applied sciences

MDPI

Article

Multi-Domain Representation Learning for Bearing Fault Diagnosis with Phase and Transient Preservation

Bingbing Hu [1], Jing Zhu [2,*], Shilei Liang [2] and Liao Ting [2]

[1] Labor Union of Henan University of Science and Technology (HAUST), Luoyang 471003, China
[2] The School of Vehicle and Traffic Engineering, Henan University of Science and Technology (HAUST), Luoyang 471003, China
* Correspondence: jingzhu244@gmail.com

Abstract

Reliable bearing fault diagnosis under complex operating conditions is often hindered by the loss of critical information during feature extraction, particularly for weak fault signatures embedded in vibration signals. To address this challenge, this work proposes a parallel multi-domain deep learning framework that emphasizes the preservation and complementary exploitation of time-domain, frequency-domain, and time–frequency representations. The proposed framework integrates temporal modeling to capture long-range signal evolution, phase-aware frequency-domain analysis to preserve amplitude–phase coherence, and transient-enhanced time–frequency representations to highlight weak impulsive features in noisy environments. To effectively integrate heterogeneous representations, a dynamic self-attention-based fusion strategy is introduced, enabling adaptive interaction and importance reweighting among multi-domain features. Experimental studies conducted on bearing datasets from Huazhong University of Science and Technology and the University of Cincinnati demonstrate that the proposed method achieves diagnostic accuracies of 99.63% and 99.82%, respectively, significantly outperforming state-of-the-art deep learning and multi-domain diagnostic methods, with accuracy improvements exceeding 20% compared to representative baseline models. Furthermore, ablation and robustness analyses confirm that the coordinated preservation and fusion of multi-domain information significantly enhance diagnostic reliability and generalization performance under complex operating conditions.

Keywords: rolling bearing fault diagnosis; multi-domain feature preservation; parallel deep learning framework; phase-aware frequency representation; dynamic attention-based feature fusion

Academic Editor: Ephraim Suhir

Received: 13 January 2026
Revised: 3 February 2026
Accepted: 4 February 2026
Published: 12 February 2026

1. Introduction

As a key component in mechanical equipment, rolling bearings often operate under high speed, heavy load, and variable conditions, making them highly susceptible to failure [1]. Statistics indicate that approximately 30% to 40% of mechanical failures originate from bearing defects, leading to significant safety hazards and economic losses [2,3]. In modern industrial applications, the significance of bearing diagnostics extends to advanced robotic systems and articular mechanisms, where the overall system reliability is heavily dictated by the condition of the bearings [4]. Recent research has highlighted that monitoring diagnostic parameters—such as vibration, temperature, and axial/radial clearances—is crucial for maintaining the operational stability of these complex systems [5]. Consequently, developing robust fault diagnosis technology based on vibration signals has become a focal point in ensuring the long-term reliability of industrial machinery [6].

With the development of artificial intelligence technology, intelligent fault diagnosis methods have attracted the attention of many researchers [7]. The traditional intelligent fault diagnosis process usually includes the following steps: first, collect and preprocess the vibration signal; then, extract features from the processed signal; finally, use a machine learning-based classifier to identify the health status of the bearing [8]. However, the feature extraction in this traditional method depends on the experience of experts and advanced signal processing techniques, which makes the selection of a suitable set of features very difficult, and machine learning cannot mine deep-level features [9]. Thus, deep learning, an end-to-end approach to fault diagnosis, has been introduced into bearing fault diagnosis because it can automatically learn effective feature representations from raw data [10,11].

In recent years, transforming one-dimensional time series into high-dimensional features has become a prevalent signal processing strategy in deep learning-based diagnosis. Jiang et al. [12] combined variational mode decomposition with cyclic graph coding to convert signals into two-dimensional images for residual network analysis. Li et al. [13] transformed filtered frequency-domain signals into coarse-grain lattice feature graphs for fault detection. Similarly, Cui et al. [14] utilized symmetric dot pattern images integrated with attention-based residual networks to extract distinct fault characteristics. Lu et al. [15] and Wang et al. [16] encoded vibration signals into two-dimensional feature images or color images to leverage the feature extraction capabilities of ResNet and LeNet-5, respectively. Furthermore, time–frequency analysis has been widely adopted; for instance, Fu et al. [17] and Li et al. [18] converted original signals into time–frequency representations via wavelet transforms, subsequently using convolutional neural networks or attention mechanisms to identify sensitive fault features.

Although effective, these methods typically rely on a single domain input (e.g., only time-domain images or only time–frequency maps), which often overlooks the inherent temporal dynamics of the original vibration signals. This reliance on a single-domain perspective creates a significant research gap: the feature extraction process remains incomplete and fails to fully capture the complex, multi-faceted fault signatures present in non-stationary signals. Given that vibration characteristics often vary across different domains due to fluctuating operating conditions [19], relying on a single domain may miss critical diagnostic information. Consequently, there is an urgent need for a framework that can synergistically fuse multi-domain features to improve diagnostic accuracy and robustness.

To overcome these limitations, this paper proposes a Parallel Multi-Dimensional Feature Deep Learning (PMFDL) framework, which synergistically models temporal, frequency, and time–frequency characteristics and adaptively fuses heterogeneous features. This approach improves diagnostic accuracy and robustness by comprehensively capturing fault information that single-domain methods may miss.

Specifically, the raw vibration signals are processed by three parallel branches: a Residual Bidirectional Gated Recurrent Unit (Res-BiGRU) module for temporal modeling, a Complex-Valued Frequency-Domain Convolutional Neural Network (CV-FCNN) for phase-preserving spectral feature extraction, and a Discrete Wavelet Transform–based Separable Deep Convolutional (DWT-S-DSC) module for transient feature enhancement. The extracted features are then adaptively fused using a Dynamic Multi-Head Self-Attention Mechanism (D-MHSAM) before final classification.

The features extracted from these three branches are subsequently subjected to deep interaction and adaptive fusion through a Dynamic Multi-Head Self-Attention Mechanism (D-MHSAM), and the fused representation is finally fed into a classifier for fault identifi-

https://doi.org/10.3390/app16041846

cation. Based on the above design, the main contributions of this paper are summarized as follows:

(1) A parallel multi-domain feature learning framework for bearing fault diagnosis is proposed. A synergistic PMFDL architecture is designed by integrating three heterogeneous branches to jointly model temporal dynamics, spectral phase information, and time–frequency transient characteristics. This parallel design enables comprehensive representation learning from raw vibration signals and improves generalization under complex operating conditions.
(2) A phase- and transient-preserving multi-domain feature extraction strategy is developed. For frequency-domain analysis, a complex-valued convolutional neural network (CV-FCNN) is introduced to simultaneously process real and imaginary components, preserving critical phase information for enhanced sensitivity to weak fault signatures. In addition, a residual-connected BiGRU and a DWT-S-DSC module enhanced with a high-pass SRM filter are employed to capture long-range temporal dependencies and high-frequency transient details, respectively.
(3) A dynamic gated multi-head self-attention mechanism is proposed for adaptive feature fusion. A Dynamic Multi-Head Self-Attention Mechanism (D-MHSAM) is designed to enable deep interaction among heterogeneous features. By incorporating learnable dynamic scaling factors and an adaptive head-selection gating strategy, the proposed fusion module effectively suppresses cross-domain redundancy and enhances the robustness of the fused representations for fault diagnosis.

2. Basic Methods

2.1. PMFDL Model

The architecture of the proposed Parallel Multi-Fusion Deep Learning (PMFDL) model for rolling bearing fault diagnosis is illustrated in Figure 1. The PMFDL model primarily consists of a residual BiGRU module, a complex-valued FCNN module, a DWT-S-DSC module, a dynamic feature interaction fusion layer, and a classification layer.

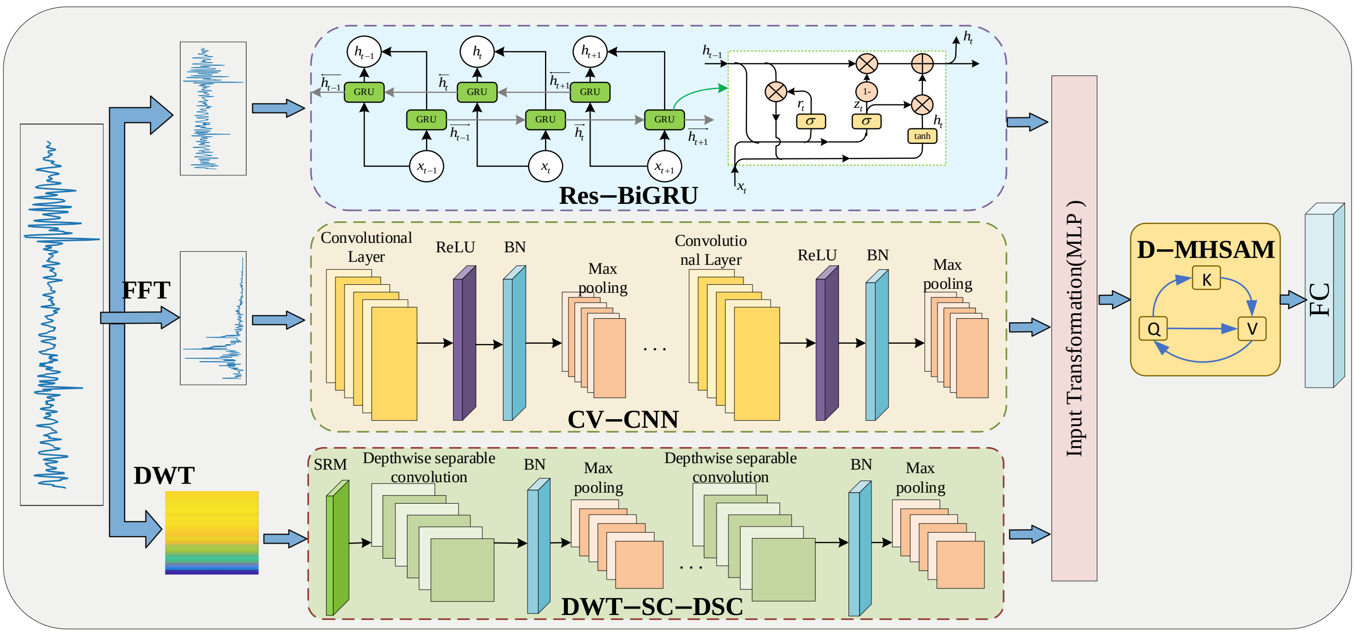


Figure 1. PMFDL network model.

