Journal of Industrial and Management Optimization

JIMO, 22(6): 2988–3022.
DOI: 10.3934/jimo.2026110
Received: 26 January 2026
Revised: 26 April 2026
Accepted: 15 May 2026
Published: 27 May 2026

https://www.aimspress.com/journal/jimo

Research article

Research on bearing fault diagnosis based on multi-scale adaptive residual network combined with three-domain attention

Xiaobo Huang[1], Zhansi Jiang[1,3], Yulong Tan[1], Peixin Chen[1] and Hui Jiang[2,*]

[1] School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, 541000, China
[2] School of Future Transportation, Guangzhou Maritime University, Guangzhou, 510725, China
[3] School of Intelligent Manufacturing, Guangzhou Maritime University, Guangzhou, 510725, China

* **Correspondence:** Email: jianghui@gzmtu.edu.cn.

Abstract: To address the challenges of weak fault feature coupling, overlapping mixed faults, and low diagnosis accuracy for rolling bearings in strong noise environments, this paper presents a novel fault diagnosis method named multi-scale adaptive residual network with a three-domain attention mechanism (MSADC-TDA). The method enables robust end-to-end diagnosis under both single and mixed fault conditions. First, variational mode decomposition (VMD) is applied to adaptively denoise the vibration signal. Then, continuous wavelet transform (CWT) is employed to convert the one-dimensional signal into a two-dimensional time-frequency image for comprehensive fault feature extraction. An MSADC layer is proposed by combining multi-scale convolution and attention mechanisms. To learn the features of time-frequency images, MSADC can dynamically adjust the weights of convolutional layers at different scales. In addition, a TDA module is constructed to adaptively focus on extracting fault features. To enhance the feature learning capability of the proposed method, residual blocks are constructed using MSADC and residual connections. Multiple residual blocks are combined, and dropout layers are utilized to prevent overfitting. Experimental results on the self-built bearing dataset and Case Western Reserve University (CWRU) bearing dataset demonstrate that the proposed model maintains high diagnostic accuracy under strong noise and mixed fault interference, and possesses excellent generalization ability.

Keywords: multi-scale adaptive residual network; three-domain attention; fault diagnosis; deep learning; rolling bearings
Mathematics Subject Classification: 68T07

1. Introduction

Rolling bearings are key fundamental components of core industrial equipment including rotating machinery, wind power installations, rail transit systems, and aerospace equipment. Their operating conditions directly govern the safety, reliability, and operational efficiency of the entire assembly. However, when subjected to demanding working environments—such as those involving high pressure, elevated temperatures, and rapid rotational speeds—rolling bearings are susceptible to diverse failure modes. Such faults not only compromise the operational efficiency of machinery but also pose potential safety hazards, risking both equipment integrity and personnel safety [1]. Therefore, developing a high-precision, end-to-end bearing fault diagnosis method that can adapt to complex working conditions with strong noise, mixed faults, and feature coupling is of great engineering value for ensuring the safe and stable operation of industrial equipment and reducing maintenance costs. Feature extraction and fault classification are key aspects of bearing fault diagnosis, and consequently, fault diagnosis technologies based on signal processing have received extensive attention.

Mixed fault refers to a multi-fault mode in rotating machinery characterized by simultaneous occurrence and mutual coupling of multiple faults. Mixed fault feature coupling refers to the superposition and modal mixing of impact signals from different faults in a mixed fault scenario. In view of the complex structure and internal correlation of mechanical equipment, actual engineering operation shows that rotating machinery faults mostly occur in mixed forms. Compound faults may trigger structural damage with diverse types and severities, greatly increasing the difficulty of accurate fault detection. Moreover, multiple fault features are coupled and mutually interfered. The resulting interference effects and mode mixing will obscure effective fault characteristic information, making it difficult to accurately identify and separate independent fault patterns [2–4]. Therefore, it is critical to develop signal processing methods that can effectively identify and extract the characteristics of mixed faults. The traditional mainstream single fault feature extraction methods, such as Fourier transform [5], envelope demodulation analysis [6], variational mode decomposition [7], empirical mode decomposition [8], wavelet transform [9], and so on, although effective in single fault diagnosis, rely more on professional knowledge and have limitations when dealing with mixed faults.

Wavelet transform has seen extensive application in fault diagnosis over the past decades due to its efficacy in signal processing. Representative efforts include an optimized wavelet packet transform for extracting discriminative fault features [10] and an adaptive redundant multiwavelet for mixed-fault identification [11]. A primary drawback of these methods, however, is their heavy reliance on expert knowledge, which hinders end-to-end diagnostic automation. Furthermore, the feature extraction capability of traditional wavelet transform faces greater challenges when dealing with mixed bearing faults. The coupling and superposition of multiple fault features in mixed faults lead to mutual overlap and interference in their time-frequency representations. Traditional wavelet transform usually relies on a single predefined wavelet basis function and fixed scale parameters. Such a fixed analysis paradigm cannot adaptively decouple and separate these complex coupled features, thus restricting the clear extraction of independent fault features from mixed signals. Consequently, its performance heavily depends on the operator's prior knowledge of the matching relationship between fault features and wavelet bases, which further increases the reliance on expert experience.

As a prominent architecture in deep learning, the convolutional neural network (CNN) has been extensively applied to fault diagnosis, leading to the development of numerous advanced models [12]. Liang et al. [13] proposed an attention-enhanced separable residual network, which reduces the number of parameters through deep separable convolutions and integrates the convolutional block attention module (CBAM) attention mechanism to improve the performance of the network by adaptively focusing on fault feature regions. Yan et al. [14] developed a CNN model incorporated with a CBAM. This architecture, characterized by its depth and wide convolutional kernels, was designed to more efficiently capture and perceive multi-peak features from various signals. Han et al. [15] proposed an automated framework that integrates feature selection, oversampling, ensemble learning, and genetic algorithm optimization to address the class imbalance problem in the multi-fault diagnosis of planetary gear carrier sets. This approach significantly improves diagnostic accuracy and reduces the risk of misclassification. Cui et al. [16] proposed a fusion model of Senet and bidirectional temporal convolutional network (BITCN) that significantly improves the timing-dependent capture ability and cross-dataset adaptability of non-stationary signals. Hong et al. [17] designed a novel one-dimensional CNN strategy to enhance the tracking performance by extracting the frequency-phase features of vibration signals while generating customized reference signals combined with system dynamic characteristics, which can improve the tracking performance, and indeed, the robustness of the tracking process was significantly improved. However, these methods based on convolutional neural networks only utilize single-scale convolutional kernels and single-domain attention mechanisms, which are unable to effectively extract the multi-scale and multi-dimensional fault features hidden in mixed fault bearings under strong noise environments. Yan et al. [18] combined the advantages of residual network (ResNet) and CNN in deep feature learning to design a new convolutional residual neural network architecture integrating multiple attention mechanisms. As an enhanced architecture of CNN, residual convolutional network (RESCNN) effectively alleviates gradient attenuation and reduces the difficulty of deep model optimization through its structural design. However, its identity mapping mechanism may induce a gradient explosion under certain circumstances, and then interfere with the convergence stability of the model. Although residual connections alleviate the vanishing gradient problem, the fixed convolutional scale and single-domain attention mechanism still limit the model's performance under strong noise and mixed fault conditions. By integrating a self-attention mechanism into a CNN framework, Wei et al. [19] proposed an approach that utilized a one-dimensional window-based representation and incorporated a local inductive bias to collectively enhance the model's robustness. Although CNN has been proven effective and widely applied in bearing fault diagnosis, the aforementioned methods all rely on single-scale feature learning schemes, which makes them unable to effectively capture the multi-scale characteristics inherent in vibration signals and thus hinders the improvement of diagnostic performance.

In practical operation, bearings often encounter interference from varying working conditions and environmental noise, causing their vibration signals to exhibit significant broadband characteristics, with their complex patterns distributed across multiple time scales. In response, some scholars have begun to adopt multi-scale learning strategies to effectively extract fault features at different scales, thus enhancing the feature recognition and anti-interference capabilities of neural networks. Xu et al. [20] proposed an attention-based multi-scale noise reduction residual convolutional neural network (AM-DRCN), which, through the lightweight module design, addresses the challenge of effectively extracting fault features from rotating machinery under strong noise interference, and the diagnosis accuracy and model robustness under complex working conditions are significantly improved.

However, this method employs a fixed multi-scale fusion mechanism and lacks the capability to adaptively adjust the weights of each scale, rendering it inadequate for fault diagnosis tasks under mixed fault conditions. Wang et al. [21] proposed a fault diagnosis framework that integrates a multiscale wavelet convolutional residual network (DSWTC-Res) with an optimized light gradient boosting machine to address the diagnostic bottleneck of rolling bearings under strong noise. Huang et al. [22] proposed a multi-scale cascade CNN (MC-CNN) to enhance feature extraction in bearing fault diagnosis. By mitigating the loss of classification information inherent in single-scale approaches, this model proves effective under various noise interferences. To tackle the domain shift problem in fault diagnosis under varying operating conditions of complex industrial processes, Zhu et al. [23] introduced a multi-scale domain adaptive network. This network mitigates distribution discrepancy by integrating global-local feature extraction with multi-scale distribution alignment. However, this method mainly targets domain transfer problems caused by changes in operating conditions and cannot simultaneously handle coupled fault diagnosis tasks involving strong noise interference and mixed fault characteristics. Song et al. [24] proposed a fault diagnosis model based on multi-scale convolutional neural network (MSCNN) and matrix graph fusion. Through time series data visualization and multi-scale feature extraction, the proposed model can effectively diagnose faults, and the problem of fault classification for Galway nonlinear data of complex chemical processes is solved. Hei et al. [25] effectively addressed the challenges of domain shift and multimodal data alignment in fault diagnosis of rotating machinery under variable working conditions by introducing a multi-scale conditional adversarial network (MCAN), which utilizes its internal deep multi-level attention mechanism and dual bidirectional long short-term memory (BILSTM) domain discriminators. Liu et al. [26] proposed a multi-scale residual network-long short-term memory fusion model (MRCNN-LSTM) to realize the joint learning of industrial process fault diagnosis through spatiotemporal feature coupling and a residual optimization mechanism.

While the aforementioned model demonstrates enhanced diagnostic performance for bearing faults, it is primarily limited to single-fault scenarios and one-dimensional signal processing. Its reliance on fixed-value convolution for extracting single-scale or multi-scale features hinders its ability to achieve effective end-to-end diagnosis under mixed-fault conditions. Although attention mechanisms integrated with multi-scale learning have advanced fault diagnosis, they remain inadequate in handling mixed faults. Most existing methods are limited to enhancing single- or dual-domain features, and conventional attention often relies on rigid weighting schemes, restricting their adaptability in complex conditions. In addition, numerous cutting-edge achievements have emerged in the field of fault diagnosis in recent years. Sun et al. [27] proposed an autoregressive data generation method (CSQ-VAE) based on wavelet packet transform and cascaded stochastic quantization. By means of signal frequency-domain decomposition, multi-scale feature learning, and attention-guided quantization fusion, this method addresses the dataset imbalance issue caused by scarce fault samples in practical engineering, and effectively improves the bearing fault diagnosis performance under small-sample conditions. Its core contribution resides in tackling the challenges of scarce data and imbalanced class distributions, with a primary emphasis on single fault types. Nevertheless, it presents difficulties in dealing with the feature coupling of concurrent mixed faults. Djordjevic et al. [28] developed a robust sensor fault estimation framework using dual sliding-mode observers and coordinate transformation, which provides a new insight for fault perception in dynamic systems. For the domain shift issue under variable working conditions, Sun et al. [29] presented a single-source domain generalization framework on the basis of wavelet packet augmentation and pseudo-domain

generation, which improves the diagnostic generalization under variable working conditions. Although the aforementioned methods have achieved encouraging progress in sensor fault estimation, data augmentation, and domain generalization, they still suffer from three common limitations: (1) The majority of these methods are designed to optimize for a single issue (e.g., sample imbalance or domain shift) and are thus unable to simultaneously tackle the three core challenges encountered in industrial scenarios: severe noise interference, coupled and mixed fault characteristics, and inadequate extraction of weak fault features. (2) Deep learning models predominantly rely on fixed-scale convolution or single-domain attention mechanisms, which are incapable of adaptively capturing the intricate fault features concealed within bearing vibration signals. (3) The generalization capability and noise-resistant robustness of these methods are inadequate to fulfill the practical demands imposed by complex industrial operating conditions. In addition, maintenance costs, downtime, and operational efficiency of industrial systems directly rely on the accuracy and timeliness of fault diagnosis. Traditional methods struggle with mixed faults and high noise, which degrades diagnostic reliability and limits the implementation of predictive maintenance.

