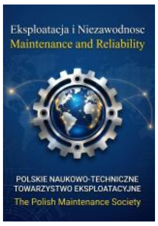


Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 1

https://ein.org.pl

The Polish Maintenance Society

TMDANN-A: Boundary-aware discriminative alignment and condition-guided multi-source attention for vibration-based bearing fault diagnosis under variable operating conditions

Changming Zhang[1], Yiping Yuan[1,*], Bo Chen[2], Jiaqi Wang[1], Shaoke Yuan[3]

[1] Xinjiang University, China
[2] Chengdu Institute of Special Equipment Inspection and Testing, China
[3] Shanghai University of Electric Power, China
* Corresponding Author: 2962501153@qq.com

Received: 20 May 2026
Revised: 1 June 2026
Accepted: 11 July 2026
Online: 28 July 2026

Abstract

Rolling bearing fault diagnosis is important for condition monitoring of rotating machinery. In practice, variations in operating conditions and heterogeneity among source domains often cause distribution shifts in vibration signals, while labeled target-domain data are usually unavailable. To address this issue, this paper proposes TMDANN-A, a multi-source domain adaptation framework for vibration-based bearing fault diagnosis. The framework combines adversarial domain alignment, boundary-aware discriminative learning, and condition-guided source weighting to improve domain invariance and class separability. A hierarchical CNN-Transformer encoder is used to capture local impulsive features and long-range temporal dependencies. Experiments on the JNU and PU datasets achieve 98.51% average accuracy over six transfer tasks. On an industrial test-rig dataset, the method further achieves 95.42% average accuracy, 95.44% Macro-F1, and 95.22% mean fault recall. The results verify its effectiveness under complex operating conditions.

Keywords: *rolling bearing, fault diagnosis, variable operating conditions, multi-source domain adapta-tion, condition monitoring*

Article citation:
Zhang C, Yuan Y, Chen B, Wang J, Yuan S, TMDANN-A: Boundary-aware discriminative alignment and condition-guided multi-source attention for vibration-based bearing fault diagnosis under variable operating conditions, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(1) http://doi.org/10.17531/ein/225749

Highlights

- The proposed framework combines adversarial alignment and boundary-aware learning.
- Condition-guided source attention improves multi-source transfer under varying conditions.
- CNN-Transformer fusion captures both local fault features and global dependencies.
- The method achieves high accuracy on six public multi-source transfer tasks.
- Its effectiveness is further verified on an industrial test-rig dataset.

1. Introduction

Rolling bearings are key components of rotating machinery, and their condition directly affects equipment safety and production efficiency. Under complex operating conditions, fault diagnosis is often hindered by domain shift and limited labeled data, which reduces the effectiveness of deep models built on the i.i.d. assumption in real applications [1,2]. Transfer learning, particularly unsupervised domain adaptation (UDA), has therefore been widely used to improve generalization in unlabeled target domains through domain-invariant representation learning [3]. When multiple source domains are available, multi-source domain adaptation (MSDA) provides richer prior knowledge and better matches industrial practice. However, inter-source discrepancy, class-conditional mismatch, and data imbalance also increase the difficulty of alignment and the risk of negative transfer [4].

Existing cross-domain fault diagnosis methods can generally be grouped into three categories. The first includes statistical discrepancy-based methods, such as MMD, LMMD, CORAL, and JMMD, which align marginal or class-conditional distribu-tions to reduce the gap between source and target domains in feature space [5–7]. In multi-source settings, some studies further employ subdomain-level alignment or multi-branch structures to improve intra-class consistency and reduce errors caused by global matching. Although these methods are easy to implement and usually stable, they are less effective in han-dling class boundaries and nonstationary characteristics under complex operating conditions.

Adversarial adaptation learns domain-invariant features through competition between the feature extractor and the do-main discriminator. Typical methods include DANN, CDAN, and MDD. Recent studies have incorporated conditional guid-ance and dynamic weighting to improve robustness in multi-source heterogeneous settings [8].

Cross-domain learning can also be strengthened by addressing long-tailed and imbalanced data. Generative methods such as diffusion-based denoising help balance data distributions, while source-sample refinement and self-supervised learning improve pseudo-label reliability and feature discrimination in unlabeled target domains [9,10]. These strategies reduce nega-tive transfer and improve adaptability under complex conditions [11].

Recent deep learning-based fault diagnosis methods have further improved transferability, feature representation, robust-ness, and data augmentation from different perspectives. Li et al. proposed an interpretable dynamic weighted domain ad-aptation network for cross-domain bearing fault diagnosis under time-varying speeds, which improves transfer performance through dynamic distribution alignment and attention-based weighting [12]. Li et al. developed a multimodal non-Gaussian denoising diffusion generative adversarial network to enhance fault sample generation and alleviate data scarcity [13]. Wang et al. proposed a spatial-channel collaborative multi-scale graph interaction deep transfer learning method to strength-en multi-source feature interaction through graph-based modeling [14]. Dong et al. developed an entropy-oriented semi-supervised dynamic prototype contrastive learning framework to improve fault representation under limited labeled samples [15]. Li et al. introduced a convolutional-transformer reinforcement learning agent for rotating machinery fault diagnosis, showing the potential of combining convolutional feature extraction, Transformer-based temporal modeling, and decision optimization [16]. These studies demonstrate the effectiveness of causal meta-learning, diffusion-based data augmentation, graph interaction, prototype contrastive learning, and CNN-Transformer modeling in intelligent fault diagnosis.

Despite these advances, several challenges remain in multi-source cross-condition bearing fault diagnosis. First, existing alignment strategies often emphasize global or subdomain distribution matching, but the joint optimization of domain invar-iance and class-level decision-boundary separability is still insufficient. Second, although dynamic weighting and graph-based interaction improve source-domain utilization, how to continuously evaluate source relevance and suppress negative transfer under heterogeneous multi-source conditions remains challenging. Third, data augmentation and semi-supervised contrastive learning can alleviate sample scarcity, but the robustness of multi-source transfer diagnosis under limited samples, strong noise, and complex operating-condition shifts still requires further investigation. These gaps motivate the proposed TMDANN-A framework, which integrates adversarial alignment, boundary-aware discriminative learning, and condition-guided multi-source attention.

To address these issues, we propose TMDANN-A, an attention-enhanced adversarial alignment framework for multi-source fault diagnosis. It is designed to handle source heterogeneity and weak class separability. A gradient reversal layer (GRL) transfers adversarial signals from the domain discriminator to the feature extractor, encouraging domain-invariant representations. A Transformer encoder is introduced to capture both local and global temporal information, providing stronger features for cross-domain alignment.

We further develop a margin-based discriminative alignment (MBDA) module to reduce boundary collapse and class overlap. By enlarging inter-class margins and improving intra-class compactness, MBDA strengthens decision boundaries and reduces negative transfer caused by boundary mismatch. In addition, a condition-guided multi-source attention (CG-MSA) mechanism is designed to assign dynamic weights according to source-target conditional similarity. This allows soft fusion of multiple source domains instead of hard selection. A monotonicity constraint is further used to avoid extreme weights and unstable fluctuations, suppressing low-quality sources while retaining useful complementary information.

This study targets multi-source transfer learning for bearing fault diagnosis under complex operating conditions. Our main contributions are as follows:

We replace a single alignment signal with joint optimization of domain-adversarial alignment, margin-aware discrimination, and condition-guided attention, systematically improving both domain invariance and class separability.

We design a multi-source fusion strategy in which continuous, monotonic attention weights balance source contributions, mitigating information loss and negative transfer.

We introduce MBDA to explicitly regularize decision-boundary geometry by enlarging margins and promoting intra-class compactness, improving discriminative robustness under long-tailed and few-shot classes.

We conduct systematic experiments on multi-source transfer tasks across loads and speeds and benchmark against representative methods, demonstrating improved accuracy, stability, and generalization.

The remainder of this paper is organized as follows. Section 2 reviews related research on cross-domain diagnosis for rotating machinery. Section 3 details the proposed method, including implementation specifics. Section 4 presents comparative and ablation experiments on multiple datasets and operating conditions. Section 5 concludes the paper and discusses direc-tions for future work.

2. Related work

2.1. Transfer learning and domain adaptation

In cross-condition and cross-equipment diagnosis of rotating machinery, the source domain $D_S = (x_s, y_s)$and target domain $D_t = (x_t)$often differ in both marginal and class-conditional distributions, and the target domain is unlabeled. UDA aims to learn a decision function f from labeled source data that generalizes to the target domain by reducing domain discrepancy while maintaining class separability. It has become an important approach in intelligent fault diagnosis for dealing with the breakdown of the deep-learning i.i.d. assumption in real engineering settings.

In recent years, cross-domain fault diagnosis has become an active topic in intelligent maintenance. Existing methods can be roughly divided into three categories. The first includes distribution-alignment methods, which reduce source-target discrepancy for feature transfer. Ma et al. used differential matching with multi-task optimization for cross-condition transfer [17]. Si et al. introduced multi-moment matching to improve feature distillation and class discrimination [18]. Cao et al. proposed a two-stage alignment strategy combining marginal and class-conditional matching [19]. Peng et al. developed M3SDA to align moments across multiple sources and the target domain [20]. The second category is adversarial domain adaptation. Ganin et al. proposed DANN, which uses a gradient reversal layer to learn domain-invariant features [21]. Shen et al. developed WDGRL by introducing Wasserstein distance to improve adversarial stability [22]. Zhu et al. extended MFSAN with multiple domain discriminators and classifier-consistency constraints for multi-source transfer [23]. Schwendemann et al. combined multi-layer alignment and adversarial training for bearing diagnosis [24]. The third category includes generation- and weighting-based strategies. Xu et al. combined pairwise source-target adversarial alignment with score fusion to reduce negative transfer [25]. Lu et al. introduced fuzzy weighting to improve the stability of multi-source fusion [26].

