machines

MDPI

Article

MFETA-Net: Multi-Branch Frequency Enhancement and Temporal Attention for Small-Sample Rolling Bearing Fault Diagnosis

Chiming Wang [1], Yiying Zhou [1], Dongke Zheng [1], Chengming Huang [1], Shunzhi Zhu [1,*], Zhenjun Li [2], Bingkun Wu [3] and Liangqing Guan [4]

[1] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China; wangchiming009@163.com (C.W.)
[2] Fujian Province Shipbuilding Industry Group Co., Ltd., Fuzhou 350501, China
[3] Zhongshu Xinke (Zhongshu (Xiamen) Information Technology Co., Ltd.), Xiamen 361000, China
[4] Fujian Fuchuan Marine Engineering Technology Research Institute Co., Ltd., Fuzhou 350501, China
* Correspondence: szzhu@xmut.edu.cn

Abstract

In practical industrial applications, rolling bearing fault samples are often scarce and costly to annotate, making small-sample fault diagnosis challenging. To address this issue, this paper proposes a frequency-aware time–frequency representation learning framework, named Multi-branch Frequency Enhancement Temporal Attention Network (MFETA-Net). In the proposed framework, dual-channel vibration signals are first transformed into time–frequency representations using the Short-Time Fourier Transform (STFT). Then, a multi-branch frequency enhancement encoder is used to extract local frequency-band patterns, cross-band correlations, and frequency variation features. A temporal-frequency dependency modeling mechanism preserves the correspondence between temporal positions and frequency distributions during sequential modeling, while a temporal attention aggregation module emphasizes diagnostically important regions. Extensive experiments on the CWRU and HUST bearing datasets show that MFETA-Net achieves accuracies of 77.26% and 79.60% under the smallest training setting, respectively, indicating its capability to learn discriminative fault representations from limited labeled samples. Ablation studies further verify the effectiveness of each proposed module, while noise experiments confirm the robustness of the proposed framework under controlled noisy conditions.

Keywords: rolling bearing fault diagnosis; small-sample learning; STFT; time–frequency representation; frequency-aware learning; attention mechanism

Check for updates

Academic Editor: Ahmed Abu-Siada

Received: 16 July 2026
Revised: 9 August 2026
Accepted: 14 August 2026
Published: 20 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Rolling bearings are critical components in rotating machinery, and their health conditions directly affect the safety and reliability of industrial equipment, including wind turbines, rail transit systems, and aero-engines [1]. During long-term operation under high-speed, heavy-load, and complex working conditions, rolling bearings are prone to various failures, such as inner-race faults, outer-race faults, and rolling-element faults, which may cause unexpected shutdowns and significant economic losses [2–4]. Therefore, reliable rolling bearing fault diagnosis is essential for ensuring operational safety and reducing maintenance costs [5,6].

Traditional bearing fault diagnosis methods mainly depend on handcrafted features extracted from vibration signals and machine learning classifiers, such as support vector machines (SVMs) [7] and random forests [8]. Although these methods have achieved satisfactory results under specific conditions, their performance strongly relies on expert knowledge and manually designed features, limiting their ability to characterize complex nonlinear fault patterns.

With the rapid development of deep learning, convolutional neural networks (CNNs) [9], recurrent neural networks (RNNs) [10], and Transformer-based models [11] have been widely applied in intelligent fault diagnosis. Compared with traditional approaches, deep learning methods can automatically learn hierarchical fault representations from raw signals or transformed features, reducing the dependence on manual feature extraction. However, most existing models require sufficient labeled samples for training, while fault samples are usually limited in practical applications. Under small-sample conditions, deep learning models often suffer from overfitting and insufficient feature learning, resulting in unstable diagnostic performance and poor generalization ability [12–14].

To preserve fault-related information, time–frequency representations have been increasingly introduced into deep learning-based diagnosis methods. Techniques such as short-time Fourier transform (STFT) [15], continuous wavelet transform (CWT) [16] and empirical mode decomposition (EMD) [17] transform one-dimensional vibration signals into two-dimensional representations by simultaneously describing temporal variations and frequency distributions. Guo et al. [18] constructed continuous wavelet transform scalograms (CWTSs) and utilized CNNs for fault classification, demonstrating the effectiveness of combining time–frequency representations with deep feature learning. Djaballah et al. [19] employed continuous wavelet transform (CWT) to generate two-dimensional time–frequency images and utilized CNNs for bearing fault diagnosis, further confirming the effectiveness of image-based deep feature learning. These studies indicate that time–frequency representations preserve richer temporal and spectral information than one-dimensional vibration signals, providing a more informative input for deep feature learning.

CNN-based methods have been extensively investigated due to their powerful spatial feature extraction capability. Eren et al. [20] applied a one-dimensional CNN to raw vibration signals for end-to-end fault detection. Jia et al. [21] proposed a deep normalized CNN (DNCNN) to automatically learn representative fault features under imbalanced conditions. Zhang et al. [22] developed a wide first-layer kernel CNN (WDCNN) to improve robustness against noise and variable operating conditions. Meng et al. [23] introduced a multi-scale CNN based on local central moment discrepancy to capture fault characteristics at different scales. Fu et al. [24] combined CNN and data augmentation for diagnosis under limited training samples, while Hou et al. [25] further verified the effectiveness of CNNs in extracting features from STFT-based time–frequency images. Nevertheless, conventional CNN-based methods generally treat time–frequency representations as ordinary images and mainly focus on local spatial feature extraction. As a result, frequency-dependent fault characteristics and cross-band spectral correlations are only implicitly learned rather than explicitly modeled, which limits their discriminative capability under small-sample conditions.

To enhance temporal feature learning, recurrent networks have been integrated with CNN-based frameworks. Han et al. [26] proposed a CNN-LSTM-GRU model to extract spatial features and temporal dependencies from time–frequency images. Zhao et al. [27] developed an LSTM-based framework for modeling temporal correlations in monitoring data. Bian et al. [28] proposed a WDCNN-BiLSTM Siamese network for small-sample rolling bearing fault diagnosis, where WDCNN and BiLSTM were combined to extract local features and temporal dependencies, while metric learning was introduced to improve

feature discrimination under limited samples. Da et al. [29] proposed a CNN-BiGRU architecture by combining convolutional extraction with bidirectional temporal modeling. Song et al. [30] designed a CNN-BiLSTM network for bearing diagnosis under multiple operating conditions and limited samples. However, most CNN-RNN methods first flatten two-dimensional time–frequency representations into one-dimensional sequences before recurrent modeling. Such a transformation inevitably mixes temporal and frequency information, weakening the natural correspondence between temporal positions and frequency structures. Consequently, frequency-related characteristics cannot be effectively preserved during sequence modeling, limiting the learning of discriminative fault representations under small-sample conditions.

Recently, attention mechanisms have been introduced into fault diagnosis frameworks to improve feature representation capability. Wang et al. [31] proposed an SDP-based SE-CNN method, where channel attention was used to enhance fault-sensitive features. Ding et al. [32] developed a time–frequency Transformer to capture long-range dependencies in STFT representations. Xu et al. [33] integrated GRU, ResNet, and CBAM for multimodal feature fusion, while Chen et al. [34] proposed a cross-space multiscale CNN-Transformer architecture to capture local and global fault characteristics.

Despite the progress of existing deep learning-based methods, three limitations remain under small-sample conditions. First, CNN-based approaches mainly learn local spatial patterns from time–frequency representations, while frequency-dependent characteristics are not explicitly modeled. Second, CNN–RNN frameworks often transform two-dimensional time–frequency maps into one-dimensional sequences, which may weaken the correspondence between temporal positions and frequency distributions. Third, existing attention-based methods mainly focus on feature reweighting or global dependency modeling, while their coordination with frequency-oriented representation learning remains limited. These limitations motivate a unified framework that jointly enhances frequency representations, preserves temporal–frequency correspondence, and adaptively emphasizes diagnostically important responses.

To address these limitations, this paper proposes a Multi-branch Frequency Enhancement Temporal Attention Network (MFETA-Net) for small-sample rolling bearing fault diagnosis. Specifically, MFFE enhances complementary frequency characteristics, TFDE preserves temporal–frequency correspondence during sequential modeling, and MTAA adaptively emphasizes diagnostically important temporal responses.

The main contributions of this paper are summarized as follows:

- A unified MFETA-Net is proposed for small-sample rolling bearing fault diagnosis.
- MFFE captures complementary local frequency patterns, cross-band interactions, and adjacent-frequency variations.
- TFDE preserves temporal–frequency correspondence, while MTAA adaptively emphasizes diagnostically important temporal responses.
- Experiments on the CWRU and HUST datasets demonstrate the effectiveness and noise robustness of MFETA-Net under small-sample conditions.

The remainder of this paper is organized as follows. Section 2 presents the proposed MFETA-Net framework, including the overall architecture, the Multi-branch Frequency Feature Enhancement Encoder (MFFE), the Temporal–Frequency Dependency Enhancement (TFDE) module, and the Multi-head Temporal Attention Aggregation (MTAA) module. Section 3 describes the experimental datasets, implementation details, and comparative analyses, including ablation studies, feature visualization, and noise robustness evaluation. Finally, Section 4 concludes this paper and discusses its limitations as well as future research directions.

2. Methodology

2.1. Overall Architecture of MFETA-Net

The proposed MFETA-Net is built upon a frequency-aware time–frequency representation learning framework, as shown in Figure 1. Starting from STFT spectrograms, the proposed framework progressively performs frequency structure enhancement, frequency-consistent temporal dependency learning, and adaptive temporal–frequency feature aggregation, enabling discriminative fault representations to be learned in a unified manner. Unlike conventional CNN-based methods that mainly treat time–frequency representations as ordinary images, MFETA-Net explicitly preserves frequency-related characteristics and their temporal correspondence during feature extraction and sequence modeling, thereby improving the discriminative capability of learned representations under limited training samples.

