



ARTICLE

Dual-Classifier Label Correction Network for Carotid Plaque Classification on Multi-Center Ultrasound Images

Louyi Jiang^{1,#}, Sulei Wang^{1,#}, Jiang Xie¹, Haiya Wang² and Wei Shao^{3,*}

¹School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China

²Gerontology Department, Shanghai Jiaotong University School of Medicine Affiliated Ninth Hospital, Shanghai, 200025, China

³Scientific Research Management Department, Shanghai University, Shanghai, 200444, China

*Corresponding Author: Wei Shao. Email: shaow@shu.edu.cn

#Equal contribution

Received: 02 December 2024; Accepted: 01 April 2025; Published: 19 May 2025

ABSTRACT: Carotid artery plaques represent a major contributor to the morbidity and mortality associated with cerebrovascular disease, and their clinical significance is largely determined by the risk linked to plaque vulnerability. Therefore, classifying plaque risk constitutes one of the most critical tasks in the clinical management of this condition. While classification models derived from individual medical centers have been extensively investigated, these single-center models often fail to generalize well to multi-center data due to variations in ultrasound images caused by differences in physician expertise and equipment. To address this limitation, a Dual-Classifier Label Correction Network model (DCLCN) is proposed for the classification of carotid plaque ultrasound images across multiple medical centers. The DCLCN designs a multi-center domain adaptation module that leverages a dual-classifier strategy to extract knowledge from both source and target centers, thereby reducing feature discrepancies through a domain adaptation layer. Additionally, to mitigate the impact of image noise, a label modeling and correction module is introduced to generate pseudo-labels for the target centers and iteratively refine them using an end-to-end correction mechanism. Experiments on the carotid plaque dataset collected from three medical centers demonstrate that the DCLCN achieves commendable performance and robustness.

KEYWORDS: Deep learning; medical image processing; carotid plaque classification; multi-center data

1 Introduction

Cardiovascular disease (CVD), a leading cause of morbidity and mortality globally, is primarily caused by the continuous development of carotid atherosclerosis [1]. One of the main causes of CVD is the continuous development of carotid atherosclerosis [2]. Thus, the concept of vulnerable plaques was introduced, and researchers have focused on CVD prevention by identifying atherosclerotic plaques with increased risk [3]. Clinically, due to the shallow position of the carotid artery, B-mode ultrasound, a noninvasive detection method, is one of the main methods for the preoperative diagnosis of carotid plaque in clinical practice [4]. However, its imaging efficacy is strongly affected by the operator's technique and instrument. The use of different operators and different instruments can impact the final imaging effect. Therefore, it is necessary to develop a computer-aided diagnosis (CAD) system based on multicenter B-mode ultrasound plaque data.

Deep learning, as an efficient auxiliary diagnostic method, has been gradually applied in carotid atherosclerosis-related fields and has achieved good results, especially for plaque data from a single center [5].



However, deep learning methods need a large amount of labeled data. Unfortunately, the collection of medical data itself is different from that of natural images, requiring much time and consultation with hospitals and adhering to pertinent ethical review procedures. Furthermore, the need for data annotation by domain-specialized medical professionals with extensive experience is highly important. These problems limit the development of multicenter plaque risk CAD to some extent. Therefore, using unsupervised learning has been proposed to solve this kind of problem. The aim of unsupervised learning is to train a domain-invariant classifier by narrowing the data features of two different domains, which also means that unsupervised learning does not require target center data labels for classification [6]. Recent advances in domain adaptation have demonstrated substantial progress in natural image classification tasks [7,8]. However, due to the inherent challenges of medical images, domain adaptation has garnered significant attention in the field of medical imaging. Common issues include unsupervised learning constraints, difficulties in adapting to diverse patient populations, and challenges in handling complex medical datasets [9]. To achieve this goal in multicenter CAD for carotid plaque risk prediction, there are some difficult problems that still need to be solved.

Establishing an appropriate deep learning classification model for multicenter carotid artery plaque ultrasound images still encounters two major challenges: 1) Differences in operators and equipment contribute to variations in ultrasound imaging features, implying substantial disparities in plaque ultrasound data acquired from different centers, resulting in divergent diagnostic outcomes. Consequently, these differences present challenges in unifying unique features within datasets from various centers. 2) Inherent deficiencies in ultrasound imaging, such as low resolution and artifacts, pose significant challenges [10]. Additionally, during examinations, factors like patient swallowing and the operator's proficiency impact the final image quality of ultrasound [11]. The resulting noise makes it difficult to identify and classify certain data features, leading to the misallocation of erroneous labels by the classifier. Such labels hinder the effective transfer of features across multiple centers. These two difficulties are shown in more detail in Fig. 1. Therefore, it is imperative to address the aforementioned challenges to facilitate seamless integration and accurate analysis within such systems.

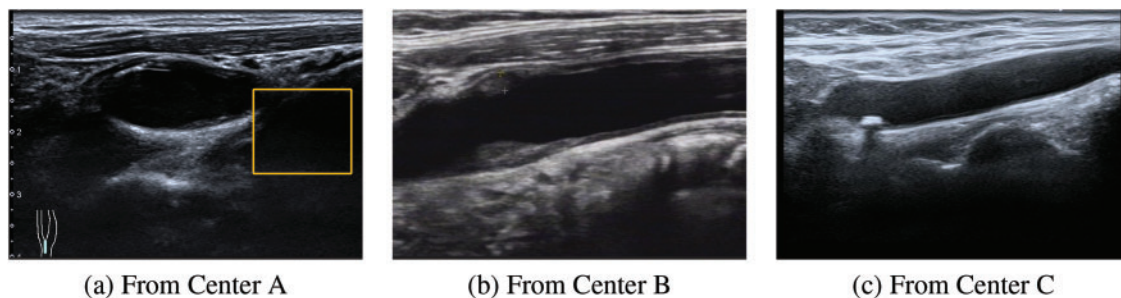


Figure 1: An example of the differences in the ultrasound images of plaques from multiple centers. These three images are from three different centers. (1) Plaque ultrasound image noise. (a) The yellow box indicates an ultrasound artifact, which covers part of the useful information and produces noise interference. (b) It is evident that the resolution of this ultrasound image is very low. (2) Feature inconsistency. It can be observed that in (a), (b) and (c) there are obvious differences among the three images, causing certain difficulties for follow-up research

