10.37965/jdmd.2026.1435

Mode entropy knowledge machine: a fully automated bearing fault diagnosis model for complex operating conditions

Hongchuang Tan[a], Yiheng Su[a], Jiang Ding[a], Enci Yan[a], Wenbin Chen[a], Yao Zeng[a], Lingli Jiang[b], Yun Kong[c*]*

a. Guangxi Key Lab of Manufacturing System and Advance Manufacturing Technology, Guangxi University, Nanning, Guangxi, China

b. School of Mechanical Engineering, Foshan University, Foshan, Guangdong, China

c. School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China

** Corresponding author: hctan@gxu.edu.cn (H. Tan), kongyun@bit.edu.cn (Y. Kong)*

Received 22 May 2026; Revised 29 June 2026; Accepted 07 July 2026; Published online 28 July 2026

Abstract: Accurate diagnosis of bearing faults is critical for industrial equipment reliability. However, bearings commonly operate under complex conditions, and coupled with data imbalance and noise interference, fault diagnosis of them remains extremely challenging. To address these issues, a novel mode entropy knowledge machine (MEKM) framework for robust bearing fault diagnosis is proposed in this study. For MEKM, the mode entropy space is first constructed to decompose the vibration signal into intrinsic mode components, and the noise-resistant feature extraction and dimensionality reduction are realized by principal component analysis. Secondly, a fast classifier based on extreme learning machines is introduced, and its parameters are automatically adjusted through a particle swarm optimization to establish an adaptive extreme learning machine diagnosis model, ensuring optimal generalization under different load and speed levels. Then, a collaborative optimization paradigm is developed to coordinate mode entropy features and classifier parameters through fully automated learning, in which entropy-driven feature characterization guides the iterative refinement of decision boundaries, while classifier feedback dynamically improves the selectivity of entropy features. Finally, validation is performed on bearings with multiple operating conditions, and the results indicated that the MEKM outperformed conventional deep learning methods across both diagnostic accuracy and generalization ability. The work provides a theoretical basis and an industrially feasible solution for health monitoring of mechanical equipment.

Keywords: Rolling bearing; Mode entropy; Mode entropy space; Mode entropy knowledge machine; Fault diagnosis.

1. Introduction

In industrial systems, bearings serve as critical rotational components and are extensively employed in motors, pumps, fans, compressors, and steel manufacturing equipment [1-3]. Owing to the function of supporting rotating parts, reducing friction, and minimizing energy losses, bearings are critical for maintaining the reliable functioning of machinery [4]. However, prolonged operation often leads to wear and degradation due to friction, fatigue, and inadequate lubrication. These factors can ultimately result in bearing faults, which, if not detected and mitigated in a timely manner, may cause unexpected equipment shutdowns, production disruptions, and even catastrophic mechanical faults, incurring significant economic losses [5]. Therefore, early and accurate fault detection of bearings is essential for reliable industrial condition monitoring.

In actual industrial practice, bearing operating conditions are frequently complicated, causing the acquired vibration signals to exhibit nonstationary and nonlinear features [6]. These properties hamper the accurate extraction of fault features by conventional signal processing techniques [7, 8]. Consequently, the development of robust feature extraction approaches capable of effectively analysing complex vibration patterns has turned into a key research topic. Traditional vibration analytical methods involve frequency-domain, time-domain, and time–frequency techniques [9]. Among them, the common time-frequency domain analysis methods, including short-time Fourier transform [10], empirical mode decomposition (EMD) [11,12], ensemble empirical mode decomposition [13,14], and variational mode decomposition (VMD) [15], aim to extract discriminative signal features that differentiate fault categories. However, due to numerous interferences in the actual environment, time-frequency domain technology alone cannot fully express the characteristics of the signal. With the advancement of nonlinear dynamics and complexity theory, entropy-based measurement methods have emerged, which enable effective evaluation of signal complexity and have been utilized in fault diagnosis research [16]. Representative entropy metrics, including approximate entropy (AE) [17], sample entropy (SE) [18], permutation entropy (PE) [19], fuzzy entropy (FE) [20], and dispersion entropy (DE) [21], have demonstrated promising capabilities in capturing the intrinsic complexity of vibration signals. Zhao et al. [22] integrated EMD with AE for effective fault feature extraction in bearings. To address the sensitivity of AE to data length and parameter settings, SE was introduced. Shang et al. [23] propounded a modified SE-based approach for early fault detection in lithium-ion battery packs. Nevertheless, due to its step-function-based similarity measure, SE suffers from discontinuities in entropy profiles. FE was thus proposed to improve continuity and smoothness through fuzzy similarity metrics. Zheng et al. [24] advanced a framework based on local feature scale decomposition and FE, while Qin et al. [25] combined FE with a neural network for UAV aileron fault detection. In addition, PE and DE have also been employed for mechanical fault diagnosis. Ye et al. [26] designed an adaptive method based on PE and manifold dynamic time regulations. Zhou et al. [27] developed an approach employing improved dispersion entropy with machine learning techniques.

The aforementioned entropy metrics can quantify signal complexity from a signal on a single scale and identify fault characteristics. However, feature extraction relying on a single type of entropy is clearly not comprehensive enough. Inspired by the concept of knowledge spaces, recent studies have explored feature extraction through the fusion of multiple descriptors, aiming to construct a richer and more informative feature space. Su et al. [28] proposed a cognitive space method that integrates both entropy and frequency features to better address complex signal patterns under diverse operating conditions. Similarly, Ma et al. [29] presented a multi-entropy space by extracting multiple types of entropy and integrating them with machine learning to achieve fault diagnosis. Although these high-dimensional features can provide more comprehensive information, there may be nonlinear factors and complex relationships among different features, which inevitably introduce some redundancy and irrelevant information. Such redundancy not only increases computational overhead but may also adversely affect diagnosis accuracy. Therefore, it becomes imperative to perform dimensionality reduction to retain the most informative features while discarding noise and irrelevant components.

Common dimensionality reduction techniques involve principal component analysis (PCA) [30,31], local linear embedding (LLE) [32], linear discriminant analysis, and locality preserving projections (LPP) [33], all of which have demonstrated effectiveness in fault diagnosis applications. Zhao et al. [34] employed a deep learning model combined with PCA to extract features for fault diagnosis. Building on this, Liu et al. [35] proposed a distributed PCA method for fault diagnosis. In another study, Liu et al. [36] utilized LLE in conjunction with wavelet transform and signal decomposition to enhance fault identification. Feng et al. [37] further improved the diagnosis performance by integrating an enhanced VMD with LLE. Wang et al. [38] adopted LPP for dimensionality reduction and integrated it with a multivariate prediction model to achieve accurate bearing fault classification. As the diversity and number of extracted features increase, complex nonlinear relationships often emerge among feature dimensions. Kernel principal component analysis (KPCA) [39,40], as a nonlinear extension of PCA, has gained wide recognition for its capability to uncover nonlinear structures in data and to project them into lower-dimensional spaces. Therefore, effectively handling data with nonlinear characteristics and preserving local low-dimensional manifolds becomes essential for extracting compact and discriminative features during the fault diagnosis process.

Currently, deep learning methods including CNN [41], CNN-long short-term memory (LSTM) [42], Transformer [43], ResNet18 [44], AlexNet [45], LeNet5 [46], and multilayer perceptron (MLP) [47] are extensively used for fault diagnosis. Huang et al. [48] presented a CNN model incorporating an attention mechanism to evaluate the status of the bearings. Luo et al. [49] achieved cross-condition intelligent detection of small modular reactors using CNN-LSTM framework. Zhang et al. [50] applied ResNet18 for bearing fault diagnosis. Sinitsin et al. [51] improved the model's performance by combining hybrid input with MLP. However, these methods generally adopt end-to-end diagnosis architecture, which enables direct classification but still lacks in controllability and interpretability of feature extraction. In contrast, entropy-based feature extraction methods show significant advantages in the process of fault characterization by virtue of strong signal evaluation capability. However, traditional entropy feature extraction methods typically rely on a single indicator and cannot adequately capture signal complexity. At the same time, the lack of fusion of multi-dimensional information makes the extracted features redundant and reduces diagnostic accuracy. In addition, under multiple interference conditions and unbalanced samples, the feature extraction process becomes more difficult, causing in a substantial decline in diagnostic accuracy. To address these challenges, this study proposes a fault diagnosis model termed the mode entropy knowledge machine (MEKM). This method firstly constructs a mode entropy feature space, performs mode decomposition on the normalized vibration signals, and combines principal component analysis to achieve feature extraction and dimensionality reduction, thereby enabling the effective integration of multi-dimensional mode entropy features. Subsequently, an extreme learning machine (ELM) [52] is utilized to construct the fault classification model, while particle swarm optimization (PSO) [53] is applied for adaptive parameter adjustment, leading to improved model generalization. Finally, a collaborative optimization paradigm is designed, in which the feature selection process is dynamically adjusted through a classifier feedback mechanism, achieving joint optimization of mode entropy features and classifier parameters to strengthen the diagnostic performance for bearing faults. The primary contributions presented in the study can be outlined below:

(1) Construct a modal entropy space to characterize signal complexity through modal decomposition and entropy features, and combine principal component analysis to reduce feature redundancy, thereby improving the representation ability of fault features.

(2) A collaborative optimization paradigm is designed to achieve automatic coordination and feedback-driven optimization of mode entropy features and classifier parameters.

(3) The bearing fault diagnosis method utilizing the MEKM is proposed, with its performance is verified on multi-condition bearing experiments.

The remainder of this paper is structured as follows. Section 2 describes the methodology, including the construction of the mode entropy space, the architecture of the mode entropy knowledge machine, and the collaborative adaptive parameter optimization strategy. Section 3 presents the diagnosis method. Section 4 demonstrates the validity of the developed approach, which is validated on three single condition bearing datasets and one multi-condition bearing dataset. Section 5 is the conclusion.

2. Methodology

2.1. Construction of the mode entropy space

Entropy serves as a commonly employed metric in assessing the complexity of signal, and it has found extensive application in signal detection tasks. It exhibits strong noise resistance and can sensitively capture the nonstationary characteristics of signals, thereby revealing the intrinsic patterns of dynamic changes within the system. To comprehensively characterize the multidimensional complexity of vibration signals, four commonly used entropy features are introduced: SE, FE, PE, and DE. These metrics extract effective information from time series from different perspectives, including local predictability, fuzzy similarity, symbolic ordering structures, and amplitude distribution trends. By calculating these entropy features for each intrinsic mode function under various operating conditions, a high-dimensional set of feature vectors is constructed. By extracting the aforementioned entropy features from signal sequences under different states and integrating them to constitute the set of high-dimensional feature vectors, a fundamental mode entropy feature space is established.

Firstly, the raw signals acquired from the bearing are normalized to reduce amplitude discrepancies. The signal then undergoes decomposition using VMD to produce several intrinsic mode functions (IMFs) with bandpass characteristics. For each IMF, the corresponding mode entropy value is computed to characterize its dynamic behavior across different complexity scales. Through this process, the original time series is effectively transferred into a structured high-dimensional feature dataset consisting of multiple IMF channels and corresponding entropy features, denoted as $Q \in \mathbb{R}^{n\times 4m}$, where n and m denote the counts of samples and IMF channels. The overall procedure is illustrated in Fig. 1. From each IMF, four types of entropy-based indicators are extracted, and these features are fused to build a multi-dimensional mode entropy feature matrix. The entropy parameters used are shown in Table 1 [54] [55].