To address the above core challenges of mixed fault feature coupling, strong noise interference, and insufficient adaptive feature extraction, this paper proposes a bearing fault diagnosis method named multi-scale adaptive dynamic convolution with three-domain attention (MSADC-TDA). The MSADC module realizes adaptive extraction of multi-scale features, while the TDA module enhances the perception of key fault features in time-frequency-space domains, jointly solving the core pain points of traditional methods. This approach employs variational mode decomposition (VMD) for adaptive noise reduction of the fault signal. Subsequently, a two-dimensional image representation of the one-dimensional signal is generated by applying the CWT to extract more comprehensive fault information. The core of MSADC-TDA involves constructing multi-scale adaptive dynamic convolution layers, dynamic adjustment of convolutional weights is achieved by integrating a channel attention mechanism, which enhances feature selectivity. The TDA module extracts and models features from the time, frequency, and spatial domains simultaneously. It incorporates a domain-aware weight generator for dynamic weight adjustment via backpropagation and employs a feature fusion strategy based on element-wise summation with Softmax normalization. This design effectively integrates multi-domain features and enhances the discernibility of complex fault characteristics under strong noise. The model constructs residual blocks using MSADC and residual connections. Additionally, a dropout layer is incorporated to mitigate redundant information, prevent overfitting, and improve diagnostic efficiency. Consequently, the proposed method not only facilitates effective adaptive feature extraction but also ensures highly robust, end-to-end diagnosis for both single and mixed bearing faults under noisy operating conditions. It provides a technical foundation for the optimization of maintenance scheduling, reduction of downtime, and improvement of operational efficiency in industrial systems. The main contributions are summarized as follows:

(1) The TDA mechanism is proposed to synergistically model the time-domain, frequency-domain, and spatial-domain features of time-frequency images. This multi-domain collaborative attention scheme alleviates the perceptual limitations inherent in single-domain attention mechanisms, enabling accurate focusing and extraction of critical fault features under complex fault scenarios.

(2) The MSADC module is designed, which dynamically recalibrates and fuses multi-scale features by incorporating a channel attention mechanism. This module overcomes the constraints of fixed-scale convolution in traditional CNNs, thus achieving adaptive learning of signal characteristics.

(3) The MSADC module is integrated with residual connections to construct the MSADC residual block, which mitigates the gradient vanishing problems during deep network training. A dropout layer suppresses overfitting from redundant information, enhancing model generalization and stability under complex conditions.

2. Basic principles

2.1. Basic principles of VMD

Variational mode decomposition (VMD) is an adaptive signal processing method that decomposes a signal into multiple intrinsic mode functions (IMFs). Its mathematical foundation involves formulating and solving a constrained variational optimization problem:

$$\begin{cases} s.t.\sum_{k} u_k(t) = f(t), \\ \min\limits_{\{u_k\},\ \{\omega_k\}} \left\{ \sum_{k} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\}. \end{cases} \tag{1}$$

In the formula, $u_k(t)$ denotes the k-th IMF; ω_k is the central frequency corresponding to each modal component; ∂_t denotes the partial derivative operator; $\delta(t)$ denotes the impulse function; $f(t)$ denotes the original gear fault signal; and K denotes the preset number of modes.

Compared with traditional empirical mode decomposition (EMD), VMD offers distinct advantages: a rigorous mathematical foundation, superior ability to mitigate mode mixing, enhanced robustness to noise, and reduced susceptibility to endpoint effects. Specifically, the constrained variational problem is reformulated by introducing a Lagrange multiplier and a quadratic penalty term:

$$L = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 +$$

$$\left\| u(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), u(t) - \sum_{k=1}^{K} u_k(t) \right\rangle. \tag{2}$$

In the formula, L is the Lagrangian function, α is the quadratic penalty factor, and $\lambda(t)$ is the Lagrangian multiplier about time t. Through loop iteration, the optimal eigenmodal function $\{u_k\}$ and its central frequency $\{\omega_k\}$ are obtained.

The mode number K is a central parameter in VMD, as its value is a decisive factor influencing the decomposition outcome. This paper uses an adaptive modal number selection method based on spectral kurtiness. Calculate the spectral kurtiness of the signal:

$$SK(f) = \frac{\langle | S(t,f) |^4 \rangle}{\langle | S(t,f) |^2 \rangle^2} - 2 \tag{3}$$

In this expression, $S(t,f)$ denotes the time-frequency representation of the signal. Calculate the overall spectral kurtitude index:

$$\bar{SK} = \frac{1}{N}\sum_{f} SK(f) \tag{4}$$

Determine the number of modals based on the spectral silence:

$$K = max\left(3, min\left(8, \left\lfloor 3 + 0.5\log_{10}\left(\bar{SK}\right)\right\rfloor\right)\right) \tag{5}$$

2.2. VMD denoising process

The VMD denoising process employed in this study is illustrated in Figure 1. The noisy signal is first decomposed into multiple IMF components using VMD. Subsequently, the autocorrelation coefficient of each IMF is calculated. The denoising process retains and reconstructs only those components whose autocorrelation coefficient exceeds 0.25.

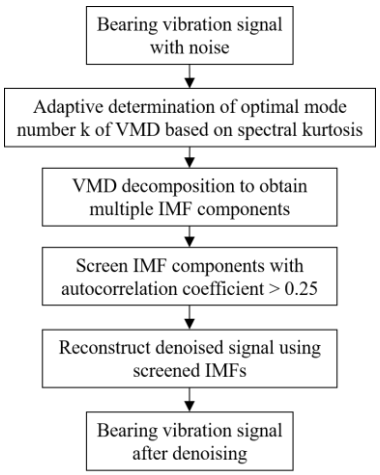


Figure 1. VMD denoising flowchart.

2.3. Continuous wavelet transformation

A two-dimensional image representation of the one-dimensional signal is generated by applying the CWT to extract more comprehensive fault information. This multi-scale analysis effectively reveals the local characteristics of signals and is thus suitable for nonstationary mechanical vibration signals. The CWT of a signal $x(t)$ is expressed mathematically as

$$\mathrm{CWT}(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt, \tag{6}$$

where a is the scale factor, b the translation parameter, and ψ(t) the mother wavelet, defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{7}$$

In this study, a Morlet wavelet is used because it is similar to the fault impact signal of bearings and is often used in bearing fault diagnosis research.

2.4. Residual connection

To address the gradient vanishing and explosion problems in traditional CNNs, He et al. [30] proposed a residual network architecture. This model constructs residual blocks using two convolutional layers with skip connections, as illustrated in Figure 2. Here, x denotes the input, and the output is computed as y =F(x) + x, where F(x) =y - x represents the residual mapping. By learning the difference between input and output, this structure facilitates gradient backpropagation and enhances feature learning capabilities.

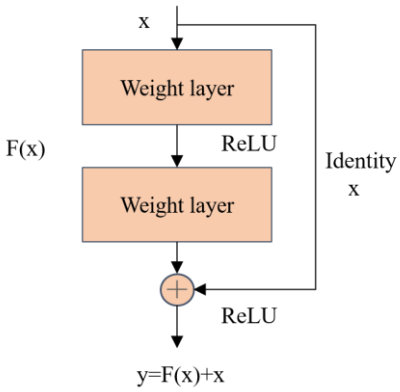


Figure 2. Residual block structure diagram.

2.5. Multi-scale adaptive residual network

2.5.1. Multi-scale adaptive dynamic convolution

Traditional CNNs, constrained by fixed architectures and single-scale convolution kernels, often capture only features at specific scales. To address this limitation, researchers have developed multi-scale convolution structures. The underlying mechanism involves kernels of diverse sizes to learn discriminative features across multiple scales, subsequently fusing them to generate the final output. When applied to bearing fault diagnosis, CWT is used to convert one-dimensional vibration signals into time-frequency images. These representations retain significant time-domain characteristics and exhibit complex temporal features. Therefore, multi-scale convolution effectively enhances a neural network's ability to recognize fault characteristics by learning features at different time scales.

Conventional multi-scale convolution does not differentiate the importance of various kernel sizes during feature learning, thereby constraining the model's efficacy in capturing time-domain information from CWT representations. To address this limitation, we have developed MSADC. The detailed architecture of our proposed module is presented in Figure 3.

First, to capture multi-scale features, the input feature map is fed into three parallel convolutional layers with kernel sizes of 3×1, 7×1, and 9×1, yielding feature maps M_1, M_2, and M_3. These multi-scale features are then fused via element-wise addition to generate a composite feature map $\mathrm{M} = \mathrm{M}_1 + \mathrm{M}_2 + \mathrm{M}_3$. This step enables multiscale feature extraction and integration within a unified representation.

Second, the fused features undergo refinement through a three-domain attention mechanism to enhance local feature extraction. This three-domain attention (TDA) mechanism is detailed in Section

2.6. The refined features are then processed by a global average pooling (GAP) layer, followed by batch normalization (BN) and rectified linear unit (ReLU) activation, resulting in feature U. The mathematical expression for U is defined as

$$\mathbf{U} = \tau\left(\beta\left(GAP\left(TDA(\mathrm{M})\right)\right)\right) \tag{8}$$

Where M represents the fused multi-scale feature map, TDA represents the three-domain attention, β represents BN processing, and τ represents the ReLU activation function.

Finally, feature U is then processed by a fully connected (FC) layer to produce the three components. The Softmax function adaptively computes their normalized weights, producing three corresponding soft attention vectors. This process is mathematically expressed as

$$a_s = \frac{e^{A_s\mathbf{U}}}{e^{A_s\mathbf{U}} + e^{B_s\mathbf{U}} + e^{C_s\mathbf{U}}}, \tag{9}$$

$$b_s = \frac{e^{B_s\mathbf{U}}}{e^{A_s\mathbf{U}} + e^{B_s\mathbf{U}} + e^{C_s\mathbf{U}}}, \tag{10}$$

$$c_s = \frac{e^{C_s\mathbf{U}}}{e^{A_s\mathbf{U}} + e^{B_s\mathbf{U}} + e^{C_s\mathbf{U}}}. \tag{11}$$

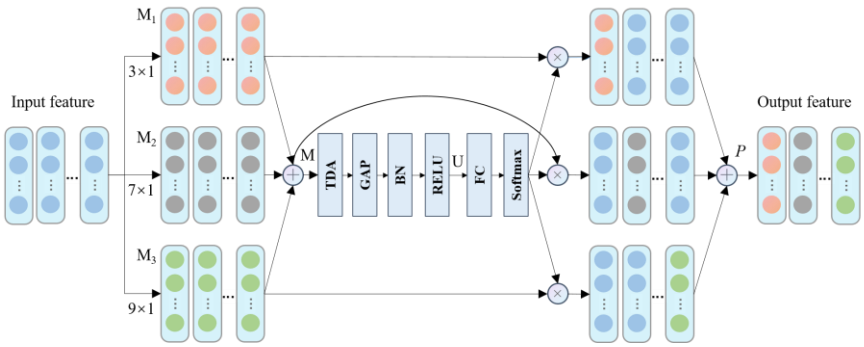


Figure 3. Multi-scale adaptive dynamic convolution structure diagram.

$$a_s + b_s + c_s = 1 \tag{12}$$

Here, A_s, B_s, and C_s denote the soft attention vectors corresponding to features U_1, U_2, and U_3, respectively. These vectors undergo element-wise multiplication and summation with the original feature map to produce the calibrated output feature P map, whose mathematical expression is defined as:

$$P = a_s \cdot M_1 + b_s \cdot M_2 + c_s \cdot M_3 \tag{13}$$

2.5.2. MSADC residual block

Residual connections effectively facilitate feature propagation while mitigating network overfitting. Consequently, we construct residual blocks by directly integrating two MSADC modules with residual connections. Since each MSADC module inherently incorporates BN layers and ReLU activation functions, no additional BN or ReLU layers are added to minimize structural redundancy. The architecture of the residual block is illustrated in Figure 4.

