Advanced Engineering & Precision Manufacturing

https://doi.org/10.63808/aepm.v2i3.425 Volume 2 , Issue 3

Research Article

A Study on Domain Generalization Methods for Rolling Bearing Fault Diagnosis Based on Mean Differences and Domain Adversarial Learning

Jiabing Zhou[1,*], Xiang Gu[1], Bo Zhang[1], Xinrui Yu[1], Yining Xu[1]

[1]School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 102488, China

[*]**Correspondence:** Jiabing Zhou, 13846403521@163.com

Abstract

To address the issue that intelligent fault diagnosis models for rolling bearings are susceptible to the effects of load and speed variations, as well as shifts in data distribution, when applied across different operating conditions, this paper proposes a fault diagnosis domain generalization method that combines mean difference constraints with domain adversarial learning. Taking raw one-dimensional vibration signals as input, this method first employs a multiscale one-dimensional convolutional network to extract fault impact features and periodic features across different time scales; Subsequently, mean difference constraints between source domains are introduced in the feature space to reduce the offset in the centers of feature distributions across different source operating conditions; simultaneously, a domain-adversarial discriminator based on gradient inversion layers is constructed to weaken the operating condition discriminative information in the feature representations, thereby prompting the model to learn domain-invariant features that possess both fault class discriminability and operating condition insensitivity. Experiments were conducted using two rolling bearing datasets from Case Western Reserve University (CWRU) and Paderborn University (PU) to establish a leave-one-out cross-condition diagnosis task. The results show that the proposed method achieves an average accuracy of 99.08% ± 0.19% on the CWRU dataset, and an average accuracy of 91.86% ± 0.34% on the PU dataset, both outperforming comparison methods such as SVM, 1D-CNN, ResNet1D, TCN, CORAL, MMD, and DANN. Furthermore, a Welch's t-test based on summary statistics indicates that the improvement over the best baseline method is statistically significant. Ablation experiments and parameter sensitivity analysis further validate the effectiveness of multiscale feature extraction, mean difference constraints, and domain adversarial learning in enhancing the model's generalization capability across different operating conditions.

Keywords

Rolling bearings, Fault diagnosis, Domain generalization, Mean difference, Domain adversarial learning

1. Introduction

Rolling bearings are critical support components in rotating machinery such as electric motors, fans, gearboxes, and CNC machine tools; their operational condition directly affects the safety and reliability of industrial systems. Because rolling bearings are subjected to high-speed rotation, alternating loads, and complex environments over extended periods, they are prone to typical failure modes such as inner ring failure, outer ring failure, and rolling element failure. If such failures are not identified in a timely manner, they can lead to equipment shutdowns, production interruptions, or even safety incidents. Therefore, intelligent fault diagnosis methods for rolling bearings have long been a major research focus in the fields of mechanical condition monitoring and predictive maintenance (Case Western Reserve University Bearing Data Center, n.d.; Lei et al., 2020).

Traditional fault diagnosis methods typically rely on manually designed time-domain, frequency-domain, or time-frequency-domain features, combined with classifiers such as Support Vector Machines (SVMs) to perform condition identification (Cortes & Vapnik, 1995). While these methods offer a degree of interpretability, feature design depends on expert experience, and their generalization capabilities are limited under complex operating conditions. In recent years, deep learning methods have demonstrated the ability to automatically learn fault signatures directly from raw vibration signals, achieving high recognition accuracy on publicly available bearing datasets (Zhang et al., 2017; Zhang et al., 2018). Among these, residual networks can mitigate the training degradation issues of deep networks through cross-layer connections (He et al., 2016). Furthermore, the role of vibration statistical features and stochastic modeling methods in bearing health assessment has garnered attention. Bhatti et al. (2024) established a stochastic analysis framework based on rolling bearing vibration features, while Bhatti et al. (2025) further utilized statistical features and stochastic models to predict equipment health status. The aforementioned studies demonstrate that fault diagnosis models must retain discriminative vibration information related to bearing degradation while mitigating distribution shifts caused by operational variations such as load and rotational speed.

However, deep learning methods typically assume that the training and test data follow the same or similar distributions. In real-world industrial settings, equipment load, rotational speed, sampling environments, and sensor mounting locations may vary, leading to significant differences in distribution across different operating conditions. If a model is trained only under a specific, fixed operating condition, its diagnostic performance under unknown conditions may deteriorate significantly. To address the issue of distribution shifts, domain adaptation methods mitigate the differences between the source and target domains by performing feature alignment using unlabeled data from the target domain (Ganin et al., 2016;

Sun & Saenko, 2016). However, in many real-world scenarios, data for unknown target operating conditions is unavailable during the training phase; therefore, domain generalization methods that rely solely on source domain data for training are better suited to engineering applications (Gretton et al., 2012; Muandet et al., 2013).

Existing domain generalization methods primarily enhance a model's generalization ability in unknown domains by focusing on feature distribution alignment, adversarial learning, data augmentation, and causally invariant representations (Chen et al., 2022; Guo et al., 2024; Kim et al., 2024; Li et al., 2017; Ren et al., 2023; Zhang et al., 2023). Among these, distribution discrepancy constraints can yield stable representations by minimizing the statistical differences between features from different source domains; the Maximum Mean Discrepancy (MMD) and its variants are commonly used to measure the differences between distributions (Gretton et al., 2012); Domain adversarial learning, on the other hand, employs domain discriminators and gradient reversal strategies to learn feature representations that are difficult to distinguish by domain origin; DANN being a representative method (Ganin et al., 2016). For bearing fault diagnosis, domain differences caused by operating condition variations manifest not only in shifts in overall amplitude and distribution centers but also in more complex domain-specific features. Therefore, relying solely on mean constraints or domain adversarial mechanisms may prove insufficient for obtaining stable cross-condition representations.