The following section provides a detailed introduction to the four entropy methods employed in this study.

Sample Entropy: SE quantifies time series complexity by evaluating the likelihood that similar patterns remain consistent as the sequence progresses. A higher probability of novel pattern generation corresponds to greater sequence complexity. The calculation formula is as follows:

$$SE = -\ln\left(\frac{A_m(r)}{B_{m+1}(r)}\right) \tag{1}$$

where $A_m(r)$ denotes the proportion of similar vector pairs of length m with tolerance r and $B_{m+1}(r)$ represents the proportion of length $m+1$ using the same tolerance.

Fuzzy entropy: FE characterizes the degree of unpredictability in a time series by quantifying the fuzzy variation in the similarity of pattern pairs as the sequence is extended. Specifically, fuzzy entropy assesses how the fuzzy similarity between pairs of similar patterns evolves with increasing pattern length, thereby revealing the intrinsic structural complexity of the sequence. The calculation formula is given by:

$$FE = \ln\left(\frac{\Phi^m(r)}{\Phi^{m+1}(r)}\right) \tag{2}$$

where $\Phi^m(r)$ denotes the average similarity for an embedding dimension of m, r represents the similarity tolerance.

Permutation entropy: PE quantifies signal complexity by examining the ordinal patterns generated from subsequence permutations, thereby capturing both its randomness and regularity. The formula is given above:

$$PE = -\sum_{j=1}^{m!} p(\pi_j)\log p(\pi_j) \tag{3}$$

where $m!$ represents the count of possible permutation types. $p(\pi_j)$ indicates the likelihood of this pattern π_j occurring.

Dispersion entropy: DE measures time series complexity by assessing local trends and the probability distribution of discrete patterns. The calculation is defined as:

$$DE = -\sum_{j=1}^{c^m}\left(\frac{n_j}{N-(m-1)\tau}\right)\log\left(\frac{n_j}{N-(m-1)\tau}\right) \tag{4}$$

where $\frac{n_j}{N-(m-1)\tau}$ denotes the relative frequency (i.e., probability p_j) of pattern j, n_j representing the count of occurrences of dispersion pattern j.

10.37965/jdmd.2026.1435

Table 1 Parameter settings for entropy feature extraction

Entropy	Embedding dimensions	Fuzzy coefficient	Similar tolerance
SE	2	/	0.2
FE	2	2	0.1
PE	3	/	0.2
DE	2	/	0.1

The mode entropy space is constructed by extracting four types of entropy to comprehensively characterize the signal properties, resulting in the formation of a high-dimensional feature dataset Q. The fusion of multiple entropy features enables a more comprehensive quantification of the nonlinear structure and dynamic variations of the original signal, which enhances sensitivity to differences in system states. The constructed feature matrix Q effectively captures the multidimensional information of the signal. However, its high dimensionality and nonlinear characteristics also increase the computational complexity and classification difficulty in subsequent fault diagnosis tasks. Thus, the application of dimensionality reduction methods is required to compress and optimize the extracted features.

$$Q=\left\{\begin{matrix} SE_1\cdots SE_m & FE_1\cdots FE_m & PE_1\cdots PE_m & DE_1\cdots DE_m \\ SE_2\cdots SE_m & FE_2\cdots FE_m & PE_2\cdots PE_m & DE_2\cdots DE_m \\ SE_3\cdots SE_m & FE_3\cdots FE_m & PE_3\cdots PE_m & DE_3\cdots DE_m \\ \vdots & \vdots & \vdots & \vdots \\ SE_n\cdots SE_m & FE_n\cdots FE_m & PE_n\cdots PE_m & DE_n\cdots DE_m \end{matrix}\right\}\in \mathbb{R}^{n\times 4m} \quad (5)$$

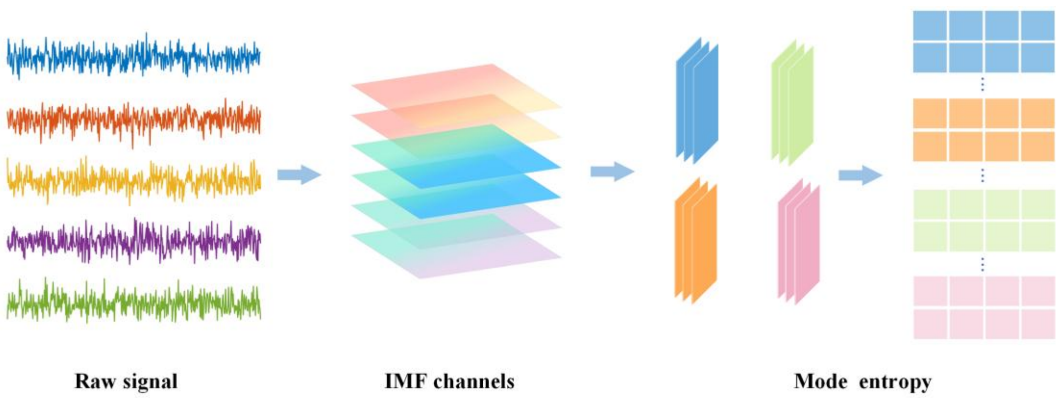


Fig. 1. Schematic diagram of mode entropy

To capture the potential nonlinear correlations among multiple entropy features, KPCA is employed to conduct nonlinear dimensionality reduction on the feature matrix $Q\in \mathbb{R}^{n\times 4m}$. The principle of constructing the mode entropy space is illustrated in Fig. 2. KPCA leverages kernel functions to implicitly project original data onto the high-dimensional feature space, where linear separability or improved principal component extraction can be achieved, thus enabling more effective extraction of the intrinsic structure among complex features.

The Gaussian radial basis function (RBF) is adopted as the kernel in view of its superior smoothness and global characteristics, which make it highly effective for capturing complex nonlinear relationships. The kernel enables the evaluation of the similarity between samples Q_i and Q_j within the implicit feature space and is formally defined as:

$$k(Q_i,Q_j)=\exp\left(-\frac{\| Q_i-Q_j\|^2}{2\sigma^2}\right) \tag{6}$$

where $\| Q_i-Q_j\|$ denotes the Euclidean distance between samples. σ represents the kernel bandwidth setting that determines both the sensitivity in similarity assessment and the function's smoothness.

Subsequently, the kernel matrix $K\in\mathbb{R}^{n\times n}$ is used to construct the kernel function, which is centralized to obtain the centralized kernel matrix $\tilde{K}$, on which eigenvalue decomposition is performed as follows:

$$\tilde{K}\alpha=\lambda\alpha \tag{7}$$

The top d=4 eigenvectors corresponding to the largest eigenvalues, $\alpha^{(1)},\alpha^{(2)},\alpha^{(3)},\alpha^{(4)}$, are selected as the projection bases. For each sample Q_i, its projection onto kernel principal component direction l is calculated as follows:

$$P_i^{(l)}=\sum_{j=1}^{n}\alpha_j^{(l)}\cdot k(Q_i,Q_j) \tag{8}$$

Finally, the projection results of all samples onto the first four principal component directions are combined to obtain the reduced-dimensional feature matrix:

$$P=\begin{bmatrix} SE_1' & FE_1' & PE_1' & DE_1' \\ SE_2' & FE_2' & PE_2' & DE_2' \\ SE_3' & FE_3' & PE_3' & DE_3' \\ \vdots & \vdots & \vdots & \vdots \\ SE_n' & FE_n' & PE_n' & DE_n' \end{bmatrix}\in\mathbb{R}^{n\times 4} \tag{9}$$

Through this process, the initial high-dimensional mode entropy features are projected into a lower-dimensional space with enhanced discriminative power and reduced redundancy, thereby effectively retaining key information pertinent to fault states.

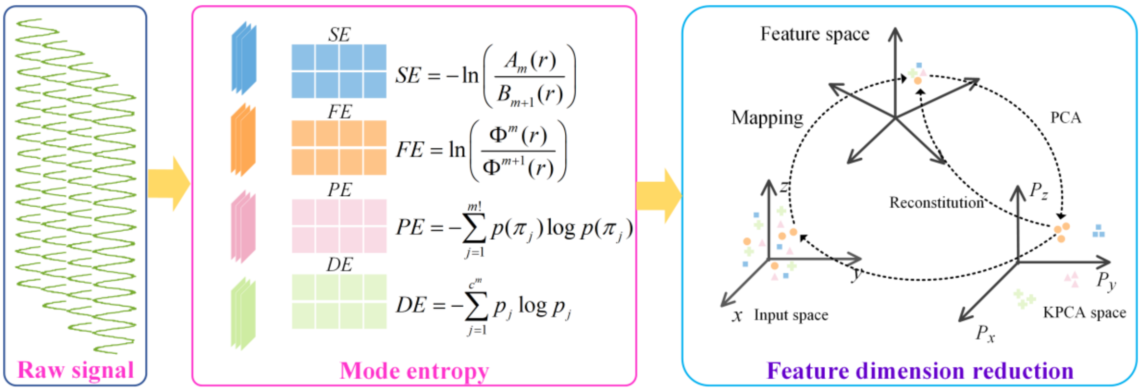


Fig. 2. Schematic diagram of the construction of mode entropy space

2.2. The proposed mode entropy knowledge machine

2.2.1. Architecture of the mode entropy knowledge machine

After the extraction and dimensionality reduction of mode entropy features, pattern recognition becomes an indispensable step. In this study, the proposed MEKM employs ELM for fault classification. Specifically, the dataset derived from the lower-dimensional mode entropy feature space is split into training and testing groups. The training group helps build the ELM model's classification structure, and the testing group assesses the model's ability to generalize and its accuracy in recognition. The basic architecture of ELM is illustrated in Fig. 3. On this basis, ELM is further combined with kernel functions by replacing the original hidden layer feature mapping with a kernel-based form, thereby strengthening the ability to capture high-dimensional and nonlinear data structures [56]. This integration improves data processing capabilities and further enhances classification accuracy.

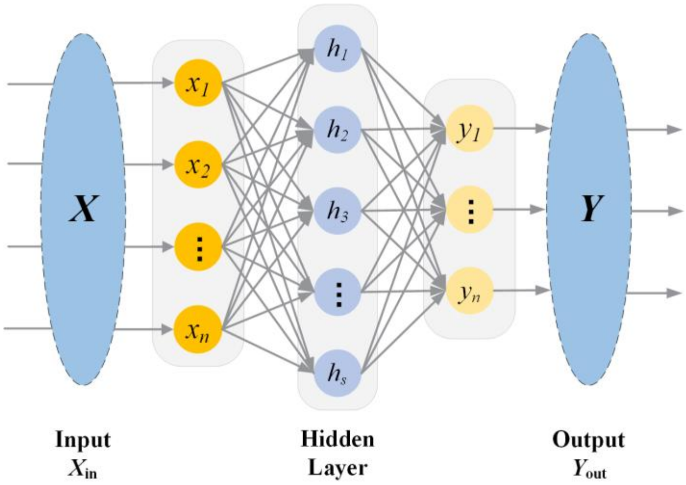


Fig. 3. Architecture of the ELM

In the ELM, the mapping between the input layer and the hidden layer is computed using the activation function $g(\cdot)$. The hidden layer features are subsequently mapped to the target output space via output weights. Let the input samples be denoted as $X_i \in \mathbb{R}^d$. The output matrix is as follows:

$$H(\omega, X_{\text{in}}, b) = \begin{bmatrix} g(\omega_1 \cdot X_1 + \beta_1) & g(\omega_2 \cdot X_1 + \beta_2) & \cdots & g(\omega_s \cdot X_1 + \beta_s) \\ g(\omega_1 \cdot X_2 + \beta_1) & g(\omega_2 \cdot X_2 + \beta_2) & \cdots & g(\omega_s \cdot X_2 + \beta_s) \\ \vdots & \vdots & \ddots & \vdots \\ g(\omega_1 \cdot X_n + \beta_1) & g(\omega_2 \cdot X_n + \beta_2) & \cdots & g(\omega_s \cdot X_n + \beta_s) \end{bmatrix} \tag{10}$$

where $\omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ indicates the weight vector linking the input and hidden layers, while $\beta = \{\beta_1, \beta_2, \cdots, \beta_n\}$ corresponds to the bias vector.