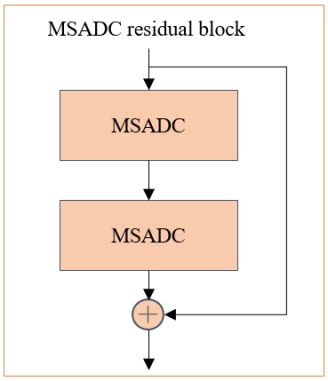


Figure 4. MSADC residual block.

2.6. Three-domain attention module

Bearing fault characteristics exhibit three intrinsic attributes: leverage information from the time, frequency, and spatial domains. (1) TD characteristics: temporal correlations in impact signals, compressing spatial features into time-series vectors along the temporal axis to preserve global sequence information. (2) FD characteristics: salient fault feature frequencies, generated by compressing feature maps along spatial dimensions to produce channel-level statistical vectors. (3) SD characteristics: localized regions of interest in time-frequency images, captured using 7×7 large convolutional kernels to extract texture patterns of fault impacts.

Traditional attention mechanisms are confined to single-dimensional processing. This paper innovatively proposes a three-domain collaborative attention model, as shown in Figure 5. After converting 1D vibration signals into 2D CWT representations, the model concurrently models temporal, spectral, and spatial dimensions to enhance focus on critical features, making it suitable for multimodal or multi-domain feature fusion tasks. During feature learning, the mechanism dynamically allocates domain-specific attention weights based on input signal characteristics, where $\otimes$ denotes element-wise multiplication.

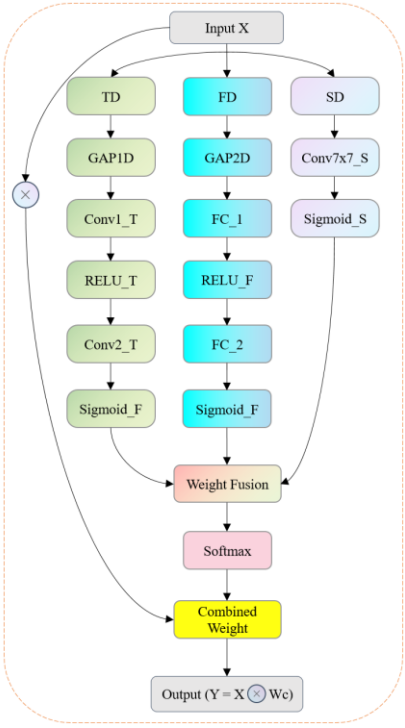


Figure 5. Three-domain attention structure.

After entering the features, the three-domain attention module performs adaptive feature extraction, and the expression can be described as

$$\mathbf{Z}_{TD} = \tau(\mathrm{GAP}(C_K(\mathbf{X})) \tag{14}$$

$$\mathbf{Z}_{FD} = \tau(FC(\mathrm{GAP}(\mathbf{X})) \tag{15}$$

$$\mathbf{Z}_{SD} = \tau(C_K(\mathbf{X})) \tag{16}$$

Here, C_K denotes the convolutional operation with kernel size K, FC represents the fully connected layer, GAP indicates global average pooling, and τ corresponds to the ReLU activation function.

Following feature extraction, the temporal, spectral, and spatial domains are each processed through convolutional and fully connected layers, followed by adaptive weight generation via Sigmoid functions. The dynamic weight generation process is defined as

$$\mathbf{W}_{TD} = \sigma(C_K(\mathbf{Z}_{TD})) \tag{17}$$

$$\mathbf{W}_{FD} = \sigma(FC(\mathbf{Z}_{FD})) \tag{18}$$

$$\mathbf{W}_{SD} = \sigma(\mathbf{Z}_{SD}) \tag{19}$$

$$\mathbf{W}_{*} = \sigma(\mathcal{G}_{\theta}(\mathbf{X})) \tag{20}$$

Here, $\mathbf{W}_{TD}$, $\mathbf{W}_{FD}$, $\mathbf{W}_{SD}$ denote the weights for the temporal, spectral, and spatial domains, respectively. $\mathcal{G}_\theta$ represents a multi-branch feature processor (TD: 1D convolution, FD: GAP+FC, SD: 7×7 convolution). σ denotes the Sigmoid function, constraining outputs to the [0,1] interval. These modality-specific weights $\mathbf{W}_*$ are functions of input features $\mathbf{X}$, dynamically optimized via backpropagation. The domain weights are then fused and normalized by the Softmax function. Finally, the calibrated feature map Y is generated through element-wise multiplication and summation of each domain weight with the original feature map.

$$\mathrm{Y} = \mathbf{X} \otimes (\mathbf{W}_{TD} + \mathbf{W}_{FD} + \mathbf{W}_{SD}) \tag{21}$$

3. Based on multi-scale adaptive residual and three-domain attention fault diagnosis model

3.1. Multi-scale dynamic adaptive residual and three-domain attention model

Owing to the operation of rolling bearings under varying working conditions and the interference of intense noise, the incidence of mixed faults in bearings is substantially elevated. This study proposes a multi-scale dynamically adaptive residual network integrated with a three-domain attention model (MSADC-TDA), as illustrated in Figure 6. Specifically, the multi-scale ResNet is enhanced with dynamic adaptability, which integrates the robust feature extraction capability of multi-scale residuals with the independent processing of information across different dimensions by each branch of the three-domain attention. For the fusion of three-domain weights, element-wise summation is adopted rather than concatenation, so as to retain the intrinsic characteristics of each dimension.

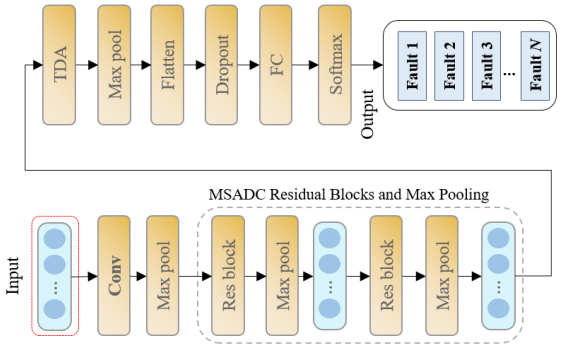


Figure 6. MSADC-TDA structure diagram.

This method primarily consists of an initial convolutional layer, a pooling layer, MSADC residual blocks, a three-domain attention layer, a subsequent pooling layer, a dropout layer, fully connected layers, and a classification layer. A key innovation is the MSADC residual block, which integrates parallel multi-convolution feature learning and a dynamic weight adjustment mechanism adaptive to input scales. By exploring the nonlinear correlation between scaling rates and channel counts, this module can adaptively optimize the scaling rate of each convolutional layer, thereby markedly enhancing the efficiency of feature learning. Subsequently, after feature extraction, the three-domain attention strengthens the model's focus on key fault features. Through multi-domain joint dynamic weighting, it further improves the model's capacity to capture multidimensional features. As a result, the MSADC-TDA network—composed of MSADC residual blocks and three-domain attention—

exhibits superior fault feature extraction performance. Furthermore, following the three-domain attention module, a max-pooling layer is incorporated into the network to reduce feature dimensionality, while a dropout layer is employed to alleviate overfitting. Ultimately, at the network's output stage, fully connected layers coupled with a Softmax function serve as the classifier for identifying different fault types.

3.2. MSADC-TDA fault diagnosis process

To address the high noise levels in the operating environment of rotating machinery—in which the collected bearing vibration signals are contaminated with noise—as well as the problem of fixed-value convolution in multi-scale feature extraction and the prevalence of mixed faults in bearings, a rolling bearing fault diagnosis model named MSADC-TDA is proposed. The fault diagnosis flowchart is presented in Figure 7, and the diagnostic process of MSADC-TDA proceeds as follows:

1. For vibration signals of bearings with single faults or mixed faults (all of which are inherently contaminated with noise), first, VMD is applied for denoising. Subsequently, the denoised fault vibration signals were converted into time-frequency images by applying the CWT.

2. The denoised and transformed signals are split into training, validation, and test sets, which are then subjected to encoding and labeling.

3. The training set is utilized to train the neural network: during this process, the model extracts key fault feature representations and gradually learns fault characteristics.

4. The validation set facilitates the optimization of the pretrained model's hyperparameters, ultimately yielding a fault diagnosis model with stable performance.

5. The trained model processes the test set to produce the final fault diagnosis results, thereby completing the entire fault diagnosis process.

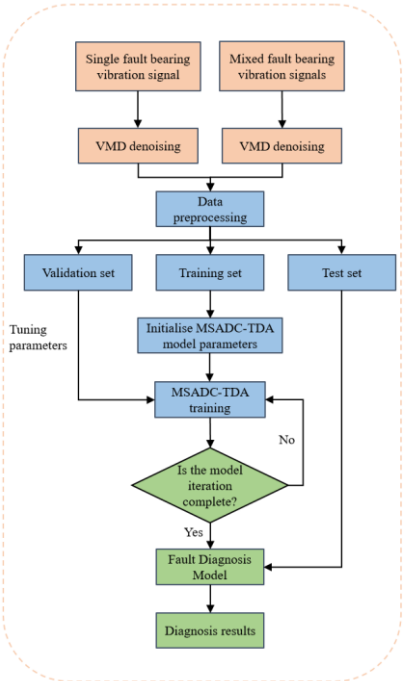


Figure 7. MSADC-TDA fault diagnosis flowchart.

3.3. Industrial application potential

The design of MSADC-TDA prioritizes feasibility for industrial deployment, and its complete end-to-end conceptual diagnostic workflow—including signal acquisition, preprocessing, intelligent diagnosis, and result feedback—can be readily integrated into industrial monitoring systems. First, real-time bearing vibration signals are collected via industrial sensors, then adaptively denoised using VMD and converted into time-frequency maps via CWT to enhance fault features, a preprocessing chain well suited to high-noise industrial environments. These maps are then fed into the MSADC-TDA model for fault classification: The MSADC module enables adaptive feature learning to reduce reliance on manual intervention, while the TDA module enhances feature discriminability to mitigate false alarms. Finally, diagnostic results are uploaded to the industrial monitoring system in real time to provide support for equipment operation, maintenance and early warning, triggering maintenance alerts and dynamically optimizing equipment operation and maintenance plans to directly support predictive maintenance practices and reduce unplanned equipment downtime. With its high lightweight property and fast inference speed, the entire workflow can be directly deployed on edge computing devices in industrial sites, satisfying the engineering requirements of online real-time fault diagnosis for rotating machinery.

4. Experimental verification and result analysis

In this section, we evaluate the fault diagnosis performance of the proposed method on a self-built rolling bearing dataset, and experimentally verify its noise robustness and mixed fault diagnosis capability. All models were trained and inferred on a platform equipped with an AMD R5 7500F CPU and NVIDIA RTX 4060Ti GPU, with PyTorch as the deep learning framework. To achieve optimal performance, the mode number K of VMD is adaptively determined via spectral kurtosis in the preprocessing stage. For the core modules of the proposed model, parameter settings are elaborated as follows: MSADC employs parallel convolutional kernels of 3×1, 7×1, and 9×1 to capture comprehensive multi-scale temporal features. The structural parameters of the TDA module are configured according to the time-, frequency- and spatial-domain characteristics of bearing faults. Specifically, the time-domain branch adopts a 1D convolution with a kernel size of 3; the frequency-domain branch utilizes a bottleneck fully-connected layer with a compression ratio of 4; and the spatial-domain branch fixedly applies a 7×7 convolutional kernel to focus on local fault textures. The remaining training hyperparameters are set in line with robust empirical conventions in the fault diagnosis field. This parameter configuration strategy effectively guarantees the high diagnostic accuracy and strong robustness of the proposed method in complex industrial scenarios.