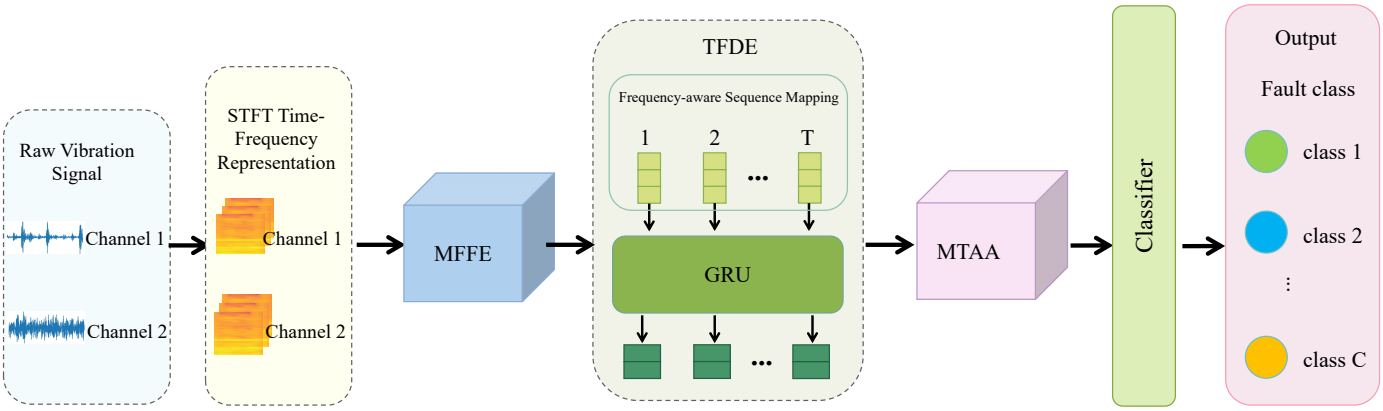


Figure 1. Overall architecture of MFETA-Net. Channel 1 and Channel 2 denote the dual-channel vibration inputs collected from different bearing datasets. The output dimension C corresponds to the number of diagnostic classes, with 10 classes for the CWRU dataset and 9 classes for the HUST dataset; the detailed class definitions are given in Tables 1 and 2, respectively.

Table 1. Classification details of the CWRU bearing dataset.

Label	Class Name	Health State	Fault Location	Fault Diameter
0	Normal	Normal	–	–
1	B_007	Fault	Ball	0.007 inch
2	B_014	Fault	Ball	0.014 inch
3	B_021	Fault	Ball	0.021 inch
4	IR_007	Fault	Inner race	0.007 inch
5	IR_014	Fault	Inner race	0.014 inch
6	IR_021	Fault	Inner race	0.021 inch
7	OR_007	Fault	Outer race	0.007 inch
8	OR_014	Fault	Outer race	0.014 inch
9	OR_021	Fault	Outer race	0.021 inch

Given a raw vibration signal x, a two-dimensional time–frequency representation X is first obtained by STFT. The MFFE module then extracts complementary frequency information from local band patterns, cross-band correlations, and frequency variation trends, producing the enhanced feature representation F. Subsequently, TFDE converts the two-dimensional features into a time-dominant sequence while preserving the frequency-related characteristics of each time step. Finally, MTAA adaptively aggregates temporal–frequency features according to their diagnostic importance, emphasizing fault-sensitive regions while suppressing redundant responses, thereby producing a more discriminative global representation for fault classification.

Table 2. Classification details of the HUST bearing dataset.

Label	Class Name	Health State	Fault Type
0	H	Healthy	Normal
1	0.5X_I	Fault	Medium inner-race fault
2	I	Fault	Severe inner-race fault
3	0.5X_O	Fault	Medium outer-race fault
4	O	Fault	Severe outer-race fault
5	0.5X_B	Fault	Medium ball fault
6	B	Fault	Severe ball fault
7	0.5X_C	Fault	Medium combination fault
8	C	Fault	Severe combination fault

Note: The prefix "0.5X" denotes the medium-damage category, whereas the corresponding class without this prefix denotes the severe-damage category. A combination fault represents the simultaneous occurrence of inner-race and outer-race faults. The terms "medium" and "severe" are categorical descriptors rather than quantitative measurements of defect size or damage ratio.

The overall feature learning process can be formulated as:

$$X = \mathrm{STFT}(x), \tag{1}$$

$$F = \mathrm{MFFE}(X), \tag{2}$$

$$H = \mathrm{TFDE}(F), \tag{3}$$

$$z = \mathrm{MTAA}(H), \tag{4}$$

$$\hat{y} = \mathrm{Classifier}(z). \tag{5}$$

Here, X denotes the STFT spectrogram, F the frequency-enhanced feature map, H the sequence representation after TFDE, and z the global discriminative feature used for fault classification.

In the implemented network, each MFFE branch outputs 32 channels, which are fused into a 64-channel representation. Deeper convolution blocks further increase the feature dimension to 128 channels. TFDE produces a temporal sequence with a 128-dimensional feature at each time step and employs a single-layer GRU for temporal dependency modeling. MTAA adopts four attention heads and outputs a 128-dimensional global feature for classification.

2.2. STFT

Bearing vibration signals are typically non-stationary, and their frequency components vary over time. Therefore, conventional time-domain or frequency-domain analysis alone is insufficient to comprehensively characterize fault features. To effectively extract the time–frequency information of vibration signals, STFT is employed to transform the raw vibration signals into time–frequency representations. STFT performs localized spectral analysis on successive windowed signal segments to obtain a joint time–frequency representation [35] and has been widely applied to rolling bearing fault analysis [36]. The signal within each short-time window is assumed to be approximately stationary. The STFT is defined as

$$X(t, f) = \int_{-\infty}^{+\infty} x(\tau) w(\tau - t) e^{-j2\pi f \tau} \, d\tau, \tag{6}$$

where $X(t, f)$ denotes the time–frequency representation of the signal, $x(\tau)$ is the original vibration signal, $w(\tau - t)$ is the window function centered at time t, and t and f represent the time and frequency variables, respectively.

Compared with one-dimensional vibration signals, STFT spectrograms preserve both temporal and spectral information, making transient responses and frequency distributions

more distinguishable. These explicit frequency structures provide a suitable basis for the subsequent frequency-aware feature extraction and temporal dependency modeling.

2.3. Multi-Branch Frequency Feature Enhancement Encoder (MFFE)

The STFT representation contains diverse frequency patterns, including local spectral responses, cross-band correlations, and frequency variations, which are difficult to fully capture using single-scale convolution. Multi-scale convolutional structures have been shown to capture complementary fault features under different receptive fields in rolling bearing diagnosis [23,37]. Therefore, a multi-branch structure is employed to enhance complementary frequency features, as illustrated in Figure 2.

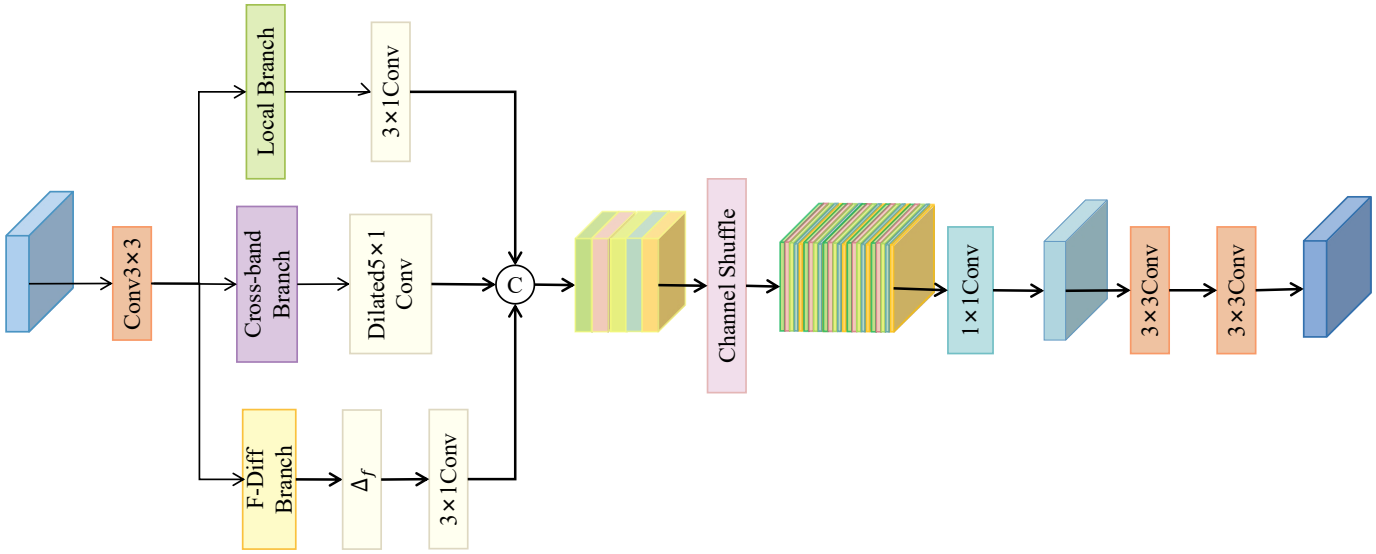


Figure 2. Architecture of the proposed MFFE.

Given the input time–frequency representation

$$X \in \mathbb{R}^{C \times H \times W}, \tag{7}$$

an initial feature embedding is first obtained through a convolutional layer:

$$F_0 = \delta(\mathrm{BN}(\mathrm{Conv}_{3\times3}(X))), \tag{8}$$

where $\mathrm{Conv}_{3\times3}$ denotes the two-dimensional convolution operation, BN represents batch normalization, and $\delta(\cdot)$ is the ReLU activation function.

Different bearing fault mechanisms generate distinct spectral patterns, such as localized frequency peaks, harmonic correlations, and abrupt energy variations. Accordingly, local frequency patterns, cross-band correlations, and adjacent-frequency variations are considered complementary characteristics for fault representation. Therefore, MFFE employs three parallel branches to model these physically meaningful frequency characteristics. The local branch employs a 3×1 convolution along the frequency axis to extract local spectral patterns from adjacent frequency bands:

$$F_l = \delta(\mathrm{BN}(\mathrm{Conv}_{3\times1}(F_0))). \tag{9}$$

The cross-band branch adopts dilated convolution to enlarge the receptive field along the frequency dimension, enabling the network to capture interactions over a wider frequency range [37]:

$$F_h = \delta\left(\mathrm{BN}\left(\mathrm{DConv}_{5\times1}^{d=2}(F_0)\right)\right), \tag{10}$$

The dilation rate is set to $d = 2$, which enlarges the receptive field along the frequency dimension without increasing the convolution kernel size.

To explicitly characterize local variations between adjacent frequency bins, an F-Diff branch is introduced to calculate the first-order frequency difference:

$$D_f = \nabla_f F_0, \tag{11}$$

$$\nabla_f F_0(i,j) = F_0(i+1,j) - F_0(i,j), \tag{12}$$

followed by a convolution layer to further extract discriminative frequency-difference features:

$$F_d = \delta\left(\mathrm{BN}\left(\mathrm{Conv}_{3\times1}(D_f)\right)\right). \tag{13}$$

This branch enhances the representation of spectral discontinuities and improves the sensitivity to fault-related frequency boundaries and frequency variation trends.

The outputs of the three branches are concatenated to form a unified frequency representation:

$$F_c = \mathrm{Concat}(F_l, F_h, F_d). \tag{14}$$

Considering the complementary information and potential channel redundancy among different branches, a Channel Shuffle operation is employed to promote cross-channel information interaction [38]:

$$F_s = \mathrm{Shuffle}(F_c), \tag{15}$$

followed by a 1×1 convolution for feature fusion and channel compression:

$$F_f = \delta(\mathrm{BN}(\mathrm{Conv}_{1\times1}(F_s))). \tag{16}$$

Finally, the fused features are further refined by two 3×3 convolutional blocks with downsampling, producing the final frequency-enhanced representation for subsequent temporal–frequency dependency modeling.

