



A deep targeted transfer network with clustering pseudo-label learning for fault diagnosis across different Machines

Feiyu Lu, Qingbin Tong^{*}, Xuedong Jiang, Ziwei Feng, Jianjun Xu, Xin Wang, Jingyi Huo

School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China

ARTICLE INFO

Communicated by Matthias Faes

Keywords:

Targeted transfer learning
Local maximum mean discrepancy
Fault diagnosis
Pseudo-label learning

ABSTRACT

In recent years, deep transfer models have addressed the issue of distribution shift between the source and target domains by learning domain-invariant features. However, in cross-machine fault diagnosis scenarios, existing deep learning models struggle to fit the conditional distribution of target domain samples, limiting the performance and generalization of domain adaptation models. To address these problems, we propose a deep targeted transfer network with clustering pseudo-label learning (DTTN-CPLL). DTTN-CPLL consists of three parts. First, a deep transfer network is constructed to extract cross-domain features. To reduce the intra-class distribution gap, a clustering pseudo-label learning algorithm is proposed to create subdomain labels within the target domain features. Then, we simultaneously minimize and maximize the entropy of features and the number of linearly independent vectors in the target domain to reduce the distance between subdomain features. Finally, under the constraint of local maximum mean discrepancy, we reduce the distribution discrepancy between the source and target domains at the subdomain feature level. We conducted 12 cross-machine transfer tasks on three open bearing datasets and a private high-speed train traction motor bearing dataset. The results demonstrate that, compared to other state-of-the-art models, DTTN-CPLL is effective and superior in cross-machine fault diagnosis.

1. Introduction

Intelligent fault diagnosis has received sufficient attention in many industrial scenarios, such as energy power, aerospace, and rail transit [1–4]. Condition monitoring and fault diagnosis are important measures to prevent major accidents and ensure the safety of life and property [5].

With the rapid development of deep learning, intelligent fault diagnosis models, such as autoencoders [6], convolutional neural networks [7–9], etc., have been recognized by more and more scholars. Unlike traditional methods based on fault diagnosis knowledge, intelligent fault diagnosis can automatically learn the underlying diagnosis knowledge from massive data, greatly reducing the time for expert knowledge to be deployed in the model and improving efficiency.

In practical engineering applications, the monitored data often lack sufficient labels, which poses a greater challenge to the training of the model [10]. To solve this problem, transfer learning technology is introduced into fault diagnosis. In the field of transfer

* Corresponding author.

E-mail address: qbtong@bjtu.edu.cn (Q. Tong).

learning, the source domain represents a dataset with sufficient labels, and the target domain represents a dataset with few labels or even no labels. How to use the existing label information to transfer the source domain data knowledge to the target domain is the main task of transfer learning [11,12]. In the existing research on bearing fault diagnosis, transfer learning is mainly used to solve two tasks: 1) Cross-operating conditions: domain transfer tasks of the same machine under different operating conditions. (2) Cross-machines [13]: domain transfer tasks of different machines under different operating conditions.

In addressing the domain transfer task under different operating conditions for Task 1), researchers have employed various strategies and methods to reduce the distribution differences in data. The primary challenge of cross-condition transfer lies in the fact that the data distribution of the same machine may change under different environmental and load conditions. This variation can be caused by factors such as temperature, humidity, speed, load, and various variable operating parameters. To tackle this issue, multiple transfer learning techniques have been proposed. They focus on minimizing the distribution differences between the source domain (a dataset with known conditions and labels) and the target domain (an unlabeled dataset under new operating conditions). Several models or loss functions have been introduced, including maximum mean discrepancy (MMD) [14,15], correlation alignment (CORAL) [16], joint MMD (JMMD) [17], Local MMD (LMMD) [18], adversarial loss [19], and have been successfully applied to fault diagnosis tasks.

For Task 2), the core challenge of cross-machine transfer tasks lies in the potential significant differences in feature distribution among monitoring data from different machines. Due to the heterogeneity of data from different machines, such as varying sampling frequencies, load conditions, and bearing characteristics, the cross-machine transfer task becomes complex. For instance, two identical bearing models operating under different temperature and humidity conditions may exhibit vastly different wear patterns and damage processes, causing significant differences at the data level, thereby affecting the transfer and generalization of fault diagnosis models.

Facing this challenge, unsupervised domain adaptation techniques offer researchers a new approach [20,21]. It aims to explore a strategy to bridge the feature distribution differences between the source domain (labeled data) and the target domain (unlabeled data). Unsupervised domain adaptation methods typically achieve domain alignment between the source and target domains in the feature space, allowing the model to extract more generalized feature representations. Common unsupervised domain adaptation algorithms include but are not limited to domain adversarial neural networks (DANN) and feature alignment-based transfer learning frameworks.

On the other hand, weakly supervised learning [22] takes a different path in transfer learning, allowing the utilization of relatively sparse or incomplete annotation information in the target domain. In this setting, the model is pre-trained based on sufficient annotation information from the source domain and fine-tuned using limited annotation data from the target domain. Weakly supervised transfer learning is particularly important in practice since fully labeled datasets are often unrealistic in many real-world fault diagnosis scenarios. Guiding and optimizing the model based on a small number of labeled samples is feasible.

In summary, unsupervised domain adaptation techniques attempt to transfer information without relying on target domain labels, while weakly supervised learning methods seek to maximize the use of limited fault annotation information in the target domain. The fault diagnosis scenario set in this paper falls within the weakly supervised learning category. Weakly supervised learning aims to enhance transfer effects by utilizing partial label information or easily obtainable auxiliary information, such as fault category labels, leveraging limited labeled information and abundant unlabeled data to improve the accuracy of cross-machine bearing fault diagnosis.

The remainder of the article is structured as follows: In Section 2, we discussed related work, including introductions to the state-of-the-art methods. In Section 3, we give the problem definition. In Section 4, the proposed fault classification framework is presented. Details of four bearing datasets and the corresponding experimental results are presented in Section 5. Section 6 is the conclusion.

2. Related works

To address the issue of cross-machine fault diagnosis, we summarized the current state-of-the-art algorithms from the perspectives of unsupervised domain adaptation and weakly supervised learning, highlighting their overviews and weaknesses. Based on this analysis, we then introduced the problems addressed by the proposed method in this paper and outlined its contributions.

Unsupervised domain adaptation means that when the source domain data label is known and the target domain data label is unknown, the domain adaptive paradigm in transfer learning is used to map the source domain and target domain features to the same space, to achieve feature transfer [23]. Yang et al. [24] realized the transfer of laboratory bearing data to locomotive bearing data by using multi-layer MMD and pseudo-label learning algorithm. Guo et al. [25] proposed deep convolutional transfer learning network (DCTLN) by combining MMD and conditional recognition loss function. The model realizes cross-testing in three different bearing datasets, and the better diagnostic accuracy proves that the unsupervised cross-machines transfer method is feasible. Considering the alignment of data with the same label in subdomains, Liu et al. [26] proposed the deep adversarial subdomain adaptation network (DASAN), which integrates domain adaptation, domain adversarial, and pseudo-label algorithmic functions, and has been validated in both cross-speed and cross-machines tasks. Although the unsupervised domain adaptive method is robust to some extent, when the data structures of the source domain and target domain differ greatly, negative transfer is likely to occur, which cannot meet the requirements of the intelligent fault diagnosis model for strong generalization capability.

Weakly supervised learning can alleviate the above problems. It refers to that only a small number of data labels in the target domain are known, and the model is trained with a large number of source domain data, so as to realize the diagnosis task from the source domain to the target domain. Based on the deep transfer learning model, Li et al. [27] realized the alignment of features in the source domain and the target domain through adversarial training, which was verified in bearing and coupling data. However, this method requires a large amount of label data in the target domain. To fit the distribution of target domain data, Yang et al. [28] proposed an optimal transport embedded joint distribution similarity measure (OT-JDSM) and applied it to transfer learning tasks on six different bearing datasets, among which, it is assumed that 5 % of the data labels in the target domain are known. Han et al. [29]

proposed a modified convolutional neural network (MCNN), and validated it in three different types of gearbox fault data by using parameter transfer strategy. In addition, the authors discussed the influence of the number of labeled data in the target domain (512, 256, 128, 64, and 32) on the performance of the model.

In the above work, there are still the following two problems: 1) The unsupervised domain adaptation method focuses on the subdomain adaptation problem, but ignores the inherent differences between the target domain and the source domain at the feature level, making it impossible to fully fit the data distribution in the target domain. Moreover, the use of pseudo-labels instead of real labels leads to poor feature alignment or even negative transfer [30]. 2) Most weakly supervised adaptive methods rarely consider the case of a single label in the target domain, so the overall performance of the weakly supervised model under very few labels is questionable.

To solve the above problems, this paper proposes a deep targeted transfer network with clustering pseudo-label learning (DTTN-CPLL) under the framework of weakly supervised learning. The model consists of feature extractor, label classifier and multiple loss functions. The feature extractor and the label classifier parameters are shared, the feature extractor automatically extracts the feature vectors of the source domain and the target domain, and the label classifier clusters similar features. Specifically, firstly, the source and target domain data from different devices are transformed from the time domain to the frequency domain. Then, a feature extractor based on a six-layer convolutional structure is constructed to extract common features from the frequency domain data. We develop a dynamic pseudo-label learning algorithm and embed it into a semi-supervised learning framework. Finally, the proposed domain adaptation loss function is utilized to reduce the distance between features at the level of latent discriminative features, and a linear neural network classifier is employed to distinguish fault types. The main contributions of this paper are as follows:

(1) A clustering pseudo-label learning (CPLL) algorithm is proposed, which is dynamic and can update the pseudo-label with the number of iterations. The ablation experiment also shows that CPLL has a significant effect on improving the performance of the model.