2.2. Adversarial domain adaptation

Adversarial methods learn domain-invariant features through competition between the feature extractor and the domain discriminator. DANN is a representative model that achieves end-to-end adversarial training with a Gradient Reversal Layer (GRL). It consists of a feature extractor G_f, a domain classifier G_d, and a label predictor G_y. The learned features should both support accurate classification and confuse the domain discriminator, so that domain information is suppressed. Let the input space be R^m, and let $G_f: R^m \rightarrow R^n$map samples into an n-dimensional feature space. With sigmoid activation, the output of $G_f$can be written as：

$$G_f（x; W, b） = sigm(Wx + b) \tag{1}$$

$$sigm(a) = [\frac{1}{1 + exp(-a_i)}]_{i=1}^{|a|} \tag{2}$$

where G_f treats matrix－vector parameters （W, b）$\in R^{n \times m} \times R^n$.

The label predictor G_y adopts a softmax:

$$G_y（G_f(x); V, c） = soft\,max(VG_f(x) + c) \tag{3}$$

$$\mathrm{soft}\,max(a) = [\frac{exp(a_i)}{\Sigma_{j=1}^{|a|} exp(a_j)}]_{i=1}^{|a|} \tag{4}$$

and the cross-entropy loss is:

$$L_y(p,q) = -\sum_{i=1}^{n} p(x_i)\,log\,q\,(x_i) \tag{5}$$

where p denotes the ground-truth distribution of a sample and q the model' s predicted distribution.

The empirical objective on the source domain is:

$$min_{W,b,V,c} = [\frac{1}{n}\Sigma_{i=1}^{n} L_y^i(W,b,V,c) + \lambda \cdot R(W,b)] \tag{6}$$

where L_y^i denotes the label-prediction loss for the i-th sample, $R(W,b)$ is an optional regularizer, and λ is a user-set coefficient to prevent overfitting. For the domain classifier, a sigmoid activation is used. Its output is:

$$G_d\left(G_f(x);u,z\right) = sigm(u^T G_f(x) + z) \tag{7}$$

where $G_d(\cdot)$is the domain discriminator. The discriminator loss is the binary cross-entropy:

$$L_d(G_d(G_f(x_i)),d_i) = d_i\,log\,\frac{1}{G_d(G_f(x_i))} + (1-d_i)\,log\,\frac{1}{G_d(G_f(x_i))} \tag{8}$$

with d_i the domain label of the i-th sample (used for source/target discrimination).

In summary, DANN minimizes the source-domain label loss L_y to ensure task discriminability, while simultaneously maximizing the domain loss L_d to push G_f toward domain-invariant representations.

2.3. Multi-source domain adaptation

Compared with single-source learning, multi-source learning better matches real industrial scenarios and provides richer prior knowledge. However, source heterogeneity also increases inter-source discrepancy and class-conditional mismatch, making alignment more difficult. As shown in Fig. 1, the gap between any single source and the target is usually hard to remove [27]. Aligning multiple sources with the target may further enlarge the mismatch and degrade performance. Therefore, MSDA aims to make full use of multi-source information while reducing negative transfer and improving class-conditional alignment [28–29].

Representative MSDA strategies can be summarized into five types. First, subdomain or class-level alignment methods use multi-branch or multi-head structures to learn source-specific subspaces and reduce class confusion caused by global alignment. Second, condition-guided dynamic weighting assigns source importance according to conditional discrepancy, so that weak sources are not directly discarded and relevant sources contribute more to transfer. Third, latent-space alignment methods use priors such as WAE to align source-target distributions indirectly and reduce the effect of noise and bias. Fourth, sample-refinement methods improve alignment by reweighting source samples and using self-training with high-confidence pseudo-labels. Fifth, generative balancing methods use diffusion-based models to synthesize shifted samples, which improves training stability, minority-class recognition, and decision boundaries.

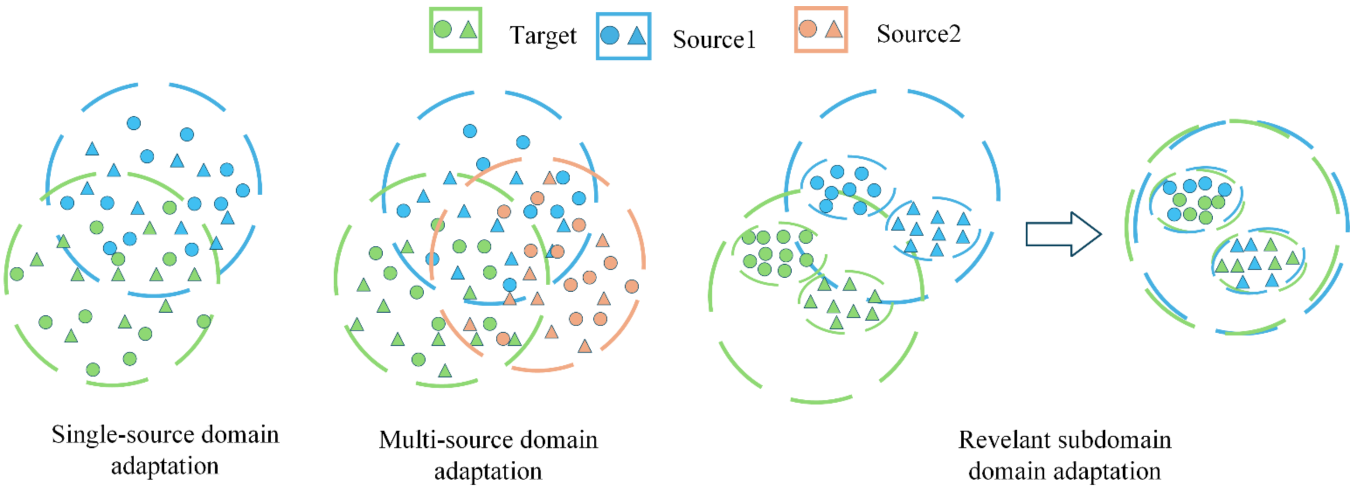


Figure 1. Domain adaptation diagram.

3. The proposed method

3.1. Overall framework

In cross-condition and multi-source transfer tasks, domain alignment alone cannot guarantee class separability. At the same time, different source domains contribute unequally and should be modeled accordingly. Therefore, our framework is built on three principles: (i) adversarial learning for domain-invariant representation; (ii) boundary-aware discrimination for better decision geometry; and (iii) condition-guided multi-source attention for continuous monotonic weighting instead of hard dropping.

The feature extractor adopts a serial CNN-Transformer structure to capture both local impulsive patterns and long-range temporal dependencies. Specifically, the raw vibration signal is first processed by three convolutional stages with kernel sizes of 7, 5, and 3, respectively, to extract local fault-related features and short-term temporal patterns. The extracted convolutional features are then projected into a 256-dimensional embedding and fed into one Transformer encoder layer with four attention heads for global dependency modeling. In this way, the feature extraction process follows a progressive pipeline from local pattern extraction to long-range temporal representation, providing compact and discriminative features for subsequent domain alignment and fault classification.

Building on these advantages, we develop an attention-enhanced adversarial alignment network for multi-source domains (TMDANN-A), whose architecture is shown in Fig. 2. Let C denote the classifier (label predictor), D the domain discriminator, and let a multi-source weight generator produce fusion weights. The forward relations are:

$$h = G(x) \tag{9}$$

$$\tilde{h} = \sum_{k=1}^{K} w_k h^{(k)}, \sum_{k} w_k = 1, w_k \geq 0 \tag{10}$$

$$\tilde{y} = C(\tilde{h}) \tag{11}$$

where $\tilde{h}$ denotes the feature after multi-source fusion and $\tilde{y}$denotes the predicted distribution.

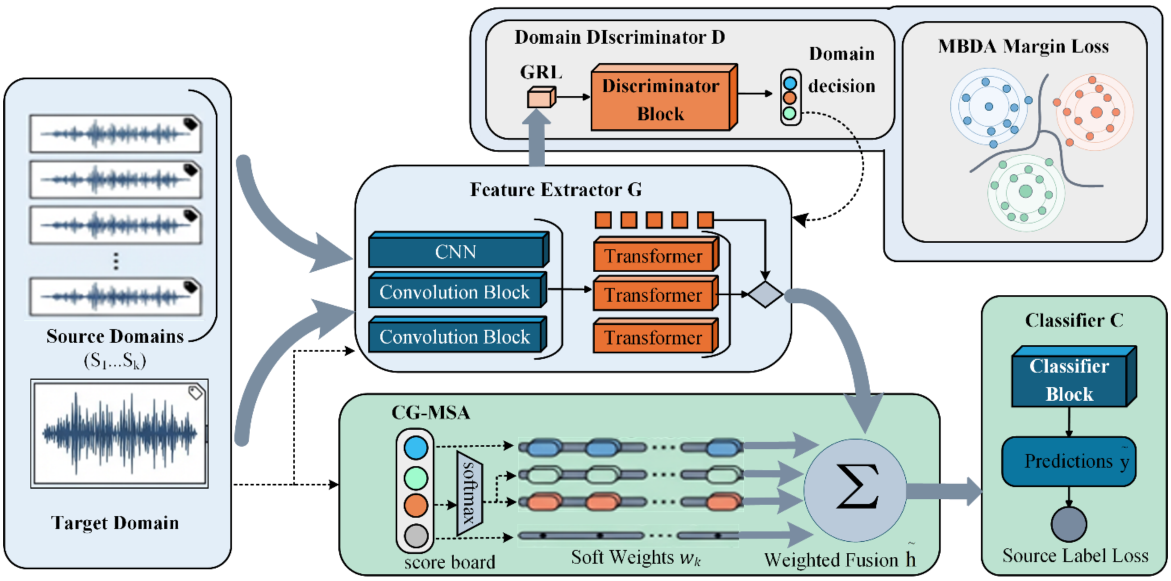


Figure 2. The network structure of TMDANN-A. The feature extractor uses a serial CNN-Transformer pipeline with three CNN stages fol-lowed by one Transformer encoder layer with four attention heads.

