CLASS-CONDITIONAL PROGRESSIVE CALIBRATION UNDER SELECTIVE DECOUPLED DOMAIN ADAPTATION FOR LONG-TAILED CROSS-CONDITION BEARING FAULT DIAGNOSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Rolling bearings are critical supporting components of rotating machinery and directly affect the safe and reliable operation of industrial equipment. Rolling bearing fault diagnosis remains challenging when operating-condition variations coexist with long-tailed fault distributions. Under such settings, uniform domain alignment may introduce unreliable transfer across operating conditions, while long-tailed class priors can bias target supervision toward head classes. To address these issues, we propose **S**elective **D**ecoupled **D**omain **A**daptation with **C**lass-**C**onditional **P**rogressive **C**alibration (SDDA-CCPC). Specifically, SDDA separates fault-discriminative and condition-related representations and performs reliability-weighted domain alignment to reduce the influence of unreliable cross-condition transfer. CCPC calibrates target predictions under long-tailed class priors and combines classifier–prototype agreement with progressive within-class selection to construct more reliable target supervision for minority classes. Across six long-tailed unsupervised cross-condition transfer tasks under different load conditions, SDDA-CCPC achieves an average Macro-F1 of **0.9178** and improves both overall diagnostic performance and tail-class recognition. For the extreme tail inner-race fault class, the average F1-score improves by **3.91%**. Ablation results further demonstrate the complementary contributions of SDDA and CCPC.

Keywords: rolling bearing fault diagnosis; cross-condition fault diagnosis; long-tailed learning; unsupervised domain adaptation; class-conditional reliability calibration

1 INTRODUCTION

Rolling bearings are critical supporting components in rotating machinery and are widely used in industrial equipment such as motors, gearboxes, wind turbines, and pumps. Their health conditions directly affect operational safety and system reliability. Deep learning-based methods have achieved substantial progress in bearing fault diagnosis owing to their ability to automatically learn nonlinear and discriminative representations from vibration signals (Chen et al., 2023; Zhang et al., 2020). However, most supervised diagnostic models implicitly assume similar training and test distributions. In practical applications, variations in load, speed, and operating environment can induce substantial distribution shifts across operating conditions, making representations learned from a source condition difficult to generalize to an unseen target condition. Reliable cross-condition diagnosis without target-domain labels therefore remains an important challenge (Chen et al., 2023; Li et al., 2025).

Unsupervised domain adaptation (UDA) provides a natural framework for mitigating distribution shifts between labeled source and unlabeled target domains. DAN reduces cross-domain feature discrepancy through maximum mean discrepancy (MMD) (Long et al., 2015), whereas DANN learns domain-invariant representations through adversarial training (Ganin et al., 2016). Subsequent approaches further incorporate category information into adaptation, including conditional adversarial alignment (Long et al., 2018) and subdomain matching (Zhu et al., 2021). In bearing fault diagnosis, recent bearing-diagnosis studies have improved conditional alignment (Wu et al., 2024a) and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

target-data utilization through pseudo-label-based strategies (Pang et al., 2024; Cai et al., 2024). More recently, MKWJDAN jointly adapts marginal and conditional distributions and introduces target pseudo-label rectification for cross-condition bearing diagnosis (Li et al., 2025). However, improved distribution alignment does not imply that all samples are equally reliable for adaptation. Operating-condition shifts can affect individual samples to different extents, and treating them equally during alignment may introduce unreliable transfer and weaken the preservation of fault-discriminative information.

The above cross-condition transfer challenge is further compounded by long-tailed class distributions. Conventional long-tailed learning methods alleviate class-frequency bias through re-weighting, class-dependent margins, or prior-based classifier correction (Cui et al., 2019; Cao et al., 2019; Menon et al., 2021). When class imbalance coexists with operating-condition shifts, however, both transferable representation learning and minority-class recognition become more difficult. Existing fault-diagnosis approaches address this problem through feature resampling, class-aware alignment, cost-sensitive learning, or decision-boundary adjustment (Ding et al., 2023; Wu et al., 2024b; Li et al., 2024; Zhou et al., 2025). Recent studies have further considered collaborative optimization of imbalanced cross-condition features and classification bias (Li et al., 2026), as well as class-aware large-margin learning together with conditional weighted adversarial adaptation under severe class imbalance (Kim et al., 2026). However, in an unlabeled target domain, supervision must ultimately be inferred from model predictions. Under long-tailed class priors, head classes tend to dominate high-confidence predictions, while useful tail-class samples may receive comparatively lower confidence. A single global confidence criterion can therefore bias target pseudo-supervision toward head classes.

Taken together, these challenges show that long-tailed cross-condition diagnosis requires more than simply reducing domain discrepancy or re-balancing class frequencies. At the representation-transfer level, individual samples can differ in their reliability for cross-condition alignment, making selective transfer preferable to indiscriminate alignment. At the target-supervision level, long-tailed class priors can bias pseudo-label selection toward head classes, requiring target-sample reliability to be evaluated in a class-conditional manner. Effective long-tailed cross-condition adaptation should therefore improve the reliability of representation transfer while reducing head-class bias in target pseudo-supervision. These two challenges are illustrated in Fig. 1.

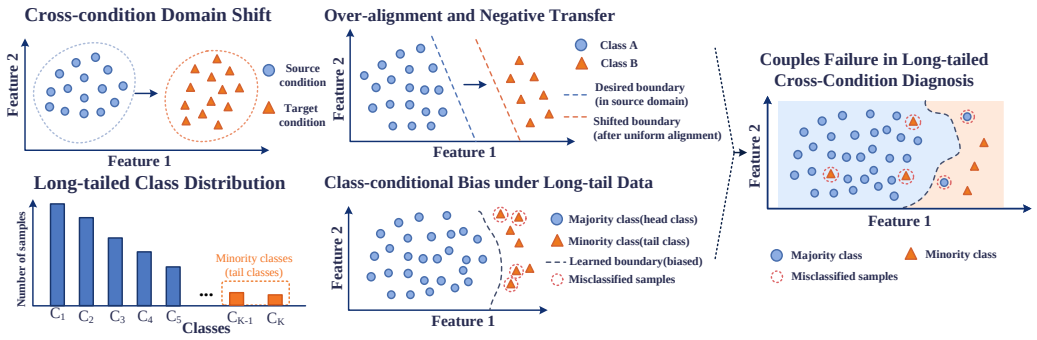


Figure 1: Key challenges in long-tailed cross-condition bearing fault diagnosis.

Motivated by these observations, we propose **S**elective **D**ecoupled **D**omain **A**daptation with **C**lass-**C**onditional **P**rogressive **C**alibration (SDDA-CCPC) for long-tailed unsupervised cross-condition bearing fault diagnosis. The proposed method addresses two coupled difficulties: heterogeneous sample transfer reliability across operating conditions and head-class-biased target pseudo-supervision under long-tailed class priors. Specifically, SDDA separates fault-discriminative and condition-related representations, uses sample transfer reliability to weight adversarial alignment on the condition-related representation, and further reduces information coupling between the two representations through an orthogonality constraint. CCPC calibrates target predictions according to the estimated target class prior, verifies prediction reliability through classifier–prototype agreement, and progressively selects reliable target samples within each predicted class. Consequently, target

samples are evaluated within their respective classes rather than competing across classes under a single confidence scale.

