



A new multi-source information domain adaption network based on domain attributes and features transfer for cross-domain fault diagnosis[☆]

Yue Yu^{a,*}, Hamid Reza Karimi^{a,*}, Peiming Shi^b, Rongrong Peng^c, Shuai Zhao^d

^a Department of Mechanical Engineering, Politecnico di Milano, via La Masa 1 Milan, 20156, Italy

^b School of Electrical Engineering, Yanshan University, Qinhuangdao, 066004, China

^c Nonlinear Dynamics and Application Research Center, Nanchang Institute of Science & Technology, Nanchang, 330108, China

^d School of Information and Artificial Intelligence, Nanchang Institute of Science & Technology, Nanchang, 330108, China

ARTICLE INFO

Communicated by Y. Lei

Keywords:

Fault diagnosis

Multi-source information

Deep transfer learning

Multi-source domain adaptation

ABSTRACT

Compared to the single-source domain adaptation fault diagnosis methods, the multi-source domain adaptation methods can not only take advantage of the rich and diverse diagnostic information of multiple source domains but also draw on the feature alignment of single-source setting to reduce the domain discrepancy. However, forcing the alignment of feature distributions is challenging and may lead to negative transfer. Meanwhile, labeled data are often scarce and difficult to collect in actual production, which can be mitigated by multi-source information, but the diagnostic performance of the model is degraded by large domain differences. To tackle the above issues, a domain attribute and feature transfer network is proposed to model multi-source information domains in a unified deep network and achieve cross-domain fault diagnosis. In the transferable attributes learning section, we adopt an attention mechanism and a domain attribute loss function to extract transferable latent attributes from multi-source information. In the transferable features learning section, we apply the local maximum mean discrepancy metric to adjust the category distribution of single-source information and target domains. Then, intra-class compactness learning and pseudo-labeling learning strategies are utilized to further obtain richer feature representations. Finally, we propose the knowledge fusion module to fuse the results of multi-source information classifiers to yield a more reliable diagnosis result. Extensive experiments on three different multi-source information datasets show the superiority of our method compared to the state-of-the-art (SOTA) methods by comparing indicators from various aspects.

1. Introduction

With the advent of the Industry 4.0 era, the mechanical equipment in industrial production (e.g., aircraft engines, rolling mills, and wind-driven generators) is gradually becoming intelligent and integrated. Mechanical components (i.e., bearings and gears) play a crucial role in the mechanical equipment and they are more prone to failure due to vibration and shock loads. Therefore, it is of utmost importance to timely and accurately monitor the operating states of mechanical components to enhance industrial production

[☆] This research is supported by the scholarship from the China Scholarship Council (CSC), China under Grant CSC N202308130067. This research is supported by the Initial Scientific Research Foundation for Talented Scholars of Nanchang Institute of Science and Technology, under Grant NGRZX-23-09.

* Corresponding authors.

E-mail addresses: yue.yu@polimi.it (Y. Yu), hamidreza.karimi@polimi.it (H.R. Karimi).

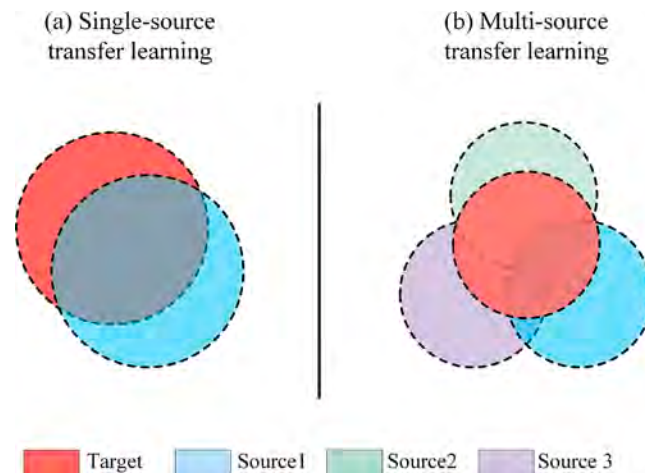


Fig. 1. Illustration of single-source and multi-source TL.

efficiency, reduce maintenance costs, and foster intelligent and efficient development [1–3]. Owing to their potent nonlinear feature extraction capabilities, deep learning (DL) methods have emerged as prominent techniques in Prognostic and Health Management (PHM), garnering considerable attention in recent research [4–6].

Traditional DL methods have achieved tremendous success in fault diagnosis, relying mainly on two assumptions: (1) the availability of sufficient labeled data under various working conditions for model training, and (2) the alignment of training and testing data under the same operating conditions. Shi et al. took advantage of multi-source sensing data fusion to diagnose the health states of the rolling mill [7]. Yan et al. applied an interpretable DL model to achieve intelligent fault diagnosis [8]. In [9], an improved DBN is proposed to detect the failures in the mechanical equipment. An et al. implemented a DL-based model to deal with the strong noise of bearings [10]. However, as the research goes deeper, it is evident that the accomplishments of traditional DL fault diagnosis methods mentioned above are challenging to apply in practical industrial environments, mainly due to three reasons [11,12]: (1) Collecting sufficient fault data necessitates prolonged exposure of mechanical equipment to fault conditions, potentially causing damage. (2) The scarcity of labeled data results in time-consuming and expensive collection processes. (3) Inevitably, there exists a distribution gap between training and testing samples due to variations in loads and environments. Recent research on transfer learning (TL) and domain adaptation (DA) in fault diagnosis suggests that these methods can effectively address the aforementioned limitations, as depicted in Fig. 1(a). DA-based fault diagnosis methods usually assume that there are sufficient training samples in the source domain and unlabeled samples in the target domain. We use sufficient labeled samples to train a model in one operation condition (a.k.a., source domain) to achieve fault diagnosis of unlabeled samples under other operating conditions (a.k.a., target domains). The number of labeled samples in the source domain is significantly smaller than the number of unlabeled samples in the other target domains, so scholars usually conclude that the number of labeled samples is scarce compared to the overall sample sizes from all the working conditions. The core idea of DA is to transfer knowledge and models from a source domain to a target domain by learning their similarities and differences. This aims to enhance fault diagnosis performance in the target domain, particularly when labeled data is limited or insufficient. Lei et al. presented a TL model based on a transport-embedded similarity statistic to classify the health states across machines [13]. Shao et al. adopted a domain-share CNN to extract transferable features to diagnose fault types at different speeds [14]. Wang et al. combined Bayesian network and active querying into the TL framework to predict classification results [15]. Zhao et al. proposed various fault descriptions to conduct zero-shot fault diagnosis in real industry environments [16]. Wu et al. developed a knowledge mapping function based on maximum classifier discrepancy to address small unlabeled samples [17]. However, such expositions are unsatisfactory as these DA-based fault diagnosis methods solely concentrate on single-source data, leading to reduced data efficiency and insufficient adaptability. This limitation restricts the performance and generalization ability of fault diagnosis models in the target domain, especially when encountering different fault modes and operating conditions.

The multi-source domain adaption (MSDA), whose structure leverage data from multiple related yet distinct domains, is helpful to alleviate the unavoidable disadvantages arising from the above studies. The methods built based on the MSDA have been successfully applied to image classification [18,19], text classification [20,21] and object detection [22,23]. MSDA-based models can employ adaptive training and transfer domain knowledge from multiple source domains to employ adaptive training and transfer domain knowledge to improve the performance of them on the target domain. In recent years, MSDA provides new insights in the research field of fault diagnosis into further improve diagnostic generalization ability under different operating conditions, as shown in Fig. 1(b). Deng et al. used a residual deep subdomain adaptation network to diagnose bearing failures [24]. Lei et al. established a knowledge fusion module to synthesize rich information from multi-source domains for unseen fault diagnosis [25]. Jiang et al. introduced various new kernel maximum mean discrepancy metrics into the reinforcement ensemble TL framework to extract multi-source features [26]. Karimi et al. proposed a complementary transferability metric to identify open-set faults based

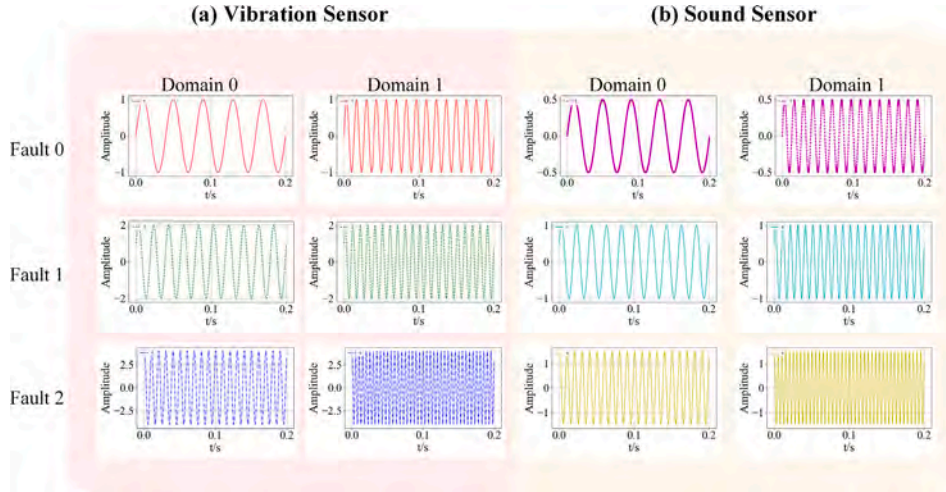


Fig. 2. Example signals from MSDA. Each column contains the same fault labels but from different domains and the red and yellow boxes symbolize different sources of information.

on multi-source domains [27]. Yang et al. designed multi-source domain class-central samples to construct anchor adapters and then achieved fault identification through the integration of results of these anchor adapters [28]. Meng et al. used two-stage multi-source domain adaption strategy to alleviate the complex differences between domains [29]. The success of existing MSDA methods in fault diagnosis is due to the feature alignment strategy, which aims to align the feature distributions of multi-source domains and the target domain in the identical feature space to minimize the distribution discrepancy between them. However, relying solely on the implementation of this strategy in MSDA appears to be less successful. This is obvious because the same fault type can exhibit entirely different signal characteristics under various working conditions. Notably, the same fault signal acquired from different sensors (e.g., vibration, sound, and current sensors) can manifest differently, as illustrated in Fig. 2. The existence of diversity and potential conflicts between the source domains may create a challenge. Forceful alignment of feature distributions between multi-source domains and the target domain can lead to negative transfer. In such cases, enforcing feature alignment across multi-source domains may result in the diagnostic performance of DL models collapsing, rendering it inferior to using a single source domain for fault identification [30,31]. In conclusion, research on cross-domain fault diagnosis based on MSDA is quite limited compared to single-source DA methods, and aligning the distributions between all source domains and the target domain poses a significant challenge.

Meanwhile, the current research in the field of MSDA-based fault diagnosis primarily concentrates on single-source information, specifically vibration signals. Single-source information in fault diagnosis has several key drawbacks, including incomplete information, susceptibility to noise and interference, uncertainty in information, and potential limitations. Conversely, multi-source information can offer a more comprehensive understanding of faults and significantly improve diagnostic capabilities by effectively leveraging data and features from various sources of information. Additionally, it can make comprehensive judgments and decisions regarding fault diagnosis results, thereby enhancing the accuracy and efficiency of fault diagnosis. Chen et al. combined multi-source information extracted by a deep coupling auto-encoder for fault detection [32]. Wu et al. adopted frequency spectra in multi-source information to obtain the final decision of the key components [33]. Xie et al. combined multisensor data and CNNs to explore failures under various operating conditions [34]. Surprisingly, very little attention has been paid to the role of multi-source information on MSDA in fault diagnosis. Hence, an essential challenge lies in effectively integrating multi-source information and MSDA to enhance the cross-domain diagnostic performance of the model.

In response to the aforementioned challenges, we present a novel method called multi-source information domain adaptation based on domain attributes and features transfer (MSIDA-Net) for cross-domain fault diagnosis. Specifically, the term ‘multi-source information domain adaption’ (MSIDA) refers to training a model to extract domain-invariant and discriminative knowledge (a.k.a., features and attributes) from various sensors in the multi-source domains to achieve fault diagnosis of a single sensor in the target domain. In this study, we introduce a novel DL-based fault diagnosis framework that effectively leverages transferable features and attributes from different yet relevant multi-source information domains to achieve end-to-end diagnosis within a unified deep network. To address the first issue, we proposed the attention mechanism and the domain attribute loss function to transfer attributes to alleviate negative transfer, and the LMMD metric is used to further relieve this problem. Then, the pseudo-labeling (PLL) strategy and the intra-class compactness learning are adopted to extract richer features from multi-source information to tackle the second problem. Finally, the fault diagnosis task is decomposed into multiple single-source information domain subtasks

and then the knowledge fusion module is to integrate the weighted scores of multiple subtasks to determine the health states of the key components in the mechanical equipment.

The primary novelties and contributions of this paper can be summarized as follows.

- The attention mechanism, the domain attribute loss function, and LMMD metric strategies are proposed to address the issue of the negative TL, aiming to enhance the discriminative attributes and features and improve the diagnostic performance.
- The intra-class compactness learning, and PLL strategies are incorporated to align the distribution discrepancy for improving both model performance and generalization ability in DA-based tasks.
- The proposed MSIDA-Net framework can effectively leverage transferable attributes and features from different yet relevant multi-source information domains to achieve cross-domain fault diagnosis. Its superiority and effectiveness can be validated by three case studies.

This paper has been organized in the following way. Section 2 will examine related works and preliminary knowledge. Section 3 is concerned with the detailed principles of the proposed MSIDA-Net used for this study. Section 4 introduces the experimental platforms and validates the effectiveness of the proposed method on different diagnostic tasks compared to the SOTA methods. Section 5 presents the discussions of the model, focusing on the four key parts. Section 6 lays out some future works of the proposed method. In Section 7, the conclusion and limitation of this paper are carried.

2. Preliminary

2.1. Problem formulation

The mathematical notations used in this study are listed in Table 1. In this study, we introduce a novel term called MSIDA which refers to MSDA based on the different sources of information under the same working conditions. The multi-source information domains can be represented as $S = \{S_1, S_2, \dots, S_M\}$, where there are M single-source information domains. Especially, the labeled samples n_{S_k} from the k th different source information domain S_k can be denoted as $\mathbb{N}^{S_k} = \left\{ \left(x_i^{S_k}, y_i^{S_k} \right) \mid i = 1, 2, \dots, n_{S_k}, x_i^{S_k} \sim P_{S_k}, y_i^{S_k} \in \mathcal{Y}^{S_k}, k \in |S| \right\}$, where \mathcal{Y}^{S_k} represents the label space from the k th different source information domain. Subsequently, the unlabeled samples from the target domain are represented as $\mathbb{N}^T = \{x_i^T \mid i = 1, 2, \dots, n_T, x_i^T \sim P_T\}$, where n_T denotes the number of unlabeled samples. This paper investigates MSIDA-based fault diagnosis across multiple different yet relevant multi-source information in mechanical equipment under different working conditions. The following assumptions are considered:

- (1) The multi-source information domains and the target domain have a common label space.
- (2) There exists a distribution discrepancy (a.k.a., probability distribution) of $P_{S_k} \neq P_T$ and $P_{S_i} \neq P_{S_j}$ ($i, j \in |S|$ and $i \neq j$) between the multi-source information domains and the target domain due to differences in working conditions.
- (3) In the multi-source information domains, there are enough labeled samples to build and train a satisfactory model. Meanwhile, there are insufficient or few labeled samples to establish effective DL-based classification models in the target domain.

These assumptions are normal as they usually follow the typical setup of DA-based fault diagnosis literature [35,36]. The main objective is to train the proposed MSIDA-Net $f_{\text{MSIDA-Net}}(\cdot) := \left\{ f_{S_k}(\cdot) : \mathbb{N}^{S_k} \rightarrow \mathcal{Y}^{S_k} \mid k \in |S| \right\}$ by leveraging both $\{\mathbb{N}^{S_k}\}_{k=1}^M$ and \mathbb{N}^T to minimize the target empirical loss function $R_T = E_{x^T \sim P_T} [f_{\text{MSIDA-Net}}(x_i^T) \neq y_i^T]$ and allowing the trained method to handle unseen testing samples from the target domain.

2.2. Local maximum mean discrepancy

The maximum mean discrepancy (MMD) metric is utilized in the TL process as a nonparametric distance matrix to adjust the discrepancies between the source and target domains, as illustrated in Fig. 3(a). Assume that the source and target domain datasets can be represented as and, so the MMD metric can be written as

$$D_H(X, Y) = \sup_{\Phi \in \mathcal{H}} \left\{ E_{x \sim p} [\Phi(\mathbb{N}^{S_k})] - E_{y \sim q} [\Phi(\mathbb{N}^T)] \right\} \quad (1)$$

where \mathcal{H} is the reproducing kernel Hilbert space (RKHS). $\Phi(\cdot)$ represents the feature mapping function. $\sup\{\cdot\}$ denotes the maximum value of the input. Particularly when the RKHS is computed using feature kernels, the MMD metric can be updated as

$$D^2(X, Y) = \frac{1}{n_{S_k}^2} \sum_{i=1}^{n_{S_k}} \sum_{j=1}^{n_{S_k}} k(x_i^{S_k}, x_j^{S_k}) - \frac{2}{n_{S_k} n_T} \sum_{i=1}^{n_{S_k}} \sum_{j=1}^{n_T} k(x_i^{S_k}, x_j^T) + \frac{1}{n_T^2} \sum_{i=1}^{n_T} \sum_{j=1}^{n_T} k(x_i^T, x_j^T) \quad (2)$$

where $k(\cdot, \cdot)$ denotes the Gaussian feature kernels.

Nonetheless, a significant limitation of the MMD metric is its exclusive focus on the global marginal distribution. Consequently, this negative limitation not only tends to hamper efficiency by leading to a chaotic distribution of related categories between the two domains but also contributes to causing undesirable negative transfer. To alleviate this deficiency, we employ the following LMMD metric to align the distribution of categories in domains from a local perspective. It can enhance discriminative information and enhance diagnostic performance in fault diagnosis tasks. The rationale behind this lies in the ability of LMMD metric to capture

Table 1
Mathematical notations.

