Adversarial Domain Adaptation Fault Diagnosis Method Based on Self-attention Graph Convolutional Network

Bo Zhang[1], Shuai Su[2], Ning Ma[3], Yingxue Wang[4] and Wei Li[5]

[1,2,3,4] *School of Computer and Technology, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China*
zbcumt@163.com
ts23170050a31@cumt.edu.cn
1390440583@qq.com
w132931@163.com

[5] *School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China*
liwei_cmee@163.com

ABSTRACT

Intelligent fault diagnosis has made significant progress with the advancements in deep learning and big data. However, the assumption of identical training and testing data distributions often fails in dynamic industrial environments, leading to performance degradation. To address this issue, we propose an Adversarial Domain Adaptation Fault Diagnosis Model Based on Self-attention Graph Convolutional Network (ADA-SAG). The model employs the k-nearest neighbors algorithm to construct graph structures that capture fault-instance relationships across source and target domains. A self-attention enhanced graph convolutional network extracts critical features, while a dual-classifier framework, combined with adversarial learning and maximum mean discrepancy regularization, ensures domain-invariant feature alignment. Experimental results on two benchmark datasets show that the proposed model achieves higher accuracy and robustness compared to existing methods, making it suitable for diverse operating conditions. Ablation studies further validate the contributions of each component to the overall effectiveness of the model.

1. INTRODUCTION

With the increasing complexity of industrial systems, advanced fault diagnosis algorithms have become essential to ensure reliability and reduce maintenance costs. Deep learning has significantly enhanced diagnostic accuracy by enabling efficient feature extraction(Z. Chen et al., 2023; Y. Chen et al., 2023; Tang et al., 2023; Liang, Deng, Yuan, & Zhang, 2023). However, its effectiveness heavily depends on stable operating conditions(H. Li et al., 2024; Guo et al., 2021; Yan et al., 2020). In real-world industrial environments, frequent variations in equipment states cause significant discrepancies between training (source domain) and testing (target domain) data distributions, which severely impact model performance.

To address domain discrepancies and enhance robustness and generalizability in fault diagnosis models, researchers have proposed various domain adaptation methods, broadly categorized into feature-based and instance-based approaches to address specific challenges in domain adaptation. This discrepancy undermines the performance of deep learning models, causing domain distribution discrepancies(L. Chen et al., 2021; W. Li, Yuan, Sun, & Liu, 2020; Zhao, Liu, Shen, & Gao, 2021a; Wang, Huang, Wang, Shen, & Zhu, 2022). Feature-based methods focus on learning domain-invariant feature representations(Du et al., 2019; C. Chen, Chen, Jiang, & Jin, 2019; Long, Cao, Wang, & Jordan, 2015). For example, Zhao et al.(Zhao, Liu, Shen, & Gao, 2021b) developed a multi-representation domain adaptation network that achieves adaptation through domain-invariant feature extraction. Lu et al.(Lu, Fan, Zeng, Li, & Chen, 2022) introduced a self-supervised domain adaptation approach that adjusts the feature distributions of two domains through domain adversarial training and MMD. Ma et al.(Ma, Zhang, Fan, & Wang, 2020) developed a diagnostic framework that focuses on feature-based domain adaptation methods to adapt to different domains. Instance-based methods reweight source samples to align with the target domain distribution(Deng et al., 2022; Gong, Yu, & Xia, 2020). Zhu et al.(Zhu, Shi, Feng, & Tang, 2023) proposed a domain adaptation method for in-

telligent bearing fault diagnosis based on cycle-consistent adversarial learning. Chen et al.(Z. Chen, Yu, Ding, Shao, & Mechefske, 2022) introduced an end-to-end fault diagnosis network that considers cross-domain discrepancies through instance-based methods. Liu et al.(Liu, Wang, Chow, & Li, 2022) proposed a deep adversarial subdomain adaptation network to enhance the generalization capability of fault diagnosis.

Although existing domain adaptation methods have achieved promising results in fault diagnosis, they often fail to leverage the spatial structures in the data, which limits their ability to address domain discrepancies. To address this issue and further improve the robustness of the model, this paper proposes an adversarial domain adaptation fault diagnosis model based on self-attention graph convolutional network (ADA-SAG), which consists of a fault sample graph construction module, a self-attention graph convolutional network module and an enhanced domain adaptation module. Finally, the model adopts a dual-classifier structure and an adversarial learning strategy, promoting competition between the feature generator and the classifiers to achieve more precise feature alignment(Saito, Watanabe, Ushiku, & Harada, 2018; Zhang, Dong, Qaid, & Wang, 2024).To further reduce the feature distribution differences between the source and target domains, the model integrates an MMD regularization term, strengthening distribution consistency across different domains. The followings are the key contributions of this paper:

(1) The KNN algorithm is employed to convert raw signals into fault sample graphs, capturing subtle similarities between samples and efficiently extracting and utilizing the spatial structure features of the data.

(2) A self-attention enhanced graph convolutional network is introduced as the feature extractor, enabling more accurate capture of critical features. (3)The effectiveness and superiority of the ADA-SAG model are demonstrated through comparative studies with state-of-the-art methods and ablation experiments on the CWRU and PHM2009 datasets.

The remainder of this paper is structured as follows: Section II provides the problem statement and theoretical background. Section III describes the proposed method. Experiments and comparative studies are given in Section IV. Section V contains the conclusion of this study.

2. PRELIMINARY

2.1. Problem statement

The domain adaptive fault diagnosis problem aims to develop a classification model capable of effective transfer between different operating conditions (source and target domains), despite significant distributional and feature differences. The source domain X_s is defined as $\{x_i^s, y_i^s\}_{i=1}^{n_s}$,where y_i^s is the label of source sample x_i^s,n_s is the number of source samples.The source domain includes labeled normal states and various fault types, representing known and representative working conditions and fault modes. The source domain dataset contains n categories,denoted as $C = \{c_1, c_2, \ldots, c_n\}$,where each category corresponds to a specific fault type or normal state. The target domain X_t consists of unlabeled data,denoted as $\left\{x_j^t\right\}_{j=1}^{n_t}$.The data distribution, operating conditions, and noise levels in the target domain differ significantly from the source domain, and obtaining sufficient fault labels in the target domain is typically challenging due to practical constraints. Notably, despite these differences, the fault types and normal state categories are identical in both domains.

