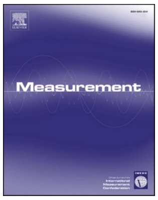
Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement



An improved dual-channel CNN-BILSTM fusion attention model for fault diagnosis of aero-engine bearings

Delin Huang [a], Xiangdong Su [a], Jinghui Yang [a,b,*], Shichang Du [c], Dexian Wang [a], Qiuyu Ran [a]

[a] *School of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, Shanghai 201209, China*
[b] *Wuyi Intelligent Manufacturing Industrial Technology Research Institute, Zhejiang 321200, China*
[c] *School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

ARTICLE INFO

Keywords:
Fault diagnosis
Aero-engine bearings
Convolutional neural network
Bidirectional long short-term memory
Feature fusion

ABSTRACT

Accurate fault diagnosis of aero-engine bearings is vital for ensuring flight safety. Existing methods still struggle with extracted features lacking multi-dimensional representation, insufficient fault information, and ineffective feature fusion under complex conditions (e.g., varying rotational speeds) and multi-source signal inputs. As such, an improved two-channel fault diagnosis model for rolling bearings is proposed, integrating a convolutional neural network and bidirectional long short-term memory (CNN–BILSTM) architecture, enhanced by multiple improved attention mechanisms.First, the raw vibration signals were directly used as time-domain inputs and processed to obtain their frequency-domain counterparts, forming a dual-channel input to the customized and optimized CNN-BILSTM feature extraction network. Then, a one-dimensional convolutional block attention module (1DECBAM) is inserted after each of the two CNNs to retain initial features while enhancing key ones critical for fault diagnosis. Moreover, the proposed Hybrid Interaction-Fusion Attention (HIFAttn) framework incorporates a Time-Frequency Interactive Attention Mechanism (T-FIAttn) and a Local-Global Adaptive Attention Module (L-GAAM) to perform multimodal feature fusion. Specifically, the T-FIAttn is employed to capture latent feature relationships across both time and frequency domains. In addition, the L-GAAM was appended after the BILSTM layers in each channel to dynamically capture essential features. Experimental results on two aero-engine datasets demonstrate that the proposed model achieves accuracies of 99.32% and 99.94%, respectively, surpassing current state-of-the-art methods.The model also demonstrates excellent stability and robustness, even under high-noise conditions.These results indicate that the proposed model achieves high accuracy and strong generalization, making it well-suited for aero-engine bearing fault diagnosis.

1. Introduction

The aero-engine serves as the core component of an aircraft, with its stable operation being essential for ensuring safe flight. In twin-rotor aero-engines, aircraft rolling bearings are critical load-bearing components, operating in harsh environments and being highly susceptible to failures that may lead to severe accidents. The operational status of bearings [1] directly impacts the overall performance of the equipment. Therefore, accurately and promptly identifying the location and type of bearing failures is significant for ensuring the safe and reliable operation of aero-engines. Traditional bearing fault diagnosis primarily relies on extracting fault features from raw vibration signals. Typically, fault features are first extracted using time–frequency signal processing techniques, including Singular Value Decomposition (SVD)[2,3], Fast Fourier Transform (FFT)[4,5], continuous wavelet transform (CWT) [3,6], and variational mode decomposition (VMD)[7–9]. These extracted features are then input into classifiers, such as Support Vector Machines (SVMs)[3,10], Extreme Learning Machine (ELM)[9,11], and Bayesian Networks (BN)[12,13], for fault diagnosis. In recent years, relatively novel approaches, such as the method proposed in [14], have introduced a small-sample fault diagnosis technique based on multi-scale perception, multi-level feature fusion, and image quadrant entropy (MPMFFIQE). This method transforms transient signals into images and extracts key fault information through feature enhancement and fusion.When integrated with the Harris-Hawk Optimized Support Vector Machine (HHOSVM), this method achieves significantly superior

* Corresponding author.
E-mail addresses: dlhuang@sspu.edu.cn (D. Huang), 20231510056@sspu.edu (X. Su), jhyang@sspu.edu.cn (J. Yang), lovbin@sjtu.edu.cn (S. Du), dxwang@sspu.edu.cn (D. Wang), 20231510016@sspu.edu.cn (Q. Ran).

https://doi.org/10.1016/j.measurement.2025.117761
Received 13 January 2025; Received in revised form 5 April 2025; Accepted 3 May 2025
Available online 4 May 2025

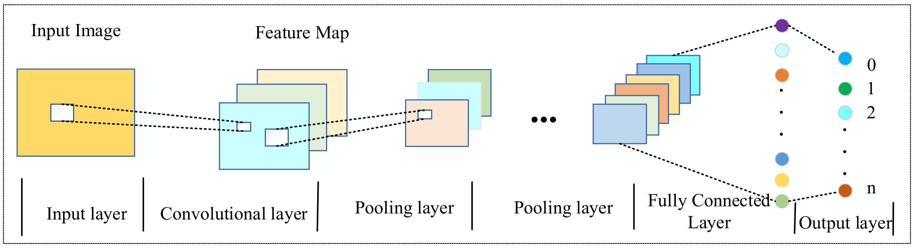


Fig. 1. Classical structural model of CNN.

experimental results compared to existing approaches.Moreover, the study in [15] introduces a fault diagnosis framework for rotating machinery based on extended time-shift multi-scale phase entropy. This approach effectively captures the equipment's operational state by generating and analyzing dynamic features across multiple scales. Experimental results demonstrate that the framework exhibits broad applicability and outstanding performance in diagnosing various types of rotating machinery.Literature [16] introduces the generalized redistribution transform (GRT), a novel time–frequency analysis (TFA) algorithm that optimizes time–frequency representation (TFR) using the time–frequency fusion extraction criterion (TFFEC) and short-time Fourier transform (STFT) to enhance resolution and suppress noise. The synchronized (SRT) and horizontal redistribution transforms (HRT) further enhance time–frequency energy aggregation and readability. Experimental results confirm GRT's effectiveness and superiority in monitoring and diagnosing rotating machinery faults.While these methods partially alleviate the limitations of manual feature selection, they continue to rely on specific signal processing and optimization steps. Furthermore, as mechanical inspection systems operate over extended periods, the volume of state data will increase, intensifying the need for models capable of adaptive feature extraction.

As data volumes and the need for fault feature extraction grow, deep learning methods have increasingly been applied to fault diagnosis for their powerful feature self-learning capabilities. Unlike traditional methods, deep learning models automatically extract features from raw signals, eliminating the tedious manual feature design process. Xu and Zhang et al. [17] introduces a rolling bearing fault diagnosis method using a 1D-Visual Transformer (1D-ViT) to achieve end-to-end diagnosis by directly inputting raw 1D vibration signals. In addition, Zhu et al. [18] streamlines signal processing and automatically extracts key fault features by combining Principal Component Analysis (PCA) with a Deep Belief Network (DBN). Shao et al. [19] integrates transfer learning to enhance CNNs, achieving high fault diagnosis accuracy even with limited target domain data. Zou et al. [20] combined multi-scale weighted entropy morphological filtering (MWEMF) with BILSTM to address issues such as modal aliasing. Liu et al. [21] proposed a transfer learning-based method with an Inception-ResNet-v2 model, improving small-sample classification through signal-to-image conversion. Chen et al. [22] introduced a neural network for automatic feature learning that employs two CNNs with different kernel sizes to extract multi-frequency features, which are then combined with LSTM to identify fault types. Further research has shown that integrating attention mechanisms with graph neural networks can enhance fault diagnosis. Guo et al. [23] proposed a fault diagnosis approach using an attention CNN and BILSTM (ACNN-BILSTM), enhancing model focus on key features and effectively reducing noise interference. Zuo et al. [24], by contrast, developed a method based on a multi-scale weighted visibility graph and multi-channel graph convolutional network (MCGCN), leveraging the global and local feature extraction capabilities of graph neural networks to address overfitting issues with limited sample data. Meanwhile, Zhao et al. [25] proposed a multi-scale perceptual graph convolutional network (MPGCTN) to address aero-engine bearing fault diagnosis under severely imbalanced data by constructing a dual-channel feature map and employing multi-scale Chebyshev graph convolution.Experimental results show that the method maintains high accuracy under imbalanced data, outperforming mainstream methods.

Although recent algorithms have advanced fault diagnosis, existing methods still suffer from extracted features lacking multi-dimensional representation under complex conditions (e.g., varying rotational speeds), insufficient fault information in multi-source signals, and ineffective feature fusion. To address these issues, a real-time fault diagnosis method for dual-channel rolling bearings is proposed, leveraging a CNN-BILSTM architecture enhanced with an attention mechanism to fuse extracted features and maximize fault information representation. Time-domain signals provide dynamic information on instantaneous changes, whereas frequency-domain signals reveal frequency components and periodic characteristics. By integrating time-series and spectral features, the model can analyze fault signals from multiple perspectives. Incorporating an attention mechanism enhances the model's focus on key features, enabling more comprehensive capture of fault characteristics in bearing vibration signals. The main contributions of this paper are summarized as follows:

1. The CNN-BILSTM feature extraction network has been designed and optimized at the structural level, enhancing the model's feature extraction capability and adaptability under varying rotational speed conditions
2. A convolutional attention module following the CNN network is added to emphasize key features in the time–frequency domain across both channel and spatial dimensions. The original inputs are combined to create the feature fusion attention module (1DECBAM), enabling the model to retain original inputs while weighting important features, thus enhancing feature extraction capabilities
3. The proposed HIFAttn framework is constructed using T-FIAttn and L-GAAM. T-FIAttn captures potential feature relationships across both time and frequency domains. A gating mechanism is introduced after the two-channel BILSTM, enabling L-GAAM to adaptively and dynamically learn key features by combining local and global attention mechanisms. Additionally, HIFAttn fuses multimodal features, enhancing the model's focus on essential features while enriching feature diversity

The subsequent sections of this paper are structured as follows. Section 2 presents the relevant theory. Section 3 presents the basic architecture for real-time fault diagnosis of two-channel rolling bearings based on CNN-BILSTM and improved attention mechanisms. Section 4 constructs the dataset and experiments are conducted to validate the effectiveness of the fault diagnosis method proposed in this paper. Section 5 presents the conclusions and future work of the paper.

2. Related work

2.1. Convolutional neural network

CNN [26] is a multi-layer feed-forward neural network that updates its parameters via the backpropagation algorithm and generally consists of convolutional, pooling, and fully connected layers. The convolutional layer primarily extracts features automatically, the pooling layer

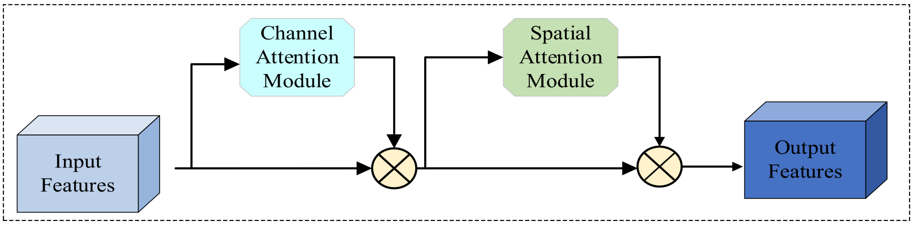
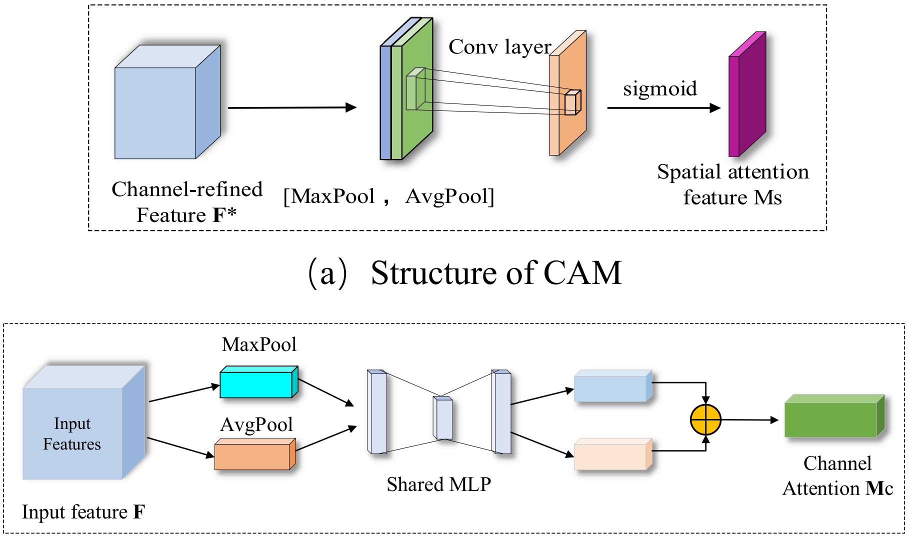


Fig. 2. Model structure of CBAM.



(a) Structure of CAM

(b) Structure of SAM

Fig. 3. Model structure of CAM and SAM.

performs subsampling to reduce overfitting, and the fully connected layer classifies the extracted features for output. Fig. 1 illustrates the classical CNN model structure.

