

Article | Received 30 July 2026; Revised 17 September 2026; Accepted 12 October 2026; Published XX October 2026
<https://doi.org/10.55092/mt20260007>



Cross-working condition fault diagnosis via multi-classifier discrepancy-driven learning and confidence pseudo-label self-supervision

Xingchi Lu¹, Yuguang Li^{1,2}, Liuyang Song¹ and Huaqing Wang^{1,*}

¹ College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing, China

² Refining, Chemicals & New Materials Company, PetroChina, Beijing, China

* Correspondence author; E-mail: hqwang@buct.edu.cn.

Highlights:

- A novel MuCD framework is developed by extending MCD for improved target-domain representation.
- A CPSL mechanism is developed to improve the boundary discrimination of fault classes in the target domain.
- A CPSL-MuCD transfer fault diagnosis model is proposed by integrating CPSL into the MuCD framework.

Abstract: The maximum classifier discrepancy domain adaptation (DA) method based on a dual-classifier structure has achieved some success in fault diagnosis of cross-working condition, yet its reliance on the dual-classifier structure may fail to effectively represent the data characteristics of the target domain and may suffers from insufficient classification boundary discrimination. To address this, this paper extends the aforementioned structure to multiple classifiers and proposes a multi-classifier discrepancy (MuCD)-driven DA approach, which enhances the representation of target domain data characteristics, through exploiting the discrepancies among multiple classifier pairs. Additionally, given the uncertainty of multi-classifier outputs, the harmonic mean is calculated and a screening threshold is set. Combined with the self-supervised learning losses for target domain, a confidence pseudo-label strategy-guided self-supervised learning (CPSL) mechanism is designed, to guide and enhance discriminative learning among different fault classes in target domain, thereby improving the model's diagnosis performance. Furthermore, by specifically integrating CPSL into the MuCD framework, a cross-working condition fault diagnosis method based on multi-classifier discrepancy-driven learning enhanced by confidence pseudo-label self-supervision (CPSL-MuCD) is proposed. Finally, comparative experiments and ablation studies on bearing fault, gear fault, and bearing-gear compound fault datasets verify the effectiveness and generalizability of the proposed CPSL-MuCD.

Keywords: cross-working condition; fault diagnosis; rotating machinery component; multi-classifier discrepancy; confidence pseudo-label



Copyright©2026 by the authors. Published by ELSP. This work is licensed under Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

1. Introduction

The health status of rotating components such as rolling bearings and gears is critical to the safe and reliable operation of rotating machinery. However, as vulnerable components, they frequently operate under complex and variable working conditions in mechanical systems [1,2]. Accordingly, it is essential to conduct research on condition monitoring and fault diagnosis for key rotating components, so as to ensure the safe and reliable operation of mechanical systems as well as economic benefits [3]. Against the current backdrop of big data and the Industrial Internet of Things (IIoT), data-driven deep intelligent fault diagnosis methods, which can automatically extract fault features and realize end-to-end diagnosis by training deep learning models solely based on monitoring data without relying on prior expert knowledge [4,5], have received extensive research and application [6,7].

However, deep intelligent diagnosis models are constrained by the inherent paradigm of deep learning, and can only perform satisfactorily under two core assumptions: the training (source domain) and testing (target domain) data follow the same distribution, and labeled data are available during model training [8,9]. In practical industrial scenarios, mechanical equipment is frequently affected by varying working conditions, and manual annotation of monitoring data is notoriously difficult, rendering the aforementioned assumptions largely unfeasible. Consequently, the diagnosis models suffer from limited generalization and fail to achieve reliable cross-working condition fault diagnosis [10].

Domain adaptation (DA), as a branch of transfer learning, provides a promising solution to the aforementioned cross-working condition fault diagnosis problem [11,12]. Existing DA approaches can be broadly categorized into two paradigms: minimizing cross-domain feature distribution discrepancy via metric functions, and generating domain-invariant features through adversarial learning [13]. Explicit-metric-based DA methods compute and reduce cross-domain distribution discrepancy by means of kernel functions or statistical metrics, thereby guiding the model to learn domain-invariant representations [14]. Typical representatives include multi-kernel maximum mean discrepancy (MK-MMD) and local maximum mean discrepancy (LMMD). For instance, Lu *et al.* [15] proposed a semi-supervised contrastive domain adaptation network (SSCDAN), which promotes LMMD learning of cross-domain features through intra-domain semi-supervised contrastive learning, thus improving the alignment of cross-domain feature distributions. Lu *et al.* [16] designed a hierarchical shared multi-kernel correlation and centroid feature discrepancy network (HSMK-CCFD), which combines MK-MMD and multi-kernel correlation alignment (MK-CORAL) to construct a joint distribution alignment mechanism, and further develops an adaptive loss weighting strategy based on kernel quality evaluation. Adversarial-training-based DA methods, by contrast, drive the feature extractor to produce indistinguishable features, that is, domain-invariant features, through a zero-sum game between the feature extractor and the domain discriminator [17]. Representative methods include the domain-adversarial neural network (DANN) and the conditional domain-adversarial network (CDAN). For example, Wen *et al.* [18] proposed a clustering graph convolutional network with multiple adversarial learning (*c*-GCN-MAL), which enhances clustering and transfer capabilities through data-adversarial and domain-adversarial learning for cross-domain bearing fault diagnosis. Chen *et al.* [19] developed a dual-adversarial-guided unsupervised multi-domain adaptation network (DAG-MDAN) that can more effectively integrate fault diagnosis knowledge from multiple source domains.

Among these, beyond the conventional adversarial training between a domain discriminator and a feature extractor, Saito *et al.* [20] proposed a dual-classifier adversarial paradigm called maximum classifier discrepancy (MCD). This paradigm replaces the domain discriminator with task-specific classifiers to conduct adversarial training with the feature extractor. Compared with conventional paradigms, MCD accounts for the relationship between the decision boundary and target samples, making it more consistent with target-domain characteristics when aligning feature distributions and implementing DA. Nevertheless, when MCD is practically applied to cross-working condition diagnosis of rotating machinery component, it still exhibits certain limitations in learning the boundaries between different fault classes. During the process of shrinking the decision boundaries, ambiguous boundary discrimination may occur, as shown in Figure 1a. Accordingly, Jiang *et al.* [21] developed a dual-classifier domain adaptation diagnosis model based on self-supervised learning (SLDDA), which alleviates the output ambiguity of classifiers via classification certainty metric and constructs target-data-clustering-based self-supervised learning to enhance discrimination among different fault classes. Jin *et al.* [22] proposed a dual-classifier adversarial learning network for cross-domain fault diagnosis, which integrates conditional entropy and local Lipschitz terms into the loss function to push decision boundaries away from data-dense regions, thus strengthening the discriminability of different fault classes.

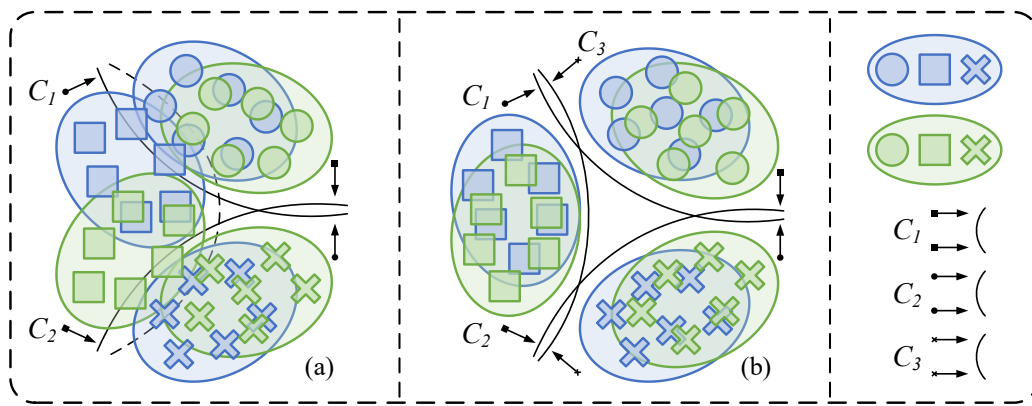


Figure 1. Motivation for the proposed method: **(a)** Limitations of MCD adversarial framework; **(b)** Multi-classifier discrepancy (MuCD)-driven and CPSL-enhanced.

This paper constructs a confidence pseudo-labeling strategy to guide self-supervised learning of the target domain, enabling targeted learning of the discrimination on target domain to improve the discrimination and shrinking of boundaries between different fault classes. While focusing on improving the discrimination of boundaries between different fault classes through self-supervised learning of the target domain, this paper further considers that MCD relies only on discrepancy learning between a single pair of classifiers, making its results susceptible to the decision bias of that classifier pair and thus posing a potential risk of insufficient representation of target-domain data characteristics. This risk may also affect the model's discrimination of classification boundaries, as shown in Figure 1a. Therefore, the dual-classifier adversarial paradigm of MCD is extended to discrepancy learning among multiple classifiers. By constructing multiple sets of decision boundaries formed by different classifiers, the uncertainty of target-domain samples can be characterized from multiple discriminative perspectives, thereby driving a more sufficient representation of target-domain data characteristics and further improving boundary discrimination and shrinking. Consequently, by organically combining the two

improvements described above, the ideal decision boundary discrimination and shrinking can be achieved, as shown in Figure 1b. The main contributions of this paper are as follows:

(1) To address the potential risk of insufficient representation of target-domain data characteristics, a MuCD-driven DA approach is designed, by extending the dual-classifier adversarial paradigm of MCD to multiple classifiers and exploiting discrepancy learning among them.

(2) To alleviate the ambiguous boundary discrimination problem, a confidence pseudo-label-guided self-supervised learning (CPSL) mechanism is designed to guide and enhance the discrimination of diagnosis model for different fault classes, thereby clarifying classification boundaries and improving diagnosis performance.

(3) According to the design objectives, CPSL is specifically integrated into MuCD, and a cross-working condition fault diagnosis method for rotating machinery components based on multi-classifier discrepancy-driven learning enhanced by confidence pseudo-label self-supervision (CPSL-MuCD) is proposed.

(4) The cross-working condition fault diagnosis performance of the proposed CPSL-MuCD is comprehensively validated on the Paderborn University (PU) bearing fault dataset, PT500mini gear fault dataset, and wind turbine gearbox (WTG) bearing–gear compound fault dataset.

The remainder of this paper is organized as follows. Section 2 describes the principle of the proposed CPSL-MuCD and its specific implementation for cross-working condition fault diagnosis. Section 3 presents comprehensive experimental validations and analyses. Section 4 concludes this work and provides an outlook on future research directions.

