Contrastive Regularization-Based Domain Generalization for Rolling Bearing Fault Diagnosis under Unseen Rotational-speed Conditions

Xu Zhang[1], XingYan Lu[1], Yuan Gao[1], Xinwei Zhao[1,] *, Jinxin Wang[2,3], Lei Su[1], Ke Li[1], Hang Ran[4]

1. School of Intelligent Manufacturing, Jiangnan University, Wuxi 214122, China

2. School of Safety Engineering, China University of Mining and Technology, Xuzhou 221116, China

3. Xuzhou Mining Business Group Co., Ltd., Xuzhou 221111, China

4. Chongqing Tsingshan Industrial, Chongqing 402760, China

*Corresponding author: xinweizhao@jiangnan.edu.cn

Received 19 August , 2026; Revised 2 October,2026; Accepted 4 November, 2026; Published online 5 November , 2026.

Abstract：To address the performance degradation of rolling bearing fault diagnosis under varying operating conditions, where data from unseen conditions are unavailable during training, a contrastive regularization-based domain generalization method is proposed. First, omni-dimensional dynamic convolution is employed to adaptively characterize vibration-feature variations across multiple source domains and extract discriminative fault representations. Second, class-specific feature perturbation is introduced to enrich intra-class variations, and a class-conditional domain-discrepancy-aware interpolation mechanism is developed to quantify the feature shifts of different fault classes across source domains and adaptively regulate the Beta interpolation distribution, thereby expanding the class-conditional feature space according to actual cross-condition variations. Subsequently, self-supervised contrastive regularization is imposed at both the feature and classifier-output levels to enhance the consistency of same-class representations and classification responses, while the classification objective preserves inter-class discriminability. Finally, experiments on rolling bearing domain generalization tasks demonstrate that the proposed method achieves an average diagnostic accuracy of 97.53%, indicating its effectiveness for fault diagnosis under unseen rotational-speed conditions.

Keywords: Rolling bearing; Domain generalization; Omni-dimensional dynamic convolution;

Contrastive regularization

I. Introduction

With the increasing application of artificial intelligence technologies, manufacturing industry is gradually

advancing toward intelligent manufacturing [1]. As critical components in manufacturing equipment, the operational reliability of rolling bearings is essential to the overall performance of mechanical systems. Artificial intelligence-based fault diagnosis of rolling bearings can reduce the risk of catastrophic accidents [2]. The application of such technologies not only accelerates the intelligent transformation of manufacturing but also promotes advances in equipment management and maintenance practices, thereby further enhancing the international competitiveness of manufacturing industry.

In recent years, deep learning has achieved remarkable progress in rolling bearing fault diagnosis owing to its automatic feature extraction capability and excellent performance in processing complex high-dimensional data [3-6]. However, under complex and varying operating conditions, the dynamic response characteristics of rolling bearing vibration signals vary significantly across different operating conditions, resulting in degraded diagnostic performance of deep learning models [7]. Domain adaptation can effectively align feature spaces across different operating domains and substantially alleviate distribution discrepancies among operating conditions, making it a mainstream approach for cross-condition fault diagnosis. Li et al. [8] jointly constrained the marginal and conditional distributions to enhance feature transfer across different operating conditions. Huang et al. [9] developed a multi-stage transfer framework that sequentially employs feature disentanglement, adversarial adaptation, and pseudo-label distillation to reduce distribution discrepancies under large domain shifts. Jin et al. [10] introduced multi-channel multi-scale feature extraction and domain-adversarial learning to improve model adaptability across varying operating conditions. Sun et al. [11] designed a universal domain adaptation strategy for noisy cross-domain scenarios, improving transfer performance by identifying and suppressing the interference of noisy samples. Bai et al. [12] extended single-target domain adaptation to multiple target domains, thereby enhancing diagnostic performance across diverse unseen rotational-speed conditions. Wang et al. [13] employed a domain-adaptive meta-learning mechanism to enable rapid adaptation to new operating conditions with limited samples. Fang et al. [14] leveraged digital twins to facilitate knowledge transfer between virtual and physical domains, enabling cross-space fault diagnosis of roller bearings. Zeng et al. [15] further reduced the feature distribution discrepancy between the source and target domains through pseudo-label refinement and maximum mean discrepancy constraints. However, the substantial potential of domain adaptation for addressing variable operating conditions relies on the assumption that target-domain data are accessible for distribution alignment. This assumption is difficult to satisfy in practical engineering applications, because domain adaptation typically requires a new model to be trained separately for each newly encountered operating condition, resulting in considerable time costs and limited adaptability to dynamically changing operating environments [16]. Therefore, an intelligent diagnostic method capable of generalizing to unseen domains without access to prior target-domain data is urgently needed.

Recent studies have begun to explore domain generalization to address the scarcity of prior knowledge about the marginal distributions of target-domain operating conditions. Domain generalization aims to train models that can generalize across different domains without direct access to target-domain data. Existing methods can be broadly categorized into three types: domain-invariant feature learning, causal disentanglement learning, and meta-learning.

For domain-invariant feature learning, Huang et al. [17] constrained feature representations from the perspectives of class-level invariance and cross-domain invariance to enhance the learning of common fault features across different source domains. Li et al. [18] employed MixStyle to diversify feature styles and dynamically weighted invariant risk minimization to reduce distribution discrepancies among source domains. Song et al. [19] utilized multi-Gaussian attention and covariance constraints to extract fault features with cross-condition stability from a single source domain. Huo et al. [20] constrained the feature space in terms of feature invariance and intra-class compactness, thereby improving fault discrimination under unseen rotational-speed conditions. For causal disentanglement learning, Guo et al. [21] separated fault-related causal features from domain-related non-causal features according to the fault signal generation mechanism, reducing the influence of operating-condition variations on diagnostic results. Sun et al. [22] introduced causal gating and causal intervention mechanisms into a capsule network to strengthen stable causal fault representations across domains. For domain augmentation and meta-learning, Chen et al. [23] generated diverse training tasks through data augmentation and employed a gradient-alignment-based meta-learning strategy to improve model adaptability to distribution shifts. Xing et al. [24] developed a meta-learning-based speed-invariant prototypical network that learns stable class prototypes across rotational speeds, enabling rolling bearing fault diagnosis under unseen speed conditions. The key to domain generalization lies in sufficiently exploiting discriminative features with cross-domain stability from limited source domains. Although the aforementioned methods have improved model generalization to unseen rotational-speed conditions to some extent, several limitations remain. (1) The feature extraction processes of some methods lack the ability to dynamically respond to variations in input features, making it difficult to adaptively adjust feature extraction according to changes in the dynamic response characteristics of vibration signals under different operating conditions, thereby limiting the effective representation of cross-condition discriminative features. (2) Although existing methods have improved cross-condition generalization through invariance constraints and feature augmentation, some methods use the same interpolation distribution for all fault classes without considering their different cross-domain feature variations. As a result, the augmented features may not adequately reflect the differences in feature distributions among fault classes across operating conditions.

