Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 13
AMMSAC 2026

Rolling Bearing Fault Diagnosis Based on Multi-Domain Feature Fusion and CBAM-Enhanced Residual Dense Network

Jiaqi Liu †, Kerui Liu †, Shunyuan Hou, Wencheng Ouyang and Tianxing Wang

Lanzhou Jiaotong University, Lanzhou, China

† These authors also contributed equally to this work

Abstract. Based on rolling bearing vibration signals, this paper proposes an intelligent fault diagnosis model that integrates multi-domain features with attention mechanisms. First, by developing a multi-domain feature fusion strategy, time-domain signals, envelope spectra, and Discrete Cosine S-Transform (DCST) features are combined to form multi-channel inputs, effectively enhancing the information-carrying capacity of the original signals. Second, a deep one-dimensional network architecture combining residual and dense connections is designed to mitigate network degradation while enabling full reuse and efficient transmission of multi-layer features. Building on this foundation, the Convolutional Block Attention Mechanism (CBAM) is introduced to adaptively weight features across both channel and spatial dimensions, thereby enhancing the expression of key failure features and improving the model's ability to identify weak features under complex operating conditions. Experimental results demonstrate that this method exhibits excellent accuracy and stability in multi-class bearing fault recognition tasks, achieving an average recognition rate of 99.19%, which significantly outperforms traditional convolutional networks and classical deep learning models. The proposed method possesses strong generalization capabilities and engineering application potential, offering an efficient and reliable solution for the fault diagnosis of rotating machinery.

Keywords: Rolling Bearing Fault Diagnosis; Multi-Domain Feature Fusion; Residual Dense Network.

1. Introduction

As industrial equipment becomes increasingly complex, rolling bearings—as critical components—play a vital role, with their operational status directly impacting system safety and stability. However, under actual operating conditions, bearing fault signals often exhibit characteristics such as non-stationarity and weak signals. Traditional methods that rely on single-signal analysis or manual feature extraction struggle to accurately capture complex vibration information, resulting in limited diagnostic performance. In recent years, although deep learning methods have made progress in fault diagnosis, most models still rely primarily on single-domain time-domain inputs, resulting in limited feature representation capabilities. Additionally, issues such as information decay and the weakening of key features within deep neural networks have hindered further improvements in model performance [1].

To address these shortcomings, this paper proposes an improved fault diagnosis model framework centered on multi-domain information fusion and feature enhancement. First, multi-channel inputs are constructed by fusing time-domain, envelope spectrum, and improved time-frequency features to enhance signal representational capacity [2]. Second, a deep network architecture combining residual and dense connections is designed to improve feature propagation and reuse efficiency. Furthermore, an attention mechanism is introduced to adaptively highlight key features, thereby enhancing the model's ability to recognize complex operating conditions [3].

Using rolling bearing vibration data as an example, the proposed method demonstrates good accuracy and stability in multi-class classification tasks, reflecting strong generalization capabilities and application potential. By optimizing both feature construction and model architecture, this paper provides an effective approach for intelligent fault diagnosis.

2. Fault Diagnosis Method Based on Multi-Domain Feature Fusion and Attention-Enhanced Network

2.1. Data Acquisition

The data used in this study come from the publicly available bearing fault dataset provided by Case Western Reserve University (CWRU). The experimental platform is a dual-motor drive system, where power is transmitted through a coupling, and a multi-dimensional vibration acquisition setup is installed. Three high-performance accelerometers are mounted in the system, located at the outer circumference of the drive-end bearing, the outer circumference of the non-drive-end bearing, and the motor base, respectively, to synchronously collect vibration signals from different positions. The sampling frequency of the accelerometer at the drive end is fixed at 12 kHz, which allows the simulation and analysis of vibration characteristics of rolling bearings under four typical conditions: normal condition, rolling element damage, inner race defect, and outer race fault.

For localized defects, three preset defect sizes are introduced on the surfaces of the rolling element, inner race, and outer race, respectively: 0.1778 mm, 0.3556 mm, and 0.5334 mm. These combinations generate test samples with different fault severities, ensuring that the dataset remains both scientifically valid and reliable. The analysis focuses on the vibration signals from the drive-end bearing, with particular attention to the characteristics of inner race defects, while also examining the variation of vibration responses under different operating conditions.

