Cost-Sensitive Distribution Alignment for Improving Automated Medical Diagnosis Using Retinal Fundus Photography

Yangqin Feng[1], Xinxing Xu[1], Zizhou Wang[1], Yan Wang[1], Huazhu Fu[1], Shaohua Li[1], Liangli Zhen[1], Tien-En Tan[2, 3, 4], Mukharram M. Bikbov[5], Jost B. Jonas[6], Chee Wai Wong[2, 3, 4], Ching-Yu Cheng[2, 3, 4], Daniel Shu Wei Ting[2, 3, 4], Rick Siow Mong Goh[1], Yong Liu[1]
[1]Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore
[2]Singapore Eye Research Institute, Singapore [3]Singapore National Eye Centre, Singapore
[4]Duke-National University of Singapore Medical School, Singapore
[5]Ufa Eye Research Institute, Ufa, Bashkortostan, Russia
[6]Department of Ophthalmology, Medical Faculty Mannheim, Heidelberg University, Germany

***Abstract*—Domain adaptation (DA) has emerged as a promising approach to address the domain shift problem in deep learning for automated medical diagnosis. However, current approaches often overlook the imbalanced nature of different categories and primarily focus on aligning the distributions of the source and target domains globally. This oversight leads to suboptimal performance on imbalanced target datasets, as the alignment process becomes dominated by the majority class during adaptation. To tackle this limitation, this paper proposes a novel domain adaptation method called cost-sensitive distribution alignment (CSDA), which aims to enhance deep learning performance on target data. The proposed approach involves collecting a limited-sized dataset from the target domain and utilising CSDA to bridge the domain gap between the source and target domains. Specifically, to address class imbalance, cost-sensitive learning is incorporated into the distribution alignment process, giving more emphasis to the misalignment cost associated with minority class samples. More importantly, CSDA aligns the cross-domain projection distribution in the feature space with the ideal geometric distribution derived from the ground-truth labels. Unlike existing methods that directly minimise or employ adversarial learning to reduce the distribution discrepancy between the source and target domains, our proposed CSDA method focuses on minimising the semantic relationship misalignment among cross-domain samples. Experimental results on the detection of high myopia and myopic macular degeneration (MMD) demonstrate the superiority of CSDA over state-of-the-art methods. The results provide empirical evidence of CSDA's efficacy in enhancing automated medical diagnosis.**

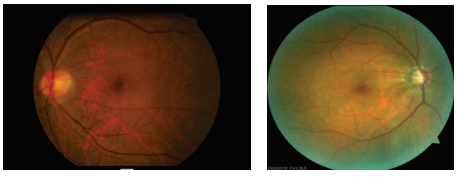
***Index Terms*—Domain adaptation, distribution alignment, medical image analysis, retinal fundus photography**

This work was supported by the Agency for Science, Technology, and Research (A*STAR) through its AME Programmatic Funding Scheme Under Project A20H4b0141, and the Agency for Science, Technology, and Research (A*STAR) through its Biomedical Engineering Programme Project C221318005, the Agency for Science, Technology, and Research (A*STAR) through its RIE2020 Health and Biomedical Sciences (HBMS) Industry Alignment Fund Pre-Positioning (IAF-PP) (grant no. H20C6a0032). The used SEED dataset comes from the SEED study, which is supported by the National Medical Research Council, Singapore (with Projects NMRC/CIRG/1417/2015, NMRC/CIRG/1488/2018, and NMRC/OFLCG/004a/2018).

Corresponding author: X. Xu (e-mail: xuxinx@ihpc.a-star.edu.sg)

I. Introduction

Vision, as the primary means of perceiving the world, is essential in all facets of life. Regrettably, eye conditions and vision impairment are prevalent and pose a significant challenge to eye care, particularly in low- and middle-income nations. The World Report on Vision by the World Health Organisation (WHO) indicates that globally, at least 2.2 billion individuals suffer from vision impairment, and of these, at least 1 billion cases could have been prevented or remain unaddressed [1]. Early diagnosis and treatment are crucial in preventing widespread vision conditions and impairments. Fundus retinal photography, a non-invasive imaging method for capturing colour images of the interior surface of the eye, is widely used for detecting eye disorders and monitoring their progress over time, as it is quick to complete and does not require any invasive procedures. However, the interpretation of fundus photographs is dependent on the expertise of experienced specialists to identify disease pathology. To increase speed and scale in interpretation, artificial intelligence (AI), particularly deep learning [2–4], has been applied to detect major ophthalmic diseases from high-quality retinal fundus images.



(a) SEED (b) UEMS

Fig. 1. Two example images from the SEED dataset and the UEMS dataset, which are captured on Malay and Russian ethnic groups using cameras of CR-DGi with 10 SLR back and VISUCAM 500, respectively.