Based on the above analysis, this study proposes a contrastive regularization-based domain generalization method for rolling bearing fault diagnosis under unseen rotational-speed conditions. The main contributions are as follows:

(1) A contrastive-regularized domain generalization framework is developed for rolling bearing fault diagnosis under unseen rotational-speed conditions, in which omni-dimensional dynamic convolution is employed to adaptively capture vibration features across different source rotational-speed conditions.

(2) A class-conditional domain-discrepancy-aware interpolation mechanism is proposed. By quantifying the cross-domain shift of each fault class and adaptively regulating the Beta interpolation distribution, the method performs class-dependent feature-space expansion according to actual operating-condition sensitivity, instead of applying uniform interpolation to all classes.

(3) A dual-level self-supervised contrastive regularization strategy is

constructed by jointly imposing consistency constraints in the feature and classifier-output spaces. Combined with the classification objective, it enhances the stability of same-class representations and classification responses while preserving inter-class discriminability, thereby improving generalization to unseen rotational-speed conditions.

II. Methodology

To address the insufficient diagnostic performance caused by complex and varying operating conditions in practical engineering, where prior knowledge of unseen rotational-speed conditions is difficult to obtain, this study proposes a contrastive regularization-based domain generalization method for rolling bearing fault diagnosis. The network architecture of the proposed method is shown in Fig. 1.

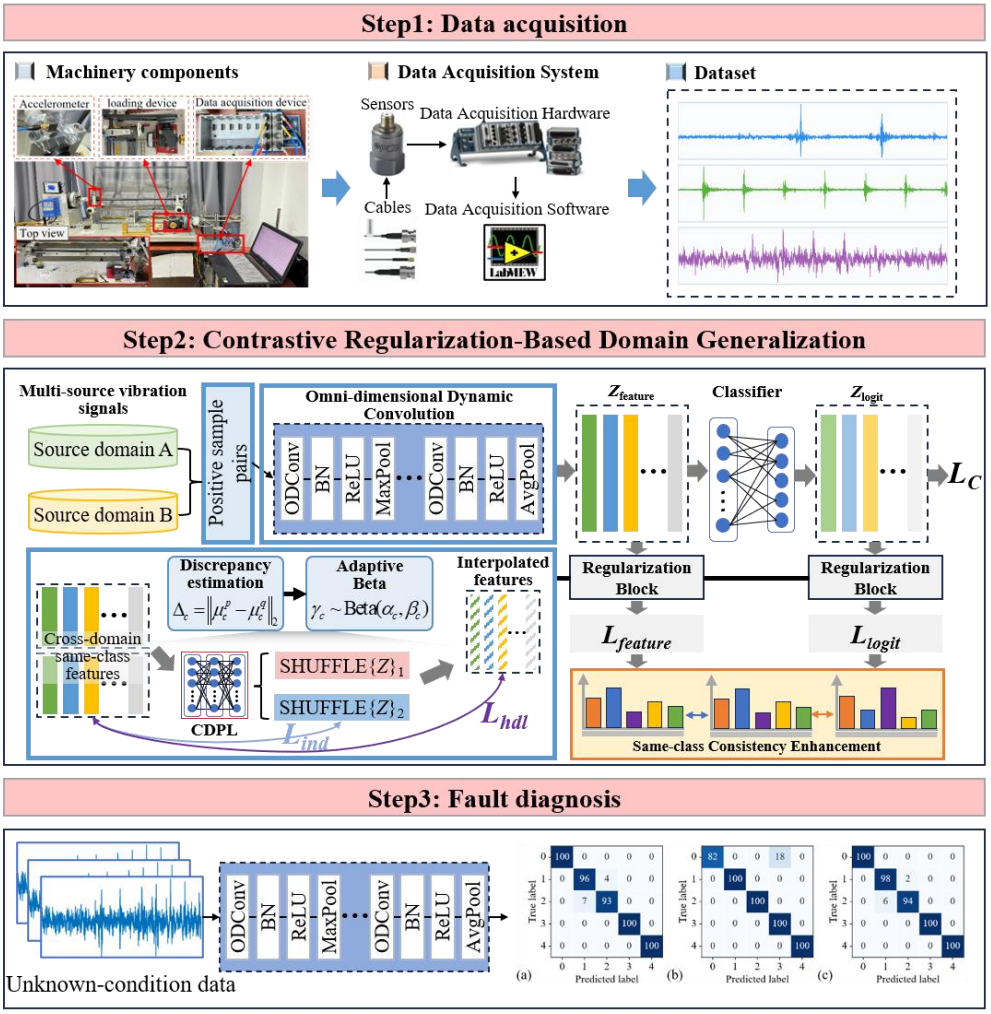


Fig. 1. Architecture of the proposed method

2.1 Omni-dimensional dynamic feature extraction

Omni-dimensional Dynamic Convolution (ODConv) [25] is a variant of the convolutional architecture that significantly enhances feature representation through a multi-dimensional attention mechanism, as illustrated in Fig. 2. ODConv learns distinct attention weights across all dimensions of the feature space, enabling dynamic weighting of different components of the input signals.

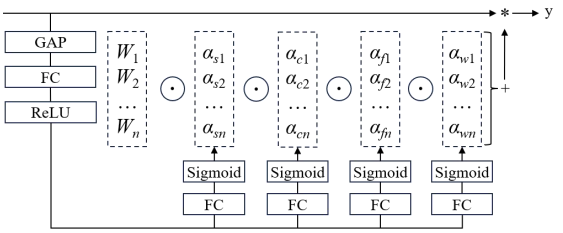


Fig. 2. Omni-dimensional dynamic convolution.

