A Novel Open-Set Domain Generalization Approach via Metalearning-Based Dual-Level Gradient Alignment for Intelligent Fault Diagnosis

Lanjun Wan, Jian Zhou, Le Huang, Jiaen Ning, Hongwei Tan, Wei Ni, and Keqin Li, *Fellow, IEEE*

***Abstract*—The effectiveness of existing domain generalization-based fault diagnosis (DGFD) methods usually relies on the assumption that the label space of the source domain (SD) is consistent with that of the unseen target domain (TD). However, in actual industrial scenarios, the unknown fault classes that do not exist in the SDs may appear in the TD, resulting in the degradation of the diagnosis accuracies of DGFD methods on the unseen TD. Therefore, a novel open-set domain generalization (OSDG) approach via metalearning-based dual-level gradient alignment (MLDGA) for intelligent fault diagnosis (FD) is proposed. First, a metalearning optimization strategy with dual-level gradient alignment is designed to optimize the gradient update directions of the interdomain and interclass tasks simultaneously by gradient matching, so as to ensure that the decision boundaries are located in the optimal positions between each fault class. Second, an entropy-guided dynamic weighting strategy is designed to improve the discrimination ability and accuracy of the model in the multiclass fault classification tasks. Finally, a classification-clustering dual-guided open decision boundary construction strategy is designed to improve the recognition capability of unknown fault classes and the adaptability of the class decision boundaries in fault classification tasks. The experimental results confirm that the proposed approach can effectively identify both known and unknown fault classes.**

***Index Terms*—Domain generalization (DG), fault diagnosis (FD), gradient matching, metalearning, open set.**

Received 19 July 2025; revised 28 September 2025; accepted 10 October 2025. Date of publication 6 November 2025; date of current version 17 November 2025. This work was supported in part by the Scientific Research Foundation of Hunan Provincial Education Department under Grant 24A0391, in part by the Natural Science Foundation of Hunan Province under Grant 2025JJ70030 and Grant 2023JJ30217, in part by the National Natural Science Foundation for Young Scientists of China under Grant 61702177, and in part by the Open Project of Hunan Key Laboratory of Intelligent Information Perception and Processing Technology under Grant 2025KF03. The Associate Editor coordinating the review process was Dr. Benkuan Wang. *(Corresponding author: Lanjun Wan.)*

Lanjun Wan, Hongwei Tan, and Wei Ni are with the School of Computer Science and Artificial Intelligence, Hunan University of Technology, Zhuzhou 412007, China (e-mail: wanlanjun@hut.edu.cn; tanhongwei@stu.hut.edu.cn; niwei@hut.edu.cn).

Jian Zhou and Le Huang are with the School of Computer Science and Artificial Intelligence, Hunan University of Technology, Zhuzhou 412007, China, and also with the School of Computer Science, Guangzhou College of Applied Science and Technology, Guangzhou 526072, China (e-mail: jianzhou@stu.hut.edu.cn; huangle@stu.hut.edu.cn).

Jiaen Ning is with the School of Computer Science and Artificial Intelligence, Hunan University of Technology, Zhuzhou 412007, China, and also with the School of Electronic Information, Central South University, Changsha 410075, China (e-mail: jiaenning@stu.hut.edu.cn).

Keqin Li is with the Department of Computer Science, State University of New York, New Paltz, NY 12561 USA (e-mail: lik@newpaltz.edu).

Digital Object Identifier 10.1109/TIM.2025.3629859

I. Introduction

ROTATING machinery is an important component in modern industrial productions, and its health states directly or indirectly affect the normal operation of mechanical systems, resulting in problems such as safety, production efficiency, and economic losses [1]. Therefore, the effective fault diagnosis (FD) is crucial to ensure the safe operation of mechanical systems and reduce economic losses. The changes of the working conditions (also known as domains) will cause the fault data of different domains to have cross-domain distribution shift, that is, domain shift [2]. Fault classes are usually closely related to working conditions. In actual industrial scenarios, some fault classes that have not been seen in the SDs may appear in the target domain (TD), leading to label shift. The samples corresponding to these unseen fault classes are called the missing unknown fault samples or open-set samples in the SDs. For this FD, it is called open-set FD (OSFD) [3]. The existence of domain and label shifts has brought great challenges to OSFD, thus a more effective intelligent FD strategy is needed.

To solve the label shift problem caused by unknown classes in the TD in FD, researchers have conducted extensive research on OSFD methods. Presently, most of the existing OSFD methods are based on feature extraction using deep learning networks and enhance open-set recognition performance by incorporating various unknown-class discrimination mechanisms (such as the extreme value theory, data augmentation, statistical modeling, and trustworthy learning). For example, Yu et al. [4] proposed an OSFD method by combining deep learning network and extreme value theory, realizing the effective recognition of unknown classes in the TD. Lundgren and Jung [5] developed a data-driven FD framework for quantitative analysis and open-set classification, which uses Kullback–Leibler divergence to model fault data and adopts the open-set classification algorithm to identify unknown classes. Peng et al. [6] designed an OSFD framework based on supervised contrastive learning with negative out-of-distribution data augmentation, effectively improving the performance of open-set classification. Mei et al. [7] studied a conditional variational encoder classifier for extracting features and exploited the empirical threshold and extreme value theory to effectively separate unknown fault classes. Wei et al. [8] put forward a trustworthy deep learning-based OSFD method, which helps the FD model effectively identify unknown fault

classes by introducing a new evidential abstention classifier. Although the aforementioned studies on OSFD provide different solutions to the label shift problem, they failed to fully consider the coupling effect of domain and label shifts. When facing the complex industrial scenarios where both domain and label shifts occur, they still have the problem of significant decline in FD accuracy, which drives researchers to turn to a new way to deal with domain and label shifts cooperatively.

To solve the problem of decline in diagnosis accuracy when OSFD methods are adopted to simultaneously handle domain and label shifts, researchers have turned their attention to open-set domain adaptation (OSDA) methods. Some OSDA methods improve the ability of cross-domain distribution alignment through the adversarial learning to solve the domain shift problem, and adjust or expand the classifier structure to identify unknown classes to solve the label shift problem. For instance, Zhu et al. [9] offered a multiadversarial learning domain adaption model, effectively solving the OSFD problem with incomplete SD diagnosis knowledge by controlling the weights of samples of known and unknown classes during adversarial training. Su et al. [10] designed a TD slanted adversarial network, which uses the TD slanted classifier to build the adaptive threshold for effectively distinguishing known and unknown faults. Zhang et al. [11] designed an intrinsic information-guided open-set domain adaptation network (IODAN), which uses a multi-information integrated weighting module to embed the weights of the target samples into the adversarial loss, so as to accurately identify the unknown fault classes. Wang et al. [12] introduced adversarial domain adaptation with double auxiliary classifiers for cross-domain open-set intelligent FD (ADDOS), which realizes the alignment of the known shared classes by constructing a private class classifier to identify private classes and using the weighted adversarial mechanism. The other OSDA methods rely on self-supervised learning and data generation to identify both known and unknown classes. For example, Wang et al. [13] devised a self-supervised-enabled OSFD approach, which extracts fault features via contrastive learning and uses the squeeze confidence rule to effectively improve the recognition accuracy of the known and unknown classes. Weng et al. [14] designed a progressive domain separation network with multimetric ensemble quantification, effectively realizing the cross-domain OSFD of motor bearings. The existing OSDA methods have made significant progress in unknown class discrimination and effectively addressed the combined impact of domain and label shifts, but they rely on the TD data to participate in the model training. However, it is hard for collecting huge data under various working conditions on the target devices beforehand for model training in actual industrial scenarios. The reliance of OSDA methods on the TD data limits their applicability in real industrial scenarios, because they are unable to address the issue of the TD data being unavailable in practice. To solve this issue, domain generalization (DG) is introduced to realize cross-domain FD without relying on the TD data.

DG aims to apply the general knowledge learned from different SDs into the unseen TD while only using the SD data [15]. For example, Chen et al. [16] studied an adversarial domain-invariant generalization (ADIG) framework for improving the DG ability of the FD model on the unseen TD through the adversarial training. Qian et al. [17] proposed a relationship transfer DG network (RTDGN), which uses the adversarial training and inverse entropy loss to enhance the universality of features in different domains, thus effectively improving the diagnosis accuracy of the FD model. Zhao and Shen [18] put forward a DG network driven by semantic-discriminative augmentation, where the minimization of the triplet loss and semantic regularization are adopted for building decision boundaries, thereby improving the DG ability of the FD model. The existing DG methods solve the issue of unavailable TD data and enhance the generalization ability of cross-domain FD model through mechanisms such as adversarial training and minimization of the triplet loss. However, their research focus is mainly on the closed-set DGFD, and they do not deal with the label shift caused by unknown classes in DGFD, which makes them difficult to address the unseen fault classes in practical applications, thus limiting their effectiveness in open-set scenarios. To cooperatively cope with the three challenges of domain shift, label shift, and the unavailability of TD data, the open-set DG FD (OSDGFD) has gradually attracted attention.

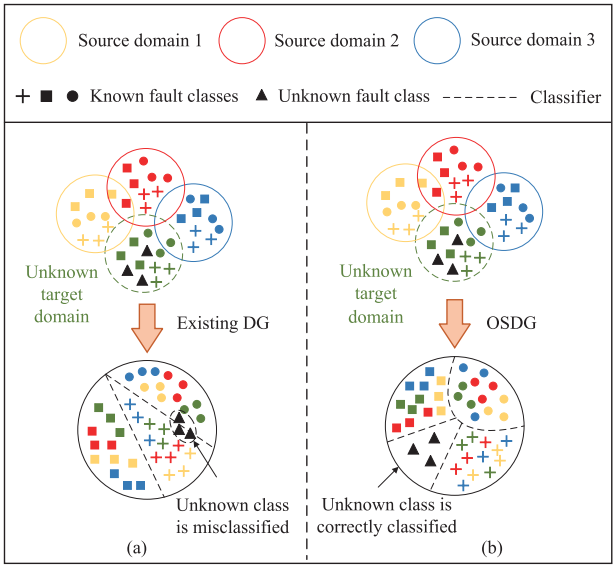


Fig. 1. Illustrations of DG and OSDG when new unknown classes appear in the unseen TD.

The goal of OSDGFD is to address the problems of domain and label shifts in OSFD and accurately identify known and unknown classes without accessing the TD data. Fig. 1 gives the illustrations of DG and OSDG when unknown classes appear in the TD. Currently, research on OSDGFD remains limited and can be broadly categorized into the following directions. First, OSDGFD based on prototype similarity and reconstruction difference. For example, Liu et al. [19] studied an adaptive feature reconstruction difference network, which adaptively constructs decision thresholds based on the feature reconstruction difference values computed from the formed class prototypes, thereby improving the recognition ability of known and unknown fault classes. Zhao and Shen [20] devised

an adaptive OSDG network (AOSDGN), which can effectively detect unknown fault modes under unknown working conditions according to the distances between the prototypes and the samples by constructing local class cluster and outlier detection modules. Second, OSDGFD based on contrastive learning. For example, Lu et al. [21] developed a multidomain contrastive coding framework for learning OSDG representations, which effectively realizes OSDGFD by exploiting both the cross-domain generalize knowledge and domain-unique knowledge. Third, OSDGFD based on data generation and distribution alignment. For instance, Jian et al. [22] designed an OSDG framework consisting of the data generation and feature learning modules, which enhances cross-domain FD performance under open-set and unseen working conditions by generating data and improving the distribution alignment of known classes.

The above studies have made valuable explorations into the application of OSDG in rotating machinery FD, they have solved the three intertwined problems of the heterogeneous label spaces, unseen TD, and difficulty in predicting fault modes in the TD in actual industrial productions to some extent, and are capable of effectively recognizing both known and unknown classes in the TD. However, the existing OSDGFD methods lack the ability to quickly adapt to new domains with different data distributions and stably distinguish known and unknown classes when encountering new FD tasks. Therefore, how to build an intelligent FD model that can accurately identify unknown faults under unseen working conditions without accessing the TD is still worth further exploration. In view of the characteristics of faster adaptability and stronger generalization of metalearning in the face of unknown and distribution-shifted tasks, a novel OSDG approach via metalearning-based dual-level gradient alignment (MLDGA) for intelligent FD is proposed.

The main contributions of this article are as follows.

1) A metalearning optimization strategy with dual-level gradient alignment is designed, where the data partitioning at the domain level and class level and the gradient matching property of metalearning are exploited to achieve interdomain gradient matching and interclass gradient matching, which can effectively alleviate the problems of domain and label shifts in OSDGFD.
2) An entropy-guided dynamic weighting strategy is designed, which can help the FD model more effectively identify different fault classes by dynamically assigning weights to the classification loss.
3) A classification-clustering dual-guided open decision boundary construction strategy is designed, which can accurately identify the potential unknown faults in the TD through the multibinary classifier and increase the compactness of the class clusters by minimizing the triplet loss.
4) Extensive experiments are performed on three bearing and one gearbox datasets, and the results show that the proposed approach has higher diagnosis accuracies compared to the other methods under OSDGFD scenarios.

The rest of this article is organized as follows. The basic theory is introduced in Section II. The proposed approach is described in Section III. The experimental results and analysis are presented in Section IV. The conclusion and future work are given in Section V.

