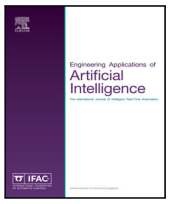
Engineering Applications of Artificial Intelligence 164 (2026) 113322

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

Prototype contrastive-based decentralized federated transfer learning for intelligent fault diagnosis

Zhaokang Li [a], Lanjun Wan [a,*], Jiaen Ning [a], Wei Ni [a], Keqin Li [b]

[a] School of Computer Science and Artificial Intelligence, Hunan University of Technology, Zhuzhou 412007, China
[b] Department of Computer Science, State University of New York, New Paltz, NY, 12561, USA

ARTICLE INFO

Keywords:
Contrastive learning
Fault diagnosis
Federated learning
Prototype learning
Transfer learning

ABSTRACT

Federated transfer learning (FTL) has been widely applied in the rotating machinery fault diagnosis (RMFD). However, how to alleviate the impact of domain shift and avoid the negative transfer problem caused by model aggregation remain the key challenges that urgently need to be solved. Therefore, a prototype contrastive-based decentralized federated transfer learning (PCDFTL) approach for intelligent fault diagnosis is proposed. Firstly, a dual-phase adapter projection optimization (DPAPO) strategy is designed to ensure that the sample features can be mapped into a unified feature space by freezing the certain layers of the model while fine-tuning the other layers in the two phases, effectively achieving the domain alignment. Secondly, a prototype-based differentiated contrastive (PDC) loss is designed to mitigate the impact of domain shift by combining the multi-kernel maximum mean discrepancy (MK-MMD) and leveraging the target prototypes to guide the source models for domain alignment. Thirdly, by abandoning the idea of aggregating a global model, a multi-source dynamic weighted inference (MSDWI) strategy is studied to avoid the negative transfer problem by dynamically weighting and fusing the inference results of multiple source models on the target domain according to the distribution discrepancies between the source and target domains, the stability of domain adaptation (DA) process, and the certainty of the entropy of the inference results. Finally, extensive experiments are conducted to verify the effectiveness of the proposed approach. The results reveal that the approach can achieve excellent cross-domain RMFD performance while protecting data privacy. The code is available at https://github.com/HutAC/PCDFTL.

1. Introduction

Rotating machines are indispensable in the industrial fields. Once they malfunction, it will severely affect production efficiency and may even lead to casualties. Recently, deep learning (DL)-based RMFD approaches have attracted extensive attention (Su et al., 2024). For example, Li et al. (2025c) designed a view attention mechanism, where the multi-view graph neural network is applied to RMFD, which can still maintain high fault diagnosis (FD) accuracy under strong noise conditions. Keshun et al. (2025) adopted a multi-modal information fusion technology and guided the DL model training by the multi-regularization constraints, improving the precision and reliability of FD. Existing research prove that DL technology has been successfully applied in RMFD. However, DL-based RMFD approaches rely on abundant high-quality labeled training data. In actual industrial productions, it will face the intricacy and changeable operating conditions, and a single enterprise usually finds it difficult to obtain enough high-quality labeled training data to build accurate and reliable RMFD models. Due to data privacy issues, it is not feasible to directly concentrate the local data of multiple enterprises together for model training.

To overcome the aforementioned issues, some scholars have applied federated learning (FL) to RMFD (Prigent et al., 2024). Such as, Sun et al. (2025) studied a periodic weights-based federated aggregation strategy to accelerate the convergence of a global model. Jiang et al. (2024) devised a diagnosis knowledge-based FL framework, improving the FD accuracy of wind turbines by optimizing a cloud aggregation strategy and local knowledge learning. Zhou et al. (2025a) devised a clustered FL framework for wind turbine FD, improving the efficiency and accuracy of FD by constructing the lightweight multi-scale residual network on each client. Mehta et al. (2023) produced a FL method for hybrid RMFD, which can diagnose faults of various rotating machinery by a dual-model classifier. Du et al. (2024) designed a low-cost FL framework for RMFD, accelerating the local model training and improving the FD accuracy by various lightweight training methods. Xiao et al. (2025) implemented Rademacher complexity-based dynamic weighted

* Corresponding author.
E-mail address: wanlanjun@hut.edu.cn (L. Wan).

https://doi.org/10.1016/j.engappai.2025.113322
Received 23 April 2025; Received in revised form 13 October 2025; Accepted 22 November 2025

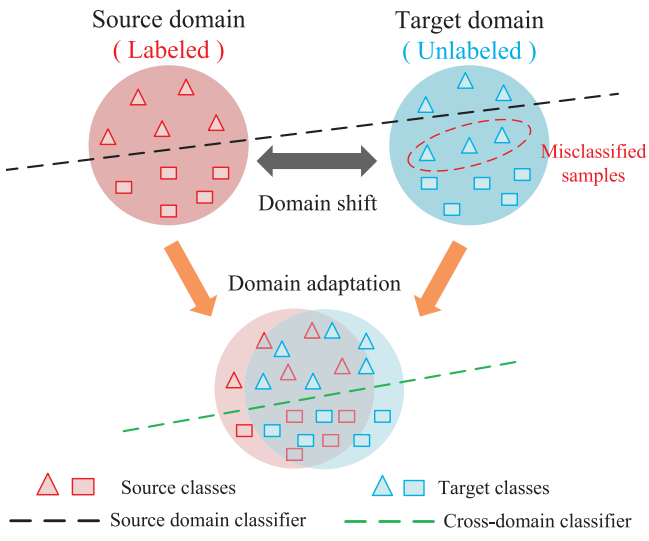


Fig. 1. Illustration of DA.

aggregation, achieving efficient feature extraction and robust RMFD performance. The above studies demonstrate that it is possible to successfully train RMFD models in a collaborative manner using the data from multiple clients while ensuring data privacy. However, existing FL-based RMFD approaches usually assume that the data from multiple clients are independent and identically distributed. In the actual FD scenarios, the data from different clients commonly follow different distributions, which originate from diverse operating conditions or rotating machines.

Transfer learning (TL) is adept at dealing with distribution discrepancies between different domains (Tang et al., 2024). TL aims to transfer the knowledge from source domain to target domain, where DA is the most commonly used approach. DA aims to improve the classification accuracy of the model on target domain by mitigating the impact of domain shift, as revealed in Fig. 1. Recently, TL-based RMFD approaches have received increasing attention. Yu et al. (2025) designed a novel multi-scale feature extractor for extracting domain-invariant features, enhancing the generalization of the model and improving the accuracy of RMFD. Yang et al. (2024) constructed a dual-network autoencoder that capable of capturing deep features and temporal information, and introduced Wasserstein divergence in the domain-adversarial process, improving the stability of model training and successfully extracting domain-invariant features. Wang et al. (2025) proposed an adversarial-based DA network for unsupervised RMFD, alleviating the impact of data distribution discrepancy and improving FD accuracy. Zhang et al. (2024) studied a dual-phase multi-source partial DA approach for RMFD, enhancing the ability to extract domain-invariant features. Liu et al. (2024a) designed an improved subdomain adaptation network, improving the capability of cross-machine FD. Zhou et al. (2025b) developed a cost-sensitive classifier for complicated cross-domain RMFD, improving the inter-class distinguishability by adjusting the weights of different classes. Li et al. (2025a) developed an unsupervised DA framework based on joint marginal-conditional distribution alignments and pseudo-label self-correcting, improving cross-domain FD accuracy. Li et al. (2025b) devised a DA network based deep feature alignment and maximum classifier discrepancy, enhancing the cross-operating condition FD ability. The previous research can deal with RMFD in the presence of data distribution discrepancy. In the traditional TL-based RMFD, it is required that both source and target domain data are publicly accessible and are centralized for model training. However, due to data privacy issues, it is difficult to centralize the data provided by different enterprises.

Recently, some researchers have focused on FTL-based RMFD (Qian et al., 2025), which aims to integrate FL and TL for privacy-preserving cross-domain RMFD. FTL not only achieves privacy-safe collaborative modeling through FL architecture, but also effectively overcomes domain shift between clients through TL, which can significantly improve the generalization performance of the global model on the target domain. For instance, Wan et al. (2024) studied a ring-based decentralized FTL approach, which not only significantly reduces communication overhead but also enhances generalization of global model via the multi-perspective aggregation strategy. Li et al. (2024a) introduced domain alignment loss in FL to eliminate domain distribution discrepancy, strengthening the FD capability of the global model. Zhao et al. (2024a) designed an FTL approach which combines multi-source DA and self-supervised learning, achieving excellent results in RMFD under data privacy. Wang et al. (2023) devised a low-quality knowledge filtering strategy in FTL with low-quality data, effectively improving the accuracy of cross-domain FD with data privacy. Zhao et al. (2023) proposed an FTL framework for RMFD, boosting the FD accuracy of the global model by federated multi-source DA and domain adversarial. He and Shen (He and Shen, 2024) designed a hybrid perception mechanism in FTL-based cross-machine FD, mitigating global domain shift. Zhan and Shen (Zhao and Shen, 2024) combined knowledge distillation and domain generalization to devise a novel FTL method, enhancing the FD capability of the model on the unseen conditions. Li et al. (2024b) constructed a shared reference domain and adopted a non-deterministic sampling strategy in FTL for RMFD, improving the generalization of the model. Liu et al. (2024b) present a FTL framework that combines differential training with dynamic weighted aggregation based on contribution degree, achieving high bearing FD accuracy. The prior studies indicate that FTL-based RMFD approaches, by deeply integrating the privacy-protection mechanism of FL and the cross-domain knowledge transfer capability of TL, can effectively solve the collaborative challenges of the data silos, domain distribution discrepancy, and privacy protection in RMFD.

In summary, FTL has been extensively used in RMFD. However, how to effectively mitigate the impact of domain shift and avoid the negative transfer problem remain the key challenges that urgently need to be solved (Qian et al., 2025). Specifically, in the traditional FTL frameworks, the data from different clients are obtained from various operating conditions or different machines, which leads to the domain shift between clients, directly degrading the generalization of the global model. Moreover, the models from different clients are more inclined to conform to the data distribution of the respective client. During the model aggregation, if the aggregation weights fail to precisely measure the data distribution discrepancies across clients, it will result in sub-optimal generalization capability of the global model, thereby inducing the negative transfer. Compared to traditional domain adversarial strategies adopted in cross-domain FD (Misbah et al., 2024), the prototype contrastive learning (PCL) has the following advantages. Firstly, by explicitly constructing discriminative prototypes, PCL enables more interpretable and fine-grained domain alignment compared with domain adversarial, and can more explicitly optimize intra-class compactness and inter-class separability, thereby learning more discriminative cross-domain feature representations. Secondly, PCL avoids the adversarial game process between the discriminator and feature extractor that is difficult to balance in the domain adversarial training, making the training process more stable and often achieving better cross-domain FD performance. Finally, PCL only requires exchanging lightweight prototype information rather than feature representations of all samples, greatly reducing communication cost while preserving data privacy. Consequently, a prototype contrastive-based decentralized FTL approach for intelligent FD is proposed. The primary contributions of this paper are as follows.

(1) A dual-phase adapter projection optimization strategy is proposed, which can map the sample features of the source and target domains into the unified feature space to enhance the effect of domain alignment.

(2) A prototype-based differentiated contrastive loss is designed, which can effectively mitigate the impact of domain shift by combining MK-MMD and leveraging the target prototypes to guide the source models for domain alignment.
(3) A multi-source dynamic weighted inference strategy is proposed, where the inference results of multiple source models on the target domain are weighted and fused dynamically according to the distribution discrepancies between the source and target domains, the stability of DA, and the certainty of the entropy of the inference results, which can effectively avoid the negative transfer problem.

The remainder sections are organized as follows. The basic theory is introduced in Section 2. The proposed approach is described in Section 3. The experimental results and analysis are presented in Section 4. The conclusions and future work are given in Section 5.

2. Basic theory

2.1. Federated transfer learning

FTL (Qian et al., 2025) is a distributed machine learning paradigm that enables cross-domain knowledge transfer while protecting data privacy. The core of FTL lies in addressing the challenge of the co-existence of data silos and domain shift through the collaborative local model training and global knowledge sharing among heterogeneous participants from multiple sources. FTL mainly includes the local model training, model transmission, and model aggregation. Specifically, firstly, the N source clients $C_s = \{C_s^1, C_s^2, \ldots, C_s^N\}$ that participate in FL conduct the local model training. Secondly, the local models $M_s = \{M_s^1, M_s^2, \ldots, M_s^N\}$ of the N source clients are uploaded to the server. Thirdly, on the server, the weight of each source model is calculated through an aggregation strategy and a weighted aggregation of N source models is performed to obtain a global model with robust adaptability. Finally, the global model is distributed to all clients, the inference evaluation is performed on the target client C_t, and the next epoch of model training continues on each source client. Repeat the aforementioned process until the preset stopping condition is met, ultimately achieving cross-domain knowledge transfer without the need for centralized data.

In FTL, the aggregation of local models is a central mechanism that directly impacts performance on the target domain. The most widely used aggregation strategy is the average aggregation, as implemented in FedAvg (McMahan et al., 2017). The global model M_g is obtained through the weighted average aggregation of local models M_s from N source clients C_s:

$$M_g = \sum_{i=1}^{N} \frac{\varsigma_i}{\sum_{j=1}^{N} \varsigma_j} M_s^i, \tag{1}$$

where ς_i is the number of samples in the ith source client C_s^i. The averaging aggregation strategy is simple and easy to use, and it strikes a good balance between efficiency and generality. However, it treats the local model of each source client in proportion to the number of samples in the client without considering domain differences or model quality, which can dilute the contributions of high-quality models and amplify the negative influences of poorly performing ones, particularly under significant domain shift. The weighted aggregation strategies attempt to assign different weights to source models based on performance metrics, which can partially mitigate the impact of negative transfer, but the performance metrics are usually static and single dimensional. More advanced dynamic aggregation strategies can adjust weights adaptively during training or inferring, but they still face challenges in effectively alleviating negative transfer caused by low-quality source models.

2.2. Prototype contrastive learning

PCL combines prototype learning and contrastive learning (Hu et al., 2024) effectively, aiming to guide the maximization of inter-class distances and the minimization of intra-class distances by class prototypes. In the prototype learning, a dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ containing n samples of the $\mathbb{C}$ classes can be divided into $\mathbb{C}$ clusters in a high-dimensional space, resulting in $\mathbb{C}$ class prototypes $P = \{P_1, P_2, \ldots, P_{\mathbb{C}}\}$. The class prototype is usually represented by the mean vector of the data points in the cluster, which is also called the cluster center. The prototype P_c of the cth class can be calculated as

$$P_c = \frac{1}{N_c} \sum_{i=1}^{N_c} f_\theta(x_c^i), \tag{2}$$

where N_c denotes the number of samples of the cth class, and $f_\theta\left(x_c^i\right)$ represents the feature representation of the ith sample of the cth class obtained by the feature extractor f_θ. The core idea of PCL is as follows. By constructing sample pairs using class prototypes, making the feature representations of positive pairs more similar and negative pairs more dissimilar. Specifically, the samples of the same class combined with the class prototype of that class serve as positive pairs, whereas those combined with the class prototypes of other different classes serve as negative pairs. The loss is defined as

$$\mathcal{L}_{\text{pcl}} = \sum_{c=1}^{\mathbb{C}} \sum_{i=1}^{N_c} -\log \frac{\exp(\text{sim}(f_\theta(x_c^i),\ P_c)/\gamma)}{\sum_{k\in\{1,2,\ldots,\mathbb{C}|k\neq c\}} \exp(\text{sim}(f_\theta(x_c^i),\ P_k)/\gamma)}, \tag{3}$$

where γ denotes the normalized temperature parameter and $\text{sim}(\cdot)$ represents the cosine similarity function.

3. Proposed approach

3.1. Overall framework

Fig. 2 illustrates the overall framework of the proposed PCDFTL approach, mainly comprising the following two stages of the model training and model inferring. In the proposed PCDFTL approach, a domain-to-domain decentralized architecture is adopted instead of the centralized server architecture used in the traditional FL, which enables the direct communication between C_s and C_t, thereby reducing the redundant data transmission. As revealed in Fig. 2, the model M_s^k of the kth source client C_s^k consists of a feature extractor F_k, a projection head Ph_k, and a classifier Cf_k.