In the experimental design, three defect sizes are combined with three motor loads (0 W, 735 W, 1470 W), forming nine typical inner race fault conditions. In addition, a normal operating condition without faults is included as a reference, resulting in a dataset with twelve fault categories (indexed as 0–11). The specific grouping is shown in Table 1.

Table 1. Dataset identification states and classification labels

Load	Inner Race			Normal
	0.1778	0.3556	0.5334	
0	0	3	6	9
735	1	4	7	10
1470	2	5	8	11

Each fault category contains 100 samples. The dataset is divided into training, validation, and test sets in a ratio of 8:1:1. Each sample consists of 1024 raw time-series points, which ensures both balance and representativeness of the classification dataset.

2.2. Multi-Domain Feature Extraction

The original vibration signals of rolling bearings contain only time-domain information. If these signals are directly used as model input, the amount of information may be insufficient. To address this, a multi-domain signal fusion strategy is introduced, where time-domain, frequency-domain, and time–frequency domain signals are combined as model inputs, allowing the network to learn more features.

2.2.1. Envelope Spectrum Feature Extraction.

When a bearing develops localized damage, it excites vibration responses at specific frequencies. For this reason, spectral analysis has long been one of the most important tools in processing vibration signals of rotating machinery. The envelope spectrum technique identifies the fault location in the bearing based on peak frequencies, and it can also demodulate high-frequency resonance signals into the low-frequency region, making it easier to obtain clearer frequency resolution.

A Hilbert transform-based method is used to compute the envelope spectrum, mainly in two steps:

(1) Hilbert Transform

Apply the Hilbert transform to the signal $x(t)$ to obtain $h(t)$:

$$h(t) = H[x(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{1}$$

Then construct the analytic signal, whose magnitude is the envelope signal:

$$x_{ht}(t) = |x(t) + jh(t)| = \sqrt{x^2(t) + h^2(t)} \tag{2}$$

(2) Fourier Transform for Envelope Spectrum

Perform the Fourier transform on the envelope signal and take the amplitude of each frequency component to obtain the envelope spectrum:

$$E(f) = \mathrm{F}\ [x_{ht}(t)] \tag{3}$$

In the envelope spectrum, the amplitude at different frequencies reflects the severity of bearing damage, while the frequency corresponding to peaks indicates the fault location. Based on this relationship, various features can be extracted to distinguish different types of bearing faults.

2.2.2. S-Transform Feature Extraction.

The S-transform combines the advantages of the short-time Fourier transform and wavelet transform. It adaptively adjusts the window function and decomposes the signal in both time and frequency domains, showing strong time–frequency concentration and multi-scale analysis capability. However, the conventional S-transform produces the same amount of data for both low-frequency and high-frequency components, which leads to some redundancy. To address this issue, the discrete orthogonal S-transform (DOST) is adopted to achieve a non-redundant multi-resolution representation while reducing computational cost:

$$DOST = \sum_{i=1}^{k} D_i \mathrm{DFT} \tag{4}$$

By replacing the discrete Fourier transform (DFT) with the discrete cosine transform (DCT), the discrete cosine S-transform (DCST) is obtained:

$$DCST = \sum_{i=1}^{k} DCT_{ni}^{-1} DCT \tag{5}$$

Through optimized partitioning of the frequency space, a signal sequence of length 2^N is divided into different frequency bands: $n_1 = 1,\quad n_i = 2^{i-2},\quad 2 \le i \le N-1$.

DCST is efficient in both computation and memory usage. It reconstructs the time-domain energy of bearing signals effectively in the frequency domain and distinguishes between low-frequency and high-frequency characteristics, making multi-frequency component extraction more efficient and supporting fault diagnosis. Compared with traditional methods, this approach shows strong potential in industrial applications.

2.3. Fault Feature Learning

As the network becomes deeper, the performance of traditional convolutional neural networks in feature extraction often degrades. In the network design, residual modules and skip connections are introduced, where the input and deep features are fused through element-wise addition. This helps control the growth of parameters, while also improving training speed and prediction accuracy.