TABLE I
COMPARISON BETWEEN OSDG AND THE OTHER RELATED TASK SETTINGS

Task setting	$y^{\mathrm{Tr}} = y^{\mathrm{Te}}$	$P^{\mathrm{Tr}}_{XY} = P^{\mathrm{Te}}_{XY}$	Need to access the test data in training
Open-set recognition	×	✓	×
Domain adaptation	✓	×	✓
OSDA	×	×	✓
DG	✓	×	×
OSDG	×	×	×

× means that this assumption is not required; ✓ means that this assumption is required; $y^{\mathrm{Tr}} = y^{\mathrm{Te}}$ means that the label space of the training data is the same as that of the test data; $P^{\mathrm{Tr}}_{XY} = P^{\mathrm{Te}}_{XY}$ means that the joint distribution of the training data is the same as that of the test data

II. BASIC THEORY

A. Metalearning

Metalearning, also known as *learning to learn*, is developed by drawing inspiration from the human learning process. Metalearning is a technique that learns the prior knowledge from multiple known tasks and relies on the acquired knowledge to improve the performance of the target task [23]. The goal of metalearning is to find a high-performance universal algorithm based on the ability to "learning to learn" metaknowledge, which enables the model to have faster adaptability and stronger generalization when applied to unknown tasks $d(\mathscr{T})$ with different distributions. The objective of the metalearning is defined as follows:

$$\min_{\delta} L\left(\mathscr{T} \sim d\left(\mathscr{T}\right); \delta\right) \tag{1}$$

where δ denotes the metaknowledge learned across different tasks. It is worth noting that δ is learned across multiple tasks during the metalearning process, and the optimal δ can minimize the loss L of the new tasks as much as possible.

B. Open-Set Recognition

The objective of open-set recognition is to address the problem of unknown-class recognition in the real world. In open-set recognition, the new classes that have not been seen in the training stage may appear during the testing stage, requiring the classifier can not only correctly identify known classes, but also effectively handle unknown classes. In open-set recognition, the SD and TD have the same known classes, while the unknown classes appear in the TD, where the SD and TD are independent and identically distribution. The comparison between OSDG and the other related task settings is shown in Table I. In OSDG, the SD and TD are not only different in the label space, but also different in the data distribution.

III. PROPOSED APPROACH

A. Problem Definition

In OSDGFD scenarios, multiple SDs with the same label space (i.e., fault classes) Y are combined to form an SD set

$\mathcal{M}$, where $\mathcal{M} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M\}$. In $\mathcal{M}$, $\mathcal{D}_m = \{(x_m^i, y_m^i)\}_{i=1}^{n_m}$ represents the mth SD, where $m \in \{1, 2, \ldots, N_s\}$, n_m is the number of samples of the mth SD, x_m^i and $y_m^i \in \mathbb{R}^{N_c}$ denote the ith sample of the mth SD and the corresponding true label, respectively, N_s is the number of SDs, and N_c is the number of fault classes. Similarly, these data with the extended label space $H = Y \cup E$ are combined to form an unseen TD $\mathcal{U} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_u\}$, where E represents the label space of unknown classes in the TD and $Y \cap E = \varnothing$. The goal of the proposed approach is to maximize the utilization of $\mathcal{M}$ to train a model that can generalize to any unseen TD containing unknown classes. To achieve this goal, according to the idea of meta-earning, $\mathcal{M}$ needs to be split into a metatraining set $\mathcal{M}_{\mathcal{V}}$ and a metatesting set $\mathcal{M}_{\mathcal{W}}$ during model training, where $\mathcal{M}_{\mathcal{V}} \cup \mathcal{M}_{\mathcal{W}} = \mathcal{M}$ and $\mathcal{M}_{\mathcal{V}} \cap \mathcal{M}_{\mathcal{W}} = \varnothing$. The proposed approach will use the data sampled from $\mathcal{M}_{\mathcal{V}}$ and $\mathcal{M}_{\mathcal{W}}$ to train the model. In the current training epoch of the model, the metatraining loss function $\mathcal{V}(\rho)$ is used for updating the parameters ρ of the model obtained after the previous training epoch to the parameters $\bar{\rho}$ in the metatraining phase, and the metatesting loss function $\mathcal{W}(\bar{\rho})$ is used to update ρ in the metatesting phase.

B. OSDGFD Framework via MLDGA

To train an intelligent FD model that can accurately identify unknown faults under unknown working conditions under the scenario where the TD cannot be accessed, an OSDGFD framework via MLDGA is constructed, as shown in Fig. 2. The OSDGFD via MLDGA mainly includes the following three steps.

1) *Step 1 (Data acquisition and preprocessing:)* The vibration data of the rotation machinery are collected via acceleration sensors. The collected vibration data are divided into the SDs and TD according to different OSDGFD tasks. To better extract the fault features from the vibration data, the vibration data are converted into 2-D time–frequency (TF) images by using short-time Fourier transform (STFT), which are used as the input of MLDGA.
2) *Step 2 (MLDGA:)* First, the SD data are divided into the metatraining sets and metatesting sets according to different tasks. Second, the metatraining set and metatesting set are used as the inputs to cooperatively and iteratively train the FD model. In each iteration, the metatraining loss is used to update the model parameters, and the metatesting loss is used to optimize the model parameters. Through multiple iterations, an OSDGFD model that can effectively identify unknown fault classes in the TD is trained.
3) *Step 3 (FD:)* First, the unseen TD data are converted into 2-D TF images by STFT. Second, the 2-D TF images are input into the trained FD model. Finally, the values of the positive output channel of the multibinary classifier are used as confidence score to determine the known and unknown fault classes.

C. Metalearning Optimization Strategy With Dual-Level Gradient Alignment

To solve the problem that the FD accuracy of the model is significantly reduced due to the phenomena of domain and label shifts under OSDGFD scenarios, a metalearning optimization strategy with dual-level gradient alignment is designed, as shown in Fig. 3. Specifically, first, to achieve interdomain gradient matching and interclass gradient matching simultaneously, $\mathcal{M}_{\mathcal{V}}$ and $\mathcal{M}_{\mathcal{W}}$ are further divided into a metatraining set $(\mathcal{M}_{\mathcal{V}_1}, \mathcal{M}_{\mathcal{W}_2})$ and a metatesting set $(\mathcal{M}_{\mathcal{V}_2}, \mathcal{M}_{\mathcal{W}_1})$, and the loss functions corresponding to $\mathcal{M}_{\mathcal{V}_1}$, $\mathcal{M}_{\mathcal{V}_2}$, $\mathcal{M}_{\mathcal{W}_1}$, and $\mathcal{M}_{\mathcal{W}_2}$ are defined as $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{W}_1$, and $\mathcal{W}_2$, respectively. Notably, the label space between $\mathcal{M}_{\mathcal{V}_1}$ and $\mathcal{M}_{\mathcal{V}_2}$ and that between $\mathcal{M}_{\mathcal{W}_1}$ and $\mathcal{M}_{\mathcal{W}_2}$ are not intersected, but the label space between $\mathcal{M}_{\mathcal{V}_1}$ and $\mathcal{M}_{\mathcal{W}_1}$ and that between $\mathcal{M}_{\mathcal{V}_2}$ and $\mathcal{M}_{\mathcal{W}_2}$ are the same. Second, the metatraining set $(\mathcal{M}_{\mathcal{V}_1}, \mathcal{M}_{\mathcal{W}_2})$ is input into the model R for training, to obtain the metatraining loss L_{Mtrain} , where R includes a feature extractor $\mathcal{F}$, a fault classifier $\mathcal{C}$, and a multibinary classifier $\|\mathbb{C}\|$ composed of multiple binary classifiers. At this time, the gradient update directions between different tasks are inconsistent. Third, the parameters of R are updated with L_{Mtrain} to obtain the model R'. Fourth, the metatesting set $(\mathcal{M}_{\mathcal{V}_2}, \mathcal{M}_{\mathcal{W}_1})$ is input into R' for training, to obtain the metatesting loss L_{Mval}. Finally, the parameters of the model R are optimized using L_{Mval}, at this point an iteration is completed. Repeating the above steps, after t training epochs, the gradient update directions between different metalearning tasks are basically the same, and the problems of the domain and label shifts can be effectively solved.

The metaobjective function of training the FD model is defined as follows:

$$L_{\text{mobj}} = \min_{\rho} \left(\mathcal{V}_1(\rho) + \mathcal{W}_2(\rho) + \gamma\left(\mathcal{V}_2(\bar{\rho}) + \mathcal{W}_1(\bar{\rho})\right)\right) \tag{2}$$

where γ represents the weight ratio between the metatraining loss and the metatesting loss, ρ denotes the parameters of the model obtained after the previous training epoch, and $\bar{\rho}$ indicates the model parameters updated after metatraining. $\bar{\rho}$ is defined as follows:

$$\bar{\rho} = \rho - \eta\left(\mathcal{V}_1'(\rho) + \mathcal{W}_2'(\rho)\right) \tag{3}$$

where $\mathcal{V}'_1(\rho)$ and $\mathcal{W}'_2(\rho)$ denote the gradients corresponding to $\mathcal{V}_1(\rho)$ and $\mathcal{W}_2(\rho)$ respectively, and η represents the learning rate of the FD model in the metatraining phase. In the metatesting phase, the model parameters $\bar{\rho}$ after metatraining are used to update the model parameters ρ

$$\rho = \rho - \mu\left(\mathcal{V}_1'(\rho) + \mathcal{W}_2'(\rho) + \gamma\left(\mathcal{V}_2'(\bar{\rho}) + \mathcal{W}_1'(\bar{\rho})\right)\right) \tag{4}$$

where μ is the learning rate of the FD model in the metatesting phase.

To further verify that the proposed metalearning optimization strategy with dual-level gradient alignment can achieve interdomain gradient matching and interclass gradient matching simultaneously, the first-order Taylor expansion [24] is performed on $\mathcal{V}_2(\bar{\rho})$ and $\mathcal{W}_1(\bar{\rho})$ in (2)

$$\mathcal{V}_2(\bar{\rho}) = \mathcal{V}_2(\rho) - \eta \cdot \mathcal{V}_2'(\rho) \cdot \left(\mathcal{V}_1'(\rho) + \mathcal{W}_2'(\rho)\right) \tag{5}$$

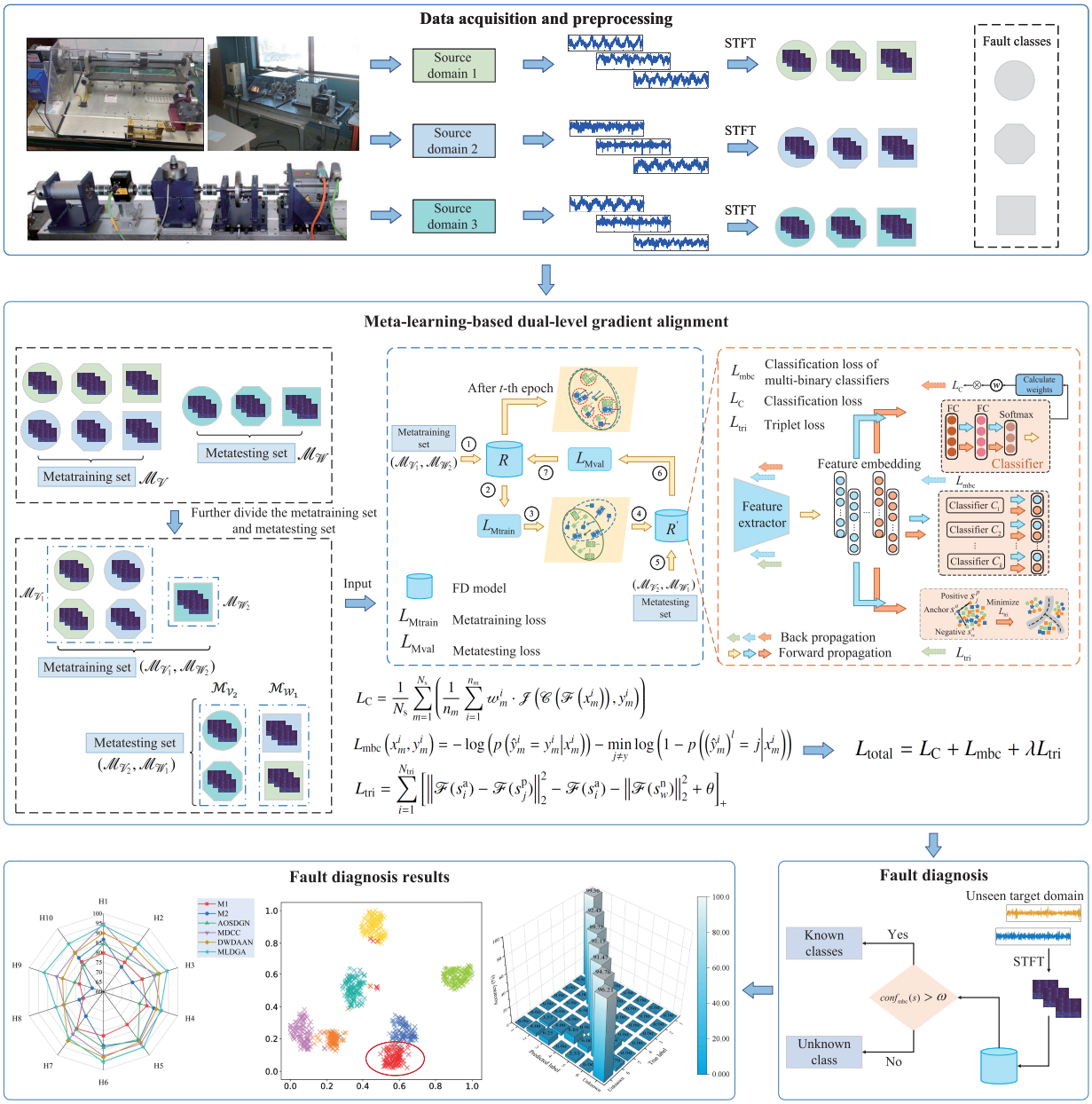


Fig. 2. OSDGFD framework via MLDGA.

and

$$\mathscr{W}_1(\bar{\rho}) = \mathscr{W}_1(\rho) - \eta \cdot \mathscr{W}_1'(\rho) \cdot (\mathscr{V}_1'(\rho) + \mathscr{W}_2'(\rho)) \tag{6}$$

where $\mathscr{V}_2'(\rho)$ and $\mathscr{W}_1'(\rho)$ are the gradients corresponding to $\mathscr{V}_2(\rho)$ and $\mathscr{W}_1(\rho)$, respectively. At this time, the metaobjective function of training the FD model is transformed into

$$\begin{aligned} L_{\text{mobj}} = \min_{\rho} & \left(\mathscr{V}_1(\rho) + \mathscr{W}_2(\rho) + \gamma\left(\mathscr{V}_2(\rho) + \mathscr{W}_1(\rho)\right)\right. \\ & -\eta\gamma\left(\mathscr{V}_1'(\rho) \cdot \mathscr{V}_2'(\rho) + \mathscr{V}_1'(\rho) \cdot \mathscr{W}_1'(\rho)\right. \\ & \left.\left. + \mathscr{W}_2'(\rho) \cdot \mathscr{V}_2'(\rho) + \mathscr{W}_2'(\rho) \cdot \mathscr{W}_1'(\rho)\right)\right). \end{aligned} \tag{7}$$

It can be seen from the first term $\mathscr{V}_1(\rho) + \mathscr{W}_2(\rho) + \gamma(\mathscr{V}_2(\rho) + \mathscr{W}_1(\rho))$ and the second term $\eta\gamma(\mathscr{V}'_1(\rho) \cdot \mathscr{V}'_2(\rho) + \mathscr{V}'_1(\rho) \cdot \mathscr{W}'_1(\rho) + \mathscr{W}'_2(\rho) \cdot \mathscr{V}'_2(\rho) + \mathscr{W}'_2(\rho) \cdot \mathscr{W}'_1(\rho))$ in (7), the optimization objectives are as follows.