In the standard ELM, the output of the hidden layer is derived by applying the following activation function:

$$H_{ij} = g(\omega_j \cdot X_i + \beta_j), \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, s \tag{11}$$

The input data are mapped into a higher-dimensional feature space, thereby removing the requirement to explicitly calculate the weight and bias parameters. To avoid the explicit calculation of the hidden layer weights and biases, a kernel function is utilized to replace the hidden layer mapping. Let $k(X_i, X_j)$ is the kernel function evaluated on X_i and X_j. The kernel matrix is subsequently constructed as follows:

$$\Omega_{\text{ELM}} = \begin{bmatrix} k(X_1,X_1) & k(X_1,X_2) & \cdots & k(X_1,X_n) \\ k(X_2,X_1) & k(X_2,X_2) & \cdots & k(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(X_n,X_1) & k(X_n,X_2) & \cdots & k(X_n,X_n) \end{bmatrix} \tag{12}$$

The RBF is chosen as the kernel function. The output weight matrix, calculated from the hidden layer outputs, is given by:

$$\beta = (H^T H + \lambda I)^{-1} H^T T \tag{13}$$

where λ denotes the regularization parameter. I is the identity matrix. Upon introduction of the kernel matrix, this expression can be further reformulated as:

$$\beta = (\Omega_{\text{ELM}} + \frac{I}{C})^{-1} T \tag{14}$$

where $\frac{I}{C}$ represents the regularization term and C refers to the regularization coefficient.

The ultimate prediction outcome is as follows:

$$Y = K \cdot \left(\frac{I}{C} + \Omega_{\text{ELM}} \right)^{-1} T \tag{15}$$

Although the RBF kernel function enhance the nonlinear model capability of the model to some extent, its performance is largely dependent on the choice of kernel parameter σ and the regularization coefficient C. Moreover, this method struggles to adequately retain the discriminative weight information from input samples, which limits the model's representational capacity and generalization performance. Therefore, optimizing the kernel function along with its related parameters is essential.

2.2.2. Collaborative adaptive parameter optimization strategy

In the proposed MEKM, the model performance depends not only on the effectiveness of entropy-based feature extraction but also on the configuration of classifier parameters. To boost the generalization and precision in fault identification, a collaborative optimization paradigm is developed to achieve joint adaptive optimization of the mode entropy feature space and the parameters of the ELM.

The core idea of this strategy is to employ PSO to facilitate collaborative evolution at both the feature selection and parameter search levels. While maintaining the diversity of entropy features, this approach adaptively adjusts the key model parameters, thus enhancing the overall stability of the diagnosis framework. Compared with the conventional paradigm where feature extraction and classifier training are conducted independently, the proposed collaborative optimization strategy enables an optimization paradigm in which feature structures drive classification performance and classification feedback guides feature adjustment.

PSO is an adaptive search algorithm. The specific implementation is as follows: Assume a group of N particles exists within a search space. Each particle i possesses its own velocity $P_i = \left(V_{i1}, V_{i2}, \cdots, V_{iN}\right)$ and position $X_i = \left(X_{i1}, X_{i2}, \cdots, X_{iN}\right)$. Each particle iteratively updates its state guided by both P_i and G_i . The update equations for particle position and velocity in iteration $m+1$ are as follows:

$$X_i^{n+1} = X_i^n + V_i^n \tag{16}$$

$$V_i^{n+1} = \omega V_i^n + c_1 r_1 (P_i - X_i^n) + c_2 r_2 (G_i - X_i^n) \tag{17}$$

where m denotes the inertia weight, c_1 and c_2 represent learning coefficients, and r_1, r_2 are uniformly sampled random values in the range [0, 1].

The particle swarm optimization uses a population size of 20 and an iteration count of 50. The fitness function of every particle is determined by its classification accuracy on the test set. In each iteration, the personal best and global best positions are updated until the maximum number of iterations or the convergence criterion is reached. Throughout the entire collaborative optimization process, PSO enables adaptive regulation across two aspects:

(1) Feature selection feedback: based on the classifier's performance with different feature combinations, the weights of entropy features are dynamically optimized to enhance the model's capacity for filtering out redundancy.

(2) Classifier parameter adjustment: by searching for the optimal scale parameter of the RBF kernel and the regularization coefficient, the learning ability of the classifier for mode entropy features is improved.

This collaborative mechanism introduces classification performance feedback within a unified optimization framework, enabling the modal entropy feature space and classifier parameters to interact during the iterative process, thereby achieving structural alignment and performance coordination. Compared to traditional optimization methods, this mechanism effectively mitigates inconsistencies between feature representations and classification decisions, further enhancing the model's accuracy and robustness under various operating conditions.

3. Fault diagnosis method based on the proposed MEKM

Firstly, the raw signals from the test bearings are collected and normalized. Then, the signals are segmented at fixed intervals to obtain sample data. Next, the mode entropy feature space is constructed for feature extraction and dimensionality reduction, resulting in a low-dimensional feature dataset. Lastly, the MEKM is employed to diagnose fault types. The flowchart is depicted in Fig. 4, and the following steps are involved:

Step 1: Construct the modal entropy feature space. Decompose the acquired bearing vibration signals to extract various modal entropy features and construct a modal entropy space that fully characterizes the fault features in the signals.

Step 2: Perform noise-resistant feature dimensionality reduction using KPCA. Utilize KPCA to map high-dimensional modal entropy features to a high-dimensional feature space, extracting key nonlinear features while removing redundant information and noise interference.

Step 3: Establish a collaborative optimization mechanism. Use the PSO algorithm to simultaneously optimize the feature combinations resulting from KPCA dimensionality reduction and the parameters of the ELM classifier, thereby achieving a synergistic improvement in both feature representation and classification performance.

Step 4: Construct an adaptive ELM fault diagnosis model. Input the feature matrix optimized by PSO into the ELM classifier for training and testing, ultimately outputting the classes for faults.

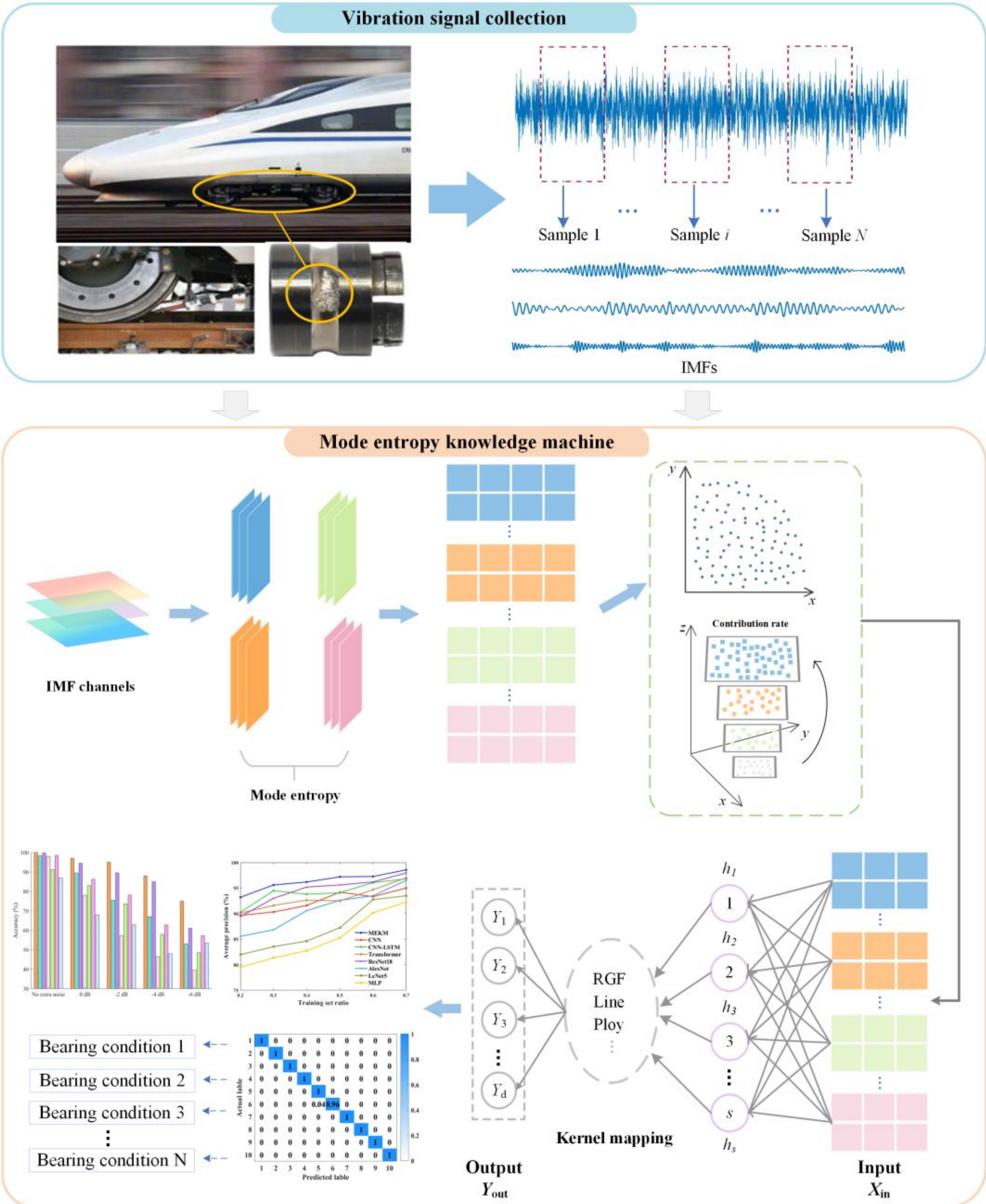


Fig. 4. Flowchart of the proposed fault diagnosis method

4. Experimental validation

4.1. Experimental validation under single operating conditions

4.1.1. Case I: CWRU bearings dataset

Case I is based on the CWRU bearings dataset [57]. The drive-end accelerator was used to generate drive-end bearing fault data at a rotational speed of 1750 rpm. According to the bearing status, the dataset includes normal conditions, inner ring fault, outer ring fault, and ball fault. The damage is classified under three types according to the diameter of the damage, resulting in ten different types of faults. Each bearing state corresponds to a raw vibration signal, as showed in Fig. 5. The distribution of sample data is presented in Table 2.