2.4. Temporal–Frequency Dependency Enhancement (TFDE)

Time–frequency features contain both temporal evolution and frequency structures. Direct sequence transformation may disrupt their correspondence; therefore, the frequency-related characteristics associated with each temporal position should be preserved during temporal modeling.

Let the output feature map of MFFE be

$$F \in \mathbb{R}^{B\times C\times H\times W}, \tag{17}$$

where B, C, H, and W denote the batch size, the number of channels, the frequency dimension, and the temporal dimension, respectively. The MFFE feature map contains both temporal evolution and frequency characteristics, but directly flattening it into a sequence may mix frequency components and weaken their correspondence with temporal positions. Therefore, TFDE organizes the feature map along the temporal dimension and aggregates

the frequency information of each temporal position through adaptive pooling and linear projection. The obtained frequency-consistent sequence can be represented as:

$$S = \{s_1, s_2, \cdots, s_T\}, \quad S \in \mathbb{R}^{B \times T \times d}, \tag{18}$$

where T denotes the sequence length and d is the feature dimension of each time step.

Although the spectral patterns of many bearing faults may remain relatively stable, local transient responses can vary across temporal positions. Therefore, the frequency-consistent sequence is modeled by GRU to capture complementary temporal dependencies when present [26], as illustrated in Figure 3.

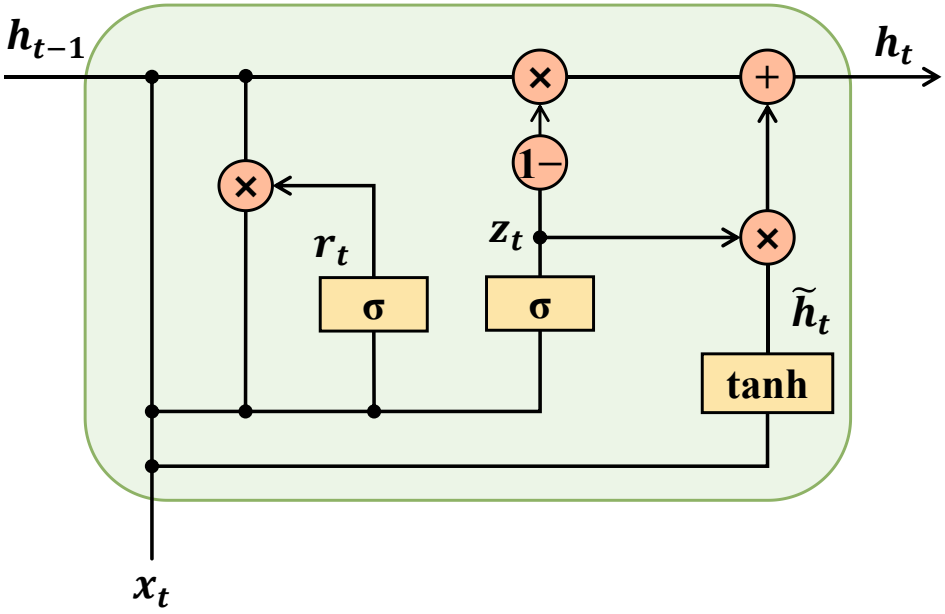


Figure 3. Architecture of the GRU-based temporal dependency modeling module.

At time step t, let x_t denote the input feature. The GRU update equations are given as follows.

Update gate:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}). \tag{19}$$

Reset gate:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}). \tag{20}$$

Candidate hidden state:

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1})). \tag{21}$$

Hidden state update:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \tag{22}$$

where $\odot$ denotes element-wise multiplication. The hidden states generated by the GRU are represented as

$$H = [h_1, h_2, \cdots, h_T] \in \mathbb{R}^{B \times T \times d}, \tag{23}$$

where H denotes the output time–frequency sequence representation, and d denotes the feature dimension of each hidden state.

Compared with conventional flatten-based sequence modeling, TFDE avoids indiscriminate mixing of temporal and frequency dimensions while preserving the frequency-

related characteristics associated with each temporal position. Therefore, the generated representation provides a more frequency-consistent feature space for subsequent attention aggregation.

2.5. Multi-Head Temporal Attention Aggregation (MTAA)

After the TFDE module, the network obtains a frequency-consistent temporal sequence representation that preserves both frequency characteristics and temporal dependencies. However, fault-related responses are usually distributed unevenly across different temporal positions, and directly aggregating all features may weaken important diagnostic patterns. Accordingly, different temporal positions are considered to contribute unequally to fault discrimination. Therefore, MTAA is introduced to refine inter-temporal relationships and adaptively aggregate discriminative temporal features through attention-based weighting. The overall architecture of MTAA is illustrated in Figure 4.

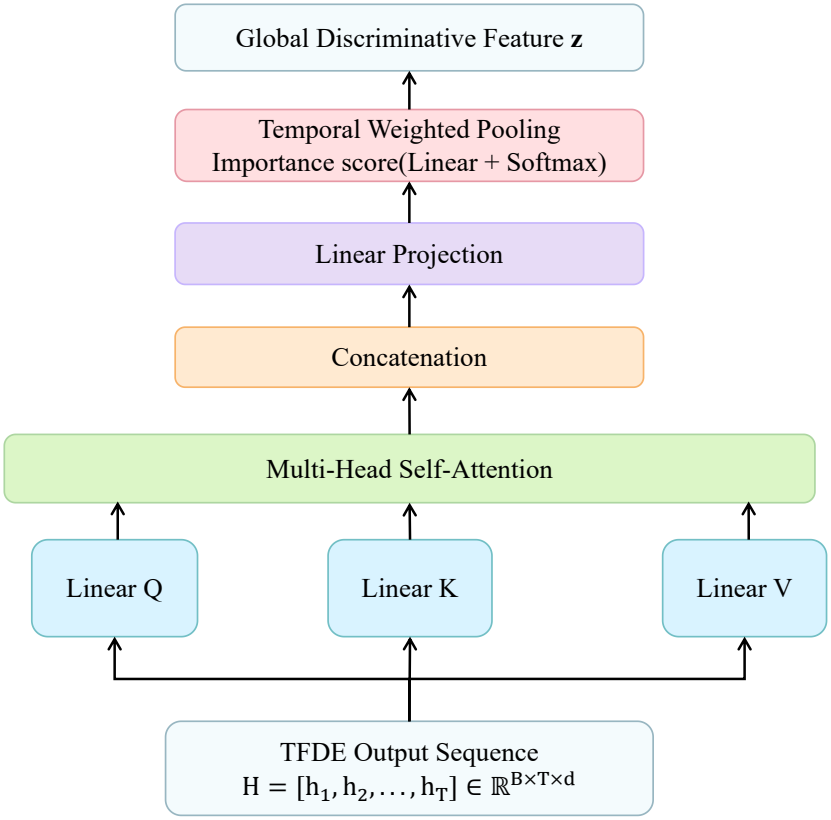


Figure 4. Architecture of the proposed MTAA.

Let the output sequence of TFDE be

$$H = [h_1, h_2, \cdots, h_T] \in \mathbb{R}^{B \times T \times d}, \tag{24}$$

where B denotes the batch size, T denotes the sequence length, and d represents the feature dimension.

Self-attention has been applied to capture long-range dependencies in time–frequency representations for rolling bearing fault diagnosis [32]. The input sequence is first projected into the query (Q), key (K), and value (V) representations:

$$Q = HW_Q, \quad K = HW_K, \quad V = HW_V, \tag{25}$$

where W_Q, W_K, W_V denote the learnable projection matrices for the query, key, and value, respectively.

For the i-th attention head, the attention output is computed as:

$$\text{head}_i = \text{softmax}\left(\frac{Q_i K_i^{\mathrm{T}}}{\sqrt{d_k}}\right) V_i, \tag{26}$$

where $d_k = d/M$ denotes the feature dimension of each attention head and M is the number of attention heads.

The outputs of all attention heads are concatenated and linearly projected to obtain the final attention representation:

$$\begin{aligned} H_a &= \text{Concat}(\text{head}_1, \cdots, \text{head}_M), \\ F_a &= H_a W_o. \end{aligned} \tag{27}$$

where W_o denotes the output projection matrix.

To further emphasize critical fault responses, a learnable temporal weighting strategy is introduced after the multi-head attention module. The importance score of each time step is first computed as

$$\alpha_t = \frac{\exp(w^{\mathrm{T}} f_t)}{\sum_{j=1}^{T} \exp(w^{\mathrm{T}} f_j)}, \tag{28}$$

where f_t denotes the feature representation of the t-th time step in F_a, and w is a learnable weight vector that measures the contribution of each time step to the final diagnosis. Here, α_t represents the importance of the t-th temporal feature, and a larger value indicates that the corresponding temporal region contributes more strongly to the final representation.

The global discriminative feature is then obtained through weighted aggregation,

$$z = \sum_{t=1}^{T} \alpha_t f_t. \tag{29}$$

Finally, the resulting feature vector z is fed into the classifier to predict the fault category.

3. Experimental Setup and Results

3.1. Datasets and Experimental Settings

3.1.1. CWRU Dataset

To evaluate the effectiveness of the proposed method, experiments were first conducted on the Case Western Reserve University (CWRU) bearing dataset, one of the most widely used benchmark datasets for bearing fault diagnosis. The experimental platform consists of an induction motor, a shaft, a torque transducer, and bearings installed at the drive end (DE) and fan end (FE), as shown in Figure 5 [39]. Vibration signals were collected using accelerometers mounted on the bearing housing.

Dual-channel vibration signals measured at the drive end (DE) and fan end (FE) were selected from the CWRU dataset and used as model inputs. The dataset covers multiple operating speeds, including 1730, 1750, and 1772 rpm, and contains ten bearing conditions, including one healthy condition and nine fault conditions with different fault locations (inner race, outer race, and rolling element) and fault sizes (0.007, 0.014, and 0.021 inch), as shown in Table 1.

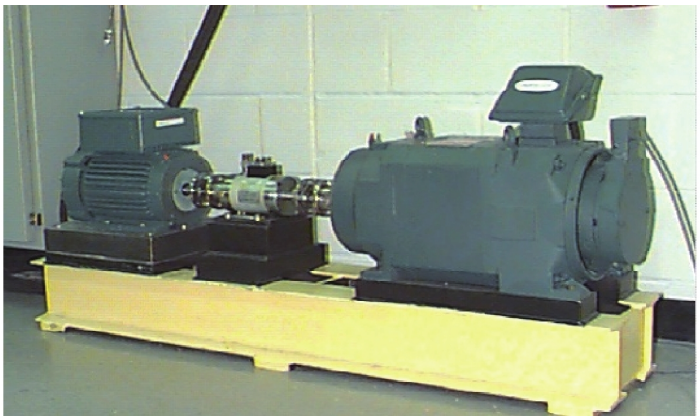


Figure 5. Experimental platform of the CWRU bearing dataset.

