machines

MDPI

Article

Adaptive Domain-Aligned Multi-Modal Feature Fusion Network for Cross-Speed Fault Diagnosis of Planetary Gearboxes

Xin Xia [1,*] and Xiaolu Wang [2]

[1] School of Mechanical and Electrical Engineering, Suqian University, Suqian 223800, China
[2] Information Construction Center, Suqian University, Suqian 223800, China; wangxiaolu@squ.edu.cn
* Correspondence: xiaxin@alumni.hust.edu.cn

Abstract

Vibration signals of planetary gearboxes under variable-speed conditions exhibit strong non-stationarity and modulation, so a single feature representation cannot comprehensively describe intricate fault patterns, and distribution discrepancies across rotating speeds degrade the cross-condition generalization of diagnostic models. To address these limitations, this paper proposes an adaptive domain-aligned multi-modal feature fusion network (ADAMFFN). Three parallel branches extract complementary features from dual-channel vibration signals: spatial coupling features from orbit images, time–frequency energy features from continuous wavelet transform (CWT) representations, and frequency-domain statistical (FreqStat) features from power and envelope spectra. Heterogeneous features are mapped into a shared latent subspace through a unified projection layer, deep cross-modal interaction is realized by a progressive fusion network, and a domain alignment mechanism based on a domain-adversarial neural network (DANN) is introduced to eliminate source–target distribution gaps via adversarial training. On eight leave-one-speed-out (LOSO) cross-speed tasks constructed on the public WT-Planetary Gearbox dataset, ADAMFFN achieves an average accuracy of 99.25%, outperforming the best single-branch and dual-branch schemes by 2.63 and 0.40 percentage points, respectively; ablation experiments verify the complementarity of the three modalities and the effectiveness of domain alignment. Cross-condition external validation on the Southeast University (SEU) gearbox dataset further demonstrates its generalization capability under a different test rig and acquisition conditions.

Keywords: planetary gearbox; fault diagnosis; multi-modal feature fusion; domain adversarial neural network; cross-speed condition; deep learning

Check for updates

Academic Editor: Christophe Simon

Received: 20 July 2026
Revised: 19 August 2026
Accepted: 20 August 2026
Published: 24 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

Planetary gearboxes possess the merits of high power density, large transmission ratio and strong load-bearing capacity, which are widely deployed in critical equipment including wind turbine generators, aero-engines and high-speed trains [1]. However, planetary gear transmissions also have inherent limitations: their structures are complex, demanding high manufacturing and assembly precision, and uneven load sharing among planet gears tends to induce dynamic imbalance [2]. Moreover, the compound motions among the sun gear, planet gears, ring gear, and carrier introduce complicated meshing modulation and transmission-path effects, making fault-induced vibration characteristics strongly speed-dependent. These characteristics render planetary gearbox fault diagnosis more challenging in terms of signal coupling than that of fixed-axis gearboxes. Operating

under complex variable-speed and variable-load conditions over long-term service, key components are prone to failures including wear, cracks and tooth breakage [3–7]. If such faults cannot be detected in a timely manner, they may lead to unplanned shutdowns and economic losses in mild cases, or trigger cascading safety accidents in severe cases. Accordingly, developing accurate and robust intelligent fault diagnosis technologies for planetary gearboxes has important theoretical significance and engineering value for ensuring the safe operation of equipment and reducing operation and maintenance costs.

Early fault diagnosis research mainly relies on signal processing technologies, which extract fault-sensitive features via time-domain statistical indicators, spectrum analysis, and time–frequency analysis tools [8–12]. Such methods achieve satisfactory performance under steady operating conditions, yet their diagnostic accuracy heavily depends on manual feature engineering and domain expertise, resulting in limited feature representation capacity when processing strongly non-stationary and heavily modulated vibration signals of planetary gearboxes [13,14]. In recent years, deep learning methods represented by convolutional neural network (CNN), long short-term memory (LSTM) and Transformers have exhibited prominent superiority in gear and bearing fault diagnosis owing to their end-to-end hierarchical feature learning capability [15–18]. Cross-condition diagnostic strategies based on transfer learning have been further developed [19–21]. Nevertheless, most existing deep diagnostic models adopt single-modality input and assume that training and test data follow identical distributions, which gives rise to two critical drawbacks in practical variable-speed scenarios. First, single feature representation shows inconsistent sensitivity to various fault types and suffers from information loss. Second, speed fluctuations induce signal distribution shift, drastically degrading model performance under unseen working conditions.

Multi-modal information fusion offers an effective solution to break through the bottleneck of single representation [22]. According to the fusion stage, fusion strategies can be categorized into feature-level fusion, decision-level fusion and hybrid fusion. Li et al. [23] performed early fusion of vibration and acoustic emission signals to improve the diagnosis accuracy of gear faults. Wang et al. [24] introduced an attention mechanism to adaptively adjust the weight of each modality and strengthen the discriminability of fused representations. It is worth noting that most existing fusion studies focus on homogeneous features collected from multi-sensor vibration signals. In contrast, heterogeneous features including images and vectors present obvious disparities in data type, dimension, scale and distribution. Direct concatenation or weighted summation fails to achieve effective information complementarity, and may even degrade diagnostic performance due to cross-modality inconsistency [25]. Therefore, the deep fusion mechanism for such heterogeneous features remains to be further investigated.

Transfer learning and domain adaptation methods have attracted extensive attention to tackle the issue of cross-condition distribution shift. A domain-adversarial neural network is proposed, which achieves adversarial alignment of feature distributions between source and target domains via a gradient reversal layer [26]. Tzeng et al. [27] further developed an adversarial discriminative domain adaptation framework. Han et al. [28] incorporated DANN into cross-load fault diagnosis of bearings, which effectively narrows the distribution discrepancy induced by varying working conditions. Nevertheless, most existing domain adaptation studies are designed for single-modality features. How to embed a domain alignment mechanism into a multi-modal fusion framework such that heterogeneous fused features simultaneously possess discriminability and domain invariance remains an open problem that has not been fully addressed. In recent years, novel deep transfer learning architectures such as Transformers, contrastive and self-supervised learning, graph neural networks, and diffusion models have developed rapidly in fault

diagnosis [17,29]. It should be clarified that the focus of this work is not to propose a new network backbone, but to address the deep fusion and domain alignment of heterogeneous multi-modal features under cross-speed conditions. These novel backbones are structurally orthogonal to the proposed framework and can be integrated as alternative encoders for the individual branches in future work.

In summary, there are three major challenges in cross-speed fault diagnosis of planetary gearboxes: (1) single-type features fail to fully capture complex fault information; (2) effective deep fusion mechanisms for heterogeneous multi-modal features are lacking; (3) distribution shifts caused by speed variations degrade the generalization performance of diagnostic models. To address the above challenges, this paper proposes an adaptive domain-aligned multi-modal feature fusion network (ADAMFFN). The main contributions are summarized as follows:

(1) A tri-modal complementary feature extraction framework is proposed. Three parallel branches corresponding to shaft orbit (Orbit), continuous wavelet transform (CWT) time–frequency images and frequency-domain statistics (FreqStat) are constructed to characterize the operating conditions of planetary gearboxes from three complementary perspectives, including spatial coupling, time–frequency energy and spectral statistics, achieving multi-dimensional coverage of fault information.

(2) A unified projection and progressive fusion mechanism for heterogeneous features is designed. The unified projection layer maps heterogeneous features with disparate dimensions and distributions into a shared subspace, and a three-layer fusion network equipped with progressive dropout regularization is adopted to learn deep cross-modal interactions, realizing collaborative enhancement of complementary features.

(3) The DANN domain alignment mechanism is embedded into the multi-modal fusion framework. Adversarial training between source and target domains is performed on fused features via the gradient reversal layer, enabling the fused representations to simultaneously possess discriminability and domain invariance and improving the generalization capacity of the model under unseen speed conditions.

(4) A systematic cross-speed verification framework is established. Eight leave-one-speed-out (LOSO) cross-condition diagnostic tasks are designed based on the WT-Planetary Gearbox dataset, and systematic comparisons with six single-branch and dual-branch methods are conducted. Experimental results demonstrate that the proposed ADAMFFN achieves an average diagnostic accuracy of 99.25%. Ablation experiments further verify the complementarity of each modality and the effectiveness of the domain alignment mechanism. Moreover, cross-condition external validation on the SEU gearbox dataset further demonstrates that the generalization capability of the proposed method is not restricted to a single test rig.

It should be clarified that the individual components (Orbit, CWT, FreqStat, and DANN) are not new in themselves. The novelty lies in their problem-oriented integration into a unified framework: heterogeneous features are made compatible through unified projection, fused progressively rather than simply concatenated, and domain-aligned on the fused representation, jointly addressing heterogeneous feature incompatibility and speed-induced domain shift in cross-speed planetary gearbox diagnosis. The ADAMFFN is built upon the closed-set assumption [30]; the recognition of unknown faults is beyond the scope of this work and is discussed in the conclusion section.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical fundamentals of shaft orbit analysis, continuous wavelet transform and frequency-domain statistical features. Section 3 elaborates on the overall architecture and module design of ADAMFFN in detail. Section 4 presents the dataset, experimental setup and result analysis. Section 5 concludes this paper and outlines future research directions.

2. Theoretical Background

2.1. Orbit Diagram Analysis

The orbit diagram serves as a vital visual analysis tool for fault diagnosis of rotating machinery. By plotting the motion trajectory of the rotor center within the plane perpendicular to the rotating shaft, it intuitively reflects the spatial motion behavior of the rotor and globally reveals the dynamic characteristics of rotary equipment, thus forming an important type of diagnostic feature [31].

Assume that two vibration sensors are mounted along mutually perpendicular directions on the rotating machinery, and the collected vibration signals are denoted as $x(t)$ and $y(t)$, respectively. The orbit diagram is generated by plotting a parametric curve on a two-dimensional plane with $x(t)$ as the horizontal axis and $y(t)$ as the vertical axis, and its mathematical formulation is given:

$$r(t) = [x(t), y(t)]^T, t \in [0, T], \tag{1}$$

where $r(t)$ represents the position vector of the rotor center at time t, and T denotes the total sampling duration.

Distinct geometric patterns can be observed on orbit diagrams under different fault conditions. Under healthy operating conditions, the orbit diagram appears as a smooth approximate circle or ellipse [31]. For missing-tooth and broken-tooth faults, the trajectory exhibits petal shapes, local sharp corners and double-layer nested structures. In the case of gear tooth wear, the orbit presents an ellipse covered with uniform ripples. Such differences in geometric morphology provide a physical basis for fault classification based on orbit diagram images [32].

Nevertheless, planet gears inside planetary gearboxes rotate about their own axes while revolving around the sun gear, introducing prominent modulation effects in vibration signals. The corresponding orbit diagrams generally contain folds, nested contours and periodic distortions induced by sideband modulation [33]. Effective fault information cannot be accurately extracted merely through manual observation or handcrafted feature design. Therefore, deep learning is adopted in this paper to mine latent fault information contained in orbit diagrams.

