



Fusion-driven fault diagnosis based on adaptive tuning feature mode decomposition and synergy graph enhanced transformer for bearings under noisy conditions

Lin Zhu^{a,*}, Jin Wang^a, Min Chen^b, Lintong Liu^a

^a School of Mechanical Engineering, Yangzhou University, Yangzhou 225001, China

^b School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China



ARTICLE INFO

Keywords:

Adaptive tuning feature mode decomposition
Synergy graph enhanced transformer
Fault diagnosis
Synergy graph representation module
Noise interference

ABSTRACT

Bearing fault diagnosis is critical for maintaining the reliability of health monitoring in electromechanical systems. However, traditional feature extraction methods often struggle to accurately capture fault information in complex environments. This paper proposes an adaptive tuning feature mode decomposition (ATFMD) method, based on the intrinsic signal characteristics, to effectively extract robust features. ATFMD dynamically adjusts to complex fault signals, providing more representative feature inputs and pruning redundant features induced by noise. Additionally, given the limitations of single-dimensional feature domains in fully revealing fault information, this study constructs feature topology graphs by mapping spatial phase and time-frequency characteristics, offering a comprehensive representation of fault information. To achieve complementary fusion of fault information within the feature topological graph, this paper proposes the synergy graph enhanced transformer (SGET). SGET optimizes the fusion process by reinforcing feature interactions through its synergy graph representation module. Additionally, a hierarchical cross-attention mechanism is employed to modulate attention distribution across feature dimensions, enhancing the sensitivity to critical features during fusion. Experimental validation was conducted on two distinct rotating machinery transmission systems. The results demonstrate that the proposed method maintains exceptional robustness and generalization, even in the presence of severe noise and complex fault conditions. Compared to the leading methods, including MS-DGCNs, CapsFormer, TFT, and ConvFormer, the proposed method achieves notable accuracy improvements of 8.13 %, 8.71 %, 11.94 %, and 6.25 %, respectively, under challenging conditions.

1. Introduction

The health status of rolling bearings is a crucial source of information for health monitoring and predictive maintenance decision-making in mechanical systems (Lin, Shao, Zhou, Cai, & Liu, 2023; Gawde et al., 2023). However, the continuous operation of bearings in complex and harsh environments makes them susceptible to failure. These unexpected failures can disrupt mechanical equipment and even cause catastrophic safety incidents (Qin, Liu, & Mao, 2024). Therefore, the robust extraction of valuable fault features and precise diagnosis of faults are essential for ensuring the operational safety of mechanical systems.

Local faults of bearings usually manifest as periodic pulses within

vibration signals. These periodic pulse components carry important fault information (Alavi, Ohadi, & Niaki, 2022). However, fault information in the raw signal is highly susceptible to noise contamination (Li et al., 2023). To address this challenge, signal decomposition techniques have been widely applied in fault feature identification due to their excellent multicomponent decoupling capabilities (Rajabi, Azari, Santini, & Flammini, 2022; Chen et al., 2023). These techniques decouple complex signals into independent multi-component modes, effectively enhancing the effectiveness of fault feature analysis (Li, Liu, Wu, & Chen, 2020). Currently, common signal decomposition techniques encompass empirical mode decomposition (EMD) (Kedadouche, Thomas, & Tahan, 2016), local mean decomposition (LMD) (Zuo, Xu, Zhang, Xiahou, & Liu, 2022), variational mode decomposition (VMD) (Gharesi, Arefi, Razavi-

* Corresponding author.

E-mail addresses: linzhu@yzu.edu.cn (L. Zhu), mz120220881@stu.yzu.edu.cn (J. Wang), min.chen@xjtlu.edu.cn (M. Chen), mz120230851@stu.yzu.edu.cn (L. Liu).

Far, Zarei, & Yin, 2020), and symplectic geometry mode decomposition (SGMD) (Pan, Zhang, Cheng, & Zheng, 2023). Many researchers effectively applied these techniques in both academic and industrial settings to enhance fault feature extraction. Cao, Sun, Xie, and Li, (2021) utilized EMD combined with feature fusion filtering to develop a fault identification method for railway point machines. Touti, Salah, Bacha, and Chaari, (2022) proposed a frequency detection method based on LMD, identifying the frequency domain fault characteristics of the motor. Kumar, Zhou, and Xiang, (2021) introduced a defect identification technique for centrifugal pump bearings based on VMD, integrating neutrosophic cross entropy to enhance the extraction of potential fault modes. Cheng, Yang, Hu, Cheng, and Cheng, (2021) proposed an adaptive weighted SGMD method that improves early gear fault diagnosis by considering both fault information and impact energy. While these traditional signal decomposition techniques have improved the efficiency of fault feature extraction, they fall short in providing sufficient decomposition mechanisms for periodic pulse signals. This limitation hinders the accurate extraction of fault feature from nonlinear signals, especially in the presence of background noise and random impacts.

Given the inherent impulsiveness and periodicity of fault signals, there is an urgent need to develop advanced methods that capable of more effectively elucidating the fundamental mechanisms of periodic pulse features. Inspired by the concept of blind deconvolution, Miao, Zhang, Li, Lin, and Zhang, (2022) proposed a novel interpretable signal decomposition theory known as feature mode decomposition (FMD). FMD employs correlation kurtosis to guide the parameter updates of finite impulse response (FIR) filter banks. The filter banks are capable of enhancing the detection of impulsive and periodic components within signals. However, one notable limitation of FMD is the need to predefine the initial parameters of the FIR filter banks, which limits its adaptability. To address this limitation, research on this deficiency has been conducted by several scholars. He et al., (2023) focused on optimizing the filter length of FMD by integrating an autoregressive model, utilizing the weighted envelope harmonic-to-noise ratio as an evaluation metric. This approach effectively extracts fault features, even amidst strong harmonic interference. Chauhan et al., (2024) proposed a sparsity impact measure index (SIMI), which leverages the sparsity and impulsiveness of fault signals to guide the adaptive FMD. This method enables accurate diagnosis of unknown defects by effectively extracting energy features. Yan, Yan, Xu, and Yuen, (2023) proposed an innovative approach by combining multivariate feature mode decomposition (MFMD) with the multichannel comprehensive index, which optimizes the extraction of multichannel modal components rich in fault information. The above study achieved promising advancements by constructing composite evaluation metrics that enhance the adaptability of FMD to impulse signals. Despite the progress achieved, there are still challenges in fully elucidating the coupling characteristics of signal components in complex conditions, as well as in enhancing sensitivity to subtle faults.

Advanced signal decomposition techniques provide more representative fault characteristic inputs for diagnostic tasks. However, relying solely on these methods for fault diagnosis presents significant limitations when handling large-scale data. With the progression of artificial intelligence, data-driven models have rapidly evolved, enabling more efficient fault analysis and causal inference (Wang, Zhuang, Tao, Paszke, & Stojanovic, 2023). Notably, research focused on deep learning network frameworks has received considerable attention (Wang, Jiang, Zhu, Wang, & Liu, 2022). Vashishtha and Kumar (2022) introduced a fault diagnosis method using the time-varying filter-based EMD with CNN. This method aids in fault identification by constructing scale diagrams of the dominant intrinsic mode functions. Dibaj, Ettefagh, Hasannejad, and Ehghaghi, (2021) proposed an intelligent diagnostic method that integrates VMD with CNN for the diagnosis of untrained compound faults. This approach, trained exclusively on single-fault data, successfully detected compound faults in bearings with high

accuracy. Song, Wu, Song, Stojanovic, and Tejado, (2024) proposed an innovative approach that combines a fuzzy wavelet neural network with signal filtering techniques to effectively manage unmanned aerial vehicle actuator faults. This method employs a self-triggering mechanism to optimize the signal update, improving the identification of fluctuations in tracking error characteristics. Tao, Shi, Qiu, Jin, and Stojanovic, (2023a) proposed a method utilizing the fast fourier transform (FFT) to process signals as graph nodes. Additionally, they developed a KNN graph construction technique based on pooling and fuzzy distance calculations. The dynamic graph attention network improved the distribution of attention weights, demonstrating superior diagnostic performance. These works have yielded promising results. The integration of deep learning models with advanced signal processing methods has successfully mitigated interference during feature recognition, thereby improving the robustness and stability of the models. However, the above methods exhibit certain limitations. Current approaches are restricted by convolutional kernels or limited receptive fields, which hinder their ability to capture the long-term dependencies of fault features (Mo, Zhang, & Tsui, 2021). This limitation results in representative fault features being overwhelmed under strong noise interference.

Nowadays, a novel model termed Transformer has garnered considerable attention due to its unique self-attention mechanism. Transformer transcends the limitations of traditional local receptive fields, enabling the exploration of global relationships between signal segments through parallel computation (Yang, Qin, Pang, & Huang, 2021). The self-attention mechanism effectively directs the model to focus on pulse signal segments critical for fault detection, enhancing the robustness of model. The application of the Transformer in the field of fault diagnosis has become a research hotspot. Zhu et al., (2023) proposed a feature enhancement method based on periodic representations for Transformers. Fault information from varying input lengths is effectively extracted by this model, which demonstrates remarkable diagnostic performance. Parvin, Faiz, Qi, Kalhor, and Akin, (2023) proposed a transformer neural network (TNN) based on multi-head attention mechanisms. The proposed TNN focuses on specific parts of the signal, enabling precise diagnostics of the severity of interturn short-circuit faults. Ding, Jia, Miao, and Cao, (2022) integrated a self-attention mechanism with residual connections into the time-frequency Transformer (TFT) architecture. The TFT focuses on the time-frequency characteristics of signals, achieving efficient fault diagnosis. Wang, Liu, Jia, Zhao, and Li, (2024) proposed the multi-scale enhanced convolution Transformer network (MSECTN), which combines multi-scale convolution and the self-attention mechanism. By leveraging Fourier transformation for efficient frequency feature modeling, MSECTN effectively addresses the high randomness and complexity of intermittent faults. The aforementioned studies demonstrate that Transformers excel in capturing temporal dependencies and global information from features, making significant contributions to the field of fault diagnosis. Nevertheless, several challenges remain that warrant further investigation.

Although traditional Transformer architectures demonstrate exceptional performance in handling sequential signal data, their efficacy in processing data within non-Euclidean spaces has not met the anticipated expectations. However, non-Euclidean space data is critical for a comprehensive understanding of the interactions among fault information (Li et al., 2022). Most existing studies focus on analyzing the time-frequency attributes of signal sequence data. However, data that rely exclusively on time-frequency information highly vulnerable to strong noise interference, and insufficiently capture the dynamic characteristics generated by vibrational impulses (Tao, Qiu, Chen, Stojanovic, & Cheng, 2023b). Additionally, current methods primarily depend on traditional attention mechanisms for the integration of complex features (Lv et al., 2022). Nevertheless, direct linear combination or concatenation typically falls short in capturing the synergy and complementarity between key features.

Inspired by these motivations, this paper proposed a novel fault

diagnosis model for rolling bearings to overcome the above challenges. To address the challenge of robust feature extraction from raw signals, this paper utilizes adaptive tuning feature mode decomposition (ATFMD) method, leveraging the periodic pulse characteristics of fault signals. Bayesian optimization is employed for adaptive tuning of complex signals, improving the environmental adaptability of ATFMD. A novel health indicator, defined as the multilayer pulse characteristic integration measure (MPCIM), is proposed to boost the sensitivity to subtle fault features. Moreover, considering the limitations of one-dimensional feature domains in revealing fault information, this study expands feature representation through image encoding and time-frequency analysis. The feature topology graph is constructed by extracting and mapping spatial phase and time-frequency features, with node interactions defined through comprehensive distance metrics. This paper introduces the synergy graph enhanced transformer (SGET) for the complementary fusion of fault information within feature topology graphs. The SGET optimizes the information fusion by reinforcing feature interactions through the synergy graph representation module (SGRM). Additionally, a hierarchical cross-attention mechanism is employed to modulate attention distribution across feature dimensions, improving sensitivity to key features during the fusion process. Our main contributions of this study are as follows:

- A signal processing method based on ATFMD is developed to adaptively tune noisy and complex fault signals. This method effectively prunes redundant features caused by noise, providing more robust inputs for diagnostic models.
- Feature expansion is achieved by employing image encoding and time-frequency analysis techniques, with feature topology graphs constructed by mapping spatial phase and time-frequency characteristics.
- This study introduces the SGET for complementary fusion of fault information in feature topology graphs. SGET enhances feature interaction through the SGRM and employs the hierarchical cross-attention mechanism to sharpen the perception to essential features during fusion.
- Experimental validation was conducted on two rotating machinery transmission systems. The results demonstrate that the proposed ATFMD-SGET method maintains exceptional robustness and generalization, even under challenging noise and fault conditions.

2. Adaptive tuning feature mode decomposition

$$\mathbf{W}_M = \begin{bmatrix} \left(\prod_{m=0}^M u_k[1-mT_s] \right)^2 & 0 & \dots & 0 \\ 0 & \left(\prod_{m=0}^M u_k[2-mT_s] \right)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \left(\prod_{m=0}^M u_k[N-L+1-mT_s] \right)^2 \end{bmatrix} \frac{1}{\sum_{n=1}^{N-L+1} u_k[n]^{M-1}}$$

2.1. Basic theory of FMD

Based on the theory of blind deconvolution, a series of finite impulse response (FIR) filters is devised by FMD to extract the distinct modes of the signal (Miao, Zhang, Li, Lin, and Zhang, 2022). Throughout the dynamic updating of FIR filter coefficients, each mode is progressively aligned closer to the deconvolution objective function. Correlated kurtosis (CK) is employed as the deconvolution objective function to deconvolve each sub-signal. For a vibration signal $x(t)$ of length N , FMD is characterized as a constrained optimization problem, which is

formulated as follows:

$$\begin{aligned} \operatorname{argmax}_{\{f_k(l)\}} \left\{ CK_M(\mathbf{u}_k) = \sum_{n=1}^N \left(\prod_{m=0}^M u_k(n-mT_s) \right)^2 / \left(\sum_{n=1}^N u_k(n)^2 \right)^{M+1} \right\} \\ \text{s.t. } u_k(n) = \sum_{l=1}^L f_k(l)x(n-l+1) \end{aligned} \quad (1)$$

where M represents the shift order of CK. $u_k(n)$ is the k^{th} mode, $f_k(l)$ is the k^{th} FIR filter with length L . T_s is the period defined by the number of samples, expressed as:

$$T_s = f_s \cdot T \quad (2)$$

where f_s is the sampling frequency, and T is the period associated with the fault.

For the FIR filter banks, an initialization strategy employing the Hanning window is applied. Hanning window is used to evenly divide the signal bandwidth into K segments, thereby enhancing the filtering performance. The lower and upper cutoff frequencies for each mode are denoted as f_l and f_u , respectively. f_l and f_u can be defined as:

$$\begin{cases} f_l = k f_s / 2K \\ f_u = (k+1) f_s / 2K \end{cases}, k = 0, 1, 2, \dots, K-1 \quad (3)$$

where K represents the total number of filters constituting the FIR filter bank.

The constrained problem outlined in Eq. (1) can be solved through iterative eigenvalue decomposition. The mode can be represented in the form of a matrix:

$$\mathbf{u}_k = \begin{bmatrix} u_k(1) \\ \vdots \\ u_k(N-L+1) \end{bmatrix} = \mathbf{X}\mathbf{f}_k = \begin{bmatrix} x(1) & \dots & x(L) \\ \vdots & \ddots & \vdots \\ x(N-L+1) & \dots & x(N) \end{bmatrix} \begin{bmatrix} f_k(1) \\ \vdots \\ f_k(L) \end{bmatrix} \quad (4)$$

The CK of the mode can be defined as:

$$CK_M(\mathbf{u}_k) = \frac{\mathbf{u}_k^Q \mathbf{W}_M \mathbf{u}_k}{\mathbf{u}_k^Q \mathbf{u}_k} \quad (5)$$

where the superscript Q denotes the conjugate transpose operation, \mathbf{W}_M is used to control the weighted correlation matrix, which is expressed as:

Therefore, the expression for Eq. (5) can be updated as:

$$CK_M(\mathbf{u}_k) = \frac{\mathbf{f}_k^Q \mathbf{X}^Q \mathbf{W}_M \mathbf{X} \mathbf{f}_k}{\mathbf{f}_k^Q \mathbf{X}^Q \mathbf{f}_k} = \frac{\mathbf{f}_k^Q \mathbf{R}_{XWX} \mathbf{f}_k}{\mathbf{f}_k^Q \mathbf{R}_{XX} \mathbf{f}_k} \quad (7)$$

where \mathbf{R}_{XWX} and \mathbf{R}_{XX} represent the weighted correlation matrix and the correlation matrix.