The operational flow is as follows: original vibration time-series data are simultaneously fed into three parallel branches. First, the residual BiGRU module leverages its dynamic adaptability and bidirectional propagation mechanism to process raw time-domain data, maximizing the retention of temporal evolutionary characteristics. Second, the CV-FCNN module employs Fast Fourier Transform (FFT) to convert signals into frequency-domain representations, where frequency components, phase information, and amplitude spectra are synchronously extracted via complex-valued convolutions. Third, the DWT-S-DSC module utilizes Discrete Wavelet Transform (DWT) to capture local transient characteristics and generate 2D image-based features; these images are then processed by an SRM high-pass filter to suppress low-frequency noise and enhance high-frequency impulse details, followed by feature refinement using Depthwise Separable Convolution (DSC). The heterogeneous features from these three branches are integrated via the D-MHSAM layer for dynamic interaction and weight redistribution. Finally, the fused representations are fed into the classification layer to execute the diagnostic task. By integrating multi-dimensional information from time, frequency, and time–frequency domains, the model significantly enhances diagnostic accuracy and generalization robustness.

The operational flow is as follows: original vibration time-series data are simultaneously fed into three parallel branches. First, the residual BiGRU module leverages its dynamic adaptability and bidirectional propagation mechanism to process raw time-domain data, maximizing the retention of temporal evolutionary characteristics. Second, the CV-CNN module employs Fast Fourier Transform (FFT) to convert signals into frequency-domain representations, where frequency components, phase information, and amplitude spectra are synchronously extracted via complex-valued convolutions. Third, the DWT-S-DSC module utilizes Discrete Wavelet Transform (DWT) to capture local transient characteristics and generate 2D image-based features. These images are then processed by an SRM high-pass filter to suppress low-frequency noise and enhance high-frequency impulse details, followed by feature refinement using Depthwise Separable Convolution (DSC). The heterogeneous features from these three branches are integrated via the D-MHSAM layer for dynamic interaction and weight redistribution. Finally, the fused representations are fed into the classification layer (FC layer) to execute the diagnostic task. By integrating multi-dimensional information from time, frequency, and time–frequency domains, the model significantly enhances diagnostic accuracy and generalization robustness.

2.2. Residual Bidirectional Gated Recurrent Unit (Res-BiGRU) Module

BiGRU is a modified recurrent neural network (RNN) consisting of two independent GRUs, one processing data in the forward direction of the time series and the other in the reverse direction of the time series, which can capture forward and backward information in the time series data. GRU is a simplified version of the Long Short-Term Memory Network (LSTM) with fewer parameters. GRU contains two gating mechanisms: update gate and reset gate. The update gate determines to what extent the hidden state of the previous time step is updated by the candidate hidden state of the current time step. The reset gate determines how much the hidden state of the previous time step affects the candidate hidden state. The candidate hidden state is the core part of the GRU, which integrates the information of the current input and the hidden state of the previous time step. The current hidden state is the final hidden state after combining the previous time step's hidden state and the current time step's candidate hidden state. Then the forward BiGRU is calculated as follows:

$$\overrightarrow{h_t} = GRU_{forward}\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{1}$$

where h_t is the hidden state of the current time step, h_{t-1} is the hidden state of the previous time step, and x_t is the input of the current time step.

The formula for the reverse BiGRU is as follows:

$$\overleftarrow{h_t} = GRU_{backward}\left(x_t, \overleftarrow{h_{t+1}}\right) \tag{2}$$

The final BiGRU output is

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{3}$$

To improve the training stability of deep BiGRU networks and enhance information retention, a residual connection is introduced in the BiGRU module. Specifically, the output of each BiGRU layer is added to its input before being passed to the subsequent layer, which alleviates the vanishing gradient problem and preserves subtle temporal features:

$$\tilde{h}_t^{(l)} = h_t^{(l)} + h_t^{(l-1)} \tag{4}$$

where $h_t^{(l)}$ and $h_t^{(l-1)}$ denote the hidden states of the l-th and $(l-1)$-th BiGRU layers, respectively.

By incorporating the residual connection, the deep BiGRU network can more effectively propagate both the original input information and high-level features. This enhances the model's sensitivity to subtle transient patterns in the time series, thereby improving the accuracy and robustness of rolling bearing fault diagnosis.

2.3. Complex-Valued Convolutional Neural Network (CV-FCNN) Module

In the proposed model, a Complex-Valued Convolutional Neural Network (CV-FCNN) is employed to process frequency-domain signals. Unlike conventional CNNs, which typically consider only the magnitude of the Fast Fourier Transform (FFT) and discard phase information, CV-FCNN preserves both magnitude and phase, which are critical for capturing the temporal distribution of impulsive components in rolling bearing fault signals.

The frequency-domain signal $X(k)$ is obtained via the Discrete Fourier Transform (DFT) and is represented as a complex-valued sequence with real and imaginary components:

$$X(k) = \mathrm{Re}[X(k)] + j\mathrm{Im}[X(k)] \tag{5}$$

In the CV-FCNN module, complex-valued convolution is performed using complex weights $W = A + jB$ and complex inputs $H = X + jY$. The convolution operation is defined as follows:

$$W * H = (A * X - B * Y) + j(B * X + A * Y) \tag{6}$$

where $*$ denotes the convolution operation. This operation enables the network to learn the interdependencies between amplitude and phase, capturing more discriminative and synchronized features of fault-related frequencies.

To introduce non-linear expressiveness in the complex domain, the Complex-ReLU (CReLU) activation function is employed, which applies the ReLU function independently to the real and imaginary parts:

$$\mathrm{CReLU}(z) = \mathrm{ReLU}(\mathrm{Re}(z)) + j\mathrm{ReLU}(\mathrm{Im}(z)) \tag{7}$$

By leveraging CV-FCNN instead of a conventional real-valued CNN, the model fully exploits both amplitude and phase information from the frequency-domain signals, enhancing its ability to detect subtle faults and high-frequency transient features in rolling bearing vibration data.

2.4. Discrete Wavelet Transform and Enhanced Depthwise Separable (DWT-S-DSC) Module

The Discrete Wavelet Transform (DWT) provides a multi-resolution representation of signals by decomposing them into a set of mutually orthogonal wavelet basis functions [20]. Unlike the Continuous Wavelet Transform (CWT), DWT operates on discrete scales and translations, typically following a dyadic structure to eliminate redundancy. The DWT coefficients $W_{j,k}$ are obtained through the inner product of the signal $x[n]$ and the discrete wavelet basis $\psi_{j,k}[n]$:

$$W_{j,k} = \sum_{n} x[n]\psi_{j,k}[n] \tag{8}$$

The basis function $\psi_{j,k}[n]$ is defined as follows:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j k}{2^j}\right) \tag{9}$$

where j is the scale parameter, which controls the frequency of the wavelet basis function; k is the translation parameter, which controls the position of the wavelet basis function; $W_{j,k}$ are the wavelet coefficients of the jth layer and kth position; $\psi_{j,k}(t)$ is the discrete wavelet basis function.

The SRM high-pass filter is an image processing technology that can inhibit low-frequency components and enhance high-frequency information. It can analyze images from different directions and at different scales, and can better capture local information in images. The seven types of high-pass filter used in this article are shown below:

$$\text{1st}: \begin{bmatrix} -1 & 1 \end{bmatrix},\ \text{2nd}: \frac{1}{2}\times\begin{bmatrix} 1 & -2 & 1 \end{bmatrix},\ \text{3rd}: \frac{1}{3}\times\begin{bmatrix} -1 & -3 & 3 & 1 \end{bmatrix},$$

$$\text{edge}_{3\times3}: \frac{1}{4}\times\begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \end{bmatrix},\ \text{square}_{3\times3}: \frac{1}{4}\times\begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix},$$

$$\text{edge}_{5\times5}: \frac{1}{12}\times\begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \end{bmatrix},\ \text{square}_{5\times5}: \frac{1}{12}\times\begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix},$$

The above seven high-pass filters can extract high-frequency feature information from seven different modes. All of the seven high-pass filters shown above are expanded to a size of 5×5 and these seven filters are used in the first convolutional layer of the DWT-S-DSC subnetwork.

Deep-separable convolution (DSC) is composed of deep convolution, which is used to extract spatial features, and point-by-point convolution, which is used to extract channel features. As shown in Figure 2, each channel of the input feature is first deeply convolved, where each convolution kernel only works with one channel; then, the point-by-point convolution is used to carry out weighted fusion of each channel output feature graph after deep convolution, thus reducing the computation amount and improving the computation efficiency. Assuming that there are M convolution kernels with the size of $D_K \times D_K$, the feature graph size is $D_F \times D_F$, and the output feature graph channel is mapped from M

to N (that is, the number of convolution kernels for point-by-point convolution), then the calculation amount of traditional convolution is

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \tag{10}$$

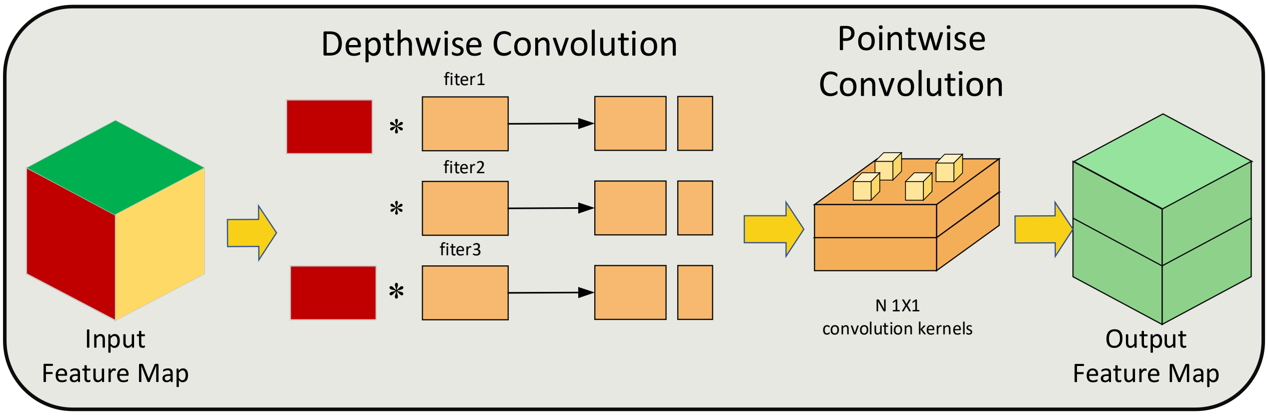


Figure 2. DSC model diagram. The symbol $*$ denotes the convolution operation.

The calculation amount of DSC is the sum of the deep convolution calculation amount and the point-by-point convolution calculation amount, and the formula is as follows:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \tag{11}$$

Comparing the differences between the two convolution operations, the results of the calculations are shown below:

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \tag{12}$$

From the results, it can be seen that as the size and number of convolution kernels increase, deep separable convolution significantly outperforms traditional convolution in terms of computational complexity. Deep separable convolution decomposes standard convolution into deep convolution and pointwise convolution, effectively reducing computational complexity and the number of required parameters. This enables deep separable convolutions to have higher efficiency and better performance when processing large-scale data and complex models.