2.2. Graph Attention Networks

Graph Attention Networks (GAT) extend the capabilities of traditional GCN by introducing an attention mechanism, allowing for the dynamic weighting of the importance of neighboring nodes during feature aggregation(Veličković et al., 2017). This approach is particularly advantageous in fault diagnosis, as different time periods or signal features may have varying relevance when identifying specific fault conditions.

In GAT, the attention mechanism assigns a weight coefficient to each neighboring node, thereby adjusting its contribution during the aggregation process. The importance of neighboring nodes is first determined by calculating the unnormalized attention score $\vec{e}_{ij}$ between pairs of nodes, as shown in the following equation.

$$e_{ij} = \text{LeakyReLU}\left(\alpha^T \left[Wh_i \| Wh_j\right]\right) \tag{1}$$

where W is the weight matrix for feature transformation; h_i and h_j are the feature vectors of nodes i and j; $||$ denotes the concatenation operation of vectors; and α is the learnable attention weight vector used to calculate the correlation between nodes.

Next, the softmax function is applied to normalize these scores so that the sum of attention weights α_{ij} from node i to all its neighboring nodes j equals 1, as illustrated in the following Eq. (2).

$$\alpha_{ij} = \text{softmax}\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in N(i)} \exp\left(e_{ik}\right)} \tag{2}$$

where α_{ij} represents the normalized attention weight; and $N(i)$ represents the set of neighboring nodes of node i.

By performing a weighted sum of the features of neighboring nodes, the new feature representation h_i' for node i is obtained, as shown in the following Eq. (3):

$$h_i' = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W h_j\right) \tag{3}$$

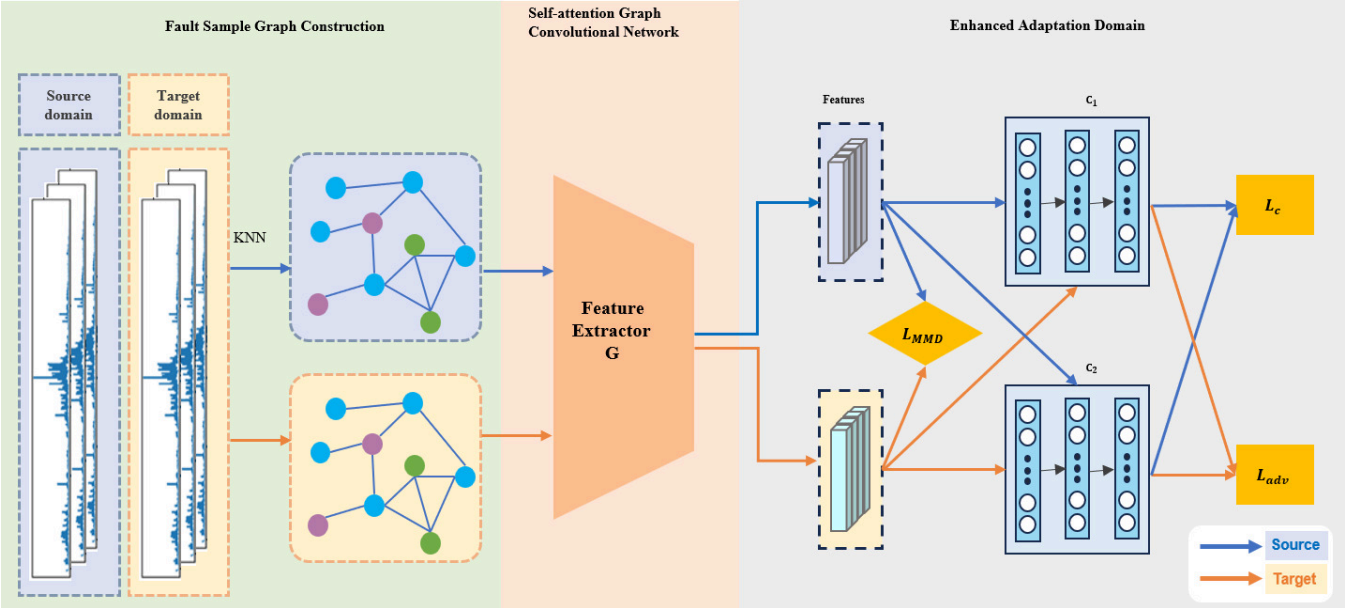


Figure 1. The model consists of three primary modules: fault sample graph construction, self-attention graph convolutional network, and enhanced adaptation domain. It incorporates three loss functions: L_c for source domain classification, L_{MMD} for feature distribution alignment, and L_{adv} for adversarial domain adaptation.

3. PROPOSED METHOD

As shown in Fig. 1, the proposed ADA-SAG mainly includes:fault sample graph construction module, self-attention graph convolutional network module and enhanced domain adaptation module.The detailed description of each module is given as follows.

3.1. Fault sample graph construction module

In the field of intelligent fault diagnosis, due to the lack of explicit dependency information, constructing association graphs can effectively reveal the similarity between fault modes. The K-Nearest Neighbors (KNN) algorithm is applied to construct fault sample graphs for the source and target domains. The process of constructing fault sample graph is shown in Fig. 2. By converting data samples into graph representations, the relationships between nodes are explored, capturing the intrinsic spatial features of the graph structure. In the fault sample graph, each node represents a fault sample, and the edges are defined based on the similarity between samples, with the edge weights indicating the degree of similarity. Therefore, the specific process of constructing a fault sample graph based on KNN is as follows.

In the source domain, the similarity between fault sample x_i and other samples $x_j (j = 1, 2, \cdots n, j \neq i)$ is measured using a Gaussian kernel function, as shown in Eq. (4)

$$\text{Sim}\left(x_i, x_j\right) = e^{-\frac{\|FFT(x_i) - FFT(x_j)\|^2}{2\sigma^2}} \tag{4}$$

where σ represents the bandwidth parameter of the Gaussian kernel, $\|FFT\left(x_i\right) - FFT\left(x_j\right)\|$ denotes the Euclidean distance between the frequency domain features of x_i and x_j. The closer $\text{Sim}\left(x_i, x_j\right)$ is to 1, the more similar x_i is to x_j.

For fault sample x_i, other fault samples x_j are sorted in descending order of similarity $\text{Sim}(x_i, x_j)$. The top K most similar fault samples are selected as the neighborhood of x_i, constructing the adjacency matrix A. If samples x_i and x_j are neighbors, the element A_{ij} in the adjacency matrix is set to the normalized similarity measure $\text{Sim}(x_i, x_j)$; otherwise, A_{ij} is set to 0. The construction of the adjacency matrix is formulated in Eq. (5):

$$A_{ij} = \begin{cases} \frac{\text{Sim}(x_i, x_j)}{\sum_{k \in N(i)} \text{Sim}(x_i, x_k)} & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $N(i)$ represents the neighborhood of fault sample x_i, which is the set of the top K most similar fault samples. The adjacency matrix A represents the edges in the graph, where a_{ij} denotes the edge weight between x_i and x_j. Each fault sample is connected to its K-nearest neighbors, forming the source domain fault sample graph $G_s\left(V_s, E_s\right)$ and the target domain fault sample graph $G_t\left(V_t, E_t\right)$ through identical processes.