To address the above difficulties, we propose a multicenter plaque risk prediction algorithm based on dual classifier noise label correction. The overarching concept is visually represented in Fig. 2. This method introduces the ideas of adversarial and label correction. On the basis of narrowing the characteristics of multicenter data, the collective knowledge of two classifiers is used to continuously correct labels to transfer

information between different centers. In more detail, we construct a feature-sharing multilayer domain adaptation layer to narrow the feature differences in different central datasets and use two independent classifiers, which can cooperate to obtain pseudo labels for the data from the target center. Based on the pseudolabels, iterative label updates are achieved by means of adversarial generation, leveraging feature learning and loss function computations. Our main contributions can be summarized as follows:

- 1 We propose a domain adaptation framework specifically designed for multicenter ultrasound plaque image classification, namely, the double classifier label correction network (DCLCN), which aims to address noise and feature transfer issues more effectively.
- 2 The DCLCN algorithm includes the multicenter feature adaptation (MFA) module, which identifies latent shared features between the source and target centers, facilitating the discovery of transferable knowledge conducive to classification.
- 3 Building upon the MFA module, the DCLCN implements the modeling and correction of noise labels (MCNL) module. This module performs deep correction by modeling the labels of the target center based on the MFA module, addressing the predictions influenced by noise in the model, and maintaining and updating the end-to-end label distributions from two center datasets.

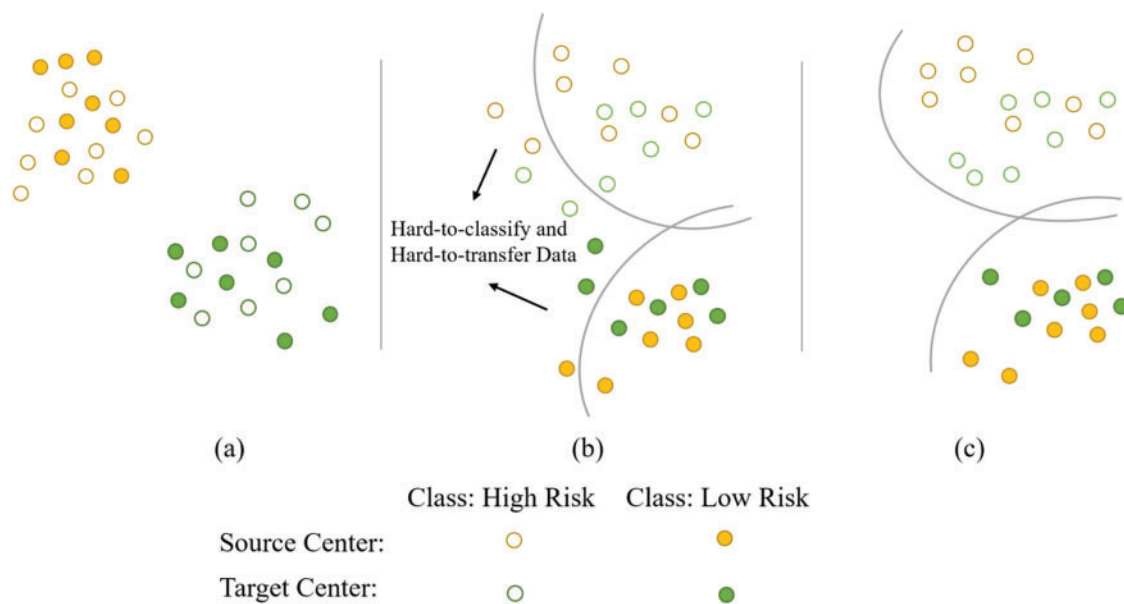


Figure 2: Schematic of feature distribution and our main ideas. (a) The initial distribution of plaque ultrasound data features, which is highly disordered and cannot be classified by the classifier. (b) After narrowing the feature distance, the majority of the data can now be classified by the classifier. The classifier has difficulty in the classification and identification of some data due to noise in certain ultrasound images, hindering feature transfer across multiple centers. (c) By further modeling and correcting noise labels, this image represents the final goal that the DCLCN aims to achieve

2 Related Work

The role of deep learning in medical image analysis is becoming increasingly evident, particularly in single-center plaque analysis. Single-center plaque classification research has made significant strides, offering valuable diagnostic tools [12,13]. A deep learning-based framework for mammogram classification is proposed to enhance the accuracy of breast cancer detection [14]. An explainable deep learning architecture,

DeepOCT, analyses macular edema in OCT images, enhancing both diagnostic accuracy and interpretability [15]. However, real-world applications span multiple hospitals, necessitating multicenter carotid plaque classification. While multicenter plaque studies are limited, progress in other fields is noteworthy.

2.1 Unsupervised Domain Adaptation

Unsupervised domain adaptation (UDA) addresses domain differences between source and target domains for enhanced image processing task performance. Collaborative unsupervised domain adaptation (CoUDA) [16] combats dataset noise but overlooks center-specific feature differences. Consequently, existing methods for multicenter ultrasound plaque risk prediction are insufficient. Relevant research integrates multi-modal information fusion and domain adaptation techniques to enhance the generalization ability across multi-center data [17]. However, they are often limited to specific modalities like CT, require high-quality data, and are computationally demanding [18]. Some studies optimize cross-domain feature alignment and semantic transfer through multi-scale adversarial learning, a mix-up discriminator, and a self-attention mechanism, demonstrating superior performance in ultrasound image tasks [19]. These studies highlight the need for further improvements in data quality, modality-specific strategies, and computational efficiency.

2.2 Pseudolabel-Based Methods

Pseudolabels enhance model performance in the target domains by aiding in feature and pattern recognition. Several studies have explored the applications in UDA. The moving semantic transfer network (MSTN) [20] investigates semantic representation learning with pseudolabels in UDA. Box-Adapt [21] which uses bounding boxes for partial labeling, was also proposed. Despite the success of these approaches, challenges remain in improving pseudolabel precision and addressing noisy labels. To address these issues, Shi et al. [22] proposed a UDA framework based on label distribution learning, which corrects noisy pseudolabels iteratively to retrain the model.

3 Materials and Methods

3.1 Datasets

We evaluate the proposed method with three ultrasound datasets collected from three collaborating centers, A, B, and C. All three datasets were annotated by radiologists with years of clinical experience. In this study, stable plaques are considered low risk, while unstable plaques are deemed high risk due to their risk of detachment [23]. All the datasets are summarized in Table 1.