Table 2 Data details for Case I

Fault type	Abbreviation	Label	Sample size	Fault diameter (inches)
Normal condition	Nor	1	100	/
	IR1	2	100	
Inner ring fault	IR2	3	100	0.007
	IR3	4	100	
	OR1	5	100	
Outer ring fault	OR2	6	100	0.014
	OR3	7	100	
	B1	8	100	
Ball fault	B2	9	100	0.021
	B3	10	100	

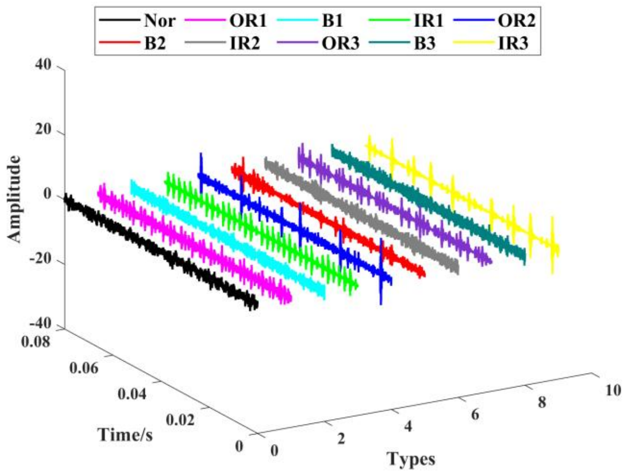


Fig. 5. The bearing vibration signal in Case I

To avoid analytical bias caused by large amplitude differences among the signals, all ten bearing signals with different fault types from Case I were first normalized. Subsequently, VMD was applied to the normalized signals, and the first three IMFs were extracted. The selection of the modal number K and the penalty factor α directly impacts the decomposition effectiveness. In [58], an optimization algorithm was employed to select K within the range [3,8] and α within [200,2000]. Multiple experiments were conducted using an iterative enumeration method. Through numerous experiments, the penalty factor to be 600 and the number of modes was determined to be 6, which effectively preserved the main features of bearing faults while suppressing the influence of noise. The extracted IMFs were then input into the MEKM for feature extraction, resulting in a high-dimensional feature dataset. Dimensionality reduction was further performed to obtain a low-dimensional dataset with the highest contribution ratios. The key parameters related to PSO, KPCA, and ELM have been consolidated. The specific parameter settings are shown in Table 3. To further evaluate the MEKM's ability to extract features, comparative experiments were conducted with CNN-LSTM, Transformer, and ResNet18. For all methods, the feature dimension was reduced to three, and the top three components were visualized by t-SNE, as depicted in Fig. 6. The visualization results indicate that MEKM exhibits significant advantages in distinguishing the ten fault states, with almost no overlap among different states, which demonstrates its effectiveness in fault feature extraction.

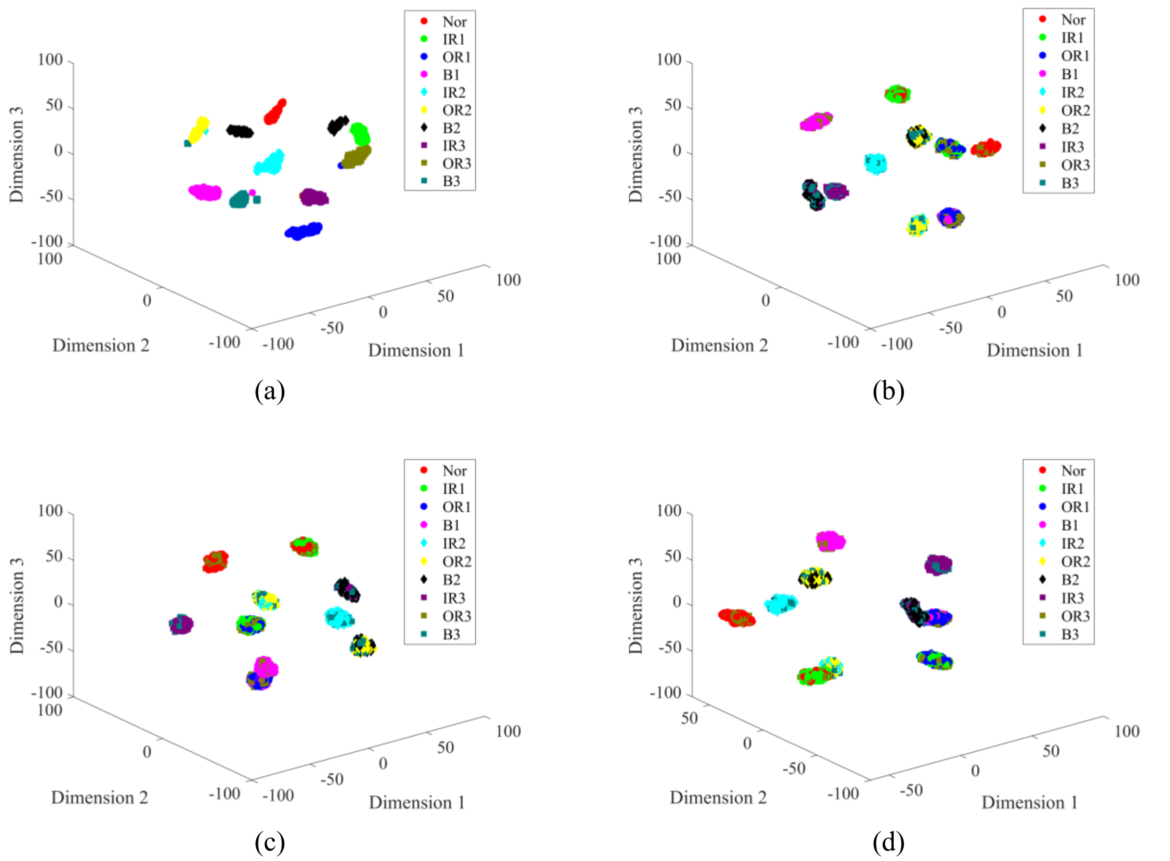


Fig. 6. Visualization comparisons of various methods: (a) MEKM; (b) CNN-LSTM; (c) Transformer; (d) ResNet18

Secondly, the MEKM is contrasted with several commonly used deep learning methods. The normalized data are input into MEKM and other networks for identification. Each bearing condition contains 100 samples, with a test ratio of 0.5. To ensure the fairness and reproducibility of the experimental results, all deep learning models under comparison were trained using the same data split and training strategy. Specifically, the maximum number of training epochs was set to 50, the batch size was set to 64, the Adam optimizer was used for all models, and the initial learning rate was uniformly set to 0.001. In terms of model architecture, LeNet-5 consists of two convolutional layers and three fully connected layers. The MLP features a three-layer fully connected structure. The CNN comprises three convolutional layers combined with pooling layers and a fully connected classifier. The CNN-LSTM consists of three one-dimensional convolutional layers and one LSTM layer, with 128 hidden units in the LSTM. ResNet18 consists of four residual stages, each containing two residual blocks, and uses global average pooling and a fully connected classification layer. AlexNet consists of a five-layer convolutional structure and two fully connected layers. The Transformer is built on an encoder architecture, comprising two encoder layers with an embedding dimension of 128 and a feedforward dimension of 256. As shown in the confusion matrix in Fig. 7, MEKM achieves the best performance in fault classification, with only one label incorrectly predicted, and its accuracy and stability are superior to those of the other methods. In comparison, although the classification results of the Transformer and ResNet18 models are similar to those of MEKM, both models exhibit a higher number of misclassifications and slightly lower stability. More severe misclassification issues and overall unstable performance are observed for CNN-LSTM and AlexNet. The CNN performs the worst, with a large number of incorrect label predictions. Overall, although deep learning networks achieve automatic feature extraction through convolutional layers and feature dimensionality reduction through pooling layers, which can directly process the data and save the time of manual feature engineering, in practical applications it is susceptible to a variety of interferences and has poor stability. The proposed MEKM performs better in terms of diagnostic performance.

Table 3 MEKM parameter settings

Module	Parameter	Value/Range
PSO	Population size	20
	Maximum iterations	50
	Inertia weight w	0.7
	Individual learning factor c_1	1.5
	Social learning factor c_2	1.7
	KPCA parameter search range	[1.1, 10]
	ELM parameter search range	[0.01, 10]
KPCA	Kernel type	RBF kernel
	Kernel parameter σ	1.4943
	Dimensionality reduction	3
ELM	Kernel type	RBF kernel
	Regularization coefficient C	2.3898
	Kernel parameter S	5.3172

To evaluate the performance of MEKM with a finite amount of training samples, experiments were conducted with different training-to-testing sample ratios. As shown by the heatmap in Fig. 8, all methods exhibit a noticeable decrease in diagnosis accuracy as the percentage of the test set increases. With the test set ratio exceeding 0.5, the diagnostic accuracy of MEKM remained above 98.5%, outperforming the other high performing models such as CNN, CNN-LSTM, Transformer, ResNet18, and AlexNet. In contrast, simpler networks such as LeNet5 and MLP demonstrate inferior classification performance. These results indicate that MEKM is able to maintain high accuracy even with fewer training samples, exhibiting better stability and robustness compared to other deep learning methods.

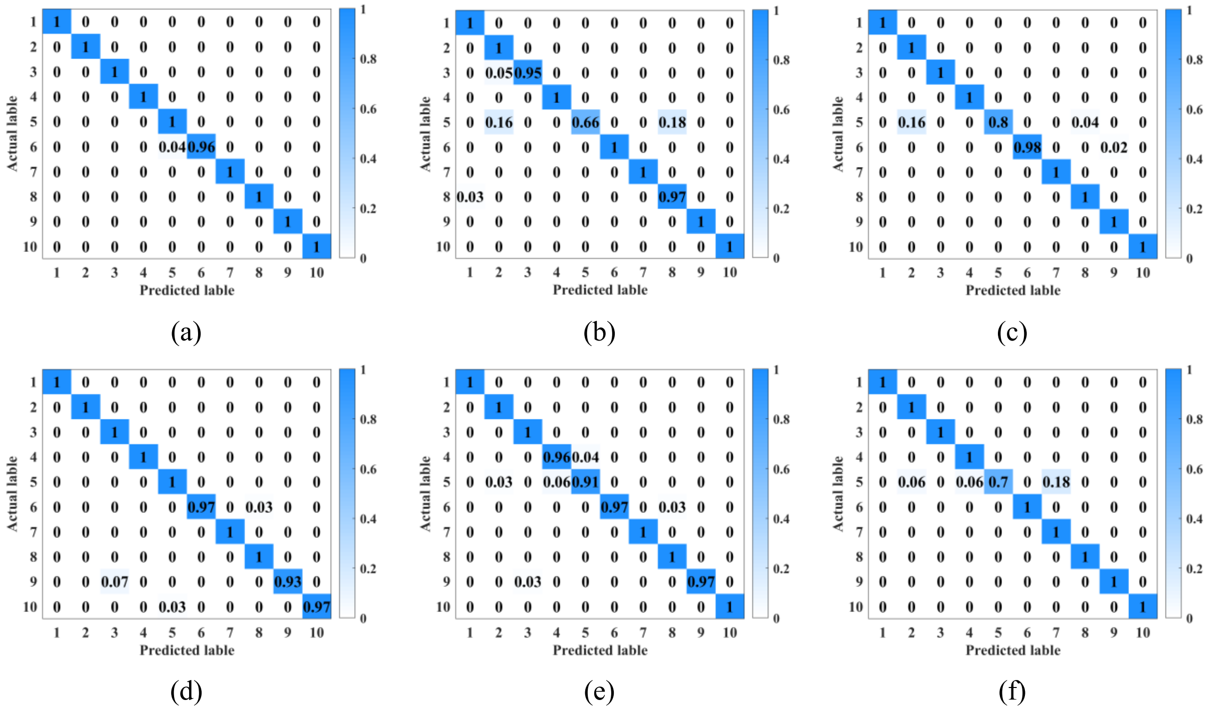


Fig. 7. Confusion matrices for Case I test set with a ratio of 0.5: (a) MEKM; (b) CNN; (c)CNN-LSTM; (d) Transformer; (e) ResNet18; (f) AlexNet