The convolution operation is represented by the following formula:

$$y_j^l = F\left(\sum\nolimits_{h\in M_j} y_h^{l-1} * w_{hj}^l + a_j^l\right) \tag{1}$$

where, y_j^l represents the j th feature of the l th layer; M_j denotes the set of input feature signals from the previous layer; y_h^{l-1} indicates the h th feature signal of layer $l-1$ in the constructed model; w_{hj}^l represents the weight matrix of the convolutional kernel; a_j^l is the bias term; * represents the convolution operation; and F () is the activation function.

2.2. Convolutional block attention module

CBAM [27], an advancement of the Bottleneck Attention Module (BAM) [28], is a highly efficient attention mechanism module commonly used in 2D Convolutional Neural Networks (CNNs), as illustrated in Fig. 2. CBAM enables the model to focus on task-relevant information by applying attention mechanisms across both channel and spatial dimensions of the feature map, thereby enhancing overall model performance. As illustrated in Fig. 3(a), the Channel Attention Module (CAM) assigns weights to the feature map according to each channel's significance, thereby highlighting essential channel features. Following channel weighting, the Spatial Attention Module (SAM) applies additional spatial weighting to emphasize critical regions, as shown in Fig. 3(b).

The feature $\mathrm{M_c(F)}$ of the CAM is calculated by equation (2).

$$\mathrm{M_c(F)} = \sigma(MLP(AvgPool(\mathrm{F})) + MLP(MaxPool(\mathrm{F})))$$

$$= \sigma\left(\mathrm{W_1}\left(\mathrm{W_0}\left(\mathrm{F^c_{avg}}\right)\right) + \mathrm{W_1}\left(\mathrm{W_0}\left(\mathrm{F^c_{max}}\right)\right)\right) \tag{2}$$

where, σ denotes a sigmoid function, $\mathrm{W_0} \in \mathbb{R}^{\mathrm{C/r\times C}}$,and $\mathrm{W_1} \in \mathbb{R}^{\mathrm{C\times C/r}}$. The MLP weights $\mathrm{W_0}$ and $\mathrm{W_1}$ are shared across both inputs, with the ReLU activation function applied after $\mathrm{W_0}$.

The feature $\mathrm{M_s(F)}$ of the SAM is calculated by equation (3).

$$\mathrm{M_s(F)} = \sigma(f^{7\times 7}([AvgPool(\mathrm{F}); MaxPool(\mathrm{F})]))$$

$$= \sigma\left(f^{7\times 7}\left(\left[\mathrm{F^s_{avg}}; \mathrm{F^s_{max}}\right]\right)\right) \tag{3}$$

where σ denotes the sigmoid function and $f^{7\times 7}$ denotes the convolution operation with a filter size of 7 × 7.

2.3. Bidirectional long and short-term memory network

The core unit of a long short-term memory network (LSTM) comprises three gates: the forget gate f_t, the input gate i_t, and the output gate o_t. The forget gate evaluates information in the storage unit and determines what to retain and discard based on this analysis. Subsequently, the input gate updates the memory cell state, while the output gate regulates the LSTM's output. In an LSTM model, the input vector comprises the hidden state h_{t-1} from the previous time step and the current input X_t, with the output being the hidden state h_t at the current moment. The memory units $\widehat{s}_t$ and s_{t-1} form the primary memory structure within the LSTM. During training, the previous memory s_{t-1} is

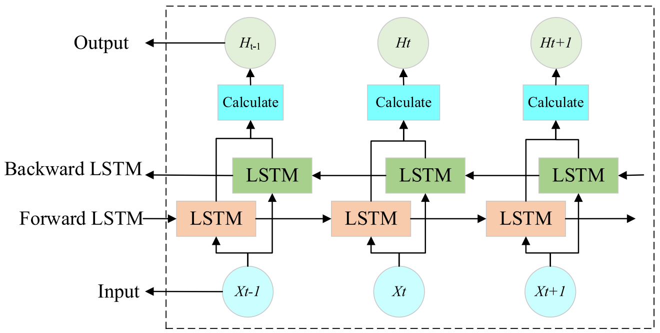


Fig. 4. Structure of BILSTM.

partially decayed by the forget gate, while new memory $\hat{s}_t$ is generated through input gate replenishment. Finally, the output gate controls the generation of the current output h_t. Below are the mathematical expressions for each LSTM gating mechanism.

The information processed by the forgetting gate is expressed as equation (4):

$$f_t = \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right) \tag{4}$$

where W represents the weight matrix and b represents the bias term.

After processing by the input gate, the information is represented as equations (5)-(6):

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{5}$$

$$\hat{s}_t = \tanh(W_{\hat{s}x}x_t + W_{\hat{s}h}h_{t-1} + b_{\hat{s}}) \tag{6}$$

The process of updating the final output hidden state by the output gate is expressed by equation (7):

$$h_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)\tanh(i_t\hat{s}_t + f_t s_{t-1}) \tag{7}$$

However, the unidirectional nature of LSTM may limit its ability to effectively utilize critical information from future time steps. The BILSTM [29] addresses this by incorporating LSTM layers in both forward and backward directions, enabling the model to access information from both past and future moments. This bidirectional structure allows for more comprehensive feature capture, enhancing the accuracy of sequence modeling. The architecture of BILSTM is illustrated in Fig. 5, with the corresponding operations shown in equations (8)–(10).

$$\overrightarrow{h}_t = M\left(x_t, \overrightarrow{h}_{t-1}\right) \tag{8}$$

$$\overleftarrow{h}_t = M\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{9}$$

$$H_t = \overrightarrow{h}_t + \overleftarrow{h}_t + b_y \tag{10}$$

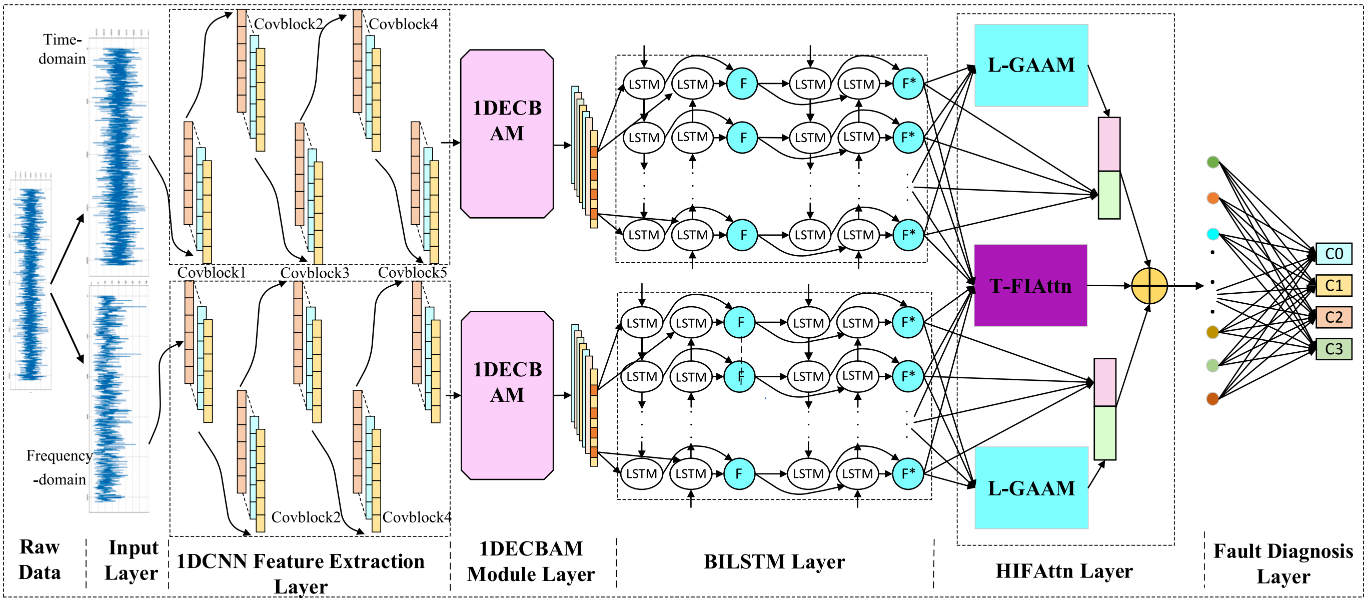


Fig. 5. Structure of the proposed FCNN-1DECBAM-BiHIFAttn.

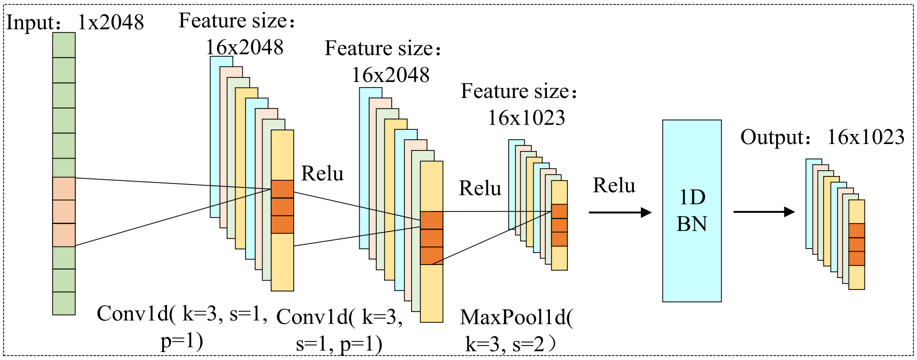


Fig. 6. The structure of ConvBlock1.

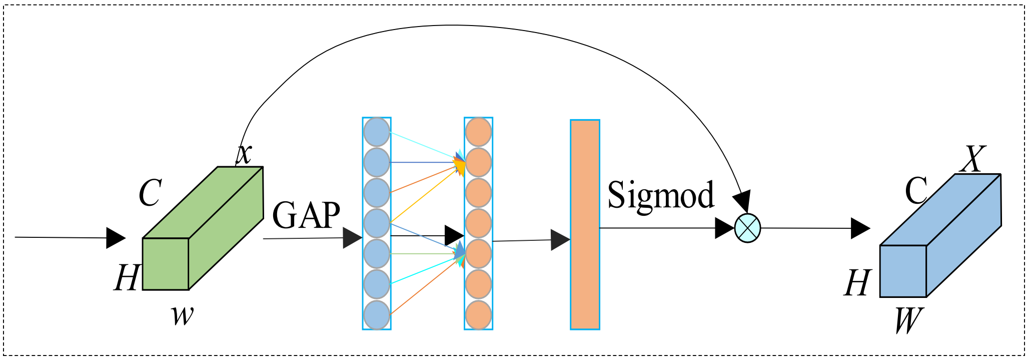


Fig. 7. The structure of channel attention mechanism.

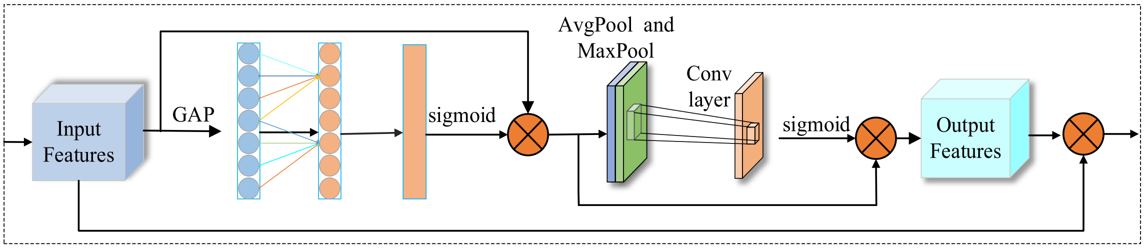


Fig. 8. The structure of proposed 1DECBAM.

where, $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ denote the outputs of the forward and backward layers at time t, b_y represents the output layer bias vector, M denotes the LSTM network model, and H_t is the output of the BILSTM network. Fig. 4 illustrates the standard structural diagram of BILSTM.

3. The proposed FCNN-1DECBAM-BiHIFAttn model

The proposed FCNN-1DECBAM-BiHIFAttn model here is built upon CNN-BILSTM and integrates multiple attention mechanisms. It processes both time- domain and frequency-domain signals for end-to-end fault diagnosis. The architecture consists of an input layer, CNN feature extraction layer, 1DECBAM layer, BILSTM layer, HIFAttn layer, and fault diagnosis layer. The structure of FCNN-1DECBAM-BiHIFAttn, which is illustrated in Fig. 5, demonstrates the integration of multiple attention mechanisms.

3.1. *Optimal feature extracting network CNN-BILSTM*

Given that the raw vibration signals obtained from bearing failures are inherently one-dimensional time series data, traditional two-dimensional convolutional neural networks (2DCNN) often require a transformation of this data into representations like time–frequency diagrams to enable effective application. To enable direct processing of raw data while minimizing information loss, this paper enhances the 1DCNN architecture and develops a foundational CNN-BILSTM feature extraction network utilizing BILSTM. The 1DCNN-based feature extraction network employed in this paper comprises a series of five identical convolutional blocks. The configuration of ConvBlock1 is illustrated in Fig. 6.

