Supplemental Material

A. Problem Definition

In this supplemental material, we provide a detailed explanation of the problem definition process.

Intelligent fault diagnosis is a classification problem in machine learning. Assuming that (x^S, y^S) and (x^T, y^T) represent the features and labels in the source domain D^S and the target domain D^T, respectively, the initial problem is to train a classifier $\mathcal{F}$ to minimize the loss of classification in the target domain:

$$\min_{\mathcal{F}} Loss(\mathcal{F}(x^T), y^T). \tag{S.1}$$

In FTL since there is only a small number of labeled samples in the target domain, training is performed at the intersection of the source and target domains $D^{S,T}$:

$$\begin{aligned} &\Rightarrow \min_{\mathcal{F}} Loss(\mathcal{F}(x^{S,T}), y^{S,T}), \\ &(x^{S,T}, y^{S,T}) \in D^{S,T} = D^S \cap D^T. \end{aligned} \tag{S.2}$$

Due to the system model, there are no overlapping samples in the source and target domains, which means that $D^{S,T}$ does not exist. Under this restriction, when the samples in the source and target domains are similar, such as the same equipment under similar working conditions, the distance between the source domain and the target domain is relatively close, and the model-based transfer learning method can be used to directly transfer the fault diagnosis model trained in the source domain to the target domain, and then fine-tune with a small number of labeled data $(x^{T'}, y^{T'})$ in the target domain:

$$\begin{aligned} &\Rightarrow \min_{\mathcal{F}} Loss(\mathcal{F}(x^S), y^S) + Loss(\mathcal{F}(x^{T'}), y^{T'}), \\ &dist(x^S, X^T) < \epsilon. \end{aligned} \tag{S.3}$$

In practical industrial scenarios, especially among different agents, it is difficult to ensure the consistency of working conditions, which conflicts with $dist(x^S, x^T) < \epsilon$. Under the restriction of feature heterogeneity, the model-based transfer learning method is not suitable, and the feature-based domain adaptation method can be used. Build feature extractors to map the source domain and target domain to a latent common space. Combine the source domain data and a small number of labeled samples in the target domain for training on the latent common space:

$$\begin{aligned} \Rightarrow \min_{\mathcal{F_C}, \mathcal{F_S}, \mathcal{F_T}} & Loss(\mathcal{F_C}(\mathcal{F_S}(x^S)), y^S) + Loss(\mathcal{F_C}(\mathcal{F_T}(x^{T'})), y^{T'}) \\ & + \lambda dist(\mathcal{F_S}(X^S), \mathcal{F_T}(X^T)). \end{aligned} \tag{S.4}$$

A problem with the above approach is that a small number of labeled samples in the target domain are still needed in the process of fault diagnosis, which fail to meet the need for some practical industrial scenarios like the cold start problem. Based on this setting, there is zero fault label in the target domain. Training with only source domain samples may lead to overfitting. Since the label spaces in the source and target domains are assumed to be the same, a related method of unsupervised domain adaptation can be used to align the source and target domain output label distributions. Ultimately our research question is defined as follows:

$$\begin{aligned} \Rightarrow \min_{\mathcal{F_C}, \mathcal{F_S}, \mathcal{F_T}} & Loss(\mathcal{F_C}(\mathcal{F_S}(x^S)), y^S) + \lambda dist(\mathcal{F_S}(x^S), \mathcal{F_T}(x^T)) \\ & + \beta dist(\mathcal{F_C}(\mathcal{F_S}(x^S)), \mathcal{F_C}(\mathcal{F_T}(x^T))). \end{aligned} \tag{S.5}$$

B. Model Structure

According to Eq.(1), our research problem can be regarded as an unsupervised domain adaptation problem, and the objective function mainly consists of three parts: classification loss, feature alignment loss and output label alignment loss.

Combining feature-based domain adaptation and unsupervised learning, the main methods are discrepancy-based and adversarial-based. We propose a vertical federated joint domain adversarial adaptation, based on the Adversarial-based method CDAN and vertical federated scheme, to calculate the classification loss and feature alignment loss. The key is a novel conditional domain discriminator conditioned on the cross-covariance of domain-specific feature representations and classifier predictions, which can map heterogeneous source and target feature spaces to a latent common space.