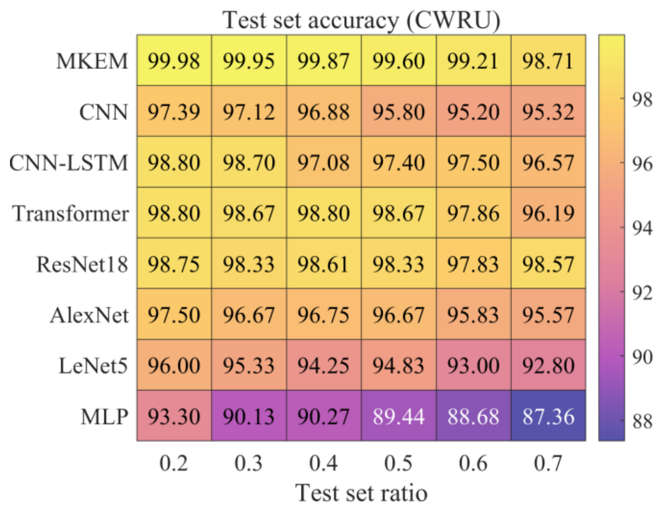


Fig. 8. Heatmap of classification accuracy for Case I at different test set ratios

In addition, three additional metrics are used for further analysis, including recall, precision, and F1-score. A total of 30 experiments were conducted, and the average values for these three metrics were calculated for each method. As shown in Table 4, the MEKM achieved scores above 99.5% for all three evaluation metrics, with minimal fluctuation, further demonstrating the excellent generalization capability of MEKM.

To assess the interference resistance of the MEKM, white noise with different signal-to-noise ratios (SNRs) was injected into the signals, and the specific SNR values were 0, -2, -4, and -6 dB. The mean diagnostic accuracy of each method was compared across various SNR conditions. The SNR, which quantifies the noise intensity, is defined below:

$$SNR = 10\log_{10}(\frac{P_1}{P_2}) \tag{18}$$

where P_1 and P_2 represent the power of the raw signal and noise.

Table 4 Average values of multiple evaluation metrics for various methods

Methods	Recall	Precision	F1- score
MEKM	99.72±0.22	99.82±0.13	99.63±0.11
CNN	94.24±0.25	93.64±0.63	93.31±0.52
CNN-LSTM	96.78±0.32	97.22±0.85	97.18±0.62
Transformer	98.81±0.31	98.78±0.26	98.83±0.34
ResNet18	97.37±0.16	96.84±0.13	97.42±0.16
AlexNet	96.29±0.43	95.54±0.23	96.34±0.18
LeNet5	94.60±0.63	96.05±0.52	94.34±0.23
MLP	87.65±0.26	88.20±0.63	88.63±0.54

In the experiments, 100 samples were used for each method, with a test ratio of 0.2. In Fig. 9, the accuracy of all methods declines significantly as the SNR decreases. Even after adding noise with an SNR of -6 dB, the diagnosis accuracy of MEKM remains above 98%, and no significant change in accuracy is observed as the SNR decreases further. In contrast, among other methods, Transformer and ResNet18, which exhibit better noise resistance, also show a significant decrease in accuracy under noise conditions with an SNR of -2 dB. The experimental results indicate that under strong noise interference, MEKM demonstrates excellent noise resistance and outperforms other methods. MEKM enhances fault-sensitive features through modal entropy space and uses PSO for collaborative optimization of KPCA and ELM, thereby achieving adaptive adjustment of feature extraction and classification parameters, which improves the model's generalization ability under complex operating conditions and in noisy environments.

In the above experiments, all datasets used were balanced. However, in practical applications, the training dataset is often imbalanced, particularly due to the relative paucity of fault samples compared to normal samples. To simulate the typical imbalance where healthy bearing samples greatly outnumber fault samples, this study adopts a positive-to-negative imbalance ratio for dividing the dataset during training. A smaller ratio indicates a higher degree of imbalance. To assess MEKM's effectiveness under imbalanced sample conditions, the CWRU dataset was categorized into normal and fault cases. For each fault type, the ratio between normal and faulty samples in the training set was kept consistent, and the count of test samples was set equal for all classes. The specific data partitioning is shown in Table 5. This data division enables further validation of the stability and performance of MEKM under positive-to-negative sample imbalance conditions.

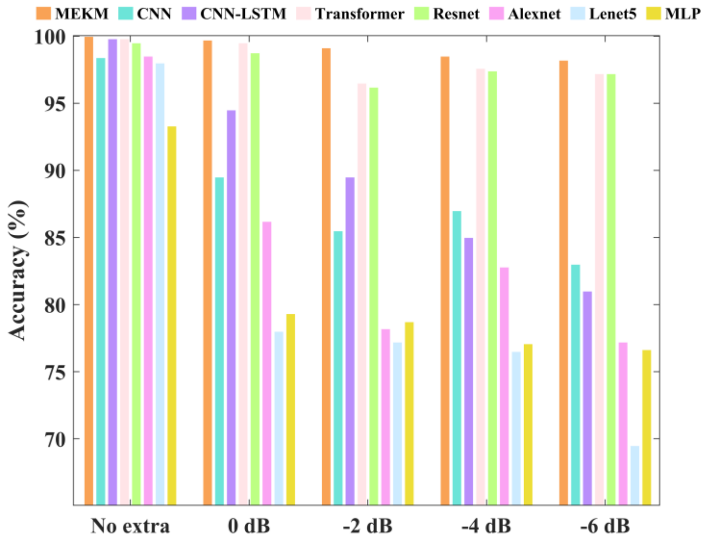


Fig. 9. Diagnosis accuracy under noise-free and noisy conditions

For imbalanced samples, using only accuracy cannot fully reflect the quality of the response method. Therefore, an imbalanced multi-class indicator, G-means, is further adopted to assess the model's classification effectiveness. In multi-class bearing fault diagnosis tasks, the number of samples across different fault classes may be uneven. G-means can comprehensively evaluate the ability to identify each class in a balanced manner, thereby preventing evaluation results from being unduly influenced by samples from the majority class. Specifically, the original samples are divided into two categories, G_i and G_j , and the G-mean (G_i, G_j) index is calculated for each pair. All computed G-means values are then summed with weights to obtain the overall evaluation metric G-means as follows:

$$G-mean=\sqrt{\frac{f_{ii}f_{jj}}{f_i f_j}} \tag{19}$$

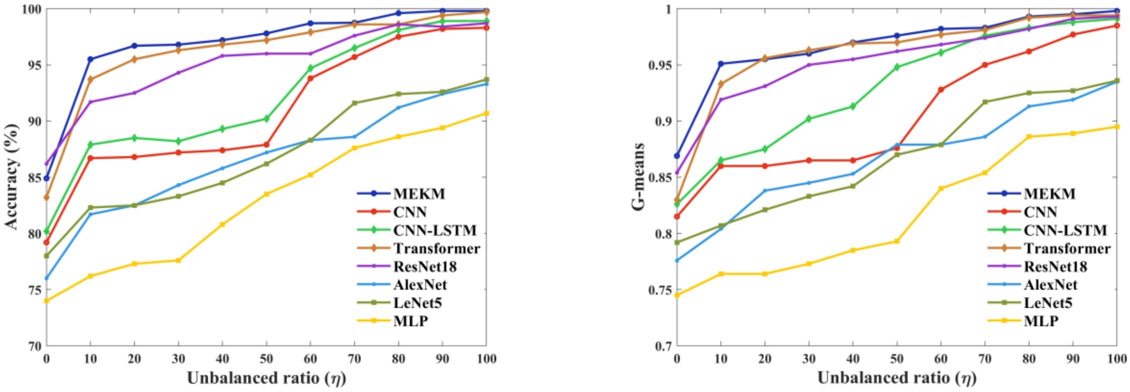
$$G-means=\frac{2}{k(k-1)}\sum_{0\le i<j} G-mean(G_i,G_j) \tag{20}$$

where, f_i and f_j represent the count of samples with true labels i and j. f_{ii} denotes the counts of samples correctly classified as i, and k represents the total class count.

The positive-to-negative imbalanced sample dataset was input into the MEKM and various deep learning networks. The test accuracy results are depicted in Fig. 10(a). With limited samples of faulty bearings, MEKM is still able to maintain high accuracy by training on samples from normal bearings, making it applicable to real industrial scenarios where faulty bearing samples are scarce. Compared to the other methods, MEKM can achieve 99% accuracy even when the imbalance ratio reaches 20%, and it maintains extremely high accuracy with minimal subsequent fluctuation. Fig. 10(b) illustrates that MEKM steadily achieves the top G-means score, reaching above 0.95 when the imbalance ratio is 10% and tending to stabilize thereafter. In comparison with other deep learning methods, MEKM demonstrates superior diagnosis stability when the number of faulty samples is limited.

Table 5 Information on the positive-to-negative imbalance of fault samples

η	Training dataset		Testing dataset	
	Fault	Normal	Fault	Normal
1%	1	100	100	100
10%	10	100	100	100
20%	20	100	100	100
40%	40	100	100	100
60%	60	100	100	100
100%	100	100	100	100



(a) (b)

Fig. 10. Test results of different methods under positive-to-negative imbalance conditions: (a) accuracy; (b) G-means

4.1.2. Case II: XJTU bearing dataset

Case II utilizes the Xi'an Jiaotong University (XJTU) dataset [59]. The experimental conditions are a motor speed of 2100 rpm and a radial load of 12 kN. Under these conditions, five types of bearing degradation data were collected, including three types of outer ring faults, a cage fault, and a combined inner and outer ring fault. The corresponding raw signals are shown in Fig. 11. This dataset comprehensively documents the degradation process of a bearing from normal operation to failure. Since fault features are weak and the signal-to-noise ratio is low during the early stages of degradation, using this data directly for analysis may introduce uncertainty; therefore, the latter half of the signal, where fault features are more pronounced, was selected for analysis. During this phase, all comparison methods were tested using the same data interval to ensure consistency in experimental conditions and comparability of results. Detailed data descriptions are provided for Table 6.

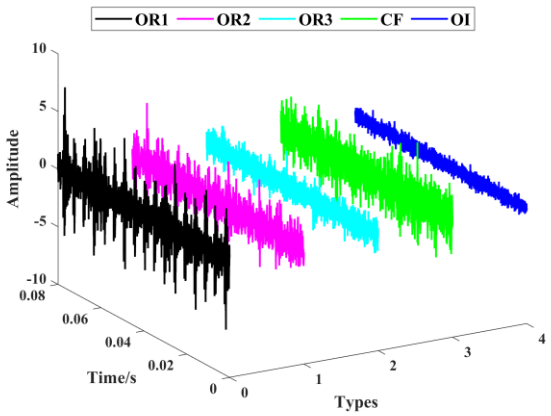


Fig. 11. The bearing vibration signal in Case II

Table 6 Data details for Case II

Fault type	Abbreviation	Label	Sample size
Outer ring fault	OR1	1	100
	OR2	2	100
	OR3	3	100
Cage fault	CF	4	100
Outer and Inner ring fault	OI	5	100

The five bearing signals were input into each model for fault diagnosis, with the test set ratio set to 0.5. Fig. 12 summarizes the results. From the confusion matrices, MEKM misclassified few samples and achieved the highest overall accuracy. Although MEKM achieved perfect classification for only one of the five fault types, the number of misclassifications for the remaining four types was relatively low. In contrast, the other five deep learning methods exhibited inferior overall performance.