Optimizing the filter coefficients to maximize Eq. (7) essentially constitutes solving a generalized eigenvalue problem. The solution to this problem is the eigenvector corresponding to the maximum

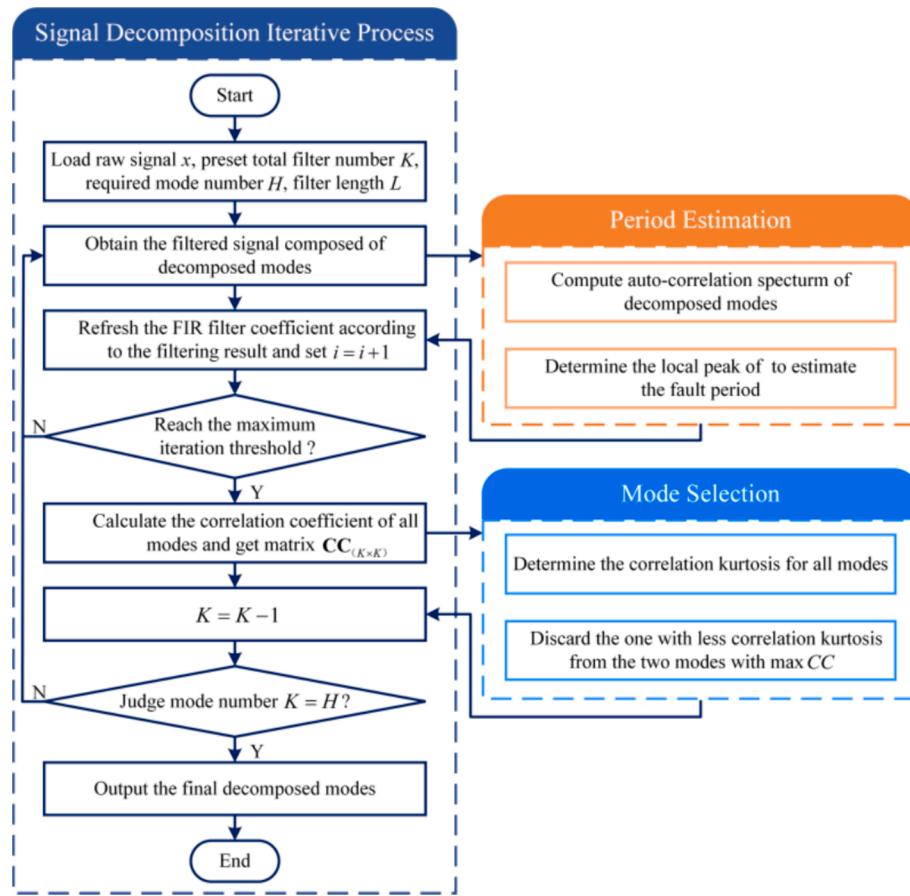


Fig. 1. Flowchart of FMD technique.

eigenvalue λ .

$$\mathbf{R}_{\mathbf{xw}x}\mathbf{f}_k = \mathbf{R}_{\mathbf{xx}}\mathbf{f}_k\lambda \quad (8)$$

where λ corresponds to maximum CK. The filter coefficients \mathbf{f}_k are updated by solving Eq. (8), to progressively approach the filtered signal with maximum CK.

The FIR filter banks provide an initial method for decomposition. Furthermore, the number of filters (K) and the number of required modes (H) need to be preset in FMD. Consequently, the correlation coefficient (CC) is utilized to filter modes until the required number of modes is achieved. The correlation matrix $\mathbf{CC}_{(KxK)}$ with $K \times K$ is constructed by the correlation coefficient between all mode pairs.

The detailed procedural flowchart of FMD technique is illustrated in Fig. 1. It can be seen that the implementation of FMD involves several key steps including the design of FIR filters, the updating of filters, the estimation of fault period, and mode selection. From the above formulas, it is evident that the modulus K of the filter and the filter length L are crucial parameters for achieving optimal performance in FMD. However, without prior information, inappropriate parameter settings for FMD can significantly impact its performance and efficiency. Consequently, this study aims to identify the optimal filter bank parameters, enabling FMD to adaptively adjust to complex signals in noisy environments.

2.2. Bayesian optimization

In previous studies, metaheuristic algorithms have been employed to optimize FMD parameters (Chauhan et al., 2024; Yan, Yan, Xu, & Yuen, 2023). However, the computational cost for fitness function evaluations in FMD make it difficult for metaheuristic algorithms to balance efficiency and accuracy. Bayesian optimization is characterized by its high

efficiency, attributable to its systematic exploitation of existing information (Wang, Wang, & Peng, 2021). This study employs Bayesian optimization to adaptively learn the optimal parameter combination (\hat{K} , \hat{L}) for FMD. In this study, the Gaussian process (Singh, Nagar, Sharma, & Kotiyal, 2021) is utilized as the probabilistic surrogate model for Bayesian optimization:

$$\begin{cases} f(x) \sim GP[m(x), k(x, x')] \\ m(x) = E[f(x)] \\ k(x, x') = E\{[f(x) - m(x)][f(x') - m(x')]^T\} \end{cases} \quad (9)$$

where $f(x)$ is the objective function, $m(x)$ represents the mean value function, $k(x, x')$ represents covariance functions, which are used to characterize Gaussian processes. $k(x, x')$ is a kernel function. In this paper, the squared exponential kernel (Ruan, Wang, Yan, & Gühmann, 2023) is selected as the kernel function, which is constructed as:

$$k(x, x') = \exp\left[-\frac{1}{2\sigma_l^2}(x - x')^T(x - x')\right] \quad (10)$$

where σ_l is the length-scale parameter.

Assume there exists a prior distribution with a mean of 0:

$$p(\psi|X, \theta) = N(0, \Sigma) \quad (11)$$

where X represents the training set, ψ denotes the set of function values of f , Σ is the covariance matrix composed of $k(x, x')$, θ signifies the hyperparameters.

Assuming the presence of noise ϵ , which follows an independent and identically distributed Gaussian distribution $p(\epsilon) = N(0, \sigma^2)$. The likelihood function is obtained as:

$$p(Y|\psi) = N(\psi, \sigma^2 I) \quad (12)$$

where Y represents the observed outputs, and I is the identity matrix, indicating that the noise is independent across different observations. The joint distribution is constructed as:

$$\begin{bmatrix} Y \\ \psi_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma + \sigma^2 I & K_* \\ K_*^T & K_{**} \end{bmatrix}\right) \quad (13)$$

where ψ_* represents the predicted function values, K_* denotes the covariance from training set X to prediction point X_* . The covariance matrix is defined as:

$$\begin{aligned} K_*^T &= \{k(x_1, X_*), k(x_2, X_*), \dots, k(x_t, X_*)\} \\ K_{**} &= k(X_*, X_*) \end{aligned} \quad (14)$$

where X_* is the predictive input. The resulting predictive distribution is calculated by:

$$\begin{cases} p(\psi_* | X, Y, X_*) = N(\langle \psi_* \rangle, \text{cov}(\psi_*)) \\ \langle \psi_* \rangle = K_*^T [\Sigma + \sigma^2 I]^{-1} Y \\ \text{cov}(\psi_*) = K_{**} - K_*^T [\Sigma + \sigma^2 I]^{-1} K_* \end{cases} \quad (15)$$

where $\langle \psi_* \rangle$ represents the predicted mean, and $\text{cov}(\psi_*)$ denotes the predicted covariance.

Given the advantage of the parsimonious parameters of the Expected Improvement (EI) strategy, this study employs EI as the acquisition function in Bayesian optimization. By integrating the cumulative density function with the density function of the standard normal distribution, EI calculates the potential enhancement values (Lyu et al., 2017):

$$\eta = \Phi\left(\frac{\nu^* - \mu_t(x)}{\sigma_t(x)}\right) \quad (16)$$

$$\gamma = \phi\left(\frac{\nu^* - \mu_t(x)}{\sigma_t(x)}\right) \quad (17)$$

$$\alpha_t(x; D_{1:t}) = \begin{cases} [\nu^* - \mu_t(x)]\eta + \sigma_t(x)\gamma, & \sigma_t(x) > 0 \\ 0, & \sigma_t(x) = 0 \end{cases} \quad (18)$$

where ν^* represents the optimal value of the current objective function, $\mu_t(x)$ and $\sigma_t(x)$ denote the predicted mean and the predicted standard deviation at point x , respectively. $\Phi(\cdot)$ is the standard cumulative distribution function, and $\phi(\cdot)$ is the standard normal distribution density function.

2.3. Adaptive parameter tuning

This study applies Bayesian optimization to adaptively adjust FMD, ensuring optimal signal adaptation. During the optimization process, it is crucial to establish an evaluation function. Inspired by envelope entropy and envelope characteristic frequency ratio (Wu et al., 2024), a novel comprehensive health indicator termed multilayer pulse characteristic integration measure (MPCIM) is presented. MPCIM is leveraged as the fitness value guiding the parameter optimization of ATFMD, effectively capturing weak pulse characteristics and complex fault features. Additionally, recognizing that the instantaneous power of fault signals in rotating machinery often demonstrates periodic fluctuations, we propose the cyclic squared envelope entropy (CSEE) to mitigate this issue.

For a signal $x(t)$ with a period T , the autocorrelation function of the squared envelope signal is defined as:

$$\begin{cases} R_{SE(x)}(\tau) = \int SE(x(t))SE(x(t+\tau))dt \\ SE(x) = |x|^2 = |x + j \cdot \text{Hilbert}(x)|^2 \end{cases} \quad (19)$$

where $SE(x)$ is the squared envelope signal, $\text{Hilbert}(\cdot)$ denotes the Hilbert transform, and $R_{SE(x)}(\tau)$ describes the autocorrelation of $SE(x)$ at the time delay τ .

For each period iT , the ratio of the autocorrelation amplitude for that period to the total autocorrelation amplitude can be expressed as:

$$p_i(x, T) = \frac{R_{SE(x)}(iT)}{\sum_{i=1}^M R_{SE(x)}(iT)} \quad (20)$$

where M represents the number of sampling points within one period, and $i = 1, 2, \dots, M$.

The CSEE is defined as:

$$\begin{aligned} \text{CSEE}(x, T) &= -\sum_{i=1}^M (p_i(x, T) \log_2 p_i(x, T)) \\ &= -\sum_{i=1}^M \left(\frac{R_{SE(x)}(iT)}{\sum_{i=1}^M R_{SE(x)}(iT)} \log_2 \frac{R_{SE(x)}(iT)}{\sum_{i=1}^M R_{SE(x)}(iT)} \right) \end{aligned} \quad (21)$$

where $\sum_{i=1}^M p_i(x, T) = 1$.

CSEE inherits the advantages of envelope entropy in evaluating the subtle fault characteristics of signals. Additionally, it enhances the functionality of envelope entropy through cyclic analysis.

In addition, the envelope characteristic frequency ratio (ECFR) is introduced as a core metric for assessing the richness of fault information. ECFR focuses on quantifying the proportion of characteristic fault frequencies and their harmonics within the signal.

For the signal $x(t)$, the Hilbert Transform can be expressed as:

$$\tilde{X}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (22)$$

where τ is the integral variable.

The analytical calculation for the signal $x(t)$ can be expressed as:

$$Z(t) = x(t) + j\tilde{X}(t) = H(t)e^{j\omega(t)} \quad (23)$$

where $H(t)$ and $\omega(t)$ denote the amplitude and phase, respectively, can be calculated as:

$$H(t) = \sqrt{x^2(t) + \tilde{X}^2(t)} \quad (24)$$

$$\omega(t) = \arctan \frac{\tilde{X}(t)}{x(t)} \quad (25)$$

The ECFR represents the proportion of fault-related information within the overall envelope signal and is defined as:

$$ECFR = \frac{\sum_{i=1}^M \sum_{j=1}^N E_n(f_{ij})}{E_n(A)} \quad (26)$$

where $E_n(f_{ij})$ denotes the energy at frequency f_{ij} , $E_n(A)$ indicates the total energy of the entire envelope signal, $\sum_{i=1}^M \sum_{j=1}^N E_n(f_{ij})$ represents the energy of fault-related characteristic information. M and N represent the number of fundamental frequency harmonics and the number of fault categories, respectively. ECFR provides a quantified perspective for assessing the total energy $E_n(A)$ of envelope signals.

Considering the fluctuations caused by minor changes in the ECFR, this study introduces an exponential function adjustment mechanism (He, Li, Li, Dong, & Zhu, 2022). Consequently, to ensure the processed signal retains optimal fault characteristics, the MPCIM is defined as:

$$MPCIM = \frac{CSEE}{e^{ECFR}} \quad (27)$$

The lower MPCIM value indicates that the ATFMD exhibits improved perceptibility towards fault features, coupled with a more effective filtration of noise and non-critical information. Consequently, the fitness function during the adaptive tuning process of ATFMD is defined as:

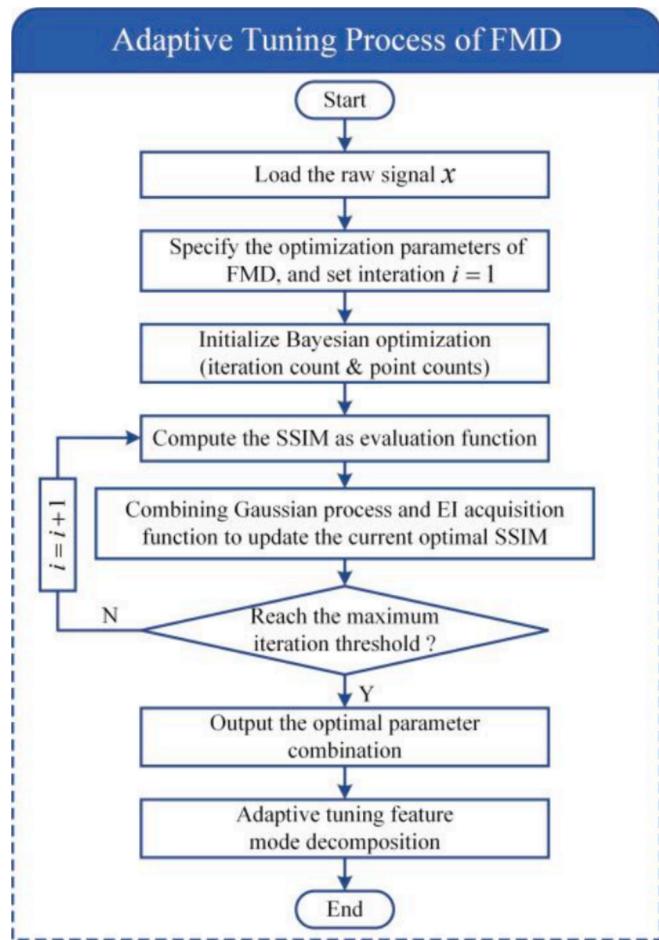


Fig. 2. Adaptive tuning process of FMD.

$$\langle \hat{K}, \hat{L} \rangle = \operatorname{argmin}_{\Psi_i=(K,L)} \{MPCIM_{\Psi_i}\}, \text{subject to } K \in [2, 8] \text{ and } L \in [10, 100] \quad (28)$$

where $\langle \hat{K}, \hat{L} \rangle$ is the optimal parameter combination, and $MPCIM_{\Psi_i}$ denotes the MPCIM of the mode components, evaluated across various sets of combination parameters $\Psi_i = (K, L)$. K and L denote the total filter count and filter length, respectively. Based on empirical evidence (Yan, Yan, Xu, & Yuen, 2023), the search range for K is set from 2 to 8, while the range for L is established between 10 and 100. The adaptive tuning process of FMD is illustrated in Fig. 2. As shown, after defining the FMD parameters, iteration counts and point numbers are identified as core elements of the optimization cycle driven by Bayesian optimization. The MPCIM serves as the evaluative function, propelling the iterative loop through its computations. The data generated from these computations are utilized by the Gaussian process to predict and map the parameter performance. At each iteration, the next set of parameters to be evaluated is selected based on the expected improvement (EI) strategy, aiming to minimize MPCIM. The iterative process continues until the maximum iteration threshold is reached, culminating in the determination of the optimal parameter combination.

3. Preprocessing case study

In this study, a case study process is designed based on the autonomously experimental bench for rotating machinery to verify the performance of the proposed preprocessing method. The flowchart of the preprocessing case study is presented in Fig. 3. The comprehensive assessment of the ATFMD is conducted in this study using multidimensional signal processing techniques, including frequency band distribution analysis, power spectral computation, envelope spectrum analysis, and continuous wavelet transform. Through comparative analysis with various signal decomposition methods, the efficacy of ATFMD in analyzing weak fault features and compound fault characteristics has been validated. Furthermore, a baseline model based on deep learning was applied to quantitatively evaluate the contribution of ATFMD to the diagnostic model.