Based on the above analysis, this paper proposes a domain generalization method for rolling bearing fault diagnosis that combines mean difference constraints with domain adversarial learning. Unlike CORAL, which aligns second-order covariance statistics; MMD, which measures the difference in mean embedding distributions; IRM, which constrains classifier invariance; and DANN, which relies solely on adversarial obfuscation, our method explicitly averages the first-order distances between the centers of features from different source domains and combines this with a domain adversarial mechanism, thereby simultaneously constraining stable first-order shifts and more complex domain-specific information. The main contributions of this paper include: first, the design of a lightweight multi-scale convolutional feature extraction network; second, the integration of mean difference constraints with domain adversarial learning for learning domain-invariant features; and third, the validation of the proposed method through comparative experiments, ablation studies, parameter sensitivity analysis, and t-SNE visualization on the CWRU and PU datasets.

2. Method

2.1. Problem Definition

Suppose that during the training phase, K labeled source domains are obtained, with each source domain corresponding to a known operating condition. The set of source domains is represented as:

$$D_s=\{D_1,D_2,\ldots,D_k\},\quad D_k=\{(x_i^k,y_i^k,d_i^k)\}_{i=1}^{n_k} \tag{1}$$

Here, x_i^k denotes the i-th vibration signal sample in the k-th source domain, $y_i^k \in \{1,2,...,C\}$ denotes the fault category label, $d_i^k=k$ denotes the domain label, n_k denotes the number of samples in the k-th source domain, and C denotes the number of fault categories. The goal of the domain generalization task is to train a diagnostic model using data from multiple source domains without the involvement of the unknown target domain D_t during training, so that the model can maintain high fault recognition performance under target operating conditions not encountered during the training phase.

Compared to conventional supervised learning, domain generalization not only requires the model to distinguish between different fault categories but also demands that the model reduce its dependence on specific operating condition distributions, such as load, rotational speed, or radial load. To this end, this paper imposes constraints on the feature space at two levels—first-order statistical differences and complex domain discriminative information—ensuring that the learned features possess both class discriminability and domain invariance.

2.2. Overall Framework

The method described in this paper consists of a multi-scale one-dimensional convolutional feature extractor, a fault classifier, a mean-difference constraint module, and a domain adversarial discriminator. Let the feature extractor, fault classifier, domain discriminator, and gradient reversal layer(GRL) be denoted as $G_f(\cdot\ ;\theta_f),G_y(\cdot\ ;\theta_y),G_d(\cdot\ ;\theta_d), and\ R_\eta(\cdot\ )$, respectively. For an input sample x_i, the model first extracts shared features:

$$f_i=G_f(x_i;\theta_f) \tag{2}$$

The fault classifier then outputs the predicted probability of the fault category:

$$\hat{y}_i=G_y(f_i;\theta_y) \tag{3}$$

The domain classifier predicts the source domain of a sample based on the features that have passed through the gradient reversal layer:

$$\hat{d}i=G_d(R_\eta(f_i);\theta_d) \tag{4}$$

A gradient reversal layer maintains an identity mapping during forward propagation and multiplies the gradients passed to the feature extractor by a negative coefficient during backpropagation. Its function can be expressed as:

$$R_\eta(f)=f,\quad \frac{\partial R_\eta(f)}{\partial f}=-\eta I \tag{5}$$

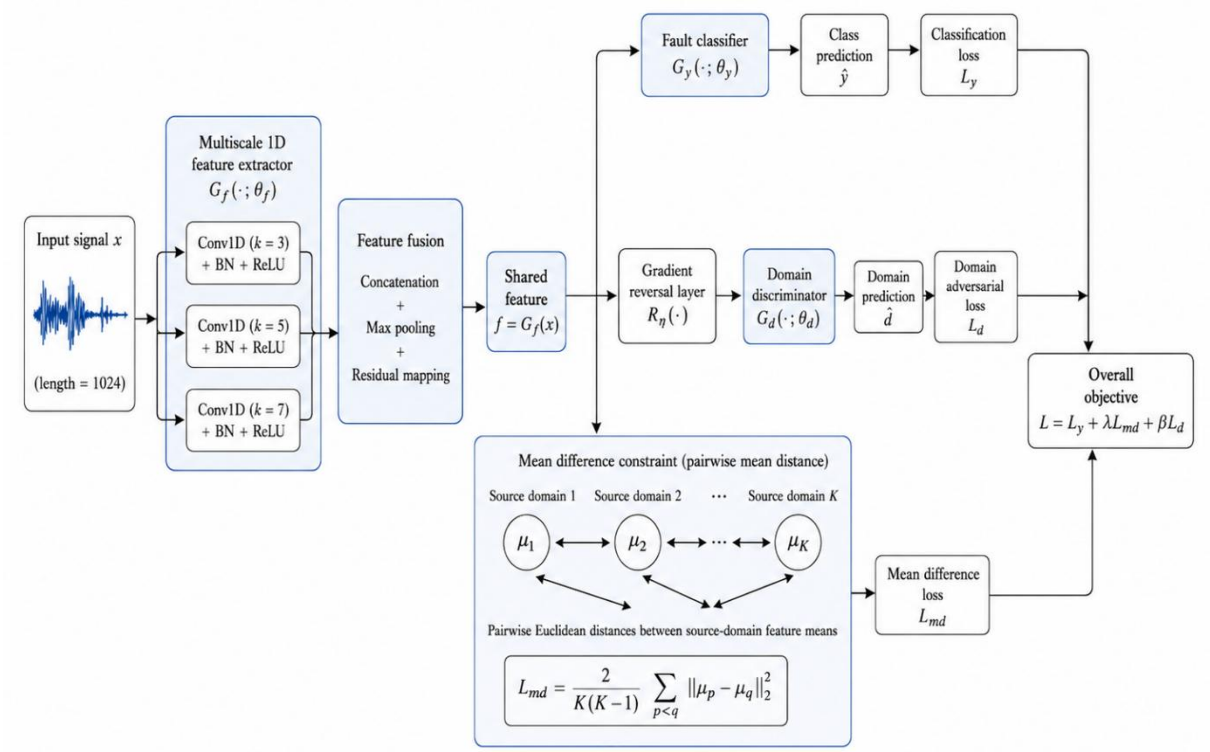
Therefore, the fault classification branch is used to preserve the model's ability to distinguish between categories, the mean difference module is used to constrain first-order statistical shifts between the centers of features in different source domains, and the domain adversarial branch weakens the information used to