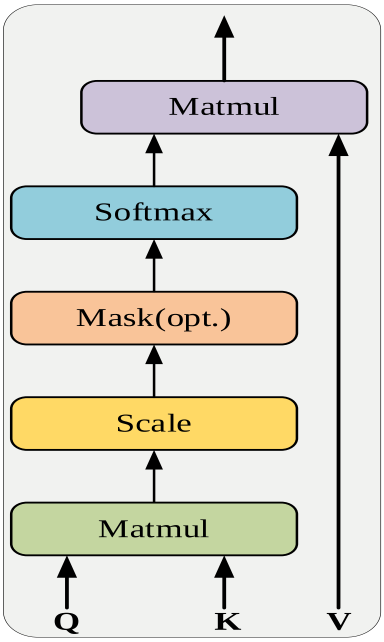
2.5. Dynamic Multi-Headed Attention Mechanism (D-MHSAM)

The multi-headed attention mechanism is a core component of the Transformer architecture, as illustrated in Figure 3b, consisting of multiple parallel self-attention heads as shown in Figure 3a. To fully exploit the complementary information from the time, frequency, and time–frequency domains, this study develops a Dynamic Multi-Headed Self-Attention Mechanism (D-MHSAM) to facilitate deep interactive fusion among the features Ftime, Ffreq, and Ftf extracted from the three independent branches.

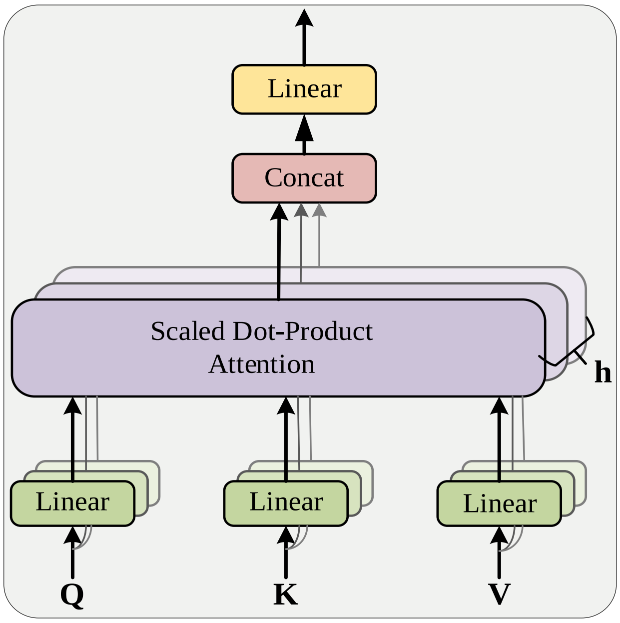
To integrate information from multiple domains, this study proposes a Dynamic Multi-Headed Self-Attention Mechanism (D-MHSAM). Before the attention computation, the complex-valued features from the CV-FCNN module are projected into the real-valued space by concatenating their magnitude, real part, and imaginary part, followed by a non-linear Multi-Layer Perceptron (MLP) transformation:

$$F_{freq} = \mathrm{MLP}([|F_{complex}|; \mathrm{Re}(F_{complex}); \mathrm{Im}(F_{complex})]) \tag{13}$$

All branch features are then aligned via Global Average Pooling (GAP) to obtain a unified feature representation, which serves as the input to the attention layer. In the attention layer, this unified feature is linearly projected to generate the Query (Q), Key (K), and Value (V) matrices, which respectively represent the feature queries, feature keys, and feature values used to compute the attention weights.



(**a**)



(**b**)

Figure 3. Structure of attention mechanisms: (**a**) structure of self-attention mechanisms; (**b**) structure of multiple attention mechanisms.

In conventional attention mechanisms, a fixed scaling factor is typically adopted. To adapt to the non-stationary energy fluctuations in vibration signals, D-MHSAM replaces the fixed scaling factor with a learnable parameter λ. This parameter adaptively modulates the Query (Q) and Key (K) representations.

To adapt to the non-stationary energy fluctuations in vibration signals, D-MHSAM replaces the fixed scaling factor with a learnable parameter λ. This parameter adaptively modulates the Query (Q) and Key (K) representations:

$$Q' = \lambda \cdot Q, \quad K' = \lambda \cdot K \tag{14}$$

Here, Q' and K' denote the dynamically scaled Query and Key, respectively, and λ is a learnable parameter that controls the concentration of attention weights, allowing the model to adaptively modulate the attention distribution for signals under different operating conditions.

The adjusted Q', K', and Value (V) are divided along the feature dimension into hhh attention heads, each with a dimension of dk. For each attention head, scaled dot-product attention is computed, followed by Softmax normalization, and then weighted summation of the Value vectors produces the attention output. To further enhance the adaptability of multi-headed attention under complex operating conditions, an adaptive attention head selection strategy is introduced. Each attention head is assigned a learnable gating weight:

$$\alpha_i = \sigma\left(W_g \cdot f_i + b_g\right) \tag{15}$$

where f_i denotes the feature representation of the i-th attention head, $\sigma(\cdot)$ is the Sigmoid activation function, and W_g and b_g are learnable parameters. The gating weight $\alpha_i \in (0, 1)$ dynamically adjusts the contribution of each attention head to the final output.

Finally, the outputs of all attention heads are weighted by their gating coefficients, concatenated, and linearly transformed to generate the final output of the D-MHSAM module. Through the synergistic effect of dynamic scaling and adaptive head selection, D-MHSAM can more effectively capture critical features of rolling bearing vibration signals, providing robust and discriminative high-level representations for subsequent fault diagnosis.

2.6. Loss Function

In multi-class classification tasks, the cross-entropy loss is widely adopted to quantify the discrepancy between the predicted class probability distribution and the ground-truth labels. It is typically used in conjunction with the Softmax function, which maps the raw network outputs into a normalized probability distribution over all fault categories. While the cross-entropy loss provides effective supervision at the classification level, it primarily focuses on output prediction accuracy and does not explicitly constrain the consistency of feature representations learned by different branches. The formulation of the cross-entropy loss is given as follows:

$$L_{CE} = -\sum_{i=1}^{N} y_i \log(p_i(x)) \tag{16}$$

where y_i is the true label of the ith fault category; $p_i(x)$ is the probability corresponding to the ith fault category.

Although the cross-entropy loss is effective for multi-class classification, relying solely on it may lead to overfitting when the network architecture becomes relatively complex. To further enhance the generalization capability of the proposed framework, an auxiliary inter-branch feature consistency loss is introduced as an additional regularization term.

Rather than constructing an explicit teacher–student framework, the proposed consistency loss enforces the alignment of intermediate feature representations learned by different parallel branches. By encouraging cross-branch feature agreement, this constraint guides the network toward more compact and discriminative representations, stabilizes the training process, and improves generalization performance without introducing additional inference complexity. The formulation of the inter-branch feature consistency loss is defined as follows:

$$L_{FD} = -\sum_{i=1}^{N} p_i^T \log\left(q_i^T(x)\right) \tag{17}$$

where p_i^T q_i^T denote the normalized feature representations extracted from two different parallel branches of the proposed PMFDL framework, respectively.

In order to adjust the weights between different loss terms and prevent overfitting, a balance factor β is introduced, and the formula for the total loss is

$$L = \beta L_{CE} + (1-\beta) L_{FD} \tag{18}$$

3. Methodology

The bearing fault diagnosis process based on PMFDL proposed in this paper is shown in Figure 4.

(1) Systematically collected vibration signal data of rolling bearings under various operating conditions using professional data acquisition software. The data were then divided into training and test sets according to scientific criteria. The expected out-

come of this step is to ensure sufficient coverage of different fault types and guarantee the accuracy and reliability of model learning and evaluation.

(2) Traditional statistical methods, such as time-domain and frequency-domain indicators, are limited in capturing weak, non-linear fault signatures and often fail to fully utilize multi-domain information present in rolling bearing vibration signals. To overcome these limitations, the PMFDL framework integrates time-domain, frequency-domain, and time–frequency-domain features in a parallel deep learning architecture, enabling the automatic and adaptive extraction of complementary multi-domain fault information.

(3) The raw signals are simultaneously input into the Res-BiGRU module for capturing long-distance temporal dependencies, the CV-FCNN module for extracting frequency-domain features while preserving phase information, and the DWT-S-DSC module (enhanced by an SRM high-pass filter) for capturing refined high-frequency transient details in the time–frequency domain.

(4) The extracted features from these three parallel branches are input into the Dynamic Multi-head Self-Attention Mechanism (D-MHSAM), which utilizes a gated strategy and dynamic scaling to perform interactive learning and achieve the adaptive fusion of cross-domain information. This step aims to combine complementary multi-domain features to improve fault discrimination capability.

(5) The fused features are flattened into one-dimensional vectors and input to the classification layer for fault diagnosis. The model weights are updated using the reverse communication algorithm during training. After training, an independent test dataset is used to evaluate the generalization ability and diagnostic accuracy of the model. This step validates the model's effectiveness in correctly identifying and classifying various bearing fault types.

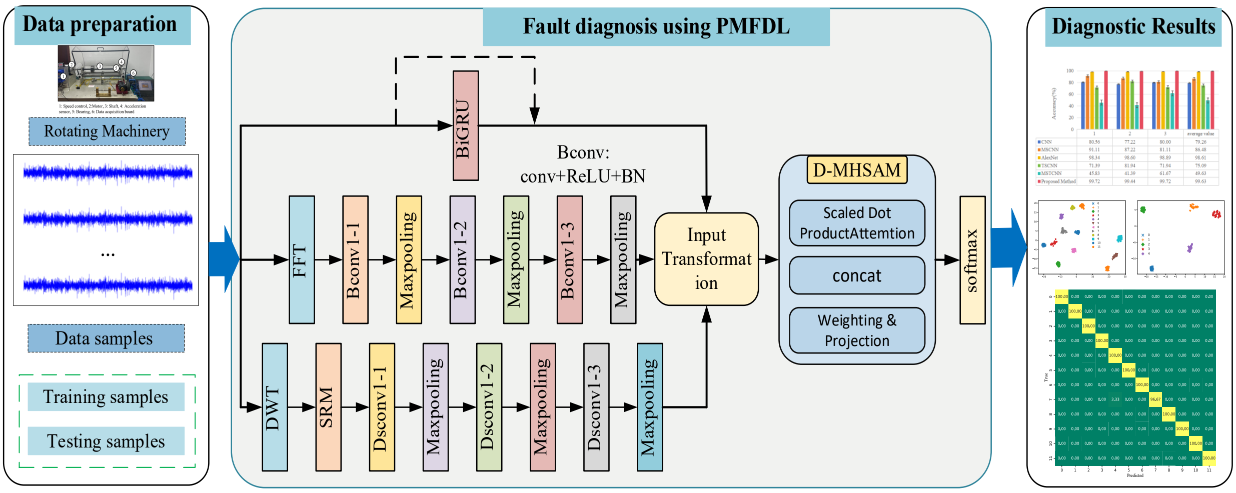


Figure 4. PMFDL rolling bearing fault diagnosis flow chart.