4.1. Single fault bearing fault diagnosis

4.1.1. Data acquisition and experimental design

The test bench used in this experiment is the Mechanical Fault Simulator (MFS) from Spectra Quest Co, as shown in Figure 8. The key components of this system include a drive motor, a tachometer, an accelerometer, faulty bearings, and a software system for signal acquisition and speed control. Among them, the model of the faulty bearing is ER-12K. Under the constant speed condition, the

sampling frequency for the bearing test is set to 12.8 kHz. Bearing faults occur in the inner race, rolling element, and outer race, respectively, and the rotational frequencies under each fault state are 19.88 Hz, 29.87 Hz, and 39.84 Hz in sequence. There are 10 types of bearings available for testing, including bearings under normal operating conditions.

The single fault types and codes of bearings operating at constant speed are shown in Table 1. Each signal with 327,680 points was segmented into 1024-point samples, with 320 samples per bearing and a total of 3,200 samples in the dataset.

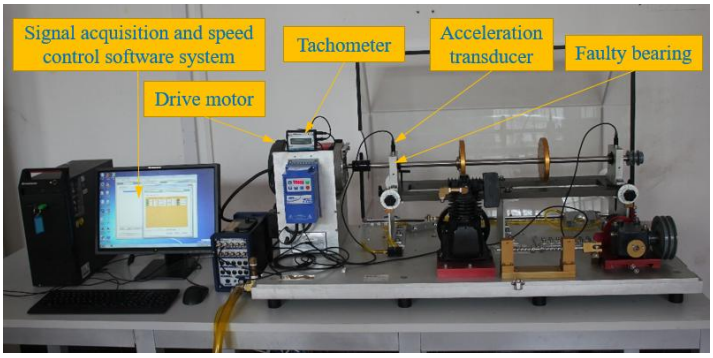


Figure 8. Bearing experimental platform device diagram.

Table 1. Single fault bearing type and code.

Fault type	Rotating frequency	Tags
Normal	19.88Hz	0
Rolling element fault	19.88Hz	1
Rolling element fault	29.87Hz	2
Rolling element fault	39.84Hz	3
Inner ring failure	19.88Hz	4
Inner ring failure	29.87Hz	5
Inner ring failure	39.84Hz	6
Outer ring fault	19.88Hz	7
Outer ring fault	29.87Hz	8
Outer ring fault	39.84Hz	9

4.1.2. Parameter settings

We adopted a 70/20/10 split for the dataset, with 70% of the samples allocated to the training set, 20% to the validation set, and the remaining 10% to the test set. We set the Adam optimizer with a learning rate of 0.001 and a weight decay of 0.01. The classification task utilizes the Softmax activation function, and training proceeds with a batch size of 64. A dropout rate of 0.25 is applied before the fully connected layer to prevent overfitting by randomly disabling neurons. The model architecture incorporates maximum pooling and ReLU activation functions. The specific parameters of the model are listed in Table 2.

Table 2. MSADC-TDA model parameters.

Network structure	Kernel size/stride	Kernel channel size	Output size
Convolution	64×1/8×1	64	64×64
Pooling	2×1/2×1	—	32×64
Res block 1	MSADC	64	32×64
	MSADC	64	32×64
Pooling	2×1/2×1	—	16×64
Res block 2	MSADC	128	16×128
	MSADC	128	16×128
Pooling	2×1/2×1	—	8×128
FC	128	1	128×1
Softmax	10	1	10

4.1.3. Experimental results

The accuracy and loss of bearing training sets and verification sets under a single fault condition are shown in Figure 9.

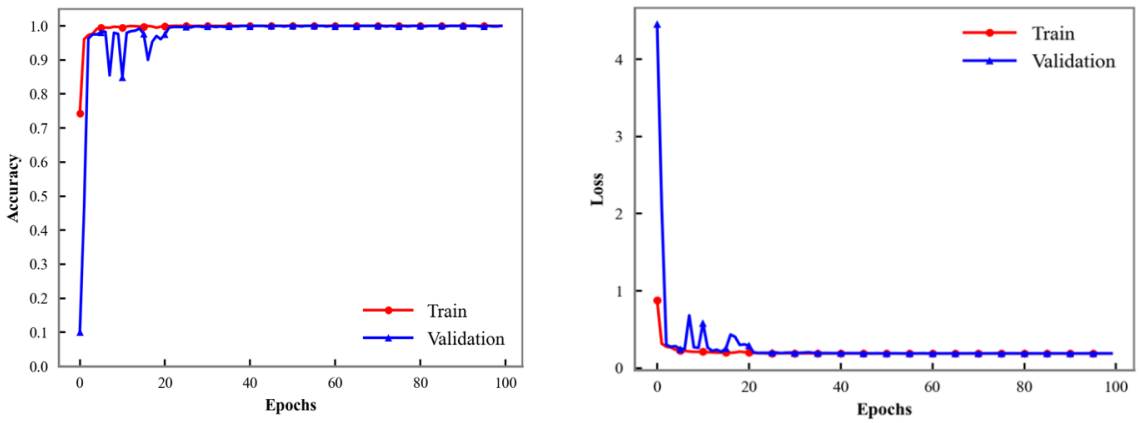


Figure 9. Results of a single fault bearing in the training set and verification set.

After 20 iterations of the model, the training and validation accuracy both approached 100%, with a stable loss function, demonstrating full model convergence. For a clear demonstration of the model's performance in identifying various faults on the test set, the prediction results are visually presented using the confusion matrix in Figure 10. The results show that the MSADC-TDA's determination of bearing failure status is highly consistent with the real label. Under a single fault condition, the identification error occurs only in the rolling element fault sample with a rotation frequency of 29.87 Hz.

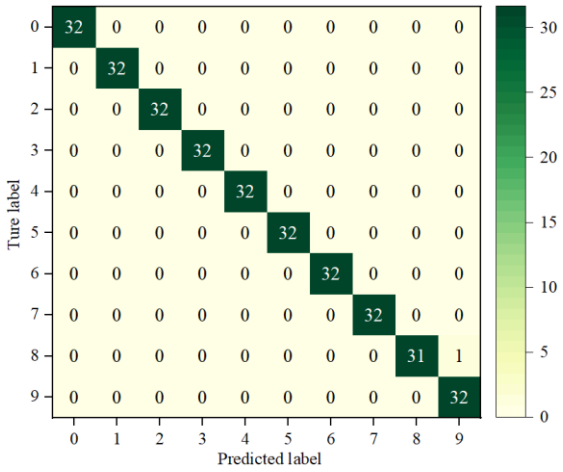
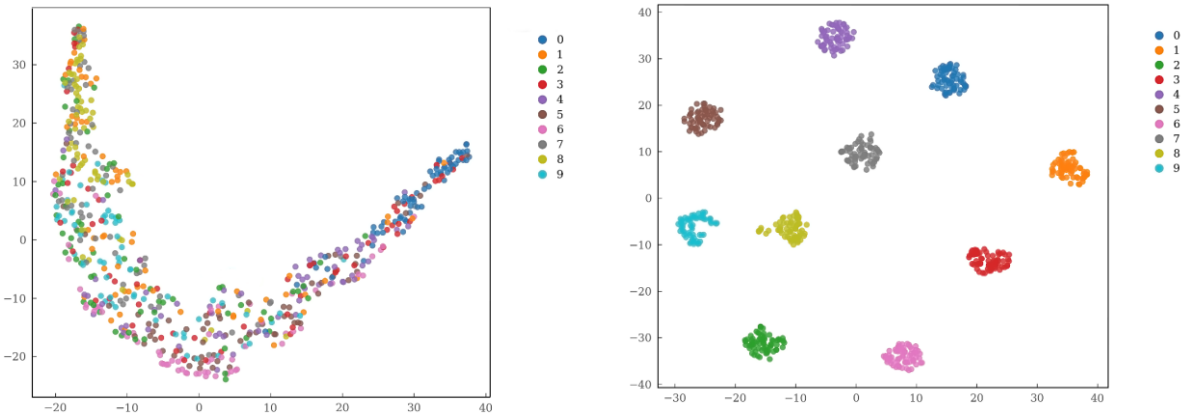


Figure 10. Single fault bearing confusion matrix.

The t-distributed stochastic neighbor embedding (t-SNE) method is further employed to validate the feature learning capability of MSADC-TDA on bearing fault data, as shown in Figure 11. Comparison of the input layer data distribution in Figure 11(a) reveals severe intermixing of different fault types, exhibiting a near-chaotic state that impedes clear differentiation. In contrast, the output layer feature visualization shown in Figure 11(b) demonstrates exceptional separability, and all 10 fault categories are distinctly separated while samples within each category form tightly clustered groups. This conclusively verifies MSADC-TDA's robust high-dimensional feature extraction and discrimination capabilities.



(a) (b)

Figure 11. (a) Single fault bearing input layer t-SNE diagram; (b) Single fault bearing output layer t-SNE diagram.

Fast training convergence and near-perfect test accuracy confirm the high reliability of MSADC-TDA. This stability enables its deployment in industrial online monitoring systems. It supports continuous and accurate bearing health assessment. It provides critical support for shifting from scheduled to predictive maintenance, avoiding under- or over-maintenance and optimizing maintenance scheduling.

4.1.4. Comparative experiment

To verify the effectiveness of MSADC-TDA for bearing fault diagnosis under single-fault conditions, we compare its performance against several established methods: CNN [31], WDCNN [32], MS-CNN [33], attention-CNN [34], MSCNN with an efficient convolutional module (ECM) [35], adaptive multiscale convolutional neural network (AMCNN) [36], and ResNet-18. To ensure the fairness and reproducibility of the comparison, all comparative models are constructed based on their original papers or widely accepted standard implementations in the field. Specifically, CNN consists of four 3x3 convolutional blocks, with increasing channel numbers of 16, 32, 64, and 128. WDCNN adopts a wide first-layer convolution kernel of size 32×1 and a shallow network structure; MS-CNN incorporates three parallel convolutional branches with different kernel sizes of 3×3, 5×5, and 7×7; Attn-CNN shares the same architecture as the single-scale CNN, with an squeeze-and-excitation channel attention module inserted after each convolutional layer; ResNet-18 employs its standard 18-layer residual structure; MSCNN-ECM uses three parallel branches with convolutions of 1×1, 7×7, and 9×9, respectively, and integrates efficient channel attention modules into the residual block; and the AMCNN model employs an ensemble of four parallel convolutional branches with kernel sizes of 1×1, 3×3, 5×5, and 7×7, and incorporates a hybrid attention module that combines both channel and spatial attention mechanisms. In all experiments, every model is trained and tested under exactly identical settings, including the same preprocessing pipeline, dataset split ratio, optimizer, learning rate, batch size, and number of training epochs. Experimental results, presented in Table 3, show the average diagnostic accuracy across five trials. As shown in the table, MSADC-TDA achieves a superior average accuracy of 99.38% under single-fault conditions, outperforming all other compared methods.

Table 3. Diagnostic accuracy of different methods under a single fault condition (100%).

Methods	Average accuracy（%）
CNN	96.86
WDCNN	97.4
MS-CNN	98.1
Attn-CNN	97.25
ResNet-18	96.34
MSCNN-ECM	97.66
AMCNN	98.21
MSADC-TDA	99.38

4.1.5. Noise robustness analysis

To evaluate the robustness of the MSADC-TDA model to noise under mixed fault conditions, the signal-to-noise ratios are set to -5dB, -3dB, and -1dB, respectively. The classification accuracy in the experiment was averaged by 5 tests.

As shown in Figure 12, despite a decline in fault diagnosis accuracy due to noise interference, the model maintains superior performance. Under harsh noise conditions where the signal-to-noise ratio (SNR) was set to -5 dB, -3 dB, and -1 dB, the model achieves average diagnostic accuracies of 88.69%,

91.13%, and 93.43%, respectively, outperforming all other comparative methods. Furthermore, the model demonstrates superior stability across five experimental trials compared to alternative methods.

