machines

MDPI

Article

An Uncertainty-Aware Cross-Modal Fusion Network for Domain Generalization in Mechanical Fault Diagnosis

He Ren [1], Manman Bian [1], Weifeng Lian [2], Chaoyang Wu [2], Qiuyu Song [1,*] and Yi Zhang [1,*]

[1] School of Mechanical Engineering and Rail Transit, Changzhou University, Changzhou 213164, China; hren@cczu.edu.cn (H.R.); s24050858001@smail.cczu.edu.cn (M.B.)
[2] Changzhou NRB Corporation, Changzhou 213022, China; weifeng.lian@nrb.com.cn (W.L.); owen.wu@nrb.com.cn (C.W.)
* Correspondence: qysong@cczu.edu.cn (Q.S.); zy@cczu.edu.cn (Y.Z.)

Abstract

Condition monitoring of modern industrial machinery increasingly relies on data-driven deep learning methods. However, domain shifts caused by dynamic working conditions often lead to substantial accuracy degradation in unseen target domains. Single-modality sensing also suffers from inherent perception limitations under heavy interference. To address these issues, this paper proposes a multi-modal domain generalization framework named the uncertainty-aware cross-modal fusion network (UCFNet). The framework incorporates three key modules. Specifically, a parallel mixed attention strategy captures fine-grained cross-modal interactions and intra-modal context, thereby extracting domain-invariant representations. In addition, a dynamic collaborative gating mechanism filters out feature redundancy and suppresses interference caused by working condition shifts. Furthermore, an uncertainty-aware feature fusion module estimates the prediction confidence of each modality and uses it to guide decision fusion in unseen domains. Comprehensive experiments are conducted on reciprocating compressor and wheelset bearing datasets. Experimental tasks evaluate both multi-source domain generalization and single-source domain generalization. The results demonstrate that UCFNet achieves superior performance compared with state-of-the-art methods.

Keywords: domain generalization; mechanical fault diagnosis; dynamic working conditions; cross-modal fusion

1. Introduction

Data-driven fault diagnosis plays an important role in the safe and stable operation of modern industrial equipment. Traditional vibration-based diagnostic approaches often rely on statistical indices and discriminant analysis to distinguish different operating states or fault types [1]. Deep learning methods excel at feature representation and end-to-end classification [2–4]. In practice, machinery often operates under fluctuating speeds and varying loads. These variations induce severe domain shifts between training and deployment data distributions [5]. A deep learning model trained under a specific stable working condition often relies on condition-specific statistical correlations rather than intrinsic physical fault characteristics. When such a model is deployed under unseen or dynamic working conditions, its diagnostic accuracy often drops sharply [6,7]. Effective diagnosis also requires intuitive interaction with on-site operators for timely decision-making, yet such deployment still depends on the model's reliability under unseen working conditions [8].

Check for updates

Academic Editor: Maciej Sułowicz

Received: 13 October 2026
Revised: 26 October 2026
Accepted: 27 October 2026
Published: 2 November 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.

To reduce the performance degradation caused by distribution shifts, domain adaptation has been widely studied in mechanical fault diagnosis [9]. Domain adaptation aims to align feature distributions between a fully labeled source domain and an unlabeled target domain. Typical techniques include maximum mean discrepancy minimization, adversarial domain discriminators, and optimal transport alignment [10–12]. Although domain adaptation can reduce domain discrepancy, it relies on a restrictive assumption. It requires access to unlabeled target domain samples during the model training phase [13,14]. In real-world industrial scenarios, target working conditions are inherently unpredictable. Target domain fault data is often impossible to collect prior to actual equipment failure [15].

To overcome these limitations, domain generalization has become an important diagnostic paradigm. Domain generalization aims to learn domain-invariant representations solely from source domain data without accessing any target domain information during training [16]. Once trained, a generalized model can be deployed directly to completely unknown target domains with robust diagnostic capabilities [17–19]. However, it remains difficult to collect labeled fault samples under multiple working conditions in practical applications. Therefore, diagnostic algorithms must demonstrate strong generalization capabilities under both multi-source conditions and constrained single-source conditions [20–22].

Apart from distribution shifts, single-modality sensing also limits intelligent fault diagnosis [23]. Most existing domain generalization methods rely exclusively on single-modality signals, such as vibration signals. Although vibration signals effectively capture mechanical impacts, single-modality sensors provide limited physical perspectives [24,25]. In complex industrial environments, vibration data remains highly vulnerable to heavy background interference, electromagnetic interference, and structural resonance [26]. With a single modality, important fault features may be masked by environmental interference. Multi-modal data fusion offers a compelling solution by combining complementary information from distinct physical dimensions, such as vibration signals and acoustic emission signals [27–29]. Acoustic emission sensors capture high-frequency transient elastic waves, providing early detection of micro-cracking and friction, whereas vibration sensors capture low-frequency structural dynamics. Integrating multi-modal signals substantially enhances diagnostic reliability and system tolerance against sensor degradation [30].

Common multi-modal fusion strategies use feature concatenation, element-wise addition, or static tensor products. Under severe domain shifts, dynamic working conditions corrupt different sensor modalities unevenly [31]. For instance, heavy load variations may introduce severe interference into vibration signals while acoustic emission signals remain relatively stable. Simple fusion strategies may mix corrupted feature channels with clean ones. This propagates condition-dependent interference across modalities and distorts the unified feature space, leading to diagnostic failure in unseen target domains. Furthermore, existing multi-modal models lack explicit mechanisms to evaluate prediction confidence when encountering unknown distribution shifts, often producing overconfident erroneous predictions [32,33].

To address these challenges, this paper proposes a multi-modal domain generalization framework called the uncertainty-aware cross-modal fusion network (UCFNet). The proposed framework achieves robust mechanical fault diagnosis across unseen working conditions by explicitly coordinating feature extraction, feature selection, and decision fusion. First, a parallel mixed attention (PMA) strategy establishes fine-grained bidirectional cross-modal semantic interactions while capturing intra-modal context dependencies. PMA effectively isolates domain-invariant fault features from condition-specific amplitude variations. Second, a dynamic collaborative gating (DCG) mechanism operates at the feature channel level. DCG adaptively filters out feature redundancy and suppresses

domain-shift interference generated by dynamic working conditions across modalities. Third, an uncertainty-aware feature fusion (UFF) module calculates prediction uncertainty for each modality based on information entropy. UFF dynamically adjusts fusion confidence weights in real time, preventing corrupted modalities from compromising the final decision in unseen target domains. The main contributions of this paper are as follows:

1. This paper develops a multi-modal domain generalization architecture, UCFNet, for mechanical fault diagnosis under variable working conditions. The network achieves robust performance across unseen target domains without requiring target domain data during training.
2. The parallel mixed attention strategy is designed to capture bidirectional cross-modal dependencies and intra-modal context at the same time. This strategy extracts domain-invariant synergistic features and mitigates condition-specific signal distortion.
3. A dynamic collaborative gating mechanism is introduced to perform channel-wise feature filtering. The gating mechanism dynamically eliminates redundant features and suppresses domain-shift interference across heterogeneous sensor modalities.
4. An uncertainty-aware feature fusion module is designed to estimate the diagnostic confidence of each modality using information entropy. The module dynamically modulates fusion weights under unseen target domains to ensure decision resilience.

The remainder of this paper is organized as follows. Section 2 presents the detailed mathematical formulation of the proposed UCFNet framework. Section 3 describes the experimental datasets and comprehensive experimental results. Section 4 concludes the paper.

2. Methodology

2.1. Problem Formulation

Multi-modal domain generalization for fault diagnosis aims to learn a robust mapping function from labeled multi-modal data in source domains. As shown in Figure 1, multiple-source working conditions define the multi-source domains used during training. The trained model must generalize directly to an unseen target domain under unseen working conditions without leveraging any target samples.

Let $X^v \subset \mathbb{R}^{L_{in}}$ and $X^a \subset \mathbb{R}^{L_{in}}$ denote the raw sample spaces of vibration modality V and acoustic emission modality A, respectively, where L_{in} represents the time-series signal length. Let $Y = \{1,2,\ldots, N_C\}$ denote the label space containing N_C distinct mechanical health categories. Under multi-source domain generalization scenarios, the training set consists of M distinct source domains denoted as $\mathcal{D}_s = \{\mathcal{D}_{s,1},\mathcal{D}_{s,2},\ldots,\mathcal{D}_{s,M}\}$. $M = 1$ corresponds to single-source domain generalization. The m-th source domain is defined as $\mathcal{D}_{s,m} = \left\{(x^v_{s,m,i}, x^a_{s,m,i}, y_{s,m,i})\right\}_{i=1}^{N_{s,m}}$, where $x^v_{s,m,i} \in X^v$ and $x^a_{s,m,i} \in X^a$ are vibration and acoustic emission signals collected simultaneously, $y_{s,m,i} \in Y$ is the health state label, and $N_{s,m}$ is the sample count in domain $\mathcal{D}_{s,m}$. When $M = 1$, the framework reduces to a single-source domain generalization task.

