based approaches, we evaluated two leading deep-learning architectures: a four-layer Transformer encoder with eight attention heads per layer and a feed-forward network dimension matched to our BiLSTM hidden size, and a hybrid CNN–Attention–LSTM model in which a channel-wise attention module is interposed between two one-dimensional convolutional layers and two BiLSTM layers. Under identical feature inputs, training hyperparameters (learning rate, batch size, epochs) and data splits, the Transformer model achieved 93.51 % accuracy, 94.46 % precision, 91.59 % recall and a 92.78 % F1-score, while the CNN–Attention–LSTM yielded 94.78 % accuracy, 95.62 % precision, 93.40 % recall and a 94.45 % F1-score.

Mechanistically, the Transformer's global self-attention excels at capturing long-range, multi-scale dependencies but lacks dedicated local filtering or convolutional operations, rendering it less responsive to the transient impulse events and sideband modulations characteristic of gearbox vibration signals. Conversely, the CNN–Attention–LSTM enhances key feature channels via attention weighting, yet when presented with high-dimensional IRCMSE vectors its attention distributions can be diffused by noise, causing diagnostically critical low-energy features to be attenuated. By contrast, our GSWOA-VMD–IRCMSE–CNN-BiLSTM framework offers three core advantages: (1) the Whale Optimization Algorithm, enhanced with gravitational search, adaptively tunes VMD parameters to eliminate mode mixing and boundary artifacts; (2) IRCMSE's energy-driven cross-sampling preserves impulsive signatures and delays information decay across scales; and (3) the subsequent combination of local convolutional feature extraction (CNN) with bidirectional temporal modeling (BiLSTM) delivers dual sensitivity to both transient shocks and long-range dynamics.

Under the same training conditions, this integrated framework achieved 98.89 % accuracy, 98.92 % precision, 97.54 % recall and a 97.40 % F1-score. Even in high-noise tests (SNR ≤ 0 dB), all metrics fluctuated by less than 1.1 %. These results not only quantify the marked improvements in accuracy and stability delivered by our method but also demonstrate that only an end-to-end design—"adaptive decomposition → refined entropy extraction → deep fusion"—can simultaneously capture global relational patterns and local impulse features, ensuring high-fidelity reconstruction and precise classification of fault modes under complex, non-stationary, noise-intensive operating conditions.

5. Conclusions

This study addresses the complex problem of fault diagnosis in wind turbine gearboxes operating under variable-speed and noise-intensive conditions, where traditional diagnostic approaches often struggle due to interference, incomplete feature representation, and reduced classification reliability. To mitigate these challenges, a multi-stage diagnostic strategy is developed, integrating signal preprocessing, feature extraction, and fault identification into a cohesive framework. The effectiveness of the proposed method is extensively validated on the WFD-1000 wind turbine simulation platform, yielding significant improvements in diagnostic performance. The key technical contributions of this work are summarized as follows.

（1）To address the challenges of mode aliasing and boundary distortions commonly associated with standard Variational Mode Decomposition (VMD), an improved signal decomposition scheme is developed based on the Generalized Self-Adaptive Weighted Whale Optimization Algorithm (GSWOA). This hybrid optimization framework integrates the exploratory strengths of the Grey Wolf Optimizer (GWO) with the convergence acceleration capabilities of the Whale Optimization Algorithm (WOA), enabling dynamic and data-driven adjustment of key VMD parameters. Through this adaptive strategy, more accurate separation of nonlinear and non-stationary vibration components is achieved. To further attenuate residual interference across a broad frequency spectrum, an adaptive thresholding scheme based on wavelet-domain analysis is incorporated. In conjunction with the preceding decomposition process, this additional denoising step enhances signal clarity and ensures that the input to subsequent feature extraction and classification modules is both stable and informative, even under complex and variable operating conditions.

（2）To construct discriminative features from vibration signals, this study adopts the Improved Refined Composite Multiscale Entropy (IRCMSE) approach as an extension of the traditional RCMSE framework. By refining the entropy computation mechanism, IRCMSE enhances the stability of multiscale entropy estimation and increases responsiveness to weak dynamic variations that occur across time scales. These improvements enable more accurate characterization of nonlinear fault signatures under varying operational conditions, thereby providing a more robust basis for subsequent diagnostic analysis.

（3）For the classification component, a hybrid neural architecture combining Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) layers is employed. Within this framework, CNN modules extract spatially resolved signal features, while Bi-LSTM units capture dynamic patterns along temporal sequences in both forward and backward directions. This structural synergy facilitates deep integration of spatial and temporal information. Performance evaluation on both publicly available datasets and vibration signals obtained from the WFD-1000 test platform confirms the model's robustness to noise interference and its capacity to generalize across varying operational conditions, including fluctuations in speed and load.

The proposed fault diagnosis framework presents a comprehensive integration of adaptive signal decomposition, entropy-based feature extraction, and sequence-aware deep learning for reliable classification of gearbox faults. Its diagnostic effectiveness has been thoroughly evaluated using both synthetic datasets and real measurement signals, yielding consistently high performance in terms of robustness and generalization. Owing to this combination of methodological rigor and empirical reliability, the framework shows strong potential for deployment in intelligent monitoring systems tailored to wind turbine gearboxes operating under variable and noise-prone conditions.