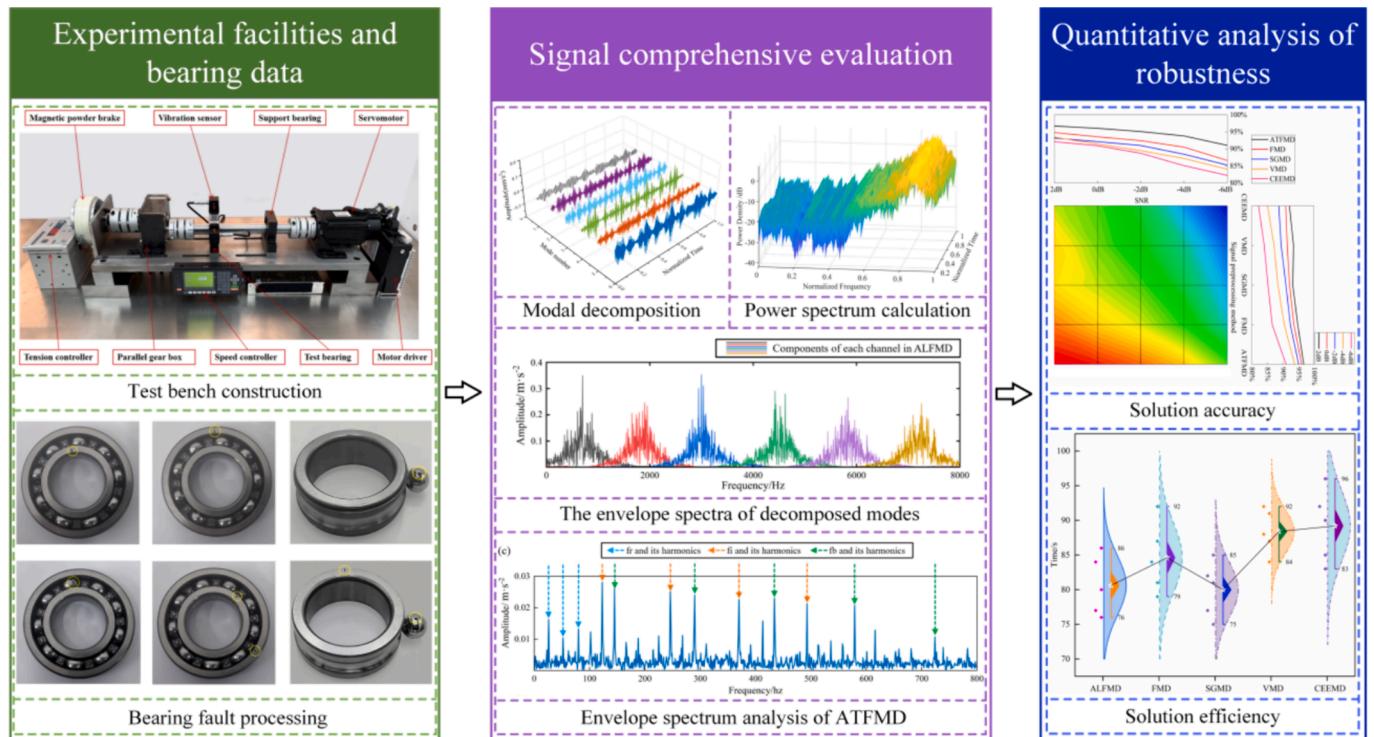


Fig. 3. Flowchart of the preprocessing case study.

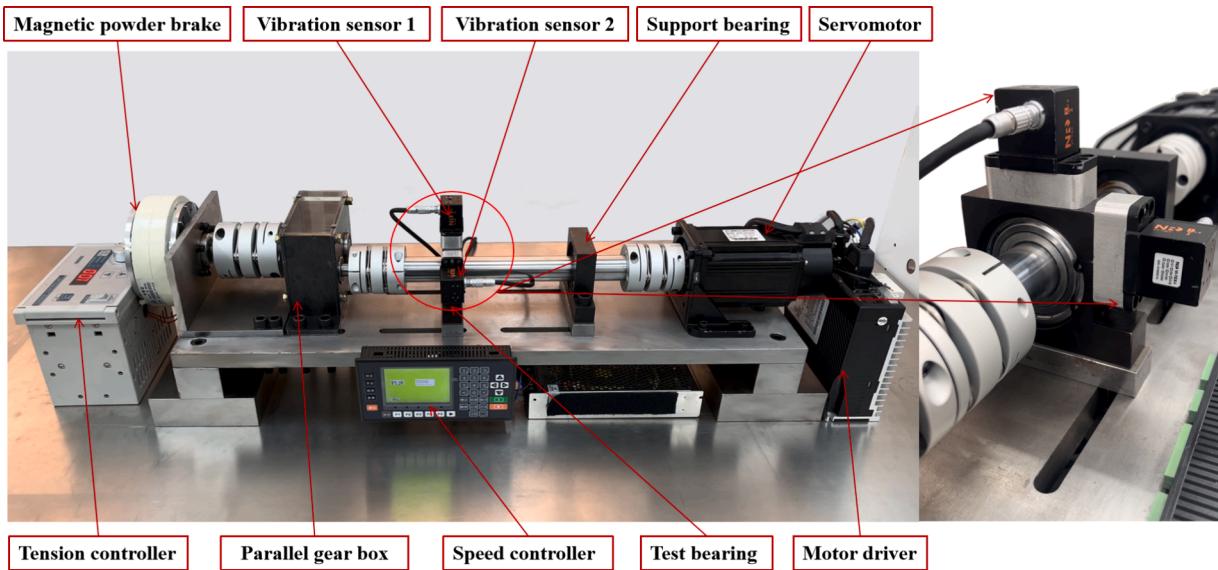


Fig. 4. Rotating machinery test bench.

3.1. Experimental facilities and bearing data

The bearing fault experiment is carried out on the autonomously rotating machinery test bench. Fig. 4 shows a schematic diagram of the experimental facility and its corresponding structure. The experimental platform is mainly composed of a drive motor, a bearing test module, a reduction gearbox, a magnetic particle brake, an electrical control system, and a data acquisition system. The electrical control system regulates the motor speed, while the magnetic particle brake provides the load. The data acquisition system consists of an NI-9234 acquisition card, AP-TEX piezoelectric acceleration sensors, and a computer equipped with AP-TEX assistant software. Two AP-TEX piezoelectric acceleration sensors are installed at the 9 and 12 o'clock positions on the bearing test module, with the sampling frequency of 16 kHz.

Electrical discharge machining technology was employed to introduce various defects into the test bearing (6206-ZNR SKF). These defects included an inner ring fault (IRF), outer ring fault (ORF), ball fault (BF),

Table 1
Size of the testing bearing.

Bearing type	Roller diameter	Pitch diameter	Number of balls	Contact angle
6206-ZNR SKF	7.75 mm	39 mm	9	0°

cage break fault (CF), inner-outer ring combined fault (IORF), and combined inner ring-ball combined fault (IRBF). The failure mode of different damaged bearings is shown in Fig. 5.

The size of the testing bearing is illustrated in Table 1. During the experiments, the motor speed is set at 1600 rpm ($f_r = 26.67\text{Hz}$, $1600/60 = 26.67$), and the sampling frequency is established to 16 kHz. Based on the bearing characteristic frequency formula (Ni, Ji, Feng, & Halkon, 2022), the fault characteristic frequencies for the outer race, inner race,

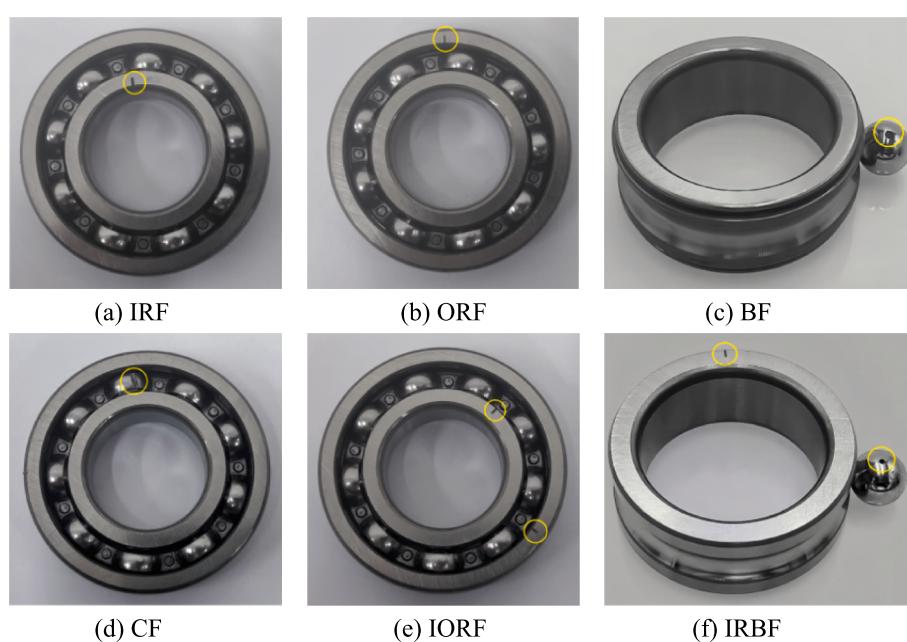


Fig. 5. Failure mode of different damaged bearings.

and balls are $f_o = 95.2$ Hz, $f_i = 144.8$ Hz and $f_b = 123.2$ Hz, respectively. Among the fault types described, the signal from the rolling element fault (BF) is the weakest. The accurate extraction of features from rolling element and inner ring faults under combined fault conditions is considered more challenging. In this section, the inner ring-ball combined fault (IRBF) under combined fault conditions is utilized to validate the capability of the ATFMD technique in extracting features from weak faults.

3.2. Signal comprehensive evaluation

The original signal from the combined inner ring-ball fault (IRBF) was analyzed using the ATFMD technique. During this process, Bayesian optimization was employed to adaptively tune the key combinatorial parameters $\langle \hat{K}, \hat{L} \rangle$ of the finite impulse response (FIR) filter bank, specifically the number of modes K and the filter length L . The results of the Bayesian optimization were evaluated using the multilayer pulse characteristic integration measure (MPCIM) as a comprehensive evaluation metric. For the IRBF signal in this experiment, the optimal parameter combination for the FIR filter bank in the ATFMD technique was determined to be $\langle 6, 40 \rangle$. Corresponding to the original IRBF signal processed using ATFMD, six channel signals are obtained, as shown in Fig. 6. It can be seen that the amplitude variations across different modalities demonstrate nonlinear behavior over time, indicating that the signal exhibits complex time-varying characteristics.

The original signal is composed of multiple frequency components, influenced by the interaction of composite faults and nonlinear effects. Spectral analysis transforms the signal into the frequency domain, revealing the frequency components of each channel signal. The spectral distribution of each channel signal is shown in Fig. 7. It can be observed that the frequency spectrum is well-distributed across the channels with clear delineation between them, indicating an effective decomposition with no evident mode aliasing among adjacent frequency bands.

Three-dimensional power spectral density (Yi et al., 2022) plots are commonly used to represent the dynamic variations of signals across three dimensions: time, frequency, and energy. The three-dimensional power spectral density plot of decomposed modes is shown in Fig. 8. From the perspective of modal broadening, it is evident that the power spectra of the decomposed modes are broadly distributed across a wide frequency range, rather than being concentrated at a single frequency. This phenomenon demonstrates the effectiveness of ATFMD in identifying multiple frequency components within the signal, showcasing its broad frequency recognition capability. Further quantitative analysis reveals significant differences in the energy distribution across the frequency domain of each mode. ATFMD provides essential quantitative

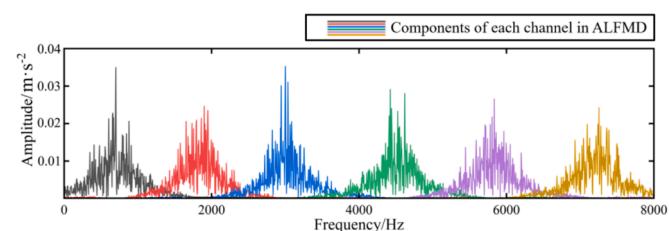


Fig. 7. The envelope spectrum of decomposed modes.

data for fault identification and severity assessment, derived from peak power density and bandwidth measurements in specific frequency bands.

In this study, the original signals and the signals processed through VMD and ATFMD are compared through envelope spectrum analysis, as shown in Fig. 9. Fig. 9(a) shows that in original signals, only fundamental rotational frequencies f_r and fault frequencies f_i and f_b , including their first harmonics, are detectable. The amplitudes of these frequencies are severely constrained by background noise and composite faults. Fig. 9(b) shows that VMD identifies the f_r , as well as f_i and f_b , along with their first and second harmonics. However, the higher-order harmonics of f_i and f_b are less prominent. This suggests that while VMD effectively identifies primary fault characteristics, it struggles with higher-order harmonics. Fig. 9(c) shows that for signals processed through ATFMD, the delineation of frequency components achieves outstanding clarity. The envelope spectrum demonstrates that the f_r and its corresponding harmonics emerge distinctly up to the fifth order. Similarly, f_i and f_b are sharply defined, with their harmonics clearly distinguishable to the fourth order. ATFMD distinguishes harmonics from the noise floor, illustrating enhanced filtering capabilities. The integration of MPCIM within the ATFMD framework clearly enhances efficacy, particularly evident as the ECFR identifies characteristics across various fault modes.

3.3. Quantitative analysis of robustness

Compared to laboratory data, real data collected in the industrial environment will contain heavy noise and more interference components. In order to assess the performance of ATFMD in noise suppression, Gaussian noise was added to the YZU dataset from the autonomously rotating machinery test bench illustrated in Section 3.1. The evaluation was performed by comparing ATFMD with existing signal decomposition methods, including CEEMD (Chen, Lin, Yao, Yang, & Ge, 2023), VMD, SGMD, and FMD. The dataset processed by these various different signal decomposition methods were respectively input into the benchmark diagnostic model, ResNet (Zhang, Tang, Deng, & Liu, 2021), for both training and testing. The proportion of training set and test set is 7:3. This process involved adding Gaussian noise with the signal-to-noise ratio(SNR) from -6 dB to 2 dB to the raw signals and quantifying the degree of interference:

$$\text{SNR}_{\text{dB}} = 10 \lg(P_{\text{signal}}/P_{\text{noise}}) \quad (29)$$

where P_{signal} and P_{noise} denote the energy of signal and noise.

Five independent experiments were conducted under each noise condition to ensure the reliability of the data. The corresponding average diagnostic accuracy results are displayed in Table 2.

The influence of each preprocessing method on the average diagnostic accuracy and efficiency of the ResNet is presented in Fig. 10. Fig. 10(a) shows that with the increase in noise, the average diagnostic accuracy of each method correspondingly decreased. Under the SNR of 2 dB, the average accuracy of ATFMD-ResNet was recorded at 96.75% , slightly exceeding the 94.52% achieved by FMD-ResNet. With increasing noise levels, the performance differences between the methods were observed to become more distinct. When the SNR was reduced to -6 dB, the average accuracy of ATFMD-ResNet was found to

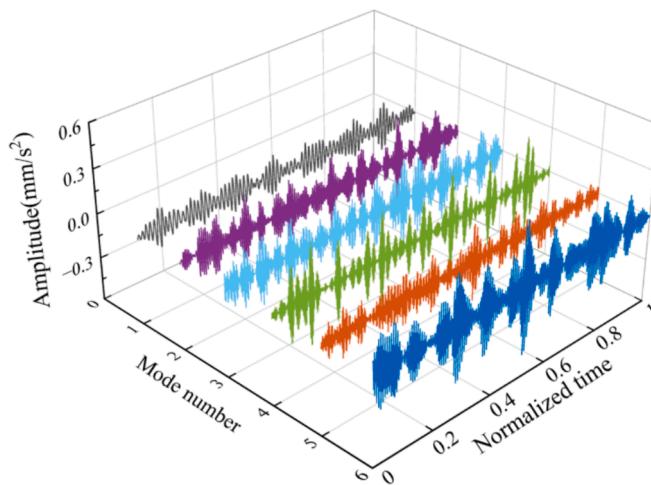


Fig. 6. Decomposed modes of ATFMD.