Our main contributions are summarized as follows:

- We propose **S**elective **D**ecoupled **D**omain **A**daptation (SDDA) to address heterogeneous sample transfer reliability during cross-condition adaptation. SDDA separates fault-discriminative and condition-related representations, uses sample-level transfer reliability to weight domain adversarial alignment, and employs an orthogonality constraint to reduce coupling between the two representations. This design limits the influence of unreliable samples during cross-condition alignment while preserving fault-discriminative information.
- We propose **C**lass-**C**onditional **P**rogressive **C**alibration (CCPC) to alleviate head-class bias in target pseudo-supervision under long-tailed distributions. CCPC integrates target-prior calibration, classifier–prototype agreement, and progressive within-class selection, shifting target-sample reliability assessment from cross-class confidence comparison to within-class evaluation and thereby constructing more reliable target supervision for under-represented classes.
- Systematic experiments on six long-tailed cross-condition transfer tasks demonstrate that SDDA-CCPC improves both overall diagnostic performance and the recognition of under-represented fault classes. Ablation results further confirm the individual effectiveness and complementary contributions of SDDA and CCPC.

2 Related Work

2.1 Domain Adaptation for Cross-Condition Fault Diagnosis

Unsupervised domain adaptation aims to learn transferable representations from a labeled source domain and an unlabeled target domain under distribution shift. Early deep UDA methods primarily focused on reducing global domain discrepancy. DAN minimizes distribution differences through maximum mean discrepancy (Long et al., 2015), whereas DANN learns domain-invariant representations through adversarial domain discrimination (Ganin et al., 2016). Later methods increasingly incorporated category information into adaptation. CDAN conditions adversarial alignment on classifier predictions (Long et al., 2018), DSAN performs fine-grained matching between class-related subdomains (Zhu et al., 2021), and MCC improves target discrimination by reducing class confusion in unlabeled target predictions (Jin et al., 2020).

Beyond direct distribution alignment, representation separation provides another perspective on transferable learning. Domain Separation Networks explicitly organize representations into domain-shared and domain-private components (Bousmalis et al., 2016), showing that domain-related factors and task-relevant information need not be modeled identically. This distinction is particularly relevant to vibration-based diagnosis, where changes in operating conditions and fault characteristics may affect learned representations in different ways. Therefore, reducing overall domain discrepancy alone does not necessarily guarantee reliable preservation of fault-discriminative information.

Recent studies have further specialized UDA to cross-condition bearing fault diagnosis. GCDAN introduces gradual conditional adversarial adaptation (Wu et al., 2024a), while pseudo-label-based approaches improve the utilization of unlabeled target samples (Pang et al., 2024; Cai et al., 2024). More recently, MKWJDAN dynamically combines marginal and conditional distribution alignment and rectifies target pseudo-labels for cross-condition diagnosis (Li et al., 2025). These studies have improved adaptation through more effective distribution matching, category information, and target-data utilization. However, most alignment mechanisms primarily focus on *how to reduce domain discrepancy*, while comparatively less attention is paid to whether different samples should contribute equally to alignment. Since operating-condition shifts can affect individual samples to different extents, indiscriminate alignment may allow unreliable samples to interfere with fault-discriminative representation learning. This motivates sample-reliability-aware representation transfer across operating conditions.

2.2 Long-Tailed Adaptation and Reliable Target Supervision

Long-tailed learning aims to mitigate classification bias caused by highly imbalanced class frequencies. Class-Balanced Loss re-weights categories according to the effective number of samples (Cui et al., 2019), whereas LDAM introduces class-dependent margins to improve minority-class decision boundaries (Cao et al., 2019). Decoupled classifier learning, Balanced Softmax, and Logit Adjustment further show that representation learning, class frequencies, and label priors can substantially affect long-tailed prediction bias (Kang et al., 2020; Ren et al., 2020; Menon et al., 2021).

When class imbalance coexists with operating-condition shifts, the problem becomes more difficult because transferable representation learning and minority-class recognition must be addressed simultaneously. DIDA combines cost-sensitive learning, categorical alignment, and margin regularization (Ding et al., 2023), while IRDDDA separates cross-condition representation learning from feature-resampling-based classifier calibration (Wu et al., 2024b). DACA introduces class-aware adaptation for extreme cross-domain imbalance (Li et al., 2024), and CIDSAN further addresses class-imbalanced cross-condition fault diagnosis through subdomain adaptation (Zhou et al., 2025). Recent studies continue this direction: Li et al. jointly consider imbalanced feature extraction, distribution alignment, and classification bias in UDA (Li et al., 2026), whereas Kim et al. combine class-aware large-margin learning with weighted adversarial adaptation under class imbalance (Kim et al., 2026). These methods demonstrate the importance of explicitly considering class imbalance during cross-condition adaptation, but mainly address it through re-balancing, class-aware alignment, or decision-boundary adjustment.

A complementary line of research focuses on obtaining reliable supervision from unlabeled target data. SHOT, FixBi, and AdaMatch exploit pseudo-labels, prediction confidence, or distribution alignment for target adaptation (Liang et al., 2020; Na et al., 2021; Berthelot et al., 2022), while prototype-based methods use class-level structure to support transferable prediction (Pan et al., 2019). In bearing diagnosis, uncertainty estimation has also been introduced to improve pseudo-label reliability (Liu et al., 2024). More recently, Wu et al. study source-free bearing fault diagnosis and show that both reliable and unreliable target predictions contain useful adaptation information (Wu et al., 2025). Together, these studies highlight the importance of prediction reliability and class structure when constructing target supervision.

Long-tailed cross-condition adaptation introduces an additional difficulty: prediction confidence is not necessarily comparable across classes. Under long-tailed class priors, head classes are more likely to dominate high-confidence predictions, whereas correct tail-class samples may remain relatively less confident. Consequently, selecting target samples using a single cross-class confidence scale can still bias pseudo-supervision toward frequent classes. This motivates class-conditional target supervision, where prediction bias is first calibrated, target candidates are verified using classifier–prototype agreement, and reliable samples are progressively selected within each predicted class rather than directly competing across classes.

3 Method

Long-tailed unsupervised cross-condition bearing fault diagnosis is challenged by two coupled issues. At the representation-transfer level, samples affected by operating-condition shifts exhibit heterogeneous transfer reliability, making indiscriminate domain alignment potentially harmful to fault-discriminative representations. At the target-supervision level, long-tailed class priors bias prediction confidence toward head classes, causing global confidence-based pseudo-label selection to favor head classes and under-select tail-class samples. We therefore propose **S**elective **D**ecoupled **D**omain **A**daptation with **C**lass-**C**onditional **P**rogressive **C**alibration (SDDA-CCPC). SDDA improves the reliability of representation transfer across operating conditions, while CCPC reduces head-class bias in target pseudo-supervision through class-conditional calibration and selection.The overall framework is illustrated in Fig. 2.

3.1 Selective Decoupled Domain Adaptation

Conventional domain-adversarial learning generally aligns source and target samples without explicitly distinguishing their sample-level transfer reliability. SDDA instead separates fault discrim-

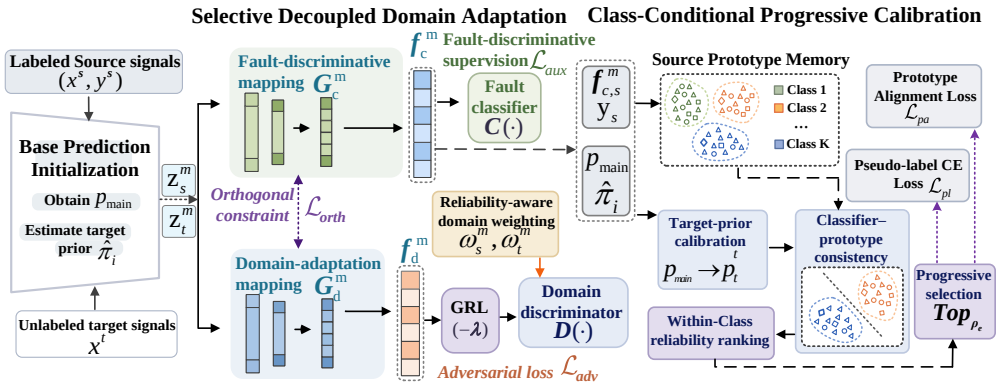


Figure 2: Overall framework of the proposed SDDA-CCPC.

ination from condition adaptation and further modulates domain alignment according to sample transfer reliability.