3.2. Self-attention graph convolutional network module

For accurate fault diagnosis, the self-attention graph convolutional network module integrates GAT and GCN. The module employs a GAT layer to capture local feature importance by assigning dynamic weights to neighboring nodes through

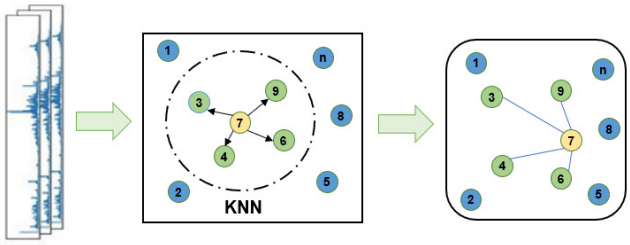


Figure 2. Schematic diagram of the fault sample graph construction method based on KNN.The circle represents the fault sample, and the straight line represents the similarity weight between the two fault samples. Assuming k=4, the green fault sample is a neighbor of the yellow fault sample, while the blue fault sample is not.

a self-attention mechanism. This ensures that key neighboring nodes contribute more significantly to the node's feature representation, enabling precise local feature extraction. Subsequently, the GCN layer combines these locally refined features across the entire graph to capture global structural information. This combination ensures local detail sensitivity and global integration, enabling the model to adapt to variations in node influence. The computational process of this module is as follows.

By sequentially applying Eq. (1), (2), and (3), the feature matrix H' is computed using the adjacency matrix A and the initial feature matrix H. The new feature representation h'_i for each node is then derived. All updated node features are then aggregated into the new feature matrix H', as shown in Eq. (6).

$$H' = \begin{bmatrix} h'_1 \\ h'_2 \\ \vdots \\ h'_n \end{bmatrix} \tag{6}$$

where H' represents the output feature matrix.

Next, the GCN integrates global information, updating the node features as shown in Eq. (7) to enhance the representation of structural dependencies.

$$H'' = \mathrm{ReLU}(\hat{A}H'W) \tag{7}$$

where H'' represents the output feature matrix, $relu(.)$ represents the non-linear activation, and $\hat{A}$ is the normalized adjacency matrix.

3.3. Enhanced domain adaptation module

The enhanced domain adaptation module applies an adversarial learning strategy to align feature spaces by optimizing both minimization and maximization strategies, reducing distributional discrepancies between source and target domains. The self-attention enhanced graph convolutional module serves as the feature extractor G, while C_1 and C_2 classifiers with distinct initializations, designed to capture diverse feature representations.To effectively achieve this alignment, we address the problem in three distinct steps.

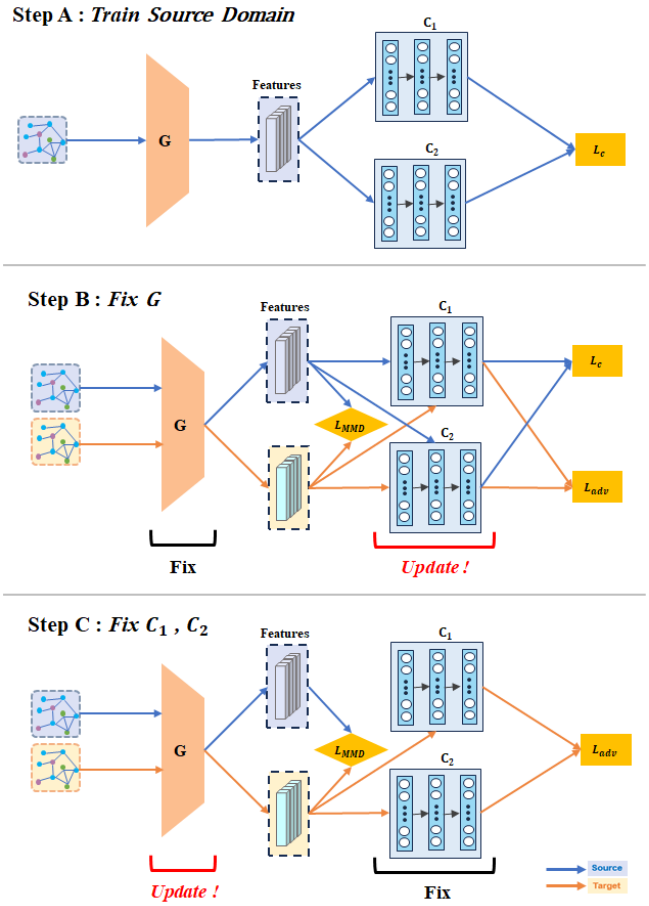


Figure 3. The classifiers in Step B learn to maximize the discrepancy on the target domain samples, while the generator in Step C learns to minimize this discrepancy, thereby aligning the feature distributions between the source and target domains.

Step A Initially, C_1 , C_2 and G are jointly trained on the source domain, ensuring correct classification of X_s through classification loss minimization. This step corresponds to Step A in Fig. 3. The goal is to ensure correct classification of source domain samples by minimizing the classification loss. The optimization objectives are as follows.

$$\min_{G,C_1,C_2} L_c(X_s, Y_s) \tag{8}$$

$$L_c(X_s, Y_s) = -E_{(x_s,y_s)-(X_s,Y_s)} \left[\sum_{k=1}^{K} 1_{[k=y_s]} \log p_1(y \mid x_s) + \sum_{k=1}^{K} 1_{[k=y_s]} \log p_2(y \mid x_s) \right] \tag{9}$$

where $p_1(y \mid x_s)$ and $p_2(y \mid x_s)$ denote the predicted probabilities by C_1 and C_2 for the input source sample x_s,respectively.