Each convolutional block comprises two one-dimensional convolutional layers, a max-pooling layer, and a normalization layer. In each convolutional layer, the stride is set to 1 and the activation function used is relu.A uniform convolution kernel size of 3 is selected, as it represents a classical and effective choice for 1D convolutional layers, balancing the extraction of local signal features and computational efficiency, thus ensuring robust and accurate feature learning performance. Leveraging the sliding window technique, the one-dimensional convolutional kernel can efficiently capture local signal features while preserving the temporal structure of the target sequence. As the number of convolutional channels directly influences the quantity of feature signals learned within each convolutional layer. Consequently, to achieve comprehensive feature extraction, this study incrementally increases the number of channels in the 1DCNN architecture from 16 to 256. Each convolutional layer isolates features within specific frequency bands, and by sequentially layering multiple convolutional layers interspersed with activation functions, the network progressively learns to capture more sophisticated patterns and hierarchical features across diverse frequency ranges present in the raw data.Pooling and normalization layers are incorporated after each convolutional block to reduce dimensionality and computational complexity during feature extraction while mitigating overfitting and enhancing the model's generalization capability. The BILSTM architecture applied in both the time and frequency domains is structured with two layers, featuring hidden dimensions of 256 and 128, respectively. The two-layer BILSTM progressively captures temporal dependencies at different scales. The first layer, with a larger hidden dimension (256), effectively extracts global temporal and spectral variations of the signal, while the second layer (128) refines feature representations and reduces redundancy.This hierarchical design, transitioning from high to low dimensions, ensures comprehensive extraction of bearing fault features while minimizing network complexity and mitigating overfitting. Consequently, the model achieves improved generalization and diagnostic accuracy across varying operating conditions.The BILSTM layer in the time domain generates a novel temporal feature representation. These features not only retain the intrinsic time-series characteristics of the original data but also enhance the model's capacity to capture global temporal patterns via bidirectional processing. In the frequency domain, BILSTM learns temporal patterns and dependencies within frequency variations by processing the raw data features along the frequency dimension. The CNN-BILSTM feature extraction network developed in this paper demonstrates the ability to learn features from any point within the signal, effectively capturing fault characteristic frequencies and abrupt changes in vibration. Furthermore, applying the CNN-BILSTM feature extraction network simultaneously in both the time and frequency domains enables the model to learn more intricate feature representations across multiple

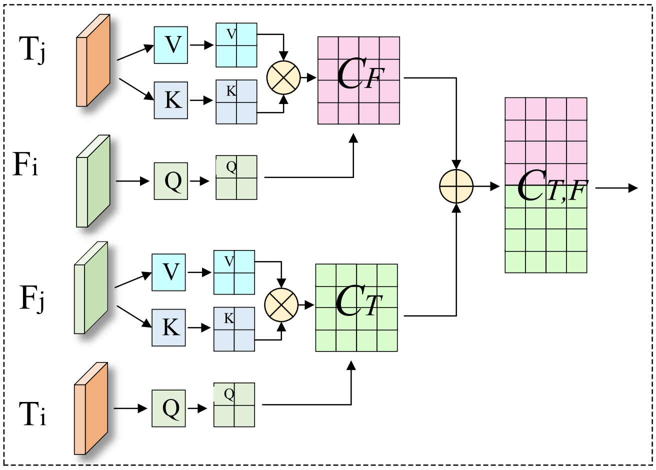


Fig. 9. The structure of T-FIAttn.

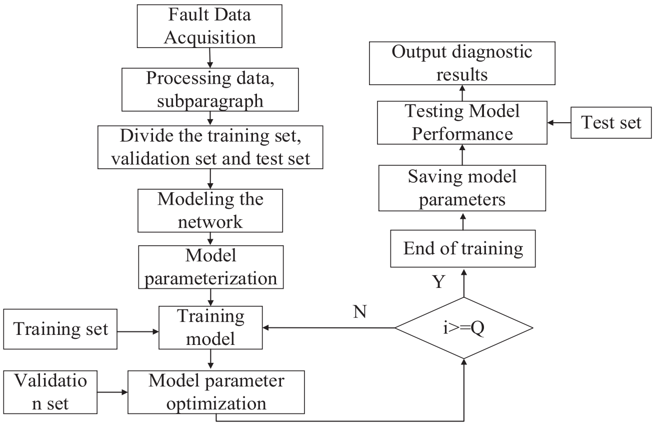

Fault Data Acquisition
Processing data, subparagraph
Divide the training set, validation set and test set
Modeling the network
Model parameterization
Training set
Training model
Validatio n set
Model parameter optimization
N
i>=Q
Y
End of training
Saving model parameters
Testing Model Performance
Test set
Output diagnostic results

Fig. 10. Fault diagnosis flowchart.

Table 1
FCNN-1DECBAM-BiHIFAttn model parameters applied.

Layer name	Type	Kernel	Channel (In, Out)	Stride	Output
Input layer	–	–	1	–	[32,1,2048]
Time domain/ frequency domain CNN feature extraction layer	ConvBlock1	3	(1,16)	1	[32, 16,1023]
	ConvBlock2	3	(16,32)	1	[32,32,511]
	ConvBlock3	3	(32,64)	1	[32,64,255]
	ConvBlock4	3	(64,128)	1	[32,128,127]
	ConvBlock5	3	(128,256)	1	[32,256,63]
1DECBAM module layer	1DECBAM	–	(256,256)	–	[32,256,63]
Time domain BILSTM/ Frequency domain BILSTM layer	BILSTM1	–	(256,512)	–	[32,63,512]
	BILSTM2	–	(512,256)	–	[32,63,256]
HIFAttn layer	T-FIAttn	–	(256,256)	–	[32,63,256]
	L-G AAM	–	(256,256)	–	[32,63,256]
Fault diagnosis layer	Linear	–	(768, 4/7)	–	[32,4/7]

dimensions. This approach significantly benefits the classification task addressed in this paper.

3.2. 1DECBAM

The channel attention module incorporated in the 1DECBAM architecture proposed in this paper is illustrated in Fig. 7. Initially, Global Average Pooling (GAP) is applied to each channel following the convolutional operation, enabling the extraction of global information from each channel. Subsequently, cross-channel interaction is facilitated by a one-dimensional convolutional layer with a kernel size of kize, where the kernel size is adaptively determined according to the number of channels. The 1D convolutional operation captures both local and

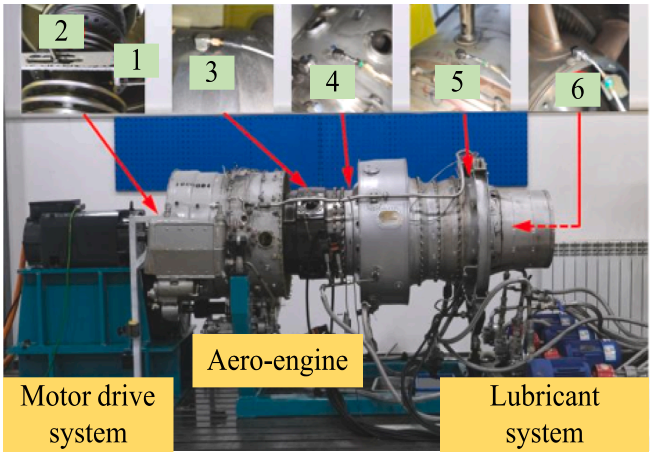


Fig. 11. Aero-engine experimental platform.

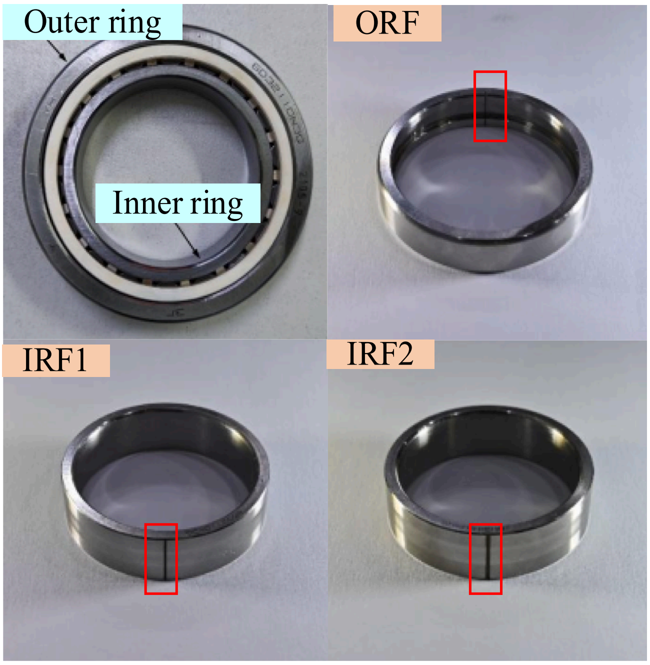


Fig. 12. Fault components.

Table 2
Details of the dataset.

Data label	Fault type	Damage Size /mm	LP and HP speed (r/min)	Sample length	Training set/ validation set/ test set
C1	NOR	0*0	LP speed range: 1000~5000 HP speed range: 1200~6000	2048	2016/672/672
C2	IRF1	0.5*0.5		2048	2016/672/672
C3	IRF2	0.5*1.0		2048	2016/672/672
C4	ORF	0.5*0.5		2048	1800/600/600

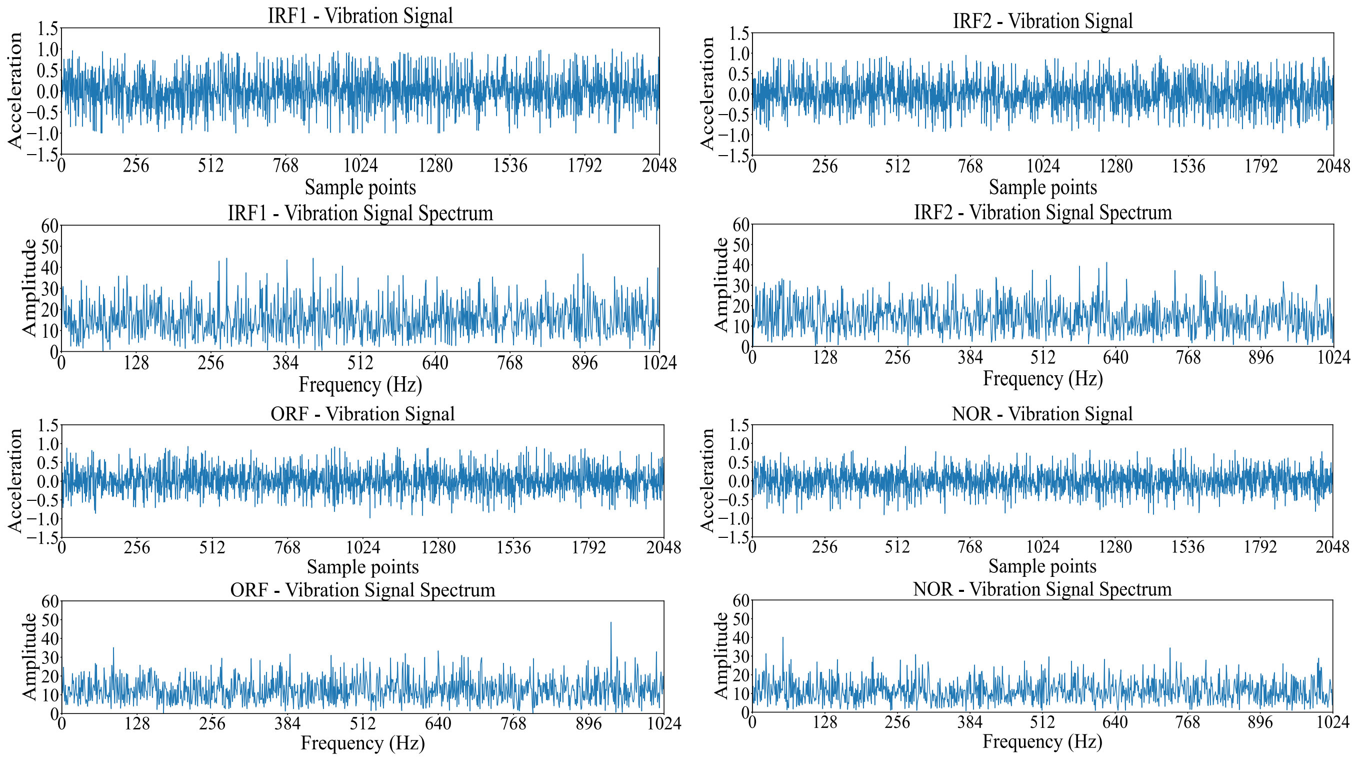


Fig. 13. Raw signal and frequency domain waveforms for the four health states.