(2) Unlike many weakly supervised learning, this paper simulates a more rigorous engineering application: the number of labels available in the target domain is only one for each health state.

(3) A domain adaptive function is designed based on LMMD and the proposed CPLL, which can effectively reduce intra class and inter class intervals, thereby reducing domain offset and improving model classification performance.

(4) A cross device fault diagnosis method based on DTTN-CPLL is proposed, and the effectiveness and superiority of DTTN-CPLL are verified in four different device fault datasets.

3. Problem definition

In this paper, the mechanical fault diagnosis is studied under the following assumptions and definitions.

(1) The source domain and target domain data come from different mechanical equipment, and the types of equipment being tested

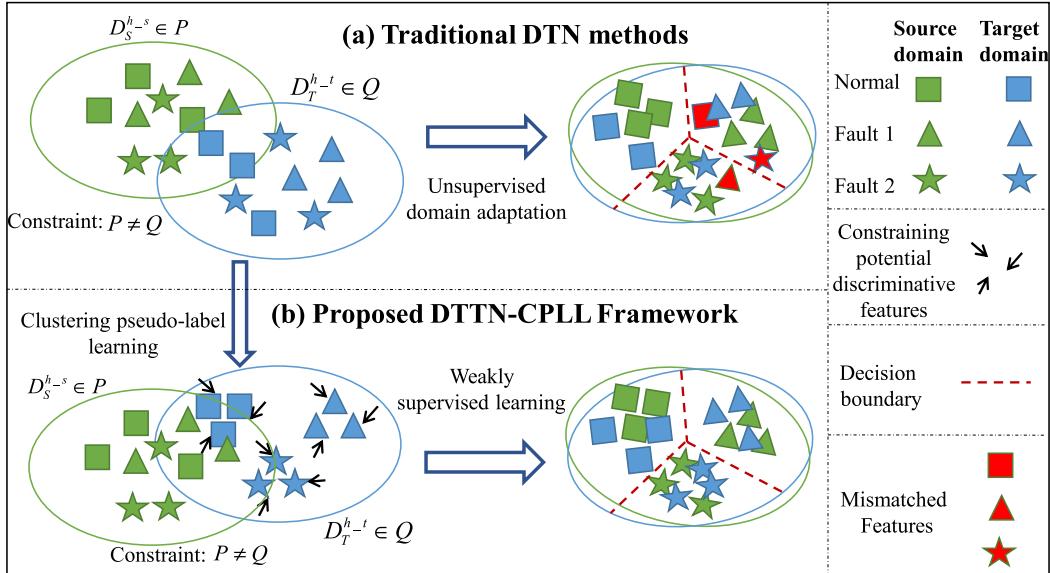


Fig. 1. The comparison of DTN methods and the proposed DTTN-CPLL framework. Using normal data and two types of fault data as examples, assuming the source domain $D_S^{h-s} \in P$ and target domain data $D_T^{h-t} \in Q$, $P \neq Q$ indicate differences in data distribution. (a) Traditional DTN ignores exploring the underlying discriminative features of the target data, which may lead to suboptimal performance. (b) DTTN-CPLL considers the differences in the distribution of subdomain features and utilizes clustering pseudo-label learning algorithms to implement the label propagation process, constraining potential discriminative features and greatly improving the diagnostic performance of the model.

are not the same, resulting in differences in data distribution, such as different test environments, different structures, different speeds, or loads.

(2) The source domain data labels are known, and a supervised classification model can be constructed.

(3) Only a single data label is known for each health state in the target domain data, and the remaining data labels are unknown, so it is impossible to construct an effective classifier alone.

Specifically, the source domain $D_S^{h-s} = \{(x_s^{(i)}, y_s^{(i)})\}_{i=1}^N$ and the target domain $D_T^{h-t} = \{(x_t^{(i)})\}_{i=1}^M$ are from different mechanical equipment, N, M are the number of samples in the source domain and the target domain, respectively, h_s, h_t are the corresponding number of health status. In addition, the labels of $D_S^{h-s} = \{(x_s^{(i)})\}_{i=1}^N \{(y_s^{(i)})\}_{i=1}^N$ are known, and $D_T^{h-t} = \{(x_t^{(i)})\}_{i=1}^M$ has n_t unlabeled samples and h_t labeled samples (where $n_t + h_t = M$). Let P, Q be the feature discrepancy of D_S^{h-s} and D_T^{h-t} . Usually, affected by the difference in data distribution between different devices, domain shift $P \neq Q$ occurs, which makes unsupervised models unreliable.

To this end, we propose a DTTN-CPLL model under the weakly supervised framework. Different from the traditional deep transfer network (DTN) model, DTTN-CPLL embeds CPLL and distance constraint (DC) algorithms to constrain the probability distribution of the target domain data, and propagates the label information to the unlabeled data in the target domain. Because this process is embedded in the iterative learning of neural networks, automation is achieved. The comparison of DTN and DTTN-CPLL is shown in Fig. 1.

4. The proposed fault diagnosis framework

In this section, the bearing fault diagnosis framework based on DTTN-CPLL is introduced. The core idea of DTTN-CPLL is to constrain the data in the target domain and propagate the limited labels to the entire data space to achieve the alignment of subdomain features, thereby completing cross-machines fault diagnosis from different device data.

4.1. Deep targeted transfer network with clustering pseudo-label learning

The proposed DTTN-CPLL model consists of feature extractor, classifier and four loss functions. Loss functions include classification loss, DC, LMMD, and CPLL. The feature extractor and classifier are twin neural network structures, and the detailed structural parameters are shown in Table 1. The feature extractor automatically extracts highly abstract low-dimensional feature data, and the classifier predicts the health category corresponding to the data.

Under the constraint of multiple loss functions, the weight and bias are adjusted through the back propagation algorithm. Specifically, supervised classification loss is used to fit the distribution of source domain data. The DC loss narrows the intra-class spacing and increases the inter-class spacing by enhancing the predictive discriminability of the target domain data. CPLL dynamically adjusts the clustering of target domain data and spreads limited labels to the entire target domain data. Finally, the data features are aligned in the subdomain with the LMMD.

4.1.1. Feature extractor and classifier

The feature extractor G_f is used to extract similar features. It consists of six layers of convolution and pooling structures, as shown in Table 1. For Conv-Pool-1, it is composed of 16 convolution kernels, each with a size of 64×1 . The padding number is 1, BN is the abbreviation of batch normalization, ReLU is the nonlinear activation function, and Max-Pool is the maximum pooling. Other layers have the same meaning.

The classifier G_c consists of two basic fully connected neural networks, of which No. Category is the number of fault states. For source domain data, using labeled supervised learning can ensure the separability between source domain data, thus improving the diagnostic accuracy. The process is realized by cross entropy loss function, and its expression is as follows:

$$L_C = \mathbb{E}_{(x_s^{(i)}, y_s^{(i)}) \in D_S^{h-s}} [-\log(\hat{y}_c^{(n)})] = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^K y_c^{(n)} \log \frac{\exp(\hat{y}_c^{(n)})}{\sum_{\tilde{c}=1}^C \exp(\hat{y}_{\tilde{c}}^{(n)})} \quad (1)$$

Table 1
Structure parameters of DTTN.

Networks	Layers	Operations
Feature extractor	Conv-Pool-1	Kernel 16-64 $\times 1$, Stride 8, Padding 1; BN; ReLU; Max-Pool 2 $\times 1$, Stride 2
	Conv-Pool-2	Kernel 32-3 $\times 1$, Stride 1, Padding 1; BN; ReLU; Max-Pool 2 $\times 1$, Stride 2
	Conv-Pool-3	Kernel 64-3 $\times 1$, Stride 1, Padding 1; BN; ReLU; Max-Pool 2 $\times 1$, Stride 2
	Conv-Pool-4	Kernel 64-3 $\times 1$, Stride 1, Padding 1; BN; ReLU; Max-Pool 2 $\times 1$, Stride 2
	Conv-Pool-5	Kernel 64-3 $\times 1$, Stride 1, Padding 1; BN; ReLU; Max-Pool 2 $\times 1$, Stride 2
	Conv-Pool-6	Kernel 1024-3 $\times 1$, Stride 1, BN; ReLU; Max-Pool 2 $\times 1$, Stride 2
Classifier	Linear-1	Node: 256
	Linear-2	Node: No. category; Softmax

where, N is the number of samples, K is the category corresponding to the sample. $y_c^{(n)}$ is a symbolic function. If the true category of sample n is c , take 1, otherwise take 0. $\hat{y}_c^{(n)}$ is the characteristic value of the n -th sample in FC at the c label.