The unseen target domain is denoted as $\mathcal{D}_t = \{(x^v_{t,j}, x^a_{t,j}, y_{t,j})\}_{j=1}^{N_t}$, where N_t represents the target sample count. Due to fluctuating motor speeds and loads, the joint probability distributions differ significantly across domains, such that $P(\mathcal{D}_{s,m}) \neq P(\mathcal{D}_t)$ for any $m \in \{1,2,\ldots,M\}$. The objective of UCFNet is to optimize a mapping function $f: X^v \times X^a \to Y$ using source domain data $\mathcal{D}_s$ to achieve minimum prediction error on target domain $\mathcal{D}_t$.

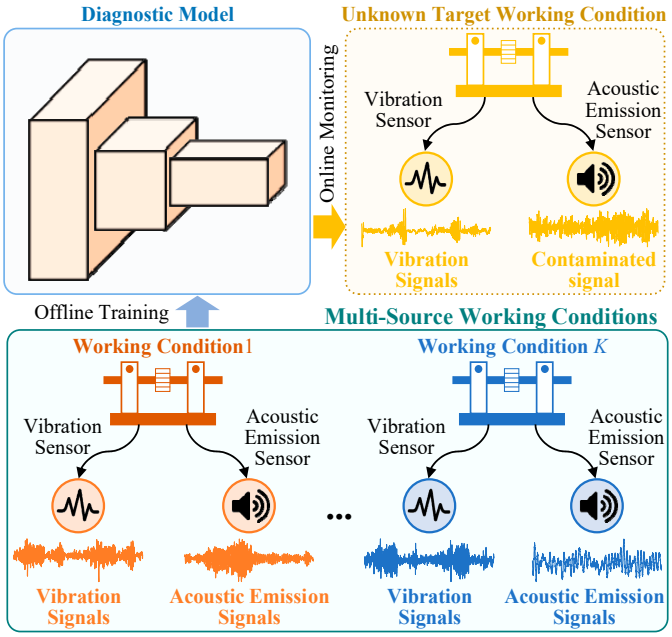


Figure 1. Illustration of multi-modal domain generalization fault diagnosis.

2.2. Architecture of UCFNet

The proposed UCFNet is designed to mitigate feature degradation and domain shifts under unseen working conditions. As illustrated in Figure 2, UCFNet uses a dual-stream feature extractor as the backbone for feature extraction, integrating attentional refinement, dynamic gating, and uncertainty-guided aggregation into an end-to-end framework. The architecture extracts deep diagnostic representations from raw 1-D vibration and AE signals, suppresses domain-irrelevant interference, and merges feature streams according to their predictive reliability. To achieve this, UCFNet operates through several sequentially linked core components: the parallel mixed attention, the dynamic collaborative gating, and the uncertainty-aware feature fusion.

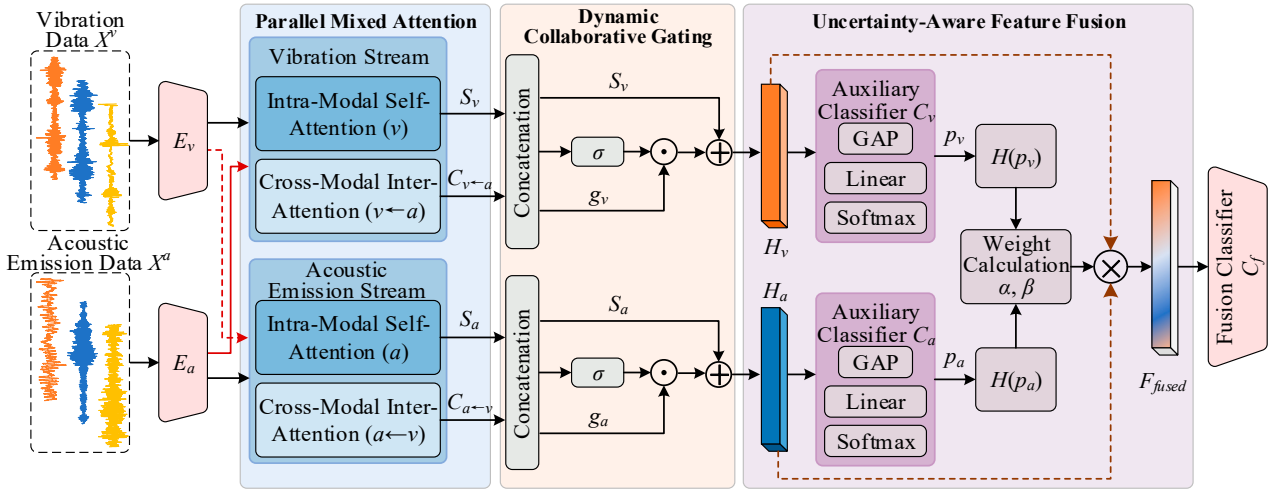


Figure 2. Architecture of the proposed UCFNet.

Typical multi-modal fusion strategies often rely on static feature concatenation or element-wise addition, which can inadvertently propagate condition-specific interference across modalities when facing severe domain shifts. In contrast, the proposed parallel mixed attention captures fine-grained bidirectional cross-modal dependencies while preserving

intra-modal context, rather than treating modalities independently. Furthermore, whereas many existing domain generalization approaches depend on adversarial training or distribution alignment, UCFNet employs the dynamic collaborative gating mechanism to explicitly filter out condition-induced feature redundancy at the channel level without accessing target domain data. Finally, departing from fixed-weight fusion schemes, the uncertainty-aware feature fusion module dynamically adjusts the contribution of each modality based on information entropy. By coordinating these three components, the architecture effectively isolates domain-invariant representations and ensures decision resilience under unseen working conditions.

2.2.1. Feature Extractor

To extract deep semantic representations from raw heterogeneous signals, UCFNet constructs a structurally symmetric dual-stream feature extractor, comprising a vibration feature extraction stream E_v and an acoustic emission (AE) feature extraction stream E_a. Both branches are built using 1-D convolutional neural networks (Conv1D). For the l-th layer of either branch, the feature mapping process is defined as:

$$F_l = \delta(BN(Conv1D(F_{l-1}))) \tag{1}$$

where F_l denotes the feature map output by the l-th layer, δ represents the ReLU activation function, $BN(\cdot)$ denotes Batch Normalization, and $Conv1D(\cdot)$ indicates the 1-D convolution operation. After multi-layer feature extraction, the deep feature representation of the vibration signal F_v and that of the AE signal F_a are obtained, respectively.

2.2.2. Parallel Mixed Attention

Traditional cross-modal interaction methods often ignore correlations across different semantic subspaces. Therefore, this strategy adopts a symmetric dual-branch multi-head attention mechanism. Features are divided along the semantic dimension into h independent attention heads to compute attention weights in parallel across distinct feature subspaces, thereby capturing richer intra-modal and cross-modal semantic interactions.

(1) Intra-Modal Multi-Head Self-Attention

To enforce the retention of critical domain-specific periodic characteristics and localized impact features, single-modality deep features are first enhanced via self-attention, as shown in Figure 3. Taking the vibration feature F_v as an example, in the i-th semantic subspace, it is linearly projected into query vector $Q^i_{v,self}$, key vector $K^i_{v,self}$, and value vector $V^i_{v,self}$:

$$Q^i_{v,self} = F_v W^{v,i}_{Q,self}, \quad K^i_{v,self} = F_v W^{v,i}_{K,self}, \quad V^i_{v,self} = F_v W^{v,i}_{V,self} \tag{2}$$

where $W^{v,i}_{Q,self}$, $W^{v,i}_{K,self}$, and $W^{v,i}_{V,self}$ are learnable projection matrices for the i-th self-attention head.