4. Huazhong University of Science and Technology Rolling Bearing Data Analysis

4.1. Data Introduction

The publicly available rolling bearing fault dataset from Huazhong University of Science and Technology (HUST) [21,22] was employed to evaluate the diagnostic performance of the proposed model. As illustrated in Figure 5, the vibration signals were acquired using a Spectra Quest mechanical fault simulator (MFS) (Spectra Quest, Inc., Richmond, VA, USA). The experimental setup was powered by a 2 HP (1.5 kW) three-phase induction motor (2), with its rotational speed regulated by a variable frequency speed controller (1).

The motor was coupled to a 1-inch (25.4 mm) diameter transmission shaft (3) supported by two ER-16K rolling bearings (manufactured by MB/Rexnord, Milwaukee, WI, USA) (5). These bearings feature a bore diameter of 25.4 mm and nine rolling elements. To simulate real-world industrial loading, a magnetic powder brake was utilized to apply adjustable radial torque. Vibration signals were captured at a sampling frequency of 25.6 kHz via TREA331 acceleration sensors (4) and recorded through a high-speed data acquisition board (6).

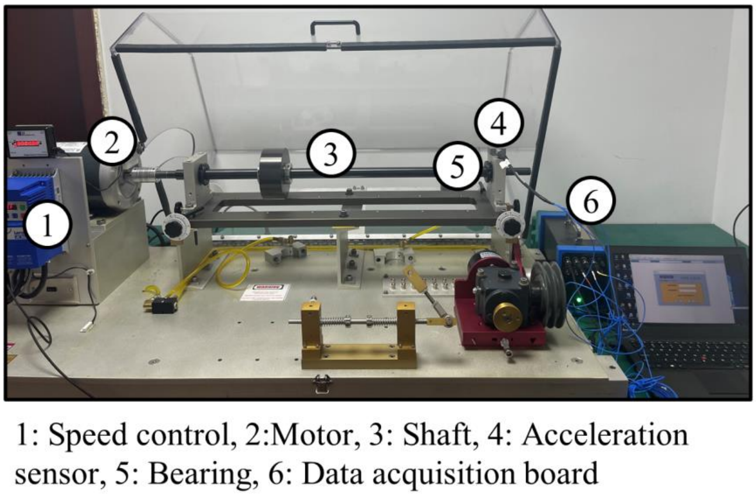


Figure 5. HUST-bearing dataset test stand.

To ensure the authenticity of the fault signatures, single-point defects—including inner ring, outer ring, and rolling element faults—were artificially induced via electrical discharge machining (EDM). This precision technique created micro-cracks with controlled dimensions to represent both moderate and severe damage levels. The specific morphologies of these induced damages are visualized in Figure 6, where the red markers highlight the material removal areas on the inner race, outer race, and balls. The resulting dataset encompassed four rotational frequencies (65 Hz, 70 Hz, and 75 Hz). Detailed specifications of the fault categories, including combined faults, and their corresponding labels are summarized in Table 1.

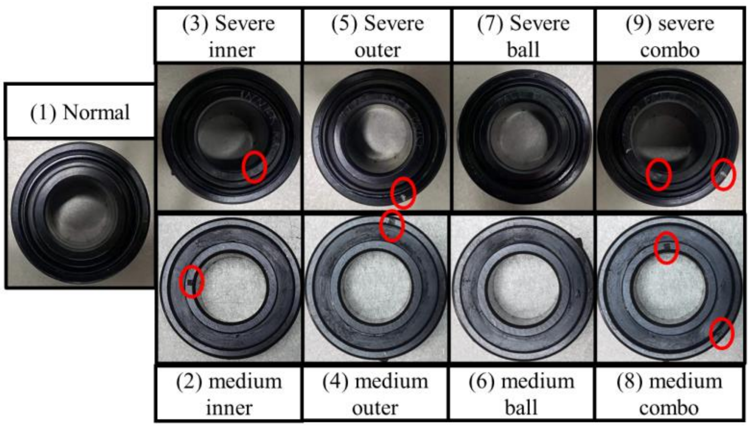


Figure 6. Microscopic morphology of rolling bearing damages induced by EDM: (1) normal, (2–3) inner ring faults, (4–5) outer ring faults, (6–7) ball faults, and (8–9) combined faults at medium and severe levels.

Table 1. Fault labels of the tested bearings in the HUST dataset.

Working Condition	Fault Type	Degree of Failure	Label
65 HZ	Health	Normal	0
	Inner Ring	Severe	1
	Outer Ring	Severe	2
	Rolling Element	Severe	3
70 HZ	Health	Normal	4
	Inner Ring	Severe	5
	Outer Ring	Severe	6
	Rolling Element	Severe	7
75 HZ	Health	Normal	8
	Inner Ring	Severe	9
	Outer Ring	Severe	10
	Rolling Element	Severe	11

4.2. Data Selection

The acquisition time of each set of signals was 10.2 s, and a total of 262,144 data points were acquired, covering the X, Y, and Z axial directions. For each axial direction, the data were analyzed separately for fault diagnosis, and the fusion of axial data was also considered, including the fusion of the X and Y axes, the X and Z axes, and the Y and Z axes, as well as the fusion of data from all three axial directions. These diagnostic results are recorded in detail in Table 2. In the data processing stage, 50,000 data points were selected from each set of signals (25,000 points per axis for the two-axis fusion dataset and 16,667 points per axis for the three-axis fusion dataset) and divided into 100 samples, each containing 500 consecutive data points. For all fault signals, a total of 1200 samples were obtained, constituting the sample dataset. In order to scientifically evaluate the performance of the method in this paper, these sample datasets are divided into a training set and a test set with a ratio of 70% and 30%, respectively. The evaluation criteria were accuracy, F1 score, precision and recall. In order to reduce the effect of randomness, each experiment was conducted three times and the average result of the three experiments is the final result.

Table 2. Results of the evaluation criteria for each axis.

Axle	Average Accuracy/%	Average F1 Score/%	Average Precision Rate/%	Average Recall Rate/%
X	99.63	99.63	99.64	99.63
Y	99.17	99.17	99.20	99.17
Z	97.96	97.97	98.04	97.96
X, Y	97.41	97.41	97.54	97.40
X, Z	96.39	96.37	96.54	96.39
Y, Z	95.00	95.03	95.41	95.00
X, Y, Z	92.68	92.68	92.93	92.68

It can be seen from the experimental results that the model of the X-axis data shows excellent performance in the diagnosis of the bearing fault. The accuracy rate has reached 99.63%, the F1 score is 99.63%, the accuracy rate is 99.64%, and the recall rate is 99.63%. These indicators not only exceed the results of the Y-axis or Z-axis data alone, but also exceed the results when using multi axis data. This shows that the data of the Y-axis and the Z-axis contribute less to the diagnosis of fault. Specifically, the results of the four evaluation indicators using the X-axis increased by 0.46%, 0.46%, 0.44% and 0.46% respectively compared with the Y-axis data. Compared with the Z-axis data, the results increased by 1.67%, 1.66%, 1.60% and 1.67% respectively. Compared with the X- and Y-axis fusion data, the results increased by 2.22%, 2.22%, 2.10% and 2.23% respectively. Compared with the X- and Z-axis fusion data, the results increased by 3.24%, 3.26%, 3.10% and 3.24% respectively. Compared with the Y- and Z-axis fusion data, the results increased by 4.63%, 4.60%, 4.23% and 4.63% respectively. Compared with the fusion data results of the X-, Y- and Z-axes, the results increased by 6.95%, 6.95%, 6.71% and 6.95% respectively. These improvements indicate that although X-axis data is single-axis data, its performance in bearing fault diagnosis is superior to multi-axis data. This phenomenon proves that the X-axis data can more effectively capture the bearing fault characteristics, so the X-axis data was selected for subsequent tests.

4.3. Signal Conversion

After dividing the data, fast Fourier transform is applied to the sample signals to convert the time-domain signals into frequency-domain signals, while the DWT is applied to the sample signals to convert them into time–frequency-domain signals. Figure 7 shows the spectrograms of some samples converted to frequency-domain signals and the time–frequency plots converted to time–frequency-domain signals.

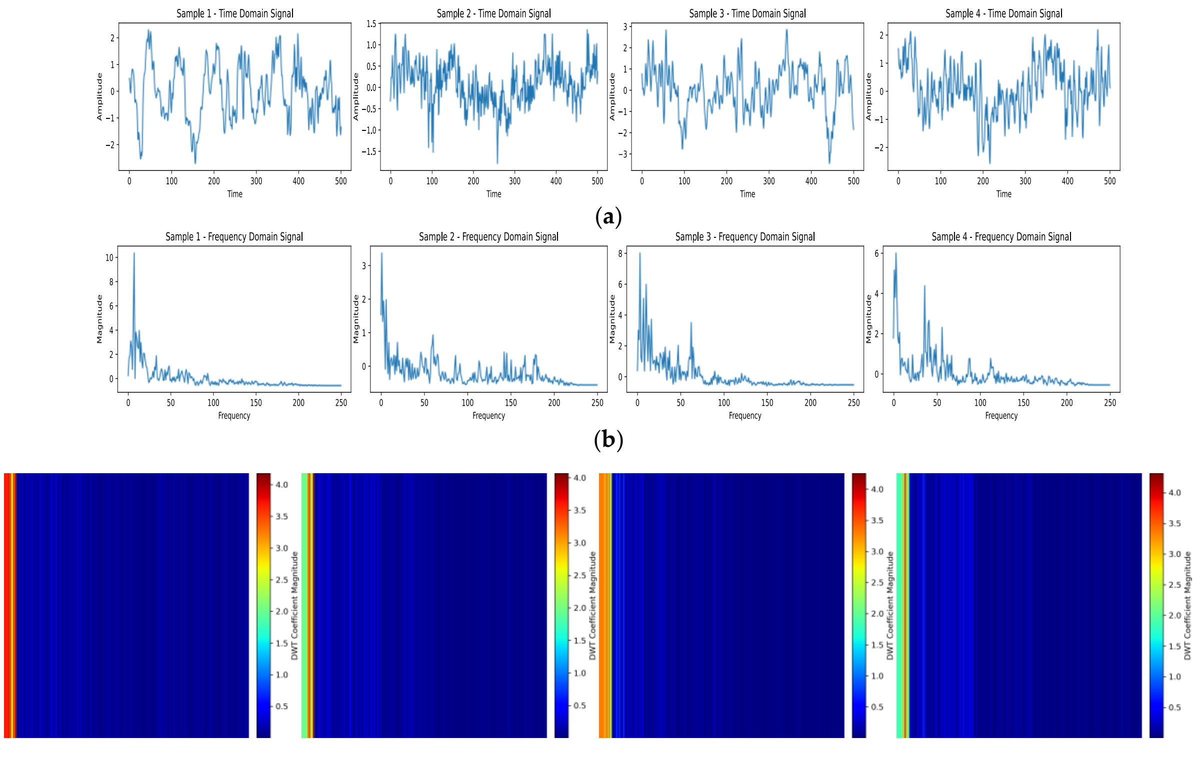


Figure 7. Preprocessed images of vibration signals from the HUST bearing dataset: (**a**) original vibration signal; (**b**) spectrum diagram; (**c**) DWT time frequency diagram.