3.2. Collaborative alignment mechanism

3.2.1. Domain-level adversarial alignment

We follow the DANN paradigm and introduce a Gradient Reversal Layer (GRL) between the feature extractor G and the domain discriminator D to set up an adversarial game, so that $G(\cdot)$learns domain-invariant features that are indistinguishable to D. This line of work has been widely validated for cross-domain diagnosis of rotating machinery. In the multi-source setting, the adversarial loss is accumulated over each source – target pair to shrink the global domain discrepancy, thereby creating favorable conditions for subsequent class-level alignment. The supervised source classification loss and the domain-adversarial loss are defined as:

$$L_{cls}^{(k)} = -E_{(x,y)\sim D_s^{(k)}} \sum_{c} 1[y = c] \, log \, p_c \left(C(G(x))\right) \tag{12}$$

$$L_{adv}^{(k)} = -E_{x\sim D_s^{(k)}} logD(G(x)) - E_{x\sim D_t} \, log(1 - D(G(x))) \tag{13}$$

3.2.2. Margin-Based Discriminative Alignment (MBDA)

Relying on domain indistinguishability alone is insufficient to avoid mismatches near class boundaries. On top of the adversarial signal, TMDANN-A introduces a boundary-aware discriminative constraint that aims to enlarge inter-class margins while compressing intra-class dispersion—explicitly shaping the decision geometry in feature - decision space so that class-level separability and domain invariance are synergistic rather than conflicting. As illustrated in Fig. 3, MBDA implements this geometry-aware constraint by pulling features toward their class centers while pushing different centers apart, thereby enlarging inter-class margins and compacting intra-class clusters. The boundary-aware discriminative loss is:

$$L_{margin} = \sum_{c} \sum_{x \in c} ||G(x) - \mu_c||_2^2 - \alpha \sum_{c \neq c'} ||\mu_c - \mu_{c'}||_2^2 \quad (14)$$

where α controls the "class-separation" strength in the margin term.

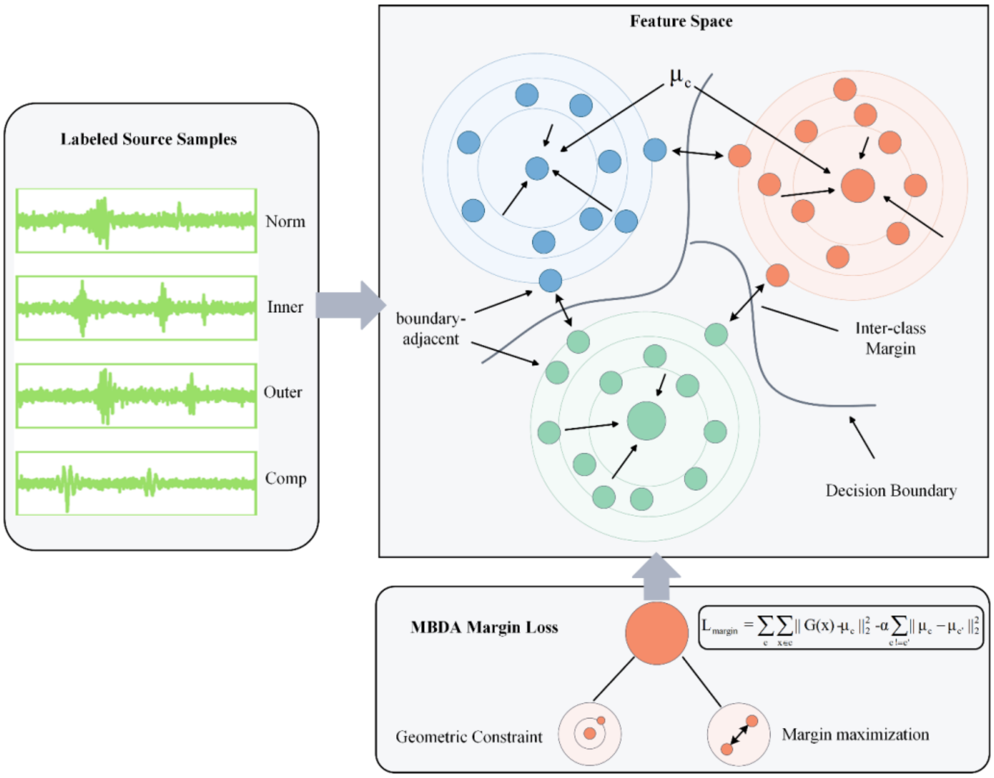


Figure 3. MBDA margin geometry and decision-boundary shaping.

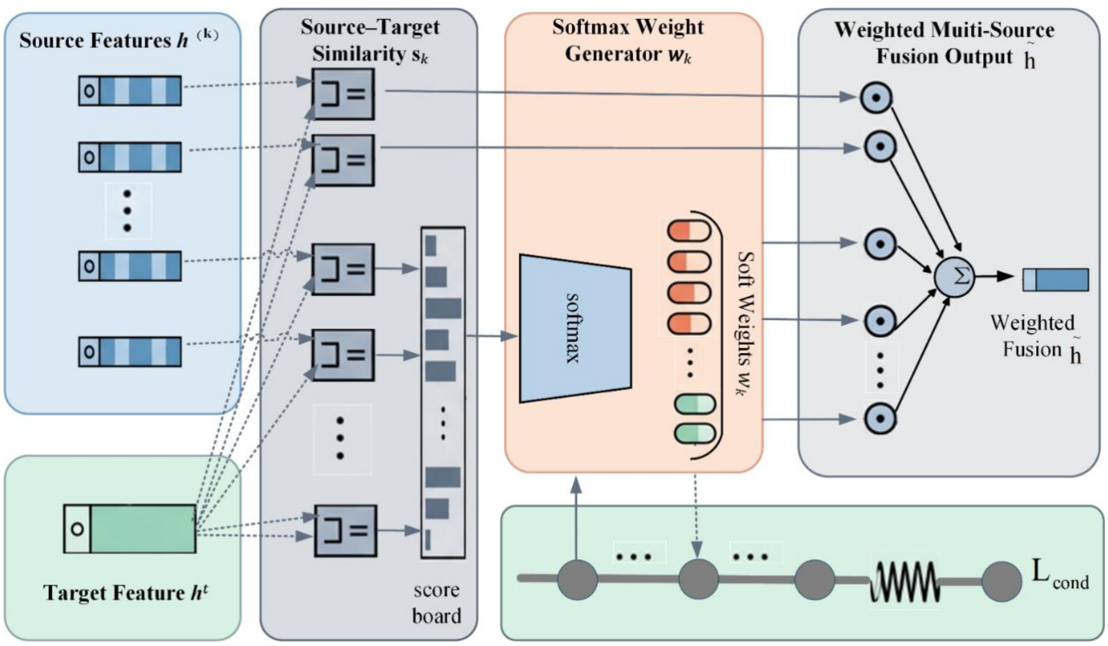


Figure 4. CG-MSA: condition-guided multi-source attention weighting and monotonicity-regularized fusion.

3.2.3. Condition-Guided Multi-Source Attention (CG-MSA)

Multi-source heterogeneity can improve transfer performance, but it can also introduce negative transfer. To handle this, we design a condition-distribution-guided attention mechanism with three parts. First, class-conditional discrepancy is used to estimate the transferability of each source domain, rather than relying only on marginal distributions. Second, each source is assigned a continuous weight instead of being selected or discarded directly. A monotonic regularization term is added to avoid extreme weights and preserve useful low-weight sources while suppressing noise. Third, these weights are used as attention coefficients for end-to-end fusion of source features or source-specific logits, producing a target-oriented adaptive representation. As shown in Fig. 4, CG-MSA calculates source-target class-conditional similarity scores s_k, transforms them into attention weights $w_k$through a temperature-scaled softmax, and performs weighted fusion under the regularization term L_{cond} . The attention weight for the k-th source and the corresponding regularization term are defined as follows:

$$w_k = \frac{exp(\tau s_k)}{\sum_j exp(\tau s_j)} \tag{15}$$

$$L_{cond} = \beta \sum_k w_k log w_k + \gamma \sum_k |w_k - \varphi(s_k)| \tag{16}$$

where w_k is the attention weight of the k-th source, τ is a temperature parameter, s_k denotes the class-conditional similarity score between the k-th source and the target, $\varphi(\cdot)$ is a single-parameter mapping, and β/γ are smoothing/guidance coefficients for the condition attention.

4. Experimental analysis

4.1. Data description

Dataset 1: Jiangnan University Bearing Dataset (JNU)

The JNU bearing dataset from Jiangnan University was collected on the platform shown in Fig. 5. It includes four conditions: inner-race fault, outer-race fault, rolling-element fault, and normal condition. Vibration signals were recorded at 600, 800, and 1000 r/min with a sampling frequency of 50 kHz. The corresponding diagnosis tasks are denoted as J0, J1, and J2, respectively (Table 1).