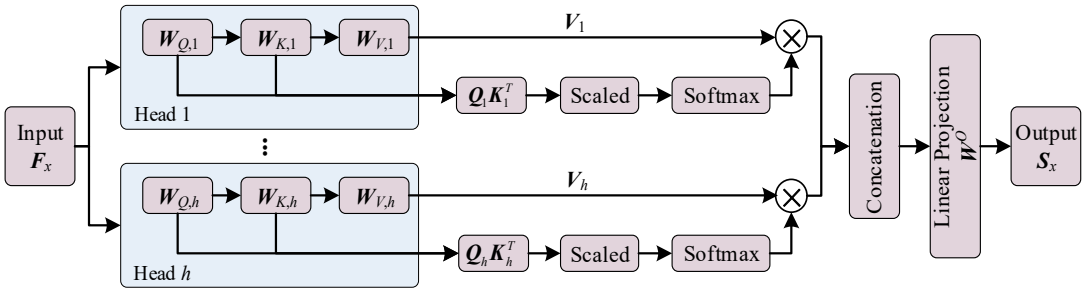


Figure 3. Illustration of Multi-Head Self-Attention.

Subsequently, scaled dot-product attention is calculated independently within this subspace:

$$\text{Head}_{i,self}^{v} = \text{Softmax}\left(\frac{Q_{v,self}^{i}(K_{v,self}^{i})^{T}}{\sqrt{d_k}}\right)V_{v,self}^{i} \tag{3}$$

where d_k is the feature dimension of each attention head. Finally, the outputs from all h subspaces are concatenated and projected via W_{self}^{O} to obtain the complete self-attention enhanced vibration feature S_v:

$$S_v = \text{Concat}(\text{Head}_{1,self}^{v}, \ldots, \text{Head}_{h,self}^{v})W_{self}^{O} \tag{4}$$

Similarly, the self-attention enhanced AE feature S_a is derived through a symmetric process.

(2) Cross-Modal Multi-Head Cross-Attention

To establish fine-grained complementary relationships between different modalities, a symmetric cross-modal cross-attention branch is developed. Taking vibration-guided AE as an example, the vibration feature F_v serves as the query to retrieve relevant fault semantics from the AE feature F_a. In the i-th subspace, the projection process is formulated as:

$$Q_{v,cross}^{i} = F_v W_{Q,cross}^{v,i}, \quad K_{a,cross}^{i} = F_a W_{K,cross}^{a,i}, \quad V_{a,cross}^{i} = F_a W_{V,cross}^{a,i} \tag{5}$$

The cross-modal dependencies and the final complementary feature $C_{v\leftarrow a}$ are computed as follows:

$$\text{Head}_{i,v\leftarrow a}^{cross} = \text{Softmax}\left(\frac{Q_{v,cross}^{i}(K_{a,cross}^{i})^{T}}{\sqrt{d_k}}\right)V_{a,cross}^{i} \tag{6}$$

$$C_{v\leftarrow a} = \text{Concat}(\text{Head}_{1,v\leftarrow a}^{cross}, \ldots, \text{Head}_{h,v\leftarrow a}^{cross})W_{cross}^{O} \tag{7}$$

Similarly, the AE-guided vibration cross-modal complementary feature $C_{a\leftarrow v}$ is obtained symmetrically.

2.2.3. Dynamic Collaborative Gating

Since AE and vibration signals may contain redundant interference under unseen target conditions, a gating residual mechanism is used when fusing cross-modal features $C_{v\leftarrow a}$ with intra-modal features S_v. The dynamic gating weight g_v is calculated to yield the hybrid enhanced vibration feature H_v:

$$g_v = \sigma(W_{g,v}[S_v, C_{v\leftarrow a}]) \tag{8}$$

$$H_v = S_v + g_v \odot C_{v\leftarrow a} \tag{9}$$

Symmetric operations are executed for the AE branch to obtain H_a:

$$g_a = \sigma(W_{g,a}[S_a, C_{a\leftarrow v}]) \tag{10}$$

$$H_a = S_a + g_a \odot C_{a\leftarrow v} \tag{11}$$

where $W_{g,v}$ and $W_{g,a}$ are learnable linear projection matrices that map the concatenated intra-modal and cross-modal features into the gating space, and $\sigma(\cdot)$ denotes the Sigmoid activation function. When the cross-modal features contain strong non-target interference, g_v approaches 0 and blocks contamination. Otherwise, the gate opens and g_v approaches 1, allowing complementary features to pass through.

2.2.4. Uncertainty-Aware Feature Fusion

Under single-source domain generalization or varying working conditions, different modalities are affected by interference to different degrees. Therefore, the model uses predictive entropy to estimate the confidence of each modality and performs reliability-weighted fusion.

(1) Auxiliary Classifiers and Entropy Calculation

To obtain independent information entropy for each modality, identical auxiliary classifiers are connected in parallel after the gated output features H_v and H_a. Each auxiliary classifier consists of a Global Average Pooling (GAP) layer, a linear mapping layer, and a Softmax function. Taking the vibration modality as an example, its predicted probability distribution p_v is calculated as:

$$p_v = \text{Softmax}(W_{c,v} \cdot \text{GAP}(H_v) + b_{c,v}) \tag{12}$$

The predicted probability distribution for the AE modality p_a is computed as:

$$p_a = \text{Softmax}(W_{c,a} \cdot \text{GAP}(H_a) + b_{c,a}) \tag{13}$$

where W_c and b_c denote the weight matrix and bias vector of the corresponding auxiliary linear layer.

Next, the predictive information entropies $H(p_v)$ and $H(p_a)$ are calculated. A lower entropy indicates higher confidence in the prediction of a modality.

$$H(p_v) = -\sum_{j=1}^{N_c} p_{v,j} \log(p_{v,j}) \tag{14}$$

$$H(p_a) = -\sum_{j=1}^{N_c} p_{a,j} \log(p_{a,j}) \tag{15}$$

where N_c represents the total number of fault categories, and $p_{v,j}$ and $p_{a,j}$ are the predicted probabilities of belonging to class j for vibration and AE modalities, respectively.

The reliability weights α and β derived from the information entropy are formulated as:

$$\alpha = \frac{\exp(-H(p_v))}{\exp(-H(p_v)) + \exp(-H(p_a))} \tag{16}$$

$$\beta = \frac{\exp(-H(p_a))}{\exp(-H(p_v)) + \exp(-H(p_a))} \tag{17}$$

Finally, the two features are weighted and concatenated to obtain the fused representation F_{fused}:

$$F_{fused} = \text{Concat}(\alpha H_v, \beta H_a) \tag{18}$$

(2) Overall optimization objective

To improve classification at the fusion level and encourage the auxiliary classifiers to learn discriminative features, a multi-task optimization framework is used. The primary classification loss adopts standard cross-entropy, applied to the output probability distribution p_{fused} from the fusion classifier C receiving F_{fused}:

$$\mathcal{L}_{cls} = -\sum_{i} \sum_{c=1}^{N_c} y_{i,c} \log(p_{fused,i,c}) \tag{19}$$

where y_i is the one-hot ground-truth label and $p_{fused,i,c}$ is the predicted probability of class c.

Simultaneously, auxiliary supervision losses $\mathcal{L}_v$ and $\mathcal{L}_a$ are introduced for the auxiliary classifiers to independently optimize the feature extraction network of each modality.

$$\mathcal{L}_v = -\sum_i \sum_{c=1}^{N_c} y_{i,c} \log(p_{v,i,c}) \tag{20}$$

$$\mathcal{L}_a = -\sum_i \sum_{c=1}^{N_c} y_{i,c} \log(p_{a,i,c}) \tag{21}$$

The overall loss function $\mathcal{L}_{total}$ of the model is defined as the weighted sum of the primary and auxiliary classification losses:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda(\mathcal{L}_v + \mathcal{L}_a) \tag{22}$$

where λ balances the fusion classification loss and the auxiliary modality losses. Through joint optimization, the model drives feature interaction and reliability assessment of heterogeneous signals, significantly enhancing generalization capabilities and diagnostic accuracy under unseen working conditions.

It should be noted that the overall loss function contains only classification losses, without explicit domain-invariance constraints. The domain generalization capability of UCFNet is primarily achieved through structural inductive biases rather than loss-level alignment. Specifically, the classification loss forces the network to extract discriminative features. The PMA module captures cross-modal complementary information to suppress condition-specific amplitude variations, while the DCG mechanism explicitly filters out redundant channels that are highly correlated with working condition shifts. Moreover, the UFF module adapts to unseen domains during the inference phase by estimating prediction confidence.

3. Experiment and Analysis

3.1. Description of Datasets and Diagnosis Tasks

To evaluate the cross-condition generalization of UCFNet under unseen working conditions, experiments are conducted on two rotating machinery datasets: the private reciprocating compressor dataset from Changzhou University and the wheelset bearing dataset from Soochow University [34].