3.1.2. HUST Dataset

To further evaluate the proposed method under more complex fault conditions, experiments were also conducted on the public bearing dataset released by Huazhong University of Science and Technology (HUST). The test platform consists of a driving motor, a transmission shaft, test bearings, and a tri-axial accelerometer, as shown in Figure 6 [40]. The vibration signals were collected at a sampling frequency of 25.6 kHz.

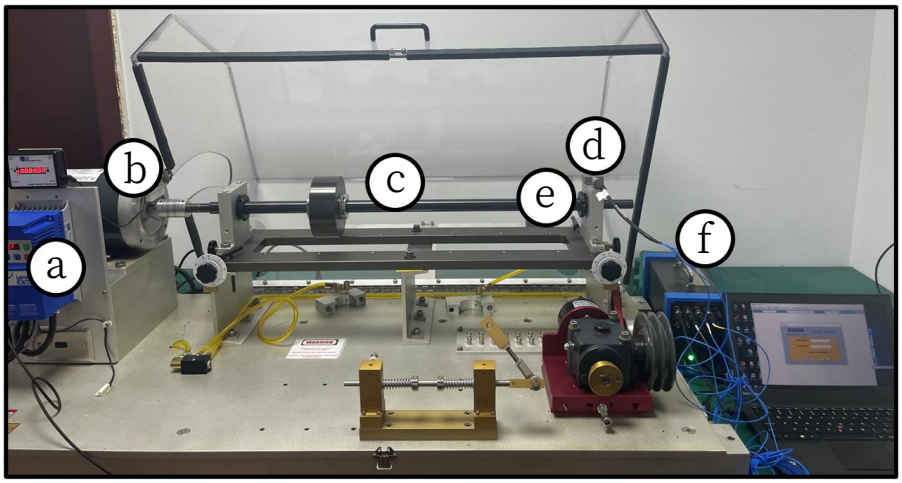


Figure 6. Experimental platform of the HUST bearing dataset: (a) speed controller; (b) motor; (c) shaft; (d) accelerometer; (e) bearing; (f) data acquisition board.

Among the operating conditions available in the HUST dataset, four constant-speed conditions (65, 70, 75, and 80 Hz) were selected to construct the small-sample fault diagnosis tasks. The vibration signals measured along the X and Y directions were used as dual-channel inputs. The dataset contains nine bearing health states, including one healthy state and medium and severe inner-race, outer-race, ball, and combination faults, as summarized in Table 2. Compared with the CWRU dataset, the inclusion of paired damage categories and combination faults makes fine-grained fault classification more challenging.

3.1.3. Small-Sample Partition Strategy

Each raw vibration record was first divided chronologically into two equal parts before window segmentation. The first half was used for model development, including training and validation, while the second half was reserved for testing. Overlapping windows with a length of 2048 and a stride of 80 were generated from the first half to provide sufficient candidate samples under limited-data conditions, with 10% assigned to a validation candidate pool for early stopping and model selection. In contrast, non-

overlapping windows with a stride of 2048 were generated from the second half for testing to reduce sample redundancy. Therefore, the model-development and test samples were obtained from temporally separated portions of the original signals.

Different training scales indicate the total number of labeled samples randomly selected from the training candidate set in a class-balanced manner. For the CWRU dataset, an equal number of samples was selected from each diagnostic class, resulting in 60, 90, 120, 200, 300, and 400 labeled training samples. For the HUST dataset, which contains nine diagnostic classes, the corresponding actual training sizes were 54, 90, 117, 198, 297, and 396 for the preset sizes of 60, 90, 120, 200, 300, and 400, respectively. Each experimental setting was repeated five times with independent random sampling, and the mean test accuracy was reported.

3.1.4. Experimental Settings

For a fair comparison, all methods were trained using the same data partition strategy, training protocol, and hyperparameter settings. The detailed experimental configurations are summarized in Table 3.

Table 3. Experimental parameter settings.

Hyperparameter	Setting
Optimizer	AdamW
Initial learning rate	1×10^{-4}
Learning rate scheduler	Cosine Annealing
Batch size	32
Number of epochs	100
Weight decay	5×10^{-4}
Early stopping patience	10
Dropout	0.30
Attention heads	4
Data augmentation	SpecAugment
Time masking width	30
Frequency masking width	8

3.2. Small-Sample Fault Diagnosis Performance

3.2.1. Results on the CWRU Dataset

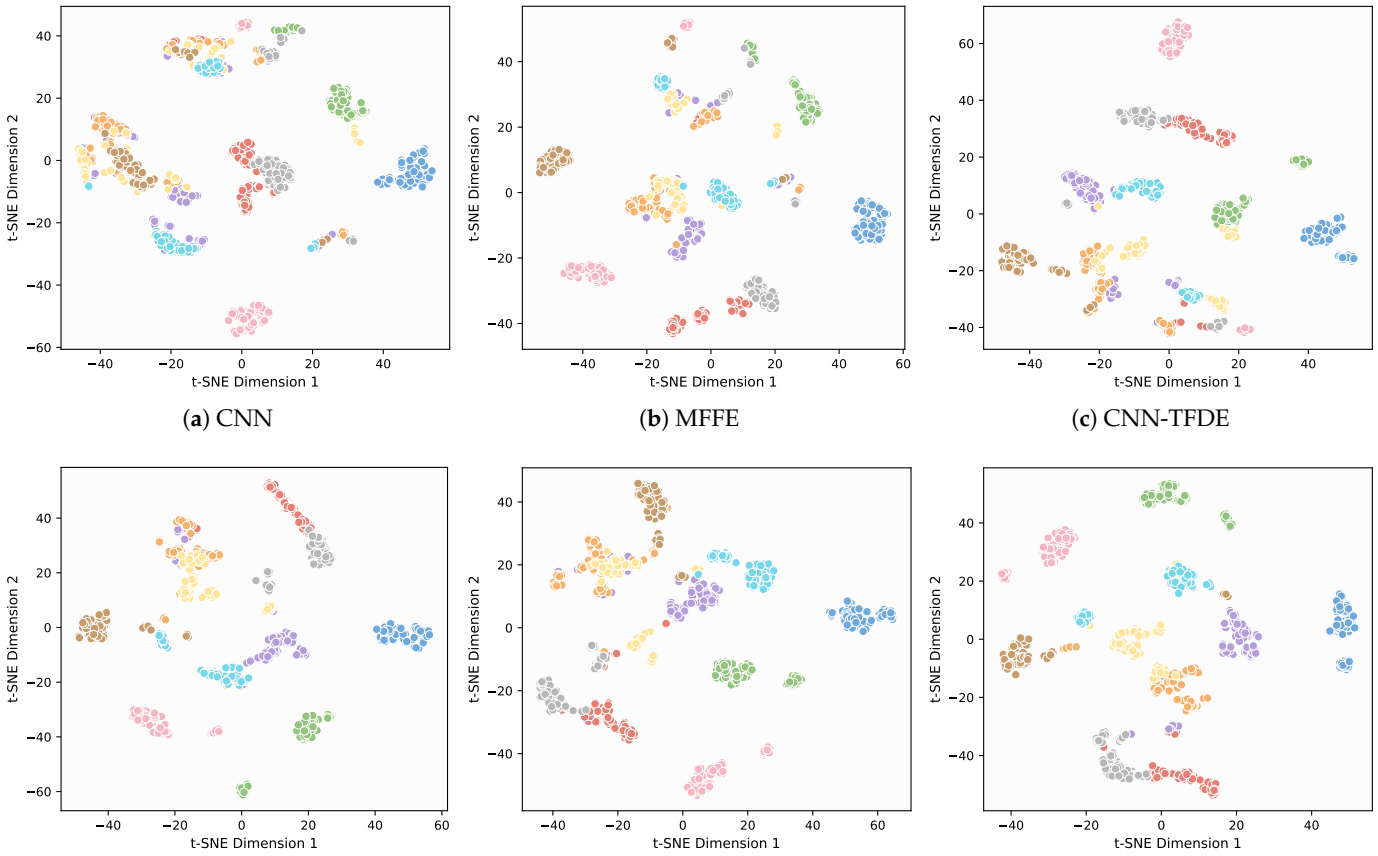
Table 4 summarizes the average test accuracy of different methods on the CWRU dataset under various training sample sizes. Overall, the performance of all methods improves as more labeled samples become available. Compared with the baseline CNN, the proposed variants generally achieve higher accuracies, indicating the benefit of frequency-aware representation learning and temporal–frequency modeling under limited-data conditions.

MFETA-Net achieves the highest accuracy in most settings, reaching 77.26% with only 60 training samples, which exceeds the baseline CNN by 52.67 percentage points, and 98.62% with 400 samples. Under limited-data conditions, faults with similar locations or defect sizes may exhibit overlapping time–frequency patterns, making fine-grained discrimination difficult. The proposed frequency enhancement and temporal–frequency dependency modeling help exploit complementary spectral and transient characteristics, which may explain the larger performance gains at smaller training sizes. Although MFFE-TFDE slightly outperforms MFETA-Net at 300 samples, the difference is marginal, and MFETA-Net remains competitive across the evaluated settings.

Table 4. Average test accuracy (%) of different methods on the CWRU dataset under different training sample sizes.

Method	60	90	120	200	300	400
CNN	24.59	34.21	38.92	52.20	61.24	65.12
MFFE	49.53	71.00	70.65	73.48	85.49	89.66
CNN-TFDE	55.40	66.40	79.43	91.56	96.88	97.99
MFFE-TFDE	60.77	75.24	84.37	95.95	98.37	98.51
CNN-TFDE-MTAA	69.93	85.80	91.34	95.73	97.66	97.80
MFETA-Net	77.26	89.17	93.97	97.38	98.33	98.62

To qualitatively evaluate the learned feature representations, Figure 7 presents the t-SNE visualizations of different implemented variants using 90 training samples. The features learned by the baseline CNN are widely scattered with noticeable overlap among several fault categories. As frequency enhancement and temporal–frequency dependency modeling are introduced, intra-class compactness and inter-class separability are gradually improved. MFETA-Net produces relatively compact clusters and clear class boundaries under this setting, indicating improved feature discriminability.