4.1.2. Distance constraint

Most existing methods consider the display feature alignment of source domain and target domain data, ignoring the implicit alignment of models at the prediction level. Specifically, we can reduce the intra class distance indirectly by strengthening the discriminability of the model prediction results. High discriminability corresponds to low uncertainty of prediction results, so Shannon entropy [31] is used to measure uncertainty, which is denoted as follows:

$$H(\mathbf{Y}) = -\frac{1}{B} \sum_{i=1}^B \sum_{j=1}^C Y_{i,j} \log(Y_{i,j}) \quad (2)$$

where, $\sum_{j=1}^C Y_{i,j} = 1 \forall i \in 1 \dots B$, $\mathbf{Y} \in \mathbb{R}^{B \times C}$ is the prediction result matrix, B is batch size of model input. Let C be the number of labels. Frobenius-norm of output matrix \mathbf{Y} is as follows:

$$\|\mathbf{Y}\|_F = \sqrt{\sum_{i=1}^B \sum_{j=1}^C |Y_{i,j}|^2} \leq \sqrt{\sum_{i=1}^B \left(\sum_{j=1}^C Y_{i,j} \right) \cdot \left(\sum_{j=1}^C Y_{i,j} \right)} = \sqrt{\sum_{i=1}^B 1 \cdot 1} = \sqrt{B} \quad (3)$$

According to [32], the monotonicity of $\|\mathbf{Y}\|_F$ and $H(\mathbf{Y})$ is opposite. Therefore, maximizing $\|\mathbf{Y}\|_F$ is equivalent to minimizing $H(\mathbf{Y})$. Thus, we can maximize $\|\mathbf{Y}\|_F$ to enhance the discriminability. In [33,34], the Frobenius-norm $\|\mathbf{Y}\|_F$ and the nuclear-norm $\|\mathbf{Y}\|_\circ$ have the following relationship:

$$\frac{1}{\sqrt{D}} \|\mathbf{Y}\|_\circ \leq \|\mathbf{Y}\|_F \leq \|\mathbf{Y}\|_\circ \leq \sqrt{D} \cdot \|\mathbf{Y}\|_F \quad (4)$$

where, $D = \min(B, C)$. And (14) shows that $\|\mathbf{Y}\|_\circ$ tends to be larger with the increase of $\|\mathbf{Y}\|_F$. So we can use the nuclear-norm $\|\mathbf{Y}\|_\circ$ to enhance prediction discriminability. The nuclear norm can express the diversity of the prediction results of the model [32]. Based on the above analysis, the distance constraint loss function is denoted as follows:

$$L_{DC} = -\frac{1}{B} \|\mathbf{Y}\|_\circ \quad (5)$$

where, $\|\mathbf{Y}\|_\circ$ is the nuclear-norm of the model prediction matrix.

4.1.3. Clustering pseudo-label learning (CPLL)

For the case where each health state in the target domain data has only a single label, we propose a new pseudo-label learning algorithm, which consists of five steps.

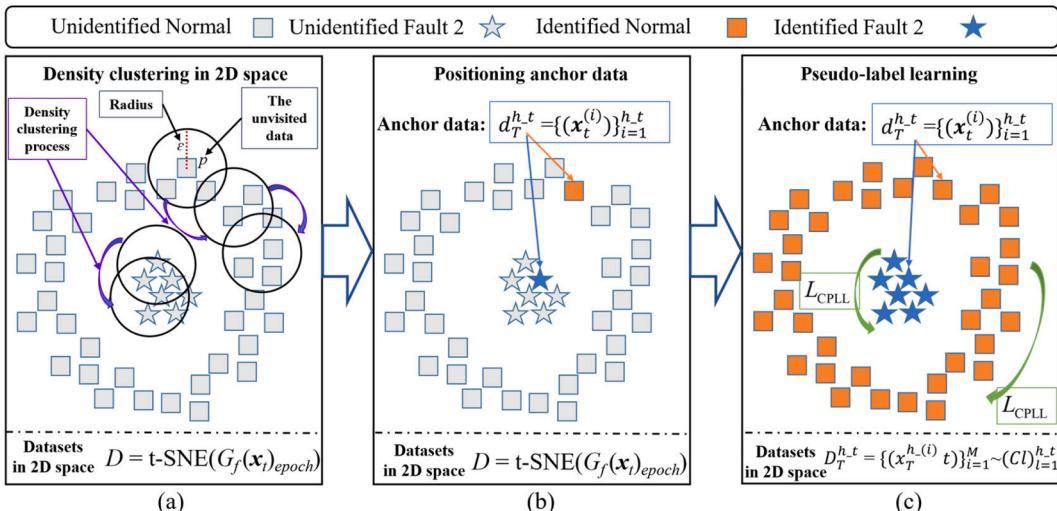


Fig. 2. Illustration of CPLL. Assume that there are two fault type data, namely unidentified normal and unidentified fault 2. First, obtain the output features t-SNE($G_c^{Linear-1}(G_f(\mathbf{x}_t))_{epoch}$) through Step 1 and Step 2, then perform Step 3 ((a) Density clustering in 2D space) to obtain l category data, and use Step 4 ((b) Positioning anchor data) to obtain the anchor point in the target domain. Finally, the labels of each category are obtained through Step 5 ((c) Pseudo-label learning).

Step 1: The output result of target domain data $D_T^{h-t} = \{(\mathbf{x}_t^{(i)})\}_{i=1}^M$ through the feature extractor G_f is $G_c^{Linear-1}(G_f(\mathbf{x}_t))_{epoch}$. epoch is the iteration number of model training.

Step 2: Data dimensionality reduction based on t-distributed stochastic neighbor embedding (t-SNE): t - SNE($G_c^{Linear-1}(G_f(\mathbf{x}_t))_{epoch}$), The algorithm principle can refer to [35].

Step 3: Density based clustering: Randomly select a point in t - SNE($G_f(\mathbf{x}_t)_{epoch}$), take the point as the origin, and establish a neighborhood with a radius of ϵ . If the amount of data in the neighborhood is greater than or equal to MinPts, then the points in the neighborhood form a cluster, and then recurse the points in the neighborhood in the same way to form a cluster. Use the same method to process the data in t - SNE($G_f(\mathbf{x}_t)_{epoch}$) that has not been analyzed to form l cluster.

Step 4: Set anchor point: set corresponding anchor points according to h_t labeled samples in the target domain.

Step 5: Pseudo-label learning: determine the label type in each cluster according to the position of anchor $\{(\mathbf{x}_t^{(i)})\}_{i=1}^{h-t}$ in κ cluster.

Fig. 2 shows the algorithm concept diagram of CPLL. Algorithm 1 is the pseudo code of CPLL. The target domain data is supervised by using the pseudo-label learned above. The pseudo-label loss function is defined as follows:

$$L_{CPLL} = -\frac{1}{M} \sum_{j=1}^M \sum_{m=1}^C p(\hat{y}_j^t = m | \mathbf{x}_j^t) \log p(\hat{y}_j^t = m | \mathbf{x}_j^t) \quad (6)$$

Where \hat{y}_j^t is the pseudo-label of the j -th target sample calculated by CPLL.

4.1.4. Local maximum mean discrepancy (LMMD)

Although CPLL can solve the problem that the data distribution in the target domain is difficult to fit, to achieve feature alignment based on the subdomain, we need to train the model under the driving of the subdomain adaptive. To this end, LMMD algorithm is introduced. Its principle is as follows.

$$d_{\mathcal{H}}(p, q) \triangleq \mathbf{E}_c \| \mathbf{E}_{p^{(c)}} [\phi(\mathbf{x}_s)] - \mathbf{E}_{q^{(c)}} [\phi(\mathbf{x}_t)] \|_{\mathcal{H}}^2 \quad (7)$$

where \mathbf{x}_s and \mathbf{x}_t are the samples in source domain and target domain. $p^{(c)}$ and $q^{(c)}$ are the distributions of D_S^{h-s} and D_T^{h-s} , respectively. Let w_i^c be the weight of each sample belonging to each category. (7) can be rewritten as (8).

$$L_{LMMD} = \frac{1}{K} \sum_{c=1}^K \left\| \sum_{\mathbf{x}_i \in \mathcal{D}_s^s} w_i^{sc} \phi(\mathbf{x}_i^s) - \sum_{\mathbf{x}_j \in \mathcal{D}_t} w_j^{tc} \phi(\mathbf{x}_j^t) \right\|_{\mathcal{H}}^2 \quad (8)$$

Since $\sum_{i=1}^N w_i^{sc}$ and $\sum_{i=1}^M w_i^{tc}$ are both equal to one, $\sum_{\mathbf{x}_i \in \mathcal{D}_s^s} w_i^c \phi(\mathbf{x}_i)$ is the weighted sum of category c . And the calculation formula of w_i^c is as follows.

$$w_i^c = \frac{y_{ic}}{\sum (\mathbf{x}_i, \mathbf{y}_j) \in \mathcal{D}^{y_{ic}}} \quad (9)$$

where y_{ic} is the c -th label of vector \mathbf{y}_i . LMMD is embedded in the output layer of the feature extractor G_f . Therefore, the LMMD loss function is calculated as follows:

$$\begin{aligned} L_{LMMD} = & \frac{1}{C} \sum_{c=1}^C \left[\sum_{i=1}^{n_s} \sum_{j=1}^{n_t} w_i^{sc} w_j^{tc} k\left(G_c^{Linear-1}(G_f(\mathbf{x}_i^s)), G_c^{Linear-1}(G_f(\mathbf{x}_j^t))\right) \right. \\ & + \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} w_i^{tc} w_j^{sc} k\left(G_c^{Linear-1}(G_f(\mathbf{x}_i^s)), G_c^{Linear-1}(G_f(\mathbf{x}_j^t))\right) \\ & \left. - 2 \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} w_i^{sc} w_j^{tc} k\left(G_c^{Linear-1}(G_f(\mathbf{x}_i^s)), G_c^{Linear-1}(G_f(\mathbf{x}_j^t))\right) \right] \end{aligned} \quad (10)$$

where $G_f(\mathbf{x}_i^s)$ and $G_f(\mathbf{x}_j^t)$ are the output value of G_f . $G_c^{Linear-1}(\bullet)$ is the output value of the first linear layer of the classifier.

Algorithm 1: Clustering Pseudo-label Learning (CPLL)

Input:

- Dataset $D = t$ - SNE($G_c^{Linear-1}(G_f(\mathbf{x}_t))_{epoch}$) which containing h_t labeled samples
- MinPts: the threshold of field density.
- Radius ϵ .