(1) Changzhou University Reciprocating Compressor Dataset

As shown in Figure 4, the reciprocating compressor dataset was collected from a compressor test rig located at Changzhou University. The suction and discharge valves are key components of the compressor. To reproduce valve damage, a scratch measuring 2 mm in width was introduced on the valve. A rigid element was mounted onto the valve to restrict the lift of the valve plate, thereby mimicking spring failure. Vibration and acoustic emission (AE) sensors were attached to the compressor cylinder block to collect condition data at sampling rates of 12.8 kHz and 200 kHz, respectively. These signals were collected under three working conditions with different rotational speeds. The dataset covers typical health conditions: normal condition (N), intake valve wear (I), exhaust valve wear (E), exhaust valve spring failure (S), and the compound fault of intake valve wear and exhaust valve spring failure (IS), which are labeled from 0 to 4, respectively. The sample length is 4096 and there is no overlap between consecutive segments. Before being fed into the UCFNet model, both vibration and acoustic emission signals are normalized to the range of $[-1, 1]$ using min-max normalization. Under the domain generalization scheme, the model is trained exclusively on labeled data from single or multi-source domains. It is then evaluated directly on the remaining unseen target domain. Details of the working conditions are shown in Table 1.

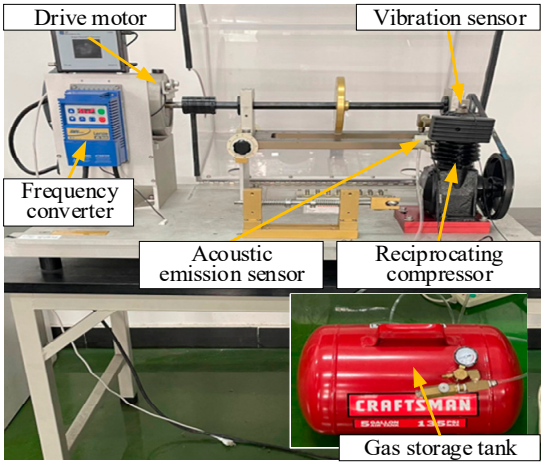


Figure 4. Reciprocating compressor fault simulation test rig.

Table 1. Description of the compressor dataset.

Domain	Rotational Speed/(rpm)	Number of Samples
D1	720	5 × 300
D2	960	5 × 300
D3	1200	5 × 300

(2) As shown in Figure 5, the test rig for wheelset bearing fault simulation at Soochow University consists primarily of a drive motor, a wheel-rail simulation section, and a radial loading apparatus. The large wheelset and small wheelset in the wheel-rail simulation section serve to emulate the rail and train wheels, respectively. A test bearing is fitted to one side of the small wheelset. To capture condition signals, vibration and acoustic emission sensors were deployed, with sampling rates set at 32,768 Hz and 1 MHz, respectively. Signals were collected under three working conditions by adjusting motor speed and load, forming three domains, as shown in Table 2. Each domain involves four types of wheelset bearing health conditions: normal condition (N), inner race fault (I), outer race fault (O), and roller fault (R), labeled 0 to 3, respectively. The length of each sample is 4096.

The vibration and acoustic emission signals are acquired simultaneously via a synchronized data acquisition system, ensuring temporal consistency. The original sampling rates are retained without resampling to preserve the high-frequency transient features of the acoustic emission signals. Instead of raw signal interpolation, the two modalities are aligned in the deep feature space by PMA, which establishes fine-grained bidirectional semantic interactions across different time scales.

To evaluate cross-domain diagnosis, 14 tasks were designed on the two datasets, as listed in Table 3. During training, only the designated source domain data are used, and the target domain is used only for testing. To minimize the influence of random fluctuations, all final metrics are reported as the average accuracy obtained over five separate experimental trials.

Table 2. Description of the wheelset bearing dataset.

Domain	Rotational Speed/(rpm)	Radial Load/(N)	Number of Samples
M1	400	1600	4 × 400
M2	800	800	4 × 400
M3	400~700	2400	4 × 400

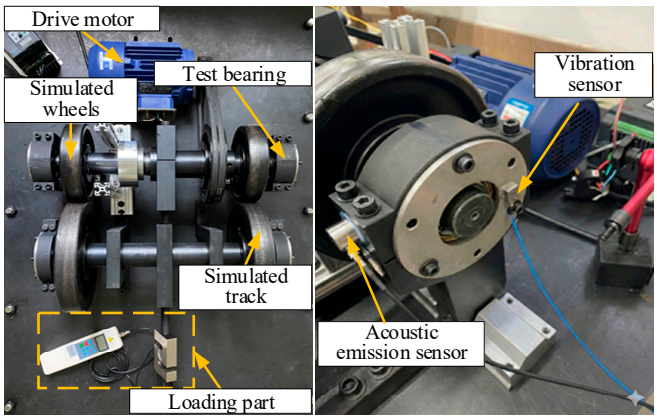


Figure 5. Wheelset bearing fault simulation test rig.

Table 3. Fault diagnosis tasks.

Dataset	Task	Source	Target	Dataset	Task	Source	Target
Compressor	T1	D1/D2	D3	Wheelset bearing	S1	M1/M2	M3
	T2	D1/D3	D2		S2	M1/M3	M2
	T3	D2/D3	D1		S3	M2/M3	M1
	T4	D1	D2		S4	M1	M2
	T5	D2	D3		S5	M2	M3
	T6	D3	D1		S6	M3	M1
	T7	D3	D2		S7	M3	M2

The proposed method is compared with the following baselines. (1) CCDG: Conditional contrastive domain generalization [15]; (2) MFDACN: where multi-modal (MF) data is fed into dual adversarial and contrastive network (DACN) for recognition [21]; (3) MFIDAN: where multi-modal (MF) data is fed into Intra-domain adversarial network (IDAN) for recognition [22]; (4) DCMFC [28]: Dual-input CNN for multi-modal fault classification; (5) w/o DCG: in which the dynamic collaborative gating mechanism is removed; (6) w/o PMA: in which the parallel mixed attention is removed; (7) w/o UFF: in which the uncertainty-aware feature fusion is removed. (8) Conventional attention fusion (CAF): replace PMA with standard multi-head self-attention.

Only vibration signals are used as inputs in CCDG, which provides a reference point for single-modality domain generalization. Tables 4 and 5 present the parameters of the UCFNet and the multi-head attention, respectively. For model training, we utilize the Adam optimizer with a batch size of 32, a learning rate equal to 0.001, and the training process runs for 400 epochs. All experiments are conducted on a workstation equipped with an Intel i5-13600KF CPU at 3.5 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX 4070 SUPER GPU. The software environment is Ubuntu 20.04 with Python 3.7.

Table 4. Network parameters of UCFNet.

Module	Layer	Kernel Size/Node	Kernel Number	Stride	Activation Function
Feature extractor E_v, E_a	Conv_1	9 × 1	8	4	ReLU
	Conv_2	9 × 1	32	4	ReLU
	Conv_3	9 × 1	64	4	ReLU
	Conv_4	9 × 1	128	4	ReLU
	Conv_5	16 × 1	256	1	-
Auxiliary Classifier C_v, C_a	FC_1	256	-	-	ReLU
	FC_2	256	-	-	Softmax
Fusion Classifier C_f	FC_1	512	-	-	ReLU
	FC_2	N_C	-	-	Softmax

Table 5. Multi-head attention parameters.

Parameter	Value/Dimension
Input	256
Head Number	8
Key/Value	16
Output	256

3.2. Analysis of Diagnostic Results

Tables 6 and 7 summarize the average accuracy, standard deviation, and 95% confidence interval of UCFNet, the baseline models, and the ablation variants on the two datasets. Conventional single-modality domain generalization methods perform poorly under dynamic working conditions because a single sensor provides limited physical information. For instance, CCDG optimizes class boundaries to enhance feature discriminability; however, it yields lower overall performance, obtaining average diagnostic accuracies of only 75.77% on the reciprocating compressor dataset and 85.12% on the wheelset bearing dataset. Multi-modal approaches adapted for domain shift mitigation, such as MFDACN and MFIDAN, achieve improved performance through multi-scale adversarial feature alignment and domain representation learning. MFIDAN achieves average accuracies of 84.18% on the compressor tasks and 93.59% on the wheelset bearing tasks. Although the DCMFC model incorporates multi-modal feature fusion, it lacks explicit constraints for out-of-distribution generalization scenarios, resulting in average diagnostic accuracies of 80.36% and 92.09% across the two datasets.