Table 1: Dataset details

	Stable	Unstable	Total
A	506	253	759
B	132	178	310
C	248	386	634

3.2 Methods

In this study, the domain adaptation challenge within the context of B-mode ultrasound multicenter plaques is addressed. From the source center, we obtain the image x_s and its corresponding label y_s , resulting in a dataset denoted as $\{X_s, Y_s\}$ from the source center. Similarly, the unlabeled images x_t from the target

center are in a dataset denoted as $\{X_t\}$. The DCLCN first adopts a shared deep neural network with feature generator G, which is capable of extracting high-level feature representations from ultrasound images. The ultrasound image data of the source and target centers are input into G to obtain the corresponding feature representation. After passing through G, the DCLCN will undergo a two-stage adjustment.

As shown in Fig. 3, the first stage involves an MFA module to address the issue of multicenter feature inconsistency. This module includes a domain feature adaptation layer and two independent classifiers (C_1 and C_2). The main goal of this stage is to narrow the adjustment decision boundary through the feature adaptation layer, further expanding the classifier differences while reducing the data differences between different centers.

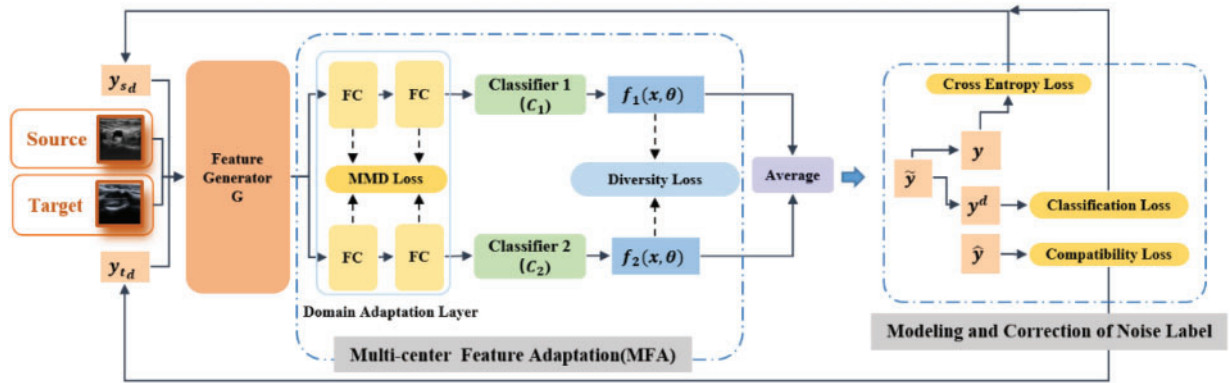


Figure 3: The overall framework of the DCLCN, which consists of two main parts: the multicenter feature adaptation (MFA) and modeling and correction of noise label (MCNL) modules. Images from different centers are input to the model as sources and targets. Following the two parts, the label distribution of the images is continuously updated and refined in the subsequent training iterations and utilized as the input in the next iteration

In the second stage, the MCNL module is introduced. Here, the model simultaneously provides label distribution sets y_s^d and y_t^d for the source and target centers, respectively, where y_t^d from the target center is initialized by the pretrained model containing “noise”. In the second stage, the purpose of the DCLCN is to continuously update y^d through backpropagation during the training process to process the “noise”.

3.2.1 Multicenter Feature Adaptation

To solve the multicenter data alignment problem in the first stage, the DCLCN introduces the MFA module based on the dual classifier. We designed a domain adaptation layer after the feature generator G. For model implementation, this component comprises two fully connected layers. The primary objective of the MTA module is to enhance feature alignment. To effectively quantify feature disparities, the DCLCN uses the Maximum Mean Difference (MMD) [24] as the chosen metric. The MMD projects features into a high-dimensional space using the kernel function and computes the expected difference between the random variables of the two feature distributions after mapping. The maximum value of this calculation serves as the standard for measuring the distance. By minimizing the MMD loss, the DCLCN can identify invariant features among multiple centers.

$$L_{MMD}^2(\theta_F, X_s, X_t) = \left\| \frac{1}{n} \sum_{i=1}^n \phi(x_i^s) - \frac{1}{n} \sum_{j=1}^m \phi(x_j^t) \right\|_{\mathcal{H}}^2 \quad (1)$$

To further measure the difference between the two classifiers, the DCLCN introduces diversity loss, which uses the difference in absolute values of the output of the classifier for the target center data to perform quantitative measurements. The specific definition is shown in Eq. (2).

$$L_{diversity}(\theta_F, \theta_G, \theta_C) = \frac{1}{K} \sum_{k=1}^K |f_1(x, \theta_F, \theta_G)_k - f_2(x, \theta_F, \theta_G)_k| \quad (2)$$

where f_1 and f_2 represent the outputs of the two classifiers and k represents the corresponding plaque category, which is a stable plaque or vulnerable plaque in this study. We attempted to use the classifier determinacy disparity (CDD) [25] distance rather than the absolute values, and the impact was not prominently evident.

By maximizing $L_{diversity}$, the differences are maximized by the two classifiers, enhancing the model's discriminative ability for the target center. The goals of this stage is shown in Eq. (3).

$$\min(\gamma L_{MMD} - L_{diversity}) \quad (3)$$

Through multicenter domain adaptation, the model can obtain prediction results $f_1(x, \theta)$ and $f_2(x, \theta)$ from the two classifiers. Data that are difficult to correctly classify as well as those for which it is difficult to transfer knowledge are also difficult to classify using the two classifiers. Therefore, averaging the output results $f_1(x, \theta)$ and $f_2(x, \theta)$ can further reduce the interference of ultrasonic noise on the results. In addition, the model can obtain a prediction distribution that is convenient for label modeling, the definition is shown in Eq. (4).

$$\tilde{y}(x, \theta) = \frac{f_1(x, \theta) + f_2(x, \theta)}{2} \quad (4)$$

3.2.2 Modeling and Correction of the Noise Label

For plaque images, simply reducing the distance does not significantly enhance the accuracy because the inherent noise can make data transfer between centers challenging. Hence, the DCLCN initially models the labels and subsequently refines them to reduce the impact of noise; thus, the MCNL is proposed. The MCNL contains two parts. In the following sections, we will elucidate these two components separately.