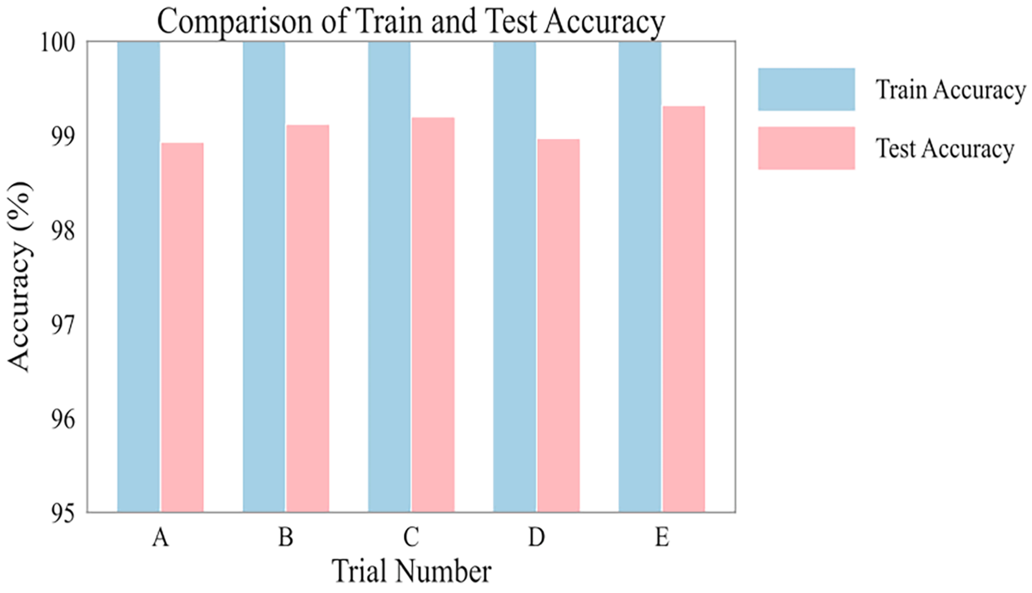


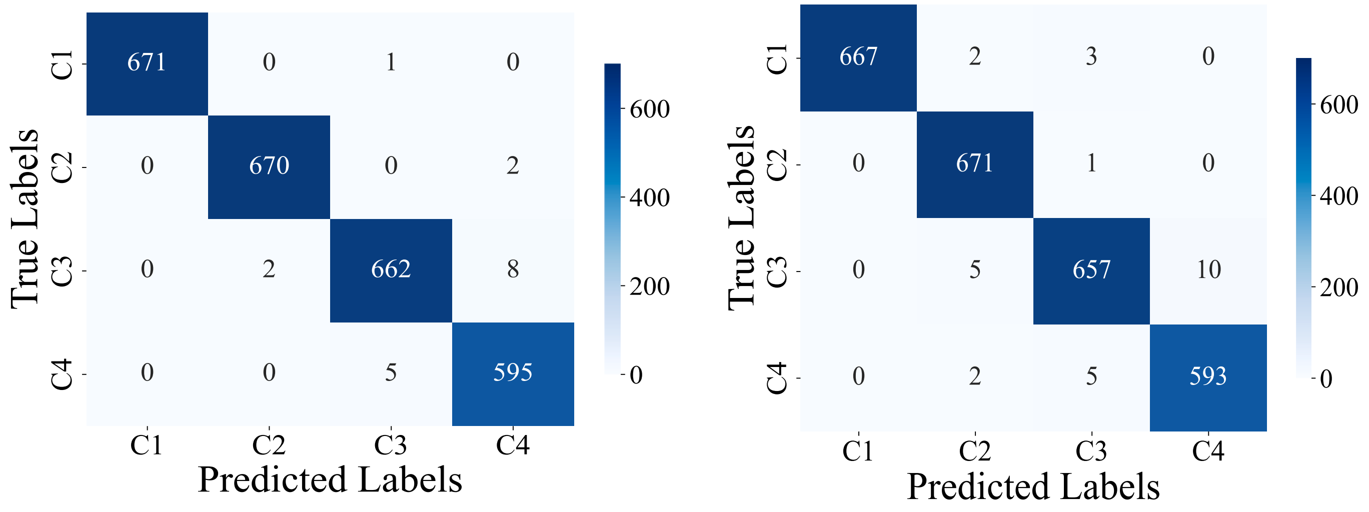
Fig. 14. Accuracy of five experiments.

Table 3
Rolling bearing fault diagnosis results using FCNN-1DECBAM-BiHIFAttn.

Fault label	Precision (%)	Recall (%)	F1-score (%)
C1	99.81	99.48	99.65
C2	98.97	99.55	99.31
C3	98.84	98.03	98.43
C4	98.98	99.17	98.86
Accuracy	99.15		

global feature information within each channel, assigning greater weights to significant channels to enhance the extraction of fine-grained features. Finally, a sigmoid function is applied to generate a weight for each channel, which is then combined with the original features to produce channel-attentive feature representations.

The channel attention module depicted in Fig. 7 is employed to replace the original channel attention mechanism within the CBAM. This module is then integrated with the initial features extracted by the CNN to construct the 1DECBAM module, as illustrated in Fig. 8. The input and output channels of 1DECBAM are maintained at 256 to ensure consistency. Since the attention module selectively enhances or



(a)

(b)

Fig. 15. Confusion matrix of FCNN-1DECBAM-BiHIFAttn: (a) highest accuracy of 99.32%, (b) lowest accuracy of 98.93%.

Table 4
Ablation experiments.

Model	1DCNN	BILSTM	FFT	1DECBAM	HIFAttn	Accuracy	Precision	Recall	F1-score
A(Classic)	√	√				85.78	86.19	85.69	85.85
B	√	√				95.45	95.44	95.41	95.42
C	√		√			88.26	88.29	88.27	88.26
D	√	√	√			96.87	96.82	96.91	96.84
E	√	√		√		96.79	96.79	96.81	96.80
F	√		√	√		90.37	90.31	90.34	90.30
G	√	√	√	√		97.86	97.85	97.86	97.86
H	√	√	√	√	√	99.15	99.16	99.14	99.15

suppresses features extracted by the CNN, it adopts the same number of channels as the final CNN layer. The 1DECBAM module assigns weights to the features extracted by the CNN, selectively enhancing critical channel and spatial features that are filtered out. Furthermore, the original feature information is preserved to prevent the loss of crucial features. The optimized 1DECBAM module improves the model's capacity to focus on essential features, while reducing computational complexity through an efficient channel attention mechanism. Additionally, by incorporating the features extracted by the CNN, the module enables the model to better capture multi-scale features, thereby enhancing performance in complex tasks.

3.3. Hybrid-Interaction fusion attention

In the context of bearing fault diagnosis, signals in the time and frequency domains each encapsulate distinct critical information. Time-domain features capture instantaneous variations in the signal, enabling the detection of short-term shocks and transient dynamics in vibration data. In contrast, frequency-domain features uncover the periodicity and inherent frequency components of the signal, which are instrumental in identifying long-term periodic failure modes. To fully leverage and integrate these two types of features, this study introduces a Hybrid Interaction-Fusion Attention (HIFAttn) layer designed to merge multiple feature sets. The architecture of the HIFAttn layer is presented in Fig. 5. The HIFAttn layer comprises two main components: T-FIAttn and L-GAAM, which are discussed in detail in subsections 3.3.1 and 3.3.2.

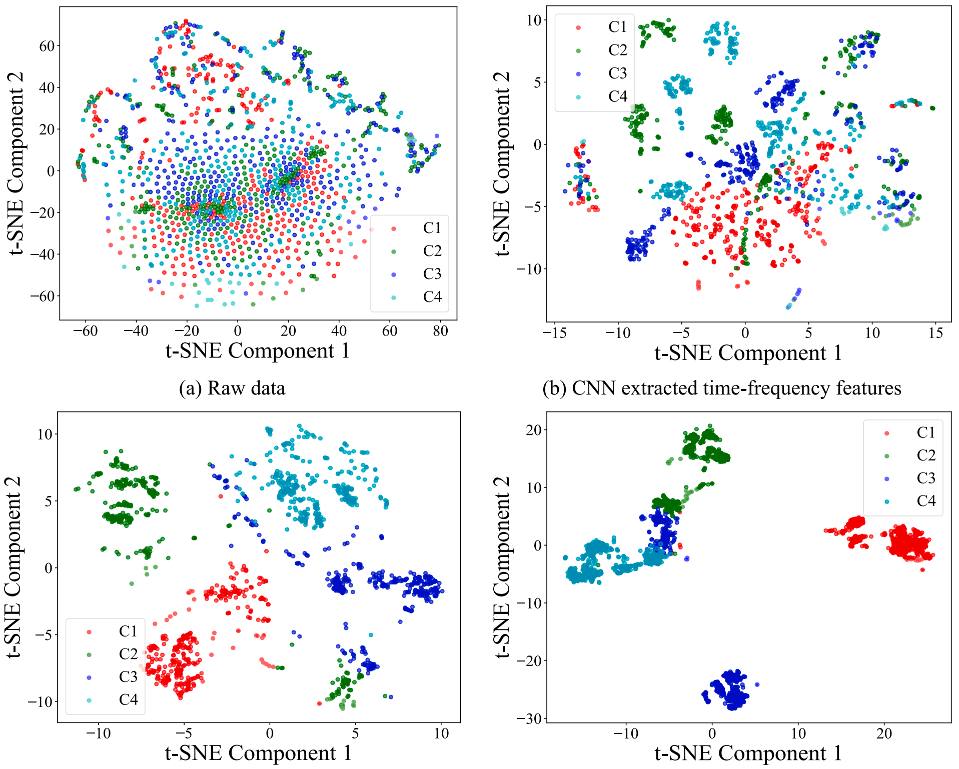
The HIFAttn layer concatenates and integrates three distinct types of feature information: time-domain channel features, frequency-domain channel features, and frequency-domain interaction representations. This multimodal feature fusion effectively mitigates the limitations associated with unidimensional information while enhancing the model's capability to discriminate bearing faults. The concatenated feature vectors preserve the time and frequency domain information along with their interactions, thereby reducing the risk of information loss and enhancing model robustness. Finally, these fused features are input into the fully connected layer to complete the fault classification task.

3.3.1. Time-Frequency interactive attention mechanisms

In the field of fault diagnosis based on vibration signals, most existing research focuses on fault feature extraction within a single domain, with limited consideration of inter-domain correlations. However, fault characteristics are often exhibited across both time and frequency domains, making it essential to capture the interrelationship between these modes for precise diagnosis. To address this limitation, this paper presents the T-FIAttn, as depicted in Fig. 9.

Initially, the similarity between the time-domain feature T and the frequency-domain feature F is computed using a dot product attention mechanism, providing a measure of the correlation between each time-domain and frequency-domain feature. For instance, time-domain features are employed to compute correlations with frequency-domain features, as shown in equation (11).

$$\alpha_{i,j}^{T\rightarrow F} = \mathrm{softmax}\left(\frac{\left(T_i \cdot F_j^T\right)}{\sqrt{d}}\right) \tag{11}$$



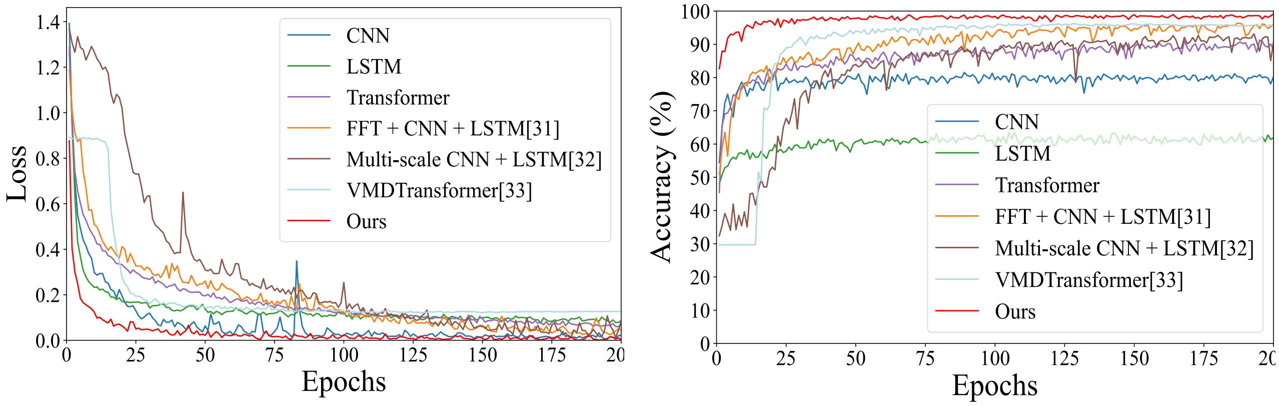
(a) Raw data

(b) CNN extracted time-frequency features

(c) Timing features extracted by BILSTM

(d)Fusion features after HIFAttn layer

Fig. 16. Feature visualization Based on t-SNE.