1) Minimizing the losses of the FD model on the metatraining set and metatesting set.
2) Maximizing the dot products between the gradients corresponding to the metatraining loss and metatesting loss, so as to find a position in the weight space where the included angle between the gradients corresponding to different losses is small. The small included angle between gradients means that the FD tasks corresponding to the two losses will not conflict with each other. In the second term in (7), the dot products (including $\mathcal{V}'1(\rho) \cdot \mathcal{V}'2(\rho)$, $\mathcal{V}'1(\rho) \cdot \mathcal{W}'1(\rho)$, $\mathcal{W}'2(\rho) \cdot \mathcal{V}'2(\rho)$, $\mathcal{W}'2(\rho) \cdot \mathcal{W}'1(\rho)$) between the gradients corresponding to any two losses that either comes from different domains or contains different fault classes are calculated and summed, which proves that the proposed met-

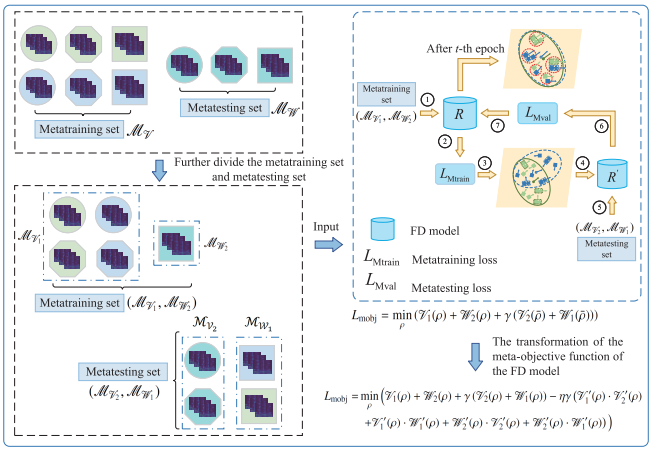


Fig. 3. Illustration of the metalearning optimization strategy with dual-level gradient alignment.

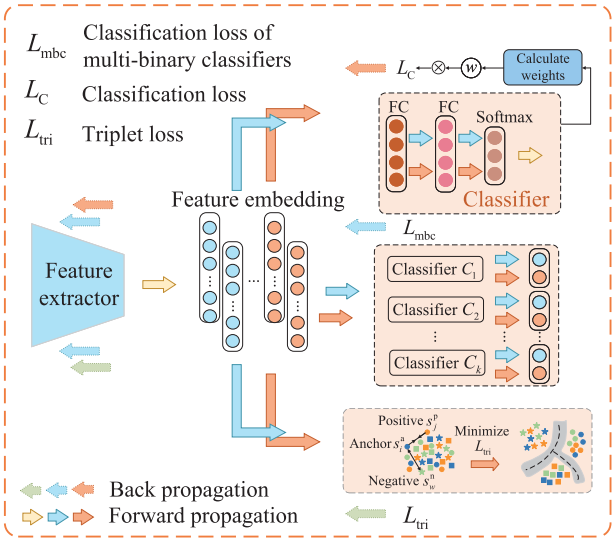


Fig. 4. Illustration of the entropy-guided dynamic weighting strategy.

alearning optimization strategy with dual-level gradient alignment can achieve interdomain gradient matching and interclass gradient matching simultaneously, thereby enabling the FD model to learn the common decision boundaries suitable for different fault classification tasks under open-set recognition scenarios.

D. Entropy-Guided Dynamic Weighting Strategy

In the actual industrial scenarios, there will inevitably be some indistinguishable fault samples [15], which may be located near the decision boundaries and are easy to be predicted as the wrong fault classes. Therefore, an entropy-guided dynamic weighting strategy is designed, as shown in Fig. 4.

The strategy first calculates the entropy according to the confidences of the prediction classes of each sample, and then uses the calculated entropy to dynamically assign the confidence weight for the corresponding sample. These weights will be used to weight the classification loss to help the model effectively distinguish the features of different fault classes, so as to improve the FD ability of the model. Specifically, first, the entropy corresponding to each sample is calculated by using the probability distribution of the prediction classes of each sample through (8), so as to measure the uncertainty of each sample and give a smaller confidence weight to the sample with large entropy (i.e., the sample with low confidence). The entropy corresponding to the ith sample x_m^i of the mth SD is defined as follows:

$$\mathcal{Q}\left(\mathscr{C}\left(\mathscr{F}\left(x_m^i\right)\right)\right) = -\sum_{c=1}^{N_c} p\left(\hat{y}_m^i = c\right) \log\left(p\left(\hat{y}_m^i = c\right)\right) \tag{8}$$

where c denotes the class label and $\hat{y}_m^i$ is the prediction label of x_m^i. Second, according to the entropy corresponding to x_m^i obtained by (8), the confidence weight w_m^i corresponding to x_m^i is calculated by

$$w_m^i = 1 - \frac{\mathcal{Q}\left(\mathscr{C}\left(\mathscr{F}\left(x_m^i\right)\right)\right)}{\log\left(N_c + 1\right)} \tag{9}$$

and compressed to $(0, 1]$, where $\log(N_c + 1)$ denotes the maximum entropy value corresponding to the samples containing N_c fault classes. According to (9), the samples with large entropy values represent the samples that are difficult to be distinguished, and the weights are close to 0. The samples with small entropy values represent the samples that are easy to be identified, and the weights are close to 1. Finally, the calculated w_m^i corresponding to x_m^i is applied to the classification loss

$$L_C = \frac{1}{N_s}\sum_{m=1}^{N_s}\left(\frac{1}{n_m}\sum_{i=1}^{n_m} w_m^i \cdot \mathscr{J}\left(\mathscr{C}\left(\mathscr{F}\left(x_m^i\right)\right), y_m^i\right)\right) \tag{10}$$

where $\mathscr{J}(\cdot)$ denotes the cross-entropy loss function. By using the entropy-guided dynamic weighting strategy, the separation between classes can be promoted effectively and the FD ability of the model can be enhanced.

E. Classification-Clustering Dual-Guided Open Decision Boundary Construction Strategy

In OSDGFD, there may be some unknown fault classes in the TD that have not been seen in the SDs. To effectively identify unknown classes in the TD, a classification-clustering dual-guided open decision boundary construction strategy is proposed. In the OSDGFD framework via MLDGA shown in Fig. 2, this strategy uses the triplet loss in the metric learning method to increase the compactness of the clusters, and accurately identifies the potential unknown faults in the TD through the multibinary classifier. In the FD model, the fault feature representations are first obtained from the SD data through the feature extractor $\mathscr{F}$, and then the fault feature representations are input into the multibinary classifier $\|\mathbb{C}\|$ for classification. $\|\mathbb{C}\|$ contains k binary classifiers, where each binary classifier is trained to detect whether the sample belongs to the corresponding fault class. For the sample (x_m^i, y_m^i), its loss L_{mbc} on multiple binary classifiers is defined as follows:

$$\begin{aligned} L_{\text{mbc}}\left(x_m^i, y_m^i\right) = & -\log\left(p\left(\hat{y}_m^i = y_m^i \middle| x_m^i\right)\right) \\ & -\min_{j\neq y}\log\left(1 - p\left(\left(\hat{y}_m^i\right)^l = j \middle| x_m^i\right)\right) \end{aligned} \tag{11}$$

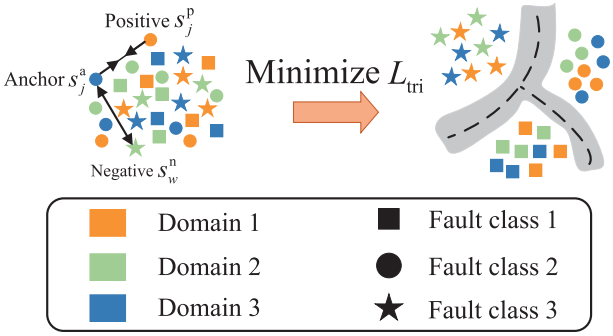


Fig. 5. Illustration of improving the adaptability of the class decision boundaries in fault classification tasks by minimizing the triplet loss.

where $p((\hat{y}_m^i)^l = j|x_m^i)$ represents the output probability of the lth $(1 \le l \le k)$ binary classifier in $\|\mathbb{C}\|$ for x_m^i. During the FD phase, k binary classifiers are used and the values of their positive output channels are selected as the confidence score of x_m^i

$$\text{conf}_{\text{mbc}}\left(x_m^i\right) = p\left(\left(\hat{y}_m^i\right)^{\text{argmax}_{l=1}^{k}\left(p\left((\hat{y}_m^i)^l=c\right)\right)}\middle| x_m^i\right). \tag{12}$$

If the confidence score is greater than the preset threshold value ω, the sample is determined as a known class and a specific class label is generated for it. Otherwise, the sample is determined as an unknown class. The candidate values of ω are traversed in the threshold interval thre_range, and the optimal value of ω is determined by evaluating and screening according to the diagnosis accuracy. thre_range is defined as follows:

$$\text{thre\_range} = \left\{p_{\min} + \frac{p_{\max} - p_{\min}}{h-1} \times i \;\middle|\; i = 0, 1, \ldots, h-1\right\} \tag{13}$$

where $p_{\min}$ and $p_{\max}$ denote the minimum value and the maximum value of the predicted class probabilities of the samples, respectively, and h represents the number of divisions of thre_range.

The proposed MLDGA improves the adaptability of the class decision boundaries in fault classification tasks by minimizing the triplet loss, as depicted in Fig. 5. By minimizing the triplet loss, the distances between samples of different classes can be increased, so as to form a clearer discrimination boundary [25]. In the triplet loss, each triplet includes three kinds of samples: anchor sample s_i^a, positive sample s_j^p, and negative sample s_w^n. The triples are split into the simple, semidifficult, and difficult triplets according to the distances between different kinds of samples, which are defined as follows:

$$\begin{cases} \|\mathscr{F}\left(s_i^a\right) - \mathscr{F}\left(s_j^p\right)\|_2^2 + \theta < \|\mathscr{F}\left(s_i^a\right) - \mathscr{F}\left(s_w^n\right)\|_2^2 \\ \|\mathscr{F}\left(s_i^a\right) - \mathscr{F}\left(s_j^p\right)\|_2^2 < \|\mathscr{F}\left(s_i^a\right) - \mathscr{F}\left(s_w^n\right)\|_2^2 + \theta \\ \|\mathscr{F}\left(s_i^a\right) - \mathscr{F}\left(s_w^n\right)\|_2^2 < \|\mathscr{F}\left(s_i^a\right) - \mathscr{F}\left(s_j^p\right)\|_2^2 \end{cases} \tag{14}$$

TABLE II
WORKING CONDITIONS OF HUST, PU, AND PHM DATASETS

HUST		PU				PHM		
Working condition	Rotating speed (Hz)	Working condition	Rotating speed (rpm)	Load torque (Nm)	Radial force (N)	Working condition	Rotating speed (Hz)	Load
W_1	65	W_5	900	0.7	1000	W_9	30	Low
W_2	70	W_6	1500	0.1	1000	W_{10}	35	Low
W_3	75	W_7	1500	0.7	400	W_{11}	40	Low
W_4	80	W_8	1500	0.7	1000	W_{12}	45	Low

where θ denotes a certain margin to be positive. The optimization objective of the FD model is to form clear decision boundaries in the feature space by reducing the distances between samples of the same classes and increasing the distances between samples of different classes. During model training, it is necessary to optimize the distance between the samples s_i^a and s_j^p of the same classes to be closer than the distance between the samples s_i^a and s_w^n of different classes. The triplet loss L_{tri} is defined as follows:

$$L_{\text{tri}} = \sum_{i=1}^{N_{\text{tri}}}\left[\|\mathscr{F}\left(s_i^a\right) - \mathscr{F}\left(s_j^p\right)\|_2^2 - \mathscr{F}\left(s_i^a\right) - \|\mathscr{F}\left(s_w^n\right)\|_2^2 + \theta\right]_+ \tag{15}$$

where the training set contains N_{tri} triples and $[\cdot]_+ = \max(0, \cdot)$.

F. Parameter Updating

The total objective loss function of the proposed OSDGFD model via MLDGA in the training process is

$$L_{\text{total}} = L_C + L_{\text{mbc}} + \lambda L_{\text{tri}} \tag{16}$$

where λ is a trade-off parameter. The training process of the proposed OSDGFD model via MLDGA is described in Algorithm 1.