For the latent representation $z = E(x)$, given a labeled source domain $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and an unlabeled target domain $\mathcal{D}_t = \{x_j^t\}_{j=1}^{N_t}$,two mappings are introduced:

$$f_c = G_c(z), \qquad f_d = G_d(z), \tag{1}$$

where f_c is used for fault classification and f_d is used for domain adaptation. To reduce redundant information between the two branches, an orthogonality constraint $\mathcal{L}_{\text{orth}}$ is imposed on f_c and f_d.

To account for heterogeneous sample transferability, source and target reliability are defined as

$$r_i^s = p_{i,y_i^s}^s, \qquad r_j^t = 1 - \frac{H(\mathbf{p}_j^t)}{\log K}, \qquad \omega = \frac{r^\gamma}{\mathbb{E}[r^\gamma]}, \tag{2}$$

where the source reliability is the probability assigned to the ground-truth class, while the target reliability is measured by prediction certainty. The normalization is performed independently for the two domains.

These weights modulate domain-adversarial learning:

$$\mathcal{L}_{\text{adv}}^{\text{rel}} = -\mathbb{E}_s[\omega^s \log D(f_d^s)] - \mathbb{E}_t\left[\omega^t \log(1 - D(f_d^t))\right]. \tag{3}$$

Hence, samples with higher transfer reliability contribute more strongly to domain-adversarial alignment across operating conditions, whereas uncertain samples exert less influence. The SDDA objective is

$$\mathcal{L}_{\text{SDDA}} = \mathcal{L}_{\text{cls}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}}^{\text{rel}} + \lambda_{\text{orth}}\mathcal{L}_{\text{orth}}, \tag{4}$$

where $\mathcal{L}_{\text{cls}}$ denotes the standard supervised source classification loss.

3.2 Class-Conditional Progressive Calibration

SDDA improves the reliability of representation transfer, but target pseudo-supervision remains biased under long-tailed distributions. Since head classes tend to produce higher-confidence predictions, directly ranking all target samples on a common confidence scale can suppress useful tail-class samples. CCPC therefore evaluates target-sample reliability conditionally within each predicted class rather than relying on a global cross-class confidence criterion.

Let $\boldsymbol{\pi}_s$ denote the source prior and $\hat{\boldsymbol{\pi}}_t$ the target prediction prior tracked from unlabeled target predictions. Target probabilities are calibrated as

$$\tilde{p}_{i,k} = \frac{p_{i,k}^t(\pi_{s,k}/\hat{\pi}_{t,k})^\beta}{\sum_{\ell=1}^K p_{i,\ell}^t(\pi_{s,\ell}/\hat{\pi}_{t,\ell})^\beta}. \tag{5}$$

Meanwhile, source class prototypes $\mathcal{P} = \{\boldsymbol{\mu}_k\}_{k=1}^{K}$ are maintained from f_c using momentum updates. A target sample is considered a candidate only when the calibrated classifier prediction agrees with its nearest source prototype:

$$\mathcal{C}_k = \{i \mid \hat{y}_i^c = \hat{y}_i^p = k\}\,. \tag{6}$$

Within each candidate class, reliability is evaluated relatively rather than by an absolute cross-class confidence:

$$R_i = \alpha_c\ \mathrm{Rank}_k(c_i) + \alpha_m\ \mathrm{Rank}_k(m_i), \qquad \mathcal{R}_k^{(e)} = \mathrm{Top}_{\rho_e}\{R_i : i \in \mathcal{C}_k\}, \tag{7}$$

where c_i is the calibrated classifier confidence, m_i is the top-1/top-2 prototype-similarity margin, and $\mathrm{Rank}_k(\cdot)$ denotes relative ranking within class k. The retention ratio ρ_e gradually increases during training, allowing target supervision to expand from conservative to broader reliable sets.

The selected samples are used for confidence-weighted pseudo-label learning and class-conditional prototype alignment:

$$\mathcal{L}_{\mathrm{CCPC}} = \lambda_{\mathrm{pl}}\mathcal{L}_{\mathrm{pl}} + \lambda_{\mathrm{pa}}\mathcal{L}_{\mathrm{pa}}. \tag{8}$$

Thus, CCPC avoids direct cross-class confidence competition and progressively constructs more reliable class-conditional target pseudo-supervision.

3.3 Joint Optimization

SDDA and CCPC address the two coupled challenges from complementary perspectives: SDDA improves the reliability of representation transfer across operating conditions, whereas CCPC reduces head-class bias in target pseudo-supervision. CCPC is progressively activated using $a_e \in [0, 1]$, with $\rho_e = \rho_{\min} + a_e(\rho_{\max} - \rho_{\min})$. The overall objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{SDDA}} + a_e\mathcal{L}_{\mathrm{CCPC}}. \tag{9}$$

The core optimization procedure is summarized in Algorithm 1.

Algorithm 1 Core optimization of SDDA-CCPC

Require: $\mathcal{D}_s$, $\mathcal{D}_t$, classes K, epochs E
Ensure: Optimized SDDA-CCPC model
Initialize the model, target-prior tracker $\hat{\boldsymbol{\pi}}_t$, and prototype memory $\mathcal{P}$
for $e = 1, \ldots, E$ **do**
  Obtain $(f_c^s, f_d^s, f_c^t, f_d^t)$ by Eq. (1)
  Compute sample reliability by Eq. (2)
  Compute $\mathcal{L}_{\mathrm{SDDA}}$ by Eq. (4)
  Update $\hat{\boldsymbol{\pi}}_t$ and $\mathcal{P}$
  Calibrate target predictions by Eq. (5)
  for $k = 1, \ldots, K$ **do**
    Construct $\mathcal{C}_k$ by Eq. (6)
    Select $\mathcal{R}_k^{(e)}$ by Eq. (7)
  end for
  Compute $\mathcal{L}_{\mathrm{CCPC}}$ by Eq. (8)
  Update the model by minimizing Eq. (9)
end for

4 Experiments

4.1 Experimental Setup

Dataset and transfer tasks. We evaluate SDDA-CCPC on the Case Western Reserve University (CWRU) bearing dataset under four operating loads, denoted as T_0, T_1, T_2, and T_3 for 0, 1, 2, and 3 HP, respectively. The load variation produces cross-condition distribution shifts while preserving the same fault categories.

To evaluate adaptation under severe class imbalance, six health conditions are organized into a long-tailed distribution with an imbalance ratio of 40:1. In addition to the extreme tail class 21 mil IRF,

14 mil ORF is retained as a moderately under-represented class, allowing us to examine whether the method benefits minority faults at different imbalance levels. The resulting class distribution is reported in Table 1.

Using the four operating conditions, we construct six source-to-target transfer tasks covering different load shifts. For every task, training uses only labeled source data and unlabeled target data; target ground-truth labels are excluded from optimization and are used only for final evaluation. This protocol ensures that improvements cannot result from target-label information leakage.

Table 1: Long-tailed class distribution and cross-condition transfer tasks on CWRU.

(a) Long-tailed class distribution

	Normal	Ball	ORF-7	ORF-14	IRF-7	IRF-21
No. samples	50,000	25,000	12,500	5,000	10,000	1,250
Relative ratio	40	20	10	4	8	1

(b) Operating conditions and transfer tasks

Domain	T_0	T_1	T_2	T_3		
Load (HP)	0	1	2	3		
Tasks:	$T_0 \to T_1$	$T_0 \to T_2$	$T_0 \to T_3$	$T_1 \to T_2$	$T_1 \to T_3$	$T_2 \to T_3$

Table 2: Main experimental settings.