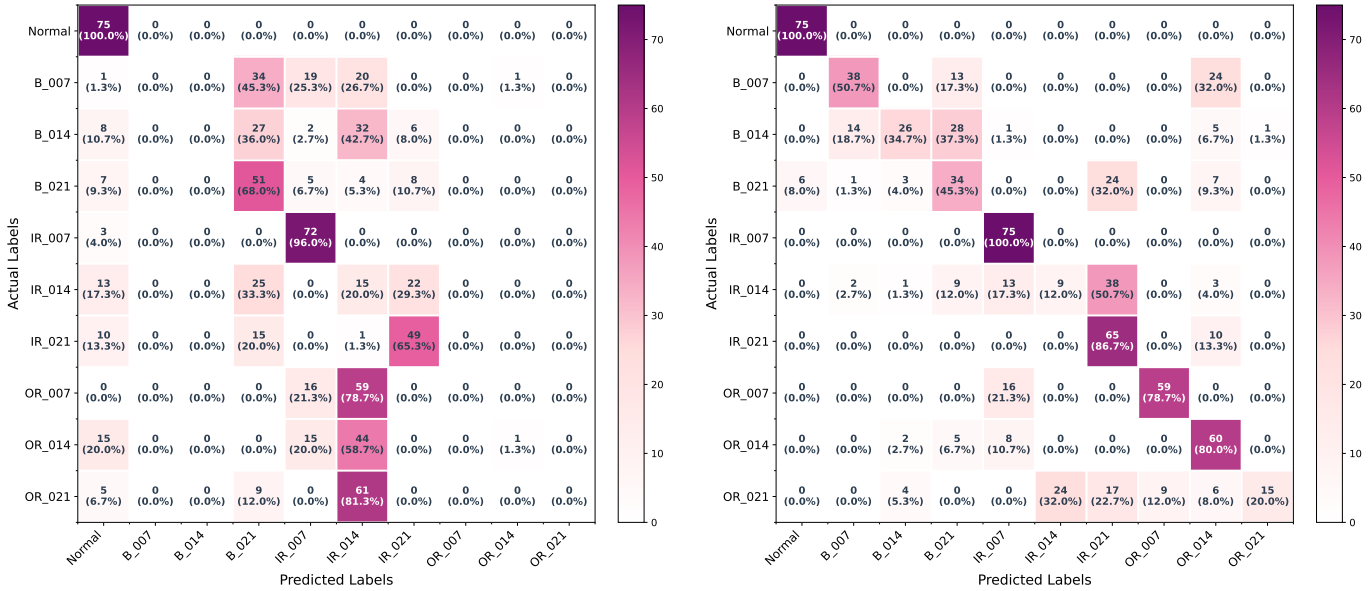
(**a**) CNN (**b**) MFFE (**c**) CNN-TFDE

(**d**) MFFE-TFDE (**e**) CNN-TFDE-MTAA (**f**) MFETA-Net

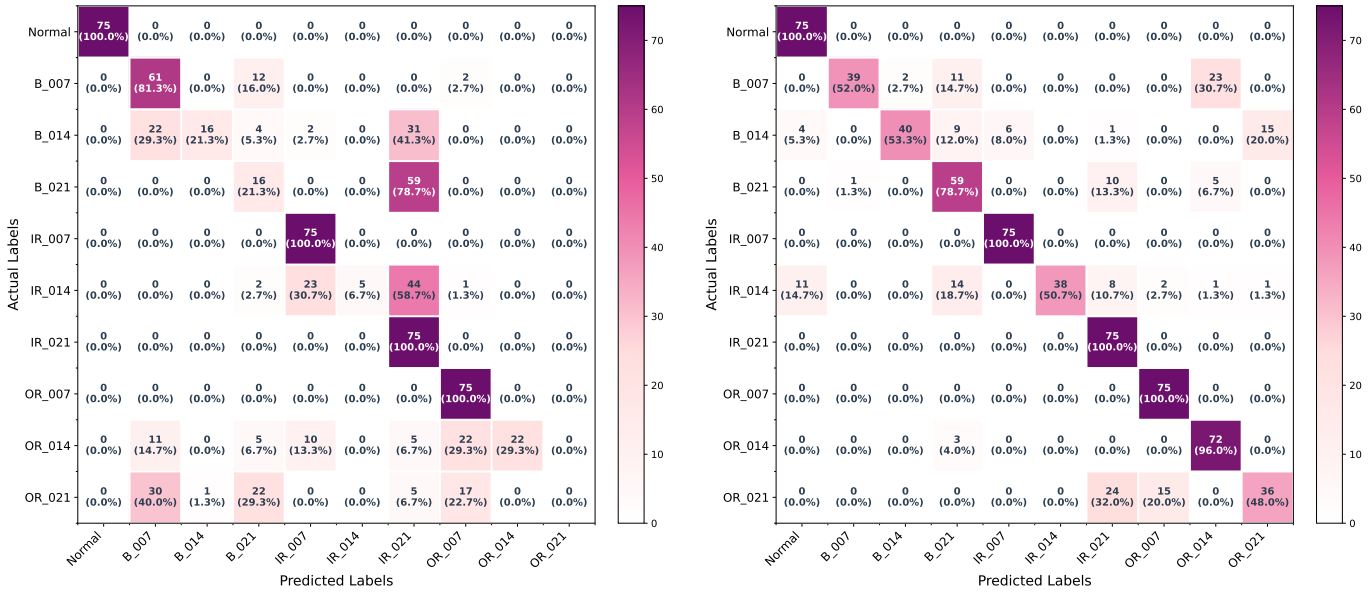
Figure 7. t-SNE visualization of different methods on the CWRU dataset with 90 training samples. (**a**) CNN; (**b**) MFFE; (**c**) CNN-TFDE; (**d**) MFFE-TFDE; (**e**) CNN-TFDE-MTAA; (**f**) MFETA-Net. Different colors denote different fault classes.

The corresponding confusion matrices are shown in Figure 8. Compared with the competing methods, MFETA-Net achieves higher classification accuracy for most fault categories, with more concentrated diagonal elements and fewer off-diagonal misclassifications. These results further demonstrate the superior discriminative capability of the

proposed method. A small number of misclassifications still occur between fault classes with similar fault sizes, mainly because they exhibit similar time–frequency characteristics, making fine-grained fault identification more challenging.

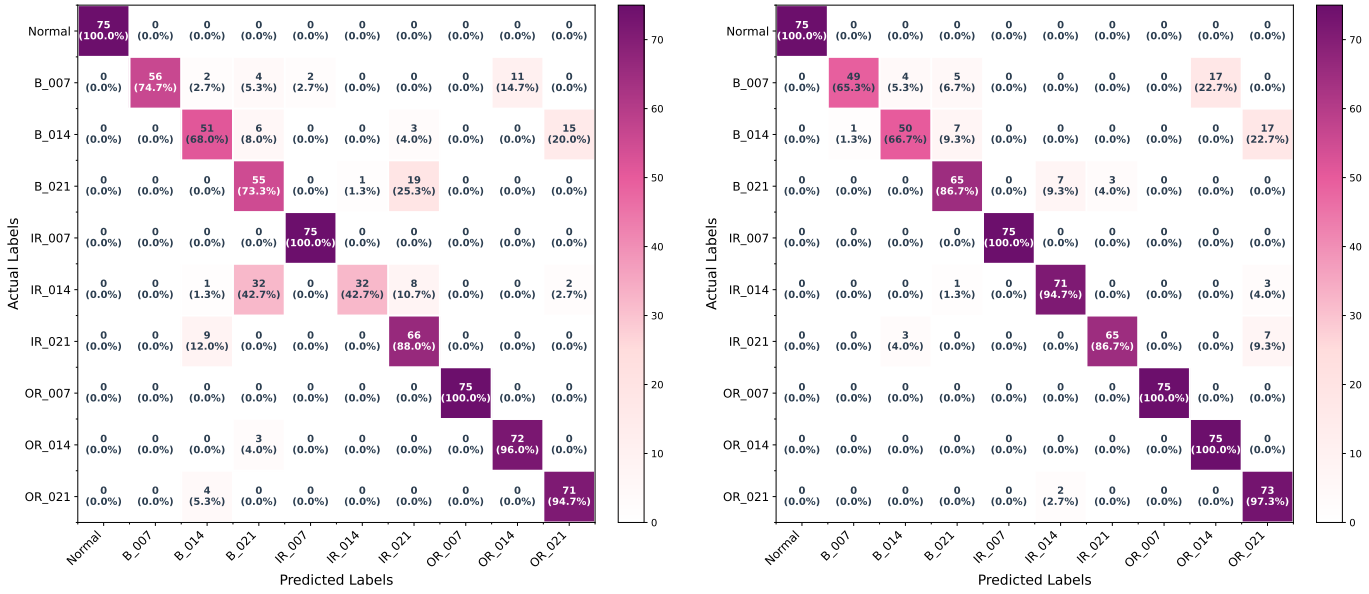


(**a**) CNN (**b**) MFFE



(**c**) CNN-TFDE (**d**) MFFE-TFDE

Figure 8. *Cont.*



(**e**) CNN-TFDE-MTAA (**f**) MFETA-Net

Figure 8. Confusion matrices of different methods on the CWRU dataset with 90 training samples. (**a**) CNN; (**b**) MFFE; (**c**) CNN-TFDE; (**d**) MFFE-TFDE; (**e**) CNN-TFDE-MTAA; (**f**) MFETA-Net.

3.2.2. Results on the HUST Dataset

To further evaluate the proposed method under more challenging conditions, experiments were conducted on the HUST bearing dataset using the same small-sample settings. Compared with CWRU, HUST contains multiple fault severity levels and combination faults, resulting in more complex vibration characteristics and higher inter-class similarity. As shown in Table 5, MFETA-Net achieves the highest accuracy across the evaluated training sample sizes.

