frontiers | Frontiers in Physiology

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fphys.2022.918929

Multi-Model Domain Adaptation for Diabetic Retinopathy Classification

Guanghua Zhang[1,2†], *Bin Sun*[3†], *Zhaoxia Zhang*[3†], *Jing Pan*[4], *Weihua Yang*[5*] and *Yunfang Liu*[6*]

[1]Department of Intelligence and Automation, Taiyuan University, Taiyuan, China, [2]Graphics and Imaging Laboratory, University of Girona, Girona, Spain, [3]Shanxi Eye Hospital, Taiyuan, China, [4]Department of Materials and Chemical Engineering, Taiyuan University, Taiyuan, China, [5]Affiliated Eye Hospital, Nanjing Medical University, Nanjing, China, [6]The First Affiliated Hospital of Huzhou University, Huzhou, China

Diabetic retinopathy (DR) is one of the most threatening complications in diabetic patients, leading to permanent blindness without timely treatment. However, DR screening is not only a time-consuming task that requires experienced ophthalmologists but also easy to produce misdiagnosis. In recent years, deep learning techniques based on convolutional neural networks have attracted increasing research attention in medical image analysis, especially for DR diagnosis. However, dataset labeling is expensive work and it is necessary for existing deep-learning-based DR detection models. For this study, a novel domain adaptation method (multi-model domain adaptation) is developed for unsupervised DR classification in unlabeled retinal images. At the same time, it only exploits discriminative information from multiple source models without access to any data. In detail, we integrate a weight mechanism into the multi-model-based domain adaptation by measuring the importance of each source domain in a novel way, and a weighted pseudo-labeling strategy is attached to the source feature extractors for training the target DR classification model. Extensive experiments are performed on four source datasets (DDR, IDRiD, Messidor, and Messidor-2) to a target domain APTOS 2019, showing that MMDA produces competitive performance for present state-of-the-art methods for DR classification. As a novel DR detection approach, this article presents a new domain adaptation solution for medical image analysis when the source data is unavailable.

Keywords: diabetic retinopathy classification, multi-model, domain adaptation, convolutional neural network, deep learning

OPEN ACCESS

Edited by:
Wei Wang,
Huzhou Maternity and Child Health Care Hospital, China

Reviewed by:
Xuefei Song,
Shanghai Ninth People's Hospital, China
Somayah Abdullah Albaradei,
King Abdullah University of Science and Technology, Saudi Arabia

***Correspondence:**
Yunfang Liu
panlele0701@139.com
Weihua Yang
benben0606@139.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

Received: *12 April 2022*
Accepted: *12 May 2022*
Published: *01 July 2022*

Citation:
Zhang G, Sun B, Zhang Z, Pan J, Yang W and Liu Y (2022) Multi-Model Domain Adaptation for Diabetic Retinopathy Classification. Front. Physiol. 13:918929. doi: 10.3389/fphys.2022.918929

1 INTRODUCTION

Diabetic retinopathy (DR) is a complication of diabetic patients and a significant cause of blindness globally among the working population (Antonetti et al., 2021). There are 451 million suffering from DR in the world, and this is projected to increase to 639 million in 2045 (Cho et al., 2018). In diabetics, blood is provided to all retina layers through micro blood vessels that are sensitive to unrestricted blood sugar levels. DR may cause no symptoms or only mild vision problems at first, but it can cause blindness eventually. When substantial glucose or fructose is collected in the blood, blood vessels begin to collapse due to insufficient oxygen supply to the cells. Occlusion in these blood vessels can cause serious eye damage. As a result, metabolic rate decreases, and abnormal blood vessels accumulate in DR (Dai et al., 2021). Microaneurysms (MAs) are the early signs of DR, which

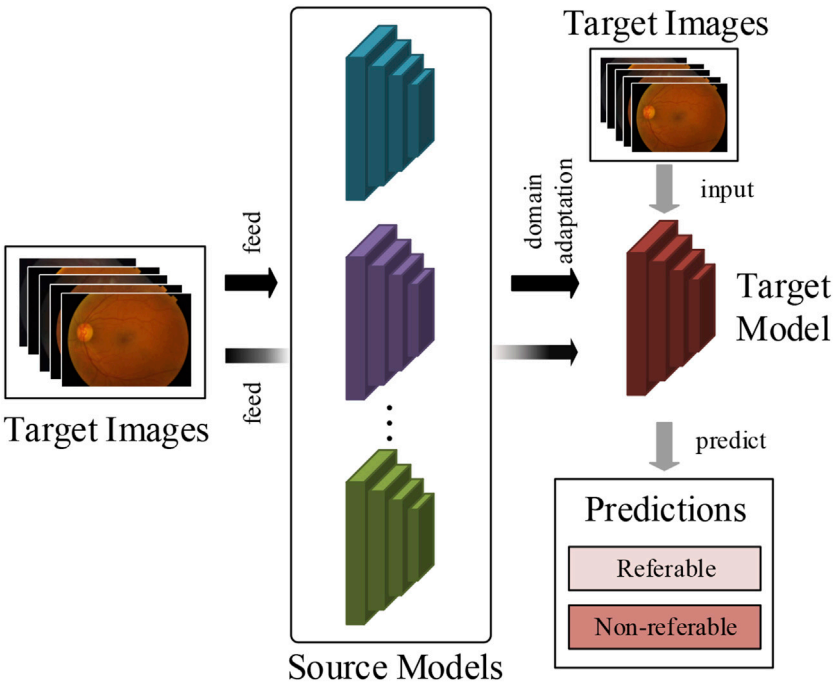

FIGURE 1 | The work flow of our method. We train the target prediction model to simply use pre-trained multiple source models and unlabeled target retinal images.

cause changes in the size (swelling) of the blood vessels. Moreover, hemorrhages (HMs), exudates (EXs), and abnormal blood vessel growth are the symptoms of DR. The International Clinical Diabetic Retinopathy (ICDR) scale is one of the most commonly used clinical scales and is composed of five levels of DR: normal, mild, moderate, severe and proliferative (Bodapati et al., 2021). Generally, diabetic retinopathy is divided into referable diabetic retinopathy (RDR) and non-referable diabetic retinopathy (NRDR).

Blindness can be completely avoided by early diagnosis. Annual regular clinical examination for diabetics is strongly recommended, especially for middle-aged and older adults (Mohamed et al., 2007; Ferris, 1993). Nevertheless, researchers find that a considerable number of people with diabetes failed to have annual eye examinations due to very mild symptoms, long examination time, and a shortage of ophthalmologists (Owsley et al., 2006; MacLennan et al., 2014; Chou et al., 2014). Therefore, it is necessary to adopt automatic DR diagnosis methods to lighten the workload on eye specialists and shorten the detection time, making patients understand the condition and get treatment in time.