2.2. Continuous Wavelet Transform

The continuous wavelet transform (CWT) provides an effective framework for characterizing non-stationary signals jointly in the time and frequency domains. By introducing wavelet basis functions with variable scales, it enables multi-resolution analysis of signals. CWT possesses adaptive time–frequency resolution: it achieves high time resolution for high-frequency components and high frequency resolution for low-frequency components, which makes it highly suitable for analyzing non-stationary mechanical vibration signals [34].

The continuous wavelet transform of a signal $x(t)$ is defined as:

$$w(a, b) = f(x(t), \psi_{a,b}(t)) = \frac{1}{\sqrt{a}} \int x(t) \psi^*(\frac{t-b}{a}) dt, \tag{2}$$

where a denotes the scale factor, b denotes the translation factor, $\psi_{a,b}(t)$ represents the mother wavelet function, and ψ^* is the complex conjugate of the mother wavelet function.

The Morlet wavelet, widely adopted in the field of mechanical fault diagnosis, is formulated as:

$$\psi(t) = \pi^{-\frac{1}{4}} e^{j\omega_0 t} e^{-\frac{t^2}{2}}, \tag{3}$$

where ω_0 stands for the central angular frequency.

In this paper, the Morlet wavelet is selected as the basis function to conduct CWT processing on dual-channel vibration signals, and deep learning is utilized to extract fault features from the resulting time–frequency images.

2.3. Frequency-Domain Statistical Features

Frequency-domain analysis is a classical technique for mechanical fault diagnosis. By analyzing the frequency components and their statistical properties of signals, periodic faults such as gear meshing faults can be effectively identified [35]. Two frequency-domain analysis approaches are mainly adopted in this work, namely power spectral density analysis and envelope spectrum analysis.

Power spectral density (PSD) characterizes the distribution of signal power over the frequency domain. The Welch algorithm is employed to estimate the PSD, which can efficiently reduce the variance of spectral estimation. The Welch method divides the signal $x(t)$ into K segments, with each segment containing M sampling points, applies windowing to each segment, and averages the periodograms of all segments. Its estimation formula is given as follows:

$$P_{xx}(f) = \frac{1}{KMU}\sum_{k=1}^{K}|X_k(f)|^2, \tag{4}$$

where U denotes the energy normalization factor of the window function, and $X_k(f)$ represents the Fourier transform of the windowed signal of the k-th segment.

Envelope spectrum analysis is an effective method for extracting modulation features contained in signals, which is especially suitable for fault diagnosis of rotating machinery. The analytic signal is obtained via the Hilbert transform of the original signal $x(t)$, as expressed below:

$$z(t) = x(t) + jH\{x(t)\} = x(t) + \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d\tau, \tag{5}$$

where $H\{\cdot\}$ denotes the Hilbert transform operator.

The modulus of the analytic signal yields the envelope signal:

$$e(t) = |z(t)| = \sqrt{x^2(t) + H\{x(t)\}^2}, \tag{6}$$

The envelope spectrum is obtained by performing the Fourier transform on the envelope signal $e(t)$.

In this paper, the logarithmic power spectral density features and logarithmic envelope spectrum features are taken as the feature vectors for fault diagnosis.

3. Proposed Method

3.1. Overall Framework

This paper proposes an adaptive domain-aligned multi-modal feature fusion network (ADAMFFN) for planetary gearbox fault diagnosis, and its overall architecture is illustrated in Figure 1.

The entire processing pipeline of ADAMFFN consists of three sequentially connected modules:

(1) Multi-modal feature extraction module: The dual-channel signals are processed by three parallel branches, i.e., shaft orbit (Orbit) imaging, continuous wavelet transform (CWT) time–frequency imaging, and frequency-domain statistics (FreqStat) feature calculation. Each branch extracts deep features via its dedicated branch network.

(2) Unified projection and progressive fusion module: The three groups of features with disparate dimensions and distributions are projected into a 128-dimensional shared

subspace. After concatenation into a 384-dimensional fused vector, the vector is fed into a three-layer progressive fusion network to output 128-dimensional fused features.

(3) Domain-adversarial alignment module: A label predictor and a domain classifier equipped with a gradient reversal layer (GRL) are arranged in parallel after the fused features. Adversarial training is conducted to endow the fused representations with both fault discriminability and cross-speed domain invariance.

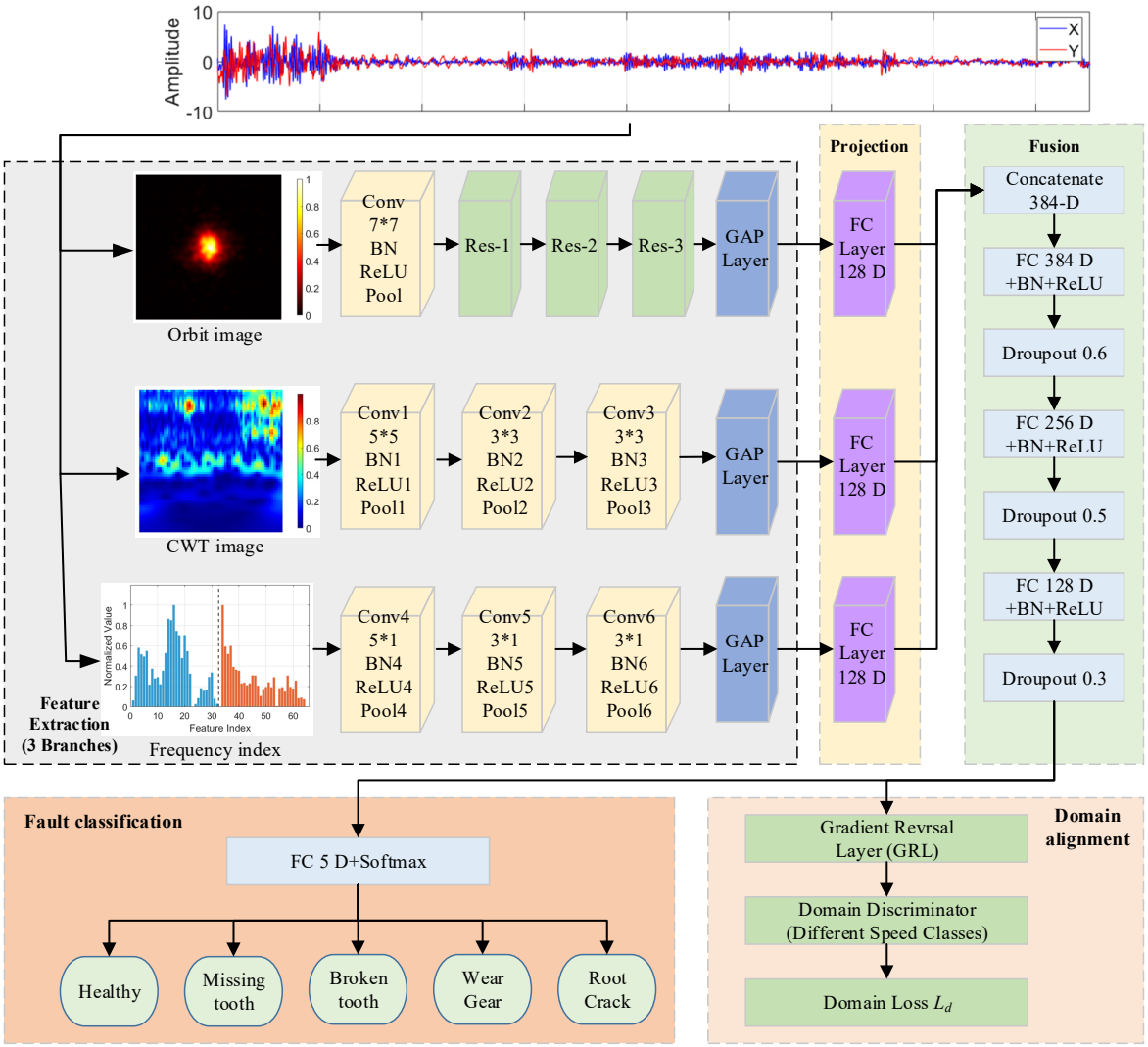


Figure 1. Framework of the proposed ADAMFFN.

3.2. Tri-Modal Feature Extraction

3.2.1. Orbit Image Generation

The Orbit branch is designed to extract spatial coupling features from dual-channel vibration signals. Its detailed processing procedures are presented as follows:

Step 1: Data normalization: Min–max normalization is performed on dual-channel vibration signals to scale signal amplitudes into the range of [0, 1], eliminating the influence of dimensional differences on trajectory shapes.

Step 2: Two-dimensional histogram construction: The normalized signals are mapped onto a 128 × 128 two-dimensional grid, and the data point frequency within each grid cell is counted to construct a 2D histogram $H \in R^{(128 \times 128)}$. To improve the separability of features, the single-channel histogram is replicated into three channels to generate a red–green–blue (RGB) image with the size of 128 × 128 × 3.

Step 3: Gaussian smoothing: A 3 × 3 Gaussian filter is applied to the two-dimensional histogram to smooth noisy points while preserving the overall morphological features of trajectories. The processed shaft orbit diagrams corresponding to various fault conditions are shown in Figure 2.

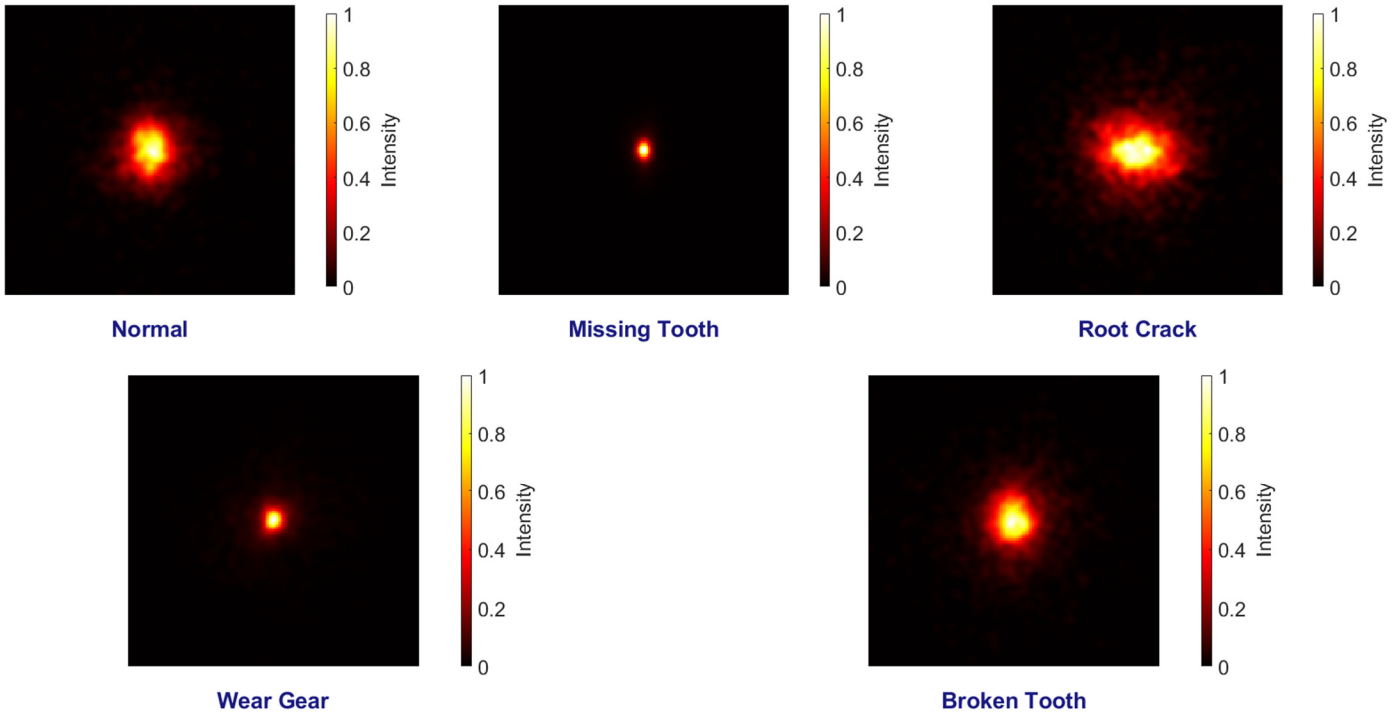


Figure 2. Shaft orbit diagrams under different operating conditions.