WISDOM ACADEMIC https://doi.org/10.63808/aepm.v2i3.425 Volume 2 , Issue 3

distinguish operating conditions through gradient inversion. The overall structure of the model is shown in Figure 1.

Figure 1

Overall structure of the proposed domain-generalized fault diagnosis model



Note. G_f= feature extractor; G_y = fault classifier; G_d= domain discriminator; GRL = gradient reversal layer;$L_c ls$ = fault classification loss; $L_m d$ = mean-difference loss; $L_a dv$ = domain-adversarial loss. Solid arrows indicate forward propagation, and dashed arrows indicate loss-based optimization.

The main network architecture is shown in the table 1 below.

Table 1

Main network architecture

Module	Configuration	Output/Function
Input	A one-dimensional vibration window of length 1024	Raw signal clip
Multiscale branching	1D convolution with kernels of sizes 3, 5, and 7 + Batch Normalization + ReLU Activation Function	Short-term shocks and cross-cycle characteristics
Fusion Encoder	Feature concatenation + max pooling + residual mapping	Domain-shared feature f
Fault Classifier	Fully Connected Layer + Softmax	Fault Category Probability $\hat{y}$
Domain Discriminator	GRL + two fully connected layers + Softmax	Domain probability $\hat{d}$

Note. 1D = one-dimensional; BN = batch normalization; ReLU = rectified linear unit; GRL = gradient reversal layer. f denotes the shared feature representation, ŷ denotes the predicted fault-category probability, and $\hat{d}$ denotes the predicted domain probability.

2.3. Multiscale One-Dimensional Convolutional Feature Extraction

Vibration signals from rolling bearing failures typically contain components such as periodic impacts, local abrupt changes, and noise disturbances. Different failure types exhibit variations in the time intervals between impacts, energy distribution, and local patterns; a single-scale convolution kernel may struggle to fully capture failure information across different time scales. To address this, this paper employs a multi-scale one-dimensional convolution structure for feature extraction.

Given an input vibration signal x, features are extracted using one-dimensional convolution branches with kernel sizes of 3, 5, and 7, respectively:

$$F_m=\mathrm{ReLU}\big(\mathrm{BN}(\mathrm{Conv1D}_m(x))\big),\quad m\in 3,5,7 \tag{6}$$

In particular, F_3, F_5, and F_7 represent features extracted from different receptive fields, respectively. Subsequently, features from different scales are concatenated, and the final shared features are obtained through pooling and residual mapping:

$$F_{ms}=[F_3;F_5;F_7],\quad f=G_r(\mathrm{Pool}(F_{ms})) \tag{7}$$

Here, $[\cdot;\cdot]$ denotes channel-wise concatenation, and $G_r(\cdot)$ denotes the subsequent residual mapping layer. The multiscale structure preserves both short-term impulse responses and periodic variations over longer time scales, thereby enhancing the model's ability to capture failure modes under various load and speed conditions.

2.4. Mean difference constraint

The amplitude, frequency components, and noise levels of vibration signals may vary under different operating conditions, leading to shifts in the distributions of different source domains in the feature space. To mitigate this shift, this paper introduces a constraint on the mean difference between source domains, ensuring that the centers of the feature distributions for samples from different source domains remain close to one another.

For a sample set $\mathscr{B}_k$ from the k-th source domain in a mini-batch, the mean of its features is defined as:

$$\mu_k=\frac{1}{N_k}\sum_{i\in\mathscr{B}_k} f_i \tag{8}$$

Here, $N_k=|\mathscr{B}_k|$ denotes the number of samples from the k-th source domain in the current mini-batch. The inter-source mean difference loss is defined as:

$$L_{md}=\frac{2}{K(K-1)}\sum_{1\le p<q\le K}\left\|\mu_p-\mu_q\right\|_2^2 \tag{9}$$

Equation (9) averages over all unordered source-domain pairs, where $K(K-1)/2$ is the number of source-domain pairs; thus, the coefficient $2/[K(K-1)]$ prevents double-counting and maintains consistency in the loss scale across different tasks. By minimizing L_{md}, the model reduces the first-order statistical differences between the centers of the feature distributions across different source domains.

2.5. Domain-Adversarial Learning

While mean difference constraints primarily focus on first-order statistical shifts in source-domain features, domain-specific information resulting from variations in load, rotational speed, and radial load may also be present in more complex feature dimensions. To further mitigate operating condition-discriminative information in the feature representations, this paper introduces a domain-adversarial learning mechanism based on gradient inversion layers. This mechanism aligns with the fundamental concept of DANN, wherein the domain discriminator attempts to distinguish the source domain of a sample, while the feature extractor learns domain-invariant features capable of confusing the domain discriminator under the influence of reverse gradients [8].

The domain discrimination loss is formulated as cross-entropy:

$$L_d(\theta_f,\theta_d)=-\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbf{1}[d_i=k]\log\hat{d}_{ik} \tag{10}$$

Here, $\mathbf{1}[d_i=k]$ denotes the one-hot label indicating that sample i belongs to the k-th source domain, and $\hat{d}_{ik}$ denotes the probability predicted by the domain discriminator that sample i belongs to the k-th source domain.