Artificial intelligence (AI) is a popular technique for computer-aided automatic DR diagnosis to overcome these obstacles and deep learning has achieved progress in biomedical image analysis (Meng et al., 2021b; Preston et al., 2021; Meng et al., 2021a). Yoo and Park (2013) utilized ridge, elastic net, and LASSO to perform validation on 1052 DR patients. Roychowdhury et al. (2013) proposed a novel two-step approach for DR detection, where non-lesions or normal images are rejected in the first step, and bright and red lesions are classified as hard exudates and hemorrhages, respectively in the second step. In addition to the machine learning methods, the deep learning method becomes very popular in DR screening in recent years. For instance, Vo and Verma (2016) used a deep neural network improved upon GoogLeNet and VGGNet for DR recognition, aiming to learn fine-grained features of retinal images. Moreover, He et al. (2020) combined two attention blocks with a backbone network to solve the imbalanced DR data distribution problem and capture more detailed lesion information, respectively. Ai et al. (2021) proposed an algorithm adopting deep ensemble learning and attention mechanism to detect DR. However, both traditional machine learning methods (Yoo and Park, 2013; Roychowdhury et al., 2013) and supervised deep learning methods (Vo and Verma, 2016; He et al., 2020; Mohamed et al., 2021; Ai et al., 2021) require a large amount of labeled retinal images to train their models, which fail to new data from other domains. As an effective solution, domain adaptation always requires source data, which is usually difficult to access in practical applications because of the strict privacy rules in medical image management agencies.

To tackle this critical problem in supervised deep learning methods, this article attempts to develop a multi-model domain adaptation (MMDA) to conduct transfer learning for DR classification without access to source data. As shown in **Figure 1**, the proposed method can sufficiently utilize the knowledge of source models and unlabeled target images to improve the DR detection performance.

In the MMDA framework, the target model is initially parameterized, and the trained source models are provided. We propose a model weight determination module to estimate

the importance of each source model by measuring the average distance between two retinal feature groups extracted from the source models and target model. This module is optimized by a weight determination loss to output realistic model weights in target feature learning. By using the weights of source models, the pseudo label of the target images is obtained in a feature-level clustering-based way. Finally, we optimize the target model by cross-entropy loss and information maximization loss to guarantee the performance of diabetic retinopathy detection.

To evaluate the performance of MMDA, we conduct extensive experiments on five publicly available retinal image datasets: DDR, IDRiD, Messidor, Messidor-2, and APTOS 2019, obtaining excellent performance without access to source data. The results demonstrate that our proposed method can effectively complete the DR diagnosis task with only unlabeled target data.

2 MATERIALS AND METHODS

2.1 Data Acquisition

In order to validate our method for diabetic retinopathy diagnosis, we trained four source models from publicly available datasets (DDR, IDRiD, Messidor, and Messidor-2) and employed APTOS 2019 as the target domain.

DDR dataset (Diabetic retinopathy detection, 2015) involves 12,522 fundus images from a 45° field of view. In detail, it has 6,266 normal fundus images and 6,256 abnormal samples. Moreover, the class distribution of the dataset is imbalanced in that the normal images are more than the abnormal data.

The IDRiD dataset (Porwal et al, 2018) contains 516 fundus images which were captured by an ophthalmologist from an Indian eye clinic. It provides adequate quality and clinically relevant fundus images with ground truths.

The Messidor dataset (Decencière et al., 2014) is a publicly available diabetic retinopathy dataset provided by the Messidor program partners, which consists of 1,200 retinal images, and for each image, two grades, retinopathy grade, and risk of macular edema, are provided.

The Messidor-2 dataset (Decencière et al., 2014) has been globally used by researchers for DR detection algorithm analysis, which is an extension of Messidor. It contains 1,748 retinal images of 874 examinations. Although there are no official annotations for this dataset, the third-party grades for 1,744 out of the 1,748 images adjudicated by a panel of three retina specialists are available for researchers (Messidor-2 dr grades, 2018).

The APTOS 2019 dataset (Khalifa et al., 2019) is the most recent publicly available Kaggle dataset from the APTOS Blindness Detection competition on Kaggle for DR detection. It contains 3,662 labeled fundus photography images.

The above datasets are graded into five stages from 0 to 4 for no DR, mild DR, moderate DR, severe DR, and proliferative DR, respectively, according to the ICDR severity scale. The label distribution of the datasets and the division of the referable and non-referable DR are shown in **Table 1**. Moreover, the APTOS 2019 dataset is regarded as the target domain, and the other four datasets are used as source datasets to train source models.

TABLE 1 | Label distributions of DDR, IDRiD, Messidor, Messidor-2, and APTOS 2019 datasets.

Dataset	Type	Non-referable			Referable	
		No	Mild	Moderate	Severe	Proliferative
DDR	Source	6,266	630	4,477	236	913
IDRiD	Source	168	25	168	93	62
Messidor	Source	546	153	247	254	—
Messidor-2	Source	1,017	270	347	75	35
APTOS 2019	Target	1,805	370	999	193	295

2.2 Data Preprocessing

When collecting retinal images, the differences in lighting conditions and camera types may cause a large data inconsistency (Graham, 2015). Data preprocessing mitigates noise and enhances image details, reducing inconsistency and playing a significant role in improving performance.

In order to eliminate these negative effects and make data consistent, we perform data preprocessing in the following two steps (**Figure 2**):

Step 1: Resize and crop.

Considering various resolutions of retinal images in different datasets, we resize all images to 1,024 pixels if their width or height is bigger than that size. Then, we crop as much of the black space by identifying the center and radius of the circle in the retinal images.

Step 2: Image enhancement.

In DR detection, the observation of hard exudates, hemorrhages, and cotton wools is significant for eye specialists to diagnose. However, the variations of brightness and resolution not only make ophthalmologists produce misdiagnoses but also make it difficult for a model to compose robust features. To address this problem, we perform image enhancement after resizing and cropping by the following formula:

$$I_o(x, y; \sigma) = \lambda G(x, y; \sigma) * I(x, y) + \omega I(x, y) + \delta \quad (1)$$

where $I(x, y)$ denotes the input retinal image, $G(x, y; \sigma)$ is a Gaussian filter with standard deviation σ, "$*$" represents the convolution operator. λ, ω and δ are manually adjusted variables. In our study, λ, ω and δ are set to 4, -4 and 128, respectively. By improving image contrast with **Eq. 1**, the lesion area is easier to distinguish.

2.3 Multi-Model Domain Adaptation Architecture

This subsection elaborates on our proposed MMDA method, which aims to address the central problem that the labeled image data cannot always be obtained in automatic DR detection.

