

Article

# OCT Retinopathy Classification via a Semi-Supervised Pseudo-Label Sub-Domain Adaptation and Fine-Tuning Method

Zhicong Tan <sup>1</sup>, Qinqin Zhang <sup>2</sup>, Gongpu Lan <sup>2,3</sup>, Jingjiang Xu <sup>2,3</sup>, Chubin Ou <sup>3</sup>, Lin An <sup>3</sup>, Jia Qin <sup>3</sup> and Yanping Huang <sup>2,3,\*</sup> 

<sup>1</sup> School of Mechatronic Engineering and Automation, Foshan University, Foshan 528225, China; zhicongtan0228@gmail.com

<sup>2</sup> Guangdong-Hong Kong-Macao Joint Laboratory for Intelligent Micro-Nano Optoelectronic Technology, School of Physics and Optoelectronic Engineering, Foshan University, Foshan 528225, China; qinqin.zhang87@gmail.com (Q.Z.); langongpu@fosu.edu.cn (G.L.); jaysonxu25@gmail.com (J.X.)

<sup>3</sup> Innovation and Entrepreneurship Teams Project of Guangdong Provincial Pearl River Talents Program, Guangdong Weiren Meditech Co., Ltd., Foshan 528015, China; cou@connect.ust.hk (C.O.); wranlin@weirenmeditech.com (L.A.); wrqinjia@weirenmeditech.com (J.Q.)

\* Correspondence: yale.huangyp@fosu.edu.cn; Tel.: +86-13660026462

**Abstract:** Conventional OCT retinal disease classification methods primarily rely on fully supervised learning, which requires a large number of labeled images. However, sometimes the number of labeled images in a private domain is small but there exists a large annotated open dataset in the public domain. In response to this scenario, a new transfer learning method based on sub-domain adaptation (TSLDA), which involves a first sub-domain adaptation and then fine-tuning, was proposed in this study. Firstly, a modified deep sub-domain adaptation network with pseudo-label (DSAN-PL) was proposed to align the feature spaces of a public domain (labeled) and a private domain (unlabeled). The DSAN-PL model was then fine-tuned using a small amount of labeled OCT data from the private domain. We tested our method on three open OCT datasets, using one as the public domain and the other two as the private domains. Remarkably, with only 10% labeled OCT images (~100 images per category), TSLDA achieved classification accuracies of 93.63% and 96.59% on the two private datasets, significantly outperforming conventional transfer learning approaches. With the Gradient-weighted Class Activation Map (Grad-CAM) technique, it was observed that the proposed method could more precisely localize the subtle lesion regions for OCT image classification. TSLDA could be a potential technique for applications where only a small number of images is labeled in a private domain and there exists a public database having a large number of labeled images with domain difference.

**Keywords:** optical coherence tomography; retinopathy classification; deep learning; transfer learning; sub-domain adaptation

**MSC:** 68T07



**Citation:** Tan, Z.; Zhang, Q.; Lan, G.; Xu, J.; Ou, C.; An, L.; Qin, J.; Huang, Y. OCT Retinopathy Classification via a Semi-Supervised Pseudo-Label Sub-Domain Adaptation and Fine-Tuning Method. *Mathematics* **2024**, *12*, 347. <https://doi.org/10.3390/math12020347>

Academic Editor: Faheim Sufi

Received: 25 December 2023

Revised: 15 January 2024

Accepted: 18 January 2024

Published: 21 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Optical coherence tomography (OCT) has become a de facto standard for guiding the diagnosis and treatment of several leading diseases of blindness worldwide, such as age-related macular degeneration (AMD) and diabetic macular edema (DME) [1]. However, current manual diagnosis of retinopathies using OCT images are labor-intensive, time-consuming and easily affected by the subjective experience of ophthalmologists.

Recently, with the fast development of hardware computing resources and the availability of a large amount of data, deep learning (DL) has achieved great success in various tasks, including medical image processing and analysis [2–5]. For classification, popular deep learning approaches firstly conduct feature extraction using convolutional neural networks (CNN) and then build neural network classifiers using fully connected layers. In

OCT retinopathy classification, a number of studies have focused on using fully supervised DL method, which requires a large amount of labeled data. For example, Li et al. [6] trained VGG-16 to classify OCT images with AMD and DME, which achieved a high accuracy of 98.6%, with a sensitivity of 97.8% and a specificity of 99.4%. Lu et al. [7] used the ResNet-101 network for multi-categorical retinopathy classification. In their study, the accuracies in discriminating normal, cystoid macular edema, serous macular detachment, epiretinal membrane and macular hole were 97.3%, 84.8%, 94.7%, 95.7% and 97.8%, respectively. Alqudah et al. [8] proposed a CNN architecture which has fewer layers compared with AlexNet to classify five classes of retinopathies. The overall accuracy in their study was 95.3%. These fully supervised DL models were normally trained with hundreds or thousands of images for each class. Therefore, it is generally seen that satisfactory classification can be achieved when training examples are sufficient.

Collecting and annotating large-scale datasets are time-consuming and expensive in the real world. The deep learning model is prone to being overfit in a lack of annotated data. To address this problem, semi-supervised and unsupervised learning methods have recently attracted great attention. For semi-supervised deep learning, it intends to learn visual features based on a small amount of labeled data. Sedai et al. [9] introduced a semi-supervised uncertainty-guided student–teacher deep learning framework to improve the segmentation of retinal structures in OCT images. For unsupervised learning, an effective technique is the domain adaptation. It intends to overcome the difference between two different but closely related domains with and without labeled data, i.e., the model trained using the labeled dataset is able to work well on the unlabeled dataset. Wang et al. [10] proposed a generative network-based domain adaptation model to address the cross-domain OCT images classification task. They applied the model to OCT images obtained from two different device manufacturers and achieved a cross-domain classification accuracy of 95.53%. Luo et al. [11] proposed a novel domain alignment method with adversarial learning and entropy minimization to train a model based on a labeled source domain and then adapted it to the unlabeled target domain, which achieved retinopathy classification accuracies of 91.5%, 95.9% and 99.0% in three cross-domain scenarios, respectively.

However, to our knowledge, there are few studies on testing whether an existing public OCT dataset can be used to train a model that can work well on a private dataset having a large domain difference but a small number of labeled data, which is critical to clinical OCT applications lacking sufficiently labeled data from qualified ophthalmologists. In this paper, we discuss this new scenario, i.e., assuming that we only have a private OCT database named as the private domain, which has a few labeled but a large number of unlabeled images, with another fully labeled public OCT database named as the public domain. This situation exists commonly in a product development cycle, e.g., at the early stage of commercializing a new medical imaging such as OCT device when only a small amount of labeled data are collected at the beginning of its own clinical trial. Our purpose is to train a network that is able to work well on the private domain with a small number of labeled images. To achieve this objective, we propose a new transfer-learning approach based on the sub-domain adaptation (TLSDA) method for the automatic classification of retinopathy using OCT images. Specifically, the proposed TLSDA method consists of two steps. The first step is to use a new sub-domain adaptation method to align the feature spaces of the public domain and the private domain. The second step is to use a small percentage, e.g., 10% labeled data, in the private domain to further fine-tune the domain adapted model. Experiments showed that our method could obtain remarkable results for OCT image classification even with very few labeled OCT images. Details of this study are described as follows.