(continued on next page)

(continued)

Algorithm 1: Clustering Pseudo-label Learning (CPLL)

Output: The labeled dataset D_T^{h-t}

- 1: Mark all samples in D as unvisited data
- 2: **Do**
- 3: Randomly select a sample p from the unvisited data. M is the number of samples in field $p-\epsilon$. ϕ is a collection within the field $p-\epsilon$.
- 4: **if** $M \geq \text{MinPts}$
- 5: Create l -th cluster C_l , and put p into C_l .
- 6: **for** p' in ϕ **do**
- 7: **if** p' is unvisited
- 8: Mark p' as a visited data.
- 9: **if** the number of $p'-\epsilon \geq \text{MinPts}$
- 10: Put the samples in the field $p'-\epsilon$ into ϕ .
- 11: **if** $p' \notin C$
- 12: Put p' into C_l .
- 13: **end for**
- 14: Output C_l
- 15: **else** p is noise
- 16: **until** there are no unvisited samples in D .
- 17: $d_T^{h-t} = \{(x_t^{(i)})\}_{i=1}^{h-t}$
- 18: $D_T^{h-t} = \{(x_t^{(i)})\}_{i=1}^M \sim (C_l)_{l=1}^{h-t}$

4.2. Training process

The proposed fault diagnosis method is shown in Fig. 3. The overall loss function of the DTTN-CPLL model is as follows:

$$L_{\text{Total}} = L_C + \alpha L_{\text{CPLL}} + \beta(L_{\text{DC}} + L_{\text{LMMD}}) \quad (11)$$

where α is the weight coefficient with a variable value, and its calculation equation is as follows.

$$\beta = \frac{2}{1 + \exp(-10 \times q/Q)} - 1 \quad (12)$$

where q is the number of model training iterations and Q is the total number of training iterations.

Let θ_f and θ_c are the parameters of the feature extractor and classifier. In the process of DTTN-CPLL model training, the back propagation algorithm is used to update the model parameters. The formula is as follows:

$$\theta_f = \theta_f - \lambda \frac{\partial L_{\text{Total}}}{\partial \theta_f} \quad (13)$$

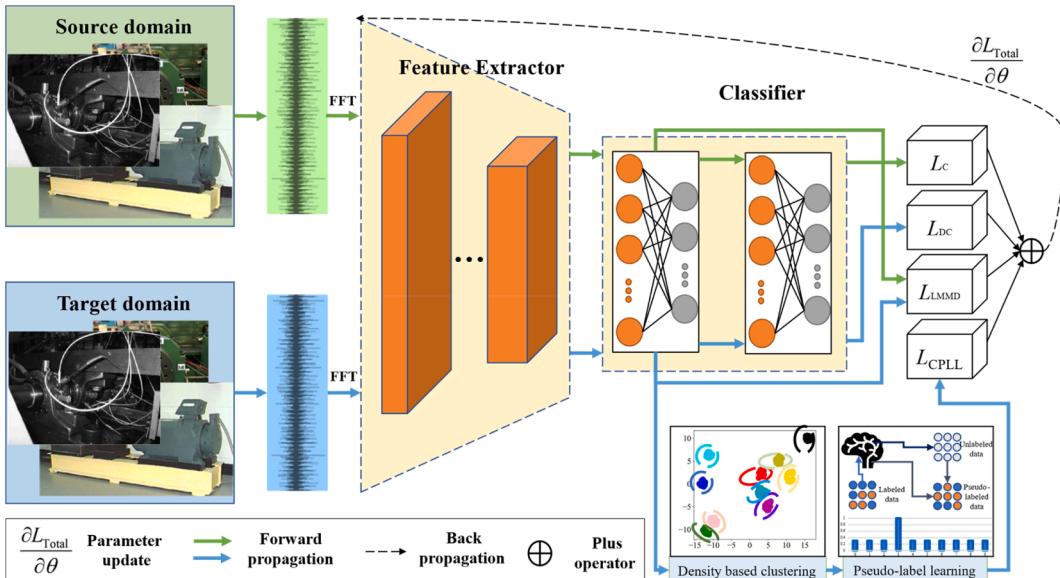


Fig. 3. Illustration of the proposed fault diagnosis method.

$$\theta_c = \theta_c - \lambda \frac{\partial L_{\text{Total}}}{\partial \theta_c} \quad (14)$$

where ∂ represents the partial derivative operator and λ denotes the learning rate. And the pseudocode of the DTTN-CPLL model training process is shown in Algorithm 2.

5. Experiments

5.1. Dataset description

We evaluated the performance of DTTN-CPLL in 12 fault diagnosis tasks on four mechanical equipment, including Case Western Reserve University (CWRU) dataset [36] for dataset A, dataset B comes from Intelligent Maintenance System (IMS) [37], dataset C comes from DataCastle [38], and dataset D comes from Beijing Jiaotong University (BJTU). For each dataset, there are 1000 samples in each health state, and a total of 4000 samples., And the length of each sample is 2400. 50 % of them are training sets. The details of the above dataset are as follows.

Algorithm 2: The training process for the proposed DTTN-CPLL

Input:

- Initial feature extractor G_f and a classifier G_c .
- The source domain $D_S^{b-s} = \{(x_s^{(i)}, y_s^{(i)})\}_{i=1}^N$ and the target domain $D_T^{b-t} = \{(x_t^{(i)})\}_{i=1}^M$ after fast Fourier transform (FFT).
- Learning rate $\lambda = 0.02$, batchsize = 64, epoch = 200, $\alpha = 0.6$ and Ranger optimizer.

- 1: **for** epoch **do**
- 2: Calculate the classification loss using (1).
- 3: Calculate the distance constraint loss using (5).
- 4: Calculate the pseudo-label loss with (6).
- 5: Calculate the subdomain adaptive loss with (10).
- 6: Obtain the overall objective with (11).
- 7: Train and update model parameters with (13) and (14).
- 8: **end for**

9: **Output:** The DTTN-CPLL model with trained parameters.

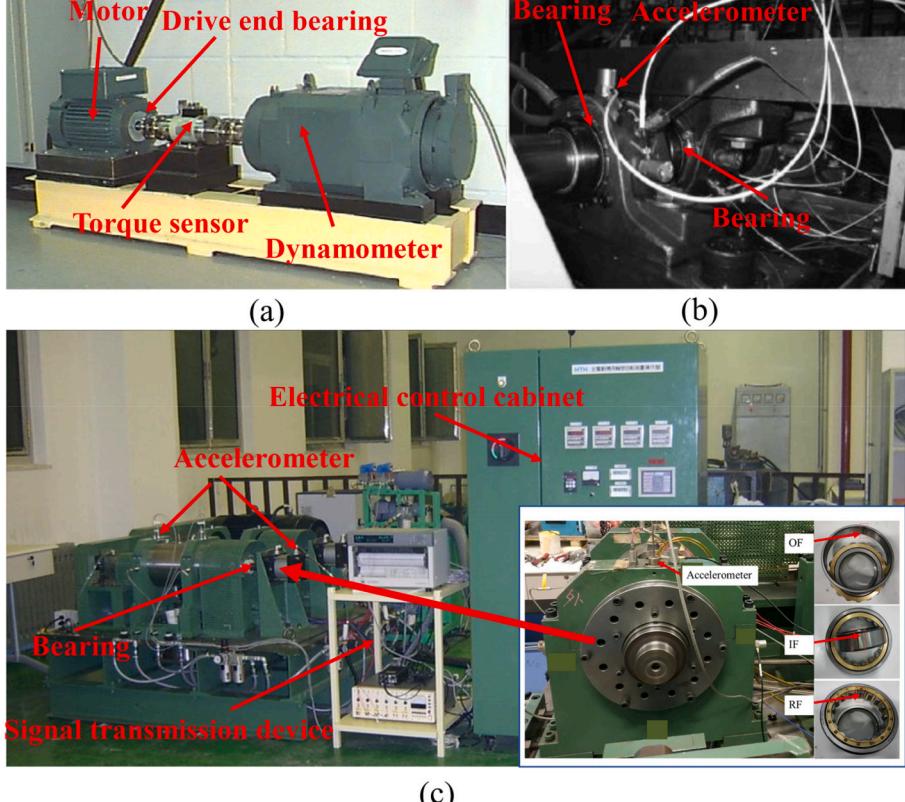


Fig. 4. The test rig of the (a) CWRU, (b) IMS and (c) BJTU dataset.

(1) CWRU datasets: CWRU datasets [39] that were collected under constant speed are utilized for analysis. The sampling frequency at the driving end is 12 kHz. Its test rig is shown in Fig. 4(a). There are four health states: including normal condition (NC), outer race faults (OF), inner race faults (IF), roller faults (RF). The data are obtained at 1750 rpm.

(2) IMS datasets refer to the life cycle data from the Center for Intelligent Maintenance Systems. Its test rig is shown in Fig. 4(b). The sampling frequency is 20 kHz. Similar to [25], we have built dataset B. There are four health states: NC, OF, IF, and RF.

(3) DataCastle datasets come from the global developer competition [38]. However, this data has no details, only labels and data. Similarly, we use four health states (NC, OF, IF, RF) to create datasets.

(4) BJTU datasets: We use the BJTU dataset which were collected under constant speed to test the performance of the method proposed, and the experiments are conducted using the high-speed EMU traction motor bearing test bench from BJTU, as shown in Fig. 4(c). The bearing failure points are made by wire cutting. The sampling frequency is 100 kHz, and there are four health states: NC, OF, IF, and BF. The details of the dataset are summarized in Table 2.