(a) Loss curves

(b) Accuracy curves

Fig. 17. Experimental results of different models.

where, T_i represents the eigenvector at the i th time step of the time-domain feature, while F_j^T denotes the transpose of the eigenvector at the j th time step of the frequency-domain feature.Here, d refers to the dimensionality of the feature vectors at each respective time step. $\alpha_{i,j}^{T\to F}$ represents the attention weight between the i th time domain feature and the j th frequency domain feature. The T-FIAttn proposed exhibits reciprocity. This mechanism enables time-domain features to attend to significant features in the frequency domain via the dot product attention mechanism, while the reverse interaction is equally effective. Attention weights $\alpha^{F\to T}$ on the time domain feature T are also derived from the frequency-domain feature F. By calculating these attention weights, a weighted representation C_T for time-domain features derived from frequency-domain features, and C_F for frequency-domain features derived from time-domain features, can be obtained. This process is illustrated in Fig. 9, with the corresponding equations shown below:

$$C_T = \sum_j \alpha_{i,j}^{T\to F} F_j, C_F = \sum_i \alpha_{j,i}^{F\to T} T_i \tag{12}$$

$C_{T,F} = C_T + C_F$ (13).

Ultimately, the time–frequency interaction yields a fused contextual representation, $C_{T,F}$, encapsulating all critical correlation information between the two modes.The dimension of the Time-Frequency Interactive Attention Module (T-FIAttn) is set to 256 to ensure adequate capacity for capturing interactions and complementary information between time- and frequency-domain features. Additionally, this setting maintains dimensional consistency between the upper and lower layers

Table 5
Experimental results.

Model	Accuracy (%)					Average accuracy (%)
	A	B	C	D	E	
CNN	79.13	79.09	80.66	80.92	79.85	79.93 ± 0.78
LSTM	64.63	64.00	63.55	65.68	63.58	64.29 ± 0.64
Transformer	89.69	90.04	90.12	89.21	90.98	90.01 ± 0.34
FFT + CNN + LSTM [31]	96.52	95.68	95.72	96.33	96.71	96.19 ± 0.47
Multi-scale CNN + LSTM [32]	91.93	92.43	92.81	92.51	92.20	92.38 ± 0.33
VMTransformer [33]	95.56	95.60	94.89	95.31	95.67	95.40 ± 0.12
Ours	**99.01**	**99.12**	**99.20**	**99.08**	**99.32**	**99.15 ± 0.11**

of the CNN-BILSTM network. The T-FIAttn proposed in this paper enables the model to identify correlations between specific time steps and frequency-domain bands. For instance, fault signals within particular frequency bands may display distinct time-domain vibration patterns.

This interaction mechanism effectively integrates transient changes in the time domain with periodic information in the frequency domain, thereby enhancing the model's capacity to recognize complex failure modes. The bidirectional interactive attention mechanism proposed fully captures deep-level associations between these two types of features, compensating for the lack of associative information resulting from the single cross-attention mechanisms used in previous studies. By incorporating an interactive attention mechanism for both time-domain and frequency-domain features, the model proposed in this paper achieves effective integration of temporal and frequency information, thereby enhancing the model's overall performance.

3.3.2. *Local-Global adaptive attention module*

In the field of vibration signal fault diagnosis, local attention focuses on signal variations within specific time intervals or frequency ranges, making it well-suited for capturing short-term, abrupt fault characteristics such as transient shocks or localized spectral peaks. In contrast, global attention captures long-term dependencies across the entire time series or spectrum. This is particularly beneficial for identifying important periodic patterns associated with recurring faults in the

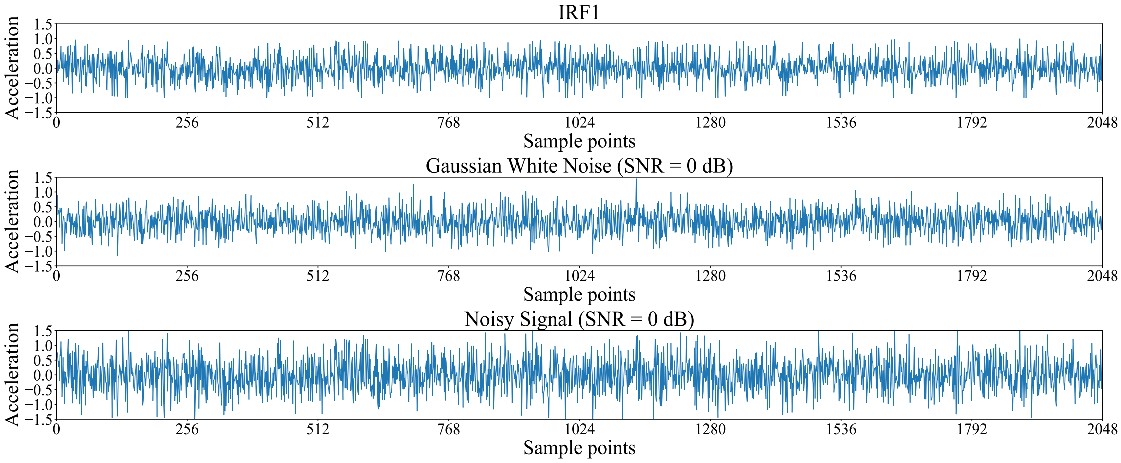


Fig. 18. Waveform of IRF1, Gaussian white noise, and composite noise signal with SNR = 0.

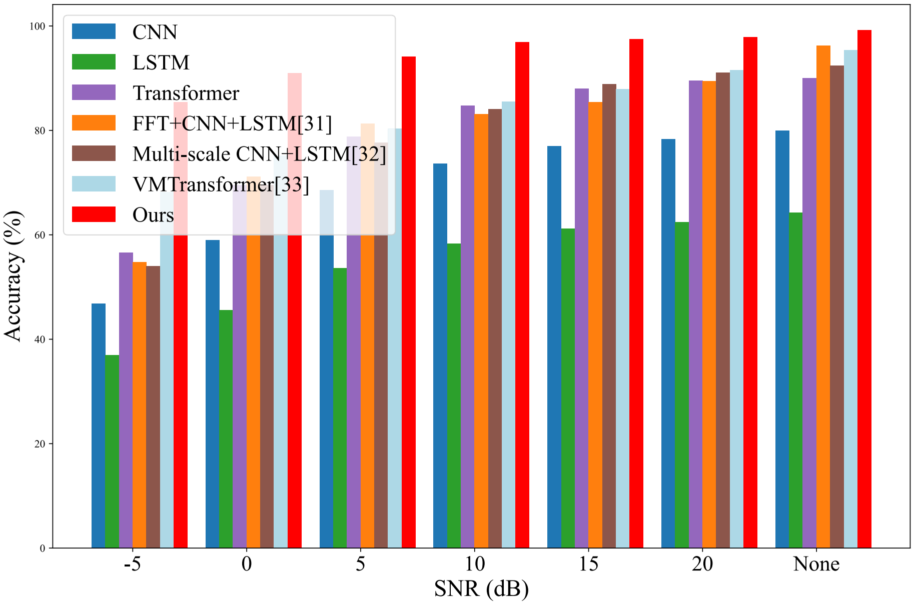


Fig. 19. Experimental results of different methods under different noise conditions.

Table 6
Accuracy of each model under different noise conditions.

MODEL	SNR						
	−5	0	5	10	15	20	None
CNN	46.86 ± 0.34	58.93 ± 3.18	68.53 ± 1.86	73.64 ± 0.91	76.98 ± 0.33	78.34 ± 0.33	79.93 ± 0.78
LSTM	36.95 ± 0.86	45.53 ± 0.21	53.62 ± 1.35	58.32 ± 0.58	61.21 ± 0.49	62.39 ± 0.56	64.29 ± 0.64
Transformer	56.61 ± 1.31	69.54 ± 0.18	78.78 ± 0.25	84.77 ± 0.30	88.03 ± 0.84	89.57 ± 0.02	90.01 ± 0.34
FFT + CNN + LSTM [31]	54.72 ± 2.30	71.14 ± 3.28	81.33 ± 6.91	83.14 ± 0.63	85.40 ± 1.25	89.44 ± 3.46	96.19 ± 0.47
Multi-scale CNN + LSTM [32]	54.04 ± 2.53	69.48 ± 6.78	77.66 ± 1.73	84.06 ± 1.53	88.90 ± 1.06	91.08 ± 0.28	92.38 ± 0.33
VMTransformer[33]	68.20 ± 1.63	75.20 ± 0.76	80.31 ± 1.08	85.51 ± 0.88	87.90 ± 0.57	91.50 ± 0.91	95.40 ± 0.12
Ours	**85.45 ± 0.37**	**90.97 ± 0.08**	**94.13 ± 0.01**	**96.89 ± 0.01**	**97.46 ± 0.09**	**97.86 ± 0.40**	**99.15 ± 0.11**

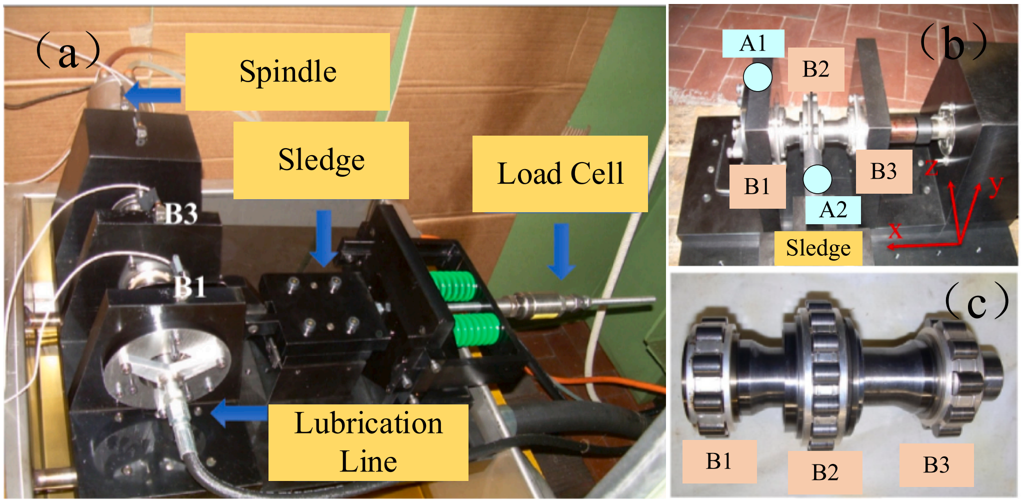


Fig. 20. (a) Test bench tested (b) Two acceleration test points (c) Bearing tested.

Table 7
Details of the dataset.

Data label	Fault type	Damage Size /mm	LP and HP speed (r/min)	Sample length	Training set/ validation set/ test set
C1	NOR	–	Speed range: 6000~30000	2048	720/240/240
C2	IRF1	450		2048	720/240/240
C3	IRF2	250		2048	720/240/240
C4	IRF2	150		2048	720/240/240
C5	ORF1	450		2048	720/240/240
C6	ORF2	250		2048	720/240/240
C7	ORF3	150		2048	720/240/240

signal. However, most previous approaches rely on a single attention mechanism. To achieve a more comprehensive understanding of the input data while capturing both long-range dependencies and detailed information, this study proposes a L-GAAM, as depicted in Fig. 5. The module incorporates an adaptive gating mechanism that facilitates the dynamic, weighted fusion of local and global features through a learnable gating parameter γ. This module is governed by the following equations (14–15):

$$\gamma = \sigma\left(W_{\gamma} \cdot F + b_{\gamma}\right) \tag{14}$$

$$C_{\text{fusion}} = \gamma \cdot \sum_{i} \alpha_{\text{local},i} H_i + (1-\gamma) \cdot \sum_{i} \alpha_{\text{global},i} H_i \tag{15}$$

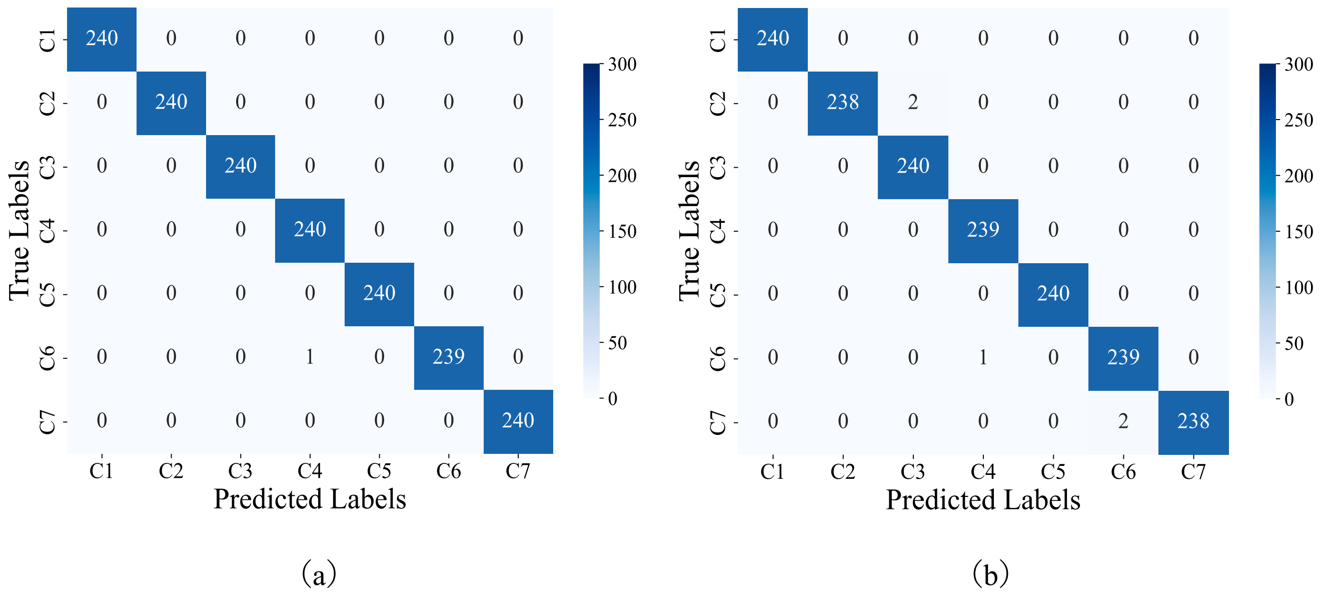
where, σ represents a sigmoid function that constrains the output γ value between 0 and 1. The γ parameter dynamically adjusts the model's reliance on local and global features. W_{γ} denotes the weight matrix, F is the feature vector derived from the BILSTM output, and b_{γ} represents the bias term. H_i represents a time-domain or frequency-domain feature from the BILSTM output. $\alpha_{\text{local},i}$ and $\alpha_{\text{global},i}$ denote the attentional weights at the *ith* position, computed by the local and global attention mechanisms, respectively.