IV. EXPERIMENTS

A. Experimental Setup

1) Description of Experimental Datasets: To verify the effectiveness of the proposed MLDGA in OSFD, extensive experiments are performed on the Huazhong University of Science and Technology (HUST) dataset [26], Paderborn University (PU) dataset [27], and Prognostics and Health Management (PHM) 2009 dataset [28]. The vibration data of HUST, PU, and PHM datasets are collected from the test rigs shown in Fig. 6(a)–(c), and their sampling frequencies are 25.6, 64, and 66.67 kHz, respectively. Table II gives the working condition information of HUST, PU, and PHM datasets. Tables III and IV show the information of fault classes of HUST, PU, and PHM datasets, respectively. In this experiment, the vibration data of different fault classes collected under each working condition shown in Table II are selected from HUST, PU, and PHM datasets, and split into nonoverlapping samples with a length of 2048, which are converted into 2-D TF images by STFT.

Algorithm 1 Training Process of the Proposed OSDGFD Model via MLDGA

Require: The SDs $\mathscr{M}$, the fault classes Y, the learning rate η and μ, the weight ratio γ, the trade-off parameter λ, and the maximum number of epochs n_t.

Randomly initialize the training parameters ρ of the model;
for $t = 1$**to**n_t **do**
  Randomly split $\mathscr{M}$ and Y into $(\mathscr{M}_{\mathscr{Y}}, \mathscr{M}_{\mathscr{W}})$ and (Y, Y), respectively;
  Divide $\mathscr{M}_{\mathscr{Y}}$ and $\mathscr{M}_{\mathscr{W}}$ into the metatraining set $(\mathscr{M}_{\mathscr{Y}_1}, \mathscr{M}_{\mathscr{W}_2})$ and the metatesting set $(\mathscr{M}_{\mathscr{Y}_2}, \mathscr{M}_{\mathscr{W}_1})$, respectively;
  Metatraining phase:
  Randomly choose the samples $\mathscr{S}_{\mathscr{Y}_1}$ and $\mathscr{S}_{\mathscr{W}_2}$ from $(\mathscr{M}_{\mathscr{Y}_1}, Y_1)$ and $(\mathscr{M}_{\mathscr{W}_2}, Y_2)$, respectively;
  Calculate the gradients $\mathscr{V}'_1(\rho) + \mathscr{W}'_2(\rho)$ with $\mathscr{S}_{\mathscr{Y}_1}$ and $\mathscr{S}_{\mathscr{W}_2}$;
  Calculate the losses L_C, L_{mbc}, and L_{tri} by Eqs. (10), (11), and (15) with $\mathscr{S}_{\mathscr{Y}_1}$ and $\mathscr{S}_{\mathscr{W}_2}$, respectively;
  Calculate the total metatraining loss: $L_{Mtrain} = L_C + L_{mbc} + \lambda L_{tri}$;
  Update the parameters: $\bar{\rho} \leftarrow \rho - \eta(\mathscr{V}'_1(\rho) + \mathscr{W}'_2(\rho))$;
  Metatesting phase:
  Randomly choose the samples $\mathscr{S}_{\mathscr{Y}_2}$ and $\mathscr{S}_{\mathscr{W}_1}$ from $(\mathscr{M}_{\mathscr{Y}_2}, Y_1)$ and $(\mathscr{M}_{\mathscr{W}_1}, Y_2)$, respectively;
  Calculate the gradients $\mathscr{V}'_2(\bar{\rho}) + \mathscr{W}'_1(\bar{\rho})$ with $\mathscr{S}_{\mathscr{Y}_2}$ and $\mathscr{S}_{\mathscr{W}_1}$;
  Calculate the losses L_C, L_{mbc}, and L_{tri} by Eqs. (10), (11), and (15) with $\mathscr{S}_{\mathscr{Y}_2}$ and $\mathscr{S}_{\mathscr{W}_1}$, respectively;
  Calculate the total metatesting loss: $L_{Mval} = L_C + L_{mbc} + \lambda L_{tri}$;
  Calculate the total loss: $L_{total} = L_{Mtrain} + L_{Mval}$;
  Update the parameters: $\rho \leftarrow \rho - \mu(\mathscr{V}'_1(\rho) + \mathscr{W}'_2(\rho) + \gamma(\mathscr{V}'_2(\bar{\rho}) + \mathscr{W}'_1(\bar{\rho})))$;
end for
Return the trained model

2) Evaluation Metrics: In this experiment, the three evaluation metrics, namely OS*, UK, and H-score, are used as the evaluation criteria for the performance of different FD methods. OS* = E_k/N_k denotes the prediction accuracy of known classes, where E_k is the number of test samples of known classes correctly predicted and N_k is the number of test samples of all known classes. UK = E_u/N_u represents the prediction accuracy of unknown classes, where E_u is the number of test samples of unknown classes correctly predicted and N_u is the number of test samples of all unknown classes. H-score = $2 \cdot \text{OS}^* \cdot \text{UK}/(\text{OS}^* + \text{UK})$ indicates the harmonic mean of OS* and UK. When the prediction accuracies of both the known and unknown classes are high, H-score is larger.

3) Comparison Methods: To better evaluate the effectiveness of the proposed MLDGA, the comparative experiments between the MLDGA and the following five different OSDG methods and three different OSDA methods are conducted on different transfer tasks listed in Table V: M1, M2, AOSDGN [20], MDCC [21], DWDAAN [22], OSBP [29], IODAN [11], and ADDOS [12]. The key difference between the OSDG and OSDA methods is that, during model training, the former cannot access the TD data, whereas the latter can. In Table V, the experimental settings of the transfer tasks are mainly designed according to the distribution differences between the SDs and TD and the core index (i.e., openness) for measuring the FD ability of the model under the open-set scenarios, where openness = $1 - (Y/H)$ and the definitions of Y and H are given in Section III-A. Here, the openness is set between 0.14 and 0.43 to avoid that too low openness degenerates into the closed-set DG problem or too high openness causes the problem of ignoring diagnosis of known classes. The relevant experimental settings of eight different comparison methods are described as follows.

1) *M1:* It adopts ADIG [16] as the DG method and uses OpenMax [30] for open-set recognition. Specifically, ADIG is a DGFD method, which uses the adversarial training between the feature extractor and the domain classifier to obtain the domain invariant features related to the faults, thereby improving the generalization of the FD model. OpenMax uses the extreme value theory to calculate the classification probability. The classes whose classification probabilities are lower than the preset threshold are regarded as unknown classes. The main optimization parameters of ADIG are given in [16].
2) *M2:* It adopts RTDGN [17] as the DG method and uses OpenMax for open-set recognition. Specifically, RTDGN is a DGFD method, which reduces the

TABLE III
INFORMATION OF FAULT CLASSES OF HUST AND PU DATASETS

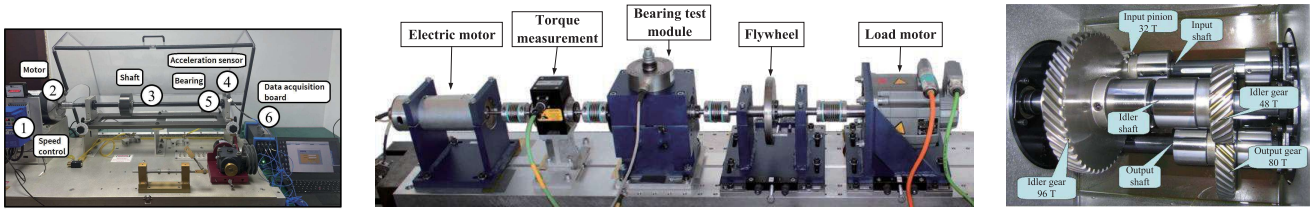
HUST			PU		
Health state	Fault degree	Class label	Health state	Bearing code	Class label
HC	-	1	HC	K004	1
IF1	Medium	2	OF1	KA04	2
IF2	Severe	3	OF2	KA16	3
OF1	Medium	4	IF1	KI21	4
OF2	Severe	5	IF2	KI18	5
BF1	Medium	6	IF+OF1	KB27	6
BF2	Severe	7	IF+OF2	KB23	7

HC = Healthy condition; IF = Inner-race fault; OF = Outer-race fault; BF = Ball fault

TABLE IV
INFORMATION OF FAULT CLASSES OF PHM DATASET

Class label	Gear				Bearing						Shaft	
	32T	96T	48T	80T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output
1	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC	HC
2	CH	HC	EC	HC	HC	HC	HC	HC	HC	HC	HC	HC
3	HC	HC	EC	HC	HC	HC	HC	HC	HC	HC	HC	HC
4	HC	HC	EC	BR	BF	HC	HC	HC	HC	HC	HC	HC
5	CH	HC	EC	BR	IF	BF	OF	HC	HC	HC	HC	HC
6	HC	HC	HC	BR	IF	BF	OF	HC	HC	HC	IM	HC
7	HC	HC	HC	HC	IF	HC	HC	HC	HC	HC	HC	KS

KS = Keyway sheared; BR = Broken; EC = Eccentric; CH = Chipped; IM = Imbalance; :IS = Input side; IS = Input shaft; :OS = Output side; OS = Output shaft; ID = Idler shaft



(a) (b) (c)

Fig. 6. Test rigs for HUST, PU, and PHM. (a) HUST test rig [26]. (b) PU test rig [27]. (c) PHM test rig [28].

TABLE V
TRANSFER TASK INFORMATION ON HUST, PU, AND PHM DATASETS

Dataset	Task	SDs → Unknown TD	Source classes	Target classes	Openness
HUST	H1	$[W_2, W_3, W_4] \rightarrow [W_1]$	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6, **7**	0.14
	H2	$[W_2, W_3, W_4] \rightarrow [W_1]$	1, 2, 3, 4, 5, 7	1, 2, 3, 4, 5, **6**, 7	0.14
	H3	$[W_2, W_3, W_4] \rightarrow [W_1]$	1, 2, 5, 6, 7	1, 2, **3**, **4**, 5, 6, 7	0.29
	H4	$[W_2, W_3, W_4] \rightarrow [W_1]$	1, 2, 3, 6, 7	1, 2, 3, **4**, **5**, 6, 7	0.29
	H5	$[W_1, W_3, W_4] \rightarrow [W_2]$	1, 2, 3, 4, 5, 7	1, 2, 3, 4, 5, **6**, 7	0.14
	H6	$[W_1, W_3, W_4] \rightarrow [W_2]$	1, 4, 5, 6, 7	1, **2**, **3**, 4, 5, 6, 7	0.29
	H7	$[W_1, W_2, W_4] \rightarrow [W_3]$	1, 2, 5, 6, 7	1, 2, **3**, **4**, 5, 6, 7	0.29
	H8	$[W_1, W_2, W_4] \rightarrow [W_3]$	1, 2, 3, 7	1, 2, 3, **4**, **5**, **6**, 7	0.43
	H9	$[W_1, W_2, W_3] \rightarrow [W_4]$	1, 5, 6, 7	1, **2**, **3**, **4**, 5, 6, 7	0.43
	H10	$[W_1, W_2, W_3] \rightarrow [W_4]$	1, 2, 3, 4	1, 2, 3, 4, **5**, **6**, 7	0.43
HUST	P1	$[W_6, W_7, W_8] \rightarrow [W_5]$	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6, **7**	0.14
	P2	$[W_5, W_7, W_8] \rightarrow [W_6]$	1, 2, 3, 6, 7	1, 2, 3, **4**, **5**, 6, 7	0.29
	P3	$[W_5, W_6, W_8] \rightarrow [W_7]$	1, 4, 5, 6, 7	1, **2**, **3**, 4, 5, 6, 7	0.29
	P4	$[W_5, W_6, W_7] \rightarrow [W_8]$	1, 2, 3, 5, 6, 7	1, 2, 3, **4**, 5, 6, 7	0.14
PHM	G1	$[W_{10}, W_{11}, W_{12}] \rightarrow [W_9]$	1, 3, 4, 5, 6	1, **2**, 3, 4, 5, 6, **7**	0.29
	G2	$[W_9, W_{11}, W_{12}] \rightarrow [W_{10}]$	1, 2, 3, 4, 7	1, 2, 3, 4, **5**, **6**, 7	0.29
	G3	$[W_9, W_{10}, W_{12}] \rightarrow [W_{11}]$	1, 2, 3, 4, 5	1, 2, 3, 4, 5, **6**, **7**	0.29
	G4	$[W_9, W_{10}, W_{11}] \rightarrow [W_{12}]$	1, 4, 5, 6, 7	1, **2**, **3**, 4, 5, 6, 7	0.29

distribution discrepancies between the SDs and TD via the adversarial training between the feature extractor and multiple domain classifiers, thereby improving the generalization performance of the model on the TD. The main optimization parameters of RTDGN are given in [17].

3) *AOSDGN:* It is a classic OSDGFD method. It improves the compactness of clusters by minimizing the triplet loss, thereby learning the domain-invariant representations. Meanwhile, the unknown classes are identified according to the distances between the samples and the constructed class prototypes. In AOSDGN, the batch size and maximum training epochs are set to 100 and 1024, respectively, and the hyperparameters μ, β, and δ are set to 1, 1, and 1.2, respectively.
4) *MDCC:* It is an advanced OSDGFD method, which achieves multidomain contrastive coding by introducing a new contrastive encoding task and loss, thereby reducing interdomain and intraclass discrepancies and facilitating the separation of private classes. In MDCC, the number of iterations and the mini-batch in the initialization phase are set to 1000 and 128, respectively. After the initialization phase, the learning rate and the maximum number of iterations are adjusted to 0.001 and 4000, respectively, and the hyperparameter λ is set to 0.45.
5) *DWDAAN:* It is an advanced OSDGFD method, which uses ACGAN to generate open-set data for simulating unknown faults, thereby alleviating the phenomena of label shift. Moreover, it adopts a dual-level weighted mechanism to reduce distribution discrepancies. The main optimization parameters of DWDAAN are given in [22].
6) *OSBP:* It is a classic OSDA method based on adversarial training, which is mainly used to solve the domain adaptation problem that the TD class set contains the SD class set. The main optimization parameters of OSBP are provided in [29].
7) *IODAN:* It is an advanced OSDA-based FD (OSDAFD) method, mainly solving the domain adaptation problem that the TD class set contains the SD class set. The method realizes the diagnosis of unknown and known fault classes through the weighted domain adversarial training between the feature extractor and the classifier. In IODAN, the batch size, maximum training epochs, and learning rate are set to 64, 150, and 0.001, respectively.
8) *ADDOS:* It is an advanced OSDAFD method. In ADDOS, the weighted domain adversarial training is carried out between the feature extractor and the private class classifier, and the dual auxiliary classifier module is constructed to realize the simultaneous recognition of unknown and known fault classes. The main optimization parameters of ADDOS are provided in [12].