Step 4: Feature extraction network: A ResNet-like architecture is adopted to process orbit images. The network starts with a 7 × 7 convolutional layer with 64 filters and a stride of 2. After batch normalization, rectified linear unit (ReLU) activation and 3 × 3 max pooling with stride 2, the spatial resolution is reduced from 128 × 128 to 32 × 32. Three residual blocks are stacked subsequently: the first residual block retains 64 filters and utilizes an identity shortcut; the second and third residual blocks double the number of filters to 128 and 256 in sequence, and adopt 1 × 1 projection shortcuts to match dimension variations. Finally, global average pooling is implemented to output a 256-dimensional feature vector.

3.2.2. CWT Time–Frequency Image Generation

The CWT branch aims to extract time–frequency energy features from non-stationary vibration signals, and its detailed processing steps are given as follows:

Step 1: Continuous wavelet transform: The Morlet wavelet is selected as the basis function to perform CWT on dual-channel vibration signals, and the frequency range is set to cover the major frequency bands corresponding to characteristic fault frequencies of planetary gearboxes. The output of CWT is a complex matrix of time–frequency coefficients.

Step 2: Magnitude calculation and logarithmic compression: The magnitudes of CWT coefficients are calculated to derive the time–frequency amplitude representation. Logarithmic compression is then applied to boost the visibility of low-energy fine details while compressing the dynamic range of high-energy regions.

Step 3: Resizing: The logarithmically compressed time–frequency representation is resized to a standard dimension of 64 × 64 × 1 for subsequent processing by the CNN. The

processed CWT time–frequency maps of different operating conditions are illustrated in Figure 3.

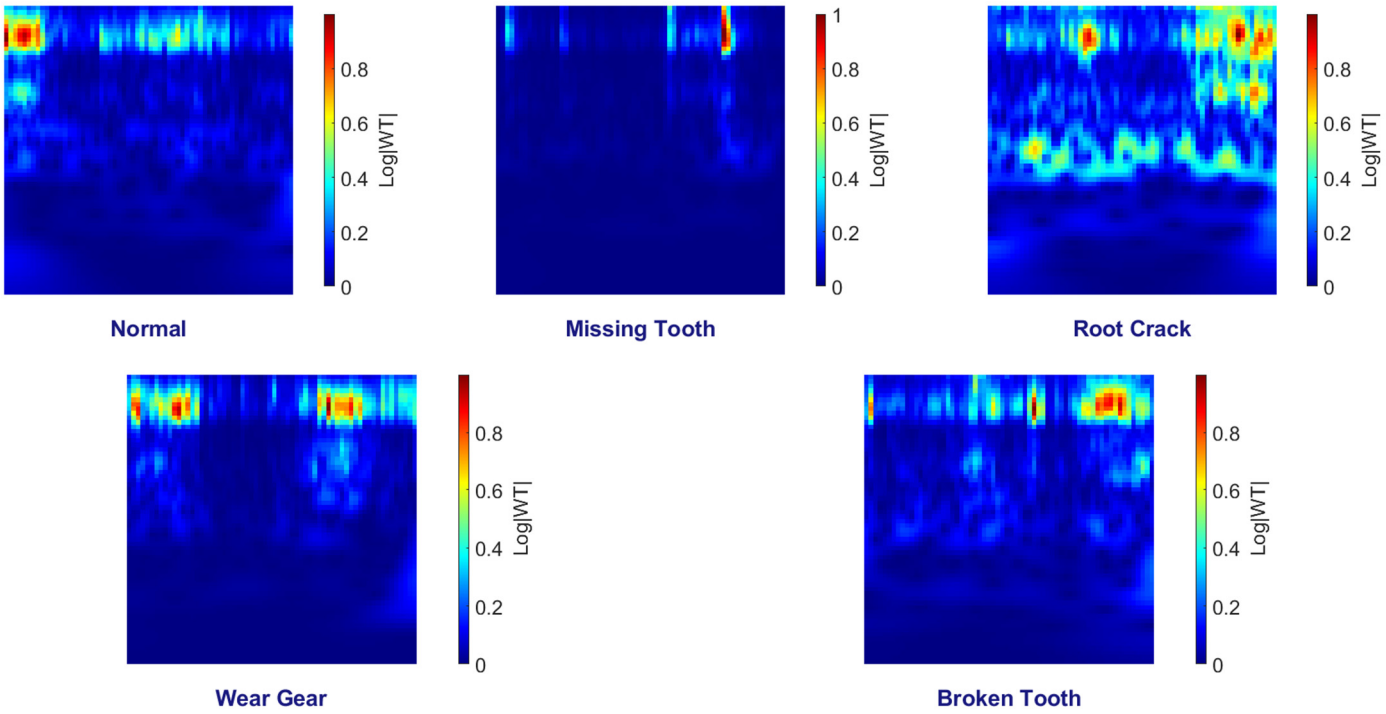


Figure 3. CWT time–frequency images of different operating conditions.

Step 4: Feature extraction network. A lightweight CNN architecture is designed to process CWT time–frequency images. The network consists of three convolutional blocks, each composed of a convolutional layer, batch normalization and ReLU activation, followed by 2 × 2 max pooling after the first two blocks. The first convolutional block adopts a large 5 × 5 kernel with 32 filters to capture coarse-grained time–frequency patterns. The subsequent two blocks utilize 3 × 3 kernels, with the number of filters increased to 64 and 128 sequentially for extracting finer time–frequency features. The two pooling layers reduce the spatial size from 64 × 64 to 16 × 16. Considering the translation invariance of time–frequency representations, residual connections are not deployed in this branch. Finally, global average pooling is performed to output a 128-dimensional feature vector.

3.2.3. Frequency-Domain Statistical Feature Vector

The FreqStat branch is designed to extract frequency-domain statistical regularities and characterize fault features from the frequency perspective. Its specific processing procedures are detailed as follows:

Step 1: Welch power spectrum estimation: The Welch power spectral density estimation is implemented on vibration signals with a 512-point Hamming window and 50% overlap. The power spectrum covering the main characteristic frequency range is calculated. After logarithmic compression, 32 points are uniformly sampled to form a 32-dimensional logarithmic power spectral feature.

Step 2: Envelope spectrum analysis: The Hilbert transform is applied to vibration signals to extract the amplitude envelope. The same Welch spectral estimation is carried out on the envelope signal to compute the envelope spectrum within the main characteristic frequency band. After logarithmic compression, 32 uniformly sampled points generate a 32-dimensional logarithmic envelope spectrum feature.

Step 3: Feature concatenation: The 32-dimensional power spectrum feature and 32-dimensional envelope spectrum feature are concatenated into a frequency-domain statistical feature vector with a dimension of $64 \times 1 \times 1$. The concatenated feature vector is fed into the subsequent network as illustrated in Figure 4.

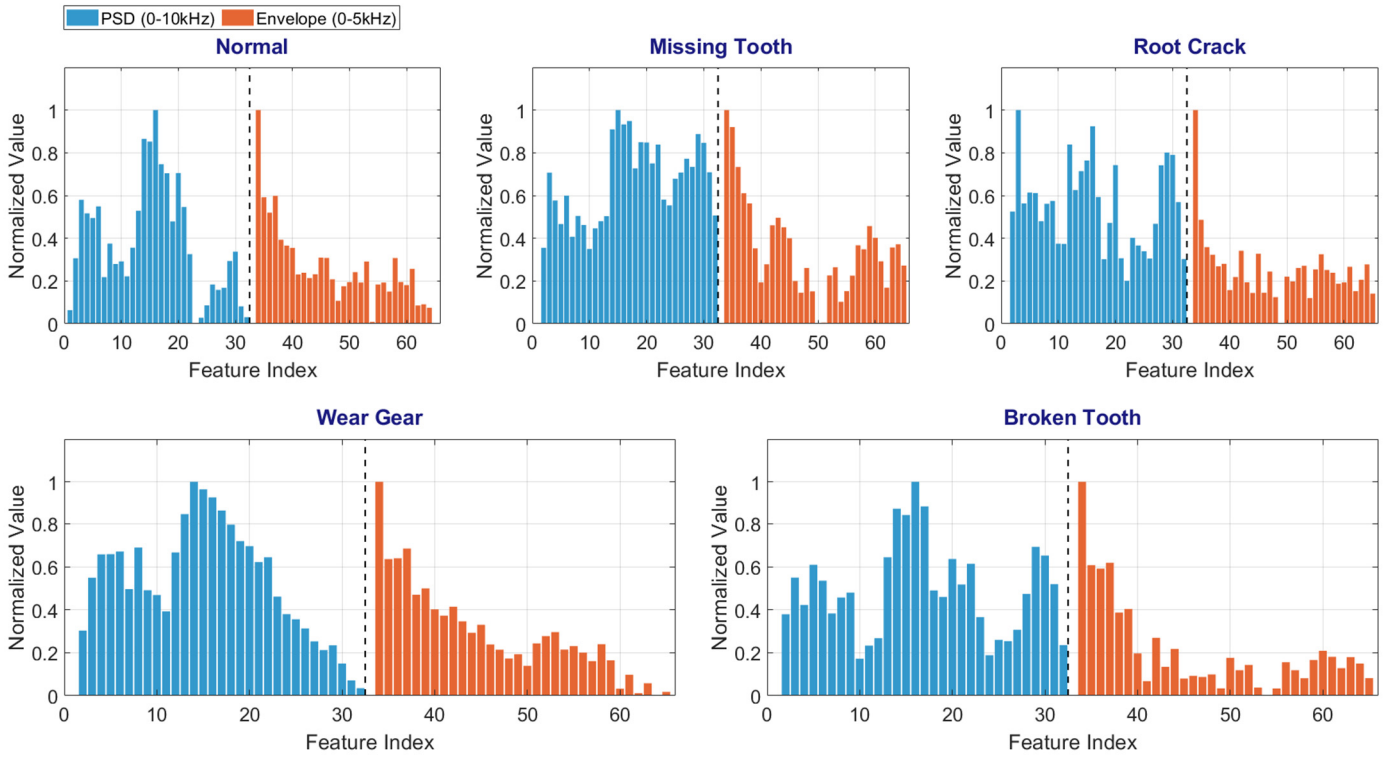


Figure 4. Frequency-domain statistical features under various conditions.