This mechanism allows more effective extraction and representation of discriminative features
from multiple source domains. ODConv is defined as follows:

$$y = \left(\alpha_{w1} \ulcorner \alpha_{f1} \ulcorner \alpha_{c1} \ulcorner \alpha_{s1} \ulcorner W_1 + \ldots \right) * x \quad (1)$$

Where α_{wi} denotes the attention scalar assigned to the convolution kernel W_i , which is consistent with the attention scalar in Eq. (1). α_{si}, α_{ci} , and α_{fi} denote three newly introduced attention terms, which are computed along the spatial, input-channel, and output-channel dimensions of W_i, respectively. ω represents the multiplication operation across different dimensions of the kernel space. The attention weights α_{si}, α_{ci} , and α_{fi} are calculated by the attention module $\pi_{wi}(x)$. For the convolution kernel W_i, α_{si} assigns different attention scalars to different spatial positions of the kernel; α_{ci} assigns different attention scalars to the input features of each W_{mi} kernel; α_{fi} assigns different attention scalars to the output features of each W_{mi} kernel; and α_{wi} assigns an attention scalar to the entire convolution kernel.

2.2 Self-supervised contrastive regularization

Variations in operating conditions can alter the dynamic responses of rolling bearing vibration signals and induce feature distribution shifts for the same fault class. When only the classification loss is used for supervised training, the model may capture source-domain-specific information, resulting in insufficient intra-class consistency across operating conditions. To address this issue, self-supervised contrastive regularization is introduced by integrating class-specific feature perturbation, class-conditional domain-discrepancy-aware interpolation, and feature- and output-level consistency constraints to enhance the cross-condition stability of fault representations.

2.2.1 Class-specific feature perturbation

Samples belonging to the same fault class share consistent fault semantics, whereas their feature representations may vary with operating conditions. Directly enforcing the similarity of the original same-class features may cause the model to overfit the limited feature patterns observed in the source domains. Therefore, feature perturbation is introduced while preserving class semantics, enabling the model to learn stable representations from more diverse intra-class variations.

For an input sample x_i , its latent representation is extracted by the feature extractor $f_\theta(\cdot)$:

$$z_i = f_\theta(x_i) \quad (2)$$

The features within a mini-batch are grouped according to their class labels and randomly shuffled within each class to construct positive pairs. Let $\pi_c(\cdot)$ denote a random permutation within class c. A class-specific domain perturbation layer (CDPL) is then employed to asymmetrically transform the paired feature:

$$\tilde{z}_i^c = f_{\mathrm{CDPL}}\left(z_{\pi_c(i)}^c \right) \quad (3)$$

Where $\tilde{z}_i^c$ denotes the perturbed same-class feature. The individualized dissimilarity loss is further constructed using the Euclidean distance between the original and perturbed features:

$$L_{\mathrm{ind}} = \frac{1}{N} \sum_{i=1}^{N} \left\| z_i^c - \tilde{z}_i^c \right\|_2^2 \quad (4)$$

By minimizing L_{ind} , same-class features are encouraged to maintain consistent representations under feature perturbations,

thereby reducing the sensitivity of the model to condition-dependent local variations.

2.2.2 Class-Conditional Domain-Discrepancy-Aware Same-Class Feature Interpolation

Different from conventional feature interpolation or Mixup-like augmentation strategies using a fixed sampling distribution, the proposed DDA Interpolation explicitly estimates class-conditional cross-domain discrepancies and uses them to adaptively regulate the Beta interpolation distribution. Since different fault classes may exhibit different degrees of cross-domain feature variation, this mechanism adjusts the interpolation proportions according to the observed discrepancy of each class, rather than applying the same interpolation distribution to all fault categories.

Class-specific feature perturbation enriches intra-class variations; however, the limited source conditions may still cause the model to rely on source-specific feature patterns. Interpolating perturbed same-class features across different source domains generates additional representations that combine feature characteristics from different operating conditions, while the discrepancy-aware interpolation strategy adjusts these combinations according to class-dependent cross-domain variations. Combined with subsequent feature- and output-level consistency regularization, these augmented representations encourage the model to maintain stable fault features and classification responses under different feature variations, thereby reducing its dependence on specific source conditions and improving generalization to unseen rotational-speed conditions.

For class c in source domains D_p and D_q, the class-conditional domain discrepancy is defined as the Euclidean distance between their class centers μ_c^p and μ_c^q:

$$\Delta_c = \left\| \mu_c^p - \mu_c^q \right\|_2 \tag{5}$$

A larger Δ_c indicates that the corresponding fault class is more sensitive to operating-condition variations. The discrepancy is normalized to obtain the domain-discrepancy-aware coefficient $r_c = \dfrac{\Delta_c}{\max_{k\in\mathbb{C}} \Delta_k + \varepsilon}$. The parameters of the Beta distribution are then adaptively adjusted according to r_c:

$$\alpha_c = \alpha_0 + \kappa r_c,\ \beta_c = \beta_0 + \kappa r_c,\ \gamma_c \sim \mathrm{Beta}(\alpha_c, \beta_c) \tag{6}$$

Where $\alpha_0 = \beta_0$ denote the basic distribution parameters, κ controls the influence of the domain discrepancy, and γ_c is the class-specific interpolation coefficient. Two perturbed features u_i^c and u_j^c with the same class label but from different source domains are then selected and interpolated as

$$\overline{u}_i^c = \gamma_c u_i^c + (1-\gamma_c) u_j^c,\ y_i = y_j = c,\ d_i \neq d_j \tag{7}$$

As the cross-domain discrepancy increases, the symmetric Beta distribution becomes more concentrated around 0.5, encouraging the generated features to cover the intermediate region between the two source-domain distributions. For classes with relatively small domain shifts, greater interpolation diversity is retained. In this manner, the feature expansion process is adaptively matched to the actual cross-condition shift of each fault class. The heterogeneous dissimilarity loss is further defined between the original and interpolated representations:

$$L_{\mathrm{hdl}} = \frac{1}{N} \sum_{i=1}^{N} \left\| z_i^c - \overline{u}_i^c \right\|_2^2 \tag{8}$$

By minimizing L_{hdl}, the generated intermediate representations are encouraged to preserve the original class semantics while enriching the class-conditional feature distribution, thereby improving the coverage of potential feature variations under unseen conditions.