Even if the discriminator is completely obfuscated, there is no guarantee that the feature extractor can extract domain-invariant features. This risk arises from the equilibrium challenges that exist in adversarial learning, since we cannot guarantee that the two distributions are sufficiently similar even if the discriminator is completely confused . Since the domain adversarial adaptation has already aligned the feature space, we additionally add the discrepancy-based joint alignment method as the joint domain alignment to calculate the output label alignment loss, which minimizes the distance between the source label distribution and the target classification result distribution, fundamentally different from the conventional pseudo label method that does not comprehensively leverage the target domain information. The overall structure of the model is shown in Fig.2.

C. Supplementary Experimental Results

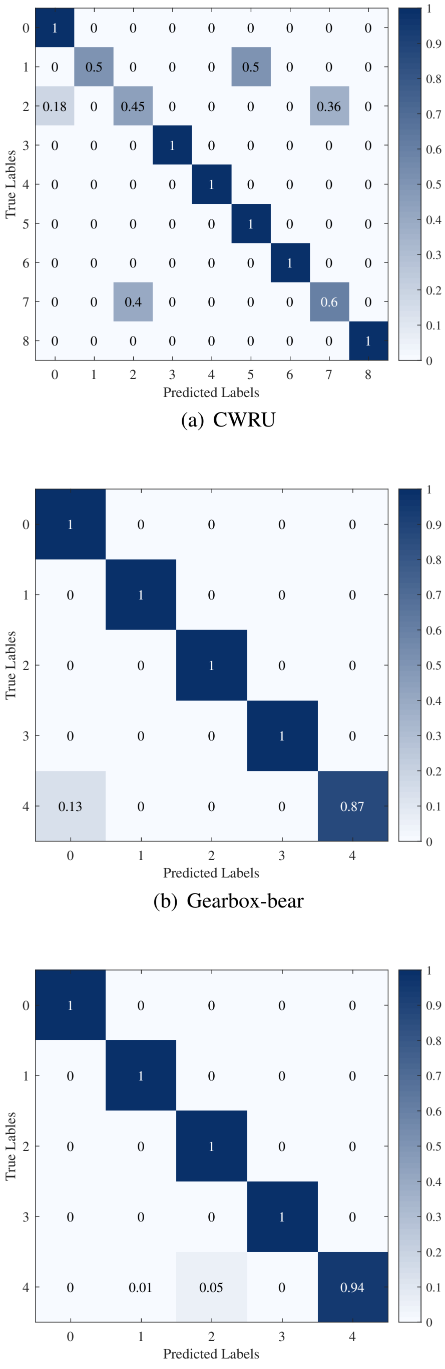
In supplemental material C, we list some supplementary materials for the experiment, including sample extraction, confusion matrixes, and table results for experiments.

1) Sample Extraction:

We extracted learning samples from monitoring signals using the non-overlapping sliding window method, where both the window width and the step size were 1024. Normalization was applied to each sample to reduce the within-sample variability. Extracted samples were divided into the source and target domains following the aforementioned tasks. Samples in both the source and target domains were further divided as training and testing sets with a 7:3 ratio.

2) Fault Diagnosis Accuracy:

In this section, we list the confusion matrixes and detailed numerical results of the experiments in Section 5. Confusion matrixes of all tasks are exhibited in Fig.S.1 It can be found from the confusion matrix that FedLED can effectively diagnose most faults except those labeled 1, 2, and 7 in CWRU dataset.



(a) CWRU

(b) Gearbox-bear

(c) Gearbox-gear

Fig. S.1. Confusion Matrixes of FedLED under all Tasks

TABLE S.1
FAULT DIAGNOSIS ACCURACY ON CWRU

Method	T1	T2	T3	T4	T5	T6	AVG
Baseline	0.57	0.63	0.43	0.43	0.41	0.56	0.50
Coral	0.42	0.58	0.66	0.31	0.45	0.40	0.47
Mk-MMD	0.57	0.56	0.51	0.68	0.23	0.59	0.52
CDA_E	0.72	0.67	0.32	0.49	0.20	0.78	0.53
DANN	0.15	0.15	0.13	0.15	0.12	0.17	0.15
SFL-multi	0.59	0.68	0.72	0.65	0.67	0.54	0.64
Abl Exp 1	0.74	0.77	0.70	0.70	0.68	0.64	0.71
Abl Exp 2	0.75	0.80	0.39	0.47	0.22	0.29	0.49
Ours	**0.87**	**0.93**	**0.56**	**0.73**	**0.73**	**0.81**	**0.77**