Step 4: Feature extraction network: A 1D-CNN architecture is adopted to process the frequency-domain statistical feature vectors. The network consists of three 1D convolutional blocks with kernel sizes of 5×1, 3×1 and 3×1 respectively, which only operate along the frequency axis to capture local spectral patterns. Each convolutional block includes convolution, batch normalization, ReLU activation and max pooling. The resolution along the frequency axis is gradually reduced from 64 to 32 and then to 16. The number of filters increases sequentially from 32 to 64 and 128 to realize hierarchical abstraction of frequency-domain information. Finally, global average pooling outputs a 128-dimensional feature vector.

3.3. Feature Projection and Fusion

There are discrepancies in dimension and data distribution among the features output by the three branches. A projection layer is introduced to standardize these features, where fully connected layers project the feature vectors of each branch into a unified dimension. After projection processing, each branch outputs a 128-dimensional feature vector, which unifies the dimensions of heterogeneous features.

After unified projection, the three 128-dimensional feature vectors are concatenated along the channel dimension to form a 384-dimensional fused vector. This fused vector simultaneously integrates spatial trajectory information (Orbit), time–frequency energy distribution information (CWT), and frequency-domain statistical regularity information (FreqStat), achieving comprehensive coverage of fault information from multiple perspectives. The fused vector is then fed into the progressive fusion network, which comprises three fully connected layers with the dimension transition structure of $384 \rightarrow 256 \rightarrow 128$.

Each fully connected layer is followed sequentially by batch normalization, ReLU activation and dropout regularization. Notably, the dropout ratio is designed in a progressive manner: 0.6 for the first layer, 0.5 for the second layer, and 0.3 for the third layer. This dropout strategy, which decreases from shallow layers to deep layers, possesses the following merits: shallow layers carry high-dimensional fused features, so strong regularization helps suppress overfitting; deep layers contain fully abstracted features, and moderately reduced regularization intensity facilitates the retention of subtle discriminative information.

The fusion network finally outputs a 128-dimensional fused feature vector. Through multi-layer nonlinear transformations, deep interaction and complementary enhancement of three modal features are realized, providing high-quality representations for subsequent fault classification and domain alignment.

To facilitate understanding of the branch architectures, Table 1 summarizes the input dimensions, main structures, output dimensions, and trainable parameter counts of all modules in ADAMFFN.

Table 1. Summary of architectures of ADAMFFN modules.

Module	Input Dimension	Main Structure	Output Dimension
Orbit branch	128 × 128 × 3	7 × 7 Conv (64) + three residual blocks (64/128/256) + GAP	256
CWT branch	64 × 64 × 1	Conv blocks 5 × 5/32, 3 × 3/64, 3 × 3/128 + GAP	128
FreqStat branch	64 × 1 × 1	1D-Conv blocks 5 × 1/32, 3 × 1/64, 3 × 1/128 + GAP	128
Unified projection	256/128/128	Fully connected layers	128 × 3
Fusion network	384	FC 384→256→128, BN + ReLU + Dropout (0.6/0.5/0.3)	128
Label classifier	128	Fully connected layer	5
Domain classifier (training only)	128	Fully connected layers	2

3.4. DANN-Based Domain Alignment Mechanism

In variable-speed fault diagnosis scenarios, the distribution of vibration signals varies significantly under different rotating speeds. Such distribution shift leads to performance degradation of the model trained on source domains (training speeds) when deployed on target domains (testing speeds). To tackle this issue, a domain alignment mechanism based on the domain-adversarial neural network (DANN) is embedded after the fusion network in this work.

The core idea of DANN is to learn domain-invariant features via adversarial training [26], and its structure is illustrated in Figure 5. Two branches are arranged in parallel after the feature extractor $G_f(.,.,\theta_f)$. The label predictor $G_l(.,.,\theta_l)$ is used to predict fault categories from fused features by minimizing the classification loss. The domain classifier $G_d(.,.,\theta_d)$ minimizes the domain classification loss, whereas the feature extractor maximizes it through the gradient reversal layer GRL. The total loss function of DANN is defined as:

$$L_{DANN}(\theta_f, \theta_l, \theta_d) = L_l(\theta_f, \theta_l) - \lambda L_d(\theta_f, \theta_d), \tag{7}$$

where L_l denotes the cross-entropy loss for label prediction, L_d denotes the binary cross-entropy loss for domain classification, and λ represents the domain adversarial weight.

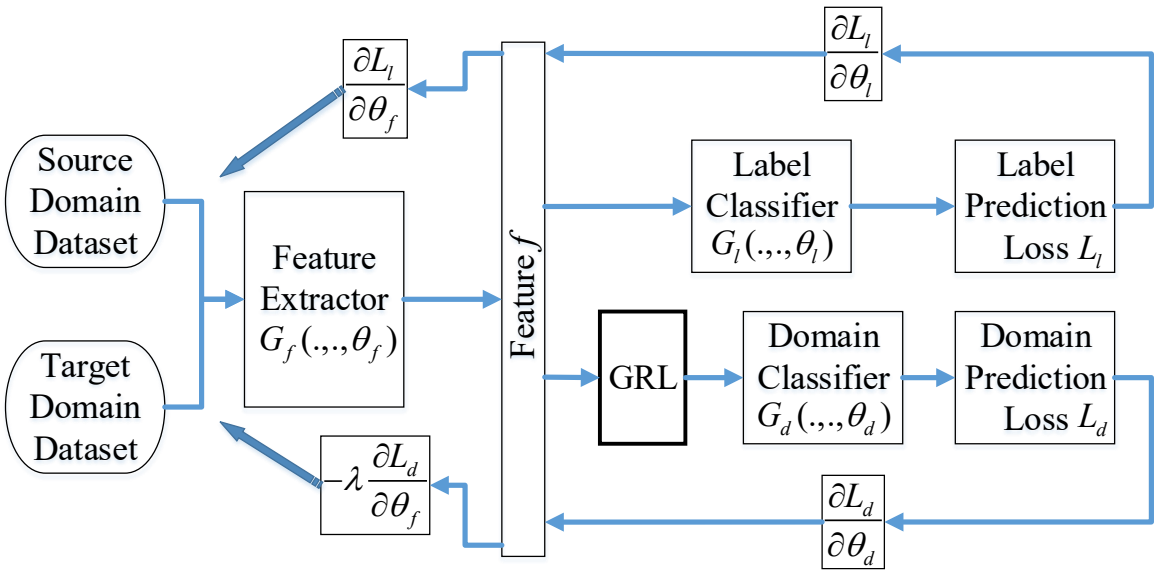


Figure 5. Schematic diagram of the domain alignment mechanism.

The saddle-point solution for optimizing the objective function is expressed as:

$$\begin{cases} \left(\hat{\theta}_f, \hat{\theta}_l\right) = \arg\min_{\theta_f,\theta_l} L_{DANN}\left(\theta_f, \theta_l, \hat{\theta}_d\right) \\ \hat{\theta}_d = \arg\max_{\theta_d} L_{DANN}\left(\hat{\theta}_f, \hat{\theta}_l, \theta_d\right) \end{cases}, \tag{8}$$

During adversarial training, the domain classifier aims to boost its capability to distinguish the domain origin of features. In contrast, the feature extractor receives reversed gradients through the GRL and learns to generate domain-invariant features that confuse the domain classifier. The two modules converge to a saddle-point equilibrium through adversarial competition, enabling the fused features to retain fault discriminability while possessing cross-speed invariance.

The detailed training procedure is as follows: in each iteration, samples from the source domain and target domain are fed into the network simultaneously. The label prediction loss is calculated solely using labeled source-domain samples, whereas the domain classification loss is computed based on the domain labels of all samples. At the test stage, the domain classifier is discarded, and only the feature extractor and label predictor are retained to implement fault diagnosis for target-domain samples.

4. Experimental Study

4.1. Experimental Dataset

To verify the effectiveness of the proposed method, the WT-Planetary Gearbox dataset released by Liu et al. [36] is adopted in this work. The dataset is available at the following link: https://github.com/Liudd-BJUT/WT-planetary-gearbox-dataset (accessed on 5 November 2024). This public dataset is widely utilized in the field of planetary gearbox fault diagnosis. The dataset is collected on a planetary gearbox fault simulation test rig, which consists of a driving motor, a planetary gearbox, a magnetic powder brake and a data acquisition system. Two mutually perpendicular accelerometers are mounted on the gearbox bearing pedestal to collect vibration signals in horizontal and vertical directions, with a sampling frequency of 48 kHz.

The dataset covers five operating conditions: (1) normal condition (N); (2) wear gear (W); (3) root crack (R); (4) missing tooth (M); (5) broken tooth (B). Data for each health

condition are collected at eight rotational speeds, namely 20 Hz, 25 Hz, 30 Hz, 35 Hz, 40 Hz, 45 Hz, 50 Hz and 55 Hz, covering the typical operating speed range of planetary gearboxes.

The data preprocessing pipeline is described as follows. Each sample consists of dual-channel vibration signals with 4096 sampling points (approximately 0.085 s), and there is no overlap between adjacent samples. After preprocessing, around 2500 samples can be obtained for each health condition under each rotational speed. To control the experimental scale and guarantee class balance, 40 samples are randomly selected for each health condition at each speed, yielding a total of 1600 samples (5 health conditions × 8 speeds × 40 samples) for subsequent experiments. The collected dual-channel vibration signal samples are presented in Figure 6.

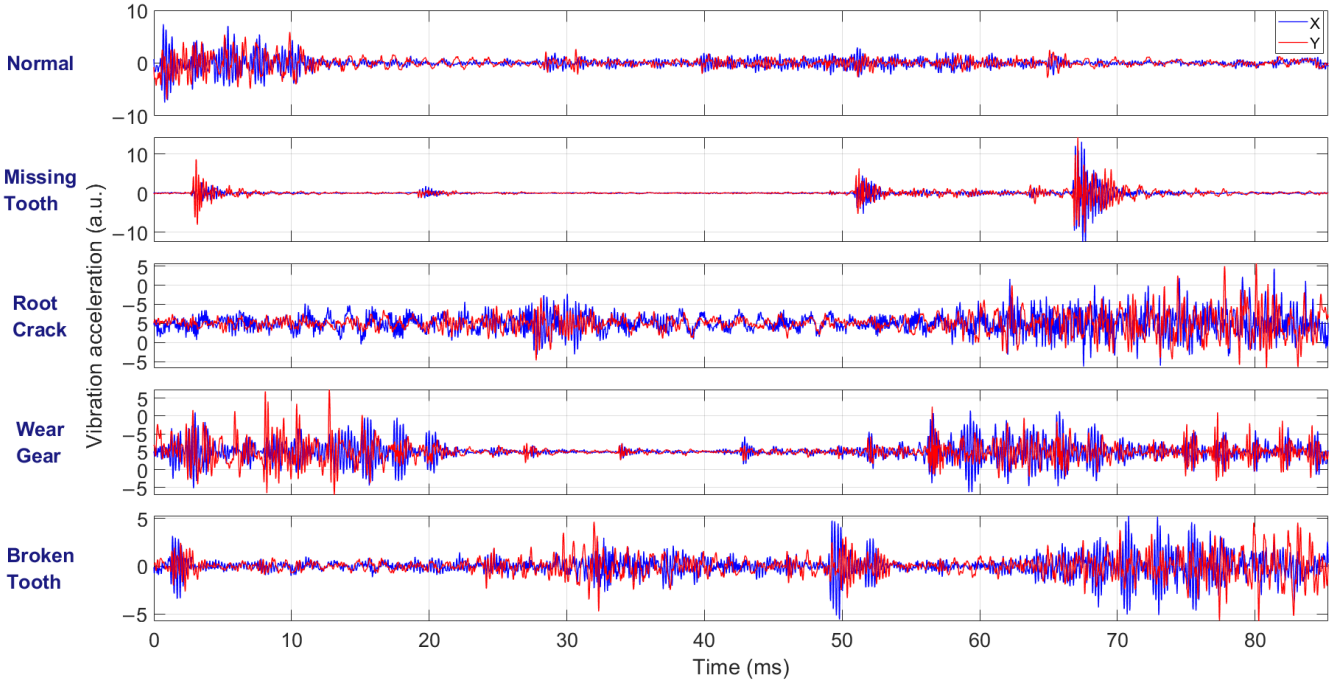


Figure 6. Dual-channel vibration signals of five operating conditions.