Step B In this phase, G is fixed, while C_1 and C_2 are adjusted to maximize prediction inconsistency on target domain samples, thus highlighting distributional differences between domains. This step corresponds to Step B in Fig. 3. At the same time, to maintain high classification accuracy on source domain samples during target domain alignment, the classifiers must also incorporate source domain classification loss in the optimization process. Furthermore, the MMD regularization term serves to align feature distributions between source and target domains, enhancing consistency across domain boundaries. The optimization objective is as follows.

$$\min_{C_1,C_2} L_c\left(X_s, Y_s\right) - L_{adv}\left(X_t\right) + \beta L_{MMD}\left(X_s, X_t\right) \quad (10)$$

$$L_{adv}\left(X_t\right) = E_{x_t - X_t}\left[\operatorname{dis}\left(p_1(y \mid x), p_2(y \mid x)\right)\right] \quad (11)$$

where X_t represents the distribution of target domain data; $L_{adv}(X_t)$ represents the classifier difference loss, which is measured by the maximum absolute difference between the probability distributions of C_1 and C_2 on the target domain samples; β is a hyperparameter controlling the strength of the MMD regularization.

Step C In the final step, C_1 and C_2 remain fixed, while G is optimized to minimize both the classifier discrepancy loss and the MMD loss thereby generating features tailored to the target domain. This step corresponds to Step C in Fig. 3. The optimization objective is as follows.

$$\min_{G} L_{adv}\left(X_t\right) + \beta L_{MMD}\left(X_s, X_t\right) \quad (12)$$

By iteratively optimizing these three steps, the model's fundamental performance on the source domain is enhanced, while class alignment and domain alignment are achieved in the target domain. The ultimate goal of training G is to develop robust generalized features, thereby enhancing target domain performance. The entire process simultaneously addresses both class and domain alignment, utilizing a strategy that integrates adversarial learning between the feature extractor and classifiers with MMD regularization to achieve effective domain adaptation. This approach ensures the discriminative ability and adaptability of the model in the target domain.

4. EXPERIMENTS

The proposed method was evaluated on the Case Western Reserve University Bearing Dataset and the PHM2009 Data Challenge Gearbox Dataset to validate its effectiveness and generalization capabilities. The model was developed in Python 3.9 using PyTorch 1.12.0 and tested in a computing environment with an Intel Core i7-7700HQ CPU (2.80 GHz), an NVIDIA GeForce GTX 2080Ti GPU (12 GB), and 32 GB of RAM. The experimental model employs cross-entropy loss as the loss function and updates its learnable parameters using the Adam optimizer.

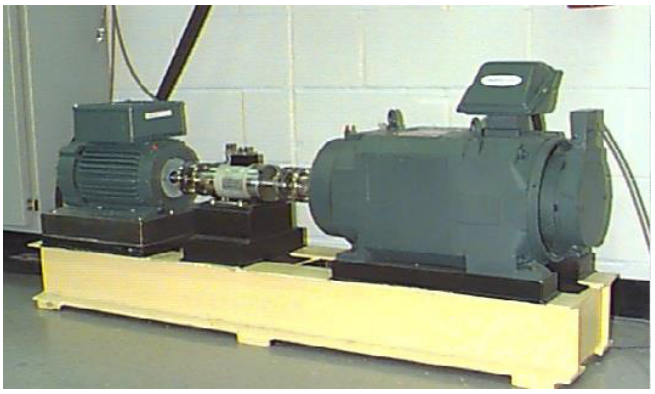


Figure 4. CWRU bearing test bench

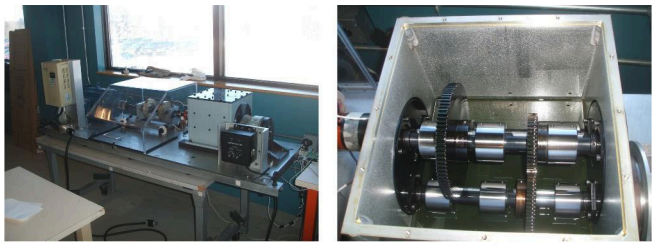
Table 1. CWRU DATASET INFORMATION

DataSet	Class Label	0	1	2	3	4	5	6	7	8	9
CWRU	Fault Type	H	BF	BF	BF	IF	IF	IF	OF	OF	OF
	Fault Size(mil)	0	7	14	21	7	14	21	7	14	21
	Load	0hp,1hp,2hp,3hp									

4.1. Data description

(1) Case Western Reserve University Dataset(Smith & Randall, 2015): The first dataset was obtained from the Bearing Data Center of Case Western Reserve University (CWRU). The layout of the test bench is shown in Fig .4, consisting of a 2-horsepower motor (left), a torque sensor (center), and a testing machine (right). Vibration signals are collected from the drive end of the motor at four motor loads and speeds: 0hp/1797rmp, 1hp/1772rmp, 2hp/1750rmp, and 3hp/1730rmp. There are four bearing health states in this dataset for each operating condition, including normal state (H), inner ring failure (IF), outer ring failure (OF), and rolling element failure (RF). Each fault state has 3 damage sizes (7, 14, and 21 inches (mils)). Each sample consists of 1024 sampling points. The detailed information of CWRU is shown in Table1. In order to construct a domain adaptation problem, two conditions were randomly selected from the above four conditions, one as the source domain and the other as the target domain, forming twelve domain adaptation problems.

(2) PHM2009 dataset: This dataset is an experimental dataset used for gearbox diagnosis research and can be obtained on the PHM (Data Analysis Competition PHM Society) official website. The variable speed device is shown in Fig .5, including three shafts, four gears, and six bearings. The vibration signals were collected at five axis speeds of 30, 35, 40, 45, and 50Hz, forming five different operating domains. The sampling frequency is 66.67kHz and the collection time is 4 seconds. There are six different health states under each operating condition, including normal state and five fault states.



(a) (b)

Figure 5. PHM2009 dataset setup fig(a) gearbox, fig(b) gearbox interior

Table 2. PHM2009 Dataset Information

DataSet	Label	Gear		Bearing			Shaft	
		24T	Others	Input Shaft: Output Side	Idler Shaft: Output side	Others	Input	Output
PHM2009	0	Good	Good	Good	Good	Good	Good	Good
	1	Chipped	Good	Good	Good	Good	Good	Good
	2	Broken	Good	Combination	Inner	Good	Bent Shaft	Good
	3	Good	Good	Combination	Ball	Good	Imbalance	Good
	4	Broken	Good	Good	Inner	Good	Good	Good
	5	Good	Good	Good	Good	Good	Bent Shaft	Good

Divide the vibration signal of each axis speed into segments with a length of 6144 sampling points. The detailed information of PHM2009 is shown in Table2. The experimental setup of this dataset involves randomly selecting two operating conditions from five different scenarios, one of which is considered as the source domain and the other as the target domain, and constructing twenty domain adaptation problems.