2.3.1 Overview

Domain adaptation is one of the branches of transfer learning in computer science. For a vanilla multi-source unsupervised domain adaptation task, we have n source domains with fundus images, and N_s^i labeled samples $\{x_i^j, y_i^j\}_{j=1}^{N_s^i}$ from the ith source domain are given, where $x_i^j \in \mathcal{X}_i$, $y_i^j \in \mathcal{Y}_i$, and also N_t unlabeled retinal images $\{x_t^j\}_{j=1}^{N_t}$

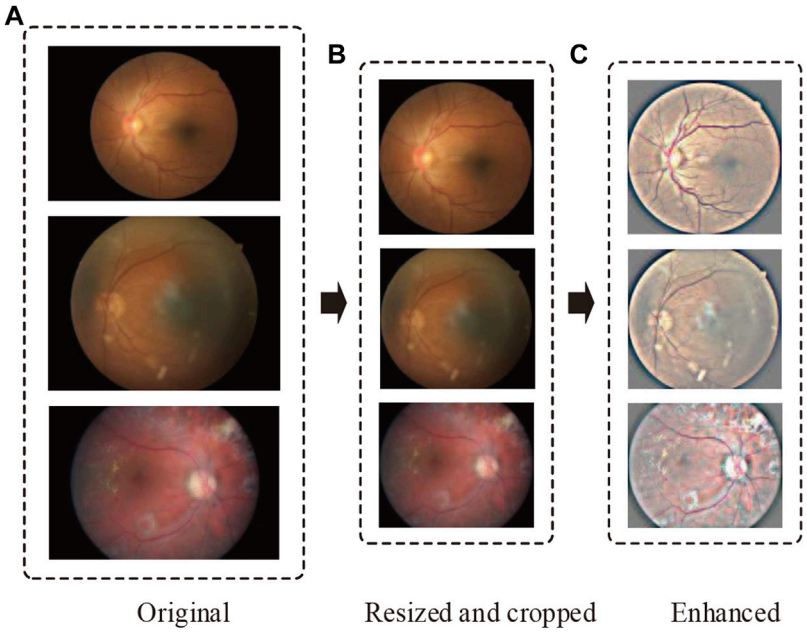

FIGURE 2 | Representative retinal images adopting our preprocessing techniques. From top to bottom, the representative images are sampled from no DR, moderate DR, and proliferative DR, respectively. The parts **(A–C)** denote the original, resized and cropped, and enhanced retinal images.

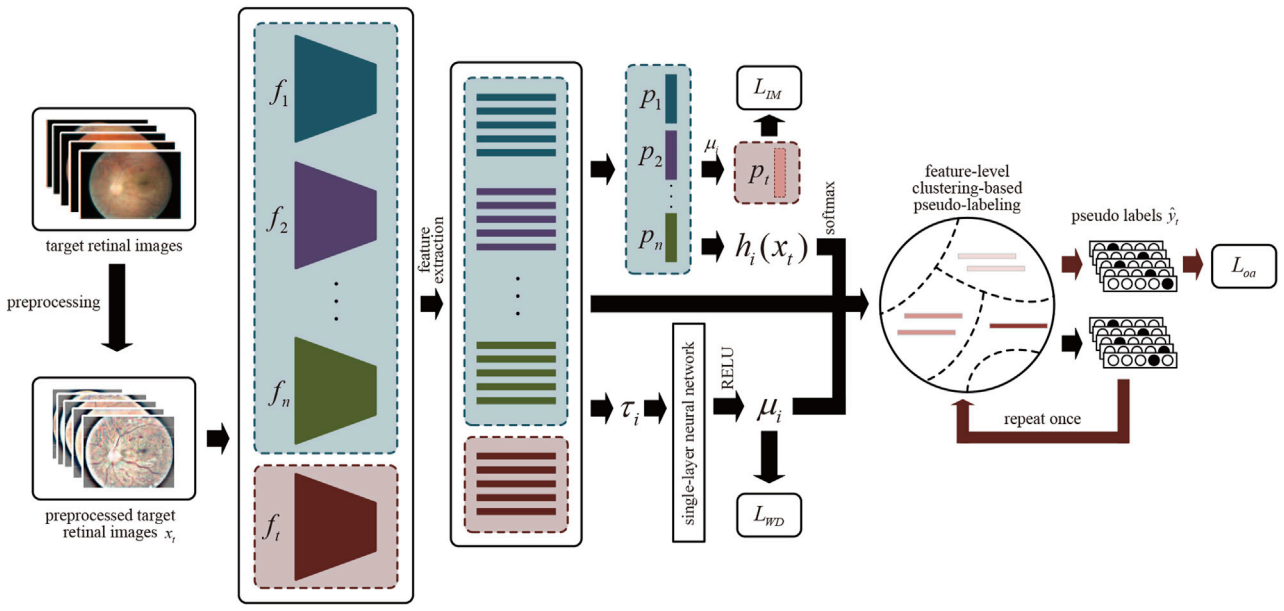

FIGURE 3 | The overview of MMDA architecture. After preprocessing, we obtain the features of target retinal images by source models $f_i|_{i=1}^n$ and target model f_t, and calculate the weight of each model μ_i using the single-layer neural network. The output of the target classifier is defined by the source classifiers with fixed parameters. Pseudo labels $\hat{y}_t$ for each retinal image x_t are obtained after the process of feature-level clustering-based pseudo-labeling.

from the target domain $\mathcal{D}_t$ where $x_t^j \in \mathcal{X}_t$. Domain adaptation aims to obtain a target model to predict the labels $\{y_t^j\}_{j=1}^{N_t}$, where $y_t^j \in \mathcal{Y}_t$. Here, the goal of MMDA is to learn a target prediction model for function h_t: $\mathcal{X}_t \to \mathcal{Y}_t$ and infer $\{y_t^i\}_{i=1}^{N_t}$, with only $\{x_t^i\}_{i=1}^{N_t}$ and the source prediction models for function: h_i: $\mathcal{X}_i \to \mathcal{Y}_i$ available. Note that, only trained source models can be utilized, without access to

their data. **Figure 3** illustrates the overview of our MMDA model for referable DR detection.

Suppose that we have multiple trained source models for DR classification and an ImageNet pre-trained target model. Each model contains two modules: the feature encoding module f_i: $\mathcal{X}_i \to \mathbb{R}^d$ and the classifier module p_i: $\mathbb{R}^d \to \mathbb{R}^K$, i.e., $h_i(x) = p_i\,(f_i(x))$. Here, d is the dimension of the feature, and K represents the number of categories. We extract the features of retinal images using the source and target feature encoding modules firstly. A single-layer neural network is integrated to determine the weights of each source model using euclidean distance and weight determination loss. By employing the weights and source classifiers, target prediction logits are obtained. Then the pseudo labels $\hat{y}_t$ of the target images are generated in a feature-level clustering-based way. Finally, the whole network is optimized using cross-entropy loss and information maximization loss to make the target feature encoding module has excellent DR diagnosis capability.