Through the introduction of adaptive gating, the L-GAAM allows the model to more accurately capture elements critical for decision-making. The model can dynamically adjust the weighting of local and global attention based on the characteristics of varying input signals, enhancing its flexibility in handling diverse failure modes. The dimension of the Local-Global Adaptive Attention Module (L-GAAM) is also set to 256 to preserve uniformity within the CNN-BILSTM structure while providing sufficient capacity to dynamically emphasize key local and global features. Building on this, this study fuses features processed by the local–global adaptive attention mechanism with the original BILSTM output to create a richer feature set, as illustrated in Fig. 5. This combination encapsulates global dependencies along with essential local information, offering the model a more expressive feature representation and thereby enhancing its performance in subsequent classification tasks.

3.4. Fault diagnosis process

The fault diagnosis flow of the FCNN-1DECBAM-BiHIFAttn model, incorporating the enhanced attention mechanism, is illustrated in Fig. 10. The detailed fault diagnosis procedure is as follows:

1. Acquire raw fault signals and divide them into training, validation, and test sets.
2. Construct the FCNN-1DECBAM-BiHIFAttn fault diagnosis model. Initialize model weights, set the Dropout rate, and configure hyperparameters such as training batch size, learning rate, and the number of hidden units in the BILSTM layer.
3. Train the model built in Step 2 using the training set.



(a)

(b)

Fig. 21. Confusion matrix of FCNN-1DECBAM-BiHIFAttn: (a) highest accuracy of 99.94%, (b) lowest accuracy of 99.70%.

Table 8
Ablation experiments.

编号	1DCNN	BILSTM	FFT	1DECBAM	HIFAttn	Accuracy	Precision	Recall	F1-score
A(Classic)	√	√				85.36	85.83	85.36	85.34
B	√	√				94.23	94.27	94.29	94.26
C	√		√			86.66	86.68	86.67	86.57
D	√	√	√			97.98	97.99	97.98	97.98
E	√	√		√		97.08	97.09	97.08	97.08
F	√		√	√		91.19	91.37	91.19	91.20
G	√	√	√	√		98.81	99.84	99.81	99.81
H	√	√	√	√	√	99.83	99.84	99.83	99.83

4. Evaluate model performance on the validation set and adjust model parameters accordingly.
5. Check if the number of training iterations i has reached the predefined iteration limit Q If reached, save the model weights; otherwise, return to Step 3.
6. Evaluate the trained model on the test set and output the fault diagnosis results.

4. Experimental verification

4.1. Experimental setup

In this paper, the proposed model is trained and evaluated within the Python 3.9-based PyTorch framework, utilizing an Intel 13th generation i9 CPU (2.2 GHz), 32 GB DDR5 RAM, and an NVIDIA GeForce RTX 4060 GPU. During training, dropout is applied to mitigate the risk of overfitting, and backpropagation is used to update model weights. Cross Entropy Loss (CEL) and the ADM optimization algorithm were employed, with an initial learning rate of 0.0003, a batch size of 32, and 200 epochs. To reduce fluctuations during training and accelerate model convergence, a learning rate decay strategy is applied, whereby the learning rate is multiplied by 0.9 if there is no decrease in loss for more than 10 epochs. As initial weights and bias settings significantly impact neural network results, five independent experiments were conducted using the same initial parameters to minimize the effect of initial choices on classification accuracy.

In this paper, four performance metrics—overall accuracy, precision, recall, and F1 score—are employed to evaluate the diagnostic performance. Accuracy serves as a quantitative metric for model fault diagnosis, while the loss function value is utilized to measure the proximity of the model's predicted fault labels to the true fault labels.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{17}$$

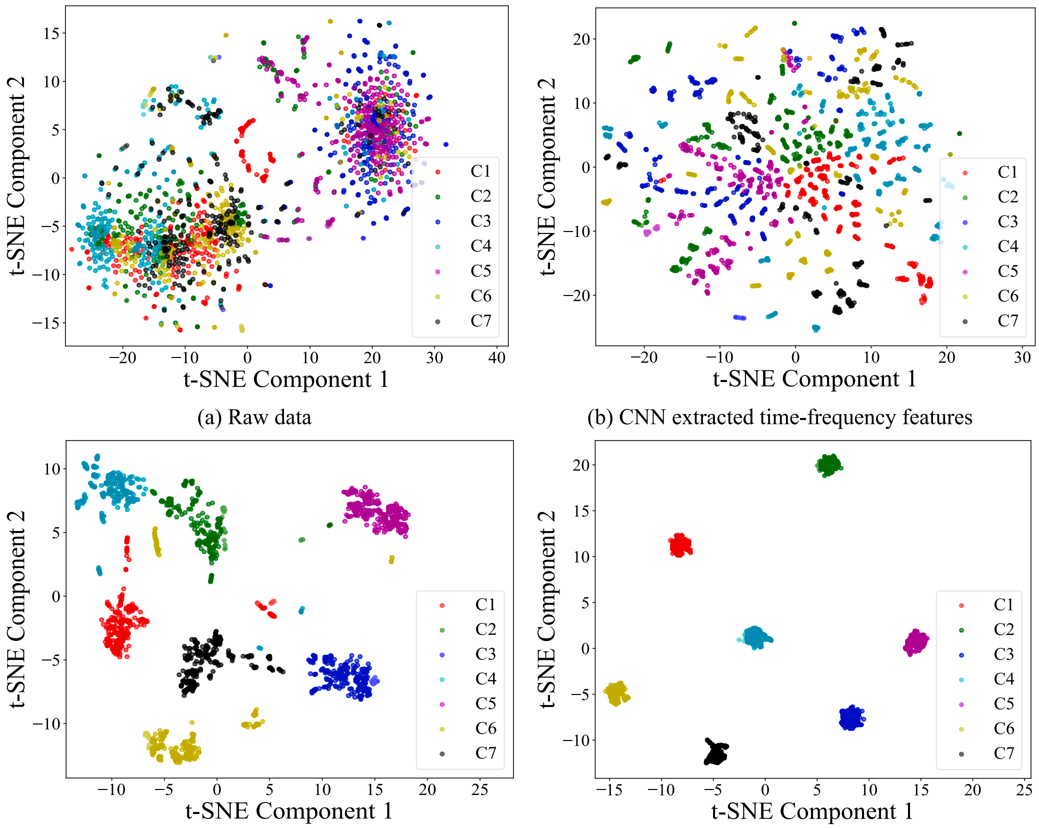
$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{18}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \tag{19}$$

where, TP denotes the number of true positive samples, FP represents the number of false positive samples, TN indicates the number of true negative samples, and FN denotes the number of false negative samples. Accuracy, recall, and precision all range from 0 to 1, with higher values indicating better fault diagnosis performance.Table 1 presents the detailed parameters of the proposed FCNN-1DECBAM-BiHIFAttn model, as discussed in Section III.

4.2. Evaluation on aero engine bearing dataset 1

In practical diagnostic scenarios, the operating conditions of engine bearings may vary randomly, and the measurement points may not be consistently located. To validate the model's performance under complex operating conditions, this study conducts variable operating



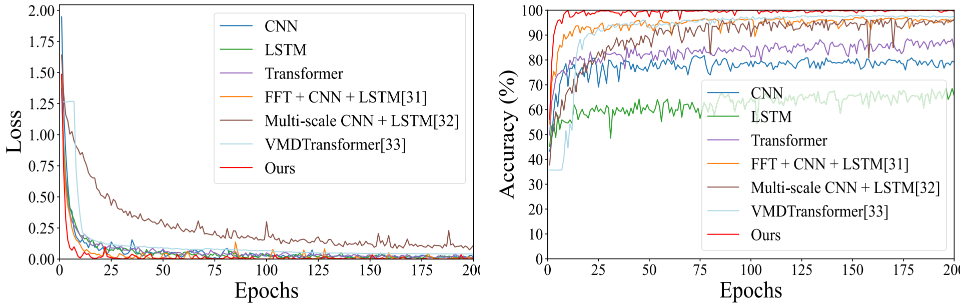
(a) Raw data

(b) CNN extracted time-frequency features

(c) Timing features extracted by BILSTM

(d)Fusion features after HIFAttn layer

Fig. 22. Feature visualization Based on t-SNE.



(a) Loss curves

(b) Accuracy curves

Fig. 23. Experimental results of different models.

condition tests using space engine bearing fault data [30].

The experimental setup primarily comprises an aero-engine, a motor drive system, and a lubrication system, as depicted in Fig. 11. The aero-engine was experimentally modified by removing rotor blades, combustion chambers, and selected accessory casings. Only key components of the twin-rotor structure, such as the low-pressure (LP) and high-pressure (HP) pressurizers, along with the LP and HP turbines, are retained. The experiment includes four operating states of the interaxial bearing: normal (NOR), inner ring failure 1 (IRF1), inner ring failure 2 (IRF2), and outer ring failure (ORF), as shown in Fig. 12. Figs. 11 and 12 are both derived from space engine bearing failure data [30]. Six measurement points are established in the experiment: two displacement sensors (No. 1 and No. 2) capture displacement vibration signals from the low-pressure rotor, while the remaining four acceleration sensors record the acceleration vibration signals of the magazine. The sampling frequency is set to 25,000 Hz.

Considering the realistic scenario where bearing failures occur at varying speeds, this study incorporates a mixture of data from different measurement points and operating speeds to simulate complex and variable detection conditions. Specifically, raw data for NOR, IRF1, IRF2, and ORF conditions were selected, with two samples randomly taken from each condition across 28 different high-speed and low-speed recordings. Due to partial loss of ORF data, only 50 samples were extracted for the Data5 category. A total of 218 new sample sets were extracted. The displacement and acceleration vibration signals were then extracted from these samples. Each sample consists of 2048 data points across four categories, resulting in a total of 13,080 new samples. These samples are divided into training, validation, and test sets in a 6:2:2 ratio. The specific data parameters are presented in Table 2, while the raw signal and frequency domain waveforms for the four health

Table 9
Experimental results.

Model	Accuracy (%)					Average accuracy (%)
	A	B	C	D	E	
CNN	81.79	82.56	82.86	81.19	81.25	81.93 ± 0.45
LSTM	71.41	72.16	72.74	71.30	72.39	71.80 ± 0.35
Transformer	89.40	89.81	90.06	89.76	89.82	89.77 ± 0.04
FFT + CNN + LSTM [31]	95.95	96.07	96.55	95.45	96.61	96.12 ± 0.18
Multi-scale CNN + LSTM [32]	93.33	94.23	93.15	93.69	94.35	93.75 ± 0.22
VMTransformer [33]	96.71	96.87	97.48	97.49	97.02	97.11 ± 0.10
Ours	**99.70**	**99.94**	**99.82**	**99.94**	**99.76**	**99.83 ± 0.01**

states are illustrated in Fig. 13.

4.2.1. Results of the proposed model

Fig. 14 illustrates the accuracy of the proposed method across five experimental runs for both the training and test sets, demonstrating 100 % training accuracy and over 98 % test accuracy in each case. Table 3 presents the average accuracy, precision, recall, and F1 score of the proposed FCNN-1DECBAM-BiHIFAttn model over five replicate runs. All test metrics exceeded 98 %, with recall surpassing 99 % for labels 0, 1, and 3, and an overall average accuracy of 99.15 %. These results indicate that the proposed model demonstrates high accuracy and stability on the aero-engine bearing dataset.