In the model training stage, firstly, the local model training is conducted on C_s^k. The total loss $\mathcal{L}_1$ in the training process is defined as

$$\mathcal{L}_1 = \mathcal{L}_{\text{cls}} + \alpha\mathcal{L}_{\text{pdc}}, \tag{4}$$

where $\mathcal{L}_{\text{cls}}$ denotes the classification loss of the kth source domain X_s^k, $\mathcal{L}_{\text{pdc}}$ represents PDC loss obtained with the target prototypes P_t and the projected features H_s^k of X_s^k, and α is the adjustable weight for $\mathcal{L}_{\text{pdc}}$. It should be noted that the parameters of Ph_k and Cf_k in M_s^k are frozen and only the parameters of F_k are optimized in the training process. Specifically, the parameters of F_k are optimized by minimizing the loss $\mathcal{L}_1$ on C_s^k, making F_k capable of extracting domain-invariant features, thereby facilitating the sharing and transfer of the cross-domain knowledge and further enhancing the generalization ability of M_s^k on the target domain X_t. Secondly, after the local model training is completed, the parameters of F_k in M_s^k are uploaded from C_s^k to C_t. The inference is performed on C_t according to the parameters of F_k from C_s^k to obtain the sample features Z_t of X_t, which are sent to C_s^k for fine-tuning M_s^k. When fine-tuning M_s^k on C_s^k, the parameters of F_k in M_s^k are frozen, and only the parameters of Ph_k and Cf_k in M_s^k are optimized. The Z_t and the sample features Z_s^k of X_s^k are mapped to a high-dimensional projection space by Ph_k and the domain alignment

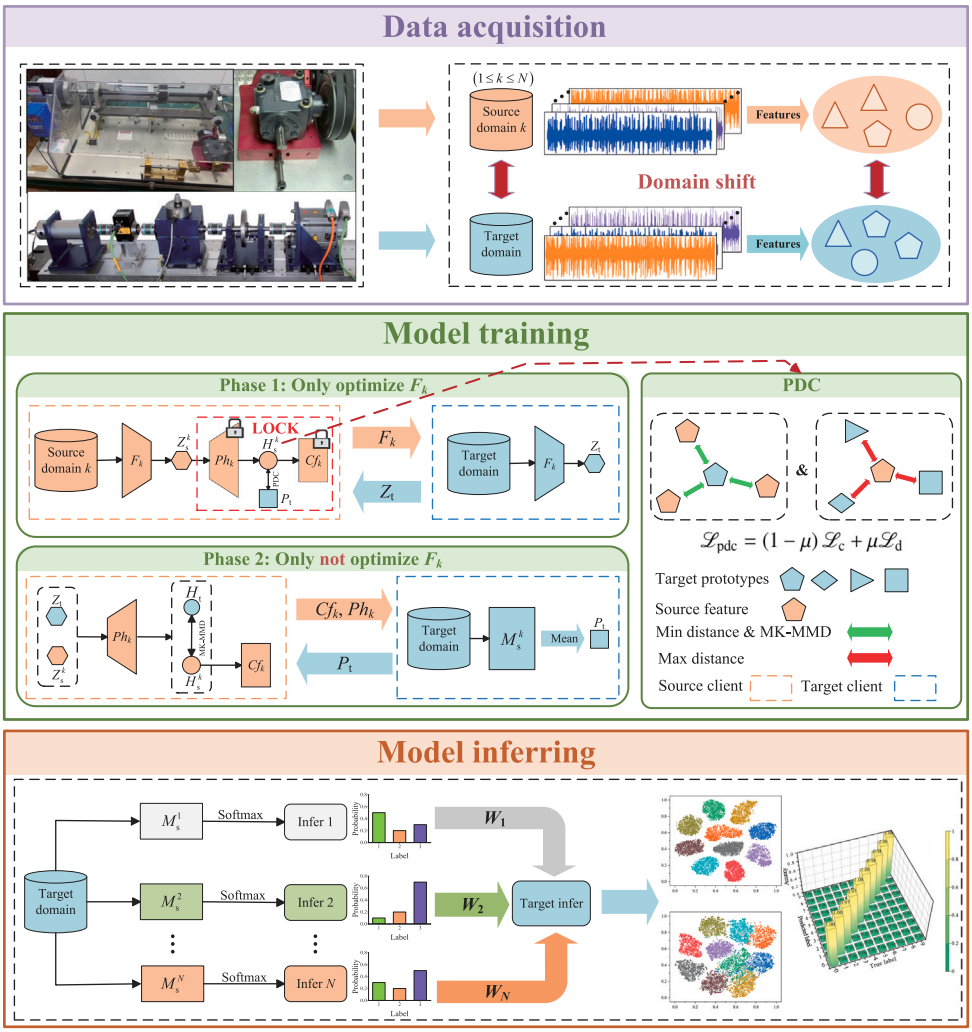


Fig. 2. The overall framework of the proposed PCDFTL.

is performed based on MK-MMD. The overall loss $\mathcal{L}_2$ of the fine-tuning process is defined as

$$\mathcal{L}_2 = \mathcal{L}_{\text{cls}} + \beta\mathcal{L}_{\text{dist}}, \tag{5}$$

where $\mathcal{L}_{\text{dist}}$ denotes MK-MMD loss between H_{s}^k and the projected features H_{t} of X_{t}, and β is the adjustable weight for $\mathcal{L}_{\text{dist}}$. Thirdly, after the M_{s}^k fine-tuning is completed, the updated parameters of Ph_k and Cf_k in M_{s}^k are uploaded from C_{s}^k to C_{t}. The X_{t} is inferred by M_{s}^k on C_{t}, obtaining H_{t} and the inference results of X_{t}. The target prototypes $P_{\text{t}} = \{P_{\text{t}}^1, P_{\text{t}}^2, \ldots, P_{\text{t}}^{\mathbb{C}}\}$ can be calculated as

$$P_{\text{t}}^c = \frac{1}{N_c}\sum_{i=1}^{N_c} H_{\text{t}}^{c_i}, \tag{6}$$

where P_{t}^c denotes the prototype of the cth class in X_{t}, and $H_{\text{t}}^{c_i}$ represents the projected feature of the ith sample of the cth class in X_{t}. Finally, all prototypes in the P_{t} are sent from C_{t} to C_{s}^k for the next epoch of model training.

In the model inferring stage, the M_{s}^k is uploaded from C_{s}^k to C_{t}, where $1 \le k \le N$. The samples of X_{t} are inferred by each source model, and the inference results of all source models are weighted and fused according to the distribution discrepancies between the source and target domains, the stability of DA, and the certainty of the entropy of inference results, thus the optimal inference results of X_{t} are obtained. The process can be described as

$$\hat{y}_t = \sum_{k=1}^{N} W_k \hat{y}_s^k, \tag{7}$$

where $\hat{y}_s^k$ indicates the inference results of M_{s}^k on X_{t}, W_k represents the weight of the inference results of M_{s}^k, $\hat{y}_t$ denotes the final inference results of X_{t}, and N is the number of source clients.

3.2. Dual-phase adapter projection optimization strategy

By reducing the distribution discrepancies between the source and target domains, the impact of domain shift can be effectively alleviated. In the traditional DA, MK-MMD is commonly used as a statistical distance metric to quantify the distribution discrepancies between the source and target domains. The distribution discrepancy $dist_k$ between X_{s}^k and X_{t} can be calculated based on MK-MMD as

$$dist_k = \left\| E_{p(X_{\text{s}}^k)}\left[\phi(F_k(X_{\text{s}}^k))\right] - E_{p(X_{\text{t}})}\left[\phi(F_k(X_{\text{t}}))\right] \right\|_{\mathcal{H}}^2, \tag{8}$$

where $p(X_{\text{s}}^k)$ and $p(X_{\text{t}})$ denote the probability distributions over all fault classes in X_{s}^k and X_{t} respectively, $\mathcal{H}$ is the reproducing kernel Hilbert space, and $\phi(\cdot)$ indicates the kernel function that can map data to a high-dimensional feature space.

Without considering data privacy, a feature extractor can be shared between X_{s}^k and X_{t}. The DA approaches based on MK-MMD can

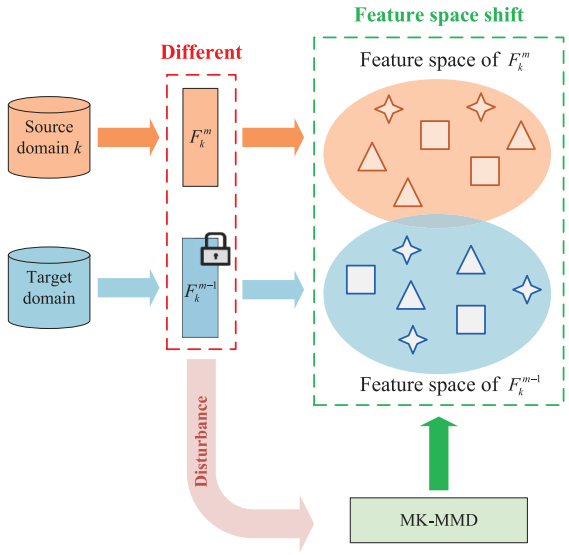


Fig. 3. Illustration of the shift between feature spaces of X_s^k and X_t.

effectively reduce distribution discrepancies between domains, thus achieving better FD results on the transfer tasks. However, when protecting data privacy, the parameters of the feature extractor used by X_s^k and that used by X_t are often inconsistent, leading to a shift between the feature spaces of X_s^k and X_t, as displayed in Fig. 3. F_k^m denotes the feature extractor of C_s^k at the mth epoch, and F_k^{m-1} represents the feature extractor uploaded from C_s^k to C_t at the $(m-1)$-th epoch. During model training, the parameters of F_k^{m-1} are frozen, whereas the parameters of F_k^m keep changing as the optimization progresses at the mth epoch, leading to the feature spaces of F_k^m and F_k^{m-1} are deviated. Under the aforementioned conditions, $dist_k$ can be calculated based on MK-MMD as

$$dist_k = \left\| E_{p(X_s^k)}\left[\phi(F_k^m(X_s^k))\right] - E_{p(X_t)}\left[\phi(F_k^{m-1}(X_t))\right] \right\|_{\mathcal{H}}^2. \tag{9}$$

The differences in the feature spaces of F_k^m and F_k^{m-1} make it difficult to measure $dist_k$, thus making the DA approaches based on MK-MMD less effective in alleviating the impact of domain shift. Therefore, DPAPO strategy is proposed, as shown in Fig. 4. The strategy divides the DA based on MK-MMD into the following two training phases.

In the first training phase, according to Eq. (4), the parameters of F_k are optimized by the losses $\mathcal{L}_{pdc}$ described in Section 3.3 and $\mathcal{L}_{cls}$, minimizing $dist_k$ after Z_s^k and Z_t extracted by F_k pass the projection of Ph_k, the optimized parameters of F_k are uploaded from C_s^k to C_t, and Z_t are obtained by using F_k to extract the features from X_t on C_t.

In the second training phase, Z_s^k are extracted from X_s^k by the frozen F_k on C_s^k and projected into a high-dimensional projection space by the unfrozen Ph_k. The statistical distance $dist_{k_i}$ can be calculated based on MK-MMD as

$$dist_{k_i} = \left\| E_{p(X_s^{k_i})}\left[\phi(Ph_k(F_k(X_s^{k_i})))\right] - E_{p(X_t^i)}\left[\phi(Ph_k(F_k(X_t^i)))\right] \right\|_{\mathcal{H}}^2, \tag{10}$$

where $X_s^{k_i}$ and X_t^i denote the ith sample of X_s^k and that of X_t, respectively. H_s^k are classified by the unfrozen Cf_k to improve the classification performance of M_s^k on X_s^k. The loss $\mathcal{L}_{cls}$ can be expressed as

$$\mathcal{L}_{cls} = -\frac{1}{N_s^k}\sum_{i=1}^{N_s^k} y_s^{k_i} \log\left(Cf_k\left(H_s^{k_i}\right)\right), \tag{11}$$

where N_s^k denotes the number of samples of X_s^k, $H_s^{k_i}$ represents the projected feature of $X_s^{k_i}$, and $y_s^{k_i}$ indicates the true label of $X_s^{k_i}$. The loss $\mathcal{L}_{dist}$ can be defined as

$$\mathcal{L}_{dist} = \frac{1}{N_s^k}\sum_{i=1}^{N_s^k} dist_{k_i}. \tag{12}$$

The total loss is further defined as

$$\mathcal{L}_2 = -\frac{1}{N_s^k}\sum_{i=1}^{N_s^k} y_s^{k_i} \log\left(Cf_k\left(H_s^{k_i}\right)\right) + \beta\left(\frac{1}{N_s^k}\sum_{i=1}^{N_s^k}\left\| E_{p(H_s^{k_i})}\left[\phi(H_s^{k_i})\right] - E_{p(H_t^i)}\left[\phi(H_t^i)\right]\right\|_{\mathcal{H}}^2\right), \tag{13}$$

where H_t^i represents the projected feature of X_t^i. The optimization objective of Cf_k is defined as

$$\hat{\theta}^{Cf_k} = \arg\min_{\theta^{Cf_k}} \mathcal{L}_{cls}, \tag{14}$$

where θ^{Cf_k} denotes the parameters of Cf_k. The optimization objective of Ph_k is defined as

$$\hat{\theta}^{Ph_k} = \arg\min_{\theta^{Ph_k}} \mathcal{L}_2, \tag{15}$$

where θ^{Ph_k} represents the parameters of Ph_k. The M_s^k is uploaded from C_s^k to C_t for the inference of X_t, and the pseudo-label $\tilde{y}_t^i$ is set for X_t^i by

$$\tilde{y}_t^i = \arg\max_{j\in\{1,2,\ldots,\mathbb{C}\}}\left(\frac{e^{z_j}}{\sum_{l=1}^{\mathbb{C}} e^{z_l}}\right), \tag{16}$$

where z_j indicates the output of the jth neuron in the last layer of M_s^k. The P_t can be calculated on C_t using the pseudo-labels $\tilde{y}_t$ of X_t according to Eq. (6), and all prototypes in the P_t are sent from C_t to C_s^k, which will be used in the first training phase of the next epoch.

The proposed DPAPO strategy enables fine-grained domain alignment while mitigating negative transfer by separating the learning of domain-invariant and domain-specific features.

3.3. Prototype-based differentiated contrastive loss

PCL can improve the generalization of the model and reduce the impact of domain shift under cross-domain tasks. However, the traditional federated PCL approaches are based on label-rich clients. When these approaches are used in the unlabeled target client, the statistical distances between the source and target domains cannot be accurately measured solely by the prototype contrastive loss. Therefore, PDC loss is proposed, as revealed in Fig. 5, which consists of the contrastive loss $\mathcal{L}_c$ between H_s^k and P_t and the subdomain alignment loss $\mathcal{L}_d$ based on MK-MMD between H_s^k and P_t.

In the proposed PDC loss, $\mathcal{L}_c$ is defined as

$$\mathcal{L}_c = \sum_{c=1}^{\mathbb{C}}\sum_{i=1}^{N_c} -\log\frac{\exp\left(\mathrm{sim}\left(H_s^{k_{ci}}, P_t^c\right)/\gamma\right)}{\sum_{j\in\{1,2,\cdots,\mathbb{C}|j\neq c\}}\exp\left(\mathrm{sim}\left(H_s^{k_{ci}}, P_t^j\right)/\gamma\right)}, \tag{17}$$

where $H_s^{k_{ci}}$ denotes the projected feature of the ith sample of the cth class of X_s^k. To alleviate the issue of the misalignment between subdomains, the subdomain alignment will be further carried out to minimize the distribution discrepancies between the same fault classes within X_s^k and X_t. The subdomain alignment loss $\mathcal{L}_d$ is defined as

$$\mathcal{L}_d = \frac{1}{\mathbb{C}}\sum_{c=1}^{\mathbb{C}}\left\| E_{p(H_s^{k_c})}\left[\phi\left(H_s^{k_c}\right)\right] - E_{p(P_t^c)}\left[\phi\left(P_t^c\right)\right]\right\|_{\mathcal{H}}^2, \tag{18}$$

where $H_s^{k_c}$ denotes the projected features of all samples of the cth class in X_s^k. The total loss of the proposed PDC is defined as

$$\mathcal{L}_{pdc} = (1-\mu)\mathcal{L}_c + \mu\mathcal{L}_d, \tag{19}$$

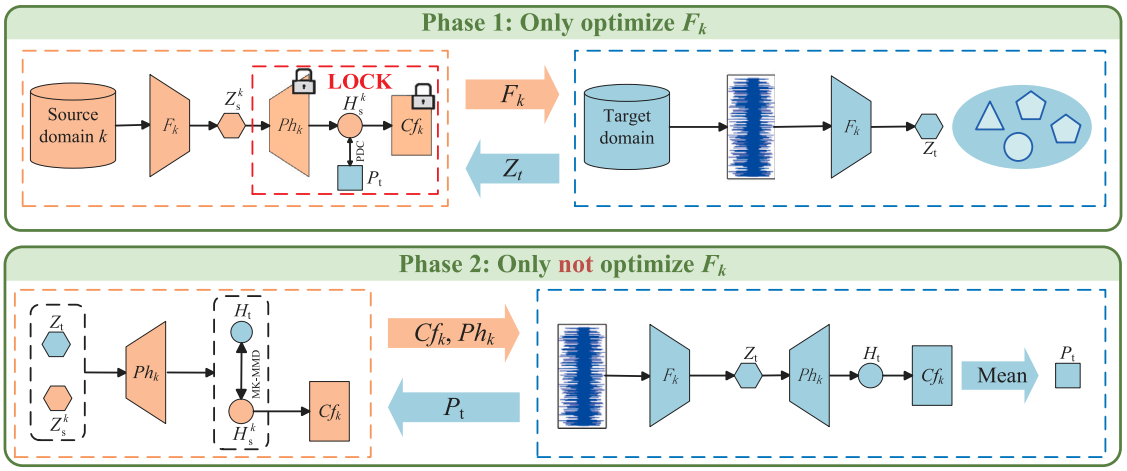


Fig. 4. The proposed DPAPO strategy.