Label modeling: To model the labels, the DCLCN uses two sets of label distributions y_s^d and y_t^d to describe the data from the two centers. In the specific model learning process, during each training iteration, y_s^d and y_t^d are input into the model as pseudoground truth labels (especially for the target center).

For label initialization, the source center uses its ground truth label y_s . For the target center data, since there is no initial ground truth label for initialization, the DCLCN utilizes the source center's pretrained model to provide an initial estimation of the distribution. This initial distribution estimate for pseudolabels at the target center includes factors such as image noise and the presence of noisy labels.

Correction of noise labels: y_s^d and y_t^d are two sets of label distributions that include the influence of noise. The DCLCN progressively mitigates the effect of this aspect during ongoing learning. To achieve this goal, the parameters θ of the network are continuously updated. The label distributions y_s^d and y_t^d for the source and target centers are also added to the training process and are continuously updated during the training process. Furthermore, there are four types of labels, namely, $\tilde{y}$, $\hat{y}$, $\check{y}$, and y . Among them, $\tilde{y}$ is the prediction result from the previous part of the classifier, which is the probability distribution of a category; y is the real label of the dataset (note that in this article, the data label of the target center does not participate in the training); $\hat{y}$ is the pseudolabels assigned to the target center by the pretrained classifier; and $\check{y}$ is a set

of hard labels that are continuously updated by backpropagation. For the initialization of $\check{y}$, the source center uses y , which is the ground truth label. The target center uses the initial noise pseudolabel $\hat{y}$:

$$\check{y}_s = My \quad (5)$$

$$\check{y}_d = M\hat{y} \quad (6)$$

where M is a large constant. The softmax function's rule inherently diminishes the distinction between results approaching values of 0 and 1. Therefore, M is used to make the hard label $\check{y}$ more discriminative after being activated by the softmax function later.

Thus, the softmax function is applied to $\check{y}$ to obtain a set of continuously updated probability distributions y^d needed by the model.

$$y^d = \text{softmax}(\check{y}) \quad (7)$$

Cross-entropy loss: The data from the source center have ground truth labels. The DCLCN utilizes the cross-entropy loss to enhance the classification capability of the source center.

$$\mathcal{L}_{CE}(\theta_c, \theta_G, \theta_F) = \text{CrossEntropy}(\check{y}_s, y) \quad (8)$$

Classification loss: To reduce the impact of noise, the model needs to continuously update the label y^d and the parameter θ . The update in this section relies mainly on the guidance of y^d and the deviation of the prediction result $\tilde{y}$ of the dual classifier in the previous section. The DCLCN uses the Kullback Leibler (KL) loss to measure bias. This loss is shown in Eq. (9).

$$\mathcal{L}_C = \frac{1}{K} \sum_{k=1}^K KL(\tilde{y}(x, \theta), y^d) \quad (9)$$

Compatibility loss: Although the pseudolabels predicted by the model include hard-to-classify and hard-to-transfer plaque predictions, in each subsequent iteration of training, the model's outcomes should largely align with the initial pseudolabels. This viewpoint has also been confirmed in prior research [26]. Therefore, a compatibility loss function is added to the model. This loss is shown in Eq. (10).

$$\mathcal{L}_O = -\frac{1}{K} \sum_{k=1}^K \hat{y} \log y_d \quad (10)$$

Finally, the overall loss in the second stage is shown in Eq. (11).

$$\mathcal{L}_{label} = \mathcal{L}_{CE} + \alpha \mathcal{L}_C + \beta \mathcal{L}_O \quad (11)$$

where α and β are two hyperparameters whose values are determined based on the experimental results.

The goal of this stage is shown in Eq. (12).

$$\min \mathcal{L}_{label} \quad (12)$$

3.2.3 Training Steps

DCLCN training is divided into three steps.

Step A To obtain a preliminary model that is more suitable for multicenter plaque classification, the DCLCN uses plaque data from the source center to train the model. The training goal of this step is to minimize the cross-entropy loss, and the goal is shown in Eq. (13).

$$\min_{\theta_F, \theta_G, \theta_C} \mathcal{L}_{CE} \quad (13)$$

Step B The DCLCN adds data from the target center for training. For a clearer algorithm flow, please refer to Algorithm 1. First, the multicenter field adaptation layer is used to shorten the discrepancy between the target center and source center, and the L_{MMD} loss is calculated accordingly. Moreover, the difference between the two classifiers is enlarged by maximizing $L_{Diversity}$. Second, after these two steps, for the modeling and updating part of the label, the three losses L_{CE} , L_C and L_o are calculated to update the network parameters and pseudolabels. The specific parameters of each loss function are explained in the experimental section.

$$\min_{\theta_F, \theta_G, \theta_C} (\mathcal{L}_{CE} + \gamma \mathcal{L}_{MMD} + \alpha \mathcal{L}_C + \beta \mathcal{L}_O - \mathcal{L}_{Diversity}) \quad (14)$$

Finally, the previous feature extraction part and the multicenter domain adaptation layer are trained to minimize $L_{Diversity}$ in these two parts. α , β and γ are hyperparameters.

$$\min_{\theta_F, \theta_G} \mathcal{L}_{Diversity} \quad (15)$$

Step C Through Step B, the model obtains a set of pseudolabels from the target center. The model subsequently used this set of final fixed pseudolabels to perform the final fine-tuning of the target center.

Algorithm 1: The proposed DCLCN

Input: Source data $\{x_s, y_s\}$; Target data $\{X_t\}$ and the number of training epochs T ; Feature generator G ; Classifiers $\{C_1, C_2\}$; Multicenter adaptation layer $\{F_1, F_2\}$.