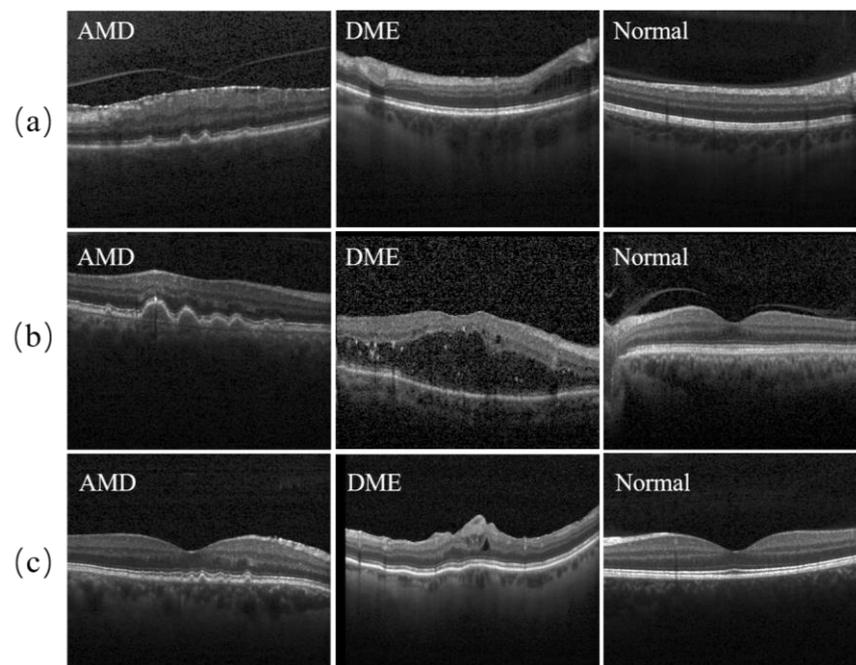
## 2. Materials and Methods

### 2.1. Datasets and Processing Method

In this paper, three publicly open OCT retinopathy datasets are utilized to demonstrate the effectiveness of the proposed algorithm. The first dataset (denoted as Dataset A) was

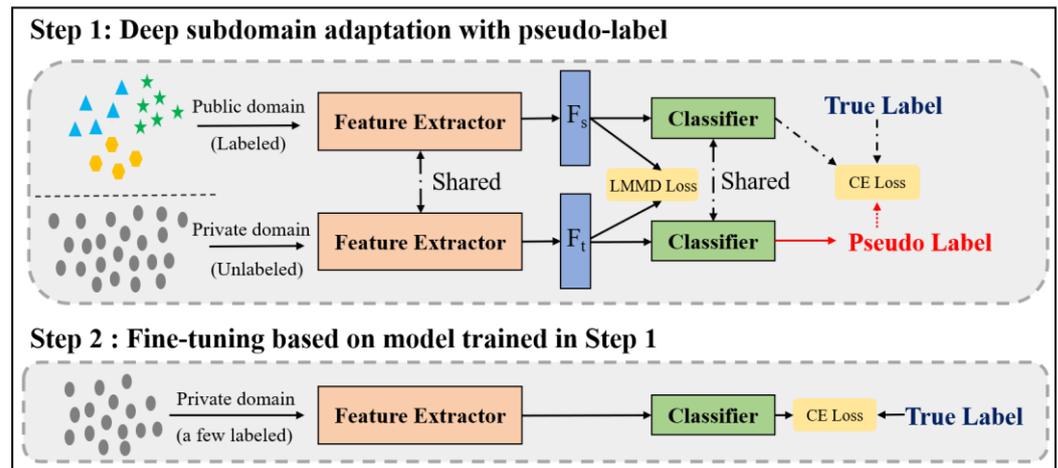
acquired from 45 subjects in different locations of the USA, which includes 723 AMD, 1101 DME and 1407 normal images [12]. The second dataset (denoted as Dataset B) was obtained at Noor Eye Hospital in Tehran, Iran [13]. Specifically, Dataset B was acquired from 148 subjects, which includes 1565 AMD, 1104 DME and 1585 normal images. The third dataset was collected from six different hospitals in the USA and China [1], which includes 37,206 CNV, 11,349 DME, 8617 drusen and 51,140 normal images from 4686 subjects.

Since Dataset A and Dataset B only consider “dry” AMD and drusen usually present in early AMD, to keep consistency, we discarded the category of CNV (“wet” AMD) and treated drusen as AMD in the third dataset. Furthermore, to balance the total numbers in each dataset, we randomly selected 1000 AMD images, 1000 DME images and 1000 normal images in the third dataset and named it as Dataset C for experiment in this study. Although those three datasets were all imaged using the same brand of OCT imaging system (Spectralis, Heidelberg Engineering GmbH., Heidelberg, Germany), subject characteristics were found to be quite different. For instance, subjects in Dataset A and Dataset C were predominantly Caucasian and those in Dataset B were predominantly Asian. In addition, Dataset A, Dataset B and Dataset C were collected in 2014, 2017 and 2013~2017, respectively. This indicated different datasets might come from different versions of the same brand of OCT machine, which might bring variations to the acquired OCT images. Figure 1 shows some typical examples of different retinopathies in the three datasets.



**Figure 1.** Some typical OCT images of AMD, DME and normal eyes in (a) Dataset A, (b) Dataset B and (c) Dataset C.

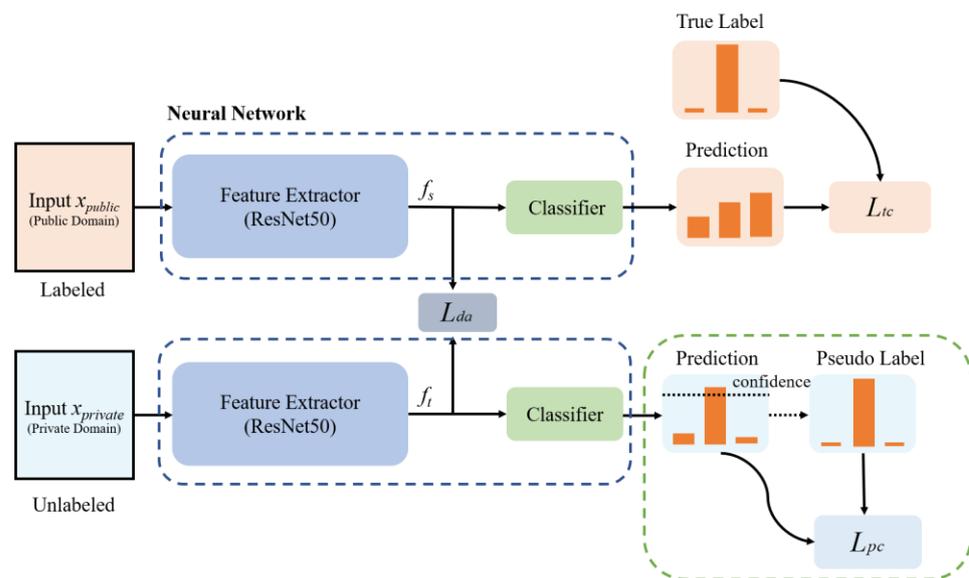
The proposed TLSDA method is shown in Figure 2. It is assumed that we have a public (source) domain and a private (target) domain. The data in the public domain are fully labeled, while only 10% of the private domain data (~100 images per category) are labeled. Although the two domain datasets include the same type of cross-sectional B-mode OCT retinal images, their data distribution is significantly different. The proposed TLSDA method consists of two steps. In Step 1, we assume all the data in the private domain are not labeled and a new sub-domain adaptation algorithm is used to reduce the discrepancy of feature distribution between the public domain  $D_{public} = \{x_{public}, y_{public}\}$  and the private domain  $D_{private} = \{x_{private}\}$ . In Step 2, we fine-tune the Step 1 pretrained network using 10% of the labeled data in the private domain. More details are provided in the following sections.