To reflect practical industrial settings, the dataset is class-imbalanced. The normal class contains 1466 samples, whereas each fault class contains 488 samples. Each sample consists of 1024 points, as shown in Fig. 6.

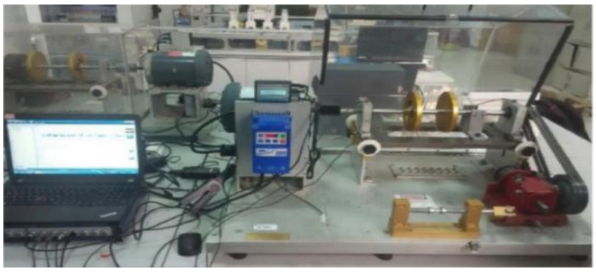


Figure 5. Test rig of JNU dataset.

Table 1. Operating conditions of the JNU dataset.

subset	speed/rpm
J0	600
J1	800
J2	1000

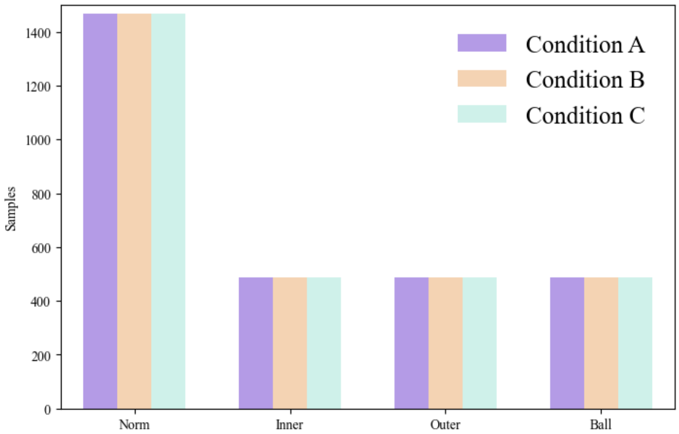


Figure 6. Sample counts under different operating conditions (JNU dataset).

Dataset 2: Paderborn University (PU) Dataset

The Paderborn University (PU) dataset is a benchmark dataset for bearing fault diagnosis [30]. It contains vibration and motor current signals, and only the vibration signals are used in this study. The experimental platform is shown in Fig. 7, and the sampling frequency is 64,000 Hz.

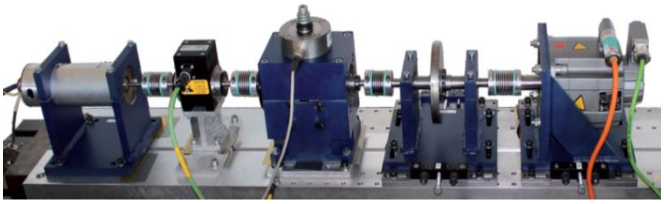


Figure 7. Test rig of PU dataset

Three operating conditions are selected and denoted as P0, P1, and P2 (Table 2). Four health states are considered: normal, outer-race fault, inner-race fault, and compound fault. Each class contains about 500 samples, and each sample consists of 1,024 points, as shown in Fig. 8.

Table 2. Operating conditions of the PU dataset.

Subset	Speed (rpm)	Load torque (Nm)	Radial force (N)
P0	1500	0.7	1000
P1	1500	0.7	400
P2	1500	0.1	1000

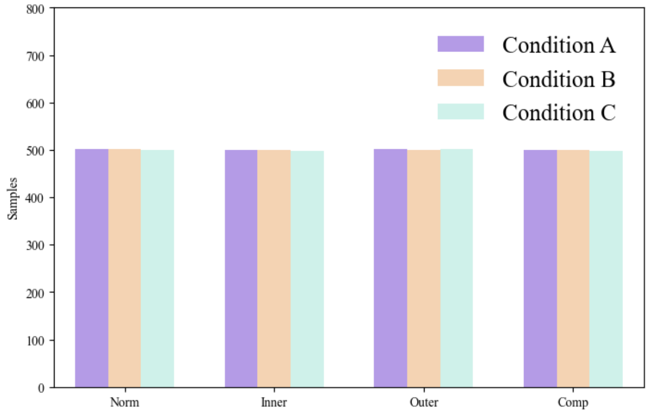


Figure 8. Sample counts under different operating conditions (PU dataset).

Based on the above datasets, we construct the multi-source transfer tasks summarized in Table 3.

Table 3. The illustrations of diagnostic task construction.

Task	Source domain	Target Domain
T1	J0 J1	J2
T2	J0 J2	J1
T3	J1 J2	J0
T4	P0 P1	P2
T5	P0 P2	P1
T6	P1 P2	P0

4.2. Experimental results

To ensure a fair evaluation, TMDANN-A is compared with several representative baselines. All methods use comparable backbones, and hyperparameters are set according to the original literature. For multi-source transfer, methods designed for single-source adaptation report the average of the two corresponding single-source transfer results. The baselines are listed below:

DANN: uses adversarial training with a Gradient Reversal Layer (GRL) to learn domain-invariant features [21].

CORAL: aligns source and target features by matching their covariance matrices [31].

MDD: optimizes margin disparity between two classifiers to improve target discriminability [32].

CDAN: performs conditional adversarial alignment using the joint information of features and class predictions [33].

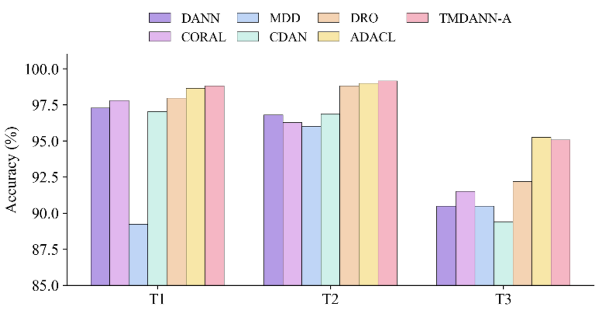
DRO: improves robustness under distribution shift by minimizing worst-case risk over an uncertainty set [34].

ADACL: decomposes multi-source adaptation into multiple single-source tasks and integrates source-specific predictors for target diagnosis [35].

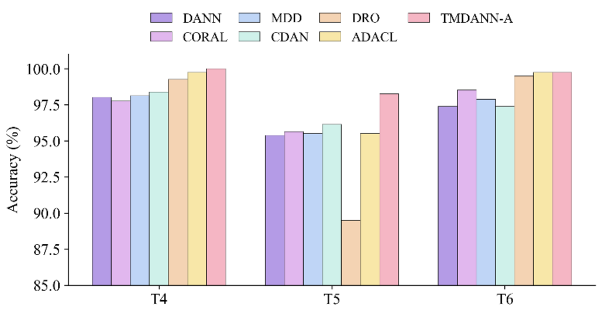
The results are reported in Table 4 and Fig. 9. Overall, the proposed method performs better on multi-source transfer tasks. The t-SNE visualizations and confusion matrices are shown in Fig. 10.

Table 4. Diagnostic accuracy of each method under different transfer tasks.

Method	DANN	CORAL	MDD	CDAN	DRO	ADACL	TMDANN-A
T1(%)	97.28	97.79	89.29	97.03	97.96	98.64	**98.81**
T2(%)	96.78	96.26	96.01	96.86	98.81	98.98	**99.15**
T3(%)	90.48	91.50	90.48	89.38	92.18	**95.24**	95.07
T4(%)	98.01	97.76	98.13	98.38	99.25	99.75	**100**
T5(%)	95.38	95.63	95.50	96.13	89.50	95.50	**98.25**
T6(%)	97.38	98.51	97.88	97.38	99.50	**99.75**	**99.75**
Average	95.88	96.24	94.47	95.86	96.04	98.48	**98.51**