In contrast, UCFNet achieves better performance in all evaluated scenarios on both datasets. Specifically, UCFNet obtains overall average diagnostic accuracies of 90.29% ± 0.98% on the reciprocating compressor dataset and 97.12% ± 0.86% on the wheelset bearing dataset, outperforming all baseline models. Notably, UCFNet exhibits the lowest standard deviation among all methods. Compared with the strongest baseline MFIDAN, UCFNet reduces the standard deviation from 1.43% to 0.98% on the compressor dataset and from 1.23% to 0.86% on the bearing dataset, indicating highly stable diagnostic performance across multiple runs. Moreover, the 95% confidence intervals of UCFNet are strictly separated from those of the baseline methods, demonstrating that the performance gains are statistically significant rather than random fluctuations. In multi-source tasks, UCFNet reaches peak accuracies up to 94.34% on compressor task T3 and 99.14% on bearing task S1. Furthermore, under challenging single-source domain generalization tasks, UCFNet retains high decision stability with accuracies of 85.16% on compressor task T7 and 93.58% on bearing task S5. This demonstrates that the coordination of parallel mixed attention, dynamic collaborative gating, and uncertainty-aware feature fusion enables the network to extract domain-invariant diagnostic representations resilient to operational variations.

Table 6. Diagnostic accuracy on the compressor dataset.

Method	Accuracy/% ± Standard Deviation/%								
	T1	T2	T3	T4	T5	T6	T7	Average	95% CI
CCDG	74.21 ± 2.54	79.07 ± 1.98	81.61 ± 3.26	72.78 ± 1.55	75.64 ± 1.84	71.57 ± 2.39	75.49 ± 1.26	75.77 ± 2.12	[72.62, 78.90]
MFDACN	80.54 ± 0.98	87.21 ± 1.37	88.43 ± 1.68	80.56 ± 1.74	85.25 ± 1.28	77.30 ± 2.65	81.32 ± 2.24	82.94 ± 1.71	[79.14, 86.74]
MFIDAN	79.32 ± 0.81	90.18 ± 0.83	92.44 ± 1.25	79.55 ± 1.61	87.70 ± 1.98	78.83 ± 1.69	81.27 ± 1.83	84.18 ± 1.43	[78.57, 89.79]
DCMFC	81.26 ± 1.62	85.42 ± 1.57	87.19 ± 1.68	77.62 ± 2.33	79.53 ± 2.16	75.15 ± 1.35	76.38 ± 1.85	80.36 ± 1.79	[76.00, 84.72]
w/o DCG	87.14 ± 2.36	88.65 ± 1.67	89.81 ± 0.76	82.94 ± 1.95	84.17 ± 2.01	80.40 ± 1.24	84.35 ± 0.64	85.35 ± 1.52	[82.13, 88.57]
w/o PMA	80.37 ± 2.09	85.47 ± 2.37	89.08 ± 1.62	78.39 ± 1.17	82.12 ± 1.58	78.46 ± 1.95	79.35 ± 1.66	81.89 ± 1.78	[78.10, 85.68]
w/o UFF	84.30 ± 0.82	89.41 ± 0.94	90.85 ± 1.05	81.20 ± 1.55	85.17 ± 1.61	81.74 ± 1.75	82.53 ± 1.23	85.03 ± 1.28	[81.33, 88.73]
CAF	82.06 ± 1.95	87.25 ± 1.75	88.64 ± 1.34	80.03 ± 1.28	84.21 ± 1.31	80.39 ± 1.84	81.52 ± 1.25	83.44 ± 1.53	[80.31, 86.57]
UCFNet	90.81 ± 0.74	93.29 ± 0.82	94.34 ± 1.27	87.63 ± 1.21	92.78 ± 0.87	88.05 ± 1.05	85.16 ± 0.93	90.29 ± 0.98	[87.10, 93.50]

Table 7. Diagnostic accuracy on the wheelset bearing dataset.

Method	Accuracy/% ± Standard Deviation/%								
	S1	S2	S3	S4	S5	S6	S7	Average	95% CI
CCDG	90.25 ± 1.35	87.50 ± 1.81	85.07 ± 1.64	84.49 ± 1.52	79.65 ± 2.14	84.32 ± 1.03	84.58 ± 1.85	85.12 ± 1.62	[81.92, 88.32]
MFDACN	93.91 ± 0.92	92.63 ± 0.84	92.75 ± 1.57	91.58 ± 1.60	89.30 ± 1.55	88.06 ± 1.20	88.54 ± 1.70	90.97 ± 1.34	[88.87, 93.07]
MFIDAN	95.84 ± 0.86	95.69 ± 0.78	96.13 ± 1.01	94.24 ± 1.34	88.78 ± 1.57	92.33 ± 1.69	92.09 ± 1.37	93.59 ± 1.23	[91.07, 96.11]
DCMFC	96.20 ± 1.03	95.03 ± 0.91	91.98 ± 1.24	92.46 ± 1.56	88.20 ± 1.60	88.37 ± 1.52	92.37 ± 1.14	92.09 ± 1.29	[89.30, 94.88]
w/o DCG	94.92 ± 0.84	95.98 ± 0.76	93.50 ± 1.14	94.47 ± 1.12	87.06 ± 1.35	93.44 ± 1.27	90.45 ± 1.09	92.83 ± 1.08	[89.86, 95.80]
w/o PMA	91.73 ± 1.25	91.75 ± 0.81	90.42 ± 0.93	90.56 ± 1.38	90.73 ± 1.84	92.80 ± 1.16	90.74 ± 1.14	91.25 ± 1.22	[90.46, 92.04]
w/o UFF	97.36 ± 0.67	92.39 ± 0.92	92.05 ± 0.97	96.87 ± 1.30	91.41 ± 1.34	95.18 ± 1.42	94.16 ± 1.57	94.20 ± 1.17	[91.93, 96.49]
CAF	93.75 ± 0.95	91.88 ± 0.76	91.37 ± 1.04	93.65 ± 1.20	90.82 ± 1.47	93.16 ± 0.96	91.74 ± 1.42	92.34 ± 1.11	[91.26, 93.42]
UCFNet	99.14 ± 0.83	98.76 ± 0.59	97.62 ± 0.71	97.21 ± 1.06	93.58 ± 1.25	97.15 ± 0.86	96.38 ± 0.72	97.12 ± 0.86	[95.35, 98.89]

To verify the explicit contribution of each structural component within UCFNet, comprehensive ablation experiments were conducted under identical parameter settings. Among the variants, removing PMA causes the largest performance drop, with average diagnostic accuracies falling to 81.89% on the reciprocating compressor dataset and 91.25% on the wheelset bearing dataset. This decrease shows that bidirectional cross-modal interaction and intra-modal context modeling are important for preventing condition-induced distortions from masking fault features.

The remaining ablation variants further demonstrate the necessity of feature filtering and uncertainty-guided fusion mechanisms for robust domain generalization. The w/o DCG leads to accuracy drops, yielding average diagnostic accuracies of 85.35% and 92.83% across the compressor and wheelset bearing datasets, respectively. This indicates that channel-wise dynamic gating plays a critical role in filtering feature redundancy and suppressing domain-shift interference across heterogeneous modalities. Similarly, the w/o UFF variant, which essentially acts as an equal-weight fusion strategy with fixed 0.5 weights, results in average accuracies of 85.03% on the compressor dataset and 94.20% on the wheelset bearing dataset, proving that evaluating prediction confidence via information entropy successfully prevents corrupted sensor modalities from compromising final decision-making in unseen working environments. In addition, the CAF variant, which replaces the proposed PMA with a conventional multi-head self-attention mechanism, achieves average accuracies of only 83.44% and 92.34% on the two datasets, respectively. This confirms that conventional attention lacks the ability to capture fine-grained bidirectional cross-modal interactions compared with our parallel mixed attention design.

To directly validate the interference suppression capabilities of DCG and UFF, Gaussian white noise at different signal-to-noise ratios (SNRs) from 4 dB to −4 dB was added to the test signals. As shown in Figure 6, under low SNR conditions, removing DCG or UFF leads to a more severe performance drop compared to the full UCFNet. This confirms that DCG effectively filters out noise-induced redundancy, while UFF dynamically reduces the weight of corrupted modalities, thereby ensuring robust decision-making.