In addition to the WT-Planetary Gearbox dataset, the Southeast University gearbox dataset is further employed for external validation. The dataset description, task settings, and result analysis are presented in Section 4.6.

4.2. Cross-Speed Evaluation Protocol

To systematically evaluate the generalization capability of the proposed method under various rotating speed conditions, eight leave-one-speed-out (LOSO) cross-speed fault diagnosis tasks are designed. For each task, one rotational speed is treated as the target domain (test set), while the remaining seven speeds serve as the source domain (training set + validation set). The detailed partition of the eight tasks is listed in Table 2.

In each task, the source-domain data are split into a training set and a validation set at a ratio of 3:1. The training set is utilized for model parameter optimization, and the validation set is used for hyperparameter tuning and early stopping strategy. All experiments are independently repeated 10 times, and the mean accuracy and standard deviation are reported as the final evaluation metric to guarantee the statistical reliability of results.

The Adam optimizer is adopted for network training with an initial learning rate of 0.001, a batch size of 64, and a maximum training epoch of 100. The early stopping strategy is deployed, where training terminates if the validation loss fails to decline for

20 consecutive epochs. The domain adversarial weight λ is set to 0.5. All the above hyperparameters were determined based on source-domain validation performance, and the target-domain test data were not involved in tuning. Specifically, the learning rate of 0.001 is a common default for Adam, the batch size of 64 balances gradient stability against memory cost, and the projection dimension of 128 balances feature capacity against parameter count; the rationale for the progressive dropout is given in Section 3. An excessively small λ leads to insufficient domain alignment, whereas an excessively large λ weakens class discriminability; λ = 0.5 is adopted as the balance between the two.

Table 2. Partition of cross-speed transfer tasks.

Task	Training Set + Validation Set	Test Set
1	25 Hz, 30 Hz, 35 Hz, 40 Hz, 45 Hz, 50 Hz, 55 Hz	20 Hz
2	20 Hz, 30 Hz, 35 Hz, 40 Hz, 45 Hz, 50 Hz, 55 Hz	25 Hz
3	20 Hz, 25 Hz, 35 Hz, 40 Hz, 45 Hz, 50 Hz, 55 Hz	30 Hz
4	20 Hz, 25 Hz, 30 Hz, 40 Hz, 45 Hz, 50 Hz, 55 Hz	35 Hz
5	20 Hz, 25 Hz, 30 Hz, 35 Hz, 45 Hz, 50 Hz, 55 Hz	40 Hz
6	20 Hz, 25 Hz, 30 Hz, 35 Hz, 40 Hz, 50 Hz, 55 Hz	45 Hz
7	20 Hz, 25 Hz, 30 Hz, 35 Hz, 40 Hz, 45 Hz, 55 Hz	50 Hz
8	20 Hz, 25 Hz, 30 Hz, 35 Hz, 40 Hz, 45 Hz, 50 Hz	55 Hz

The software platforms adopted in this paper are MATLAB R2024b, and all calculations are accelerated by a GPU. The main hardware configuration includes a GeForce RTX 3070 graphics card with 8 GB GDDR6 video memory and 16 GB system memory.

4.3. Comparative Analysis

To verify the effectiveness of the proposed ADAMFFN for planetary gearbox fault diagnosis, systematic ablation experiments are conducted to analyze the contributions of multi-modal feature fusion and domain adversarial alignment. Given the large number of cross-speed tasks, Task 2 is selected as the representative case for analysis in this subsection.

4.3.1. Comparison of Different Feature Extraction Schemes

To validate the superiority of multi-modal fusion, six baseline schemes and the proposed ADAMFFN are devised as follows: (1) Orbit-ResNet: only the shaft orbit branch is utilized; (2) CWT-CNN: only the CWT time–frequency branch is utilized; (3) FreqStat-1D-CNN: only the frequency-domain statistics branch is utilized; (4) Orbit+CWT: dual-branch fusion of orbit and CWT features; (5) Orbit+FreqStat: dual-branch fusion of orbit and frequency statistical features; (6) CWT+FreqStat: dual-branch fusion of CWT and frequency statistical features without the orbit branch; (7) ADAMFFN (the proposed method): complete fusion of all three branches.

Each method is independently trained 10 times with random data sampling. The average classification accuracy of all schemes is illustrated in Figure 7.

As shown in Figure 7, among the three single-branch methods, Orbit-ResNet only achieves an average accuracy of 74.15% with the largest standard deviation of 1.62%. In contrast, CWT-CNN and FreqStat-1D-CNN reach 97.15% and 97.50% respectively, which demonstrates that time–frequency features and frequency-domain statistical features possess significantly stronger robustness against rotating speed variations than shaft orbit features. This phenomenon can be attributed to two factors. First, each sample is a fixed 4096-point window (approximately 0.085 s), covering only about 1.71 revolutions at 20 Hz but about 4.69 revolutions at 55 Hz; the trajectory morphology and density therefore vary inherently with speed, making Orbit images strongly speed-dependent. Second, the min–max normalization used in Orbit construction weakens amplitude information related to

fault-induced impact energy. By contrast, the time–frequency energy distributions and spectral structures extracted by CWT and FreqStat are relatively insensitive to speed variations, resulting in more stable cross-speed performance.

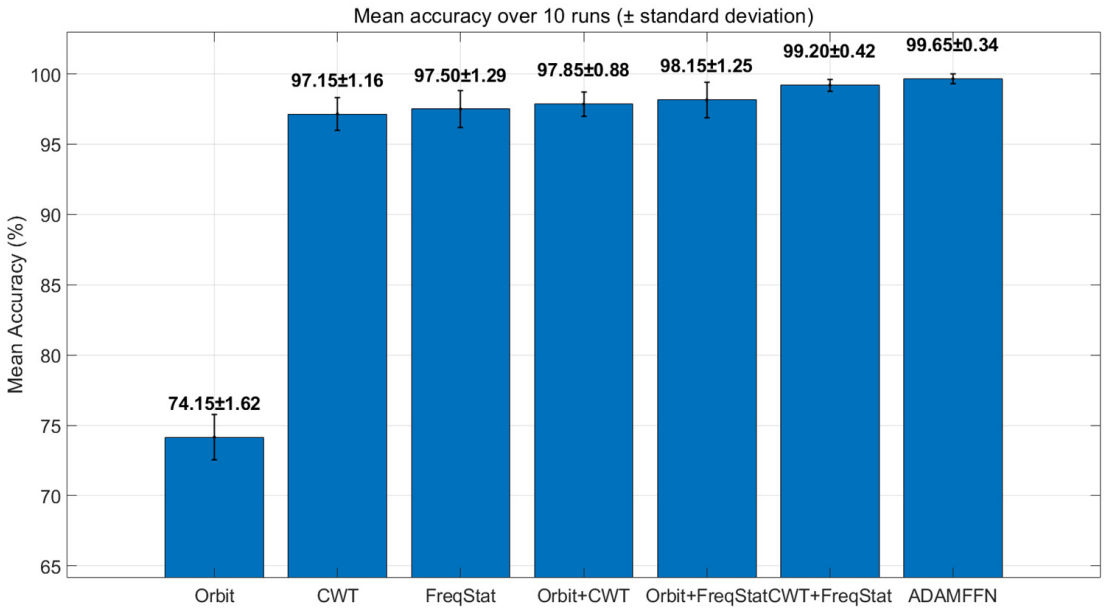
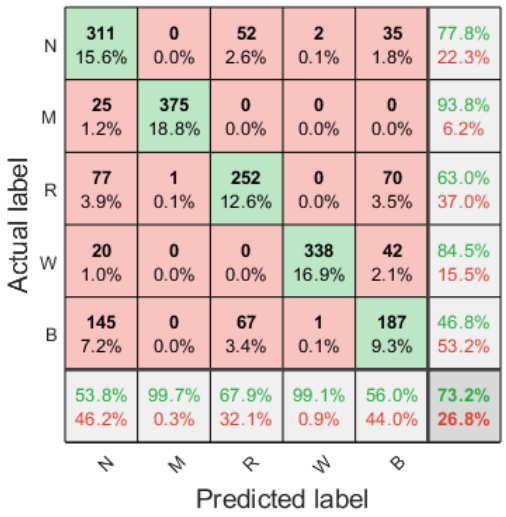


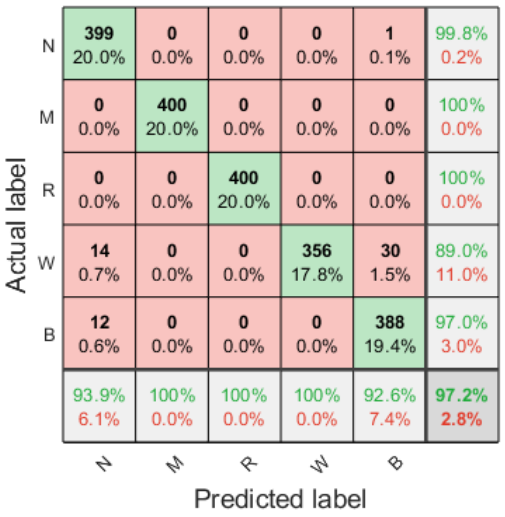
Figure 7. Bar chart of mean diagnostic accuracy for different feature extraction schemes on Task 2.

All dual-branch fusion methods yield higher average accuracy ranging from 97.85% to 99.20% than any single-branch counterpart, accompanied by an overall decline in standard deviation. This reveals that multi-modal fusion can boost diagnostic accuracy and simultaneously improve prediction stability. The proposed ADAMFFN with full three-branch fusion achieves the highest average accuracy of 99.65%, outperforming the best dual-branch scheme (CWT+FreqStat, 99.20%) by 0.45 percentage points, while its standard deviation further drops to 0.34%. The results verify that the three types of feature information are mutually complementary, and full fusion can synchronously promote diagnostic accuracy and stability.