If the gradient reversal layer is not explicitly included, the domain adversarial optimization can be expressed in the maximum-minimum form:

$$\min_{\theta_d} L_d\,,\min_{\theta_f,\theta_y}\left(L_y+\lambda L_{md}-\beta L_d\right) \tag{11}$$

Equation (11) shows that the domain classifier improves domain classification performance by minimizing L_d, while the feature extractor learns domain-invariant features by maximizing the domain discriminative loss. In practice, the gradient reversal layer has already reversed the gradient signs during backpropagation; therefore, the domain adversarial term can be incorporated into the overall loss with a positive sign.

2.6. Overall Optimization Goals

The classification loss is expressed in terms of cross-entropy:

$$L_y=-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}y_{ic}\log\hat{y}_{ic} \tag{12}$$

Here, y_{ic} is the one-hot label indicating that sample i belongs to class c, and $\hat{y}_{ic}$ represents the model's predicted probability that sample i belongs to class c.

Combining the fault classification loss, mean difference loss, and domain adversarial loss, the final training objective is:

$$L=L_y+\lambda L_{md}+\beta L_d \tag{13}$$

Here, λ and β represent the weight coefficients for the mean difference constraint and the domain adversarial loss, respectively. In Equation (13), L_d is added with a positive sign because the adversarial direction is already established by the gradient reversal layer during backpropagation. L_y ensures that the features are discriminative across fault categories, L_{md} constrains first-order statistical shifts between source domains, and L_d weakens complex domain-related information. The synergistic effect of these three loss terms enables the model to achieve more stable diagnostic performance under unknown operating conditions. Based on the results of the parameter sensitivity analysis, the main experiments in this paper are set with λ=0.2 *and* β=0.2.

To further illustrate the differences between the method proposed in this paper and existing domain generalization methods, we compare them below in terms of mathematical constraints and optimization objectives. CORAL primarily aligns second-order covariance statistics across different domains, and its core objective can be expressed as $\|C_s - C_t\|_F^2$ (Sun & Saenko, 2016); MMD measures the distance between different distributions via the kernel mean embedding, essentially aligning distribution means in a regenerative kernel Hilbert space (Gretton et al., 2012); IRM focuses more on the stability of classifier optimality across different environments, emphasizing invariant prediction mechanisms (Arjovsky et al., 2019); DANN primarily relies on adversarial obfuscation by a domain discriminator to obtain domain-invariant features (Ganin et al., 2016).

Unlike the aforementioned methods, the mean difference constraint in this paper directly averages the pairwise distances between the centers of all domain features across multiple source domains, explicitly constraining first-order feature center shifts; simultaneously, a domain adversarial discriminator is further introduced to suppress more complex operational condition-related information. Therefore, our method does not rely solely on first-order moment alignment or adversarial obfuscation; instead, it combines interpretable first-order statistical alignment with adversarial domain-invariant learning, thereby achieving a better balance between class discriminability and domain invariance.

3. Experiments and Analysis

3.1. Introduction to the Dataset and Experimental Setup

To validate the effectiveness of the proposed method across different operating conditions, this paper conducts experiments using two publicly available rolling bearing datasets: CWRU and PU. The CWRU dataset primarily reflects variations in rotational speed and load under different motor load conditions, while the PU dataset includes multiple operating variables such as rotational speed, load torque, and radial load. Therefore, these two datasets can validate the model’s cross-domain generalization capability from different perspectives.

The CWRU dataset is publicly available from the Case Western Reserve University Bearing Data Center and includes data for normal bearings as well as bearings with inner ring, outer ring, and rolling element failures induced by electrical discharge machining (Case Western Reserve University Bearing Data Center, n.d.). This paper selects 12 kHz sampling vibration signals from the drive end and treats the four load conditions of 0 HP, 1 HP, 2 HP, and 3 HP as four domains, denoted as A, B, C, and D, respectively.

WISDOM ACADEMIC https://doi.org/10.63808/aepm.v2i3.425 Volume 2 , Issue 3

The PU dataset is provided by Paderborn University and is used for condition monitoring research on rolling bearing damage in motor drive systems (Lessmeier et al., 2016). This dataset simultaneously captures bearing vibration signals and motor current signals, while also recording auxiliary information such as rotational speed, load torque, radial load, and temperature. This paper selects four typical operating conditions as four domains.

Table 2

Operating Conditions for the CWRU Dataset

Domain ID	Load/HP	Approximate speed/rpm	Domain type
A	0	1797	Load domain
B	1	1772	Load domain
C	2	1750	Load domain
D	3	1730	Load domain

Note. CWRU = Case Western Reserve University; HP = horsepower; rpm = revolutions per minute. Domains A–D represent four load conditions.

Table 3

Operating Conditions for the PU Dataset

Domain ID	Operating Condition Name	Approximate speed/rpm	Load torque / Nm	Radial force / N
A	N15_M07_F10	1500	0.7	1000
B	N09_M07_F10	900	0.7	1000
C	N15_M01_F10	1500	0.1	1000
D	N15_M07_F04	1500	0.7	400

Note. PU = Paderborn University; rpm = revolutions per minute; N·m = newton-meter; N = newton. Domains A–D represent four operating conditions.

In terms of diagnostic categories, the CWRU dataset includes four categories: normal condition, inner ring failure, outer ring failure, and rolling element failure; the PU dataset includes 10 bearing condition categories that can be obtained under all four operating conditions. To ensure experimental fairness and sample balance, 200 one-dimensional samples were selected for each category in each domain of the CWRU dataset, resulting in a total of 800 samples per domain and 3,200 samples across all four domains; In the PU dataset, 200 one-dimensional samples were selected for each category in every domain, resulting in a total of 2,000 samples per domain and 8,000 samples across all four domains. The diagnostic categories for the two datasets are shown in Tables 4 and 5, respectively.