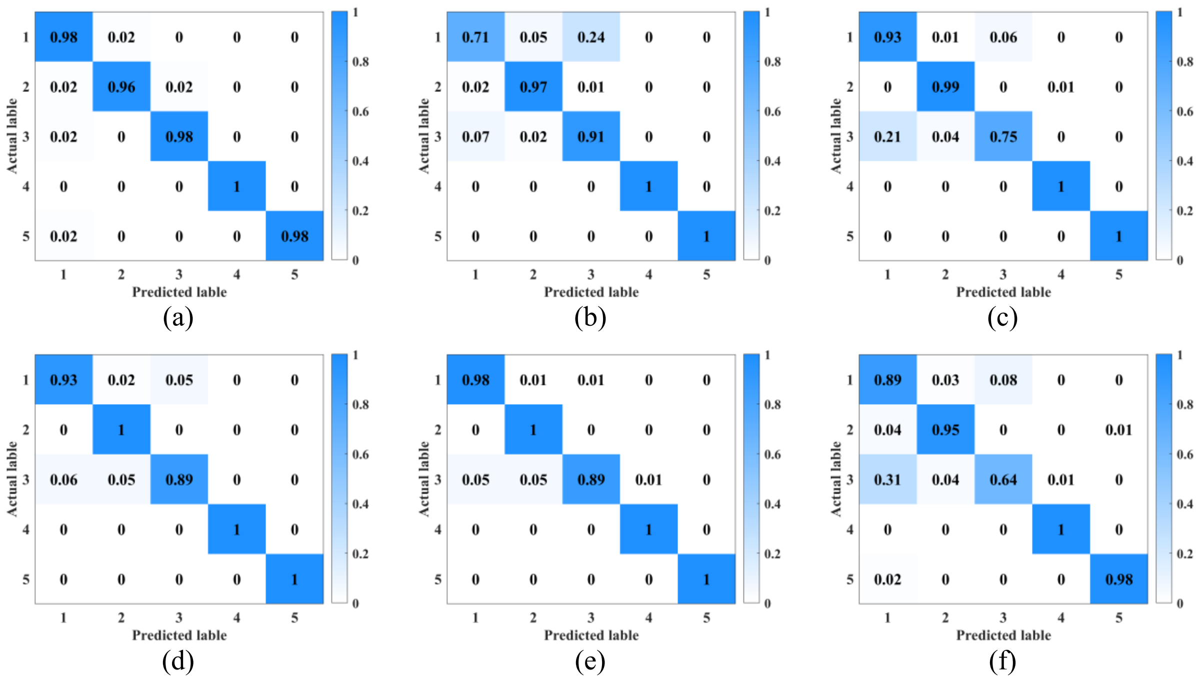


Fig. 12. Confusion matrices for the Case II test set with a ratio of 0.5: (a) MEKM; (b) CNN;

(c)CNN-LSTM; (d) Transformer; (e) ResNet18; (f) AlexNet

To further evaluate the effectiveness of MEKM, additional tests were performed to assess its accuracy under different test set ratios, specifically 0.2, 0.3, 0.4, 0.5, and 0.6. Fig. 13 displays the results for each method across various scenarios. The results indicate that MEKM achieved an accuracy of 97.2% when the test set ratio was 0.6, a result that notably surpasses those of the other approaches. Overall, MEKM demonstrated better performance and stability across various test set ratios.

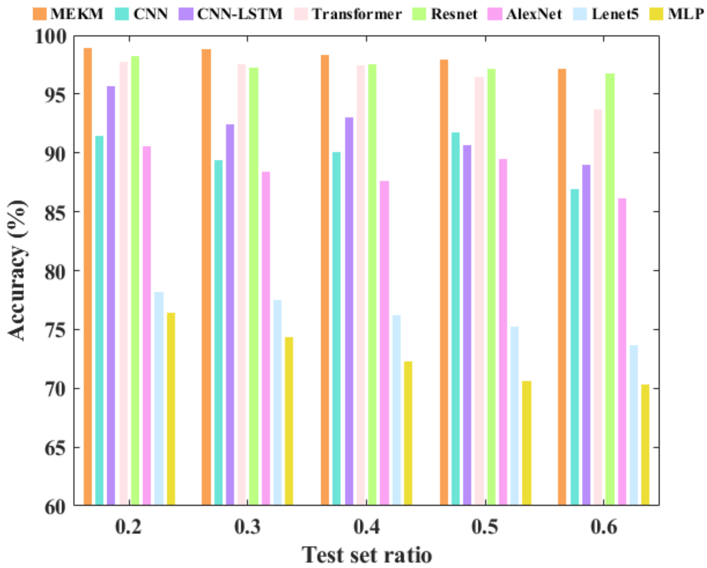


Fig. 13. Accuracy of different methods on the Case II under various test set ratios

4.1.3. Case III: BJTU bogie bearing dataset

Case III utilizes the Beijing Jiaotong University (BJTU) bogie bearing dataset [60]. The experimental system is illustrated in Fig. 14. The bogie transmission system primarily consists of a traction motor, drive gearbox, and axle box. The motor speed was set to 20 Hz, and the lateral load was 10 kN. Under this condition, five types of bearing data were collected, as detailed in Table 7. The test bearing is illustrated in Fig. 15. For each type, 200 samples were extracted, each with a length of 2048, and the training set ratios were set to 0.2, 0.3, 0.4, 0.5, and 0.6, respectively. The raw signals collected under the five types are illustrated in Fig. 16. Observations indicate that the time domain features across various fault types exhibit significant resemblance, which makes it challenging to differentiate them. Further processing is required for identification.

10.37965/jdmd.2026.1435

Fig. 14. The experimental system of the train bogie

Fig. 15. Test bearings

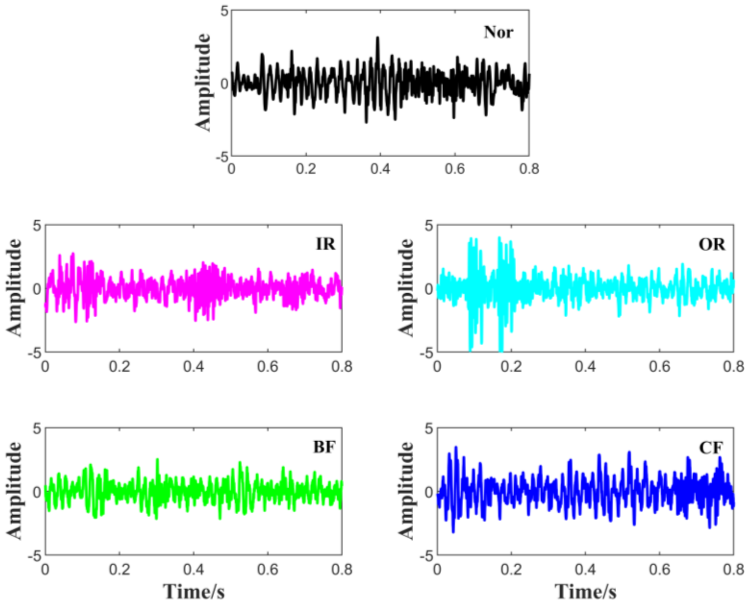


Fig. 16. The bearing vibration signal in Case III

The BJTU bearings dataset was used to compare MEKM with several other deep learning models in fault diagnosis. Each method's effectiveness was assessed using four metrics: precision, recall, specificity, and F1-score. Fig. 17 presents the average diagnosis outcomes for each approach. The results show that MEKM consistently outperforms advanced models such as Transformer and ResNet18 across all metrics, especially when the available data is limited. Additionally, MEKM's performance remains stable as the number of training samples increases, further validating its strong generalization ability.

Table 7 Data details for Case III

Fault type	Abbreviation	Sample size	Label
Normal condition	Nor	100	1
Inner ring fault	IR	100	2
Outer ring fault	OR	100	3
Ball fault	BF	100	4
Cage fault	CF	100	5

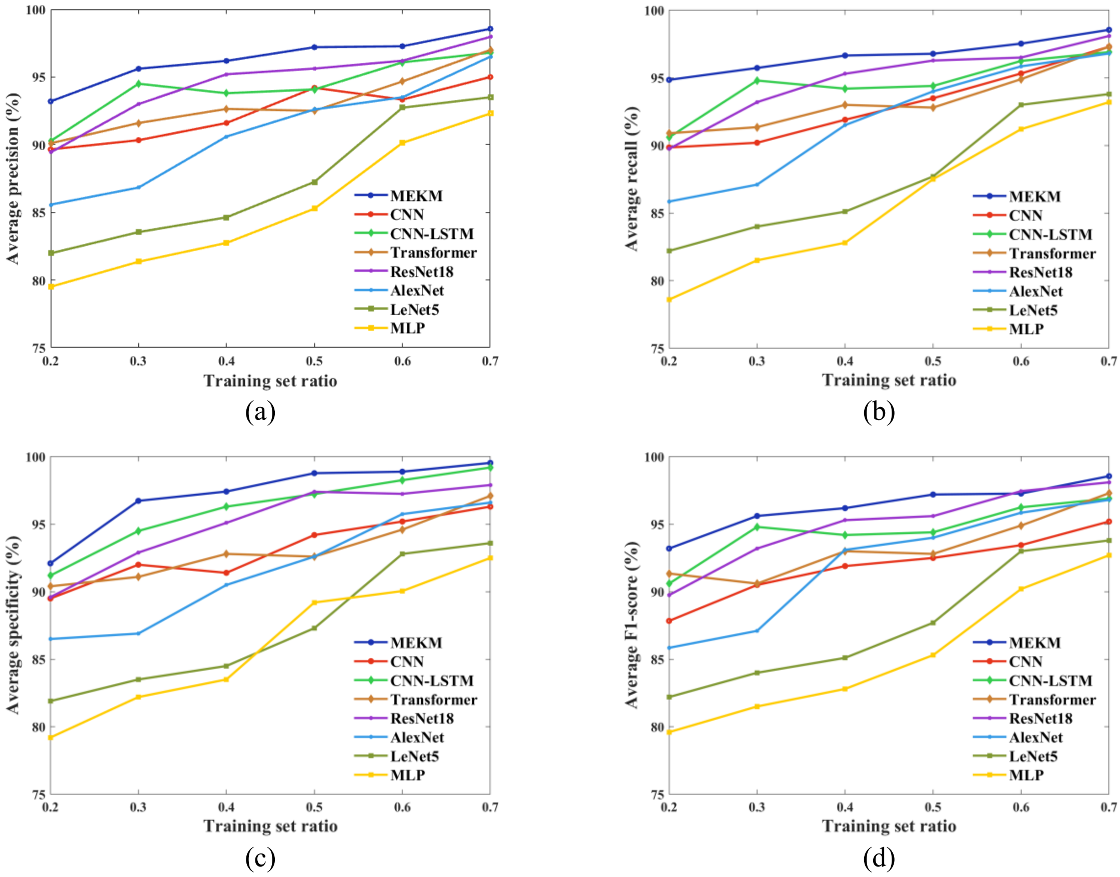


Fig. 17. Comparison of different methods in Case III based on four evaluation metrics: (a) accuracy; (b) recall; (c) specificity; (d) F1-score

4.2. Experimental validation under multiple operating conditions

In Section 4.1, experiments were carried out on three datasets at identical operating conditions, demonstrating the feasibility of MEKM in single-condition scenarios. However, in practical applications, bearings are often installed in different positions, resulting in varying speeds and loads during operation. Therefore, it is necessary to consider bearing faults under different working environments and conditions when performing fault diagnosis. To better evaluate and validate MEKM, this section employs the Case II to conduct multi-condition fault classification experiments. The specific conditions are summarized in Table 8. A comprehensive summary of the multi-condition data is shown in Table 9.