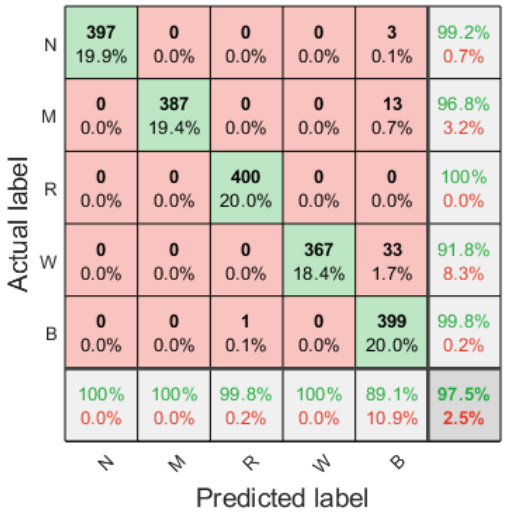
The confusion matrices summarizing the 10 repeated diagnostic results of all comparative methods are displayed in Figure 8. It can be observed from Figure 8 that misclassifications of the single Orbit branch mainly occur among normal (N), root crack (R) and broken-tooth (B), where the recall rate of class B is merely 46.8%. This indicates that shaft orbit features suffer severe distortion under varying rotating speeds and exhibit weak inter-class separability. For the single CWT and FreqStat branches, misclassifications are primarily concentrated between wear gear (W) and broken-tooth (B). As multi-modal fusion is introduced, off-diagonal elements in confusion matrices gradually decrease: dual-branch fusion remarkably alleviates mutual confusion among N, R and B. After full three-branch fusion, only seven misclassified samples exist among all 2000 test samples. This indicates that the three modalities characterize fault information from distinct perspectives with prominent complementarity, and full fusion can eliminate inter-class confusion to the greatest extent.



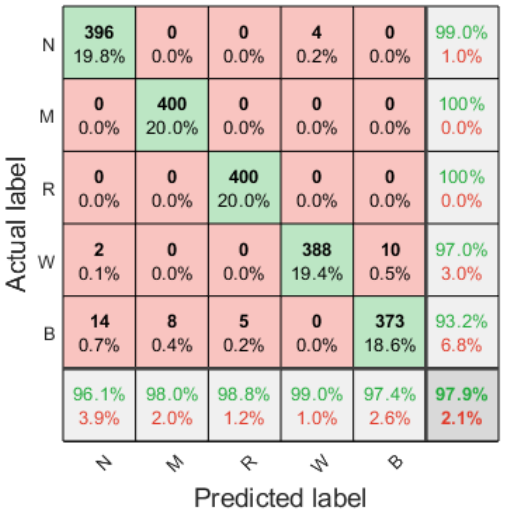
(**a**)



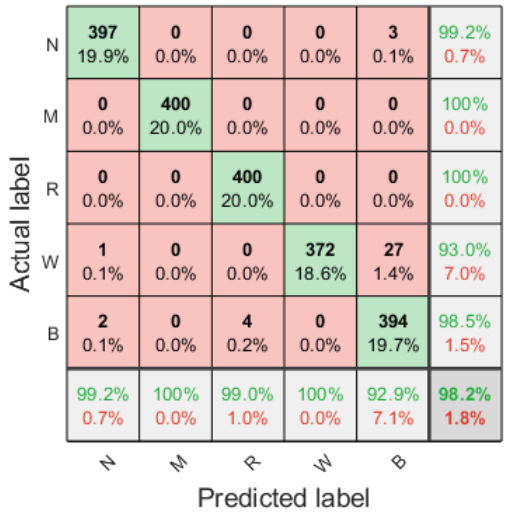
(**b**)



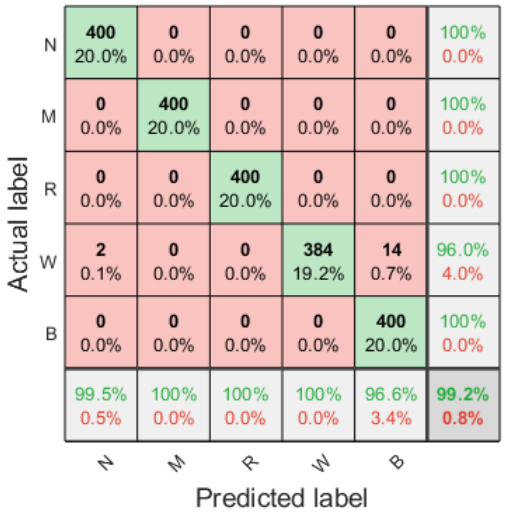
(**c**)



(**d**)

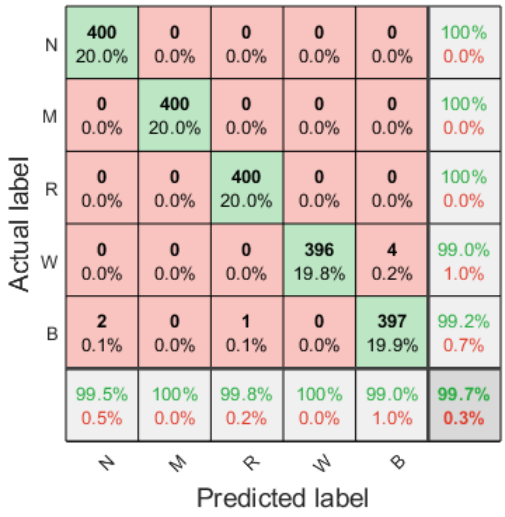


(**e**)



(**f**)

Figure 8. *Cont.*



(**g**)

Figure 8. Comparison of confusion matrices under different feature combinations (Task 2, aggregated results of 10 independent runs). (**a**) Orbit; (**b**) CWT; (**c**) FreqStat; (**d**) Orbit+CWT; (**e**) Orbit+FreqStat; (**f**) CWT+FreqStat; (**g**) ADAMFFN.

4.3.2. Effect of Domain Alignment Mechanism

To validate the effectiveness of the domain alignment mechanism, diagnostic performances with and without DANN domain alignment are compared based on the three-branch feature extraction framework of ADAMFFN. The confusion matrices aggregated from 10 independent runs are presented in Figure 9.

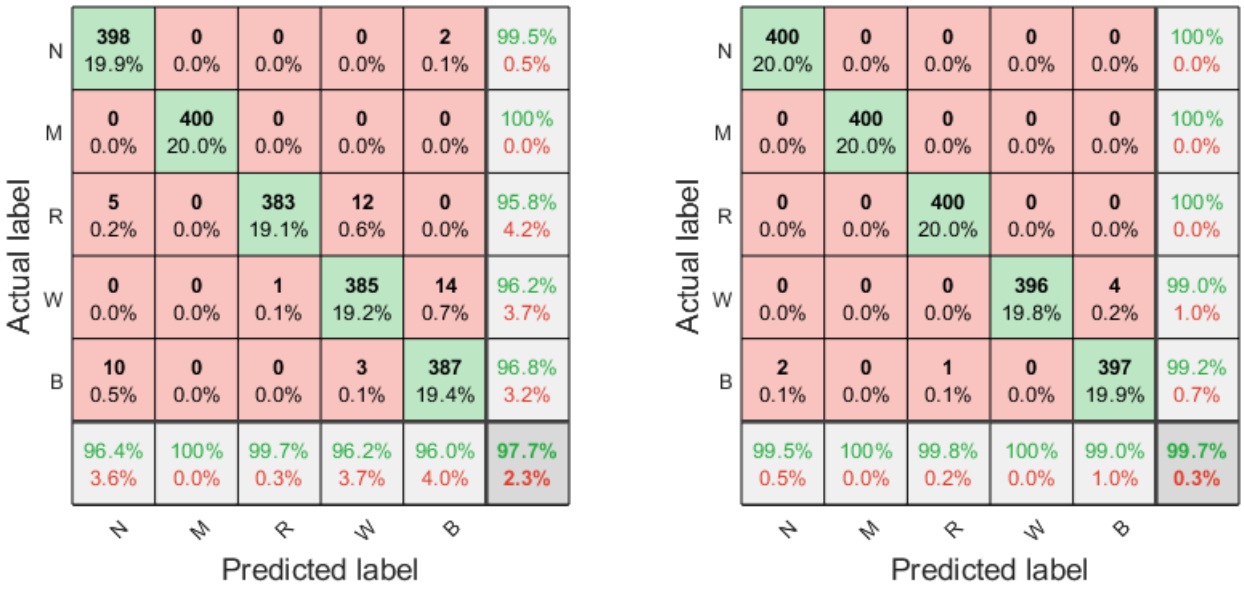


Figure 9. Influence of domain alignment mechanism on diagnostic performance (Task 2, aggregated results of 10 independent runs). (**a**) Without domain alignment; (**b**) with DANN domain alignment.

By comparing Figures 9a and 9b, it can be seen that misclassifications are scattered across multiple categories, including R, W and B, without domain alignment, with a total of 47 misjudged samples and the recall rate of class R dropping to 95.8%. This demonstrates that the feature distribution discrepancy between source and target domains degrades the

generalization ability of fused features. After introducing DANN-based domain alignment, the number of misclassified samples decreases from 47 to 7, and the overall accuracy rises from 97.7% to 99.7%. In addition, all samples of class R are correctly identified. The above results verify that the domain alignment mechanism effectively narrows the inter-domain distribution gap under variable rotating speeds and strengthens the transferability of fused multi-modal features.

4.4. Cross-Speed Generalization Analysis

To comprehensively evaluate the generalization performance of the proposed method across various cross-speed transfer tasks, all comparative approaches are tested on all eight LOSO tasks. The mean accuracy across ten independent runs for each method across the eight tasks is tabulated in Table 3.

Table 3. Comparison of mean accuracy (%) of different methods on eight LOSO tasks.

Task	Orbit	CWT	FreqStat	Orbit+CWT	Orbit+FreqStat	CWT+FreqStat	Proposed
1	78.10	94.85	89.85	95.65	96.35	95.95	97.40
2	74.15	97.15	97.50	97.85	98.15	99.20	99.65
3	81.65	99.05	95.65	99.30	97.70	99.40	99.55
4	75.75	99.60	97.90	99.70	99.85	100.00	100.00
5	73.55	98.35	98.65	99.60	99.65	99.80	99.90
6	72.55	97.35	97.40	97.85	98.50	99.50	99.80
7	83.80	90.45	97.00	96.75	98.75	98.95	99.25
8	77.70	96.15	82.85	96.25	94.60	98.00	98.45
Mean	77.16	96.62	94.60	97.87	97.94	98.85	99.25

It can be observed from Table 3 that the proposed method achieves the highest accuracy, ranging from 97.40% to 100% across all eight tasks, with an average task accuracy of 99.25%, which outperforms the best comparative dual-branch method CWT+FreqStat (98.85%) by 0.40 percentage points. The performance of single-branch methods fluctuates drastically across different tasks. The Orbit single branch only delivers an average accuracy of 77.16%. Although the CWT and FreqStat single branches reach 96.62% and 94.60% respectively, they suffer severe performance degradation on several individual tasks. For instance, the accuracy of CWT drops to 90.45% on Task 7, while FreqStat declines to 82.85% on Task 8. This reveals that single-modal features fail to adapt to all variable-speed transfer scenarios. Dual-branch fusion methods consistently outperform all single-branch counterparts on every task, with average accuracy improved to 97.87–98.85%. Full three-branch fusion further achieves the best results among the compared schemes for each task with minimal performance fluctuations across tasks. This validates that the proposed method simultaneously delivers high diagnostic accuracy and strong generalization capability under variable-speed operating conditions.