2. Methodology

2.1. Problem definition

In cross-domain fault diagnosis scenarios, the available labeled data are defined as the source domain $\mathcal{D}_S = \{(x_S^i, y_S^i)\}_{i=1}^{n_S}$ containing n_S samples and labels (x_S, y_S) . The available unlabeled data are defined as the target domain $\mathcal{D}_T = \{(x_T^j)\}_{j=1}^{n_T}$ containing n_T samples $(x_T,)$. Affected by cross-working condition, distribution discrepancy exist between different domains, *i.e.*, $X_S \neq X_T$ and $\mathcal{P}_S(X_S) \neq \mathcal{P}_T(X_T)$. The purpose of constructing cross-working condition diagnosis model is to learn domain-invariant feature satisfying $\mathcal{P}_S(Y_S|X_S) \approx \mathcal{P}_T(Y_T|X_T)$ from source and target domain data, thereby realizing cross-working condition fault diagnosis.

2.2. Overview

To address the potential deficiencies of the dual-classifier structure in MCD in representing the data characteristics of target domain, as well as its insufficient discrimination of classification boundaries between different fault classes, this paper proposes a multi-classifier discrepancy-driven DA approach enhanced by confidence pseudo-label and self-supervised learning (CPSL-MuCD) for cross-working condition fault diagnosis. The network architecture of CPSL-MuCD is shown in Figure 2. The main components of CPSL-MuCD are described as follows.

First, the backbone network consisting of five one-dimensional convolutional layers is employed to extract features from raw input signals, and it is shared by the source and target domains. Second, the

MuCD-driven DA approach composed of three classifiers with identical architectures and parameters is developed to learn the classification of different fault classes and domain-invariant features. Third, the CPSL mechanism consists of a confidence pseudo-label generation strategy and target-domain self-supervised learning losses. Specifically, the confidence pseudo-label generation process primarily involves calculation of the harmonic mean of class prediction probabilities and threshold screening. Furthermore, the multilayer perceptron (MLP) is used to encode the features for contrastive learning.

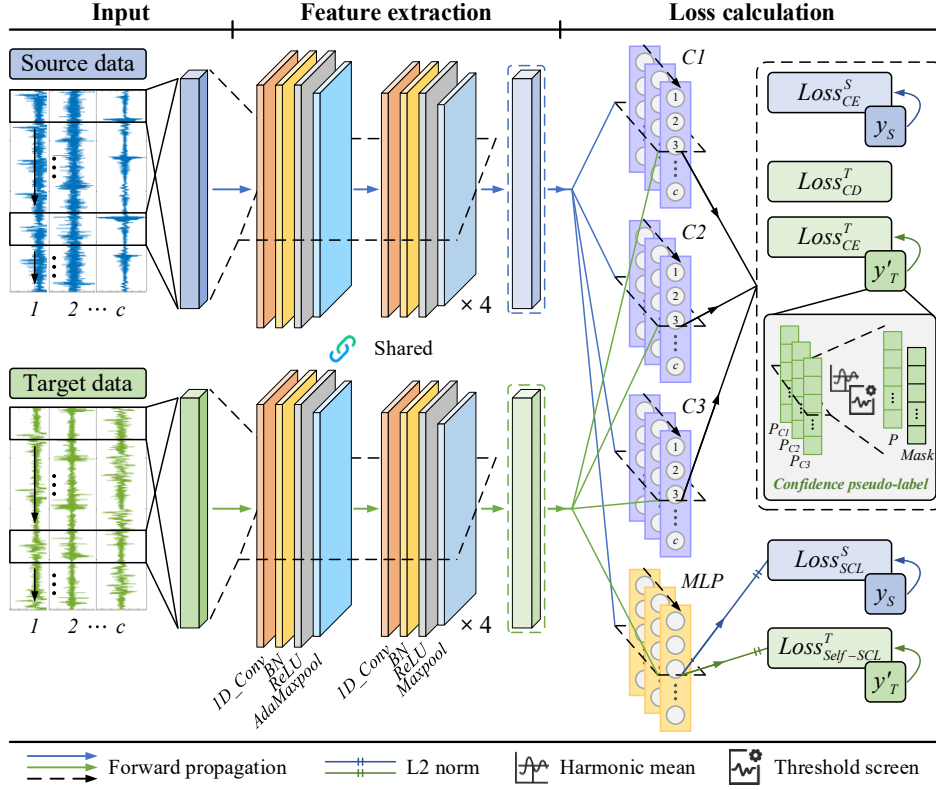


Figure 2. The network structure of the proposed CPSL-MuCD.

2.3. Multi-classifier discrepancy approach

Regarding the potential risk that the dual-classifier structure of MCD is susceptible to the decision bias of a single classifier pair, which may lead to insufficient representation of target-domain data characteristics, thereby affecting classification boundary discrimination and limiting the model's diagnosis performance. This paper proposes a MuCD approach by increasing the number of classifiers to three. MuCD aims to exploit discrepancy learning among multiple classifiers to characterize the uncertainty of target-domain samples from multiple discriminative perspectives, thereby enhancing the representation of target-domain data characteristics and sufficiently learning the correlation information among different fault classes. In this way, the discrimination of classification boundaries is indirectly strengthened and the diagnosis performance of model is improved. MuCD follows the three-stage adversarial training paradigm of MCD, and its implementation process within a mini-batch is as follows:

The first step involves the initial training of the model, with the objective of training the feature extractor and classifiers so that the initial model can classify faults and determine the decision boundary in source domain. During this step, the parameters of the feature extractor and classifiers are updated by calculating and minimizing the source-domain cross-entropy loss $\mathcal{L}_{CE}^S$.

$$\mathcal{L}_{\text{Step1}} = \min\{\mathcal{L}_{\text{CE}}^S\} \quad (1)$$

$$\mathcal{L}_{\text{CE}}^S = \sum_{c=1}^C \sum_{k=1}^B \text{CE}\left(G_y^c\left(G_E(x_S^k)\right), y_S^k\right) \quad (2)$$

where, C represents the number of classifiers, B is the batch size, $\text{CE}(\cdot)$ is the cross-entropy loss function, $G_E(\cdot)$ is the feature extractor, and G_y^c is the c th classifier.

The second step focuses on maximizing classification discrepancy. Its objective is to detect samples in the target domain that lie outside the decision boundary and adjust the boundary. During training in this step, the parameters of feature extractor are fixed, and only the classifiers are trained and their parameters updated. Meanwhile, the source domain cross-entropy loss is still calculated and minimized to ensure the classification performance.

$$\mathcal{L}_{\text{Step2}} = \min\{\mathcal{L}_{\text{CE}}^S - \mathcal{L}_{\text{CD}}^T\} \quad (3)$$

$$\mathcal{L}_{\text{CD}}^T = \frac{1}{B} \sum_{k=1}^B \left(\left| p_{C1T}^k - p_{C2T}^k \right| + \left| p_{C1T}^k - p_{C3T}^k \right| + \left| p_{C2T}^k - p_{C3T}^k \right| \right) \quad (4)$$

where, p_{C1} , p_{C2} , and p_{C3} respectively represent the predicted probabilities output by three classifiers, $p_{Cc} = \text{Softmax}\left(G_y^c(G_E(x))\right)$.

The third step focuses on minimizing classification discrepancy. Its objective is to train and confuse the feature extractor, thereby enabling it to generate target domain features that fall within the corresponding decision boundaries, thus adjusting the feature distribution and achieving DA. During training in this step, the parameters of classifiers are fixed, and only the feature extractor is trained and its parameters updated.

$$\mathcal{L}_{\text{Step3}} = \min\{\mathcal{L}_{\text{CD}}^T\} \quad (5)$$

Finally, by repeatedly performing adversarial training between Step 2 and Step 3, the cross-domain feature distribution is gradually adjusted, thereby achieving DA for target domain features. Based on prior experience and experimental validation, the number of repetitions is set to 4.

2.4. Self-supervised learning based on confidence pseudo-label

Although MuCD can improve the classification boundary discrimination problem to some extent through learning from the discrepancies among multiple classifiers, the lack of targeted learning for the discrimination of different fault classes in target domain still limits the final diagnosis performance. Therefore, this paper designs a self-supervised learning mechanism to clarify the classification boundaries by setting up self-supervised classification learning loss and self-supervised contrastive learning loss in target domain. Additionally, considering that the self-supervised medium for aforementioned losses originates from generated pseudo-labels, a confidence pseudo-label strategy is designed to enhance the effectiveness of the self-supervised learning mechanism. This forms a CPSL approach, enabling MuCD to further clarify classification boundaries. The specific implementation of CPSL is as follows:

In a mini-batch, based on the target class prediction probabilities p_{C1T} , p_{C2T} , and p_{C3T} output by the three classifiers, the pseudo-labels y'_T are constructed by calculating the harmonic mean p_T , as shown in Equation (6). A threshold screening mechanism is then applied according to the probability

values to form confidence pseudo-labels $y'_T[n]$, *i.e.*, $y'_T \xrightarrow{\text{if } p_T[n] > 0.8} y'_T[n]$. Specifically, compared with the arithmetic mean, constructing pseudo-labels via the harmonic mean suppresses extreme deviation values and balances the uncertainty among the outputs of three classifiers, thereby ensuring the initial reliability of pseudo-labels at the fusion mechanism level [23]. Subsequently, a pseudo-label filtering threshold is imposed to enforce a threshold constraint on the maximum predicted probability after harmonic averaging, which further filters out low-confidence samples. This elevates the quality of target domain pseudo-labels and provides reliable supervision medium for self-supervised learning. The self-supervised learning losses for target domain based on confidence pseudo-labels, $\mathcal{L}_{CE}^T$ and $\mathcal{L}_{Self-SCL}^T$, are shown in Equations (7) and (9). Additionally, to ensure clear and distinct classification boundaries in source domain, a supervised contrastive learning loss $\mathcal{L}_{SCL}^S$ is introduced, as shown in Equation (10).

$$p_T = \frac{3 \times p_{C1_T} \times p_{C2_T} \times p_{C3_T}}{(p_{C1_T} + p_{C2_T}) + (p_{C1_T} + p_{C3_T}) + (p_{C2_T} + p_{C3_T})} \Rightarrow y'_T \quad (6)$$

$$\mathcal{L}_{CE}^T = \sum_{c=1}^C \sum_n \text{CE}(G_y^c(G_E(X_T))[n], y'_T[n]), \text{ if } p_T[n] > 0.8 \quad (7)$$

$$\mathcal{L}_{SCL} \left(\{(z^k, y^k)\}_k^B \right) = -\log \frac{\sum_{y=y^+} \exp(\text{sim}(\|z^k\|_2, \|z^{+i}\|_2) / \tau)}{\sum_{y \neq y^-} \exp(\text{sim}(\|z^k\|_2, \|z^{-j}\|_2) / \tau)} \quad (8)$$

$$\mathcal{L}_{Self-SCL}^T = \mathcal{L}_{SCL} \left(\{(Z_T[n], y'_T[n])\}_n \right) \quad (9)$$

$$\mathcal{L}_{SCL}^S = \mathcal{L}_{SCL} \left(\{(z_S^k, y_S^k)\}_k^B \right) \quad (10)$$

where, C represents the number of classifiers, n represents the index obtained by threshold screening. B is the batch size, $k \in \{1, 2, \dots, B\}$, $k \neq i \neq j$, $\text{sim}(\cdot, \cdot)$ denotes the similarity calculation between feature vectors, $\|\cdot\|_2$ denotes the L_2 norm of feature vectors, and τ is the temperature coefficient. z represents the feature representation encoded by MLP, $Z = \{(z^k)\}_{k=1}^B = \text{MLP}(G_E(\{(z^k)\}_{k=1}^B))$. z^+ represents the feature representation belonging to the same fault class as z^k within a mini-batch, *i.e.*, $y = y^+$. Conversely, z^- represents the feature representation belonging to the different fault classes than z^k within a mini-batch, *i.e.*, $y = y^-$.