2.3.2 Source Model Generation

We consider producing several source backbone pre-trained source models, i.e., $h_i = p_i \circ f_i$ ($i = 1, 2, \cdots n$), by optimizing them using the following cross-entropy loss:

$$L^i_{src}\left(f_i; \mathcal{X}_i, \mathcal{Y}_i\right) = -\mathbb{E}_{\left(x_i, y_i\right)\in\mathcal{X}_i\times\mathcal{Y}_i} \sum_{k=1}^{K} q_k \log\, \sigma_k\left(h_i\left(x_i\right)\right) \quad (2)$$

where $\sigma_k(a) = \frac{\exp(a_k)}{\sum_i \exp(a_i)}$ denotes the kth element in the softmax output of a K-dimensional vector a, and q is the one-hot encoding of y_i where q_k is set to "1" if y_i is the kth class and the rest is set to "0". In order to learn more discriminative feature representations and further enhance the following target data alignment, we adopt the label smoothing technique because it prevents the model from becoming over-confident thus improving generalization and performance (Müller et al., 2019). With label smoothing, the objective loss function is modified as below:

$$L^i_{src,ls}\left(f_i; \mathcal{X}_i, \mathcal{Y}_i\right) = -\mathbb{E}_{\left(x_i, y_i\right)\in\mathcal{X}_i\times\mathcal{Y}_i} \sum_{k=1}^{K} q_{k,ls} \log\, \sigma_k\left(h_i\left(x_i\right)\right) \quad (3)$$

where $q_{k,ls} = (1 - \alpha)q_k + \alpha/K$ represents the smoothed label and α is the smoothing factor which is set to 0.1 experientially.

2.3.3 Information Maximization Loss for Target Model

Due to the source classifier modules encoding the distribution information of unseen source data, our framework is proposed to learn the domain-specific feature encoding module while the source classifier modules are fixed. Specifically, MMDA employs the weighted source classifier modules during the target model learning process:

$$p_t(\cdot) = \sum_{i=1}^{n} \mu_i p_i(\cdot) \quad (4)$$

where μ_i is the weight for the ith source model, which will be explained in detail in the following subsection.

In essence, our goal is to obtain an optimal target feature extractor f_t: $\mathcal{X}_t \to \mathbb{R}^d$ for target retinal images, in order that the extracted target features can match source distributions well. However, it is noteworthy that the source images are not accessible in our study. As a result, there's no way to perform feature-level alignment since it is unfeasible to estimate the source distribution in the absence of source data. We look at the problem from a different angle that the expected output logits of the target model should seem like one-hot encoding but differ from each other if the domain gaps are mitigated. To this end, we employ the information maximization (IM) loss (Hu et al., 2017), which enhance the certainty and diversity of target outputs. Specifically, we optimize f_t by IM loss L_{IM} that consists of two objective functions L_{ce} and L_{div}:

$$\begin{aligned} L_{ce}\left(f_t; \mathcal{X}_t\right) &= -\mathbb{E}_{x_t\in\mathcal{X}_t} \sum_{k=1}^{K} \sigma_k\left(h_t\left(x_t\right)\right) \log\, \sigma_k\left(h_t\left(x_t\right)\right) \\ L_{div}\left(f_t; \mathcal{X}_t\right) &= \sum_{k=1}^{K} P_k \log\, P_k = D_{KL}\left(P, \frac{1}{K}\mathbf{1}_K\right) - \log\, K \\ L_{IM} &= Lce + \beta L_{div} \end{aligned} \quad (5)$$

where $h_t(x) = p_t\,(f_t(x))$ is the K-dimensional output logits of each retinal images, $\mathbf{1}_K$ is an all-ones vector with K elements, and $P = \mathbb{E}_{x_t\in\mathcal{X}_t}[\sigma(f_t(x_t))]$ represents the average output probabilities of the whole target domain, β is the balance factor. Information maximization would work better than conditional entropy minimization (Grandvalet and Bengio, 2005) commonly used in traditional domain adaptation works, since it can circumvent the trivial solution where all unlabeled fundus images have the same one-hot encoding via the fair diversity-promoting objective function L_{div}.

2.3.4 Model Weight Determination

In the MMDA framework, a robust target feature encoder is learned by bridging the domain gap between each source domain and the target domain. However, the feature discrepancies between each source domain and target domain are different. To measure the feature discrepancies, we propose a Model Weight Mechanism (MWM). Precisely, we first calculate the average Euclidean Distance between the ith source domain and the target domain:

$$\tau_i = \frac{1}{N_t} \sum_{j=1}^{N_t} \left| f_i\left(x_j\right) - f_t\left(x_j\right) \right| \quad (6)$$

The closer the τ_i is, the more important the source model, i.e., the greater the weight is. To this end, we integrate a single-layer neural network, which is parameterized by a weight vector $w = (w_1, w_2, \cdots w_d) \in \mathbb{R}^{d\times 1}$. Formally, we learn model weight μ_i as

$$\mu_i = \frac{\exp\left(\mathrm{ReLU}\left(w^T\tau_i\right)\right)}{\sum_{j=1}^{N} \exp\left(\mathrm{ReLU}\left(w^T\tau_j\right)\right.} \quad (7)$$

where $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$ is an activation function, which guarantees the nonnegativity of μ_i. The role of the above softmax

operation is to guarantee the model weight satisfy the following property:

$$\sum_{i=1}^{n} \mu_i = 1, \mu_i \geq 0 \tag{8}$$

We optimize the weight vector of the single-layer neural network w by minimizing the following loss function:

$$L_{WD}(w; \mathcal{X}_t) = \frac{1}{N_t} \sum_{i=1}^{n} \sum_{j=1}^{N_t} \mu_i \left\| f_i(x_j) - f_t(x_j) \right\|_2^2 \tag{9}$$

That is, a larger distance $\|f_i(x_j) - f_t(x_j)\|_2$ between features $f_i(x)$ and $f_t(x)$ enforces a smaller value μ_i.

2.3.5 Feature-Level Clustering-Based Pseudo-Labeling

The role of IM loss in **Eq. 5** is to enforce the similarity of one-hot encoding output. Therefore, an accurate prediction network is crucial to reduce this impact. For this purpose, a pseudo-labeling strategy at the feature level is applied for better supervision during the adaptation process.

First, the weighted features centroid of target retinal images for each class is obtained, similar to weighted k-means clustering.

$$m_k^{(0)} = \frac{\sum_{i=1}^{n} \sum_{x_t \in \mathcal{X}_t} \mu_i \sigma_k(h_i(x_t)) f_i(x_t)}{\sum_{x_t \in \mathcal{X}_t} \sigma_k(h_i(x_t))} \tag{10}$$

These centroids can robustly and more reliably characterize the distribution of different categories within the target domain. Then, the pseudo labels can be attained via the nearest centroid classifier:

$$\hat{y}_t = \arg\min_k D_c\left(f_t(x_t), m_k^{(0)}\right) \tag{11}$$

where $D_c(a, b) = 1 - \frac{a \cdot b}{\|a\| \|b\|}$ represents the cosine distance between vector a and b.

Based on generated $\hat{y}_t$ previously , new centroids $m_k^{(1)}$ and pseudo labels are computed:

$$\begin{aligned} m_k^{(1)} &= \frac{\sum_{i=1}^{n} \sum_{x_t \in \mathcal{X}_t} 1_{[\hat{y}_t = k]} \mu_i f_i(x_t)}{\sum_{x_t \in \mathcal{X}_t} 1_{[\hat{y}_t = k]}} \\ \hat{y}_t &= \arg\min_k D_c\left(f_t(x_t), m_k^{(1)}\right) \end{aligned} \tag{12}$$

We refer to $\hat{y}_t$ in **Eq. 12** as the final pseudo labels.