Figure 10 presents the aggregated confusion matrices of ten repeated diagnostic results for all eight tasks. The cumulative accuracy of the proposed method is no less than 97.4% on all eight tasks, among which Task 4 achieves 100% correct classification. Only sporadic misclassifications exist in the off-diagonal entries for Task 2 to Task 6. The relatively concentrated misclassification on Task 1 and Task 8 mainly occurs between classes B, N and R. A plausible explanation is that the target speeds of these two tasks lie at the boundaries of the overall speed range, leading to limited transferable information provided by unilateral source domains and weakened inter-class feature discrepancies. Overall, the proportion of misclassified samples remains extremely low and scattered for all tasks, demonstrating that the proposed method maintains stable and reliable diagnostic performance under diverse cross-speed transfer scenarios.

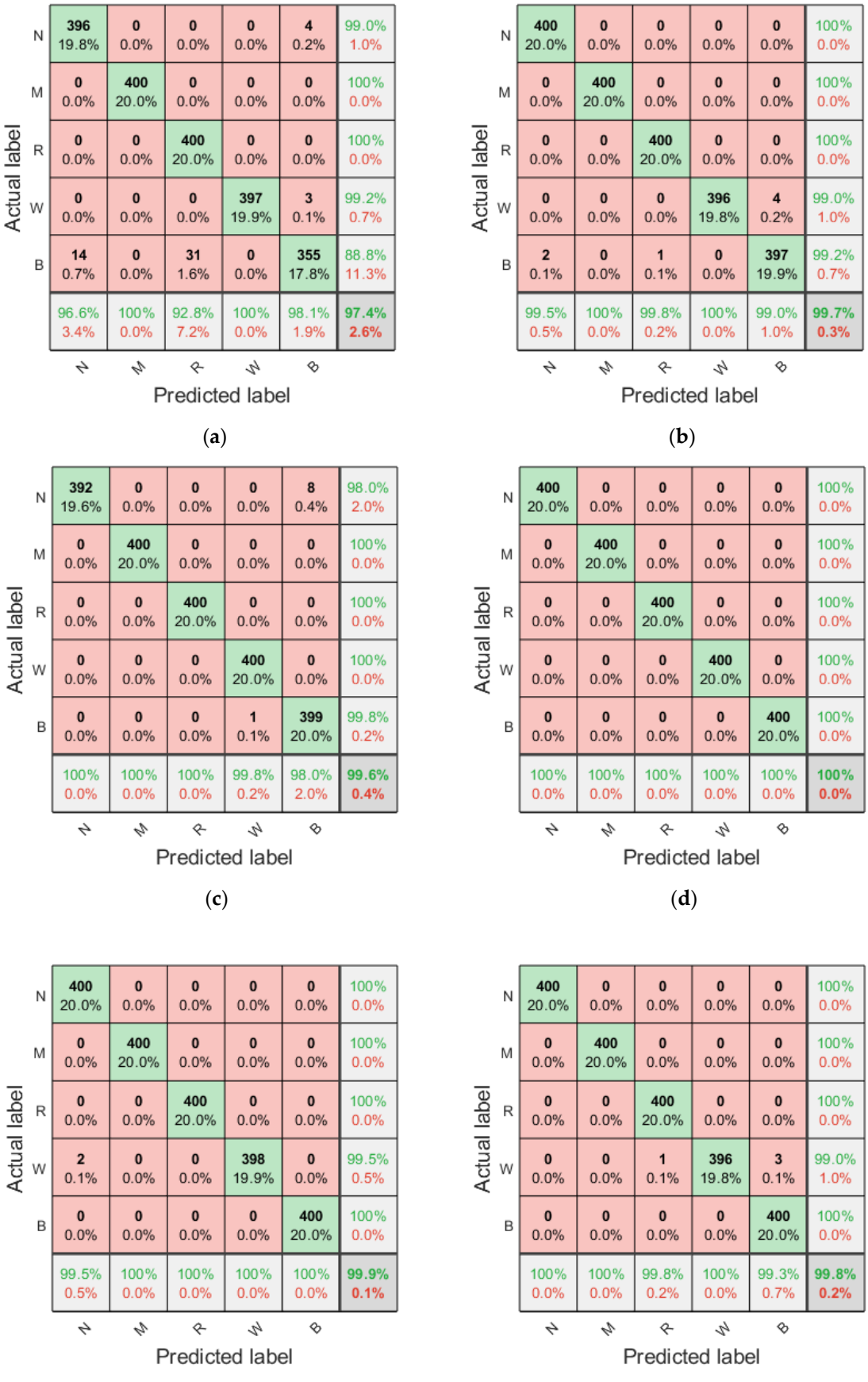
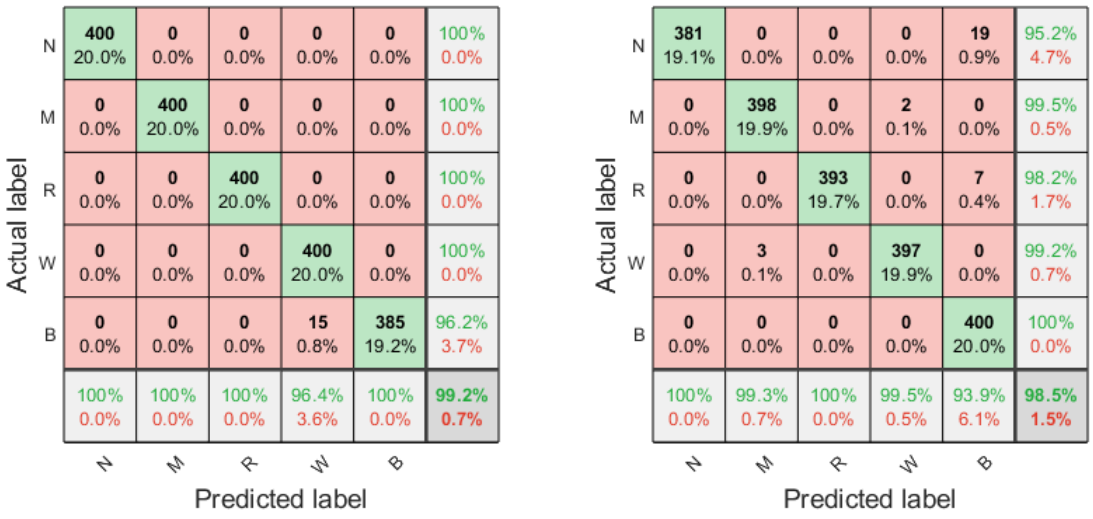


Figure 10. *Cont.*



(**g**) (**h**)

Figure 10. Confusion matrices of the proposed method on eight LOSO cross-speed tasks (aggregated results of 10 independent runs). (**a**–**h**) Correspond to Task 1–Task 8, respectively.

The above LOSO experiments verify the transfer capability of ADAMFFN among discrete steady-state speeds. In practical industrial scenarios, however, planetary gearboxes also undergo continuously varying-speed processes caused by startup, shutdown, and load fluctuations. Under such strongly non-stationary conditions, the three input modalities are affected to different extents.

4.5. Computational Complexity and Real-Time Feasibility

To evaluate the computational overhead and industrial real-time diagnostic potential of ADAMFFN, this section conducts quantitative analysis from three dimensions: model scale, training cost and inference efficiency. The evaluation metrics include trainable parameters, multiply-accumulate operations (MACs), model file size, single-sample feature preprocessing time, training time, and model inference time. All metrics are measured under the hardware and software environment described in Section 4.2. The computational overhead of ADAMFFN and comparative methods is listed in Table 4. It should be emphasized that the domain classifier of DANN is completely removed during testing and deployment. Thus, the domain adversarial mechanism only increases training overhead and does not introduce additional computational cost during inference. The inference and preprocessing times were measured over 10 repeated runs after 10 warm-up runs, and the mean values are reported.

Table 4. Comparison of computational complexity and efficiency among different methods.

Methods	Params /M	MACs /M	Model Size /MB	Feature Preprocessing Time /ms (per Sample)	Training Time/min	Inference Time per Sample/ms
Orbit	1.23	231.47	4.94	2.5	2.82	5.0
CWT	0.11	41.04	0.44	22.1	1.27	2.4
FreqStat	0.05	0.61	0.20	3.9	0.63	1.6
Orbit+CWT	1.46	272.63	5.84	24.3	3.31	7.1
Orbit+FreqStat	1.40	232.20	5.59	5.7	3.05	6.8
CWT+FreqStat	0.23	41.72	0.90	25.2	1.41	4.2
ADAMFFN	1.67	273.41	6.69	27.8	3.77	8.2

As shown in Table 4, ADAMFFN contains 1.67 M parameters with a model file of only 6.69 MB. The Orbit branch accounts for 73.7% of the parameters and 84.7% of the MACs,

constituting the main computational burden, whereas the CWT and FreqStat branches are extremely lightweight, indicating that the accuracy gain of fusion stems from low-cost complementary information. In terms of efficiency, the per-sample preprocessing time is 27 ms, the inference time is 8 ms, and the total end-to-end latency is about 35 ms, accounting for only 41.2% of the 0.085 s signal acquisition window. This leaves a time margin of approximately 2.4 times with an equivalent throughput of about 28 samples/s, demonstrating online real-time diagnostic capability; the 6.69 MB model size also meets the deployment requirements of mainstream edge devices. Compared with the best dual-branch scheme CWT+FreqStat, ADAMFFN gains 0.40 percentage points of accuracy at the cost of 1.44 M additional parameters and 4 ms additional latency, indicating a reasonable accuracy–cost trade-off.

4.6. External Validation on the SEU Gearbox Dataset

To verify that the generalization capability of ADAMFFN is not limited to a single test rig, external validation is conducted on the public gearbox dataset released by Southeast University (SEU) [37]. The SEU gearbox dataset is publicly available at https://github.com/cathysiyu/Mechanical-datasets (accessed on 5 November 2024). The dataset was collected on a drivetrain dynamics simulator containing a planetary gearbox, and covers five gear health states under two speed–load operating configurations as shown in Table 5. Vibration signals were acquired from multiple channels mounted on the planetary gearbox housing. In this work, two mutually perpendicular vibration channels of the planetary gearbox are selected to maintain consistency with the dual-channel input setting of the WT dataset.

Table 5. The operation conditions of the SEU gearbox dataset.

Operation Conditions	Operation States Description	Rotating Frequency	Load
C1	Healthy, Cracks on the gear, Gear tooth breakage, Cracks at the tooth root, Wear on gear tooth surface	20 Hz	0 V (0 Nm)
C2	Healthy, Cracks on the gear, Gear tooth breakage, Cracks at the tooth root, Wear on gear tooth surface	30 Hz	2 V (7.32 Nm)

The preprocessing pipeline is kept identical to that described in Section 4.1: non-overlapping samples of 4096 points are segmented from each channel, and 200 samples per operation state per condition are randomly selected for class balance. Since the two operating configurations differ in both rotational speed and load, they naturally form two cross-condition transfer tasks: C1 $\rightarrow$ C2 and C2 $\rightarrow$ C1. In each task, one configuration serves as the labeled source domain and the other as the unlabeled target domain, following the same closed-set LOSO-style domain adaptation protocol as in Section 4.2. All training hyperparameters are kept unchanged, and each task is independently repeated 10 times.