2.5. The fault diagnosis model based on CPSL-MuCD

To achieve cross-working condition fault diagnosis, a fault diagnosis model based on CPSL-MuCD has been constructed. When building the CPSL-MuCD model, this paper specifically embedded the losses of $\mathcal{L}_{SCL}^S$, $\mathcal{L}_{CE}^T$, and $\mathcal{L}_{Self-SCL}^T$ into the three-step adversarial training paradigm of MuCD, in accordance with their design objectives. The losses of $\mathcal{L}_{SCL}^S$ and $\mathcal{L}_{CE}^T$ are integrated into the first training step of MuCD to provide clear decision boundaries for subsequent feature distribution adjustments, while the loss of $\mathcal{L}_{Self-SCL}^T$ is embedded into the third training step to ensure that the final generated target-domain features better align with the corresponding decision boundaries. Therefore, the optimization objectives for the three training steps of the CPSL-MuCD model are as follows:

$$\begin{cases} \mathcal{L}_{\text{Step1}} = \min\{\mathcal{L}_{\text{CE}}^{\text{S}} + \mathcal{L}_{\text{SCL}}^{\text{S}} + \mathcal{L}_{\text{CE}}^{\text{T}} \times \lambda\} \\ \mathcal{L}_{\text{Step2}} = \min\{\mathcal{L}_{\text{CE}}^{\text{S}} - \mathcal{L}_{\text{CD}}^{\text{T}}\} \\ \mathcal{L}_{\text{Step3}} = \min\{\mathcal{L}_{\text{CD}}^{\text{T}} + \mathcal{L}_{\text{Self-SCL}}^{\text{T}} \times \lambda\} \end{cases} \quad (11)$$

$$\lambda = \frac{2}{1 + e^{\frac{-5 \times \text{epoch}}{\text{epochs}}}} - 1 \quad (12)$$

where, λ represents an exponentially increasing penalty coefficient, which is introduced to mitigate the adverse impact of low-quality pseudo-labels in the early training stage. When training the CPSL-MuCD model, the `batch_size` is set to 64, the Adam optimizer is used with an initial learning rate of 0.001 and a weight decay rate of $1e-5$, the learning rate follows an exponential decay curve with a decay rate of 0.95, and the maximum number of training iterations *epochs* is set to 200. The detailed training procedure of the CPSL-MuCD model is described as follows:

(1) The source and target domain data are constructed according to different working conditions in the dataset, followed by data sampling. The source and target domains are then independently divided into training and test sets at a ratio of 4:1, and the samples in the training and test sets are preprocessed separately.

(2) The network architecture and model parameters are specified, training-related hyperparameters are set, and the network weight parameters are initialized.

(3) The source and target domain training data are fed into the model. The parameters of the feature extractor and classifiers are updated and optimized, according to $\mathcal{L}_{\text{Step1}} = \min\{\mathcal{L}_{\text{CE}}^{\text{S}} + \mathcal{L}_{\text{SCL}}^{\text{S}} + \mathcal{L}_{\text{CE}}^{\text{T}} \times \lambda\}$ and the backpropagation algorithm.

(4) The source and target domain training data are fed into the model. The classifier parameters are updated and optimized, according to $\mathcal{L}_{\text{Step2}} = \min\{\mathcal{L}_{\text{CE}}^{\text{S}} - \mathcal{L}_{\text{CD}}^{\text{T}}\}$ and the backpropagation algorithm.

(5) The target domain training data are fed into the model. The parameters of the feature extractor are updated and optimized, according to $\mathcal{L}_{\text{Step3}} = \min\{\mathcal{L}_{\text{CD}}^{\text{T}} + \mathcal{L}_{\text{Self-SCL}}^{\text{T}} \times \lambda\}$ and the backpropagation algorithm. This step is repeated four times.

(6) Steps (3), (4), and (5) are repeated to iteratively train the CPSL-MuCD model. Training terminates and the model weights are saved, when the number of iterations *epoch* reaches the maximum value *epochs*.

(7) The target domain test data are input into the trained CPSL-MuCD model for inference, and the fault diagnosis results under the target working condition are output.

3. Experiment validation and analysis

To verify the effectiveness and generalizability of the proposed CPSL-MuCD, cross-working condition fault diagnosis experiments are conducted on three datasets: the PU rolling bearing dataset [24], the PT500mini gearbox fault dataset [25], and the WTG compound fault dataset [15]. CPSL-MuCD is validated through comparative studies and ablation analysis. Specifically, both the PU and PT500mini datasets contain single-type faults, while the WTG dataset comprises compound faults. In addition, to ensure the objectivity of experimental results, each group of experiments is repeated five times to report the average diagnosis accuracy and corresponding standard deviation.

3.1. Case study1: PU dataset

3.1.1. Dataset description

The PU dataset is a publicly available dataset widely used in research on rolling bearing fault diagnosis. The experimental bearings in this dataset are 6203-type rolling bearings, exhibiting two types of bearing faults: artificially induced damage and natural degradation damage. The bearings with natural degradation damage are obtained from the accelerated degradation testing platform, as shown in Figure 3a. This paper uses bearing data from 13 classes of single-type faults in accelerated degradation tests (see reference [24] for specific fault descriptions) to conduct cross-working condition fault diagnosis experiments and validate the proposed method.

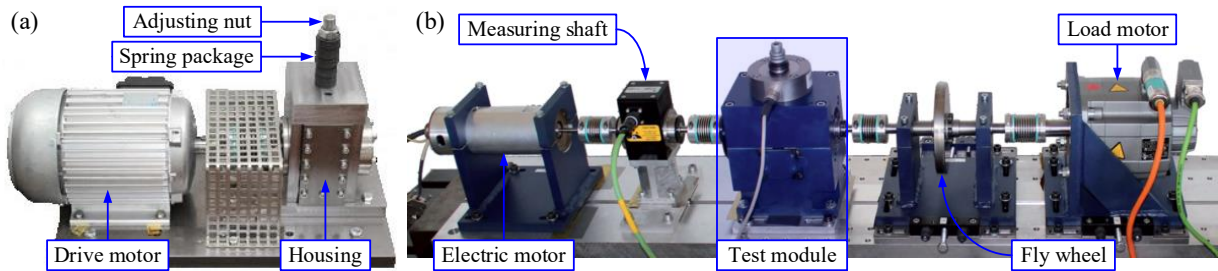


Figure 3. PU dataset test bench: (a) Accelerated degradation test platform; (b) Bearing test platform.

The data in this dataset is collected from the bearing test platform, as shown in Figure 3b. By replacing different faulty bearings in the test module and adjusting various working condition parameters (rotational speed, radial load, and torque load), bearing vibration signals under four different working conditions are collected. The samples are constructed using non-overlapping sliding-window sampling, followed by separate fast Fourier transform (FFT) and Z-score standardization for the training and test samples [26]. The working condition parameters and diagnosis tasks are shown in Table 1.

Table 1. Dataset details and transfer tasks of PU.

Working condition	Speed (rpm)	Load (Nm)	Radial force (N)	Transfer task (S-T)
0	1500	0.7	1000	0-1, 0-2
1	900	0.7	1000	0-3, 1-2
2	1500	0.1	1000	1-3, 2-3
3	1500	0.7	400	

3.1.2. Quantitative comparison analysis

To validate the effectiveness and superiority of CPSL-MuCD, the DA methods DAN, DSAN, DANN, DCTLN, MCD [20], and SLDDA [21] are selected for comparison. Meanwhile, the model consisting of the backbone network and a single classifier from the proposed method is employed as the baseline model. All comparison methods use the same backbone network and classifier architecture. Table 2 presents the quantitative results of overall transfer fault diagnosis performance on the PU dataset. The results indicate that, compared to the Baseline, the diagnosis performance of CPSL-MuCD has been significantly improved, with an increase of 33.96% in average accuracy. This demonstrates that the proposed method can effectively alleviate the distribution discrepancy between different working conditions and improve the cross-domain generalization. Compared with the other DA methods,

CPSL-MuCD still achieves higher diagnosis accuracy on all six transfer tasks and also exhibits better diagnosis stability. Furthermore, compared with MCD, which serves as the basis for improvement, and the second-best method SLDDA, CPSL-MuCD improves the overall average diagnostic accuracy by 7.04% and 5.22%, respectively, and exhibits lower accuracy fluctuations on most transfer tasks. Although the training time of CPSL-MuCD is higher than that of MCD and SLDDA, the additional computational cost is still acceptable considering the performance gains in diagnosis accuracy and stability, which further demonstrates the effectiveness of the designed MuCD approach and CPSL mechanism in improving cross-working condition diagnosis performance. In addition, the superior diagnosis performance of MuCD over MCD validates the effectiveness of the improvements made to the dual-classifier structure of MCD in this paper.

Table 2. Transfer diagnosis accuracy on PU (%).