Notations	Descriptions
S, \mathcal{T}	Multi-source information domains and target domain
M	Number of information domains
$\mathbb{N}^{S_k}, \mathbb{N}^{\mathcal{T}}$	Labeled and unlabeled samples
$n_{S_k}, n_{\mathcal{T}}$	Number of labeled and unlabeled samples
\mathcal{Y}^{S_k}	Label space
$P_{S_k}, P_{\mathcal{T}}$	Multi-source information and target domain distribution probabilities
$R_{\mathcal{T}}$	Empirical loss function
\mathcal{H}	Reproducing kernel Hilbert space
$\Phi(\cdot), f(\cdot)$	Feature mapping function
$k(\cdot, \cdot)$	Gaussian feature kernels
C	Number of categories
$\omega_i^{sc}, \omega_j^{sc}$	Weights of the source and target domain categories
y_i^{sc}, y_i^{st}	Probability of the category
$F \in R^{C \times H \times W}$	Feature map
Cl, H, W	Number of channels, height, and width
$g(F) \in R^{C \times 1 \times 1}$	Channel attention vectors
r	Reduction ratio
$\left\{ f_i^{S_k} \right\}_{i=1}^{n_{S_k}}, \left\{ f_i^{\mathcal{T}} \right\}_{i=1}^{n_{\mathcal{T}}}$	Multi-source information and target domain feature maps
B	Batch size
$m_{S_k}, m_{\mathcal{T}}$	Attention weights of the multi-source information and target domains
$\phi(\cdot), F(\cdot)$	Shared feature extractor and domain feature extractor
H	Number of domain feature extractors
$\theta_d, \theta_s, \theta_p, \text{ and } \theta_c$	Learnable parameters
$\hat{\theta}_d, \hat{\theta}_s, \hat{\theta}_p, \text{ and } \hat{\theta}_c$	Updated parameters
$\partial(\cdot)$	Partial derivative algorithm
η	Learning rate
k, N	Number of algorithms and diagnostic tasks
Kl	Kernel length
C_{in}, C_{out}	Number of input and output channels
D_{in}, D_{out}	Input feature and output feature dimensions

subtle differences in local data distributions. By focusing on local variations rather than global statistics, the LMMD metric is effective in identifying subtle patterns associated with different fault modes [37–39], which can be shown in Fig. 3(b).

$$D^2(X, Y) = \frac{1}{C} \sum_{c=1}^C \left[\sum_{i=1}^{n_{S_k}} \sum_{j=1}^{n_{S_k}} w_i^{sc} w_j^{sc} k \left(f(x_i^{S_k}), f(x_j^{S_k}) \right) - 2 \sum_{i=1}^{n_{S_k}} \sum_{j=1}^{n_{\mathcal{T}}} w_i^{sc} w_j^{tc} k \left(f(x_i^{S_k}), f(x_j^{\mathcal{T}}) \right) + \sum_{i=1}^{n_{\mathcal{T}}} \sum_{j=1}^{n_{\mathcal{T}}} w_i^{tc} w_j^{tc} k \left(f(x_i^{\mathcal{T}}), f(x_j^{\mathcal{T}}) \right) \right] \quad (3)$$

with

$$\omega_i^{sc} = \frac{y_i^{sc}}{\sum_{(x_i^{S_k}, y_i^{S_k}) \in \mathbb{N}^{S_k}} y_i^{sc}} \quad \omega_j^{sc} = \frac{y_j^{sc}}{\sum_{(x_j^{S_k}, y_j^{S_k}) \in \mathbb{N}^{S_k}} y_j^{sc}} \quad (4)$$

$$\omega_i^{st} = \frac{y_i^{st}}{\sum_{(x_i^{\mathcal{T}}) \in \mathbb{N}^{\mathcal{T}}} y_i^{st}} \quad \omega_j^{st} = \frac{y_j^{st}}{\sum_{(x_j^{\mathcal{T}}) \in \mathbb{N}^{\mathcal{T}}} y_j^{st}} \quad (5)$$

where C denotes the category number assigned to each health state. ω_i^{sc} and ω_j^{tc} represent the weights of the source and target domain categories, respectively. y_i^{sc} and y_i^{st} are the probability of the category. $f(\cdot)$ represents the feature mapping function.

3. Methodology

Based on the proposed MSIDA-Net, the cross-domain diagnostic procedure of the key components in the mechanical equipment can be developed in Fig. 4. Then, the concrete process of the proposed method can be described as follows.

- (1) Step I: Collect multi-source information data (i.e., vibration signals, current signals, and acoustic signals) from the key components in experimental mechanical rigs.
- (2) Step II: Standardize the gathered multi-source information data.
- (3) Step III: Randomly sample the standardized multi-source information data, and then segment them into the training datasets and the testing datasets.
- (4) Step IV: Implement the multi-source information training samples in three case studies to train the proposed MSIDA-Net.
- (5) Step V: Use the multi-source testing samples to validate the diagnostic performance of the proposed method.

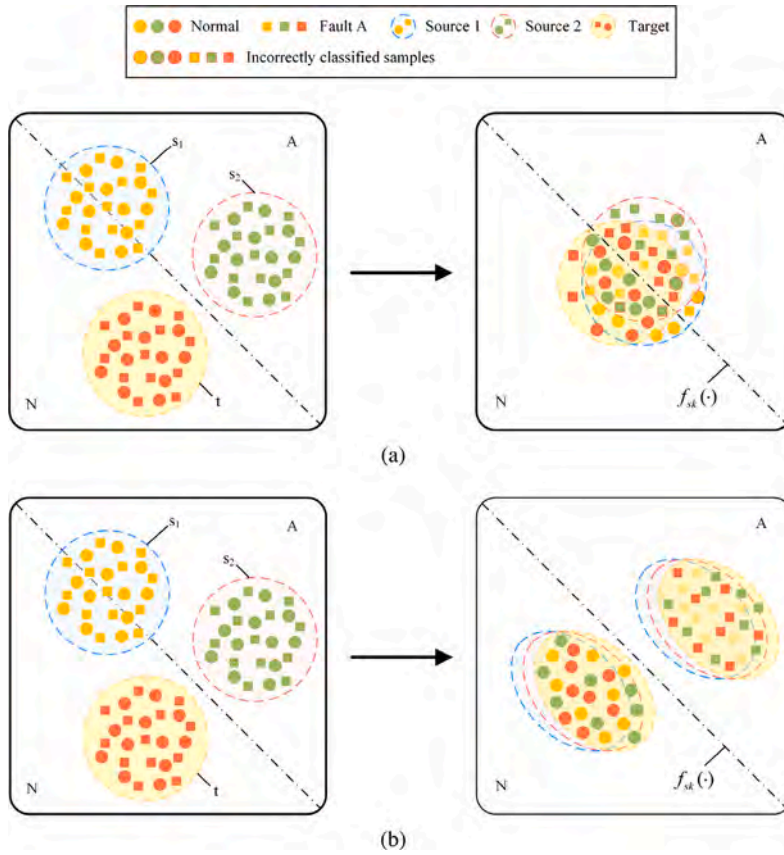


Fig. 3. The difference of the MMD and LMMD metrics: (a) MMD metric; (b) LMMD metric.

The proposed MSIDA-Net comprises three modules: transferable attributes learning, transferable features learning, and knowledge fusion, as depicted in Fig. 5. In the transferable attributes learning section, we proposed the attention mechanism and the domain attribute loss function to exploit multi-source information domain attributes for alleviating the issue of negative TL. In the transferable features learning section, the feature learning strategy is applied to obtain the diagnostic labels in the target domain. Then, the LMMD metric is adopted to achieve feature alignment. After that, the PLL strategy is used to improve generalization of the proposed method under the limited data. Next, the intra-class compactness learning strategy is introduced to improve the clustering effectiveness. Finally, the knowledge fusion module is proposed to combine diagnostic results from multi-source information classifiers to enhance the overall generalization. The following sections describe each innovation in detail.

3.1. Transferable attributes learning

In this section, we propose the attributes learning strategy to overcome the limitations of relying solely on feature transfer in MSDA. This strategy can extract transferable attributes from the multi-source information domains to the target domain, thereby enhancing the diagnostic performance of the proposed MSIDA-Net.

A detailed process of the transferable attributes learning strategy implementation can be shown in Fig. 6. The motivation behind this strategy is straightforward: despite different domains (a.k.a., different working conditions) may exhibit highly different fault forms, they contain the same health states categories, which can be explained using a common set of attributes. As shown in Fig. 2, although Fault 0 may exhibit different peaks across Domain 0 in various information domains, it consistently maintains the same frequency. Furthermore, it can be observed that within the same information domain, Fault 0 can be described by different frequencies across Domain 0 and Domain 1, but it consistently exhibits the same peak. In essence, fault sources can cause vibrations or shocks that result in the signal repeating at certain frequencies or peaks owing to the inherent periodicity of mechanical equipment operation. Consequently, specific attributes, such as peaks and frequencies, are shared across fault types in multi-source information domains under various working conditions. Based on this observation, we proposed that the MSIDA-Net method should prioritize learning the most transferable attributes from multi-source information domains to the target domain. This can be accomplished by using the attention mechanism to extract attributes in different multi-source information domains. After that, we enforce the domain attribute loss function for transferable attribute learning.

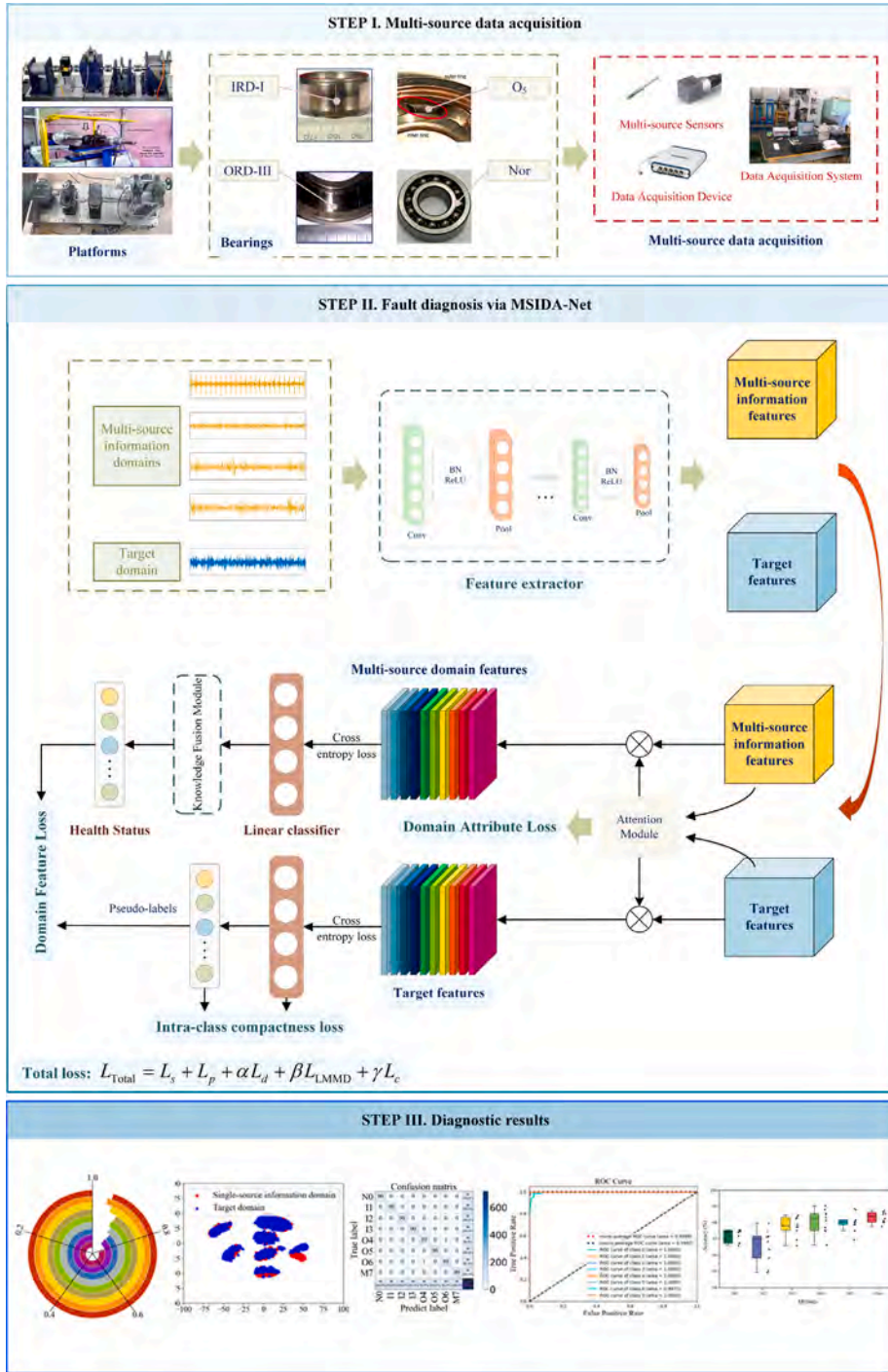


Fig. 4. Procedure of the proposed MSIDA-Net.

3.1.1. Attention mechanism

The attention mechanism offers a key advantage in extracting significant and transferable attributes from multi-source information domains. This is achieved by leveraging the channel relationships among features. The feature map $F \in \mathbb{R}^{C \times H \times W}$ is fed into the attention mechanism module, with C , H , and W representing the number of channels, height, and width, respectively. Then, the feature map F undergoes the average pooling and max pooling operations to generate two different representations, namely the average pooling feature and max pooling feature. The two features are forwarded to a shared multi-layer perceptron network

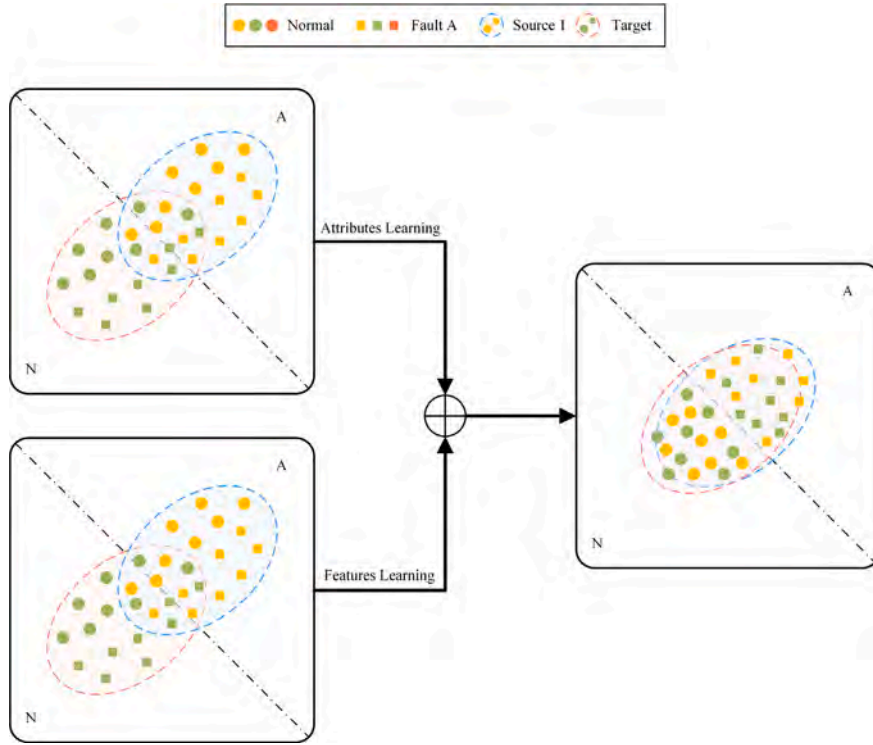


Fig. 5. Procedure of the proposed MSIDA-Net.

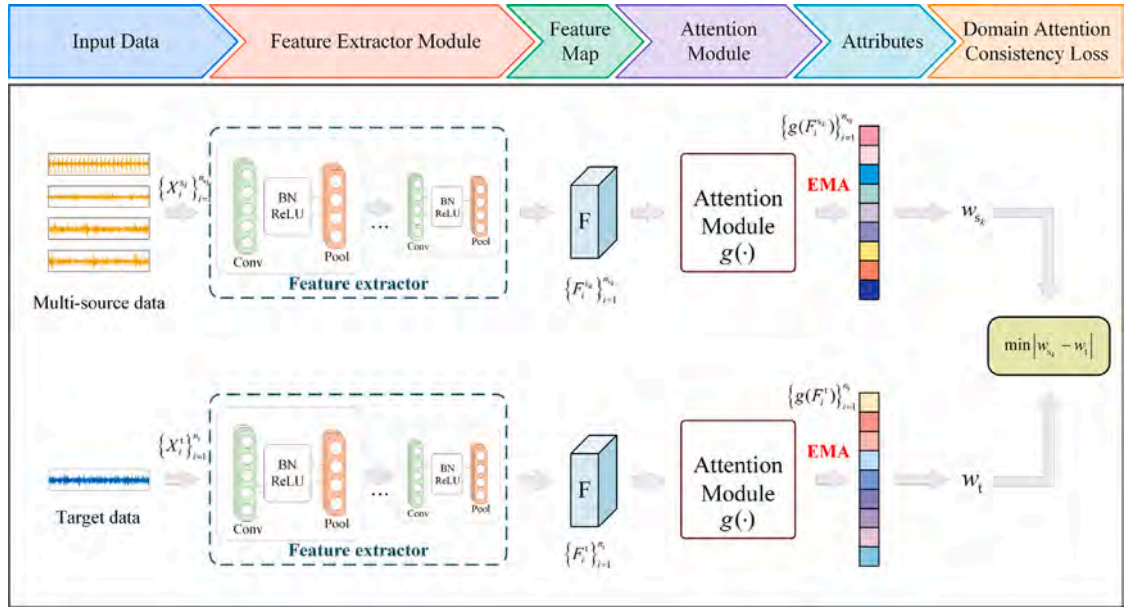


Fig. 6. Process of transferable attributes learning.

(MLP) that comprises two fully connected layers with matching input and output dimensions to generate channel attention vectors $g(F) \in \mathbb{R}^{C \times 1 \times 1}$, dubbed Vector_1 and Vector_2. It is important to note that the activation size of the hidden layer is determined by $\mathbb{R}^{C/r \times 1 \times 1}$, where r represents the reduction ratio. Subsequently, the two vectors are element-wise summed, and the resulting output feature vectors are constrained within the range of 0 to 1 using the Sigmoid function to obtain the domain attributes. Fig. 7 illustrates the detailed implementation process of the attention mechanism. By employing attention modeling on data from

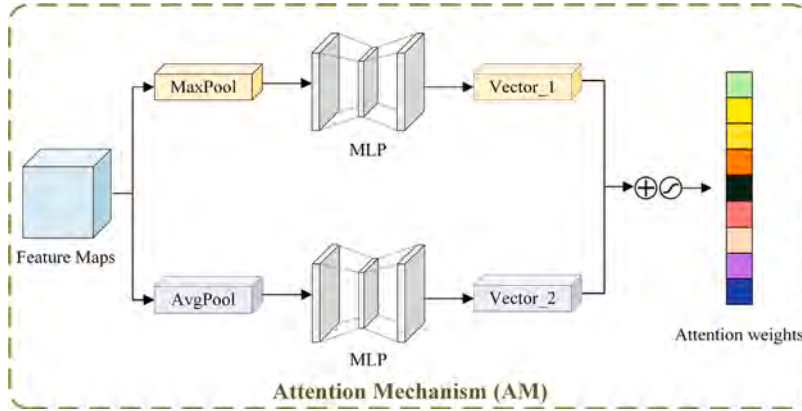


Fig. 7. Architecture of the attention mechanism.

multi-source information domains, the proposed method prioritizes meaningful and potential attributes while suppressing irrelevant background information.