To ensure fair comparison, the same data preprocessing is performed for all comparison methods. The model training and FD are conducted on NVIDIA RTX 2070 Super GPU. These comparison methods adopt the same backbone network structures as MLDGA. Table VI shows the network structure of the proposed MLDGA. Fig. 7 shows the model structure and parameters of MLDGA.

4) Setting of Hyperparameters: The proposed MLDGA contains eight hyperparameters, and the specific parameter values are set as follows. The metatraining learning rate η, metatesting learning rate μ, learning rate ε, weight ratio γ between the metatraining and metatesting losses, batch size, and maximum training epochs are set to 0.0001, 0.0001, 0.01, 1.0, 32, and 300, respectively. For the hyperparameter θ used in calculating the triplet loss, refer to [25], the value of θ is searched from {0.1, 0.5, 1.0, 1.5, 2.0, 2.5} according to the grid search method. As shown in Fig. 8, when θ = 2.0, MLDGA obtains the best average H-score on different transfer tasks. Therefore, θ is set to 2.0. For the hyperparameter λ used in

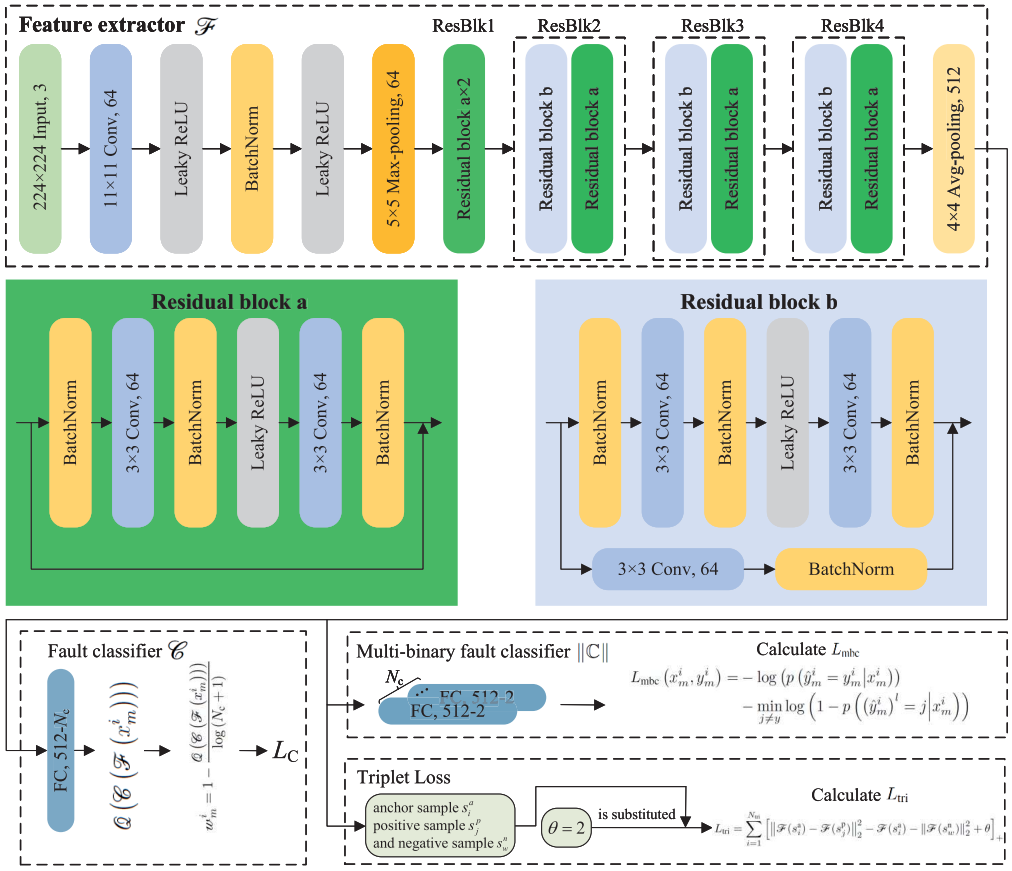


Fig. 7. Model structure and parameters of the proposed MLDGA.

TABLE VI
NETWORK STRUCTURE OF THE PROPOSED MLDGA

Component name	Layer name	Output size	Channels × Kernel size
Feature extractor	Input	$3 \times 224 \times 224$	–
	Conv + LR	$64 \times 112 \times 112$	$64 \times 11 \times 11$
	BN + LR	$64 \times 112 \times 112$	–
	Max-pooling	$64 \times 56 \times 56$	$64 \times 5 \times 5$
	ResBlk1 + LR	$64 \times 56 \times 56$	$\begin{bmatrix} 64 \times 3 \times 3 \\ 64 \times 3 \times 3 \end{bmatrix} \times 2$
	ResBlk2 + LR	$128 \times 28 \times 28$	$\begin{bmatrix} 128 \times 3 \times 3 \\ 128 \times 3 \times 3 \end{bmatrix} \times 2$
	ResBlk3 + LR	$256 \times 14 \times 14$	$\begin{bmatrix} 256 \times 3 \times 3 \\ 256 \times 3 \times 3 \end{bmatrix} \times 2$
	ResBlk4 + LR	$512 \times 7 \times 7$	$\begin{bmatrix} 512 \times 3 \times 3 \\ 512 \times 3 \times 3 \end{bmatrix} \times 2$
	Avg-pooling	$512 \times 1 \times 1$	$512 \times 4 \times 4$
Fault classifier	FC	N_c	–
Multi-binary fault classifier	FC	$N_c \times 2$	–

Conv = Convolution; LR = Leaky ReLU; BN = Batch normalization; ResBlk = Residual block; FC = Fully connection

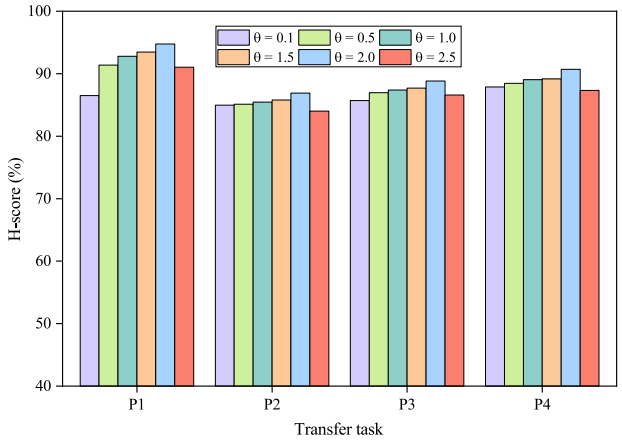


Fig. 8. H-scores obtained on different transfer tasks of PU dataset with different values of θ.

calculating the total objective loss function, the value of λ is searched from {0.1, 0.5, 0.8, 1.0, 2.0} according to the grid search method. As shown in Fig. 9, when $\lambda = 0.8$, MLDGA achieves the best average H-score on different transfer tasks. Therefore, λ is set to 0.8.

B. Comparison With Different OSDGFD Methods

The H-scores obtained by the proposed MLDGA and five different OSDGFD methods on HUST, PU, and PHM datasets are shown in Tables VII –IX, respectively. On HUST dataset, the average H-score of MLDGA is 17.34%, 14.97%, 9.14%, 6.22%, and 4.51% higher than those of M1, M2, AOSDGN, MDCC, and DWDAAN, respectively, indicating that MLDGA has high diagnosis accuracies on both known and unknown

TABLE VII
H-SCORES (%) OF DIFFERENT OSDGFD METHODS ON HUST DATASET

Method	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	Avg.
M1	77.67 ± 0.2	72.05 ± 0.7	75.40 ± 1.4	80.57 ± 0.6	80.13 ± 1.2	75.90 ± 0.5	74.35 ± 2.2	71.02 ± 0.4	68.16 ± 1.3	64.91 ± 0.6	74.02
M2	81.30 ± 1.1	73.22 ± 0.7	74.62 ± 0.8	81.56 ± 1.1	85.06 ± 0.2	80.40 ± 1.5	75.14 ± 1.2	70.31 ± 1.5	69.54 ± 0.9	72.75 ± 1.5	76.39
AOSDGN	87.46 ± 0.6	82.27 ± 1.3	85.09 ± 0.9	82.63 ± 1.2	83.13 ± 1.7	91.07 ± 0.5	79.66 ± 0.6	78.32 ± 0.8	74.49 ± 1.8	78.07 ± 1.2	82.22
MDCC	88.58 ± 2.2	86.43 ± 1.5	84.40 ± 1.1	86.02 ± 1.8	89.22 ± 1.0	91.79 ± 2.5	83.89 ± 0.8	80.50 ± 1.3	79.61 ± 1.3	80.95 ± 0.9	85.14
DWDAAN	91.79 ± 0.3	89.84 ± 0.8	88.76 ± 0.6	85.79 ± 0.5	86.27 ± 1.5	92.46 ± 0.2	87.53 ± 0.9	83.15 ± 0.6	80.07 ± 1.7	82.83 ± 0.6	86.85
MLDGA	**95.83 ± 0.5**	**92.06 ± 1.2**	**93.60 ± 0.8**	**91.59 ± 0.3**	**91.40 ± 0.9**	**96.22 ± 0.7**	**91.30 ± 1.4**	**88.02 ± 0.4**	**86.46 ± 2.4**	**87.12 ± 0.4**	**91.36**

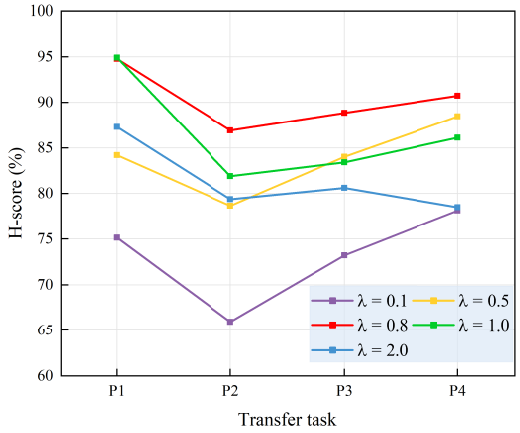


Fig. 9. H-scores obtained on different transfer tasks of PU dataset with different values of λ.

TABLE VIII
H-SCORES (%) OF DIFFERENT OSDGFD METHODS ON PU DATASET

Method	P1	P2	P3	P4	Avg.
M1	71.62 ± 1.9	68.69 ± 1.2	72.66 ± 0.7	72.06 ± 0.3	71.26
M2	74.06 ± 0.4	69.17 ± 2.8	76.18 ± 2.8	74.85 ± 2.0	73.57
AOSDGN	79.87 ± 0.8	76.85 ± 1.0	80.08 ± 1.2	76.68 ± 0.6	78.37
MDCC	83.03 ± 3.9	82.47 ± 0.8	84.75 ± 0.6	84.47 ± 1.9	83.68
DWDAAN	88.45 ± 0.1	78.21 ± 0.6	85.44 ± 0.3	85.88 ± 0.2	84.50
MLDGA	**94.76 ± 0.3**	**86.91 ± 1.0**	**88.84 ± 0.7**	**90.71 ± 0.5**	**90.31**

TABLE IX
H-SCORES (%) OF DIFFERENT OSDGFD METHODS ON PHM DATASET

Method	G1	G2	G3	G4	Avg.
M1	77.13 ± 0.8	71.57 ± 1.4	78.89 ± 0.8	70.91 ± 0.4	74.63
M2	78.52 ± 0.9	73.36 ± 1.2	79.40 ± 1.3	70.10 ± 1.0	75.85
AOSDGN	82.43 ± 1.5	80.58 ± 1.4	84.32 ± 0.9	72.70 ± 1.2	80.01
MDCC	88.02 ± 1.0	86.51 ± 1.3	88.76 ± 0.7	77.98 ± 0.9	85.32
DWDAAN	88.59 ± 0.1	85.25 ± 0.7	87.60 ± 0.3	79.20 ± 0.4	85.16
MLDGA	**90.25 ± 0.7**	**89.08 ± 0.9**	**92.72 ± 1.0**	**85.12 ± 0.6**	**89.29**

classes when facing different OSDG tasks. On the more complex PU dataset, the average H-score of MLDGA is 19.05%, 16.74%, 11.94%, 6.63%, and 5.81% higher than those of M1, M2, AOSDGN, MDCC, and OSDG-DGM-FLM, respectively, showing that the FD accuracy of MLDGA is still better than those of these comparison methods on the more complex bearing fault datasets. The main reason is that MLDGA can maintain high discrimination ability in the face of class confusion caused by combined faults through the gradient matching between different tasks, the proposed entropy-guided dynamic weighting strategy, and the

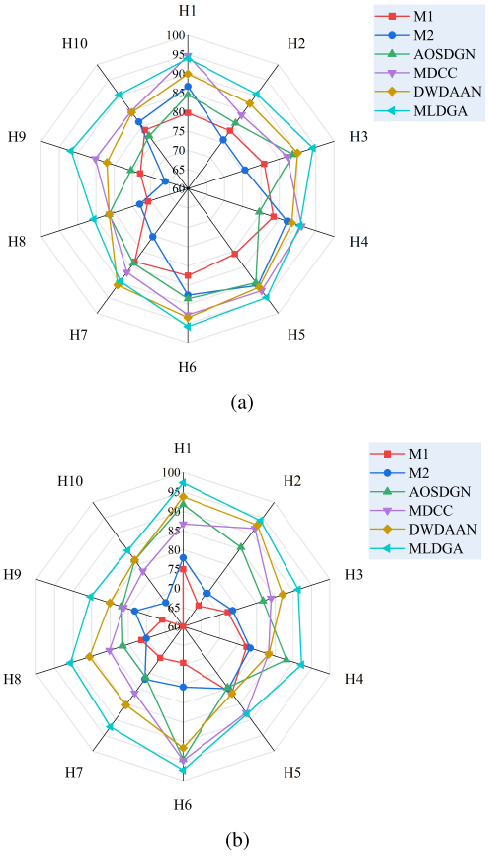
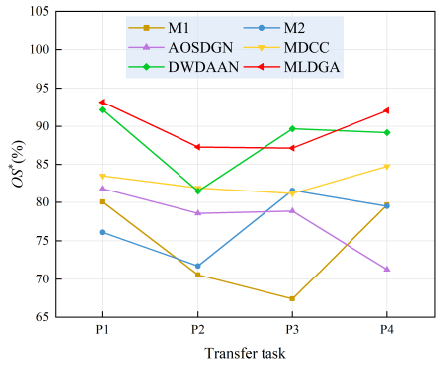