Method	0-1	0-2	0-3	1-2	1-3	2-3	Average	Time (s)
Baseline	38.68 ± 2.97	93.43 ± 1.29	57.61 ± 4.48	51.74 ± 1.39	34.25 ± 1.33	68.96 ± 1.08	57.44	43.5
DAN	47.40 ± 1.76	93.94 ± 0.90	63.33 ± 1.84	54.53 ± 1.47	36.46 ± 1.69	71.07 ± 2.45	61.12	52.0
DSAN	55.76 ± 0.50	95.57 ± 0.88	67.59 ± 2.41	59.82 ± 1.30	43.60 ± 1.31	73.83 ± 2.43	66.03	59.4
DANN	49.62 ± 0.90	93.61 ± 0.71	63.51 ± 2.55	56.79 ± 1.22	40.00 ± 2.92	71.23 ± 1.89	62.46	53.7
DCTLN	49.31 ± 2.37	95.08 ± 0.82	64.18 ± 3.70	56.39 ± 0.73	41.33 ± 1.25	72.53 ± 2.22	63.13	56.6
MCD	68.82 ± 2.82	98.35 ± 0.15	91.65 ± 1.08	90.92 ± 0.73	62.45 ± 2.55	93.98 ± 0.44	84.36	112.0
SLDDA	74.69 ± 1.24	98.69 ± 0.18	92.59 ± 0.77	91.22 ± 0.34	64.45 ± 1.29	95.46 ± 0.58	86.18	125.7
CPSL-MuCD	79.42 ± 0.89	99.05 ± 0.15	96.70 ± 0.26	97.06 ± 0.73	79.73 ± 0.57	96.43 ± 0.21	91.40	190.0
MuCD	72.10 ± 1.33	98.62 ± 0.31	92.59 ± 4.53	93.10 ± 2.03	68.99 ± 4.41	95.16 ± 0.29	86.76	136.5

To further quantify models' diagnosis performance on target domain, this paper takes task 0-3 as an example and comprehensively presents the correct and incorrect classifications of different fault classes by various models in the form of a confusion matrix, as shown in Figure 4. As shown in the results presented in the figure, compared with other DA methods, CPSL-MuCD demonstrates the best fault diagnosis performance. Furthermore, compared with MCD, which serves as the basis for improvement, CPSL-MuCD achieves better performance in terms of both the number and severity of misclassifications. Specifically, the numbers of correctly identified samples for KA16 (2), KI16 (9), and KI18 (11) are increased from 44, 31, and 37 to 50, 35, and 43, respectively, indicating improved recognition of these fault classes. Meanwhile, compared with the second-best method SLDDA, CPSL-MuCD still shows better recognition performance for some easily confused fault classes. These results indicate that, compared to the dual-classifier structure of MCD, the multi-classifier discrepancy learning constructed by CPSL-MuCD can more effectively represent the data characteristics in target domains, while the target-domain self-supervised learning guided by confidence pseudo-label can effectively enhance the discrimination between different fault classes, thereby effectively improving the model's cross-working condition diagnosis performance.

To provide a more comprehensive quantification, this paper evaluates the model's diagnosis performance for each fault class via the F1-score metric [15], in addition to calculating the overall accuracy. Figure 5 shows the F1-score for each fault class on different diagnosis tasks. As shown in the figure, except for a few fault classes on certain tasks, such as KI18 (11) on task 1-3, CPSL-MuCD

maintains relatively high F1-scores for different fault classes compared with the other methods. In particular, most fault classes on tasks 0–3, 1–2, and 2–3 achieve high F1-score values. Combined with the confusion matrix results in Figure 4, it can be further observed that CPSL-MuCD significantly improves the recognition capability for easily confused fault classes such as KA16 (2), KI16 (9), and KI18 (11), demonstrating its relatively ideal diagnosis accuracy and robust diagnosis stability. The comparative analysis of the above quantitative results indicates that the proposed CPSL-MuCD can robustly identify various fault samples under cross-working conditions, which further validates its diagnosis performance.

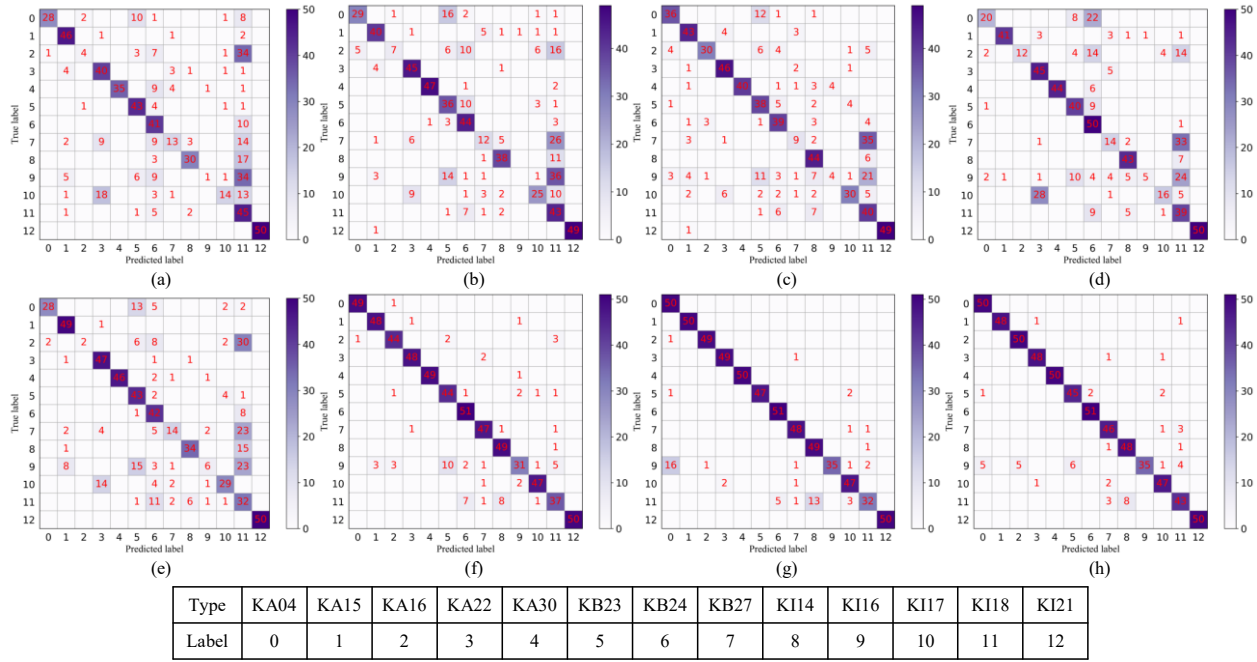


Figure 4. Confusion matrix on PU (task 0–3): (a) Baseline; (b) DAN; (c) DSAN; (d) DANN; (e) DCTLN; (f) MCD; (g) SLDDA; (h) CPSL-MuCD.

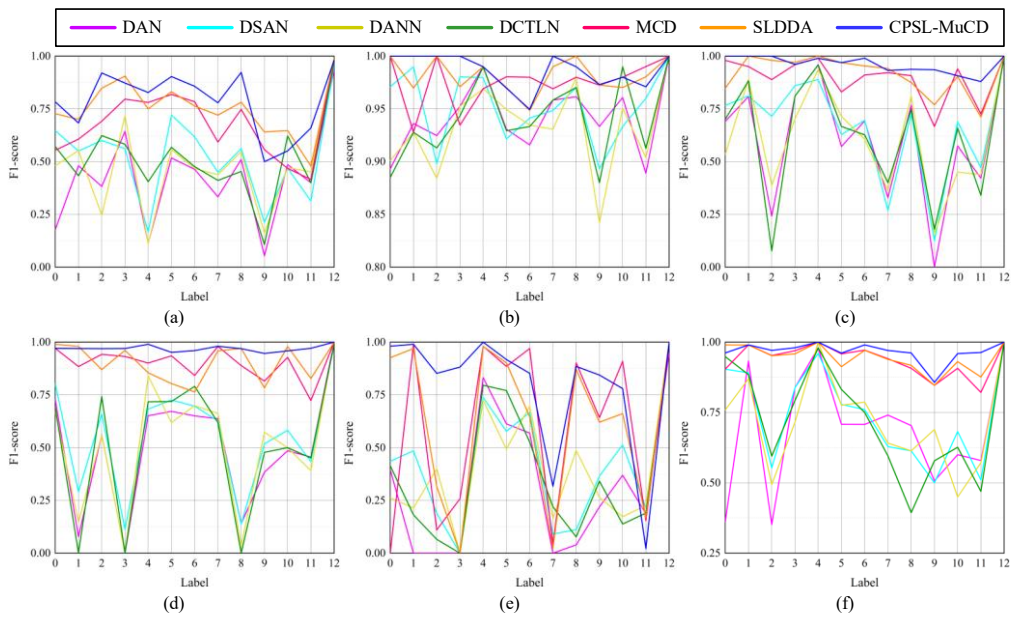


Figure 5. F1-score on PU: (a) task 0–1; (b) task 0–2; (c) task 0–3; (d) task 1–2; (e) task 1–3; (f) task 2–3.

3.1.3. Visual comparison analysis

Furthermore, to more intuitively analyze and discuss the relationship between the cross-domain feature distribution alignment effects of different models and their cross-working condition diagnosis performance, this paper employs t-distributed stochastic neighbor embedding (t-SNE) to perform dimensionality reduction on the extracted features and visualize their distributions, as shown in Figure 6. As illustrated in the figure using task 0–3 as an example, both the MCD, which serves as the basis for improvement, and the second-best method SLDDA, exhibit better feature distribution alignment than the other DA methods. However, MCD still shows dispersed feature distributions for some fault classes and relatively close boundaries between different classes, while SLDDA still suffers from a certain degree of inter-class overlap for some fault classes. In contrast, CPSL-MuCD achieves a higher degree of clustering between source-domain and target-domain features of the same class, more distinct separation between different fault classes, and a more compact overall feature distribution. CPSL-MuCD further optimizes the representation of target-domain data characteristics through multi-classifier discrepancy learning and enhances the discrimination among different fault classes through confidence pseudo-label-guided self-supervised learning, thereby further improving feature distribution alignment and cross-working condition diagnosis performance.