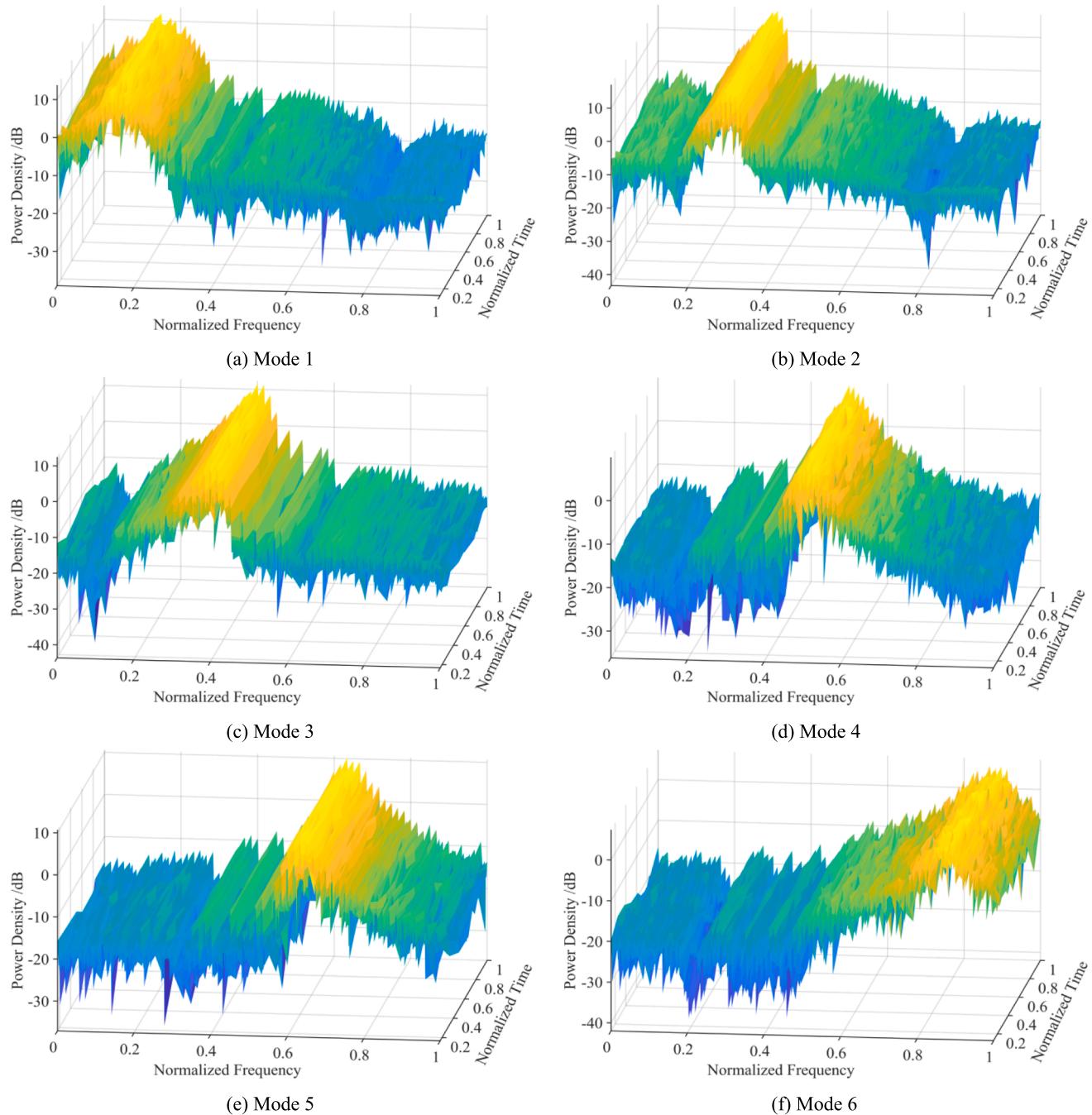


Fig. 8. Three-dimensional spectral density plot of decomposed modes.

decline to 87.36 %. However, the average precision for all comparative methods was noted to significantly deteriorate, dropping below 85 %. In the presence of -6 dB noise interference, ATFMD demonstrated a significant superiority in enhancing the performance of the ResNet model when compared with FMD, SGMD, VMD, and CEEMD methods. Specifically, the average accuracy of ATFMD-ResNet exceeded that of FMD-ResNet by 4.72 %, SGMD-ResNet by 6.09 %, VMD-ResNet by 8.30 %, and CEEMD-ResNet by 11.31 %. Fig. 10(b) demonstrates that ATFMD significantly enhances the computational efficiency of the diagnostic model. Specifically, it outperforms FMD, VMD, and CEEMD in terms of average computation time, while maintaining a performance level comparable to SGMD.

This improvement is attributed to the Bayesian optimization within ATFMD, which determines the optimal parameter combination $\langle \hat{K}, \hat{L} \rangle$. In

summary, ATFMD demonstrates excellent robustness, facilitating the extraction of critical fault information from significant channels in the signal.

4. Construction of feature topological graph

In scenarios with complex mechanical structures and frequent operational interferences, fault features from singular dimension fail to fully reflect the health status of the equipment. Traditional methods are confined to characterizing features within a single metric space, which fails to adequately reveal the underlying synergistic interactions and topological structures among features. This shortcoming limits both the precision and completeness of fault diagnosis. Consequently, this research enhances the feature array for fault diagnosis by leveraging

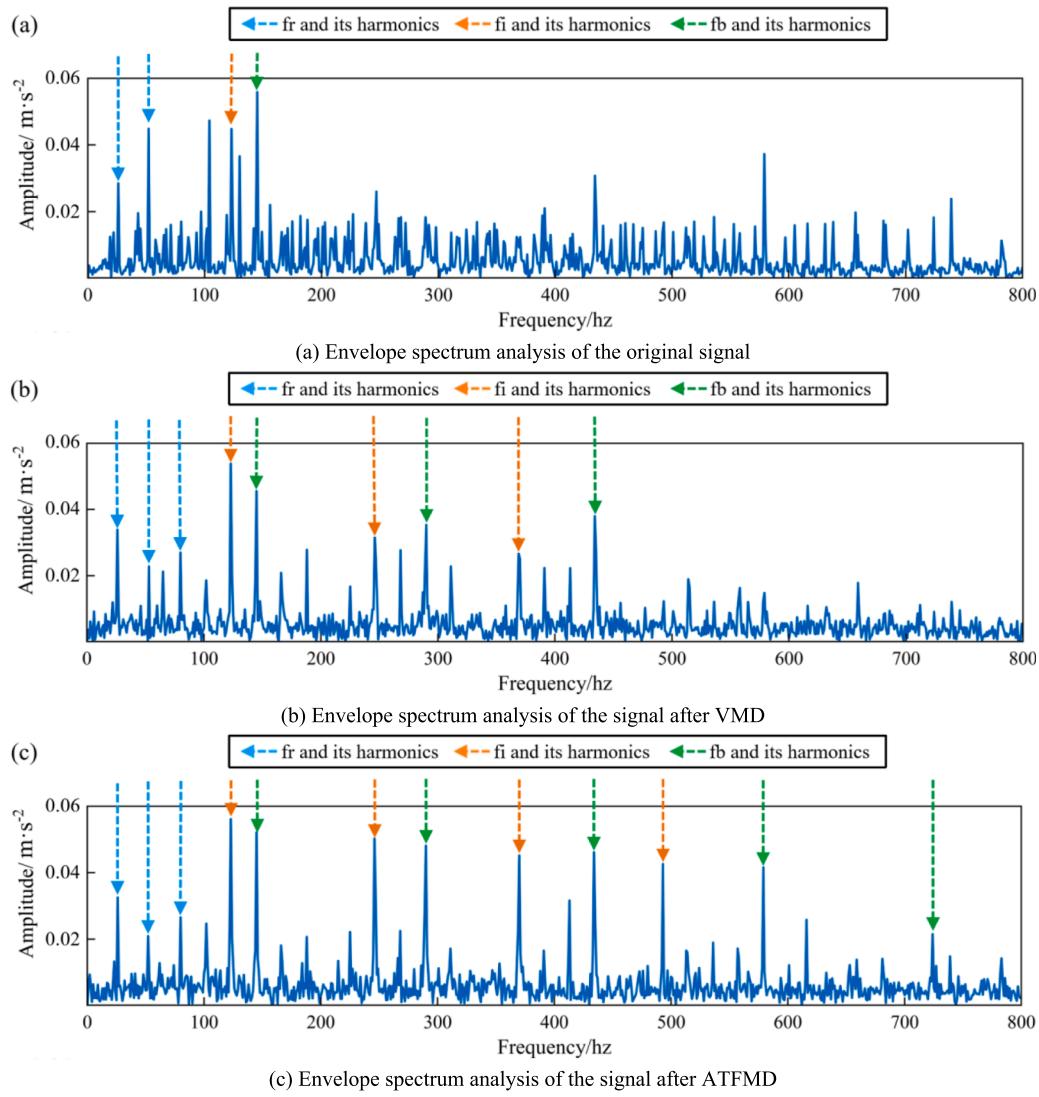


Fig. 9. Envelope spectrum analysis of different methods.

Table 2
Performance of different preprocessing methods under noisy conditions (%).

Method	SNR				
	-6dB	-4dB	-2dB	0 dB	2 dB
ATFMD-ResNet	87.36	92.83	95.12	96.53	96.75
FMD-ResNet	82.64	88.53	92.24	93.89	94.52
SGMD-ResNet	81.27	86.45	91.02	93.17	93.21
VMD-ResNet	79.06	84.22	88.49	91.45	93.47
CEEMD-ResNet	76.25	83.12	88.73	91.96	92.84

images as dense data carriers. The vibration signals preprocessed by ATFMD are encoded into gramian angular field matrices (Li et al., 2024) and analyzed using continuous wavelet transform. The signals processed by ATFMD effectively avoids errors and redundant edges that may arise from noise when directly constructing feature topologies from raw samples. During the construction of feature topological graphs, the primary objective is to fully represent the multidimensional characteristics inherent in the fault signals.

4.1. Gramian angular field

The GAF method constructs matrices by computing the cosine and

sine values across time points to identify the periodicity and structural features of the signal. In this study, it is applied to transform phase information into a visual matrix representation. For the signal x_t , it is scaled to be between [-1,1] according to Eq. (30) to eliminate dimensional inconsistencies. The normalized result is defined as follows:

$$\tilde{x}_i = \frac{(x_i - \max\{x_t\}) + (x_i - \min\{x_t\})}{\max\{x_t\} - \min\{x_t\}}, \quad i \in t \quad (30)$$

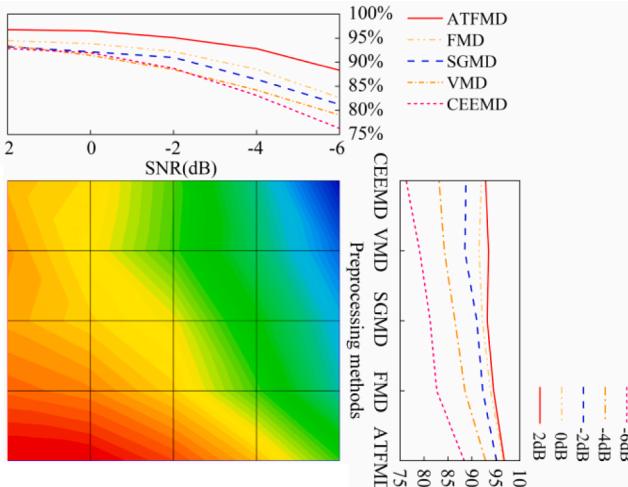
where x_t represents the vibration signal data presented as a time series, x_i is the sample value in x_t , and \tilde{x}_i represents the normalized result of x_i after preprocessing.

The scaled time series data are converted from a Cartesian to a polar coordinate system, defined as:

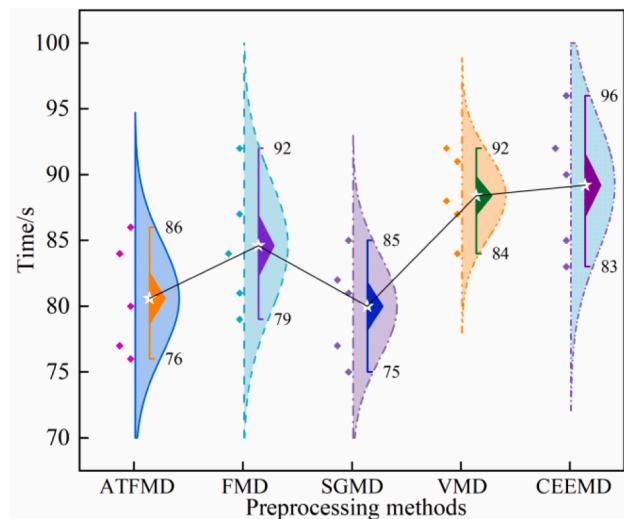
$$\begin{cases} \phi = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X}_t \\ r_i = \frac{t_i}{N}, & t_i \in N, \end{cases} \quad (31)$$

where ϕ is the polar angle, \tilde{X}_t represents the entire set of normalized time series, and \tilde{x}_i represents the i^{th} value of \tilde{X}_t . r_i is the polar axis, t_i is the time step, and N is the constant factor for the regularized polar coordinate system.

GASF is a cosine-based expression of GAF, which can more clearly reveal the phase structure and patterns within the signal. The GASF



(a) Solution accuracy



(b) Solution efficiency

Fig. 10. Comparison of signal preprocessing methods.

matrix is defined as:

$$GASF = \begin{bmatrix} \cos(\theta_1 + \theta_1) & \cos(\theta_1 + \theta_2) & \dots & \cos(\theta_1 + \theta_n) \\ \cos(\theta_2 + \theta_1) & \cos(\theta_2 + \theta_2) & \dots & \cos(\theta_2 + \theta_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\theta_n + \theta_1) & \cos(\theta_n + \theta_2) & \dots & \cos(\theta_n + \theta_n) \end{bmatrix} \quad (32)$$

$$= (\tilde{X}_t)^T \tilde{X}_t - \sqrt{E - (\tilde{X}_t)^2} \sqrt{E - (\tilde{X}_t)^2}$$

where E is the unit row vector.

The matrix images resulting from the GAF encoding of the sample signals are shown in Fig. 11. It is observed that the GAF clearly manifests the temporal variations of vibration signals and load speed within the phase space.

4.2. Continuous wavelet transform

The CWT is an analytical tool employed for signal analysis, which decomposes signals across multiple time scales to reveal local features within the time-frequency domain (Nashed, Renno, Mohamed, & Reuben, 2023). In this study, CWT is utilized to identify transient characteristics and frequency shifts in fault signals. For the signal $x(t)$, defined as $x(t) \in L^2(R)$, the CWT is expressed as:

$$CWT_f(s, \tau) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t) \psi_{s,\tau}^*(\frac{t-\tau}{s}) dt \quad (33)$$

where $\psi^*(\bullet)$ denotes the complex conjugate of the mother wavelet function with scale parameter s and translation parameter τ .

The Morlet wavelet function (Shao, Xia, Wan, & de Silva, 2021) exhibits an exponentially decaying oscillatory form, which is similar to the transient shock attenuation component generated by bearing faults.

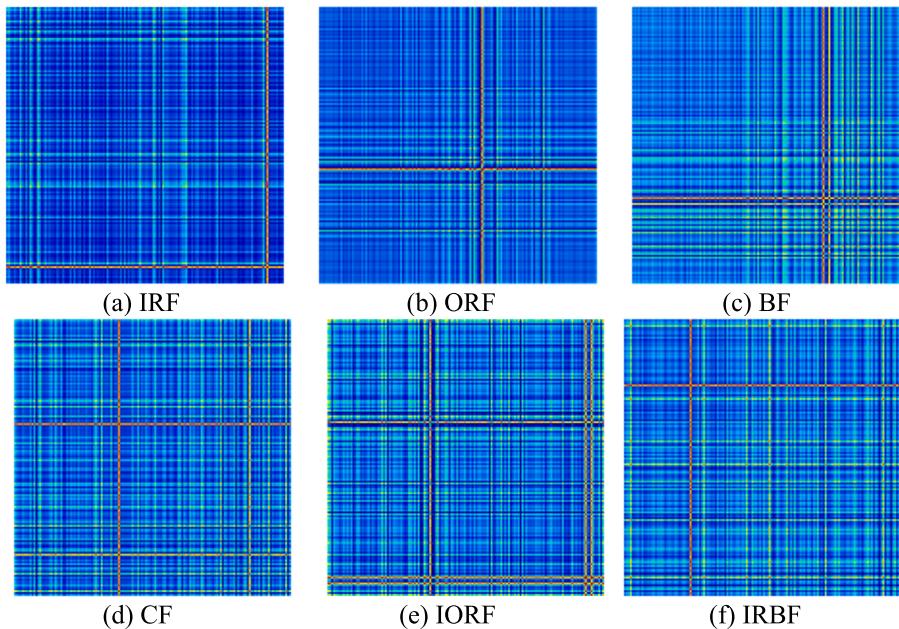


Fig. 11. Signals encoded by GAF.

The corresponding time-domain expression is defined as:

$$\psi^*(t) = \pi^{-\frac{1}{4}} e^{-\frac{t^2}{2}} \cos(5t) \quad (34)$$

The images obtained through the CWT of the sample signals are shown in Fig. 12. It is observed that the CWT distinctly captures the evolution of the signal characteristics, with clear representation of both transient and persistent features in the time-frequency domain.

4.3. Definition of topological graph node

The construction of the feature topological graph is accomplished through the integration of GAF and CWT, which reveals the interaction features inherent in fault signals across the time–frequency spectrum and spatial phase domain.

In the framework of GAF, each node is represented as a phase space point derived from the time series, capturing the consistency of phase information. The distance between nodes (Tang et al., 2021) in terms of spatial features is defined as:

$$D_{spatial(i,j)} = \sqrt{(\phi_i - \phi_j)^T S_\phi^{-1} (\phi_i - \phi_j) + (\varpi_i - \varpi_j)^T S_\varpi^{-1} (\varpi_i - \varpi_j) + (\Delta M_{ij})^T S_M^{-1} (\Delta M_{ij})} \quad (35)$$

where ϕ_i and ϕ_j represent the angular phase vectors of nodes i and j , respectively. ϖ_i and ϖ_j denote the phase change rate vectors of nodes i and j . ΔM_{ij} is the spatial scale vector between nodes i and j , indicating the disparity in embedding dimensions. S_ϕ^{-1} , S_ϖ^{-1} , and S_M^{-1} correspond to the inverses of the covariance matrices for the phase values, phase change rates, and spatial scales, respectively.