**Figure 2.** An overview of the proposed transfer learning method based on sub-domain adaptation (TLSDA).

2.1.1. Deep Sub-Domain Adaptation with Pseudo-Label

In this step, we propose a new deep sub-domain adaptation network with pseudo-label (DSAN-PL) to train a neural network, which consists of a feature extractor and a classifier. The basic feature extractor is ResNet-50 with all the fully connected layers removed, which aims at domain-invariant feature extraction. The classifier is a fully connected layer with three output neurons for classifying three types of retinopathies. The structure of the sub-domain adaptation method is presented in Figure 3. The overall loss function contains three types of loss functions, which is defined as:



**Figure 3.** The proposed sub-domain adaptation method used for cross-domain retinopathy classification of OCT images.  $x_{public}$  and  $x_{private}$  are image samples from the public and private domains, respectively.  $f_s$  and  $f_t$  are the extracted features for the public and private domains, respectively.

$$L = L_{tc} + \lambda L_{da} + \alpha(t) L_{pc} \tag{1}$$

where  $L_{tc}$  is the true label classification loss,  $L_{da}$  is the sub-domain adaptation loss,  $L_{pc}$  is the pseudo-label classification loss,  $\lambda$  and  $\alpha(t)$  are weighting coefficients, and  $t$  indicates the epoch number in training. The pseudo-label is used for labeling the unlabeled data, which indicates the predicted class with the condition that the classification probability for the unlabeled image is greater than a confidence threshold, e.g., 96% used in this study. The detailed information of the three sub-loss functions is presented as follows.

**Sub-domain adaptation loss:** The sub-domain adaptation loss is designed for reducing the discrepancy in sub-domain features distribution between the public and the private domain. We adopted the local maximum mean discrepancy (LMMD) loss [14], which is specifically used to quantitatively evaluate the sub-domain feature distance and calculated as follows:

$$L_{da} = \frac{1}{C} \sum_{c=1}^C [\sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \omega_{1-c}^i \omega_{1-c}^j k(f_s^i, f_s^j) + \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \omega_{2-c}^i \omega_{2-c}^j k(f_t^i, f_t^j) - 2 \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \omega_{1-c}^i \omega_{2-c}^j k(f_s^i, f_t^j)] \tag{2}$$

where  $f_s$  and  $f_t$  are image feature vectors generated by the feature extractor,  $k$  is a kernel function to compute the dot product of two image feature vectors,  $\omega_{1-c}^i$  and  $\omega_{2-c}^j$  are the weight of  $x_{public}^i$  and  $x_{private}^j$  belonging to class  $c$ ,  $C$  is the numbers of total classes, and  $n_s$  and  $n_t$  are the sample size for the public and private domains, respectively. The kernel function is used in association with transforming features into sparse spaces so that the chances of linear separability become higher. Readers can find more details of LMMD in [14]. According to [14], the true label  $y_{public}^i$  can be used as a one-hot vector to compute  $\omega_{1-c}^i$  in the public domain:

$$\omega_{1-c}^i = \frac{y_{public-c}^i}{\sum_{(x_{public}^j, y_{public}^j) \in D_{public}} y_{public-c}^j} \tag{3}$$

Since the data in the private domain are unlabeled, the classification output can be set as the probability of assigning  $x_{private}^j$  to each of the  $C$  classes. Then,  $\omega_{2-c}^j$  can be computed similarly for each target sample in the private domain.

**True and pseudo-label classification losses:** The discrepancy in feature distributions across different domains is evaluated using the sub-domain adaptation loss. The classifier in [14] was only trained using a loss function in the labeled domain with true labels, which is defined as follows:

$$L_{tc} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{public}^{ic} \log \hat{y}_{public}^{ic} \tag{4}$$

where  $y_{public}^{ic}$  represents the corresponding true probability of the instance  $i$  for class  $c$  (either 0 or 1),  $\hat{y}_{public}^{ic}$  represents the output probability of the classifier in the public domain,  $N$  is the total number of instances and  $C$  is the number of classes. However, training the classifier using the labeled image data in the public domain may lead to reduced classification performance in the private domain. To solve this problem, we adopted the idea of pseudo-label, a typical semi-supervised learning skill [15]. Specifically, we also used the representations with pseudo-labels in the private domain to train the classifier, which can be defined as follows:

$$L_{pc} = -\frac{1}{M} \sum_{i=1}^M \sum_{c=1}^C y_{pseudo}^{ic} \log \hat{y}_{pseudo}^{ic} \tag{5}$$

where  $\hat{y}_{pseudo}^{ic}$  represents the predicted probability of the classifier,  $y_{pseudo}^{ic}$  represents the corresponding pseudo-probability of the instance  $i$  for class  $c$  (either 0 or 1) in the private domain,  $M$  is the total number of instances and  $C$  is the number of classes.

### 2.1.2. Model Fine-Tuning Based on Deep Sub-Domain Adaptation (TLSDA)

We employed the fine-tuning technique to further train the neural network following the pseudo-label-based sub-domain adaptation. Specifically, we initialized the network

parameters from Step 1, which was trained with the deep sub-domain adaptation with pseudo-label, and fine-tuned it using a few labeled OCT data in the private domain. To preserve the knowledge previously acquired by the model, we used a smaller learning rate during fine-tuning, which prevented significant weight changes.

## 2.2. Evaluation Metrics and Model Implementation

To quantify the performance of different methods, we used six classification evaluation metrics, including Accuracy (ACC), Precision, Recall, Specificity, Areas Under the Receiver Operating Characteristic (ROC) Curve (AUC) and Matthews correlation coefficient (MCC), which are broadly used in machine learning for classification applications. Precision, Recall and Specificity are calculated separately for each class being treated as positive and regarding other classes as negative and then averaged for all the classes in a multi-category classification. We trained the network using the stochastic gradient descent (SGD) with a momentum of 0.9, a batch size of 32, a weight decay of 0.0005, a learning rate of 0.01 in Step 1 and 0.0001 in Step 2. The kernel adopted in Equation (2) was the Gaussian kernel.  $\lambda$  in Equation (1) is set to be 0.5 and  $\alpha(t)$  in Equation (1) is set as shown in Equation (6), where  $\alpha_0 = 0.3$ . The code was written using PyTorch 1.5.0 with Python 3.7 and run on a personal computer with an NVIDIA GeForce GTX 1080 GPU. For each method, we run it for 100 epochs in training.

$$\alpha(t) = \begin{cases} 0 & t < 20 \\ \frac{t-20}{40}\alpha_0 & 20 \leq t < 60 \\ \alpha_0 & t \geq 60 \end{cases} \quad (6)$$

## 2.3. Experiments

### 2.3.1. Domain Bias Experiment

In order to illustrate the domain difference existing in the three datasets, three basic DL models using ResNet50 with randomized initial parameters were trained. Specifically, we trained the model with 90% labeled OCT images on Dataset A, which was named as Model A, and then tested and compared the performance of Model A on the rest of Dataset A, full Dataset B and full Dataset C, respectively. This domain bias experiment was repeated for models trained on 90% Dataset B and 90% Dataset C, which were named as Model B and Model C, respectively.

### 2.3.2. Unsupervised DSAN-PL Experiment

A neural network was trained using the proposed DSAN-PL algorithm. To demonstrate the superior performance of the proposed domain adaptation method, we compared it with some popular and state-of-the-art domain adaptation methods, including DAN [16], DANN [17], DeepCoral [18] and DSAN [14]. Specifically, DAN, DeepCoral and DSAN are statistic moment-matching-based methods, and DANN is an adversarial-based method. For fair comparison, we have performed three domain adaptation tasks, i.e., A to B, A to C and B to C. For A to B, the labeled domain is Dataset A, while the unlabeled domain is Dataset B. The scenarios of A to C and B to C are similarly defined.