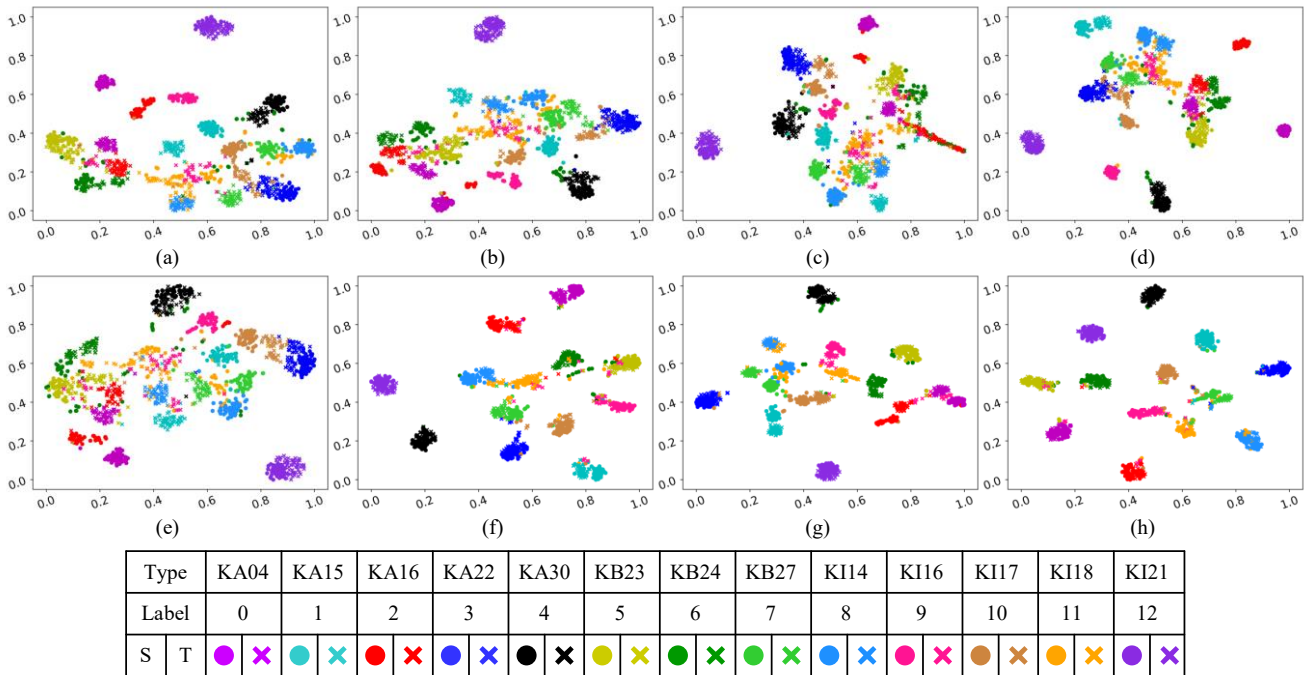


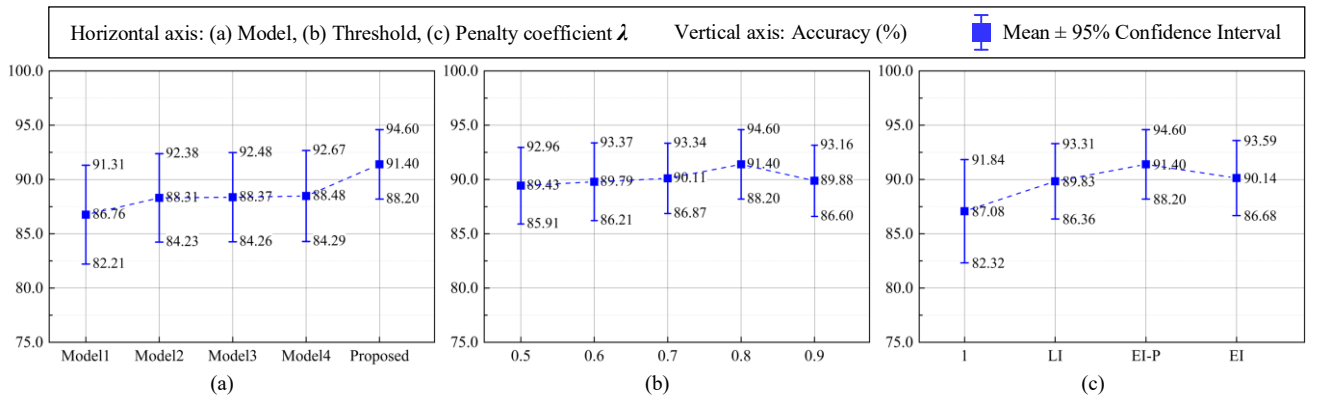
Figure 6. t-SNE visualization on PU (task 0–3): (a) Baseline; (b) DAN; (c) DSAN; (d) DANN; (e) DCTLN; (f) MCD; (g) SLDDA; (h) CPSL-MuCD.

3.1.4. Ablation analysis

In this section, ablation studies and analyses are conducted to investigate the rationality and effectiveness of the proposed CPSL-MuCD in terms of model architecture, pseudo-label screen threshold, and sensitivity of the self-supervised learning penalty coefficient. Table 3 summarizes the detailed settings of ablation experiments with different architectures, screen thresholds, and penalty coefficients, while Figure 7 presents the quantitative diagnosis performance of each ablated model on the PU dataset.

Table 3. Details of structure, threshold and penalty coefficient ablation experiments.

Abolition items	Details				
Model	Model1	Model2	Model3	Model4	Proposed
Model	MuCD	CPSL-MuCD (no CP)	CPSL-MuCD (no $\mathcal{L}_{Self-SCL}^T$)	CPSL-MuCD (no $\mathcal{L}_{CE}^T$)	CPSL-MuCD
Threshold	0.5	0.6	0.7	0.8	0.9
Penalty coefficient (λ)	Fixed type	Linearly increase type (λ_{LI})	Exponential increase type (this paper, λ_{EI-P})	Exponential increase type (λ_{EI})	
	1	$\frac{epoch}{epochs}$	$\frac{2}{1 + e^{\frac{-5 \times epoch}{epochs}}} - 1$	$\frac{2}{1 + e^{\frac{-10 \times epoch}{epochs}}} - 1$	

**Figure 7.** Quantitative results of structure, threshold and penalty coefficient ablation experiments.

The architecture ablation results indicate that without confidence pseudo-label guided target domain self-supervised learning (no CP), the diagnosis model only outperforms MuCD, which verifies the effectiveness of the proposed self-supervised learning mechanism and confirms the enhancement role of the confidence pseudo-label strategy. When the confidence pseudo-label strategy is retained but $\mathcal{L}_{Self-SCL}^T$ and $\mathcal{L}_{CE}^T$ are removed separately, the diagnosis performance degrades compared with the complete CPSL-MuCD, further validating the effectiveness of the self-supervised learning mechanism. Regarding the screen threshold ablation, different threshold values exert distinct influences on the model. An excessively low threshold may introduce erroneous pseudo-labels and mislead self-supervised learning, thereby degrading performance. Conversely, an overly high threshold may result in over-filtering of pseudo-labels and insufficient self-supervised learning. The model achieves a local optimum when the threshold is set to 0.8; accordingly, the filtering threshold is fixed at 0.8 in the confidence pseudo-label strategy. For the penalty coefficient ablation, the model with an exponentially increasing penalty coefficient outperforms those with fixed or linearly increase coefficients. This is attributed to the faster increase rate of the exponential function, which enables timely guidance of self-supervised learning. Nevertheless, an excessively rapid increase may insufficiently mitigate the adverse impact of low-quality pseudo-labels in the early training stage, as evidenced by the inferior performance of λ_{EI} compared with the selected λ_{EI-P} . This observation confirms the rationality of the chosen penalty coefficient λ_{EI-P} .

3.2. Case study2: PT500mini dataset

3.2.1. Dataset description

The PT500mini test bench mainly consists of a drive motor, support bearings, a gearbox, and a magnetic powder brake, as shown in Figure 8. During gear fault testing, the working conditions, including rotational speed and torque load, are configured by adjusting the parameters of the drive motor and magnetic powder brake. Meanwhile, gear vibration signals are acquired by accelerometers mounted at the input and output sides of the gearbox. This dataset contains three typical gear fault types, namely tooth broken, tooth missing, and tooth wear, which are collected under four different working conditions at a sampling frequency of 100 kHz. Six groups of transfer diagnosis tasks are constructed according to different condition combinations, as shown in Table 4. Similar to Case 1, sample construction in this case adopts a non-overlapping sliding window sampling strategy on the input-side signals, and the samples in the training and test sets are separately standardized by Z-Score standardization.

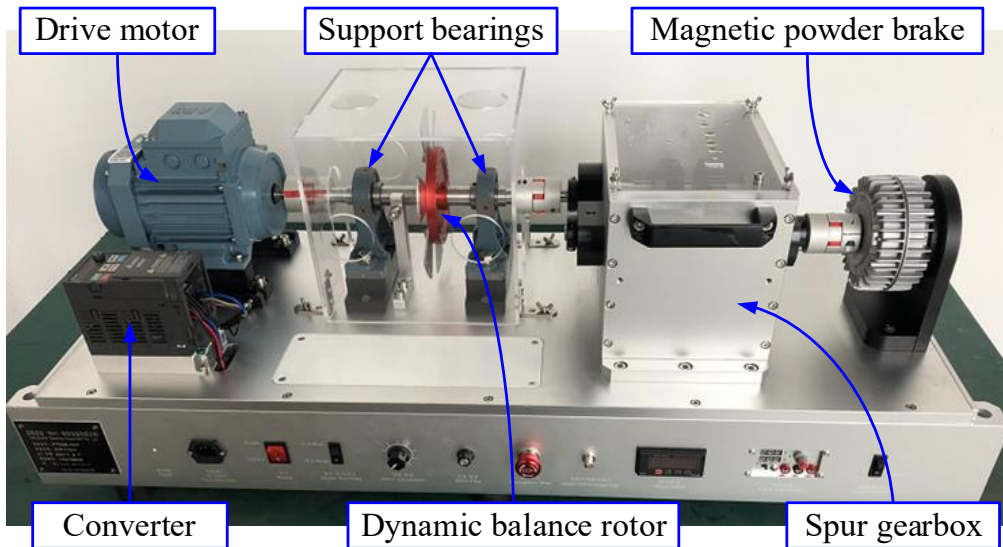


Figure 8. PT500mini dataset test bench.

Table 4. Dataset details and transfer tasks of PT500mini.

Working condition	Speed (rpm)	Torque load (N/m)	Fault types	Transfer task (S–T)
0	3000	1	Normal (0, NC),	0–1, 0–2
1	2800	1.5	Tooth broken (1, BF),	0–3, 1–2
2	2600	2	Tooth missing (2, MF),	1–3, 2–3
3	2400	2.5	Tooth wear (3, WF)	

3.2.2. Comparison analysis

To verify the generalizability of the proposed CPSL-MuCD, comparative experiments are conducted with the same methods and identical backbone network architecture and parameters as those in Case 1, followed by quantitative analysis of the results. The quantitative diagnosis performance of each method on the PT500mini dataset is summarized in Table 5. As can be observed, the proposed CPSL-MuCD

achieves the highest diagnosis accuracy on all six transfer tasks. Specifically, compared with MCD and the second-best method SLDDA, CPSL-MuCD improves the average diagnosis accuracy by 4.75% and 2.86%, respectively, and also exhibits an even more pronounced advantage over the Baseline and other DA methods, with an average accuracy improvement of over 20%. Moreover, on the more challenging transfer tasks 0–2 and 0–3, CPSL-MuCD attains diagnosis accuracies of 92.50% and 92.85%, respectively, representing a nearly 9% improvement compared with MCD. This validates its strong feature alignment capability under scenarios with significant working condition discrepancies. In terms of training time, although CPSL-MuCD requires slightly longer computation than the compared methods, the gains in diagnosis accuracy and stability outweigh the increase in time cost, yielding a more favorable overall performance. These results further confirm the generalizability and effectiveness of CPSL-MuCD on different datasets.

Table 5. Transfer diagnosis accuracy on PT500mini (%).