The application of deep learning in automated medical diagnosis has shown promising results. However, achieving a high

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00079

level of accuracy in predictions relies heavily on a substantial amount of labelled data. In the medical field, obtaining such a large sample size is often hindered by the high cost of medical data collection and annotation, resulting in limited availability of samples from a single institution. Furthermore, even with a significant collection of fundus photographs for AI model training, its performance may still decrease when applied to fundus images from diverse ethnic groups or captured using different imaging cameras and protocols. This discrepancy between the training and test data distributions is commonly known as the "domain shift" problem [5]. To illustrate the domain shift problem, we consider the example of high myopia detection using the Singapore Epidemiology of Eye Disease (SEED) study dataset and the Ural Eye and Medical Study (UEMS) dataset. As depicted in Fig. 1, the image samples from SEED and UEMS exhibit a significant difference, originating from diverse ethnic groups and cameras. This leads to a decline in the performance of AI models trained on SEED, as indicated by a reduction of $10\%-17\%$ in the AUC score when tested on UEMS. To address this issue, two common solutions exist: (1) merging all available datasets into one for training, or (2) fine-tuning pre-trained deep models with annotated target samples. However, both solutions necessitate a considerable amount of high-quality, well-annotated, and clinically verified data, which is often time-consuming and costly to acquire from multiple domains.

In recent years, domain adaptation (DA) has emerged as a promising solution for tackling the domain shift problem. Nevertheless, existing approaches frequently disregard the imbalanced distribution across various categories, instead prioritising the alignment of source and target domain distributions on a global scale. Unfortunately, this oversight results in subpar performance when dealing with imbalanced target datasets, as the adaptation process becomes heavily influenced by the majority class. In this paper, to tackle this limitation, we present a novel domain adaptation method, cost-sensitive distribution alignment (CSDA), to enhance the performance of models for imbalanced target data. Specifically, we tackle this problem in two-fold. On the one hand, we integrate cost-sensitive learning [6] into the distribution alignment process, assigning higher importance to the misalignment cost associated with minority class samples. On the other hand, we align the cross-domain projection distribution in the feature space with the desired geometric distribution derived from the ground-truth labels. Since the ideal geometric distribution is constructed using the category labels of the data samples, it can accurately represent the relationships of the samples across the two domains from a semantic perspective. Consequently, our proposed distribution alignment enforces the deep neural network to learn a feature space where their semantics across the two domains measures the relationships of the samples. The novelty and main contributions of this work are summarised as follows:

- A novel domain adaptation strategy is proposed to enhance the performance of models on imbalanced target data. We construct a two-branch neural network that utilises fundus photographs for automated medical diagnosis. The entire network can be trained end-to-end using the stochastic gradient descent (SGD) algorithm.
- In contrast to existing domain adaptation methods that minimise the distance between the distributions of the two domains directly or employ adversarial learning, our proposed method focuses on minimising the misalignment of **semantic relationships** among cross-domain samples. Additionally, we integrate cost-sensitive learning into the distribution alignment process to address the issue of class imbalance in domain adaptation.
- Extensive experiments have been conducted to evaluate the performance of CSDA. The experimental results for detecting high myopia and myopic macular degeneration (MMD) demonstrate that CSDA surpasses the current state-of-the-art methods. This confirms the effectiveness of CSDA in improving automated medical diagnosis.

II. Related Work

Our proposed method is closely relevant to domain adaptation and cost-sensitive learning. In this section, we discuss some of representative algorithms in these two areas and highlight the key difference between our method and existing ones.

A. Domain adaptation

A common problem when applying deep models to handle medical images is the lack of large-scale well-annotated dataset [7, 8]. Moreover, changes in distributions between different datasets even thorough they are for the same task can occur due to several reasons [9], such as different imaging cameras, different lighting conditions, and parameter settings. Transfer learning [10] has shown some potential in dealing with this challenge.

In transfer learning, one strategy is pre-training the model on the source dataset and fine-tuning the trained model using the target dataset. This strategy is widely used in automated medical diagnosis applications [7]. For instance, Kermany *et al.* applied the transfer learning of pre-training on ImageNet [11] for age-related macular degeneration and diabetic macular edema diagnoses using optical coherence tomography images [12]. The empirical results demonstrate that the prediction accuracy of the fine-tuned model usually can be improved with a considerable margin. However, this general purpose transfer learning strategy ignores the connection between the two domains from a semantic perspective, suffering from the domain shift problem [13] and resulting in a significant accuracy drop.