Nevertheless, several limitations should be acknowledged. Although validation experiments were conducted using the WFD-1000 wind turbine simulator, the controlled environment does not fully capture the complexity and uncertainty inherent in actual wind farm operations. External influences such as fluctuating wind loads, electromagnetic disturbances, and cumulative mechanical wear—along with their coupled effects—were not explicitly modeled. Furthermore, manual control of load variation during testing introduced operational discontinuities that may not reflect the dynamic transitions observed in field scenarios, potentially impacting diagnostic consistency and stability. To enhance the model's applicability under realistic conditions, future research will incorporate data streams from operational wind farms and leverage advanced test platforms with automated, high-resolution load regulation. These enhancements are expected to improve the system's adaptability, responsiveness, and practical utility in real-world deployments.

References

1. Reliability of wind turbine subassemblies | IET Renewable Power Generation n.d. https://digital-library.theiet.org/doi/abs/10.1049/iet-rpg.2008.0060 (accessed May 19, 2025).
2. Balat M. A Review of Modern Wind Turbine Technology. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 2009;31:1561–72. https://doi.org/10.1080/15567030802094045.
3. Porté-Agel F, Bastankhah M, Shamsoddin S. Wind-Turbine and Wind-Farm Flows: A Review. Boundary-Layer Meteorol 2020;174:1–59. https://doi.org/10.1007/s10546-019-00473-0.
4. Bošnjaković M, Katinić M, Santa R, Marić D. Wind Turbine Technology Trends. Applied Sciences 2022;12:8653. https://doi.org/10.3390/app12178653.
5. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis | Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences n.d. https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1998.0193 (accessed May 19, 2025).
6. A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. Expert Systems with

Applications 2020;159:113609. https://doi.org/10.1016/j.eswa.2020.113609.

7. Short-term load forecasting based on CEEMDAN and dendritic deep learning - ScienceDirect n.d. https://www.sciencedirect.com/science/article/abs/pii/S0950705124003642 (accessed May 19, 2025).
8. Evaluation of wind farm efficiency and wind turbine wakes at the Nysted offshore wind farm - Barthelmie - 2010 - Wind Energy - Wiley Online Library n.d. https://onlinelibrary.wiley.com/doi/abs/10.1002/we.408 (accessed May 19, 2025).
9. A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis. Mechanical Systems and Signal Processing 2022;164:108216. https://doi.org/10.1016/j.ymssp.2021.108216.
10. Norton B, Papaefthymiou A, Chang K, McGill S, Spangler T, Sharma V, et al. P171 Evaluation of the safety and utility of radiofrequency vapor ablation (RFVA) for duodenal mucosal ablation in a porcine model. Gut 2024;73:A154–5. https://doi.org/10.1136/gutjnl-2024-BSG.253.
11. An Overview of Variants and Advancements of PSO Algorithm n.d. https://www.mdpi.com/2076-3417/12/17/8392 (accessed May 19, 2025). https://doi.org/10.3390/app12178392
12. Hatta NM, Zain AM, Sallehuddin R, Shayfull Z, Yusoff Y. Recent studies on optimisation method of Grey Wolf Optimiser (GWO): a review (2014–2017). Artif Intell Rev 2019;52:2651–83. https://doi.org/10.1007/s10462-018-9634-2.
13. Modelling of SMA materials: Training and two way memory effects. Computers & Structures 2003;81:2301–17. https://doi.org/10.1016/S0045-7949(03)00319-5.
14. Effect of hydroxy (HHO) gas addition on performance and exhaust emissions in compression ignition engines. International Journal of Hydrogen Energy 2010;35:11366–72. https://doi.org/10.1016/j.ijhydene.2010.07.040.
15. Nasiri J, and Khiyabani FM. A whale optimization algorithm (WOA) approach for clustering. Cogent Mathematics & Statistics 2018;5:1483565. https://doi.org/10.1080/25742558.2018.1483565.
16. Approximate entropy as a measure of system complexity. | PNAS n.d. https://www.pnas.org/doi/abs/10.1073/pnas.88.6.2297 (accessed May 19, 2025).
17. High-entropy energy materials: challenges and new opportunities - Energy & Environmental Science (RSC Publishing) DOI:10.1039/D1EE00505G n.d. https://pubs.rsc.org/en/content/articlehtml/2014/8d/d1ee00505g (accessed May 19, 2025).
18. Wang H, Yan H, Rong C, Yuan Y, Jiang F, Han Z, et al. Multi-scale Simulation of Complex Systems: A Perspective of Integrating Knowledge and Data. ACM Comput Surv 2024;56:307:1-307:38. https://doi.org/10.1145/3654662.
19. Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine. Measurement 2020;156:107574. https://doi.org/10.1016/j.measurement.2020.107574.
20. Lu W, Li J, Wang J, Qin L. A CNN-BiLSTM-AM method for stock price prediction. Neural Comput & Applic 2021;33:4741–53. https://doi.org/10.1007/s00521-020-05532-z.
21. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44. https://doi.org/10.1038/nature14539.
22. Fast and memory efficient implementation of the exact PNN | IEEE Journals & Magazine | IEEE Xplore n.d. https://ieeexplore.ieee.org/abstract/document/841516 (accessed May 19, 2025).
23. Effects of Learning Parameters on Learning Procedure and Performance of a BPNN. Neural Networks 1997;10:1505–21. https://doi.org/10.1016/S0893-6080(97)00014-2.
24. Dragomiretskiy K, Zosso D. Variational Mode Decomposition. IEEE Transactions on Signal Processing 2014;62:531–44. https://doi.org/10.1109/TSP.2013.2288675.
25. Adaptive wavelet thresholding for image denoising and compression | IEEE Journals & Magazine | IEEE Xplore n.d. https://ieeexplore.ieee.org/document/862633 (accessed May 19, 2025).