Method	0–1	0–2	0–3	1–2	1–3	2–3	Average	Time (s)
Baseline	67.05 ± 1.95	57.20 ± 1.11	53.80 ± 0.70	71.40 ± 1.51	65.65 ± 1.84	91.20 ± 1.22	67.72	28.5
DAN	70.40 ± 1.83	65.15 ± 1.37	60.60 ± 1.37	77.75 ± 0.61	73.40 ± 1.91	91.55 ± 0.51	73.14	34.3
DSAN	71.40 ± 1.61	65.65 ± 2.49	62.70 ± 2.02	78.90 ± 2.87	74.20 ± 1.08	91.70 ± 0.70	74.09	38.8
DANN	67.75 ± 2.25	62.80 ± 1.33	60.25 ± 4.34	77.65 ± 2.88	72.70 ± 0.62	91.35 ± 0.82	72.08	35.0
DCTLN	69.55 ± 1.78	64.85 ± 2.40	60.60 ± 1.91	78.90 ± 2.03	74.00 ± 1.31	91.25 ± 1.76	73.19	38.5
MCD	84.20 ± 2.20	83.60 ± 1.20	84.80 ± 1.69	94.75 ± 0.57	92.45 ± 0.83	98.30 ± 0.40	89.68	68.2
SLDDA	85.60 ± 1.89	86.00 ± 1.93	87.65 ± 2.21	96.10 ± 0.30	95.35 ± 0.75	98.70 ± 0.29	91.57	78.0
CPSL-MuCD	88.15 ± 0.54	92.50 ± 0.76	92.85 ± 1.06	97.20 ± 0.19	96.40 ± 0.20	99.50 ± 0.27	94.43	111.1

To further quantify the diagnosis performance of each model on different tasks at a fine-grained level, confusion matrices are employed, consistent with Case 1, to illustrate the correct and incorrect classification of each fault class in the target domain. Meanwhile, the F1-score metric is adopted to evaluate the per-class diagnosis performance of each model, as shown in Figures 9 and 10. Taking task 0–2 as an example, Figure 9 presents the classification results of each model for different fault classes. It can be observed that, compared with the Baseline and other DA methods, MCD, which serves as the basis for improvement, and the second-best method SLDDA significantly reduce severe misclassifications, while CPSL-MuCD further decreases the number of misclassified samples for different fault classes. Specifically, compared with MCD, the numbers of correctly identified samples for NC (0), BF (1), and WF (3) are increased from 84, 78, and 73 to 93, 85, and 90, respectively. Compared with SLDDA, the numbers of correctly identified samples for BF (1) and WF (3) are increased from 80 and 69 to 85 and 90, respectively. As shown by the per-class F1-score results of different tasks in Figure 10, CPSL-MuCD maintains a higher overall F1-score level on the transfer tasks and shows more pronounced performance advantages for BF (1) and WF (3). These results demonstrate that, by extending the dual-classifier architecture of MCD to multi-classifier discrepancy learning, CPSL-MuCD more sufficiently represents target domain data characteristics. Concurrently, the construction of confidence pseudo-label guided target domain self-supervised learning effectively enhances the cross-working condition diagnosis performance of model.

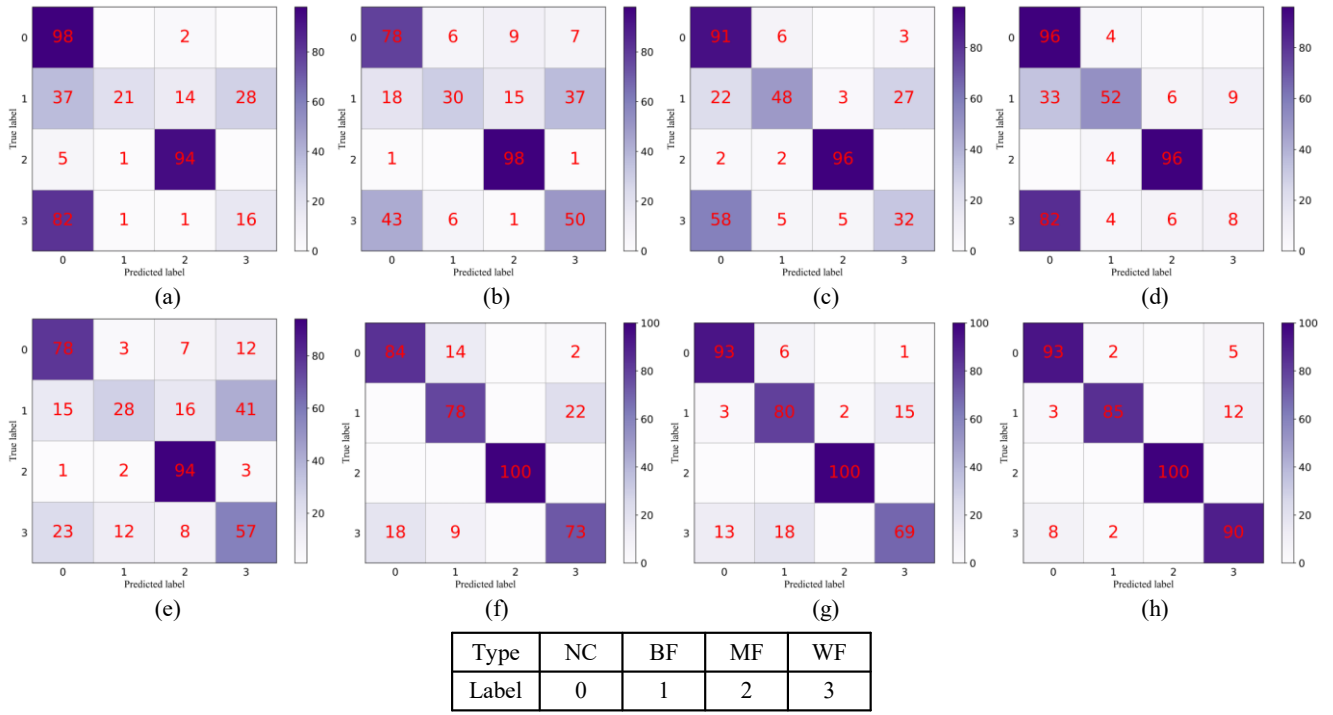


Figure 9. Confusion matrix on PT500mini (task 0-2): (a) Baseline; (b) DAN; (c) DSAN; (d) DANN; (e) DCTLN; (f) MCD; (g) SLDDA; (h) CPSL-MuCD.

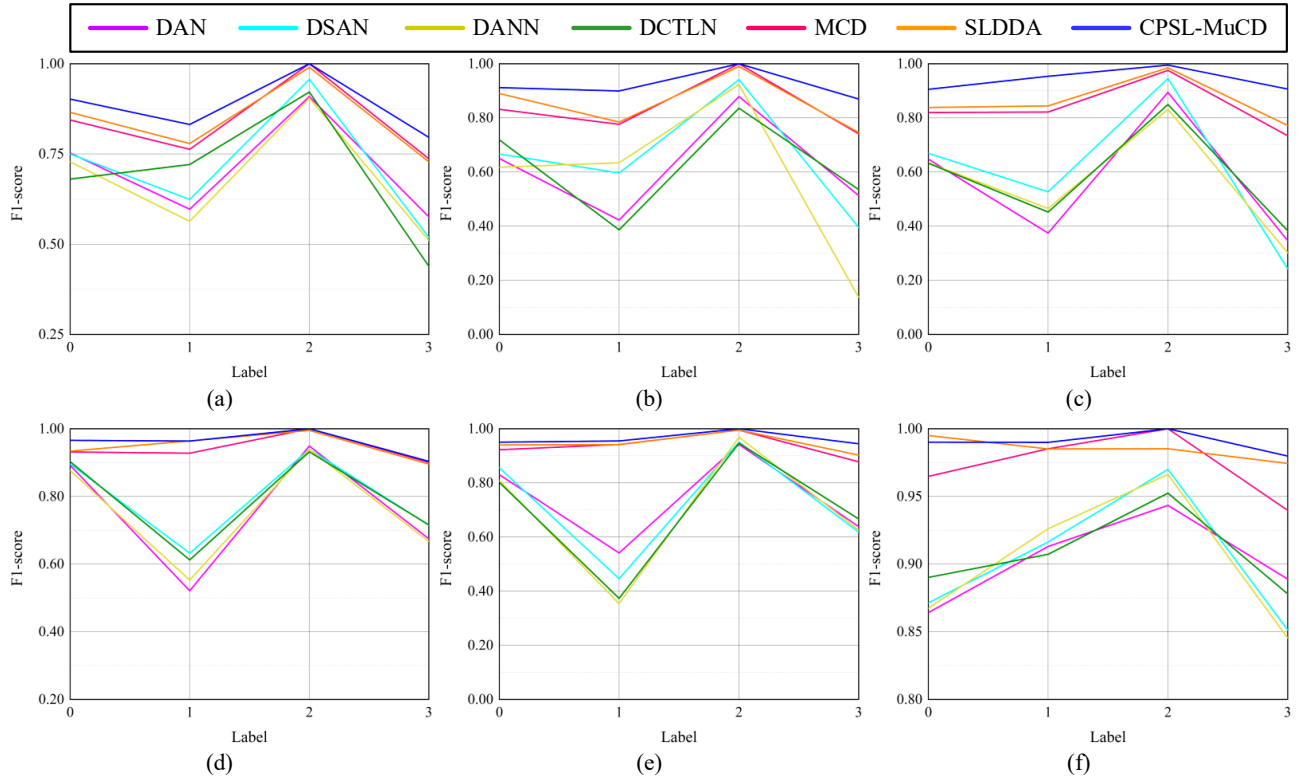


Figure 10. F1-score on PT500mini: (a) task 0-1; (b) task 0-2; (c) task 0-3; (d) task 1-2; (e) task 1-3; (f) task 2-3.

3.3. Case study3: WTG dataset

3.3.1. Dataset description

The WTG dataset test bench mainly comprises a drive motor, a gearbox, transmission bearings, and an electromagnetic brake, as shown in Figure 11. On this test bench, working condition parameters are adjusted by controlling the drive motor and electromagnetic brake, while vibration signals from the gearbox and transmission bearings are collected via accelerometers. This dataset contains six types of bearing-gear compound faults tested and acquired under four different working conditions at a sampling frequency of 100 kHz, as detailed in Table 6. Consistent with Case 2, non-overlapping sliding-window sampling is adopted for sample construction in Case 3, and the samples in the training and test sets are separately standardized by Z-Score standardization. The configured transfer diagnosis tasks are also shown in Table 6.