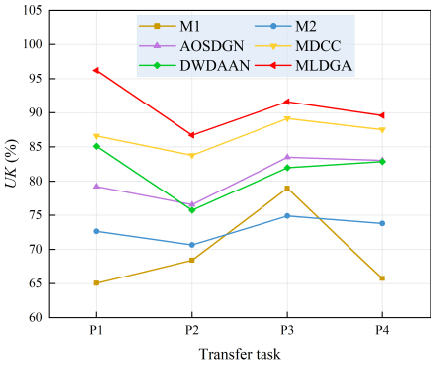
Fig. 10. OS* and UK achieved by different OSDGFD methods on the TD of HUST dataset. (a) OS*. (b) UK.

classification-clustering dual-guided open decision boundary construction strategy. On PHM dataset, the average H-score of MLDGA is 14.66%, 13.44%, 9.28%, 3.97%, and 4.13% higher than those of M1, M2, AOSDGN, MDCC, and DWDAAN, respectively, indicating that MLDGA still has superior diagnosis performance in OSDGFD of the gearbox. There are the compound faults in both PU and PHM datasets, resulting in the obvious distribution discrepancies between the SDs and TD. However, it can be seen from Tables VIII and IX that the average H-score of MLDGA is higher than those of the other comparison methods on eight different transfer tasks, showing that MLDGA can effectively deal with complex OSDGFD scenarios and has strong robustness and FD ability.

The OS* and UK obtained by MLDGA and five different OSDGFD methods on HUST, PU, and PHM datasets are shown in Figs. 10–12, respectively. It can be seen that the

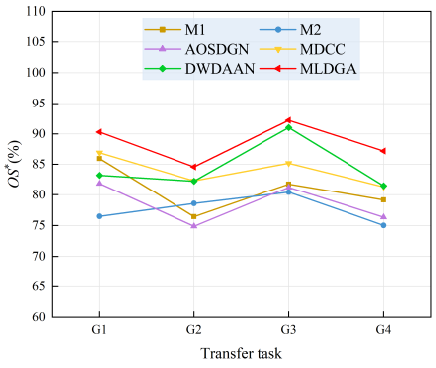


(a)

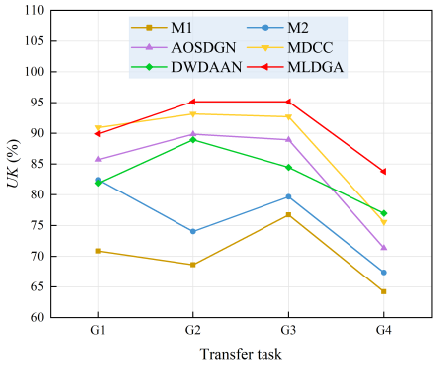


(b)

Fig. 11. OS* and UK achieved by different OSDGFD methods on the TD of PU dataset. (a) OS*. (b) UK.



(a)



(b)

Fig. 12. OS* and UK achieved by different OSDGFD methods on the TD of PHM dataset. (a) OS*. (b) UK.

OS* and UK obtained by MLDGA on different transfer tasks are better than those obtained by the other methods on the whole. For example, as shown in Fig. 10(a) and (b), on the transfer tasks H3 and H9, the OS* and UK of MLDGA are 94.25% and 91.33%, and 95.79% and 97.32%, respectively, which are significantly better than the other methods. This is because MLDGA forms the decision boundaries in the middle region between the class clusters in the decision space through dual-level gradient alignment of the interdomain and interclass, and the unknown samples are more likely to appear near the decision boundaries, thereby improving the probability of the known and unknown classes to be correctly identified. As shown in Fig. 11(a) and (b), on the transfer task P3, the OS* of MLDGA is 87.14%, which is 2.54% lower than that of DWDAAN. However, the UK of MLDGA is 91.63%, which is 9.64% higher than that of DWDAAN. This shows that MLDGA can more effectively distinguish the known and unknown fault classes at the same time.

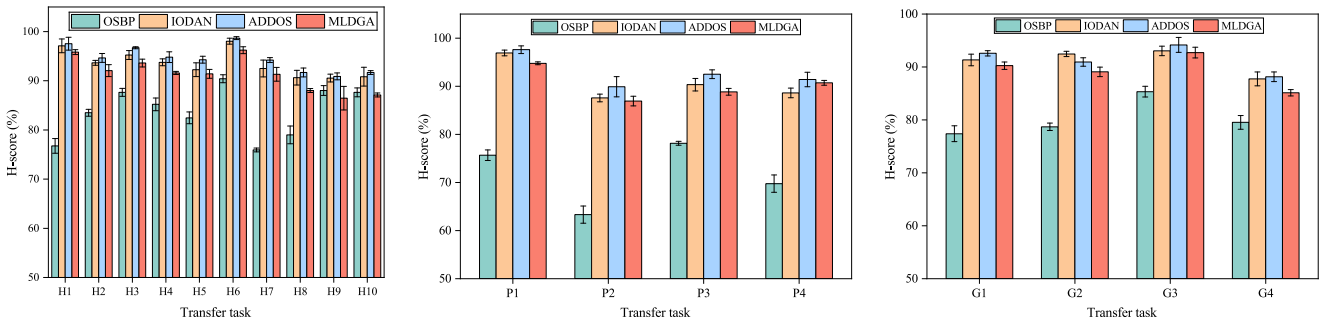
C. Comparison With Different OSDAFD Methods

The H-scores obtained by the proposed MLDGA and three different OSDAFD methods on the HUST, PU, and PHM datasets are depicted in Fig. 13(a)–(c), respectively. The H-scores of MLDGA are higher than those of OSBP on HUST, PU, and PHM datasets, respectively. For example, on the task H1, the H-score of MLDGA is 95.83%, which is 19.08% higher than that of OSBP, showing that MLDGA can still effectively identify known and unknown classes without accessing the TD data. Compared with IODAN and ADDOS, the average H-score of MLDGA on HUST, PU, and PHM datasets is slightly lower. For instance, on the task G3, the H-score of MLDGA is 92.72%, which is 0.32% and 1.45% lower than that of IODAN and ADDOS, respectively. This is mainly because IODAN and ADDOS let the TD data participate in the model training, while MLDGA does not. Although MLDGA does not access the TD data during model training, the diagnosis performance gap between MLDGA and the two advanced OSDAFD methods, IODAN and ADDOS, is small, showing that MLDGA has strong cross-domain transfer ability. MLDGA introduces a dual-level gradient alignment strategy in the gradient updating, so that the model can not only learn the SD features, but also reduce the impact of the interdomain and interclass discrepancies, so as to maintain better classification performance under the unseen TD.

D. Cross-Machine OSDGFD on PU and HUST Datasets

To further verify the OSDGFD capability of MLDGA under cross-machine scenarios, the comparative experiments are carried out under the two different cross-machine scenarios listed in Table X.

Fig. 14 indicates the H-scores of different OSDGFD methods under different cross-machine scenarios. As can be seen from Fig. 14, since the OSDGFD tasks under the cross-machine scenarios are more complex than those under the cross-working condition scenarios, the H-scores obtained by all methods under the cross-machine scenarios are significantly



(a) (b) (c)

Fig. 13. H-scores of the proposed MLDGA and three different OSDAFD methods on HUST, PU, and PHM datasets. (a) On HUST dataset. (b) On PU dataset. (c) On PHM dataset.

TABLE X
CROSS-MACHINE FD TASK INFORMATION

Cross-machine scenario	Task	SDs → Unknown TD	Source classes	Target classes	Openness
HUST → PU	CH1	$[W_1, W_2, W_3] \rightarrow [W_5]$	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6, **7**	0.14
	CH2	$[W_1, W_2, W_3] \rightarrow [W_6]$	1, 3, 4, 7	1, **2**, 3, 4, **5**, **6**, 7	0.43
	CH3	$[W_2, W_3, W_4] \rightarrow [W_7]$	1, 2, 5, 6, 7	1, 2, **3**, **4**, 5, 6, 7	0.29
	CH4	$[W_2, W_3, W_4] \rightarrow [W_8]$	2, 4, 5, 6, 7	**1**, 2, **3**, 4, 5, 6, 7	0.29
PU → HUST	CP1	$[W_5, W_6, W_7] \rightarrow [W_1]$	1, 2, 3, 4, 5	1, 2, 3, 4, 5, **6**, **7**	0.29
	CP2	$[W_5, W_6, W_7] \rightarrow [W_2]$	2, 4, 5	**1**, 2, **3**, 4, 5, **6**	0.50
	CP3	$[W_6, W_7, W_8] \rightarrow [W_3]$	1, 3, 5, 7	1, **2**, 3, **4**, 5, **6**, 7	0.43
	CP4	$[W_6, W_7, W_8] \rightarrow [W_4]$	3, 4, 5, 6	**1**, **2**, 3, 4, 5, 6	0.33

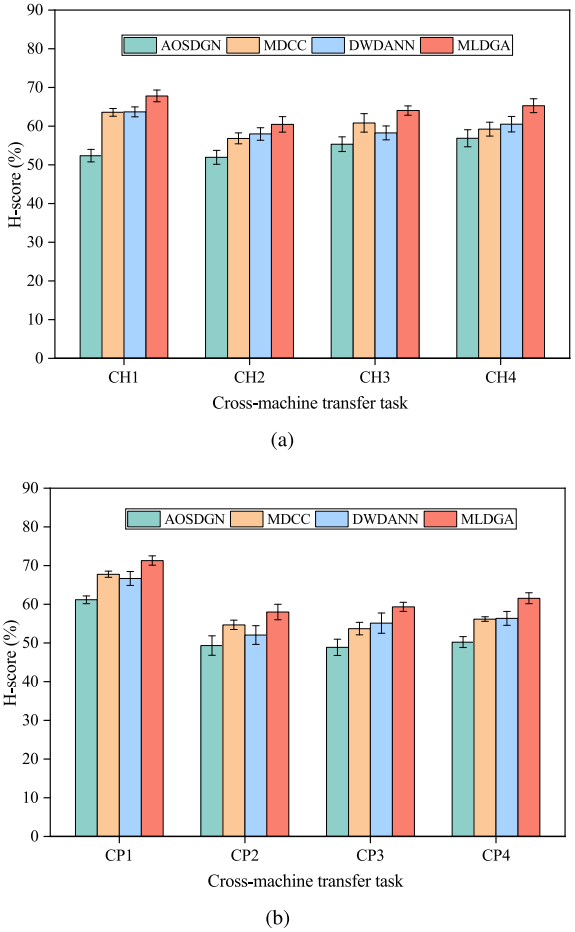
TABLE XI
WORKING CONDITIONS OF RFB DATASET

Working condition	Rotating speed (rpm)
W_{13}	100
W_{14}	200
W_{15}	300
W_{16}	400

reduced. However, the proposed MLDGA achieves the best performance on all cross-machine FD tasks. For example, the average H-score obtained by MLDGA is superior to other OSDGFD methods under HUST → PU. Taking the transfer task CH2 with an openness of 0.43 as an example, the average H-score obtained by MLDGA is 60.45% on CH2, which is 8.5%, 3.61%, and 4.12% higher than AOSDGN, MDCC, and DWDANN, respectively. Under PU → HUST, the average H-score obtained by MLDGA is 71.29% on the transfer task CP1 with an openness of 0.29, which is 10.12%, 3.54%, and 4.62% higher than AOSDGN, MDCC, and DWDANN, respectively. The experimental results achieved under two cross-machine scenarios prove that the proposed MLDGA can still effectively balance the known class recognition and unknown class detection in the face of significant distribution discrepancies, showing a strong cross-machine OSDGFD capability.

E. Comparison With Different OSDGFD Methods on Real Factory Bearing Dataset

To further verify the practicability of the proposed MLDGA in real industrial scenarios, the additional experiments are



(a)

(b)

Fig. 14. H-scores obtained with four different OSDGFD methods under two different cross-machine scenarios. (a) HUST → PU. (b) PU → HUST.

carried out on a real factory bearing (RFB) dataset [31] from a real factory environment. This experiment aims to evaluate the OSDGFD ability of MLDGA under the real distribution discrepancies and compare it with five different OSDGFD methods. The RFB dataset consists of naturally developed defective bearings obtained at four different rotating speeds in the actual production, as shown in Table XI. The RFB dataset includes four different health states: HC, rolling-element deviation (RD), rolling-element missing (RM),

TABLE XII
INFORMATION OF FAULT CLASSES OF RFB DATASET

Health state	Class label
HC	1
RD	2
RM	3
YS	4

TABLE XIII
TRANSFER TASK INFORMATION ON RFB DATASET

Task	SDs → Unknown TD	Source classes	Target classes	Openness
R1	$[W_{14}, W_{15}, W_{16}] \rightarrow [W_{13}]$	1, 3, 4	1, **2**, 3, 4	0.25
R2	$[W_{14}, W_{15}, W_{16}] \rightarrow [W_{13}]$	1, 2, 4	1, 2, **3**, 4	0.25
R3	$[W_{13}, W_{15}, W_{16}] \rightarrow [W_{14}]$	1, 2, 3	1, 2, 3, **4**	0.25
R4	$[W_{13}, W_{15}, W_{16}] \rightarrow [W_{14}]$	1, 4	1, **2**, **3**, 4	0.50
R5	$[W_{13}, W_{14}, W_{16}] \rightarrow [W_{15}]$	2, 3	**1**, 2, 3, **4**	0.50
R6	$[W_{13}, W_{14}, W_{16}] \rightarrow [W_{15}]$	1, 2	1, 2, **3**, **4**	0.50
R7	$[W_{13}, W_{14}, W_{15}] \rightarrow [W_{16}]$	1, 2, 4	1, 2, **3**, 4	0.25
R8	$[W_{13}, W_{14}, W_{15}] \rightarrow [W_{16}]$	1, 3	1, **2**, 3, **4**	0.50

and yarn stick (YS), as seen in Table XII. The proposed MLDGA and five different OSDG methods carried out comparative experiments on different transfer tasks shown in Table XIII.