To sum up, given n source models $h_i = p_i \circ f_i$ $(i = 1, 2, \cdots n)$ and the final pseudo labels $\hat{y}_t$ generated from **Eq. 12**, MMDA fixes the parameters of sources classifiers, $p_t(\cdot) = \sum_{i=1}^{n} \mu_i p_i(\cdot)$ and optimizes the feature extractor f_t with the overall loss as:

$$L_{oa}(f_t) = L_{IM} - \gamma \mathbb{E}_{(x_t, \hat{y}_t) \in \mathcal{X}_t \times \hat{\mathcal{Y}}_t} \sum_{k=1}^{K} 1_{[k=\hat{y}_t]} \log \sigma_k(p_t(f_t(x_t))) \tag{13}$$

where $\gamma > 0$ is a balancing hyper-parameter. The whole implementation of MMDA model is shown in Algorithm 1.

Algorithm 1. Pseudo-code of MMDA training process

Input: Pre-trained n source models $p_i \circ f_i (i = 1, 2, \cdots n)$, ImageNet pre-trained target model f_t, unlabeled retinal images $\mathcal{X}_t$, β, γ, batch size B, learning rates ζ
Output: target model f_t
Initialize parameters w of the single-layer neural network;
for $epoch = 1$ to N **do**
  Randomly sample x_t of size B from $\mathcal{X}_t$
  for each mini-batch **do**
    Extract features $f_t(x_t)$;
    Calculate L_{IM} and update f_t using Adam optimizer;
    if starting adaptation **then**
      Extract features $f_t(x_t)$ and $f_i(x_t)$ from f_t and $f_i (i = 1, 2, \cdots n)$;
      Calculate μ_i via Eq. (6), Eq. (7), Eq. (9) and update w via L_{WD};
      Calculate L_{IM} via Eq. (4) and Eq. (5);
      Obtain pseudo labels $\hat{y}_t$ via Eq. (10) to Eq.(12);
      Calculate overall loss (Eq. (13)) by obtained pseudo labels;
      Update f_t using Adam optimizer.
    end if
  end for
end for

2.4 Implementation Details

In the experiments we first train the source models by corresponding source retinal datasets, and each model is designed following ResNet-50 (He et al., 2016). As for the target model, it also employs ResNet-50, initialized by pre-trained parameters from ImageNet (Deng et al., 2009). We perform data argumentation by applying random horizontal flips, vertical flips, and random rotation to prevent overfitting. The input size of the MMDA is 224 × 224. We trained 40 epochs for all the source models using the Stochastic Gradient Descent (SGD) optimization algorithm (Kingma and Ba, 2014) with a learning rate decay factor of $1e^{-4}$. The learning rates for DDR, IDRiD, Messidor, and Messidor-2 datasets are 5×10^{-3} equally. For the target training, we adopt a mini-batch SGD with momentum 0.9, weight decay $1e^{-4}$, and learning rate $\zeta = 1e^{-2}$. The balance factor for IM loss β and the overall loss γ are set to 0.3 and 0.3, respectively. In addition, a batch size of 64 is set for the entire experimental process. MMDA is implemented on two NVIDIA RTX 2080Ti GPUs with 2 × 11 GB RAM using the PyTorch framework.

To validate the effectiveness of MWM (**Section 2.3.4**), we adjust μ_i in **Eqs 4**, **10**, **12**. Moreover, hyper-parameters β in L_{IM} and γ are fine-tuned to analyze their influence on DR detection performance. Details are described in **Section 3.3** and **Section 3.4**. Note that we do not integrate model weight mechanism and pseudo-labeling into the target model until training the target model several epochs with IM loss. That means we attach the model weight and pseudo-labeling modules when the target model has a certain diagnosing capability.

3 RESULTS

3.1 Evaluation Metrics

To measure the performance of the MMDA model, we employ accuracy and sensitivity as the measurements. The accuracy can be defined as the percentage of correctly classified images. Sensitivity measures the ability of a test to correctly identify samples with referable DR, which is an effective metric to measure the DR diagnosis.

This metric is calculated as follows. First, we compute the accuracy by $\frac{TP+TN}{TP+FP+TN+FN}$, where TP is the correctly predicted positive samples, TN denotes the correctly predicted negative images, and FP represents the false predicted positive samples and FN means the false predicted negative images. For the sensitivity, it follows the formula,

TABLE 2 | Accuracy and sensitivity of MMDA for diabetic retinopathy diagnosis compared with state-of-the-art supervised learning approaches on the APTOS 2019 dataset.

Method	Accuracy (%)	Sensitivity (%)
Xie et al. (2017)	92.8	86.8
Vives-Boix and Ruiz-Fernández (2021)	94.5	90.0
Narayanan et al. (2020)	98.4	98.9
Farag et al. (2022)	97.0	97.0
MMDA (Ours)	90.6	98.5

TABLE 3 | The DR classification results of MMDA with different backbones on the APTOS 2019 dataset.

Backbone	Method	Accuracy	Sensitivity	δ Accuracy (%)
VGG-16	ACDA	0.873	0.965	↓2.8
	APDA	0.882	0.973	↓1.9
	MMDA	**0.901**	**0.980**	-
ResNet-50	ACDA	0.880	0.972	↓ 2.6
	APDA	0.902	0.960	↓ 0.4
	MMDA	**0.906**	**0.985**	-

The best results are in bold.

TABLE 4 | DR classification results using different β on the APTOS 2019 dataset.

β	0.1	0.2	0.3	0.4	0.5
Accuracy	0.899	0.903	**0.906**	0.902	0.851
Sensitivity	0.984	0.962	**0.985**	0.987	0.986

The best results are in bold.

TABLE 5 | DR classification results using different γ on the APTOS 2019 dataset.

γ	0.1	0.2	0.3	0.4	0.5
Accuracy	0.877	0.896	**0.906**	0.905	0.902
Sensitivity	0.976	0.982	**0.985**	0.966	0.963

The best results are in bold.

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

3.2 Performance Compared With Supervised Learning Methods

We first compare MMDA with the existing supervised learning methods on the APTOS2019 dataset. Specifically, Xie et al. (2017) present a simple, highly modularized network architecture for image classification, which is often employed in DR detection; Vives-Boix and Ruiz-Fernández (2021) conducted automated detection of DR by directly interfering in both learning and memory by reinforcing less common occurrences during the learning process; Narayanan et al. (2020) proposed a hybrid machine learning architecture to detect and grade the level of diabetic retinopathy; Farag et al. (2022) proposed an automatic deep-learning-based model for severity detection by utilizing a single color fundus photograph. From **Table 2**, it is observed that MMDA achieves approving results with 90.6% accuracy and 98.5% sensitivity. Our method performs a relatively excellent accuracy compared with the compared methods, which only remains a distance of 2.2%, and it presents the second-best sensitivity of 98.5%, only weaker than Narayanan et al. (2020). Although the results of MMDA are lower than these supervised learning methods, huge amounts of labeled data are essentially required in their training process. In contrast, we train MMDA simply to utilize unlabeled retinal images and obtain satisfactory performance, showing the superiority of our framework for DR diagnosis.