In the framework of CWT, each node corresponds to a signal segment characterized by specific time–frequency attributes, with its magnitude representing the energy distribution at a particular time and frequency. The distance between nodes in terms of time–frequency features is defined as follows:

$$D_{temporal(i,j)} = \sqrt{(f_i - f_j)^T S_f^{-1} (f_i - f_j) + (t_i - t_j)^T S_t^{-1} (t_i - t_j) + (\Delta \alpha_{ij})^T S_a^{-1} (\Delta \alpha_{ij})} \quad (36)$$

where f_i and f_j denote the frequency coordinate vectors of nodes i and j , respectively. t_i and t_j represent the temporal coordinate vectors of nodes i and j , respectively. $\Delta \alpha_{ij}$ is the amplitude difference vector between nodes i and j , indicating the difference in energy intensity. S_f^{-1} , S_t^{-1} , and S_a^{-1} are the inverses of the covariance matrices for the frequency, time, and amplitude dimensions, respectively.

For the construction of feature graph, the comprehensive distance between nodes is formulated as:

$$\bar{D}_{ij} = \sqrt{(D_{temporal(i,j)})^2 + (D_{spatial(i,j)})^2} \quad (37)$$

where $D_{temporal(i,j)}$ and $D_{spatial(i,j)}$ represent the distances between nodes in time–frequency features and spatial features, respectively.

5. Synergy graph enhanced transformer

This study introduces a synergy graph enhanced transformer (SGET), designed to address the limitations of traditional Transformer and CNN models in capturing multidimensional features and modeling complex dependency structures. The SGRM within the SGET iteratively updates the adjacency matrix and employs convolutional fusion to adaptively capture dependencies between feature domains. By integrating time–frequency and spatial feature graphs, it enables multi-dimensional joint feature representation. The hierarchical cross-attention mechanism enhances feature fusion by facilitating interactions across time–frequency and spatial domains. Initially, time–frequency features capture local correlations via attention mechanisms, and these are then fused with spatial features to extract cross-domain dependencies, improving the comprehensiveness of information fusion. The overall SGET model

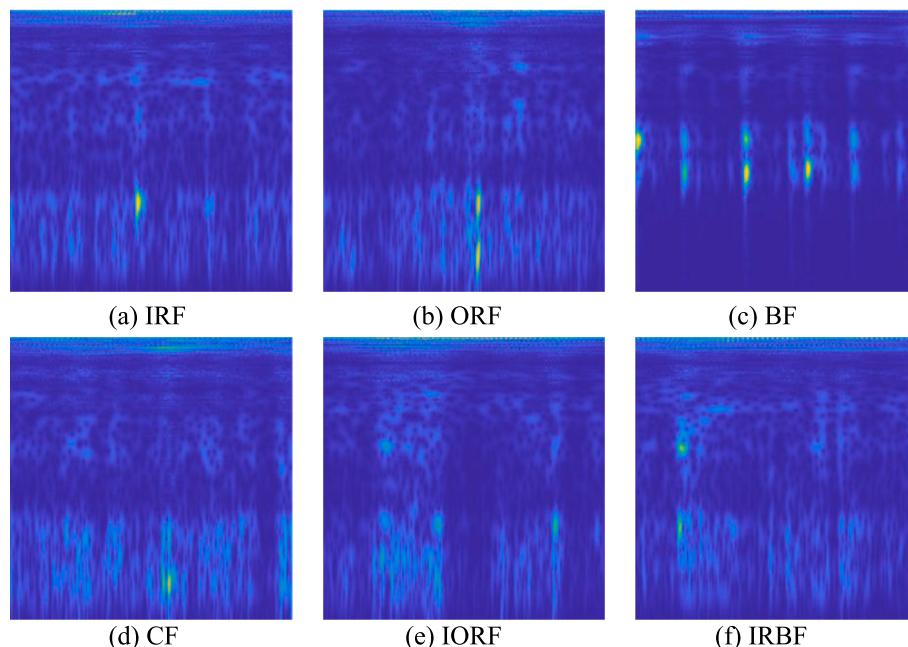


Fig. 12. CWT time-frequency distribution.

follows an encode-process-decode architecture, leveraging feature distillation and residual connections to ensure stable extraction of critical features and efficient training, ultimately classifying faults through a softmax layer.

5.1. Synergy graph representation module

The predefined adjacency matrix is constructed based on fault information, and the original adjacency matrix (Kong et al., 2024) is defined as follows:

$$G_0 = \begin{cases} \exp\left(-\frac{(\bar{D}_{ij})^2}{\delta^2}\right), & \text{if } \bar{D}_{ij} \geq \Theta \\ 0, & \text{if } \bar{D}_{ij} < \Theta \end{cases} \quad (38)$$

where \bar{D}_{ij} represents the comprehensive distance between nodes, and Θ defines the node sparsity threshold. δ is the standard deviation of the distance.

Given the multiscale nature of fault features, the adaptive mechanism of the adjacency matrix is updated iteratively as:

$$G_{1n} = \text{Softmax}(\text{ReLU}(ST^T + \text{Diag}(E))) \quad (39)$$

where n denotes the iteration count, S and T are the source and target node embedding matrices, respectively. $\text{Diag}(E)$ represents the learned adjacency matrix describing the graph structure.

Convolutional fusion is employed to integrate the initial and updated matrices. The integration process is expressed as:

$$S = \text{Sigmoid}(\text{Conv}([G_0, G_{1n}])) \quad (40)$$

$$G_{2n} = \lambda \otimes G_0 + (1 - \lambda) \otimes G_{1n} \quad (41)$$

$$G = \sum_{i=1}^N G_{2n}(i,:) \quad (42)$$

where λ is the selection coefficient, \otimes denotes element-wise multiplication. G_{2n} denotes the adjacency matrix resulting from the integrated updates, and G represents the final aggregated adjacency matrix.

The details of the synergy graph representation module (SGRM) is illustrated in Fig. 13. As shown, the module consists of two key components: the time-frequency feature representation (TFR) and the spatial feature representation (SFR).

The time-frequency adjacency matrix G_t , constructed from Eqs. (38) – (42), undergoes both a linear and a non-linear transformation to refine the feature representation. The weighted time-frequency feature graph G_{tem} and the resulting similarity graph G_T are expressed as:

$$G_T = W_{tb} \circ \tanh(W_{ta} \cdot G_t + b_{ta}) + b_{tb} \quad (43)$$

where W_{ta} and W_{tb} represents the respective weight matrices, b_{ta} and b_{tb} are the bias vectors.

The time-frequency similarity graph merges with the batch data matrix to form the final time-frequency adjacency matrix:

$$G_{TR} = G_{BD} \cdot G_T \quad (44)$$

where G_{BD} represents the batch data matrix, incorporating information from multiple time-frequency sequences and provides a broader data perspective. The SFR section is similar to the TFR in both structure and approach, and is used to process data related to spatial dimensions.

5.2. Hierarchical cross-attention

Multi-head attention captures feature information from different subspaces in parallel, enabling comprehensive analysis of multidimensional features. To enhance feature fusion, we propose hierarchical cross-attention (HCAtt), which extends multi-head attention by capturing interactions across time-frequency and spatial domains. This allows for a more fine-grained and comprehensive feature fusion.

The formula for multi-head attention (Pan, Chen, Ye, & Li, 2022) is expressed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (45)$$

where $Q \in \mathbb{R}^{n \times d_k}$, $K \in \mathbb{R}^{n \times d_k}$, $V \in \mathbb{R}^{m \times d_k}$, Q , K and V are abbreviations for the query, key and value matrices, respectively. d_k is the dimension of the key vectors.

The processed time-frequency features from the attention layer are defined as:

$$\bar{R}^t = \text{LayerNorm}(R^t + \text{Att}^t(Q_t, K_t, V_t)) \quad (46)$$

where R^t is the input time-frequency feature, and Q_t , K_t and V_t are the query, key, and value vectors, respectively. Att^t represents the time-frequency attention layer calculation.

These time-frequency features are further refined using a feed-forward network and layer normalization, producing the final time-frequency feature representation:

$$\hat{R}^t = \text{LayerNorm}(\bar{R}^t + \text{FNN}(\bar{R}^t)) \quad (47)$$

where FNN stands for feed-forward network.

The time-frequency feature \hat{R}^t is used as the query vector for the spatial attention layer, interacting with the spatial features. The initial representation of spatial features is defined as:

$$\bar{R}^s = \text{Att}^s(\hat{R}^t, K_s, V_s) \quad (48)$$

where K_s and V_s are the key and value vectors of the spatial features,

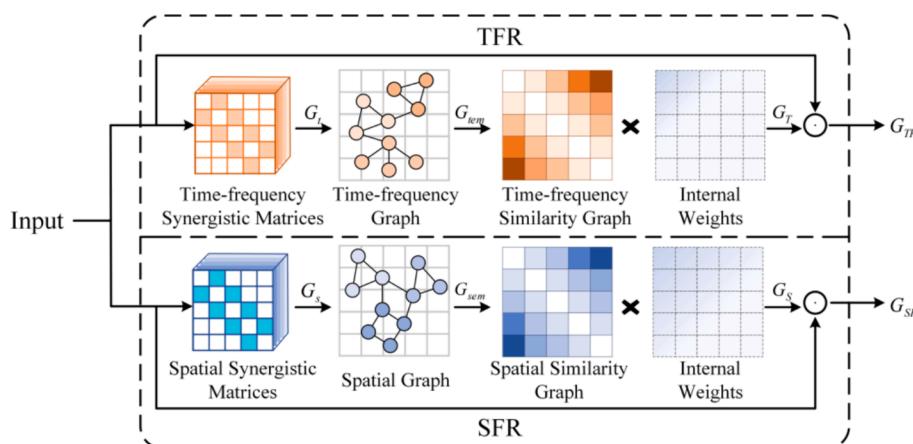


Fig. 13. Design of the synergy graph representation module.

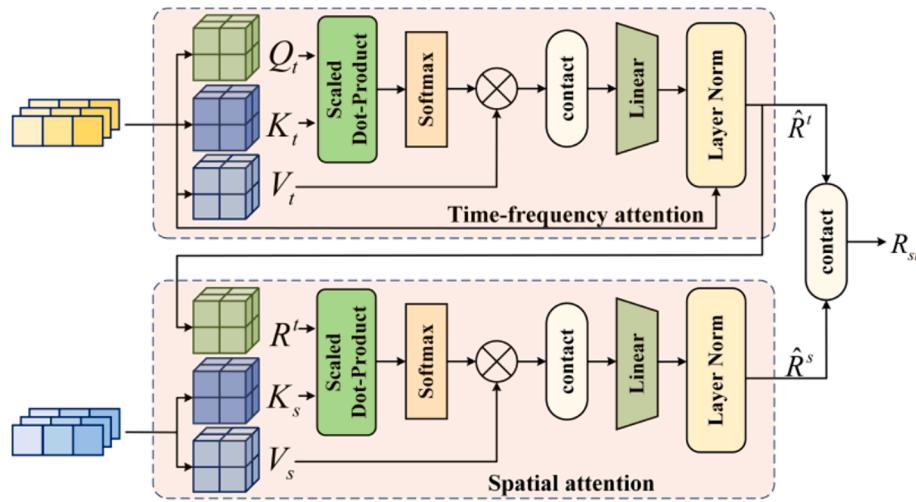


Fig. 14. Design of the hierarchical cross-attention.

respectively, and Att^s represents the spatial attention layer calculation.

The enhanced spatial features are defined as follows:

$$\tilde{R}^s = \text{LayerNorm}(\hat{R}^t + \bar{R}^s) \quad (49)$$

The feed-forward network and layer normalization further refine these features, culminating in the final spatial features:

$$\hat{R}^s = \text{LayerNorm}(\tilde{R}^s + \text{FNN}(\tilde{R}^s)) \quad (50)$$

The fault features obtained from hierarchical cross-attention are combined, and the interaction features are defined as:

$$R_{ts} = \text{Concat}(\hat{R}^t, \hat{R}^s) \quad (51)$$

As illustrated in Fig. 14, the hierarchical cross-attention (HCAtt) module uniquely facilitates the interaction between time–frequency and spatial attention, resulting in more comprehensive feature representations.

5.3. SGET architecture

The SGET is integrated within an encode-process-decode scheme, as shown in Fig. 15. The encoder processes time–frequency and spatial features through the HCAtt module, transforming them into encoded feature vectors. Through the hierarchical cross-attention and feature distillation, the encoder is able to extract critical features. The decoder interacts with the encoder's output and additional information to extract hidden states.

The input to the encoder is sourced from the output of the SGRM, which is defined as follows:

$$F_{en}^l = \text{Encoder}(F_{en}^{l-1}) \quad (52)$$

where F_{en}^l denotes the output of the l^{th} encoder layer.

The distillation mechanism (Guan, Qiao, Zhai, & Wang, 2022) is employed between the encoder layers to highlight the most performance-critical features. This process is defined as:

$$F_{en}^l = \text{Pooling}(\text{ReLU}(C_1d(HCAtt(F_{en}^{l-1})))) \quad (53)$$

where $HCAtt$ represents hierarchical cross-attention, and C_1d represents the convolutional layer for feature extraction. Pooling reduces information redundancy by performing multi-layer feature distillation.

Functional process of the decoder is defined as:

$$F_{de}^l = \text{Decoder}(\hat{R}^s, F_{en}^l) \quad (54)$$

where \hat{R}^s denotes the spatial feature representation and F_{en}^l denotes

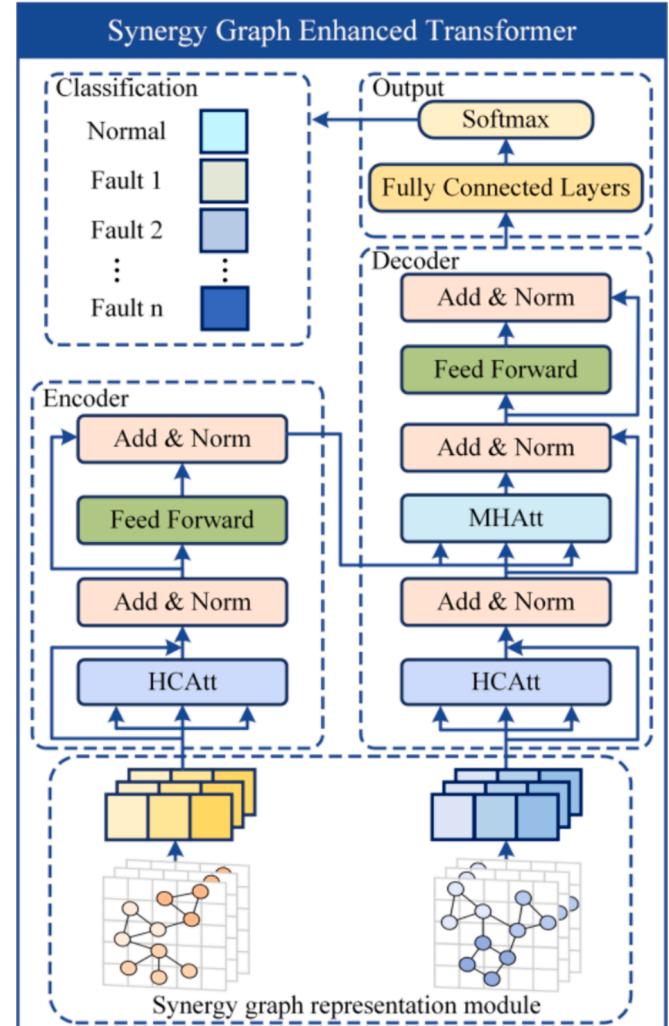


Fig. 15. The overall architecture of SGET.

the output of the l^{th} encoder layer.

Each module employs residual connections and layer normalization to improve training stability and efficiency. The processing workflow of decoder is defined as:

$$\begin{aligned} F_{de}^l &= HCAAtt(R_{ts}, R_{ts}, R_{ts})^l = MHAtt(\tilde{R}^s, F_{en}^l, F_{en}^l)\tilde{F}_{de}^l \\ \bar{F}_{de}^l &= \text{LayerNorm}(\tilde{R}^s + \tilde{F}_{de}^l)\tilde{F}_{de}^l = \text{LayerNorm}(\tilde{F}_{de}^l + \text{FNN}(\tilde{F}_{de}^l)) \end{aligned} \quad (55)$$

where R_{ts} stands for multidimensional synergy feature representation, and $MHAtt$ denotes multi-head attention. \bar{F}_{de}^l indicates the features extracted during the encoding phase, \tilde{F}_{de}^l represents the features combined through layer normalization, and \tilde{F}_{de}^l is the final decoder output.

The final output features of decoder are processed through the fully connected layer, expressed as:

$$F_{fc} = \text{ReLU}(W_{fc} \cdot \tilde{F}_{de}^l + b_{fc}) \quad (56)$$

where F_{fc} is the final output of the decoder, and W_{fc} and b_{fc} represent the weights and biases, respectively.

The output of the fully connected layer is converted into classification probabilities by the softmax layer:

$$Y_{out} = \text{Softmax}(W_{cls} \cdot F_{fc} + b_{cls}) \quad (57)$$

where W_{cls} and b_{cls} are the weights and biases of the softmax layer, respectively.

Cross-entropy loss function (Wang, Zeng, Zhang, Barros, & Miao, 2022) is defined as the loss function, quantifying the difference between the predicted output and the true labels. The function is defined as:

$$\mathcal{L}(\theta) = \mathcal{L}_{CE}(Y_{label}, Y_{out}) \quad (58)$$

where \mathcal{L}_{CE} denotes the cross-entropy loss function, Y_{label} represents the one-hot encoding of the true class labels for the samples, and θ refers to the trainable parameters of the SGET.