### 2.3.3. Semi-Supervised TLSDA Experiment

We further tested the scenario where there was a small percentage (10%) of data with true labels existing in the private domain and how sub-domain adaptation could help improve the classification performance. Following the previous experiment, four different types of experiments were conducted on Dataset B and Dataset C as the private domain where 10% data were labeled. The first method is the basic one without transfer learning (No-TL), i.e., the model trained using 10% of the private OCT dataset first and then tested on the remaining 90% data, with random network parameter initialization. The second one is the transfer learning with ImageNet (TL-ImageNet). This model was obtained by fine-tuning the basic ResNet-50 model pretrained on the ImageNet dataset

using 10% of the private OCT dataset. The third one is similar to the TL-ImageNet, with the difference that the model was fine-tuned using a whole public OCT dataset (here indicating Dataset A) and it was named TL-OCT. The last one is the transfer learning with the sub-domain adaptation model (TLSDA) proposed in this study, which was trained using the 100% source domain data from Dataset A. Our codes for the above three experiments are available at: <https://github.com/tzc123456/OCT-retinopathy-classification> (accessed on 25 December 2023).

### 3. Results

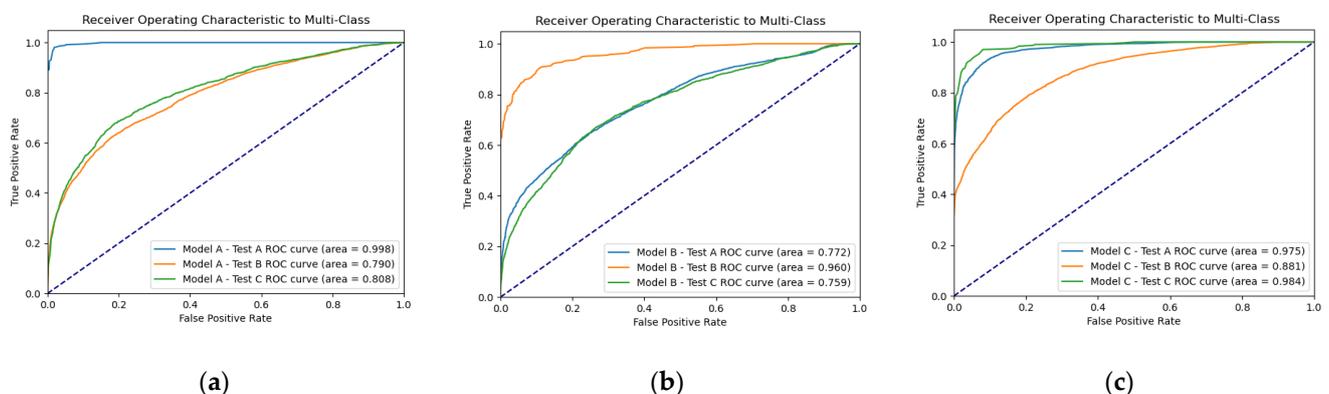
#### 3.1. Domain Bias Experiment Results

Raw results of domain bias experiment using various evaluation metrics and ROC curves are shown in Table 1 and Figure 4, respectively. Figure 5 further shows a comparison of the mean results for the same and different domains. The results clearly show that the model trained on one domain could not be well generalized to another domain directly. The average accuracy was  $91.62 \pm 3.20\%$  for models trained and tested in the same domain, while it was only  $65.34 \pm 11.00\%$  for those trained in one domain and tested in another domain. It clearly indicates the problem of domain bias, which is a well-known problem in computer vision [19]. Based on these results, it could also be found that the generalization result seemed to be the best for C→A as there was only an 8% decrease for this domain change, while the decrease almost approached ~30% for other cross-domain cases. Similar trends of results were found for other evaluation metrics, including Precision, Specificity, Recall, AUC and MCC in all results of this study in comparison with Accuracy (ACC). Therefore, typical results were described using Accuracy here and in all the following text.

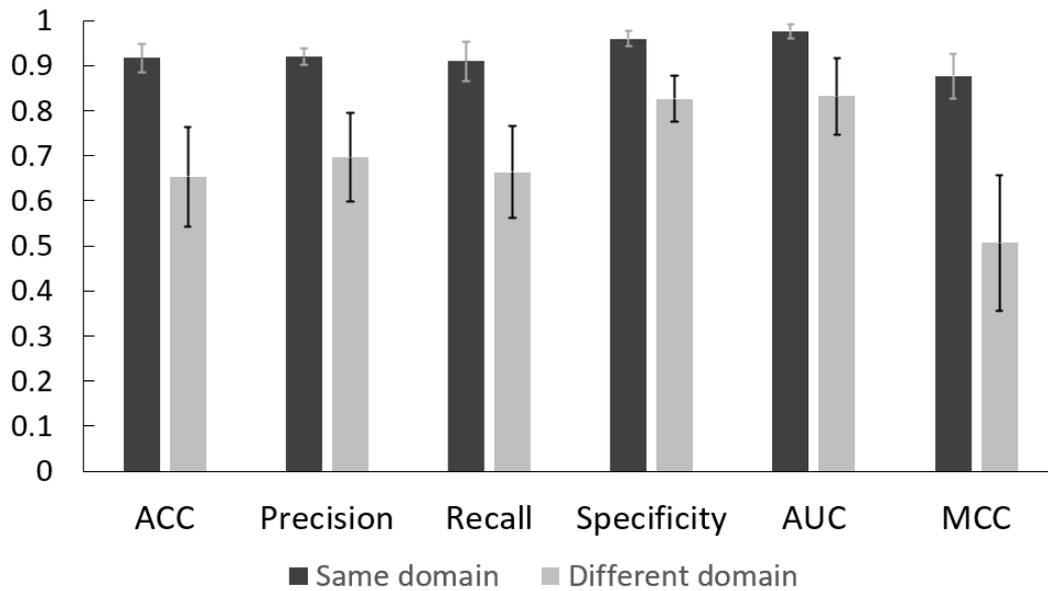
**Table 1.** Results using various evaluation metrics for the domain bias experiment.

Models	Test	ACC (%)	Precision (%)	Recall (%)	Specificity (%)	AUC	MCC
Model A	<b>A</b>	<b>93.79</b>	<b>93.01</b>	<b>94.00</b>	<b>97.13</b>	<b>0.998</b>	<b>0.909</b>
	B	60.06	62.88	61.44	79.43	0.790	0.436
	C	64.37	66.69	64.37	82.18	0.808	0.470
Model B	A	57.91	66.62	61.55	80.99	0.772	0.439
	<b>B</b>	<b>87.74</b>	<b>89.72</b>	<b>86.14</b>	<b>93.53</b>	<b>0.960</b>	<b>0.816</b>
	C	56.20	63.22	56.20	78.10	0.759	0.363
Model C	A	85.58	88.87	84.92	91.90	0.975	0.788
	B	67.89	69.42	69.64	84.46	0.881	0.537
	<b>C</b>	<b>93.33</b>	<b>93.41</b>	<b>93.33</b>	<b>96.67</b>	<b>0.984</b>	<b>0.900</b>

Bold indicates the best result among all tests. Model X means the DL model trained using Dataset X, and Test Y means the trained model is tested on Dataset Y.



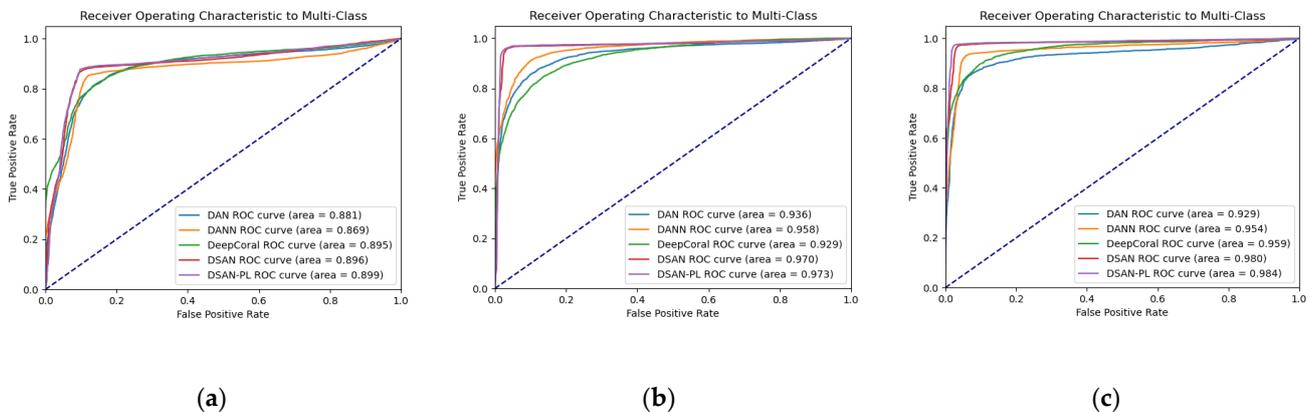
**Figure 4.** ROC curves for testing the domain bias using a model trained using one specific dataset and tested on all the three datasets. (a) ROC curves for Model A; (b) ROC curves for Model B; (c) ROC curves for Model C. See text for the details of different models.