To solve the domain shift problem, domain adaptation, a more effective way of transfer learning, proposes to align the distributions of the source and target domains [9]. Pioneering DA methods directly minimise the discrepancy between the two domains in the feature space. For instance, Long et al. proposed to minimise the maximum mean discrepancy

(MMD) [14] or the Joint MMD [15] between the two domains. Conjeti *et al.* proposed a supervised DA method to adapt decision forests in the presence of the distribution shift between the two domains [16]. Recently, adversarial learning has been applied to mitigate the domain gap between the two domains. The basic idea is to use the generator to mitigate the domain gap and fool the domain discriminator to classify which domain the input comes from. For example, Conditional Domain Adversarial Networks (CDAN) [5] is proposed to exploit discriminative information implicitly in the classifier predictions to achieve adversarial adaptation. Thomas *et al.* [17] developed a class-aware unsupervised domain adaptation (UDA) method by estimating the label hypothesis of target samples via clustering. Thomas *et al.* proposed the Test-time Unsupervised Domain Adaptation (TTUDA), which contains two training phases [18]. All of these DA methods have achieved promising performance for image classification tasks. Our proposed CSDA considers the domain adaptation from a different perspective compared with existing DA methods. Specifically, CSDA conducts domain adaption by minimising the misalignment of the semantic relationships among the cross-domain samples instead of directly minimising the distance between the distributions of the two domains.

B. Cost-sensitive learning

Cost-sensitive learning [6] considers the different costs associated with majority-class and minority-class data samples, which is a high-performing strategy for handling class-imbalance data. The pioneering cost-sensitive learning work [19] provides some foundations of cost-sensitive learning. Various studies have demonstrated the effectiveness of cost-sensitive learning for class-imbalance classification problems. For instance, Kukar and Kononenko proposed to apply the cost-sensitive modification to the probabilistic estimate of the neural network in the testing stage [20]. It can maintain the original structure of the network and strengthen the original estimates on the minority-class samples with cost consideration. Also, Kukar and Kononenko explored applying the cost modifications to the outputs of the network during the training stage, to the learning rate, or to the expected cost for the misclassified samples [20]. The empirical results show that all these four ways of cost modifications can improve the base classifiers' accuracy. The cost-sensitive learning strategy also has been applied to algorithms of decision trees by conducting cost-sensitive adjustments to the decision threshold, the split criteria, or the pruning schemes [21], Bayesian classifiers [22], and support vector machines [23]. Unlike these methods, our CSDA considers the cost-sensitive learning in the domain alignment instead of in the classification loss or the output of the network. The domain alignment quality has a high impact on effective knowledge transfer.

III. Our Proposed Method

A. Problem formulation

Assume that we have a source dataset that includes n_s samples $\mathbf{X}^s = \left[\mathbf{x}_1^s, \mathbf{x}_2^s, \ldots, \mathbf{x}_{n_s}^s\right]$ and the corresponding data labels as $\mathbf{Y}^s = \left[\mathbf{y}_1^s, \mathbf{y}_2^s, \ldots, \mathbf{y}_{n_s}^s\right]$. Specifically, each data sample $\mathbf{x}_i^s$ in $\mathbf{X}^s$ has a semantic label vector $\mathbf{y}_i^s$, where $\mathbf{y}_i^s \in \{0,1\}^{c_s}$ and c_s is the number of source data categories. The target dataset includes n_t samples $\mathbf{X}^t = \left[\mathbf{x}_1^t, \mathbf{x}_2^t, \ldots, \mathbf{x}_{n_t}^t\right]$. We denote its associated label matrix as $\mathbf{Y}^t = \left[\mathbf{y}_1^t, \mathbf{y}_2^t, \ldots, \mathbf{y}_{n_t}^t\right]$, where $\mathbf{y}_i^t \in \{0,1\}^{c_t}$ and c_t is the number of target categories. In this paper, we consider the scenario where each sample $\mathbf{x}_i^t$ is labelled as a positive or negative case: $\mathbf{y}_i^t = [0,1]^T$ or $\mathbf{y}_i^t = [1,0]^T$. By denoting the probability distributions of the source and target domains as $D^s(\mathbf{x}^s, \mathbf{y}^s)$ and $D^t(\mathbf{x}^t, \mathbf{y}^t)$, we have $D^s \neq D^t$ if there is a domain gap. The goal of domain adaptation is to align D^s and D^t in the feature space according to their underlying semantics in an interactive way.

B. Framework of CSDA

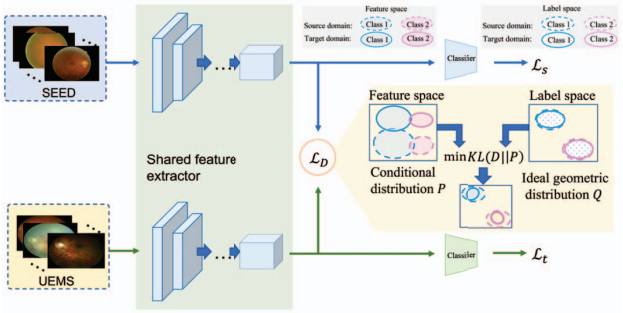


Fig. 2. The framework of CSDA. It contains two branches that share the weights of convolutional feature extraction layers. The whole network can be optimised jointly by minimising the entropy losses of classifying the source and target samples and the distribution misalignment loss simultaneously.