6. Experimental validation and analysis

6.1. Experimental setup and data preparation

During the training process, we utilize the PyTorch deep learning framework, with PyTorch Geometric specifically used to construct the synergy graph representation module. The optimizer and loss function is Adam (Mystkowski et al., 2024) and cross-entropy respectively. We implement a fixed step learning rate decay strategy, starting with an initial learning rate of 0.001, which is reduced by 20 % every 5 epochs. During the training of SGET, the initial learning rate for all network layers, except for the final classification layer, is set to 0.00001. The batch size is 32, the dropout is 0.2, and the epoch is 100. The proposed ATFMD-SGET framework was tested and validated in the hardware and software environment listed in Table 3.

In this study, the proposed method was validated and analyzed using two datasets to assess its performance in rolling bearing fault diagnosis under noisy and harsh conditions. Dataset I was collected from a self-developed rotating machinery test bench at Yangzhou University, and Dataset II was sourced from Southeast University (Shao, McAleer, Yan, & Baldi, 2018).

1) Dataset I was obtained from the rotating machinery test bench, developed by Yangzhou University. The test bench is shown in Section 3.1 regarding its experimental device and settings. This dataset is collected from the bearing test module of the experimental platform, with a sampling frequency of 16 kHz. The detailed information of Dataset I is illustrated in Table 4. It can be seen that Dataset I is classified

Table 3
Software and hardware parameters.

Parameters	Detail	Parameters	Detail
GPU	NVIDIA RTX 3090	CPU	AMD 5950X
RAM	64G	Coding language	Python 3.9
PyTorch Version	PyTorch 1.10.2	Geometric Library	Geometric 2.0.4

Table 4
Detailed description of Dataset I.

Health states of the rotating machinery	Defect size(mm)	Labels of conditions
Inner ring fault	0.5 × 0.5	0
Outer ring fault	0.5 × 0.5	1
Ball fault	0.5 × 0.5	2
Cage break	/	3
Inner-outer ring combined fault	0.5 × 0.5	4
Inner ring-ball combined fault	0.5 × 0.3	5
Healthy state	0	6

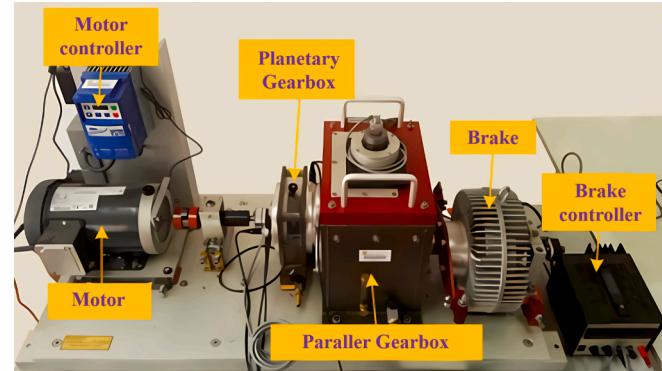


Fig. 16. Experimental setup for Dataset II.

Table 5
Detailed description of Dataset II.

Health states of the transmission system	Labels of conditions
Inner ring fault	0
Outer ring fault	1
Roller fault	2
Compound fault	3
Health state	4

into seven categories based on different fault types. In addition to the healthy state, the defects include inner ring fault (IRF), outer ring fault (ORF), ball fault (BF), cage break fault (CF), inner-outer ring combined fault (IORF), and inner ring-ball combined fault (IRBF). For simplicity, each condition has been assigned a unique label.

2) Dataset II was obtained in the drivetrain dynamics simulator, developed by Southeast University. The experimental setup is shown in Fig. 16. The dataset contains two sub-datasets, each corresponding to specific operating conditions: 20 Hz-0 V and 30 Hz-2 V speed system loads. Within the bearing dataset sampled at a frequency of 5120 Hz, five data files represent four fault types in addition to the health status.

The detailed information of Dataset II is illustrated in Table 5. The bearing faults include inner ring fault, outer ring fault, roller fault, and compound faults. In this study, we apply the bearing dataset under the 20 Hz-0 V load condition to validate the performance of ATFMD-SGET.

6.2. Applications of the ATFMD-SGET method

This paper introduces a novel ATFMD-SGET model, specifically designed for robust fault diagnosis in the presence of complex noise and harsh environments. The detailed process is outlined in Fig. 17.

Step 1: Vibration signals from bearings are first collected and transmitted to a computational device. The ATFMD method is employed to denoise the noisy signals and trim irrelevant features. Feature representation is further enhanced by image encoding and time-frequency analysis, constructing a feature topology through the extraction and mapping of spatial phase and time-frequency characteristics.

Step 2: The training dataset is input into the SGET model, which

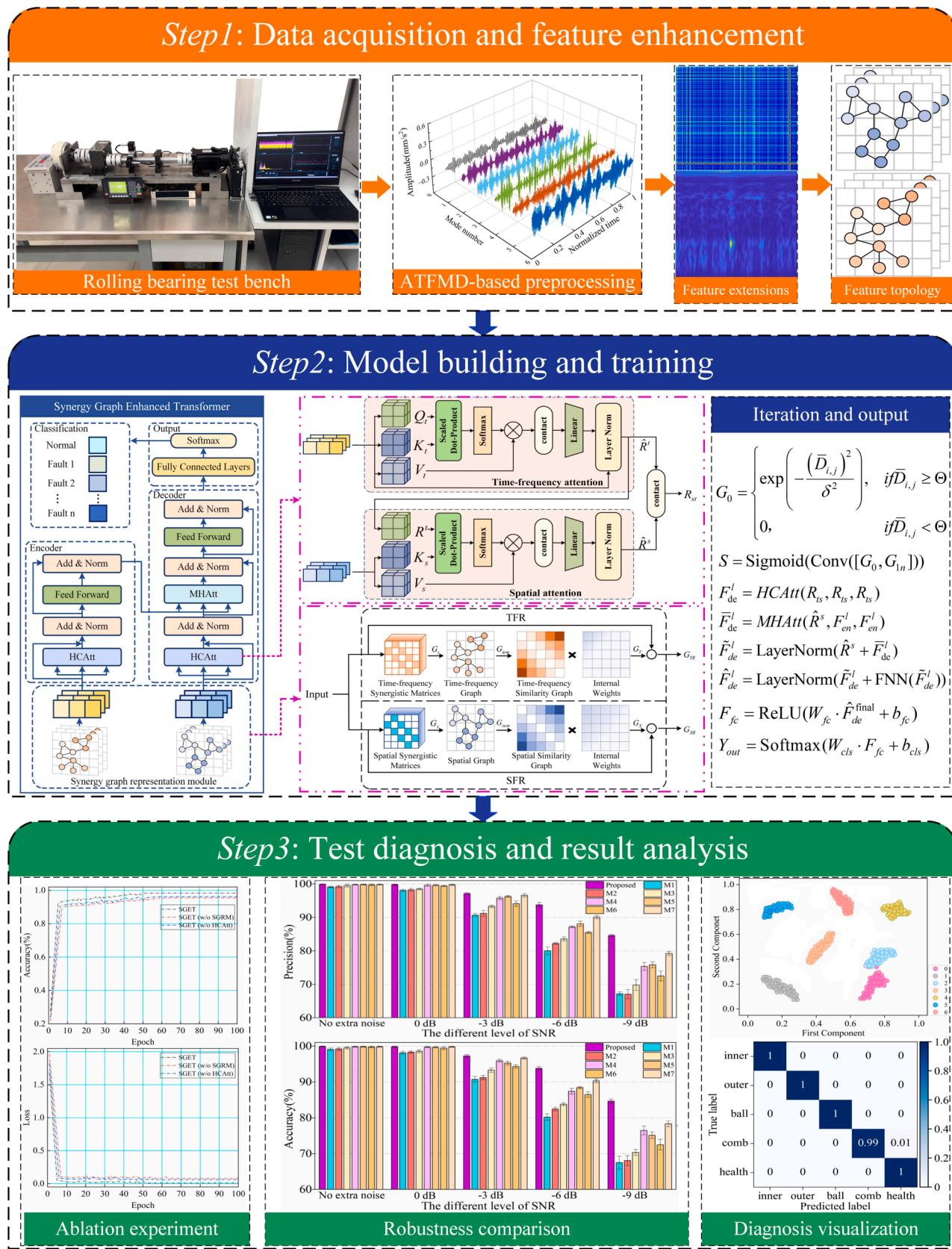


Fig. 17. Running flow of the proposed method.

integrates the SGRM, hierarchical cross-attention modules, and output modules. The model operates under an encode-process-decode framework, utilizing cross-entropy loss for reverse optimization to ensure robustness and generalization.

Step 3: Ablation experiments are conducted to evaluate the contribution of each model component to the fault diagnosis task. The trained model is then tested with the dataset, followed by a robustness analysis and visualization of the results.

6.3. Ablation analysis

In order to evaluate the impact of different modules on model performance, we conducted ablation experiments on the SGRM and STCAtt modules in the SGET model. The performance metrics used for evaluation include accuracy, precision, recall, and F1-score (Min et al., 2023). Accuracy quantifies the overall classification success, precision measures the proportion of correct positive predictions, recall evaluates the model's effectiveness in detecting true positives, and F1-score harmonizes precision and recall, offering a comprehensive performance measure, especially in cases of class imbalance. The definitions of these four metrics are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (59)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (60)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (61)$$

$$\text{F1 - Score} = \frac{2TP}{2TP + FP + FN} \quad (62)$$

The experimental design is as follows: (1) SGET (w/o SGRM): In this setup, the SGRM was removed, and the multidimensional data were directly fed into the diagnostic model without graph-based information enhancement. (2) SGET (w/o HCAtt): This configuration removed the HCAtt and replaced it with traditional multi-head attention.

The experiments were conducted on Dataset I, with 70 % of the data used for training and 30 % for testing. To test model robustness, Gaussian noise with a SNR of -3 dB was introduced. Each method was repeated 10 times, and the mean and standard deviation of the results were computed to ensure consistency. These ablation studies aimed to investigate the contribution and effectiveness of each module and assess their impact on diagnostic performance. The results are summarized in Table 6.

As demonstrated in Table 6, the removal of either module leads to a decline in the diagnostic performance of the SGET model. The most significant impact comes from the exclusion of the SGRM, which results in a 2.94 % decrease in average accuracy, a 2.98 % drop in precision, a 2.97 % reduction in recall, and a 2.97 % decrease in F1-score. The absence of SGRM hinders the ability of the model to construct feature graphs and facilitate multidimensional feature interaction, leading to a noticeable drop in overall performance. In contrast, the impact of removing the HCAtt is slightly smaller but remains notable, with accuracy dropping by 2.09 %, precision by 2.11 %, recall by 2.13 %, and F1-

score by 2.12 %. Moreover, the observed increase in performance variance suggests a decline in model stability. To further illustrate these effects, we visualized the loss and accuracy curves for SGET and its two variants, SGET (w/o SGRM) and SGET (w/o HCAtt), as shown in Fig. 18.

As illustrated in Fig. 18, removing the two modules leads to increased fluctuations in the loss of model and negatively impacts convergence. These findings indicate that both the SGRM and HCAtt modules play a vital role in improving model robustness and optimizing the utilization of multidimensional feature interactions.

6.4. Comparison experiment

6.4.1. Case 1: Impact of training sample ratios on model performance

Preliminary validation of the ATFM-SGET was first performed using Dataset I. To evaluate the feasibility and effectiveness of the proposed method under normal noise conditions with insufficient training samples, Gaussian white noise with a SNR of 0 dB was added to create various sub-datasets. Detailed information on the sub-datasets is presented in Table 7. These datasets feature varying ratios of training to testing samples, which are utilized to validate the impact of changes in training sample sizes on the generalization performance of various diagnostic models. In Case 1, each data sample consists of 1024 consecutive points, and there is no overlap between two consecutive samples.

We compared the proposed model with seven deep learning models. First, we selected three classical models: DCNN, DenseNet, and Transformer. Additionally, we selected four advanced models developed in recent years to further investigate the diagnostic capabilities of more complex architectures: multiscale deep graph convolutional networks (MS-DGCNs) (Zhao et al., 2022), which utilize multiscale graph convolution to process complex structured data; CapsFormer (Xu, Tao, Li, & Zhong, 2023), which combines the strengths of capsule networks (CapsNet) and Transformer; time-frequency Transformer (TFT) (Ding, Jia, Miao, & Cao, 2022), designed for feature extraction in the time-frequency domain; and Convformer (Han, Shao, Cheng, Yang, & Cai, 2022), which excels in both local feature capture and global information fusion. Moreover, these advanced models incorporate the signal processing techniques used in the original studies, as detailed in Table 8.

The performance of different methods in case 1 is presented in Table 9. The experimental results demonstrate that when sufficient training samples are available, all methods exhibit strong robustness against small random noise. This robustness arises from the effective noise suppression by signal processing techniques and the powerful data handling capabilities of deep learning models. However, as the number of training samples decreases, especially in cases of significant data shortage, the performance of traditional methods declines markedly across four evaluation metrics. Notably, three traditional models show the most pronounced performance degradation due to their lack of deep feature extraction capabilities. In contrast, ATFM-SGET maintains superior performance, even with only 40 % of the training data, achieving an average accuracy of 96.87 %, a precision of 96.45 %, a recall of 96.36 %, and an F1-score of 96.40 %. Compared to advanced methods M4, M5, M6, and M7, ATFM-SGET improves average F1-score by 2.01 %, 0.83 %, 3.32 %, and 0.98 %, respectively.

Furthermore, we conducted a visualization of the feature distribution to provide a more intuitive evaluation. Given the effectiveness of t-distributed stochastic neighbor embedding (t-SNE) in dimensionality reduction and clustering, we applied t-SNE to evaluate the quality of extracted features in Sub-dataset 3, which has limited training samples. As shown in Fig. 19, the comparative methods suffer from feature overlap and suboptimal clustering in specific regions, compromising the accuracy. Specifically, the features extracted by the traditional models (M1, M2, and M3) show significant overlap, severely impairing fault category discrimination. In contrast, the features extracted by the advanced methods (M4, M5, M6, and M7) demonstrate improved clustering performance, though some degree of entanglement between

Table 6
Ablation study of the proposed model.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SGET (w/o SGRM)	95.52 ± 0.43	95.50 ± 0.47	95.51 ± 0.42	95.51 ± 0.45
SGET (w/o HCAtt)	96.27 ± 0.38	96.27 ± 0.43	96.25 ± 0.39	96.26 ± 0.41
SGET	98.26 ± 0.33	98.28 ± 0.41	98.28 ± 0.28	98.28 ± 0.33

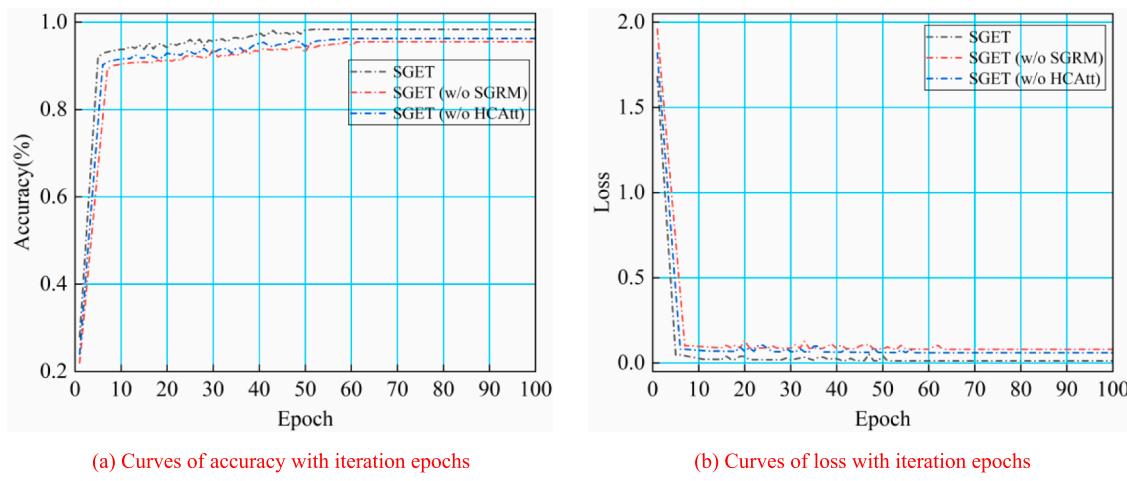


Fig. 18. Ablation analysis in terms of SGET.

Table 7
Detailed information on sub-datasets in case 1.

Sub-datasets	Training size (%)	Testing size (%)	SNR of the noise
Sub-dataset 1	70	30	0 dB
Sub-dataset 2	60	40	0 dB
Sub-dataset 3	40	60	0 dB

Table 8
Detailed description of different methods.