**Figure 5.** Averaged domain bias evaluation results for ACC, Precision, Recall, Specificity, AUC and MCC. Results tested on the same and different domain with respect to training and testing dataset difference are shown in two bars. Error bars indicate the standard deviations of results.

### 3.2. Unsupervised DSAN-PL Results

Table 2 and Figure 6 show the results of the domain adaptation experiment using various evaluation metrics and ROC curves, respectively. It could be found that the basic model ResNet-50 [20] without domain adaptation had the poorest performance. ACCs were 60.06%, 64.37% and 56.20% for the scenarios of A to B, A to C and B to C, respectively, demonstrating the above-mentioned problem of domain differences. With domain adaptation methods, the classification performance in the unlabeled domain was significantly improved. Among various domain adaptation methods, the performance of DANN, which is based on adversarial domain adaptation, outperformed domain adaptation methods based on statistical feature transformation such as DeepCoral and DAN. This indicates that adversarial domain adaptation has a strong domain alignment ability. However, adversarial domain adaptation methods did not consider fine-grained information, so their performance was still inferior to sub-domain adaptation methods such as DSAN [14] and the proposed DSAN-PL. Compared with DSAN, the proposed DSAN-PL further considered the benefit of high-quality pseudo-labels in serving as effective training samples and intrinsically increasing the number of training data, thus improving the model performance.



**Figure 6.** ROC curves of different domain adaptation models across different domains. (a) Domain adaptation for A → B; (b) domain adaptation for A → C; (c) domain adaptation for B → C. See text for the details of different models.

**Table 2.** Comparison results of three scenarios using various domain adaptation methods.

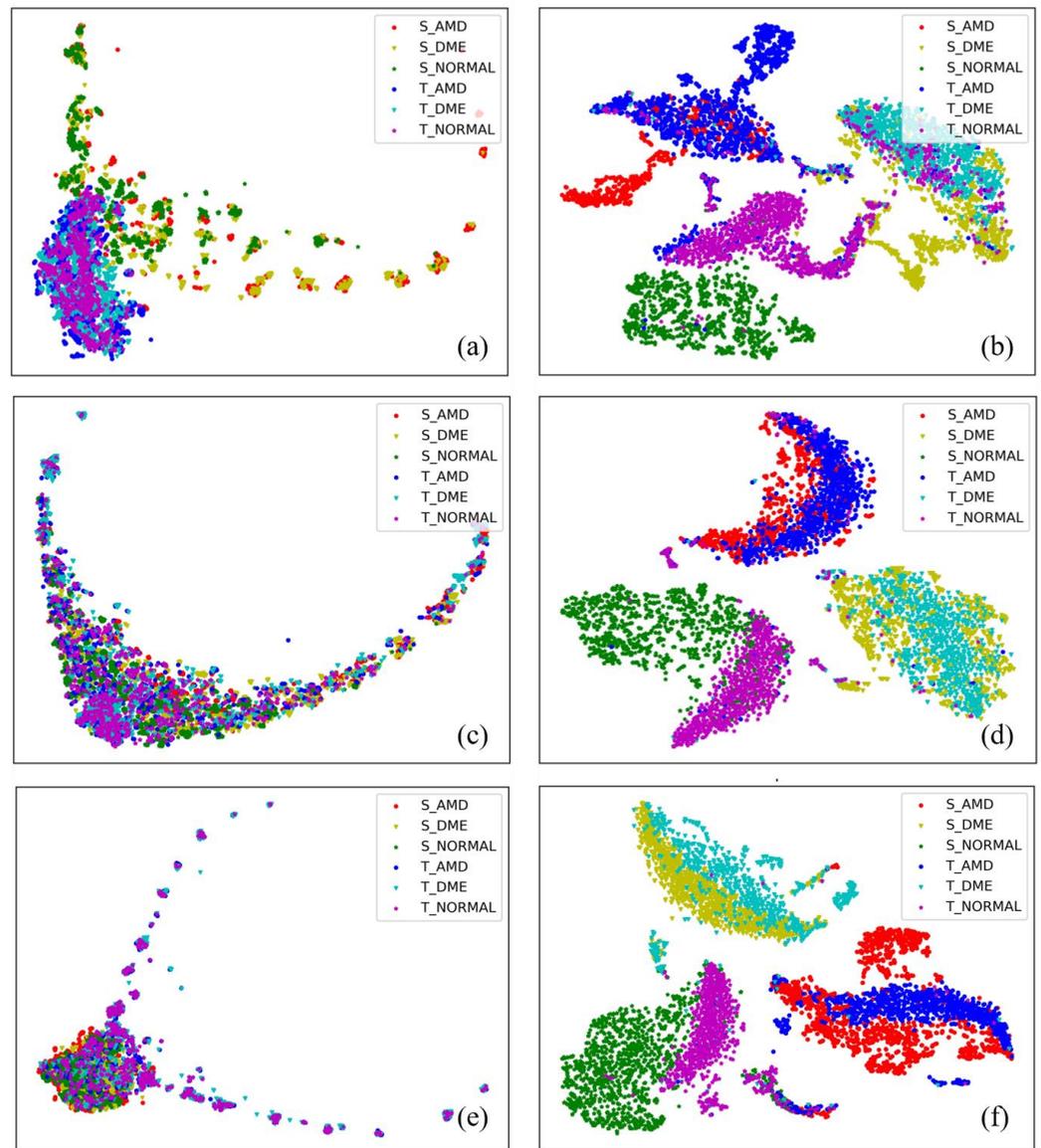
Scenarios	Methods	ACC (%)	Precision (%)	Recall (%)	Specificity (%)	AUC	MCC
A to B	ResNet-50	60.06	62.88	61.44	79.43	0.790	0.436
	DAN	75.04	78.05	77.19	88.05	0.881	0.651
	DANN	82.86	84.90	84.21	91.67	0.869	0.762
	DeepCoral	70.29	73.46	73.10	85.88	0.895	0.590
	DSAN	83.69	84.22	85.06	92.14	0.896	0.764
	<b>DSAN-PL</b>	<b>84.20</b>	<b>84.60</b>	<b>85.52</b>	<b>92.42</b>	<b>0.899</b>	<b>0.771</b>
A to C	ResNet-50	64.37	66.69	64.37	82.18	0.808	0.470
	DAN	82.83	84.49	82.83	91.42	0.936	0.750
	DANN	87.30	89.02	87.30	93.65	0.958	0.818
	DeepCoral	80.27	81.71	80.27	90.13	0.929	0.711
	DSAN	95.17	95.18	95.17	97.58	0.970	0.928
	<b>DSAN-PL</b>	<b>95.35</b>	<b>95.33</b>	<b>95.33</b>	<b>97.67</b>	<b>0.973</b>	<b>0.930</b>
B to C	ResNet-50	56.20	63.22	56.20	78.10	0.759	0.363
	DAN	81.50	84.85	81.50	90.75	0.929	0.741
	DANN	89.77	90.96	89.77	94.88	0.954	0.854
	DeepCoral	80.83	84.14	80.83	90.42	0.959	0.729
	DSAN	95.43	95.71	95.43	97.72	0.980	0.933
	<b>DSAN-PL</b>	<b>96.20</b>	<b>96.36</b>	<b>96.20</b>	<b>98.10</b>	<b>0.984</b>	<b>0.944</b>

Bold indicates the best results.