4.2. Analysis and discussion

To verify the effectiveness and superiority of the proposed ADA-ASG model, several state-of-the-art methods were compared to measure performance differences between the model and other domain adaptation strategies. These methods cover 1D-CNN baseline models trained using a single source domain data (directly applied to the target domain) and other domain adaptive methods. These methods adjust the distribution differences between the source domain and the target domain through different strategies. The detailed information of the comparison methods is as follows:

1D-CNN: As a baseline model, it is trained directly on the source domain data and applied directly to the target domain without any domain adaptation steps, with the aim of verifying the contribution of adversarial steps to improving model performance.

MK-MMD(Gretton et al., 2012) and CORAL: These methods narrow the differences between the source and target domains by adjusting and matching statistical features between them.

DANN(Ganin & Lempitsky, 2015): By introducing a domain discriminator to distinguish different domains and adopting a domain obfuscation training strategy, the feature extractor learns domain-invariant features.

MCD(Saito et al., 2018): Using a dual classifier adversarial mechanism, the inconsistency in predicting unlabeled samples in the target domain is maximized, thus adjusting the feature distribution and reducing prediction differences between classifiers.

Table 3. The accuracy of domain adaptation tasks on the CWRU dataset using different methods

Task	1D-CNN	MK-MMD	CORAL	DANN	MCD	MWDAN	ADA-SAG
0HP→1HP	98.67	99.67	98.33	99.33	100	100	**100**
0HP→2HP	97	99.67	98.33	100	100	100	**100**
0HP→3HP	90	94.67	93.67	92.67	93.33	95.67	**100**
1HP→0HP	94.33	99.00	98.33	99.67	100	100	**100**
1HP→2HP	93	99.00	100	100	100	100	**100**
1HP→3HP	95.77	99.33	99.67	98.33	99.33	99.67	**100**
2HP→0HP	90.67	98.33	99.00	98.33	99.33	99.67	**100**
2HP→1HP	95.67	99.67	99.67	100	100	100	**100**
2HP→3HP	89	99.00	100	100	100	100	**100**
3HP→0HP	91.33	94.33	92.33	93.33	92.00	94.33	**100**
3HP→1HP	94	99.00	92.00	94.00	100	99.67	**100**
3HP→2HP	96	99.67	99.67	98.67	100	100	**100**
Average	93.79	98.45	97.58	97.86	98.67	99.08	**100**

MWDAN(Song et al., 2021): An innovative multi-weight domain adversarial network that distinguishes and adapts source and target domains with partially shared label spaces by implementing weight mechanisms at the class and instance levels.

Table 3 shows the performance comparison of the proposed ADA-SAG model with other domain adaptation techniques, using the CWRU dataset. The experiment includes 12 load change migration tasks, representing different workload conditions (0HP, 1HP, 2HP, and 3HP), to evaluate the performance of the model under different operating conditions.

The ADA-SAG model outperformed all the compared methods, achieving 100% diagnostic precision in migration tasks due to its robust graph-based feature extraction and effective domain adaptation strategies. These strategies enable the model to capture intricate spatial relationships in fault data and align feature distributions between source and target domains more effectively. In contrast, traditional CNN models have average performance in reducing distribution differences between the source and target domains. When combined with methods based on distribution distance constraints such as MK-MMD and CORAL, its performance is improved, further verifying the importance of reducing inter-domain differences in improving model performance. In addition, although the DANN, MCD, and WMDAN methods based on adversarial learning have improved classification efficiency, their accuracy significantly decreases when there is a significant difference between the two domains (such as from 3HP to 0HP), which has been verified in experiments. These comparative experimental results not only validate the effectiveness of the proposed ADA-SAG model, but also highlight its ability to maintain high accuracy under various working conditions.

In order to further evaluate its generalization performance, the experimental scope was extended to the PHM2009 dataset. Unlike the relatively single fault type in the CWRU dataset,

Table 4. The accuracy of domain adaptation tasks on the PHM2009 dataset using different methods

Task	1D-CNN	MK-MMD	CORAL	DANN	MCD	MWDAN	ADA-SAG
30hz→35hz	52.01	52.71	61.11	69.44	92.43	98.82	**100**
30hz→40hz	49.31	54.86	56.25	63.19	77.43	97.22	**98.61**
30hz→45hz	41.67	46.39	50	53.96	70.83	85.76	**87.43**
30hz→50hz	27.78	45.9	52.15	64.58	66.67	85.71	**86.04**
35hz→30hz	53.96	51.32	45.9	50.83	92.5	96.53	**98.68**
35hz→40hz	52.08	69.1	66.74	72.92	88.89	92.57	**98.61**
35hz→45hz	48.61	57.64	55.9	59.72	83.33	86.04	**89.58**
35hz→50hz	25.42	41.67	45.97	50.35	73.61	80.56	**82.64**
40hz→30hz	41.67	50.35	46.04	52.08	77.22	83.33	**87.22**
40hz→35hz	46.25	65.97	62.5	67.92	93.64	98.54	**99.3**
40hz→45hz	50.14	69.31	66.74	68.4	86.8	92.36	**94.44**
40hz→50hz	34.72	69.38	67.92	65.97	76.39	86.81	**87.43**
45hz→30hz	26.94	47.78	48.61	63.33.	73.61	83.33	**86.04**
45hz→35hz	29.38	59.03	55.56	63.89	77.22	99.3	**100**
45hz→40hz	48.61	66.67	62.15	65.28	78.13	90.27	**92.71**
45hz→50hz	55.56	61.67	67.92	66.74	82.99	96.53	**97.22**
50hz→30hz	23.06	50.07	61.8	66.67	68.06	72.92	**73.33**
50hz→35hz	24.38	47.22	61.88	68.06	67.36	76.39	**79.17**
50hz→40hz	40.97	53.96	67.5	70.83	77.15	84.72	**85.07**
50hz→45hz	53.96	59.72	69.38	73.61	85.42	86.11	**89.58**
Average	41.32	56.04	58.6	63.89	79.48	88.69	**90.66**

the multi class mixed fault of PHM2009 presents a more challenging diagnostic scenario.