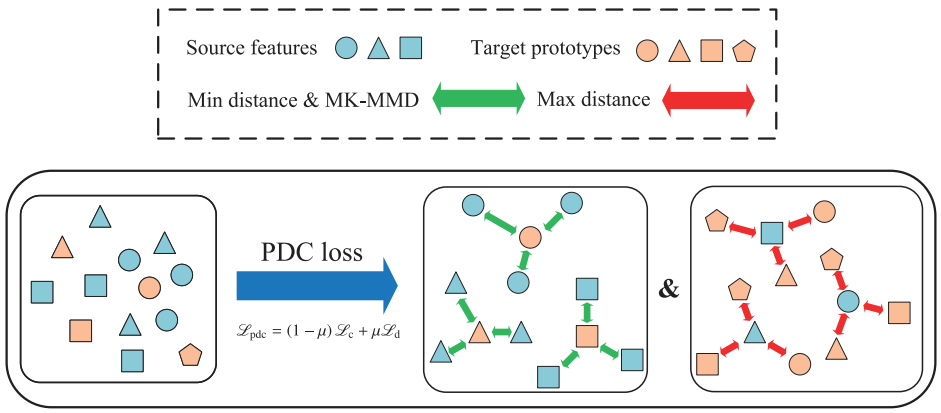


Fig. 5. The proposed PDC loss.

where μ is the adjustable weight and $0 < \mu < 1$. The total loss for the first training phase of DPAPO strategy is further defined as

$$\mathcal{L}_1 = \mathcal{L}_{\text{cls}} + \alpha\left((1-\mu)\mathcal{L}_{\text{c}} + \mu\mathcal{L}_{\text{d}}\right). \tag{20}$$

The optimization objective of F_k in the first training phase of DPAPO strategy is defined as

$$\hat{\theta}^{F_k} = \arg\min_{\theta^{F_k}}\left(\mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{c}} + \mathcal{L}_{\text{d}}\right), \tag{21}$$

where θ^{F_k} denotes the parameters of F_k.

The proposed PDC loss not only aligns the subdomains in the source and target domains at the class level but also mitigates negative transfer by ensuring that irrelevant features do not interfere with the prototype representations.

3.4. Multi-source dynamic weighted inference strategy

In traditional FTL, the model aggregation strategy almost determines the generalization of the global model. If the aggregation weights of the local models cannot accurately measure the distribution discrepancies among different domains, it will lead to the poor generalization of the global model, causing the negative transfer problem and even making the training difficult to converge. Therefore, MSDWI strategy is proposed, as illustrated in Fig. 6.

Different from traditional FTL using a global model to infer on X_{t}, MSDWI strategy can use multiple local models to perform the dynamic weighted fusion inference of X_{t}, thereby avoiding the negative transfer problem. Specifically, MSDWI strategy evaluates the weights of inference results of different source models on X_{t} according to the following three indicators: the statistical distances between the source and target domains, the stability of DA, and the certainty of the entropy of the inference results. The inference weight W_k of M_{s}^k can be formally expressed as

$$W_k = \eta W_N^D + (1-\eta)W_k^E, \tag{22}$$

where W_N^D represents the weight derived from the statistical distances between the source and target domains and the stability of DA process, W_k^E represents the weight derived from the certainty of the entropy of the inference results, and $\eta \in (0,1)$ is an adjustable hyperparameter which is used for controlling the balance between these indicators. The optimal inference results $\hat{y}_{\text{t}}$ of X_{t} are then obtained according to Eq. (7). The following three indicators with corresponding normalization methods are used to compute the inference weights.

The statistical distances between the source and target domains: The statistical distances based on MK-MMD can be used to measure the degree of discrepancies between the source and target domains. The statistical distance $dist_{k_i}^c$ between $H_{\text{s}}^{k_{ci}}$ and P_{t}^c is defined as

$$dist_{k_i}^c = \left\| E_{p(H_{\text{s}}^{k_{ci}})}\left[\phi\left(H_{\text{s}}^{k_{ci}}\right)\right] - E_{p(P_{\text{t}}^c)}\left[\phi\left(P_{\text{t}}^c\right)\right]\right\|_{\mathcal{H}}^2. \tag{23}$$

Typically, the model corresponding to the source domain with less distribution discrepancy from X_{t} has the better inference performance on X_{t}.

The stability of DA process: The stability can also reflect the FD performance of the source model on X_{t}. The variance var_k of the

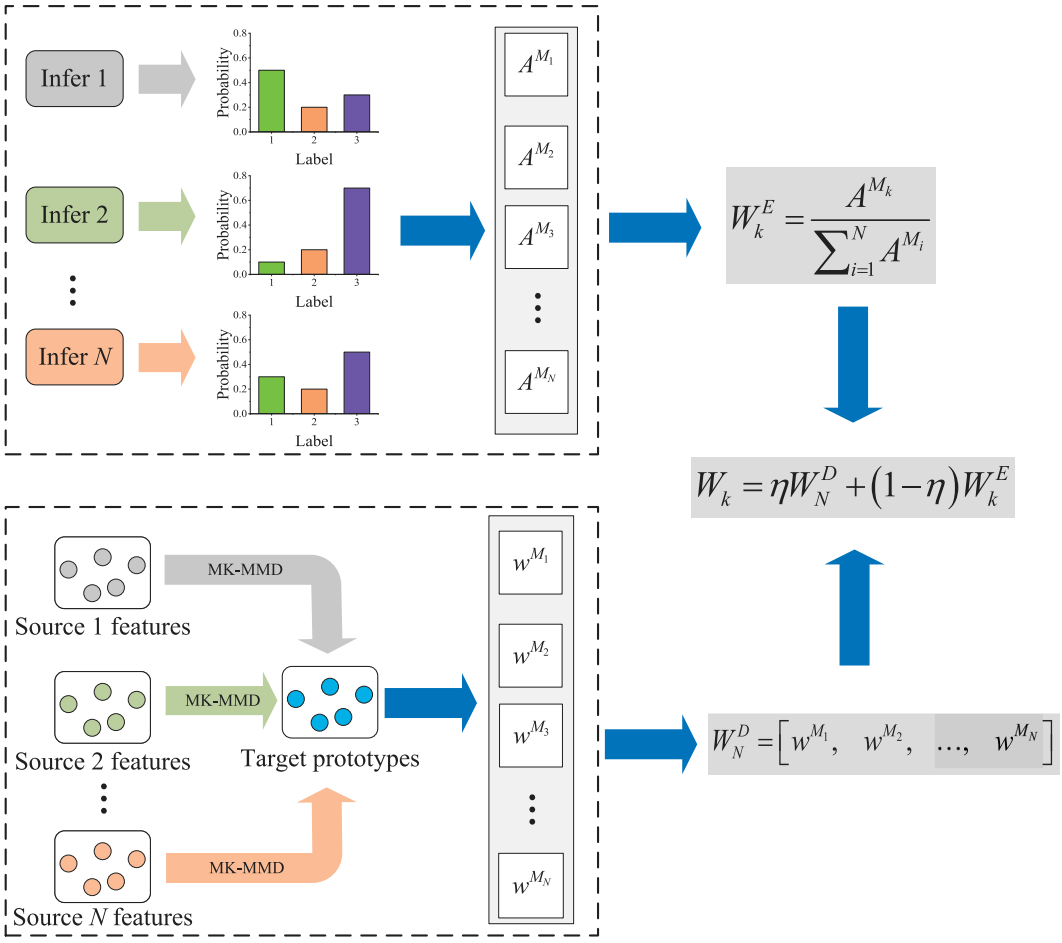


Fig. 6. The proposed MSDWI strategy.

statistical distances between H_s^k and P_t is defined as

$$var_k = \frac{1}{N_s^k} \sum_{c=1}^{\mathbb{C}} \sum_{i=1}^{N_c} \left(dist_{k_i}^c - \frac{\mathbb{C}\mathcal{L}_d}{N_s^k} \right)^2. \tag{24}$$

To avoid the model corresponding to the source domain with larger distribution discrepancy from X_t obtains a larger inference weight, the inference weight includes the variance of the statistical distances. The weight w^{M_k} of M_s^k can be calculated by

$$w^{M_k} = \mathrm{softmax}\left(\frac{\sum_{i=1}^{N} var_i}{\left(b \sum_{c=1}^{\mathbb{C}} \sum_{j=1}^{N_c} dist_{k_j}^c var_k \right) / N_s^k} \right), \tag{25}$$

where b is the smoothing factor and $b > 1$. The weight matrix W_N^D of M_s can be calculated by

$$W_N^D = \left[w^{M_1}, w^{M_2}, \ldots, w^{M_N} \right]. \tag{26}$$

The certainty of the entropy of the inference results: In the classification tasks, the outputs of the model are usually the probability distributions that represent how likely the input data belong to different classes. If the model is highly confident about a particular class, the probability distributions will be concentrated on that class. In this case, the entropy values of the inference results are smaller. Conversely, if the model exhibits almost the same degree of uncertainty for all classes, the probability distributions will approach a uniform distribution, and the entropy values of the inference results will be larger. Therefore, the entropy is selected as an important index to evaluate the certainty of inference results of the model. The entropy $S_i^{M_k}$ of the ith inference result of M_s^k can be calculated by

$$S_i^{M_k} = -\sum_{j=1}^{\mathbb{C}} p\left(j | \hat{y}_s^{k_i} \right) \log\left(p\left(\hat{y}_s^{k_i} \right) \right). \tag{27}$$

When M_s^k infers on X_t, the entropy matrix A^{M_k} of M_s^k can be calculated by

$$A^{M_k} = \left[S_1^{M_k}, S_2^{M_k}, \ldots, S_{N_t}^{M_k} \right], \tag{28}$$

where N_t represents the total number of samples in X_t. The entropy weight matrix W_k^E of M_s^k on X_t can be calculated by

$$W_k^E = \frac{A^{M_k}}{\sum_{i=1}^{N} A^{M_i}}. \tag{29}$$

The proposed MSDWI strategy is revealed in Algorithm 1.

The proposed MSDWI strategy dynamically fuses the inference results of multiple source models on the target domain by assigning weights based on the three aforementioned indicators, thereby mitigating negative transfer and improving cross-domain FD performance.

The comparisons between the proposed PCDFTL and other FTL approaches across multiple dimensions are succinctly illustrated in Table 1. Firstly, PCDFTL employs a decentralized federated architecture to minimize communication overhead. Secondly, unlike traditional FTL approaches that are simply based on MK-MMD or domain adversarial, PCDFTL ingeniously combines MK-MMD, adapter fine-tuning, and contrastive learning to perform domain alignment. Finally, PCDFTL does not rely on the model aggregation in FTL but instead utilizes multiple models for weighted inference on the target domain.

Table 1
Comparisons between the proposed PCDFTL and other FTL approaches.

Approach	Federated architecture	Aggregation strategy	Domain alignment strategy	Mitigate negative transfer strategy
FedAvg (McMahan et al., 2017)	Centralization	Average aggregation	–	–
FTLDAN (Zhang and Li, 2022)	Centralization	Average aggregation	Introducing a deep adversarial network	Introducing consistency constraint of multiple classifiers
FMDAAN-V (Zhao et al., 2023)	Centralization	Average aggregation	Combining MK-MMD and adversarial learning	Fine-tuning the global model based on pseudo-labels
LQKF (Wang et al., 2023)	Centralization	Dynamic aggregation based on contribution degree	Introducing a distance metric based on batch normalization layer statistics	Low-quality knowledge filtering
SPDFT (Zhao et al., 2024a)	Decentralization	Dynamic aggregation based on contribution degree	Combining hash mapping and MK-MMD	Introducing self-paced learning
FTL-A (Li et al., 2024a)	Centralization	Average aggregation	Introducing an alignment loss	Introducing an alignment loss
RDFTL (Wan et al., 2024)	Decentralization	Dynamic aggregation based on multiple indicators	Combining global domain discriminator and subdomain aligner	Suppressing the contribution of low-quality models
AFL-SA (Yuan et al., 2024)	Centralization	Dynamic aggregation based on sample size	Introducing local MMD-based subdomain DA	Introducing an active learning query mechanism
CECFTL (Liang et al., 2025)	Decentralization	Hierarchical federated aggregation	Introducing an adaptive model fine-tuning strategy	Introducing an adaptive model fine-tuning mechanism
Proposed PCDFTL	Decentralization	–	Combining DPAPO strategy and PDC loss	Introducing MSDWI strategy

Algorithm 1 The proposed MSDWI strategy

Input: The source domain datasets $\{D_s^i\}_{i=1}^N$, the target domain dataset D_t, the maximum number of epochs $maxEpochs$, and the number of iterations $iterPhase1$ and $iterPhase2$ in the first and second phases of the DPAPO strategy during each training epoch.
Output: The inference results $\hat{y}_t$ of the target domain.
Initialize M_s^i on C_s^i, where $1 \le i \le N$;
for $e = 1$ **to** $maxEpochs$ **do**
  for C_s^i, $1 \le i \le N$, **in parallel do**
    for $k = 1$ **to** $iterPhase1$ **do**
      if P_t is not empty **then**
        Train M_s^i on C_s^i by optimizing $\mathcal{L}_1$ calculated by Eq. (20) while freezing Ph_i and C_i;
      else
        Train M_s^i on C_s^i only using D_s^i;
      end if
    end for
    Transmit F_i from C_s^i to C_t;
    Extract Z_t by F_i without using gradient on C_t;
    Transmit Z_t from C_t to C_s^i;
    for $j = 1$ **to** $iterPhase2$ **do**
      Train M_s^i on C_s^i by optimizing $\mathcal{L}_2$ calculated by Eq. (13) while freezing F_i;
    end for
    Transmit M_s^i from C_s^i to C_t;
    Extract H_t and perform the inference by M_s^i to get $\tilde{y}_t$ on C_t;
    Calculate P_t by Eq. (6) on C_t;
    Transmit P_t from C_t to C_s^i;
  end for
  Calculate W_N^D by Eq. (26) on C_t;
  Calculate $\{A^{M_i}\}_{i=1}^N$ by Eq. (28) on C_t when using $\{M_s^i\}_{i=1}^N$ to infer D_t;
  Calculate $\{W_i^E\}_{i=1}^N$ by Eq. (29) on C_t;
  Calculate $\{W_i\}_{i=1}^N$ by Eq. (22) on C_t;
  Calculate $\hat{y}_t$ by Eq. (7) on C_t;
end for

4. Experiments

4.1. Experimental setup

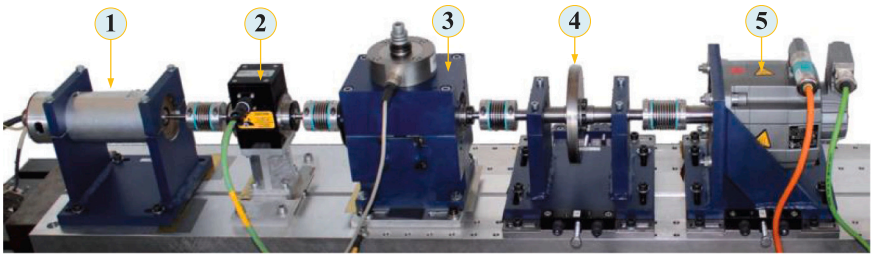
To better evaluate the proposed PCDFTL approach, a series of experiments are carried out on the Paderborn University (PU) bearing dataset (Lessmeier et al., 2016), Huazhong University of Science and Technology (HUST) gearbox dataset (Zhao et al., 2024b), real factory bearing (RFB) dataset (Chen and Xiao, 2024), and Jiangnan University (JNU) bearing dataset (Li et al., 2019), where RFB dataset is collected from the real industrial production line.

Table 2
Operating conditions of PU dataset.

Operating condition	Rotation speed (rpm)	Load torque (Nm)	Radial force (N)
P1	1500	0.7	1000
P2	900	0.7	1000
P3	1500	0.1	1000
P4	1500	0.7	400

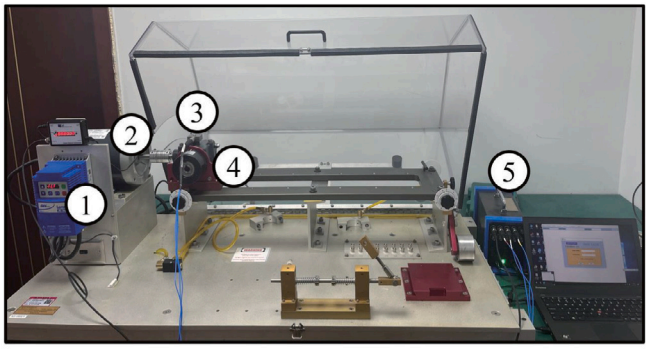
Fig. 7 gives the PU bearing test platform. In PU dataset, the vibration signals of the following ten different health states collected at 64 kHz sampling frequency under different operating conditions listed in Table 2 are adopted: normal (N), three inner-race faults (i.e., IF1, IF2, and IF3), three outer-race faults (i.e., OF1, OF2, and OF3), and three compound faults (i.e., IF+OF1, IF+OF2, and IF+OF3), whose class labels are 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively. Fig. 8 shows the HUST gearbox test platform. In HUST dataset, the vibration signals of the following three different health states collected at 25.6 kHz sampling frequency under different operating conditions listed in Table 3 are adopted: healthy, broken tooth, and missing tooth. Fig. 9 displays the RFB test platform. In RFB dataset, the vibration signals are sampled at a frequency of 20.48 kHz from naturally degraded bearings in the actual production line under different operating conditions listed in Table 4 and cover four different health states: healthy, rolling-element deflection, rolling-element missing, and yarn stuck. Fig. 10 presents the JNU bearing test platform. In JNU dataset, the vibration signals of the following four different health states collected at 50 Hz sampling frequency under different operating conditions listed in Table 5 are selected: healthy, IF, OF, and roller fault. In this experiment, the data in each operating condition from PU, HUST, RFB, and JNU datasets are overlapped, whose sliding window lengths are 2048, 2048, 1024, and 1024 sample points, respectively, and the step sizes are 128, 1024, 256, and 128 sample points, respectively. The training and test sets are split in an 8:2 ratio for all datasets.