From the time-domain diagram, it can be seen that the vibration signal presents non-linear and unstable characteristics, and this characteristic leads to the traditional fault diagnosis methods being unable to effectively extract features from the vibration signal. From the spectrum plot, it can be seen that the signal retains features such as frequency components, phase information, amplitude spectrum, and so on, after Fast Fourier Transform, which provides valuable information for fault diagnosis and feature extraction. The DWT converts the time-domain signals into the time–frequency domain in the process of retaining the features such as time, frequency and local information. By combining the original time-domain signal with the frequency-domain signal and time–frequency-domain signal, more diversified feature information can be extracted and richer feature vectors can be constructed.

It should be noted that all vibration signals in the HUST dataset were collected under constant-speed operating conditions (65–75 Hz) and can therefore be regarded as stationary. This property facilitates feature extraction and contributes to the high performance of the proposed model under these conditions.

4.4. Result Discussion

The proposed model is trained and tested using this dataset, with an initial learning rate of 0.001, a batch size of 32, an optimizer of Adam, and 50 iterations. To minimize the effect of randomness, each trial is performed three times and the final average accuracy is 99.63%. In order to more intuitively show the accurate precision of the proposed model for each classification, the confusion matrix of the fault classification results is plotted. This is shown in Figure 7a, where the horizontal axis indicates the diagnostic state and the vertical axis indicates the actual fault state. In order to increase the interpretability of the model, the final classification results are visualized with T-SNE, as shown in Figure 8b.

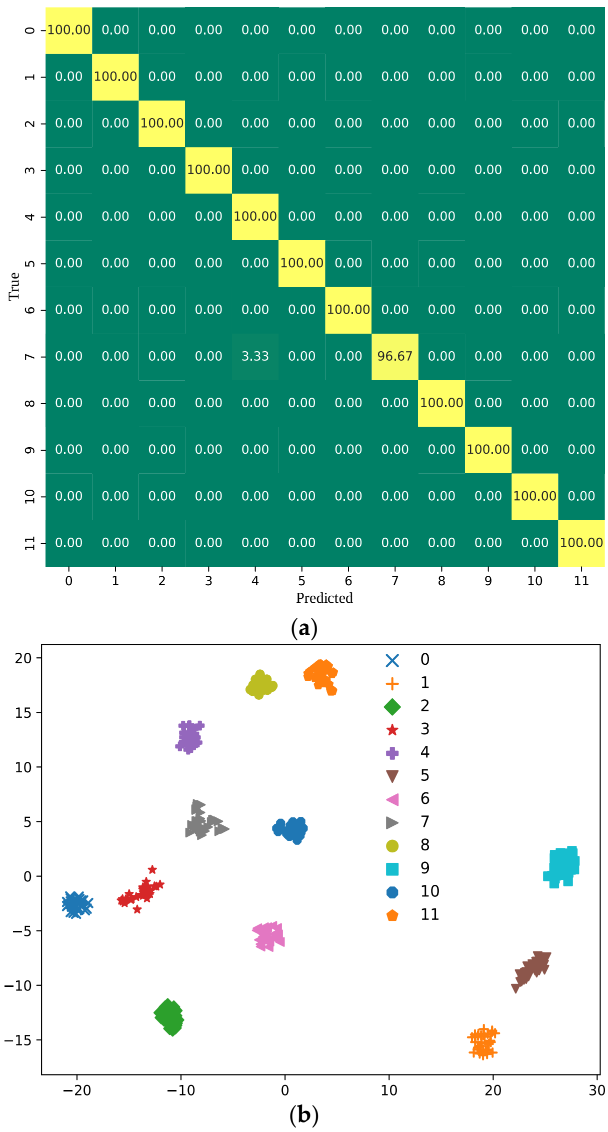


Figure 8. Experimental results on the HUST dataset: (**a**) confusion matrix; (**b**) T-SNE feature visualization.

In the figure, 0–11 correspond to 12 categories of state, with the values on the diagonal representing the recognition accuracy of each state and the remaining values representing the recognition error rate. The results of the confusion matrix show that except for category 7, which had 3.33% misdiagnosis as category 4, the diagnostic rates for all other categories were 100%, with no misdiagnosis. The feature visualization of T-SNE shows clear boundaries between each category, and even after reducing from high-dimensional features to a two-dimensional view, each fault category still has extremely high distinguishability. The method proposed in this article demonstrated extremely high classification accuracy and stability in the field of rolling bearing fault recognition, as well as the ability to effectively extract and maintain classification differences in complex feature spaces, effectively reducing confusion between categories.

While the above results demonstrate the effectiveness of the proposed method in terms of diagnostic accuracy and feature separability, computational efficiency and response time are also critical factors for practical maintenance applications. Although the proposed method integrates vibration signal information from the time domain, frequency domain, and time–frequency domain, the associated computational burden remains within a reasonable range. The signal preprocessing operations, such as fast Fourier transform (FFT)-based spectral analysis and DWT, have relatively low computational complexity and can be efficiently implemented.

Furthermore, the proposed DWT-S-DSC module adopts depthwise separable convolution, which significantly reduces computational complexity and parameter numbers compared with conventional convolution operations. As analyzed in Section 2.4, this design effectively improves inference efficiency while maintaining high feature representation capability.

It should be noted that the training process of the proposed model is performed offline, whereas only the forward inference process is required during deployment. Therefore, the response time for fault diagnosis is sufficiently short to meet the requirements of near-real-time condition monitoring, indicating that the proposed method is suitable for implementation in practical maintenance programs.

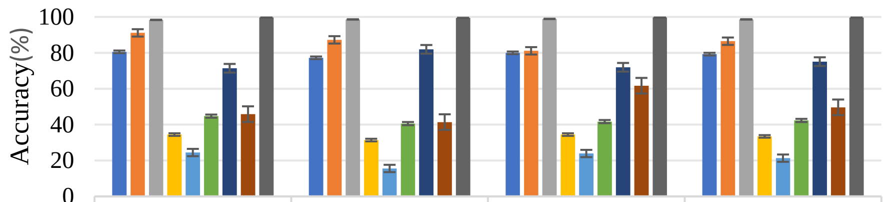
4.5. Contrast Experiment

All comparison models were implemented following their original architectures and hyperparameter settings, and adapted to vibration signal inputs without task-specific redesign, in order to ensure a consistent and fair comparison. To verify the performance of the proposed method, it was compared with traditional CNN, MSCNN, AlexNet, CapsNet, GCN, Transformer, and advanced TSCNN [23], MSTCNN [24] network model, as follows: (1) CNN, MSCNN, AlexNet, CapsNet, GCN, and Transformer models are direct inputs of the original vibration signals; (2) TSCNN is a parallel network model with inputs from two different domains: the time domain and time–frequency domain; (3) MSTCNN: a parallel network model with inputs from three different domains: the time domain, frequency domain, and time–frequency domain. To ensure fairness, all models are trained using the same initial parameters as the method proposed in this paper. The evaluation criteria are accuracy, F1 score, precision and recall. The results are shown in Figure 8 and Table 3. To ensure the statistical reliability of the results, three independent trials (labeled as 1, 2, and 3 in Figure 9) were conducted for each model under the same experimental conditions. The final evaluation is based on the ‘average value’ derived from these three groups of data.

https://doi.org/10.3390/app16041846

Table 3. Results of the evaluation criteria for each model.

Model	Average F1 Score/%	Average Precision Rate/%	Average Recall Rate/%
CNN	78.73	81.65	79.26
MSCNN	86.17	87.87	86.48
AlexNet	98.61	98.73	98.61
CapsNet	34.37	42.49	33.42
GCN	17.15	20.25	21.30
Transformer	41.54	43.60	42.60
TSCNN	69.69	80.23	75.09
MSTCNN	39.28	47.56	49.63
PMFDL	**99.63**	**99.64**	**99.63**



	1	2	3	average value
CNN	80.56	77.22	80.00	79.26
MSCNN	91.11	87.22	81.11	86.48
AlexNet	98.34	98.60	98.89	98.61
CapsNet	34.44	31.39	34.44	33.42
GCN	24.44	15.56	23.89	21.30
Transformer	44.72	40.56	41.67	42.32
TSCNN	71.39	81.94	71.94	75.09
MSTCNN	45.83	41.39	61.67	49.63
PMFDL	99.72	99.44	99.72	99.63

Figure 9. Comparison of the accuracy of each model.

Based on the experimental results, the model proposed in this paper demonstrates superior performance over the compared models in all four evaluation metrics. Among the eight models involved in the comparison, CapsNet, GCN, Transformer and MSTCNN models failed to break the 50% threshold in performance, while the others achieved more than 70% on all metrics except TSCNN's average F1 score. The model proposed in this paper achieved significant improvement in all four evaluation metrics compared to the remaining eight models, as shown in the following: the model proposed in this paper achieves 20.37%, 13.15%, 1.02%, 66.21%, 78.33%, 57.31%, 24.54%, and 50% performance improvements in terms of the average diagnosis rate relative to the comparison models; in terms of the average F1 score, PMFDL achieves 20.90%, 13.46%, 1.02%, 65.26%, 82.48%, 58.09%, 29.94%, and 60.35% performance improvements; and in precision accuracy, it achieves 17.99%, 11.77%, 0.91%, 57.15%, 79.39%, 56.04%, 19.41% and 52.08% performance improvements, while 20.37%, 13.15%, 1.02%, 66.21%, 78.33%, 57.03%, 24.54%, and 50% performance improvements were achieved in average recall. Compared with MSTCNN, which also has input signals from three different domains—the time domain, frequency domain, and time–frequency domain—the evaluation metrics of the proposed method in this paper are more than 50% higher than those of MSTCNN, due to the fact that MSTCNN's time-domain signals are extracted with features using RestNet34, frequency-domain signals are extracted using VGG16, and time–frequency-domain signals are extracted with features using RestNet18. All three models are made of convolutional stacks with different kernel sizes, which are effective but not optimized for each domain. In this paper, the method uses BiGRU to extract features for time-domain signals, CNN to extract features for frequency-domain signals, and Enhanced Depth Separable Module for time–frequency-domain signals, which is a more suitable model to be selected for feature extraction of the signal characteristics in different domains. Additionally, MSTCNN in the final feature fusion, when using the direct two–two fusion strategy, may lead to the loss of important features or interference from secondary features. In this paper, through the integration of dynamic multi-head attention mechanisms, important features can be made more prominent and unimportant features can be suppressed. The necessity and effectiveness of feature extraction from the three dimensions of the time domain, frequency domain and time–frequency domain are verified so that the complexity and diversity of bearing faults can be more accurately captured, thus significantly improving the accuracy and reliability of fault diagnosis. This result proves that the proposed method not only performs well in terms of effectiveness, but also has significant advantages in terms of generalization ability and robustness.