2.2.3 Feature- and Output-Level Dual Regularization

The class-specific perturbation and domain-discrepancy-aware interpolation generate feature representations with unchanged class semantics but different distribution characteristics. However, consistency in the intermediate feature space alone cannot guarantee stable classification responses. Therefore, consistency constraints are imposed simultaneously in both the feature space and the classifier-output space. At the feature level, the individualized dissimilarity loss constrains the consistency between the original and perturbed features, while the heterogeneous dissimilarity loss constrains the consistency between the original and interpolated features. The feature-level regularization is defined as

$$L_{\text{feature}} = L_{\text{ind}}^{f} + L_{\text{hdl}}^{f} \tag{9}$$

Where L_{ind}^{f} constrains the consistency between the original and perturbed features, while L_{hdl}^{f} constrains the consistency between the original and interpolated features. This regularization reduces intra-class feature variations caused by operating-condition changes.

The original, perturbed, and interpolated features are further fed into the shared classifier, and the corresponding consistency constraints are imposed in the classifier-output space:

$$L_{\text{logit}} = L_{\text{ind}}^{q} + L_{\text{hdl}}^{q} \tag{10}$$

Where L_{ind}^{q} constrains the classification responses of the original and perturbed features, while L_{hdl}^{q} constrains those of the original and interpolated features. Hence, feature-level regularization stabilizes the latent representations, whereas output-level regularization further stabilizes the corresponding classification decisions. Finally, the classification loss and the two regularization terms are jointly optimized:

$$L_{\text{total}} = L_{c} + \lambda_{\text{feature}} L_{\text{feature}} + \lambda_{\text{logit}} L_{\text{logit}} \tag{11}$$

Where L_c preserves inter-class discriminability, while λ_{feature} and λ_{logit} control the contributions of the feature- and output-level regularization terms, respectively. Their joint optimization encourages the model to maintain fault discriminability while enhancing cross-condition consistency, thereby improving its generalization capability under unseen rotational-speed conditions.

From the perspective of rotational-speed generalization, the above procedure progressively reduces the dependence of fault representations on speed-specific variations. CDPL introduces class-preserving perturbations to enrich the intra-class variations observed under the source-speed conditions, while DDA Interpolation further expands the class-conditional feature space according to the actual cross-speed discrepancy of each fault class. The feature- and output-level regularization then constrains these expanded representations to maintain consistent feature structures and classification responses, while the classification objective preserves inter-class discriminability. Consequently, the model is encouraged to learn fault-related representations and decision boundaries that are less dependent on the discrete source-speed distributions, thereby improving its ability to accommodate new feature shifts under unseen rotational-speed conditions.

III. Experimental analysis

3.1 Experimental data

The experimental data were collected from a rolling bearing fault simulation test rig developed by our research group, as shown in Fig. 3. The test rig mainly consists of a frequency converter, a motor, a faulty bearing, a coupling, and a load. An acceleration sensor was used to collect rolling bearing vibration data under three operating conditions at a sampling frequency of 25.6 kHz. The three operating conditions, denoted as A, B, and C, correspond to rotational speeds of 1600 rpm, 1800 rpm, and 2000 rpm, respectively. The dataset contains five health states, including inner race fault, combined inner and outer race

fault, outer race fault, rolling element fault, and normal condition, corresponding to labels 0, 1, 2, 3, and 4, respectively. Each sample contains 2048 data points, and the dataset was divided into training and test sets at a ratio of 8:2.

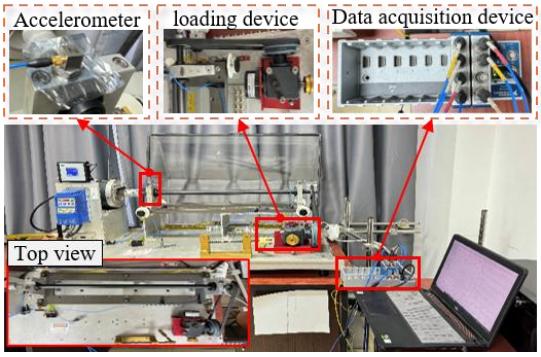


Fig . 3. Bearing fault test rig

3.2 Parameter settings

The parameters of the proposed method are set as follows. The initial learning rate is 0.0001, and a StepLR learning-rate scheduling strategy with a step size of 10 is adopted. The Adam optimizer is used for model optimization. The number of training epochs and batch size are set to 50 and 64, respectively. For the self-supervised contrastive regularization, the feature-level and output-level loss weights λ_{feature} and λ_{logit} are set to 0.3 and 1.0, respectively. The base parameters of the Beta distribution are set to $\alpha_0 = \beta_0 = 0.5$. All experiments are implemented using PyTorch 1.10 and conducted on a computer equipped with an Intel Core i7-11800H CPU and an NVIDIA RTX 3060 Ti GPU. The detailed model parameters of the proposed method are listed in Table 1.

Table 1. Model Parameter

Layers	Operation	Size/Stride
Layer 1	ODConv-BN-ReLU	32×7×1/1
Layer 2	MaxPool	2/2
Layer 3	ODConv-BN-ReLU	64×3×1/1
Layer 4	MaxPool	2/2
Layer 5	ODConv-BN-ReLU	128×3×1/1
Layer 6	MaxPool	2/2
Layer 7	ODConv-BN-ReLU	256×3×1/1
Layer 8	AvgPool	-
Layer 9	Fully-connected layer	5

Table 2. Domain generalization diagnostic task

Task	Training conditions	Test condition	Training samples	Test samples
R1	A+B	C	400×5	100×5
R2	A+C	B	400×5	100×5
R3	B+C	A	400×5	100×5

To ensure a fair comparison, all baseline methods were evaluated under the same domain generalization protocol. For each task, only the two source operating conditions were used for training, while the remaining condition was completely excluded from training and used only as the unseen target domain for testing.

(1) AGFCN[4]: Samples from the two source operating conditions were jointly used for supervised training, and the trained model was directly evaluated on the unseen target condition.

(2) MKWJDAN[8]: The two source operating conditions were used as the two domains for joint distribution alignment, while the unseen target condition was excluded from both distribution alignment and pseudo-label generation.

(3) OSSMCL[27]: The momentum contrastive encoder was trained using only source-domain samples, and the support set for similarity-based classification was constructed from labeled source-domain data. The unseen target-domain samples were used only for final evaluation.

(4) DANN[28]: The domain discriminator was used to distinguish the two source operating conditions, and adversarial training was performed between these source domains to learn condition-invariant features. No target-domain samples were involved in training.

Based on the three operating conditions of the rolling bearing, three domain generalization fault diagnosis tasks were designed, as shown in Table 2.