The experiment platform consists of three source clients and one target client, where the source domains are labeled and the target domain is unlabeled. Moreover, the cross-operating condition FD tasks are designed on PU, HUST, RFB, and JNU datasets, respectively, as



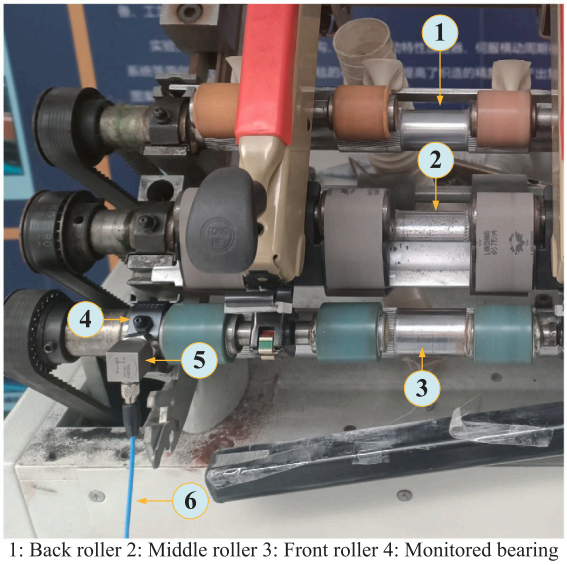
1: Test motor 2: Torque measuring shaft 3: Bearing module 4: Flywheel 5 : Load motor

Fig. 7. PU bearing test platform (Lessmeier et al., 2016).



1: Speed control 2: Motor 3: Acceleration sensor 4: Gearbox 5: Data acquisition board

Fig. 8. HUST gearbox test platform (Zhao et al., 2024b).



1: Back roller 2: Middle roller 3: Front roller 4: Monitored bearing
5: Acceleration sensor 6: Sensor cable

Fig. 9. RFB test platform (Chen and Xiao, 2024).

Table 3
Operating conditions of HUST dataset.

Operating condition	Operating speed (Hz)	Operating load (Nm)
H1	20	0
H2	25	0.113
H3	30	0.226
H4	35	0.339

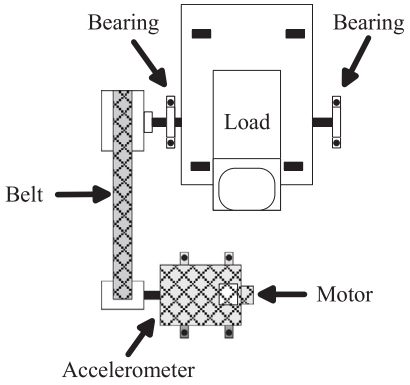


Fig. 10. JNU bearing test platform (Li et al., 2019).

Table 4
Operating conditions of RFB dataset.

Operating condition	Rotation speed (rpm)
R1	100
R2	200
R3	300
R4	350

Table 5
Operating conditions of JNU dataset.

Operating condition	Rotation speed (rpm)
J1	600
J2	800
J3	1000

Table 6
Tasks of cross-operating condition FD.

Dataset	Task	Source client 1	Source client 2	Source client 3	Target client
PU	T1	P2	P3	P4	P1
	T2	P1	P3	P4	P2
	T3	P1	P2	P4	P3
	T4	P1	P2	P3	P4
HUST	T5	H2	H3	H4	H1
	T6	H1	H3	H4	H2
	T7	H1	H2	H4	H3
	T8	H1	H2	H3	H4
RFB	T9	R2	R3	R4	R1
	T10	R1	R3	R4	R2
	T11	R1	R2	R4	R3
	T12	R1	R2	R3	R4
JNU	T13	J2	J3	–	J1
	T14	J3	J1	–	J2
	T15	J1	J2	–	J3

presented in Table 6. Each client is configured with an Intel i5-13400F CPU, 16 GB of RAM, an NVIDIA RTX 4060 Ti GPU with 8 GB of GPU memory, PyTorch 2.6.0, and Ubuntu 18.04 LTS.

Table 7 gives the network structure of the source model in the proposed PCDFTL, which mainly contains the convolution (Conv) layer,

Table 7
Network structure of the source model in the proposed PCDFTL.

Module	Layer	Kernel/Stride/Ouput channel
Feature extractor	Conv	7×7 / 2 / 64
	ReLU	–
	Max-pooling	3×3 / 2 / 64
	ResBlk 1	$\begin{bmatrix}3\times3 / 1 / 64\\3\times3 / 1 / 64\end{bmatrix}\times2$
	ResBlk 2	$\begin{bmatrix}3\times3 / 2 / 128\\3\times3 / 1 / 128\end{bmatrix}$ $\begin{bmatrix}3\times3 / 1 / 128\\3\times3 / 1 / 128\end{bmatrix}$
Projection head	ResBlk 1	$\begin{bmatrix}3\times3 / 2 / 256\\3\times3 / 1 / 256\end{bmatrix}\times2$
	ResBlk 2	$\begin{bmatrix}3\times3 / 2 / 512\\3\times3 / 1 / 512\end{bmatrix}$ $\begin{bmatrix}3\times3 / 1 / 512\\3\times3 / 1 / 512\end{bmatrix}$
	Avg-pooling	7×7 / 7 / 512
	Flatten	–
Classifier	FC1	–
	FC2	–

Table 8
Hyper-parameter settings adopted in the source model training.

Name	Value
$maxEpochs$	50
$iterPhase1$	1
$iterPhase2$	2
$batchSize$	128
$learningRate$	0.001
α	1.0
β	0.5
γ	0.1
μ	0.5
b	2
η	0.9

ReLU activation function, max-pooling layer, avg-pooling layer, residual blocks (ResBlks), and fully connected (FC) layers. Table 8 lists the hyper-parameter settings adopted in the training of the source model.

The detailed component configurations of the proposed PCDFTL and its seven variants are presented in Table 9, where the symbols ✓and ✕ indicate that the component is included and that has been removed, respectively.

4.2. Validation of the effectiveness of the proposed DPAPO strategy

To verify the effectiveness of the projection optimization and dual-phase adapter optimization in the proposed DPAPO strategy, the comparative experiments with TM1 and TM2, two variants of the proposed PCDFTL approach, are carried out on the tasks of different cross-operating condition FD. As shown in Table 9, in TM1, the proposed DPAPO strategy is not adopted, whereas performing directly the domain alignment of the sample features. In TM2, only the projection optimization in DPAPO strategy is used.

Table 10 gives the accuracies gained using different domain alignment strategies on the different tasks of cross-operating condition FD. As observed from Table 10, the accuracies achieved by TM1 on all cross-operating condition FD tasks of PU and RFB datasets are significantly lower than those of TM2 and the proposed PCDFTL. Especially on the tasks of T2 and T9, TM1 only gains an FD accuracies of 41.56% and 57.48%, respectively. There are the following two reasons. Firstly, because the large data distribution discrepancies of PU and RFB datasets, TM1 directly performs domain alignment without first using the projection head to align the feature space, resulting in significant changes in the parameters of the feature extractor after one training epoch. Due to the source and target domains cannot be performed domain alignment in the unified feature space, making it difficult to measure the statistical distances between the source and target domains. Therefore, TM1 achieves significantly lower accuracies than TM2 and PCDFTL on the cross-operating condition FD tasks of PU and RFB datasets. Secondly, on the cross-operating condition FD tasks of PU and RFB datasets, the distribution discrepancies between the source and target domains of the task T2 are larger than those of the other tasks, thus the FD accuracy of TM1 on the task T2 is significantly lower than those of TM2 and PCDFTL. Different from PU and RFB datasets, the accuracies obtained by TM1 on the cross-operating condition FD tasks of HUST and JNU datasets are only about 1% lower than those obtained by TM2 and PCDFTL, whose main reason is that the distribution discrepancies between different domains of HUST and JNU datasets are smaller than those of PU and RFB datasets. As a result, even if TM1 does not use the projection head to align the feature space, the parameters of the feature extractor do not change much after one training epoch, and the source and target domains can be performed domain alignment in the unified feature space, thereby effectively performing DA.

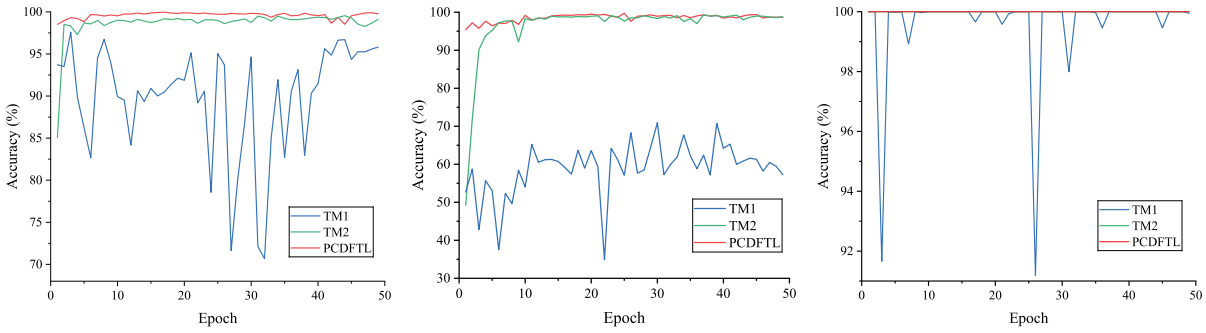
As also can be seen from Table 10, PCDFTL obtains higher FD accuracies than TM2 on all tasks. The vital reason is that the dual-phase adapter optimization is added to PCDFTL compared with TM2. In the first phase, by freezing the projection head and classifier, the target prototypes inferred by the source models of the previous training epoch can guide the source models for domain alignment within the unified feature space, effectively enhancing the ability of the feature extractor to extract domain-invariant features. In the second phase, unlike TM2 that directly performs projection alignment for the feature spaces of different feature extractors, PCDFTL first freezes the feature extractors and then performs the projection alignment of the feature spaces, which ensures that both the source and target domains use the feature extractors with the same parameters to extract features, thereby enabling the projection head to find a more suitable projection space. The projection space can make MK-MMD measure the statistical distances between the source and target domains more accurately, effectively improving the FD ability of the source models on the target domain.

Taking the tasks T1, T4, and T8 as examples, the accuracy changes of TM1, TM2, and PCDFTL in the model training are further analyzed. Figs. 11(a), 11(b), and 11(c) illustrate the changes in accuracies of TM1, TM2, and PCDFTL on the tasks of T1, T4, and T8, respectively. As displayed in Figs. 11(a) and 11(b), on the tasks of T1 and T4, PCDFTL converges fastest. Both TM2 and PCDFTL exhibit smaller training fluctuations compared to TM1, and their final FD accuracies exceed 99%. TM2 slightly inferior in the final accuracy and convergence speed to PCDFTL. TM1 has the lowest final FD accuracy, the slowest convergence speed, and significant training fluctuations. As depicted in Fig. 11(c), on the task T8, both TM2 and PCDFTL gain a final FD accuracy of 100% with smaller training fluctuations compared to TM1. Although TM1 also reaches a final FD accuracy of 100%, it shows larger training fluctuations than TM2 and PCDFTL. The above results demonstrate that whether facing the cross-operating condition FD tasks with smaller data distribution discrepancies or those with larger ones, the projection optimization in the proposed DPAPO strategy fully addresses the feature space shift issue during domain alignment while protecting data privacy. It effectively aligns the feature spaces of the source and target domains, significantly enhancing the performance of DA based on MK-MMD. Moreover, the dual-phase adapter optimization in DPAPO strategy can thoroughly align the feature spaces of the source and target domains, effectively strengthening the stability of the model training.

The average computation and communication times spent on different tasks of different datasets during model training with DPAPO strategy are depicted in Fig. 12. As shown in Fig. 12, the communication overhead is much lower than the computation overhead in the proposed DPAPO strategy, indicating that DPAPO has high communication efficiency in the FTL scenario. It can be seen that while maintaining model performance, DPAPO can better support parallel training and improve the resource utilization, thus facilitating practical deployment.

Table 9
Component configurations of the proposed PCDFTL and its variants adopted in the ablation experiments.

Approach	DPAPO strategy		PDC loss		MSDWI strategy		
	Adapter optimization	Projection optimization	MK-MMD	Prototype aggregation	Statistical distances	Stability of DA	Certainty of the entropy
PCDFTL	✓	✓	✓	Average	✓	✓	✓
TM1	×	×	✓	Average	✓	✓	✓
TM2	×	✓	✓	Average	✓	✓	✓
PM1	✓	✓	×	×	✓	✓	✓
PM2	✓	✓	✓	Cosine similarity	✓	✓	✓
M1	✓	✓	✓	Average	×	×	×
M2	✓	✓	✓	Average	✓	×	×
M3	✓	✓	✓	Average	✓	✓	×



(a) Task T1 (b) Task T4 (c) Task T8

Fig. 11. Accuracies obtained by different domain alignment strategies on different tasks.

Table 10
Accuracies (%) obtained by adopting different domain alignment strategies on the different tasks of cross-operating condition FD.

Dataset	Task	TM1	TM2	PCDFTL
PU	T1	97.23 ± 0.37	99.55 ± 0.01	99.66 ± 0.11
	T2	41.56 ± 0.21	94.87 ± 0.53	95.30 ± 0.15
	T3	95.67 ± 0.28	99.70 ± 0.03	99.97 ± 0.02
	T4	70.73 ± 0.20	99.32 ± 0.01	99.32 ± 0.02
	Avg.	76.30	98.36	98.56
HUST	T5	99.50 ± 0.40	100.00 ± 0.00	100.00 ± 0.00
	T6	99.40 ± 0.50	99.83 ± 0.17	100.00 ± 0.00
	T7	99.28 ± 0.52	99.77 ± 0.23	100.00 ± 0.00
	T8	99.17 ± 0.83	100.00 ± 0.00	100.00 ± 0.00
	Avg.	99.34	99.90	100.00
RFB	T9	57.48 ± 0.40	93.61 ± 0.96	96.12 ± 0.86
	T10	79.40 ± 0.50	98.33 ± 0.17	99.74 ± 0.11
	T11	72.74 ± 0.42	98.17 ± 0.24	99.86 ± 0.14
	T12	85.44 ± 0.71	98.54 ± 0.32	99.16 ± 0.28
	Avg.	73.77	97.16	98.72
JNU	T13	99.50 ± 0.45	99.84 ± 0.15	99.99 ± 0.01
	T14	99.44 ± 0.51	99.82 ± 0.12	99.98 ± 0.02
	T15	99.55 ± 0.24	99.83 ± 0.12	99.99 ± 0.01
	Avg.	99.50	99.83	99.99
	Total avg.	87.23	98.81	99.32

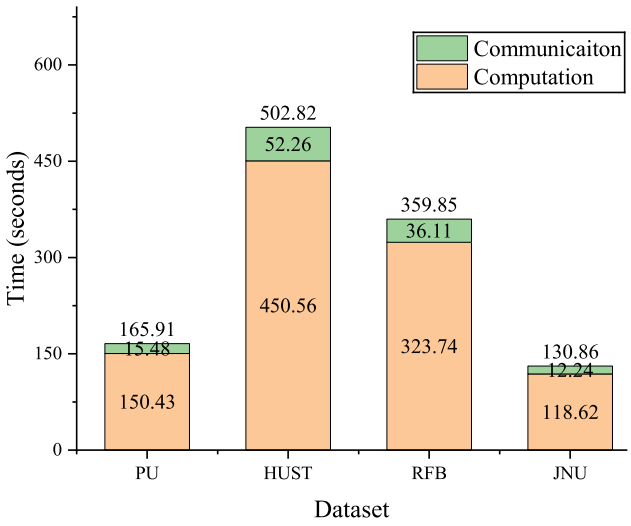


Fig. 12. Average computation and communication times spent on different tasks of different datasets during model training with DPAPO strategy.