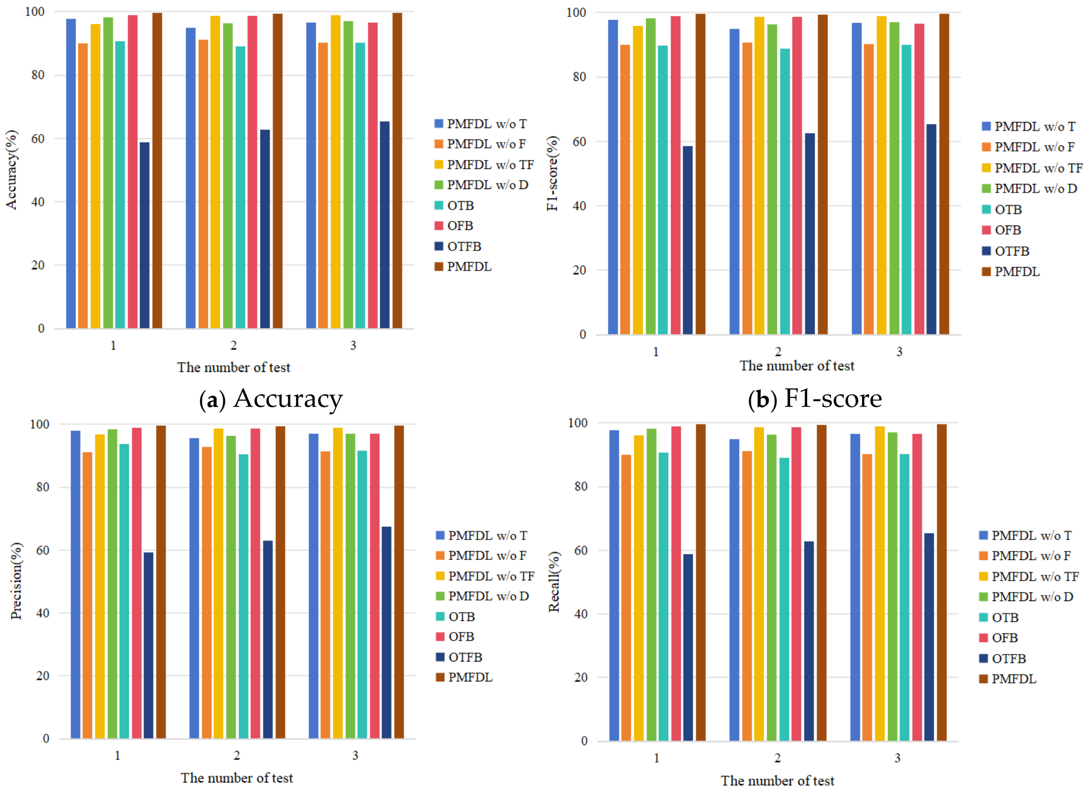
4.6. Ablation Test

Ablation experiments were conducted to verify the effectiveness of each branch in the proposed method. The models analyzed for comparison are as follows: (a) no time-domain branching (PMFDL w/o T); (b) no frequency-domain branching (PMFDL w/o F); (c) no time–frequency-domain branching (PMFDL w/o TF); (d) no D-MHSAM (PMFDL w/o D) (e) only the time-domain branching (OTB); (f) only the frequency-domain branching (OFB); (g) only time–frequency-domain branch (OTFB); (h) the method proposed in this paper. The network parameters are all set as in the proposed method. Table 4 and Figure 10 record detailed results of the ablation experiments.

Table 4. Results of evaluation criteria for each model.

Model	Average Accuracy/%	Average F1 Score/%	Average Precision Rate/%	Average Recall Rate/%
PMFDL w/o T	96.48	96.49	96.96	96.48
PMFDL w/o F	90.46	90.35	91.87	90.46
PMFDL w/o TF	97.87	97.63	98.13	97.87
PMFDL w/o D	97.31	97.31	97.34	97.31
OTB	90.09	89.54	92.02	90.09
OFB	98.06	98.07	98.26	98.06
OTFB	62.32	62.18	63.26	62.32
PMFDL	**99.63**	**99.63**	**99.64**	**99.63**

From Table 4 and Figure 10, it can be seen that removing the time-domain branch (a) leads to a decrease of 3.15% in accuracy, 3.15% in F1 score, 2.68% in precision, and 3.15% in recall, which shows the critical role of time-domain features in fault diagnosis and indicates that time-domain information is crucial in capturing the dynamics of the signal. Removing the frequency-domain branch (b) decreases the accuracy by 9.17%, F1 score by 9.28%, precision by 7.77%, and recall by 9.17%, which highlights the central value of frequency-domain features in identifying failure modes and reveals the importance of the frequency component in understanding signal characteristics. Removing the time–frequency-domain branch (c) decreases the accuracy by 1.76%, F1 score by 2.00%, precision by 1.51%, and recall by 1.76%, which demonstrates that time–frequency-domain features, although contributing to the model performance, have a lesser impact compared to time- and frequency-domain features. Removing D-MHSAM (d) decreased the accuracy by 2.32%, F1 score by 2.32%, precision by 2.30%, and recall by 2.32%, indicating that this mechanism plays an important role in enhancing the feature expression ability, which in turn affects the diagnostic performance of the model. Removing the frequency- and time–frequency-domain branch (e) decreased the accuracy by 9.45%, F1 score by 10.09%, precision by 7.62%, and recall by 9.45%, showing the importance of the frequency- and time–frequency-domain features for the diagnostic capability of the model. The frequency-domain features can reveal the periodic components of the signal, while the time–frequency-domain features provide the localized information of the signal in time and frequency, and the combination of the two can provide a more comprehensive analysis of the signal characteristics. The simultaneous removal of time-domain and time–frequency-domain branch (f) decreases the accuracy by 1.57%, the F1 score by 1.56%, the precision by 1.38%, and the recall by 1.58%, demonstrating the importance of the time-domain signals along with the time–frequency-domain signals. Time-domain features capture the dynamic changes in the signal, while time–frequency-domain features provide the local frequency information of the signal, and the combination of the two helps to improve the model's sensitivity and recognition of signal changes. The simultaneous removal of the time-domain and time–frequency-domain branch (g) decreases the accuracy by 37.31%, the F1 score by 37.45%, the precision by 36.38%, and the recall by 37.31%, which highlights that the time-domain and frequency-domain features have a significant impact on the model. Time-domain features reflect the original fluctuation pattern of the signal, while frequency-domain features reveal the frequency components and energy distribution of the signal. The combination of the two is essential for constructing a model capable of accurately diagnosing bearing faults, and together they provide the model with a full picture of the signal, including its dynamic and frequency characteristics. In summary, these results not only validate the importance of comprehensive feature extraction from the three dimensions of time domain, frequency domain and time–frequency domain, but also reveal the specific impact of different signal types on model performance.



(**a**) Accuracy (**b**) F1-score

(**c**) Precision (**d**) Recall

Figure 10. Experimental results of the ablation experiment.

5. University of Cincinnati Rolling Bearing Data Analysis

5.1. Data Introduction

The rolling bearing fault dataset from the University of Cincinnati [25] was utilized to further validate the diagnostic performance of the proposed model. As illustrated in the schematic of the test stand in Figure 11, four ZA-2115 double-row rolling bearings (manufactured by Rexnord) were mounted on a single shaft driven by an AC motor at a constant speed of 2000 rpm. A radial load of 6000 lbs was applied to the bearings through a spring mechanism, and all bearings were force-lubricated. High-precision quartz accelerometers were employed to measure the vibration signals in both the X and Y directions.

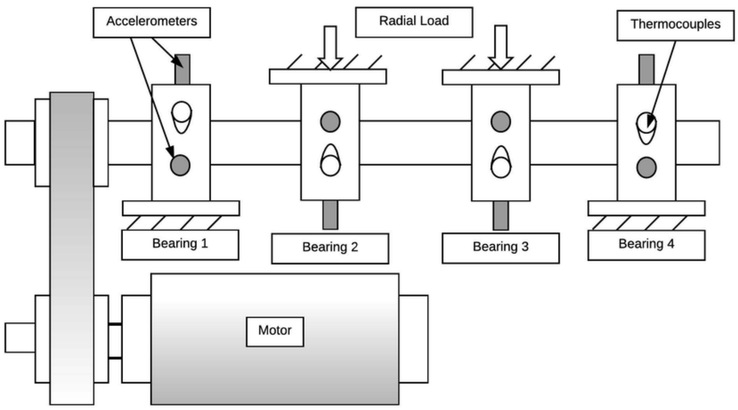


Figure 11. Schematic diagram of the University of Cincinnati (IMS) bearing test rig showing the spatial arrangement of Bearings 1, 2, 3, and 4.

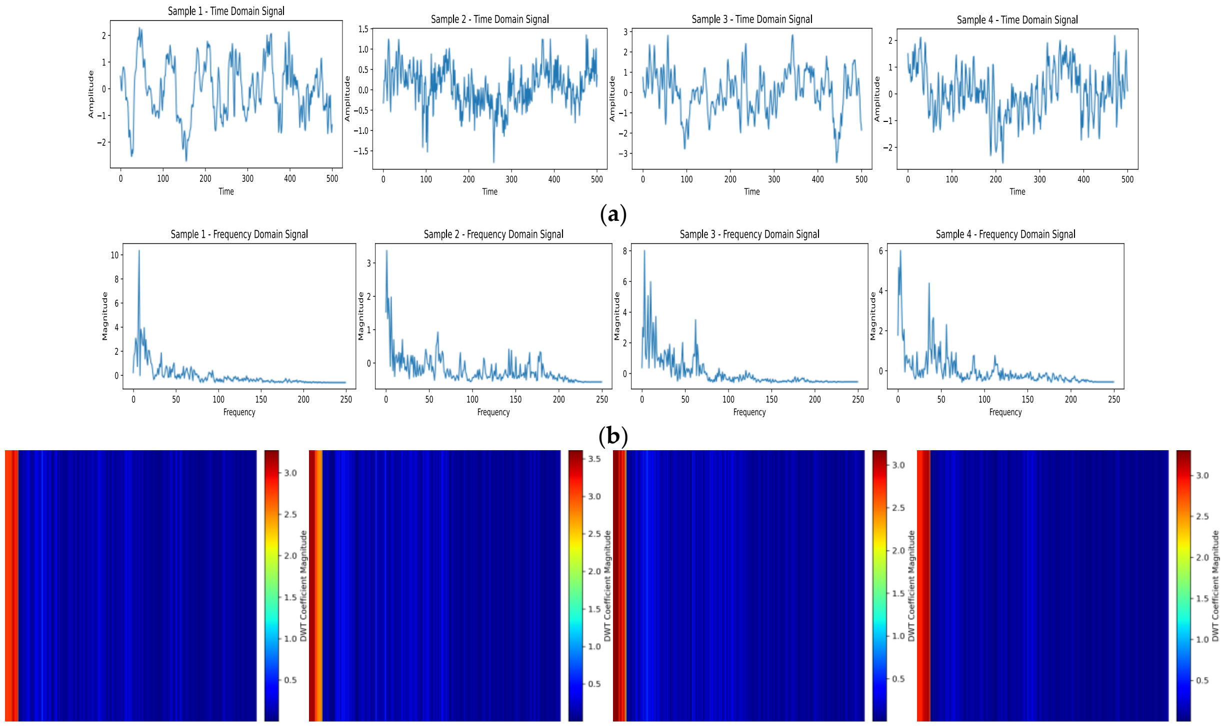
A DAQ 6062E data acquisition card was used for signal acquisition with a sampling frequency of 20 kHz. Three distinct datasets were collected: Dataset 1 included vibration signals in both X and Y directions, while Datasets 2 and 3 recorded signals in a single direction only. Regarding the fault occurrences, Bearing 3 in Dataset 1 exhibited an inner ring failure, and Bearing 4 experienced a rolling element failure by the end of the test. In Dataset 2, an outer-ring fault was identified in Bearing 1, while Bearing 3 in Dataset 3 similarly demonstrated an outer-ring fault. The detailed technical specifications for the ZA-2115 bearings, including a pitch diameter of 71.5 mm, a roller diameter of 11.112 mm, and 16 rollers per row, were incorporated to support reproducibility. The fault categories and their corresponding labels for the selected data were summarized in Table 5.