To solve the above problem, we propose a novel method CSDA to mitigate the domain gap and present a framework, as shown in Fig. 2. From the figure, we can see that our proposed framework contains a two-branch neural network. One branch is to classify the source photographs, and another is to classify the target photographs. The two branches share the weights of the convolutional feature extraction layers, which makes the network mapping input samples of the two branches into a shared common space. Note that we explicitly improve the compactness of intra-class samples and the separability of inter-class samples according to their underlying semantics for both two domains. Unlike the existing methods, the proposed method adjusts the compatibility of the data distributions of the two domains based on the ideal geometric distributions obtained from the data labels in an interactive way. As shown in Fig. 2, during the training, we compute the conditional distribution $\mathbf{P}$ between the samples across domains and an ideal geometric distribution $\mathbf{Q}$. Then we minimise the discrepancy between the two distributions by the KL divergence function with cost-sensitive learning. In summary, we impose a new distribution alignment constraint on the two domains explicitly besides minimising the prediction errors of the two classifiers. In this manner, we can learn compatibility features for the two domains, thus improving the model's performance for target data.

C. Objective function

For automated medical diagnosis, we design and minimise the following objective function:

$$\mathcal{L}(\Theta|\mathbf{X}^s, \mathbf{X}^t, \mathbf{Y}^s, \mathbf{Y}^t) = \mathcal{L}_t + \lambda_1 \mathcal{L}_s + \lambda_2 \mathcal{L}_D, \quad (1)$$

where Θ denotes the model's weights. $\mathcal{L}_s$ and $\mathcal{L}_t$ stand for the classification losses for source and target data, respectively. $\mathcal{L}_D$ is the cost-sensitive distribution alignment loss, and λ_1 and λ_2 are the hyper-parameters that control the contributions of the three terms. The details of $\mathcal{L}_s$, $\mathcal{L}_t$ and $\mathcal{L}_D$ are as followings.

By considering the imbalance issue of the datasets of the two domains, we use the weighted cross entropy loss (WCEL) for the two classifiers. Specifically, for each target training sample x_i^t, the loss is defined as:

$$\mathcal{L}_t(x_i^t, y_i^t) = -\left(y_{1i}^t \log(\hat{y}_{1i}^t) + u^t y_{2i}^t \log(\hat{y}_{2i}^t)\right), \quad (2)$$

where u^t is a manual re-scaling weight to emphasise the positive class in the target dataset, and we set u^t as the ratio of the numbers of negative and positive target samples. $\hat{\mathbf{y}}_i^t$ is the output prediction for $\mathbf{x}_i^t$.

For the whole target dataset, we sum up the losses for all the training samples and obtain the classification loss

$$\mathcal{L}_t = \sum_{i=1}^{n_t} \mathcal{L}_t(x_i^t, y_i^t). \quad (3)$$

Similarly, for a training sample in the source domain, we have

$$\mathcal{L}_s(x_i^s, y_i^s) = -\left(y_{1i}^s \log(\hat{y}_{1i}^s) + u^s y_{2i}^s \log(\hat{y}_{2i}^s)\right), \quad (4)$$

where u^s is a manual re-scaling weight of the source domain, $\hat{\mathbf{y}}_i^s$ is the output prediction of the network for $\mathbf{x}_i^s$.

Thus, we compute the classification loss of the source domain as

$$\mathcal{L}_s = \sum_{i=1}^{n_s} \mathcal{L}_s(x_i^s, y_i^s). \quad (5)$$

The CSDA loss is proposed to improve the compatibility of the distributions D^s and D^t by performing iterative interactions between the two domains. Specifically, we introduce the cross-domain projection to learn discriminative source-target representations. CSDA aligns the cross-domain projection distributions to the corresponding ideal geometric distributions by minimising their Kullback–Leibler (KL) divergence. Mathematically, by giving a batch with n source samples $\mathbf{X}^s$ and another batch of n target samples $\mathbf{X}^t$, we compute their representations $\mathbf{Z}^s = [\mathbf{z}_1^s, \mathbf{z}_2^s, \ldots, \mathbf{z}_n^s]$ and $\mathbf{Z}^t = [\mathbf{z}_1^t, \mathbf{z}_2^t, \ldots, \mathbf{z}_n^t]$.