2.3.1. Feature Extraction with Residual Dense Blocks.

To learn useful fault features from multi-channel signals, the core of the model is a deep one-dimensional network, where each layer progressively abstracts signal features. Simply stacking residual blocks may lead to the loss of some useful features during transmission [4]. To address this, residual modules are combined using dense connections to form a residual dense structure (Fig. 1).

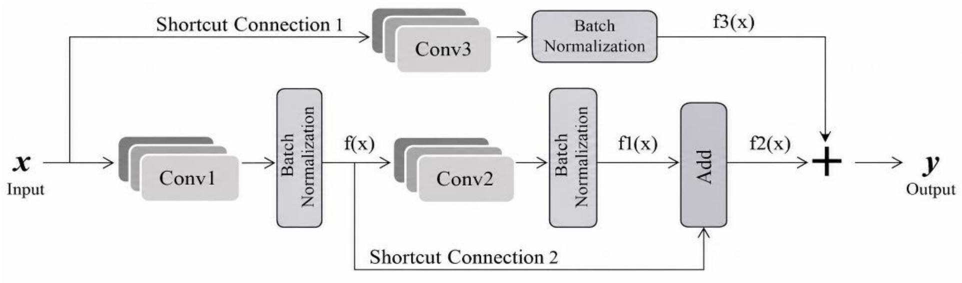

Fig. 1 Schematic diagram of the residual dense structure

In this structure, low-level features $f(x)$ and high-level features $f_1(x)$ are arranged in parallel. Through the linear operation of convolution 3, $f_3(x)$ is generated. Then $f_2(x)$ and $f_3(x)$ are concatenated to form a residual block. The network output can be expressed as:

$$x_1 = \sigma_r\left(w_1 * x + b_1\right),\ x_2 = \sigma_r\left(w_2 * x_1 + b_2\right),\ x_3 = \sigma_r\left(w_3 * \mathrm{Concat}\left(x_1, x_2\right) + b_3\right) \tag{6}$$

$$y = x + x_3 \tag{7}$$

where $w_1, b_1, w_2, b_2, w_3, b_3$ are the weights and biases of the three convolutional layers, $Concat(\cdot)$ denotes channel-wise feature concatenation and σ_r denotes the ReLU activation function:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{8}$$

2.3.2. Channel-Wise Feature Importance.

The channel attention mechanism reflects the importance of features across different channels and applies corresponding weights. The residual dense module includes convolutional layers, batch normalization, nonlinear activation, and shortcut connections, which improve information flow and alleviate gradient vanishing in deep networks. By reusing convolution results from previous layers, redundant learning of irrelevant features is reduced, and more vibration details are preserved for subsequent layers.

2.3.3. Attention Mechanism for Optimized Convolution.

To improve classification effectiveness, a convolutional block attention module (CBAM) is introduced into the residual dense network [5][6]. It combines channel attention and spatial attention in a sequential manner to reassign feature weights and achieve finer feature adjustment (Fig. 2).

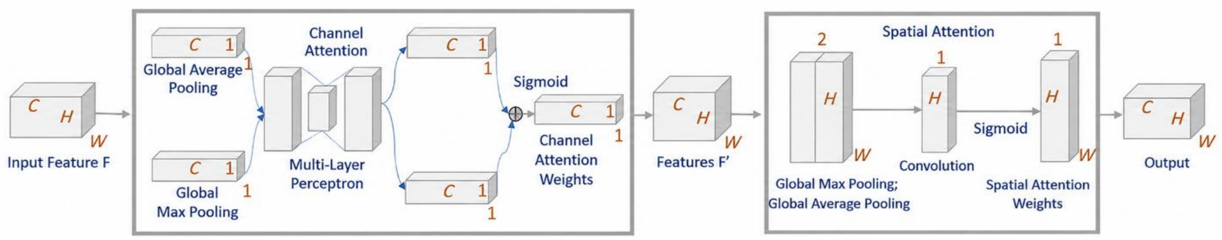

Fig. 2 Structure of CBAM

Given the input feature tensor F, global average pooling and global max pooling are first applied to generate two spatial context descriptors F_{avg}^{c} and F_{max}^{c}. These are then fed into a shared network for dimensionality reduction and expansion. The resulting feature maps are summed element-wise, followed by a Sigmoid activation to obtain the channel attention weights $M_c(F)$:

$$M_c(F)=\sigma\left\{W_1\left[W_0\left(F_{avg}^{c}\right)\right]+W_1\left[W_0\left(F_{\max}^{c}\right)\right]\right\} \tag{9}$$

where $W_0 \in \mathbb{R}^{c/r\times c}$ and $W_1 \in \mathbb{R}^{c\times c/r}$, r is the reduction ratio, and σ denotes the Sigmoid function (denoted as f_c in this paper).

2.3.4. Spatial Feature Importance.

The output of channel attention F' is used as the input for spatial attention. Average pooling and max pooling are applied to generate F'_{avg} and $F'_{\max}$, which are combined into a two-channel feature map. A convolution with stride 2 and kernel size 1×7 is then used to compress the channel dimension. Finally, a Sigmoid mapping produces the spatial attention weights $M_s(F')$:

$$M_s\left(F'\right)=\sigma\left\{f^{1\times 7}\left\{\left[\mathrm{Avgpool}\left(F'\right);\mathrm{Maxpool}\left(F'\right)\right]\right\}\right\}=\sigma\left\{f^{1\times 7}\left(F_{\mathrm{avg}}^{s};F_{\max}^{s}\right)\right\} \tag{10}$$

The final spatial attention map is obtained by element-wise multiplication of $M_s(F')$ and F'.

By separating and evaluating feature importance in both channel and spatial dimensions, information can be extracted more effectively, and the learning efficiency of the model is significantly improved.

2.4. Diagnosis Method with Attention-Enhanced Network

Traditional deep convolutional neural networks may be limited in feature extraction, especially when the network becomes very deep, where key features can be overlooked. To address this, a bearing fault diagnosis method is proposed that combines a residual dense architecture with an attention mechanism. The network mainly consists of a multi-channel feature preprocessing module and three levels of nested residual dense units. With this design, rich fault features can be learned from multi-channel signals, while different features are weighted accordingly.

The overall process is as follows. First, the original vibration signals are transformed into multiple domains (time domain, envelope spectrum, and DCST). The transformed signals are then combined with the original signal to form multi-channel inputs. Next, deep features are adaptively extracted through stacked residual dense blocks, and the key features are further enhanced using the CBAM module. Finally, a softmax classifier is used to output the fault category probabilities for each sample.

Let the combined signal be x_i. Feature extraction is performed using a convolution with stride 1 and kernel size $1*3$, which can be expressed as:

$$x_i^{inp}=f(x_i;\theta_{inp})=\sigma_r(x_i*k_{inp}+b_{inp}) \tag{11}$$

where $\theta_{inp}=(k_{inp},b_{inp})$ represents the trainable parameters.

In the final residual dense unit, CBAM is introduced to differentiate a large number of features. Let the features extracted through multi-level residual dense units be x. After integrating CBAM, the output is:

$$x_1=\sigma_r\left(w_1*x_i+b_1\right),\ x_2=\sigma_r\left(w_2*x_1+b_2\right),\ x_3=\sigma_r\left(w_3*\mathrm{Concat}\left(x_1,x_2\right)+b_3\right),\ x_i^4=f_s\left(f_c\left(x_3\right)\right) \tag{12}$$

where (w_1, b_1), (w_2, b_2), and (w_3, b_3) are the trainable parameters of convolution 1, convolution 2, and convolution 3, respectively.

To reduce the number of parameters, the features x_i^1 from each residual dense unit and the CBAM-enhanced features x_i^4 are divided into non-overlapping segments, and max pooling is applied to each segment:

$$v_i^{l,m,n} = \max_{s(m-1)+1 \le j \le sm} x_{(i),j,n}^{l}, \quad v_i^{l+1,m,n} = \max_{s(m-1)+1 \le j \le sm} x_{(i),j,n}^{4} \tag{13}$$

where s is the segment length, set to 2. In this way, the most significant feature response in each segment is preserved, while redundant information is reduced.