To further demonstrate the effectiveness of the domain adaptation method, we also plot its learnt features by using the t-distributed stochastic neighbor embedding (T-SNE) technique [21] with some typical examples before and after the sub-domain adaptation shown in Figure 7. The T-SNE is a nonlinear dimensionality reduction method to lower the rank of high-dimensional data to two dimensions for visualization. It could be found that the features in different domains were not aligned well without domain adaptation method. In contrast, the features were aligned quite well after the application of our sub-domain adaptation method.

### 3.3. Semi-Supervised TLSDA Results

Table 3 and Figure 8 show the evaluation results for the transfer learning experiment using various evaluation metrics and ROC curves, respectively. It could be found that No-TL, a model trained with random parameter initialization, had the poorest performance in the private domain. ACCs were 43.19% and 39.15% for Dataset B and Dataset C, respectively. This clearly indicated the probable over-fitting problem in lack of sufficient annotated data for training. Using transfer learning techniques, the model classification performance could be significantly improved. For TL-ImageNet and TL-OCT, ACC increased to 83.63% and 88.90% for Dataset B and to 82.37% and 88.56% for Dataset C, respectively. In other words, transfer learning can be a highly effective technique for a private domain with limited annotated data. Among the four different experiments, the proposed TLSDA method significantly outperformed other methods. It achieved classification accuracies of 93.63% and 96.59% for Dataset B and Dataset C, respectively. It clearly indicates fine-tuning the network with sub-domain adaptation can achieve remarkably improved classification results. For TLSDA, we also conducted a further experiment with different training sample ratios increasing from 5% to 30% in fine-tuning and the experiment results are included in Supplementary Figure S1 and Table S1. The results showed that classification performance generally increased as the training sample ratio increased up to 30%.

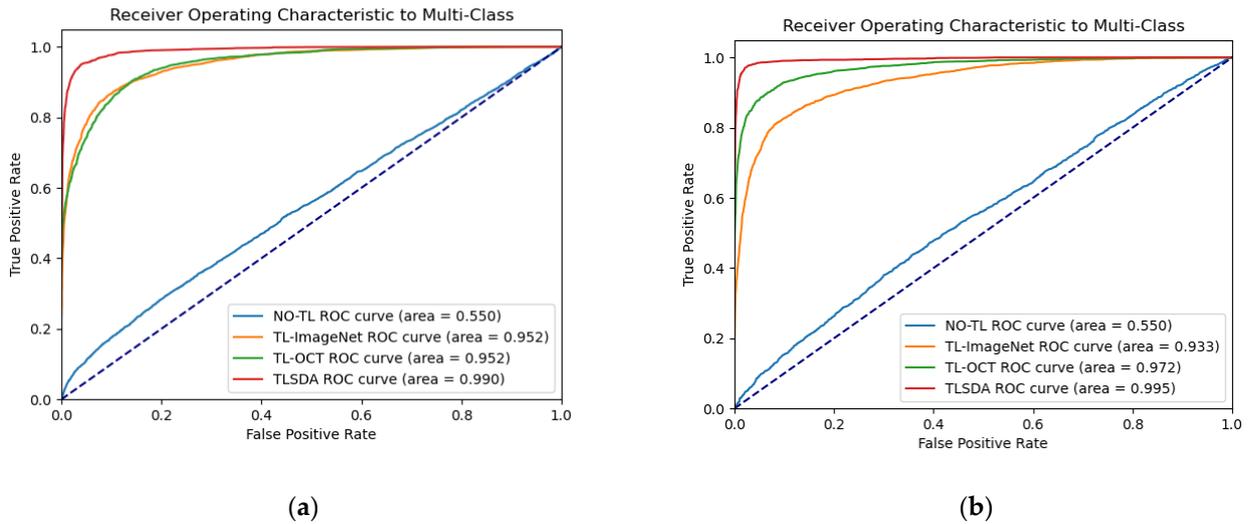


**Figure 7.** T-SNE plots for features of the source and target domain datasets. (a,c,e) are the visualizations of the learned representations for ResNet-50 without domain adaptation on tasks of A→B, A→C and B→C, respectively. (b,d,f) are the visualizations for the proposed domain adaptation on tasks A→B, A→C and B→C, respectively, where better sub-domain feature alignments are clearly seen. S and T stand for the public and private domains, respectively.

**Table 3.** Results using various evaluation metrics for various transfer learning methods.

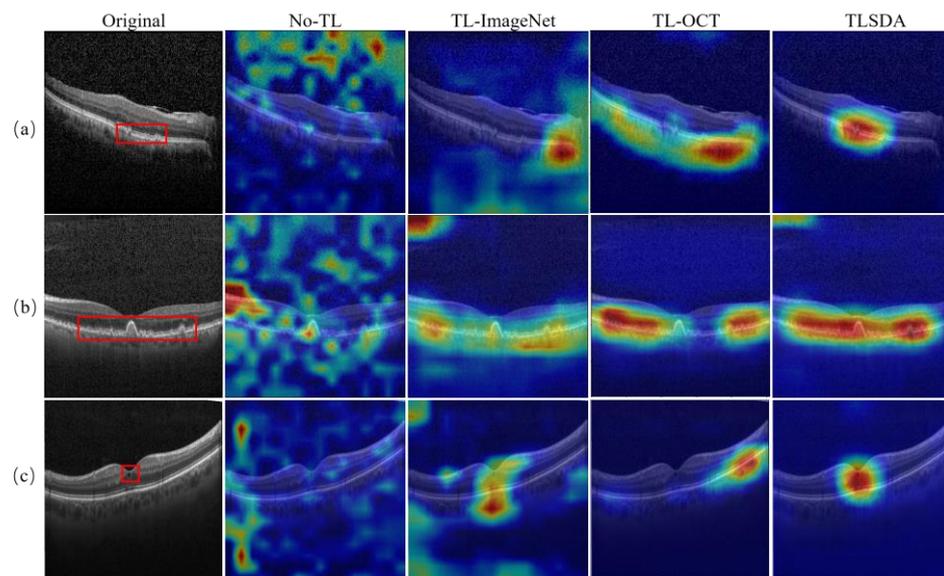
Dataset	Test	ACC (%)	Precision (%)	Recall (%)	Specificity (%)	AUC	MCC
B	No-TL	43.19	52.00	38.74	69.83	0.550	0.156
	TL-ImageNet	83.63	83.75	82.05	91.60	0.952	0.752
	TL-OCT	88.90	89.47	89.15	94.31	0.952	0.834
	<b>TLSDA</b>	<b>93.63</b>	<b>93.73</b>	<b>93.74</b>	<b>96.74</b>	<b>0.990</b>	<b>0.903</b>
C	No-TL	39.15	30.73	39.15	69.57	0.550	0.103
	TL-ImageNet	82.37	83.12	82.37	91.19	0.933	0.739
	TL-OCT	88.56	89.08	88.56	94.28	0.972	0.831
	<b>TLSDA</b>	<b>96.59</b>	<b>96.61</b>	<b>96.59</b>	<b>98.30</b>	<b>0.995</b>	<b>0.949</b>

Bold indicates the best results.



**Figure 8.** ROC curves of different transfer learning models. (a) Testing on Dataset B; (b) testing on Dataset C. See text for the details of different models.