Table 8 Bearing speed and radial load under different operating conditions

Condition	Speed (r/min)	Radial load (kN)
A	2100	12
B	2250	11
C	2400	10

Table 9 Description of multi-condition data in Case II

Condition A		Condition B		Condition C	
Fault type	Label	Fault type	Label	Fault type	Label
Outer ring (123 min)	1	Inner ring (491min)	6	Outer ring (2538 min)	11
Outer ring (161min)	2	outer ring (161min)	7	Composite (2496 min)	12
Outer ring (158 min)	3	Cage (533 min)	8	Inner ring (371 min)	13
Cage (122 min)	4	Outer ring (42 min)	9	Inner ring (1515 min)	14
Inner and outer ring (52 min)	5	Outer ring (339 min)	10	Outer ring (114 min)	15

To further compare the performance differences between MEKM and mainstream deep learning methods under multiple operating conditions, the four best-performing methods were selected for testing. As shown in Table 10, there are differences in accuracy and stability among the methods. MEKM demonstrated both high accuracy and low standard deviation across repeated experiments, reflecting its excellent stability and robustness. Specifically, MEKM attained an average accuracy of 97.93%, reflecting highly reliable performance with minimal fluctuation. ResNet18, CNN-LSTM, and Transformer also exhibited strong performance, achieving high accuracies and low standard deviations. Notably, ResNet18 had an average accuracy of 97.29%, reflecting good stability. In contrast, LeNet5 and MLP exhibited lower accuracies and greater variability. In addition, Fig. 18 presents the confusion matrices of MEKM, CNN-LSTM, Transformer, and ResNet18 for the multi-condition experiments on Case II. The results show that MEKM has fewer misclassified labels, whereas the other three methods each exhibit more than two classes with a large number of misclassifications. These findings further confirm that MEKM possesses stronger generalization ability and robustness under complex multi-condition scenarios.

Table 10 Comparison of average accuracy and standard deviation for each method (%)

Methods	Number of experiments					Average accuracy	Standard deviation
	1	2	3	4	5		
MEKM	97.89	97.67	98.2	97.87	98.2	97.93	0. 36
CNN	92.11	91.78	90.78	92.44	92.67	90.16	0.811
CNN-LSTM	93.56	92.79	94.78	93.67	94.89	92.94	0.663
Transformer	96.67	97.78	95.67	97.11	96.89	96.82	0.393
ResNet18	97.33	96.56	96.45	97.44	96.75	97.29	0.151
AlexNet	91.58	91.41	90.72	92.67	91.68	91.21	0.349
LeNet5	67.01	62.78	67.23	64.56	62.4	62.28	4.318
MLP	59.84	60.47	61.54	63.50	61.26	61.28	1.325

10.37965/jdmd.2026.1435

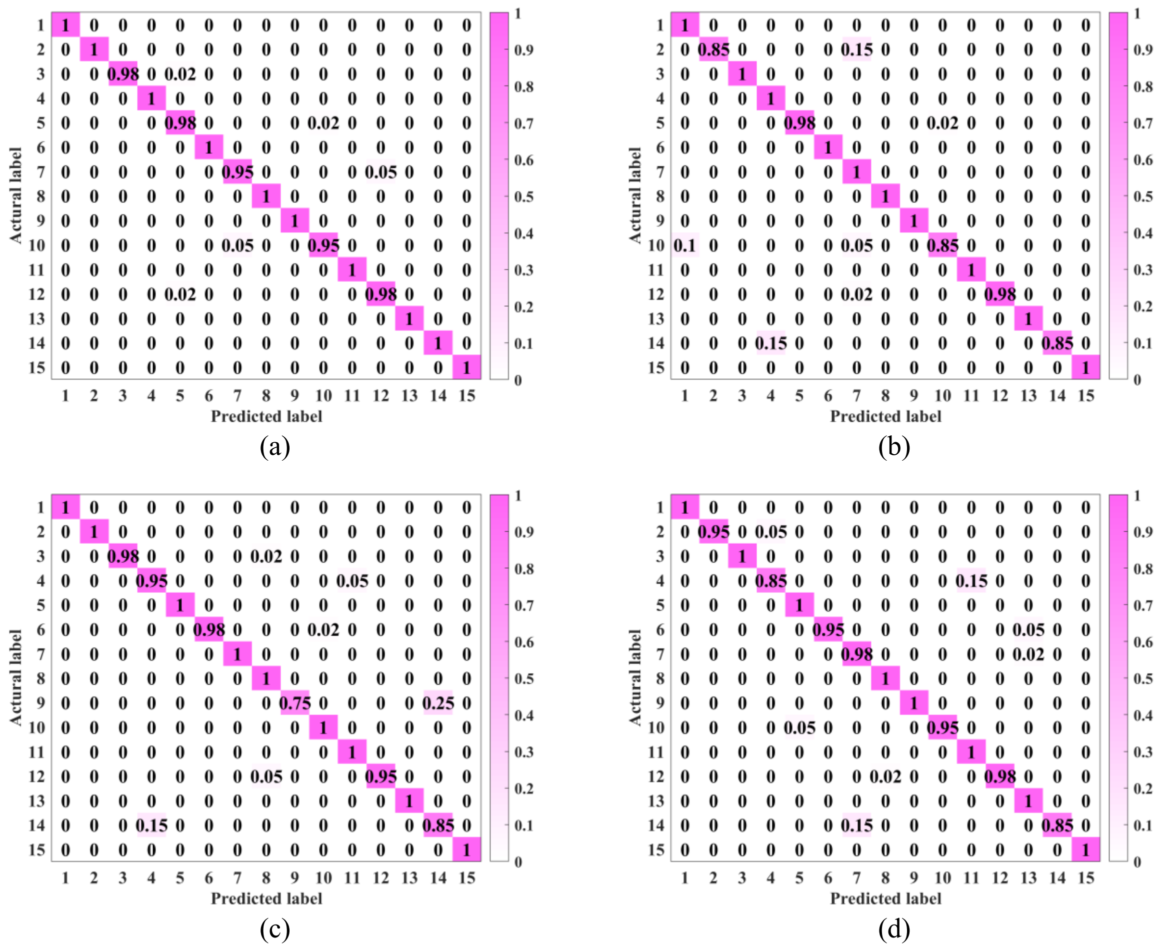


Fig. 18. Confusion matrices of different methods: (a) MEKM; (b) CNN-LSTM; (c) Transformer; (d) ResNet18

5. Conclusion

To improve the robustness and precision of fault diagnosis, this study introduces a novel method utilizing the MEKM. The method integrates multi-dimensional entropy features within the mode entropy space, effectively capturing the complex dynamic information of vibration signals. The performance of MEKM was validated through comparative experiments on multiple publicly available bearing datasets and multi-condition scenarios. Experimental findings confirm MEKM's superior performance in feature extraction and fault diagnosis relative to deep learning models such as CNN, CNN-LSTM, Transformer, ResNet18, AlexNet, LeNet5, and MLP. MEKM not only provides a more comprehensive representation of the complex dynamic characteristics of vibration signals but also effectively adapts to practical conditions, including high noise interference, limited samples, and severe class imbalance. Further analysis shows that MEKM maintains the highest diagnosis accuracy with minimal fluctuation and excellent robustness across different test set ratios and complex operating conditions. It also excels in various evaluation metrics, including recall, precision, and F1-score, showcasing its strong generalization capability. Overall, MEKM exhibits superior performance and enhanced robustness under various challenging conditions.

In future research, we will further validate the effectiveness of the proposed method in fault diagnosis tasks for rotating machinery other than bearings, such as gearboxes, planetary gear systems, and multi-component coupled systems under complex operating conditions. At the same time, we will focus on improving the computational efficiency and real-time performance of MEKM to enhance its applicability in practical online engineering diagnostics.

Acknowledgements

This research was supported by the Guangxi Natural Science Foundation (Grant No. 2026GXNSFBA00640230) and the Guangxi Science and Technology Program (Grant No. GuikeAD25069080). This paper is also supported by the National Natural Science Foundation of China (52275094).

Conflict of Interest Statement

The authors declare no conflicts of interest.

References

[1] B. Hou, Y. Wang, D. Wang, Investigations on Multiclass Classification Model-Based Optimized Weights Spectrum for Rotating Machinery Condition Monitoring, Journal of Dynamics, Monitoring and Diagnostics 4 (2025) 194–202.

[2] Y. Zhang, C. Shen, X. Ye, L. Chen, J. Shi, Z. Zhu, A Dual-Component Elastic Adaptation Network for Rotating Machinery Incremental Fault Diagnosis under Variable Operating Conditions, Journal of Dynamics, Monitoring and Diagnostics 5 (2026) 49–63.

[3] X. Li, B. Kang, J. Tang, Q. Li, T. Wang, M. Zhang, A Single-Device Environment-Adaptive Mixed Reality Framework for Real-Time Industrial Fault Diagnosis, Journal of Dynamics, Monitoring and Diagnostics 5 (2026) 64–73.

[4] J. Wang, X. Zhang, Z. Zhang, B. Han, X. Jiang, H. Bao, X. Jiang, Attention guided multi-wavelet adversarial network for cross domain fault diagnosis, Knowl-Based Syst, 284 (2024) 111285.

[5] J. Wang, Z. Zhang, Z. Liu, B. Han, H. Bao, S. Ji, Digital twin aided adversarial transfer learning method for domain adaptation fault diagnosis, Reliab Eng Syst Safe, 234 (2023) 109152.

[6] D. Hou, H. Qi, C. Wang, D. Han, High-speed train wheel set bearing fault diagnosis and prognostics: Fingerprint feature recognition method based on acoustic emission, Mech Syst Signal Pr, 171 (2022) 108947.

[7] H. Tan, S. Xie, D. Yang, J. Cheng, Z. Zhou, Health condition monitoring and diagnosis of rotating machinery based on Manhattan entropy, Measurement : Journal of the International Measurement Confederation, 227 (2024) 114243.

[8] G. Xu, D. Hou, H. Qi, L. Bo, High-speed train wheel set bearing fault diagnosis and prognostics: A new prognostic model based on extendable useful life, Mech Syst Signal Pr, 146 (2021) 107050.

[9] Y. Liu, J. Chen, T. Pan, J. Xie, A long-short-term feature extraction network based on soft-parameter-sharing for high-speed train bogies multi-object fault diagnosis under long-tailed distribution, Expert Syst Appl, 269 (2025) 126409.