Output: The label corrected label *prediction*^d.

for $t = 1$ to T **do**

 Generate features f_s and f_t based on G , F_1 and F_2

 Obtain the prediction $f_1(x, \theta)_s$ and $f_2(x, \theta)_s$ of source data X_s based on C_1 and C_2

 Obtain the prediction $f_1(x, \theta)_t$ and $f_2(x, \theta)_t$ of source data X_t based on C_1 and C_2

 Compute $\tilde{y}_s$ and $\tilde{y}_t$ by Eq. (3)

 Compute L_{MMD} based on f_s and f_t by Eq. (1)

 Compute $L_{Diversity}$ based on f_1 and f_2 by Eq. (2)

 Compute L_{CE} of source data based on $\tilde{y}_s$ and y_s by Eq. (5)

 Compute L_C of source data and target data based on $\tilde{y}$ and y^d by Eq. (6)

 Compute L_o based on $\hat{y}$ and y^d by Eq. (7)

for $n = 1$ to N **do**

 Compute $L_{Diversity}$ based on f_1 and f_2 by Eq. (2)

end for

end for

4 Results

4.1 Performance Evaluation Measures

This study uses accuracy, regression precision and AUC to evaluate the performance of the DCLCN. These four metrics are shown in Eqs. (16)–(20).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

$$AUC = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{1}(T_i > F_j) \quad (19)$$

$$F_{1score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (20)$$

where TP , TN , FP , and FN represent the true positive, true negative, false-positive, and false-negative classifications, respectively. T_i is the predicted score of the i -th positive sample. F_j is the predicted score of the j -th negative sample. N_1 is the number of positive samples. N_2 is the number of negative samples.

Classification performance metrics with a 95% confidence interval provide robust performance measures while quantifying uncertainty, which is critical for high-stakes applications such as medical imaging. They also enable meaningful comparisons between models, ensuring that statistically significant improvements are accurately identified.

4.2 Experimental Results

4.2.1 Effect Validation on Six Classification Tasks

To prove the feasibility and advancement of the proposed DCLCN, we conduct experiments for all possible combinations of the three centers (six potential source-to-target center mappings). The results of these efforts are presented in Table 2.

Table 2: Accuracy (%), precision (%) and recall (%) of different tasks on the DCLCN

	Accuracy	Precision	Recall
A → B	73.55	78.90	72.96
B → A	73.79	79.89	73.42
A → C	71.54	78.99	72.43
B → C	72.44	71.55	72.28
C → A	70.28	72.75	69.75
C → B	69.37	70.37	71.95

Note: The bold entries represent the best-performing results among the compared methods.

The experimental results show that the accuracy, precision and recall metrics are consistently approximately 70% or greater across the six specified tasks.

4.2.2 Model Comparison

In addition, we select one of the 6 tasks ($B \rightarrow A$) and compare the proposed method with classical and current mainstream methods. The detailed results are shown in Table 3. The following is a brief overview of the comparison methods. Classification performance metrics provide essential insights into a model's accuracy and its ability to generalize to new data. The 95% confidence interval quantifies the uncertainty around these performance estimates, ensuring statistical rigor in the analysis.

Table 3: Comparison of the DCLCN with several other methods

	Accuracy	Precision	Recall	F1-score	AUC
Resnet-50 (Source only)	45.47 $\pm$ 3.73%	41.41 $\pm$ 2.73%	41.20 $\pm$ 3.73%	41.30 $\pm$ 3.11%	53.32 $\pm$ 3.73%
Resnet-50 (Fully supervised)	77.35 $\pm$ 0.79%	80.90 $\pm$ 0.62%	76.33 $\pm$ 0.58%	78.55 $\pm$ 0.58%	85.54 $\pm$ 0.39%
DANN [27]	68.32 $\pm$ 2.33%	64.57 $\pm$ 1.76%	62.30 $\pm$ 2.03%	63.42 $\pm$ 1.85%	74.34 $\pm$ 2.00%
DAN [28]	74.29 $\pm$ 1.57%	75.37 $\pm$ 1.33%	72.95 $\pm$ 1.75%	74.14 $\pm$ 1.49%	78.84 $\pm$ 1.56%
MCD [6]	71.44 $\pm$ 1.69%	68.13 $\pm$ 1.53%	73.05 $\pm$ 1.70%	70.52 $\pm$ 1.57%	77.46 $\pm$ 1.68%
DMCD [29]	73.06 $\pm$ 1.32%	72.89 $\pm$ 1.78%	72.00 $\pm$ 1.54%	72.71 $\pm$ 1.64%	76.27 $\pm$ 1.64%
(AD ² A) [30]	72.89 $\pm$ 1.53%	73.56 $\pm$ 1.63%	71.33 $\pm$ 1.60%	72.43 $\pm$ 1.60%	79.84 $\pm$ 1.60%
CoUDA [16]	73.38 $\pm$ 1.73%	78.34 $\pm$ 1.75%	72.47 $\pm$ 1.55%	75.31 $\pm$ 1.63%	80.45 $\pm$ 1.75%
TRANS-DA [31]	74.30 $\pm$ 1.02%	77.83 $\pm$ 1.03%	73.78 $\pm$ 1.10%	75.78 $\pm$ 1.05%	82.63 $\pm$ 1.08%
DCLCN (ours)	73.79 $\pm$ 0.98%	79.89 $\pm$ 1.05%	73.42 $\pm$ 1.48%	76.29 $\pm$ 1.25%	81.35 $\pm$ 1.32%

Note: All classification performance metrics with a 95% confidence interval.

These methods aim to improve cross-domain transfer learning performance. DANN reduces feature distribution discrepancies with adversarial networks; DAN [28] aligns feature distributions for better model generalization; MCD minimizes classifier output differences for improved cross-domain performance; DMCD [29] combines maximum classifier discrepancy and denoising techniques without requiring source domain data; AD²A [30] uses attention-guided domain adaptation for multisite data harmonization; CoUDA [16] introduces domain discriminators and noise adaptation layers to improve classification, focusing on labeling errors; TRANS-DA [31] uses Vision Transformers and self-supervised learning to enhance cross-domain feature mapping.

The DCLCN has demonstrated good performance across the metrics compared to the other semi-supervised approaches for addressing multicenter problems. Although the accuracy of the DCLCN is slightly lower than that of the DAN, the difference in accuracy is minimal (0.5%). However, the DCLCN achieves substantial increases in precision and recall, with a notable 4.52% (79.89% vs. 75.37%) improvement in precision. Additionally, the p -values of the model are all less than 0.05. It can be observed that DCLCN performs lower than TRANS-DA in terms of Accuracy (Acc) and AUC. DCLCN outperforms the second-best model by 2.06 in precision and 0.51 in F1-score. For other metrics, DCLCN ranks second and is very close to the best-performing model. DCLCN is in an optimal or near-optimal position, demonstrating the best performance across all the metrics.