3.3 Experimental analysis

3.3.1 Comparison with existing methods

To verify the effectiveness of the proposed method in domain generalization fault diagnosis tasks, comparative experiments were conducted against the selected baseline methods. Owing to the characteristics of domain generalization, target-domain data are unavailable during model training. Therefore, to ensure a fair comparison, all transfer strategies were applied only between the two source domains, while data from the unseen target operating condition were used exclusively for testing. All experiments were independently repeated five times, and the reported results are presented as the mean ± standard deviation. The results are presented in Table 3 and Fig. 4. Table 3 reports the diagnostic accuracies of different methods on the three domain generalization tasks, while Fig. 4 shows the accuracy curves of different methods over the training epochs for task R2. The target-domain test set was evaluated after each training epoch only to visualize the evolution of generalization performance.

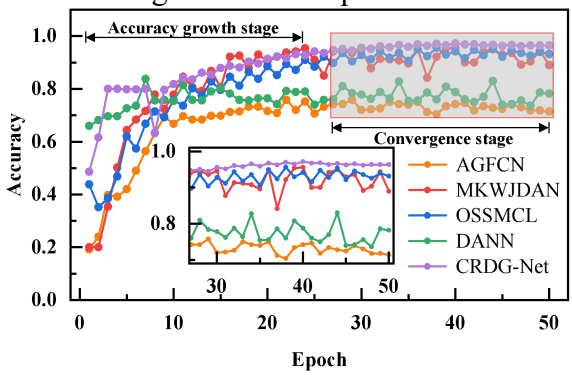


Fig . 4. Accuracy iteration curves of various methods for R2 diagnostic task

Table 3. Domain generalization fault diagnosis results

Methods	Task			
	R1	R2	R3	Average
AGFCN	71.44 ± 0.48	78.56 ± 0.57	66.16 ± 0.57	72.05 ± 0.54
MKWJDAN	94.20 ± 0.32	89.04 ± 0.48	93.56 ± 0.38	92.27 ± 0.39
OSSMCL	95.72 ± 0.23	93.32 ± 0.23	94.80 ± 0.32	94.61 ± 0.26
DANN	80.40 ± 0.32	79.00 ± 0.63	82.72 ± 0.23	80.71 ± 0.39
CRDG-Net	**97.76 ± 0.26**	**96.52 ± 0.27**	**98.32 ± 0.23**	**97.53 ± 0.24**

As shown in Table 3 and Fig. 4 the diagnostic performance of different methods varies significantly under unseen rotational-speed conditions. AGFCN, which does not introduce any cross-domain transfer mechanism, achieves an average diagnostic accuracy of only 72.05%, indicating that distribution discrepancies among different operating conditions can substantially degrade the generalization performance of conventional feature extraction models. By introducing cross-domain adaptation, DANN and MKWJDAN improve the average accuracy to 80.71% and 92.27%, respectively, demonstrating that cross-domain distribution constraints can alleviate feature shifts caused by operating-condition variations to some extent. OSSMCL further enhances feature consistency through contrastive learning and achieves an average accuracy of 94.61%. In comparison, CRDG-Net obtains the highest accuracies of 97.76%, 96.52%, and 98.32% on tasks R1, R2, and R3, respectively, with an average accuracy of 97.53%, outperforming the second-best OSSMCL by 2.92 percentage points. Meanwhile, Fig. 4 shows that all methods exhibit certain fluctuations during the accuracy improvement stage and gradually stabilize as training proceeds, whereas CRDG-Net maintains a higher and more stable diagnostic accuracy during the convergence stage. These results demonstrate that the proposed method can more effectively learn discriminative features with cross-condition stability from limited source domains and enhance the consistency of same-class samples across different source domains, thereby improving diagnostic accuracy and

generalization performance when encountering unseen rotational-speed conditions.

3.3.2 Feature distribution visualization

To further evaluate the feature representation capability of different methods under unseen rotational-speed conditions, t-SNE is employed to project the learned high-dimensional features into a two-dimensional space for visualization.

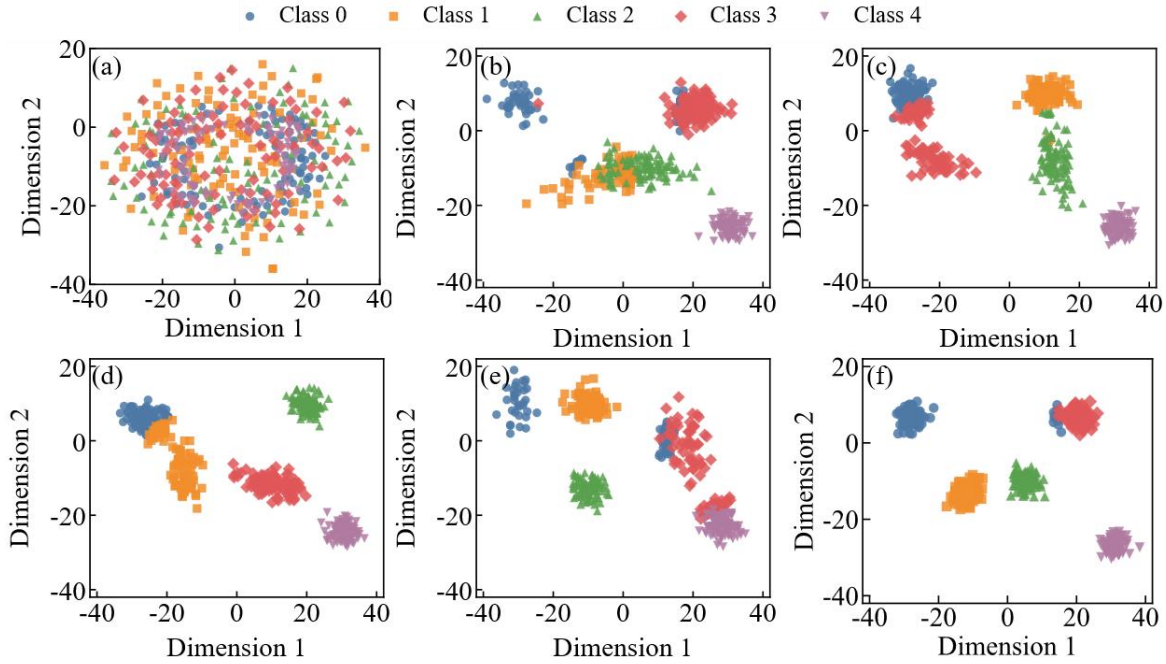


Fig . 5. Feature distribution visualization of different methods. (a) Original input features; (b) AGFCN; (c) MKWJDAN; (d) OSSMCL; (e) DANN; (f) CRDG-Net

As shown in Fig. 5, the raw input features exhibit substantial inter-class overlap, with no clear clustering boundaries among different fault categories. After feature extraction by different models, the separability among fault classes is improved to varying degrees. AGFCN still shows obvious inter-class overlap, while MKWJDAN and OSSMCL further improve class separability, although some neighboring categories remain partially mixed. DANN improves the compactness of several fault clusters, but certain feature distributions remain dispersed. In comparison, CRDG-Net produces more compact intra-class distributions, clearer class boundaries, and larger inter-class margins, demonstrating its ability to learn more domain-invariant and discriminative fault representations under unseen rotational-speed conditions.