4.3. Validation of the effectiveness of the proposed PDC loss

To verify the effectiveness of the proposed PDC loss, the comparative experiments are carried out on different cross-operating condition FD tasks with the two variants of the proposed PCDFTL approach, namely PM1 and PM2. As listed in Table 9, in PM1, the PDC loss is not adopted. In PM2, MK-MMD and cosine similarity-based prototype aggregation are adopted in PDC loss. The accuracies obtained by adopting different domain alignment losses on the different tasks of cross-operating condition FD are listed in Table 11. As shown in Table 11, PM2 achieves higher average FD accuracies than PM1 on different tasks. This is because PM1 cannot effectively mitigate the impact of domain shift, whereas PM2 alleviates the impact of domain shift by combining cosine similarity-based prototype aggregation with MK-MMD to some extent. In contrast, PCDFTL achieves higher average FD accuracies than PM2 on different tasks. The key reason is that there is a significant directional bias in the local prototypes generated by each client in FTL due to the data distribution discrepancies between clients. The cosine similarity-based prototype aggregation used in PM2 amplifies this bias, causing model training to fall into local optimum. However, the average aggregation adopted in PCDFTL can effectively smooth out the specific noise between clients, thereby approaching the true class center in a more robust way and achieving better performance, especially in RFB dataset with high noises.

Table 11
Accuracies (%) obtained by adopting different domain alignment losses on the different tasks of cross-operating condition FD.

Dataset	Task	PM1	PM2	PCDFTL
PU	T1	88.16 ± 0.57	95.95 ± 0.15	99.66 ± 0.11
	T2	70.25 ± 0.95	90.59 ± 0.50	95.30 ± 0.15
	T3	84.61 ± 0.96	94.79 ± 0.13	99.97 ± 0.02
	T4	80.92 ± 0.48	94.32 ± 0.11	99.32 ± 0.02
	Avg.	80.99	93.91	98.56
HUST	T5	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	T6	99.01 ± 0.31	99.33 ± 0.67	100.00 ± 0.00
	T7	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	T8	99.99 ± 0.01	100.00 ± 0.00	100.00 ± 0.00
	Avg.	99.75	99.83	100.00
RFB	T9	71.94 ± 0.95	81.54 ± 0.20	96.12 ± 0.86
	T10	79.88 ± 0.19	85.14 ± 0.17	99.74 ± 0.11
	T11	87.99 ± 0.92	90.38 ± 0.24	99.86 ± 0.14
	T12	85.44 ± 0.84	91.91 ± 0.32	99.16 ± 0.28
	Avg.	81.31	87.24	98.72
JNU	T13	96.19 ± 0.04	98.93 ± 0.05	99.99 ± 0.01
	T14	97.74 ± 0.10	98.44 ± 0.20	99.98 ± 0.02
	T15	95.77 ± 0.54	97.82 ± 0.31	99.99 ± 0.01
	Avg.	96.57	98.40	99.99
	Total avg.	89.66	94.85	99.32

To further verify the robustness of the pseudo-label-based prototype generation mechanism adopted in the proposed PDC loss, an additional ablation experiment is conducted by introducing different levels of additive white Gaussian noises into the vibration signals. Specifically, 25%, 50%, 75%, and 100% noise intensities are added into the vibration signals to simulate varying degrees of label uncertainty and signal degradation. Table 12 presents accuracies of PCDFTL under different pseudo-label noise levels on the different tasks of cross-operating condition FD. As observed in Table 12, the pseudo-label-based prototype generation mechanism used in PDC loss demonstrates strong robustness under varying levels of pseudo-label noises. Although the accuracies slightly decrease with increasing noise, the overall performance remains high across different tasks of different datasets. For instance, on PU dataset, the average accuracy of PCDFTL shows a moderate decline from 98.56% to 97.49%. On HUST dataset, it maintains 100.00%. On JNU dataset, it drops marginally from 99.99% to 99.64%. On RFB dataset, it exhibits a more noticeable decrease from 98.72% to 96.76%, which can be attributed to the fact that RFB dataset is collected from naturally degraded bearings in a real production line, making it more susceptible to signal degradation and environmental interference. Nevertheless, even under severe noise conditions, PCDFTL maintains high FD accuracies across different tasks of different datasets, further confirming that PCDFTL effectively generates reliable prototypes from noisy pseudo-labels for PDC loss and ensures stable FD performance under cross-operating conditions.

4.4. Validation of the effectiveness of the proposed MSDWI strategy

To verify the effectiveness of three key indicators to evaluate the inference weights in the proposed MSDWI strategy, including the statistical distances between the source and target domains, the stability of DA, and the certainty of the entropy of the inference results, the comparative experiments are carried out on different cross-operating condition FD tasks with the three variants of the proposed PCDFTL approach, namely M1, M2, and M3. As listed in Table 9, in M1, the mean weighting is adopted instead of MSDWI strategy, where each source model has the same inference weight. In M2, the inference results of the source models are weighted and fused only according to the statistical distances between the source and target domains. In M3, the inference results of the source models are weighted and fused according to the statistical distances between the source and target domains and the stability of DA.

Table 13 presents the accuracies obtained by different inference result aggregation strategies on the cross-operating condition FD tasks. As observed in Table 13, M2, M3, and PCDFTL achieve higher average accuracies on the cross-operating condition FD tasks across different datasets than M1, respectively. The primary reason is that M1 uses the mean weighting to aggregate the inference results of different source models, which fails to adequately reflect the distribution discrepancies between the source and target domains, thereby declining the accuracy of the final inference results. The average accuracies of M3 and PCDFTL are higher than those of M2. The main reason is that although M2 weights and fuses the inference results of the source models according to the distribution discrepancies between the source and target domains, the training fluctuations during DA can easily lead M2 to assign larger inference weights to the models of the source domains with excessively large distribution discrepancies from the target domain. This ultimately degrades the inference results. PCDFTL outperforms M3 on all tasks. This is primarily because PCDFTL not only takes into account the distribution discrepancies between the source and target domains and the stability of DA, but also focuses on the certainty of the entropy of the inference results. This effectively prevents the low-certainty inference results from occupying the larger weights in the final inference results, thereby substantially enhancing the accuracy of the final inference results.

As also indicated in Table 13, the accuracy gaps among M1, M2, M3, and PCDFTL on the cross-operating condition FD tasks of PU dataset are larger than those on HUST and JNU datasets. Because when confronted with the cross-operating condition FD tasks featuring larger data distribution discrepancies, the FD performance of different source models on the target domain varies greatly. In such cases, the inappropriate inference result aggregation weights can lead to subpar FD performance on the target domain. In contrast, when dealing with the cross-operating condition FD tasks featuring smaller data distribution discrepancies, the FD performance of all source models on the target domain is relatively excellent, rendering the impact of inference result aggregation weights on FD accuracy less pronounced.

The changes in accuracies achieved by different inference result aggregation strategies on the tasks of T2 and T3 are further analyzed. As depicted in Fig. 13(a) and 13(b), PCDFTL attains the highest accuracies with the smallest training fluctuations. M3 gets higher accuracies than M2 and M1, coupled with relatively smaller training fluctuations. M2 outperforms M1 in accuracies and exhibits smaller training fluctuations. Although M1 can reach a high level of accuracy, it experiences the most significant training fluctuations. The results reveal that when facing with the cross-operating condition FD tasks characterized by substantial data distribution discrepancies, MSDWI strategy can precisely determine the aggregation weights of the inference results from different source models. This effectively enhances the FD accuracies of the source models on the target domain and boosts the stability of the model training.

The average computation and communication times spent on different tasks of different datasets during model inferring with MSDWI strategy are depicted in Fig. 14. It can be seen from Fig. 14 that the proposed MSDWI strategy incurs almost no communication overhead. This is primarily because MSDWI strategy is centered on the model inference, with the sole communication overhead arising from the final transmissions of the source models. This indicates that MSDWI strategy is highly suitable for the FTL scenario.

4.5. Comparison with other FD approaches

The comparative experiments are conducted on the proposed PCDFTL with MSSA (Tian et al., 2022), FMDAAN-V (Zhao et al., 2023), LQKF (Wang et al., 2023), SPDFT (Zhao et al., 2024a), FTL-A (Li et al., 2024a), RDFTL (Wan et al., 2024), AFL-SA (Yuan et al., 2024), and CECFTL (Liang et al., 2025) to further verify the effectiveness of PCDFTL, where MSSA is an advanced TL approach for RMFD, whereas

Table 12
Accuracies (%) of PCDFTL under different pseudo-label noise levels on the different tasks of cross-operating condition FD.

Dataset	Task	0% noise	25% noise	50% noise	75% noise	100% noise
PU	T1	99.66 ± 0.11	99.60 ± 0.11	99.53 ± 0.12	99.27 ± 0.21	99.06 ± 0.42
	T2	95.30 ± 0.15	94.89 ± 0.63	94.37 ± 0.17	93.85 ± 0.25	92.96 ± 0.52
	T3	99.97 ± 0.02	99.77 ± 0.13	99.51 ± 0.39	99.38 ± 0.17	99.19 ± 0.37
	T4	99.32 ± 0.02	99.24 ± 0.05	99.07 ± 0.09	98.97 ± 0.14	98.74 ± 0.28
	Avg.	98.56	98.38	98.12	97.87	97.49
HUST	T5	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	T6	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	T7	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	T8	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	Avg.	100.00	100.00	100.00	100.00	100.00
RFB	T9	96.12 ± 0.86	95.98 ± 0.12	95.36 ± 0.68	93.48 ± 0.67	91.89 ± 0.78
	T10	99.74 ± 0.11	99.52 ± 0.10	99.15 ± 0.09	98.89 ± 0.28	98.67 ± 0.18
	T11	99.86 ± 0.14	99.67 ± 0.14	99.39 ± 0.12	99.17 ± 0.10	98.77 ± 0.11
	T12	99.16 ± 0.28	99.05 ± 0.17	98.87 ± 0.11	98.53 ± 0.27	97.71 ± 0.15
	Avg.	98.72	98.55	98.69	97.52	96.76
JNU	T13	99.99 ± 0.01	99.99 ± 0.01	99.98 ± 0.02	99.72 ± 0.17	99.68 ± 0.12
	T14	99.98 ± 0.02	99.98 ± 0.02	99.97 ± 0.03	99.82 ± 0.04	99.66 ± 0.33
	T15	99.99 ± 0.01	99.99 ± 0.01	99.98 ± 0.02	99.77 ± 0.13	99.59 ± 0.39
	Avg.	99.99	99.99	99.98	99.77	99.64
	Total avg.	99.32	99.23	99.20	98.79	98.47

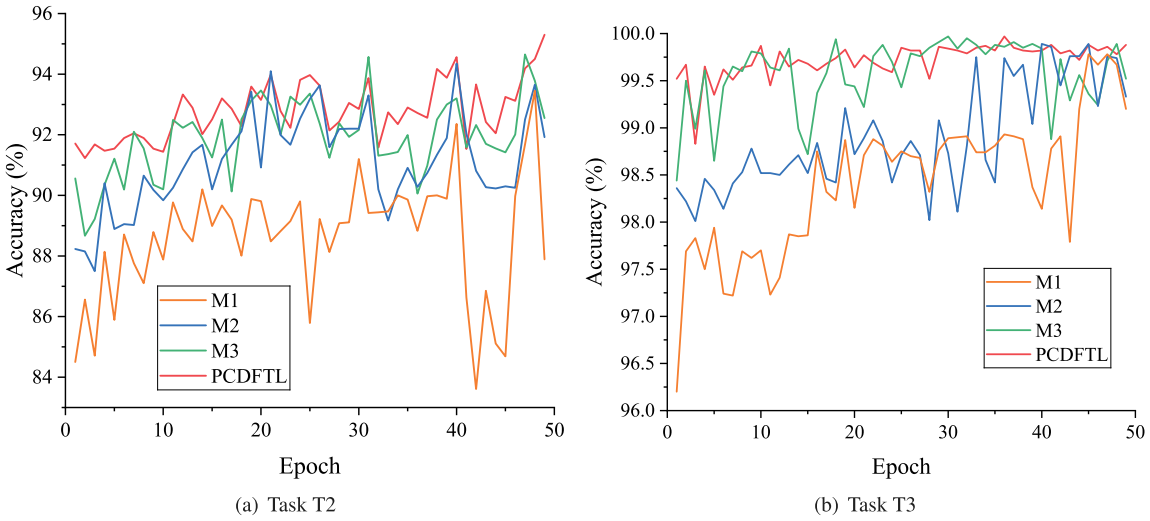


Fig. 13. Accuracies realized by different inference result aggregation strategies on different tasks.

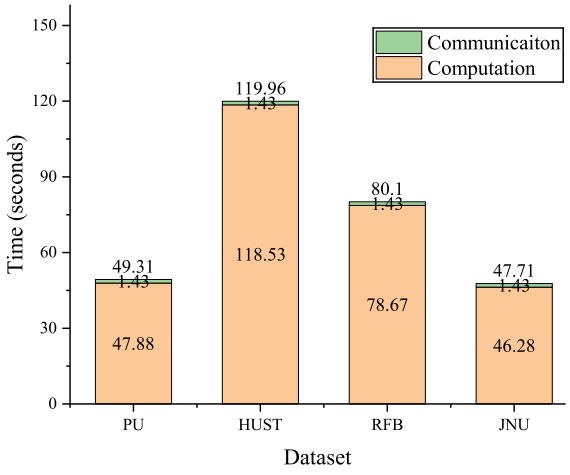


Fig. 14. Average computation and communication times spent on different tasks of different datasets during model inferring with MSDWI strategy.

FMDAAN-V, LQKF, SPDFT, FTL-A, RDFTL, AFL-SA, and CECFTL are advanced FTL approaches for RMFD. MSSA (Tian et al., 2022) is a non-federated multi-source DA approach that can improve the ability of the model to extract domain-invariant features by sharing feature extraction layers and mitigate the impact of domain shift using local MMD. FMDAAN-V (Zhao et al., 2023) addresses the negative transfer issue during feature alignment with a feature discriminator and fine-tunes the global model using the shared knowledge of the source models on the target domain, thereby improving the generalization of the global model. LQKF (Wang et al., 2023) utilizes a low-quality knowledge filter to select the high-confidence pseudo-labeled datasets on the target domain and employs a dynamic model aggregation strategy based on the distribution discrepancy. SPDFT (Zhao et al., 2024a) trains a high-quality auxiliary model based on contrastive self-supervised learning to enhance the FD performance of the global model. FTL-A (Li et al., 2024a) uses the mean square loss in the classification layer to measure the domain distribution discrepancies to alleviate the impact of domain shift. RDFTL (Wan et al., 2024) adopts a multi-perspective aggregation strategy to alleviate the negative transfer problem to improve FD performance of the global model on the target domain. AFL-SA (Yuan et al., 2024) employs an improved Transformer, active learning, and subdomain adaptation to address the compound

Table 13
Accuracies (%) attained by different inference result aggregation strategies on different tasks of cross-operating condition FD.

Dataset	Task	M1	M2	M3	PCDFTL
PU	T1	99.28 ± 0.50	99.32 ± 0.41	99.40 ± 0.31	99.66 ± 0.11
	T2	93.05 ± 0.40	94.12 ± 0.34	94.25 ± 0.32	95.30 ± 0.15
	T3	98.21 ± 0.12	99.51 ± 0.02	99.96 ± 0.02	99.97 ± 0.02
	T4	89.06 ± 0.31	98.10 ± 0.12	98.46 ± 0.04	99.32 ± 0.02
	Avg.	94.90	97.76	98.02	98.56
HUST	T5	99.77 ± 0.23	99.84 ± 0.16	99.92 ± 0.08	100.00 ± 0.00
	T6	99.73 ± 0.27	99.80 ± 0.20	99.89 ± 0.11	100.00 ± 0.00
	T7	99.80 ± 0.17	99.85 ± 0.15	99.96 ± 0.04	100.00 ± 0.00
	T8	99.77 ± 0.23	99.83 ± 0.17	99.91 ± 0.09	100.00 ± 0.00
	Avg.	99.77	99.83	99.92	100.00
RFB	T9	70.30 ± 0.31	88.97 ± 0.57	91.70 ± 0.13	96.12 ± 0.86
	T10	97.15 ± 0.17	98.84 ± 0.73	99.11 ± 0.43	99.74 ± 0.11
	T11	96.92 ± 0.64	97.93 ± 0.25	99.09 ± 0.26	99.86 ± 0.14
	T12	94.89 ± 0.55	96.75 ± 0.36	98.48 ± 0.57	99.16 ± 0.28
	Avg.	89.82	95.63	97.10	98.72
JNU	T13	99.98 ± 0.02	99.99 ± 0.01	99.99 ± 0.01	99.99 ± 0.01
	T14	99.54 ± 0.46	99.71 ± 0.07	99.90 ± 0.10	99.98 ± 0.02
	T15	99.84 ± 0.12	99.85 ± 0.03	99.94 ± 0.06	99.99 ± 0.01
	Avg.	99.79	99.85	99.94	99.99
	Total avg.	96.07	98.27	98.75	99.32

challenges of data silos, poor labeling, and distribution discrepancies to boost the performance of FTL-based RMFD. CECFTL (Liang et al., 2025) adopts cloud–edge hierarchical FedAvg to generate a source consensus, and selectively fine-tunes residual blocks, alleviating domain shift and negative transfer.