Table 4

Diagnosis Categories in the CWRU Dataset

Category Code	Fault Type	Note	Number of samples per domain
C0	Normal	Normal state	200
C1	Inner Race	Inner ring failure	200
C2	Outer Race	Outer ring failure	200
C3	Ball	Rolling element failure	200

Note. CWRU = Case Western Reserve University. C0–C3 denote the diagnostic category codes. Each domain contains 200 samples per category.

Table 5

Diagnostic Category Settings for the PU Dataset

Category Code	Bearing Number	Fault Location/Type	Description	Number of Samples per Domain
P0	K001	Health	Non-destructive reference bearing	200
P1	KA04	outer ring	Actual pitting damage	200
P2	KA15	outer ring	Plastic deformation/indentation	200
P3	KA16	outer ring	Actual pitting damage	200
P4	KA22	outer ring	Actual pitting damage	200
P5	KA30	outer ring	Plastic deformation/indentation	200
P6	KI04	inner ring	Actual pitting damage	200
P7	KI14	inner ring	Actual pitting damage	200
P8	KI16	inner ring	Actual pitting damage	200
P9	KI18	inner ring	Actual pitting damage	200

Note. PU = Paderborn University. P0–P9 denote the diagnostic category codes. KA denotes outer-ring damage, KI denotes inner-ring damage, and K001 denotes the normal condition.

Both datasets employ a domain generalization setup using leave-one-domain testing. For the CWRU and PU datasets, four task pairs were constructed: A, B, C → D; A, B, D → C; A, C, D → B; and B, C, D → A. In each pair, the domain on the left side of the arrow serves as the source domain for training, while the domain on the right side serves as the unknown target domain and is used only for final testing; it does not participate in model training, validation, or parameter updates.

During the data preprocessing stage, the raw vibration signals were first Z-score normalized, then segmented into one-dimensional samples of length 1024 using a sliding window with a step size of 512. To avoid potential data leakage caused by overlapping sliding windows, this study first divides the training and target domains according to operating conditions, and then performs window segmentation within each domain. Consequently, overlapping windows from the same raw recording will not appear simultaneously in both the training and test sets. All deep learning models were trained using the Adam optimizer, with an initial learning rate of 0.001, a batch size of 64, and 100 training epochs. Each set of experiments was independently repeated five times using different random seeds, and the results are reported as the mean ± standard deviation. The evaluation metric used was classification accuracy (Accuracy):

$$Acc = N_{correct}/N_{total} \times 100 \tag{14}$$

Here, $N_{correct}$ denotes the number of samples correctly classified in the target domain test set, and N_{total} denotes the total number of test samples in the target domain. To meet statistical significance requirements, this paper employs Welch's two-sample t-test based on summary statistics—specifically, the mean and standard deviation from the five repeated experiments—to compare the differences between the proposed method and the best baseline method.

3.2. Comparative experiment

To comprehensively validate the effectiveness of the proposed method, this paper compares it with traditional machine learning methods, typical deep learning methods, and domain-invariant learning methods, including SVM, 1D-CNN, ResNet1D, TCN, CORAL, MMD, and DANN. SVM uses common time-domain statistical features for classification (Cortes & Vapnik, 1995); 1D-CNN takes the raw vibration signal directly as input (Zhang et al., 2017; Zhang et al., 2018); ResNet1D extends the residual architecture to the one-dimensional vibration signal classification task (He et al., 2016); TCN is used to model temporal dependencies (Bai et al., 2018); CORAL, MMD, and DANN represent second-order statistical alignment, kernel mean difference constraints, and domain adversarial learning methods, respectively (Ganin et al., 2016; Gretton et al., 2012; Sun & Saenko, 2016). All deep learning methods employ the same data split, number of training epochs, batch size, and learning rate settings.

Table 6

Accuracy of different methods on the CWRU cross-load task (mean ± standard deviation)

Method	A,B,C→D	A,B,D→C	A,C,D→B	B,C,D→A	Average
SVM	94.12±0.58	94.83±0.46	95.27±0.52	94.36±0.61	94.65±0.54
1D-CNN	96.18±0.43	96.72±0.39	97.04±0.34	95.89±0.47	96.46±0.41
ResNet1D	97.02±0.36	97.43±0.32	97.88±0.30	96.75±0.40	97.27±0.35
TCN	97.26±0.34	97.61±0.30	98.03±0.27	96.92±0.38	97.46±0.32
CORAL	97.48±0.31	98.01±0.28	98.26±0.24	97.22±0.34	97.74±0.29
MMD	97.53±0.30	98.06±0.27	98.31±0.23	97.25±0.33	97.79±0.28
DANN	97.65±0.29	98.17±0.25	98.44±0.22	97.36±0.33	97.91±0.27

Advanced Engineering & Precision Manufacturing
WISDOM ACADEMIC https://doi.org/10.63808/aepm.v2i3.425 Volume 2 , Issue 3

Ours	98.75±0.21	99.18±0.18	99.46±0.16	98.93±0.22	99.08±0.19

Note. Values are presented as mean ± standard deviation across five independent runs. SVM = support vector machine; 1D-CNN = one-dimensional convolutional neural network; ResNet1D = one-dimensional residual network; TCN = temporal convolutional network; CORAL = correlation alignment; MMD = maximum mean discrepancy; DANN = domain-adversarial neural network. The domain before the arrow represents the source domain, and the domain after the arrow represents the target domain. "Ours" denotes the proposed method.