5.2. Approaches for comparison and implementation details

To fully verify the superiority of the proposed method, we use the current advanced model for comparative analysis, as follows:

- (1) The CNN model consists of feature extractor and classifier in Table I. And only the label data in the source domain is supervised, and the trained model is used to test the test set, that is, there is no domain adaptive loss function.
- (2) The domain adversarial neural network (DANN) [27] adds an additional domain discriminator on the basic CNN, and DANN can map the source domain and target domain data features to the same feature space.
- (3) DCTLN is an unsupervised cross-machines fault diagnosis model. We use the model in the literature [25] to compare with the proposed method.
- (4) DASAN [26] is also an unsupervised fault diagnosis model, which is consistent with the super parameter equipment and literature.
- (5) Joint distribution alignment (JDA) [40]: The JDA model is a joint distribution adaptation model proposed by Han et al. It is used in cross-machine fault diagnosis tasks later in this paper. The parameters and model configurations are consistent with the original paper.
- (6) DDTLN [41]: DDTLN is a recently proposed domain adaptation model for cross-machine tasks that improves the conditional distribution alignment mechanism.
- (7) DCTLN + weak supervision learning (WSL) is based on DCTLN and uses the labeled data existing in the target domain to do weak supervision training.
- (8) DASAN + WSL is based on DASAN and uses the tag data existing in the target domain for weak supervision training.

5.3. Results

To verify the superiority of the proposed method in cross equipment bearing fault diagnosis, we conducted cross validation on the four datasets mentioned above. There are 12 fault diagnosis tasks in total. The fault diagnosis accuracy of the proposed method and

Table 2
Bearing dataset details.

Datasets	Equipment	Faults	Labels	Working conditions
A	Motor bearing	Normal	NC	1750 r/min 2 HP
		Outer	OF	1750 r/min 2 HP
		Inner	IF	1750 r/min 2 HP
		Roller	RF	1750 r/min 2 HP
B	Shaft support bearing	Normal	NC	2000 r/min 26.6 kN
		Outer	OF	2000 r/min 26.6 kN
		Inner	IF	2000 r/min 26.6 kN
		Roller	RF	2000 r/min 26.6 kN
C	Bearing	Normal	NC	Unknown [38]
		Outer	OF	Unknown [38]
		Inner	IF	Unknown [38]
		Roller	RF	Unknown [38]
D	Traction motor bearing	Normal	NC	2873 r/min 26.2 kN
		Outer	OF	2873 r/min 30.9 kN
		Inner	IF	2766 r/min 26.0 kN
		Roller	RF	2765 r/min 25.7 kN

other six comparison methods in each task is shown in [Table 3](#).

The results in [Table 3](#) show that the proposed method has the highest average diagnostic accuracy of 99.98 % across machines, and the worst performance is DANN (39.89 %) rather than the base-line CNN model (49.88 %). DCTLN and DASAN are advanced cross-machines models, and their accuracies are 41.60 % and 55.78 %, respectively. This is because unsupervised cross-machines fault diagnosis is difficult to fit the target domain data distribution. After considering weakly supervised learning, the diagnostic performance of DCTLN + WSL and DASAN + WSL is greatly improved, which is 72.92 % and 92.32 % respectively. It is proved that weakly supervised learning can promote accuracy. It is worth noting that the diagnosis results of DTTN-CPLL fluctuate the least, which indicates that DTTN-CPLL has strong generalization ability for different transfer tasks. Besides, [Fig. 5](#) is the histogram of experimental results of all methods.

In addition, in order to intuitively see the ability of the proposed method in subdomain feature alignment, we use t-SNE algorithm to visually analyze A → B task. The results are shown in [Fig. 6](#). The data points are derived from the data features after domain adaptive analysis of each method. Specifically, the output features of the source domain and target domain data after passing through the feature extractor and Linear-1 in the classifier are discrete points in the figure. It can be seen that DANN, DCTLN and DASAN do not align the source domain and target domain data well, and the same features between different domains cannot be aggregated into the same subdomain space. DCTLN + WSL and DASAN + WSL can aggregate data in the target domain, which is the result of weakly supervised learning. Under the joint action of CPLL, DC, and LMMD, the proposed DTTN-CPLL aligns the same fault features in the source domain and target domain, and achieves better fault diagnosis results.

5.4. Analysis and discussion

5.4.1. Parameters sensitivity of loss function weight

To determine the optimal loss function weight parameter, we have done parameters sensitivity experiments. For (11), the α value is 0 to 1 with an interval of 0.1. The across-machines diagnosis results under 11 parameters are shown in [Fig. 7](#). It can be seen that the proposed method has the best effect in all diagnostic tasks when $\alpha = 0.6$.

5.4.2. Parameters sensitivity for CPLL

In this section, the sensitivity of two parameters of CPLL is analyzed. Specifically, the value range of Radius is: [0.5, 1, 5, 10, 20], MinPts range: [1,2,5,8,10]. We use the grid search method to select appropriate parameter values. For task A → B, the impact of each parameter combination on the final diagnosis accuracy is shown in [Fig. 8](#). It can be seen that the diagnosis effect of the model is optimal when Radius = 10 and MinPts = 2.

5.4.3. Ablation study

In order to prove the necessity of the existence of each loss function in the proposed model, we carried out ablation tests, namely none L_{CPLL} (without (w/o) L_{CPLL}), none L_{DC} (w/o L_{DC}), none L_{LMMD} (w/o L_{LMMD}). The results in [Table 4](#) show that no matter which loss function is deleted under the 12 tasks, the results will be affected. Importantly, deletion L_{CPLL} has the greatest impact on the diagnostic accuracy, which fully proves the effectiveness of the proposed CPLL algorithm.

5.4.4. The effect of the loss function combination selection

To select the best combination of loss functions, all combinations are listed, as shown in [Table 5](#). Cross-machine fault diagnosis research is carried out for the loss functions under eight combinations. The results are shown in [Fig. 9](#). It can be seen that the proposed method (combination 1) has the highest diagnostic accuracy and the best robustness in all transfer tasks.

5.4.5. Performance of subdomain feature alignment

The performance of the proposed method in subdomain feature alignment is evaluated quantitatively. For A → B tasks, we use intra class spacing to measure the degree of aggregation of the same label data in the target domain. The calculation formula of intra class

Table 3

Diagnosis accuracy of the proposed method and other compared methods.

tasks	CNN	DANN	DCTLN	DASAN	JDA	DDTLN	DCTLN + WSL	DASAN + WSL	DTTN-CPLL
A → B	49.68 %	49.95 %	25.00 %	25.00 %	25.10 %	42.20 %	47.70 %	89.95 %	99.95 %
A → C	27.75 %	25.00 %	50.00 %	50.45 %	50.00 %	43.75 %	68.25 %	99.75 %	99.90 %
A → D	64.95 %	51.95 %	100.00 %	100.00 %	100.00 %	98.00 %	100.00 %	100.00 %	100.00 %
B → A	46.50 %	25.00 %	13.95 %	24.35 %	25.00 %	39.05 %	72.70 %	99.60 %	100.00 %
B → C	66.30 %	50.05 %	31.10 %	50.00 %	50.00 %	64.75 %	65.05 %	100.00 %	100.00 %
B → D	25.00 %	25.00 %	24.65 %	1.50 %	25.00 %	41.45 %	25.00 %	99.50 %	100.00 %
C → A	50.00 %	50.00 %	50.20 %	50.00 %	52.55 %	50.85 %	75.00 %	89.40 %	100.00 %
C → B	43.10 %	25.40 %	25.00 %	31.20 %	25.00 %	47.35 %	50.85 %	75.00 %	100.00 %
C → D	50.00 %	50.00 %	50.00 %	56.20 %	53.85 %	50.35 %	57.25 %	100.00 %	100.00 %
D → A	98.30 %	60.80 %	50.00 %	100.00 %	100.00 %	99.85 %	100.00 %	100.00 %	100.00 %
D → B	35.50 %	35.80 %	29.25 %	50.00 %	50.00 %	47.60 %	49.60 %	100.00 %	99.95 %
D → C	41.45 %	29.75 %	50.00 %	50.00 %	26.50 %	60.55 %	75.00 %	100.00 %	100.00 %
Average	49.88 %	39.89 %	41.60 %	49.06 %	48.58 %	57.15 %	65.53 %	96.10 %	99.98 %

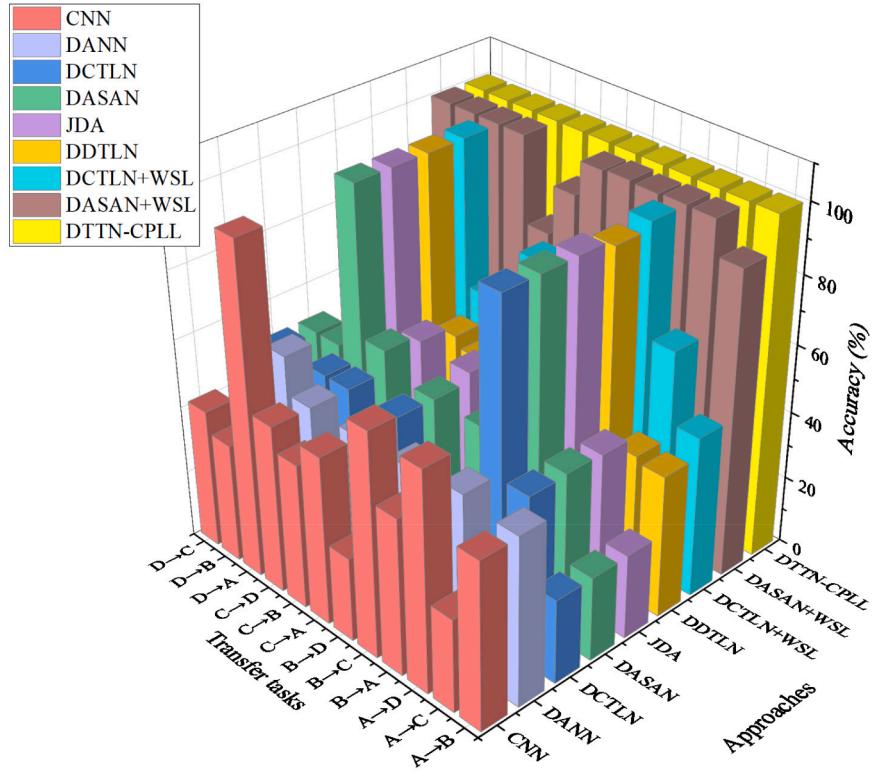


Fig. 5. The fault diagnosis results of all approaches.