4.5.1. FD on cross-operating condition tasks

To validate the effectiveness of the proposed PCDFTL approach for the cross-operating condition FD, a series of comparative experiments are conducted on the 15 cross-operating condition tasks listed in Table 6 with MSSA, FMDAAN-V, LQKF, SPDFT, FTL-A, RDFTL, AFL-SA, CECFTL, and PCDFTL. Table 14 presents the accuracies of different FD approaches on the different cross-operating condition tasks. On the cross-operating condition FD tasks of PU dataset, the average FD accuracies of MSSA, FMDAAN-V, LQKF, SPDFT, FTL-A, RDFTL, AFL-SA, CECFTL, and PCDFTL are 98.96%, 95.54%, 96.87%, 96.75%, 95.84%, 97.46%, 97.48%, 96.31%, and 98.56%, respectively. In particular, the average FD accuracy of PCDFTL is 3.02% higher than that of FMDAAN-V. The improvement is primarily attributed to the following two key factors. Firstly, FMDAAN-V employs the mean-weighted model aggregation strategy, struggling to precisely measure the distribution discrepancies between the source and target domains during the aggregation phase. In contrast, PCDFTL adopts the proposed MSDWI strategy, which not only effectively avoids the impact of negative transfer by utilizing multiple source models to perform weighted fused inference on the target domain, but also effectively enhances the stability of model training and the accuracies of cross-operating condition FD tasks by evaluating the contributions of different source models to the final inference results from three different perspectives. Secondly, FMDAAN-V uses a simple voting strategy to generate the pseudo-labels of the target domain. Unlike LQKF and SPDFT adopting more complex pseudo-labeling strategies, the quality of the pseudo-labels generated by the pseudo-labeling strategy based on simple voting is generally low, which can easily provide the incorrect guidance when fine-tuning the global model using these pseudo-labels. It is noteworthy that the FD performance of PCDFTL is almost identical to that of the non-federated MSSA without considering data privacy, and even slightly better than MSSA on the tasks of T3 and T4, demonstrating that the proposed PCDFTL approach can maintain strong FD capability while protecting data privacy.

Figs. 15(a) and 15(b) illustrate the accuracy changes of different FD approaches on the task T2 of PU dataset and task T13 of JNU dataset, respectively. As depicted in Fig. 15, PCDFTL achieves rapid convergence to high accuracies on the tasks of T2 and T13. Notably, PCDFTL demonstrates the convergence speed and fluctuation magnitude comparable to those of the non-federated MSSA. Compared to the other FTL-based FD approaches, PCDFTL not only converges faster but also exhibits smaller fluctuations during training. There are the following two main reasons. Firstly, PCDFTL adopts the proposed MSDWI strategy, which conducts the one-to-one transfer training between the source and target clients during the model training phase. In the target domain inference phase, it leverages multiple source models for joint weighted inference on the target domain to avoid the negative transfer problem, effectively enhancing the stability of the model training. Secondly, PCDFTL adopts the proposed DPAPO strategy, which successfully maps the sample features of the source and target domains into the unified feature space, effectively performing domain alignment based on MK-MMD, thereby enhancing the generalization of the source models on the target domain.

Fig. 16 more distinctly illustrates the performance differences of various FD approaches across the cross-operating condition FD tasks from different datasets. As revealed in Fig. 16, the average accuracies of all FD approaches on the cross-operating condition FD tasks of PU dataset is lower than those on HUST, RFB, and JNU datasets. On the cross-operating condition FD tasks of PU dataset, FMDAAN-V achieves the lowest average accuracy, whereas the average accuracy of PCDFTL surpasses all the comparison approaches except MSSA. Overall, FTL-A has the lowest total average accuracy, whereas the total average accuracy of PCDFTL closely approaches that of the non-federated MSSA. These results reveal that PCDFTL outperforms the other FTL-based approaches in FD accuracy and nearly matches the non-federated MSSA, thereby validating its effectiveness on the cross-operating condition FD tasks.

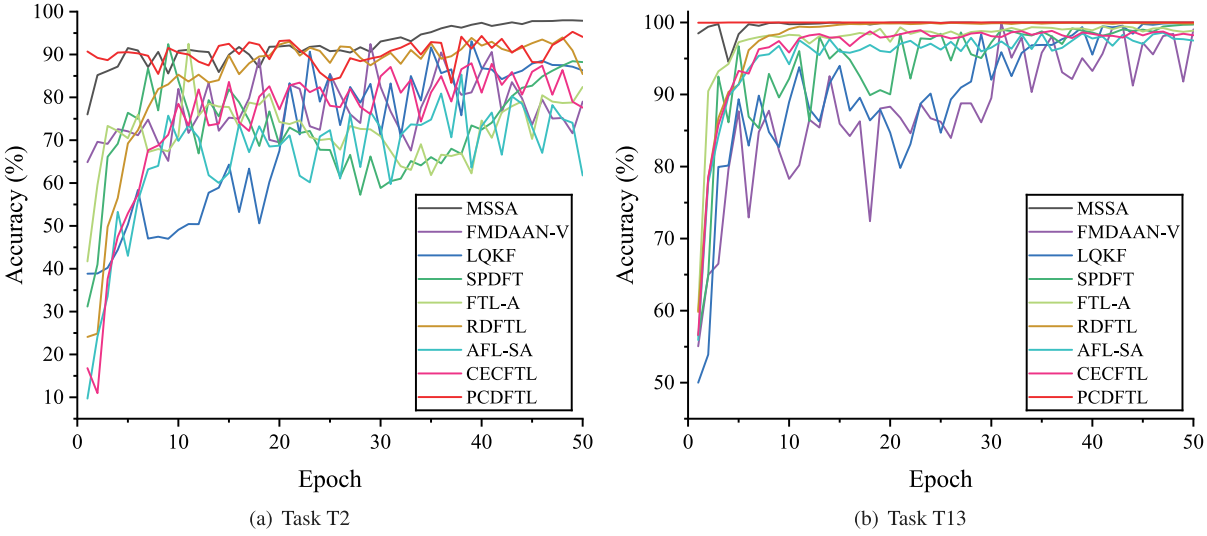
The further analysis of the feature extraction capabilities of different FD approaches on the target domain on the task T3 is conducted. Fig. 17 presents the t-SNE visualizations of the output features of the previous layer of the classification layer of the models trained by different FD approaches on the task T3. Meanwhile, the KL divergences of all t-SNE visualizations are measured in order as 1.26, 1.56, 1.32, 1.42, 1.40, 1.35, 1.20, 1.16, and 1.29, respectively. It can be observed that the final KL divergences of all t-SNE visualizations fall within the range of 1.00 to 1.60, indicating that the relatively low information loss during dimensionality reduction and high fidelity in preserving local structures. Therefore, the t-SNE visualizations can be considered a reliable reflection of the feature extraction capabilities of different models. As revealed in Fig. 17, the models trained using FMDAAN-V, FTL-A, SPDFL, and CECFTL exhibit weak feature separability, with some features remaining inseparable or mixed. In contrast, the models trained using MSSA, LQKF, RDFTL, AFL-SA, and PCDFTL produce highly separable features that form distinct clusters. This indicates that PCDFTL effectively extracts high-quality features of the target domain, thereby improving FD accuracy.

Figs. 18(a)–18(i) display the confusion matrices obtained by MSSA, FMDAAN-V, LQKF, SPDFT, FTL-A, RDFTL, AFL-SA, CECFTL, and PCDFTL on the task T1, respectively. For the three kinds of IF faults, the average misdiagnosis rates of the nine approaches are 0.67%, 6.00%, 2.33%, 2.33%, 5.00%, 5.00%, 3.00%, 3.33%, and 2.00%, respectively. For the three types of OF faults, the average misdiagnosis rates of the nine approaches are 0.67%, 3.33%, 5.00%, 1.67%, 2.67%, 1.67%, 1.00%, 1.67%, and 0.67%, respectively. For the three types of IF+OF faults, the average misdiagnosis rates of the nine approaches are 0.00%, 3.33%, 3.00%, 0.33%, 2.00%, 1.00%, 1.00%, 1.67%, and 0.67%, respectively. The above results indicate that the proposed PCDFTL approach can still accurately diagnose different types of faults while ensuring data privacy protection.

Table 15 presents the F1-scores of different FD approaches across all cross-operating condition tasks. The total average F1-scores of MSSA, FMDAAN-V, LQKF, SPDFT, FTL-A, RDFTL, AFL-SA, CECFTL,

Table 14
Accuracies (%) obtained by different FD approaches on different cross-operating condition tasks.

Dataset	Task	MSSA (Tian et al., 2022)	FMDAAN-V (Zhao et al., 2023)	LQKF (Wang et al., 2023)	SPDFT (Zhao et al., 2024a)	FTL-A (Li et al., 2024a)	RDFTL (Wan et al., 2024)	AFL-SA (Yuan et al., 2024)	CECFTL (Liang et al., 2025)	PCDFTL
PU	T1	99.74 ± 0.15	96.45 ± 0.06	97.16 ± 0.33	98.86 ± 0.03	97.52 ± 0.54	98.14 ± 0.01	98.55 ± 0.05	97.98 ± 0.28	99.66 ± 0.11
	T2	97.99 ± 0.01	92.44 ± 0.05	93.10 ± 0.09	92.40 ± 0.02	92.49 ± 0.17	93.98 ± 0.03	94.11 ± 0.04	92.74 ± 0.32	95.30 ± 0.15
	T3	99.85 ± 0.01	96.69 ± 0.07	99.37 ± 0.05	97.36 ± 0.08	96.75 ± 0.15	98.85 ± 0.02	98.75 ± 0.05	97.28 ± 0.09	99.97 ± 0.02
	T4	98.26 ± 0.01	96.57 ± 0.08	97.86 ± 0.08	98.37 ± 0.30	96.60 ± 0.25	98.85 ± 0.02	98.52 ± 0.02	97.25 ± 0.02	99.32 ± 0.02
	Avg.	98.96	95.54	96.87	96.75	95.84	97.46	97.48	96.31	98.56
HUST	T5	100.00 ± 0.00	99.13 ± 0.34	99.34 ± 0.01	99.90 ± 0.09	99.66 ± 0.34	99.70 ± 0.30	99.95 ± 0.05	99.95 ± 0.05	100.00 ± 0.00
	T6	100.00 ± 0.00	99.89 ± 0.05	100.00 ± 0.00	99.77 ± 0.52	99.15 ± 0.45	99.80 ± 0.20	99.72 ± 0.28	99.63 ± 0.33	100.00 ± 0.00
	T7	100.00 ± 0.00	99.95 ± 0.05	100.00 ± 0.00	99.58 ± 0.42	99.78 ± 0.21	100.00 ± 0.00	99.77 ± 0.13	99.58 ± 0.17	100.00 ± 0.00
	T8	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	Avg.	100.00	99.74	99.84	99.81	99.65	99.88	99.86	99.79	100.00
RFB	T9	99.42 ± 0.40	94.11 ± 0.88	94.89 ± 0.31	94.71 ± 0.47	93.10 ± 0.12	95.67 ± 0.14	94.98 ± 0.14	94.57 ± 0.10	96.12 ± 0.86
	T10	99.89 ± 0.10	97.98 ± 0.08	98.88 ± 0.56	98.41 ± 0.44	97.43 ± 0.77	99.00 ± 0.32	98.91 ± 0.12	98.45 ± 0.10	99.74 ± 0.11
	T11	99.56 ± 0.10	98.19 ± 0.82	98.73 ± 0.52	98.56 ± 0.50	98.11 ± 0.15	99.23 ± 0.02	98.83 ± 0.02	98.41 ± 0.05	99.86 ± 0.14
	T12	99.86 ± 0.01	98.17 ± 0.63	98.43 ± 0.07	98.68 ± 0.25	97.77 ± 0.41	98.76 ± 0.08	98.89 ± 0.05	98.78 ± 0.10	99.16 ± 0.28
	Avg.	99.68	97.11	97.73	97.59	96.60	98.17	97.90	97.55	98.72
JNU	T13	99.50 ± 0.50	99.45 ± 0.55	99.65 ± 0.35	99.51 ± 0.46	99.65 ± 0.20	99.57 ± 0.42	99.15 ± 0.05	99.37 ± 0.01	99.99 ± 0.01
	T14	99.46 ± 0.54	99.45 ± 0.55	99.55 ± 0.33	99.52 ± 0.45	99.55 ± 0.45	99.57 ± 0.20	99.18 ± 0.40	99.25 ± 0.05	99.98 ± 0.02
	T15	99.98 ± 0.02	99.69 ± 0.01	99.74 ± 0.21	99.72 ± 0.15	99.62 ± 0.24	99.58 ± 0.01	99.43 ± 0.10	99.07 ± 0.07	99.99 ± 0.01
	Avg.	99.65	99.53	99.65	99.58	99.61	99.57	99.25	99.23	99.99
	Total avg.	99.57	97.98	98.52	98.43	97.92	98.77	98.62	98.22	99.32



(a) Task T2 (b) Task T13

Fig. 15. Accuracies obtained by different FD approaches on different tasks.

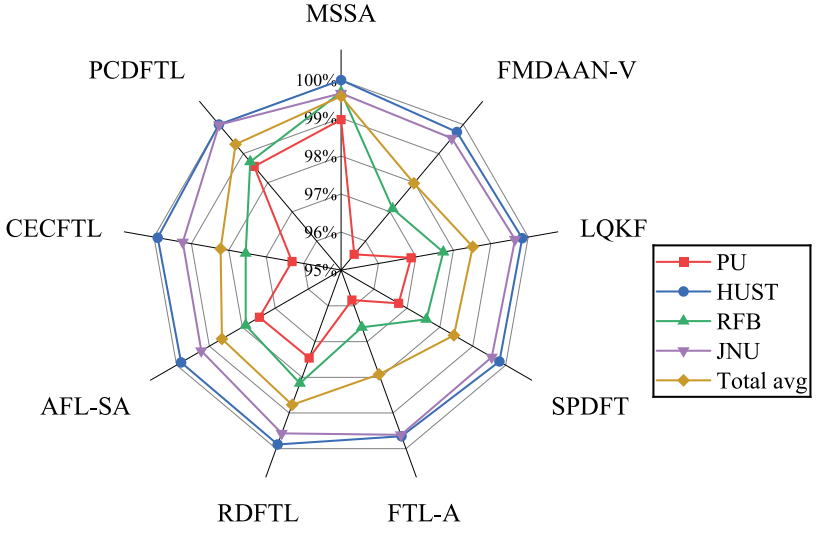


Fig. 16. Average accuracies achieved by different FD approaches on the tasks of the cross-operating condition FD from different datasets.

and PCDFTL are 99.47%, 97.68%, 98.16%, 97.99%, 97.40%, 98.32%, 97.99%, 97.80%, and 99.16% on PU, HUST, RFB, and JNU datasets, respectively. Notably, the average F1-score of PCDFTL on each dataset is generally higher than that of most other approaches, mainly due to its synergistic advantages in the model aggregation, domain alignment, and negative transfer mitigation. Compared with these approaches such as FedAvg, FTLDAN, FMDAAN-V, and FTL-A which rely on the average aggregation of the source models, PCDFTL abandons the model

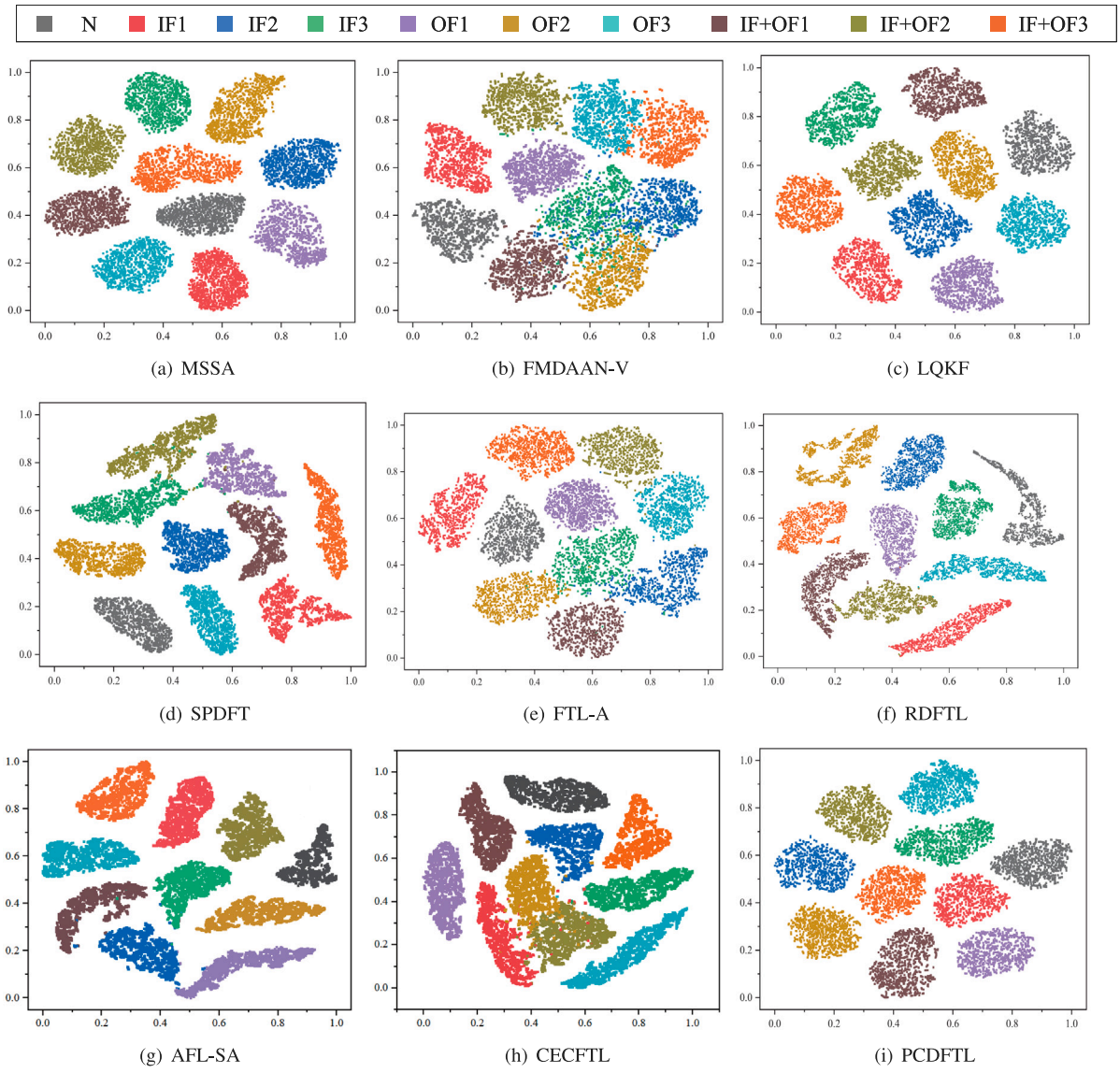


Fig. 17. The t-SNE visualizations of the output features of the previous layer of the classification layer of the models trained by different FD approaches on the task T3.