3.3.3 Confusion matrix analysis

As shown in Fig. 6, CRDG-Net demonstrates high classification performance across all three domain generalization tasks. In R1, misclassifications mainly occur between Classes 1 and 2, resulting in an overall accuracy of 97.8%. In R2, noticeable confusion appears only for Class 0, with 18 samples misclassified as Class 3, yielding an accuracy of 96.4%. In R3, the class boundaries are clearer, with only a small number of misclassifications between Classes 1 and 2, and the accuracy reaches 98.4%. Overall, the model maintains stable discriminative performance under different unseen rotational-speed conditions, indicating that the proposed method can effectively learn fault representations with cross-condition stability.

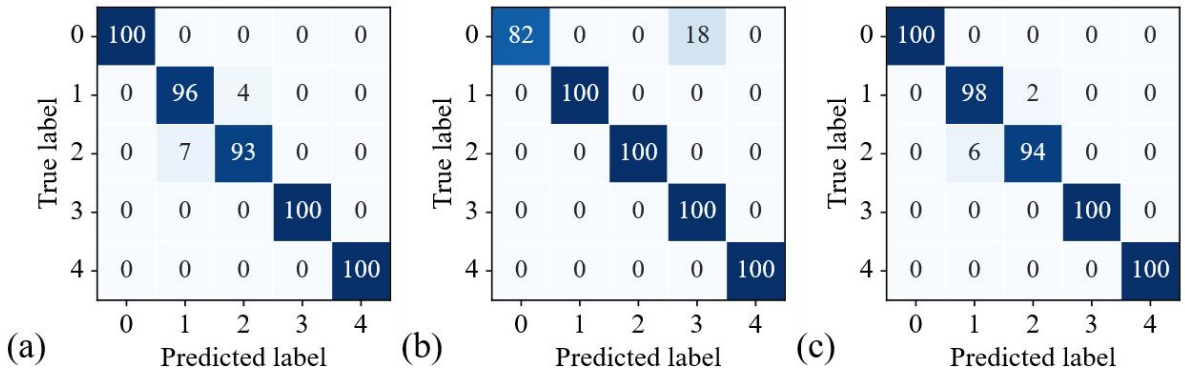


Fig . 6. Confusion matrices of CRDG-Net under different domain generalization tasks. (a) R1; (b) R2; (c) R3

3.3.4 Ablation study

It should be noted that CDPL and DDA Interpolation constitute a sequential feature-expansion process, where CDPL first generates class-preserving perturbed features and DDA Interpolation subsequently performs discrepancy-aware cross-domain interpolation. Therefore, the perturbation pathway is kept unchanged in the following ablation settings, while the effects of DDA Interpolation and the dual-level regularization are separately evaluated. As shown in Table 4, each component contributes to the domain generalization performance. Introducing ODConv improves the accuracies from 67.24%, 65.12%, and 69.32% to 79.96%, 77.20%, and 82.12% on R1–R3, respectively, confirming its effectiveness in enhancing cross-condition feature representation. When DDA Interpolation is not employed, a fixed Beta distribution with $\alpha = \beta = 0.5$ is used as the baseline same-class interpolation strategy. This setting provides a symmetric interpolation distribution independent of class-specific domain discrepancies, allowing the contribution of the proposed discrepancy-aware adaptive interpolation to be evaluated separately. With fixed interpolation and dual-level regularization, the accuracies further increase to 94.48%, 92.56%, and 95.68%, indicating that consistency constraints effectively enhance cross-domain intra-class stability.

Under DDA Interpolation, feature-level regularization alone achieves 96.28%, 94.92%, and 96.84%, while output-level regularization alone achieves 95.72%, 94.36%, and 96.20%. Their combination yields the best performance of 97.76%, 96.52%, and 98.32%, demonstrating the complementary effects of feature- and output-level constraints. Overall, the results confirm that DDA Interpolation, combined with dual-level regularization, enables more effective class-conditional feature expansion and improves generalization to unseen rotational-speed conditions.

Table 4. Results of the Ablation Study

ODConv	DDA Interpolation	Feature-level Reg	Output-level Reg	R1	R2	R3
×	×	×	×	67.24 ± 0.86	65.12 ± 0.95	69.32 ± 0.95
√	×	×	×	79.96 ± 0.89	77.20 ± 0.86	82.12 ± 0.83
√	×	√	√	94.48 ± 0.48	92.56 ± 0.57	95.68 ± 0.48
√	√	√	×	96.28 ± 0.41	94.92 ± 0.58	96.84 ± 0.48
√	√	×	√	95.72 ± 0.41	94.36 ± 0.46	96.20 ± 0.58
√	√	√	√	97.76 ± 0.26	96.52 ± 0.27	98.32 ± 0.23

3.3.5 Feature distance evolution analysis

To further investigate feature evolution during training, the R2 task is taken as an example. The cross-domain intra-class distance is calculated as the average Euclidean distance between same-class feature centers from the two source domains, while the inter-class distance measures the average distance between different class centers. Smaller intra-class distance indicates stronger cross-condition consistency, whereas larger inter-class distance reflects better class separability. As shown in Fig. 7, the intra-class distance decreases from 0.73 to 0.34, while the inter-class distance increases from 0.76 to 1.24 and gradually stabilizes. This indicates that the proposed method enhances cross-condition intra-class consistency while maintaining clear separation among different fault classes.