3.1.2. Domain attribute loss

The multi-source information domain feature maps $\{f_i^{S_k}\}_{i=1}^{n_{S_k}}$ and the target domain feature maps $\{f_i^T\}_{i=1}^{n_T}$ can be extracted from k th source information domain data $\{x_i^{S_k}\}_{i=1}^{n_{S_k}}$ and target domain data $\{x_i^T\}_{i=1}^{n_T}$ through the feature extractor module, respectively. Subsequently, the feature maps $\{f_i^{S_k}\}_{i=1}^{n_{S_k}}$ and $\{f_i^T\}_{i=1}^{n_T}$ are input into the attention module $g(\cdot)$ to obtain the corresponding attention weight vectors $\{g(f_i^{S_k})\}_{i=1}^{n_{S_k}}$ and $\{g(f_i^T)\}_{i=1}^{n_T}$. In this study, we employ the exponential moving average (EMA) technique on the weights to compute a more precise measurement of domain attribute statistics [40]. Hence, the attention weights calculated by EMA based on the multi-source information and target domains are expressed as

$$m_{S_k} = \alpha m_{S_k} + (1 - \alpha) \frac{1}{B} \sum_{i=1}^B g(f_i^{S_k}) \quad (6)$$

$$m_T = \alpha m_T + (1 - \alpha) \frac{1}{B} \sum_{i=1}^B g(f_i^T) \quad (7)$$

where B denotes the batch size. means the fixed hyperparameter of 0.999. Despite the computational burden imposed by EMA, we choose to adopt it for a more accurate estimation of domain attention weights.

Finally, we adopt the Manhattan distance ℓ_1 to estimate the discrepancy of attention weights between single source information domain m_{S_k} and target domain m_T . Therefore, the domain attribute loss can be written as

$$\mathcal{L}_d = \frac{1}{M} \sum_{m=1}^M |m_{S_k} - m_T| \quad (8)$$

where M denotes the number of multi-source information domains.

3.2. Transferable features learning

In this section, to further achieve effective MSIDA, feature transfer is also imperative. By employing different learning strategies, we can not only mitigate negative TL, but also can further improve model generalization capability. The feature learning strategy is proposed to obtain a promising diagnostic result. Structure alignment learning strategy is used to align the features structure of multi-source information domains and target domain. The PLL strategy is utilized to improve model performance. The intra-class compactness learning strategy is introduced to better capture and distinguish patterns within individual categories for improved accuracy and robustness of the proposed method.

3.2.1. Feature learning strategy

To ensure that the framework can learn valuable features using multi-source information, we use the cross-entropy loss function to optimize the feature extraction parameters, which can be written as

$$\mathcal{L}_s = -\frac{1}{MB} \sum_{m=1}^M \sum_{i=1}^B \log p_{i,y_i}^{S_k} \quad (9)$$

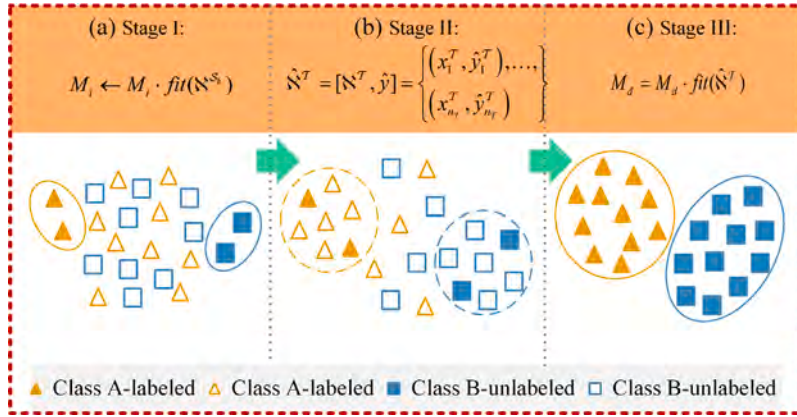


Fig. 8. Implementation process of PLL strategy.

where M denotes the number of multi-source information domains. B means the batch size. $p_{i,y_i}^{S_k}$ represents that assigns to the predicted probability of in the multi-source information domain S_k .

3.2.2. Structure alignment learning strategy

The LMMD metric aligns the various health state categories of the multi-source information and target domains in the unified feature space to lessen the correlation between local distribution differences. The DA loss function can be denoted by

$$\mathcal{L}_{\text{LMMD}} = \sum_{m=1}^M \hat{D}_H \left(F \left(\phi_m \left(x_i^{S_k} \right) \right), F \left(\phi_m \left(x_i^T \right) \right) \right) \quad (10)$$

where M is the number of information domains. $\phi(\cdot)$ and $F(\cdot)$ denote the shared feature extractor and domain feature extractor, respectively.

3.2.3. PLL

In this paper, the PLL strategy is introduced to alleviate the issue of data scarcity in the target domain and then to improve the cross-domain diagnostic accuracy and robustness of the proposed MSIDA-Net. The proposed PLL strategy consists of three stages. In Stage I, an initial model M_i is trained using labeled data from multi-source information domains $\{x_i^{S_k}\}_{i=1}^{n_{S_k}}$. Subsequently, Stage II involves utilizing a new deep model to predict unlabeled data in the target domain $\{x_i^T\}_{i=1}^{n_T}$ and obtain the pseudo-labels. In the final Stage III, the deep model is fine-tuned using both the pseudo-labeled data and the labeled samples. The specific implementation process is illustrated in Fig. 8, and its detailed description is provided below.

Let $\mathbf{N}^{S_k} = \{x_i\}_{i=1}^{n_{S_k}} = \{(x_1^{S_k}, y_1^{S_k}), \dots, (x_{n_{S_k}}^{S_k}, y_{n_{S_k}}^{S_k})\}$ and $\mathbf{N}^T = \{x_i^T\}_{i=1}^{n_T} = \{x_1^T, \dots, x_{n_T}^T\}$ denote the labeled samples of the single-source information domain and unlabeled samples of the target domain, where n_{S_k} and n_T ($n_{S_k} \ll n_T$) mean the number of labeled samples and the number of unlabeled samples, respectively.

First, the PLL model is trained and optimized by \mathbf{N}^{S_k} , which can be defined as

$$M_i \leftarrow M_i \cdot \text{fit}(\mathbf{N}^{S_k}) \quad (11)$$

where $\text{fit}(\cdot)$ represents the training and updating process.

Second, the pseudo-labels for the unlabeled data can be obtained by picking U into M_i to obtain the predicted class with the maximum probability, which is denoted as

$$\hat{y} = \{\hat{y}_1^T, \dots, \hat{y}_{n_T}^T\} = \{\text{argmax} M_i(x_1^T), \dots, \text{argmax} M_i(x_{n_T}^T)\} \quad (12)$$

where $\text{argmax} M_i(\cdot)$ denotes the optimization function.

Third, \mathbf{N}^T can be updated to a new labeled dataset $\hat{\mathbf{N}}^T = [\mathbf{N}^T, \hat{y}] = \{(x_1^T, \hat{y}_1^T), \dots, (x_{n_T}^T, \hat{y}_{n_T}^T)\}$ and the pre-training and fine-tuning process of the deep model M_d using $\hat{\mathbf{N}}^T$ and \mathbf{N}^T can be written as

$$M_d \leftarrow M_d \cdot \text{fit}(\hat{\mathbf{N}}^T) \quad (13)$$

$$M_d \leftarrow M_d \cdot \text{fit}(\mathbf{N}^T) \quad (14)$$

Finally, the low confidence values can be filtered out to make accurate predictions for categories by using a manually set threshold τ and the cross-entropy loss function is adopted to update the parameters of the M_d , which is denoted as follows:

$$\mathcal{L}_p = -\frac{1}{n_T} \sum_{i=1}^{n_T} \sum_{c=1}^C \Phi \left[P \left(y_c^i = 1 \mid x_{n_T}^T \right) \right] \log P \left(y_c^i = 1 \mid x_{n_T}^T \right) \quad (15)$$

where $P(y_m^i = 1 | x_i^u)$ represents the predicted probability of the pseudo-labeled class. $\Phi(\cdot)$ denotes the indicator function.

3.2.4. Intra-class compactness learning strategy

Inspired by the class prototype, we introduce the intra-class compactness loss to eliminate irrelevant features and encourage them to approach the category centers (a.k.a. prototype vectors). The higher intra-class compactness indicates increased similarity among samples within the same category, thereby promoting the learning of transferable features in the target domain. Assume that $f_{n_u}^u$ denotes the features in the last fully-connected layer of $x_{n_T}^T$ and W signifies the weight between the output features and categories, the intra-class compactness loss can be written as

$$\mathcal{L}_c = \frac{1}{B} \sum_{i=1}^B \sum_{c=1}^C \Phi \left[P(y_c^i = 1 | x_{n_T}^T) \right] \left\| f_{n_u}^u - W_{P(y_c^i = 1 | x_{n_T}^T)} \right\|_2^2 \quad (16)$$

where B denotes the batch size.

3.3. Knowledge fusion module

We denote the predicted probability of the single-source information domain as $Q_m(F_m(\phi(x^T)))$, which not only indicates whether the unlabeled samples $\{x_i^T\}_{i=1}^{n_T}$ are assigned to the health states in the target domain but also reflects the similarity between the multi-source information domains \mathbb{N}^{S_k} and the target domain \mathbb{N}^T . Finally, the knowledge fusion module is proposed to fuse the results from M pairs of the single-source information domain and the target domain to yield a more reliable diagnosis result, which can be written as

$$C_T(\mathbb{N}^T) = \sum_{m=1}^M \sum_{i=1}^{n_T} \omega_m \cdot Q_m(F_m(\phi(x_i^T))) \quad (17)$$

where M denotes the number of the multi-source information domains. $\phi(\cdot)$ and $F(\cdot)$ represent the shared feature extractor and domain feature extractor, respectively. ω_m can be calculated as follows:

$$\omega_m = \frac{\exp(-\hat{D}_H(F_m(\phi(x^{sm})), F_m(\phi(x^T))))}{\sum_{m=1}^M \exp(-\hat{D}_H(F_m(\phi(x^{sm})), F_m(\phi(x^T))))} \quad (18)$$

It is essential to note that each information domain contributes differently to fault diagnosis in the target domain. Consequently, the source and target pair with a high similarity will significantly influence the diagnostic decision, whereas the pair with low similarity may yield results with low confidence. A larger value of ω_m indicates a greater contribution of the single-source information domain to the target domain.

3.4. Model training

In this study, we proposed the MSIDA-Net model for extracting domain-invariant, discriminative, and transferable knowledge to reduce distribution shift between multi-source information domains and the target domain, embedding attributes learning and features learning into the unified deep network to achieve cross-domain fault diagnosis. Synthesizing the above loss functions, the total objective function of the MSIDA-Net can be given as

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_s + \mathcal{L}_p + \alpha \mathcal{L}_d + \beta \mathcal{L}_{\text{LMMD}} + \gamma \mathcal{L}_c \quad (19)$$

where α , β , and γ denote the trade-off parameters of the proposed MSIDA-Net.

Based on the overall objective function, let θ_d , θ_s , θ_p , and θ_c represent the learnable parameters for the transferable attributes learning, features learning, PLL, and intra-class compactness learning, respectively. Hence, the formula for the optimization problem can be described as:

$$\hat{\theta}_d = \arg \left\{ \begin{array}{l} \min_{\theta_d} \mathcal{L}_s(\theta_d, \hat{\theta}_s, \hat{\theta}_p, \hat{\theta}_c), \min_{\theta_d} \mathcal{L}_p(\theta_d, \hat{\theta}_s, \hat{\theta}_p, \hat{\theta}_c), \\ \min_{\theta_d} \mathcal{L}_d(\theta_d, \hat{\theta}_s, \hat{\theta}_p, \hat{\theta}_c), \min_{\theta_d} \mathcal{L}_{\text{LMMD}}(\theta_d, \hat{\theta}_s, \hat{\theta}_p, \hat{\theta}_c), \min_{\theta_d} \mathcal{L}_c(\theta_d, \hat{\theta}_s, \hat{\theta}_p, \hat{\theta}_c) \end{array} \right\} \quad (20)$$

$$\hat{\theta}_s = \arg \left\{ \begin{array}{l} \min_{\theta_s} \mathcal{L}_s(\theta_s, \hat{\theta}_d, \hat{\theta}_p, \hat{\theta}_c), \min_{\theta_s} \mathcal{L}_p(\theta_s, \hat{\theta}_d, \hat{\theta}_p, \hat{\theta}_c), \\ \min_{\theta_s} \mathcal{L}_d(\theta_s, \hat{\theta}_d, \hat{\theta}_p, \hat{\theta}_c), \min_{\theta_s} \mathcal{L}_{\text{LMMD}}(\theta_s, \hat{\theta}_d, \hat{\theta}_p, \hat{\theta}_c), \min_{\theta_s} \mathcal{L}_c(\theta_s, \hat{\theta}_d, \hat{\theta}_p, \hat{\theta}_c) \end{array} \right\} \quad (21)$$

$$\hat{\theta}_p = \arg \left\{ \min_{\theta_p} \mathcal{L}_p(\theta_p, \hat{\theta}_d, \hat{\theta}_s, \hat{\theta}_c) \right\} \quad (22)$$

$$\hat{\theta}_c = \arg \left\{ \min_{\theta_c} \mathcal{L}_c(\theta_c, \hat{\theta}_d, \hat{\theta}_s, \hat{\theta}_p) \right\} \quad (23)$$

where $\arg(\cdot)$ means the abbreviation for the independent variable. $\hat{\theta}_d$, $\hat{\theta}_s$, $\hat{\theta}_p$, and $\hat{\theta}_c$ mean the updated parameters of θ_d , θ_s , θ_p , and θ_c , respectively.

During the model training process, the parameters can be iterated and updated through the stochastic gradient descent (SGD) algorithm, which can be calculated by

$$\theta_d = \theta_d - \eta \frac{\partial \mathcal{L}_{\text{Total}}}{\partial \theta_d} \quad (24)$$

Table 2

Algorithm 1: Flowchart of the proposed method.

Multi-Source Information Domain Adaption Based on Domain Attributes and Features Transfer for Cross-domain Fault Diagnosis
--

Input: - Labeled multi-source information domain samples $\{x_i^{S_k}\}_{i=1}^{n_{S_k}}$; Unlabeled target domain samples;
 - Trade-off parameters $\tau, \alpha, \beta, \gamma$, and batch size B ;
 - Learning rate η .
 Output: - Trained features extractor F and ϕ ;
 - Trained attributes extractor g ;
 - Trained label classifier C ;
 - Predicted labeled samples \hat{Y}_T .
 While not convergent do
 1) for $[x_i^{S_k}, x_i^T]$ in $[N^{S_k}, N^T]$ randomly **sample**
 2) $[H_i^{S_k}, H_i^T] \leftarrow \text{MSIDA-Net}([x_i^{S_k}, x_i^T])$
 3) $\mathcal{L}_{\text{Total}}(H_i^{S_k}, H_i^T, \mathcal{Y}^{S_k}) = \mathcal{L}_s(H_i^{S_k}, H_i^T, \mathcal{Y}^{S_k}) + \mathcal{L}_p(H_i^{S_k}, H_i^T) + \alpha \mathcal{L}_d(H_i^{S_k}, H_i^T, \mathcal{Y}^{S_k}) + \beta \mathcal{L}_{\text{LMMD}}(H_i^{S_k}, H_i^T) + \gamma \mathcal{L}_c(H_i^{S_k}, H_i^T)$
 4) Update model parameters through the SGD algorithm
 $\theta_d = \theta_d - \eta \frac{\partial \mathcal{L}_{\text{Total}}}{\partial \theta_d}$
 $\theta_s = \theta_s - \eta \frac{\partial \mathcal{L}_{\text{Total}}}{\partial \theta_s}$
 $\theta_p = \theta_p - \eta \frac{\partial (\mathcal{L}_p)}{\partial \theta_p}$
 $\theta_c = \theta_c - \eta \frac{\partial (\mathcal{L}_c)}{\partial \theta_c}$
end while

$$\theta_s = \theta_s - \eta \frac{\partial \mathcal{L}_{\text{Total}}}{\partial \theta_s} \quad (25)$$

$$\theta_p = \theta_p - \eta \frac{\partial (\mathcal{L}_p)}{\partial \theta_p} \quad (26)$$

$$\theta_c = \theta_c - \eta \frac{\partial (\mathcal{L}_c)}{\partial \theta_c} \quad (27)$$

where $\partial(\cdot)$ denotes the partial derivative algorithm. η means the learning parameter of the model. The SGD optimizer can demonstrate stronger generalization capabilities and is more suitable for multiple diagnostic tasks from different datasets in this study [41]. Meanwhile, the SGD optimizer can better inhibit fluctuations in the learning rate during the later training stage to obtain an optimal solution [42,43].

The parameters of the proposed MSIDA-Net are updated by minimizing the total loss function, which aims to reduce domain discrepancies between the multi-source information and the target domains. In this way, the trained method can identify the unlabeled health states of the target domain by extracting discriminative and transferable attributes and features. The detailed algorithm for cross-domain fault diagnosis of key components in the mechanical equipment can be presented as Table 2.

4. Experimental verification

4.1. Case studies

In this section, we have formulated 16 MSDA diagnosis tasks aimed at evaluating the performance of MSIDA-Net across three multi-source information datasets. Meanwhile, we employ the following SOTA methods to demonstrate the effectiveness and superiority of the proposed MSIDA-Net in case studies to facilitate a comprehensive evaluation. To mitigate the effect of randomization in the three cases, we repeated each experiment ten times. It is worth noting that a single sensor type (i.e., one of the vibration sensor, acoustic sensor, and current sensor) in each target domain was chosen alternatively to reduce redundancy in the experimental part.

4.1.1. Case study 1: Fault diagnosis of roller bearing platform

In this study, we utilize public multi-source information fault datasets sourced from the Paderborn Data Center [44]. The experimental platform comprises five key components: a drive motor, a torch-measuring shaft, a rolling bearing test module, a flywheel, and a load motor as illustrated in Fig. 9.