To further demonstrate the effectiveness of the proposed TLSDA method, we visualized the important regions in the OCT image that were vital in classification of various diseases using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique [22]. Grad-CAM is a technique utilizing the weighted average of features and their gradients in the form of a heatmap to visualize the key region of an image in the decision of category classification. Three typical results for classification of two cases of AMD and one case of DME where the lesions are still subtle are shown in Figure 9. In these three examples where the lesion features were not obvious, No-TL, TL-ImageNet and TL-OCT methods did not find the lesion regions well, while the proposed TLSDA could localize these regions accurately. The highlighted regions in TLSDA well include typical lesions of small drusen and edema, which are important symptoms for diagnosis of AMD (Figure 9a,b) and DME (Figure 9c), respectively.



**Figure 9.** Grad-CAM results for three typical examples of (a) AMD Case 1; (b) AMD Case 2; (c) DME Case. The first column shows the original images, while the other three columns show the results of No-TL, TL-ImageNet, TL-OCT and TLSDA, respectively. Red blocks in the first column represent the manually marked key image regions of drusen or cysts for disease prediction.

#### 4. Discussion

Currently, supervised deep learning has achieved remarkable success in OCT retinopathy classification. However, to our knowledge, there are scarce studies focusing on disease classification with a small amount of training data. In this study, we proposed a novel method to solve this problem by firstly utilizing a public dataset through an improved sub-domain adaptation method and then fine-tuning (TLSDA) method. Experimental results showed that the proposed TLSDA method outperformed other popular transfer learning algorithms. TLSDA could be potentially recommended as an effective solution in the application scenario of semi-supervised learning with a small amount of training data for the task of OCT retinopathy classification or other similar applications.

The first step of our proposed TLSDA is the utilization of domain adaption technique. Domain bias generally exists among datasets, which can be caused by various factors including but not limited to measurement bias and sampling bias [23]. In this study, the measurement bias could be induced by the different versions of the image acquisition device, although the same brand of OCT machine was used for the three datasets. The sampling bias referred to the difference of subject populations, for which there might be a significant difference among the three datasets, particularly the ethnicity. According to the domain bias experiment results, the mean classification ACC was  $91.62 \pm 3.20\%$  and  $65.34 \pm 11.00\%$  when the model was applied to the test data of the same and different domains, respectively. Obviously, the domain difference caused the classification performance to be significantly degraded when the DL model was applied to a different domain dataset in comparison with the one used for training. Therefore, our domain bias experiment clearly showed that the model trained in the public domain could not be directly used in the private domain due to domain bias. Further observation showed that the classification model trained on Dataset C was generally more applicable to the other two datasets, which might be due to a broader ethnic range for the tested subjects in Dataset C, which included both Caucasian and Asians. Other factors, such as age and gender, might also play a role in the domain bias but these were not specifically analyzed in this study, which warrants further investigation.

In this study, a modified deep sub-domain adaptation network with pseudo-label (DSAN-PL) was proposed to realize the domain adaptation for the classification task. Sub-domain adaptation was adopted because this technique not only considers the overall cross-domain alignment but also specifically takes account of the sub-domain feature alignment, which is beneficial for improving the classification performance [14]. Our DSAN-PL network further utilized pseudo-labels in training to update the weights of models. Pseudo-labels with high confidence in class prediction should be used in loss calculation; otherwise, low-quality ones may induce error and, therefore, bring adverse rather than beneficial effect to model training. Pseudo-label became more and more reliable along with training time. Therefore, a time-dependent coefficient  $\alpha(t)$  which increases from 0 and a plateau at a constant  $\alpha_0$  with the training time  $t$  was used in the part of the pseudo-label-related loss function to control this timing effect. Among various domain adaptation methods, DSAN and our method performed much better than others. DSAN and our method focused on sub-domain adaptation, which emphasized the objective of learning a local sub-domain shift. Other domain adaptation methods including DAN [16] and DeepCoral [18] mainly learned a global domain shift, i.e., they did not consider the relationships among sub-domains of different classes, which might fail to extract fine-grained features for each class. DANN [17] achieved the task of domain adaptation by incorporating an adversarial structure which is one of the most important breakthroughs in recent deep learning field [24]. With this advanced technique, DANN achieved significantly better results than DAN and DeepCoral. However, its performance was still inferior to DSAN and our method, showing the sub-domain feature alignment as an extraordinarily important point to be considered in improving domain adaptation performance. Compared with DSAN, our method further considered the benefit of high-quality pseudo-labels in serving as effective training samples and intrinsically increasing the number of training

data, thus improving the model performance [15]. It could be seen that the increase in accuracy after using pseudo-labels compared to DSAN without pseudo-labels was 0.18% and 0.77% for the cross-domain learning of  $A \rightarrow C$  and  $B \rightarrow C$ , respectively, which was not so large. It should be noted that a fixed confidence threshold of 96% was used for defining pseudo-labels associated with calculation of pseudo-label classification loss in this study, which could be improved in further investigation using other strategies such as the curriculum-labeling-based pseudo-labeling method [25].

Our transfer learning experiment showed that the proposed TLSDA method with DSAN-PL as the first step for domain adaptation and a fine-tuning with 10% labeled data as the second step for transfer learning was effective to achieve good performance for OCT retinopathy classification. This method was much better than the deep learning model which was pretrained on ImageNet or a public OCT Dataset. The feature heatmap obtained using Grad-CAM also showed the TLSDA could well locate the critical lesion parts such as drusen and cysts, which are also clinical symptoms used by doctors in making a classification decision. Compared to our DSAN-PL, the second step of fine-tuning significantly improved the classification performance, although the extent of improvement was quite different on Dataset B (+9.43%) and Dataset C (+1.24%). The 10% labeled OCT image seemed to be more important in providing new information in the transfer learning of Dataset B than Dataset C in terms of improving classification accuracy, which might be due to the different cross-domain distances in domain adaptation ( $A \rightarrow B$  vs.  $A \rightarrow C$ ). Extra experiments for the training sample ratio changing from 5% to 30% showed that, when 10% labeled data were used, the model was not overfit yet, as the result constantly increased up to 30% sample ratio. It is important to note that, while increasing the proportion of labeled data can improve model accuracy, it does not necessarily mean that more labeled data is always required. In practical applications, labeled data are often expensive and difficult to obtain. The overall transfer learning results have demonstrated that our proposed TLSDA method has the potential to be used in a real scenario where only a small number of labeled images may exist in a private domain, e.g., at the beginning of commercializing a new or new generation OCT or other medical diagnosis device, but there are open clinical datasets regarded as the public domain from other brands of the same type of device or its past generation. In this case, the public domain dataset can be fully utilized through our proposed sub-domain adaptation method to accelerate the establishment of a performant DL model in the private domain.

There are some limitations in this study. First, we only considered the most common retinal diseases including AMD and DME here. Other retinal diseases were overlooked due to too few training samples and, therefore, whether the current conclusions could be well generalized to other retinal diseases is of some concern. Second, ResNet-50 was used as the basic feature extractor in this study. The performance of our method would likely be enhanced when more advanced deep learning architectures such as attention module [26,27] or dense block [28] are used. Third, the training time was still relatively long, which was not particularly considered in this study. Light neural networks such as MobileNet [29] or ShuffleNet [30] or specific model simplification techniques such as knowledge distillation [31] could be considered if small deep learning or some conventional machine learning models [32] are deployed in a mobile terminal or an embedded system with limited computing resources. In general, the problem we tried to solve in this study belongs to a field of meta-learning [33,34] where the specific question is to optimize model parameter initialization in the case of transfer learning with a small sample number. The objective is to adjust the model parameters to be nearer the center of solution before transfer learning with a small amount of data for the target dataset. So, any technique which focuses on solving the problem of small sample such as the few-shot learning [35,36] could be applied to our problem, which will be investigated in future studies. Lastly, the generalizability of our domain adaptation method to other OCT datasets still needs to be evaluated due to different characteristics such as imaging conditions, imaging devices and population demographics existing in different datasets. This domain difference may be

defined using a quantitative index [37], the relationship of which with domain adaptation technique generalizability may be investigated in future studies. Clinical validation studies are also in planning for testing the clinical value of the current method, where various cautions such as recording of imaging conditions, sample size, subject demographics and manual annotation should be excised in close collaboration with doctors to obtain and comprehend the final clinical results.