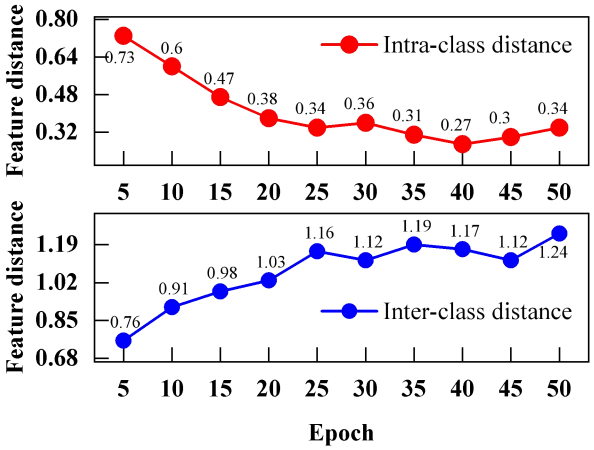


Fig . 7. Evolution of cross-domain intra-class and inter-class distances during training for task R2

3.3.6 Grad-CAM visualization analysis

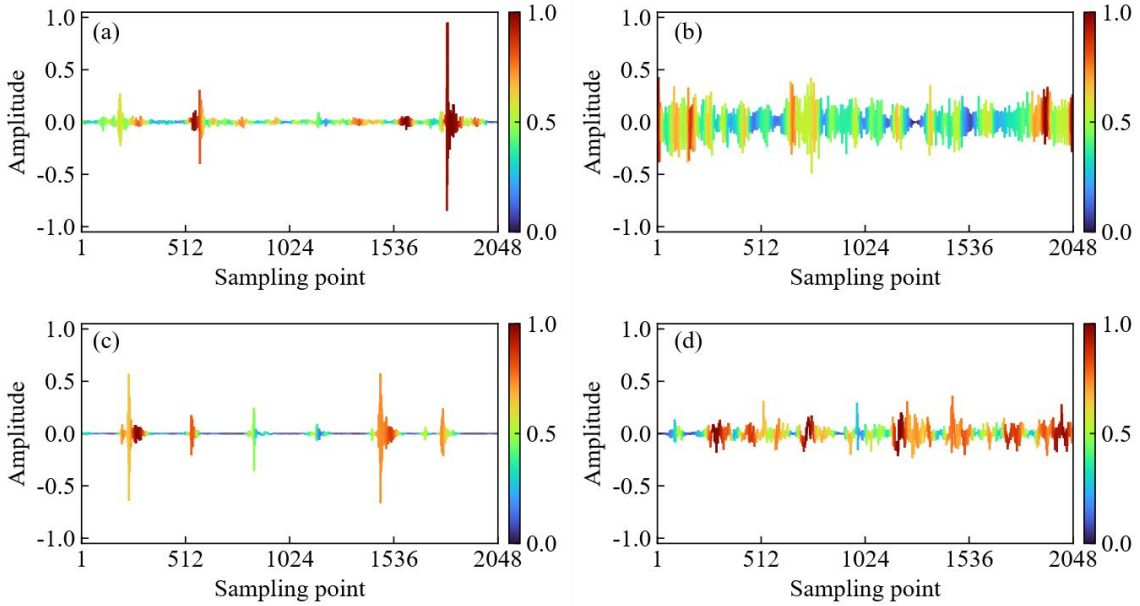


Fig . 8. Grad-CAM visualization of vibration signals under different health conditions. (a) Inner race fault; (b) Normal condition; (c) Outer race fault; (d) Rolling element fault

To further verify the ability of the proposed model to focus on fault-discriminative features under unseen rotational-speed conditions and to analyze the interpretability of its diagnostic decisions, the R2 domain generalization task is taken as an example. Grad-CAM visualization is performed on inner race fault, normal condition, outer race fault, and rolling element fault samples from the unseen test condition B (1800 rpm). As shown in Fig. 8, the model exhibits distinct attention patterns across different health states. For inner race and outer race faults, high-response regions are mainly concentrated around several pronounced local impacts, indicating that the model can effectively capture fault-related impulsive characteristics. For rolling element faults, the

responses are distributed across multiple locally enhanced segments, whereas the normal condition shows a relatively uniform attention pattern. These differences indicate that the model can still preserve class-specific fault response patterns under unseen rotational-speed conditions. CRDG-Net expands intra-class variations across different source conditions through class-conditional feature interpolation and suppresses excessive shifts in feature representations and classification responses through feature- and output-level regularization, thereby encouraging the model to retain fault-related characteristics that are relatively stable across operating conditions. Therefore, the stable attention to locally discriminative vibration regions in the unseen domain is consistent with the learned cross-condition stable fault representations.

3.3.7 Probability Density Analysis

As shown in Fig. 9, the original data exhibit noticeable probability density differences across rotational speeds, mainly reflected in variations in distribution peaks and curve shapes. These differences indicate that changes in rotational speed induce distribution shifts, making the model prone to learning features associated with specific operating conditions and thus limiting its generalization to unseen conditions. After feature learning using the proposed method, the distributions across different rotational speeds exhibit greater overlap, indicating that the proposed method can alleviate cross-condition distribution discrepancies. The key lies in expanding the variations of same-class features under the source conditions, enabling the model to learn from diverse combinations of features across different source conditions. Meanwhile, consistency constraints guide the model to maintain stable feature representations and classification responses, reducing its reliance on speed-specific feature patterns. Consequently, the model can extract fault-related features that remain stable across operating conditions, thereby improving its domain generalization capability to unseen rotational speeds.

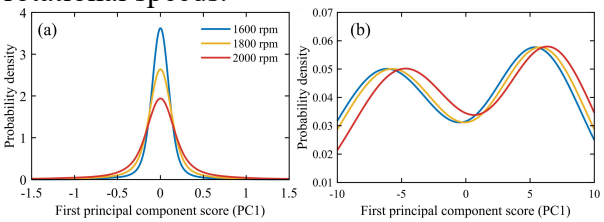


Fig . 9. Probability density distributions at different rotational speeds: (a) original data; (b) features learned by the proposed method.