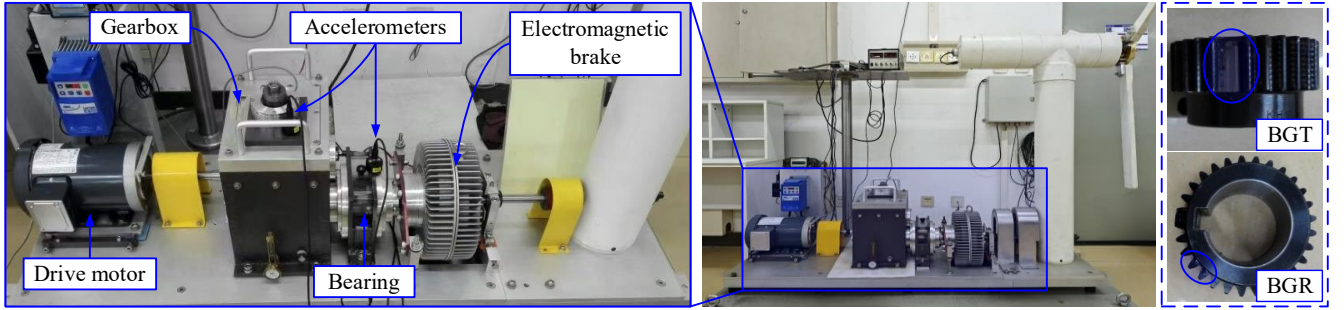


Figure 11. WTG dataset test bench.

Table 6. Dataset details and transfer tasks of WTG.

Types	Components				
Labels	Bearing		Gear		
	IR	OR	BGT	BGR	GN
(0, BGT)	✓	-	✓	-	-
(1, BGR)	✓	-	-	✓	-
(2, BIGN)	✓	-	-	-	✓
(3, BOGT)	-	✓	✓	-	-
(4, BOGR)	-	✓	-	✓	-
(5, BOGN)	-	✓	-	-	✓
Notes	IR: Inner race fault, OR: Outer race fault		BGT: Broken gear teeth, BGR: Broken gear root, GN: Gear normal		
Working conditions	0	1	2	3	Transfer tasks (S-T)
Speed (r/min)	600	900	1200	1500	0-1, 0-2, 0-3, 1-2, 1-3, 2-3

3.3.2. Comparison analysis

The comparative experiment setup of Case 3 is consistent with that of Cases 1 and 2, aiming to further validate the effectiveness and generalizability of the proposed CPSL-MuCD. Table 7 presents the

quantitative diagnosis performance of different methods on the WTG dataset. As can be observed, the proposed CPSL-MuCD achieves the best diagnosis performance on all six transfer diagnosis tasks. Compared with the Baseline, CPSL-MuCD improves the average diagnosis accuracy by 15.83%. Furthermore, compared with MCD and the second-best method SLDDA, CPSL-MuCD improves the average diagnosis accuracy by 3.29% and 2.09%, respectively. In particular, on the more challenging tasks 0–2 and 0–3, CPSL-MuCD shows more pronounced performance advantages, achieving improvements of 7.07% and 6.01% over MCD, respectively. In terms of diagnosis stability, the standard deviations of CPSL-MuCD on all six tasks are controlled within 0.69%, and lower accuracy fluctuations are observed on most tasks. Although CPSL-MuCD incurs additional training overhead due to the introduction of the MuCD approach and CPSL mechanism, the improvements in diagnosis accuracy and stability demonstrate the reasonableness of the additional computational cost to some extent. These results indicate that the MuCD approach and CPSL mechanism can effectively improve the cross-working condition diagnostic accuracy and stability of the model, while further verifying the favorable cross-working condition generalizability of CPSL-MuCD.

Table 7. Transfer diagnosis accuracy on WTG (%).

Method	0–1	0–2	0–3	1–2	1–3	2–3	Average	Time (s)
Baseline	80.56 ± 2.20	71.68 ± 2.21	67.39 ± 0.77	87.14 ± 0.93	84.66 ± 1.20	88.41 ± 1.21	79.97	42.2
DAN	88.13 ± 1.35	79.24 ± 1.23	77.21 ± 2.59	89.90 ± 0.86	89.36 ± 0.80	90.54 ± 0.76	85.73	48.3
DSAN	88.16 ± 0.93	78.72 ± 1.78	73.50 ± 1.89	88.91 ± 1.37	88.83 ± 1.13	91.13 ± 0.90	84.88	58.7
DANN	86.84 ± 0.71	78.16 ± 1.21	76.45 ± 3.41	89.77 ± 1.21	88.11 ± 0.93	91.63 ± 1.21	85.16	51.2
DCTLN	87.63 ± 1.69	80.16 ± 0.65	78.88 ± 2.01	90.56 ± 0.77	88.54 ± 1.26	90.91 ± 1.58	86.11	53.7
MCD	96.25 ± 0.32	87.47 ± 2.47	85.78 ± 2.84	95.10 ± 0.12	94.88 ± 0.24	95.60 ± 0.19	92.51	113.4
SLDDA	96.68 ± 0.50	91.12 ± 1.13	88.01 ± 1.93	95.30 ± 0.47	95.36 ± 0.45	95.76 ± 0.12	93.71	126.4
CPSL-MuCD	97.86 ± 0.29	94.54 ± 0.35	91.79 ± 0.69	96.15 ± 0.25	97.21 ± 0.23	97.27 ± 0.25	95.80	183.1

Figures 12 and 13 present the confusion matrices and per-class F1-score statistics, respectively, for further comprehensive quantification of the diagnosis performance of each model. The confusion matrix in Figure 12 illustrates the correct and incorrect classification of each fault class for all models on task 0–3. It can be observed that, compared with the Baseline and other DA methods, MCD reduces some misclassifications through the dual-classifier adversarial mechanism, but relatively obvious confusion still exists in classes such as BIGT (1), BOGT (3), and BOGR (4). The second-best method SLDDA further improves the recognition results of some fault classes. On this basis, the proposed CPSL-MuCD achieves better classification performance. Compared with MCD, which serves as the basis for improvement, the numbers of correctly identified samples for BIGT (1), BOGT (3), and BOGR (4) are increased to 89, 102, and 72, respectively, while the overall performance is also further improved compared with SLDDA. The F1-score results in Figure 13 further show that, except for BOGT (3) on task 1–2, BIGT (1) on task 1–3, and BOGT (3) on task 2–3, CPSL-MuCD achieves the highest F1-score for the remaining fault classes and maintains a relatively high and stable overall F1-score level. These results further verify the effectiveness of multi-classifier discrepancy learning and the confidence pseudo-label self-supervised mechanism in enhancing the discriminative capability of target-domain classes and improving cross-working condition fault diagnosis performance.

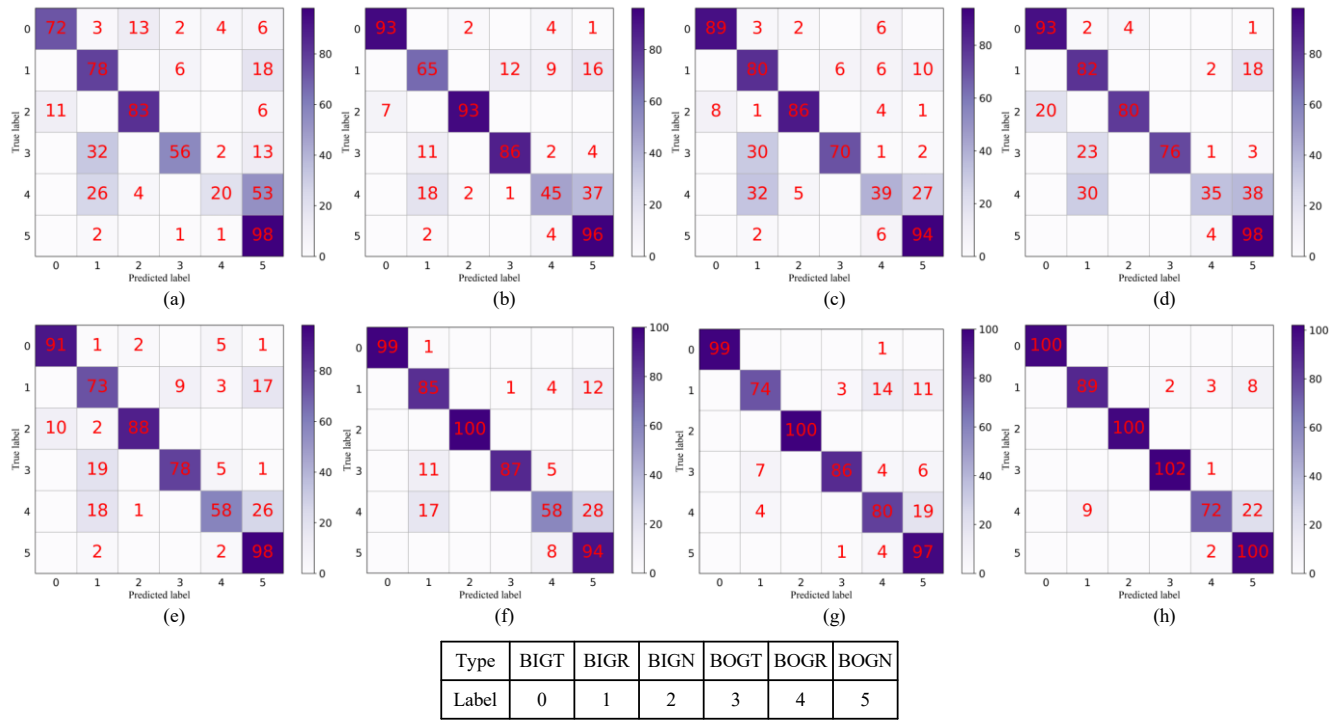


Figure 12. Confusion matrix on WTG (task 0–3): (a) Baseline; (b) DAN; (c) DSAN; (d) DANN; (e) DCTLN; (f) MCD; (g) SLDDA; (h) CPSL-MuCD.