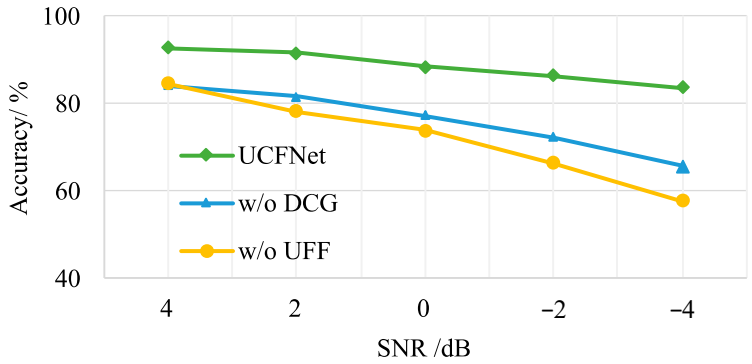


Figure 6. Robustness against noise on task T5.

To evaluate multi-modal feature fusion, performance comparisons were conducted between single-modality models and the multi-modal UCFNet, as illustrated in Figure 7. Single-modality models were constructed using a single-branch feature extractor with multi-head attention. Under cross-condition generalization scenarios, these models exhibited noticeable accuracy fluctuations. On the compressor dataset, models relying solely on vibration or acoustic emission signals showed marked accuracy drops in tasks T3 and T6. On the wheelset bearing dataset, single-modality models also displayed reduced robustness in tasks S2 and S3. These results demonstrate the physical sensing limitations of relying on a single sensor modality. In contrast, UCFNet integrates complementary information across modalities, achieving higher diagnostic stability and outperforming single-modality models across all cross-condition tasks.

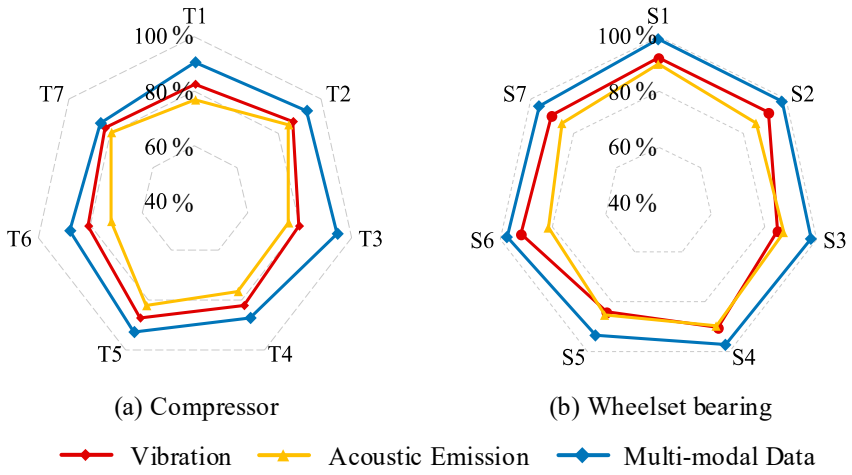


Figure 7. Diagnostic accuracy of different modalities.

3.3. Visualization Analysis

To provide a detailed analysis of the diagnostic results, Figure 8 presents the confusion matrices of several comparative methods for Task T1. Evidently, both MFIDAN and DCMFC exhibit severe misclassification between health states N (label 0) and I (label 1). Furthermore, as depicted in Figure 8a, MFDACN frequently misclassifies state S (label 3) as state E (label 2). This indicates that while MFDACN improves the recognition of certain states, it still suffers from significant confusion between states E and S. Although UCFNet exhibits minor confusion, it demonstrates a distinct advantage in overall diagnostic accuracy, with the vast majority of samples being correctly classified.

To intuitively evaluate the feature distributions learned by UCFNet and baseline methods, the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique was employed to project the learned feature representations of diagnosis task T5 into a 3-D space, as shown in Figure 9. In the visualization, S_N and T_N represent the normal samples from the source domain and the target domain, respectively. As depicted in Figure 9a–c, the features extracted by MFDACN, MFIDAN, and DCMFC exhibit substantial overlap between the source and target domain distributions for certain fault categories. Specifically, DCMFC displays severe overlapping clusters between the intake valve wear and exhaust valve wear. By contrast, Figure 9d illustrates that UCFNet achieves compact intra-class alignment between source and target domain samples, while simultaneously maintaining distinct decision boundaries among different fault states. The feature overlap is significantly reduced, demonstrating that UCFNet effectively extracts domain-invariant and highly discriminative representations under unseen working conditions.

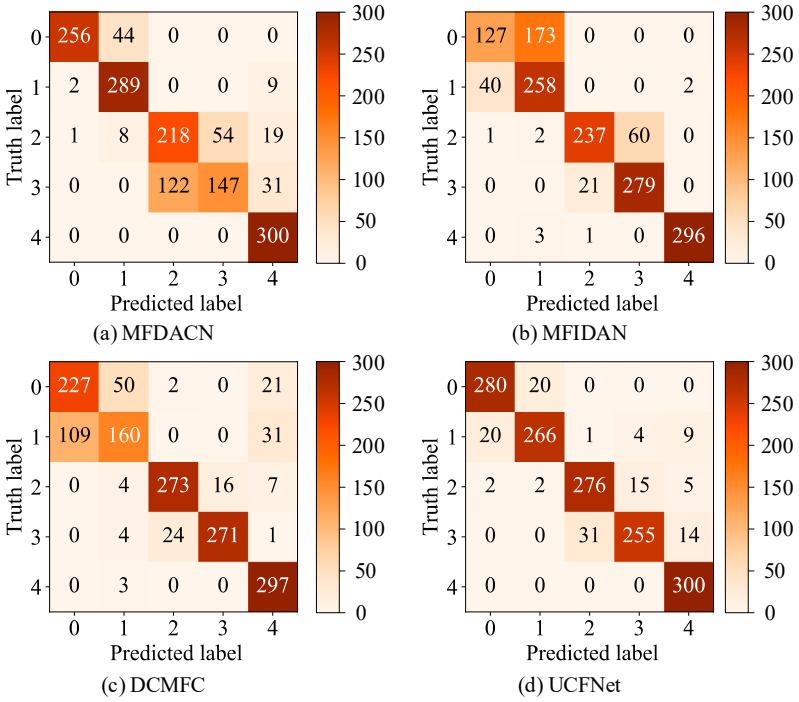


Figure 8. Confusion matrices of the diagnostic results of task T1.

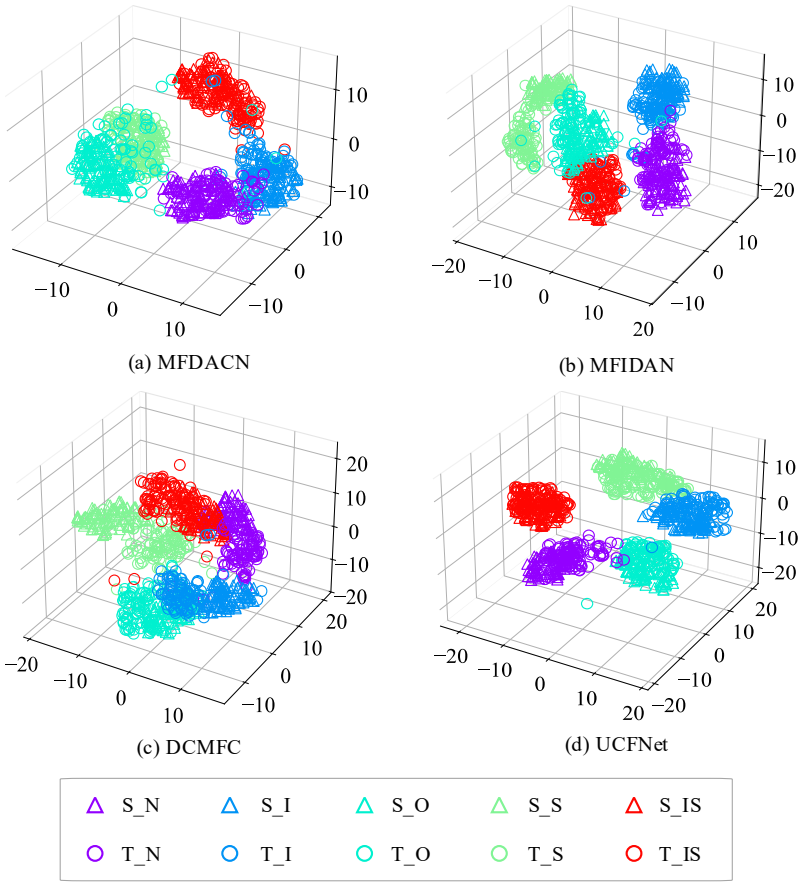


Figure 9. Feature visualization of task T5.