Table4 shows the comparison of experimental results between the proposed method and other methods on the PHM2009 dataset. The ADA-SAG model achieved 100% performance in tasks 30Hz → 35Hz and 45Hz → 35Hz. In tasks 30hz → 40hz, 30hz → 45hz, and 30hz → 50hz, the performance of the ADA-SAG model was 98.61%, 87.43%, and 86.04%, respectively. These performances are still the highest compared to other models, especially compared to the baseline 1D-CNN model, demonstrating significant improvements. For other migration tasks, such as 35Hz → 30Hz, 35Hz → 40Hz, etc., the performance of the ADA-SAG model has remained above 89.58%, with most tasks exceeding 90%. This demonstrates its stability and efficiency in handling more challenging tasks. In relatively difficult tasks such as 50Hz → 30Hz, 50Hz → 35Hz, 50Hz → 40Hz, and 50Hz → 45Hz, the performance of ADA-SAG has decreased, but still remains between 73.33% and 89.58%. This means that even in these more challenging tasks, the ADA-SAG model still has good generalization ability. In summary, the ADA-SAG model performs better than other models on the PHM2009 dataset, demonstrating its superior cross domain fault diagnosis ability.

4.3. Feature visualization

The performance of the proposed method was visually validated through feature visualization, as shown in Fig. 5. In Fig. 6(a), the visualization displays tightly clustered intra-class points and clear inter-class separations, demonstrating effective feature alignment for minor domain discrepancies (e.g., 0HP → 1HP). Fig. 6(b) shows similar trends, with increased load differences (two units), where the model maintains distinct class boundaries and consistent intra-class clustering. In Fig. 6(c), despite the significant load differences (three units), the ADA-SAG model continues to exhibit robust feature alignment, with high intra-class cohesion and inter-class separations. These results highlight the model's stability and adaptability to varying domain discrepancies. The distribution between categories indicates that the model can maintain accurate feature representation in the face of increasing load changes, which reflects the model's ability to adapt to different working conditions under dynamic changes.

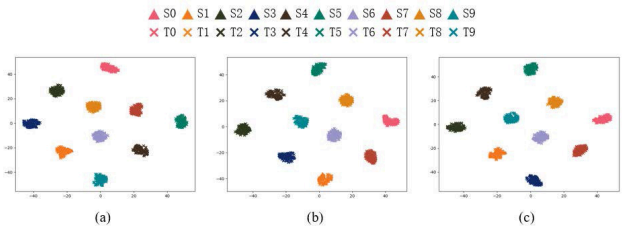


Figure 6. Visualization Results of Migration Tasks on the CWRU Dataset. (a), (b), and (c) represent the feature visualization results of load differences. (a) Task 0HP → 1HP. (b) Task 0HP → 2HP. (c) Task 0HP → 3HP.

For the PHM2009 dataset, Fig. 7 shows the distribution of migration task characteristics with speed differences of 5hz, 10hz, 15hz, and 20hz. From Fig. 7 (a) to (d), it can be observed that under a speed difference of 5Hz, the distribution of feature points exhibits high clustering and clear inter class boundaries, reflecting the effectiveness of the model under slight speed changes. When the speed difference increases to 10Hz, 15Hz, or even 20Hz, although the overlap of feature distributions slightly increases, the model still maintains a high degree of classification ability. Especially at a speed change of 20Hz, although there is some overlap between classes, clusters with high separation can still be observed, indicating that even under significant speed changes, the model has good feature discrimination and adaptability.

4.4. Ablation study

In order to comprehensively verify the effectiveness of the proposed model (ADA-SAG), this chapter conducted systematic ablation experiments on the CWRU dataset and the PHM2009 dataset, respectively. Evaluate the impact of MMD loss in the fault sample graph construction module, self attention mechanism enhanced GCN, and domain adaptation enhancement module on model performance by constructing three different model variants. The following is a detailed description and analysis of each variant:

ADA-SAG-NoGraph: This variant explores the performance of removing the fault sample graph construction module from the model, where the model only utilizes frequency domain data for domain adaptation through CNN.

ADA-SAG-NoAttention: In this variant, the self attention mechanism enhanced GCN is replaced with the traditional GCN architecture.

ADA-SAG-NoMMD: The variant after removing MMD loss aims to evaluate the role of this loss in domain adaptation.

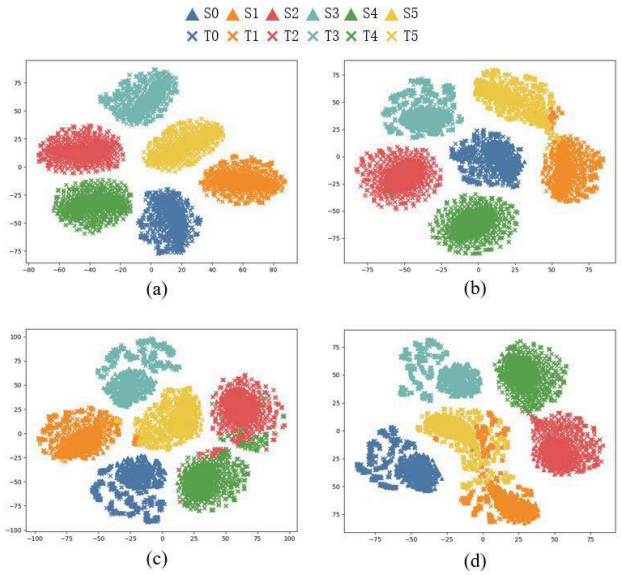


Figure 7. Visualization Results of Migration Tasks on the PHM2009 Dataset. (a), (b), (c) and (d) show the feature visualization results of speed differences. (a) Task 30hz → 35hz. (b) Task 30hz → 40hz. (c) Task 30hz → 45hz. (d) Task 30hz → 50hz.

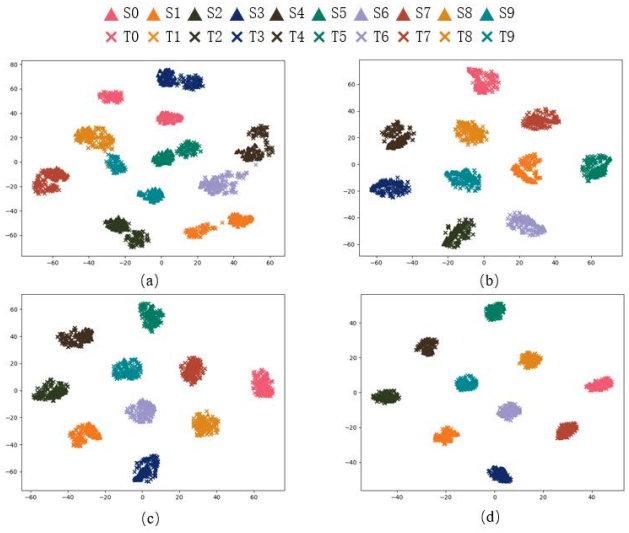


Figure 8. Comparison of Visualization Results of CWRU Dataset in 0HP → 3HP Ablation Experiments. (a) Visualization result of ADA-SAG NoGraph. (b) Visualization result of ADA-SAG-NoAttention. (c) Visualization result of ADA-SAG-NoMMD. (d) Visualization result of complete model.