MSADC-TDA maintains high diagnostic accuracy even in extremely low signal-to-noise ratio environments. This makes it suitable for harsh industrial scenes such as manufacturing workshops. It reduces noise-induced misdiagnoses and unnecessary shutdown inspections. This is critical for stable and continuous production line operation, and helps lower unplanned downtime.

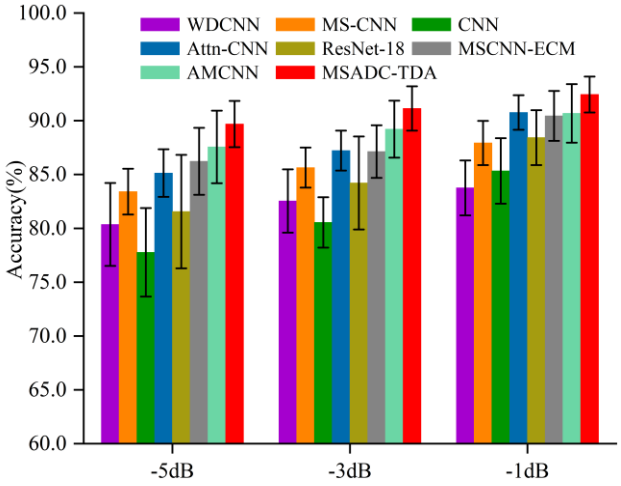


Figure 12. Diagnostic accuracy of different diagnostic methods of single fault bearings in noise environments.

4.2. Fault diagnosis of mixed fault bearings

4.2.1. Data acquisition and experimental design

The fault diagnosis experimental setup for mixed fault bearings is the same as the fault diagnosis experimental setup for a single fault bearing, using the experimental data were collected on a mechanical fault simulator (MFS) manufactured by Spectra Quest Co. Four types of mixed-fault vibration signals were collected from bearing experiments in the laboratory. These included outer ring faults, outer ring faults combined with rolling element faults, inner ring faults combined with rolling element faults, and faults involving all three components. The sampling frequency is 25.6 kHz, and the rotation frequency in various types of fault states is 19.88 Hz and 29.87 Hz in turn. Table 4 shows the fault types and codes for running mixed fault bearings. Each faulty bearing vibration signal, comprising 327,680 points, was segmented into 320 samples of 1024 points each, resulting in a total dataset of 2,880 samples. There are 9 types of bearings available for testing, including bearings under normal operating conditions.

Table 4. Mixed fault bearing types and codes.

Fault type	Rotating frequency	Tags
Normal	19.88Hz	0
Rolling element + inner race fault	19.88Hz	1
Rolling element + inner race fault	29.87Hz	2
Inner race + outer race fault	19.88Hz	3
Inner race + outer race fault	29.87Hz	4
Outer race + rolling element fault	19.88Hz	5
Outer race + rolling element fault	29.87Hz	6
Rolling element + inner race + outer race fault	19.88Hz	7
Rolling element + inner race + outer race fault	29.87 Hz	8

4.2.2. Parameter settings

For bearing fault diagnosis under mixed-fault conditions, the dataset is allocated as follows: 70% to the training set, 20% to the validation set, and 10% to the test set. The Adam optimizer is configured with a learning rate of 0.005 and a weight decay rate of 0.01. The classification task employs the Softmax activation function, with a training batch size of 64. Similarly, dropout is applied before the fully connected layer at a rate of 0.35 to prevent overfitting by randomly discarding neurons. After averaging ten experimental trials, diagnostic accuracy peaks at this dropout rate of 0.35, as shown in Table 5.

Table 5. The test results of different dropout rates.

Dropout drop rates	0.1	0.25	0.35	0.5
Accuracy (%)	97.89	98.41	98.82	98.6

4.2.3. Experimental results

The accuracy and loss of bearing training sets and verification sets under mixed fault conditions are shown in Figure 13.

After 30 iterations of the model, the training accuracy approached 100%, while the validation set attained an accuracy of 98.82%, and the accuracy curve and the loss function value remain relatively stable, indicating that the model has fully converged.

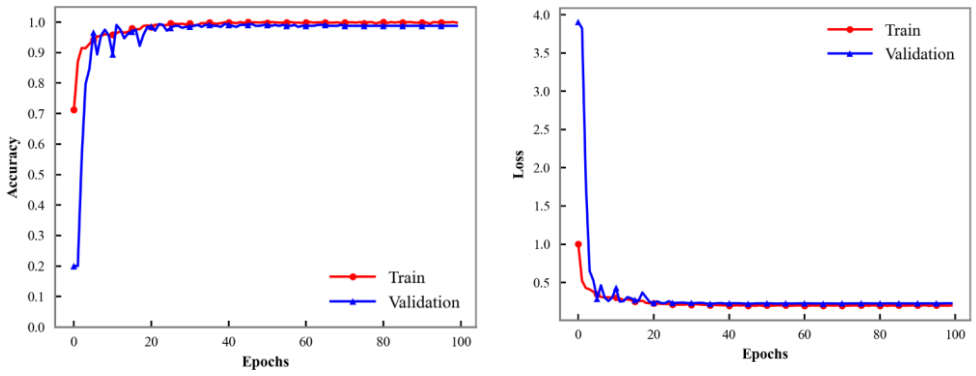
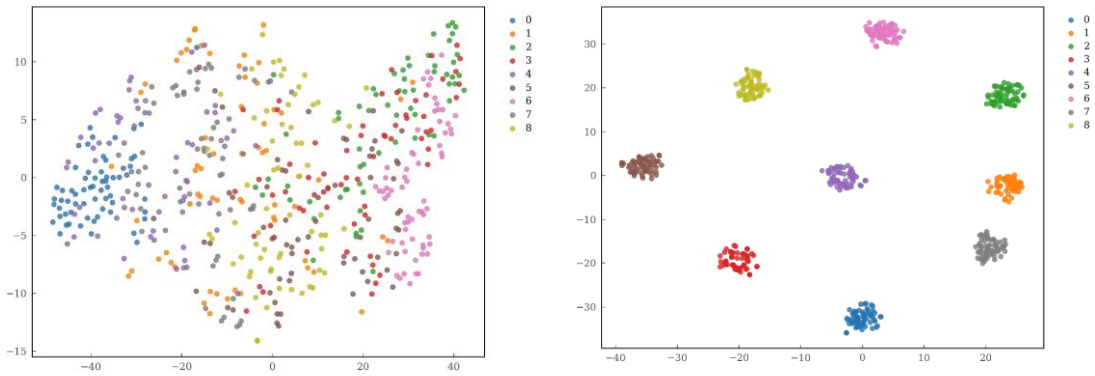


Figure 13. Results of mixed fault bearings in training set and verification set.

The t-SNE method is further employed to validate the feature learning capability of MSADC-TDA on bearing fault data, as shown in Figure 14. Comparison of the raw condition data distribution in Figure 14(a) reveals severe intermingling of different fault types, exhibiting an approximately chaotic distribution that hinders effective differentiation. In contrast, the output layer feature visualization presented in Figure 14(b) demonstrates remarkable separability: All 9 fault categories are distinctly separated, while samples of the same category form tightly clustered groups. This conclusively verifies MSADC-TDA's robust high-dimensional feature learning and discriminative capabilities under mixed fault scenarios.



(a) (b)

Figure 14. (a) Mixed fault bearing input layer t-SNE diagram; (b) Mixed fault bearing output layer t-SNE diagram.

High-precision identification of compound faults is a key industrial value of the proposed method. In engineering practice, bearing faults often appear in mixed forms with coupled and indistinguishable features. MSADC-TDA can effectively identify these complex conditions. It allows maintenance personnel to accurately evaluate the severity of compound faults and develop targeted maintenance strategies. This optimizes maintenance resource allocation and improves handling efficiency.

4.2.4. Comparative experiment

To verify the effectiveness of MSADC-TDA diagnosis under mixed fault conditions, verification was carried out through comparative experiments with other methods. The classification results of

CNN, WDCNN, MS-CNN, attention-CNN, ResNet-18, MSCNN-ECM, AMCNN, and MSADC-TDA, derived from their confusion matrices, are shown in Figure 15(a)–(h). Only ResNet-18 and the proposed MSADC-TDA maintain high accuracy for fault state 3, while other models misclassify samples from this category. However, ResNet-18 shows considerable misclassification for states 5 and 6. While other comparative methods achieve higher overall accuracy than ResNet-18, they still suffer from substantial misclassification under mixed fault scenarios. This is mainly due to mutual interference between fault features from the outer race, inner race, and rolling elements. Notably, most models exhibit a marked decline in accuracy for fault state 3 and struggle to distinguish easily confused fault states, limiting their utility in mixed fault diagnosis. In contrast, MSADC-TDA correctly classifies most fault states, with only three states being misclassified slightly. These results demonstrate the clear performance advantage of MSADC-TDA in mixed fault recognition tasks.

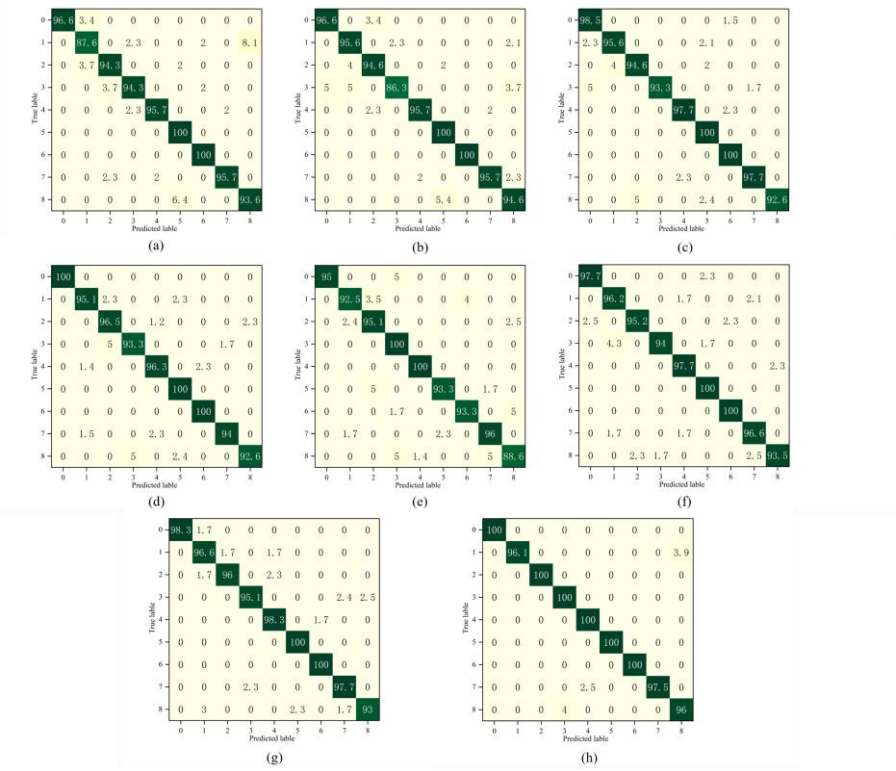


Figure 15. Confusion matrix results for mixed fault classification. (a) CNN, (b) WDCNN, (c) MS-CNN, (d) attention-CNN, (e) ResNet-18, (f) MSCNN-ECM, (g) AMCNN, and (h) MSADC-TDA.

To further evaluate the diagnostic performance of the proposed model under mixed fault conditions, accuracy, precision, and recall are selected as the evaluation metrics. Accuracy reflects the overall proportion of correctly classified samples in the dataset, which serves to measure the global diagnostic effectiveness of the model and represents the most fundamental indicator for assessing the overall performance in balanced fault diagnosis tasks. Precision represents the proportion of true fault samples among those predicted as faulty. In industrial scenarios, high precision can effectively reduce false alarms as well as avoid unnecessary equipment downtime and excessive maintenance costs. Recall represents the proportion of true fault samples that are correctly identified. High recall is critical to reducing missed alarms, ensuring that potential faults can be detected in advance and avoiding

equipment damage or safety incidents caused by fault omission. As shown in Table 6, MSADC-TDA obtains values exceeding 98.8% in all metrics, outperforming other approaches by at least 1.6 percentage points. Furthermore, MSADC-TDA exceeds MS-CNN across every evaluation indicator, which demonstrates that dynamically adjusting convolutional kernel weights is effective in enhancing network performance. These findings not only reflect an improvement in the fault diagnosis performance of MSADC-TDA but also underscore the practical usefulness of the proposed framework.