The H-scores obtained by MLDGA and five different OSDGFD methods on RFB dataset are shown in Table XIV. On the RFB dataset, the average H-score obtained by MLDGA is 17.28%, 18.18%, 12.29%, 6.3%, and 6.99% higher than that obtained by M1, M2, AOSDGN, MDCC, and DWDAAN, respectively. This indicates that MLDGA can more fully capture and balance the distribution discrepancies of complex data in real industrial scenarios through the metalearning mechanism with dual-level gradient alignment, the entropy-guided dynamic weighting strategy, and the classification-clustering dual-guided open decision boundary construction strategy, which is significantly superior to the other OSDGFD methods on RFB dataset.

F. Ablation Experiments

To validate the role of different components in MLDGA, the eight variants A1, A2, A3, A4, A5, A6, A7, and A8 of MLDGA are adopted for conducting ablation experiments on PU and PHM datasets. A1 is the base version of MLDGA without adopting any strategy. A2 is that the metalearning optimization strategy with dual-level gradient alignment is only adopted in MLDGA. A3 is that the entropy-guided dynamic weighting strategy is only adopted in MLDGA. A4 is designed to identify unknown faults using only the multibinary classifier in MLDGA. A5 is that the triplet loss is only adopted to increase the compactness of the class clusters in MLDGA. A6 is that the entropy-guided dynamic weighting strategy is not adopted in MLDGA. A7 is that the multibinary classifier is not used to identify unknown faults in MLDGA. A8 is that the triplet loss is not adopted to increase the compactness of the class clusters in MLDGA. Tables XV and XVI present the H-scores obtained from MLDGA and its eight variants on PU and PHM datasets, respectively.

Compared with A1, the average H-score of MLDGA on PU and PHM datasets is increased by 34.41% and 36.85%, respectively, which shows that by using the entropy-guided dynamic weighting strategy, the multibinary classifier that can identify unknown classes, and the triplet loss that can increase the compactness of the cluster, MLDGA can effectively identify known and unknown classes. Compared with A1, the average H-score of A2 on PU and PHM datasets is increased by 9.59% and 16.76%, respectively, which indicates that the dual-level gradient alignment can alleviate the gradient direction conflict between different tasks and make the gradient update directions more consistent, thus enhancing the generalization ability of the FD model under the domain and label shifts. Compared with A1, the average H-score of A3 on PU and PHM datasets is increased by 3.38% and 5.19%, respectively, which shows that MLDGA can enhance the discriminative ability of the FD model by using the entropy to dynamically allocate the reliability weight for the corresponding samples. Compared with A1, the average H-score of A4 on PU and PHM datasets is increased by 16.39% and 21.40%, respectively, which indicates that the identification of unknown fault classes by using multibinary classifier can avoid misjudging unknown classes to known classes, thereby improving the FD accuracy of the model under the open-set scenario. Compared with A1, the average H-score of A5 on PU and PHM datasets is increased by 4.24% and 5.80%, respectively, which shows that the feature boundaries of known classes can be optimized by minimizing the triplet loss between fault features, making it easier for the FD model to identify known and unknown fault classes. Compared with A6, the average H-score of MLDGA on PU and PHM datasets is increased by 12.26% and 10.55%, respectively, which indicates that the FD model can identify different fault classes more effectively by dynamically assigning weights to the classification loss. Compared with A7, the average H-score of MLDGA on PU and PHM datasets is increased by 21.37% and 16.36%, respectively, this is because MLDGA can more accurately identify the unknown fault classes in the TD by using the multibinary classifier to calculate the confidence scores of the samples to determine the unknown classes. In the face of unseen fault classes, the multibinary classifier can dynamically adjust the decision boundary of the classifier, avoiding the degradation of diagnosis performance caused by the existence of unknown fault classes in the TD. Compared with A8, the average H-score of MLDGA on PU and PHM datasets is increased by 9.78% and 9.44%, respectively, which indicates that MLDGA can minimize the triplet loss between fault features, allowing the FD model to bring the samples of the same classes closer and push the samples of different classes farther, thereby enhancing the discriminative ability of fault features.

G. Visualization Analysis of FD Results

To analyze the learning ability of different OSDGFD methods for domain-invariant features, t-SNE technology is adopted

TABLE XIV
H-SCORES (%) OF DIFFERENT OSDGFD METHODS ON RFB DATASET

Method	R1	R2	R3	R4	R5	R6	R7	R8	Avg.
M1	69.05 ± 0.8	67.81 ± 2.6	71.62 ± 1.8	65.92 ± 2.0	61.30 ± 1.4	68.80 ± 1.8	73.37 ± 2.2	60.15 ± 3.0	67.25
M2	66.38 ± 1.2	68.22 ± 1.5	70.14 ± 1.8	64.36 ± 0.7	62.75 ± 1.0	65.52 ± 1.9	73.10 ± 0.8	60.33 ± 2.4	66.35
AOSDGN	70.70 ± 1.3	75.47 ± 0.5	75.80 ± 0.8	68.63 ± 0.5	70.47 ± 0.5	71.35 ± 1.0	78.98 ± 1.3	66.50 ± 0.6	72.24
MDCC	79.45 ± 2.2	81.89 ± 0.7	84.79 ± 1.6	75.92 ± 0.4	73.43 ± 2.0	76.97 ± 1.9	81.25 ± 1.1	72.11 ± 2.0	78.23
DWDAAN	78.56 ± 2.0	80.08 ± 1.1	84.15 ± 1.5	74.64 ± 0.8	73.10 ± 0.6	75.38 ± 2.5	80.66 ± 1.8	73.75 ± 2.2	77.54
MLDGA	**86.24 ± 1.4**	**90.04 ± 0.6**	**87.48 ± 1.2**	**80.97 ± 0.9**	**81.12 ± 1.7**	**81.40 ± 1.8**	**86.48 ± 1.0**	**82.47 ± 2.4**	**84.53**

TABLE XV
H-SCORES (%) OF THE PROPOSED MLDGA AND ITS EIGHT VARIANTS ON PU DATASET

Method	P1	P2	P3	P4	Avg.
A1	57.64 ± 0.6	51.60 ± 1.4	55.40 ± 0.9	58.96 ± 0.7	55.90
A2	65.38 ± 1.6	61.55 ± 2.0	66.91 ± 1.8	68.12 ± 1.5	65.49
A3	61.94 ± 0.8	53.13 ± 1.3	59.37 ± 0.8	62.67 ± 0.3	59.28
A4	73.37 ± 1.0	68.13 ± 1.9	74.15 ± 1.2	73.49 ± 1.0	72.29
A5	62.78 ± 1.6	53.95 ± 0.8	60.33 ± 0.6	63.51 ± 1.2	60.14
A6	80.95 ± 0.8	73.20 ± 1.0	78.66 ± 0.5	79.37 ± 0.6	78.05
A7	68.38 ± 2.0	65.90 ± 1.8	70.56 ± 1.6	70.92 ± 1.7	68.94
A8	82.17 ± 1.3	78.49 ± 1.2	80.39 ± 1.0	81.06 ± 0.8	80.53
MLDGA	**94.76 ± 0.3**	**86.91 ± 1.0**	**88.84 ± 0.7**	**90.71 ± 0.5**	**90.31**

TABLE XVI
H-SCORES (%) OF THE PROPOSED MLDGA AND ITS EIGHT VARIANTS ON PHM DATASET

Method	G1	G2	G3	G4	Avg.
A1	52.21 ± 1.0	50.59 ± 0.7	57.68 ± 2.0	49.26 ± 1.2	52.44
A2	71.10 ± 1.2	66.25 ± 2.4	72.75 ± 2.8	66.68 ± 1.4	69.20
A3	55.80 ± 0.9	57.22 ± 0.6	63.90 ± 0.5	53.58 ± 2.4	57.63
A4	74.95 ± 1.1	71.02 ± 2.1	78.58 ± 1.5	70.80 ± 1.3	73.84
A5	54.89 ± 1.7	59.10 ± 1.9	63.72 ± 1.0	55.25 ± 0.5	58.24
A6	81.62 ± 0.6	79.92 ± 1.0	81.93 ± 1.1	71.50 ± 1.4	78.74
A7	75.16 ± 1.4	70.03 ± 1.8	77.48 ± 0.8	69.05 ± 0.9	72.93
A8	84.58 ± 0.8	79.49 ± 1.6	81.68 ± 1.5	73.65 ± 0.7	79.85
MLDGA	**90.25 ± 0.7**	**89.08 ± 0.9**	**92.72 ± 1.0**	**85.12 ± 0.6**	**89.29**

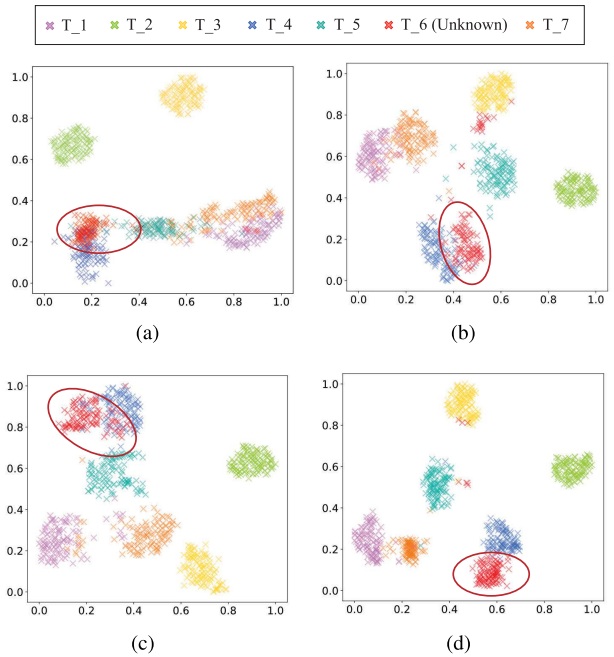


Fig. 15. Feature visualizations of the diagnosis results of different OSDGFD methods on the TD under the transfer task H2 of HUST dataset. (a) M1. (b) AOSDGN. (c) MDCC. (d) MLDGA.

to map the fault features extracted by different OSDGFD methods into the 2-D space for visualizations. Fig. 15 illustrates the feature visualizations of the diagnosis results of different OSDGFD methods on the TD under the transfer task H2 of HUST dataset. As shown in Fig. 15(a), in the FD results of M1, the features of the five fault classes, including the unknown class, all exhibit varying degrees of interclass overlap, which makes it difficult to form the decision boundaries of the fault classes. It can be observed from Fig. 15(b) and (c) that in the FD results of AOSDGN and MDCC, although the features belonging to the same fault class are relatively concentrated, there is a significant interclass overlap between the features of the unknown class (i.e., class 6) and known classes (e.g., class 4), reducing the classification performance of AOSDGN and MDCC. It can be observed from Fig. 15(d) that MLDGA performs best in terms of clustering effect, with more compact intraclass structure and more obvious interclass separation phenomenon, and the overlap between the features of the unknown class (i.e., class 6) and other known classes is small. This indicates that MLDGA has superior OSDGFD performance and clearer classification boundaries.

To further analyze the FD performance of different OSDGFD methods on each fault class, the confusion matrix is introduced to analyze the FD results on the transfer task P1 of PU dataset. Fig. 16 shows the confusion matrices of different OSDGFD methods on the transfer task P1. As seen in Fig. 16, MLDGA has higher FD accuracies in both known and unknown fault classes. For instance, on the second fault class (i.e., OF1), the FD accuracy of MLDGA is 92.43%, which is 14.02%, 12.33%, and 9.28% higher than those of M1, AOSDGN, and MDCC, respectively. This indicates that MLDGA can well handle the domain and label shift problems between multiple SDs and the unseen TD under the scenarios where the unknown fault classes occur in the unseen TD. On the unknown fault class (i.e., IF+OF2), the FD accuracy of MLDGA is 96.21%, which is 31.16%, 16.98%, 9.60% higher than those of M1, AOSDGN, and MDCC, respectively. This is because MLDGA can construct a more reasonable classification decision boundary, thus reducing the possibility of known fault classes being incorrectly classified as unknown fault classes and improve the detection ability of the model for unknown fault classes.