The multi-source information data, encompassing vibration and current data, were acquired using a piezoelectric accelerometer and a current sensor, and they were operating at a sampling frequency of 64,000 Hz. It is worth noting that the experiment was conducted under two different speed, load torque, and radial force conditions, namely 15 Hz/25 Hz, 0.1 N m/ 0.7 N m, and 400 N/1000 N. In this case, there are eight health conditions for the outer ring and inner ring of the rolling bearings, including normal (N_0), early single-point pitting fatigue in the inner race (I_1), repetitive single-point pitting fatigue in the inner race (I_2), multiple

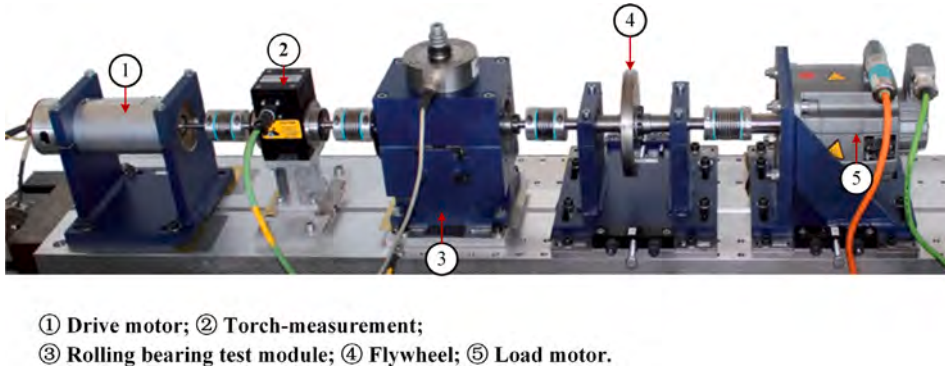


Fig. 9. Roller bearing platform.

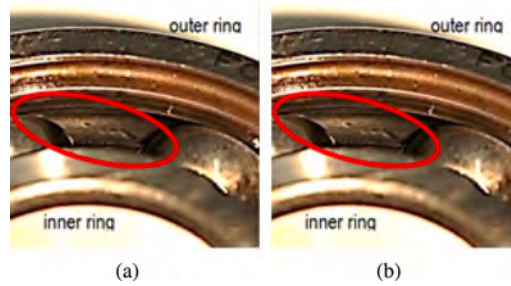


Fig. 10. Fault pictures of the bearings. (a) outer ring. (b) inner ring.

Table 3

Rolling bearing fault diagnosis based on Case Study 1.

Task	Multi-source information domains	Target domain	Labels
A_1	15 Hz/0.7 N m/1000 N (2400 samples)	25 Hz/0.7 N m/1000 N (960 samples, vibration domain)	N_0 : 0; I_1 : 1; I_2 : 2; I_3 : 3; O_4 : 4; O_5 : 5; O_6 : 6; M_7 : 7.
A_2	15 Hz/0.7 N m/1000 N (2400 samples)	25 Hz/0.1 N m/1000 N (960 samples, current domain)	
A_3	15 Hz/0.7 N m/1000 N (2400 samples)	25 Hz/0.7 N m/400 N (960 samples, vibration domain)	
A_4	25 Hz/0.7 N m/1000 N (2400 samples)	25 Hz/0.1 N m/1000 N (960 samples, current domain)	
A_5	25 Hz/0.7 N m/1000 N (2400 samples)	25 Hz/0.7 N m/400 N (960 samples, vibration domain)	
A_6	25 Hz/0.1 N m/1000 N (2400 samples)	25 Hz/0.7 N m/400 N (960 samples, current domain)	

single-point pitting fatigue damage in the inner race (I_3), early single-point pitting fatigue in the outer race (O_4), repetitive single-point pitting fatigue in the outer race (O_5), early single-point plastic deform in the outer race (O_6), and multiple single-point pitting fatigue damage in the inner race and outer race (M_7), as shown in Fig. 10, to simplify the representation, different labels are assigned for each health condition, as outlined in Table 3. For each category, 300 samples are randomly selected for training, and 120 samples are used to participate in model testing.

4.1.2. Case study 2: Fault diagnosis of cylindrical roller bearing platform

The second multi-source information signals (a.k.a., vibration and acoustic signals) were collected from Sant Longowal Institute of Engineering and Technology shown in Fig. 11. It is composed of a 346-Watt AC motor, two bearings, a shaft, a load, a break, etc. The shaft speed was set to 2050 Rpm and the vertical load was set to 200 N during experiments. The vibration and acoustic signals are acquired by a microphone and a tri-accelerometer which are mounted on the top of the bearing housing at a frequency of 70,000 Hz. In this simulation, three cylindrical roller bearing failures with different defect sizes were preset on the test stand, including inner race (IRD), outer race (ORD), and roller (RD), as demonstrated in Fig. 12.

Correspondingly, four multi-source information diagnosis tasks are formulated using the cylindrical roller bearing dataset, as depicted in Table 4. Here, a single defect size (I, II, III, and IV) functions as the target domain (a.k.a., working condition), while the remaining sizes are designated as the multi-source information domains. Then, 300 samples are generated per health condition, resulting in a total of 2400 samples in every defect size. Moreover, to constitute the training dataset for a working condition, 80% of the samples are randomly chosen from each health condition, while the remaining 20% are allocated to the testing dataset. Consequently, the training phase incorporates 2400 samples, and 600 samples are reserved for the testing phase.

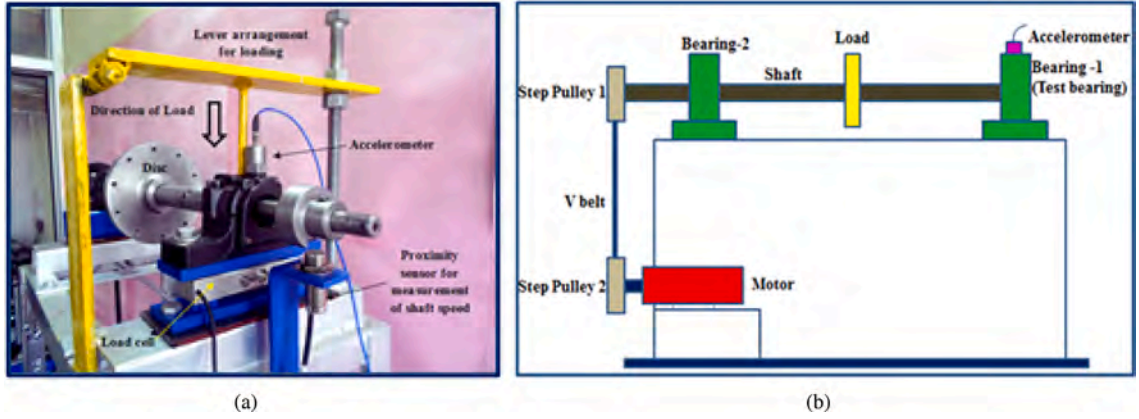


Fig. 11. Cylindrical roller bearing test rig. (a) typical photograph. (b) schematic diagram.

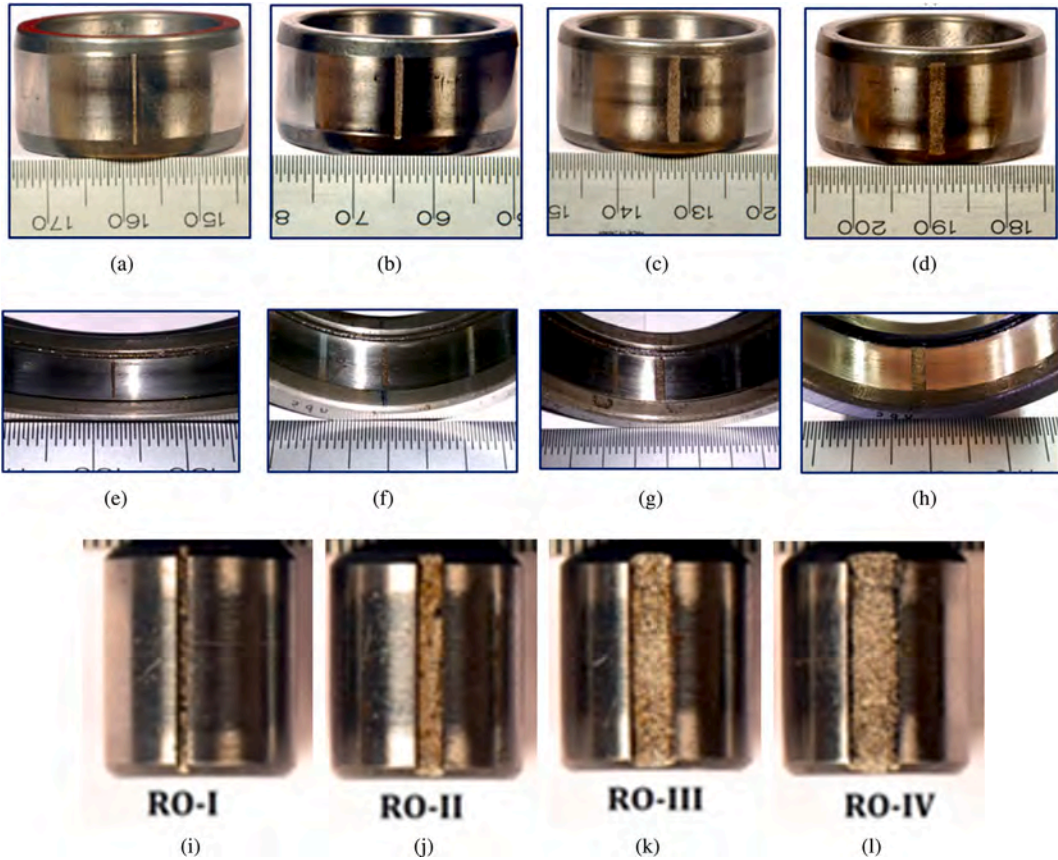


Fig. 12. Three cylindrical roller bearing health conditions. (a) IRD-I. (b) IRD-II. (c) IRD-III. (d) IRD-IV. (e) ORD-I. (f) ORD-II. (g) ORD-III. (h) ORD-IV. (i) RD-I. (j) RD-II. (k) RD-III. (l) RD-IV.

4.1.3. Case study 3: Fault diagnosis of rolling mill platform

The multi-source information data (a.k.a, vibration and acoustic data) were collected over a period of two months at Politecnico di Milano and Yanshan University, whose laboratory rig and mechanical drawing are presented in Fig. 13(a) and Fig. 13(b). The rolling mill platform consists of ① driving motor, ② coupling, ③ reduction gearbox, ④ planetary gearbox, ⑤ cross universal joint, ⑥ loading drive motor, ⑦ rolls, ⑧ loading handwheel, ⑨ vertical shaker, ⑩ horizontal shaker, and ⑪ data acquisition system. The original multi-source information signals were captured using vibration and sound sensors. The sampling time during experiments is 20 s, and the sampling frequency is set at 10, 240 Hz. Additionally, the data length is fixed at 1024. During the simulation,

Table 4
Cylindrical roller bearing fault diagnosis based on Case Study 2.

Task	Multi-source information domains	Target domain	Labels
B_1	I (2400 samples)	II (600 samples, vibration domain)	IRD: 0; ORD: 1; RD: 2.
B_2	I (2400 samples)	III (600 samples, acoustic domain)	
B_3	I (2400 samples)	IV (600 samples, vibration domain)	
B_4	II (2400 samples)	III (600 samples, acoustic domain)	
B_5	II (2400 samples)	IV (600 samples, vibration domain)	
B_6	III (2400 samples)	IV (600 samples, acoustic domain)	

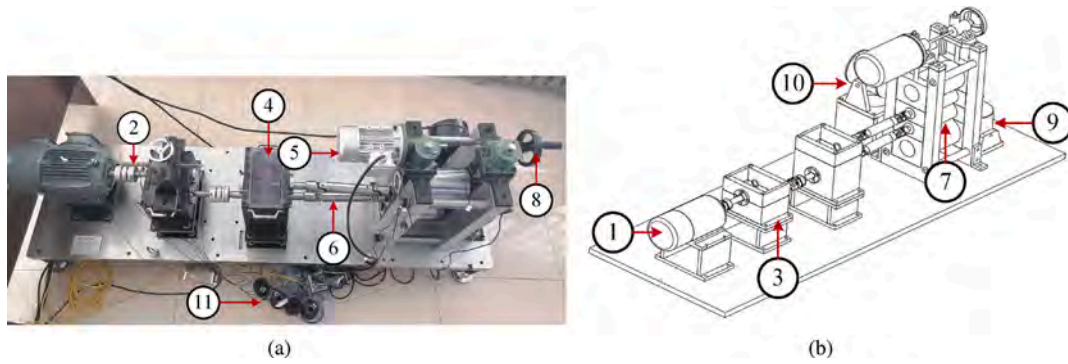


Fig. 13. Rolling mill experimental platform. (a) typical photograph. (b) mechanical drawing.

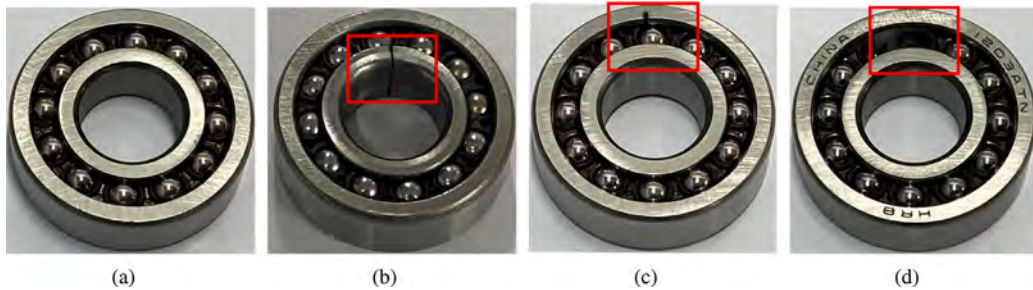


Fig. 14. Fault picture of experimental bearings. (a) N. (b) IRC. (c) ORC. (d) M.

Table 5
Bearing fault diagnosis based on Case Study 3.

Task	Multi-source information domains	Target domain	Labels
C_1	N0 (840 samples)	N1 (360 samples, vibration domain)	N: 0; IRC: 1; ORC: 2; M: 3.
C_2	N0 (840 samples)	N1 (600 samples, acoustic domain)	
C_3	N1 (840 samples)	N2 (360 samples, vibration domain)	
C_4	N1 (840 samples)	N2 (360 samples, acoustic domain)	

we preconfigured four distinct health statuses for the bearings in the upper working roll, as illustrated in Fig. 14. These states are detailed below.

- (1) Normal working condition (N).
- (2) Inner race crack in the bearing (IRC).
- (3) Outer race crack in the bearing (ORC).
- (4) Missing one ball in the bearing (M).

The experiment was conducted across three different working conditions, characterized by varying loads: N0, N1, and N2. Based on this rolling mill platform, three multi-source information diagnostic tasks (C_1 , C_2 , and C_3) are formulated in Table 5, where one load serves as the target domain, and the remaining loads are designated as multi-source information domains. Then, 300 samples are yielded for each health state and then 1200 samples are available for each task in total. Furthermore, for model training and evaluation, 70% of the samples within each working condition are randomly allocated to the training dataset, while the remaining samples constitute the testing dataset.

4.2. Comparison methods and implementation details

We introduce three standards, namely no transfer learning, best single-source transfer learning, and best multi-source transfer learning.

(1) CNNs: CNNs with no transfer learning are conventional fault diagnosis methods that can be trained using a single source data and directly test unlabeled data in the target domain.

(2) Deep adaption network (DANs) with best single-source transfer learning: DAN can use the multi-kernel maximum mean discrepancy (MK-MMD) to measure the distance between multiple high-level features between the source domain and the target domain. It can achieve distribution alignment by minimizing domain discrepancy loss [45]. The penalty parameter λ of the MMD metric loss can be set as 1 to enhance feature transferability.

(3) Joint adaptation networks (JANs) with best single-source transfer learning: JANs introduce the joint distribution inconsistency and then use Hilbert space embedding theory to measure the differences of the joint distribution. The goal is to achieve transfer learning by minimizing the joint distribution inconsistency [46]. The adaptation factor λ_p of the joint maximum mean discrepancy loss can be changed from 0 to 1 by the progressive strategy: $\lambda_p = 2/1 + \exp(-\gamma p) - 1$, and $\gamma = 10$.

(4) Conditional adversarial domain adaptation networks (CDANs) with best single-source transfer learning: CDANs can use domain adversarial learning strategy to adjust the feature representation of the source and target domains while reducing domain differences and preserving class information [47]. The conditional adversarial domain loss λ can be increased by multiplied the $1 - \exp(-\delta p)/1 + \exp(-\delta p)$, $\delta = 10$ to achieve progressive training.

(5) Deep coral correlation alignment for deep domain adaptation networks (CORALs) with best single-source transfer learning: CORALs can minimize domain differences by aligning the correlation matrices of source and target domain features achieved by second-order statistics [48]. The penalty parameter is adopted to balance the classification loss and CORAL loss, which can be defined as $2/1 + \exp(-10 * (e - 1)/E - 1) - 1$ with e and E denote the current training epoch and the max training epoch.

(6) Integrating expert knowledge with domain adaptation networks (ACDANNs) with best single-source transfer learning: ACDANNs can leverage both labeled data from the source domain and expert knowledge to achieve fault diagnosis in the target domain [49]. The domain classification loss parameter λ_d changes from 0 to 1 through $2/\exp(-10 \times p) - 1$, where p varies linearly according to the training epoch. The augmented distribution parameter λ follows the prior distribution.

(7) ADACLs with best multi-source transfer learning: ADACLs can leverage an adversarial learning strategy to align the feature distributions between the source domains and the target domain, while also ensuring that the class information is preserved to enable accurate fault diagnosis [50]. The trade-off parameters α and β denote domain distribution alignment loss and domain classifier alignment loss, respectively, which can be calculated by $\alpha = \beta = 2/1 + \exp(-\gamma(i/E)) - 1$ with $\gamma = 10$ and E means the training epoch. The SGD algorithm can be selected as the optimization function with a momentum of 0.09.

(8) MSSAs with best multi-source transfer learning: MSSAs can use TL to transfer knowledge from multi-source domains to the target domain while incorporating subdomain adaptation to handle the variations within each subdomain [51]. The value of trade-off parameter λ (structure alignment loss) is equal to $2/\exp(-10 \times l) - 1$, where l changes linearly from 0 to 1. Then, the rectified linear unit (ReLU) is used as the activation function during the training process and the input sample length is set as 1024.

(9) The proposed MSIDA-Net.

All the methods are composed using the underlying network architecture given in Table 6, and the other training settings are determined by the related studies. The network learning rate is set to 0.001, and the epochs of all methods are 300 during the hyperparameter configuration phase. Throughout the experiments, the trade-off parameters of α , β , and γ in the proposed MSIDA are set to 0.3, 0.1, and 0.8, respectively. These methods were tested on an Intel i5-10400F CPU, a GeForce RTX 3070 GPU, and a Pytorch framework.

4.3. Fault diagnosis results of three case studies

4.3.1. Diagnostic accuracy analysis

The statistical results based on three case studies are documented in Table 7 and then three diagnostic result diagrams can be shown in Fig. 15, Fig. 16, and Fig. 17, respectively (see Table 7).