TABLE S.2
FAULT DIAGNOSIS ACCURACY ON GEARBOX-BEAR

Methods	T1	T2	T3	T4	T5	T6	Average
Baseline	0.71	0.51	0.77	0.96	0.93	0.96	0.81
CORAL	0.86	0.84	0.90	0.99	0.94	0.96	0.91
Mk-MMD	0.78	0.90	0.98	0.99	0.96	0.98	0.93
CDA+E	0.73	0.85	**0.99**	0.99	0.98	0.99	0.92
DANN	0.43	0.65	0.25	0.33	0.95	0.37	0.50
SFL-multi	0.72	0.88	0.89	0.89	0.87	0.87	0.85
Abl Exp 1	0.76	0.86	0.97	0.97	0.97	0.96	0.92
Abl Exp 2	0.64	0.76	0.98	0.99	0.96	0.97	0.89
Ours	**0.82**	**0.93**	0.98	**1**	**0.98**	**0.99**	**0.95**

TABLE S.3
FAULT DIAGNOSIS ACCURACY ON GEARBOX-GEAR

Methods	T1	T2	T3	T4	T5	T6	Average
Baseline	0.49	0.67	0.49	0.92	0.91	0.94	0.74
CORAL	0.91	0.93	0.97	0.93	0.95	0.96	0.94
Mk-MMD	0.96	0.97	0.97	0.98	0.98	0.99	0.97
CDA+E	0.70	0.88	0.91	0.98	0.99	0.98	0.91
DANN	0.91	0.74	0.31	0.29	0.89	0.49	0.60
SFL-multi	0.86	0.90	0.84	0.80	0.88	0.90	0.89
Abl Exp 1	0.94	0.97	0.97	0.98	0.98	0.98	0.97
Abl Exp 2	0.83	0.93	0.40	0.98	0.99	0.98	0.84
Ours	**0.97**	**0.97**	**0.97**	**0.99**	**0.99**	**0.99**	**0.98**

TABLE S.4
FAULT DIAGNOSIS ACCURACY WITH DIFFERENT SAMPLE OVERLAPPING RATIOS ON CWRU

Methods	0	0.2	0.5	1
Baseline	0.55	0.58	0.55	0.58
CORAL	0.47	0.54	0.56	0.58
Mk-MMD	0.52	0.53	0.56	0.56
CDA+E	0.53	0.52	0.55	0.57
DANN	0.15	0.43	0.46	0.47
SFL-multi	0.64	0.69	0.73	0.78
Abl Exp 1	0.71	0.73	0.67	0.79
Abl Exp 2	0.49	0.54	0.49	0.53
Ours	**0.77**	**0.78**	**0.79**	**0.79**

TABLE S.5
FAULT DIAGNOSIS ACCURACY WITH DIFFERENT SAMPLE OVERLAPPING RATIOS ON GEARBOX-GEAR

Methods	0	0.2	0.5	1
Baseline	0.74	0.77	0.80	0.77
CORAL	0.94	0.94	0.95	0.96
Mk-MMD	0.97	**0.98**	0.98	0.98
CDA+E	0.91	0.92	0.92	0.94
DANN	0.60	0.87	0.89	0.94
SFL-multi	0.86	0.91	0.93	0.94
Abl Exp 1	0.97	0.98	0.97	0.98
Abl Exp 2	0.84	0.87	0.86	0.89
Ours	**0.98**	0.98	**0.99**	**0.99**

TABLE S.6
FAULT DIAGNOSIS ACCURACY WITH DIFFERENT SAMPLE OVERLAPPING RATIOS ON GEARBOX-BEAR

Methods	0	0.2	0.5	1
Baseline	0.81	0.79	0.82	0.83
CORAL	0.91	0.92	0.91	0.93
Mk-MMD	0.93	0.95	**0.98**	0.97
CDA+E	0.92	0.95	0.96	0.98
DANN	0.50	0.91	0.89	0.88
SFL-multi	0.85	0.94	0.94	0.96
Abl Exp 1	0.92	0.93	0.96	0.96
Abl Exp 2	0.89	0.93	0.94	0.95
Ours	**0.95**	**0.96**	0.97	**0.99**

TABLE S.7
FAULT DIAGNOSIS ACCURACY WITH FULLY OVERLAPPED FEATURES

Methods	CWRU	GEAR	BEAR
Baseline	0.73	0.94	0.95
CORAL	0.54	0.9573	0.93
Mk-MMD	0.63	0.98	0.96
CDA+E	0.76	**1**	0.99
DANN	0.31	0.99	0.88
SFL-multi	0.84	0.98	0.99
Abl Exp 1	0.77	0.98	0.94
Abl Exp 2	0.56	0.85	0.99
Ours	**0.97**	0.99	**0.99**