The visualization results in Fig. 8(a) and Fig. 9(a) show that without the support of the ADA-SAG NoGraph, the boundaries between categories are blurred, and there is a lack of tight clustering within each category, reflecting the limitations of CNN in independently capturing complex relationships between samples. Fig. 8(b) and Fig. 9(b) show the visualization results of replacing the self attention mechanism

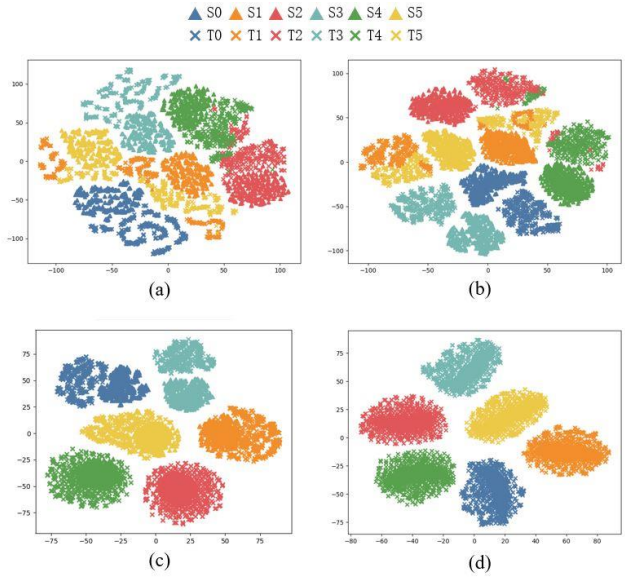


Figure 9. Comparison of Visualization Results of PHM2009 Dataset in 30hz → 35hz Ablation Experiments. (a) Visualization result of ADA-SAG NoGraph. (b) Visualization result of ADA-SAG-NoAttention. (c) Visualization result of ADA-SAG-NoMMD. (d) Visualization result of complete model.

enhanced GCN with the traditional GCN architecture, where category clustering is tighter, but compared to the complete model, there is still a certain separation between the source domain and the target domain, indicating that the importance of the self attention mechanism lies in its enhanced model's capture of key information. Fig. 8(c) and 9(c) show the visualization results after removing the MMD loss. Although the boundaries between categories are relatively clear, the intra class cohesion is not as good as the complete model, proving the significant role of MMD loss in reducing the differences between source and target domains and improving the generalization ability of the target domain. Finally, Fig. 8(d) and Fig. 9(d) show the feature distribution after domain adaptation in the complete model. It was observed that data points of different categories exhibit significant clustering, and the boundaries between categories are very clear. In contrast, other variant models perform worse than the complete model in these aspects.

The comparison of experimental results is shown in Fig. 10 and Fig. 11, which respectively demonstrate the accuracy performance of different model variants and the complete model on all transfer tasks on the CWRU and PHM2009 datasets. It can be observed that the complete model achieved the highest accuracy in the vast majority of migration tasks, verifying its strong adaptability and excellent fault diagnosis performance. In contrast, the accuracy of each variant model has decreased in certain tasks, which further confirms the important role of the fault sample graph construction module, the self attention mechanism enhanced GCN, and the MMD loss in the domain

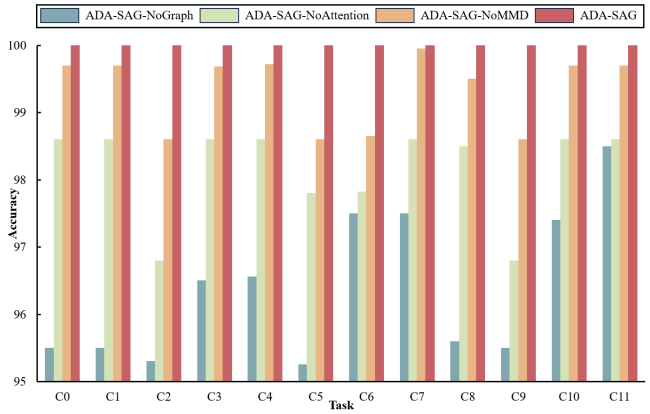


Figure 10. Comparison of accuracy of ablation experiments for various migration tasks in the CWRU dataset

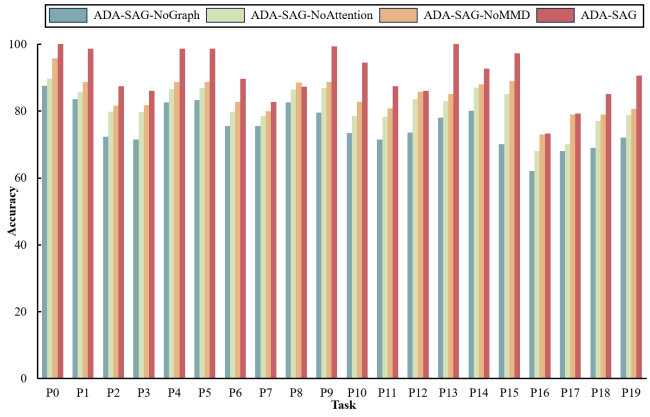


Figure 11. Comparison of accuracy of ablation experiments for various migration tasks in the PHM2009 dataset

adaptation enhancement module in improving the model domain adaptation ability and fault diagnosis accuracy.

5. CONCLUSION

This study introduces an adaptive fault diagnosis method using self-attention graph convolutional networks to address data distribution differences between source and target domains. The method employs KNN and Gaussian kernel similarity to construct a fault sample graph, enhancing feature representation by capturing sample relationships. By integrating a self-attention mechanism enhanced GCN as a feature extractor, the model captures critical features effectively, improving diagnostic accuracy and robustness. The use of MMD loss and adversarial training further aligns feature distributions between source and target domains, mitigating domain transfer challenges. Validated on two benchmark datasets, CWRU and PHM2009, the ADA-SAG model demonstrates superior diagnostic accuracy and outperforms existing methods. In complex scenarios with drastic changes in load and speed, the model shows remarkable stability and efficiency, achieving over 90% accuracy on average, with some tasks reaching 100%. Ablation experiments validate the necessity and effectiveness of each module in the ADA-SAG model, confirming the rationality of the overall design. However, current research assumes data balance, which is often unrealistic in practical applications where normal state data far exceeds fault data. Future research could focus on approaches like data augmentation, cost-sensitive learning, or class imbalance handling to optimize the proposed model for imbalanced datasets, enhancing its practicality in real-world applications.