Model	Signal processing	Abbreviated method
SGET	ATFMD	Proposed
DCNN	VMD	M1
DenseNet	VMD	M2
Transformer	VMD	M3
MS-DGCNs	FFT	M4
CapsFormer	STFT	M5
TFT	SWT	M6
Convformer	Original signal	M7

certain classes remains. Through comparative analysis, the proposed ATFMD-SGET exhibits superior performance. As illustrated in Fig. 19(a), the feature clusters become more compact, and the boundaries clearer. These results show that SGET effectively extracts key features under limited sample conditions, even in high-dimensional spaces. This advantage stems from the SGRM, which effectively enhances feature interaction modeling by refining feature maps, leading to more

representative features that excel in low-sample scenarios.

6.4.2. Case 2: Impact of heavy noise on model performance

Dataset II from Southeast University was employed to further assess the robustness of ATFMD-SGET against severe noise. In Case 2, 70 % of the dataset was randomly selected as the training set, while the remaining 30 % was used as the test set. Gaussian noise with varying levels was introduced (SNR of 0, -3, -6, and -9 dB) was introduced to evaluate model robustness. Under these conditions, the proposed method was compared to methods listed in Table 7, with each method evaluated through 10 repeated experiments. The mean and standard deviation across the trials were calculated to assess performance.

As shown in Fig. 20, increasing noise levels result in a performance decline across methods, with the most noticeable differences at -9 dB. Conventional baseline models, particularly DCNN, exhibit significant fluctuations, with average accuracy, precision, and recall of 67.48 %, 67.18 %, and 67.23 %, respectively, under -9 dB noise. Among the advanced techniques, M4 leverages multi-scale graph convolution to process complex structured data. However, due to the absence of an attention mechanism for key features, it achieves an average accuracy of 76.43 %. M5, despite integrating CapsNet and Transformer, shows poor robustness with 75.85 % accuracy. M6, limited by its focus on time-frequency domain features, struggles to capture fault characteristics in high-noise environments, with an accuracy of 72.62 %. M7, the second-best in robustness tests, integrates both local and global spatial information, achieving an accuracy of 78.31 %, but still struggles with complex fault modes. In contrast, the method proposed outperforms all others across all metrics, particularly under high-noise conditions.

Table 9
Performance of different methods in case 1 (%).

Sub-datasets	Metric	Proposed	M1	M2	M3	M4	M5	M6	M7
Sub-dataset 1	Accuracy	99.85	99.10	99.21	99.52	99.83	99.83	99.76	99.85
	Precision	99.86	99.10	99.18	99.48	99.79	99.82	99.74	99.83
	Recall	99.82	99.02	99.16	99.49	99.81	99.82	99.74	99.81
	F1-score	99.84	99.06	99.17	99.48	99.80	99.82	99.74	99.82
Sub-dataset 2	Accuracy	99.37	95.45	96.88	98.12	99.20	99.33	98.95	99.32
	Precision	99.36	95.42	96.88	98.08	99.18	99.28	98.93	99.25
	Recall	99.38	95.34	96.75	98.02	99.22	99.30	98.95	99.36
	F1-score	99.37	95.38	96.81	98.05	99.20	99.29	98.94	99.30
Sub-dataset 3	Accuracy	96.87	83.59	86.78	88.76	94.36	95.95	93.25	95.88
	Precision	96.45	83.24	87.16	88.65	94.42	95.83	93.22	95.65
	Recall	96.36	82.87	85.92	88.54	94.36	95.32	92.94	95.19
	F1-score	96.40	83.05	86.53	88.59	94.39	95.57	93.08	95.42

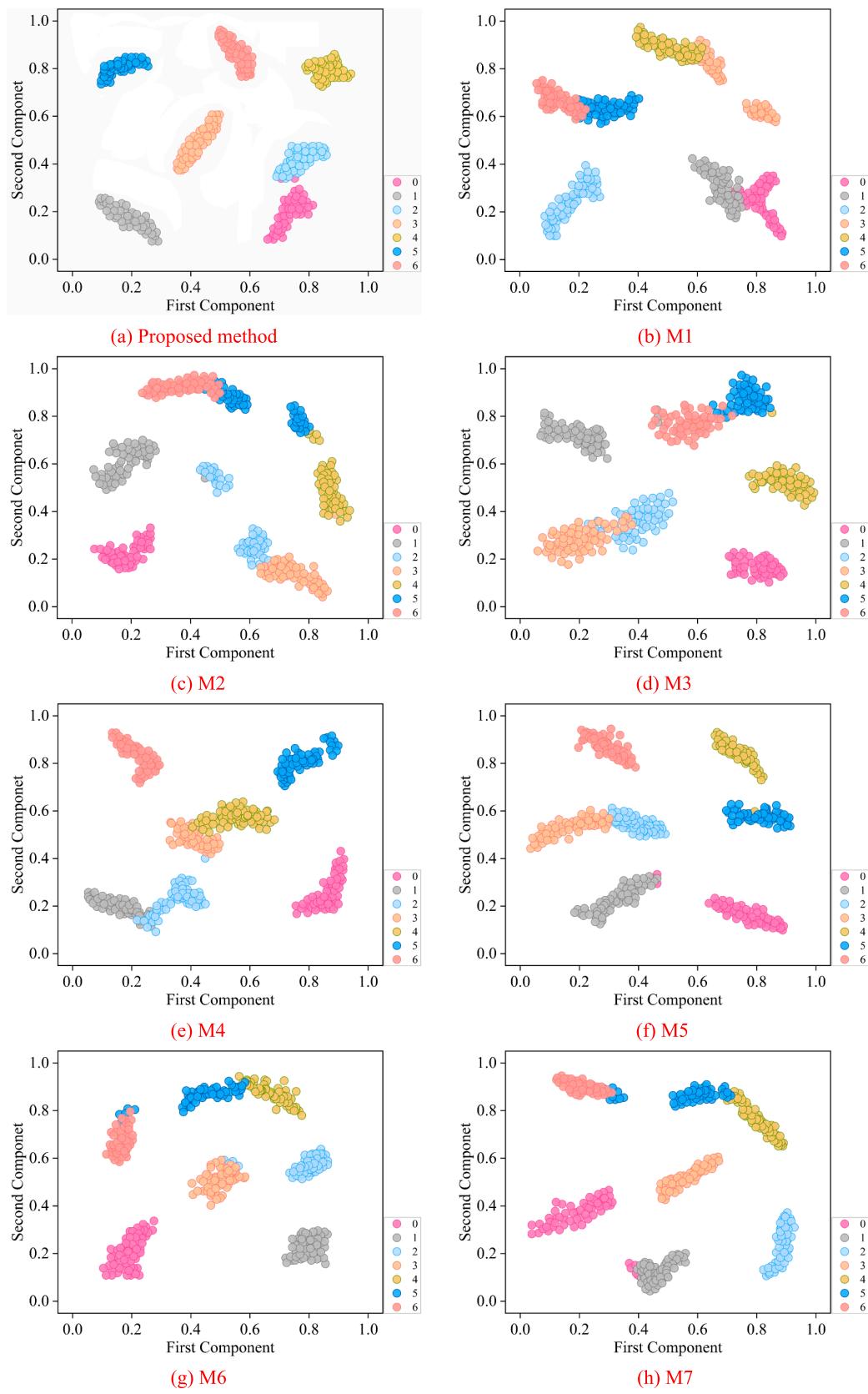


Fig. 19. Feature visualization results of different methods sufficient samples.

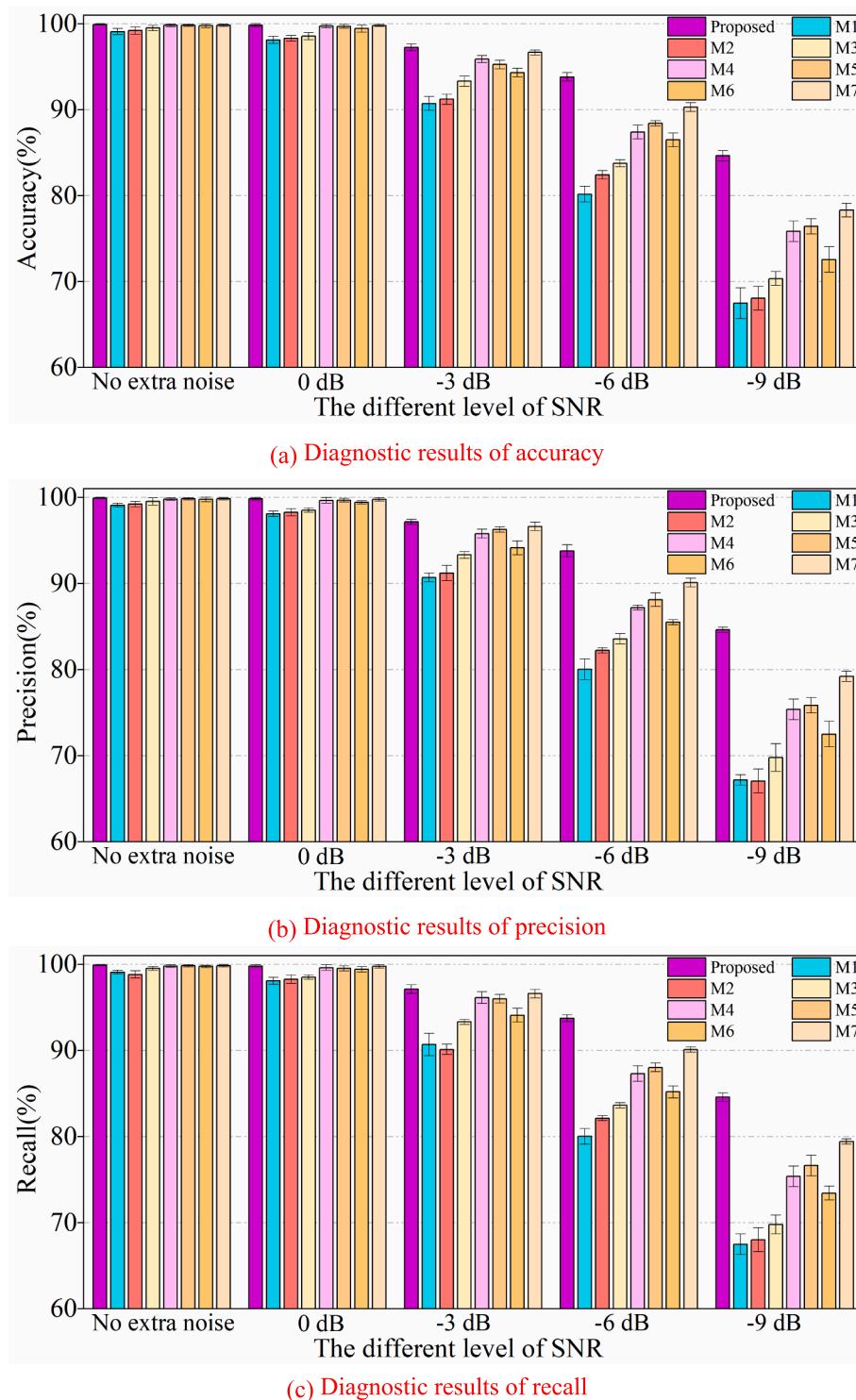


Fig. 20. Diagnostic results of methods under different SNR.

Specifically, it attains an average accuracy, precision, and recall of 84.56 %, 84.58 %, and 84.58 % at -9 dB noise. Compared to M4, M5, M6, and M7, accuracy improved by 8.13 %, 8.71 %, 11.94 %, and 6.25 %, respectively. Moreover, the proposed method exhibits lower standard deviations across these diagnostic metrics, indicating superior stability under different noise conditions. This further underscores the high robustness of the ATFMD-SGET method against noise interference.

This advantage stems from the precise frequency separation of ATFMD, which effectively suppresses disruptive noise components.

ATFMD improves data quality by avoiding noise-induced erroneous edges when constructing graphs directly from raw samples. Additionally, the hierarchical cross-attention sharpens the perception to essential features during fusion. This enables a more granular and thorough fusion of information. The classification confusion matrix is utilized in this study to further illustrate the diagnostic details of proposed method, as shown in Fig. 21.

The experimental results show that the proposed method can accurately detect the health status in most samples. Under Gaussian noise

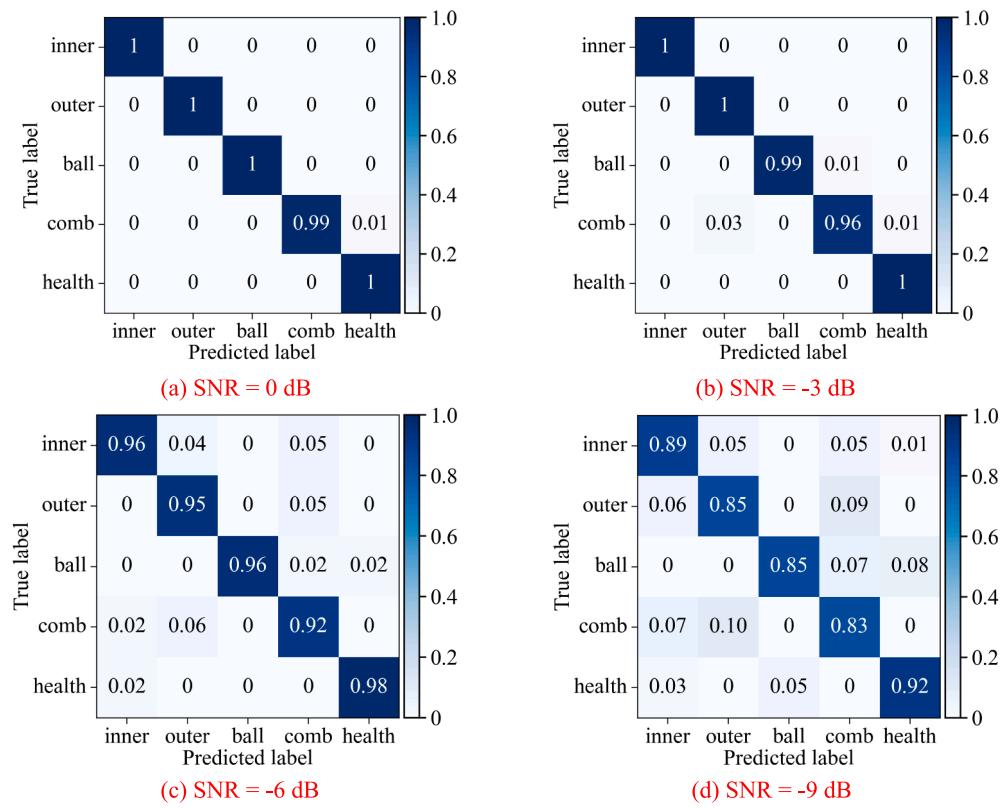


Fig. 21. The confusion matrix of proposed method in case 2.

with SNRs of 0 and -3 dB, the robustness detection rate for state 4 (the healthy state) reached 100 %, clearly demonstrating the method's outstanding capability in health state identification. Even under an SNR of -6 dB, the accuracy of fault diagnosis for all states remained above 90 %, further confirming the efficiency and resilience. Although the diagnostic accuracy declines under extreme noise conditions, this challenge can be attributed to the inherent similarity between composite and single fault modes in noisy environments, which hinders precise fault diagnosis. These findings suggest that the proposed method leads in diagnostic performance under complex noise conditions, but precise diagnosis of composite faults in harsher environments will require further investigation.

7. Conclusion

In this study, we propose a fusion-driven fault diagnosis method based on ATFMD and SGET, specifically tailored for diagnosing rolling bearing faults in noise-interfered environments. This method is particularly effective in addressing the diagnostic challenges posed by complex and harsh conditions. By integrating the periodic pulse characteristics of fault signals with Bayesian optimization, ATFMD efficiently extracts robust features from raw signals, significantly reducing the impact of noise. Given the challenge of extracting sufficient fault information from one-dimensional feature domains, we introduce image encoding and time-frequency analysis to expand the feature space. This method maps spatial phase and time-frequency features, constructing feature topology graphs that captures comprehensive fault information. Building on the feature topology graph, the SGET model further enhances fault information fusion by leveraging the SGRM. This module facilitates the interaction between features, while a hierarchical cross-attention mechanism is employed to prioritize critical features during the fusion process. Case studies on public and experimental datasets show that the ATFMD-SGET exhibits exceptional adaptability in

handling noisy signals, surpassing existing advanced diagnostic models in both robustness and accuracy. In future research, we will focus on enhancing adaptability to composite fault scenarios with limited training samples and lightweight model optimization to reduce computational complexity.

CRediT authorship contribution statement

Lin Zhu: Methodology, Conceptualization, Investigation, Writing – review & editing. **Jin Wang:** Methodology, Conceptualization, Software, Writing – original draft. **Min Chen:** Visualization, Writing – review & editing. **Lintong Liu:** Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This article was supported by National Natural Science Foundation of China (51805447), Natural Science Foundation of the Jiangsu Higher Education (22KJB460010), Innovation and Promotion Project of Forestry Science and Technology for Jiangsu Province (LYKJ[2023]06), Yangzhou Science and Technology Plan (City School Cooperation Project) (YZ2022193). Cyan blue Project of Yangzhou University.