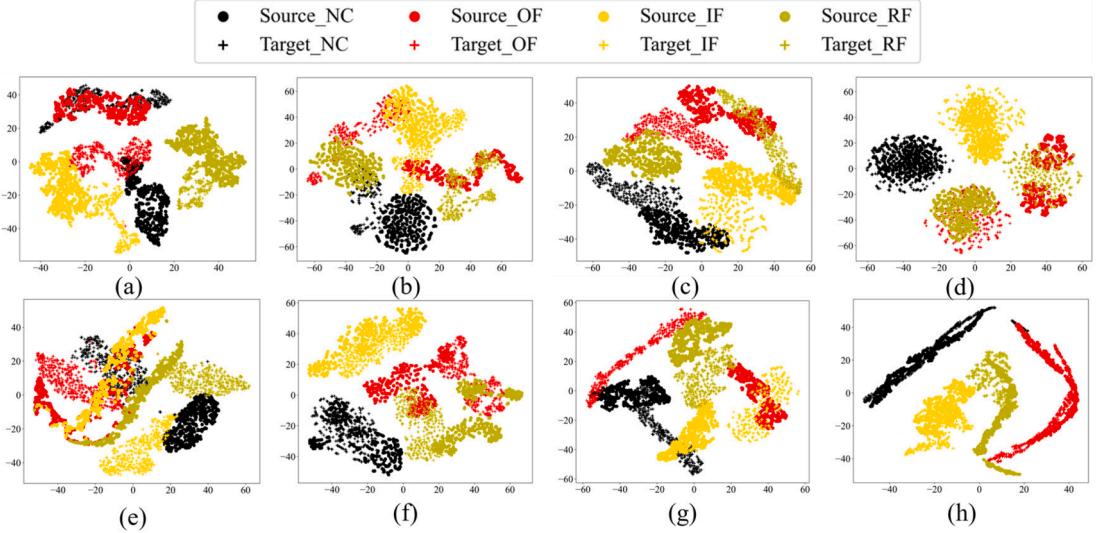


Fig. 6. Visual results for different methods using t-SNE: (a) DANN. (b) DCTLN. (c) DASAN. (d) JDA. (e) DDTLN. (f) DCTLN + WSL. (g) DASAN + WSL. (h) DTTN-CPLL.

spacing is as follows.

$$d_{\text{within}} = d^2(\Omega_i) = \frac{1}{N_i N_i} \sum_{k=1}^{N_i} \sum_{l=1}^{N_i} d^2(X_k^{(i)}, X_l^{(i)}) \quad (15)$$

$X_1^{(i)}, X_2^{(i)} \dots X_{N_i}^{(i)}$ is the sample in Ω_i . N_i is the number of samples in the category. There are four health states, so $i = 4$. The box plot of

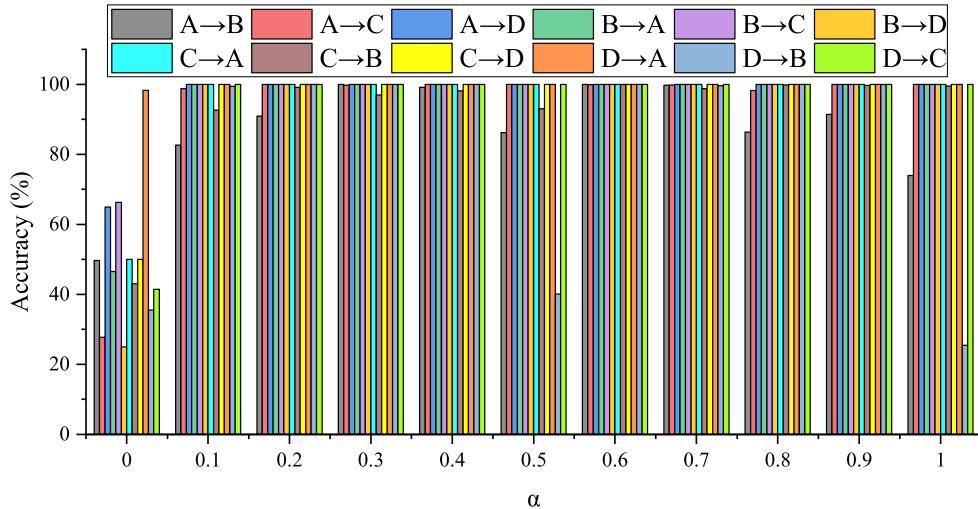


Fig. 7. Parameters sensitivity analysis for loss function weight.

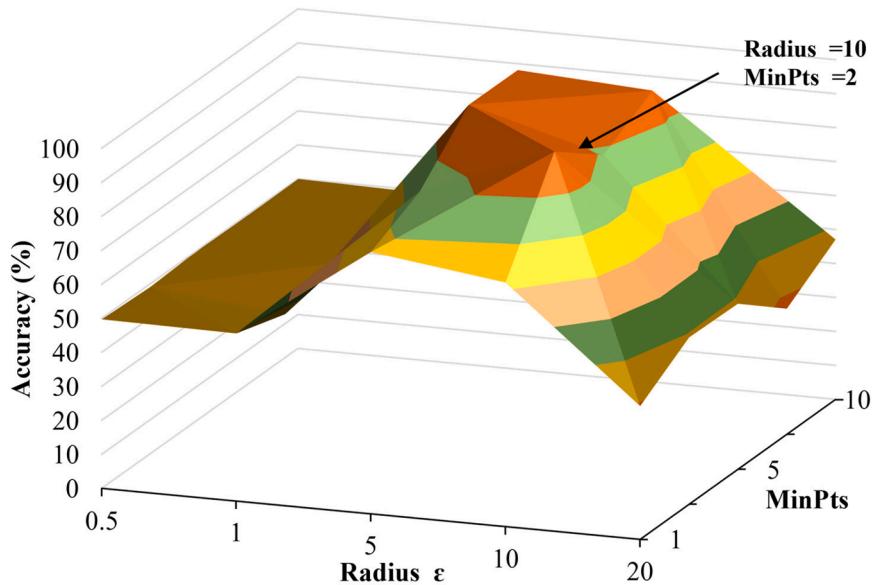


Fig. 8. Parameters sensitivity analysis for CPLL.

Fig. 10 can be obtained through calculation. For the box corresponding to each method, each black point represents the intra class spacing of different health status data. In statistics, IQR is equal to the upper quartile minus the lower quartile. A-distance [35] measures the degree of separation of different fault data. It is defined as $d_A = 2(1 - 2\epsilon)$, where, ϵ is the generalization error between two distributions. The smaller the value d_A , the greater the difference between distributions. There are four health states in this paper, so for each method, there are four results, $d_A(NC \rightarrow others)$, $d_A(OF \rightarrow others)$, $d_A(IF \rightarrow others)$, $d_A(RF \rightarrow others)$, respectively, as shown in Fig. 11. For the box corresponding to each method, the four black dots indicate $d_A(NC \rightarrow others)$, $d_A(OF \rightarrow others)$, $d_A(IF \rightarrow others)$, $d_A(RF \rightarrow others)$ separately. From the above results, we can see that DTTN-CPLL has the minimum distance between classes and within classes, which proves that it has superior effect on subdomain feature alignment.

5.4.6. Statistical significance test

By comprehensively comparing the performance of 9 methods under 12 across-machines fault diagnosis tasks, we use the critical difference (CD) algorithm to analyze, and the results are shown in Fig. 12. In the CD diagram, the smaller the value, the better the comprehensive performance. It can be seen that the proposed method DTTN-CPLL has the best effect.

Table 4

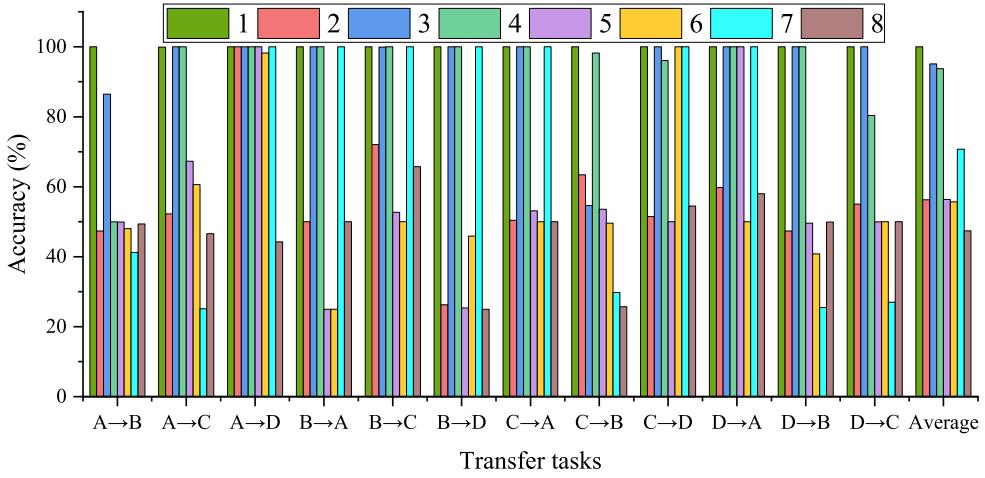
Ablation test results.

tasks	w/o L_{CPLL}	w/o L_{DC}	w/o L_{LMMD}	DTTN-CPLL
A → B	47.35 %	86.45 %	49.95 %	99.95 %
A → C	52.25 %	100.00 %	100.00 %	99.90 %
A → D	100.00 %	100.00 %	100.00 %	100.00 %
B → A	50.00 %	100.00 %	100.00 %	100.00 %
B → C	72.05 %	99.90 %	100.00 %	100.00 %
B → D	26.30 %	100.00 %	100.00 %	100.00 %
C → A	50.40 %	100.00 %	100.00 %	100.00 %
C → B	63.40 %	54.65 %	98.25 %	100.00 %
C → D	51.45 %	100.00 %	96.05 %	100.00 %
D → A	59.75 %	100.00 %	100.00 %	100.00 %
D → B	47.35 %	100.00 %	100.00 %	99.95 %
D → C	55.05 %	100.00 %	80.40 %	100.00 %
Average	56.28 %	95.08 %	93.72 %	99.98 %

Table 5

The loss function combination.