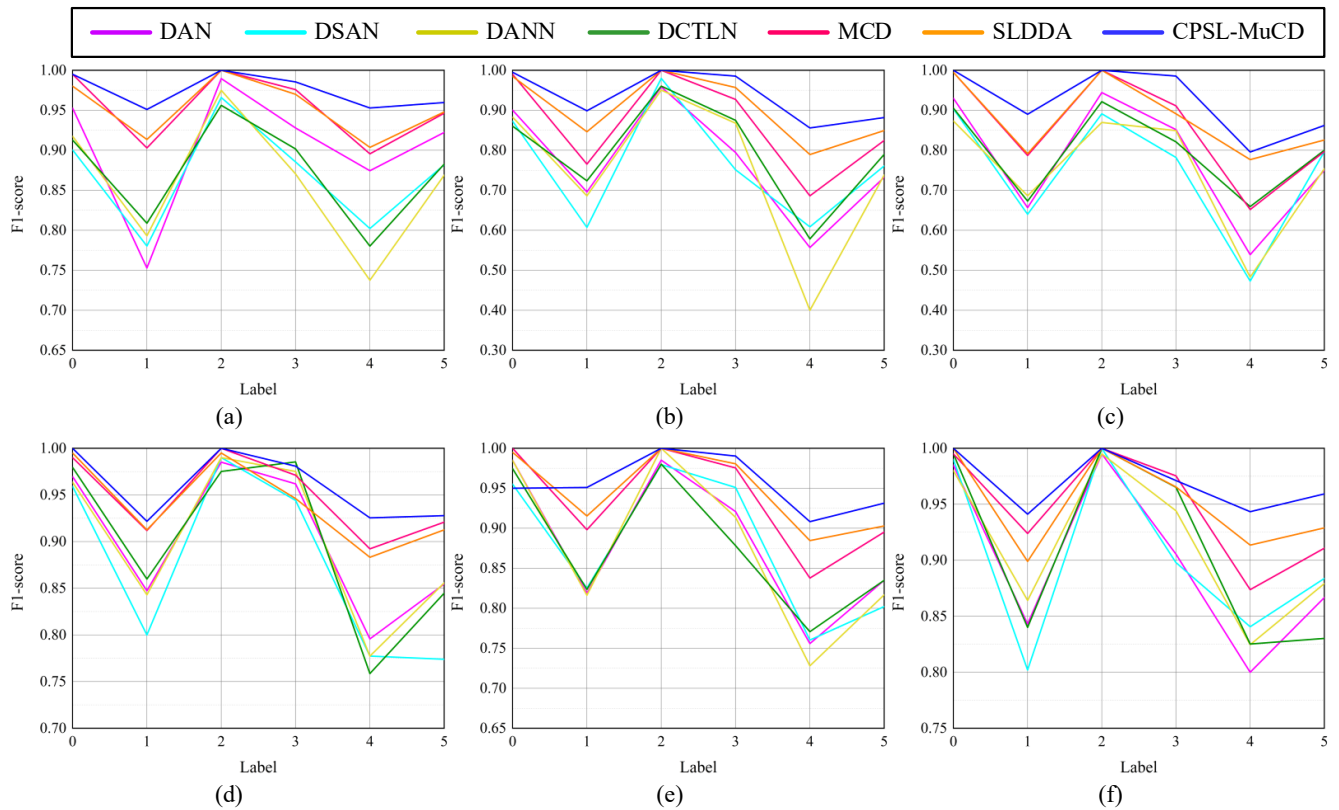


Figure 13. F1-score on WTG: (a) task 0–1; (b) task 0–2; (c) task 0–3; (d) task 1–2; (e) task 1–3; (f) task 2–3.

4. Conclusion

Aiming at the limitations of diagnosis performance of cross-working condition caused by the inadequate classification boundary discrimination and potential insufficient representation of target domain data characteristics in the dual-classifier adversarial structure of MCD, this paper proposes a CPSL-MuCD-based DA method for rolling bearing fault diagnosis. The proposed method designs a MuCD approach by constructing the discrepancy learning among multiple classifiers, to enhance the representation of target domain data characteristics, as well as the discrimination and shrinking of classification boundaries. Concurrently, a CPSL mechanism is established by designing a confidence pseudo-label generation strategy and using it to guide the self-supervised learning in target domain. This mechanism aims to perform targeted learning on the discrimination of different fault classes in target domain, enabling MuCD to further clarify the identification of classification boundaries and thereby improve the model's diagnosis performance of cross-working condition. Then, in accordance with the design objectives of the CPSL mechanism, it is specifically integrated into the various training steps of MuCD to construct the CPSL-MuCD diagnosis model. To validate the effectiveness and generalizability of CPSL-MuCD, a comprehensive diagnosis performance validation and analysis was conducted through 18 cross-working condition fault diagnosis tasks on PU bearing fault, PT500mini gear fault, and WTG bearing-gear compound fault datasets. Compared to MCD, CPSL-MuCD achieved improvements in overall accuracy of 7.04%, 4.75%, and 3.29% on the three datasets, respectively, and demonstrated better standard deviation performance during repeated experiments for each task. This not only validates the effectiveness of its improvements over MCD but also demonstrates its ability to effectively and generally achieve cross-working condition fault diagnosis.

Although CPSL-MuCD has performed well in cross-working condition fault diagnosis on three different fault datasets of rotating machinery components, the increased model training cost cannot be ignored, as this will hinder its engineering applications. Therefore, future work will focus on optimizing MuCD for lighter models while exploring more challenging cross-working condition fault diagnosis scenarios.

Data availability statement

No supplementary or additional data were generated in this study.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools only to improve language and readability. Specifically, the authors used ChatGPT and Grammarly for language polishing only for drafting in limited sections. The authors take full responsibility for the content of the manuscript.

Acknowledgments

This work was financially supported by the Joint Project of BRC-BC (BUCT-ZR-2606), and the National Natural Science Foundation of China (Grant No. 52475084 and 52375076).

Authors' contribution

Methodology, software and writing—original draft preparation, X.L.; visualization, investigation and formal analysis, Y.L.; resources, validation and writing—review and editing, L.S.; writing—review and editing, funding acquisition, project administration and supervision, H.W. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

Yuguang Li is employee of Refining, Chemicals & New Materials Company. The author declares that this affiliation does not influence the design, analysis, interpretation, or presentation of the results reported in this manuscript. 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.

References

- [1] Zhang F, Huang J, Chu F, Cui L. Mechanism and method for the full-scale quantitative diagnosis of ball bearings with an inner race fault. *J. Sound Vib.* 2020, 488:115641.
- [2] Jia N, Shen C, Zhu Z, Huang W. A single-source domain generalization method based on frequency-domain statistical priors and class-conditional representation alignment for bearing fault diagnosis. *Mechatronics Tech.* 2026, 3(2):1–13.
- [3] Li W, Huang R, Li J, Liao Y, Chen Z, *et al.* A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: theories, applications and challenges. *Mech. Syst. Signal Process.* 2022, 167:108487.
- [4] Shi M, Ding C, Chang S, Shen C, Huang W, *et al.* Cross-domain class incremental broad network for continuous diagnosis of rotating machinery faults under variable operating conditions. *IEEE Trans. Ind. Inf.* 2024, 20:6356–6368.
- [5] Zhang Y, Shen C, Yang H, Shi J, Huang W, *et al.* A new adaptive representation dual classifier residual network for continuous fault diagnosis of rotating machinery with domain increments. *Adv. Eng. Inf.* 2025, 67:103561.
- [6] Ren H, Xing J, Zhang Y, Wang J, Zhu Z. Feature semantic alignment network for rotating machinery domain generalization fault diagnosis with limited labeled samples. *Neurocomputing* 2026, 664:132085.
- [7] Li S, Ji J, Xu Y, Feng K, Zhang K, *et al.* Dconformer: a denoising convolutional transformer with joint learning strategy for intelligent diagnosis of bearing faults. *Mech. Syst. Signal Process.* 2024, 210:111142.
- [8] Shao X, Kim C. Adaptive multi-scale attention convolution neural network for cross-domain fault diagnosis. *Expert Syst. Appl.* 2024, 236:121216.
- [9] Liu Q, Wang D, Yang J, Xu X, Huang D, *et al.* Cross-condition fault diagnosis of wind turbine gearboxes based on improved domain adversarial network. *Measurement* 2026, 279:121833.
- [10] Tang S, Ma J, Yan Z, Zhu Y, Khoo B. Deep transfer learning strategy in intelligent fault diagnosis of rotating machinery. *Eng. Appl. Artif. Intell.* 2024, 134:108678.

- [11] Zhang S, Su L, Gu J, Li K, Zhou L, *et al.* Rotating machinery fault detection and diagnosis based on deep domain adaptation: a survey. *Chin. J. Aeronaut.* 2023, 36:45–74.
- [12] Qian C, Yang Z, He J, Liu Y, Yang M, *et al.* Few-shot cross-domain fault diagnosis of complex mechanical equipment based on multiangle Taylor metric. *IEEE Internet Things J.* 2026, 13(6):11926–11943.
- [13] Sheng C, Zhang M, Rao X, Huang M, Zhang X. Adaptive filtering and multi-scale MixStyle network for cross-domain fault diagnosis of marine electric thruster bearings. *Mech. Syst. Signal Process.* 2026, 246:113910.
- [14] Qian C, Yang Z, He J, Wu C, Ma C, *et al.* A new indirect transfer fault diagnosis method based on feature separation. *Knowledge-Based Syst.* 2024, 300:112058.
- [15] Lu X, Song L, Han C, Jiang Q, Xu W, *et al.* Semi-supervised contrastive domain adaptation network for fault diagnosis of rotating machinery under cross-working conditions. *IEEE Internet Things J.* 2025, 12(12):21552–21564.
- [16] Lu Y, Ruan Y, Tan D, Peng X, Zhong W. Domain adaptation network based on hierarchical shared multi-kernel for cross-condition fault diagnosis of rotating machinery. *Measurement* 2026, 288:122547.
- [17] Chen S, Zheng W, Xiao H, Han P, Luo K. A residual convolution transfer framework based on slow feature for cross-domain machinery fault diagnosis. *Neurocomputing* 2023, 546:126322.
- [18] Wen H, Guo W, Li X. A novel deep clustering network using multi-representation autoencoder and adversarial learning for large cross-domain fault diagnosis of rolling bearings. *Expert Syst. Appl.* 2023, 225:120066.
- [19] Chen X, Shao H, Xiao Y, Yan S, Cai B, *et al.* Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network. *Mech. Syst. Signal Process.* 2023, 198:110427.
- [20] Saito K, Watanabe K, Ushiku Y, Harada T. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceeding of the 2018 IEEE conference on computer vision and pattern recognition*, Salt Lake City, USA, June 18–23, 2018, pp. 3723–3732.
- [21] Jiang Q, Lin X, Lu X, Shen Y, Zhu Q, *et al.* Self-supervised learning-based dual-classifier domain adaptation model for rolling bearings cross-domain fault diagnosis. *Knowledge-Based Syst.* 2024, 284:111229.
- [22] Jin T, Chen C, Guo J, Liu Z, Zhang Y. Double-classifier adversarial learning for fault diagnosis of rotating machinery considering cross domains. *Mech. Syst. Signal Process.* 2024, 216:111490.
- [23] Lu X, Song L, Han C, Jiang Q, Xu W, *et al.* Maximum prediction discrepancy domain adaptation network for rolling bearing RUL prediction under multiple cross-domain scenarios. *Mech. Syst. Signal Process.* 2026, 249:114077.
- [24] Lessmeier C, Kimotho JK, Zimmer D, Sextro W. 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 *Proceeding of the 2016 PHM Society European Conference*, Bilbao, Spain, July 5–8, 2016.
- [25] Yao X, Wang B, Shen Y, Fu G, Zhang D, *et al.* Semi-Supervised Class Adaptive Prototype Network for Cross-Working Rolling Bearing Fault Diagnosis Under Limited Samples. *IEEE Internet Things J.* 2025, 12(17):35589–35601.

-
- [26] Al Shalabi L, Shaaban Z, Kasasbeh B. Data mining: a preprocessing engine. *J. Comput. Sci.* 2006, 2(9):735–739.