Table 15
F1-scores (%) obtained by different FD approaches on different cross-operating condition tasks.

Dataset	Task	MSSA (Tian et al., 2022)	FMDAAN-V (Zhao et al., 2023)	LQKF (Wang et al., 2023)	SPDFT (Zhao et al., 2024a)	FTL-A (Li et al., 2024a)	RDFTL (Wan et al., 2024)	AFL-SA (Yuan et al., 2024)	CECFTL (Liang et al., 2025)	PCDFTL
PU	T1	99.47 ± 0.04	96.05 ± 0.09	97.16 ± 0.33	98.01 ± 0.34	97.10 ± 0.47	97.85 ± 0.18	98.15 ± 0.57	98.04 ± 0.13	99.43 ± 0.17
	T2	97.57 ± 0.04	91.72 ± 0.05	92.11 ± 0.09	92.00 ± 0.21	91.53 ± 0.27	92.17 ± 0.23	93.67 ± 0.04	91.84 ± 0.03	95.13 ± 0.04
	T3	99.74 ± 0.01	95.71 ± 0.07	96.83 ± 0.17	98.36 ± 0.07	95.98 ± 0.04	97.74 ± 0.12	97.65 ± 0.10	96.77 ± 0.05	99.64 ± 0.03
	T4	99.01 ± 0.04	95.87 ± 0.18	97.26 ± 0.08	97.74 ± 0.26	95.73 ± 0.25	97.85 ± 0.02	97.62 ± 0.07	96.73 ± 0.23	99.11 ± 0.04
	Avg.	98.95	94.84	95.84	96.53	95.09	96.40	96.77	95.85	98.33
HUST	T5	100.00 ± 0.00	98.78 ± 0.46	98.94 ± 0.01	98.56 ± 0.90	99.06 ± 0.00	99.47 ± 0.12	99.01 ± 0.00	99.47 ± 0.00	100.00 ± 0.00
	T6	100.00 ± 0.00	99.67 ± 0.05	100.00 ± 0.00	98.98 ± 0.12	98.98 ± 0.00	99.47 ± 0.17	98.95 ± 0.00	99.12 ± 0.00	100.00 ± 0.00
	T7	100.00 ± 0.00	99.55 ± 0.05	100.00 ± 0.00	98.94 ± 0.15	99.45 ± 0.00	100.00 ± 0.00	98.98 ± 0.00	99.24 ± 0.00	100.00 ± 0.00
	T8	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
	Avg.	100.00	99.50	99.74	99.12	99.37	99.74	99.24	99.46	100.00
RFB	T9	99.42 ± 0.40	94.11 ± 0.88	94.89 ± 0.31	94.71 ± 0.47	92.10 ± 0.12	95.35 ± 0.13	93.95 ± 0.14	94.05 ± 0.19	95.84 ± 0.29
	T10	99.33 ± 0.07	97.46 ± 0.07	98.27 ± 0.16	96.13 ± 0.02	96.34 ± 0.48	98.61 ± 0.02	97.97 ± 0.02	97.93 ± 0.10	99.37 ± 0.17
	T11	99.14 ± 0.11	97.88 ± 0.27	98.33 ± 0.21	98.04 ± 0.17	97.87 ± 0.52	98.97 ± 0.03	97.98 ± 0.05	97.95 ± 0.07	99.43 ± 0.09
	T12	99.56 ± 0.01	97.84 ± 0.06	98.01 ± 0.07	98.27 ± 0.09	97.23 ± 0.11	98.43 ± 0.13	97.92 ± 0.09	97.89 ± 0.10	99.00 ± 0.13
	Avg.	99.36	96.82	97.38	96.79	95.89	97.84	96.96	96.96	98.41
JNU	T13	99.39 ± 0.18	99.23 ± 0.01	99.49 ± 0.01	99.53 ± 0.09	99.59 ± 0.07	98.89 ± 0.42	99.06 ± 0.05	99.05 ± 0.01	99.90 ± 0.01
	T14	99.32 ± 0.23	99.80 ± 0.01	99.83 ± 0.01	99.61 ± 0.17	99.04 ± 0.04	99.50 ± 0.20	98.98 ± 0.10	98.97 ± 0.05	99.90 ± 0.02
	T15	99.98 ± 0.02	99.59 ± 0.01	99.79 ± 0.01	99.46 ± 0.21	99.16 ± 0.07	99.47 ± 0.01	98.95 ± 0.01	98.77 ± 0.01	99.85 ± 0.05
	Avg.	99.56	99.54	99.70	99.53	99.26	99.29	99.00	98.93	99.88
	Total avg.	99.47	97.68	98.16	97.99	97.40	98.32	97.99	97.80	99.16

aggregation and instead employs the proposed MSDWI strategy during the inference stage to dynamically weight the contributions of different source models on the target domain based on multiple criteria, effectively leveraging high-quality source models while suppressing the negative influence of low-quality source models, thereby enhancing the predictive accuracy on the target domain. In terms of domain

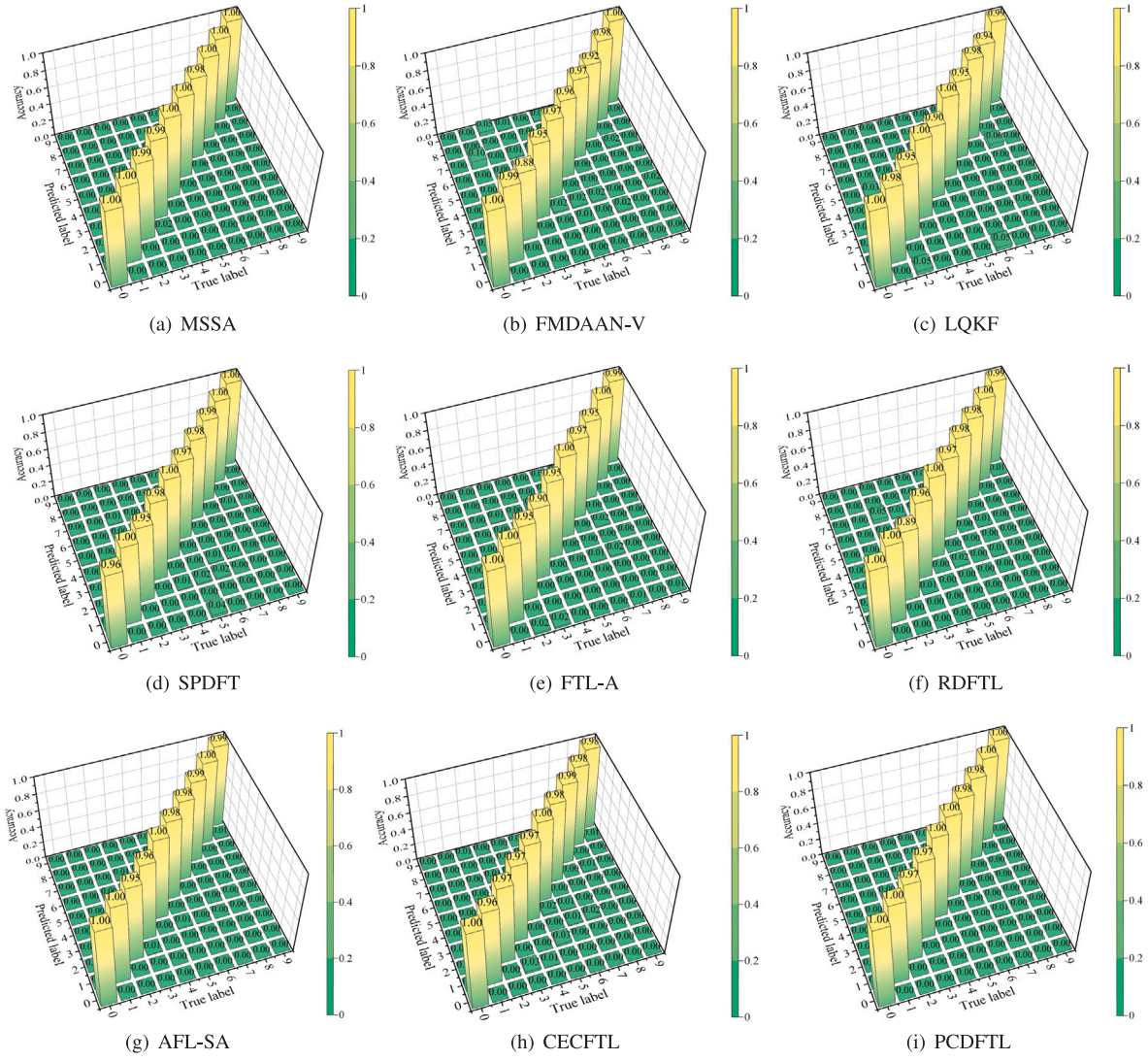


Fig. 18. Confusion matrices of different FD approaches on the task T1.

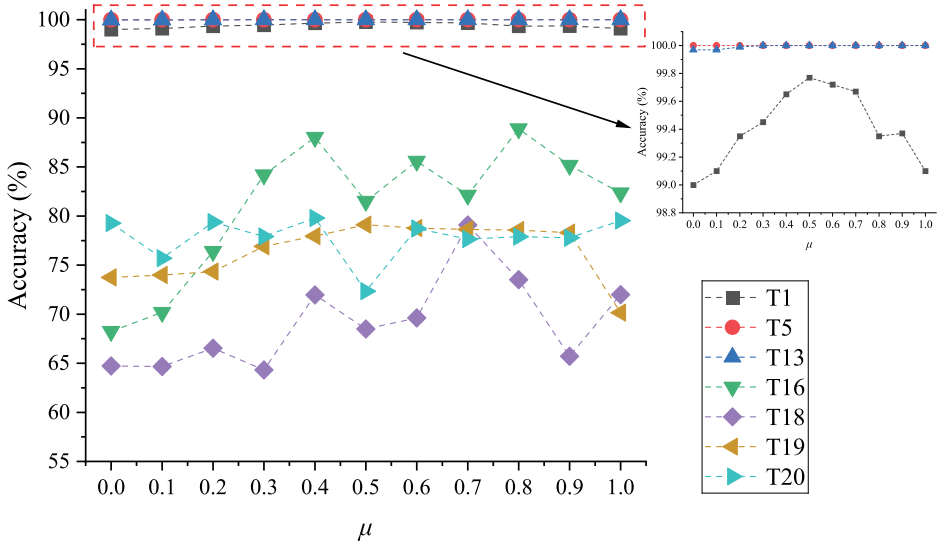


Fig. 19. Influence of the parameter μ on FD accuracy.

Table 16
Tasks of cross-machine FD.

Transfer	Task	Source client 1	Source client 2	Source client 3	Target client
JNU → PU	T16	J1	J2	J3	P1
	T17	J1	J2	J3	P2
	T18	J1	J2	J3	P3
PU → JNU	T19	P1	P2	P3	J1
	T20	P1	P2	P3	J2
	T21	P1	P2	P3	J3

alignment, FTLDAN employs deep adversarial networks, FMDAAN-V combines MK-MMD with adversarial learning, LQKF relies on the distance metric based on batch normalization layer statistics, and SPDFT integrates hash mapping with MK-MMD. These approaches often address only partial distribution shifts and struggle to fully align feature distributions across varying operating conditions and machines. In contrast, PCDFTL achieves more precise and stable cross-domain feature alignment by adopting the proposed DPAPO strategy. In terms of negative transfer mitigation, FMDAAN-V adopts pseudo-label fine-tuning, LQKF employs low-quality knowledge filtering, SPDFT introduces self-paced learning, RDFTL used the low-quality model suppression strategy, AFL-SA introduces the active learning query mechanism, and CECFTL adopts the adaptive model fine-tuning mechanism, these approaches can alleviate negative transfer to some extent. However, they are mostly static or rely on the single-metric guidance, making them difficult to completely avoid interference from unrelated source domains. In contrast, PCDFTL avoids the impact of negative transfer by using MSDWI strategy.

4.5.2. FD on cross-machine tasks

To validate the effectiveness of the proposed PCDFTL approach for the cross-machine FD, the six different cross-machine FD tasks are designed for JNU and PU datasets, as listed in Table 16, where the healthy, IF, and OF samples are included under each operating condition. Table 17 presents the accuracies of different FD approaches across these cross-machine tasks. As observed in Table 17, on the tasks of T16, T17, and T18, the average FD accuracies of MSSA, FMDAAN-V, LQKF, SPDFT, FTL-A, RDFTL, AFL-SA, CECFTL, and PCDFTL are 85.89%, 77.34%, 85.43%, 84.70%, 70.43%, 85.72%, 85.60%, 84.15%, and 86.20%, respectively. Among them, PCDFTL achieves the highest average accuracy on the three transfer tasks of JNU → PU. The primary reason is that on the cross-machine FD tasks, the distribution discrepancies between the source and target domains are significantly larger. This exacerbates the impact of the negative transfer caused by model aggregation, thereby reducing the stability of the source model training. In contrast, PCDFTL employs MSDWI strategy for target domain inference, effectively avoiding the impact of the negative transfer.

As also observed in Table 17, on the tasks of T19, T20, and T21, PCDFTL achieves an average FD accuracy of 79.92%. Compared with FTL-A achieving the lowest average accuracy, the average accuracy of PCDFTL is increased by 15.40%. The main reason is that FTL-A relying solely on the mean square error between the probability distributions at the output of the classification layer and regularization is insufficient to effectively measure the distribution discrepancies between the source and target domains. In contrast, PCDFTL uses MK-MMD and PDC loss to alleviate the impact of domain shift, which can better measure the statistical distances between the source and target domains under the scenarios with substantial data distribution discrepancies, thereby enhancing the generalization of the model. Overall, the results prove that PCDFTL can maintain a high FD accuracy even on the cross-machine FD tasks with substantial data distribution discrepancies.

4.6. Analysis of parameter sensitivity

The sensitivity analysis experiments are conducted for the key weighting parameters μ and η that are respectively employed in PDC loss and MSDWI strategy of the proposed PCDFTL approach across the three cross-operating condition FD tasks of T1, T5, and T13, as well as four cross-machine FD tasks of T16, T18, T19, and T20. To thoroughly evaluate the impacts of μ and η on the final inference results, the values of μ and η are incrementally varied from 0 to 1 with a step size of 0.1. Fig. 19 illustrates the influence of μ on FD accuracy. The FD accuracies achieved on the tasks of T5 and T13 barely changes slightly with the variation of μ, consistently remaining at 99.99% or higher. The FD accuracy achieved on T1 experiences slight variations with the change in μ, with the best value of μ occurring at 0.5, corresponding to an accuracy of 99.77%. For the tasks T16 and T18, the best range of μ for achieving the best FD accuracies is [0.7, 0.8]. For the tasks T19 and T20, the best range of μ for achieving the best FD accuracies is [0.4, 0.5]. Fig. 20 reveals the influence of η on FD accuracy. The tasks of T5 and T13 exhibit near insensitivity to variations in η. For the task T1, the best FD accuracy of 99.77% is achieved when η is set to 0.9. For the task T16, the best FD accuracy of 84.29% is attained when η is set to 0.2. For the tasks T18, T19, and T20, the best FD accuracies are 70.75%, 79.79%, and 79.70% when η is set to 0.9, 0.4, and 0.9, respectively.