[10] S. Zhao, L. Liang, G. Xu, J. Wang, W. Zhang, Quantitative diagnosis of a spall-like fault of a rolling element bearing by empirical mode decomposition and the approximate entropy method, Mech Syst Signal Pr, 40 (1) (2013) 154-177.

[11] Y. Hao, C. Zhang, Y. Lu, L. Zhang, Z. Lei, Z. Li, A novel autoencoder modeling method for intelligent assessment of bearing health based on Short-Time Fourier Transform and ensemble strategy, Precision Engineering, 85 (2024) 89-101.

[12] B. Pang, Y. Zhao, C. Yu, Z. Hao, Z. Sun, Z. Xu, P. Li, Empirical variational mode extraction and its application in bearing fault diagnosis, Appl Acoust, 228 (2025) 110349.

[13] C. Yin, Y. Wang, G. Ma, Y. Wang, Y. Sun, Y. He, Weak fault feature extraction of rolling bearings based on improved ensemble noise-reconstructed EMD and adaptive threshold denoising, Mech Syst Signal Pr, 171 (2022) 108834.

[14] B. Sun, H. Li, J. Wang, S. Lv, Z. Ma, Differential time-frequency mode decomposition and its application in rotating machinery fault diagnosis, Mech Syst Signal Pr, 229 (2025) 112571.

[15] S. Gao, T. Li, Y. Zhang, Z. Pei, Fault diagnosis method of rolling bearings based on adaptive modified CEEMD and 1DCNN model, Isa T, 140 (2023) 309-330.

[16] Y. Tian, J. Ma, C. Lu, Z. Wang, Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine, Mech Mach Theory, 90 (2015) 175-186.

[17] H. Tan, S. Xie, R. Liu, J. Cheng, K. Jing, Fatigue condition diagnosis of rolling bearing based on normalized balanced multiscale sample entropy, Int J Fatigue, 172 (2023) 107642.

[18] S.M. Pincus, Approximate entropy as a measure of system complexity, P Natl Acad Sci Usa, 88 (6) (1991) 2297-2301.

[19] D. Kong, H. Xie, Use of modified sample entropy measurement to classify ventricular tachycardia and fibrillation, Measurement, 44 (4) (2011) 653-662.

[20] C. Ma, X. Wang, Y. Li, Z. Cai, Fault causes identification of rotating machinery based on multiphase zoom permutation entropy, Measurement, 225 (2024) 114028.

[21] W. Chen, J. Zhuang, W. Yu, Z. Wang, Measuring complexity using FuzzyEn, ApEn, and SampEn, Med Eng Phys, 31 (1) (2009) 61-68.

[22] C. Zhou, L. Xing, W. Zou, Z. Wang, Y. Yang, S. Zhao, H. Bai, Y. Yang, EMGGDFuRDE: Enhancive multiscale generalized Gaussian distribution fuzzy Renyi dispersion entropy and its application in wind turbines fault diagnosis, Mech Syst Signal Pr, 231 (2025) 112636.

[23] S. Zhao, L. Liang, G. Xu, J. Wang, W. Zhang, Quantitative diagnosis of a spall-like fault of a rolling element bearing by empirical mode decomposition and the approximate entropy method, Mech Syst Signal Pr, 40 (1) (2013) 154-177.

[24] Y. Shang, G. Lu, Y. Kang, Z. Zhou, B. Duan, C. Zhang, A multi-fault diagnosis method based on modified Sample Entropy for lithium-ion battery strings, J Power Sources, 446 (2020) 227275.

[25] J. Zheng, J. Cheng, Y. Yang, A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy, Mech Mach Theory, 70 (2013) 441-453.

[26] S. Qin, Y. Zhang, K. Sun, F. Chen, Aileron fault detection with dynamic resampling based on fuzzy entropy and multi-branch neural network, Measurement, 241 (2025) 115773.

[27] Y. Tian, Z. Wang, C. Lu, Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping, Mech Syst Signal Pr, 114 (2019) 658-673.

[28] C. Zhou, J. Li, Q. Zhou, J. Liu, H. Li, Y. Li, T. Tang, Y. Sun, A fault diagnosis approach using improved multiscale fluctuation Rényi dispersion entropy and optimized binary tree LSTSVM, Expert Syst Appl, 283 (2025) 127820.

[29] Y. Su, L. Shi, K. Zhou, G. Bai, Z. Wang, Knowledge-informed deep networks for robust fault diagnosis of rolling bearings, Reliab Eng Syst Safe, 244 (2024) 109863.

[30] S. Ma, G. Cheng, Y. Li, Y. Huang, D. Zhuang, Research on multi-granularity imbalanced knowledge condition monitoring for mechanical equipment based on hierarchical ELM in multi-entropy space, Expert Syst Appl, 238 (2024) 121817.

[31] S.A. Hussain Basha, P. R. Kshirsagar, P. Srinivasa Rao, T.K. Tak, D.B. Sivaneasan, PCA-F-SHCNNet: Principal Component Analysis-Fused-Shepard Convolutional Neural Networks for lung cancer detection and severity level classification, Biomed Signal Proces, 107 (2025) 107843.

[32] E. Akbari, M. Samady Shadlu, An intelligent ANFIS-based fault detection, classification, and location model for VSC-HVDC systems based on hybrid PCA-DWT signal processing technique, Computers and Electrical Engineering, 120 (2024) 109763.

[33] Nonlinear Dimensionality Reduction by Locally Linear Embedding.

[34] M.Z.H. Shah, L. Hu, Z. Ahmed, Modified LPP based on Riemannian metric for feature extraction and fault detection, Measurement, 193 (2022) 110923.

[35] K. Zhao, J. Xiao, C. Li, Z. Xu, M. Yue, Fault diagnosis of rolling bearing using CNN and PCA fractal based feature extraction, Measurement, 223 (2023) 113754.

[36] Q. Liu, D. Kong, S.J. Qin, Q. Xu, Map-Reduce Decentralized PCA for Big Data Monitoring and Diagnosis of Faults in High-Speed Train Bearings, Ifac-Papersonline, 51 (18) (2018) 144-149.

[37] Y. Liu, Z. Yu, M. Zeng, Y. Zhang, LLE for submersible plunger pump fault diagnosis via joint wavelet and SVD approach, Neurocomputing, 185 (2016) 202-211.

[38] N. Lu, M. Li, G. Zhang, R. Li, T. Zhou, C. Su, Fault feature extraction method for rotating machinery based on a CEEMDAN-LPP algorithm and synthetic maximum index, Measurement, 189 (2022) 110636.

[39] J. Liu, Y. Song, X. Yu, Risk assessment study of hydrogen energy storage system based on KPCA-TSO-LSSVM, Int J Hydrogen Energ, 79 (2024) 931-942.

[40] J. Shen, F. Xu, Method of fault feature selection and fusion based on poll mode and optimized weighted KPCA for bearings, Measurement, 194 (2022) 110950.

[41] P. Liu, S. Zhao, L. Kang, Y. Yin, CNN Intelligent diagnosis method for bearing incipient faint faults based on adaptive stochastic resonance-wave peak cross correlation sliding sampling, Digit Signal Process, 156 (2025) 104871.

[42] S. Zhao, Y. Duan, N. Roy, B. Zhang, A deep learning methodology based on adaptive multiscale CNN and enhanced highway LSTM for industrial process fault diagnosis, Reliab Eng Syst Safe, 249 (2024) 110208.

[43] K. Liu, Y. Li, Z. Cui, G. Qi, B. Wang, Adaptive frequency attention-based interpretable Transformer network for few-shot fault diagnosis of rolling bearings, Reliab Eng Syst Safe, 263 (2025) 111271.

[44] P. Liang, W. Wang, X. Yuan, S. Liu, L. Zhang, Y. Cheng, Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment, Eng Appl Artif Intel, 115 (2022) 105269.

[45] J. Shen, Z. Wang, Y. Wang, H. Zhu, L. Zhang, Y. Tang, AGWO-PSO-VMD-TEFCG-AlexNet bearing fault diagnosis method under strong noise, Measurement : Journal of the International Measurement Confederation, 242 (2025) 116259.

[46] J. Zhong, Y. Zheng, C. Ruan, L. Chen, X. Bao, L. Lyu, M-IPISincNet: An explainable multi-source physics-informed neural network based on improved SincNet for rolling bearings fault diagnosis, Inform Fusion, 115 (2025) 102761.

[47] S. Xie, Y. Li, H. Tan, R. Liu, F. Zhang, Multi-scale and multi-layer perceptron hybrid method for bearings fault diagnosis, Int J Mech Sci, 235 (2022) 107708.

[48] D. Huang, X. Su, J. Yang, S. Du, D. Wang, Q. Ran, An improved dual-channel CNN-BILSTM fusion attention model for fault diagnosis of aero-engine bearings, Measurement, 253 (2025) 117761.

[49] R. Luo, Y. Li, H. Guo, Q. Wang, X. Wang, Cross-operating-condition fault diagnosis of a small module reactor based on CNN-LSTM transfer learning with limited data, Energy, 313 (2024) 133901.

[50] X. Zhang, X. Zhang, W. Liang, F. He, Research on rolling bearing fault diagnosis based on parallel depthwise separable ResNet neural network with attention mechanism, Expert Syst Appl, 286 (2025) 128105.

[51] V. Sinitsin, O. Ibryaeva, V. Sakovskaya, V. Eremeeva, Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model, Mech Syst Signal Pr, 180 (2022) 109454.

[52] T. Hu, H. Ma, D. Duan, W. Ge, Novel GIL mechanical fault diagnosis method based on multi-sensor data feature fusion and TDEAVOA-ELM, Computers and Electrical Engineering, 119 (2024) 109573.

[53] L. Deng, C. Zhao, X. Yan, Y. Zhang, R. Qiu, A novel approach for bearing fault diagnosis in complex environments using PSO-CWT and SA-FPN, Measurement, 249 (2025) 117027.

[54] Y. Li, Q. Ding, Geometric entropy: A novel nonlinear measure for bearing fault diagnosis, Adv Eng Inform, 69 (2026) 104031.

[55] T. Wang, S.Y. Khoo, Z.C. Ong, P.Y. Siow, T. Wang, Distance similarity entropy: A sensitive nonlinear feature extraction method for rolling bearing fault diagnosis, Reliab Eng Syst Safe, 255 (2025) 110643.

[56] J. Gong, X. Yang, H. Wang, J. Shen, W. Liu, F. Zhou, Coordinated method fusing improved bubble entropy and artificial Gorilla Troops Optimizer optimized KELM for rolling bearing fault diagnosis, Appl Acoust, 195 (2022) 108844.

[57] Case Western Reserve University Bearing Data Center 2022 Website http:// csegroups.case.edu/bearingdatacenter/home.

[58] S. Wang, Z. Feng, Multi-sensor fusion rolling bearing intelligent fault diagnosis based on VMD and ultra-lightweight GoogLeNet in industrial environments, Digit Signal Process, 145 (2024) 104306.

[59] B. Wang, Y. Lei, N. Li, N. Li, A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings, Ieee T Reliab, 69 (1) (2020) 401-412.

[60] A. Ding, Y. Qin, B. Wang, L. Guo, L. Jia, X. Cheng, Evolvable graph neural network for system-level incremental fault diagnosis of train transmission systems, Mech Syst Signal Pr, 210 (2024) 111175.