2.5. Fault Classification

During deep network training, weights and biases are dynamically updated through backpropagation. This process optimizes layer parameters and allows error signals to propagate through the network, helping improve the model. The cross-entropy loss function is used to measure the difference between the predicted probability distribution and the true labels. In each iteration, the loss is first computed through forward propagation, and then the parameters are optimized using mini-batch stochastic gradient descent, gradually aligning the output probabilities with the true labels. This improves classification accuracy and also makes the model more stable across different batches of samples.

3. Model Performance Evaluation and Comparative Analysis

3.1. Network Parameters and Diagnostic Results

The proposed method was tested multiple times on the self-constructed dataset. The classification accuracy is measured by the ratio of correctly classified samples to the total number of test samples. The average recognition accuracy over multiple experiments reaches 99.19%.

The confusion matrix of the 12-class classification under multi-domain fusion input is shown in Fig.3. The horizontal axis represents predicted labels, and the vertical axis represents true labels. The values on the diagonal indicate the number of correctly classified samples in each class, while the off-diagonal entries represent misclassified samples. It can be observed that the model shows very strong classification performance, with few misclassification cases and a very small number of affected samples. Overall, this indicates that the constructed network can effectively distinguish different fault categories when processing multi-channel bearing vibration signals.

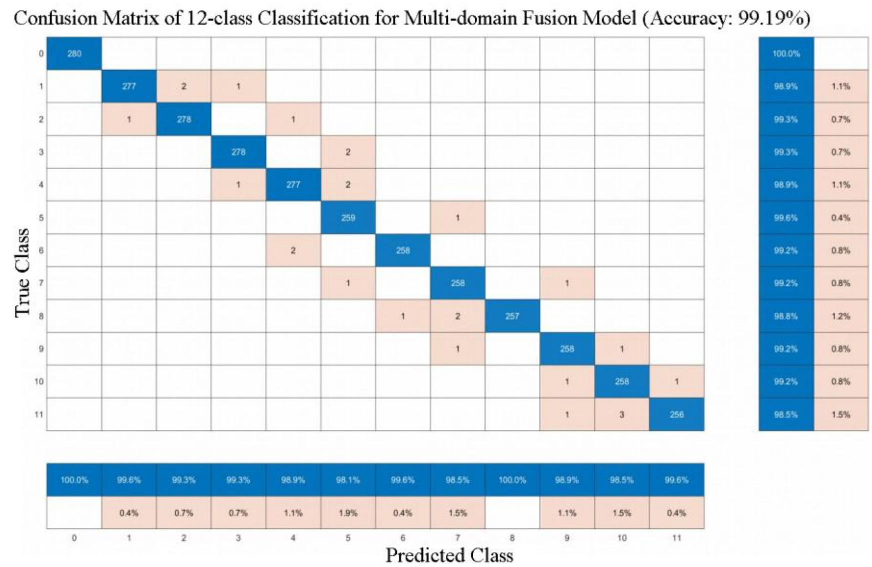

Fig. 3 Confusion matrix of one classification result

3.2. Recognition Performance and Analysis

The proposed diagnostic method mainly consists of two parts: multi-channel signal input processing and a residual dense network combined with an attention mechanism. Most existing deep learning methods for bearing fault diagnosis directly use raw time-domain vibration signals as input. In contrast, the input here is formed by combining multiple transformed domains, allowing the model to capture richer fault features. The network is composed of three residual dense units and one CBAM module to enhance the learning of important features.

3.2.1. Comparison of Different Input Data.

The impact of different input data on fault recognition performance is first examined. Fig. 4 shows the average recognition accuracy under different input settings.