Methods	L_{CPLL}	L_{DC}	L_{LMMD}
1 (proposed)	✓	✓	✓
2	✗	✓	✓
3	✓	✗	✓
4	✓	✓	✗
5	✗	✗	✓
6	✗	✓	✗
7	✓	✗	✗
8	✗	✗	✗

**Fig. 9.** The diagnostic results under different loss function combinations.

5.4.7. Calculation cost analysis

To compare the computational costs of the proposed methods, the runtime for each fault diagnosis method on 12 tasks is summarized in [Table 6](#). It can be observed that JDA has the shortest computation time, while DCTLN + WSL has the longest runtime. This is a result influenced by both the structure of the transfer model itself and the size of the input data. In word, the computation time of the proposed DTTN-CPLL is not the shortest, but it exhibits the best diagnostic performance.

5.4.8. Statistical analysis based on box plots

The use of statistical tests is crucial to confirm whether the observed results extend beyond the range of random chance variations. We employed commonly used box plots in statistics to observe the distribution of data from the experimental results table ([Table 3](#)). The results are depicted in [Fig. 13](#). Each box represents tasks where the diagnostic accuracy rankings of the method fall between the 25th and 75th percentiles, reflecting the overall diagnostic performance of the analyzed method. The interquartile range (IQR) is equal to the upper quartile minus the lower quartile. Black dots indicate outliers deviating from the 1.5IQR range, reflecting the stability of

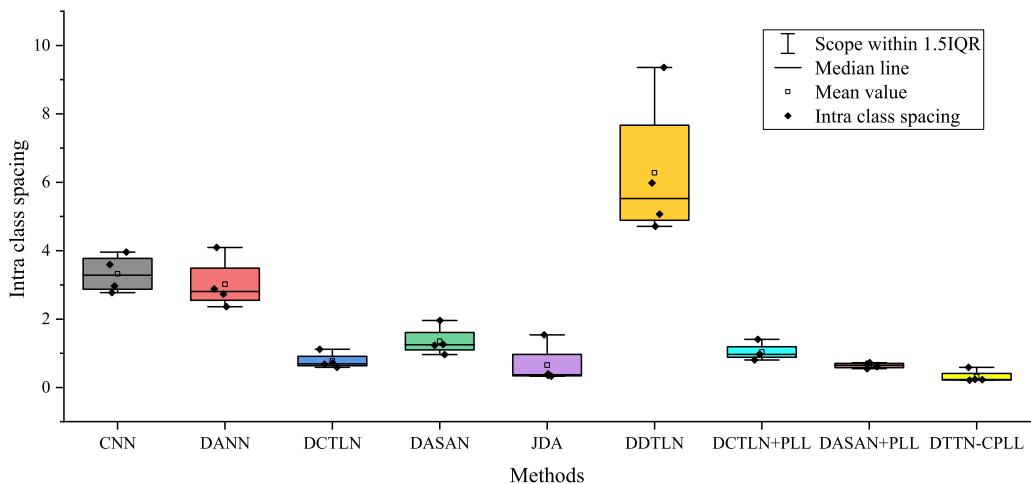


Fig. 10. Performance of DTTN-CPLL: Intra class spacing.

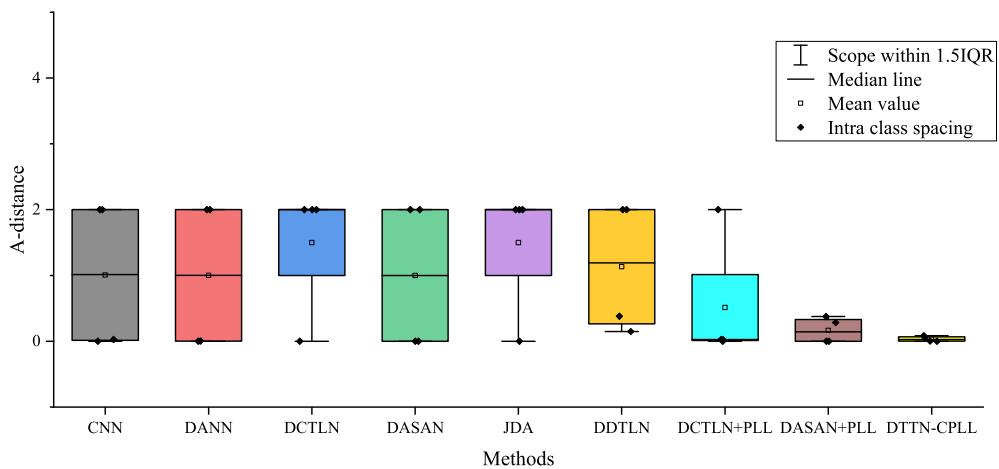


Fig. 11. Performance of DTTN-CPLL: A-distance.

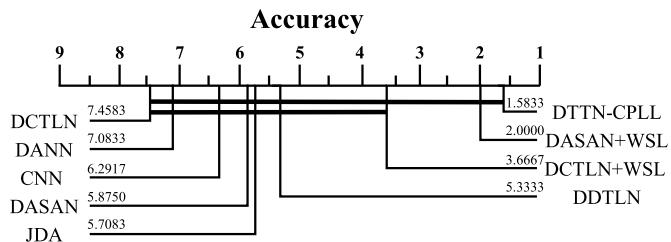


Fig. 12. CD diagram based on accuracy.

Table 6
The time consumption of different models.

Methods	Time (s)	Methods	Time (s)
CNN	4520.04	DDTLN	9505.64
DANN	6054.61	DCTLN + WSL	22220.29
DCTLN	21119.97	DASAN + WSL	22131.28
DASAN	18922.15	DTTN-CPLL	9185.80
JDA	1090.69	-	-

the method in 12 fault diagnosis tasks. From Fig. 13, it can be observed that the diagnostic accuracy and stability of the proposed DTTN-CPLL are the best.

5.4.9. Advantages and disadvantages

12 cross-machine transfer experiments have demonstrated the effectiveness of DTTN-CPLL. Its advantages can be summarized as follows:

- 1) Based on the clustering concept, CPLL fits the distribution of target domain data very well, greatly enhancing the generalization ability in cross-machine tasks and expanding the model's applicability.
- 2) DTTN-CPLL enhances the cross-domain transfer learning ability at the subdomain feature level, allowing it to fully learn and utilize source domain knowledge.

However, the proposed DTTN-CPLL also has drawbacks. Firstly, CPLL embeds clustering and iterative operations internally, resulting in significant time consumption for DTTN-CPLL. Secondly, if there are no labeled samples in the target domain, the model performs poorly. Finally, the proposed method is unable to identify fault types when the label types in the source and target domains do not match. In the future, we will explore methods based on the label assumption to replace single-label semi-supervised learning. For instance, the fault label template idea can replace a single fault label.

6. Conclusion

In this paper, a new model called DTTN-CPLL is proposed for mechanical fault diagnosis of across machines. DTTN-CPLL is composed of a feature extractor, a label classifier, and a loss function based on CPLL. It extracts common features from the source domain and target domain through the feature extractor, and then identifies fault states using the label classifier. The main conclusions are as follows.

- 1) The CPLL algorithm, driven by the single fault label in the target domain, successfully demonstrates excellent dynamic characteristics. Its unique label diffusion mechanism enables the model to accurately adapt to changes in the target domain data distribution, providing strong support for the dynamic adaptability in practical mechanical fault diagnosis. This feature allows DTTN-CPLL to exhibit more precise and flexible characteristics when dealing with different machine fault features.
- 2) Through the utilization of intra-class and inter-class distance metrics, we quantitatively verified the significant advantages of DTTN-CPLL in terms of subdomain feature alignment. By quantifying distance metrics, the model can more accurately adjust the feature differences between the source domain and the target domain, thereby enhancing the ability to capture fault features in the target domain. This provides a reliable theoretical foundation for the accuracy and robustness of the model.
- 3) After conducting cross-validation on 12 transfer tasks across four different fault datasets, the experimental results clearly demonstrate the superiority of the DTTN-CPLL method compared to the current state-of-the-art methods. This indicates the generality and robustness across different machines, providing strong support for its widespread application in practical engineering scenarios.

Although the proposed method has achieved significant performance improvement, we also fully recognize the substantial computational time consumption of DTTN-CPLL. Therefore, in future research, we will focus on addressing the computational efficiency issues during the model iteration process of CPLL. A feasible exploration direction is to introduce Monte Carlo algorithms to optimize the computation process of CPLL, aiming to reduce computational costs and enhance the real-time performance and operability of the model. This effort will provide more concrete and practical solutions for the engineering application of DTTN-CPLL, making it more suitable for practical production environments in mechanical fault diagnosis.