4.2.3 Ablation Study

Since the DCLCN involves five loss functions, to prove the rationality of the loss function selected by the method, we conduct ablation experiments. The combined utilization of domain adaptation and two label correction losses significantly contributes to the ultimate classification accuracy. Hence, integrating these two stages proves to be effective in addressing multicenter plaque risk prediction. The findings of these experiments are displayed in [Table 4](#).

Table 4: Comparison to determine the optimal layer count

Multicenter Feature Adaptation (MFA)		Label Modeling and Correction (MCNL)		Accuracy
L_{MMD}	$L_{Diversity}$	L_O	L_C	
	✓			67.58
✓				69.23
		✓		71.89
✓	✓		✓	70.55
	✓	✓	✓	72.03
✓	✓	✓	✓	73.79

4.2.4 Domain Adaptation Layer Determination

Due to the varying effects of multicenter domain adaptation layers with different numbers on model performance, a series of systematic experiments were conducted to determine the optimal number of layers. The results of these experiments are shown in [Table 5](#). The data from these tests indicate that the inclusion of multicenter domain adaptation layers improves domain adaptation performance.

Table 5: Comparison to determine the optimal layer count

Layers	Accuracy	Precision	Recall
0	73.03	78.80	72.80
1	74.20	73.91	65.55
2	73.79	79.89	73.42
3	71.20	79.54	71.24

4.2.5 Pseudo-Label Prediction Accuracy Trend Analyze

In this section, we collected the ratio of correct predictions for pseudo-labels to the total number of pseudo-labels generated during the training process on the B–A dataset and plotted the trend (as shown in [Fig. 4](#)). Whether for stable or unstable plaques, the overall trend of pseudo-label accuracy is upward with fluctuations, particularly showing a stable upward trend in the later epochs. This suggests that the pseudo-label generation has strong adaptability and stability. The model's accuracy curve does align with the quality of pseudo-labels in [Fig. 4](#), further demonstrating that changes in pseudo-labels influence the results. This result validates the effectiveness of the unsupervised domain adaptation method in handling multi-center data, ensuring the model's predictive stability.

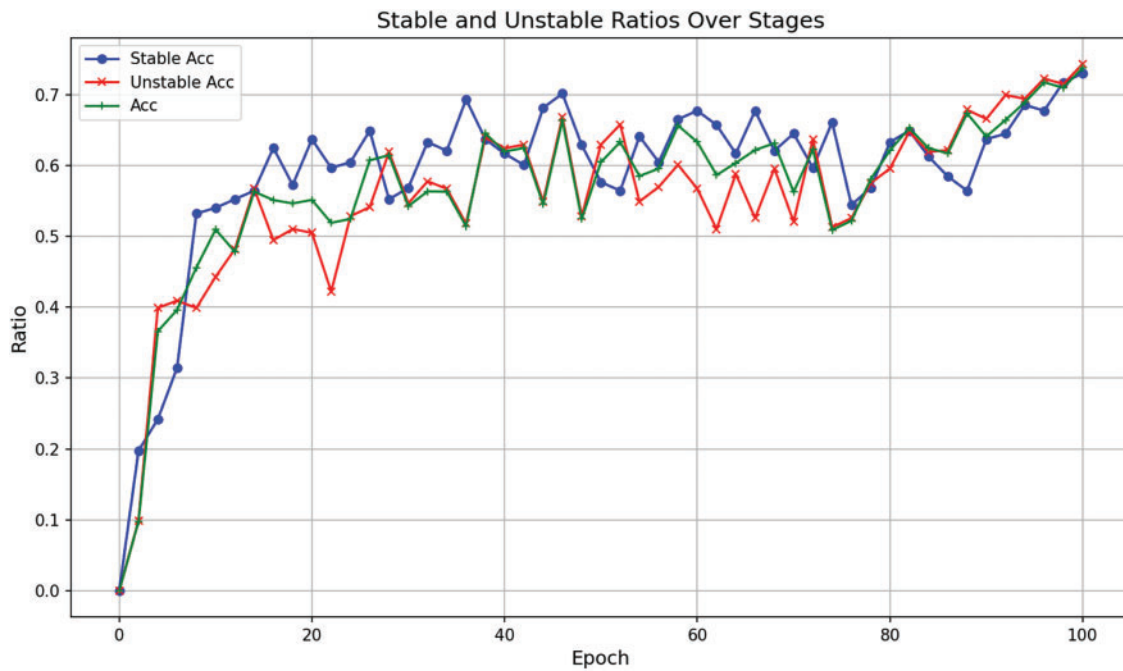


Figure 4: Trend of correct predictions for pseudo-labels in the B-A dataset during training

4.2.6 Visualization Study

To validate the classification performance of DCLCN and enhance model interpretability, this section selects the B-A center pairing and employs Grad-CAM [32] to visualize the feature representations of the target center A. Additionally, a comparative visualization analysis is conducted on two baseline models, DANN [27] and MCD [6], as shown in Fig. 5. It can be observed that, compared to the ground truth annotations by medical experts, the DANN model primarily focuses on information outside the blood vessel, failing to adequately attend to the core vascular region. In contrast, while the MCD model demonstrates some degree of focus on the vascular area, its attention remains partially misaligned. DCLCN, however, shows a more accurate and focused attention on the core vascular region, closely matching the ground truth.

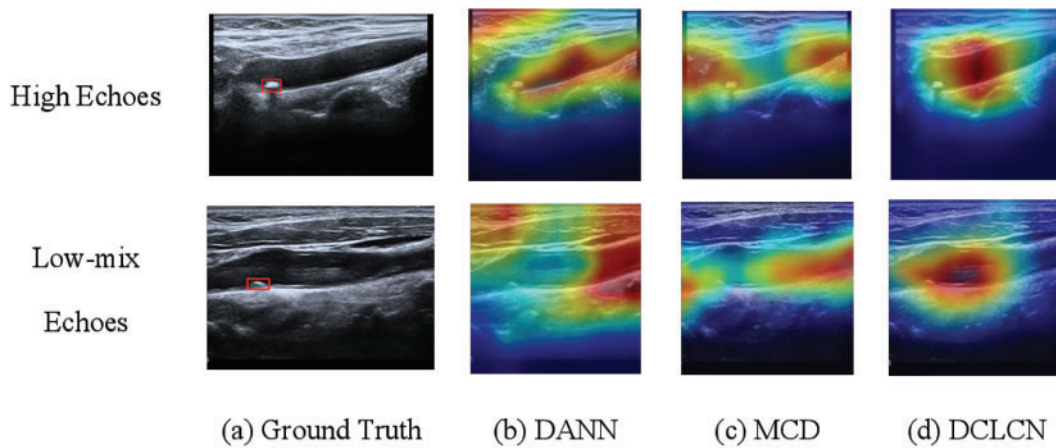


Figure 5: Feature visualization and comparative analysis of DCLCN, DANN, and MCD on target center A