## 5. Conclusions

This study proposes a novel semi-supervised method using domain adaptation and fine-tuning method to establish an effective deep learning model for classifying retinopathies in OCT images. The superior performance of the proposed method was demonstrated by comparison with state-of-the-art domain adaption methods and also popular transfer learning methods. The proposed method has the potential to be generalized to similar application scenarios in a private domain where training data are insufficient but there exists a public domain with sufficiently labeled data.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/math12020347/s1>, Figure S1. ROC curves for TLSDA using different labeled training sample ratios (5%, 10%, 20% and 30%) for tests in (a) Dataset B and (b) Dataset C. Table S1. Evaluation results of TLSDA with different training sample ratios.

**Author Contributions:** Conceptualization, Z.T., Q.Z., C.O. and Y.H.; Methodology, Z.T., Q.Z., C.O. and Y.H.; Software, Z.T.; Validation, Z.T., Q.Z. and Y.H.; Formal analysis, all authors; Investigation, Z.T., Q.Z. and Y.H.; Resources, Q.Z., L.A., J.Q. and Y.H.; Data curation, Z.T., Q.Z., C.O. and Y.H.; Writing—original draft preparation, Z.T., Q.Z. and Y.H.; Writing—review and editing, all authors; Visualization, Z.T., Q.Z. and Y.H.; Supervision, Q.Z. and Y.H.; Project administration, Q.Z. and Y.H.; Funding acquisition, Q.Z., G.L., J.X., C.O., L.A., J.Q. and Y.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was partially supported by the National Natural Science Foundation of China (62001114, 61871130), Guangdong-Hong Kong-Macao Intelligent Micro-Nano Optoelectronic Technology Joint Laboratory (No. 2020B1212030010) and Innovation and Entrepreneurship Teams Project of Guangdong Pearl River Talents Program (No. 2019ZT08Y105). The funding bodies provided some financial support in conducting the research reported in this study but they had no role in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

**Data Availability Statement:** All data used in this article are available in public databases. Dataset A is available at [12], Dataset B is available at [13], and Dataset C is available at [1].

**Conflicts of Interest:** G.L., J.X. and Y.H. are consultants at Weiren Meditech Co., Ltd. J.Q. and L.A. are currently working at Weiren Meditech Co., Ltd. The remaining authors have no disclosure of conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)]
2. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 142–158. [[CrossRef](#)]
3. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
4. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
5. Gherardini, M.; Mazomenos, E.; Menciassi, A.; Stoyanov, D. Catheter segmentation in X-ray fluoroscopy using synthetic data and transfer learning with light U-nets. *Comput. Methods Programs Biomed.* **2020**, *192*, 105420. [[CrossRef](#)] [[PubMed](#)]
6. Li, F.; Chen, H.; Liu, Z.; Zhang, X.; Wu, Z. Fully automated detection of retinal disorders by image-based deep learning. *Graefes Arch. Clin. Exp.* **2019**, *257*, 495–505. [[CrossRef](#)] [[PubMed](#)]
7. Lu, W.; Tong, Y.; Yu, Y.; Xing, Y.; Chen, C.; Shen, Y. Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. *Transl. Vis. Sci. Technol.* **2018**, *7*, 41. [[CrossRef](#)] [[PubMed](#)]

8. Alqudah, A.M. AOCT-NET: A convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Med. Biol. Eng. Comput.* **2020**, *58*, 41–53. [[CrossRef](#)]
9. Sedai, S.; Antony, B.; Rai, R.; Jones, K.; Ishikawa, H.; Schuman, J.; Gadi, W.; Garnavi, R. Uncertainty guided semi-supervised segmentation of retinal layers in OCT images. In Proceedings of the Medical Image Computing and Computer Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 282–290.
10. Wang, J.; Chen, Y.; Li, W.; Kong, W.; He, Y.; Jiang, C.; Shi, G. Domain adaptation model for retinopathy detection from cross-domain OCT images. In Proceedings of the Third Conference on Medical Imaging with Deep Learning, Montreal, QC, Canada, 6–8 July 2020; pp. 795–810.
11. Luo, Y.; Xu, Q.; Hou, Y.; Liu, L.; Wu, M. Cross-domain retinopathy classification with optical coherence tomography images via a novel deep domain adaptation method. *J. Biophotonics* **2021**, *14*, e202100096. [[CrossRef](#)]
12. Srinivasan, P.P.; Kim, L.A.; Mettu, P.S.; Cousins, S.W.; Comer, G.M.; Izatt, J.A.; Farsiu, S. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* **2014**, *5*, 3568–3577. [[CrossRef](#)]
13. Rasti, R.; Rabbani, H.; Mehridehnavi, A.; Hajizadeh, F. Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Trans. Med. Imaging* **2017**, *37*, 1024–1034. [[CrossRef](#)]
14. Zhu, Y.; Zhuang, F.; Wang, J.; Ke, G.; Chen, J.; Bian, J.; Xiong, H.; He, Q. Deep subdomain adaptation network for image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 1713–1722. [[CrossRef](#)] [[PubMed](#)]
15. Lee, D.-H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the Workshop on Challenges in Representation Learning, ICML, Atlanta, GA, USA, 16–21 June 2013; p. 896.
16. 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, Lille, France, 6–11 July 2015; pp. 97–105.
17. Ganin, Y.; Lempitsky, V. Unsupervised Domain Adaptation by Backpropagation. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; Volume 37, pp. 1180–1189.
18. Sun, B.; Saenko, K. Deep coral: Correlation alignment for deep domain adaptation. In Proceedings of the Computer Vision—ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–10 October 2016; pp. 443–450.
19. Torralba, A.; Efros, A.A. Unbiased look at dataset bias. In Proceedings of the Conference on Computer Vision and Pattern Recognition 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 1521–1528.
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
21. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
22. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
23. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A survey on bias and fairness in machine learning. *ACM Comput. Surv.* **2021**, *54*, 115. [[CrossRef](#)]
24. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
25. Cascante-Bonilla, P.; Tan, F.; Qi, Y.; Ordonez, V. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; pp. 6912–6920.
26. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
27. Manzari, O.N.; Ahmadabadi, H.; Kashiani, H.; Shokouhi, S.B.; Ayatollahi, A. MedViT: A robust vision transformer for generalized medical image classification. *Comput. Biol. Med.* **2023**, *157*, 106791. [[CrossRef](#)] [[PubMed](#)]
28. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
29. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
30. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
31. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
32. Sundas, A.; Badotra, S.; Bharany, S.; Almogren, A.; Tag-ElDin, E.M.; Rehman, A.U. HealthGuard: An Intelligent Healthcare System Security Framework Based on Machine Learning. *Sustainability* **2022**, *14*, 11934. [[CrossRef](#)]
33. Vanschoren, J. Meta-learning: A survey. *arXiv* **2018**, arXiv:1810.03548.
34. Vettoruzzo, A.; Bouguelia, M.-R.; Vanschoren, J.; Rögnvaldsson, T.; Santosh, K. Advances and Challenges in Meta-Learning: A Technical Review. *arXiv* **2023**, arXiv:2307.04722.
35. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **2020**, *53*, 63. [[CrossRef](#)]

36. Song, Y.; Wang, T.; Cai, P.; Mondal, S.K.; Sahoo, J.P. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Comput. Surv.* **2023**, *55*, 271. [[CrossRef](#)]
37. Stacke, K.; Eilertsen, G.; Unger, J.; Lundström, C. Measuring domain shift for deep learning in histopathology. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 325–336. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.