The above experimental results reveal that the proposed MSIDA-Net can attain the highest average diagnostic accuracy and the smallest standard deviation in three case studies. It can be seen that the proposed method attains an average diagnostic accuracy of 96.94% across 16 MSIDA tasks, outperforming the second-highest accuracy (95.98%) achieved by MSSAs and the worst-performing accuracy (82.08%) obtained by CNNs by a margin of 0.96% and 14.86%, respectively. Meanwhile, the performance of all the methods can exhibit a degree of increase in Case Study 2 because of the less domain variance and limited category types among data under different working conditions, which makes it easy to achieve TL fault diagnosis. These achievements further highlight the advantages of MSIDA-Net in fault diagnosis based on multi-source information domains and its potential value in engineering applications. The analysis of the above diagnosis results reveals several promising trends. Hence, the following conclusions can be drawn:

(1) Regarding fault diagnosis based on different transfer standards, the MSIDA-Net method outperforms the SOTA methods examined in this study. The proposed method exhibits the highest average diagnostic performance at 96.94%, while none of the alternative methods surpasses 95.98% (as achieved by MSSAs). Furthermore, it presents a mean standard deviation of 0.67%, representing the most favorable level when compared to all other methods. This result signifies the robustness and acceptable transfer stability.

Table 6

Underlying network architecture of all the methods. Conv, BN, and FC stand for Convolutional layer, batch normalization, and fully connected layer, respectively.

Network component	Layer name	Filter
Input	Input layer	/
	Conv_Layer_1/BN/ReLU	$16 \times 15 \times 1$
Shared feature extractor (F)	Conv_Layer_2/BN/ReLU	$32 \times 3 \times 1$
	Pooling_2/ReLU	2×1 max-pool
	Conv_Layer_3/BN	$64 \times 3 \times 1$
	Pooling_3/ReLU	1×1 adaptive avg-pool
Domain attribute extractor (g)	Conv_Layer_3_1/ReLU	$64 \times 1 \times 1/16$
	Conv_Layer_3_2/ReLU	1×1 adaptive max-pool
	Output_attribute_1/Sigmoid	/
	Conv_Layer_4/BN	$128 \times 3 \times 1$
Domain feature extractor (ϕ)	Pooling_5/ReLU	1×1 adaptive avg-pool
	Conv_Layer_5_1/ReLU	$64 \times 1 \times 1/16$
	Conv_Layer_5_2/ReLU	1×1 adaptive max-pool
	Output_attribute_2/Sigmoid	/
Classifier	ReLU/Pooling_6	4×1 adaptive max-pool
	Linear_Layer_7/BN/ReLU	512×128
	Linear_Layer_8	128×10

Table 7

Experimental results of different methods on three case studies.

Task	Accuracy (%)								
A_1	85.73 \pm 2.58	92.67 \pm 1.76	93.54 \pm 2.15	93.08 \pm 1.76	95.45 \pm 1.52	98.24 \pm 1.46	97.41 \pm 1.25	98.30 \pm 0.98	99.75 \pm 0.58
A_2	80.22 \pm 3.14	88.56 \pm 2.54	89.54 \pm 1.88	92.08 \pm 2.01	91.15 \pm 1.77	90.91 \pm 1.48	93.11 \pm 1.63	92.42 \pm 1.59	93.83 \pm 0.67
A_3	82.94 \pm 2.82	91.74 \pm 1.70	93.11 \pm 1.99	94.55 \pm 1.87	95.86 \pm 1.46	97.37 \pm 0.85	96.81 \pm 1.37	97.15 \pm 0.99	98.92 \pm 0.59
A_4	80.54 \pm 2.58	89.12 \pm 1.95	88.52 \pm 2.18	90.63 \pm 1.56	91.73 \pm 1.43	91.29 \pm 1.48	93.97 \pm 1.46	95.84 \pm 0.61	94.71 \pm 0.63
A_5	78.57 \pm 2.41	93.58 \pm 1.46	94.31 \pm 2.10	92.56 \pm 1.80	97.42 \pm 1.55	99.14 \pm 0.98	97.85 \pm 2.19	98.24 \pm 1.31	99.02 \pm 0.65
A_6	75.57 \pm 3.08	86.59 \pm 2.49	87.94 \pm 1.79	86.54 \pm 2.19	89.28 \pm 1.92	88.57 \pm 1.96	91.15 \pm 0.97	91.32 \pm 0.65	92.40 \pm 0.88
B_1	90.71 \pm 1.37	92.04 \pm 0.31	92.79 \pm 1.15	95.35 \pm 0.14	94.64 \pm 1.21	98.17 \pm 0.09	99.52 \pm 0.20	99.17 \pm 0.11	100 \pm 0.0
B_2	88.55 \pm 1.43	91.64 \pm 0.90	91.93 \pm 0.73	95.65 \pm 1.36	95.54 \pm 0.97	97.12 \pm 0.30	99.67 \pm 0.13	99.56 \pm 0.47	99.6 \pm 0.15
B_3	91.74 \pm 1.53	95.25 \pm 0.70	95.12 \pm 1.09	96.42 \pm 0.99	98.07 \pm 1.15	97.48 \pm 0.82	99.64 \pm 0.32	99.35 \pm 0.44	99.9 \pm 0.01
B_4	90.44 \pm 1.64	91.66 \pm 1.82	93.84 \pm 2.53	96.56 \pm 1.58	98.48 \pm 1.04	99.17 \pm 0.77	98.44 \pm 1.12	99.32 \pm 0.55	100 \pm 0.0
B_5	89.72 \pm 1.53	92.84 \pm 1.20	95.47 \pm 1.12	98.10 \pm 0.99	97.95 \pm 0.87	99.09 \pm 0.97	98.68 \pm 0.88	99.47 \pm 0.45	99.3 \pm 0.23
B_6	85.55 \pm 2.33	87.67 \pm 1.55	88.49 \pm 1.33	90.68 \pm 1.22	91.14 \pm 0.97	90.52 \pm 1.27	93.03 \pm 1.44	93.14 \pm 1.53	94.4 \pm 0.99
C_1	81.50 \pm 2.51	88.32 \pm 1.48	90.73 \pm 1.54	94.12 \pm 1.03	91.84 \pm 1.26	93.67 \pm 0.99	97.71 \pm 0.64	97.49 \pm 0.71	98.36 \pm 0.61
C_2	70.99 \pm 2.98	78.62 \pm 1.56	81.48 \pm 1.29	84.47 \pm 1.35	81.95 \pm 1.22	88.01 \pm 1.18	89.64 \pm 0.85	87.94 \pm 1.03	91.16 \pm 0.90
C_3	80.87 \pm 2.15	85.37 \pm 1.62	89.78 \pm 2.00	92.15 \pm 1.29	88.91 \pm 1.47	93.24 \pm 1.03	98.35 \pm 0.72	97.54 \pm 0.99	98.11 \pm 0.66
C_4	71.43 \pm 2.81	80.22 \pm 1.44	83.65 \pm 1.30	88.44 \pm 1.35	88.31 \pm 1.33	90.75 \pm 1.15	92.54 \pm 0.88	93.19 \pm 1.20	94.46 \pm 0.70
Average Results	82.08	88.45	90.17	92.28	92.40	94.20	95.92	95.98	96.94
Methods	CNNs	DANs	JANs	CDANs	CORALs	ACDANNs	ADACLs	MSSAs	Ours

(2) The performance of the MSIDA-Net is better than the comparative methods in most transfer diagnostic tasks. Notably, within the most challenging task C_2 , the proposed method demonstrated a noteworthy 16.9% improvement when contrasted with the least effective method, while also achieving a 1.1% enhancement over the top-performing model among the SOTA methods. Therefore, it is deducible that the attainment of acquired transferable features and attributes can be capable of demonstrating invariance to multi-domain shifts, while simultaneously maintaining discriminative properties aligned with the learning objective.

(3) In these fault diagnosis tasks, the MSDA methods are generally better than the single-source DA methods, mainly because the single-source DA methods may lead to the emergence of suboptimal solutions in multi-source scenarios. Meanwhile, the MSIDA methods can excel in more comprehensive representations of mechanical equipment fault transfer.

(4) Interestingly, the effectiveness of all TL diagnosis methods is notably inadequate when the target domain involves the current signal. This inadequacy can be ascribed to the relatively minor impact of faults on the current signal in comparison to the vibration signal. Meanwhile, the addition of multiple source information domains sometimes does not provide any discernible advantage when compared to the single source domain. For example, the best single-source ACDANNs outperform the MSIDA methods in the diagnostic task A_5 . Therefore, diagnostic methods using the single-source information domain can achieve the best results.

(5) It can be seen that the comparative multi-source information methods outperform the proposed method in some tasks. This is because that only the proposed method is designed to transfer both the attributes and features. Sometimes, there are biases in the properties between the different information domains (i.e., different signals may have inconsistent frequencies), which can affect the transfer performance of the proposed method. In most diagnostic tasks, benefiting from the transfer of attributes and features based on multi-source information simultaneously, the proposed MSIDA-Net is still advantageous compared to the SOTA methods.

4.3.2. Feature visualization

This paper employs the t-distributed Stochastic Neighbor Embedding (t-SNE) to visualize features and conduct exploratory analysis on the final hidden layer in the proposed method for a more intuitive visualization [52]. The analysis results of the 16 diagnostic tasks of the proposed method are presented in Figs. 18–20.

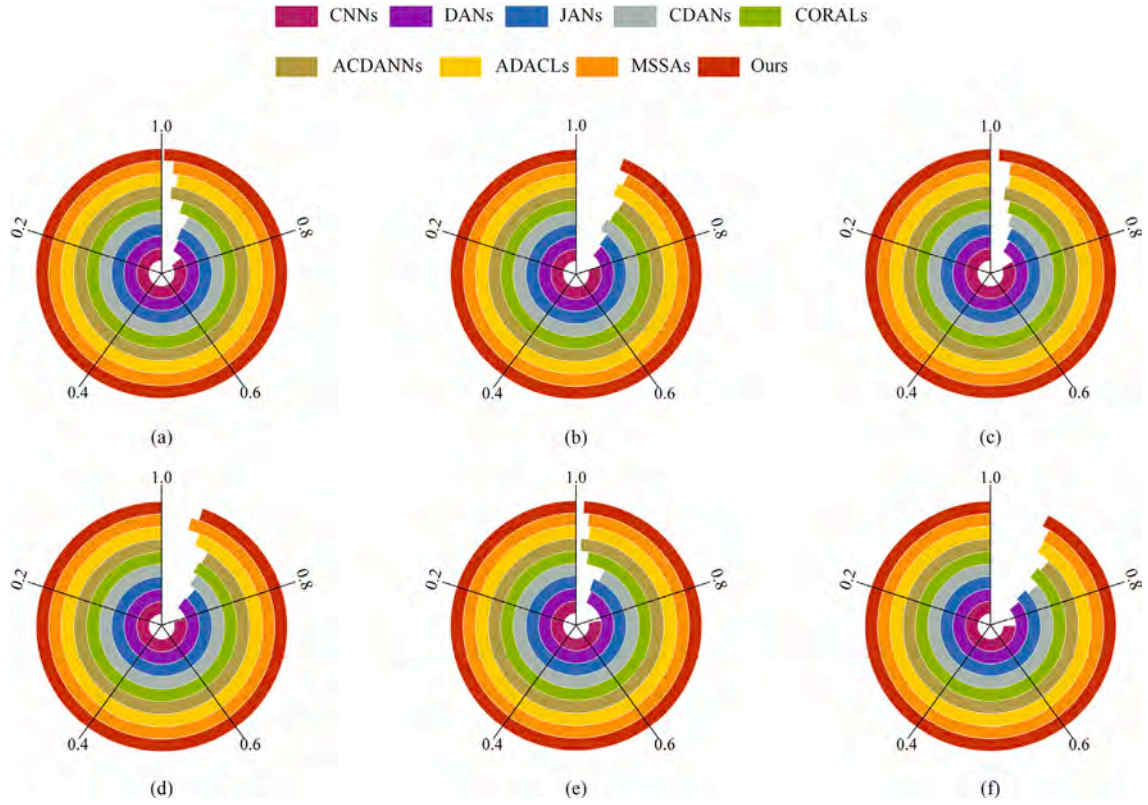


Fig. 15. Accuracy rings for each method on Case Study 1. (a) A_1 . (b) A_2 . (c) A_3 . (d) A_4 . (e) A_5 . (f) A_6 .

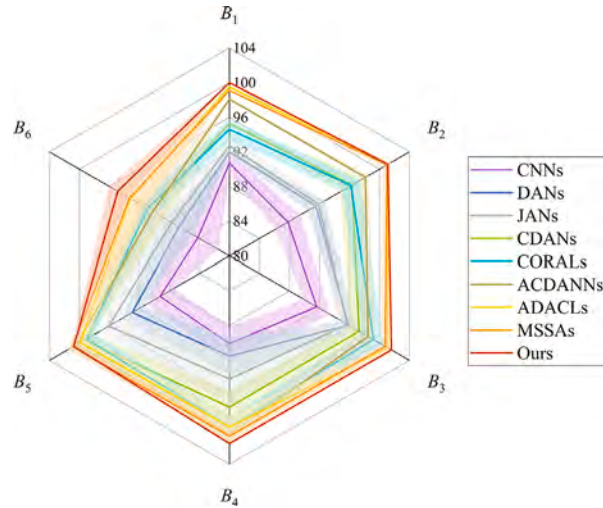


Fig. 16. Radar chart of all the methods on Case Study 2.

From Figs. 18–20, despite the minor observed mismatch, the proposed MSIDA-Net adeptly consolidates the majority of same-label features, enabling discernible decision boundaries across distinct categories and signifying successful DA. Evidently, the features of the same category between the single-source information and the target domains can exhibit strong alignment, which underscores the effectiveness of the LMMD metric in minimizing the distribution discrepancies among related subdomains. Overall, the visualization results validate the applicability and superiority of MSIDA-Net in TL fault diagnosis.

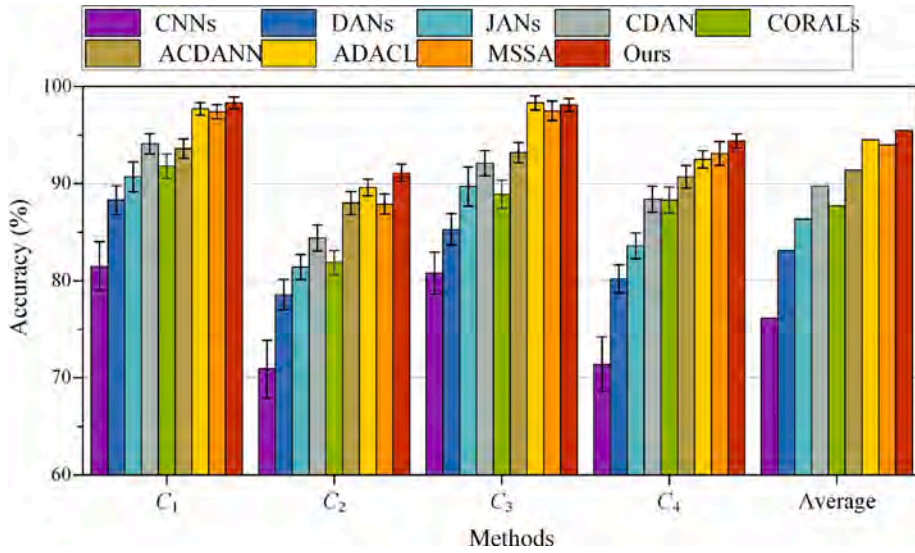


Fig. 17. Experimental results on Case Study 3 under 4 cross-domain tasks.

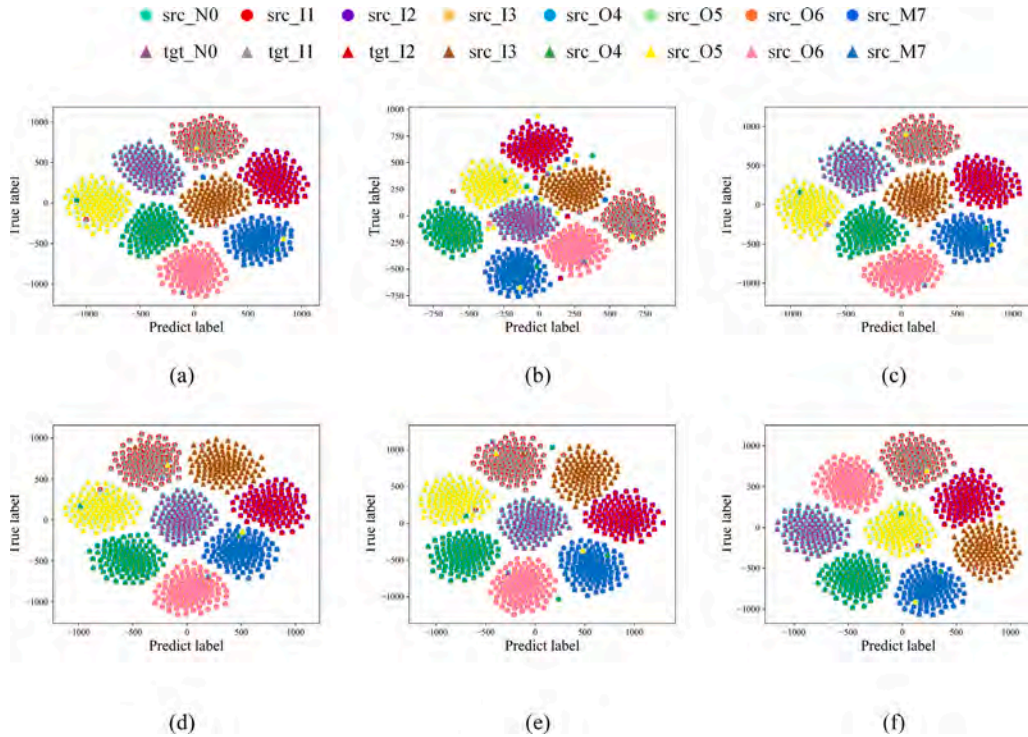


Fig. 18. T-SNE results of Case Study 1.

4.3.3. Classification performance visualization

Then, the confusion matrix is adopted to compare the output of the proposed method to the true labels of the datasets for analyzing the accuracy and effectiveness of it [53]. Hence, it can help to offer a structured framework for analyzing diagnostic results, identifying strengths and weaknesses, and making informed decisions about model tuning, feature engineering, and other improvements. The results of the confusion matrix based on 16 TL diagnostic results can be shown in Fig. 21.