3.4. Statistical Test Analysis

To evaluate performance differences, the non-parametric Friedman test [35] was conducted across all tasks. The test yielded a statistical value of 47.15 ($p = 2.02 \times 10^{-6}$),

confirming significant performance disparities among the evaluated methods. A post hoc Wilcoxon signed-rank test with Holm adjustment was subsequently performed, as detailed in Table 8. All pairwise comparisons between UCFNet and baseline algorithms produced a p-value of 1.22×10^{-4}, demonstrating statistically significant superiority. The critical difference diagram in Figure 10 illustrates the rank relationships, where connected horizontal lines indicate no significant difference. UCFNet achieves the best average rank of 1.0000 without any connecting lines, statistically confirming its superior generalization performance under unseen working conditions.

Table 8. Wilcoxon signed-rank test results.

Pairwise Comparison	p-Value	Pairwise Comparison	p-Value
UCFNet vs. CCDG	1.22×10^{-4}	CCDG vs. MFIDAN	1.22×10^{-4}
UCFNet vs. MFDACN	1.22×10^{-4}	CCDG vs. DCMFC	1.22×10^{-4}
UCFNet vs. MFIDAN	1.22×10^{-4}	MFDACN vs. MFIDAN	5.25×10^{-3}
UCFNet vs. DCMFC	1.22×10^{-4}	MFDACN vs. DCMFC	3.91×10^{-1}
CCDG vs. MFDACN	1.22×10^{-4}	MFIDAN vs. DCMFC	5.25×10^{-3}

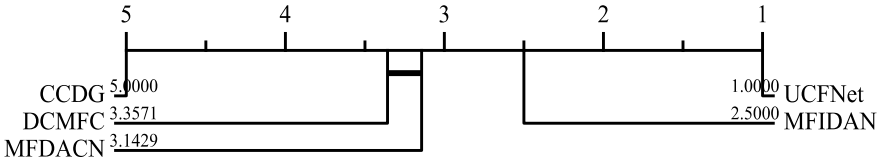


Figure 10. Critical difference diagram.

3.5. Domain Discrepancy Analysis

To measure cross-domain feature alignment, the proxy A-distance (PAD) [36] was calculated on the learned representations. Lower PAD values indicate smaller cross-domain discrepancies, facilitating feature generalization to unseen working conditions. Comparative PAD results between UCFNet and DCMFC across all 14 diagnostic tasks are depicted in Figure 11.

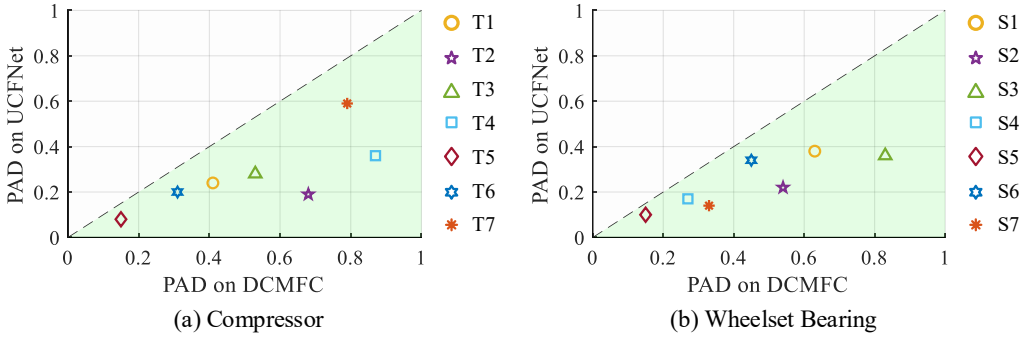


Figure 11. Feature distribution discrepancy between source and target domains.

Across all evaluation tasks on both datasets, the PAD values of UCFNet fall strictly below the diagonal line, outperforming the corresponding DCMFC metrics. On the compressor dataset, UCFNet reduces the domain shift clearly in tasks T1, T2, and T4. Similar improvements are observed in tasks S2 and S3 on the wheelset bearing dataset. These quantitative metrics confirm that UCFNet effectively minimizes feature distribution discrepancy between source and target domains, thereby enhancing diagnostic performance in unseen working environments.

3.6. Computational Complexity Analysis

To evaluate the practical deployment potential, the computational complexity of UCFNet is analyzed and compared with the representative multi-modal baselines on task T1. As reported in Table 9, UCFNet achieves the lowest parameter number of 0.27 M, which is significantly smaller than MFIDAN and DCMFC. In terms of computational cost, UCFNet requires 22.84 M FLOPs, which is substantially lower than DCMFC and comparable to MFIDAN. Furthermore, the training time per epoch of UCFNet is 1.58 s, which is much faster than MFIDAN. These results demonstrate that UCFNet achieves an excellent trade-off between diagnostic performance and computational efficiency, making it highly suitable for resource-constrained industrial deployment.

Table 9. Computational complexity of the UCFNet on Task T1.

Evaluation Criterion	MFIDAN	DCMFC	UCFNet
Parameters Number/M	0.41	0.84	0.27
FLOPs/M	19.86	63.95	22.84
Training time per epoch/s	2.53	1.32	1.58

4. Conclusions

This paper proposed UCFNet, a multi-modal domain generalization framework for mechanical fault diagnosis under unseen working conditions. The key novelty of this work lies in explicitly coordinating feature extraction, selection, and decision fusion. Specifically, the parallel mixed attention captures domain-invariant representations, the dynamic collaborative gating suppresses condition-induced interference, and the uncertainty-aware feature fusion ensures decision resilience. Experiments on compressor and bearing datasets demonstrate that UCFNet achieves superior diagnostic accuracy and statistical stability compared with state-of-the-art methods.

However, the current framework relies on synchronized multi-modal data and lacks explicit domain-invariance constraints in the training objective. Future work will focus on extending UCFNet to asynchronous or incomplete multi-modal settings, introducing explicit domain-invariance regularization, and validating its effectiveness under more complex conditions.