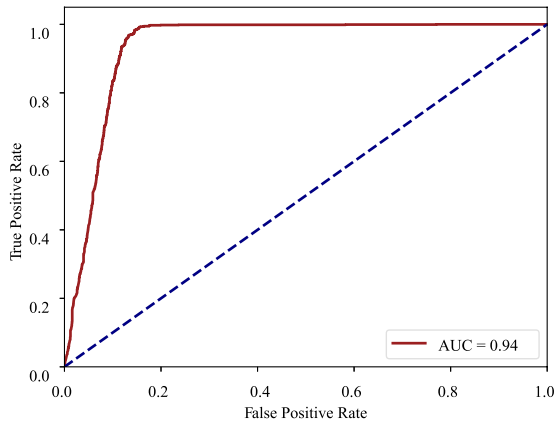

FIGURE 4 | ROC curve of DR diagnosis on the APTOS 2019 dataset.

3.3 Performance Analysis on Model Weight Mechanism

We design a novel model weight mechanism (MWM) to assign a learnable weight to each model. To verify the effect of the MWM, we perform ablation studies to analyze the MWM for the source classifier modules and the pseudo-labeling process using different backbones.

MWM for source classifier modules: We fix the source classifier modules, so we can fully utilize the source distribution information in the modules when the source data is not available. Meanwhile, the discrepancy between each source domain and the target domain cannot be ignored. Specifically, we integrate the weight mechanism into the classifiers by **Eq. 4**. To verify the effect of MWM in utilizing the source distribution information, we conduct a study using Average-weighted Classifier Multi-model Domain Adaptation (ACDA), which is the MMDA model with $\mu_i = \frac{1}{n}\,(i = 1, 2, \cdots n)$ in **Eq. 4**. As shown in **Table 3**, we obtain accuracy and sensitivity of 0.873 and 0.965 for VGG-16, 0.880 and 0.972 for RestNet-50, with accuracy drops of 2.8%, 2.6% for VGG-16 and ResNet-50, respectively, which demonstrates that MWM contributes huge effectiveness on multi-model source distribution learning.

MWM for pseudo-labeling: To evaluate the contribution of MWM in features centroid determination, we carry out an experiment named APDA, which is a modified MMDA with $\mu_i = \frac{1}{n}\,(i = 1, 2, \cdots n)$ in **Eqs 10–12**. With this setting, this model performs at accuracies of 0.882 and 0.902 for VGG-16 and ResNet-50

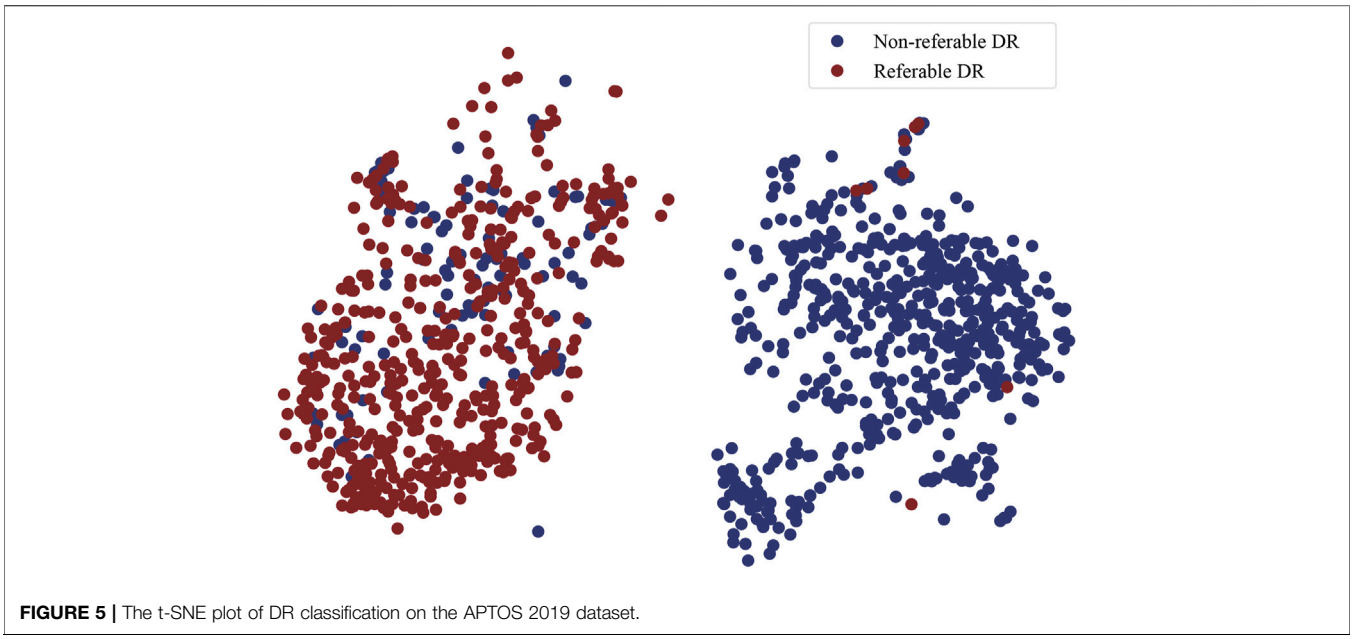

FIGURE 5 | The t-SNE plot of DR classification on the APTOS 2019 dataset.

respectively. All results are lower than the original MMDA model. This is because the average model weight cannot determine accurate centroids of features, which results in incorrect pseudo labels and failing to bridge the domain gap between each source domain and target domain. With MWM, the importance of each source model can be determined, which helps to obtain a more accurate pseudo label and improve the model performance.

3.4 Performance Analysis on Hyper-Parameters

To further validate the effectiveness of each component in MMDA, we explore the influence of hyper-parameters on the performance of our model.

The choice of β in L_{IM}: β is a balancing factor that adjusts the contribution of fair diversity-promoting objective L_{div}. The DR classification performance of MMDA with β from 0.1 to 0.5 is shown in **Table 4**. As reported, both accuracy and sensitivity is improved with the increase of β, and MMDA achieves the best performance when β is set to 0.3. However, when we further increase the value of β, the results start to decrease. We consider that the high β value weakens the effect of L_{ce}, which leads the decision boundary to go through the high-density region.

The choice of γ: γ is a balancing factor of the information maximization loss L_{IM} and the cross-entropy loss in the overall loss L_{oa}. In this section, we investigate the effectiveness of this hyper-parameter. The results shown in **Table 5** demonstrate that MMDA achieves the highest effectiveness when γ is set to 0.3. The cross-entropy loss in the overall loss L_{oa} acts as a guide of the target model. If γ is too small, the effect of the pseudo labels is reduced. If γ is too large, the generalization of the target model will be limited. In order to learn more discriminative features in the target domain and enhance the DR diagnosis performance of the model, it is necessary to adjust the best value of γ.

3.5 Visual Analysis of Model Performance

Furthermore, in order to prove the superiority of the MMDA framework for practical applications, the ROC curve, and t-SNE plot are adopted to visualize our model.