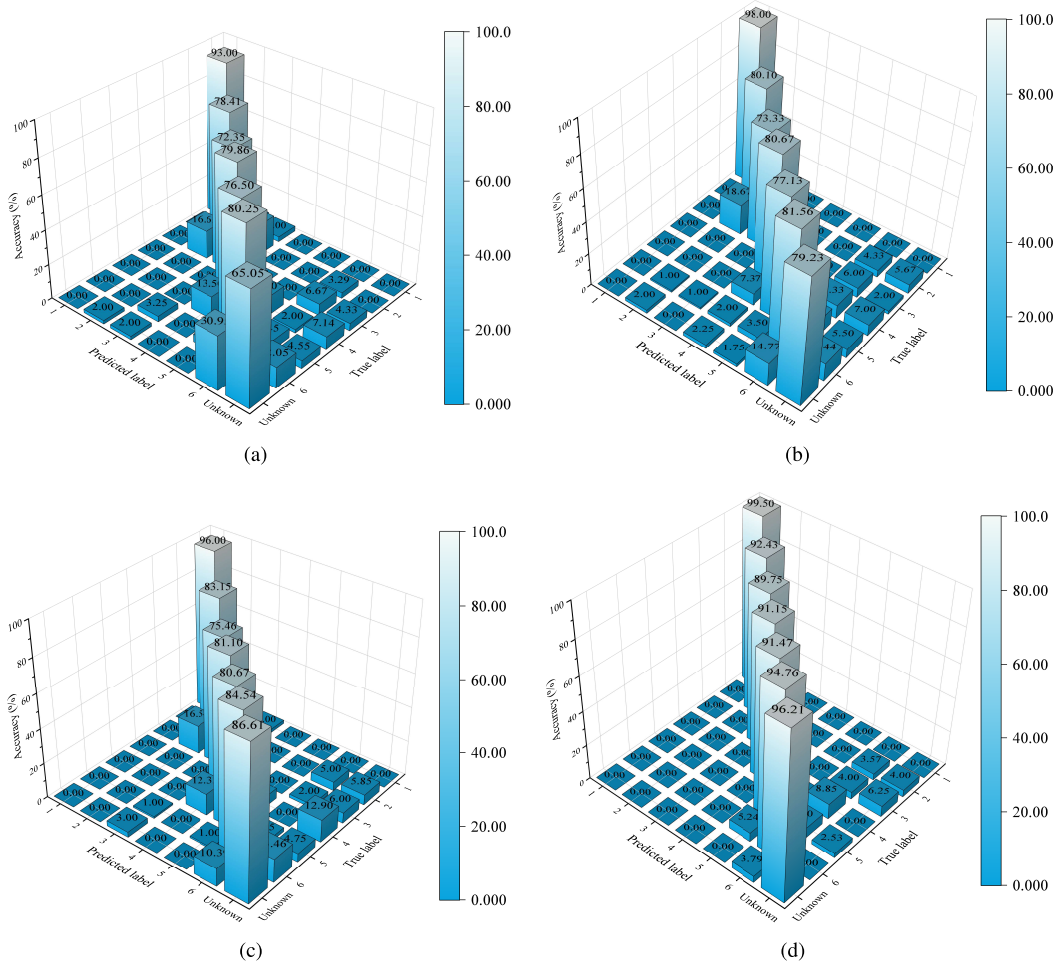


Fig. 16. Confusion matrices of different OSDGFD methods on the transfer task P1 of PU dataset. (a) M1. (b) AOSDGN. (c) MDCC. (d) MLDGA.

V. Conclusion

In this article, a novel OSDG approach via MLDGA for intelligent FD is proposed, which successfully addresses the problems of domain and label shifts caused by unknown fault classes on the unseen TD. The metalearning optimization strategy with dual-level gradient alignment is adopted, and the gradient update directions of the interdomain and interclass tasks are simultaneously optimized by gradient matching to effectively realize the interdomain gradient matching and interclass gradient matching, thereby ensuring that the class decision boundaries are reasonably located in the optimal positions of different fault classes. Moreover, the entropy-guided dynamic weighting strategy and the classification-clustering dual-guided open decision boundary construction strategy are adopted, significantly improving the ability to distinguish known and unknown fault classes. Extensive experiments are performed on HUST, PU, PHM, and RFB datasets to verify the effectiveness of the proposed approach. The average H-scores of the proposed approach reach 91.36%, 90.31%, 89.29%, and 84.53%, respectively. The OS* and UK of the proposed approach are better than those of the other comparison methods on the whole.

In practical industrial applications, the proposed MLDGA still has some limitations. First, due to the high resource cost and data privacy protection, it is often difficult for a single user to collect enough data to build a data-driven FD model with reliable performance. Therefore, federated learning can be considered to alleviate this problem. Second, not only the unknown fault classes may appear in the unseen TD, but also the label space between the SDs and that between the SDs and TD are uncertain, that is, universal DG. Therefore, the FD method combining federated learning with universal DG will be further explored in the future work to better meet the actual industrial application requirements.

References

[1] D. Neupane, M. R. Bouadjenek, R. Dazeley, and S. Aryal, "Data-driven machinery fault diagnosis: A comprehensive review," *Neurocomputing*, vol. 627, Apr. 2025, Art. no. 129588.

[2] I. Misbah, C. K. M. Lee, and K. L. Keung, "Fault diagnosis in rotating machines based on transfer learning: Literature review," *Knowl.-Based Syst.*, vol. 283, Jan. 2024, Art. no. 111158.

[3] A. U. Rehman, W. Jiao, J. Sun, H. Pan, and T. Yan, "Open set recognition methods for fault diagnosis: A review," in *Proc. 15th Int. Conf. Adv. Comput. Intell. (ICACI)*, May 2023, pp. 1–8.

[4] X. Yu et al., "Deep-learning-based open set fault diagnosis by extreme value theory," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 185–196, Jan. 2022.

[5] A. Lundgren and D. Jung, "Data-driven fault diagnosis analysis and open-set classification of time-series data," *Control Eng. Pract.*, vol. 121, Apr. 2022, Art. no. 105006.

[6] P. Peng, J. Lu, T. Xie, S. Tao, H. Wang, and H. Zhang, "Open-set fault diagnosis via supervised contrastive learning with negative out-of-distribution data augmentation," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2463–2473, Mar. 2023.
[7] J. Mei, M. Zhu, W. Liu, M. Fu, and Q. Tang, "Conditional variational encoder classifier for open set fault classification of rotating machinery vibration signals," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 3038–3049, Mar. 2024.
[8] D. Wei, M. Zuo, and Z. Tian, "Open-set fault diagnosis for industrial rotating machines based on trustworthy deep learning," *IEEE Trans. Ind. Cyber-Phys. Syst.*, vol. 3, pp. 181–189, Feb. 2025.
[9] Z. Zhu, G. Chen, and G. Tang, "Domain adaptation with multi-adversarial learning for open-set cross-domain intelligent bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3533411.
[10] Z. Su, W. Jiang, Y. Zhao, S. Feng, S. Wang, and M. Luo, "Cross-domain open-set fault diagnosis based on target domain slanted adversarial network for rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
[11] Y. Zhang, H. Zhang, B. Chen, J. Zheng, and H. Pan, "Integrating intrinsic information: A novel open set domain adaptation network for cross-domain fault diagnosis with multiple unknown faults," *Knowl.-Based Syst.*, vol. 299, Sep. 2024, Art. no. 112100.
[12] B. Wang, M. Zhang, H. Xu, C. Wang, and W. Yang, "Adversarial domain adaptation with dual auxiliary classifiers for cross-domain open set intelligent fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 3529014.
[13] L. Wang, Y. Gao, X. Li, and L. Gao, "Self-supervised-enabled open-set cross-domain fault diagnosis method for rotating machinery," *IEEE Trans. Ind. Informat.*, vol. 20, no. 8, pp. 10314–10324, Aug. 2024.
[14] C. Weng, B. Lu, L. Chen, X. Zhao, and W. Huang, "A novel progressive domain separation network with multi-metric ensemble quantification for open set fault diagnosis of motor bearings," *Adv. Eng. Informat.*, vol. 64, Mar. 2025, Art. no. 103060.
[15] Y. Xiao, H. Shao, S. Yan, J. Wang, Y. Peng, and B. Liu, "Domain generalization for rotating machinery fault diagnosis: A survey," *Adv. Eng. Informat.*, vol. 64, Mar. 2025, Art. no. 103063.
[16] L. Chen, Q. Li, C. Shen, J. Zhu, D. Wang, and M. Xia, "Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1790–1800, Mar. 2022.
[17] Q. Qian, J. Zhou, and Y. Qin, "Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions," *IEEE Trans. Ind. Informat.*, vol. 19, no. 9, pp. 9898–9908, Sep. 2023.
[18] C. Zhao and W. Shen, "Imbalanced domain generalization via semantic-discriminative augmentation for intelligent fault diagnosis," *Adv. Eng. Informat.*, vol. 59, Jan. 2024, Art. no. 102262.
[19] M. Liu, Z. Cheng, Y. Yang, N. Hu, G. Shen, and Y. Yang, "Adaptive reconstruct feature difference network for open set domain generalization fault diagnosis," *Eng. Appl. Artif. Intell.*, vol. 142, Feb. 2025, Art. no. 109895.
[20] C. Zhao and W. Shen, "Adaptive open set domain generalization network: Learning to diagnose unknown faults under unknown working conditions," *Rel. Eng. Syst. Saf.*, vol. 226, Oct. 2022, Art. no. 108672.
[21] B. Lu, Y. Zhang, Q. Sun, M. Li, and P. Li, "A novel multidomain contrastive-coding-based open-set domain generalization framework for machinery fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 6369–6381, Apr. 2024.
[22] C. Jian, Y. Peng, G. Mo, and H. Chen, "Open-set domain generalization for fault diagnosis through data augmentation and a dual-level weighted mechanism," *Adv. Eng. Informat.*, vol. 62, Oct. 2024, Art. no. 102703.
[23] A. Vettoruzzo, M.-R. Bouguelia, J. Vanschoren, T. Rögnvaldsson, and K. Santosh, "Advances and challenges in meta-learning: A technical review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 4763–4779, Jul. 2024.
[24] Y. Shi et al., "Gradient matching for domain generalization," 2021, *arXiv:2104.09937*.
[25] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*.
[26] C. Zhao, E. Zio, and W. Shen, "Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study," *Rel. Eng. Syst. Saf.*, vol. 245, May 2024, Art. no. 109964.
[27] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. Eur. Conf. PHM Soc. (PHME)*, 2016, vol. 3, no. 1, pp. 1–17.
[28] B. Zhang, W. Li, X.-L. Li, and S.-K. Ng, "Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks," *IEEE Access*, vol. 6, pp. 66367–66384, 2018.
[29] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada, "Open set domain adaptation by backpropagation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jun. 2018, pp. 153–168.
[30] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1563–1572.
[31] Y. Chen and L. Xiao, "A multisource–multitarget domain adaptation method for rolling bearing fault diagnosis," *IEEE Sensors J.*, vol. 24, no. 3, pp. 3406–3419, Feb. 2024.

Lanjun Wan received the B.S. and M.S. degrees in computer science and technology from Hunan University of Technology, Zhuzhou, China, in 2005 and 2009, respectively, and the Ph.D. degree in circuits and systems from Hunan University, Changsha, China, in 2016.

He is currently an Associate Professor with the School of Computer Science and Artificial Intelligence, Hunan University of Technology. His research interests include industrial big data analysis, industrial equipment fault diagnosis, high-performance computing, and parallel computing.

Jian Zhou received the B.S. degree in software engineering from Huaihua University, Huaihua, China, in 2022, and the M.S. degree in computer technology from Hunan University of Technology, Zhuzhou, China, in 2025.

He is currently a Lecturer with the School of Computer Science, Guangzhou College of Applied Science and Technology, Guangzhou, China. His research interests include industrial big data analysis and industrial equipment fault diagnosis.

Le Huang received the B.S. degree in computer science and technology from Hunan Institute of Science and Technology, Yueyang, China, in 2022, and the M.S. degree in computer technology from Hunan University of Technology, Zhuzhou, China, in 2025.

She is currently a Lecturer with the School of Computer Science, Guangzhou College of Applied Science and Technology, Guangzhou, China. Her research interests include industrial big data analysis and industrial equipment fault diagnosis.

Jiaen Ning received the B.S. degree in computer science and technology from Hunan University of Technology, Zhuzhou, China, in 2022, and the M.S. degree in computer technology from Hunan University of Technology, in 2025. He is currently pursuing the Ph.D. degree in information and communication engineering with the Central South University, Changsha, China.

His research interests include industrial big data analysis and industrial equipment fault diagnosis.

Hongwei Tan received the B.S. degree in software engineering from Hunan University of Information Technology, Changsha, China, in 2023. He is currently pursuing the M.S. degree in computer science and technology with Hunan University of Technology, Zhuzhou, China.

His research interests include industrial big data analysis and industrial equipment fault diagnosis.

Wei Ni received the B.S. and Ph.D. degrees in communication engineering from Huazhong University of Science and Technology, Wuhan, China, in 2003 and 2012, respectively.

He is currently an Assistant Professor with the School of Computer Science and Artificial Intelligence, Hunan University of Technology, Zhuzhou, China. His research interests include industrial artificial intelligent and industrial equipment fault diagnosis.

Keqin Li (Fellow, IEEE) received a B.S. degree in computer science from Tsinghua University, Beijing, China, in 1985, and the Ph.D. degree in computer science from the University of Houston, Houston, TX, USA, in 1990.

He is a SUNY Distinguished Professor at the State University of New York, New Paltz, NY, USA, and the National Distinguished Professor at Hunan University (China), Changsha, China. He has authored or co-authored more than 1130 journal articles, book chapters, and refereed conference papers. He holds nearly 80 patents announced or authorized by the Chinese National Intellectual Property Administration. Since 2020, he has been among the world's top few most influential scientists in parallel and distributed computing regarding single-year impact (ranked #2) and career-long impact (ranked #4) based on a composite indicator of the Scopus citation database. He is listed in Scilit Top Cited Scholars from 2023 to 2024 and is among the top 0.02% out of over 20 million scholars worldwide based on top-cited publications. He is listed in ScholarGPS Highly Ranked Scholars from 2022 to 2024 and is among the top 0.002% out of over 30 million scholars worldwide based on a composite score of three ranking metrics for research productivity, impact, and quality in the recent five years.

Dr. Li is an AAAS Fellow, an AAIA Fellow, an ACIS Fellow, and an AIIA Fellow. He is a Member of the SUNY Distinguished Academy. He is a Member of the European Academy of Sciences and Arts. He is a Member of Academia Europaea (Academician of the Academy of Europe). He was a recipient of the International Science and Technology Cooperation Award from 2022 to 2023 and the 2023 Xiaoxiang Friendship Award of Hunan Province, China. He received the IEEE TCCLD Research Impact Award from the IEEE CS Technical Committee on Cloud Computing in 2022 and the IEEE TCSVC Research Innovation Award from the IEEE CS Technical Community on Services Computing in 2023. He won the IEEE Region 1 Technological Innovation Award (Academic) in 2023.