As illustrated in Fig. 21, the overall test accuracy of the health states in Case Study 1 is the highest in the diagnostic task A_1 , at 99.5%. Specifically, 1.1% of O4 and 1.1% of O6 are misclassified as O6 and O4, respectively. Then, the diagnostic task A_6 has the lowest overall classification performance of 92.4%. Specifically, 56.4% of O6 is misclassified as O4. As a result, the fault

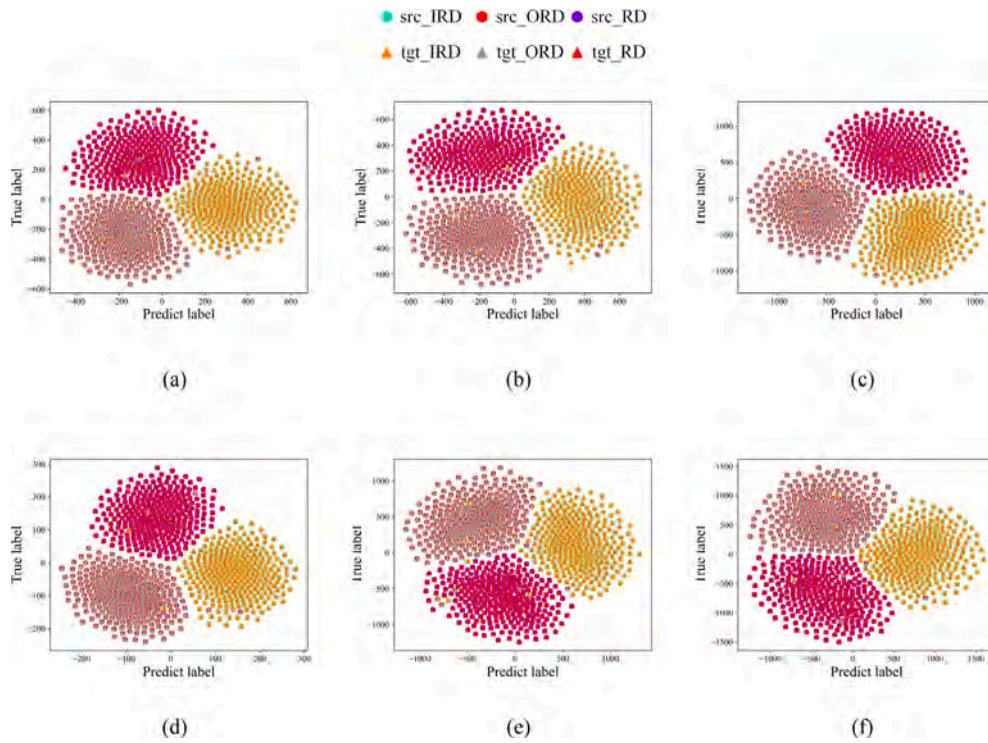


Fig. 19. T-SNE results of Case Study 2.

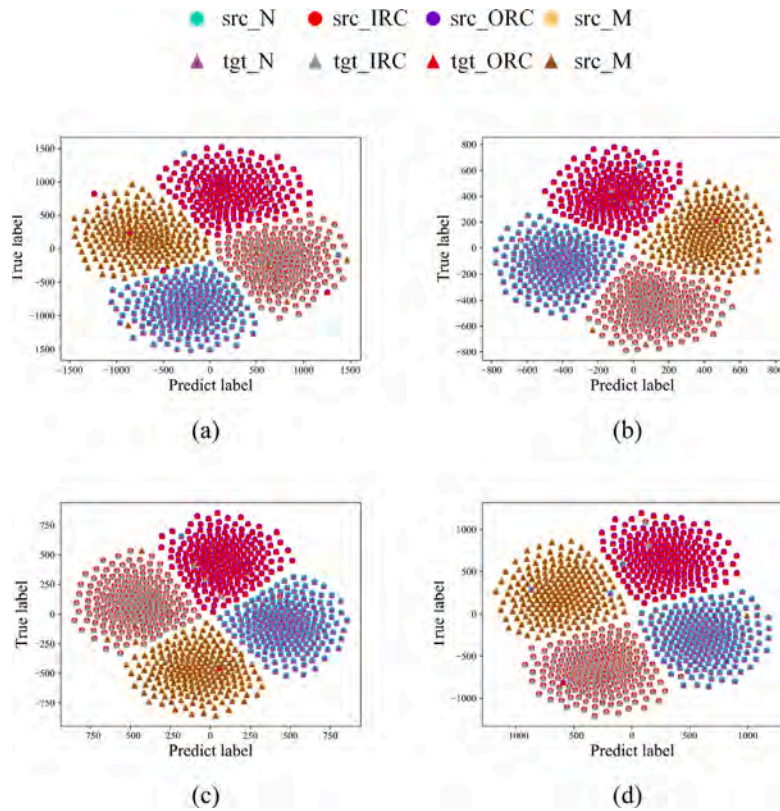


Fig. 20. T-SNE results of Case Study 3.

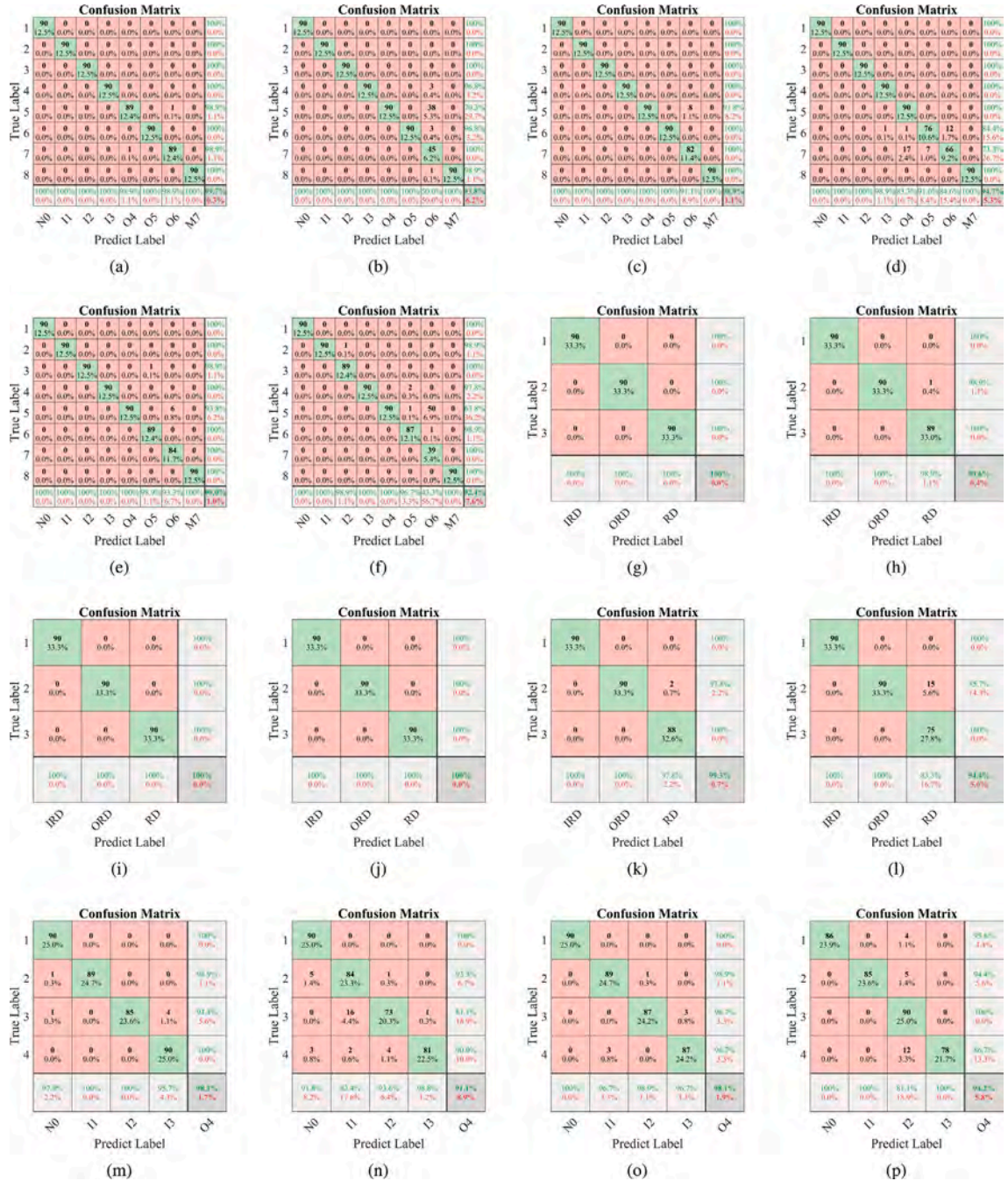


Fig. 21. Classification results of the proposed MSIDA-Net. (a) A₁, (b) A₂, (c) A₃, (d) A₄, (e) A₅, (f) A₆, (g) B₁, (h) B₂, (i) B₃, (j) B₄, (k) B₅, (l) B₆, (m) C₁, (n) C₂, (o) C₃, (p) C₄.

characteristics between O4 and O6 are too similar, leading to a slightly poorer diagnostic performance of the proposed MSIDA-Net. Especially in B₁, B₃, and B₄ from Case Study 2, 100% diagnostic accuracy was achieved for each health state. In the most challenging task B₆, 15 RD samples are erroneously classified as ORD samples. Importantly, the RD category is highly susceptible to being misclassified into the ORD category. It can be presented that the proposed method in Case Study 3 can achieve at least 81.1% diagnostic performance for each of the 4 health states in four diagnostic tasks. It is worth noting that the identification accuracy of the ORC category (81.1%) is lowest in the most challenging task C₂ and the classification performance of the ORC category (94.4%) is lowest in the simplest task C₁. In general, in 16 diagnostic tasks on three case studies, the proposed method can

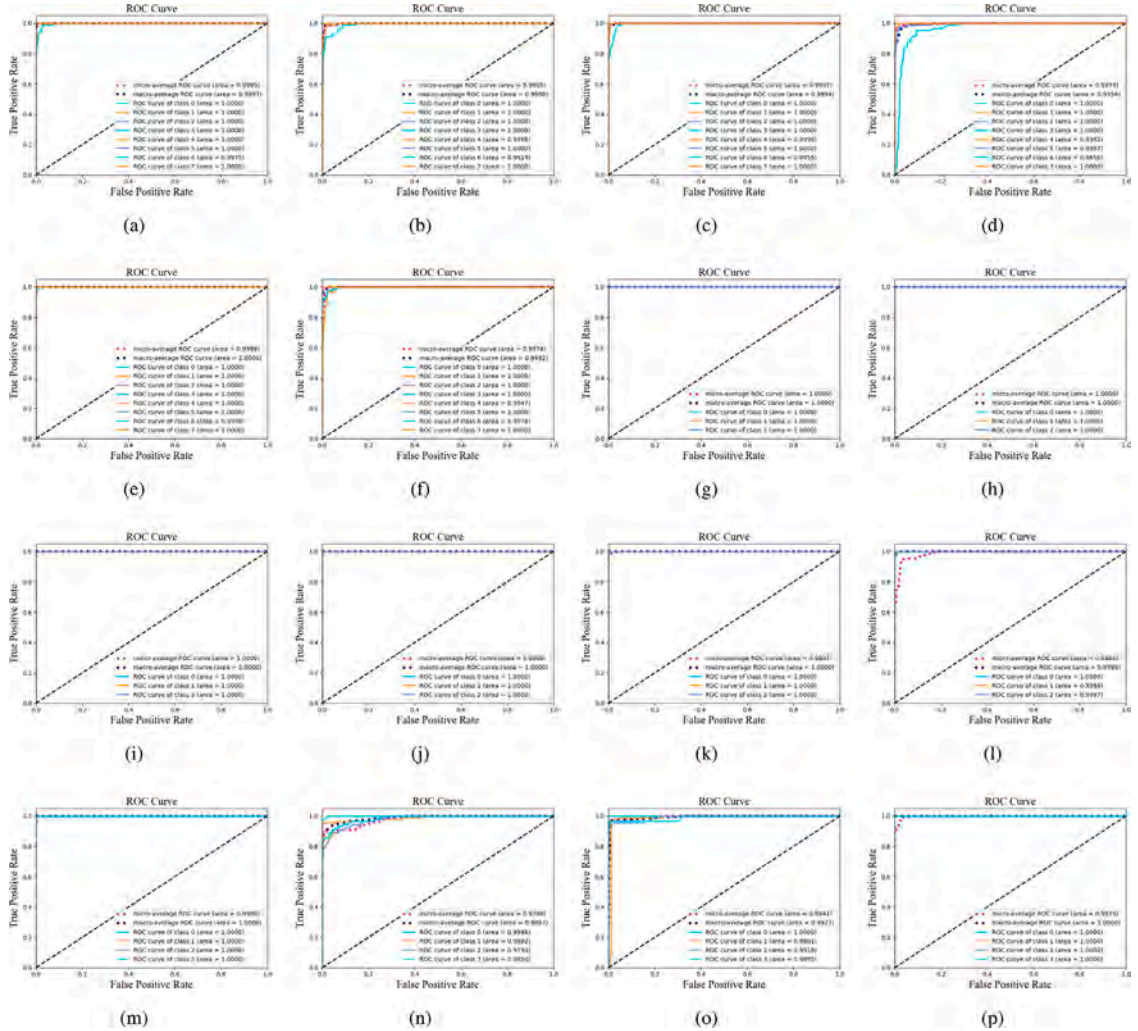


Fig. 22. Classification results of the proposed MSIDA-Net. (a) A_1 , (b) A_2 , (c) A_3 , (d) A_4 , (e) A_5 , (f) A_6 , (g) B_1 , (h) B_2 , (i) B_3 , (j) B_4 , (k) B_5 , (l) B_6 , (m) C_1 , (n) C_2 , (o) C_3 , (p) C_4 .

achieve good identification results for each fault category, which validates the outstanding generalization ability of the MSIDA-Net in cross-domain fault diagnosis.

Finally, the Receiver Operating Characteristic (ROC) curve is used in this study to assess and visualize the performance of MSIDA-Net in the trade-off between the true positive rate (sensitivity) and false positive rate as the discrimination threshold varies [54]. The x-axis represents the false positive rate (FPR), and the y-axis represents the true positive rate (TPR) or sensitivity, and the results are illustrated in Fig. 22.

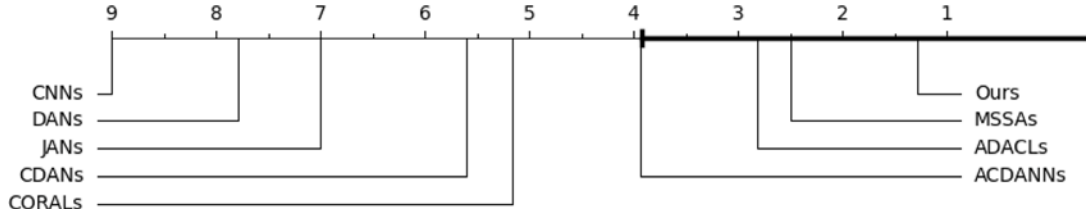
It can be observed that all the micro-average ROC curves and the macro-average ROC curves can achieve a minimum of 99.78% (obtained by A_4) and 99.54% (acquired by A_6). Then, all the area under the curve (AUC) values are more than 0.5 and indicate better-than-random performance of the proposed MSIDA-Net. As shown in Fig. 22(d), the area of class 6 under the curve is smallest, at 96.56%. The difference of the AUC between the C_2 and the other tasks was significant and the AUC of all the curves in C_2 is obviously smaller than those of the other curves. Especially, the micro-average ROC curve can only obtain 97.8%.

Finally, the diagnostic performance of all the methods is somewhat attenuated when facing Case Study 3, as this on-site dataset exhibits lower data quality compared to the other two laboratory datasets. The results in this chapter indicate that the diagnostic performance and the robustness of the proposed MSIFA-Net are superior to the SOTA methods proposed in this paper through these fundamental tools. The next chapter, therefore, moves on to discuss the performance analysis via statistical indicators.

Table 8

Average ranks of the accuracy for each diagnostic task by different methods.

	CNNs	DANs	JANs	CDANs	CORALs	ACDANNs	ADACLs	MSSAs	Ours
Average rank	9	7.78	7	5.59	5.16	3.94	2.81	2.5	1.28

**Fig. 23.** Comparison of proposed MSIDA-Net against other SOTA methods.

4.4. Performance analysis

In this section, we use four integral indicators to ensure the reliability, robustness, and generalization capability of the proposed MSIDA-Net in TL fault diagnosis tasks. Testing and performance analysis can provide insights into whether it is overfitting or underfitting. These issues can be addressed through adjustments to the complexity and training process of the proposed method.

4.4.1. Friedman test

The Friedman Test (F_i) as a non-parametric statistical analysis is employed to assess the diagnostic performance of all the methods among multiple TL tasks in this paper [55], which can be written as

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(k-1) - \tau_{\chi^2}} \quad (28)$$

$$\begin{aligned} \tau_{\chi^2} &= \frac{k-1}{k} \cdot \frac{12N}{k^2-1} \sum_{i=1}^k \left(r_i - \frac{k+1}{2} \right)^2 \\ &= \frac{12N}{k(k+1)} \left(\sum_{i=1}^k r_i^2 - \frac{k(k+1)^2}{4} \right) \end{aligned} \quad (29)$$

where k represents the number of algorithms and N stands for the number of TL diagnostic tasks.

Based on 9 methods and 16 TL diagnostic tasks, we can compute the average rank results for further analysis, as presented in Table 8. Consequently, the F_i employs the test statistic $\tau_F = 155.76$, which conforms to the F distribution with degrees of freedom $9 - 1 = 8$ and $(9 - 1) \times (16 - 1) = 120$. The null hypothesis in this paper states that the diagnosis performance of all the methods is equal across these tasks, implying no significant differences. The critical value of $F(8, 120)$ is determined based on the F_i critical values table. Specifically, at a significance level of $\alpha = 0.05$, the critical value is 2.016, and at $\alpha = 0.1$, it is 1.72. These critical values are notably lower than the computed test statistic, leading to the rejection of the null hypothesis. Subsequently, we employ the Nemenyi post-hoc test to facilitate a comparison between the SOTA methods and the proposed method as depicted in Fig. 23 and the critical difference (CD) can be calculated by

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (30)$$

where the value of q_α is 3.102 when α is 0.05.

Evidently, the methods situated on the right side of the coordinate axis demonstrate more favorable diagnostic results. As a result, the performance of the proposed method, MSSAs, ADACLs, and ACDANNs is superior to those of CORALs, CDANs, JANs, DANs, and CNNs. Especially, the methods falling outside the designated rank interval exhibit statistically significant differences from the proposed MSIDA-Net (see Fig. 23).

4.4.2. Significance test

To further compare the diagnostic significance between the proposed method and SOTA methods, we conduct the significance test (t -test) using the 10 experimental results obtained from Case Study 3 [56]. Hence, the calculation results are presented in Table 9.

Especially in Task C_3 (bold font), the p -value between the diagnostic results of ADACLs and our method exceeds 0.05, indicating that ADACLs can exhibit superior diagnostic performance to MSIDA-Net. Moreover, the p -values for alternative tests of significance are considerably below 0.05, which underscores the statistical significance of the differences between the SOTA methods and the proposed method.