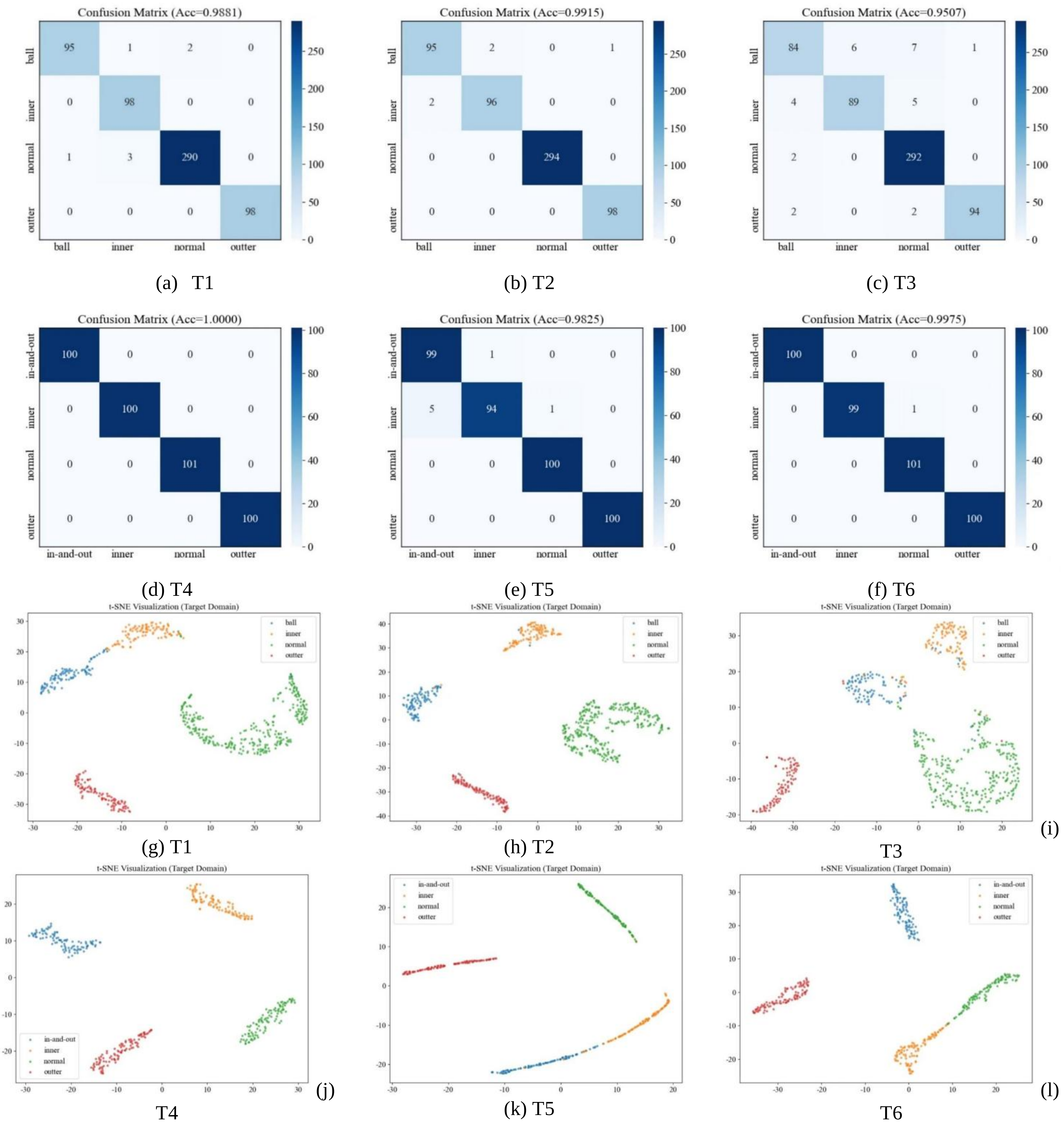


(a) T1-T3



(b) T4-T6

Figure 9. Diagnostic results for all methods.



(a) T1 (b) T2 (c) T3

(d) T4 (e) T5 (f) T6

(g) T1 (h) T2 (i) T3

(j) T4 (k) T5 (l) T6

Figure 10. Visualization Results for T1–T6.

TMDANN-A achieves an average accuracy of 98.51% across six cross-domain tasks, outperforming representative baselines such as DANN (95.88%) and CORAL (96.24%). Its task-wise accuracies range from 95.07% to 100.00%, with only about 3% variation across tasks, indicating stable generalization under different domain shifts and noise levels.

In the JNU cross-speed scenario, T1 and T2 achieve nearly or above 98% accuracy, while T3 (J1, J2→J0) is the most difficult task. In T3, conventional adversarial and boundary-discrepancy methods fall to about 89%-92%, but TMDANN-A still reaches 95.07%. The confusion matrices show that most errors occur between two fault types with similar spectral energy distributions. The t-SNE plots further indicate better source-target alignment, more compact clusters, and clearer class boundaries. This suggests that the proposed method improves class separation while reducing boundary confusion during alignment. Although the normal class is relatively dominant in JNU and may introduce prior bias, TMDANN-A still outperforms all baselines, showing good robustness.

To further examine the fault-type-level performance, we analyzed the confusion matrix of T3, which is the most difficult transfer task in the JNU dataset. The class-wise results show that TMDANN-A achieves F1 scores of 88.42%, 92.23%, 97.33%, and 97.41% for ball fault, inner-race fault, normal condition, and outer-race fault, respectively, with a macro-F1 of 93.85%. The recalls of ball fault, inner-race fault, and outer-race fault are 85.71%, 90.82%, and 95.92%, respectively, and the average fault recall reaches 90.82%. These results indicate that TMDANN-A maintains effective fault detection ability under the most difficult cross-speed transfer task. The normal and outer-race classes are classified more accurately, while most misclassifications occur between ball, inner-race, and normal samples with similar vibration responses. This suggests that the proposed model improves class-level diagnostic performance, although fine-grained confusion between similar fault types still exists.

On the PU dataset, where each class contains about 500 samples, TMDANN-A achieves an average accuracy above 99% across three multi-source tasks and reaches 100% on T4. This is consistent with the condition-guided multi-source attention mechanism, which increases the contribution of relevant sources while suppressing noisy ones. The bar-chart results further show that the proposed attention mechanism provides more stable gains than equal-weight statistical alignment and single-discriminator adversarial alignment under load and radial-force variations.

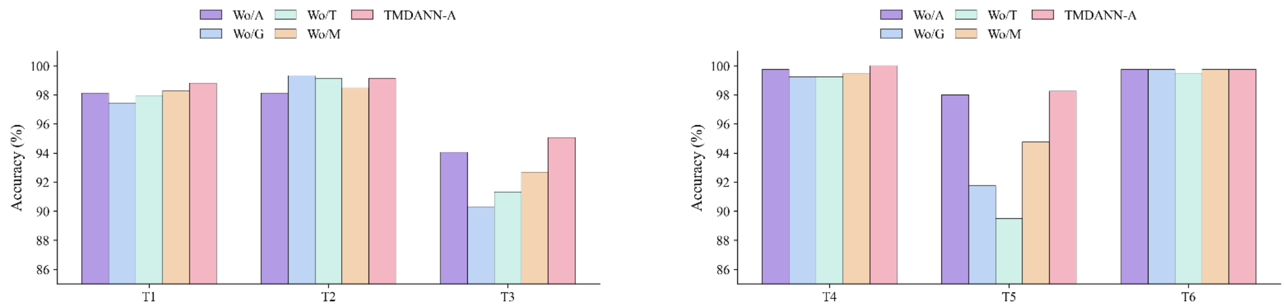
Across all six tasks, the confusion matrices show strong diagonal dominance, with only a few off-diagonal errors. Most errors occur between defect pairs with similar physical characteristics. The t-SNE plots further show good source-target alignment and clear, compact clusters with little overlap. These results suggest that the proposed method improves both domain invariance and class separability.

JNU reflects the common imbalance of many normal samples and few fault samples, whereas PU is approximately class-balanced. TMDANN-A performs well and stably on both datasets, indicating that it does not rely on a specific class-ratio structure and has good engineering applicability. In multi-source cross-speed and cross-load tasks, its adaptive weighting mechanism also supports effective fusion across different operating conditions.

Overall, TMDANN-A outperforms the baselines in accuracy, performance on difficult transfer tasks, resistance to negative transfer, and feature separability. Its consistent superiority on two datasets with different class distributions further demonstrates its generalization and robustness under complex operating conditions and multi-source scenarios.

Table 5. The diagnostic results of ablation study.

Method	Wo/A	Wo/G	Wo/T	Wo/M	TMDANN-A
T1(%)	98.13	97.45	97.96	98.30	**98.81**
T2(%)	98.13	**99.32**	99.15	98.47	99.15
T3(%)	94.05	90.31	91.33	92.69	**95.07**
T4(%)	99.75	99.25	99.25	99.50	**100**
T5(%)	98.00	91.75	89.50	94.75	**98.25**
T6(%)	**99.75**	**99.75**	99.50	**99.75**	**99.75**
Average	97.97	96.31	96.12	97.24	**98.51**



(a) T1-T3 (b) T4-T6

Figure 11. Schematic representation of ablation results.

4.3. Ablation studies

To assess the independent contribution of each module, we conduct four ablation variants by removing the attention module (WO-A), the Gradient Reversal Layer (WO-G), the Transformer encoder (WO-T), and the MBDA module (WO-M), respectively. The results are summarized in Table 5, and the accuracy histograms are shown in Fig. 11. The main findings are:

1. Without GRL, the six-task average accuracy drops from 98.51% to 96.31% (−2.20%), with the largest decline on T5, from 98.25% to 91.75%. This shows that adversarial alignment is crucial for reducing source-target discrepancy and supporting cross-domain generalization.
2. Without the Transformer, the average accuracy decreases to 96.12% (−2.39%), with larger drops on T5 (89.50%) and T3 (91.33%). This suggests that local convolution alone is insufficient, while the CNN-Transformer structure better captures global dependencies under cross-condition shifts.
3. Without MBDA, the average accuracy falls to 97.24% (−1.27%). The effect is more obvious on difficult tasks, where T3 drops from 95.07% to 92.69% and T5 from 98.25% to 94.75%. This indicates that MBDA helps improve class separation and reduce confusion between similar fault types.
4. Without the attention module, the average accuracy decreases slightly, indicating that CG-MSA has a more moderate contribution than GRL and the Transformer. To verify that this gain is not caused by random fluctuation, we further repeated TMDANN-A and WO-A ten times with different random seeds. As shown in Table 6, TMDANN-A achieves a higher average mean accuracy than WO-A, while the standard deviations remain within a small range. This indicates that the attention module provides a stable improvement for multi-source fusion, although its effect is less pronounced on easier tasks where the accuracy is already close to saturation.

The full model achieves the best or tied-best results on all six tasks, including 100% on T4. The tied results on T6 are mainly due to the ceiling effect. Overall, GRL and the Transformer provide the basis for domain alignment and global feature modeling, MBDA improves performance on difficult boundary-mixing tasks, and multi-source attention brings stable gains under stronger source heterogeneity. Together, these components improve domain invariance and class separability, resulting in the highest average accuracy of 98.51%.