References

- Alavi, H., Ohadi, A., & Niaki, S. T. (2022). A novel targeted method of informative frequency band selection based on lagged information for diagnosis of gearbox single and compound faults. *Mechanical Systems and Signal Processing*, 170, Article 108828. <https://doi.org/10.1016/j.ymssp.2022.108828>
- Cao, Y., Sun, Y., Xie, G., & Li, P. (2021). A sound-based fault diagnosis method for railway point machines based on two-stage feature selection strategy and ensemble classifier. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), 12074–12083. <https://doi.org/10.1109/TITS.2021.3109632>
- Chauhan, S., Vashishtha, G., Kumar, R., Zimroz, R., Gupta, M. K., & Kundu, P. (2024). An adaptive feature mode decomposition based on a novel health indicator for bearing fault diagnosis. *Measurement*, 226, Article 114191. <https://doi.org/10.1016/j.measurement.2024.114191>
- Chen, B., Zhang, W., Gu, J. X., Song, D., Cheng, Y., Zhou, Z., Gu, F. S., & Ball, A. D. (2023). Product envelope spectrum optimization-gram: An enhanced envelope analysis for rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 193, Article 110270. <https://doi.org/10.1016/j.ymssp.2023.110270>
- Chen, J., Lin, C., Yao, B., Yang, L., & Ge, H. (2023). Intelligent fault diagnosis of rolling bearings with low-quality data: A feature significance and diversity learning method. *Reliability Engineering & System Safety*, 237, Article 109343. <https://doi.org/10.1016/j.ress.2023.109343>
- Cheng, J., Yang, Y., Hu, N., Cheng, Z., & Cheng, J. (2021). A noise reduction method based on adaptive weighted symplectic geometry decomposition and its application in early gear fault diagnosis. *Mechanical Systems and Signal Processing*, 149, Article 107351. <https://doi.org/10.1016/j.ymssp.2020.107351>
- Dibaj, A., Etefagh, M. M., Hassannejad, R., & Eghgahgi, M. B. (2021). A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults. *Expert Systems with Applications*, 167, Article 114094. <https://doi.org/10.1016/j.eswa.2020.114094>
- Ding, Y., Jia, M., Miao, Q., & Cao, Y. (2022). A novel time-frequency Transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings. *Mechanical Systems and Signal Processing*, 168, Article 108616. <https://doi.org/10.1016/j.ymssp.2021.108616>
- Gawde, S., Patil, S., Kumar, S., Kamat, P., Kotecha, K., & Abraham, A. (2023). Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach: A review of two decades of research. *Engineering Applications of Artificial Intelligence*, 123, Article 106139. <https://doi.org/10.1016/j.engappai.2023.106139>
- Gharesi, N., Arefi, M. M., Razavi-Far, R., Zarei, J., & Yin, S. (2020). A neuro-wavelet based approach for diagnosing bearing defects. *Advanced Engineering Informatics*, 46, Article 101172. <https://doi.org/10.1016/j.aei.2020.101172>
- Guan, L., Qiao, F., Zhai, X., & Wang, D. (2022). Model evolution mechanism for incremental fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 71, 3522111. <https://doi.org/10.1109/TIM.2022.3200695>
- Han, S., Shao, H., Cheng, J., Yang, X., & Cai, B. (2022). Conformer-NSE: A novel end-to-end gearbox fault diagnosis framework under heavy noise using joint global and local information. *IEEE/ASME Transactions on Mechatronics*, 28(1), 340–349. <https://doi.org/10.1109/TMECH.2022.3199985>
- He, B., Li, J., Li, D., Dong, J., & Zhu, L. (2022). Quadratic function based price adjustment strategy on monitoring process of power consumption load in smart grid. *International Journal of Electrical Power & Energy Systems*, 134, Article 107124. <https://doi.org/10.1016/j.ijepes.2021.107124>
- He, X., Zhou, X., Li, J., Mechefske, C. K., Wang, R., Yao, G., & Liu, Q. (2023). Adaptive feature mode decomposition: A fault-oriented vibration signal decomposition method for identification of multiple localized faults in rotating machinery. *Nonlinear Dynamics*, 111(17), 16237–16270. <https://doi.org/10.1007/s11071-023-08703-4>
- Kedadouche, M., Thomas, M., & Tahan, A. J. M. S. (2016). A comparative study between Empirical Wavelet Transforms and Empirical Mode Decomposition Methods: Application to bearing defect diagnosis. *Mechanical Systems and Signal Processing*, 81, 88–107. <https://doi.org/10.1016/j.ymssp.2016.02.049>
- Kong, J., Fan, X., Zuo, M., Deveci, M., J. X., & Zhong, K. (2024). ADCT-Net: Adaptive traffic forecasting neural network via dual-graphic cross-fused transformer. *Information Fusion*, 103, 102122. <https://doi.org/10.1016/j.inffus.2023.102122>
- Kumar, A., Zhou, Y., & Xiang, J. (2021). Optimization of VMD using kernel-based mutual information for the extraction of weak features to detect bearing defects. *Measurement*, 168, Article 108402. <https://doi.org/10.1016/j.measurement.2020.108402>
- Li, G., Ao, J., Hu, J., Hu, D., Liu, Y., & Huang, Z. (2024). Dual-source gramian angular field method and its application on fault diagnosis of drilling pump fluid end. *Expert Systems with Applications*, 237, Article 121521. <https://doi.org/10.1016/j.eswa.2023.121521>
- Li, H., Liu, T., Wu, X., & Chen, Q. (2020). A bearing fault diagnosis method based on enhanced singular value decomposition. *IEEE Transactions on Industrial Informatics*, 17(5), 3220–3230. <https://doi.org/10.1109/TII.2020.3001376>
- Li, S., Ji, J. C., Xu, Y., Sun, X., Feng, K., Sun, B., Wang, Y., Gu, F., Zhang, K., & Ni, Q. (2023). IFD-MDCN: Multibranch denoising convolutional networks with improved flow direction strategy for intelligent fault diagnosis of rolling bearings under noisy conditions. *Reliability Engineering & System Safety*, 237, Article 109387. <https://doi.org/10.1016/j.ress.2023.109387>
- Li, T., Zhou, Z., Li, S., Sun, C., Yan, R., & Chen, X. (2022). The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mechanical Systems and Signal Processing*, 168, Article 108653. <https://doi.org/10.1016/j.ymssp.2021.108653>
- Lin, J., Shao, H., Zhou, X., Cai, B., & Liu, B. (2023). Generalized MAML for few-shot cross-domain fault diagnosis of bearing driven by heterogeneous signals. *Expert Systems with Applications*, 230, Article 120696. <https://doi.org/10.1016/j.eswa.2023.120696>
- Lv, H., Chen, J., Pan, T., Zhang, T., Feng, Y., & Liu, S. (2022). Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application. *Measurement*, 199, Article 111594. <https://doi.org/10.1016/j.measurement.2022.111594>
- Lyu, W., Xue, P., Yang, F., Yan, C., Hong, Z., Zeng, X., & Zhou, D. (2017). An efficient bayesian optimization approach for automated optimization of analog circuits. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(6), 1954–1967. <https://doi.org/10.1109/TCI.2017.2768826>
- Miao, Y., Zhang, B., Li, C., Lin, J., & Zhang, D. (2022). Feature mode decomposition: New decomposition theory for rotating machinery fault diagnosis. *IEEE Transactions on Industrial Electronics*, 70(2), 1949–1960. <https://doi.org/10.1109/TIE.2022.3156156>
- Min, H., Fang, Y., Wu, X., Lei, X., Chen, S., Teixeira, R., Zhu, B., Zhao, X. M., & Xu, Z. (2023). A fault diagnosis framework for autonomous vehicles with sensor self-diagnosis. *Expert Systems with Applications*, 224, Article 120002. <https://doi.org/10.1016/j.eswa.2023.120002>
- Mo, Z., Zhang, Z., & Tsui, K. L. (2021). The variational kernel-based 1-D convolutional neural network for machinery fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 70, 3523210. <https://doi.org/10.1109/TIM.2021.3105252>
- Mystkowski, A., Wolniakowski, A., Idzkowski, A., Ciężkowski, M., Ostaszewski, M., Kociszewski, R., Kotowski, A., Kulesza, Z., Dobrzański, S., & Miastkowski, K. (2024). Measurement and diagnostic system for detecting and classifying faults in the rotary hay tedder using multilayer perceptron neural networks. *Engineering Applications of Artificial Intelligence*, 133, Article 108513. <https://doi.org/10.1016/j.engappai.2024.108513>
- Nashed, M. S., Renno, J., Mohamed, M. S., & Reuben, R. L. (2023). Gas turbine failure classification using acoustic emissions with wavelet analysis and deep learning. *Expert Systems with Applications*, 232, Article 120684. <https://doi.org/10.1016/j.eswa.2023.120684>
- Ni, Q., Ji, J. C., Feng, K., & Halkon, B. (2022). A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis. *Mechanical Systems and Signal Processing*, 164, Article 108216. <https://doi.org/10.1016/j.ymssp.2021.108216>
- Pan, T., Chen, J., Ye, Z., & Li, A. (2022). A multi-head attention network with adaptive meta-transfer learning for RUL prediction of rocket engines. *Reliability Engineering & System Safety*, 225, Article 108610. <https://doi.org/10.1016/j.ress.2022.108610>
- Pan, H., Zhang, Y., Cheng, J., & Zheng, J. (2023). Symplectic Geometry Transformation-Based Periodic Segment Method: Algorithm and Applications. *IEEE Transactions on Instrumentation and Measurement*, 72, 6502708. <https://doi.org/10.1109/TIM.2023.3271006>
- Parvin, F., Faiz, J., Qi, Y., Kalhor, A., & Akin, B. (2023). A comprehensive interturn fault severity diagnosis method for permanent magnet synchronous motors based on transformer neural networks. *IEEE Transactions on Industrial Informatics*, 19(11), 10923–10933. <https://doi.org/10.1109/TII.2023.3242773>
- Qin, Y., Liu, H., & Mao, Y. (2024). Faulty rolling bearing digital twin model and its application in fault diagnosis with imbalanced samples. *Advanced Engineering Informatics*, 61, Article 102513. <https://doi.org/10.1016/j.aei.2024.102513>
- Rajabi, S., Azari, M. S., Santini, S., & Flammini, F. (2022). Fault diagnosis in industrial rotating equipment based on permutation entropy, signal processing and multi-output neuro-fuzzy classifier. *Expert systems with applications*, 206, Article 117754. <https://doi.org/10.1016/j.eswa.2022.117754>
- Ruan, D., Wang, J., Yan, J., & Gühmann, C. (2023). CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis. *Advanced Engineering Informatics*, 55, Article 101877. <https://doi.org/10.1016/j.aei.2023.101877>
- Shao, H., Xia, M., Wan, J., & de Silva, C. W. (2021). Modified stacked autoencoder using adaptive Morlet wavelet for intelligent fault diagnosis of rotating machinery. *IEEE/ASME Transactions on Mechatronics*, 27(1), 24–33. <https://doi.org/10.1109/TMECH.2021.3058061>
- Shao, S., McAleer, S., Yan, R., & Baldi, P. (2018). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 15(4), 2446–2455. <https://doi.org/10.1109/TII.2018.2864759>
- Singh, A., Nagar, J., Sharma, S., & Kotiyal, V. (2021). A Gaussian process regression approach to predict the k-barrier coverage probability for intrusion detection in wireless sensor networks. *Expert Systems with Applications*, 172, Article 114603. <https://doi.org/10.1016/j.eswa.2021.114603>
- Song, X., Wu, C., Song, S., Stojanovic, V., & Tejado, I. (2024). Fuzzy wavelet neural adaptive finite-time self-triggered fault-tolerant control for a quadrotor unmanned aerial vehicle with scheduled performance. *Engineering Applications of Artificial Intelligence*, 131, Article 107832. <https://doi.org/10.1016/j.engappai.2023.107832>
- Tang, Y., Zhang, X., Qin, G., Long, Z., Huang, S., Song, D., & Shao, H. (2021). Graph cardinality preserved attention network for fault diagnosis of induction motor under varying speed and load condition. *IEEE Transactions on Industrial Informatics*, 18(6), 3702–3712. <https://doi.org/10.1109/TII.2021.3112696>
- Tao, H., Shi, H., Qiu, J., Jin, G., & Stojanovic, V. (2023). Planetary gearbox fault diagnosis based on FDKNN-DGAT with few labeled data. *Measurement Science and Technology*, 35(2), Article 025036. <https://doi.org/10.1088/1361-6501/ad0fd>
- Tao, H., Qiu, J., Chen, Y., Stojanovic, V., & Cheng, L. (2023). Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion. *Journal of the Franklin Institute*, 360(2), 1454–1477. <https://doi.org/10.1016/j.jfranklin.2022.11.004>
- Touti, W., Salah, M., Bacha, K., & Chaari, A. (2022). Condition monitoring of a wind turbine drivetrain based on generator stator current processing. *ISA transactions*, 128, 650–664. <https://doi.org/10.1016/j.isatra.2021.10.014>

- Vashishtha, G., & Kumar, R. (2022). An amended grey wolf optimization with mutation strategy to diagnose bucket defects in Peltton wheel. *Measurement*, 187, Article 110272. <https://doi.org/10.1016/j.measurement.2021.110272>
- Wang, J., Zeng, Z., Zhang, H., Barros, A., & Miao, Q. (2022). An hybrid domain adaptation diagnostic network guided by curriculum pseudo labels for electro-mechanical actuator. *Reliability Engineering & System Safety*, 228, Article 108770. <https://doi.org/10.1016/j.ress.2022.108770>
- Wang, R., Jiang, H., Zhu, K., Wang, Y., & Liu, C. (2022). A deep feature enhanced reinforcement learning method for rolling bearing fault diagnosis. *Advanced Engineering Informatics*, 54, Article 101750. <https://doi.org/10.1016/j.aei.2022.101750>
- Wang, R., Zhuang, Z., Tao, H., Paszke, W., & Stojanovic, V. (2023). Q-learning based fault estimation and fault tolerant iterative learning control for MIMO systems. *ISA transactions*, 142, 123–135. <https://doi.org/10.1016/j.isatra.2023.07.043>
- Wang, S., Liu, Z., Jia, Z., Zhao, W., & Li, Z. (2024). Intermittent fault diagnosis for electronics-rich analog circuit systems based on multi-scale enhanced convolution transformer network with novel token fusion strategy. *Expert Systems with Applications*, 238, Article 121964. <https://doi.org/10.1016/j.eswa.2023.121964>
- Wang, Y., Wang, H., & Peng, Z. (2021). Rice diseases detection and classification using attention based neural network and bayesian optimization. *Expert Systems with Applications*, 178, Article 114770. <https://doi.org/10.1016/j.eswa.2021.114770>
- Wu, K., Tong, W., Xie, J., Wang, F., Huang, B., & Wu, D. (2024). Optimal Weighted Envelope Spectrum: An enhanced demodulation method for extracting specific characteristic frequency of rotating machinery. *Mechanical Systems and Signal Processing*, 211, Article 111165. <https://doi.org/10.1016/j.ymssp.2024.111165>
- Xu, Y., Tao, H., Li, W., & Zhong, Y. (2023). CapsFormer: A Novel Bearing Intelligent Fault Diagnosis Framework With Negligible Speed Change Under Small-Sample Conditions. *IEEE Transactions on Instrumentation and Measurement*, 72, 3532511. <https://doi.org/10.1109/TIM.2023.3318693>
- Yan, X., Yan, W. J., Xu, Y., & Yuen, K. V. (2023). Machinery multi-sensor fault diagnosis based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network. *Mechanical Systems and Signal Processing*, 202, Article 110664. <https://doi.org/10.1016/j.ymssp.2023.110664>
- Yang, D., Qin, J., Pang, Y., & Huang, T. (2021). A novel double-stacked autoencoder for power transformers DGA signals with an imbalanced data structure. *IEEE Transactions on Industrial Electronics*, 69(2), 1977–1987. <https://doi.org/10.1109/TIE.2021.3059543>
- Yi, C., Wang, H., Zhou, Q., Hu, Q., Zhou, P., & Lin, J. (2022). An adaptive harmonic product spectrum for rotating machinery fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 72, 3500612. <https://doi.org/10.1109/TIM.2022.3230462>
- Zhang, K., Tang, B., Deng, L., & Liu, X. (2021). A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement*, 179, Article 109491. <https://doi.org/10.1016/j.measurement.2021.109491>
- Zhao, X., Yao, J., Deng, W., Ding, P., Zhuang, J., & Liu, Z. (2022). Multiscale deep graph convolutional networks for intelligent fault diagnosis of rotor-bearing system under fluctuating working conditions. *IEEE Transactions on Industrial Informatics*, 19(1), 166–176. <https://doi.org/10.1109/TII.2022.3161674>
- Zhu, S., Liao, B., Hua, Y., Zhang, C., Wan, F., & Qing, X. (2023). A transformer model with enhanced feature learning and its application in rotating machinery diagnosis. *ISA transactions*, 133, 1–12. <https://doi.org/10.1016/j.isatra.2022.07.016>
- Zuo, L., Xu, F., Zhang, C., Xiahou, T., & Liu, Y. (2022). A multi-layer spiking neural network-based approach to bearing fault diagnosis. *Reliability Engineering & System Safety*, 225, Article 108561. <https://doi.org/10.1016/j.ress.2022.108561>