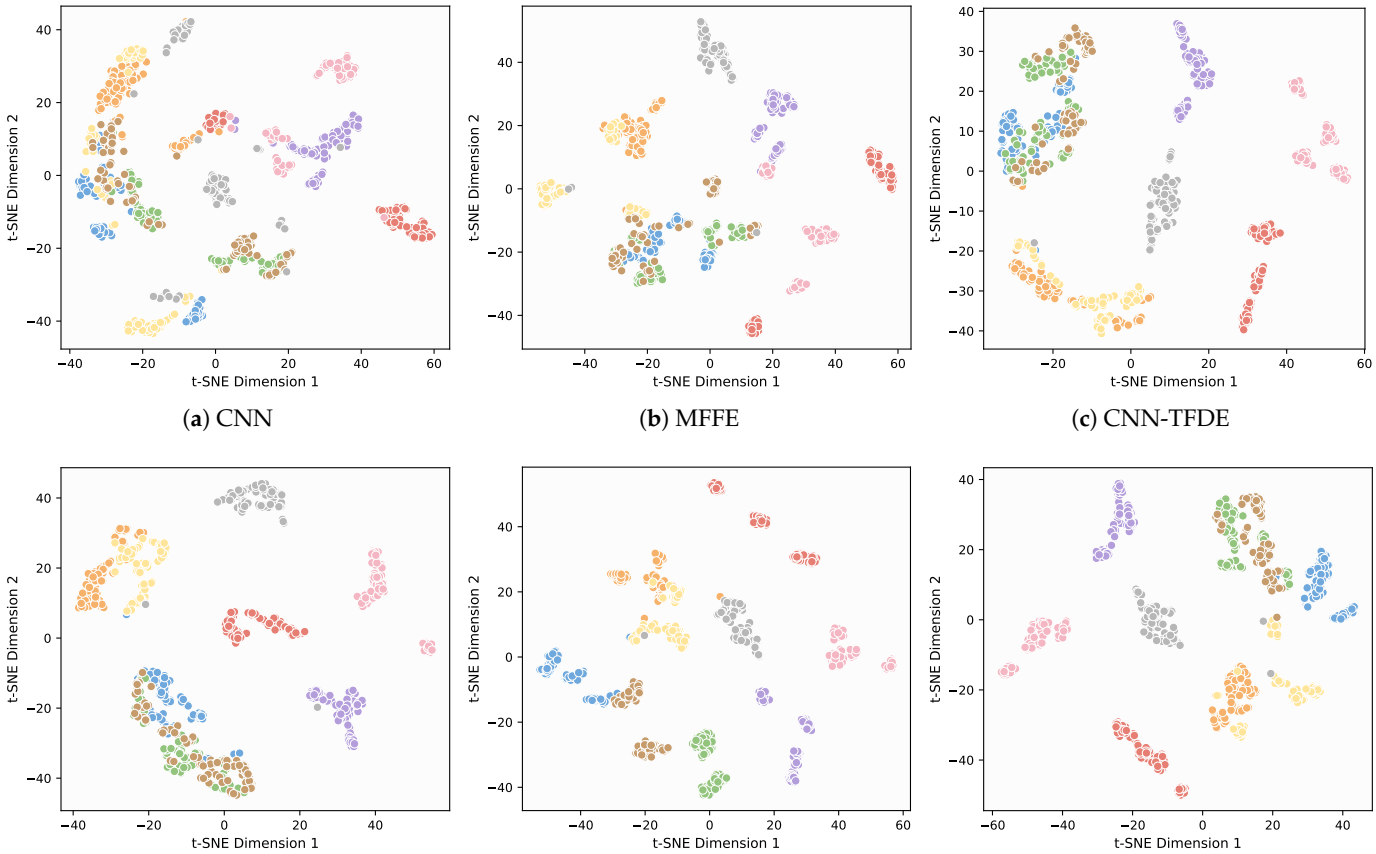
With a preset training size of 60 (54 actual samples), MFETA-Net achieves an accuracy of 79.60%, outperforming the baseline CNN by 46.48 percentage points, and its accuracy increases to 99.92% at 400 samples. The advantage is more pronounced under limited-data conditions, suggesting that frequency enhancement helps distinguish subtle spectral differences among similar fault classes, while temporal–frequency dependency modeling preserves complementary transient information. As the training set increases, the performance gap among the stronger variants gradually narrows because more representative fault patterns become available for learning.

Table 5. Average test accuracy (%) on the HUST dataset under different preset training sample sizes.

Method	60	90	120	200	300	400
CNN	33.12	43.36	45.34	57.85	65.22	63.50
MFFE	39.77	49.00	56.90	64.23	74.34	81.38
CNN-TFDE	59.61	66.15	73.60	83.60	94.17	98.10
MFFE-TFDE	63.46	71.76	82.15	94.01	99.40	99.87
CNN-TFDE-MTAA	73.92	84.02	90.84	97.32	98.62	99.29
MFETA-Net	79.60	90.58	96.28	98.19	99.79	99.92

Figure 9 presents the t-SNE visualization of the learned feature representations with 90 training samples. The baseline CNN produces scattered feature distributions with noticeable overlap among several fault categories. As the representation learning capability is enhanced, intra-class compactness and inter-class separability gradually improve. Compared with the competing methods, MFETA-Net generates more compact feature clusters

and clearer decision boundaries, indicating its superior ability to learn discriminative representations for complex fault categories.



(**a**) CNN (**b**) MFFE (**c**) CNN-TFDE

(**d**) MFFE-TFDE (**e**) CNN-TFDE-MTAA (**f**) MFETA-Net

Figure 9. t-SNE visualization of different methods on the HUST dataset with 90 training samples. (**a**) CNN; (**b**) MFFE; (**c**) CNN-TFDE; (**d**) MFFE-TFDE; (**e**) CNN-TFDE-MTAA; (**f**) MFETA-Net. Different colors denote different fault classes.

The corresponding confusion matrices are shown in Figure 10. MFETA-Net achieves higher classification accuracy for most fault categories and reduces misclassified samples compared with the implemented variants. Although slight confusion remains between fault classes with similar damage severities, the proposed method provides strong overall classification performance under this setting, further indicating its effectiveness for same-domain small-sample fault diagnosis.

3.3. Noise Robustness Analysis

To evaluate the robustness of different model variants under controlled noisy conditions, additive white Gaussian noise (AWGN) was added to the raw time-domain vibration signals of the test samples. The training samples remained unchanged. For each test segment, the noisy signal is defined as

$$\tilde{x} = x + n, \tag{30}$$

where x denotes the original time-domain vibration signal, n denotes the generated Gaussian noise, and $\tilde{x}$ denotes the noisy vibration signal. The signal-to-noise ratio (SNR) is defined as

$$\mathrm{SNR(dB)} = 10 \log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right), \tag{31}$$

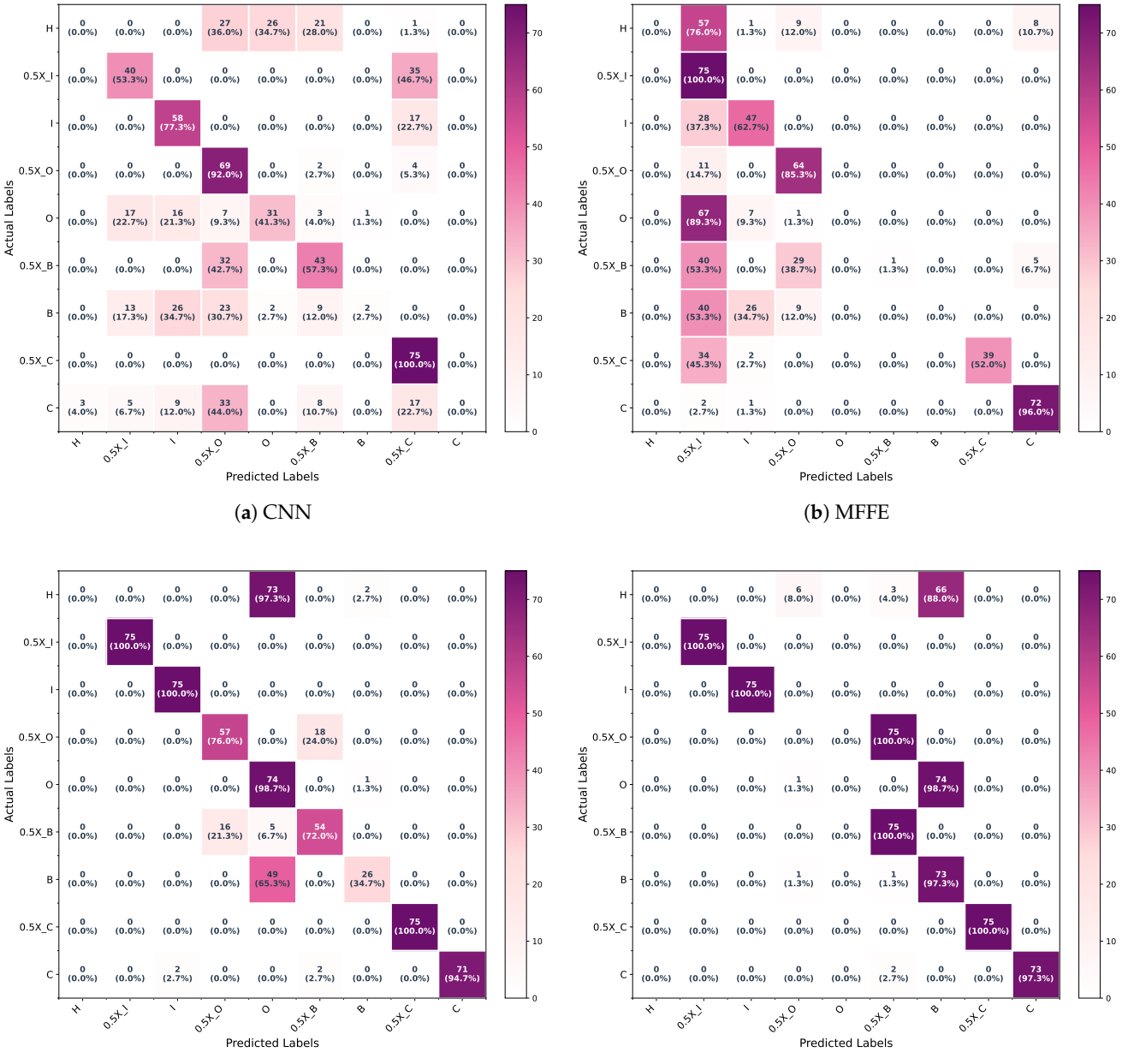
where P_{signal} and P_{noise} represent the power of the original vibration signal and the added noise, respectively. For a target SNR level, the corresponding noise power is calculated as

$$P_{\text{noise}} = \frac{P_{\text{signal}}}{10^{\text{SNR}/10}}. \tag{32}$$

After noise injection, the noisy time-domain signal $\tilde{x}$ is converted into an STFT-based time–frequency representation:

$$\tilde{X}_{\text{STFT}} = \text{STFT}(\tilde{x}). \tag{33}$$

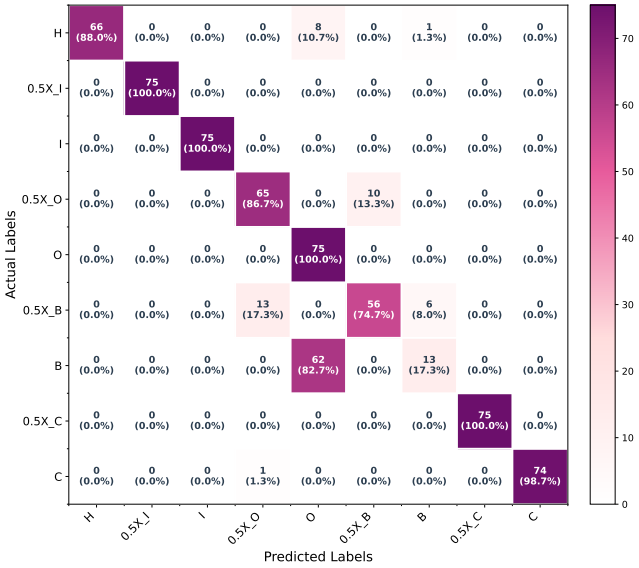
The resulting noisy STFT representation $\tilde{X}_{\text{STFT}}$ is then fed into the proposed framework for testing. In this study, AWGN was added at seven SNR levels: -6, -3, 0, 3, 6, 9, and 12 dB. Lower SNR values indicate stronger noise interference. It should be noted that AWGN is an idealized noise model and cannot fully represent non-Gaussian and non-stationary noise in real industrial environments.