For each source sample, we construct source-target cross-domain pairs as $\{(\mathbf{z}_i^s, \mathbf{z}_j^t), l_{ij}\}$, where l_{ij} indicates whether $\mathbf{z}_i^s$ and $\mathbf{z}_j^t$ is a matched pair. The conditional distribution of matching $\mathbf{z}_i^s$ and $\mathbf{z}_j^t$ is defined as

$$p_{ij} = \frac{e^{(\mathbf{z}_i^{s\,T} \mathbf{z}_j^t)}}{\sum_{k=1}^{n} e^{(\mathbf{z}_i^{s\,T} \mathbf{z}_k^t)}}. \quad (6)$$

For the above equation, we can see that the higher inner product of $\mathbf{z}_i^s$ and $\mathbf{z}_j^t$, the larger probability they will be matched.

In the most optimistic case, if x_i^s and x_j^t are intra-class samples, (for example, both of them belong to the positive class), $\mathbf{z}_i^s$ and $\mathbf{z}_j^t$ should be overlapped; otherwise, they are infinitely far away. Based on such a observation, we define an ideal geometric distribution of matching $\mathbf{z}_i^s$ and $\mathbf{z}_j^t$ as

$$q_{ij} = \frac{l_{ij}}{\sum_{j=1}^{n} l_{ij}}, \quad (7)$$

where l_{ij} is calculated based on the data labels, $l_{ij} = 1$ if $\mathbf{z}_i^s$ and $\mathbf{z}_j^t$ is a matched pair, *i.e.*, they share the same class label; otherwise $l_{ij} = 0$. Since there may exist more than one matched target samples for $\mathbf{z}_i^s$, we normalise the true matching probabilities.

Obtaining the conditional distribution $\mathbf{p}_i = [p_{i1}, p_{i2}, \ldots, p_{in}]$ and the ideal distribution $\mathbf{q}_i = [q_{i1}, q_{i2}, \ldots, q_{in}]$, we compute the matching loss. By considering the class imbalance issue, we minimise the weighted Kullback Leibler (KL) divergence of the two distributions as

$$\mathcal{L}_{s2t}(\mathbf{z}_i^s) = D_{\mathrm{KL}}(\mathbf{p}_i \| \mathbf{q}_i) = \sum_{j=1}^{n} \alpha_{ij} p_{ij} \log \frac{p_{ij}}{q_{ij} + \epsilon}, \quad (8)$$

where

$$\alpha_{ij} = \begin{cases} u^s u^t, & \text{if } \mathbf{y}_i^s = [0,1]^T \text{ and } \mathbf{y}_j^t = [0,1]^T; \\ u^t, & \text{if } \mathbf{y}_i^s = [1,0]^T \text{ and } \mathbf{y}_j^t = [0,1]^T; \\ u^s, & \text{if } \mathbf{y}_i^s = [0,1]^T \text{ and } \mathbf{y}_j^t = [1,0]^T; \\ 1, & \text{otherwise,} \end{cases} \quad (9)$$

and ϵ is a small number to avoid numerical problems [24]. In this work, we minimise $D_{\mathrm{KL}}(\mathbf{p}_i \| \mathbf{q}_i)$ instead of minimising $D_{\mathrm{KL}}(\mathbf{q}_i \| \mathbf{p}_i)$ to select a $\mathbf{p}_i$ that has low probability where $\mathbf{p}_i$ has a low probability [24]. If we minimise $D_{\mathrm{KL}}(\mathbf{p}_i \| \mathbf{q}_i)$, it makes it difficult to distinguish matched and unmatched pairs when multiple positive pairs exist in a mini-batch [25]. The distribution alignment loss from the source domain to the target domain for the whole mini-batch is calculated as

$$\mathcal{L}_{s2t} = \sum_{i=1}^{n} \mathcal{L}_{s2t}(\mathbf{z}_i^s), \quad (10)$$

Similarly, we can calculate the distribution alignment loss $\mathcal{L}_{t2s}$ from the target domain to the source domain for the whole batch samples by exchanging $\mathbf{z}_i^s$ and $\mathbf{z}_j^t$. Adopting the bi-directional distribution alignment, we have the distribution alignment loss as

$$\mathcal{L}_D = \mathcal{L}_{s2t} + \mathcal{L}_{t2s}. \quad (11)$$

By conducting CSDA, we constrain the samples in the same classes close to each other while those in different classes appear far away from each other even though the samples may be from different domains. The objective function in Equation (1) can be optimised with a stochastic gradient descent (SGD) algorithm, such as Adam [26].

IV. Experimental Study

A. Datasets

1) SEED: The Singapore Epidemiology of Eye Disease (SEED) study dataset contains $18,835$ retinal images of $9,083$ patients. In addition, it contains the annotations for high myopia and myopic macular degeneration diagnoses. The images are captured with digital retinal cameras (CRDGi with 10 D single-lens reflex camera back [Canon, Tokyo, Japan]) from three major ethnic groups (Malay, Chinese and Indian) in Singapore.