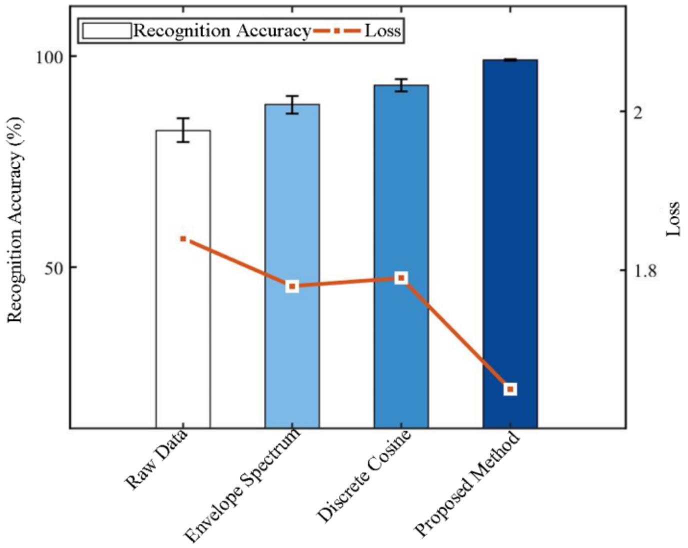

Fig. 4 Model recognition accuracy with different input data

It can be seen that, under the same network structure, multi-domain fusion input performs better than single time-domain input, achieving higher average recognition accuracy. The results from repeated experiments show only small variations, indicating good stability. The recognition results from individual experiments are also reliable, with the lowest average loss.

3.2.2. Comparison of Different Network Structures.

To evaluate the influence of CBAM placement on diagnostic performance, four network structures are compared under the same multi-channel input:

Method a (pure residual dense structure): four residual dense units stacked consecutively, without attention modules

Method b (front attention): CBAM placed at the front, arranged as one standard residual dense unit, one CBAM residual dense unit, and two standard residual units

Method c (rear attention): CBAM placed in the later part of the network, arranged as two standard residual units, one CBAM residual unit, and one standard residual unit

Method d (proposed method): CBAM embedded in the final residual dense unit, consisting of three standard residual dense units followed by one CBAM-enhanced residual dense unit.

Each structure is tested 10 times, and the result of one experiment is visualized using t-SNE, as shown in Fig. 5.

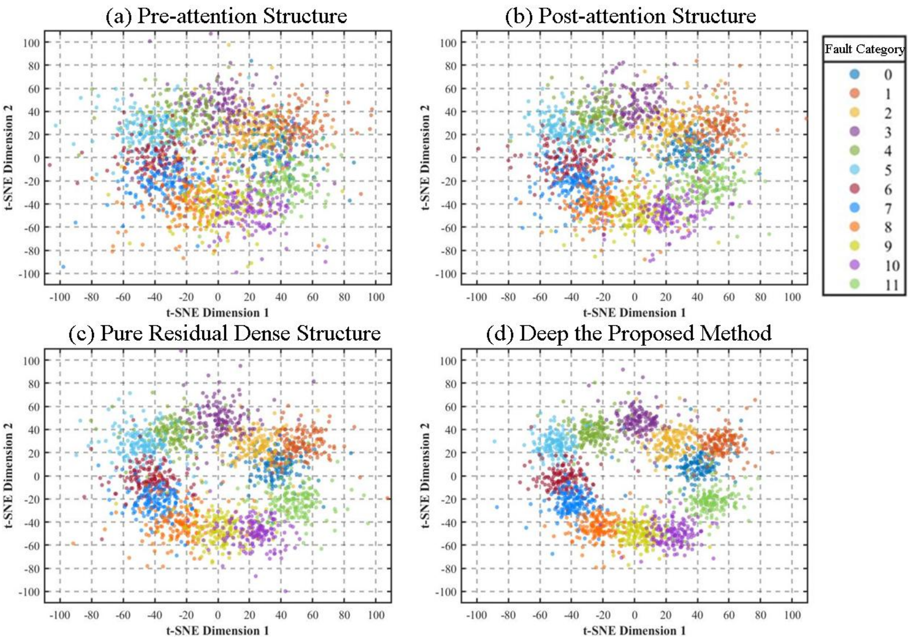

Fig. 5 t-SNE visualization of classification results (methods a–d)

From the t-SNE visualization, methods b and c show poor separation between fault categories, with clear overlap. Method a achieves relatively high accuracy, but its class separation is still weaker than method d. This indicates that embedding CBAM in the fourth residual block leads to the best classification performance. The detailed results are shown in Table 2.