Fig. 15 presents the confusion matrices for the highest and lowest accuracy results (99.32 % and 98.93 %). The primary reason for the lower accuracy is the misclassification between labels C3 and C4. This attributed to the similarity of frequency or temporal characteristics between these fault signals in the feature space, making them challenging to differentiate precisely. Additionally, noise interference in the signals or the lack of significant distinguishing features may further exacerbate this issue. Nevertheless, the proposed FCNN-1DECBAM-BiHIFAttn model achieves consistently high classification accuracy overall, demonstrating the effectiveness of this method for diagnosing bearing faults under complex working conditions.

Table 4 presents the results of the ablation experiments for the proposed method. Model A (baseline) serves as the benchmark CNN-BiLSTM network. Comparison of the baseline model (Model A) and the optimized CNN-BiLSTM network (Model B) proposed in this study reveals a marked improvement in classification accuracy (85.78 % vs. 95.45 %), indicative of the superior feature extraction capabilities of the modified CNN-BiLSTM architecture in processing vibration signals across diverse rotational speed conditions.Model D further improves upon Model B by incorporating FFT dual-channel time–frequency inputs, achieving an accuracy of 96.87 %. Meanwhile, the experimental results for Models E and G indicate that frequency-domain information supplements the shortcomings of the time-domain features. This time–frequency bimodal feature input helps the model capture more effective fault characteristics.Model G, which adds the 1DECBAM module to Model D, shows that the model can better focus on key spatio-temporal features, further increasing the accuracy to 97.86 %. Model H, built upon Model G, incorporates Hybrid Interactive Fusion Attention (HIFAttn), achieving the highest accuracy of 99.15 %. This result highlights the critical role of the HIFAttn module in multi-modal feature fusion, particularly in capturing the latent correlations between time-domain and frequency-domain features, significantly enhancing the model's diagnostic performance under complex operating conditions.

The experimental results above demonstrate that each module plays a crucial role in improving the model's performance. Notably, the synergistic effect of the 1DCNN, 1DECBAM, BiLSTM, and HIFAttn modules enables the model to effectively extract key features from both time-domain and frequency-domain dimensions, thereby significantly improving fault diagnosis accuracy and robustness. These results not only validate the independent contributions of each module but also highlight the overall advantage of their integration, showcasing the powerful potential of the proposed model in complex fault diagnosis scenarios.

To thoroughly investigate the model's feature extraction mechanism, this study randomly selects 300 samples from each of the four categories and applies t-distributed stochastic neighbor embedding (t-SNE) to visualize the features in a two-dimensional space. Fig. 16 illustrates the original data features, the time–frequency features extracted by CNN, the temporal features extracted by BILSTM, and the final fused features combining multiple features through the fused-attention mechanism in 2D space. In Fig. 16(a), the raw data clusters are dispersed, with fuzzy category boundaries. After optimized CNN processing (Fig. 16(b)), category separation improves slightly, though some overlap remains. In Fig. 16(c), the clustering performance is further improved, as the 1DECBAM module highlights important features through the attention mechanism, enhancing the model's focus on key features, while BiLSTM captures the temporal dependencies of the signal. The time–frequency features from both channels then enter the HIFAttn layer, undergoing multimodal feature fusion after separate processing by the L-GAAM and T-FIAttn modules. Finally, Fig. 16(d) presents the t-SNE visualization of the fused features. The C1 category exhibits a distinct clustering effect with the highest independence, demonstrating that the proposed model effectively extracts features of this fault type with high clarity and precision.Similarly, the C4 category forms a compact and independent feature cluster, indicating that the model effectively captures the essential distinctions of this fault type. Certain regions of the C3 category are spatially adjacent to C2 and C4, with the local features of C2 and C3 displaying significant similarity.This phenomenon is primarily attributed to the inherent similarity in signal characteristics among different fault types under real-world bearing operating conditions, making their differentiation inherently challenging.Nevertheless, the proposed model successfully distinguishes between the two fault types, as the HIFAttn module enhances the model's ability to fuse multimodal features through time–frequency interaction and local–global attention mechanisms. The effectiveness of each module has been validated through the aforementioned experiments, demonstrating the powerful

Table 10
Accuracy of each model under different noise conditions.

MODEL	SNR						
	−5	0	5	10	15	20	None
CNN	35.81 ± 0.01	51.57 ± 0.48	66.71 ± 1.26	78.37 ± 0.40	83.31 ± 0.31	85.60 ± 0.23	81.93 ± 0.45
LSTM	40.56 ± 0.13	48.44 ± 0.12	58.33 ± 0.58	68.77 ± 1.26	69.53 ± 0.80	71.89 ± 0.05	71.80 ± 0.35
Transformer	41.48 ± 0.03	57.00 ± 0.61	73.23 ± 0.41	83.13 ± 0.43	86.60 ± 0.10	88.50 ± 2.19	89.77 ± 0.04
FFT + CNN + LSTM [31]	42.60 ± 1.22	69.46 ± 0.71	82.18 ± 0.07	91.29 ± 0.21	94.64 ± 0.31	96.31 ± 0.03	96.12 ± 0.18
Multi-scale CNN + LSTM [32]	58.33 ± 0.09	75.89 ± 0.91	87.29 ± 0.07	90.60 ± 0.41	93.49 ± 0.38	94.13 ± 0.61	93.75 ± 0.22
VMTransformer[33]	76.93 ± 1.08	84.54 ± 1.78	88.45 ± 2.74	93.25 ± 0.40	95.00 ± 0.09	96.40 ± 0.02	97.11 ± 0.10
Ours	**84.70 ± 0.19**	**94.05 ± 0.15**	**97.74 ± 0.01**	**99.17 ± 0.01**	**99.53 ± 0.06**	**99.82 ± 0.01**	**99.83 ± 0.01**

feature extraction capabilities of the entire combination, which enables the model to handle fault diagnosis tasks in complex environments.

4.2.2. Comparative experiments with other models

Using the same experimental data, this study compares the FCNN-1DECBAM-BiHIFAttn model with other advanced, mainstream intelligent diagnostic methods in the field. Fig. 17 illustrates the training loss and validation accuracy curves of the seven models over 200 iterations. As shown in Fig. 17(a), the proposed model reduces the loss to approximately 0.05 within 50 iterations, achieving nearly 99 % accuracy with minimal fluctuation, which demonstrates its strong stability on the validation set. In contrast, LSTM and Transformer models show limited improvements in validation accuracy, despite faster initial loss reduction. LSTM, in particular, maintains consistently low accuracy, indicating its limited capacity for complex time-series data. Hybrid models like "FFT + CNN + LSTM [31]"、"Multi-scale CNN + LSTM [32]" and "VMTransformer[33]" exhibit greater stability, with accuracy nearing 90 %, yet still fall short of the performance achieved by the proposed method. Overall, the proposed approach is highly suited for bearing fault diagnosis tasks demanding high precision, especially in complex scenarios.

Table 5 presents the test set troubleshooting results after five independent runs for each model. Statistical analysis reveals that the proposed method achieves a maximum diagnostic accuracy of 99.32 %, with an average accuracy of 99.15 %. Compared to other models across five independent experiments under all variable working conditions, the proposed method's average accuracy is 19.22 % higher than CNN, 34.86 % higher than LSTM, 2.96 % higher than FFT + CNN + LSTM[31], 6.77 % higher than Multi-scale CNN + LSTM[32], 9.14 % higher than Transformer, and 3.75 % higher than VMTransformer[33].The standard deviation is used to quantify the variability of the model across different experiments. A smaller standard deviation indicates greater model stability. In our experimental results, the proposed model exhibits the smallest standard deviation, which demonstrates its superior stability. These results demonstrate the proposed method's excellent recognition and generalization capabilities across all 28 variable working conditions.

In real industrial environments, raw signals acquired by sensors often contain significant background noise. To evaluate the fault diagnosis performance of the proposed method under high-noise conditions, different levels of Gaussian white noise are added to the original vibration signals in this section to simulate the noise characteristics of an industrial setting. The signal-to-noise ratio (SNR) is defined as follows:

$$SNR = 10 \times \log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{20}$$

where P_{signal} denotes the effective power of the signal and P_{noise} denotes the power of the additional noise.

Fig. 18 shows the inner ring fault 1 (IRF1) fault signal with the addition of Gaussian white noise. It is obvious from Fig. 12 that when the signal-to-noise ratio is 0, the original vibration acceleration signal has been completely destroyed. When the signal-to-noise ratio is less than 0, it indicates that the energy of the noise signal is higher than that of the original signal, and under such strong noise interference, it becomes very difficult to extract useful fault signal features from the composite noise signal.

In this section, the proposed FCNN-1DECBAM-BiHIFAttn model and six other intelligent diagnostic models are evaluated across signal-to-noise ratios ranging from −5 dB to 20 dB. Each experiment was repeated five times using the same structure and parameters. Fig. 19 illustrates the average accuracy of the six methods under different noise condition, while Table 6 presents the average accuracy and standard deviation. The results show that the proposed FCNN-1DECBAM-BiHIFAttn model achieves the highest diagnostic performance across all signal-to-noise ratio scenarios. Even under strong background noise at −5 dB, the proposed method maintains an average accuracy of 85.45 %, significantly outperforming CNN (46.86 %), LSTM (36.59 %), FFT + CNN + LSTM [31] (54.72 %), Multi-scale CNN + LSTM [32] (54.04 %), Transformer (56.61 %), and VMTransformer (68.20 %). Notably, the proposed model also demonstrates the smallest standard deviation (0.37 %), indicating strong stability under noisy conditions. These findings confirm the proposed method's robustness against strong background noise.

4.3. Evaluation on aero engine bearing dataset 2

The experimental setup consists of a spindle, a load cell, and XYZ-axis sensors, as shown in Fig. 20 [34].Vibration data were collected at two acceleration measurement points (A1 and A2) under no-load conditions, with rotational speeds ranging from 6000 to 30,000 rpm.The data for each health condition were segmented into 1200 samples, with each sample containing 2048 data points.The fault types include inner race defect, roller defect, and normal (no fault) condition.Both inner race and roller defect conditions involve three defect sizes: 450 μm, 250 μm, and 150 μm.Detailed information on the dataset used in this study is provided in Table 7.

4.3.1. Results of the proposed model

Fig. 21 shows the confusion matrices corresponding to the highest and lowest accuracy results (99.94 % and 99.70 %, respectively) achieved by the proposed model on this dataset. As shown, the model exhibits extremely high classification precision, particularly in the multi-class fault diagnosis task, where nearly all fault types are correctly classified. In Fig. 21(b), the slightly lower accuracy is primarily due to misclassifications between labels C2 and C7. This can be attributed to the similarity in their time-domain or frequency-domain characteristics, which makes them more difficult to distinguish in the feature space. Nevertheless, the proposed FCNN-1DECBAM-BiHIFAttn model still achieves outstanding overall accuracy on this dataset, with very few misclassifications observed in the confusion matrix. This demonstrates the model's strong diagnostic capability under complex operating conditions. Notably, in the scenario where the model reached its highest accuracy, almost no misclassifications occurred, further validating the proposed model's ability to deliver highly accurate and stable results in complex, multi-class bearing fault diagnosis tasks.

Table 8 presents the ablation results of the proposed method. Even when the classification task becomes more challenging—expanding from four to seven classes—the proposed CNN-BiLSTM network (Model B) still achieves a high accuracy of 94.23 %, significantly outperforming the baseline Model A. With the step-by-step introduction of the FFT, 1DECBAM, and HIFAttn modules, the fault diagnosis accuracy further improves to 97.98 %, 98.81 %, and 99.83 %, respectively.This performance gain is primarily attributed to the attention mechanisms, which effectively emphasize critical spatio-temporal features under complex working conditions. Specifically, the 1DECBAM module enhances the model's ability to focus on diagnostically relevant features, while the HIFAttn module enables efficient multimodal fusion of time-domain, frequency-domain, and their interaction features. This significantly improves the model's ability to capture useful time–frequency characteristics and enhances overall robustness.Feature visualization results (Fig. 22) further confirm the model's stability and the importance of each module on the new dataset. Through t-SNE visualization, it can be observed that, although the second dataset contains more categories and exhibits greater feature overlap, the feature clustering is significantly improved after introducing the 1DECBAM and BiLSTM modules. Notably, the inclusion of the HIFAttn module further increases inter-class separability. As shown in Fig. 22(d), the seven categories are clearly separated in the feature space.

These results demonstrate that the proposed model remains capable of extracting effective features under complex conditions and maintains strong class discrimination in the time–frequency domain. The experimental findings further validate the generalization ability of the model. Each module plays a vital role on the new dataset, and in particular, the HIFAttn module shows clear advantages in multimodal feature fusion under challenging conditions. This confirms that the model not only adapts well to different data characteristics but also consistently delivers high fault diagnosis performance across varying operational environments.