As shown in Table 6, the traditional SVM method achieved an average accuracy of 94.65% ± 0.54% on the CWRU dataset, indicating that manually engineered features still possess some diagnostic capability on this dataset; however, their cross-load generalization performance is weaker than that of deep learning methods. The average accuracy rates for 1D-CNN, ResNet1D, and TCN were 96.46% ± 0.41%, 97.27% ± 0.35%, and 97.46% ± 0.32%, respectively, indicating that deep models can automatically learn more discriminative fault features from raw vibration signals. CORAL, MMD, and DANN outperform standard deep models, suggesting that distribution alignment and domain adversarial learning help mitigate feature shifts caused by load variations. Our method achieved the highest accuracy across all four task sets, with an average accuracy of 99.08% ± 0.19%, representing a 1.17 percentage point improvement over the best baseline, DANN. Results from a Welch t-test based on the mean and standard deviation of five repeated experiments indicate that this improvement is statistically significant ($t = 7.924$, $p = 8.45\times10^{-5}$).

Table 7

Accuracy of Different Methods on the PU Cross-Condition Task (%) (Mean ± Standard Deviation)

Method	A,B,C→D	A,B,D→C	A,C,D→B	B,C,D→A	Average
SVM	85.21±0.82	86.04±0.74	85.67±0.79	86.42±0.77	85.84±0.78
1D-CNN	87.62±0.66	88.49±0.61	88.05±0.64	89.14±0.58	88.33±0.62
ResNet1D	88.95±0.58	89.62±0.55	89.24±0.57	90.03±0.52	89.46±0.56
TCN	89.36±0.54	90.01±0.50	89.74±0.51	90.28±0.49	89.85±0.51
CORAL	89.84±0.49	90.53±0.45	90.18±0.47	90.67±0.44	90.31±0.46
MMD	90.05±0.47	90.76±0.43	90.34±0.45	90.91±0.42	90.52±0.44
DANN	90.24±0.45	91.02±0.41	90.59±0.43	90.99±0.40	90.71±0.42
Ours	91.42±0.37	92.08±0.32	91.75±0.35	92.19±0.31	91.86±0.34

Note. Values are presented as mean ± standard deviation across five independent runs. SVM = support vector machine; 1D-CNN = one-dimensional convolutional neural network; ResNet1D = one-dimensional residual network; TCN = temporal convolutional network; CORAL = correlation alignment; MMD =

maximum mean discrepancy; DANN = domain-adversarial neural network. The domain before the arrow represents the source domain, and the domain after the arrow represents the target domain. “Ours” denotes the proposed method.

As shown in Table 7, the overall accuracy on the PU dataset is lower than that on the CWRU dataset, indicating that the variations in multi-condition variables in the PU dataset are more complex, placing higher demands on domain generalization capabilities. Our method achieved the highest accuracy across all four cross-condition tasks on the PU dataset, with an average accuracy of 91.86% ± 0.34%, representing a 1.15 percentage point improvement over the best baseline, DANN. As shown in Table 8, the Welch t-test results indicate that this improvement is statistically significant ($p < 0.01$) compared to the second-best method, DANN. The overall accuracy for the A, C, and D→B tasks is relatively low, primarily because the target domain B operates at 900 r/min, while most source domains operate at 1500 r/min; this difference in rotational speed alters the fault impulse intervals and spectral energy distribution.

Table 8

Results of statistical significance tests comparing the proposed method with the optimal baseline method

Dataset	Comparison Group	Ours	Optimal baseline	t-value	p-value
CWRU	Ours vs. DANN	99.08±0.19	97.91±0.27	7.924	8.45×10^{-5}
PU	Ours vs. DANN	91.86±0.34	90.71±0.42	4.759	1.61×10^{-3}

Note. Values are presented as mean ± standard deviation across five independent runs. DANN = domain-adversarial neural network. The t values and *p* values were obtained using Welch’s two-sample *t* tests.

3.3. Ablation Experiment

To clarify the contribution of each module to the model’s performance, this paper conducts ablation experiments on the PU dataset for the A, C, D→B task. This task was selected because the rotational speeds in the target domain B differ significantly from those in most operating conditions in the source domain, allowing for a more thorough evaluation of the model’s generalization ability under challenging and unknown operating conditions.

Table 9

Results of ablation experiments on the PU datasets A, C, and D for the B task (in %)

Model Settings	Multiscale Convolution	Mean difference constraint	Domain-Adversarial Learning	Accuracy
Baseline	×	×	×	88.05±0.64
Model-1	√	×	×	88.72±0.56
Model-2	√	√	×	90.63±0.42

Model-3	√	×	√	90.91±0.39
Ours	√	√	√	91.75±0.35

Note. Values are presented as mean ± standard deviation across five independent runs. A checkmark indicates that the corresponding component was included, whereas a cross indicates that it was excluded. "Ours" denotes the complete proposed model.

As shown in Table 9, the baseline 1D-CNN model achieved an accuracy of 88.05% ± 0.64% on this task. After incorporating a multi-scale convolutional structure, the accuracy improved to 88.72% ± 0.56%, indicating that convolutional kernels of different scales can more effectively extract local impact features and periodic variation information from the vibration signals. After incorporating a mean difference constraint into the multi-scale model, the accuracy further improved to 90.63% ± 0.42%, indicating that reducing the variance in the feature centers of the source domains helps mitigate intra-class shifts across operating conditions. When domain adversarial learning was applied independently, the accuracy reached 90.91% ± 0.39%, indicating that adversarial domain obfuscation can further weaken complex discriminative information related to operating conditions. When all three methods were used simultaneously, the model achieved the highest accuracy of 91.75% ± 0.35%.

3.4. Parameter Sensitivity Analysis

To demonstrate the validity of the loss weights and model architecture selection, this paper further presents sensitivity analyses of λ, β, convolution kernel size, and the domain discriminator architecture. The experiments were conducted on the PU dataset for the A, C, D→B task.