Author Contributions: Conceptualization, H.R. and Q.S.; methodology, H.R.; software, M.B.; validation, W.L. and C.W.; data curation, Q.S.; writing—original draft preparation, H.R. and M.B.; writing—review and editing, Y.Z.; supervision, C.W.; project administration, W.L. and Y.Z.; funding acquisition, H.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 52605838, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20250970, in part by the Changzhou Leading Innovative Talent Introduction and Cultivation Program under Grant CQ20240109, and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 25KJB580001.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: Authors Weifeng Lian and Chaoyang Wu were employed by the company Changzhou NRB Corporation. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Niola, V.; Savino, S.; Quaremba, G.; Cosenza, C.; Nicolella, A.; Spirito, M. Discriminant Analysis of the Vibrational Behavior of a Gas Micro-Turbine as a Function of Fuel. *Machines* **2022**, *10*, 925. [CrossRef]
2. Su, H.; Xiang, L.; Hu, A. Application of Deep Learning to Fault Diagnosis of Rotating Machineries. *Meas. Sci. Technol.* **2024**, *35*, 042003. [CrossRef]
3. Zhu, Z.; Lei, Y.; Qi, G.; Chai, Y.; Mazur, N.; An, Y.; Huang, X. A Review of the Application of Deep Learning in Intelligent Fault Diagnosis of Rotating Machinery. *Measurement* **2023**, *206*, 112346. [CrossRef]
4. Pan, Z.; Pan, X.; Jiang, F.; Wu, Y.; Meng, Z.; Wang, Y.; Zhao, P.; Lu, W. Remaining Useful Life Prediction of Rolling-Element Bearings Based on Cumulative Transformation and EResNet-KSLSTM Network. *Mech. Syst. Signal Process.* **2025**, *238*, 113251. [CrossRef]
5. Elahi, M.U.; Raouf, I.; Khalid, S.; Ahmad, F.; Kim, H.S. Transfer Learning-Based Health Monitoring of Robotic Rotate Vector Reducer under Variable Working Conditions. *Machines* **2025**, *13*, 60. [CrossRef]
6. Zhang, J.; Pei, G.; Zhu, X.; Gou, X.; Deng, L.; Gao, L.; Liu, Z.; Ni, Q.; Lin, J. Diesel Engine Fault Diagnosis for Multiple Industrial Scenarios Based on Transfer Learning. *Measurement* **2024**, *228*, 114338. [CrossRef]
7. An, Y.; Zhang, K.; Chai, Y.; Liu, Q.; Huang, X. Domain Adaptation Network Base on Contrastive Learning for Bearings Fault Diagnosis under Variable Working Conditions. *Expert Syst. Appl.* **2023**, *212*, 118802. [CrossRef]
8. Li, X.; Kang, B.; Tang, J.; Li, Q.; Wang, T.; Zhang, M. A Single-Device Environment-Adaptive Mixed Reality Framework for Real-Time Industrial Fault Diagnosis. *J. Dyn. Monit. Diagn.* **2026**, *5*, 64–73. [CrossRef]
9. Qin, Y.; Qian, Q.; Luo, J.; Pu, H. Deep Joint Distribution Alignment: A Novel Enhanced-Domain Adaptation Mechanism for Fault Transfer Diagnosis. *IEEE Trans. Cybern.* **2022**, *53*, 3128–3138. [CrossRef]
10. Li, R.; Li, S.; Xu, K.; Li, X.; Lu, J.; Zeng, M.; Li, M.; Du, J. Adversarial Domain Adaptation of Asymmetric Mapping with CORAL Alignment for Intelligent Fault Diagnosis. *Meas. Sci. Technol.* **2022**, *33*, 055101. [CrossRef]
11. Zhang, S.; Su, L.; Gu, J.; Li, K.; Zhou, L.; Pecht, M. Rotating Machinery Fault Detection and Diagnosis Based on Deep Domain Adaptation: A Survey. *Chin. J. Aeronaut.* **2023**, *36*, 45–74. [CrossRef]
12. Wang, X.; Jiang, H.; Mu, M.; Dong, Y. A Dynamic Collaborative Adversarial Domain Adaptation Network for Unsupervised Rotating Machinery Fault Diagnosis. *Reliab. Eng. Syst. Saf.* **2025**, *255*, 110662. [CrossRef]
13. Chen, Y.; Zhang, D.; Yan, R.; Xie, M. Applications of Domain Generalization to Machine Fault Diagnosis: A Survey. *IEEE/CAA J. Autom. Sin.* **2025**, *12*, 1963–1984. [CrossRef]
14. Zhao, S.; Yang, Y.; Tan, M.; Zhu, J.; Sun, X.; Jing, H. Multi-Level Confusion Augmentation and Joint Domain Generalization for Fault Diagnosis under Unseen Conditions. *Meas. Sci. Technol.* **2026**, *37*, 066103. [CrossRef]
15. Ragab, M.; Chen, Z.; Zhang, W.; Eldele, E.; Wu, M.; Kwoh, C.K.; Li, X. Conditional Contrastive Domain Generalization for Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3506912. [CrossRef]
16. Rohlfs, C. Generalization in Neural Networks: A Broad Survey. *Neurocomputing* **2025**, *611*, 128701. [CrossRef]
17. Chen, G.; Ye, J.; Li, D.; Hu, L.; Wang, Z.; Zi, M.; Liang, C.; Zhang, J. Domain Generalization for Bearing Fault Diagnosis via Meta-Learning with Gradient Alignment and Data Augmentation. *Machines* **2025**, *13*, 960. [CrossRef]
18. Song, Q.; Jiang, X.; Yang, S.-K.; Ren, H.; Liu, J.; Zhu, Z. Knowledge Regroupment and Preference Calibration Framework for Unpredicted Fault Diagnosis under Unknown Working Conditions. *Reliab. Eng. Syst. Saf.* **2026**, *266*, 111775. [CrossRef]
19. Liu, Y.; Cheng, W.; Wen, W.; Fang, Q.; Wu, H. A Novel Domain Generalization Framework for Fault Diagnosis of Rotating Machinery Based on Causal Representation Learning and Causal Feature Identification. *Eng. Appl. Artif. Intell.* **2026**, *164*, 113201. [CrossRef]
20. Wang, J.; Ren, H.; Shen, C.; Huang, W.; Zhu, Z. Multi-Scale Style Generative and Adversarial Contrastive Networks for Single Domain Generalization Fault Diagnosis. *Reliab. Eng. Syst. Saf.* **2024**, *243*, 109879. [CrossRef]
21. Li, G.; Atoui, M.A.; Li, X. Dual Adversarial and Contrastive Network for Single-Source Domain Generalization in Fault Diagnosis. *Adv. Eng. Inf.* **2025**, *65*, 103140. [CrossRef]
22. Huang, K.; Ren, Z.; Zhu, L.; Lin, T.; Zhu, Y.; Zeng, L.; Wan, J. Intra-Domain Self Generalization Network for Intelligent Fault Diagnosis of Bearings under Unseen Working Conditions. *Adv. Eng. Inf.* **2025**, *64*, 102997. [CrossRef]
23. Lv, Y.; Hao, J.; Tang, M.; Wu, J. Multimodal Data Fusion-Based Intelligent Fault Diagnosis for Ship Rotating Machinery: Status Quo and Perspectives. *Eng. Appl. Artif. Intell.* **2025**, *160*, 111767. [CrossRef]
24. Tang, L.; Wu, X.; Wang, D.; Liu, X. A Comparative Experimental Study of Vibration and Acoustic Emission on Fault Diagnosis of Low-Speed Bearing. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3529211. [CrossRef]
25. Chu, Z.; Xing, S.; Han, B.; Wang, J. A Novel Cross-Domain Mechanical Fault Diagnosis Method Fusing Acoustic and Vibration Signals by Vision Transformer. *Sensors* **2024**, *24*, 5120. [CrossRef]
26. Pan, Z.; Zhang, Z.; Zhao, P.; Meng, Z.; Wang, Y.; Zheng, Y. Multi-Sensor Signal Fusion Method for Rolling Bearing Based on the Standard Relative Mean–Variance Value and Random Weighting Algorithm. *Measurement* **2024**, *235*, 114823. [CrossRef]

27. Zhou, L.; Wang, P.; Zhang, C.; Qu, X.; Gao, C.; Xie, Y. Multi-Mode Fusion BP Neural Network Model with Vibration and Acoustic Emission Signals for Process Pipeline Crack Location. *Ocean Eng*. **2022**, *264*, 112384. [CrossRef]
28. Zaman, W.; Siddique, M.F.; Khan, S.U.; Kim, J.M. A New Dual-Input CNN for Multimodal Fault Classification Using Acoustic Emission and Vibration Signals. *Eng. Fail. Anal*. **2025**, *179*, 109787. [CrossRef]
29. Li, N.; Xu, C.; Xue, B.; Ji, Z.; Wang, W.; Hu, Q.; Chen, X. Multi-Scale CNN with Efficient Channel Attention-Based Feature Fusion for Robust Diagnosis of Rolling Bearing Incipient Fault Using Vibration and Acoustic Emission Signals. *Measurement* **2026**, *279*, 121710. [CrossRef]
30. Hou, D.; Qi, H.; Luo, H.; Wang, C.; Yang, J. Comparative Study on the Use of Acoustic Emission and Vibration Analyses for the Bearing Fault Diagnosis of High-Speed Trains. *Struct. Health Monit*. **2022**, *21*, 1518–1540. [CrossRef]
31. Liu, L.; Shao, H.; Xiao, Y.; Wang, H.; Zhou, C.; Li, Z.; Liu, X. Reliability-Aware Multimodal Adaptive Fusion for Machinery Fault Diagnosis under Incomplete Data. *Reliab. Eng. Syst. Saf*. **2026**, *277*, 113355. [CrossRef]
32. Hao, J.; Lv, Y.; Liu, J.; Liu, Y.C. Dynamic Weighted Multimodal Fusion for Fault Diagnosis of Marine Rotating Machinery under Noisy and Low-Sample Conditions. *Ocean Eng*. **2025**, *339*, 122082. [CrossRef]
33. Han, Y.; Liu, T.; Li, K. Fault Diagnosis of Wind Turbine Blades under Wide-Weather Multi-Operating Conditions Based on Multi-Modal Information Fusion and Deep Learning. *Struct. Health Monit*. **2026**, *25*, 3193–3217. [CrossRef]
34. Ren, H.; Wang, J.; Zhu, Z.; Shi, J.; Huang, W. Domain Fuzzy Generalization Networks for Semi-Supervised Intelligent Fault Diagnosis under Unseen Working Conditions. *Mech. Syst. Signal Process*. **2023**, *200*, 110579. [CrossRef]
35. Friedman, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *J. Am. Stat. Assoc*. **1937**, *32*, 675–701. [CrossRef]
36. Song, Q.; Jiang, X.; Liu, J.; Shi, J.; Zhu, Z. Contrast-Assisted Domain-Specificity-Removal Network for Semi-Supervised Generalization Fault Diagnosis. *IEEE Trans. Neural Netw. Learn. Syst*. **2024**, *36*, 5403–5416. [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.