IV. CONCLUSIONS

This study proposes a contrastive regularization-based domain generalization method for rolling bearing fault diagnosis under unseen rotational-speed conditions, where target-domain information is unavailable during training. First, omni-dimensional dynamic convolution is employed to adaptively extract discriminative vibration features across different source rotational-speed conditions. Second, class-specific feature perturbation and class-conditional domain-discrepancy-aware interpolation are introduced to enrich intra-class variations and adaptively expand the class-conditional feature distribution according to the cross-domain shift of each fault class. Furthermore, self-supervised contrastive regularization is imposed at both the feature and classifier-output levels to enhance the consistency of same-class representations and classification responses while preserving inter-class discriminability. Experimental results on three domain generalization tasks show that the proposed method achieves an average diagnostic accuracy of 97.53%, outperforming the second-best method by 2.92 percentage points. Ablation, feature distribution, feature distance, and Grad-CAM analyses further demonstrate the effectiveness of the proposed method in enhancing cross-condition intra-class consistency, maintaining inter-class

separability, and capturing fault-discriminative local features. The current validation mainly considers rotational-speed variation within a relatively limited speed range on the same bearing test rig. Therefore, future work will extend the rotational-speed range and span, including lower and higher speed conditions, to systematically evaluate the generalization boundary of the proposed method, while also investigating more complex variations involving load, noise, different bearings, and practical industrial scenarios.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments. The authors also gratefully acknowledge the support from the National Natural Science Foundation of China (Grant No. 52305108), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. 3562050205264370/091), the Natural Science Foundation of Jiangsu Province, China (Grant No. BK20221117), and the Basic Research Project of Xuzhou City, China (Grant No. KC22001).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest related to this work.

REFERENCES

1. Zhu Z, Lei Y, Qi G, et al. A review of the application of deep learning in intelligent fault diagnosis of rotating machinery[J]. Measurement, 2023, 206: 112346.
2. Zhang X, Zhao X, Su L, et al. Cross-view multi-stage contrastive learning with label-guided decision boundary calibration for semi-supervised bearing fault diagnosis[J]. Structural Health Monitoring, 2026: 14759217261465874.
3. Mishra R K, Choudhary A, Fatima S, et al. A systematic review on advancement and challenges in multi-fault diagnosis of rotating machines[J]. Engineering Applications of Artificial Intelligence, 2025, 156: 111306.
4. He D, Wu J, Jin Z, et al. AGFCN: A bearing fault diagnosis method for high-speed train bogie under complex working conditions[J]. Reliability Engineering & System Safety, 2025, 258: 110907.
5. Yin Z, Zhang F, Yin C, et al. A bearing fault diagnosis method for sample imbalance[J]. Engineering Applications of Artificial Intelligence, 2025, 157: 111171.
6. Yang G, Li M, Jiao D, et al. Fault Feature Enhancement and Diagnosis Method for UAV Rotor Bearings Based on Amplitude-Modulated Graph Fourier Spectrum[J]. Journal of Dynamics, Monitoring and Diagnostics, 2026, 5(2): 199-207.
7. Yu S, Li X, Lei Y, et al. Multimodal data-enabled large model for machine fault diagnosis towards intelligent operation and maintenance[J]. Journal of Industrial Information Integration, 2026: 101061.
8. Li X, Chen H, Li S, et al. Multi-kernel weighted joint domain adaptation network for cross-condition fault diagnosis of rolling bearings[J]. Reliability Engineering & System Safety, 2025, 261: 111109.
9. Huang K, Ren Z, Zhu L, et al. A three-stage bearing transfer fault diagnosis method for large domain shift scenarios[J]. Reliability Engineering & System Safety, 2025, 254: 110641.
10. Jin Y, Song X, Yang Y, et al. An improved multi-channel and multi-scale domain adversarial neural network for fault diagnosis of the rolling bearing[J]. Control Engineering Practice, 2025, 154: 106120.
11. Sun Y, Song H, Guo L, et al. A transfer learning method: Universal domain adaptation with noisy samples for bearing fault diagnosis[J]. Advanced Engineering Informatics, 2025, 65: 103243.

12. Bai G, Xing T, Sun W, et al. Unsupervised multiple-target domain adaptation for bearing fault diagnosis[J]. Engineering Applications of Artificial Intelligence, 2025, 154: 111063.
13. Wang S, Han W, Jian J, et al. A domain adaptation meta learning method with multilayer convolutional attention for cross-domain bearing fault diagnosis[J]. IEEE Sensors Journal, 2025, 25(8): 14440-14452.
14. Fang C, Chang Q, Hu X, et al. A digital twin-enabled domain adaptation network for cross-space fault diagnosis of roller bearings[J]. Mechanical Systems and Signal Processing, 2025, 236: 113053.
15. Zeng W, Wu J, Chen Y, et al. Cross-domain bearing fault diagnosis using multi-loss deep adaptation with pseudo-label refinement and maximum mean discrepancy-based alignment[J]. Applied Soft Computing, 2026: 114942.
16. Zhao C, Zio E, Shen W. Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study[J]. Reliability Engineering & System Safety, 2024: 109964.
17. Huang K, Ren Z, Lin T, et al. A dual-perspective joint domain generalization network for bearing fault diagnosis under unseen working conditions[J]. Advanced Engineering Informatics, 2025, 66: 103447.
18. Li J, Deng W, Ding J, et al. IBN-MixStyle network with dynamic weighted invariant risk minimization for domain-generalized bearing fault diagnosis[J]. IEEE Transactions on Consumer Electronics, 2025.
19. Song Y, Zhuang Y, Wang D, et al. Fault diagnosis in rolling bearings using multi-gaussian attention and covariance loss for single domain generalization[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-10.
20. Huo T, Gong J, Liu T, et al. A novel domain generalization network integrating invariance and cohesiveness for rolling bearing fault diagnosis[J]. Structural Health Monitoring, 2025: 14759217251384919.
21. Guo C, Sun Y, Yu R, et al. Deep causal disentanglement network with domain generalization for cross-machine bearing fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-16.
22. Sun Z, Zhu X, Liu J, et al. A causality-inspired capsule network for domain generalization in cross-domain fault diagnosis of rolling bearing[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-10.
23. Chen G, Ye J, Li D, et al. Domain generalization for bearing fault diagnosis via meta-learning with gradient alignment and data augmentation[J]. Machines, 2025, 13(10): 960.
24. Xing J, Sun X, Song Y, et al. Speed-invariant prototypical network for rolling bearing fault diagnosis under variable speed conditions[J]. Expert Systems with Applications, 2026: 132111.
25. Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution[J]. arXiv preprint arXiv:2209.07947, 2022.
26. Kim D, Yoo Y, Park S, et al. Selfreg: Self-supervised contrastive regularization for domain generalization[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 9619-9628.
27. Wang W, Li C, Li A, et al. One-stage self-supervised momentum contrastive learning network for open-set cross-domain fault diagnosis[J]. Knowledge-Based Systems, 2023, 275: 110692.
28. Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J]. Journal of machine learning research, 2016, 17(59): 1-35.