Table 10

Sensitivity analysis of λ and β for the PU datasets A, C, D→B tasks / %

λ	β	Accuracy
0.0	0.0	88.72±0.56
0.1	0.1	90.84±0.41
0.2	0.2	91.75±0.35
0.3	0.2	91.28±0.38
0.2	0.3	91.10±0.40
0.5	0.5	90.36±0.47

Note. Values are presented as mean ± standard deviation across five independent runs. λ and β denote the weights assigned to the mean-difference loss and domain-adversarial loss, respectively. The domain before the arrow represents the source domain, and the domain after the arrow represents the target domain.

Table 11

Sensitivity Analysis of Convolution Kernel Size and Domain Discriminator Architecture / %

Configuration	Settings	Accuracy
Convolution kernel size	Single branch, k=5	90.42±0.44
Convolution kernel size	Two-branch tree with k=3,5	91.06±0.39
Convolution kernel size	Three branches: k=3, 5, 7	91.75±0.35
Convolution kernel size	Three branches: k=3, 5, 9	91.42±0.37
Domain Discriminator	Fully connected single layer	91.19±0.40
Domain Discriminator	Two fully connected layers	91.75±0.35
Domain Discriminator	Three fully connected layers	91.33±0.38

Note. Values are presented as mean ± standard deviation across five independent runs. k denotes the convolutional kernel size. Accuracy is reported as a percentage.

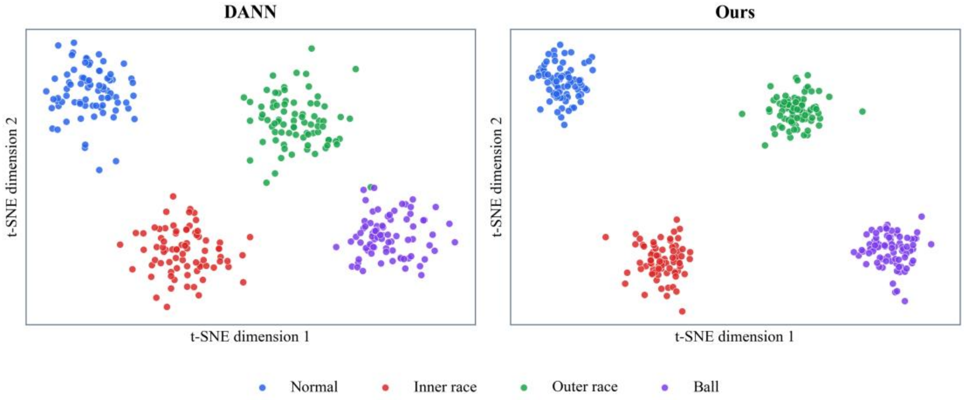
As shown in Table 10, when λ=0 and β=0, the model degenerates into a baseline model that uses only multi-scale convolutional features, with an accuracy of 88.72% ± 0.56%. As the weights of the mean difference constraint and the domain adversarial constraint increase, the model's accuracy gradually improves, reaching its optimal value when λ=0.2 and β=0.2. Further increasing the weights causes the accuracy to decline, indicating that overly strong domain alignment constraints may weaken the class separability of features. As shown in Table 11, the best results are achieved with three-branch (k=3, 5, 7) multiscale convolutions and a two-layer fully connected domain discriminator; excessively large convolutional kernels or overly deep domain discriminators may introduce redundant features or lead to instability in adversarial training.

3.5. T-SNE Visualization Analysis

To further evaluate the feature representations learned by different methods, t-distributed Stochastic Neighbor Embedding (t-SNE) is used to project the high-dimensional features into a two-dimensional space. The CWRU A,C,D→B cross-load task is selected for visualization, where domains A, C, and D are used for training and domain B is treated as the unseen target domain. We compare the target-domain features extracted before the classifier by the strong baseline DANN and the proposed method.

Figure 2

t-SNE feature visualization results on A, C, D→B tasks from the CWRU dataset



Note. DANN = domain-adversarial neural network; t-SNE = t-distributed stochastic neighbor embedding. Colors indicate the normal, inner-race fault, outer-race fault, and ball-fault categories. The left panel presents the DANN results, whereas the right panel presents the results of the proposed method. Domains A, C, and D are the source domains, and Domain B is the target domain.

As shown in Figure 2, the DANN model can form several category-related clusters in the target domain, but some samples remain relatively scattered, and the class boundaries are not sufficiently clear. In contrast, the proposed method produces more compact intra-class distributions and clearer inter-class separation among the four fault categories. The Normal, Inner race, Outer race, and Ball samples form more independent clustering regions, indicating that the learned features have stronger fault discriminability under unseen operating conditions. These visualization results provide feature-level evidence that the proposed combination of mean difference constraints and domain adversarial learning improves cross-condition generalization performance.

4. Conclusion

This paper addresses the issue of insufficient generalization capability in rolling bearing fault diagnosis models across varying loads, speeds, and operating conditions by proposing a domain-generalized fault diagnosis method that combines multi-scale convolution, mean difference constraints, and domain adversarial learning. Taking raw one-dimensional vibration signals as input, this method uses a multiscale convolutional network to extract fault impact features and periodic features across different time scales, and enhances the domain invariance of these features through a mean difference loss between source domains and a domain adversarial mechanism based on gradient inversion layers.

Results from leave-one-domain experiments on the CWRU and PU rolling bearing datasets show that the proposed method achieves an average accuracy of 99.08% ± 0.19% on the CWRU dataset, and an average accuracy of 91.86% ± 0.34% on the PU dataset, both outperforming comparison methods such as SVM, 1D-CNN, ResNet1D, TCN, CORAL, MMD, and DANN. Ablation experiments further demonstrate that multi-scale convolution, mean difference constraints, and domain adversarial learning all contribute positively to model performance improvement. Parameter sensitivity analysis indicates that moderate domain invariance constraints strike a good balance between source domain alignment and class separability.