Setting	**Value**
Sampling frequency	12 kHz
Train/test split	80% / 20%
Split strategy	Chronological
Training windows	83,000
Test windows	20,750
Window length	20
Batch size	128
Optimizer	AdamW
Learning rate	1×10^{-3}
Weight decay	1×10^{-4}
Training epochs	160

Evaluation metrics and baselines. Macro-F1 is used as the primary metric to evaluate overall transfer performance under class imbalance, while class-wise Recall and F1-score are reported for under-represented classes to assess tail-fault recognition.

The baselines cover source-only learning, class re-balancing, conventional UDA, and imbalanced domain adaptation. CNN-based variants provide the source-only and re-balancing references, while DANN (Ganin et al., 2016), CDAN (Long et al., 2018), DSAN (Zhu et al., 2021), MCC (Jin et al., 2020), and IRDDDA (Wu et al., 2024b) represent established adaptation methods. Controlled SMOTE-, cost-sensitive-, and decoupling-based variants are further included to distinguish the effects of class re-balancing and representation decoupling from the proposed reliability-aware mechanisms. For brevity, the decoupling-based SMOTE and cost-sensitive variants are denoted as Dec-SMOTE and Dec-CS. All methods use the same data split and evaluation protocol; the main training settings are listed in Table 2.

4.2 Overall Cross-Condition Performance

Table 3 summarizes the Macro-F1 results over the six long-tailed cross-condition transfer tasks. SDDA-CCPC achieves the best performance on all six transfer directions, with an average Macro-F1 of **0.9178**. As further illustrated in Fig. 3, the improvement is consistently observed across different source–target operating-condition pairs rather than being driven by a particular transfer task.

The results also show that conventional class re-balancing does not consistently improve cross-condition diagnosis. Both SMOTE- and cost-sensitive variants can underperform their corresponding base models, indicating that correcting class frequency alone is insufficient when long-tailed imbalance is coupled with operating-condition shifts. In contrast, SDDA-CCPC maintains strong performance across different transfer directions. This overall trend is consistent with the motivation of the proposed method: long-tailed cross-condition adaptation requires not only reliable representation transfer across operating conditions, but also target supervision that is less biased toward head classes.

4.3 Tail-Class Performance Analysis

To examine whether the overall improvement also extends to under-represented faults, we further evaluate 14 mil ORF and 21 mil IRF, which represent moderate and extreme tail classes, respectively. Table 4 reports their average Recall and F1-score across the six transfer tasks, while Fig. 4 shows the corresponding F1-score distributions.

Table 3: Macro-F1 comparison on six long-tailed cross-condition transfer tasks. Best results are shown in bold and second-best results are underlined.

Method	$T_0 \to T_1$	$T_0 \to T_2$	$T_0 \to T_3$	$T_1 \to T_2$	$T_1 \to T_3$	$T_2 \to T_3$	Mean
CNN-base	0.9109	0.9045	0.8581	0.9185	0.8532	0.8768	0.8870
CNN-SMOTE	0.8820	0.8714	0.8282	0.8955	0.8008	0.8416	0.8533
CNN-CS	0.8903	0.8389	0.7553	0.8435	0.7408	0.8325	0.8169
DANN-base	0.9136	0.8945	0.8798	0.9132	0.8703	0.8800	0.8919
DANN-SMOTE	0.8682	0.8562	0.8256	0.8832	0.8051	0.8366	0.8458
DANN-CS	0.8994	0.8838	0.8506	0.9008	0.8444	0.8461	0.8709
Dec-SMOTE	0.8814	0.8675	0.8190	0.8463	0.7712	0.8033	0.8314
Dec-CS	0.8850	0.8672	0.8261	0.8525	0.7786	0.8143	0.8373
CDAN	0.9223	0.8941	0.8801	0.9076	0.8737	0.8869	0.8941
DSAN	0.9113	0.7659	0.7055	0.9201	0.6658	0.8498	0.8031
MCC	0.8665	0.7702	0.6753	0.8863	0.7715	0.8125	0.7971
IRDDDA	0.9119	0.9057	0.8849	0.9187	0.8883	0.8850	0.8991
SDDA-CCPC	**0.9317**	**0.9202**	**0.9005**	**0.9300**	**0.9079**	**0.9166**	**0.9178**

For the moderately under-represented 14 mil ORF class, SDDA-CCPC maintains competitive Recall while achieving the highest average F1-score. Its advantage becomes more pronounced for the extreme 21 mil IRF class, where the average F1-score reaches **0.7627**, exceeding IRDDDA by **3.91 percentage points**. The distributional comparison in Fig. 4 further shows that the favorable tail-class performance is observed across multiple transfer tasks rather than being driven by a single source–target condition pair.

These results indicate that the overall Macro-F1 improvement is not achieved by favoring head classes at the expense of under-represented faults. Instead, the performance gain extends to the classes most affected by long-tailed learning, particularly the extreme tail class. This behavior is consistent with the design objective of reducing head-class bias in target pseudo-supervision while maintaining reliable cross-condition transfer. The specific contributions of SDDA and CCPC are further examined through the ablation study in Section 4.4.

Table 4: Average tail-class Recall and F1-score across six transfer tasks.

	14 mil ORF		21 mil IRF	
Method	**Recall**	**F1**	**Recall**	**F1**
DANN-base	0.8514	0.8848	0.6817	0.5952
DANN-CS	**0.9008**	0.8917	0.7420	0.6069
Dec-SMOTE	0.8773	0.8663	0.7520	0.5975
CDAN	0.8568	0.8888	0.6823	0.6802
DSAN	0.8631	0.8472	0.6880	0.7156
MCC	0.8770	0.8546	0.7524	0.6923
IRDDDA	0.8847	0.8914	0.6573	0.7235
SDDA-CCPC	0.8943	**0.9066**	**0.7643**	**0.7627**

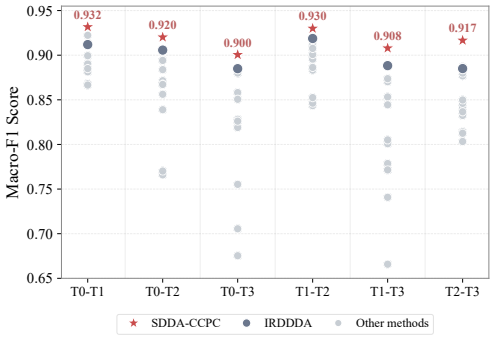


Figure 3: Performance comparison across six cross-condition transfer tasks.

4.4 Ablation and Component Analysis

To evaluate the individual and complementary contributions of SDDA and CCPC, four configurations are compared in Table 5, while Fig. 5 further reports their tail-class F1-score and Recall distributions.

Both A2 and A3 consistently outperform the base model A1, demonstrating the effectiveness of SDDA and CCPC when introduced independently. Their joint configuration further improves the overall diagnostic performance and achieves more stable results across different transfer tasks, indicating that the two components provide complementary rather than redundant benefits.