As shown in Table 6, the proposed ADAMFFN achieves an average accuracy of 94.54% on the SEU dataset, outperforming the best single-branch method (87.93%) and the best dual-branch scheme (91.45%) by 6.61% and 3.09% points, respectively. It should be noted that the SEU dataset differs from the WT dataset in gear parameters, sensor mounting, acquisition system, and background noise; the simultaneous variation in speed and load further increases the domain discrepancy. The maintained superiority of ADAMFFN under these conditions indicates that the tri-modal fusion and domain alignment mechanisms possess generalization capability beyond a single experimental platform. Nevertheless, both datasets are still collected from laboratory simulators, and validation on field industrial data remains future work.

Table 6. Comparison of mean accuracy (%) of different methods on the SEU gearbox dataset.

Task	Orbit	CWT	FreqStat	Orbit+CWT	Orbit+FreqStat	CWT+FreqStat	Proposed
C1 → C2	62.67%	89.70%	78.54%	91.24%	81.57%	93.18%	95.62%
C2 → C1	60.23%	86.16%	81.41%	87.39%	82.05%	89.71%	93.46%
Mean	61.45%	87.93%	79.98%	89.32%	81.81%	91.45%	94.54%

5. Conclusions

To address the variable-speed fault diagnosis problem of planetary gearboxes, this paper proposes an adaptive domain-aligned multi-modal feature fusion network (ADAMFFN). The developed method comprehensively characterizes the operating conditions of planetary gearboxes from multiple complementary perspectives via three parallel heterogeneous feature extraction branches: the Orbit spatial trajectory branch, the CWT time–frequency branch, and the FreqStat frequency-domain statistics branch. A unified projection layer and progressive fusion network are employed to realize effective fusion of heterogeneous features. Furthermore, a DANN-based domain alignment mechanism reduces the distribution discrepancy between training source domains and testing target domains, remarkably improving the generalization ability of the model on unseen rotating speeds. Experimental results of eight LOSO cross-speed fault diagnosis tasks conducted on the WT-Planetary Gearbox dataset demonstrate the following conclusions:

(1) The proposed ADAMFFN achieves an average accuracy of 99.25%, outperforming all single-branch and dual-branch comparative methods, and maintains 98.45% even under the most challenging 55 Hz condition.

(2) Ablation experiments verify the complementarity of the three feature modalities, and full three-branch fusion achieves superior performance over any single-branch or dual-branch combination.

(3) The DANN domain alignment mechanism improves the average accuracy by approximately two percentage points, effectively addressing the distribution shift problem under variable-speed working conditions.

(4) In the cross-condition external validation on the SEU gearbox dataset, ADAMFFN achieves an average accuracy of 94.54%, demonstrating that its generalization capability is not restricted to a single test rig or acquisition system.

Although the proposed method achieves outstanding performance on cross-speed fault diagnosis tasks, several limitations remain and will be addressed in future work. First, vibration signals in practical industrial environments are often contaminated by strong noise; preprocessing modules based on denoising autoencoders or signal augmentation networks will be investigated to enhance diagnostic performance under low signal-to-noise ratio conditions. Second, the method relies on the closed-set assumption, under which unknown faults may be forcibly assigned to known categories; prototype-based rejection or uncertainty estimation will be introduced into the fused feature space, drawing on open-set semi-supervised contrastive learning, to enable the recognition of novel faults. Third, the eight LOSO tasks are constructed at discrete steady-state speeds; angular resampling and order-tracking modules will be introduced at the front end of ADAMFFN to improve its adaptability under continuously varying-speed conditions such as startup and shutdown.

Author Contributions: Methodology, X.X.; software, X.W.; formal analysis, X.X.; data curation, X.X.; writing—original draft, X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Jiangsu Engineering Research Center of the Key Technology for Intelligent Manufacturing Equipment and the Suqian Key Laboratory of Intelligent Manufacturing (Grant No. M202108).

Data Availability Statement: The WT-Planetary Gearbox dataset is publicly available at https://github.com/Liudd-BJUT/WT-planetary-gearbox-dataset [36]. The SEU gearbox dataset is publicly available at https://github.com/cathysiyu/Mechanical-datasets [37].

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

References

1. Wang, T.; Han, Q.; Chu, F.; Feng, Z. Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review. *Mech. Syst. Signal Process.* **2019**, *126*, 662–685. [CrossRef]
2. Stanojević, M.; Tomović, A.; Ivanović, L.; Stojanović, B. Critical Analysis of Design of Ravigneaux Planetary Gear Trains. *Appl. Eng. Lett.* **2022**, *7*, 32–44. [CrossRef]
3. Yang, X.; Shen, W.; Ma, X.; Gao, L.; Zhang, X.; Huang, J. Multi-source feature fusion domain adaptation planetary gearbox fault diagnosis method. *Appl. Sci.* **2025**, *15*, 12457. [CrossRef]
4. Zhang, B.; Teng, W.; Tang, S.; Peng, D.; Fang, Q.; Ma, Z. Sparsity-assisted energy decoupling model for fault detection of planetary ring gear in wind turbines. *Mech. Syst. Signal Process.* **2025**, *238*, 113235. [CrossRef]
5. Gao, X.; Lei, Y.; Yang, X.; Liu, H.; Feng, K.; Wu, T.; Wang, Z. Dynamic interactions of surface wear and tooth crack in gear transmission systems: An investigation of progressive-degradation mechanisms. *Wear* **2025**, *571*, 205861. [CrossRef]
6. Yang, X.; He, G.; Du, C.; Xu, L.; Yu, J.; Zeng, H.; Li, Y. Superiority of fault-caused-speed-fluctuation-based dynamics modeling: An example on planetary gearbox with cracked sun gear. *Machines* **2025**, *13*, 500. [CrossRef]
7. Bukvić, M.; Vencl, A.; Milojević, S.; Skulić, A.; Gajević, S.; Stojanović, B. The Influence of Carbon Nanotube Additives on the Efficiency and Vibrations of Worm Gears. *Lubricants* **2025**, *13*, 327. [CrossRef]
8. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]
9. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3137–3147. [CrossRef]
10. Caesarendra, W.; Tjahjowidodo, T. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. *Machines* **2017**, *5*, 21. [CrossRef]
11. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [CrossRef]
12. Kong, Y.; Wang, T.; Li, Z.; Chu, F. Fault feature extraction of planet gear in wind turbine gearbox based on spectral kurtosis and time wavelet energy spectrum. *Front. Mech. Eng.* **2017**, *12*, 406–419. [CrossRef]
13. 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]
14. 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]
15. Chen, R.; Huang, X.; Yang, L.; Xu, X.; Zhang, X.; Zhang, Y. Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Comput. Ind.* **2019**, *106*, 48–59. [CrossRef]
16. Shi, J.; Peng, D.; Peng, Z.; Zhang, Z.; Goebel, K.; Wu, D. Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks. *Mech. Syst. Signal Process.* **2022**, *162*, 107996. [CrossRef]
17. Wu, R.; Liu, C.; Han, T.; Yao, J.; Jiang, D. A planetary gearbox fault diagnosis method based on time-series imaging feature fusion and a transformer model. *Meas. Sci. Technol.* **2023**, *34*, 024006. [CrossRef]
18. Xia, X.; Sun, H.; Wang, A. Fault diagnosis of planetary gearboxes based on LSTM improved via feature extraction using VMD, fusion entropy, and random forest. *Entropy* **2025**, *27*, 956. [CrossRef] [PubMed]
19. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7316–7325. [CrossRef]
20. Zhang, R.; Tao, H.; Wu, L.; Guan, Y. Transfer learning with neural networks for bearing fault diagnosis in changing working conditions. *IEEE Access* **2017**, *5*, 14347–14357. [CrossRef]
21. Ding, Y.; Jia, M.; Zhuang, J.; Cao, Y.; Zhao, X.; Lee, C.-G. Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108890. [CrossRef]
22. Kullu, O.; Cinar, E. A deep-learning-based multi-modal sensor fusion approach for detection of equipment faults. *Machines* **2022**, *10*, 1105. [CrossRef]

23. Li, X.; Li, J.; Qu, Y.; He, D. Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals. *Appl. Sci.* **2019**, *9*, 768. [CrossRef]
24. Wang, T.; Tang, Y.; He, Y.; Li, Y. A multi-task multimodal attention graph convolutional network for acoustic–vibration fusion-based rolling bearing fault diagnosis. *Appl. Sci.* **2026**, *16*, 4310. [CrossRef]
25. Chu, Y.; Zhu, L.; Lu, M. A dual-attentive multimodal fusion method for fault diagnosis under varying working conditions. *Mathematics* **2025**, *13*, 1868. [CrossRef]
26. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
27. Tzeng, E.; Hoffman, J.; Saenko, K.; Darrell, T. Adversarial discriminative domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2962–2971. [CrossRef]
28. Han, T.; Liu, C.; Yang, W.; Jiang, D. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowl.-Based Syst.* **2019**, *165*, 474–487. [CrossRef]
29. Ahmad, H.; Cheng, W.; Xing, J.; Wang, W.; Du, S.; Li, L.; Zhang, R.; Chen, X.; Lu, J. Deep learning-based fault diagnosis of planetary gearbox: A systematic review. *J. Manuf. Syst.* **2024**, *77*, 730–745. [CrossRef]
30. Cheng, Y.; Li, Y.; Di, W.; Wu, J.; Nie, Z.; Zhan, Z.; Fan, B. An Open-Set Semi-Supervised Contrastive Learning for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 2537010. [CrossRef]
31. Jeong, H.; Park, S.; Woo, S.; Lee, S. Rotating machinery diagnostics using deep learning on orbit plot images. *Procedia Manuf.* **2016**, *5*, 1107–1118. [CrossRef]
32. Jiang, X.; Wang, F.; Zhao, H.; Xu, S.; Lin, L. A Novel Orbit-based CNN Model for Automatic Fault Identification of Rotating Machines. In *Proceedings of the Annual Conference of the PHM Society*; Prognostics and Health Management (PHM): Rochester, NY, USA, 2020; Volume 12, p. 7. [CrossRef]
33. Feng, Z.; Zuo, M.J. Vibration signal models for fault diagnosis of planetary gearboxes. *J. Sound Vib.* **2012**, *331*, 4919–4939. [CrossRef]
34. Orhan, A.; Yordanov, N.; Ertarğın, M.; Zhilevski, M.; Mikhov, M. A comparative study of time–frequency representations for bearing and rotating fault diagnosis using vision transformer. *Machines* **2025**, *13*, 737. [CrossRef]
35. Feng, Z.; Ma, H.; Zuo, M.J. Amplitude and frequency demodulation analysis for fault diagnosis of planet bearings. *J. Sound Vib.* **2016**, *382*, 395–412. [CrossRef]
36. Liu, D.; Cui, L.; Cheng, W. A review on deep learning in planetary gearbox health state recognition: Methods, applications, and dataset publication. *Meas. Sci. Technol.* **2024**, *35*, 012002. [CrossRef]
37. Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Trans. Ind. Inf.* **2019**, *15*, 2446–2455. [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.