Table 6. Ten-run statistical comparison between TMDANN-A and WO-A.

Task	Wo/A	TMDANN-A
T1 (%)	98.06 ± 0.61	98.84 ± 0.25
T2 (%)	98.18 ± 0.55	99.12 ± 0.24
T3 (%)	94.12 ± 0.66	95.04 ± 0.26
T4 (%)	99.50 ± 0.49	99.85 ± 0.21
T5 (%)	97.95 ± 0.91	98.30 ± 0.24
T6 (%)	99.60 ± 0.47	99.75 ± 0.24

4.4. Sensitivity analysis of key hyperparameters

To further evaluate the influence of key hyperparameters on cross-task transfer performance, we conducted sensitivity analyses from two aspects: the CNN-Transformer feature extractor and the condition-guided multi-source attention module. A single-factor experimental setting was adopted, where only one hyperparameter was changed at a time while the others were kept at their default values. Two representative transfer tasks, T3 and T5, were selected for evaluation. T3 is the most difficult JNU cross-speed task, while T5 shows an obvious performance decrease when the Transformer encoder is removed in the ablation study.

Table 7. Sensitivity analysis of CNN-Transformer feature extractor parameters.

Config	CNN stages	Transformer layers	Heads	T3 Acc. (%)	T5 Acc. (%)
S1	2	1	4	90.82	90.00
S2	3	1	4	95.07	98.25
S3	4	1	4	92.69	95.25
S4	3	0	4	91.33	89.50
S5	3	2	4	91.67	94.25
S6	3	1	2	91.33	97.75
S7	3	1	8	90.82	92.75

As shown in Table 7, removing the Transformer encoder leads to a clear performance decrease, confirming the importance of global dependency modeling after local CNN feature extraction. Increasing the number of CNN stages or Transformer layers does not consistently improve accuracy, indicating that deeper structures may introduce redundant parameters and increase optimization difficulty under limited source-domain samples. The four-head setting provides stable performance, while too few or too many heads may weaken representation diversity or introduce redundant attention patterns. Overall, the default setting with three CNN stages, one

Transformer encoder layer, and four attention heads achieves a balanced trade-off between transfer accuracy and model complexity.

Table 8. Sensitivity analysis of CG-MSA parameters.

Config	τ	β	γ	T3 Acc. (%)	T5 Acc. (%)
C1	**0.5**	**0.1**	**0.01**	91.16	91.50
C2	**1.0**	**0.1**	**0.01**	95.07	98.25
C3	**2.0**	**0.1**	**0.01**	90.99	96.25
C4	**1.0**	**0**	**0.01**	93.37	93.75
C5	**1.0**	**0.3**	**0.01**	89.29	93.00
C6	**1.0**	**0.1**	**0**	91.50	93.25
C7	**1.0**	**0.1**	**0.05**	90.14	96.00

As shown in Table 8, the CG-MSA parameters affect the stability of multi-source fusion. The temperature parameter τ controls the sharpness of source-domain attention weights. A smaller τ tends to produce more concentrated source weights, which may overemphasize a single source domain, whereas a larger τ makes the weights smoother and may weaken source discrimination. The smoothing coefficient β stabilizes source weights across batches; however, removing β or setting it too large both reduce performance, indicating that moderate smoothing is beneficial for stable fusion. The regularization coefficient γ suppresses excessive concentration of source weights, but overly weak or overly strong regularization may degrade fusion performance. Among the tested settings, the default configuration (τ=1.0, β=0.1, and γ=0.01) achieves the best performance on both T3 and T5, indicating a stable balance between source discrimination, weight smoothing, and regularized fusion.

4.5. Additional experiments

The data in this section were collected from a bearing condition monitoring test rig at a special equipment inspection institute (Fig. 12). The rig supports adjustable load and rotational speed, allowing signal acquisition under different operating conditions. The test bearing is a cylindrical roller bearing (NJ312EM). Four health conditions are considered: normal, inner-race fault, outer-race fault, and rolling-element fault. Artificial defects were introduced by wire electrical discharge machining (WEDM). A 3 mm-wide notch with uniform size was machined at the corresponding locations to simulate localized damage (Fig. 13), ensuring controllable defect morphology and good repeatability.

Based on the results on the public datasets (JNU and PU), we further test the proposed method on the rig dataset to evaluate its generalization under more practical operating conditions. Unlike JNU, which mainly reflects cross-speed transfer with class imbalance, and PU, which focuses on cross-load transfer with relatively balanced classes, the rig dataset includes coupled variations in both speed and load. Three operating conditions are defined: Condition A (74.25 N·m, 121.5 r/min), Condition B (668.25 N·m, 151.9 r/min), and Condition C (816.75 N·m, 182.3 r/min), representing low-, medium-, and high-level operating regimes.

For each operating condition, a longer vibration signal was first collected for each health state and then segmented into samples with a length of 1024 points. In the final experimental setting, about 500 samples were selected for each health state under each condition. This setting was adopted to keep the sample scale comparable to the JNU and PU datasets used in this study, thereby avoiding an unfair advantage caused by a much larger in-house dataset. Although this setting supports controlled comparison across datasets, the limited number of health states and samples is still a limitation for evaluating large-scale industrial generalization.

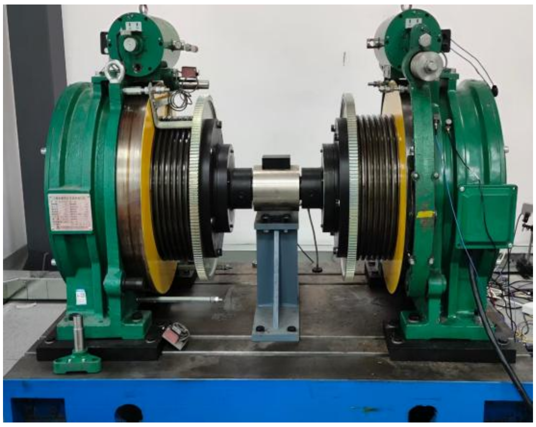


Figure 12. Experimental test rig.

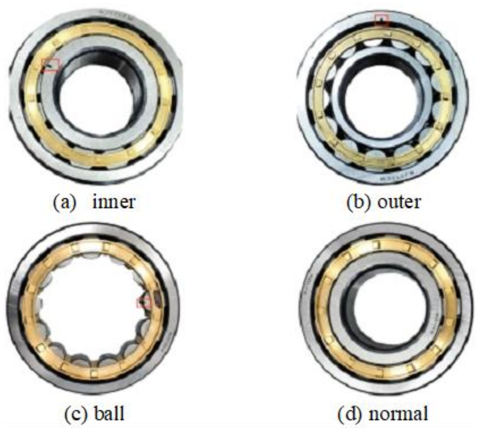


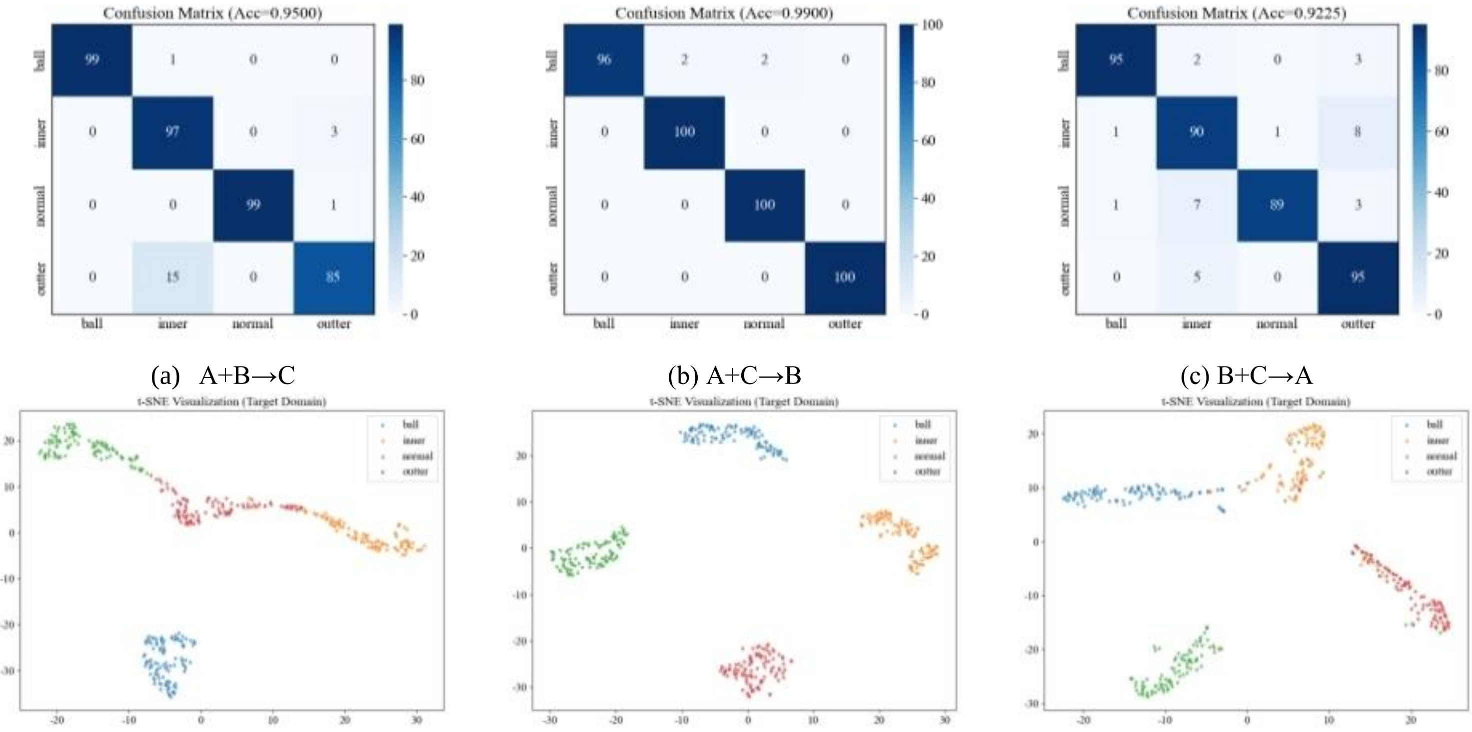
Figure 13. Examples of normal and faulty bearing conditions.