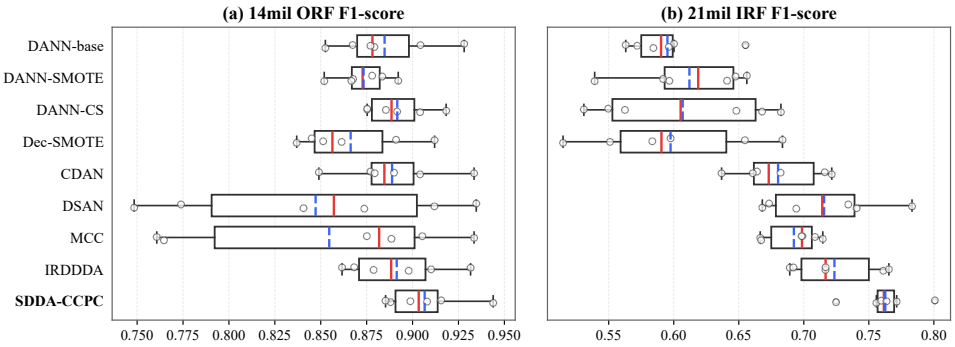
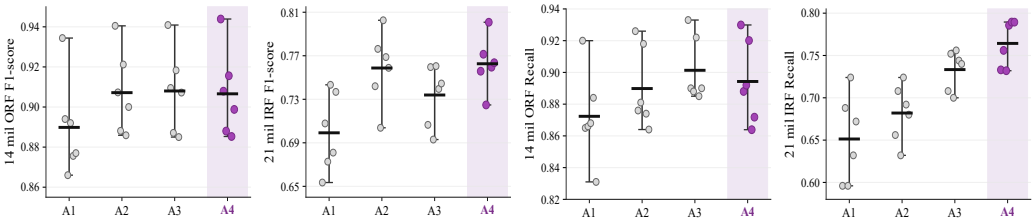


Figure 4: F1-score distributions of different methods on the tail classes 14 mil ORF and 21 mil IRF across six transfer tasks.

The tail-class results further reveal their different contributions. SDDA mainly improves the F1-score of the extreme 21 mil IRF class, while CCPC provides more evident gains in Recall. Combining both components achieves a better balance between these metrics, which is consistent with the intended roles of SDDA in improving reliable cross-condition representation transfer and CCPC in reducing head-class bias during target pseudo-supervision.

Table 5: Macro-F1 ablation results of SDDA and CCPC across six cross-condition transfer tasks.

Abl.	SDDA	CCPC	$T_0 \to T_1$	$T_0 \to T_2$	$T_0 \to T_3$	$T_1 \to T_2$	$T_1 \to T_3$	$T_2 \to T_3$	Mean
A1	–	–	0.9059	0.8967	0.8728	0.9071	0.8777	0.8901	0.8917
A2	✓	–	0.9276	0.9191	0.8938	**0.9328**	0.9019	0.9119	0.9145
A3	–	✓	0.9219	0.9171	0.8936	0.9257	0.8978	0.9129	0.9115
A4	✓	✓	**0.9317**	**0.9202**	**0.9005**	0.9300	**0.9079**	**0.9166**	**0.9178**



(a) F1-score (b) Recall

Figure 5: Tail-class performance of different ablation configurations: (a) F1-score and (b) Recall.

5 Conclusion

This work investigates long-tailed unsupervised cross-condition bearing fault diagnosis, where indiscriminate domain alignment can introduce unreliable transfer and long-tailed class priors can bias target pseudo-supervision toward head classes. To address these coupled challenges, SDDA-CCPC combines reliability-aware selective representation transfer with class-conditional progressive target calibration. Extensive experiments across multiple operating-condition shifts demonstrate consistent gains in overall diagnosis and tail-class recognition, while ablation results verify the complementary contributions of SDDA and CCPC. These findings indicate that effective long-tailed cross-condition adaptation should jointly consider sample-level transfer reliability and class-dependent target supervision rather than relying solely on domain alignment or class re-balancing.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing and LaTeX formatting. All AI-assisted edits were reviewed by the authors, who take full responsibility for the final content of the manuscript.

REFERENCES

David Berthelot, Rebecca Roelofs, Kihyuk Sohn, Nicholas Carlini, and Alex Kurakin. AdaMatch: A unified approach to semi-supervised learning and domain adaptation. In *International Conference on Learning Representations*, 2022.

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Keshen Cai, Chunlin Zhang, Pinfan Hou, Yanfeng Wang, Zhe Meng, and Wenbo Hou. A domain adaptation method for bearing fault diagnosis based on pseudo label and wavelet neural network. *IEEE Transactions on Instrumentation and Measurement*, 73:1–11, 2024. doi: 10.1109/TIM.2024.3462979.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Xiaohan Chen, Rui Yang, Yihao Xue, Mengjie Huang, Roberto Ferrero, and Zidong Wang. Deep transfer learning for bearing fault diagnosis: A systematic review since 2016. *IEEE Transactions on Instrumentation and Measurement*, 72:1–21, 2023. doi: 10.1109/TIM.2023.3244237.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9268–9277, 2019. doi: 10.1109/CVPR.2019.00949.

Yifei Ding, Minping Jia, Jichao Zhuang, Yudong Cao, Xiaoli Zhao, and Chi-Guhn Lee. Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions. *Reliability Engineering & System Safety*, 230:108890, 2023. doi: 10.1016/j.ress.2022.108890.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, Francois Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *Computer Vision – ECCV 2020*, pp. 464–480, 2020. doi: 10.1007/978-3-030-58589-1_28.

Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. In *International Conference on Learning Representations*, 2020.

Yongmin Kim, Yujun Yang, and Hyunsoo Yoon. Weighted adversarial learning with wavelet kernels and large margin networks for the cross-domain fault diagnosis of rolling bearings under a class imbalance. *Knowledge-Based Systems*, 335:115166, 2026. doi: 10.1016/j.knosys.2025.115166.

Sen Li, Xiaoqiang Zhao, and Jie Cao. An uda bearing fault diagnosis method based on synergistic optimization of unbalanced data. *Mechanical Systems and Signal Processing*, 243:113700, 2026. doi: 10.1016/j.ymssp.2025.113700.

Xin Li, Hao Chen, Shuhua Li, Dong Wei, Xiaoyu Zou, Lei Si, and Haidong Shao. Multi-kernel weighted joint domain adaptation network for cross-condition fault diagnosis of rolling bearings. *Reliability Engineering & System Safety*, 261:111109, 2025. doi: 10.1016/j.ress.2025.111109.

Yuanjiang Li, Ying Zhu, Yang Yu, Runze Mao, Linchang Ye, Yun Liu, Ruochen Liu, Tao Lang, and Jinglin Zhang. DACA: A domain adaptive fault diagnosis approach with class-aware based on cross-domain extreme imbalance data. *Expert Systems with Applications*, 256:124944, 2024. doi: 10.1016/j.eswa.2024.124944.

Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 6028–6039, 2020.

Ruiqi Liu, Wengang Ma, Feipeng Kuang, Jin Guo, and Ning Zhao. A source free robust domain adaptation approach with pseudo-labels uncertainty estimation for rolling bearing fault diagnosis under limited sample conditions. *Knowledge-Based Systems*, 304:112443, 2024. doi: 10.1016/j.knosys.2024.112443.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I. Jordan. Learning transferable features with deep adaptation networks. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pp. 97–105, 2015.

Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. In *International Conference on Learning Representations*, 2021.

Jaemin Na, Heechul Jung, Hyung Jin Chang, and Wonjun Hwang. Fixbi: Bridging domain spaces for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1094–1103, 2021.

Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2239–2247, 2019. doi: 10.1109/CVPR.2019.00234.

Bin Pang, Qiuhai Liu, Zhenduo Sun, Zhenli Xu, and Ziyang Hao. Time-frequency supervised contrastive learning via pseudo-labeling: An unsupervised domain adaptation network for rolling bearing fault diagnosis under time-varying speeds. *Advanced Engineering Informatics*, 59:102304, 2024. doi: 10.1016/j.aei.2023.102304.