2) UEMS: The Ural Eye and Medical Study (UEMS) dataset contains $7,781$ retinal images of $4,391$ patients for high myopia detection. The images are captured with another comprehensive fundus platform (VISUCAM 500, ZEISS, Jena, Germany) from individuals of Russian ethnicity in Russia.

3) SNEC-HMC: The Singapore National Eye Centre-High Myopia Clinic study dataset contains 350 retinal images of 180 patients for myopic macular degeneration detection. The images are captured with digital retinal cameras (CRDGi with 10 D single-lens reflex camera back [Canon, Tokyo, Japan]) from three major ethnic groups (Malay, Chinese and Indian) in Singapore.

TABLE I
Statistics of the three datasets used in our experiments.

Dataset	Ethnicity	# patients	# images
SEED	Malay, Indian, Chinese	$9,083$	$18,835$
UEMS	Russian	$4,391$	$7,781$
SNEC	Malay, Indian, Chinese, and others	180	350

TABLE II
Statistical results of the SEED, UEMS and SNEC datasets used in our experiments, where n_{TRAIN} and n_{VALID} are the numbers of training and primary validation set, respectively.

Datasets	Category	n_{train}	n_{valid}	Total
SEED	Negative	$12,629$	$5,413$	$18,042$
	Positive	548	235	783
UEMS	Negative	$5,347$	$2,292$	$7,639$
	Positive	99	43	142
SNEC	Negative	113	38	189
	Positive	97	32	161

Some details of the three datasets are summarised in Tab. I. We randomly split the dataset into the training, validation, and test sets, without overlap of patients from different subsets. The statistics of the three datasets are shown in Tab. II, from which we can find that it includes many more negative cases (Non-myopia) than positive cases (Myopia), leading to an imbalanced classification problem. The SEED dataset is used as the source dataset for high myopia detection and MMD detection. The UEMS dataset is used as the target dataset for high myopia detection, and the SNEC dataset is used as the target dataset for MMD detection.

B. Experimental settings

In this work, we evaluate CSDA with different backbones, including the ResNets [27], DenseNets [28] and VGGNet [29]. We connect a fully-connected layer with the Softmax activation function to the feature extractor for the source and target branches. We initial the weights of our model's backbone with the weights from a model pre-trained on ImageNet [11]. We adopt Adam [26] with the learning rate as 10^{-5} and set the batch size as 32. The optimal values of $\lambda_1 \in \{0.0001, 0.001, 0.01, 0.1, 1, 10\}$ and $\lambda_2 \in \{0.0001, 0.001, 0.01, 0.1, 1, 10\}$ are selected by using the grid search, and we select the best model on the validation set for evaluating the model's performance on the test set. The entire network is trained on two NVIDIA GeForce RTX 3090 Graphics Processing Units (GPUs) in PyTorch.

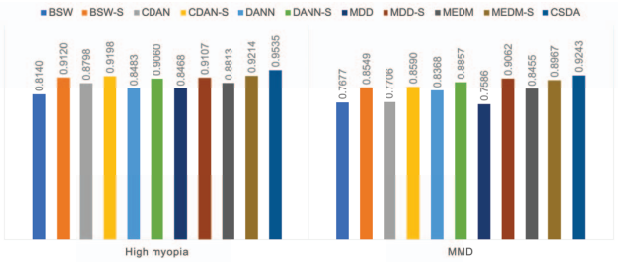


Fig. 3. Comparison of CSDA and five peer methods for the detection of high myopia and MMD in terms of the AUC scores.

C. Comparison with the peer methods

To evaluate the effectiveness of CSDA, we compare it (with the backbone of DenseNet121) with five peer DA methods, namely beyond sharing weights (BSW) [30], CDAN [5], DANN [31], margin disparity discrepancy (MDD) [32], and minimal-entropy diversity maximisation (MEDM) [33]. We also implement each compared DA method with the supervised (-S) learning setting.

Fig. 3 reports the AUC scores of CSDA and the peer methods, from which we observe that:

- Supervised domain adaptation methods (our CSDA, BSW-S, CDAN-S, DANN-S, MDD-S, and MEDM-S) perform better than unsupervised domain adaptation methods (BSW, CDAN, DANN, MDD, and MEDM) with a large margin. For example, the gap for BSW on high myopia detection can be up to 9.98%, and the gap for the MDD method on the detection of MMD can be 14.76%. It indicates that the labels of the target data are essential for model's learning.
- 2) For MMD detection, the target dataset contains a small number of training samples. The AUC scores achieved by the domain adaptation methods are higher than 76%. The score obtained by CSDA can reach 92.43%. This result implies that a large-scale source dataset can be instrumental in improving model's performance on small-scale datasets.
- 3) CSDA outperforms all other domain adaptation methods in detecting high myopia and MMD, which verifies our method's effectiveness. By comparing with other

TABLE III
THE AUC SCORES OF CSDA AND ITS THREE VARIANTS UNDER THE SETTINGS OF USING DIFFERENT BACKBONES FOR HIGH MYOPIA DETECTION.