5 Discussion

The experiments in this study are based on ResNet-50 and pretraining parameters from ImageNet [33]. The number of epochs is set to 150, of which the first 20 epochs are related to source center pretraining, the next 100 epochs are the dual classifier pseudolabel domain adaptation stage, and the remaining epochs are used to fine-tune the model. To determine the optimal values for the parameters of the loss function, we evaluated various combinations of α , β , and γ , and found that the best results were obtained with $\alpha = 0.3$, $\beta = 0.3$, and $\gamma = 0.4$ (Table 6). The parameters α and β share the same value, indicating that the classification loss and compatibility loss contribute equally. The parameter γ corresponds to the MMD layer in MFA, which suggests that, for domain adaptation, reducing the feature distance between centers plays a more critical role.

Table 6: Hyperparameter selection

α	β	γ	ACC
0.3	0.4	0.3	73.45
0.5	0.2	0.3	72.98
0.4	0.3	0.3	73.50
0.3	0.3	0.4	73.79
0.3	0.2	0.5	73.63
SGD	0.9	The learning rate	0.002
Weight decay	10^{-4}	Batch size	16

In terms of pre-processing, random vertical flipping, random rotation and random horizontal flipping were performed, enhancing the model's generalization ability and increase data utilization. Specifically, the horizontal flipping and vertical flipping operate on the image randomly with a probability of 0.5, and the random rotation of the image is between 30 and 60.

In Table 3, in contrast to ResNet-50 (source only), which has undergone no multicenter adaptation, all methods have made remarkable advancements, indicating the significance of multicenter research. Nonetheless, when juxtaposed against a fully supervised method, there exists a discernible gap that warrants further investigation. The DMCD and CoUDA are two models that account for noise. Compared with these two methods, the DCLCN yields improvements in all three metrics. CoUDA considers the impact of noisy labels on knowledge transfer, while DMCD addresses noise by incorporating high-confidence samples into the training set. However, neither of these methods considers the influence of ultrasound noise itself on multicenter knowledge transfer. AD^2A more closely considers the more valuable regions in medical images. However, its effectiveness is limited in the context of plaque ultrasound images. Noise reduction remains a crucial aspect that requires further consideration.

In terms of time complexity, Big O notation and FLOPs analysis are shown in Table 7. CoUDA [16] and TRANS-DA [31] were chosen because CoUDA is based on the CNN framework, and TRANS-DA is based on the Transformer framework, making them representative in terms of complexity. The FLOPs of DCLCN is very close to CoUDA, and an order of magnitude smaller than the Transformer framework. The big O notation of DCLCN exhibits lower complexity than TRANS-DA, while higher than CoUDA. However, DCLCN outperforms CoUDA on all metrics of classification performance (as shown in Table 3).

In this study, datasets from three different centers were used. Although these datasets provided valuable cross-center learning opportunities, the limited number of data sources may somewhat affect the model's generalizability and performance. The collection of medical images is subject to objective constraints.

Currently, the number of centers is typically limited to 2–3 in multi-center studies of ultrasound images, with an average of about 300 samples per center [18]. The data collected in our study is similar to this scale, and will be expanded to include data from more centers to further validate the model's robustness in future work.

Table 7: FLOPs analysis and Big O notation

Methods	FLOPS	Big O notation
CoUDA [16]	4.0 G	$O(n)$
TRANS-DA [31]	55 G	$O(n^3)$
DCLCN (ours)	4.2 G	$O(n^2)$

6 Conclusions

Existing deep learning models for carotid artery ultrasound images typically rely on data from a single center, and extending them to multiple centers often leads to a significant drop in accuracy due to center-specific differences and ultrasound image noise. To address this, we propose an unsupervised domain adaptation method, DCLCN, for predicting the risk of multicenter plaques. DCLCN incorporates the MFA and label modeling and correction modules to mitigate discrepancies and noise between centers. While it shows promising results on three ultrasound datasets, its performance still lags behind fully supervised models. Future work will focus on improving DCLCN by integrating multimodal and multicenter data, along with expert collaboration, to enhance clinical diagnosis.

Acknowledgement: This work is supported by Shanghai Technical Service Computing Center of Science and Engineering, Shanghai University.

Funding Statement: This research was partly funded by Science and Technology Commission of Shanghai Municipality, grant numbers 22dz1202000, 23002460100.

Author Contributions: Louyi Jiang: Study conception and design; Analysis and interpretation of results; Methodology; Verification; Draft manuscript preparation. Sulei Wang: Data collection; Analysis and interpretation of results; Draft manuscript preparation; Writing review & editing. Wei Shao: Study conception and design; Project administration; Supervision. Jiang Xie: Study conception and design; Writing review & editing; Project administration; Supervision. Haiya Wang: Data collection; Project administration; Supervision; Visualization. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to ethical and privacy concerns, it is not convenient to share this data and material.

Ethics Approval: This research was conducted with human volunteers, and all participating human volunteers provided informed consent.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. *Circulation*. 2022;145(8):e153–639.
2. Libby P. Inflammation and cardiovascular disease mechanisms. *Am J Clin Nutr*. 2006;83(2):456S–60S.
3. Alsheikh-Ali AA, Kitsios GD, Balk EM, Lau J, Ip S. The vulnerable atherosclerotic plaque: scope of the literature. *Ann Internal Med*. 2010;153(6):387–95.