It should be noted that the experiments in this paper are still primarily based on publicly available laboratory datasets and do not yet fully account for complex factors present in real industrial settings, such as sensor installation variations, non-stationary noise, composite failures, and continuous changes in operating conditions. Therefore, the conclusions of this paper should be limited to the datasets and experimental protocols used. Future work will incorporate more real-world industrial data and cross-device datasets to investigate robust domain-general diagnostic methods under conditions of noise interference, sample imbalance, and online domain adaptation. Additionally, we will combine multi-sensor fusion and interpretability analysis to enhance the reliability of the model in engineering applications.

Acknowledgments

The authors have no additional acknowledgments to declare.

Author contributions

Conceptualization, Jiabing Zhou and Xiang Gu; methodology, Jiabing Zhou, Xiang Gu, and Bo Zhang; software, Xiang Gu and Bo Zhang; validation, Bo Zhang, Xinrui Yu, and Yining Xu; formal analysis, Xiang Gu and Xinrui Yu; investigation, Xiang Gu, Bo Zhang, Xinrui Yu, and Yining Xu; resources, Jiabing Zhou; data curation, Bo Zhang, Xinrui Yu, and Yining Xu; writing—original draft preparation, Xiang Gu and Bo Zhang; writing—review and editing, Jiabing Zhou, Xiang Gu, Xinrui Yu, and Yining Xu; visualization, Xiang Gu and Xinrui Yu; supervision, Jiabing Zhou; project administration, Jiabing Zhou; funding acquisition, Jiabing Zhou. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the 2026 Undergraduate Innovation and Entrepreneurship Training Program of Beijing Technology and Business University, grant number S202610011002.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate
Not applicable.

Consent for publication
All authors have given their consent.

Additional information

Received: 22 May 2026

Accepted: 18 June 2026

Published online: 16 August 2026

Publisher's Note
Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019). Invariant risk minimization [Preprint]. arXiv. https://arxiv.org/abs/1907.02893

[2] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [Preprint]. arXiv. https://arxiv.org/abs/1803.01271

[3] Bhatti, S., Shaikh, A. A., Mansoor, A., & Hussain, M. (2024). New approaches of stochastic models to examine the vibration features in roller bearings. *Applied Sciences, 14*(4), Article 1616. https://doi.org/10.3390/app14041616

[4] Bhatti, S., Shaikh, F., Mansoor, A., & Shaikh, A. A. (2025). A unique statistical framework to predict the health of a machine by utilizing the vibration features of rolling element bearing data. *Mehran University Research Journal of Engineering and Technology, 44*(3), 37–41. https://doi.org/10.22581/muet1982.0166

[5] Case Western Reserve University Bearing Data Center. (n.d.). *Bearing Data Center dataset* [Data set]. Case Western Reserve University. https://engineering.case.edu/bearingdatacenter

[6] Chen, L., Li, Q., Shen, C., Zhu, J., Wang, D., & Xia, M. (2022). Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions. *IEEE Transactions on Industrial Informatics, 18*(3), 1790–1800. https://doi.org/10.1109/TII.2021.3078712

[7] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

[8] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research, 17*(59), 1–35. https://www.jmlr.org/papers/v17/15-239.html

[9] Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research, 13*, 723–773. https://www.jmlr.org/papers/v13/gretton12a.html

[10] Guo, Y., Ju, G., & Zhang, J. (2024). Generalization study of intelligent fault diagnosis under various working conditions and its application to planetary gearboxes. *Scientific Reports, 14*, Article 25447. https://doi.org/10.1038/s41598-024-75088-8

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

[12] Kim, I., Kim, S. W., Kim, J., Huh, H., Jeong, I., Choi, T., Kim, J., & Lee, S. (2024). Single domain generalizable and physically interpretable bearing fault diagnosis for unseen working conditions. *Expert Systems with Applications, 241*, Article 122455. https://doi.org/10.1016/j.eswa.2023.122455

[13] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing, 138*, Article 106587. https://doi.org/10.1016/j.ymssp.2019.106587

[14] Lessmeier, C., Kimotho, J. K., Zimmer, D., & Sextro, W. (2016). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. *PHM Society European Conference, 3*(1). https://doi.org/10.36001/phme.2016.v3i1.1577

[15] Li, D., Yang, Y., Song, Y.-Z., & Hospedales, T. M. (2017). Deeper, broader and artier domain generalization. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 5542–5550). IEEE. https://doi.org/10.1109/ICCV.2017.593

[16] Muandet, K., Balduzzi, D., & Schölkopf, B. (2013). Domain generalization via invariant feature representation. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th International Conference on Machine Learning* (Vol. 28, pp. 10–18). PMLR. https://proceedings.mlr.press/v28/muandet13.html

[17] Ren, H., Wang, J., Huang, W., Jiang, X., & Zhu, Z. (2023). Domain-invariant feature fusion networks for semi-supervised generalization fault diagnosis. *Engineering Applications of Artificial Intelligence, 126*, Article 107117. https://doi.org/10.1016/j.engappai.2023.107117

[18] Sun, B., & Saenko, K. (2016). Deep CORAL: Correlation alignment for deep domain adaptation. In G. Hua & H. Jégou (Eds.), *Computer vision—ECCV 2016 workshops* (pp. 443–450). Springer. https://doi.org/10.1007/978-3-319-49409-8_35

[19] Zhang, S., Su, L., Gu, J., Li, K., Zhou, L., & Pecht, M. (2023). Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey. *Chinese Journal of Aeronautics, 36*(1), 45–74. https://doi.org/10.1016/j.cja.2021.10.006

[20] Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing, 100*, 439–453. https://doi.org/10.1016/j.ymssp.2017.06.022

[21] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors, 17*(2), Article 425. https://doi.org/10.3390/s17020425