Model	DenseNet121	DenseNet201	ResNet50	ResNet101	ResNet152	VGG19
w/o $\mathcal{L}_t$	0.8344	0.8153	0.8507	0.6182	0.6927	0.7401
w/o $\mathcal{L}_D$	0.9224	0.8980	0.9163	0.9343	0.9258	0.9144
w/o $\mathcal{L}_s$	0.9321	0.9229	0.9230	0.9223	0.9197	0.9273
CSDA	**0.9535**	**0.9435**	**0.9366**	**0.9431**	**0.9407**	**0.9385**

TABLE IV
THE AUC SCORES ON THE UEMS DATASET UNDER THE SETTINGS OF USING EIGHT DIFFERENT BACKBONES WITH FOUR DIFFERENT TRAINING STRATEGIES FOR HIGH MYOPIA DETECTION.

Method	DenseNet121	DenseNet201	ResNet18	ResNet34	ResNet50	ResNet101	ResNet152	VGG19
S1	0.8204	0.8127	0.8035	0.8306	0.8711	0.8526	0.8347	0.8577
S2	0.9322	0.9295	0.9201	0.8949	0.9298	0.9153	0.9184	0.9204
S3	0.9156	0.8962	0.9154	0.8871	0.9347	0.8984	0.9135	0.9323
CSDA	**0.9535**	**0.9435**	**0.9354**	**0.9264**	**0.9366**	**0.9431**	**0.9407**	**0.9385**

supervised domain adaptation methods, CSDA obtains an improvement of 4.15%, 3.37%, 4.75%, 4.28%, and 3.21% over BSW-S, CDAN-S, DANN-S, MDD-S, and MEDM-S, respectively, in terms of the AUC score for high myopia detection. For MMD detection, CSDA improves the AUC score by 6.94%, 6.53%, 3.86%, 1.81%, and 2.76% compared with BSW-S, CDAN-S, DANN-S, MDD-S, and MEDM-S, respectively.

D. Impact of different loss terms

The objective function of CSDA includes three terms: the classification losses for the two domains and the distribution alignment loss. To investigate the impact of these three terms on CSDA's performance, we construct and evaluate its three variations: the model without $\mathcal{L}_t$ (w/o $\mathcal{L}_t$), the model without $\mathcal{L}_D$ (w/o $\mathcal{L}_D$), and the model without $\mathcal{L}_s$ (w/o $\mathcal{L}_s$).

Tab. III reports the results of CSDA and its three variants for high myopia detection, from which we can see that:

- The full CSDA achieves the highest AUC score under the settings with different backbones, which indicates that every term is essential and can contribute to the final performance.
- CSDA improves the AUC score of the model w/o $\mathcal{L}_t$ on the UEMS dataset with a large margin. For example, CSDA can improve the AUC score by 12% when we adopt DenseNet121 as the backbone. For the setting with the backbone of ResNet152, the improvement can up to 32.49%. It demonstrates that the annotation for the target dataset is essential for achieving a high AUC score.
- The CSDA model outperforms the model w/o $\mathcal{L}_s$ under the setting with all different evaluated backbones. We notice the model w/o $\mathcal{L}_s$ can achieve a high AUC score, e.g., 93.21% for the backbone of DenseNet121. The potential reason is that the label information has been used in the distribution alignment loss $\mathcal{L}_D$.
- The CSDA model outperforms the model w/o $\mathcal{L}_D$. It indicates the importance of the distribution alignment loss in the objective function. It means the iterative distribution alignment in model training is a valuable strategy for transferring knowledge.

E. Impact of backbones and training strategies

We compare CSDA under settings with eight different backbones and four different training strategies for high myopia detection. Specifically, we evaluate the performance of our model under the settings with Densenet121, Densenet201 [28], ResNet18, ResNet34, ResNet50, ResNet101, ResNet152 [27] and VGG19 [29]. We also evaluate three different training and test strategies: training on the SEED training set, then testing the trained models on the SEED test set and extending the test on the UEMS dataset (S1); training on the UEMS training set, then testing the trained models on UEMS test set and extend test on SEED dataset (S2); combining the training set of both SEED and UEMS datasets as a new training set and test on SEED and UEMS test set, respectively (S3). For S1, S2 and S3, we adopt the weighted cross-entropy loss (WCEL) as the objective function term $\mathcal{L}_s$ (or $\mathcal{L}_t$) in Equation (2) (or in Equation (4)) by following the same setting with our CSDA.