Table 9Significance test (p -value) of average accuracy on Case Study 3.

Tasks	Ours	MSSAs	ADACLs	ACDANNs	CORALs	CDANs	JANs	DANs	CNNs
C_1	Benchmark	0.0032	0.0006	$4.77e^{-10}$	$1.15e^{-8}$	$3.06e^{-8}$	$4.38e^{-15}$	$4.49e^{-15}$	$1.26e^{-14}$
C_2	Benchmark	$3.42e^{-6}$	0.0005	$1.47e^{-5}$	$7.86e^{-14}$	$6.10e^{-10}$	$1.52e^{-11}$	$8.44e^{-16}$	$5.29e^{-14}$
C_3	Benchmark	0.0151	0.1456	$8.13e^{-13}$	$4.93e^{-14}$	$8.35e^{-12}$	$2.56e^{-11}$	$1.19e^{-15}$	$1.70e^{-16}$
C_4	Benchmark	0.0177	$6.71e^{-6}$	$2.63e^{-10}$	$7.58e^{-10}$	$7.41e^{-11}$	$8.76e^{-14}$	$3.55e^{-14}$	$8.36e^{-16}$

Table 10

Quantitative evaluation metrics of the classification features for different methods.

Methods	C_1 (Case Study 3)			
	J_1	J_2	J_3	J_4
Ours	41.54	42.54	41.54	41.55
MSSAs	38.47	39.48	38.47	38.46
ADACLs	39.11	40.12	39.11	39.11
ACDANNs	21.78	22.78	22.77	22.77
CORALs	15.93	16.93	15.93	15.93
CDANs	25.29	26.30	25.29	25.28
JANs	14.36	15.36	14.36	14.37
DANs	10.84	11.84	10.84	10.84
CNNs	7.97	8.97	7.97	7.97

4.4.3. Quantitative evaluation of extracted features

To comprehensively assess the diagnostic performance and feature extraction capability of each method, we incorporate inter-class covariance (S_b) and intra-class covariance (S_w) into this paper [57–60]. The decrease in S_w typically signifies the tighter clustering of representations around the class center. Conversely, the higher S_w implies greater dispersion of representations. Similarly, the elevated representations of S_b points to enhanced class distinctiveness, facilitating ease of differentiation. Conversely, reduced inter-class covariance implies increased class similarity, potentially posing challenges for differentiation. These metrics serve to quantify the quality of feature extraction across various diagnostic tasks. Typically, the greater magnitude of S_b coupled with the diminished value of S_w indicates the enhanced feature quality extracted by the method.

$$S_b = \sum_{k=1}^C N_k (v_k - v_m) (v_k - v_m)^T \quad (31)$$

$$S_w = \sum_{k=1}^C \sum_{x \in C_k} (x - v_k) (x - v_k)^T \quad (32)$$

$$v_k = \frac{1}{N_k} \sum_{x \in C_k} x \quad (33)$$

$$v_m = \frac{1}{N} \sum_{n=1}^N x_n \quad (34)$$

where x denotes the features of the output layer. C represents the total number of health state categories. N_k means the quantity of category C_k . v_k and v_m respectively are the average values of the representation vectors for category C_k and all samples.

According to [61,62], we introduce four evaluation metrics grounded in S_b and S_w to synthetically and quantitatively assess the representation quality. A higher value across these four metrics indicates the improved diagnostic performance, as defined by

$$J_1 = \frac{S_b}{S_w} \quad (35)$$

$$J_2 = \frac{S_w + S_b}{S_w} \quad (36)$$

$$J_3 = \text{Tr} [S_w^{-1} S_b] \quad (37)$$

$$J_4 = \text{Tr} [S_w^{-1} S_b] \quad (38)$$

The results for the four metrics pertaining to task C_1 in Case Study 3 are presented in Table 10. It can be seen that results of J_1 , J_2 , J_3 , and J_4 of the proposed method are superior to the SOTA methods, which indicate the better diagnosis performance. The results of J_1 to J_4 in MSIDA-Net are 41.54, 42.54, 41.54, and 41.55, while counterparts of the best-performing method in SOTA methods are 39.11, 40.12, 39.11, and 39.11, respectively. This improvement can be attributed to the ability of the proposed MSIDA-Net to transfer both domain-variant and discriminate features and attributes from multi-source information domains to the target domain.

Table 11
Complexity of MSIDA-Net.

Network component	Layer name	Filter	Output size	TC	SC
Shared feature extractor (G)	Conv Layer_1/BN/ReLU	$16 \times 15 \times 1$	$64 \times 16 \times 1010$	242 400	16 400
	Conv Layer_2/BN/ReLU	$32 \times 3 \times 1$	$64 \times 32 \times 1008$	1 548 288	33 792
	Pooling_2/ReLU	2×1 max-pool	$64 \times 32 \times 504$	0	0
	Conv Layer_3/BN	$64 \times 3 \times 1$	$64 \times 32 \times 502$	1542144	19136

Table 12
Computational cost comparisons of different methods.

		Methods	Training time (s)	Testing time (s)	GPU memory utilization (GB)
Task C_1 as a representative	best multi-source	Ours	273.1	11.5	7.1
		MSSAs	267.9	10.8	7.1
		ADACLs	259.8	10.3	7.1
		ACDANNs	248.5	10.2	7.1
		CORALs	230.5	9.7	7.1
	best single-source	CDANs	233.8	10.0	7.1
		JANs	199.4	9.4	7.0
		DANs	182.8	9.2	6.9
	No Transfer	CNNs	169.4	9.0	6.9

4.4.4. Complexity analysis

The complexity analysis can guide model optimization and future deployment, ultimately contributing to the development of a more efficient and accountable fault diagnosis framework. Hence, we introduce time complexity (TC) and spatial complexity (SC) to assess the proposed method from. Hence, the TC and SC of the 1-D convolutional (1-D Conv) layer and fully connected layer (FC) can be defined as

$$TC_{1-DConv} = O(D_{out} \cdot Kl \cdot C_{in} \cdot C_{out}) \quad (39)$$

$$TC_{FC} = O(D_{out} \cdot W \cdot H \cdot D_{in}) \quad (40)$$

$$SC_{1-DConv} = O(K \cdot C_{in} \cdot C_{out} + D_{out} \cdot C_{out}) \quad (41)$$

$$SC_{FC} = O(D_{out} \cdot W \cdot H \cdot D_{in}) \quad (42)$$

where W and H represent the width and height of the feature maps. Kl means the kernel length. C_{in} and C_{out} denotes the number of input and output channels, respectively. D_{in} and D_{out} are the input feature dimensions and output feature dimensions.

We take the example of the shared feature extractor module in Table 6 and the results can be listed in Table 11. It can be seen that the whole TC and SC of the shared feature extractor module can be computed as 3332832 FLOPs and 69328 Bytes, respectively.

Meanwhile, we list the computational costs during the training process of methods in Table 12, including training time, testing time, and GPU memory utilization.

It can be shown that the average training time of the best multi-source information diagnostic methods is longer than that of the best single-source information diagnostic methods because the former needs to cope with multi-source information data. Then, the proposed method takes the longest time because it includes the knowledge fusion module. Especially, with the improvement of hardware, the disadvantage of long training time in the proposed method can be overcome.

5. Discussions

5.1. Ablation study

To further evaluate the importance of each innovation to the proposed MSIDA-Net in this paper, the ablation study is performed. Note that all variants are trained 10 times simultaneously using the same network and trade-off parameters for fair comparison. Five additional variants are derived from the proposed MSIDA-Net, dubbed: (1) Method 1 (M1): The MSIDA-Net without transferable attribute learning, (2) Method 2 (M2): The MSIDA-Net without transferable features learning, (3) Method 3 (M3): The MSIDA-Net with MMD metric, (4) Method 4 (M4): The MSIDA-Net without PLL strategy and (5) Method 5 (M5): The MSIDA-Net without intra-class compactness learning strategy. The first diagnostic task in each case studies was selected for comparison and display. Table 13 lists the average diagnostic results of five variations and the proposed method. Meanwhile, the ten experimental results for these methods in C_1 are shown in Fig. 24 as box and scatter diagrams.

It can be seen from the above experimental results that the proposed MSIDA-Net can exhibit greater stability in diagnostic performance and achieve the highest accuracy. Specially, it is 1.11%, 1.94%, 0.68%, 0.35%, and 0.11% higher than M1, M2, M3, M4, and M5 in diagnostic task A_1 , respectively. It is 0.74%, 1.61%, 0.49%, 0.25%, and 0.11% higher than M1, M2, M3, M4, and M5

Table 13
Experimental results of ablation study in different tasks.

Task (%)	M1	M2	M3	M4	M5	Ours
A_6	87.52	87.14	88.12	89.43	90.59	92.40
B_6	89.28	88.82	90.80	91.26	92.65	94.4
C_2	89.05	88.45	90.42	91.51	92.37	94.46

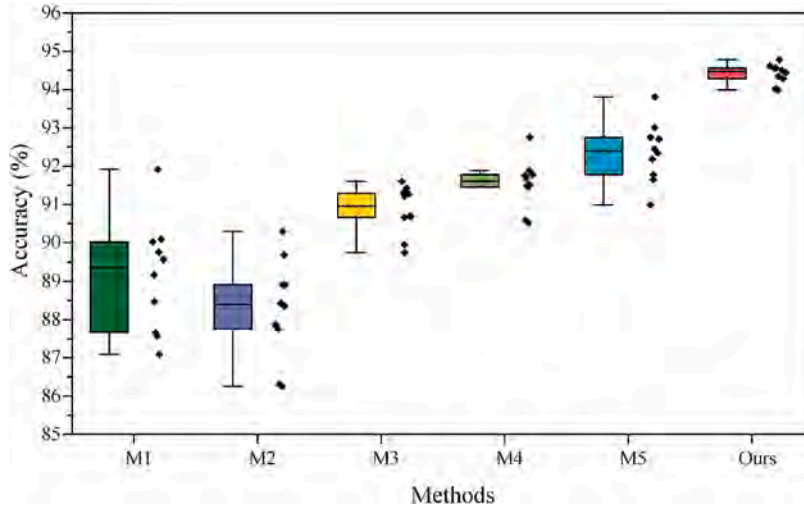


Fig. 24. Comparison of proposed MSIDA-Net against other SOTA methods.

in diagnostic task B_1 , respectively. It is 0.99%, 1.9%, 0.51%, 0.28%, and 0.23 higher than M1, M2, M3, M4, and M5 in diagnostic task C_1 , respectively. Such significant improvement indicates that it is critical to consider both transferable attributes and features learning and embed them in cross-domain fault diagnosis. In addition, the LMMD metric, PLL strategy, and intra-class compactness learning strategy can also improve the performance of the model to some extent.

5.2. Weight visualization

To vividly demonstrate the dynamic contributions of the individual single-source information domain to final fault diagnosis in the knowledge fusion module, Fig. 25 introduces the distribution distance weights from the multi-source information domains to the target domain under different tasks in Case Study 3. It can be seen that as the number of training epochs increases, the learned weights gradually stabilize and the two weights are complementary to each other. Specifically, the closer the working condition of the single-source information domain is to the target domain, the larger the weight obtained, and vice versa.

5.3. Parameter sensitivity analysis

We investigate the effect of domain attribute loss α , the LMMD factor β , and intra-class compactness parameters γ in the trade-off parameters on the performance of the proposed method. We first set β as 0 and increase α and γ linearly from 0.1 to 1 and then report the results of task C_1 in Case Study 3 in Fig. 26.

To simplify the complex process caused by three variable coefficients, we first fix β to be 0 and then we adjust the parameters between α and γ , which can be shown in Fig. 26. The diagram is plotted through the proof of exhaustion between α and γ , it can reflect that the darker the area, the higher the diagnostic accuracy. Especially when α is constant, the diagnostic performance of the proposed method first increases and then decreases as γ increases. Meanwhile, the influence of α on model effectiveness shows a concave-shaped curve when γ is fixed. Based on the experimental results, the proposed MSIDA-Net has the best performance of 98.14% when α , γ , β are 0.3, 0.1, and 0, respectively. Then, we fix α and γ to be 0.3 and 0.1 and adjust β from 0 to 1, the influence of β on model performance can be seen in Fig. 27. The result (red dashed line) shows that the diagnostic accuracy of the proposed method is a bell-shaped curve and the diagnostic performance of β remains generally increase as β goes from 0.3 to 0.8, with the optimal diagnostic at $\beta = 0.8$. In conclusion, three trade-off parameters α , γ , β can be set as 0.3, 0.1, and 0.8, respectively.

5.4. Diagnostic performance under heavy noise

In the practical application scenarios of fault diagnosis, multi-source fault signals collected by multi-source sensors are highly non-stationary due to the strong noise interference. Hence, we added additional Gaussian white noise with different signal-noise

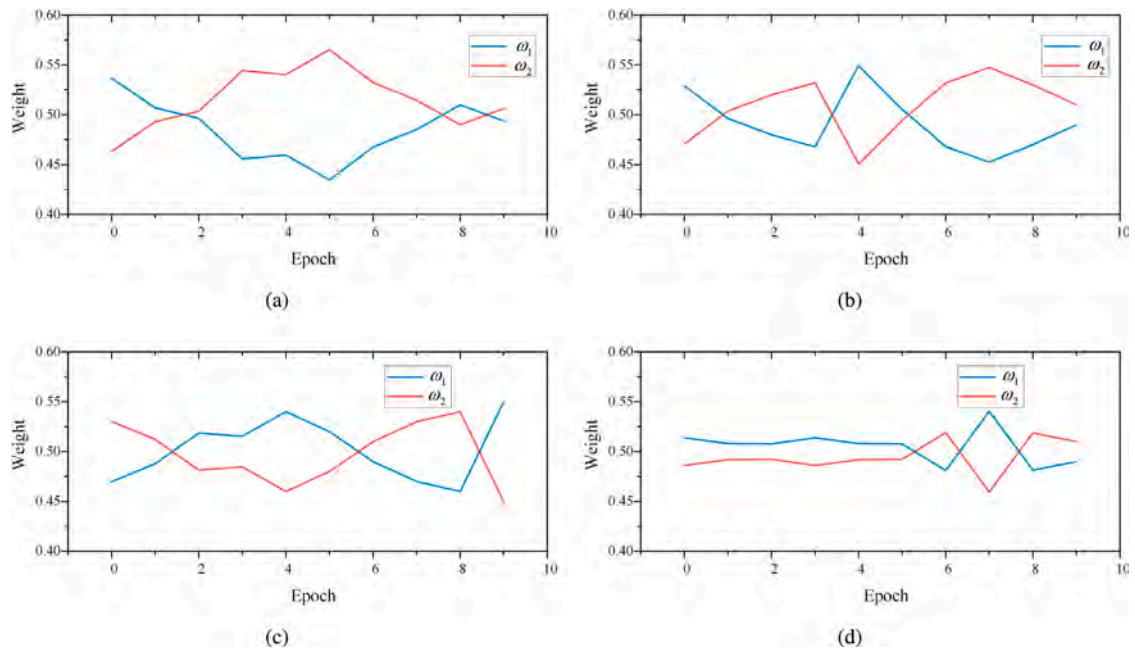


Fig. 25. Distribution distance weights learned on Case Study 3. (a) C_1 . (b) C_2 . (c) C_3 . (d) C_4 .

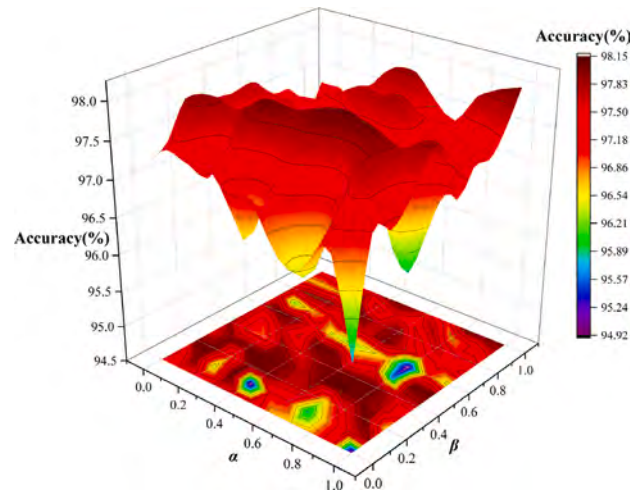


Fig. 26. Our method with different parameter settings on task C_1 in Case Study 3.

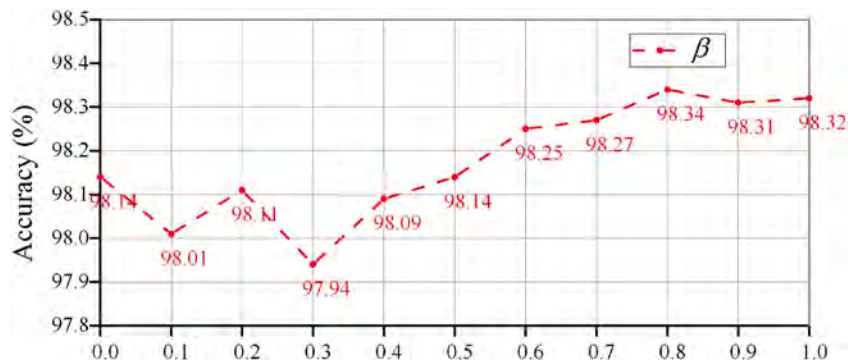


Fig. 27. Sensitivity of β when α is 0.3 and γ is 0.1 in Eq. (19).

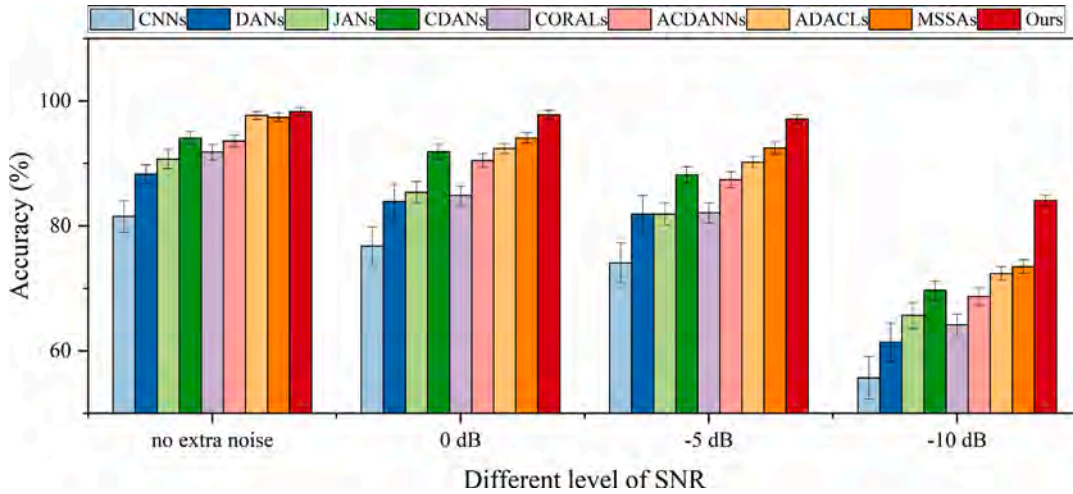


Fig. 28. Our method with different parameter settings on task C_1 in Case Study 3.

ratio (SNR) levels (including 0 dB, -5 dB, and -10 dB) to the multi-source signals acquired in both the multi-source information domain and the target domain to evaluate the diagnostic robustness of the proposed method. Fig. 28 shows the comprehensive and systematical comparisons of all methods from C_1 diagnosis task on Case Study 3.