ROC curve: The receiver operating characteristic curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. In the ROC curve, the closer the apex of the curve toward the upper left corner, the greater the discriminatory ability of the test. The ROC curve of MMDA for diabetic retinopathy classification is drawn in **Figure 4**, which obtains the area under the ROC curve of 0.94 and is above the diagonal and close to the point in the upper left corner, demonstrating that MMDA has a satisfying prediction performance.

t-SNE plot: t-Distributed Stochastic Neighbor Embedding (t-SNE) is a technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional data. It maps high-dimensional data to two or more dimensions suitable for human observation. In order to validate the effectiveness of MMDA, we perform a t-SNE plot using the target image features extracted by the trained target feature encoding module (f_t). As shown in **Figure 5**, retinal images of non-referable DR and referable DR are well separated, because MMDA can learn discriminative features to detect referable diabetic retinopathy. The relatively clear boundaries in **Figure 5** suggest that it is practical to train a robust prediction model using MMDA in the absence of labeled target data.

4 DISCUSSION

Retinal images are usually used to build an automatic diabetic retinopathy diagnosis system (Gardner et al., 1996; Acharya et al.,

2009; Ram et al., 2010; Gulshan et al., 2016; Lam et al., 2018; Jiang et al., 2019; Preston et al., 2021). However, whether using traditional machine learning methods (Gardner et al., 1996; Acharya et al., 2009; Ram et al., 2010) or deep supervised learning methods (Gulshan et al., 2016; Lam et al., 2018; Jiang et al., 2019; Preston et al., 2021), they all need a large amount of labeled data during training. In the biomedical image analysis field, labeling work is expensive and the privacy issue is highly sensitive. To tackle this challenge, we consider developing an unsupervised method that the DR diagnosis performance is excellent but labeled retinal images are unnecessary.

In this article, we present a novel MMDA that incorporates model weight mechanism into the MMDA technique. MMDA can be trained in an end-to-end manner with merely unlabeled target retinal images for DR classification. To the best of our knowledge, MMDA is the first attempt to automatically diagnose diabetic retinopathy by adopting an unsupervised domain adaptation technique with multiple source models. The main advantage of this article is that the MMDA can learn helpful knowledge only from source models without any source data, which can relieve the limitation of data privacy from different medical agencies.

Our proposed MMDA method aims to exploit the source knowledge and relationship between the source models and the target model, instead of learning from labeled retinal images directly, thus helping protect the patients' privacy and no need to label images.

In order to fully explore the discrepancy between each source domain and target domain, we propose a model weight mechanism. By incorporating the mechanism into the source classifiers and feature-level clustering-based pseudo-labeling process, the diagnosis performance of the target model is improved.

Extensive experiments and ablation studies on the APTOS2019 dataset demonstrate that MMDA achieves competitive DR diagnosis performance in comparison with state-of-the-art supervised learning methods. However, the DR classification performance still has a distance from the advanced supervised methods due to the discrepancy between source and target models, especially for the invalid access to source data.

Model visualizations (**Figures 4**, **5**) suggest that non-referable (grade 0, 1) and referable cases (from grade 2 to 4) can be diagnosed well. We will focus on the fine-grained classification of the DR grading task (Zheng et al., 2017) in the future.

5 CONCLUSION

When incorporating deep learning techniques in the automatic DR diagnosis system, time-consuming labeling work and privacy issues are critical problems. The present study is designed to exploit existing models and unlabeled retinal images for DR diagnosis to resolve these issues. Ablation studies show the effectiveness of our proposed modules, and the comparison with state-of-the-art supervised learning approaches demonstrates the superiority of our method. Moreover, model visualization indicates that our method can effectively diagnose non-referable and referable cases, with excellent diagnosing results.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

GZ, BS, and ZZ contributed to conception and design of the study. JP organized the database and performed the statistical analysis. GZ and BS wrote the first draft of the manuscript. ZZ wrote the revised manuscript. WY and YL wrote sections of the manuscript, and supervised this work. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the Research Funds of Shanxi Transformation and Comprehensive Reform Demonstration Zone (Grant No. 2018KJCX04), the Fund for Shanxi "1331 Project," and the Key Research and Development Program of Shanxi Province (Grant No. 201903D311009). The work was also partially sponsored by the Research Foundation of the Education Bureau of Shanxi Province (Grant No. HLW-20132), the Scientific Innovation Plan of Universities in Shanxi Province (Grant No. 2021L575), the Shanxi Scholarship Council of China (Grant No. 2020-149), the Innovation and entrepreneurship training program for College Students (NO.202111242006X), the Medical Science and Technology Development Project Fund of Nanjing (Grant No. YKK21262), and the Nanjing Enterprise Expert Team Project. This work was also supported by the Medical Big Data Clinical Research Project of Nanjing Medical University, Medical and Health Research Project of Zhejiang Province (2018PY066).

REFERENCES

Acharya, U. R., Lim, C. M., Ng, E. Y. K., Chee, C., and Tamura, T. (2009). Computer-based Detection of Diabetes Retinopathy Stages Using Digital Fundus Images. *Proc. Inst. Mech. Eng. H.* 223, 545–553. doi:10.1243/09544119JEIM486

Ai, Z., Huang, X., Fan, Y., Feng, J., Zeng, F., and Lu, Y. (2021). DR-IIXRN : Detection Algorithm of Diabetic Retinopathy Based on Deep Ensemble Learning and Attention Mechanism. *Front. Neuroinform* 15, 778552. doi:10.3389/fninf.2021.778552

Antonetti, D. A., Silva, P. S., and Stitt, A. W. (2021). Current Understanding of the Molecular and Cellular Pathology of Diabetic Retinopathy. *Nat. Rev. Endocrinol.* 17, 195–206. doi:10.1038/s41574-020-00451-4

Bodapati, J. D., Shaik, N. S., and Naralasetti, V. (2021). Deep Convolution Feature Aggregation: an Application to Diabetic Retinopathy Severity Level Prediction. *SIViP* 15, 923–930. doi:10.1007/s11760-020-01816-y

Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., and Ohlrogge, A. W. (2018). Idf Diabetes Atlas: Global Estimates of Diabetes Prevalence for 2017 and Projections for 2045. *Diabetes Res. Clin. Pract.* 138, 271–281. doi:10.1016/j.diabres.2018.02.023

Chou, C.-F., Sherrod, C. E., Zhang, X., Barker, L. E., Bullard, K. M., Crews, J. E., et al. (2014). Barriers to Eye Care Among People Aged 40 Years and Older with Diagnosed Diabetes, 2006–2010. *Diabetes care* 37, 180–188. doi:10.2337/dc13-1507

Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., et al. (2021). A Deep Learning System for Detecting Diabetic Retinopathy across the Disease Spectrum. *Nat. Commun.* 12, 1–11. doi:10.1155/2021/9943067

Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., et al. (2014). Feedback on a Publicly Distributed Image Database: the Messidor Database. *Image Analysis Stereology* 33, 231–234. doi:10.5566/ias.1155

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: A Large-Scale Hierarchical Image Database," in 2009 IEEE conference on computer vision and pattern recognition (IEEE), 248–255. doi:10.1109/cvpr.2009.5206848

Diabetic retinopathy detection (2015). *Diabetic Retinopathy Detection*.