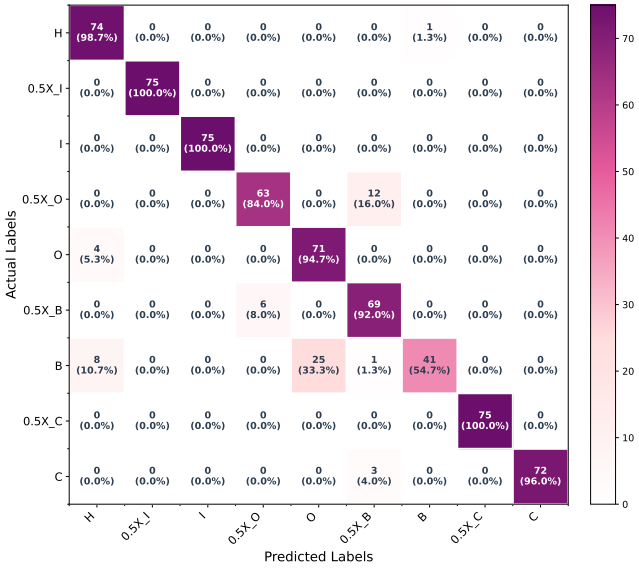
(**a**) CNN (**b**) MFFE

(**c**) CNN-TFDE (**d**) MFFE-TFDE

Figure 10. *Cont.*



(**e**) CNN-TFDE-MTAA



(**f**) MFETA-Net

Figure 10. Confusion matrices of different methods on the HUST dataset with 90 training samples. (**a**) CNN; (**b**) MFFE; (**c**) CNN-TFDE; (**d**) MFFE-TFDE; (**e**) CNN-TFDE-MTAA; (**f**) MFETA-Net.

3.3.1. Results of CWRU

Figure 11 presents the classification accuracy of different methods on the CWRU dataset under various SNR levels using 60, 120, 200, and 400 training samples. As the SNR decreases, the classification accuracy of all methods gradually declines because noise weakens the fault-related components in the time–frequency representations. This degradation is more pronounced when fewer training samples are available, since limited samples provide insufficient information for learning robust fault representations.

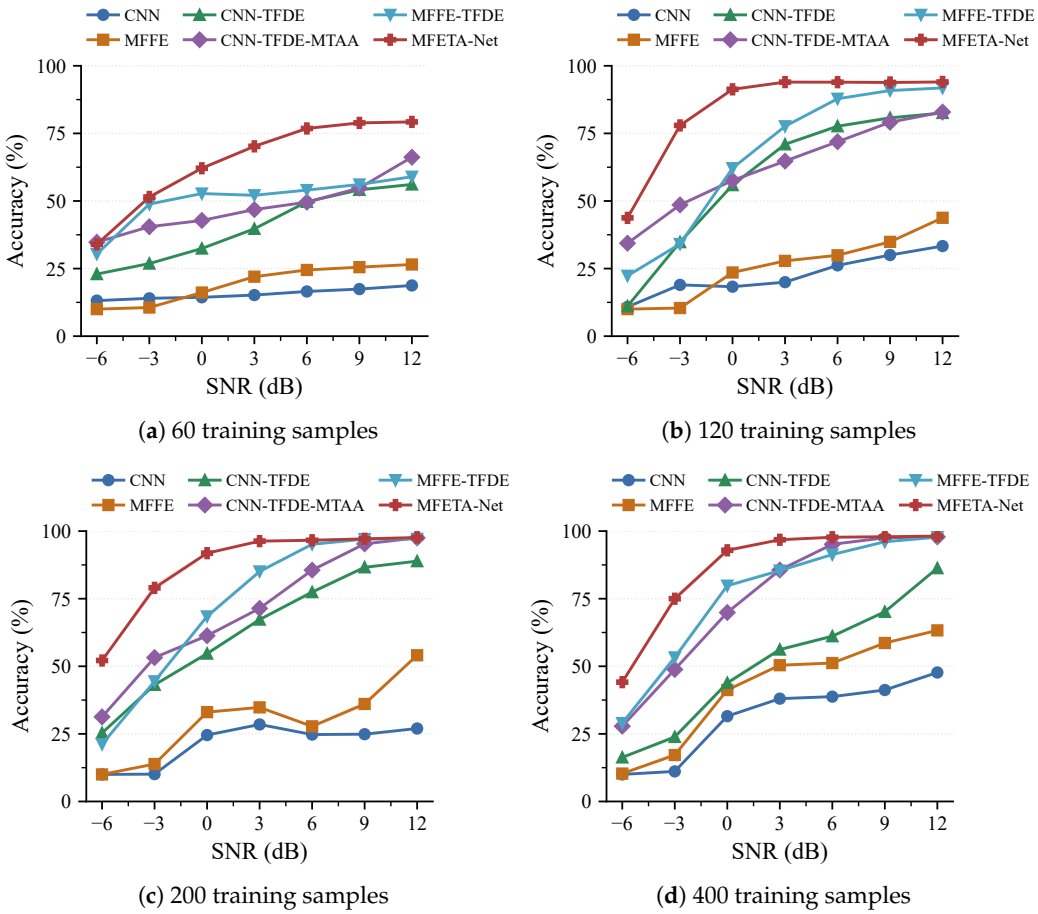
Compared with CNN and MFFE, the variants incorporating TFDE generally maintain higher accuracy under low-SNR conditions, indicating that temporal–frequency dependency modeling helps preserve the correlation between frequency characteristics and vibration responses under noise interference. Among all methods, MFETA-Net achieves higher or comparable accuracy across most SNR levels, particularly with 60 and 120 training samples. This advantage can be attributed to the complementary effects of frequency enhancement and temporal–frequency dependency modeling, which improve the extraction of discriminative fault features and enhance robustness under noisy small-sample conditions.

3.3.2. Results of HUST

Figure 12 shows the classification accuracy of different methods on the HUST dataset under various SNR levels using 60, 120, 200, and 400 training samples. Compared with the CWRU dataset, HUST contains multiple fault severity levels and combination faults, resulting in more complex vibration characteristics and higher inter-class similarity. Therefore, the classification performance decreases more significantly under noisy conditions.

As shown in Figure 12, the recognition accuracy of most variants decreases considerably at low SNRs when only a few training samples are available. Nevertheless, MFETA-Net generally achieves higher or comparable accuracy across most noise levels. As the number of training samples increases, the robustness of all variants improves, and the performance gap among different variants becomes smaller. This indicates that sufficient training samples can alleviate the influence of noise by providing more representative fault patterns. Meanwhile, MFETA-Net maintains relatively stable performance under controlled AWGN

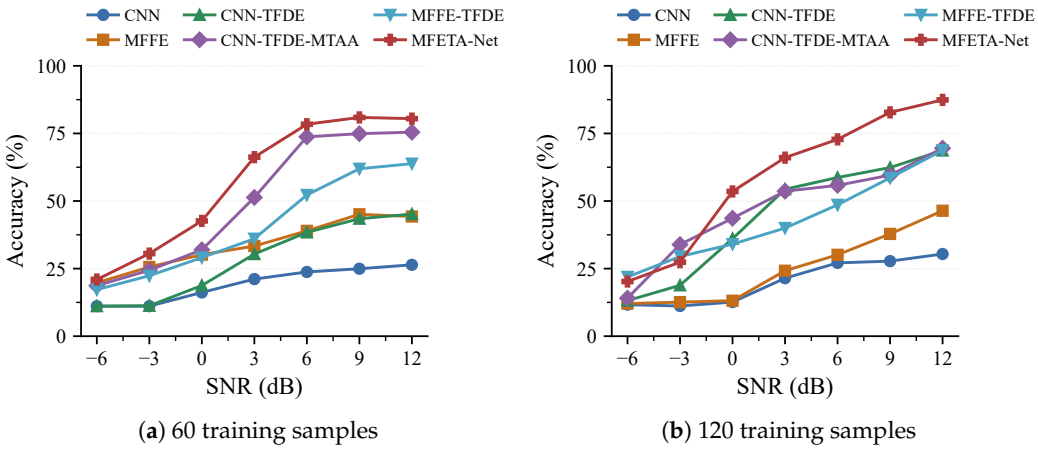
interference, demonstrating the effectiveness of frequency-aware representation learning and temporal–frequency dependency modeling for noisy complex fault diagnosis.



(**a**) 60 training samples (**b**) 120 training samples

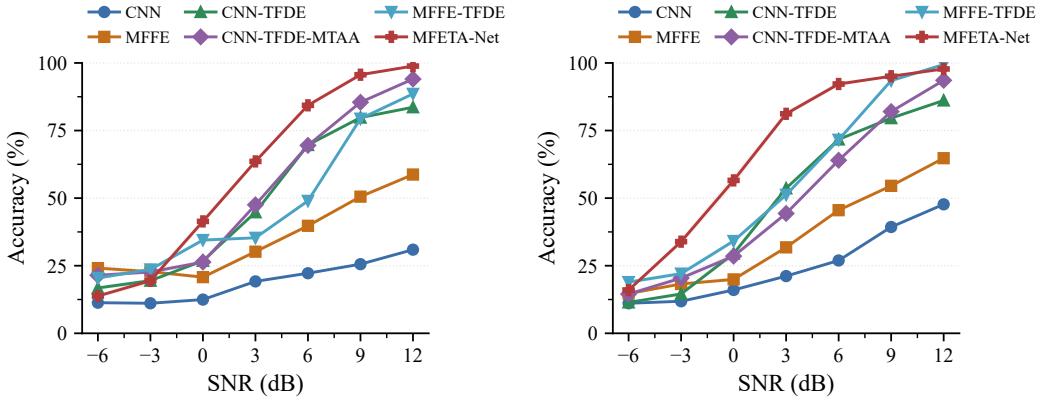
(**c**) 200 training samples (**d**) 400 training samples

Figure 11. Classification accuracy on the CWRU dataset under different noise levels: (**a**) 60 training samples; (**b**) 120 training samples; (**c**) 200 training samples; (**d**) 400 training samples.