Jiawei Ren, Cunjun Yu, Shunan Sheng, Xiao Ma, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Balanced meta-softmax for long-tailed visual recognition. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Chu-Ge Wu, Duo Zhao, Te Han, and Yuanqing Xia. Bearing fault diagnosis using gradual conditional domain adversarial network. *Applied Soft Computing*, 158:111580, 2024a. doi: 10.1016/j.asoc.2024.111580.

Wenyi Wu, Hao Zhang, Zhisen Wei, Xiao-Yuan Jing, Qinghua Zhang, and Songsong Wu. Both reliable and unreliable predictions matter: Domain adaptation for bearing fault diagnosis without source data. *Neurocomputing*, 657:131661, 2025. doi: 10.1016/j.neucom.2025.131661.

Zhenyu Wu, Juchuan Guo, Yichen Liu, Linjie Li, and Yang Ji. An iterative resampling deep decoupling domain adaptation method for class-imbalance bearing fault diagnosis under variant working conditions. *Expert Systems with Applications*, 252:124240, 2024b. doi: 10.1016/j.eswa.2024.124240.

Shen Zhang, Shibo Zhang, Bingnan Wang, and Thomas G. Habetler. Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, 8:29857–29881, 2020. doi: 10.1109/ACCESS.2020.2972859.

Jianyu Zhou, Xiangfeng Zhang, Hong Jiang, Jun Li, and Zhenfa Shao. Cross-domain transfer fault diagnosis by class-imbalanced deep subdomain adaptive network. *Measurement*, 242:115901, 2025. doi: 10.1016/j.measurement.2024.115901.

Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, and Qing He. Deep subdomain adaptation network for image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4):1713–1722, 2021. doi: 10.1109/TNNLS.2020.2988928.

A Additional Method Details

This section provides mathematical details that are omitted from the main text for conciseness. We focus only on the components required to complete the formulation of reliability-aware representation transfer and class-conditional target supervision; equations already defined in Section 3 are not repeated.

A.1 Orthogonality Regularization

As described in Section 3.1, SDDA separates the fault-discriminative representation f_c from the condition-related adaptation representation f_d. To discourage the two branches from encoding redundant information, we minimize their sample-wise cosine correlation:

$$\mathcal{L}_{\text{orth}} = \frac{1}{B_s + B_t} \sum_{x \in \mathcal{B}_s \cup \mathcal{B}_t} \left| \frac{f_c(x)^\top f_d(x)}{\|f_c(x)\|_2 \|f_d(x)\|_2} \right|. \tag{10}$$

This constraint does not force either representation to discard information independently. Instead, it reduces their directional overlap, allowing f_c to emphasize fault discrimination while f_d provides a separate space for reliability-weighted domain adaptation.

A.2 Target-Prior Tracking

The prior calibration in Eq. equation 5 requires an estimate of the evolving target prediction prior. We initialize the target prior tracker from the source class prior and update it using unlabeled target predictions:

$$\hat{\boldsymbol{\pi}}_t \leftarrow \mu_\pi \hat{\boldsymbol{\pi}}_t + (1 - \mu_\pi) \frac{1}{B_t} \sum_{j=1}^{B_t} \mathbf{p}_j^t, \tag{11}$$

followed by normalization over the K classes. The moving estimate therefore reflects the target prediction distribution without accessing target ground-truth labels. Combined with Eq. equation 5, it provides the class-prior correction used before reliable target samples are selected.

A.3 Prototype Memory and Reliability Verification

CCPC uses source prototypes as a second prediction view for verifying target reliability. For each source class k, the normalized batch centroid is

$$\bar{\boldsymbol{\mu}}_k = \text{Norm}\left(\frac{1}{|\mathcal{B}_s^k|} \sum_{i: y_i^s = k} f_{c,i}^s \right), \tag{12}$$

and the prototype memory is updated by

$$\boldsymbol{\mu}_k \leftarrow \text{Norm}\left(\mu_p \boldsymbol{\mu}_k + (1 - \mu_p) \bar{\boldsymbol{\mu}}_k \right). \tag{13}$$

For a target representation $f_{c,i}^t$, cosine similarity to the source prototype defines

$$s_{i,k} = \cos(f_{c,i}^t, \boldsymbol{\mu}_k). \tag{14}$$

The prototype margin used in the class-conditional reliability score is

$$m_i = s_i^{(1)} - s_i^{(2)}, \tag{15}$$

where $s_i^{(1)}$ and $s_i^{(2)}$ are the largest and second-largest prototype similarities. A large margin indicates that the target sample has a clearer class-level relation to the source prototype structure. Together with classifier confidence, this provides the two complementary signals used by Eq. equation 7.

A.4 Progressive Target Supervision

Target pseudo-supervision is introduced progressively because target predictions are less reliable during the early stage of adaptation. The activation coefficient is defined as

$$a_e = \mathrm{clip}\left(\frac{e - e_s}{e_f - e_s}, 0, 1\right), \tag{16}$$

where e_s and e_f denote the starting and full-activation epochs, respectively. The class-wise retention ratio described in Section 3.2 increases accordingly during training.

For the selected target set $\mathcal{R}^{(e)}$, CCPC applies a confidence-weighted pseudo-label loss:

$$\mathcal{L}_{\mathrm{pl}} = \frac{\sum_{i \in \mathcal{R}^{(e)}} c_i^{\eta} \ell_{\mathrm{CE}}(\mathbf{p}_i^t, \hat{y}_i)}{\sum_{i \in \mathcal{R}^{(e)}} c_i^{\eta}}, \tag{17}$$

where c_i denotes the calibrated classifier confidence.

To preserve class-conditional cross-domain structure, the selected target features are additionally aligned with the corresponding source class centroids:

$$\mathcal{L}_{\mathrm{pa}} = \frac{1}{|\mathcal{K}_e|} \sum_{k \in \mathcal{K}_e} \left[1 - \cos\left(\mathbf{c}_k^s, \mathbf{c}_k^t\right)\right], \tag{18}$$

where $\mathcal{K}_e$ contains classes with valid source and selected target samples. These two terms instantiate the CCPC objective defined in Eq. equation 8: the former introduces reliable target pseudo-supervision, while the latter preserves class-level structure during its progressive expansion.

B Experimental Protocol and Implementation Details

This section provides additional details of the experimental protocol to ensure reproducible evaluation. In particular, we describe the temporal data partition strategy, long-tailed distribution construction, unsupervised domain adaptation protocol, and evaluation metrics used throughout the experiments.

B.1 Chronological Train–Test Partition

To avoid temporal leakage between training and evaluation samples, the raw vibration sequence of each operating condition is first divided into training and test segments using an 80%/20% chronological split. Window generation is performed only after this partition, and no generated window is allowed to cross the train–test boundary.

A fixed window length of $L = 20$ is adopted for all experiments. Sampling positions are generated without replacement to avoid duplicated samples during window construction. Following this procedure, each operating condition contains 83,000 training windows and 20,750 test windows.

Different from randomly partitioning already generated overlapping windows, the proposed protocol prevents highly similar signal segments from appearing in both training and test sets. Therefore, the evaluation better reflects generalization to unseen vibration segments under different operating conditions.

B.2 Long-Tailed Distribution Construction

Six fault categories are considered in the experiments: Normal, Ball Fault, 7 mil ORF, 14 mil ORF, 7 mil IRF, and 21 mil IRF. To simulate realistic long-tailed fault distributions, the training samples of each operating condition are constructed according to the following class ratio:

$$40 : 20 : 10 : 4 : 8 : 1. \tag{19}$$

The resulting imbalance ratio between the largest and smallest classes is 40:1. This long-tailed construction protocol is consistently applied to all operating conditions and transfer tasks.