Farag, M. M., Fouad, M., and Abdel-Hamid, A. T. (2022). Automatic Severity Classification of Diabetic Retinopathy Based on Densenet and Convolutional Block Attention Module. *IEEE Access*. doi:10.1109/access.2022.3165193

Ferris, F. L. (1993). How Effective Are Treatments for Diabetic Retinopathy? *Jama* 269, 1290–1291. doi:10.1001/jama.269.10.1290

Gardner, G. G., Keating, D., Williamson, T. H., and Elliott, A. T. (1996). Automatic Detection of Diabetic Retinopathy Using an Artificial Neural Network: a Screening Tool. *Br. J. Ophthalmol.* 80, 940–944. doi:10.1136/bjo.80.11.940

Graham, B. (2015). *Kaggle Diabetic Retinopathy Detection Competition Report*. Coventry: University of Warwick.

Grandvalet, Y., and Bengio, Y., (2005). Semi-supervised Learning by Entropy Minimization. *CAP*, 281–296.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama* 316, 2402–2410. doi:10.1001/jama.2016.17216

He, A., Li, T., Li, N., Wang, K., and Fu, H. (2020). Cabnet: Category Attention Block for Imbalanced Diabetic Retinopathy Grading. *IEEE Trans. Med. Imaging*. 40, 143–153. doi:10.1109/TMI.2020.3023463

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778. doi:10.1109/cvpr.2016.90

Hu, W., Miyato, T., Tokui, S., Matsumoto, E., and Sugiyama, M. (2017). Learning Discrete Representations via Information Maximizing Self-Augmented Training*Int. Conf. Mach. Learn.*, 1558–1567.

Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., and Qian, W. (2019). "An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 2045–2048. doi:10.1109/embc.2019.8857160

Khalifa, N. E. M., Loey, M., Taha, M. H. N., and Mohamed, H. N. E. T. (2019). Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection. *Acta Inform. Medica* 27, 327. doi:10.5455/aim.2019.27.327-332

Kingma, D. P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*

Lam, C., Yi, D., Guo, M., and Lindsey, T. (20182018). Automated Detection of Diabetic Retinopathy Using Deep Learning. *AMIA summits Transl. Sci. Proc.*, 147.

Müller, R., Kornblith, S., and Hinton, G. (2019). When Does Label Smoothing Help? *arXiv preprint arXiv:1906.02629*

MacLennan, P. A., McGwin, G., Searcey, K., and Owsley, C. (2014). A Survey of alabama Eye Care Providers in 2010–2011. *BMC Ophthalmol.* 14, 1–10. doi:10.1186/1471-2415-14-44

Meng, Y., Zhang, H., Gao, D., Zhao, Y., Yang, X., Qian, X., et al. (2021a). "Bi-gcn: Boundary-Aware Input-dependent Graph Convolution Network for Biomedical Image Segmentation," in The British Machine Vision Conference. doi:10.48550/arXiv.2110.14775

Meng, Y., Zhang, H., Zhao, Y., Yang, X., Qiao, Y., MacCormick, I. J., et al. (2021b). Graph-based Region and Boundary Aggregation for Biomedical Image Segmentation. *IEEE Trans. Med. Imaging* 41, 690–701. doi:10.1109/TMI.2021.3123567

Messidor-2 dr grades (2018). *Messidor-2 Dr Grades*.

Mohamed, E., Abd Elmohsen, M., and Basha, T. (2021). "Improved Automatic Grading of Diabetic Retinopathy Using Deep Learning and Principal Component Analysis," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE), 3898–3901. doi:10.1109/embc46164.2021.9630919

Mohamed, Q., Gillies, M. C., and Wong, T. Y. (2007). Management of Diabetic Retinopathy: a Systematic Review. *Jama* 298, 902–916. doi:10.1001/jama.298.8.902

Narayanan, B. N., Hardie, R. C., De Silva, M. S., and Kueterman, N. K. (2020). Hybrid Machine Learning Architecture for Automated Detection and Grading of Retinal Images for Diabetic Retinopathy. *J. Med. Imaging* 7, 034501. doi:10.1117/1.JMI.7.3.034501

Owsley, C., McGwin, G., Scilley, K., Girkin, C. A., Phillips, J. M., and Searcey, K. (2006). Perceived Barriers to Care and Attitudes about Vision and Eye Care: Focus Groups with Older African Americans and Eye Care Providers. *Investigative Ophthalmol. Vis. Sci.* 47, 2797–2802. doi:10.1167/iovs.06-0107

Porwal, P., Kokare, R., and Pachade, S. (2018). Indian Diabetic Retinopathy Image Dataset (Idrid). *Data* 18, 25. doi:10.3390/data3030025

Preston, F. G., Meng, Y., Burgess, J., Ferdousi, M., Azmi, S., Petropoulos, I. N., et al. (2021). Artificial Intelligence Utilising Corneal Confocal Microscopy for the Diagnosis of Peripheral Neuropathy in Diabetes Mellitus and Prediabetes. *Diabetologia*, 1–10. doi:10.1007/s00125-021-05617-x

Ram, K., Joshi, G. D., and Sivaswamy, J. (2010). A Successive Clutter-Rejection-Based Approach for Early Detection of Diabetic Retinopathy. *IEEE Trans. Biomed. Eng.* 58, 664–673. doi:10.1109/TBME.2010.2096223

Roychowdhury, S., Koozekanani, D. D., and Parhi, K. K. (2013). Dream: Diabetic Retinopathy Analysis Using Machine Learning. *IEEE J. Biomed. health Inf.* 18, 1717–1728. doi:10.1109/JBHI.2013.2294635

Vives-Boix, V., and Ruiz-Fernández, D. (2021). Diabetic Retinopathy Detection through Convolutional Neural Networks with Synaptic Metaplasticity. *Comput. Methods Programs Biomed.* 206, 106094. doi:10.1016/j.cmpb.2021.106094

Vo, H. H., and Verma, A. (2016). "New Deep Neural Nets for Fine-Grained Diabetic Retinopathy Recognition on Hybrid Color Space," in 2016 IEEE International Symposium on Multimedia (ISM) (IEEE), 209–215. doi:10.1109/ism.2016.0049

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). "Aggregated Residual Transformations for Deep Neural Networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 1492–1500. doi:10.1109/cvpr.2017.634

Yoo, T. K., and Park, E.-C. (2013). Diabetic Retinopathy Risk Prediction for Fundus Examination Using Sparse Learning: a Cross-Sectional Study. *BMC Med. Inf. Decis. Mak.* 13, 1–14.

Zheng, H., Fu, J., Mei, T., and Luo, J. (2017). "Learning Multi-Attention Convolutional Neural Network for Fine-Grained Image Recognition," in Proceedings of the IEEE international conference on computer vision, 5209–5217. doi:10.1109/iccv.2017.557

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Sun, Zhang, Pan, Yang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.