From the graph above we can see that increasing the levels of the SNR led to a reduction in the diagnostic performance of all methods. The diagnostic results of the proposed MSIDA-Net are $97.81 \pm 0.71\%$, $97.15 \pm 0.75\%$ and $84.2 \pm 0.86\%$ when the SNR levels are 0 dB, -5 dB and -10 dB, respectively. Especially when the SNR level is -10 dB, the fault identification results are less than 80% except the proposed method. Hence, we can conclude that the diagnosis effectiveness and robustness of the proposed method is significantly superior to other comparative methods under heavy noise. This is attributed to the transferability of both the features and the attributes from the multi-source information domains to the target domain simultaneously, making the proposed MSIDA-Net can mine domain-invariant, discriminative, and transferable information from multi-source signals under heavy noise interference.

6. Future work

(1) This paper investigates the issue of cross-domain fault diagnosis in the multi-source information domains and we assume that the health state categories between the multi-source information domains and the target domain are the same. However, the key components in the mechanical equipment are prone to unseen faults in the diverse production environment. Hence, we will introduce an unknown mode detector to identify the unseen faults in the target domain.

(2) The multi-source information domains in this paper mainly refer to the current domain, the vibration domain, and the sound domain, where the vibration domain and current domain are contact signals. However, contact-based sensing technology is limited in its application in many industries due to reasons such as limited space, harsh environments, or cost restrictions. Therefore, we will introduce an infrared camera to acquire thermal images to ensure the safe operation of mechanical equipment in the future.

(3) Compared with MSDA, the multi-target DA (MTDA) refers to the process of transferring knowledge from one source domain to multiple target domains through a fault diagnosis model. It is more promising and challenging because it can solve the real challenges by minimizing the prevalence of variability and uncertainty in fault diagnosis with less cost in label and time. Hence, we want to combine multi-source information and MTDA to further improve the adaption of the diagnosis method.

(4) Labeled samples are often scarce when targeting the single-source machine. Thus, we can collect a series of limited samples from machines of different sources to form a sufficient number of labeled samples for building a cross-domain diagnostic model. However, there is a fatal problem the domain differences between different machine samples tend to be larger than the domain differences between samples from different working conditions on the same machine. Hence, we will introduce expert knowledge and domain adversarial strategy to mitigate this issue.

7. Conclusions

In this article, a MSIDA-Net based on a new term MSIDA is proposed to address some drawbacks of MSDA by transferring latent and discriminate attributes and features for achieving cross-domain fault diagnosis in a unified framework. The conclusions of this paper can be summarized as follows: (1) through the attention mechanism and the domain attribute loss function, the domain-invariant attributes from multi-source information attributes; (2) the LMMD metric is adopted to estimate category discrepancy between multi-source information domains and target domain; (3) the PLL strategy and intra-class compactness learning are leveraged to improve feature extraction capability; and (4) the knowledge fusion module is introduced to synthesize the results

from single-source domain and target domain to produce more reliable diagnostic results. In response to the above innovations, the average diagnostic accuracy on three case studies is 96.4%, 98.9%, and 95.4%, respectively, which confirms the applicability and superiority of the MSIDA-Net compared with the SOTA methods.

Meanwhile, the proposed MSIDA-Net can only achieve single-target domain TL for fault diagnosis. But as the mechanical equipment often encounters variable and complicated operating conditions, the unlabeled target data may have various distributions. Hence, the multi-target information domain adaptive (MTIDA) methods are imperative. We intend to apply it to the mechanical equipment to identify unknown faults in the future work.

CRedit authorship contribution statement

Yue Yu: Writing – original draft, Visualization, Validation, Software, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Hamid Reza Karimi:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Peiming Shi:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Rongrong Peng:** Writing – review & editing, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Shuai Zhao:** Writing – review & editing, Visualization, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Data availability

Data will be made available on request.

Acknowledgments

This research is supported by the scholarship from the China Scholarship Council (CSC), China under Grant CSC N202308130067. This research is supported by the Initial Scientific Research Foundation for Talented Scholars of Nanchang Institute of Science and Technology, China, under Grant NGRZX-23-09 and in part by the Horizon Marie Skłodowska-Curie Actions program under Grant 101073037.

References

- [1] Jun Zhu, Nan Chen, Weiwen Peng, Estimation of bearing remaining useful life based on multiscale convolutional neural network, *IEEE Trans. Ind. Electron.* 66 (4) (2018) 3208–3216.
- [2] Yi Wang, Baoping Tang, Yi Qin, Tao Huang, Rolling bearing fault detection of civil aircraft engine based on adaptive estimation of instantaneous angular speed, *IEEE Trans. Ind. Inform.* 16 (7) (2019) 4938–4948.
- [3] He Zhiyi, Shao Haidong, Jing Lin, Cheng Junsheng, Yang Yu, Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder, *Measurement* 152 (2020) 107393.
- [4] Hongru Cao, Haidong Shao, Xiang Zhong, Qianwang Deng, Xingkai Yang, Jianping Xuan, Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds, *J. Manuf. Syst.* 62 (2022) 186–198.
- [5] Baoping Cai, Chaoyang Sheng, Chuntan Gao, Yonghong Liu, Mingwei Shi, Zengkai Liu, Qiang Feng, Guijie Liu, Artificial intelligence enhanced reliability assessment methodology with small samples, *IEEE Trans. Neural Netw. Learn. Syst.* (2021).
- [6] Meng Gan, Cong Wang, et al., Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings, *Mech. Syst. Signal Process.* 72 (2016) 92–104.
- [7] Peiming Shi, Yue Yu, Hao Gao, Changchun Hua, A novel multi-source sensing data fusion driven method for detecting rolling mill health states under imbalance and limited datasets, *Mech. Syst. Signal Process.* 171 (2022) 108903.
- [8] Tianfu Li, Zhibin Zhao, Chuang Sun, Li Cheng, Xuefeng Chen, Ruqiang Yan, Robert X Gao, WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis, *IEEE Trans. Syst. Man Cybern.: Syst.* 52 (4) (2021) 2302–2312.
- [9] Yue Yu, Peiming Shi, Jinghui Tian, Xuefang Xu, Changchun Hua, Rolling mill health states diagnosing method based on multi-sensor information fusion and improved DBNs under limited datasets, *ISA Trans.* 134 (2023) 529–547.
- [10] Botao An, Shibin Wang, Ruqiang Yan, Weihua Li, Xuefeng Chen, Adaptive robust noise modeling of sparse representation for bearing fault diagnosis, *IEEE Trans. Instrum. Meas.* 70 (2020) 1–12.
- [11] Tianci Zhang, Jinglong Chen, Fudong Li, Kaiyu Zhang, Haixin Lv, Shuilong He, Enyong Xu, Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions, *ISA Trans.* 119 (2022) 152–171.
- [12] Long Wen, Xinyu Li, Liang Gao, A transfer convolutional neural network for fault diagnosis based on ResNet-50, *Neural Comput. Appl.* 32 (2020) 6111–6124.
- [13] Bin Yang, Yaguo Lei, Songci Xu, Chi-Guhn Lee, An optimal transport-embedded similarity measure for diagnostic knowledge transferability analytics across machines, *IEEE Trans. Ind. Electron.* 69 (7) (2021) 7372–7382.
- [14] Hongru Cao, Haidong Shao, Xiang Zhong, Qianwang Deng, Xingkai Yang, Jianping Xuan, Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds, *J. Manuf. Syst.* 62 (2022) 186–198.
- [15] Rong Zhu, Weiwen Peng, Dong Wang, Cheng-Geng Huang, Bayesian transfer learning with active querying for intelligent cross-machine fault prognosis under limited data, *Mech. Syst. Signal Process.* 183 (2023) 109628.
- [16] Liangjun Feng, Chunhui Zhao, Fault description based attribute transfer for zero-sample industrial fault diagnosis, *IEEE Trans. Ind. Inform.* 17 (3) (2020) 1852–1862.

- [17] Zhenghong Wu, Hongkai Jiang, Tengfei Lu, Ke Zhao, A deep transfer maximum classifier discrepancy method for rolling bearing fault diagnosis under few labeled data, *Knowl.-Based Syst.* 196 (2020) 105814.
- [18] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Jingwu Chen, Zhiping Shi, Wenjuan Wu, Qing He, Multi-representation adaptation network for cross-domain image classification, *Neural Netw.* 119 (2019) 214–221.
- [19] Morvarid Karimpour, Shiva Noori Saray, Jafar Tahmoresnezhad, Mohammad Pourmahmood Aghababa, Multi-source domain adaptation for image classification, *Mach. Vis. Appl.* 31 (2020) 1–19.
- [20] Sancheng Peng, Rong Zeng, Lihong Cao, Aimin Yang, Jianwei Niu, Chengqing Zong, Guodong Zhou, Multi-source domain adaptation method for textual emotion classification using deep and broad learning, *Knowl.-Based Syst.* 260 (2023) 110173.
- [21] Han Guo, Ramakanth Pasunuru, Mohit Bansal, Multi-source domain adaptation for text classification via distancenet-bandits, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, (no. 05) 2020, pp. 7830–7838.
- [22] Dan Zhang, Mao Ye, Yiguang Liu, Lin Xiong, Lihua Zhou, Multi-source unsupervised domain adaptation for object detection, *Inf. Fusion* 78 (2022) 138–148.
- [23] Giovanni Pasqualino, Antonino Furnari, Giovanni Maria Farinella, Unsupervised multi-camera domain adaptation for object detection in cultural sites, in: *International Conference on Image Analysis and Processing*, Springer, 2022, pp. 713–724.
- [24] Zuoyi Chen, Jun Wu, Chao Deng, Chao Wang, Yuanhang Wang, Residual deep subdomain adaptation network: A new method for intelligent fault diagnosis of bearings across multiple domains, *Mech. Mach. Theory* 169 (2022) 104635.
- [25] Bin Yang, Songci Xu, Yaguo Lei, Chi-Guhn Lee, Edward Stewart, Clive Roberts, Multi-source transfer learning network to complement knowledge for intelligent diagnosis of machines with unseen faults, *Mech. Syst. Signal Process.* 162 (2022) 108095.
- [26] Xingqiu Li, Hongkai Jiang, Min Xie, Tongqing Wang, Ruixin Wang, Zhenghong Wu, A reinforcement ensemble deep transfer learning network for rolling bearing fault diagnosis with multi-source domains, *Adv. Eng. Inform.* 51 (2022) 101480.
- [27] Jinghui Tian, Dongying Han, Hamid Reza Karimi, Yu Zhang, Peiming Shi, Deep learning-based open set multi-source domain adaptation with complementary transferability metric for mechanical fault diagnosis, *Neural Netw.* 162 (2023) 69–82.
- [28] Shengkang Yang, Xianguang Kong, Qibin Wang, Zhongquan Li, Han Cheng, Linyang Yu, A multi-source ensemble domain adaptation method for rotary machine fault diagnosis, *Measurement* 186 (2021) 110213.
- [29] Wei Cao, Zong Meng, Dengyun Sun, Jingbo Liu, Yang Guan, Lixiao Cao, Jimeng Li, Fengjie Fan, A two-stage domain alignment method for multi-source domain fault diagnosis, *Measurement* 214 (2023) 112818.
- [30] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, Bo Wang, Moment matching for multi-source domain adaptation, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1406–1415.
- [31] Zhibin Zhao, Qiyang Zhang, Xiaolei Yu, Chuang Sun, Shibin Wang, Ruqiang Yan, Xuefeng Chen, Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–28.
- [32] Meng Ma, Chuang Sun, Xuefeng Chen, Deep coupling autoencoder for fault diagnosis with multimodal sensory data, *IEEE Trans. Ind. Inform.* 14 (3) (2018) 1137–1145.
- [33] Jian Tang, Junfei Qiao, Zhiwei Wu, Tianyou Chai, Jian Zhang, Wen Yu, Vibration and acoustic frequency spectra for industrial process modeling using selective fusion multi-condition samples and multi-source features, *Mech. Syst. Signal Process.* 99 (2018) 142–168.
- [34] Tingli Xie, Xufeng Huang, Seung-Kyum Choi, Intelligent mechanical fault diagnosis using multisensor fusion and convolution neural network, *IEEE Trans. Ind. Inform.* 18 (5) (2021) 3213–3223.
- [35] Jian Lin, Haidong Shao, Xiangdong Zhou, Baoping Cai, Bin Liu, Generalized MAML for few-shot cross-domain fault diagnosis of bearing driven by heterogeneous signals, *Expert Syst. Appl.* (2023) 120696.
- [36] Xingkai Chen, Haidong Shao, Yiming Xiao, Shen Yan, Baoping Cai, Bin Liu, Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network, *Mech. Syst. Signal Process.* 198 (2023) 110427.
- [37] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, Qing He, Deep subdomain adaptation network for image classification, *IEEE Trans. Neural Netw. Learn. Syst.* 32 (4) (2020) 1713–1722.
- [38] Xue-yang Zhang, Lang He, Xiao-kang Wang, Jian-qiang Wang, Peng-fei Cheng, Transfer fault diagnosis based on local maximum mean difference and K-means, *Comput. Ind. Eng.* 172 (2022) 108568.
- [39] Pengfei Liang, Zhuozu Yu, Bin Wang, Xuefang Xu, Jiaye Tian, Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network, *Adv. Eng. Inform.* 57 (2023) 102075.
- [40] Zhongying Deng, Kaiyang Zhou, Yongxin Yang, Tao Xiang, Domain attention consistency for multi-source domain adaptation, 2021, arXiv preprint [arXiv:2111.03911](https://arxiv.org/abs/2111.03911).
- [41] Nitish Shirish Keskar, Richard Socher, Improving generalization performance by switching from adam to sgd, 2017, arXiv preprint [arXiv:1712.07628](https://arxiv.org/abs/1712.07628).
- [42] Ashia C Wilson, Rebecca Roelofs, Mitchell Stern, Nati Srebro, Benjamin Recht, The marginal value of adaptive gradient methods in machine learning, in: *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [43] Esraa Hassan, Mahmoud Y Shams, Noha A Hikal, Samir Elmougy, The effect of choosing optimizer algorithms to improve computer vision tasks: A comparative study, *Multimedia Tools Appl.* 82 (11) (2023) 16591–16633.
- [44] Wei Dong, Shuqing Zhang, Anqi Jiang, Wanlu Jiang, Liguang Zhang, Mengfei Hu, Intelligent fault diagnosis of rolling bearings based on refined composite multi-scale dispersion q-complexity and adaptive whale algorithm-extreme learning machine, *Measurement* 176 (2021) 108977.
- [45] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, Victor Lempitsky, Domain-adversarial training of neural networks, 2016, [arXiv:1505.07818](https://arxiv.org/abs/1505.07818).
- [46] Mingsheng Long, Han Zhu, Jianmin Wang, Michael I. Jordan, Deep transfer learning with joint adaptation networks, in: *International Conference on Machine Learning*, PMLR, 2017, pp. 2208–2217.
- [47] Mingsheng Long, Zhangjie Cao, Jianmin Wang, Michael I Jordan, Conditional adversarial domain adaptation, in: *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [48] Baochen Sun, Kate Saenko, Deep coral: Correlation alignment for deep domain adaptation, in: *Computer Vision—ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14*, Springer, 2016, pp. 443–450.
- [49] Qin Wang, Cees Taal, Olga Fink, Integrating expert knowledge with domain adaptation for unsupervised fault diagnosis, *IEEE Trans. Instrum. Meas.* 71 (2021) 1–12.
- [50] Yongchao Zhang, Zhaohui Ren, Shihua Zhou, Tianzhuang Yu, Adversarial domain adaptation with classifier alignment for cross-domain intelligent fault diagnosis of multiple source domains, *Meas. Sci. Technol.* 32 (3) (2020) 035102.
- [51] Jinghui Tian, Dongying Han, Mengdi Li, Peiming Shi, A multi-source information transfer learning method with subdomain adaptation for cross-domain fault diagnosis, *Knowl.-Based Syst.* 243 (2022) 108466.
- [52] Xinya Wu, Yan Zhang, Changming Cheng, Zhike Peng, A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery, *Mech. Syst. Signal Process.* 149 (2021) 107327.
- [53] Hao Su, Ling Xiang, Aijun Hu, Yonggang Xu, Xin Yang, A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions, *Mech. Syst. Signal Process.* 169 (2022) 108765.
- [54] Yu Zhang, Dongying Han, Jinghui Tian, Peiming Shi, Domain adaptation meta-learning network with discard-supplement module for few-shot cross-domain rotating machinery fault diagnosis, *Knowl.-Based Syst.* 268 (2023) 110484.

- [55] Janez Demšar, Statistical comparisons of classifiers over multiple data sets, *J. Mach. Learn. Res.* 7 (2006) 1–30.
- [56] Wentao Mao, Yamin Liu, Ling Ding, Ali Safian, Xihui Liang, A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions, *IEEE Trans. Instrum. Meas.* 70 (2020) 1–13.
- [57] Xiaoli Zhao, Jianyong Yao, Wenxiang Deng, Peng Ding, Yifei Ding, Minping Jia, Zheng Liu, Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intraclass and interclass convolutional neural network, *IEEE Trans. Neural Netw. Learn. Syst.* (2022).
- [58] Xiaoxi Ding, Qingbo He, Nianwu Luo, A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification, *J. Sound Vib.* 335 (2015) 367–383.
- [59] Shao Haidong, Jiang Hongkai, Li Xingqiu, Wu Shuaipeng, Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine, *Knowl.-Based Syst.* 140 (2018) 1–14.
- [60] Songyu Han, Haidong Shao, Junsheng Cheng, Xingkai Yang, Baoping Cai, Convformer-NSE: A novel end-to-end gearbox fault diagnosis framework under heavy noise using joint global and local information, *IEEE/ASME Trans. Mechatronics* 28 (1) (2022) 340–349.
- [61] Kaiming He, Jian Sun, Convolutional neural networks at constrained time cost, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5353–5360.
- [62] Tianfu Li, Zhibin Zhao, Chuang Sun, Ruqiang Yan, Xuefeng Chen, Domain adversarial graph convolutional network for fault diagnosis under variable working conditions, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–10.