(**a**) 60 training samples (**b**) 120 training samples

Figure 12. *Cont.*



(**c**) 200 training samples (**d**) 400 training samples

Figure 12. Classification accuracy on the HUST dataset under different noise levels: (**a**) 60 training samples; (**b**) 120 training samples; (**c**) 200 training samples; (**d**) 400 training samples.

4. Conclusions

This study proposed a Multi-branch Frequency Enhancement Temporal Attention Network (MFETA-Net) for small-sample rolling bearing fault diagnosis. By integrating the Multi-branch Frequency Feature Enhancement Encoder (MFFE), Temporal–Frequency Dependency Enhancement (TFDE), and Multi-head Temporal Attention Aggregation (MTAA), the proposed framework learns discriminative frequency-aware and temporal–frequency representations from limited labeled samples.

Experiments on the CWRU and HUST datasets demonstrate that MFETA-Net achieves competitive diagnostic performance under same-domain small-sample conditions. The frequency enhancement and temporal–frequency dependency modeling strategies improve feature representation, while temporal attention further emphasizes diagnostically important responses. Noise robustness experiments also confirm the effectiveness of the proposed framework under different SNR conditions.

Several limitations remain. The current evaluation is mainly conducted on public benchmark datasets under same-domain conditions, and cross-domain generalization across different machines, loads, and operating conditions has not yet been fully investigated. In addition, the adopted datasets do not explicitly include irregularly distributed raceway pitting or other non-periodic fault patterns, for which temporal variations may be more pronounced. The AWGN-based evaluation also cannot fully represent complex industrial noise environments. Future work will therefore investigate cross-domain adaptation, non-periodic fault conditions, more realistic industrial noise, and lightweight model optimization for practical condition monitoring.

Author Contributions: Y.Z.: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing—original draft preparation. C.W.: conceptualization, supervision, project administration, methodology, writing—review and editing. S.Z.: supervision, validation, methodology, writing—review and editing. D.Z.: software, validation, investigation. C.H.: investigation, data curation. Z.L.: resources, validation. B.W.: resources. L.G.: resources. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the project "Development and Application of Multi-Source Data Ship Intelligent Auxiliary Terminal Based on Large Model Technology" (Grant No. 2024H6020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Case Western Reserve University (CWRU) bearing dataset and the HUST bearing dataset used in this study are publicly available from their original repositories. The implementation code and processed data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the providers of the public CWRU and HUST bearing datasets for making the datasets publicly available.

Conflicts of Interest: Author Zhenjun Li was employed by the company Fujian Province Shipbuilding Industry Group Co., Ltd., author Bingkun Wu was employed by the company Zhongshu Xinke (Zhongshu (Xiamen) Information Technology Co., Ltd.), and author Liangqing Guan was employed by the company Fujian Fuchuan Marine Engineering Technology Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Cerrada, M.; Sanchez, R.; Li, C.; Pacheco, F.; Cabrera, D.; De Oliveira, J.V.; Vásquez, R.E. A review on data-driven fault severity assessment in rolling bearings. *Mech. Syst. Signal Process.* **2018**, *99*, 169–196. [CrossRef]
2. Zhou, S.Q.; Lin, L.; Chen, C.; Pan, W.; Lou, X. Application of convolutional neural network in motor bearing fault diagnosis. *Comput. Intell. Neurosci.* **2022**, *2022*, 9231305. [CrossRef] [PubMed]
3. Chen, J.; Lin, C.; Peng, D.; Ge, H. Fault diagnosis of rotating machinery: A review and bibliometric analysis. *IEEE Access* **2020**, *8*, 224985–225003. [CrossRef]
4. Zhang, X.; Zhao, B.; Lin, Y. Machine learning based bearing fault diagnosis using the Case Western Reserve University data: A review. *IEEE Access* **2021**, *9*, 155598–155608. [CrossRef]
5. Ding, S.; Chen, R.; Liu, H.; Liu, F.; Zhang, J. ASG-HOMGAT: A high-order multi-head graph attention network with adaptive small graph structure for rolling bearing fault diagnosis. *Meas. Sci. Technol.* **2024**, *35*, 065018. [CrossRef]
6. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
7. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef] [PubMed]
8. Chen, J.; Lu, C.; Yuan, H. Bearing fault diagnosis based on active learning and random forest. *Vibroeng. Procedia* **2015**, *5*, 321–326. [CrossRef]
9. Ruan, D.; Wang, J.; Yan, J.; Gühmann, C. CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis. *Adv. Eng. Inform.* **2023**, *55*, 101877. [CrossRef]
10. Chadha, G.S.; Panambilly, A.; Schwung, A.; Ding, S.X. Bidirectional deep recurrent neural networks for process fault classification. *ISA Trans.* **2020**, *106*, 330–342. [CrossRef] [PubMed]
11. Lv, J.; Xiao, Q.; Zhai, X.; Shi, W. A high-performance rolling bearing fault diagnosis method based on adaptive feature mode decomposition and Transformer. *Appl. Acoust.* **2024**, *224*, 110156. [CrossRef]
12. Tang, S.; Yuan, S.; Zhu, Y. Deep learning-based intelligent fault diagnosis methods toward rotating machinery. *IEEE Access* **2020**, *8*, 9335–9346. [CrossRef]
13. Hakim, M.; Omran, A.A.B.; Ahmed, A.N.; Al-Waily, M.; Abdellatif, A. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Eng. J.* **2023**, *14*, 101945. [CrossRef]
14. Zhao, Z.; Li, T.; Wu, J.; Sun, C.; Wang, S.; Yan, R.; Chen, X. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. *ISA Trans.* **2020**, *107*, 224–255. [CrossRef] [PubMed]
15. Kim, K.; Kim, Y.S. Vibration spectrogram analysis for bearing fault diagnosis based on Grad-CAM for feature selection and statistical approach. *J. Mech. Sci. Technol.* **2024**, *38*, 5885–5898. [CrossRef]
16. Łuczak, D. Machine fault diagnosis through vibration analysis: Continuous wavelet transform with complex Morlet wavelet and time–frequency RGB image recognition via convolutional neural network. *Electronics* **2024**, *13*, 452. [CrossRef]
17. Chen, L.; Zi, Y.; He, Z.; Cheng, W. Research and application of ensemble empirical mode decomposition principle and 1.5 dimension spectrum method. *J. Xi'an Jiaotong Univ.* **2009**, *43*, 94–98.
18. Guo, S.; Yang, T.; Gao, W.; Zhang, C. A novel fault diagnosis method for rotating machinery based on a convolutional neural network. *Sensors* **2018**, *18*, 1429. [CrossRef] [PubMed]
19. Djaballah, S.; Meftah, K.; Khelil, K.; Sayadi, M. Deep transfer learning for bearing fault diagnosis using CWT time–frequency images and convolutional neural networks. *J. Fail. Anal. Prev.* **2023**, *23*, 1046–1058. [CrossRef]

20. Eren, L.; Ince, T.; Kiranyaz, S. A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *J. Signal Process. Syst.* **2019**, *91*, 179–189. [CrossRef]
21. Jia, F.; Lei, Y.; Lu, N.; Xing, S. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mech. Syst. Signal Process.* **2018**, *110*, 349–367. [CrossRef]
22. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **2017**, *17*, 425. [CrossRef] [PubMed]
23. Meng, Z.; Cao, W.; Sun, D.; Li, Q.; Ma, W.; Fan, F. Research on fault diagnosis method of MS-CNN rolling bearing based on local central moment discrepancy. *Adv. Eng. Inform.* **2022**, *54*, 101797. [CrossRef]
24. Fu, W.; Jiang, X.; Li, B.; Tan, C.; Chen, B.; Chen, X. Rolling bearing fault diagnosis based on 2D time-frequency images and data augmentation technique. *Meas. Sci. Technol.* **2023**, *34*, 045005. [CrossRef]
25. Hou, J.; Lu, X.; Zhong, Y.; He, W.; Zhao, D.; Zhou, F. Rolling bearing fault diagnosis method by using feature extraction of convolutional time-frequency image. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2024**, *238*, 4212–4228. [CrossRef]
26. Han, K.; Wang, W.; Guo, J. Research on a bearing fault diagnosis method based on a CNN-LSTM-GRU model. *Machines* **2024**, *12*, 927. [CrossRef]
27. Zhao, H.; Sun, S.; Jin, B. Sequential fault diagnosis based on LSTM neural network. *IEEE Access* **2018**, *6*, 12929–12939. [CrossRef]
28. Bian, C.; Jia, C.; Li, J.; Chen, X.; Wang, P. Rolling bearing fault diagnosis under small sample conditions based on WDCNN-BiLSTM Siamese network. *Sci. Rep.* **2025**, *15*, 29591. [CrossRef] [PubMed]
29. Da, T.N.; Thanh, P.N.; Cho, M.Y. Novel cloud-AIoT fault diagnosis for industrial diesel generators based hybrid deep learning CNN-BGRU algorithm. *Internet Things* **2024**, *26*, 101164. [CrossRef]
30. Song, B.; Liu, Y.; Fang, J.; Liu, W.; Zhong, M.; Liu, X. An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples. *Neurocomputing* **2024**, *574*, 127284. [CrossRef]
31. Wang, H.; Xu, J.; Yan, R.; Gao, R.X. A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 2377–2389. [CrossRef]
32. Ding, Y.; Jia, M.; Miao, Q.; Cao, Y. A novel time–frequency Transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings. *Mech. Syst. Signal Process.* **2022**, *168*, 108616. [CrossRef]
33. Xu, K.; Zhang, J.; Liu, D.; Wang, C.; Wang, R.; Zhou, F. Multimodal Fault Diagnosis of Rolling Bearings Based on GRU–ResNet–CBAM. *Machines* **2026**, *14*, 318. [CrossRef]
34. Chen, Q.; Zhang, F.; Wang, Y.; Yu, Q.; Lang, G.; Zeng, L. Bearing fault diagnosis based on efficient cross space multiscale CNN transformer parallelism. *Sci. Rep.* **2025**, *15*, 12344. [CrossRef] [PubMed]
35. Allen, J.B.; Rabiner, L.R. A unified approach to short-time Fourier analysis and synthesis. *Proc. IEEE* **1977**, *65*, 1558–1564. [CrossRef]
36. Liu, D.; Cheng, W.; Wen, W. Rolling bearing fault diagnosis via STFT and improved instantaneous frequency estimation method. *Procedia Manuf.* **2020**, *49*, 166–172. [CrossRef]
37. Zhang, X.; Lv, J.; Wang, F.; He, F. A dual-channel multi-scale residual network with dilated convolution optimization for rolling bearing fault diagnosis. *J. Vib. Control* **2025**. [CrossRef]
38. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
39. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–131. [CrossRef]
40. Zhao, C.; Zio, E.; Shen, W. Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. *Reliab. Eng. Syst. Saf.* **2024**, *245*, 109964. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.