B.3 Unsupervised Domain Adaptation Protocol

For each cross-condition transfer task $T_s \rightarrow T_t$, the labeled source training set $\mathcal{D}_s^{\text{train}}$ and unlabeled target training set $\mathcal{D}_t^{\text{train}}$ are used for optimization.

The target labels are not available during training and are not used for pseudo-label generation, prototype construction, parameter updating, or loss computation. The target test set $\mathcal{D}_t^{\text{test}}$ is strictly held out and is only used for final performance evaluation.

All comparison methods are trained and evaluated under the same source–target partition and long-tailed distribution setting to ensure a fair comparison.

B.4 Evaluation Metrics

Macro-F1 is adopted as the primary metric for overall diagnostic performance, because it assigns equal importance to each fault category regardless of sample frequency. To further evaluate minority-class recognition, class-wise Recall and F1-score are additionally reported for 14 mil ORF and 21 mil IRF, which represent moderate and extreme tail conditions, respectively.

These metrics jointly evaluate whether the proposed method improves overall cross-condition transferability while maintaining reliable recognition of under-represented fault categories.

B.5 Network Configuration

The feature extractor outputs a 64-dimensional latent representation, which is further processed by the classifier for fault prediction. The classifier consists of three fully connected layers with dimensions 64-32-16-6. The remaining optimization settings follow the configurations described in Section 4.1.

C Additional Overall Cross-Condition Results

The main paper reports the complete task-wise Macro-F1 results in Table 3. Figure 6 provides an additional aggregated comparison across the six transfer tasks. Different from the task-specific evaluation, this analysis summarizes the overall adaptation behavior under multiple operating-condition shifts.

The results show that SDDA-CCPC maintains a consistent advantage over existing methods in the aggregated cross-condition setting. This observation indicates that the improvement is not limited to a particular transfer direction, but results from the joint enhancement of reliable representation transfer and class-conditional target supervision.

D Detailed Tail-Class Results

The main paper reports the six-task average Recall and F1-score for 14 mil ORF and 21 mil IRF. This section provides the corresponding task-wise results to examine whether the tail-class improvement is consistently maintained across different operating-condition shifts.

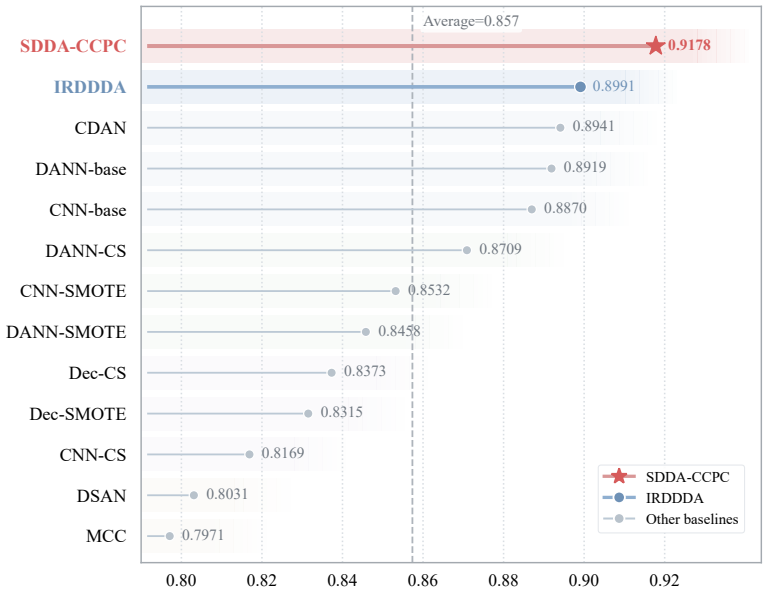


Figure 6: Average Macro-F1 comparison across six long-tailed cross-condition transfer tasks.

Table 6: Task-wise recall for the two tail classes across six transfer tasks.

Method	**Tail class**	$\mathbf{T_0 \to T_1}$	$\mathbf{T_0 \to T_2}$	$\mathbf{T_0 \to T_3}$	$\mathbf{T_1 \to T_2}$	$\mathbf{T_1 \to T_3}$	$\mathbf{T_2 \to T_3}$
DANN-base	14 mil ORF	0.9103	0.8609	0.8481	0.9083	0.8225	0.7584
	21 mil IRF	0.6278	0.6805	0.6466	0.7481	0.7068	0.6805
DANN-CS	14 mil ORF	**0.9390**	**0.9290**	**0.8970**	**0.9380**	0.8740	0.8280
	21 mil IRF	0.7000	0.7400	0.6840	0.8080	0.7560	0.7640
Dec-SMOTE	14 mil ORF	0.9380	0.9240	0.8700	0.9160	0.8110	0.8050
	21 mil IRF	0.7000	0.7400	0.7080	**0.8200**	0.7640	0.7800
CDAN	14 mil ORF	0.9142	0.8846	0.8481	0.8974	0.8343	0.7623
	21 mil IRF	0.6579	0.6090	0.6165	0.7744	0.7105	0.7256
DSAN	14 mil ORF	0.9310	0.8876	0.8511	0.9339	0.7702	0.8047
	21 mil IRF	0.6353	0.6654	0.6504	0.6992	0.7520	0.7256
MCC	14 mil ORF	0.9339	0.8817	0.8511	0.9191	0.8609	0.8156
	21 mil IRF	0.7068	**0.7744**	0.7068	0.7850	0.7519	0.7893
IRDDDA	14 mil ORF	0.9060	0.8840	0.8500	0.9160	**0.8920**	0.8600
	21 mil IRF	0.6120	0.6880	0.6560	0.6720	0.6680	0.6480
SDDA-CCPC	14 mil ORF	0.9299	0.8880	0.8920	0.9201	0.8639	**0.8718**
	21 mil IRF	**0.7330**	0.7560	**0.7320**	0.7857	**0.7895**	**0.7894**

D.1 Task-Wise Recall

Table 6 reports the Recall of the two under-represented fault classes for each transfer task. The extreme 21 mil IRF class exhibits larger variation across methods, indicating that severe class imbalance makes minority-fault recognition more sensitive to the source–target condition pair. SDDA-CCPC maintains competitive Recall across the six tasks, supporting the aggregate tail-class results reported in the main paper.

D.2 Task-Wise F1-Score

Table 7 reports the corresponding task-wise F1-scores. Compared with Recall, F1-score additionally reflects the balance between missed detections and false alarms. The results show that the advantage

Table 7: Task-wise F1-score for the two tail classes across six transfer tasks.

Method	Tail class	$T_0 \to T_1$	$T_0 \to T_2$	$T_0 \to T_3$	$T_1 \to T_2$	$T_1 \to T_3$	$T_2 \to T_3$
DANN-base	14 mil ORF	0.9281	0.8792	0.8771	0.9043	0.8674	0.8525
	21 mil IRF	0.6002	0.5962	0.5719	0.6552	0.5631	0.5843
DANN-CS	14 mil ORF	0.9183	0.8916	0.8855	0.9041	0.8753	0.8753
	21 mil IRF	0.6481	0.6679	0.5625	0.6824	0.5309	0.5496
Dec-SMOTE	14 mil ORF	0.9120	0.8910	0.8614	0.8513	0.8369	0.8451
	21 mil IRF	0.6547	0.6836	0.5510	0.5835	0.5149	0.5976
CDAN	14 mil ORF	0.9335	0.8899	0.8771	0.9041	0.8794	0.8490
	21 mil IRF	0.6368	0.6821	0.6640	0.7215	0.6608	0.7161
DSAN	14 mil ORF	0.9347	0.8406	0.7484	0.9119	0.7739	0.8737
	21 mil IRF	0.6733	0.7406	0.6681	0.7832	0.7342	0.6942
MCC	14 mil ORF	0.9335	0.7607	0.7646	0.9053	**0.8885**	0.8751
	21 mil IRF	0.6665	0.7086	0.6672	0.6989	0.7146	0.6983
IRDDDA	14 mil ORF	0.9316	0.8979	0.8616	0.9101	0.8788	0.8682
	21 mil IRF	0.6892	0.7611	0.6920	0.7654	0.7167	0.7168
SDDA-CCPC	14 mil ORF	**0.9439**	**0.9079**	**0.8880**	**0.9156**	0.8853	**0.8988**
	21 mil IRF	**0.7558**	**0.7714**	**0.7248**	**0.8008**	**0.7595**	**0.7636**

on the extreme tail class is maintained across multiple transfer directions rather than being driven by a single favorable task.