ACKNOWLEDGMENT

This research was funded by the Jiangsu Provincial Science and Technology Project on Industrial Foresight and Independent Core Key Technologies, Grant No. BE2023047.

REFERENCES

Chen, C., Chen, Z., Jiang, B., & Jin, X. (2019). Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 33, pp. 3296–3303).

Chen, L., Li, Q., Shen, C., Zhu, J., Wang, D., & Xia, M. (2021). Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions. *IEEE Transactions on Industrial Informatics*, *18*, 1790-1800. doi: 10.1109/TII.2021.3078712

Chen, Y., Tao, L., Liu, X., Ma, J., Lu, C., & Liu, H. (2023). Open-set fault recognition and inference for rolling bearing based on open fault semantic subspace. *IEEE Transactions on Instrumentation and Measurement*.

Chen, Z., Yu, W., Ding, X., Shao, Y., & Mechefske, C. K. (2022). Pair-wise orthogonal classifier based domain adaptation network for fault diagnosis in rotating machinery. *IEEE Sensors Journal*, *22*(12), 12086–12097.

Chen, Z., Yu, W., Wang, L., Ding, X., Huang, W., & Shao, Y. (2023). A dual-view style mixing network for unsupervised cross-domain fault diagnosis with imbalanced data. *Knowledge-Based Systems*, *278*, 110918.

Deng, Z., Zhou, K., Li, D., He, J., Song, Y.-Z., & Xiang, T. (2022). Dynamic instance domain adaptation. *IEEE Transactions on Image Processing*, *31*, 4585–4597.

Du, L., Tan, J., Yang, H., Feng, J., Xue, X., Zheng, Q., ... Zhang, X. (2019). Ssf-dan: Separated semantic feature based domain adaptation network for semantic segmentation. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 982–991).

Ganin, Y., & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In *International conference on machine learning* (pp. 1180–1189).

Gong, C., Yu, J., & Xia, R. (2020). Unified feature and instance based domain adaptation for aspect-based sentiment analysis. In *Proceedings of the 2020 conference on empirical methods in natural language processing (emnlp)* (pp. 7035–7045).

Gretton, A., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., Fukumizu, K., & Sriperumbudur, B. K. (2012). Optimal kernel choice for large-scale two-sample tests. *Advances in neural information processing systems*, *25*.

Guo, L., Pfohl, S., Fries, J., Johnson, A. E. W., Posada, J., Aftandilian, C., ... Sung, L. (2021). Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. *Scientific Reports*, *12*. doi: 10.1038/s41598-022-06484-1

Li, H., Liu, H., Busch, H. V., Grimm, R., Huisman, H., Tong, A., ... Lou, B. (2024). Deep learning-based unsupervised domain adaptation via a unified model for prostate lesion detection using multisite bi-parametric mri datasets. *Radiology. Artificial intelligence*, e230521. doi: 10.1148/ryai.230521

Li, W., Yuan, Z., Sun, W., & Liu, Y. (2020). Domain adaptation for intelligent fault diagnosis under different working conditions. *MATEC Web of Conferences*. doi: 10.1051/matecconf/202031903001

Liang, P., Deng, C., Yuan, X., & Zhang, L. (2023). A deep capsule neural network with data augmentation generative adversarial networks for single and simultaneous fault diagnosis of wind turbine gearbox. *ISA transactions*, *135*, 462–475.

Liu, Y., Wang, Y., Chow, T. W., & Li, B. (2022). Deep adversarial subdomain adaptation network for intelligent fault diagnosis. *IEEE Transactions on Industrial Informatics*, *18*(9), 6038–6046.

Long, M., Cao, Y., Wang, J., & Jordan, M. (2015). Learning transferable features with deep adaptation networks. In *International conference on machine learning* (pp. 97–105).

Lu, W., Fan, H., Zeng, K., Li, Z., & Chen, J. (2022). Self-supervised domain adaptation for cross-domain fault diagnosis. *International Journal of Intelligent Systems*, *37*(12), 10903–10923.

Ma, P., Zhang, H., Fan, W., & Wang, C. (2020). A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. *ISA transactions*, *99*, 465–478.

Saito, K., Watanabe, K., Ushiku, Y., & Harada, T. (2018). Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3723–3732).

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical systems and signal processing*, *64*, 100–131.

Song, Y., Liu, Z., Wang, J., Tang, R., Duan, G., & Tan, J. (2021). Multiscale adversarial and weighted gradient domain adaptive network for data scarcity surface defect detection. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–10.

Tang, X., Xu, Y., Sun, X., Liu, Y., Jia, Y., Gu, F., & Ball, A. D. (2023). Intelligent fault diagnosis of helical gearboxes with compressive sensing based non-contact measurements. *ISA transactions*, *133*, 559–574.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wang, R., Huang, W., Wang, J., Shen, C., & Zhu, Z. (2022). Multisource domain feature adaptation network for bearing fault diagnosis under time-varying working conditions. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-10. doi: 10.1109/tim.2022.3168903

Yan, W., Huang, L., Xia, L., Gu, S., Yan, F., Wang, Y., & Tao, Q. (2020). Mri manufacturer shift and adaptation: Increasing the generalizability of deep learning segmentation for mr images acquired with different scanners. *Radiology. Artificial intelligence*, *2 4*, e190195. doi: 10.1148/ryai.2020190195

Zhang, B., Dong, H., Qaid, H. A., & Wang, Y. (2024). Deep domain adaptation with correlation alignment and supervised contrastive learning for intelligent fault diagnosis in bearings and gears of rotating machinery. In *Actuators* (Vol. 13, p. 93).

Zhao, C., Liu, G., Shen, W., & Gao, L. (2021a). A multi-representation-based domain adaptation network for fault diagnosis. *Measurement*, *182*, 109650. doi: 10.1016/J.MEASUREMENT.2021.109650

Zhao, C., Liu, G., Shen, W., & Gao, L. (2021b). A multi-representation-based domain adaptation network for fault diagnosis. *Measurement*, *182*, 109650.

Zhu, W., Shi, B., Feng, Z., & Tang, J. (2023). An unsupervised domain adaptation method for intelligent bearing fault diagnosis based on signal reconstruction by cycle-consistent adversarial learning. *IEEE Sensors Journal*.