As displayed in Figs. 19 and 20, the accuracies of the cross-machine FD tasks are more sensitive to μ and η, whereas the accuracies of the cross-operating condition FD tasks exhibit less sensitivities to μ and η. Among the cross-operating condition tasks, the task T1 is most affected by μ and η. This is because the data distribution discrepancies under cross-machine FD tasks are greater than those under cross-operating condition FD tasks, leading to a greater reliance on appropriate parameter settings for cross-machine FD tasks to achieve better FD performance.

4.7. Analysis of model inference efficiency

To better evaluate the real-time inferring efficiency of the proposed PCDFTL, taking PU dataset as an example, the three well-trained source models obtained on three different source clients perform the weighted inference on different number of target samples through the proposed MSDWI strategy, where the length of each sample is 1024 sample points. Table 18 presents the model inference time on different number of samples. As seen in Table 18, PCDFTL exhibits excellent linear scalability in the model inference efficiency. When the number of samples increases from 2028 to 15240, the model inference time increases from 15 s to 117 s, that is, the model reasoning time increases by 7.8 times as the number of samples increases by 7.5 times. More importantly, the processing time of each sample remains stable in the range of 7.4 to 7.7 ms, indicating that PCDFTL maintains consistent inference efficiency across different number of samples without significant performance degradation. This stable linear growth relationship and efficient utilization of computing resources fully demonstrate the potential of PCDFTL to adapt to large-scale FD tasks and meet the stringent requirements for real-time inferring in the industrial scenarios.

5. Conclusions

In this paper, a prototype contrastive-based decentralized federated transfer learning approach for intelligent FD is proposed, which can achieve superior cross-domain RMFD performance while protecting data privacy by adopting the DPAPO strategy, PDC loss, and MSDWI strategy. A series of experiments are conducted on PU, HUST, RFB, and JNU datasets, verifying the effectiveness of the proposed PCDFTL approach. The results reveal that the average accuracies achieved by the proposed approach on the cross-operating condition and cross-machine FD tasks reach 99.32% and 83.06%, respectively. The key conclusions of this paper are as follows.

Table 17
Accuracies (%) obtained by different FD approaches on different cross-machine tasks.

Transfer	Task	MSSA (Tian et al., 2022)	FMDAAN-V (Zhao et al., 2023)	LQKF (Wang et al., 2023)	SPDFT (Zhao et al., 2024a)	FTL-A (Li et al., 2024a)	RDFTL (Wan et al., 2024)	AFL-SA (Yuan et al., 2024)	CECFTL (Liang et al., 2025)	PCDFTL
JNU → PU	T16	86.55 ± 0.30	78.83 ± 0.23	86.95 ± 0.02	85.82 ± 0.02	70.60 ± 0.59	87.85 ± 0.02	87.54 ± 0.03	85.16 ± 0.05	88.87 ± 0.02
	T17	95.02 ± 0.05	77.25 ± 0.15	90.45 ± 0.31	89.57 ± 0.11	71.42 ± 0.48	90.31 ± 0.26	90.24 ± 0.18	88.55 ± 0.11	90.69 ± 0.01
	T18	76.10 ± 0.05	75.93 ± 0.59	78.88 ± 0.13	78.72 ± 0.29	69.28 ± 0.37	79.01 ± 0.01	79.03 ± 0.01	78.73 ± 0.04	79.03 ± 0.05
	Avg.	85.89	77.34	85.43	84.70	70.43	85.72	85.60	84.15	86.20
PU → JNU	T19	75.67 ± 0.10	78.55 ± 0.19	79.30 ± 0.13	78.51 ± 0.64	60.21 ± 0.06	79.46 ± 0.31	79.59 ± 0.11	78.66 ± 0.04	79.72 ± 0.07
	T20	78.89 ± 0.15	77.57 ± 0.31	79.03 ± 0.18	75.50 ± 0.20	65.59 ± 0.19	79.04 ± 0.86	79.31 ± 0.06	75.68 ± 0.01	79.76 ± 0.03
	T21	79.49 ± 0.06	79.03 ± 0.14	78.25 ± 0.45	77.57 ± 0.26	67.75 ± 0.05	79.48 ± 0.26	79.69 ± 0.35	76.88 ± 0.42	80.27 ± 0.37
	Avg.	78.02	78.38	78.86	77.19	64.52	79.33	79.53	77.07	79.92
	Total avg.	81.96	77.86	82.12	80.95	67.48	82.53	82.57	80.61	83.06

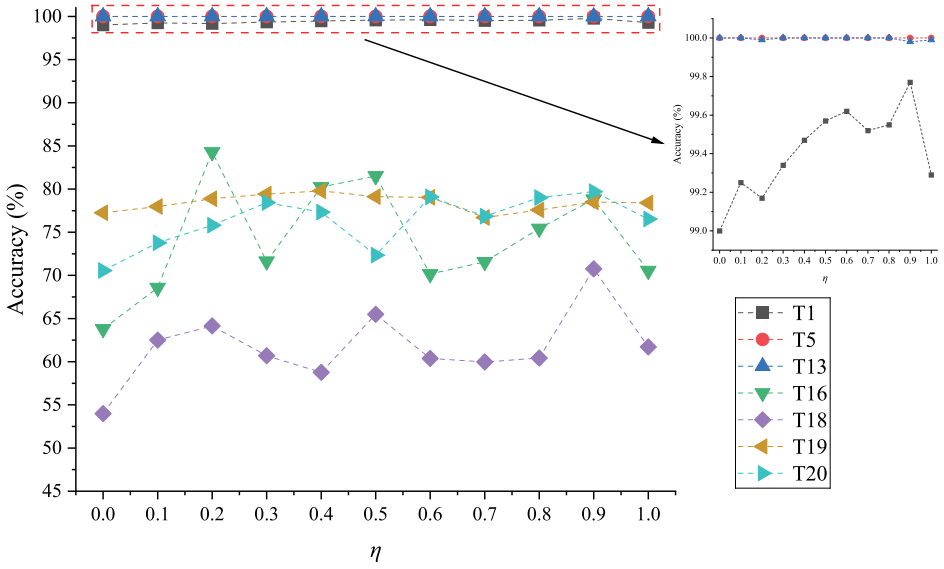


Fig. 20. Influence of the parameter η on FD accuracy.

Table 18
Model inference time on different number of samples.

Number of samples	Inferring time (seconds)
2028	15
7030	49
15240	117

1. By employing the proposed DPAPO strategy to project the sample features of the source and target domains into a unified high-dimensional feature space, the issue of the feature space shift between the source and target domains in FTL is effectively solved, significantly enhancing the effect of domain alignment.
2. By using the proposed PDC loss, where the prototype contrastive loss is combined with MK-MMD, which not only effectively alleviates the impact of domain shift, but also effectively reduces the intra-class distances and expands the inter-class distances.
3. By adopting the proposed MSDWI strategy, where the multiple source models are guided to perform dynamic weighted fused inference on the target domain according to the statistical distances between the source and target domains, the stability of DA, and the certainty of the entropy of the inference results, which effectively avoids the negative transfer problem caused by model aggregation in FTL.

Although the proposed PCDFTL demonstrates strong cross-domain FD capability under data privacy protection, it still has some limitations as follows. (1) PCDFTL requires a visible target domain to participate in model training, but the scenario where the target domain is unseen may be encountered. (2) Although the model inference time of PCDFTL increases linearly, its huge model training overhead is still a major challenge in the large-scale FTL-based FD. (3) PCDFTL assumes that the source and target domains have the same label space, but the target domain may contain fault classes that do not belong to the source domains in actual industrial scenarios. Given these limitations, the future work will mainly focus on the following three aspects. (1) A meta-contrastive learning mechanism will be incorporated to enable PCDFTL to generalize to the unseen target domain. (2) On the premise of guaranteeing feature extraction capability, the model architecture will be lightweight and the number of feature transmissions during model training will be further reduced to cut down the overall training cost to meet the requirements of the large-scale FTL-based FD. (3) A out-of-distribution detection mechanism will be introduced in PCDFTL to identify new fault classes during inference.

CRediT authorship contribution statement

Zhaokang Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Data curation. **Lanjun Wan:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Investigation, Funding acquisition, Conceptualization. **Jiaen Ning:** Visualization, Validation, Software, Methodology. **Wei Ni:** Visualization, Validation, Methodology. **Keqin Li:** Writing – review & editing, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Scientific Research Foundation of Hunan Provincial Education Department, China [grant number 24A 0391]; the Natural Science Foundation of Hunan Province, China [grant numbers 2023JJ30217 and 2025JJ70030]; and the National Natural Science Foundation for Young Scientists of China [grant number 61702177].

Data availability

Data will be made available on request.

References

Chen, Y., Xiao, L., 2024. A multisource–multitarget domain adaptation method for rolling bearing fault diagnosis. IEEE Sens. J. 24 (3), 3406–3419.

Du, J., Qin, N., Huang, D., Jia, X., Zhang, Y., 2024. Lightweight FL: A low-cost federated learning framework for mechanical fault diagnosis with training optimization and model pruning. IEEE Trans. Instrum. Meas. 73, 3328073.

He, Y., Shen, W., 2024. FedITA: A cloud–edge collaboration framework for domain generalization-based federated fault diagnosis of machine-level industrial motors. Adv. Eng. Inf. 62, 102853.

Hu, H., Wang, X., Zhang, Y., Chen, Q., Guan, Q., 2024. A comprehensive survey on contrastive learning. Neurocomputing 610, 128645.

Jiang, G., Zhao, K., Liu, X., Cheng, X., Xie, P., 2024. A federated learning framework for cloud-edge collaborative fault diagnosis of wind turbines. IEEE Internet Things J. 11, 23170–23185.

Keshun, Y., Puzhou, W., Peng, H., Yingkui, G., 2025. A sound-vibration physical-information fusion constraint-guided deep learning method for rolling bearing fault diagnosis. Reliab. Eng. Syst. Saf. 253, 110556.

Lessmeier, C., Kimotho, J.K., Zimmer, D., Sextro, W., 2016. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In: Proc. Eur. Conf. PHM Soc. (PHME16), vol. 3, (no. 1), PHM Society, pp. 1–17.

Li, X., Chen, H., Li, S., Wei, D., Zou, X., Si, L., Shao, H., 2025a. Multi-kernel weighted joint domain adaptation network for cross-condition fault diagnosis of rolling bearings. Reliab. Eng. Syst. Saf. 261, 111109.

Li, J., Deng, W., Dang, X., Zhao, H., 2025b. Cross-domain adaptation fault diagnosis with maximum classifier discrepancy and deep feature alignment under variable working conditions. IEEE Trans. Reliab. 74, 4106–4115.

Li, Z., Li, Z., Gu, F., 2024a. Intelligent diagnosis method for machine faults based on federated transfer learning. Appl. Soft Comput. 163, 111922.

Li, C., Mo, L., Kwoh, C.K., Li, X., Chen, Z., Wu, M., Yan, R., 2025c. Noise-robust multi-view graph neural network for fault diagnosis of rotating machinery. Mech. Syst. Signal Process. 224, 112025.

Li, B., Song, P., Zhao, C., 2024b. Fusing consensus knowledge: A federated learning method for fault diagnosis via privacy-preserving reference under domain shift. Inf. Fusion 106, 102290.

Li, K., Xiong, M., Li, F., Su, L., Wu, J., 2019. A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine. Neurocomputing 350, 261–270.

Liang, X., Sun, Y., Kang, S., Zhao, Z., Wang, Y., Wang, Q., 2025. A rolling bearing fault diagnosis method based on cloud-edge collaboration federated transfer learning. Meas. Sci. Technol. 36 (4), 046111.

Liu, J., Wan, L., Xie, F., Sun, Y., Wang, X., Li, D., Wu, S., 2024a. Cross-machine deep subdomain adaptation network for wind turbines fault diagnosis. Mech. Syst. Signal Process. 210, 111151.

Liu, L., Yan, Z., Zhang, T., Gao, Z., Cai, H., Wang, J., 2024b. Data privacy protection: A novel federated transfer learning scheme for bearing fault diagnosis. Knowl.-Based Syst. 291, 111587.

McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A., 2017. Communication-efficient learning of deep networks from decentralized data. In: Proc. 20th Int. Conf. Artif. Intell. Stat.. PMLR, pp. 1273–1282.

Mehta, M., Chen, S., Tang, H., Shao, C., 2023. A federated learning approach to mixed fault diagnosis in rotating machinery. J. Manuf. Syst. 68, 687–694.

Misbah, I., Lee, C.K., Keung, K.L., 2024. Fault diagnosis in rotating machines based on transfer learning: Literature review. Knowl.-Based Syst. 283, 111158.

Prigent, C., Costan, A., Antoniu, G., Cudennec, L., 2024. Enabling federated learning across the computing continuum: Systems, challenges and future directions. Futur. Gener. Comp. Syst. 160, 767–783.

Qian, Q., Zhang, B., Li, C., Mao, Y., Qin, Y., 2025. Federated transfer learning for machinery fault diagnosis: A comprehensive review of technique and application. Mech. Syst. Signal Process. 223, 111837.

Su, H., Xiang, L., Hu, A., 2024. Application of deep learning to fault diagnosis of rotating machineries. Meas. Sci. Technol. 35 (4), 042003.

Sun, Y., Kang, S., Wang, Y., Liu, L., Lv, W., Wang, H., 2025. Fault diagnosis method for harmonic reducer based on personalized federated aggregation strategy with skip cycle weight. Measurement 242, 116275.

Tang, S., Ma, J., Yan, Z., Zhu, Y., Khoo, B.C., 2024. Deep transfer learning strategy in intelligent fault diagnosis of rotating machinery. Eng. Appl. Artif. Intell. 134, 108678.

Tian, J., Han, D., Li, M., Shi, P., 2022. A multi-source information transfer learning method with subdomain adaptation for cross-domain fault diagnosis. Knowl.-Based Syst. 243, 108466.

Wan, L., Ning, J., Li, Y., Li, C., Li, K., 2024. Intelligent fault diagnosis via ring-based decentralized federated transfer learning. Knowl.-Based Syst. 284, 111288.

Wang, X., Jiang, H., Mu, M., Dong, Y., 2025. A dynamic collaborative adversarial domain adaptation network for unsupervised rotating machinery fault diagnosis. Reliab. Eng. Syst. Saf. 255, 110662.

Wang, R., Yan, F., Yu, L., Shen, C., Hu, X., Chen, J., 2023. A federated transfer learning method with low-quality knowledge filtering and dynamic model aggregation for rolling bearing fault diagnosis. Mech. Syst. Signal Process. 198, 110413.

Xiao, D., Diao, J., Ding, J., Jiang, L., Zhao, C., 2025. FedRad: A dynamic low-cost federated learning for fault diagnosis of rotating machinery. IEEE Trans. Instrum. Meas. 74, 3551914.

Yang, J.-F., Zhang, N., He, Y.-L., Zhu, Q.-X., Xu, Y., 2024. Novel dual-network autoencoder based adversarial domain adaptation with Wasserstein divergence for fault diagnosis of unlabeled data. Expert Syst. Appl. 238, 122393.

Yu, S., Pang, S., Ning, J., Wang, M., Song, L., 2025. ANC-Net: A novel multi-scale active noise cancellation network for rotating machinery fault diagnosis based on discrete wavelet transform. Expert Syst. Appl. 265, 125937.

Yuan, X., Shi, D., Shi, N., Li, Y., Liang, P., Zhang, L., Zheng, Z., 2024. Intelligent fault diagnosis of rolling bearing based on an active federated local subdomain adaptation method. Adv. Eng. Inf. 62, 102807.

Zhang, G., Kong, X., Wang, Q., Du, J., Xu, K., Wang, J., Ma, H., 2024. Multi-source partial domain adaptation method based on pseudo-balanced target domain for fault diagnosis. Knowl.-Based Syst. 284, 111255.

Zhang, W., Li, X., 2022. Federated transfer learning for intelligent fault diagnostics using deep adversarial networks with data privacy. IEEE/ASME Trans. Mechatronics 27 (1), 430–439.

Zhao, K., Hu, J., Shao, H., Hu, J., 2023. Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy. Reliab. Eng. Syst. Saf. 236, 109246.

Zhao, K., Liu, Z., Li, J., Zhao, B., Jia, Z., Shao, H., 2024a. Self-paced decentralized federated transfer framework for rotating machinery fault diagnosis with multiple domains. Mech. Syst. Signal Process. 211, 111258.

Zhao, C., Shen, W., 2024. A federated distillation domain generalization framework for machinery fault diagnosis with data privacy. Eng. Appl. Artif. Intell. 130, 107765.

Zhao, C., Zio, E., Shen, W., 2024b. Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. Reliab. Eng. Syst. Saf. 245, 109964.

Zhou, R., Li, Y., Lin, X., 2025a. A clustered federated learning framework for collaborative fault diagnosis of wind turbines. Appl. Energy. 377, 124532.

Zhou, J., Zhang, X., Jiang, H., Li, J., Shao, Z., 2025b. Cross-domain transfer fault diagnosis by class-imbalanced deep subdomain adaptive network. Measurement 242, 115901.