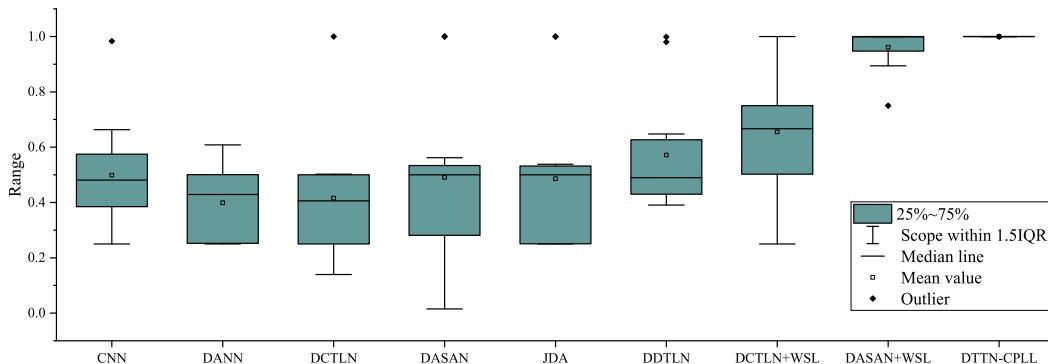


Fig. 13. Box plots corresponding to Table 3.

CRediT authorship contribution statement

Feiyu Lu: Writing – original draft, Software, Methodology, Data curation. **Qingbin Tong:** Investigation, Conceptualization. **Xuedong Jiang:** Writing – review & editing. **Ziwei Feng:** Visualization, Supervision. **Jianjun Xu:** Formal analysis. **Xin Wang:** Software. **Jingyi Huo:** Validation, Resources.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (2023JBZY039) and the Beijing Natural Science Foundation (Grant no. L211010, 3212032). The authors wish to extend their sincere thanks for the support from the Beijing Municipal Science & Technology Commission of China.

References

- [1] F. Lu, Q. Tong, Z. Feng, Q. Wan, Unbalanced bearing fault diagnosis under various speeds based on spectrum alignment and deep transfer convolution neural network, *IEEE Trans. Ind. Inf.* 19 (2023) 8295–8306.
- [2] D. Liu, L. Cui, W. Cheng, Flexible generalized demodulation for intelligent bearing fault diagnosis under nonstationary conditions, *IEEE Trans. Ind. Inf.* 1–12 (2022).
- [3] H. Su, X. Yang, L. Xiang, A. Hu, Y. Xu, A novel method based on deep transfer unsupervised learning network for bearing fault diagnosis under variable working condition of unequal quantity, *Knowl.-Based Syst.* 242 (2022).
- [4] M.A. Rahman, H. Taheri, F. Dababneh, S.S. Kargarnoudi, S. Arhamnamazi, A review of distributed acoustic sensing applications for railroad condition monitoring, *Mech. Syst. Sig. Process.* 208 (2024) 110983.
- [5] S. Kiranyaz, O.C. Devecioglu, A. Alhams, S. Sassi, T. Ince, O. Abdeljaber, O. Avci, M. Gabbouj, Zero-shot motor health monitoring by blind domain transition, *Mech. Syst. Sig. Process.* 210 (2024) 111147.
- [6] M. Sun, H. Wang, P. Liu, S. Huang, P. Wang, J. Meng, Stack autoencoder transfer learning algorithm for bearing fault diagnosis based on class separation and domain fusion, *IEEE Trans. Ind. Electron.* 69 (2022) 3047–3058.
- [7] F. Jia, Y. Lei, N. Lu, S. Xing, Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization, *Mech. Syst. Sig. Process.* 110 (2018) 349–367.
- [8] C. He, H. Shi, J. Si, J. Li, Physics-informed interpretable wavelet weight initialization and balanced dynamic adaptive threshold for intelligent fault diagnosis of rolling bearings, *J. Manuf. Syst.* 70 (2023) 579–592.
- [9] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: A survey, *Mech. Syst. Sig. Process.* 151 (2021) 107398.
- [10] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, *Mech. Syst. Sig. Process.* 138 (2020) 106587.
- [11] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, *IEEE Trans. Ind. Inf.* 15 (2019) 2446–2455.
- [12] Y. Lou, A. Kumar, J. Xiang, Machinery fault diagnostic method based on numerical simulation driving partial transfer learning, *Sci. China Technol. Sci.* 66 (2023) 3462–3474.
- [13] B. Yang, C.-G. Lee, Y. Lei, N. Li, N. Lu, Deep partial transfer learning network: A method to selectively transfer diagnostic knowledge across related machines, *Mech. Syst. Sig. Process.* 156 (2021) 107618.
- [14] Z. Zhang, H. Chen, S. Li, Z. An, Sparse filtering based domain adaptation for mechanical fault diagnosis, *Neurocomputing* 393 (2020) 101–111.
- [15] Q. Qian, Y. Wang, T. Zhang, Y. Qin, Maximum mean square discrepancy: A new discrepancy representation metric for mechanical fault transfer diagnosis, *Knowl.-Based Syst.* (2023) 110748.
- [16] X.D. Li, Y. Hu, J.H. Zheng, M.T. Li, W.Z. Ma, Central moment discrepancy based domain adaptation for intelligent bearing fault diagnosis, *Neurocomputing* 429 (2021) 12–24.
- [17] K. Zhao, H. Jiang, K. Wang, Z. Pei, Joint distribution adaptation network with adversarial learning for rolling bearing fault diagnosis, *Knowl.-Based Syst.* 222 (2021) 106974.
- [18] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, Q. He, Deep subdomain adaptation network for image classification, *IEEE Trans. Neural Networks Learn. Syst.* 32 (2021) 1713–1722.
- [19] Y. Wang, X. Sun, J. Li, Y. Yang, Intelligent fault diagnosis with deep Adversarial domain adaptation, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–9.
- [20] M. Ghorvei, M. Kavianpour, M.T.H. Beheshti, A. Ramezani, An unsupervised bearing fault diagnosis based on deep subdomain adaptation under noise and variable load condition, *Measure. Sci. Technol.* 33 (2021) 025901.
- [21] M. Kavianpour, A. Ramezani, M.T.H. Beheshti, A class alignment method based on graph convolution neural network for bearing fault diagnosis in presence of missing data and changing working conditions, *Measurement* 199 (2022) 111536.
- [22] D. Miki, K. Demachi, Bearing fault diagnosis using weakly supervised long short-term memory, *J. Nucl. Sci. Technol.* 57 (2020) 1091–1100.
- [23] M. Alabsi, L. Pearlstein, M. Franco-Garcia, Cross domain fault diagnosis based on generative adversarial networks, *J. Vibr. Control*, 0 10775463231191679.
- [24] B. Yang, Y. Lei, F. Jia, S. Xing, An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings, *Mech. Syst. Sig. Process.* 122 (2019) 692–706.
- [25] L. Guo, Y. Lei, S. Xing, T. Yan, N. Li, Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data, *IEEE Trans. Ind. Electron.* 66 (2019) 7316–7325.
- [26] Y. Liu, Y. Wang, T.W.S. Chow, B. Li, Deep adversarial subdomain adaptation network for intelligent fault diagnosis, *IEEE Trans. Ind. Inf.* 18 (2022) 6038–6046.
- [27] X. Li, W. Zhang, Q. Ding, X. Li, Diagnosing rotating machines with weakly supervised data using deep transfer learning, *IEEE Trans. Ind. Inf.* 16 (2020) 1688–1697.

- [28] B. Yang, Y. Lei, S. Xu, C.-G. Lee, An optimal transport-embedded similarity measure for diagnostic knowledge transferability analytics across machines, *IEEE Trans. Ind. Electron.* 69 (2022) 7372–7382.
- [29] T. Han, T. Zhou, Y. Xiang, D. Jiang, Cross-machine intelligent fault diagnosis of gearbox based on deep learning and parameter transfer cross-machine intelligent fault diagnosis of gearbox based on deep learning and parameter transfer, *Struct. Contr. Health Monit.* 29 (2022) e2898.
- [30] B. Yang, Y. Lei, X. Li, C. Roberts, Deep targeted transfer learning along designable adaptation trajectory for fault diagnosis across different machines, *IEEE Trans. Ind. Electron.* 1–11 (2022).
- [31] C.E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* 27 (1948) 379–423.
- [32] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, Q. Tian, Towards discriminability and diversity: batch nuclear-norm maximization under label insufficient situations. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [33] M. Fazel, T.K. Pong, D. Sun, P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, *SIAM J. Matrix Anal. Appl.* 34 (2013) 946–977.
- [34] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Rev.* 52 (2010) 471–501.
- [35] V.D.M. Laurens, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (2008) 2579–2605.
- [36] Case Western Reserve University Bearing Data Center Website [Online] Available: <http://csegroups.case.edu/bearingdatacenter/home>.
- [37] Intelligent Maintenance System Bearing Dataset [Online] Available: <https://www.nasa.gov/intelligent-systems-division/>.
- [38] Datacastle [Online] Available: https://www.datacastle.cn/dataset_description.html?type=dataset&id=539.
- [39] H. Xue, M. Wu, Z. Zhang, H. Wang, Intelligent diagnosis of mechanical faults of in-wheel motor based on improved artificial hydrocarbon networks, *ISA Trans.* 120 (2022) 360–371.
- [40] T. Han, C. Liu, W. Yang, D. Jiang, *ISA Trans.* 97 (2020) 269–281.
- [41] Q. Qian, Y. Qin, J. Luo, Y. Wang, F. Wu, Deep discriminative transfer learning network for cross-machine fault diagnosis, *Mech. Syst. Sig. Process.* 186 (2023) 109884.