Table 5. Fault labels of the tested bearings in the dataset.

Data Set	Bearing	Fault Type	Label
Data Set 1	Bearing 1	Normal	0
	Bearing 3	Inner Ring	1
	Bearing 4	Rolling Element	2
Data Set 2	Bearing 1	Outer Ring	3
Data Set 3	Bearing 3	Outer Ring	4

5.2. Signal Conversion

The acquisition time of each group of signals is 1 s, with a total of 20,480 data points. For the vibration signals collected in the two directions of the data set, the data points in the X-axis direction are selected. Each set of signals selected 61,000 data points, which were in turn divided into 122 samples with a length of 500 points. A total of 610 sample data sets can be obtained for all fault signals, with each sample containing 500 data points. In order to scientifically evaluate method performance, the data is divided into two parts—70% is used as a training set, and the remaining 30% is used as a test set—to check the generalization ability of the model. After the data is divided, the sample signal is converted into the frequency-domain signal by Fast Fourier Transform (FFT), and the sample signal is transformed into the time–frequency-domain signal by DWT. Figure 12 shows the spectrum diagram of partial sample conversion to frequency-domain signals and the time–frequency diagram of conversion to time–frequency-domain signals.

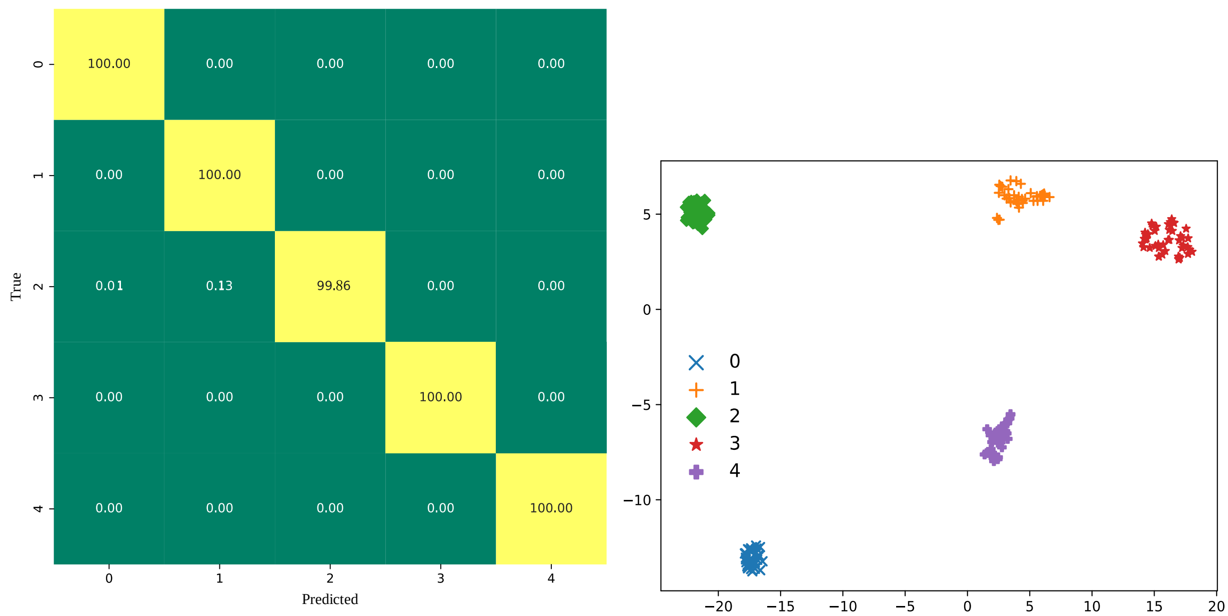


(c)

Figure 12. Preprocessed image of vibration signals: (**a**) original vibration signal; (**b**) spectrum diagram; (**c**) DWT time–frequency diagram.

5.3. Result Discussion

The proposed model is trained and tested using this dataset, with a final average accuracy of 99.82%. The confusion matrix and T-SNE visualization of its classification results are shown in Figure 13.



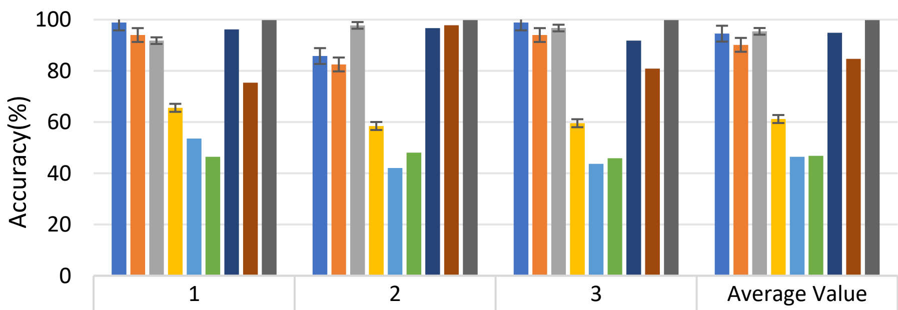
(**a**) (**b**)

Figure 13. Diagram of results: (**a**) confusion matrix; (**b**) T-SNE feature visualization.

The results of the confusion matrix showed that the diagnosis rate of all categories was 99.82% on average, with almost no misdiagnosis. The feature visualization of T-SNE shows that the boundaries of various fault categories are clearly defined, and the fault categories are still highly distinguishable after dimensionality reduction from high-dimensional features to two-dimensional views. It is verified that the proposed method shows high classification accuracy and stability in the field of rolling bearing fault identification, and the ability to effectively extract and maintain classification differences in complex feature space, effectively reducing the confusion between classes.

5.4. Contrast Experiment

Use the University of Cincinnati Bearing Data Set to compare the methods with the traditional CNN, MSCNN, AlexNet, CapsNet, GCN, Transformer and advanced TSCNN, MSTCNN network models. The results are shown in Figure 14 and Table 6.



	1	2	3	Average Value
CNN	98.91	85.79	98.91	94.54
MSCNN	93.99	82.51	93.99	90.16
AlexNet	91.8	97.76	96.73	95.43
CapsNet	65.57	58.47	59.56	61.2
GCN	53.55	42.08	43.72	46.45
Transformer	46.45	48.09	45.9	46.81
TSCNN	96.17	96.72	91.8	94.9
MSTCNN	75.41	97.81	80.87	84.7
PMFDL	99.82	99.80	99.84	99.82

Figure 14. Comparison of the accuracy of each model.

Table 6. Results of the evaluation criteria for each model.

Model	Average F1 Score/%	Average Precision Rate/%	Average Recall Rate/%
CNN	94.21	95.75	94.54
MSCNN	89.73	92.72	90.16
AlexNet	95.30	96.00	95.44
CapsNet	59.28	77.69	61.20
GCN	45.67	48.86	46.45
Transformer	45.18	46.50	46.20
TSCNN	94.90	96.20	94.90
MSTCNN	80.40	77.82	84.70
PMFDL	**99.83**	**99.80**	**99.82**

According to the experimental results, among the eight models involved in the comparison, a level of more than 80% is achieved on all the indicators except all four indicators of CapsNet, GCN and Transformer, and the average precision rate of MSTCNN. The model proposed in this paper achieves significant improvement in all four evaluation metrics compared to the remaining eight models, as shown in the following: relative to the comparison models, the model proposed in this paper achieves a performance improvement of 5.28%, 9.66%, 4.38%, 38.62%, 53.37%, 53.62%, 4.92%, and 15.12%, respectively, in the average diagnosis rate; in terms of the average F1 score, PMFDL achieves 5.62%, 10.10%, 4.53%, 40.55%, 54.16%, 54.65%, 4.93%, and 19.43% performance improvement; and in average precision, it achieves 4.05%, 7.08%, 3.80%, 22.11%, 50.94%, 53.30%, 3.60%, and 21.98%; and performance improvements of 5.28%, 9.66%, 4.38%, 38.62%, 53.37%, 53.62%, 4.92% and 15.12% are realized in the average recall rate. The necessity and effectiveness of comprehensively extracting features from three dimensions, namely, the time domain, frequency domain and time–frequency domain, are verified, enabling it to more accurately capture the complexity and diversity of bearing faults, thus significantly improving the accuracy and reliability of fault diagnosis. This result proves that the method proposed in this paper not only excels in terms of effectiveness, but also has significant advantages in generalization ability and robustness. Figure 12 shows that the stability of the method proposed in this paper is significantly improved.

5.5. Ablation Test

Ablation experiments were conducted to verify the effectiveness of each branch in the proposed method. The models analyzed for comparison are as follows: (a) no time-domain branching (PMFDL *w*/*o* T); (b) no frequency-domain branching (PMFDL *w*/*o* F); (c) no time–frequency-domain branching (PMFDL *w*/*o* TF); (d) no D-MHSAM (PMFDL *w*/*o* D) (e) only the time-domain branching (OTB); (f) only the frequency-domain branching (OFB); (g) only time–frequency-domain branching (OTFB); (h) the method proposed in this paper. The network parameters are all set as in the proposed method. Table 7 shows the detailed results of the ablation experiments.

Table 7. Results of evaluation criteria for each model.