Table 9. Comparative results under coupled speed–load variations for multi-source domain adaptation.

Source domain	Target domain	DANN	CORAL	MDD	CDAN	DRO	ADACL	TMDANN-A
A+ B	C	92.00	92.00	93.50	92.13	91.25	93.25	**95.00**
A+ C	B	97.50	89.63	98.50	96.00	97.75	98.50	**99.00**
B+ C	A	87.75	84.75	88.00	88.88	87.25	91.75	**92.25**
Average		92.42	88.79	93.33	92.33	92.08	94.50	**95.42**

Table 10. Comprehensive diagnostic metrics for multi-source cross-condition transfer on the test-rig dataset.

Source domain	Target domain	Acc	Macro-F1	Fault-Recall (ball/inner/outer)
A+ B	C	**95.00**	95.01	93.67
A+ C	B	**99.00**	98.99	98.67
B+ C	A	**92.25**	92.32	93.33
Average		**95.42**	95.44	95.22



(a) A+B→C (b) A+C→B (c) B+C→A

(d) A+B→C (e) A+C→B (f) B+C→A

Figure 14. Confusion matrices and t-SNE visualizations on the test-rig dataset.

Table 9 summarizes the results on three multi-source transfer tasks (A+B→C, A+C→B, and B+C→A). Conventional alignment-based methods, including DANN, CORAL, MDD, and CDAN, show clear performance fluctuations across different target conditions. When the target domain differs substantially from the source domains in coupled speed-load settings, these methods tend to suffer from insufficient alignment or weakened class boundaries, resulting in lower accuracy. DRO and ADACL perform better on some tasks, suggesting that robust optimization or attention mechanisms can partly reduce the effect of operating-condition shifts. However, TMDANN-A achieves the best performance on all three tasks, with accuracies of 95.00%, 99.00%, and 92.25%, respectively, and an average of 95.42%. Compared with the strongest baseline, it also shows better stability, indicating stronger generalization under coupled speed-load variations.

Because accuracy alone may hide missed detections, we further report Macro-F1 and average fault recall (Fault-Recall for ball, inner, and outer faults) as complementary metrics (Table 10). TMDANN-A achieves an average Macro-F1 of 95.44%, close to its average accuracy, indicating balanced improvement across all classes. Its mean Fault-Recall reaches 95.22%, showing strong fault detection capability under different target conditions. This is consistent with the practical requirement of reducing missed detections in condition monitoring.

Fig. 14 further shows the confusion matrices and t-SNE visualizations of TMDANN-A for the three transfer tasks. In all tasks, the normal class is clearly separable, while most errors occur among fault categories. This fine-grained confusion is likely related to differences in transmission paths and impact-response patterns under different speed-load combinations. In the most difficult setting, such as B+C→A, some fault categories still overlap, but accuracy and Macro-F1 remain high. This indicates that TMDANN-A remains robust under strong domain shifts caused by coupled operating-condition variations. Overall, the three datasets represent three typical transfer scenarios: cross-speed transfer with class imbalance (JNU), cross-load transfer with relatively balanced classes (PU), and coupled speed-load variation under compound conditions (the in-house rig dataset). TMDANN-A performs consistently well under the tested public benchmark and industrial test-rig conditions, suggesting its potential for cross-condition fault diagnosis. However, the current test-rig dataset is still based on artificially machined localized defects with relatively controlled morphology. Therefore, these results should be interpreted within the tested conditions rather than as a complete validation of real industrial deployment.

4.6. Noise robustness analysis on the industrial test-rig dataset

To further evaluate the robustness of the proposed method under industrial noise interference, additive Gaussian white noise was added to the target-domain test samples of the industrial test-rig dataset. Three noisy conditions with signal-to-noise ratios of 10 dB, 5 dB, and 0 dB were considered, together with the clean condition. The three multi-source transfer tasks, A+B→C, A+C→B, and B+C→A, were evaluated. DRO and ADACL were selected as representative comparison methods because they show relatively strong performance on the industrial test-rig dataset and are closely related to robust or adaptive transfer learning.

Table 11. Noise robustness comparison on the industrial test-rig dataset under different SNR levels.

Source domain	Target domain	Noise（dB）	DRO	ADACL	TMDANN-A
A+ B	C	Clean	91.25	93.25	**95.00**
		10	90.25	91.25	**93.75**
		5	88.25	90.50	**92.00**
		0	79.75	85.75	**87.00**
A+ C	B	Clean	97.75	98.50	**99.00**
		10	96.75	98.25	**98.75**
		5	96.00	96.00	**96.75**
		0	78.75	81.00	**87.50**
B+ C	A	Clean	87.25	91.75	**92.25**
		10	77.25	87.75	**90.00**
		5	63.75	83.00	**82.75**
		0	51.25	69.25	**70.00**
Average		Clean	92.08	94.50	**95.42**
		10	88.08	92.42	**94.17**
		5	82.67	89.83	**90.50**
		0	69.92	78.67	**81.50**

As shown in Table 11, the performance of all methods decreases as the noise intensity increases, indicating that strong noise still affects feature extraction and domain alignment. However, TMDANN-A maintains the highest average accuracy under Clean, 10 dB, 5 dB, and 0 dB conditions. In particular, under the severe 0 dB condition, TMDANN-A achieves an average accuracy of 81.50%, which is higher than ADACL and DRO by 2.83% and 11.58%, respectively. This suggests that the proposed condition-guided multi-source attention and boundary-aware alignment improve noise tolerance to some extent. Nevertheless, the results also show that the accuracy decreases under strong noise, especially in the B+C→A task, indicating that the proposed method still has limitations under extremely low signal-to-noise ratios and more complex industrial disturbances.

4.7. Computational cost and inference efficiency

To further evaluate the feasibility of the proposed method for industrial monitoring, the computational cost and inference efficiency of different methods were compared. The number of trainable parameters and the average inference time per sample were reported. All methods were tested under the same hardware environment with an NVIDIA GeForce RTX 5060

GPU. The input signal length was 1024, and the batch size was set to 1 to simulate single-sample online diagnosis. Before timing, each model was warmed up for 100 iterations, and the average inference time was calculated over 1000 forward propagations. The reported inference time only includes the forward propagation stage.

Table 12. Computational cost and inference efficiency comparison on the industrial test-rig dataset.

Method	Params (M)	Inference time (ms/sample)	Acc. (%)
DRO	0.0863	0.5117	92.08
ADACL	0.1039	0.6211	94.50
TMDANN-A	0.9194	0.9317	95.42

As shown in Table 12, TMDANN-A has more trainable parameters than the comparison methods because of the CNN-Transformer feature extractor and the condition-guided multi-source attention module. Nevertheless, its average inference time is 0.9317 ms per sample under the tested hardware environment, which remains below 1 ms. Meanwhile, TMDANN-A achieves the highest average accuracy of 95.42% on the industrial test-rig dataset. These results indicate that the proposed method improves diagnostic accuracy with an acceptable increase in computational cost under the tested GPU environment.

5. Conclusion

This paper investigates intelligent fault diagnosis of rotating machinery under complex operating conditions, with empha-sis on three major challenges in multi-source transfer learning: domain shift, source heterogeneity, and unclear class bounda-ries. To address these issues, an attention-enhanced adversarial alignment network, TMDANN-A, is proposed. The model jointly integrates domain alignment, boundary-aware discrimination, and multi-source coordination, leading to improved diagnostic performance across different conditions and domains. The main conclusions are as follows.

1. Based on the DANN framework, a hybrid CNN-Transformer feature extractor is constructed to capture both local im-pulsive patterns and long-range temporal dependencies in bearing vibration signals. With the aid of the Gradient Rever-sal Layer (GRL), domain-level adversarial learning is achieved, which helps reduce distribution shifts under different op-erating conditions.
2. A margin-based discriminative alignment (MBDA) module is introduced to regularize decision boundaries in the feature space. By enlarging inter-class margins and enhancing intra-class compactness, it reduces confusion between similar fault types and improves robustness for difficult samples, including minority and boundary samples.
3. To handle heterogeneous source domains and unequal source contributions, a condition-guided source attention mech-anism is designed to assign source weights according to class-conditional similarity. This enables continuous multi-source fusion and helps suppress negative transfer in complex transfer tasks.
4. Experiments on two public datasets (JNU and PU) and an industrial test-rig dataset show that TMDANN-A consistently outperforms representative methods, including DANN, CDAN, CORAL, MDD, DRO, and ADACL, in multi-source cross-condition transfer tasks. On the public datasets, the proposed method improves average accuracy by about 2%-4%. On the test-rig dataset with coupled speed-load variations, it also maintains strong and stable performance. The t-SNE re-sults and confusion matrices further show clearer class separation and better cross-domain alignment.
5. Ablation results indicate that adversarial alignment remains the main driver of cross-domain transfer, while the Trans-former and MBDA play important roles in global dependency modeling and boundary optimization. The attention mechanism further improves multi-source fusion by balancing source contributions. Together, these modules contribute to the overall effectiveness of TMDANN-A.