Table 2. Statistical results of four models

Method	Average Accuracy /%	Standard Deviation /%	Time /s
a	96.72	0.698	1.895
b	85.38	4.629	2.193
c	91.87	3.483	2.225
d	99.19	0.324	2.463

As shown in Table 2, method d achieves the highest average accuracy and the smallest standard deviation, indicating that the model is both accurate and stable. Although the training time is slightly longer, it remains within an acceptable engineering range.

3.3. Comparison with Common Methods

To verify the effectiveness of the proposed method, it is compared with three commonly used deep learning models, including CNN, ResNet, and convolutional autoencoder (CAE). A five-fold cross-validation strategy is adopted. The dataset is randomly divided into training, validation, and test sets in a ratio of 8:1:1, and each group of experiments is repeated five times to ensure reliability.

Table 3. Statistical comparison results with common methods

Method	Average Accuracy /%	Standard Deviation /%	Loss
Proposed method	99.19	0.324	1.692
ResNet	96.11	1.649	1.697
CNN	91.08	1.362	2.112
CAE	91.31	1.531	2.062

As shown in the Table 3, the proposed method achieves a noticeably higher average accuracy than the other three models, and it also has the smallest standard deviation, indicating the best classification stability. The paired-sample t-test gives $t = 73.65, df = 4, p < 0.00001$, which shows a highly significant advantage of the model in recognition accuracy.

4. Conclusion

This paper proposes a deep learning diagnostic method that integrates multi-domain features with attention mechanisms to address the need for fault identification in complex operating conditions of rolling bearings. By focusing on both signal representation and model architecture, the method effectively extracts multiscale information from vibration signals through the collaborative design of multi-channel feature inputs and residual-dense networks. Additionally, the introduction of attention mechanisms dynamically highlights key features, thereby enhancing the model's discrimination capability in complex environments. Experimental results demonstrate that this method achieved an average accuracy of 99.19% across 12 fault classification tasks, with a low standard deviation, indicating excellent stability and reliability. Compared to traditional methods such as CNNs and ResNets, the model exhibits significant advantages in both accuracy and consistency. Overall, the proposed framework not only holds high practical value for engineering applications but also provides a methodological reference for intelligent operation and maintenance as well as industrial big data analysis. Future research could further explore the model's adaptability in cross-operating-condition transfer and few-shot learning scenarios.

References

[1] Ge Zhuo, Xia Huameng, Wang Kailiang, et al. A Deep Learning Method for Bearing Fault Diagnosis Based on Multi-domain Information Fusion [J]. Vibration and Shock, 2024, 43(23): 47-55. DOI: 10.13465/j.cnki.jvs.2024.23.006.

[2] Liu Yang. Research on a Spindle Bearing Fault Diagnosis Method Based on Multi-Source Information Fusion [D]. Xi'an University of Technology, 2024. DOI:10.27398/d.cnki.gxalu.2024.002163.

[3] Han Zhengjie, Niu Rongjun, Ma Zikui, et al. A Bearing Fault Diagnosis Method Based on an Attention Mechanism-Enhanced Residual Neural Network [J]. Vibration and Shock, 2023, 42(16): 82-91. DOI:10.13465/j.cnki.jvs.2023.16.010.

[4] Tang Zhiyao. Research on Intelligent Fault Diagnosis Algorithms for Rolling Bearings Based on Improved Convolutional Neural Networks [D]. Lanzhou Jiaotong University, 2023. DOI:10.27205/d.cnki.gltec.2023.001089.

[5] Wu Lan, Dong Lin. A Rolling Bearing Fault Diagnosis Method Based on MTF-CBAM-IResNet [J]. Manufacturing Technology and Machine Tools, 2024, (11): 16–21. DOI: 10.19287/j.mtmt.1005-2402.2024.11.002.

[6] Lü Xiaohong, Zhang Liqiang, Li Fuxiang. Research on Rolling Bearing Fault Diagnosis Based on TVFEMD-CNN-CBAM-BiLSTM [J]. Journal of Lanzhou Jiaotong University, 2026, 45(02): 30-42.