The AUC scores for high myopia detection are reported in Tab. IV, from which we can see that:

- From the S1 strategy, the models trained on the SEED dataset significantly outperform the extended test on the UEMS dataset for all the different backbones. From the S2 strategy, the models trained on UEMS outperform the extended test on SEED for all the different backbones. These observations indicate that the two datasets have a large domain gap.
- The models trained on the combined dataset (S3) obtains a lower AUC score when compared to S2:UEMS for most of the different backbones, which demonstrates that merely combining the datasets can not help improve the performance in our task.
- Our method outperforms the other three strategies under the settings of using different backbones. It implies that CSDA can transfer knowledge effectively with the iterative distribution alignment.

V. CONCLUSION

In this paper, we proposed a novel domain adaptation method called CSDA to address the domain shift problem for

imbalanced target data in automated medical diagnosis using fundus photography. Our approach integrated cost-sensitive learning into distribution alignment, effectively mitigating the domain gap for both majority and minority classes. Furthermore, we introduced an ideal geometric distribution based on data labels and computed a condition distribution using data representations. Through an iterative bi-directional alignment process, we aligned the condition distribution with the ideal distribution. It is worth noting that the domain shift problem and imbalanced data issue are common challenges in medical diagnosis. Our CSDA framework was designed to simultaneously tackle these two challenges. The experimental results demonstrated the effectiveness of CSDA in detecting high myopia and myopic macular degeneration.

REFERENCES

[1] World Health Organization, "World report on vision," 2019.

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[3] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *Journal of Healthcare Engineering*, vol. 2019, 2019.

[4] Y. Xie, J. Zhang, and Y. Xia, "Semi-supervised adversarial model for benign–malignant lung nodule classification on chest ct," *Medical Image Analysis*, vol. 57, pp. 237–248, 2019.

[5] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Advances in neural information processing systems*, 2018, pp. 1640–1650.

[6] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[8] L. Alzubaidi, M. Al-Amidie, A. Al-Asadi, A. J. Humaidi, O. Al-Shamma, M. A. Fadhel, J. Zhang, J. Santamaría, and Y. Duan, "Novel transfer learning approach for medical imaging with limited labeled data," *Cancers*, vol. 13, no. 7, p. 1590, 2021.

[9] V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Medical Image Analysis*, vol. 54, pp. 280–296, 2019.

[10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[12] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[13] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018.

[14] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proceedings of the International Conference on Machine Learning*, 2015, p. 97–105.

[15] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proceedings of the International Conference on Machine Learning*, 2017, p. 2208–2217.

[16] S. Conjeti, A. Katouzian, A. G. Roy, L. Peter, D. Sheet, S. Carlier, A. Laine, and N. Navab, "Supervised domain adaptation of decision forests: Transfer of models trained in vitro for in vivo intravascular ultrasound tissue characterization," *Medical Image Aanalysis*, vol. 32, pp. 1–17, 2016.

[17] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4893–4902.

[18] T. Varsavsky, M. Orbes-Arteaga, C. H. Sudre, M. S. Graham, P. Nachev, and M. J. Cardoso, "Test-time unsupervised domain adaptation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 428–436.

[19] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the International Joint Conference on Artificial Intelligence*, vol. 17, no. 1, 2001, pp. 973–978.

[20] M. Kukar, I. Kononenko *et al.*, "Cost-sensitive learning with neural networks." in *Proceedings of the European Conference on Artificial Intelligence*, vol. 15, no. 27, 1998, pp. 88–94.

[21] S. Lomax and S. Vadera, "A survey of cost-sensitive decision tree induction algorithms," *ACM Computing Surveys*, vol. 45, no. 2, pp. 1–35, 2013.

[22] X. Chai, L. Deng, Q. Yang, and C. X. Ling, "Test-cost sensitive naive Bayes classification," in *Proceedings of the IEEE International Conference on Data Mining*, 2004, pp. 51–58.

[23] A. Iranmehr, H. Masnadi-Shirazi, and N. Vasconcelos, "Cost-sensitive support vector machines," *Neurocomputing*, vol. 343, pp. 50–64, 2019.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[25] Y. Zhang and H. Lu, "Deep cross-modal projection learning for image-text matching," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 686–701.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[30] A. Rozantsev, M. Salzmann, and P. Fua, "Beyond sharing weights for deep domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 4, pp. 801–814, 2018.

[31] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[32] Y. Zhang, T. Liu, M. Long, and M. Jordan, "Bridging theory and algorithm for domain adaptation," in *International Conference on Machine Learning*, 2019, pp. 7404–7413.

[33] X. Wu, S. Zhang, Q. Zhou, Z. Yang, C. Zhao, and L. J. Latecki, "Entropy minimization versus diversity maximization for domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 6, pp. 2896–2907, 2023.