4. Salem M, Bown M, Sayers R, West K, Moore D, Nicolaides A, et al. Identification of patients with a histologically unstable carotid plaque using ultrasonic plaque image analysis. *Eur J Vasc Endovasc Surg.* 2014;48(2):118–25.
5. Saba L, Sanagala SS, Gupta SK, Koppula VK, Johri AM, Sharma AM, et al. Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system. *Int J Cardiovasc Imaging.* 2021;37:1511–28.
6. Saito K, Watanabe K, Ushiku Y, Harada T. Maximum classifier discrepancy for unsupervised domain adaptation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018; Computer Vision Foundation. p. 3723–32.
7. Giglioni V, Poole J, Venanzi I, Ubertini F, Worden K. A domain adaptation approach to damage classification with an application to bridge monitoring. *Mech Syst Signal Process.* 2024;209(2):111135. doi:10.1016/j.ymssp.2024.111135.
8. Abdalla A, Wheeler TA, Dever J, Lin Z, Arce J, Guo W. Assessing fusarium oxysporum disease severity in cotton using unmanned aerial system images and a hybrid domain adaptation deep learning time series model. *Biosyst Eng.* 2024;237(6):220–31. doi:10.1016/j.biosystemseng.2023.12.014.
9. Chamarthi S, Fogelberg K, Brinker TJ, Niebling J. Mitigating the influence of domain shift in skin lesion classification: a benchmark study of unsupervised domain adaptation methods. *Inform Med Unlocked.* 2024;44(11):101430. doi:10.1016/j.imu.2023.101430.
10. Yousuf M, Nobi M. A new method to remove noise in magnetic resonance and ultrasound images. *J Sci Res.* 2011;3(1):81–1. doi:10.3329/jsr.v3i1.5544.
11. Müller HFG, Viacoz A, Kuzmanovic I, Bonvin C, Burkhardt K, Bochaton-Piallat ML, et al. Contrast-enhanced ultrasound imaging of carotid plaque neo-vascularization: accuracy of visual analysis. *Ultrasound Med Biol.* 2014;40(1):18–24. doi:10.1016/j.ultrasmedbio.2013.08.012.
12. Skandha SS, Gupta SK, Saba L, Koppula VK, Johri AM, Khanna NN, et al. 3-D optimized classification and characterization artificial intelligence paradigm for cardiovascular/stroke risk stratification using carotid ultrasound-based delineated plaque: atheromatic™ 2.0. *Comput Biol Med.* 2020;125(10):103958. doi:10.1016/j.compbimed.2020.103958.
13. Xie J, Li Y, Xu X, Wei J, Li H, Wu S, et al. CPTV: classification by tracking of carotid plaque in ultrasound videos. *Comput Med Imaging Graph.* 2023;104:102175. doi:10.1016/j.compmedimag.2022.102175.
14. Altan G. Deep learning-based mammogram classification for breast cancer. *Int J Intell Syst Appl Eng.* 2020;8:171–6. doi:10.18201/ijisae.2020466308.
15. Altan G. DeepOCT: an explainable deep learning architecture to analyze macular edema on OCT images. *Eng Sci Technol, Int J.* 2022;34(23):101091. doi:10.1016/j.jestch.2021.101091.
16. Zhang Y, Wei Y, Wu Q, Zhao P, Niu S, Huang J, et al. Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Trans Image Process.* 2020;29:7834–44. doi:10.1109/TIP.2020.3006377.
17. Yu Z, Liu Y, Dai X, Cui E, Cui J, Ma C. Enhancing preoperative diagnosis of microvascular invasion in hepatocellular carcinoma: domain-adaptation fusion of multi-phase CT images. *Front Oncol.* 2024;14:1332188. doi:10.3389/fonc.2024.1332188.
18. Seoni S, Shahini A, Meiburger KM, Marzola F, Rotunno G, Acharya UR, et al. All you need is data preparation: a systematic review of image harmonization techniques in Multi-center/device studies for medical support systems. *Comput Methods Programs Biomed.* 2024;250(9):108200. doi:10.1016/j.cmpb.2024.108200.
19. Ying X, Zhang Y, Wei X, Yu M, Zhu J, Gao J, et al. MSDAN: multi-scale self-attention unsupervised domain adaptation network for thyroid ultrasound images. In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; 2020; IEEE. p. 871–6.
20. Xie S, Zheng Z, Chen L, Chen C. Learning semantic representations for unsupervised domain adaptation. In: *International Conference on Machine Learning. Proceedings of Machine Learning Research*; 2018; PMLR Microtome Publishing. p. 5423–32.
21. Xu Y, Gong M, Xie S, Batmanghelich K. Box-adapt: domain-adaptive medical image segmentation using bounding boxsupervision. *arXiv:2108.08432.* 2021.

22. Shi P, Xin J, Zheng N. Correcting pseudo labels with label distribution for unsupervised domain adaptive vulnerable plaque detection. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021; IEEE. p. 3225–8.
23. Fuster V, Moreno PR, Fayad ZA, Corti R, Badimon JJ. Atherothrombosis and high-risk plaque: Part I: evolving concepts. *J Am Coll Cardiol*. 2005;46(6):937–54. doi:10.1016/j.jacc.2005.03.074.
24. Borgwardt KM, Gretton A, Rasch MJ, Kriegel HP, Schölkopf B, Smola AJ. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*. 2006;22(14):e49–57. doi:10.1093/bioinformatics/btl242.
25. Li S, Lv F, Xie B, Liu CH, Liang J, Qin C. Bi-classifier determinacy maximization for unsupervised domain adaptation. *Proc AAAI Conf Artif Intell*. 2021;35(10):8455–64. doi:10.1609/aaai.v35i10.17027.
26. Yi K, Wu J. Probabilistic end-to-end noise correction for learning with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019; Computer Vision Foundation. p. 7017–25.
27. Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. In: Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research; 2015; PMLR, Microtome Publishing. p. 1180–9.
28. Long M, Cao Y, Wang J, Jordan M. Learning transferable features with deep adaptation networks. In: Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research; 2015; PMLR, Microtome Publishing. p. 97–105.
29. Chu T, Liu Y, Deng J, Li W, Duan L. Denoised maximum classifier discrepancy for source-free unsupervised domain adaptation. *Proc AAAI Conf Artif Intell*. 2022;36(1):472–80. doi:10.1609/aaai.v36i1.19925.
30. Guan H, Liu Y, Yang E, Yap PT, Shen D, Liu M. Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification. *Med Image Anal*. 2021;71(3):102076. doi:10.1016/j.media.2021.102076.
31. Ye Y, Fu S, Chen J. Learning cross-domain representations by vision transformer for unsupervised domain adaptation. *Neur Comput Appl*. 2023;35(15):10847–60. doi:10.1007/s00521-023-08269-7.
32. Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: why did you say that? arXiv:161107450. 2016.
33. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009; IEEE. p. 248–55.