Model	Average Accuracy/%	Average F1 Score/%	Average Precision Rate/%	Average Recall Rate/%
PMFDL *w*/*o* T	96.90	96.86	97.42	96.90
PMFDL *w*/*o* F	93.99	93.64	94.88	93.99
PMFDL *w*/*o* TF	96.17	96.10	96.88	96.17
PMFDL *w*/*o* D	96.36	96.35	96.58	96.36
OTB	93.80	93.66	94.40	93.80
OFB	91.44	90.97	94.32	91.44
OTFB	93.61	93.67	94.01	93.62
PMFDL	**99.82**	**99.83**	**99.80**	**99.82**

From Table 7, it can be seen that removing the time-domain branch (a) leads to a decrease of 2.92% in accuracy, 2.97% in F1 score, 2.38% in precision, and 2.92% in recall, which shows the critical role of time-domain features in fault diagnosis and indicates that time-domain information is crucial in capturing the dynamics of the signal. Removing the frequency-domain branch (b) decreases the accuracy by 5.83%, F1 score by 6.19%, precision by 4.92%, and recall by 5.83%, which highlights the central value of frequency-domain features in identifying failure modes and reveals the importance of the frequency component in understanding signal characteristics. Removing the time–frequency-domain branch (c) decreases the accuracy by 3.65%, F1 score by 3.73%, precision by 2.92%, and recall by 3.65%, which demonstrates that time–frequency-domain features, although contributing to the model performance, have a lesser impact compared to time- and frequency-domain features. Removing D-MHSAM (d) decreased the accuracy by 3.46%, F1 score by 3.48%, precision by 3.22%, and recall by 3.46%, indicating that this mechanism plays an important role in enhancing the feature expression ability, which in turn affects the diagnostic performance of the model. Removing the frequency- and time–frequency-domain branch (e) decreased the accuracy by 6.02%, F1 score by 6.17%, precision by 5.40%, and recall by 6.02%, showing the importance of the frequency- and time–frequency-domain features for the diagnostic capability of the model. The frequency-domain features can reveal the periodic components of the signal, while the time–frequency-domain features provide the localized information of the signal in time and frequency, and the combination of the two can provide a more comprehensive analysis of the signal characteristics. The simultaneous removal of time-domain and time–frequency-domain branch (f) decreases accuracy by 8.38%, the F1 score by 8.86%, the precision by 5.48%, and the recall by 8.38%, demonstrating the importance of the time-domain signals along with the time–frequency-domain signals. Time-domain features capture the dynamic changes in the signal, while time–frequency-domain features provide the local frequency information of the signal, and the combination of the two helps to improve the model's sensitivity and the recognition of signal changes. The simultaneous removal of the time-domain and time–frequency-domain branch (g) decreases the accuracy by 6.21%, the F1 score by 6.16%, the precision by 5.79%, and the recall by 6.20%, which highlights that the time-domain and frequency-domain features have a significant impact on the model. Time-domain features reflect the original fluctuation pattern of the signal, while frequency-domain features reveal the frequency components and energy distribution of the signal. The combination of the two is essential for constructing a model capable of accurately diagnosing bearing faults, and together they provide the model with a full picture of the signal, including its dynamic and frequency characteristics. In summary, these results not only validate the importance of comprehensive feature extraction from the three dimensions of time domain, frequency domain and time–frequency domain, but also reveal the specific impact of different signal types on model performance.

6. Conclusions

This paper proposed a Parallel Multi-dimensional Feature Deep Learning (PMFDL) framework to improve rolling bearing fault diagnosis accuracy using vibration signals. Unlike traditional methods that rely on handcrafted features or a single neural network operating in one signal domain, the proposed approach performs parallel modeling and the fusion of time-domain, frequency-domain, and time–frequency-domain features, enabling the complementary utilization of multi-dimensional fault information.

In the PMFDL framework, raw vibration signals are processed by a Res-BiGRU module to extract temporal features, transformed into the frequency domain via Fast Fourier Transform (FFT) and analyzed using CV-FCNN to capture spectral characteristics, and converted into two-dimensional time–frequency representations through Discrete Wavelet Transform (DWT) for feature learning with the S-DSC module. An SRM high-pass filter is introduced to enhance high-frequency and transient fault components, while a Dynamic Multi-Head Self-Attention Mechanism (D-MHSAM) is employed to adaptively fuse multi-domain features.

Experimental validation on rolling bearing datasets from Huazhong University of Science and Technology and the University of Cincinnati demonstrates that the proposed PMFDL method achieves diagnostic accuracies of 99.63% and 99.82%, respectively. Compared with several representative deep learning models, the proposed method shows superior diagnostic performance and generalization ability. Ablation experiments confirm the necessity of each feature extraction branch, and visualization results further demonstrate clear class separability and effective fault feature discrimination.

The scientific justification of this study is that rolling bearing vibration signals inherently exhibit multi-domain and multi-scale characteristics, which cannot be fully captured by single-domain feature extraction strategies. By jointly preserving and exploiting complementary information across multiple domains, the PMFDL framework provides an effective solution for weak fault feature extraction. Future work will focus on fault diagnosis under conditions with limited labeled data and extending the proposed framework to variable-speed and non-stationary operating scenarios to enhance its applicability in practical industrial environments.

Author Contributions: B.H.: methodology, software, funding acquisition, writing—review and editing. J.Z.: writing—original draft, conceptualization. S.L.: validation, formal analysis, investigation. L.T.: validation, formal analysis, investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Henan Province science and technology attack key project (Grant No. 252102240042).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data included in this study are available upon request by contact with the corresponding author.

Acknowledgments: The authors would like to express their gratitude to the organizations and researchers that provided the open-source bearing fault datasets used in this study. We also sincerely thank the anonymous reviewers for their insightful comments and constructive suggestions, which have significantly improved the quality and clarity of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guo, B.; Qiao, Z.; Zhang, N.; Wang, Y.; Wu, F.; Peng, Q. Attention-based ConvNeXt with a parallel multiscale dilated convolution residual module for fault diagnosis of rotating machinery. *Expert Syst. Appl.* **2024**, *249*, 123764. [CrossRef]
2. Li, G.; Wu, J.; Deng, C.; Chen, Z. Parallel multi-fusion convolutional neural networks based fault diagnosis of rotating machinery under noisy environments. *ISA Trans.* **2022**, *128*, 545–555. [CrossRef]
3. Zhang, S.; Zhang, S.; Wang, B.; Habetler, G. Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review. *IEEE Access* **2020**, *8*, 29857–29881. [CrossRef]
4. Raj, K.K.; Kumar, S.; Kumar, R.R. Systematic review of bearing component failure: Strategies for diagnosis and prognosis in rotating machinery. *Arab. J. Sci. Eng.* **2025**, *50*, 5353–5375. [CrossRef]
5. Desnica, E.; Mikić, D.; Glavaš, H.; Palinkaš, I. Influence of diagnostics on bearing reliability on robotic systems. *Adv. Eng. Lett.* **2022**, *1*, 40–45. [CrossRef]
6. Na, K.; Kim, Y.; Yoon, H.; Youn, B.D. FASER: Fault-affected signal energy ratio for fault diagnosis of gearboxes under repetitive operating conditions. *Expert Syst. Appl.* **2024**, *238*, 122078. [CrossRef]
7. Lv, Y.; Guo, X.; Zhou, Q.; Qian, L.; Liu, J. Predictive maintenance decision-making for variable faults with non-equivalent costs of fault severities. *Adv. Eng. Inform.* **2023**, *56*, 102011. [CrossRef]
8. Lv, Y.; Zhang, X.; Cheng, Y.; Lee, C.K.M. Intelligent fault diagnosis of machinery based on hybrid deep learning with multi temporal correlation feature fusion. *Qual. Reliab. Eng. Int.* **2024**, *40*, 3517–3536. [CrossRef]
9. Hu, X.; Tang, T.; Tan, L.; Zhang, H. Fault detection for point machines: A review, challenges, and perspectives. *Actuators* **2023**, *12*, 391. [CrossRef]
10. Chao, Z.; Han, T. A novel convolutional neural network with multiscale cascade midpoint residual for fault diagnosis of rolling bearings. *Neurocomputing* **2022**, *506*, 213–227. [CrossRef]
11. Wang, X.B.; Zhang, X.; Li, Z.; Wu, J. Ensemble extreme learning machines for compound-fault diagnosis of rotating machinery. *Knowl.-Based Syst.* **2020**, *188*, 105012. [CrossRef]
12. Jiang, Y.; Xie, J. VMD–RP–CSRN based fault diagnosis method for rolling bearings. *Electronics* **2022**, *11*, 4046. [CrossRef]
13. Li, X.; Jia, B.; Liao, Z.; Wang, X. Bearing-Fault-Feature Enhancement and Diagnosis Based on Coarse-Grained Lattice Features. *Sensors* **2024**, *24*, 3540. [CrossRef]
14. Cui, W.; Meng, G.; Gou, T.; Wang, A.; Xiao, R.; Zhang, X. Intelligent rolling bearing fault diagnosis method using symmetrized dot pattern images and CBAM-DRN. *Sensors* **2022**, *22*, 9954. [CrossRef] [PubMed]
15. Lu, J.; Wang, K.; Chen, C.; Ji, W. A deep learning method for rolling bearing fault diagnosis based on attention mechanism and graham angle field. *Sensors* **2023**, *23*, 5487. [CrossRef] [PubMed]
16. Wang, X.; Wang, X.; Li, T.; Zhao, X. A fault diagnosis method based on a rainbow recursive plot and deep convolutional neural networks. *Energies* **2023**, *16*, 4357. [CrossRef]

17. 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]
18. Li, P. A Multi-scale Attention-Based Transfer Model for Cross-bearing Fault Diagnosis. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 42. [CrossRef]
19. Chen, X.; Hu, X.; Wen, T.; Cao, Y. Vibration signal-based fault diagnosis of railway point machines via double-scale CNN. *Chin. J. Electron.* **2023**, *32*, 972–981. [CrossRef]
20. Khandelwal, I.; Adhikari, R.; Verma, G. Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Procedia Comput. Sci.* **2015**, *48*, 173–179. [CrossRef]
21. 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]
22. Jian, C.; Chen, H.; Ao, Y.; Zhang, X. A two-stage learning framework for imbalanced semi-supervised domain generalization fault diagnosis under unknown operating conditions. *Adv. Eng. Inform.* **2024**, *62*, 102878. [CrossRef]
23. Hou, X.; Guo, W.; Ren, S.; Li, Y.; Si, Y.; Su, L. Bolt-loosening detection using 1D and 2D input data based on two-stream convolutional neural networks. *Materials* **2022**, *15*, 6757. [CrossRef] [PubMed]
24. Dong, H.; Lu, J.; Han, Y. Multi-Stream Convolutional Neural Networks for Rotating Machinery Fault Diagnosis under Noise and Trend Items. *Sensors* **2022**, *22*, 2720. [CrossRef]
25. Qiu, H.; Lee, J.; Lin, J.; Yu, G. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *J. Sound Vib.* **2006**, *289*, 1066–1090. [CrossRef]