Table 6. Mixed fault classification results under multiple evaluation metrics (100%).

Methods	Accuracy(%)	Precision(%)	Recall(%)
CNN	95.3	95.33	95.3
WDCNN	95.45	95.49	95.45
MS-CNN	96.6	96.66	96.6
Attn-CNN	96.42	96.52	96.42
ResNet-18	94.87	94.91	94.87
MSCNN-ECM	96.77	96.80	96.77
AMCNN	97.22	97.32	97.22
MSADC-TDA	98.82	98.84	98.82

4.2.5. Noise robustness analysis

The noise robustness of the MSADC-TDA model under mixed fault conditions is evaluated at SNRs of -5 dB, -3 dB, and -1 dB, with the classification accuracy averaged over five independent tests. As shown in Figure 16, the model still maintains excellent diagnostic performance despite the accuracy decline caused by noise interference, achieving average accuracies of 88.45%, 90.24%, and 92.56% at the above SNRs respectively. It outperforms all comparative methods in both accuracy and stability across the five tests.

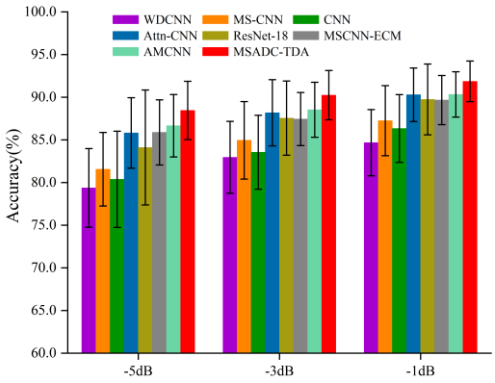


Figure 16. Diagnostic accuracy of different diagnostic methods of mixed fault bearings in noise environments.

These results further confirm that the method works reliably under the dual challenges of compound faults and noise. It ensures timely and accurate detection of potential faults in complex industrial environments. This avoids cascading failures and long equipment downtime caused by fault propagation, thus improving overall operational efficiency.

4.2.6. Computational complexity and deployment considerations

For predictive maintenance applications, computational cost is an important factor. Table 7 shows the statistics of inference time, parameter count, and floating-point operations (FLOPs) for the six methods. The metrics of MSADC-TDA are second only to ResNet-18 and higher than those of the other four methods. This is due to the attention mechanism embedded in each multi-scale convolutional layer and the multi-domain fusion of TDA, which yield a more complex structure with higher inference time, parameter count, and computational cost. However, this does not hinder its application to fault diagnosis in predictive maintenance, as MSADC-TDA achieves the highest accuracy in complex fault and strong noise environments.

Table 7. Inference time, parameter count, and FLOPs for six methods.

Methods	Time(s)	Parameters	FLOPs
CNN	6.8	8.2×105	7×108
WDCNN	4.2	2.6×105	4×108
MS-CNN	10.3	1.2×106	1×109
Attn-CNN	7.8	9.5×105	9×108
ResNet-18	13.1	1.17×107	1.55×109
MSCNN-ECM	10.5	2×106	1.16×109
AMCNN	8.9	1.6×106	9.7×108
MSADC-TDA	9.5	1.9×106	1.1×109

From the perspectives of implementation and computational efficiency, the proposed MSADC-TDA method exhibits strong feasibility for practical industrial deployment. In terms of implementation, the proposed model is developed based on the mainstream PyTorch deep learning framework with a modular architecture. Its core components, including VMD preprocessing, CWT time-frequency transformation, MSADC multi-scale adaptive convolution, and the TDA three-domain attention module, can be independently invoked and feature excellent portability. The proposed method adopts an end-to-end diagnostic pipeline, which eliminates the reliance on manual feature extraction and professional prior knowledge, thus facilitating seamless integration with existing industrial monitoring systems. In terms of computational cost, the model can be trained and inferred on a standard industrial-grade hardware platform (NVIDIA RTX 4060Ti GPU) without the need for high-performance computing clusters. The model has a parameter count of 1.9×10^6 and a single-sample inference time of 9.5 s, which are well within the mainstream allowable latency range for industrial online fault diagnosis, support deployment on edge computing devices, and meet the real-time diagnostic requirements of practical industrial scenarios. These results demonstrate that the proposed method achieves a favorable balance between diagnostic performance and computational overhead, and possesses strong engineering implementability and practical application feasibility.

4.2.7. Exploration and analysis of model lightweighting

The current model prioritizes maximum accuracy without deep model compression. To address the need for enhanced computational efficiency, we adopt the depthwise separable convolution (DSC) strategy, which is widely utilized in industrial edge deployment scenarios, to perform lightweight

optimization on the MSADC module of the proposed model. This strategy can drastically reduce the number of model parameters and computational complexity without significant loss of feature extraction capability, and has been extensively validated in the field of bearing fault diagnosis.

This experiment strictly follows the principle of controlled variables. Only the standard convolutions in the three parallel multi-scale branches of the MSADC module are replaced, while all other network structures, hyperparameters, dataset partitioning, and training environments are completely consistent with those of the original model, ensuring the fairness and reliability of the experimental results. The experimental results are shown in Table 8.

Table 8. Comparison between the original model and the lightweight model.

Methods	Time(s)	Parameters	FLOPs	Accuracy(%)	SNR=-5dB(%)
MSADC-TDA	9.5	1.9×106	1.1×109	98.82	88.45
Lightweight model	6	1.35×106	7×108	97.51	83.13
Change	↓3.5	↓29.5%	↓32%	↓1.31	↓5.32

The results demonstrate that under the noise-free mixed fault dataset, the lightweight model reduces the number of parameters by 29.5% and FLOPs by 32%, with the inference time shortened from 9.5 s to 6 s, while still retaining a high diagnostic accuracy of 97.51%. This result verifies that the proposed lightweight strategy can effectively improve the computational efficiency of the model while maintaining high diagnostic accuracy under noise-free working conditions.

Furthermore, we compared the anti-noise performance of the two models in the core research scenario of this paper, namely the strong noise environment. As shown in Table 9, under the harsh industrial condition with SNR = -5dB, the diagnostic accuracy of the lightweight model drops to 83.13%, whereas the original MSADC-TDA model still maintains a diagnostic accuracy of 88.45%. This fully proves that the original model has stronger robustness and anti-interference performance in actual complex industrial environments, which is an irreplaceable attribute for the fault diagnosis of critical core equipment

4.3. CWRU bearing dataset generalization verification

4.3.1. Data acquisition and experimental design

To verify the universal generalization capability of the proposed MSADC-TDA method, the publicly available bearing fault dataset from Case Western Reserve University (CWRU), is adopted in this section for supplementary validation of the model's diagnostic performance and anti-noise robustness. The test bench is shown in Figure 17. It comprises an electric motor, a torque sensor, and a dynamometer. A SKF 6205 bearing is mounted at the drive end, and single-point damages with diameters of 0.007, 0.014, and 0.021 inches are created on the inner race, outer race, and rolling elements through electrical discharge machining (EDM).

In this experiment, the vibration data of the drive-end bearing at a constant rotating speed of 1797 rpm from the CWRU dataset are utilized, with a sampling frequency of 12 kHz. The data cover four major health states (normal, rolling element fault, inner race fault, and outer race fault), involving a total of 10 working conditions with different fault severities. Each sample is composed of 1024 data points, and 120 independent samples are generated for each working condition, yielding a total sample

size of 1200. The fault types and labels of bearings are shown in Table 9. The dataset is split into training, validation, and test sets with a ratio of 7:2:1.

All experimental configurations are fully consistent with those of the self-constructed dataset experiment in Section 4.1 of this paper. The identical preprocessing pipeline is employed, and the model architecture, hyperparameters, and training environment are kept unchanged with only the input dataset replaced, thereby guaranteeing the single controlled variable and experimental fairness. The comparison is conducted using the methods presented in Section 4.1.4 of this paper, and all comparative methods are implemented under the exact same dataset and experimental settings.

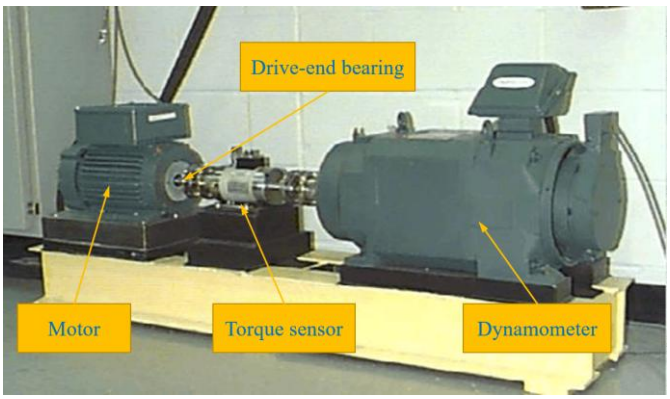


Figure 17. CWRU bearing test rig.

Table 9. Case Western Reserve University dataset.

Fault type	Fault Diameter	Tags
Normal	-	0
Ball	0.007	1
Ball	0.014	2
Ball	0.021	3
Inner race	0.007	4
Inner race	0.014	5
Inner race	0.021	6
Outer race	0.007	7
Outer race	0.014	8
Outer race	0.021	9

4.3.2. Data acquisition and experimental design

The average diagnostic accuracy, precision, and recall of different methods on the CWRU dataset under noise-free conditions are as shown in Table 10. It can be seen that the proposed MSADC-TDA method still achieves excellent values of the three metrics, with an average diagnostic accuracy of 99.22% on the public CWRU dataset. Compared with the other seven comparison methods, its accuracy is improved by 1.19%–4.21%, which fully demonstrates that the proposed method maintains excellent feature extraction capability and diagnostic performance across different bearing fault datasets.

Table 10. Diagnostic accuracy of different methods on the CWRU dataset (100%).

Methods	Accuracy(%)	Precision(%)	Recall(%)
CNN	96.37	96.40	96.37
WDCNN	97.05	97.09	97.05
MS-CNN	97.64	97.68	97.64
Attn-CNN	97.32	97.37	97.32
ResNet-18	95.01	95.05	95.01
MSCNN-ECM	97.83	97.86	97.83
AMCNN	98.03	98.08	98.03
MSADC-TDA	99.22	99.25	99.22

In addition, the feature learning capability of the model is evaluated using a confusion matrix, with the results presented in Figure 18. The model achieves outstanding classification performance, with five fault states reaching 100% accuracy and all other states maintaining accuracies above 98.1%. Minor misclassifications only occur between fault categories with highly similar vibration characteristics, and the overall misclassification rate remains below 0.8%. The results further verify the superior feature learning and generalization capabilities of MSADC-TDA for bearing fault diagnosis tasks.