In summary, TMDANN-A improves domain invariance, class separability, and robustness under multi-source heterogene-ous conditions. Its performance on public datasets and the industrial test-rig dataset suggests that TMDANN-A has potential for fault diagnosis under practical operating conditions with compounded variations. However, its robustness under stronger noise, sensor errors, compound faults, progressive degradation, and unseen industrial conditions still requires further investi-gation. Future work will extend the current vibration-based framework to multi-modal inputs, including time-frequency im-ages, infrared thermography, maintenance logs, and expert annotations, with the aim of improving fault representation and cross-domain robustness through signal-image-text fusion.

References

1. Gao Y, Li X. Fault diagnosis of centrifugal fan bearings based on I-CNN and JMMD in the context of sample imbalance. *Eksploatacja i Niezawodnosc-Maintenance and Reliability* 2024; 26(4): 191459. https://doi.org/10.17531/ein/191459.
2. Wang P, Xiong H, He H. Bearing fault diagnosis under various conditions using an incremental learning-based multi-task shared classifier. *Knowledge-Based Systems*, 2023, 266:110395. https://doi.org/10.1016/j.knosys.2023.110395.
3. Liu H, Shao M, Fu Y. Structure-preserved multi-source domain adaptation. In: 2016 IEEE 16th International Conference on Data Mining; Montreal, QC, Canada. Piscataway, NJ: IEEE; 2016. p. 1059–1064. doi: 10.1109/ICDM.2016.0136.
4. Lei X, Lu N, Jiang B, Wang C, Chen C. A multi-scale attention mechanism diagnosis method with adaptive online updating based on deep learning under variable working conditions. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2025; 27(1): 192975. https://doi.org/10.17531/ein/192975.
5. Qian W, Li S, Jiang X. Deep transfer network for rotating machine fault analysis. *Pattern Recognition*, 2019, 96:106993. https://doi.org/10.1016/j.patcog.2019.106993.
6. Lu X, Xiao Y. Fault diagnosis method of automobile rolling bearing based on transfer learning and improved DenseNet. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2025; 27(2): 194675. https://doi.org/10.17531/ein/194675.
7. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. *Communications of the ACM* 2020; 63(11): 139-144. https://doi.org/10.1145/3422622.
8. Jiao JY, Lin J, Zhao M, Liang KX. Double-level adversarial domain adaptation network for intelligent fault diagnosis. *Knowledge-Based Systems*, 2020, 205:106236. https://doi.org/10.1016/j.knosys.2020.106236.
9. Deng Y, Huang D, Du S, Li G, Zhao C, Lv J. A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis. *Computers in Industry* 2021; 127: 103399. https://doi.org/10.1016/j.compind.2021.103399.
10. Zhang J, Kong X, Cheng L, Qi H, Yu M. Intelligent fault diagnosis of rolling bearings based on continuous wavelet transform-multiscale feature fusion and improved channel attention mechanism. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 2023; 25(1): 16. https://doi.org/10.17531/ein.2023.1.16.
11. Wang X, Jiang H, Wu Z, Yang Q. Adaptive variational autoencoding generative adversarial networks for rolling bearing fault diagnosis. *Advanced Engineering Informatics* 2023; 56: 102027. https://doi.org/10.1016/j.aei.2023.102027.
12. Li X, Li S, Zhang G, Xie Y, Wang T, Chu F. A novel interpretable dynamic weighted domain adaptation network for cross-domain fault diagnosis of bearings under time-varying speeds. *Engineering Applications of Artificial Intelligence*, 2026, 170: 114173. https://doi.org/10.1016/j.engappai.2026.114173.
13. Li X, Wu X, Yu J, Wang T, Shi C, Choi J-H, Chu F. A fault diagnosis data augmentation method integrating multimodal non-Gaussian denoising diffusion generative adversarial network. *Advanced Engineering Informatics*, 2025, 68: 103776. https://doi.org/10.1016/j.aei.2025.103776.
14. Wang X, Jiang H, Dong Y, Mu M. Spatial-channel collaborative multi-scale graph interaction deep transfer learning for unsupervised rotating machinery fault diagnosis. *Engineering Applications of Artificial Intelligence,* 2026, 176: 114691. https://doi.org/10.1016/j.engappai.2026.114691.
15. Dong Y, Jiang H, Wang X, Li Z. Entropy-oriented semi-supervised dynamic prototype contrastive learning for rotating machinery fault diagnosis. *IEEE/ASME Transactions on Mechatronics*, 2025, 30(6): 4934-4945. https://doi.org/10.1109/TMECH.2025.3570186.
16. Li Z, Jiang H, Dong Y. A convolutional-transformer reinforcement learning agent for rotating machinery fault diagnosis. *Expert Systems With Applications*, 2025, 271: 126669. https://doi.org/10.1016/j.eswa.2025.126669.
17. Ma T, Li Z, Yang D, Tang J, Liu J. A multi-source subdomain adaptation fault diagnosis method based on unidirectional movement of the target domain. *Measurement Science and Technology* 2024; 35(5): 056117. https://doi.org/10.1088/1361-6501/ad2968.
18. Si J, Shi H, Han T, Chen J, Zheng C. Learn generalized features via multi-source domain adaptation: intelligent diagnosis under variable/constant machine conditions. *IEEE Sensors Journal* 2022; 22(1): 510-519. https://doi.org/10.1109/JSEN.2021.3126864.

19. Cao W, Meng Z, Sun D, Liu J, Guan Y, Cao L, Li J, Fan F. A two-stage domain alignment method for multi-source domain fault diagnosis. *Measurement* 2023; 214: 112818. https://doi.org/10.1016/j.measurement.2023.112818.
20. Peng X, Bai Q, Xia X, Huang Z, Saenko K, Wang B. Moment matching for multi-source domain adaptation. In: 2019 IEEE/CVF International Conference on Computer Vision; Seoul, Republic of Korea. Piscataway, NJ: IEEE; 2019. p. 1406-1415. https://doi.org/10.1109/ICCV.2019.00149.
21. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks. *Journal of Machine Learning Research* 2016; 17(59): 1-35.
22. Shen J, Qu Y, Zhang W, Yu Y. Wasserstein distance guided representation learning for domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence* 2018; 32(1): 4058-4065. https://doi.org/10.1609/aaai.v32i1.11784.
23. Zhu Y, Zhuang F, Wang D. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. *Proceedings of the AAAI Conference on Artificial Intelligence* 2019; 33(1): 5989-5996. doi: 10.1609/aaai.v33i01.33015989.
24. Schwendemann S, Amjad Z, Sikora A. Bearing fault diagnosis with intermediate domain based Layered Maximum Mean Discrepancy: a new transfer learning approach. *Engineering Applications of Artificial Intelligence*, 2021, 105:104415. https://doi.org/10.1016/j.engappai.2021.104415.
25. Xu R, Chen Z, Zuo W, Yan J, Lin L. Deep cocktail network: multi-source unsupervised domain adaptation with category shift. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; Salt Lake City, UT, USA. Piscataway, NJ: IEEE; 2018. p. 3964-3973. https://doi.org/10.1109/CVPR.2018.00417.
26. Lu J, Zuo H, Zhang G. Fuzzy multiple-source transfer learning. *IEEE Transactions on Fuzzy Systems* 2020; 28(12): 3418-3431. https://doi.org/10.1109/TFUZZ.2019.2952792.
27. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems 27. Red Hook, NY: Curran Associates; 2014. p. 3320-3328.
28. Li X, Su K, He Q, Wang X, Xie Z. Research on fault diagnosis of highway Bi-LSTM based on attention mechanism. *Eksploatacja i Niezawodnosc-Maintenance and Reliability*, 2023, 25(2): 162937. https://doi.org/10.17531/ein/162937.
29. Wan L, Li Y, Chen K, Gong K, Li C. A novel deep convolution multi-adversarial domain adaptation model for rolling bearing fault diagnosis. *Measurement*, 2022, 191:110752. https://doi.org/10.1016/j.measurement.2022.110752.
30. Lessmeier C, Kimotho JK, Zimmer D, Sextro W. 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* 2016; 3(1): 376-383. https://doi.org/10.36001/phme.2016.v3i1.1577.
31. Sun B, Saenko K. Deep CORAL: correlation alignment for deep domain adaptation. In: Hua G, Jégou H, editors. Computer Vision—ECCV 2016 Workshops, Part III. Lecture Notes in Computer Science, vol. 9915. Cham: Springer; 2016. p. 443-450. https://doi.org/10.1007/978-3-319-49409-8_35.
32. Zhang Y, Liu T, Long M, Jordan MI. Bridging theory and algorithm for domain adaptation. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th International Conference on Machine Learning. Proceedings of Machine Learning Research. 2019; 97: 7404-7413.
33. Long M, Cao Z, Wang J, Jordan MI. Conditional adversarial domain adaptation. In: Advances in Neural Information Processing Systems 31. Red Hook, NY: Curran Associates; 2018. p. 1647-1657.
34. Rahimian H, Mehrotra S. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization* 2022; 3: 4. https://doi.org/10.5802/ojmo.15.
35. Zhang Y, Ren Z, Zhou S, Yu T. Adversarial domain adaptation with classifier alignment for cross-domain intelligent fault diagnosis of multiple source domains. *Measurement Science and Technology* 2021; 32(3): 035102. https://doi.org/10.1088/1361-6501/abcad4.