4.3.2. Comparative experiments with other models

To benchmark the efficacy of the proposed model in fault diagnosis tasks, a comparative analysis involving six other state-of-the-art diagnostic algorithms was performed on the same dataset. Fig. 23 illustrates the training loss and validation accuracy curves of all seven models over 200 training epochs. As shown in Fig. 23(a), the proposed model rapidly reduces the training loss to approximately 0.05 within the first 50 epochs, while the validation accuracy approaches 100 % and remains consistently above 99 % throughout training. This demonstrates the model's strong stability and generalization capability on this dataset. In contrast, CNN and LSTM models converge quickly to a stable loss in the early stages of training but show limited improvement in validation accuracy, indicating bottlenecks when handling sequential signals independently. Although the Transformer and Multi-scale CNN + LSTM [32] models exhibit relatively high validation accuracy, they demonstrate significant fluctuations during training, indicating that their performance is limited when handling the more complex seven-class diagnostic task.The FFT + CNN + LSTM [31] and VMTransformer [33] hybrid models demonstrate improved stability and achieve validation accuracy close to 97 %. However, they still fall short of the performance achieved by the proposed model.

Table 9 presents the test results for each model averaged over five independent runs.Statistical analysis shows that the proposed method achieves a maximum diagnostic accuracy of 99.94 % and an average accuracy of 99.83 %, outperforming all six other methods.Table 10 reports the average accuracy and standard deviation of the seven models under various noisy conditions.The results indicate that the proposed FCNN-1DECBAM-BiHIFAttn model achieves the highest diagnostic performance across all SNR levels and consistently maintains the lowest standard deviation, highlighting its robustness and efficiency under noise interference.This superior performance is attributed to the synergistic effect of the model's components.First, the 1D CNN extracts local features from the raw signals, particularly critical patterns in both the time and frequency domains.Second, the BiLSTM module captures temporal dependencies and enhances the model's capacity to learn short- and long-range patterns in time-series data.Furthermore, the attention mechanism in the 1DECBAM module assigns higher weights to salient features, especially in the presence of noise or high inter-class similarity, enabling the model to focus more effectively on key spatio-temporal patterns.Finally, the HIFAttn module fuses time-domain, frequency-domain, and time–frequency interactive features, significantly improving the model's discriminative power across diverse signal patterns.The proposed FCNN-1DECBAM-BiHIFAttn model demonstrates remarkable robustness and accuracy under noisy conditions, confirming its effectiveness in complex scenarios and its broad applicability in real-world diagnostic tasks.

5. Conclusions

In this paper, an end-to-end FCNN-1DECBAM-BiHIFAttn model is proposed to address the complexity of aero-engine bearing operating conditions and the insufficient representation of fault features. The model integrates the strengths of FFT for time–frequency feature extraction, BILSTM for capturing temporal dependencies, and CNN combined with attention mechanisms for robust feature extraction and classification. The CNN-BILSTM framework, optimized and constructed in this study, provides a solid foundation for effective feature extraction from both time-domain and frequency-domain signals.The inclusion of the 1DECBAM module significantly enhances the model's ability to retain critical features extracted by CNNs while amplifying key spatial and channel features. The Hybrid Interaction-Fusion Attention (HIFAttn) framework, which includes the T-FIAttn and L-GAAM modules, performs multi-modal feature fusion and captures essential relationships between time and frequency domains, dynamically highlighting critical features and contributing to the model's excellent performance under diverse operating conditions. Notably, the L-GAAM module strengthens critical temporal features from the BILSTM, while the T-FIAttn module captures latent feature correlations across the time–frequency domain.Experimental results from two aero-engine datasets demonstrate that the proposed model outperforms current state-of-the-art methods, achieving diagnostic accuracies of 99.32 % and 99.94 %, respectively, and showing exceptional stability and robustness even under high-noise conditions. These results underline the model's high accuracy, strong generalization, and noise resilience, making it highly suitable for practical applications in aero-engine bearing fault diagnosis.

Future work will focus on enhancing the interpretability and visual representation of the features learned by the model to provide more insights into its decision-making process. Additionally, exploring the use of multimodal data inputs, such as vibration, acoustic, and thermal signals, is expected to further enhance the diagnostic performance, especially for more complex and challenging fault scenarios.

CRediT authorship contribution statement

Delin Huang: Writing – review & editing, Methodology, Funding acquisition. **Xiangdong Su:** Writing – original draft, Validation, Software. **Jinghui Yang:** Supervision, Resources, Investigation. **Shichang Du:** Resources, Project administration, Formal analysis, Conceptualization. **Dexian Wang:** Supervision, Software. **Qiuyu Ran:** Visualization.

Funding

This work was supported by Shanghai Sailing Program (Grant No. 21YF1400500), National Natural Science Foundation of China (Grant No. 52275499, 92467101).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Delin Huang reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by Shanghai Sailing Program (Grant No. 21YF1400500), National Natural Science Foundation of China (Grant No. 52275499, 92467101).

Data availability

Data will be made available on request.

References

[1] J. Zhou, X. Yang, L. Liu, Y. Wang, J. Wang, G. Hou, Fuzzy broad learning system combined with feature-engineering-based fault diagnosis for bearings, Machines 10 (12) (2022) 1229, https://doi.org/10.3390/machines10121229.

[2] L. Xu, S. Chatterton, P. Pennacchi, Rolling element bearing diagnosis based on singular value decomposition and composite squared envelope spectrum, Mech.

Syst. Sig. Process. 148 (2021) 107174, https://doi.org/10.1016/j.ymssp.2020.107174.

[3] H. Zhu, Z. He, J. Wei, et al., Bearing fault feature extraction and fault diagnosis method based on feature fusion, Sensors 21 (7) (2021) 2524, https://doi.org/10.3390/s21072524.

[4] Z. Wang, D. Shi, Y. Xu, et al., Early rolling bearing fault diagnosis in induction motors based on on-rotor sensing vibrations, Measurement 222 (2023) 113614, https://doi.org/10.1016/j.measurement.2023.113614.

[5] J.S. Lee, T.M. Yoon, K.B. Lee, Bearing fault detection of IPMSMs using zoom FFT, J. Electr. Eng. Technol. 11 (5) (2016) 1235–1241, https://doi.org/10.5370/JEET.2016.11.5.1235.

[6] G. Yang, Y. Hu, Q. Shi, Fault diagnosis of bearings based on SSWT, bayes optimisation and CNN, Pol. Marit. Res. 30 (3) (2023) 132–141, https://doi.org/10.2478/pomr-2023-0046.

[7] Z. Shan, Z. Wang, J. Yang, et al., Novel time–frequency mode decomposition and information fusion for bearing fault diagnosis under varying-speed condition, IEEE Trans. Instrum. Meas. 72 (2023) 1–10, https://doi.org/10.1109/TIM.2023.3260275.

[8] D. Huang, Y. Li, S. Guan, et al., A novel collaborative diagnosis approach of incipient faults based on VMD and SCN for rolling bearing, Optimal Control Appl. Methods 44 (3) (2023) 1617–1631, https://doi.org/10.1002/oca.2820.

[9] C. He, T. Wu, R. Gu, et al., Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm–extreme learning machine, Measurement 173 (2021) 108636, https://doi.org/10.1016/j.measurement.2020.108636.

[10] X. Zhang, J. Zhao, H. Teng, et al., A novel faults detection method for rolling bearing based on RCMDE and ISVM, J. Vibroeng. 21 (8) (2019) 2148–2158, https://doi.org/10.21595/jve.2019.20815.

[11] Y. Yang, C. Xi. Rolling bearing fault diagnosis based on MFDFA-SPS and ELM. Mathem. Probl. Eng. 2022(1): 4034477. Doi: 10.1155/2022/4034477.

[12] Z. Jiang, J. Zhou, Y. Ma, Fault diagnosis for rolling bearing based on parameter transfer Bayesian network, Qual. Reliab. Eng. Int. 38 (8) (2022) 4291–4308, https://doi.org/10.1002/qre.3208.

[13] Z. Liu, K. Lv, C. Zheng, et al., A fault diagnosis method for rolling element bearings based on ICEEMDAN and Bayesian network, J. Mech. Sci. Technol. 36 (5) (2022) 2201–2212, https://doi.org/10.1007/s12206-022-0404-3.

[14] Z. Wang, P. Liang, R. Bai, et al., Few-shot fault diagnosis for machinery using multi-scale perception multi-level feature fusion image quadrant entropy, Adv. Eng. Inf. 63 (2025) 102972, https://doi.org/10.1016/j.aei.2024.102972.

[15] Z. Wang, M. Zhang, H. Chen, et al., A generalized fault diagnosis framework for rotating machinery based on phase entropy, Reliab. Eng. Syst. Saf. 256 (2025) 110745, https://doi.org/10.1016/j.ress.2024.110745.

[16] D. Zhao, X. Huang, T. Wang, et al., Generalized reassigning transform: Algorithm and applications, Reliab. Eng. Syst. Saf. 255 (2025) 110677, https://doi.org/10.1016/j.ress.2024.110677.

[17] P. Xu, L. Zhang, A fault diagnosis method for rolling bearing based on 1D-ViT model, IEEE Access 11 (2023) 39664–39674, https://doi.org/10.1109/ACCESS.2023.3268534.

[18] J. Zhu, T. Hu, B. Jiang, et al., Intelligent bearing fault diagnosis using PCA–DBN framework, Neural Comput. Applic. 32 (2020) 10773–10781, https://doi.org/10.1007/s00521-019-04612-z.

[19] H. Shao, M. Xia, G. Han, et al., Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer convolutional neural network and thermal images, IEEE Trans. Ind. Inf. 17 (5) (2020) 3488–3496, https://doi.org/10.1109/TII.2020.3005965.

[20] F. Zou, H. Zhang, S. Sang, et al., Bearing fault diagnosis based on combined multi-scale weighted entropy morphological filtering and bi-LSTM, Appl. Intell. 1–18 (2021), https://doi.org/10.1007/s10489-021-02229-1.

[21] Y. Liu, J. Li, Q. Li, et al., Transfer learning with inception ResNet-based model for rolling bearing fault diagnosis, J. Adv. Mech. Des. Syst. Manuf. 16 (2) (2022), https://doi.org/10.1299/jamdsm.2022jamdsm0023. JAMDSM0023-JAMDSM0023.

[22] X. Chen, B. Zhang, D. Gao, Bearing fault diagnosis base on multi-scale CNN and LSTM model, J. Intell. Manuf. 32 (4) (2021) 971–987, https://doi.org/10.1007/s10845-020-01600-2.

[23] Y. Guo, J. Mao, M. Zhao, Rolling bearing fault diagnosis method based on attention CNN and BILSTM network, Neural Process. Lett. 55 (3) (2023) 3377–3410, https://doi.org/10.1007/s11063-022-11013-2.

[24] D. Zuo, T. Tang, M. Chen, Rolling bearing fault diagnosis based on multi-scale weighted visibility graph and multi-channel graph convolution network, Meas. Sci. Technol. 34 (11) (2023) 115019, https://doi.org/10.1088/1361-6501/ace7e5.

[25] D. Zhao, W. Cai, L. Cui, Multi-perception Graph Convolutional Tree-embedded Network for Aero-engine Bearing Health Monitoring with Unbalanced Data[J], Reliab. Eng. Syst. Saf. 110888 (2025), https://doi.org/10.1016/j.ress.2025.110888.

[26] Z. Li, F. Liu, W. Yang, et al., A survey of convolutional neural networks: analysis, applications, and prospects, IEEE Trans. Neural Networks Learn. Syst. 33 (12) (2021) 6999–7019, https://doi.org/10.1109/TNNLS.2021.3084827.

[27] S. Woo, J. Park, J.Y. Lee et al (2018) Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV) 3–19.

[28] J. Park, S. Woo, J.Y. Lee et al. Bam: Bottleneck attention module.arXiv preprint arXiv:1807.06514. (2018). Doi: 10.48550/arXiv.1807.06514.

[29] S.M. Nacer, B. Nadia, R. Abdelghani, et al., A novel method for bearing fault diagnosis based on BILSTM neural networks, Int J Adv Manuf Technol 125 (3) (2023) 1477–1492, https://doi.org/10.1007/s00170-022-10792-1.

[30] L. Hou, H. Yi, Y. Jin et al. Inter-shaft bearing fault diagnosis based on aero-engine system: a benchmarking dataset study. J. Dynam. Monitor. Diagn. (2023), 228–242. Doi: 10.37965/jdmd.2023.314.

[31] S. Cen, D.O. Kim, C.G. Lim, A fused CNN-LSTM model using FFT with application to real-time power quality disturbances recognition, Energy Sci Eng 11 (7) (2023) 2267–2280, https://doi.org/10.1007/s10845-020-01600-2.

[32] X. Chen, B. Zhang, D. Gao, Bearing fault diagnosis base on multi-scale CNN and LSTM model, J Intell Manuf 32 (2021) 971–987, https://doi.org/10.1007/s10845-020-01600-2.

[33] D. Wang, X. Xu, J. Yang, et al., Fault diagnosis of planetary gears in noisy environments using a VMTransformer model, Meas. Sci. Technol. 36 (3) (2025) 036209, https://doi.org/10.1088/1361-6501/adb2b0.

[34] A.P. Daga, A. Fasana, S. Marchesiello, et al., The Politecnico di Torino rolling bearing test rig: description and analysis of open access data, Mech. Syst. Sig. Process. 120 (2019) 252–273, https://doi.org/10.1016/j.ymssp.2018.10.010.