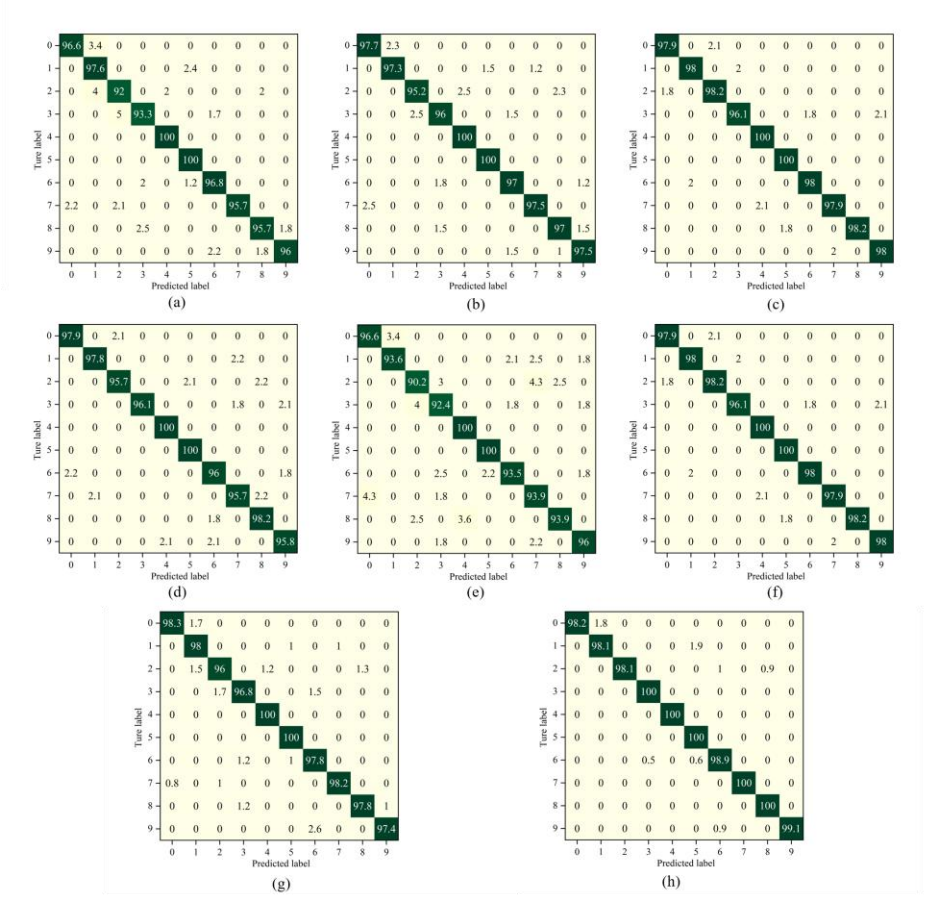


Figure 18. Confusion matrix results of CWRU dataset classification. (a) CNN, (b) WDCNN, (c) MS-CNN, (d) Attention-CNN, (e) ResNet-18, (f) MSCNN-ECM, (g) AMCNN, and (h) MSADC-TDA.

To verify the anti-noise robustness of the proposed method on the CWRU dataset, consistent with the experiments in Section 4.1.5, the signal-to-noise ratios are set to −5 dB, −3 dB, and −1 dB, respectively. The classification accuracy in each experiment is averaged over five independent tests, as shown in Figure 19.

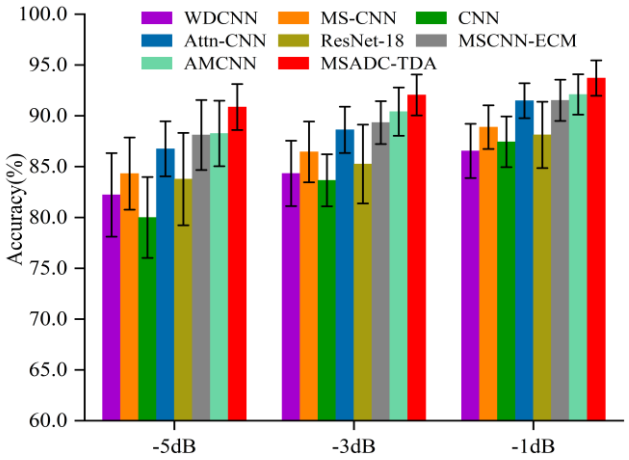


Figure 19. Diagnostic accuracy of different diagnostic methods in the CWRU dataset under noisy environments.

The experimental results illustrate that the proposed MSADC-TDA method still maintains a significant performance advantage even under strong noise interference. At a harsh noise level of −5 dB, the diagnostic accuracy of the model remains as high as 93.71%, representing a maximum accuracy improvement of 7.17% and a minimum improvement of 1.61% compared with other comparison methods. This outcome fully demonstrates that the proposed MSADC-TDA method can effectively extract fault-related features from the public dataset in strong noise environments, with superior generalization ability and anti-interference robustness. Therefore, it is well adaptable to the bearing fault diagnosis requirements in various industrial scenarios.

4.4. Qualitative and quantitative comparison between the proposed scheme and its traditional counterpart scheme

To demonstrate the advancement of the proposed MSADC-TDA, a comprehensive qualitative and quantitative comparison is performed between the proposed method and its conventional counterpart. The primary conventional baseline is established as a typical fault diagnosis framework, which maintains the exact same network depth and residual structure as the proposed MSADC-TDA. Specifically, it retains the parallel multi-scale convolutional branches with kernel sizes of 3×1, 7×1, and 9×1, but removes the adaptive dynamic weight calibration mechanism embedded in the MSADC module, and adopts conventional fixed-weight feature fusion via direct element-wise addition. This baseline is free of any adaptive multi-scale mechanism or attention enhancement strategy, representing the mainstream conventional design in the field of intelligent bearing fault diagnosis. As a supplementary comparison scheme, another widely adopted mainstream model, namely the traditional single-scale CNN, is also included in the comparison, in which the MSADC multi-scale adaptive

convolution module is replaced by a conventional fixed-scale convolution layer with a unified 3×1 kernel.

The conventional scheme relies on fixed convolution kernels and passive feature extraction, which makes it difficult to capture weak and multi-scale fault features under strong noise. They lack the ability to suppress background interference and focus on distinguishing fault information, resulting in limited performance in complex industrial scenarios. In contrast, the proposed MSADC-TDA breaks through the limitations of conventional design by using adaptive multi-scale dynamic convolution and three-domain attention mechanism. It can actively perceive and enhance critical fault features, showing stronger robustness, better anti-noise ability, and higher reliability for practical bearing fault diagnosis tasks. Quantitative comparisons are conducted on the self-built mixed fault dataset, strong noise environment, and public CWRU dataset under identical experimental settings, as shown in Table 11.

Table 11. Comparison between the proposed solution and its traditional counterpart.

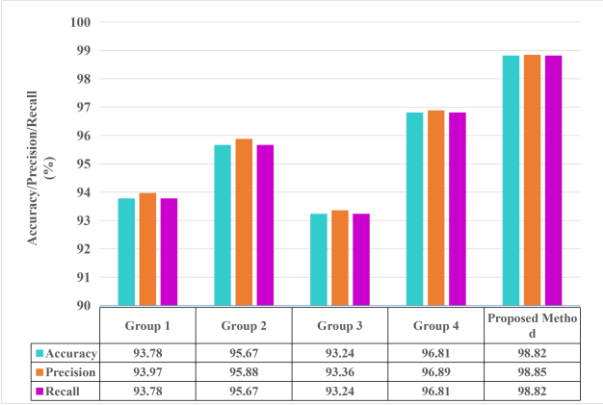
Model	Mixed fault accuracy(%)	SNR= -5dB accuracy (%)	CWRU accuracy (%)
Traditional single scale	92.21	81.33	94.88
Traditional multi scale	93.78	83.21	97.01
MSADC-TDA	98.82	88.45	99.22

The proposed MSADC-TDA achieves the optimal diagnostic accuracy across all test scenarios, with an improvement of 5.04%–6.61% over the two conventional counterparts under mixed-fault conditions. It also delivers significantly higher classification accuracy than conventional methods in strong noise environments and on the public CWRU bearing dataset. The experimental results adequately demonstrate that, benefiting from multi-scale adaptive dynamic feature extraction and the three-domain attention mechanism, the proposed method presents superior feature representation capability, anti-interference robustness and generalization performance in complex industrial bearing fault diagnosis tasks.

5. Module ablation analysis

To quantitatively assess the individual contributions of the three core innovative modules in the model proposed in this study—the multi-scale residual block, the adaptive residual mechanism, and the tri-domain attention mechanism—we designed an ablation study using different model configurations. The experiments were organized into four groups: Group 1, replacing the MSADC module with a conventional multi-scale residual network; Group 2, substituting the TDA module with a single attention module; Group 3, removing the MSADC module; and Group 4, eliminating the TDA module entirely.

The influence of each structural variation on fault diagnosis performance was quantitatively evaluated using accuracy, precision, and recall. Comparative experiments were carried out between the four ablated models and the proposed method on a collected mixed-fault bearing dataset. All results represent the average of 10 independent runs, as shown in Figure 20.



	Group 1	Group 2	Group 3	Group 4	Proposed Method
Accuracy	93.78	95.67	93.24	96.81	98.82
Precision	93.97	95.88	93.36	96.89	98.85
Recall	93.78	95.67	93.24	96.81	98.82

Figure 20. Ablation comparison experiment.

Ablation studies confirmed that eliminating any of the modules resulted in decreased performance across all metrics compared to the intact model.

The TDA mechanism is instrumental in boosting model performance. Its replacement with a single-attention mechanism (Group 2) or complete removal (Group 4) results in accuracy reductions of 3.15% and 2.01%, respectively, underscoring the critical contribution of its unique multi-domain feature fusion.

The MSADC module yields the most substantial performance gain. Direct removal of this module (Group 3) leads to a marked 5.58% drop in model accuracy, demonstrating the high effectiveness of the synergistic integration between multi-scale feature learning and adaptive weight allocation.

To evaluate the independent contribution of the adaptive mechanism, we compared the ablation results between Group 1 and Group 3. Replacing the MSADC module with a conventional multi-scale residual network led to a performance drop of 5.04%, while its complete removal resulted in a 5.58% decrease—both reductions being highly comparable. This result clearly shows that the adaptive mechanism is the key to achieving significant performance gains in our model, confirming the essential role of its dynamic kernel-weighting strategy in extracting discriminative multi-scale fault features. Furthermore, the proposed model exhibited consistently superior performance over all ablated variants in terms of both precision and recall.

The experimental results further demonstrate that each module incorporated in the model is instrumental in the fault diagnosis process, with the multi-scale adaptive dynamic convolution module having the most substantial impact on the overall model. This result demonstrates the critical role of our multi-scale adaptive dynamic convolution and three-domain attention modules in achieving the performance gains.

6. Conclusion

This paper presents MSADC-TDA, a novel method for bearing fault diagnosis under strong noise conditions. First, this method converts one-dimensional vibration signal into a two-dimensional CWT diagram to obtain more comprehensive fault information, so as to better distinguish different fault types. Second, the powerful adaptive feature extraction capability of multi-scale adaptive residual networks is combined with three-domain attention. To overcome the inherent limitation of fixed-weight convolutional layers in multi-scale architectures, a channel attention mechanism is incorporated. This module is designed to dynamically recalibrate and assign optimal weights across different

convolutional layers. Additionally, a three-domain attention module is proposed. It effectively mitigates the limitation of incomplete perception in traditional single-dimensional attention, particularly in complex fault diagnosis tasks. Moreover, it addresses the challenge of identifying essential fault features from multi-dimensional weak information under strong interference conditions. A dropout layer is added to the network to eliminate redundant information, prevent overfitting, and enhance the model's generalization ability. The proposed method exhibits high efficiency in diagnosing both single and mixed faults, while maintaining superior performance and strong robustness against interference in high-noise environments. Furthermore, it demonstrates excellent generalization capability on public datasets. The specific results are as follows:

1. The MSADC-TDA method---combining multi-scale adaptive residuals with three-domain attention---was validated on single fault datasets, achieving an average test set accuracy of 99.38%. Moreover, in the bearing data set experiment under mixed fault conditions, the average accuracy rate reached 98.82%.

2. In the bearing mixed fault experiment, the classification outcomes of nine fault types are visualized via a confusion matrix. MSADC-TDA achieves the lowest misclassification rate, with errors being only marginal. These results confirm the superior performance of MSADC-TDA in addressing mixed-fault scenarios and its efficacy in accurately distinguishing each type of mixed fault.

3. Verification on the public CWRU bearing dataset demonstrates that the proposed method still achieves 99.22% accuracy, proving its strong cross-dataset generalization ability.

4. Under operating conditions where vibration signals contain strong noise, the model can still maintain high accuracy, effectively reducing the impact of noise on classification results.

The self-built laboratory dataset used in this study covers complex scenarios close to industrial realities, including compound faults and strong noise. It provides a reliable platform for verifying the robustness of the diagnostic method. Future research will focus on multi-source verification of the proposed system. We plan to not only evaluate the model's generalization capability on more public datasets, but also validate its performance using bearing data from real industrial scenarios such as automobile manufacturing workshops and aerospace equipment. This will verify its applicability and robustness under more complex conditions, including varying operating conditions and loads.