D.3 Additional Tail-Class Visualization

Figure 7 complements the numerical results by visualizing how tail-class F1-score varies across different source–target operating-condition pairs. The larger variation observed for 21 mil IRF further illustrates the sensitivity of the extreme tail class to cross-condition adaptation.

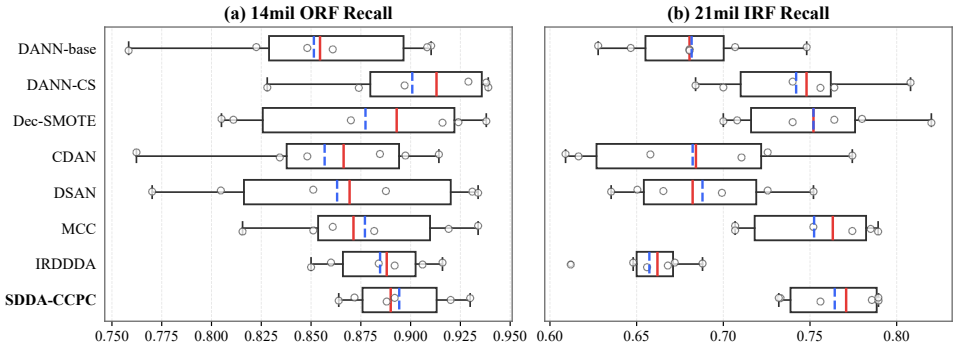


Figure 7: Task-wise F1-score comparison for the two tail classes across six transfer tasks.

E Additional Ablation Analysis

The main paper reports the aggregate ablation results of SDDA and CCPC. This section provides the complete task-wise results to further examine their individual and complementary effects across different operating-condition shifts. A1 denotes the base DANN model, A2 adds SDDA, A3 adds CCPC, and A4 denotes the complete SDDA-CCPC model.

E.1 Task-Wise Macro-F1 Ablation Results

Figure 8 compares the four configurations across the six transfer tasks. Both SDDA and CCPC improve upon the base model when introduced independently, while their relative gains vary across

source–target pairs. Their combination provides the strongest overall performance, supporting the complementary behavior observed in the main paper.

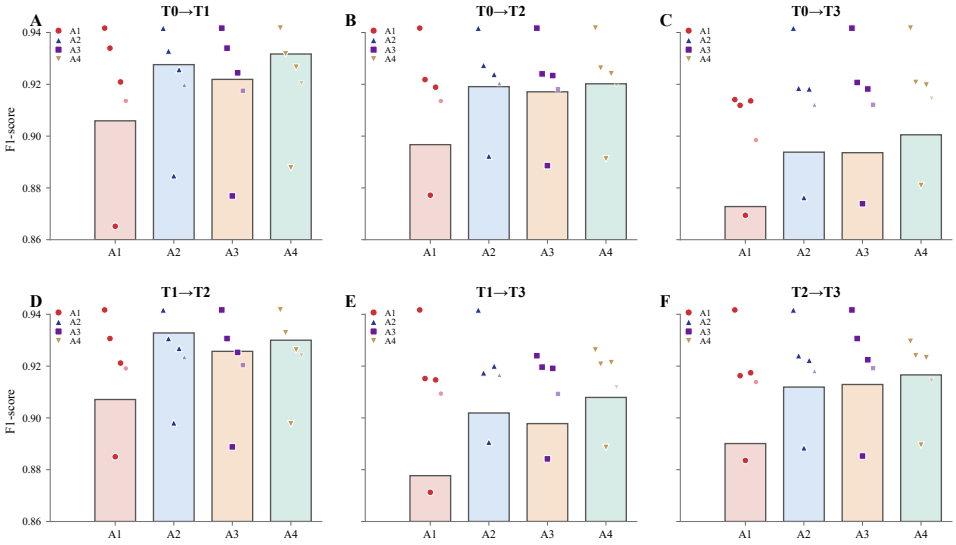


Figure 8: Macro-F1 ablation results across six cross-condition transfer tasks.

E.2 Task-Wise Tail-Class F1-Score Ablation

Table 8 reports the task-wise F1-score results for the two tail classes. The best configuration varies across tasks for 14 mil ORF, whereas the benefit of SDDA and the complete model becomes more evident for the extreme 21 mil IRF class. This behavior is consistent with the increasing difficulty of reliable minority-class learning under more severe imbalance.

Table 8: Task-wise F1-score ablation results for the two tail classes.

Abl.	**Tail class**	$\mathbf{T_0 \to T_1}$	$\mathbf{T_0 \to T_2}$	$\mathbf{T_0 \to T_3}$	$\mathbf{T_1 \to T_2}$	$\mathbf{T_1 \to T_3}$	$\mathbf{T_2 \to T_3}$
A1	14 mil ORF	0.9344	0.8940	0.8660	0.8920	0.8756	0.8770
	21 mil IRF	0.6535	0.7078	0.6727	0.7433	0.6810	0.7368
A2	14 mil ORF	0.9405	0.9073	**0.8881**	**0.9212**	**0.8859**	0.9000
	21 mil IRF	0.7420	**0.7763**	0.7038	**0.8026**	**0.7688**	0.7589
A3	14 mil ORF	0.9409	**0.9095**	0.8871	0.9183	0.8850	**0.9072**
	21 mil IRF	0.7065	0.7596	0.6930	0.7605	0.7395	0.7445
A4	14 mil ORF	**0.9439**	0.9079	0.8880	0.9156	0.8853	0.8988
	21 mil IRF	**0.7558**	0.7714	**0.7248**	0.8008	0.7595	**0.7636**

E.3 Task-Wise Tail-Class Recall Ablation

Table 9 reports the corresponding Recall results. A3 performs strongly on the moderately under-represented 14 mil ORF class, whereas A4 achieves the highest Recall for 21 mil IRF across all six transfer tasks. This pattern further supports the complementary roles of SDDA and CCPC, particularly under the most severe tail condition.

Table 9: Task-wise Recall ablation results for the two tail classes.

Abl.	Tail class	$T_0 \to T_1$	$T_0 \to T_2$	$T_0 \to T_3$	$T_1 \to T_2$	$T_1 \to T_3$	$T_2 \to T_3$
A1	14 mil ORF	0.9200	0.8650	0.8660	0.8680	0.8840	0.8310
	21 mil IRF	0.5960	0.6880	0.5960	0.7240	0.6320	0.6720
A2	14 mil ORF	0.9260	0.8760	0.8810	0.9180	0.8740	0.8640
	21 mil IRF	0.6560	0.7080	0.6320	0.7240	0.6920	0.6800
A3	14 mil ORF	**0.9330**	**0.8900**	0.8880	**0.9220**	**0.8850**	**0.8900**
	21 mil IRF	0.7080	0.7520	0.7000	0.7560	0.7440	0.7400
A4	14 mil ORF	0.9299	0.8880	**0.8920**	0.9201	0.8639	0.8718
	21 mil IRF	**0.7330**	**0.7560**	**0.7320**	**0.7857**	**0.7895**	**0.7894**