From an industrial optimization perspective, MSADC-TDA is not merely a diagnostic tool, but a core component of intelligent maintenance. Its high robustness and accuracy provide a solid technical guarantee for the dynamic optimization of maintenance scheduling, the minimization of downtime, and the maximization of operational efficiency. Future work will focus on industrial deployment cases and quantify the economic benefits of the proposed method in real-world scenarios.

Author contributions

Xiaobo Huang: conceptualization, methodology, data curation, writing–original draft, visualization, review & editing.

Zhansi Jiang: conceptualization, validation, writing–review and editing, project administration, funding acquisition.

Yulong Tan: methodology, technical validation, writing-original draft.

Peixin Chen: formal analysis, writing-original draft.

Hui Jiang: resources, writing–review and editing, supervision, project administration, funding acquisition.

Acknowledgments

This work was supported in part by the Offshore Wind Power Joint Project of Guangdong Province Basic and Applied Basic Research Fund (2024A1515240048), in part by the Special Projects in Key Fields of Higher Education Institutions in Guangdong Province (2024ZDZX3037), in part by the Tertiary Education Scientific research project of Guangzhou Municipal Education Bureau (2024312474), and in part by the Innovation Project of GUET Graduate Education (2026YCXS028).

Use of Generative-AI tools declaration

The authors declare that AI tools are only used to polish sentences, correct grammatical errors and improve linguistic clarity in this manuscript. All research content is original, and the authors take full responsibility for the final published article.

Conflict of interest

The authors declare that there are no conflicts of interest related to this manuscript.

References

1. M. Hakim, A. A. B. Omran, A N. Ahmed, M. Al-Waily, A. Abdellatif, A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations, *Ain Shams Eng. J.*, **14** (2023), 101945. https://doi.org/10.1016/j.asej.2022.101945
2. Y. Liu, H. Xiang, Z, S. Jiang, J. W. Xiang, Second-order transient-extracting S transform for fault feature extraction in rolling bearings, *Reliab. Eng. Syst. Safe.*, **230** (2023), 108955. https://doi.org/10.1016/j.ress.2022.108955
3. Y. Liu, H. Xiang, Z, S. Jiang, J. W. Xiang, Iterative synchrosqueezing-based general linear chirplet transform for time-frequency feature extraction, *IEEE Trans. Instrum. Meas.*, **72** (2023), 1–11. https://doi.org/10.1109/TIM.2022.3232090
4. J. X. Gao, R. Pan, Y. P. Yuan, J. X. Zhou, H. Y. Ding, W. Y. Kong, A novel multiaxial fatigue life prediction method based on multi-scale physical neural network, *Reliab. Eng. Syst. Saf.*, **269** (2026), 112049. https://doi.org/10.1016/j.ress.2025.112049
5. R. H. Wang, S. B. Sun, P. C. Zhao, X. G. Yang, X. J. Wei, C. Y. Hu, Multi-scale fusion network using time-division fourier transform for rolling bearing fault diagnosis, *Comput. Mater. Con.*, **84** (2025), 3519–3539. https://doi.org/10.32604/cmc.2025.066212
6. P. W. Tse, Y. A. Peng, R. Yam, Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities, *J. Vib. Acoust.*, **123** (2001), 303–310. https://doi.org/10.1115/1.1379745
7. J. Gu, Y. X. Peng, H. Lu, X. D. Chang, G. A. Chen, A novel fault diagnosis method of rotating machinery via VMD, CWT and improved CNN, *Measurement*, **200** (2022), 111635. https://doi.org/10.1016/j.measurement.2022.111635

8. Y. G. Lei, J. Lin, Z. G. He, M. J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, *Mech. Syst. Signal Process*., **35** (2013), 108–126. https://doi.org/10.1016/j.ymssp.2012.09.015
9. M. R. Sethi, A. B. Subba, M. Faisal, S. Sahoo, D. K. Raju, Fault diagnosis of wind turbine blades with continuous wavelet transform based deep learning model using vibration signal, *Eng. Appl. Artif. Intell.*, **138** (2024), 109372. https://doi.org/10.1016/j.engappai.2024.109372
10. X. Yu, Z. T. Liang, Y. J. Wang, H. S. Yin, X. W. Liu, W. L. Yu, et al., A wavelet packet transform-based deep feature transfer learningmethod for bearing fault diagnosis under different working conditions, *Measurement*, **201** (2022), 111597. https://doi.org/10.1016/j.measurement.2022.111597
11. J. L. Chen, Y. Y. Zi, Z. J. He, J. Yuan, Compound faults detection of rotating machinery using improved adaptive redundant lifting multiwavelet, *Mech. Syst. Signal Process*., **38** (2013), 36–54. https://doi.org/10.1016/j.ymssp.2012.06.025
12. J. Y. Jiao, M. Zhao, J. Lin, K. Liang, A comprehensive review on convolutional neural network in machine fault diagnosis, *Neurocomputing*, **417** (2020), 36–63. https://doi.org/10.1016/j.neucom.2020.07.088
13. C. Liang, X. L. Mu, X. G. Zhang, F. F. Lu, C. C. Wang, Y. B. Shao, Enhanced fault diagnosis of rolling bearings using attention-augmented separable residual networks, *Eng. Sci. Technol. Int. J.*, **61** (2025), 101930. https://doi.org/10.1016/j.jestch.2024.101930
14. Y. N. Yan, Z. K. Liu, J. W. Xu, H. Zhang, N. Guo, L. Zhao, et al., A temperature-decoupled impedance-based mass sensing using CBAM-CNN and adaptive weighted average preprocessing with high accuracy, *Mech. Syst. Signal Process*., **213** (2024), 111347. https://doi.org/10.1016/j.ymssp.2024.111347
15. S. Y. Han, H. J. Choi, S. K. Choi, J. S. Oh, Fault diagnosis of planetary gear carrier packs: a class imbalance and multiclass classification problem, *Int. J. Precis. Eng. Manuf.*, **20** (2019), 167–179. https://doi.org/10.1007/s12541-019-00082-4
16. J. Q. Cui, J. W. Gao, R. X. Xing, W. K. Wu, M. Y. Li, Y. P.Yang, Fault Diagnosis Method for Rolling Bearing Based on Attention Mechanism and BiTCN model, *Digit. Signal Process*., **167** (2025), 105454. https://doi.org/10.1016/j.dsp.2025.105454
17. D. W. Hong, B. Kim, 1D convolutional neural network-based adaptive algorithm structure with system fault diagnosis and signal feature extraction for noise and vibration enhancement in mechanical systems, *Mech. Syst. Signal Process*., **197** (2023), 110395. https://doi.org/10.1016/j.ymssp.2023.110395
18. X. A. Yan, W. J. Yan, Y. D. Xu, K. V. Yuen, Machinery multi-sensor fault diagnosis based on adaptive multivariate feature mode decomposition and multi-attention fusion residual convolutional neural network, *Mech. Syst. Signal Process*., **202** (2023), 110664. https://doi.org/10.1016/j.ymssp.2023.110664
19. Q. Z. Wei, X. C. Tian, L. Cui, F. Q. Zheng, L. D. Liu, WSAFormer-DFFN: A model for rotating machinery fault diagnosis using 1D window-based multi-head self-attention and deep feature fusion network, *Eng. Appl. artif. intell*., **124** (2023), 106633. https://doi.org/10.1016/j.engappai.2023.106633
20. Y. D. Xu, X. A. Yan, K. Feng, X. Sheng, B. B. Sun, Z. Liu, Attention-based multiscale denoising residual convolutional neural networks for fault diagnosis of rotating machinery, *Reliab. Eng. Syst. Saf*., **226** (2022), 108714. https://doi.org/10.1016/j.ress.2022.108714

21. L. J. Wang, Z. J. Wang, Z. X. Wu, X. F. Yang, Y. X. Xie, Rolling bearing fault diagnosis method based on improved multi-scale residual networks and light gradient boosting machine, *J. Braz. Soc. Mech. Sci. Eng*., **48** (2026), 104. https://doi.org/10.1007/s40430-025-06071-1
22. W. Huang, J. Cheng, Y. Yang, G. Guo, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, *Neurocomputing*, **359** (2019), 77–92. https://doi.org/10.1016/j.neucom.2019.05.052
23. Q. X. Zhu, Y. S. Qian, N. Zhang, Y. L. He, Y. Xu, Multi-scale Transformer-CNN domain adaptation network for complex processes fault diagnosis, *J. Process Control*., **130** (2023), 103069. https://doi.org/10.1016/j.jprocont.2023.103069
24. Q. Song, P. Jiang, A multi-scale convolutional neural network based fault diagnosis model for complex chemical processes, *Process Saf. Environ. Prot*., **159** (2022), 575–584. https://doi.org/10.1016/j.psep.2021.11.020
25. Z. D. Hei, H. Y. Yang, W. F. Sun, M. P. Zhong, G. H. Wang, A. Kumar, et al., Multiscale conditional adversarial networks based domain-adaptive method for rotating machinery fault diagnosis under variable working conditions, *ISA Trans*., **154** (2024), 352–370. https://doi.org/10.1016/j.isatra.2024.08.027
26. K. Liu, N. Y. Lu, F. Wu, R. D. Zhang, F. R. Gao, Model fusion and multiscale feature learning for fault diagnosis of industrial processes, *IEEE Trans. Cybern*., **53** (2022), 6465–6478. https://doi.org/10.1109/TCYB.2022.3176475
27. Y. W. Sun, H. F. Tao, V. Stojanovic, Autoregressive data generation method based on wavelet packet transform and cascaded stochastic quantization for bearing fault diagnosis under unbalanced samples, *Eng. Appl. Artif. Intell*., **138** (2024), 109402. https://doi.org/10.1016/j.engappai.2024.109402
28. V. Djordjevic, L. Dubonjic, M. M. Morato, D. Prsic, V. Stojanovic. Sensor fault estimation for hydraulic servo actuator based on sliding mode observer, *Math. Model. Control*., **2** (2022), 34–43. https://doi.org/10.3934/MMC.2022005
29. Y. W. Sun, H. F. Tao, Y. Z. Ni, V. Stojanovic, A generic single-source domain generalization framework for fault diagnosis via wavelet packet augmentation and pseudo-domain generation, *IEEE Internet Things J*., **12** (2025), 31629–31642. https://doi.org/10.1109/JIOT.2025.3573752
30. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, 770–778. https://doi.org/10.48550/arXiv.1512.03385
31. G. C. Nie, Z. W. Zhang, Z. H. Jiao, Y. J. Li, M. Shao, X. Dai, A novel intelligent bearing fault diagnosis method based on image enhancement and improved convolutional neural network, *Measurement*, **242** (2025), 116148. https://doi.org/10.1016/j.measurement.2024.116148
32. W. Zhang, G. L. Peng, C. H. Li, Y. H. Chen, Z. J. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors*, **17** (2017), 425. https://doi.org/10.3390/s17020425
33. B. Q. Hu, J. Liu, Y. Xu, A novel multi-scale convolutional neural network incorporating multiple attention mechanisms for bearing fault diagnosis, *Measurement*, **242** (2025), 115927. https://doi.org/10.1016/j.measurement.2024.115927
34. L. Y. Wang, W. G. Zhao, An ensemble deep learning network based on 2D convolutional neural network and 1D LSTM with self-attention for bearing fault diagnosis, *Appl. Soft Comput*., **172** (2025), 112889. https://doi.org/10.1016/j.asoc.2025.112889

35. Q. M. Shen, Z. Q. Zhang, Fault diagnosis method for bearing based on attention mechanism and multi-scale convolutional neural network, *IEEE Access*., **12** (2024), 12940–12952. https://doi.org/10.1109/ACCESS.2024.3357113
36. R. S. Qin, J. S. Zhao, Adaptive multiscale convolutional neural network model for chemical process fault diagnosis, *Chin. J. Chem. Eng*., **50** (2022), 398–411. https://doi.org/10.1016/j.cjche.2022.10.001

AIMS Press