Computers in Biology and Medicine 155 (2023) 106631

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Triple-DRNet: A triple-cascade convolution neural network for diabetic retinopathy grading using fundus images

Muwei Jian [a,b,*], Hongyu Chen [a,1], Chen Tao [a], Xiaoguang Li [c,**], Gaige Wang [d]

[a] School of Information Science and Technology, Linyi University, Linyi, China
[b] School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China
[c] Faculty of Information Tecnology, Beijing University of Technology, Beijing, China
[d] School of Computer Science and Technology, Ocean University of China, Qingdao, China

ARTICLE INFO

Keywords:
Diabetic retinopathy grading
Fundus images
Triple-DRNet
Attention mechanism

ABSTRACT

Diabetic Retinopathy (DR) is a universal ocular complication of diabetes patients and also the main disease that causes blindness in the world wide. Automatic and efficient DR grading acts a vital role in timely treatment. However, it is difficult to effectively distinguish different types of distinct lesions (such as neovascularization in proliferative DR, microaneurysms in mild NPDR, etc.) using traditional convolutional neural networks (CNN), which greatly affects the ultimate classification results. In this article, we propose a triple-cascade network model (Triple-DRNet) to solve the aforementioned issue. The Triple-DRNet effectively subdivides the classification of five types of DR as well as improves the grading performance which mainly includes the following aspects: (1) In the first stage, the network carries out two types of classification, namely DR and No DR. (2) In the second stage, the cascade network is intended to distinguish the two categories between PDR and NPDR. (3) The final cascade network will be designed to differentiate the mild, moderate and severe types in NPDR. Experimental results show that the ACC of the Triple-DRNet on the APTOS 2019 Blindness Detection dataset achieves 92.08%, and the QWK metric reaches 93.62%, which proves the effectiveness of the devised Triple-DRNet compared with other mainstream models.

1. Introduction

Diabetic Retinopathy (DR) has been regarded as a high incidence complication of diabetes; it is responsible for visual damage in diabetes patients, and even resulted in permanent blindness in severe cases [1]. Nowadays, the precise classification of DR is paid more attention in clinical investigations, because it can distinguish different stages of DR, so as to select the appropriate therapeutic schedule.

The severity of DR is able to categorize according to the number, size and type of retinal surface lesions in fundus images, such as microaneurysms, bleeding, exudates, neovascularization, etc. According to the International Clinical DR Classification System in 2003 [2,3], the clinical diagnosis of DR Grnerally is separated into the following stages: normal, non-proliferative and proliferative, in which non-proliferative is classified as mild, moderate, and severe. The typical examples of fundus images of five types of severity are shown in Fig. 1. In the non-pathological stage, the patient had no obvious retinopathy as displayed in Fig. 1 (a). In the pathological stage, the earliest DR stage is mild Non-Proliferative Diabetic Retinopathy (NPDR), which is characterized by Microaneurysm (MA) caused by retinal microvascular leakage. Then the severity of DR will further evolve to moderate NPDR, in which MA starts to swell and other lesions (Hemorrhage (HM), Exudate (EX)) caused by MA appear. When severe NPDR develops, the number of MA, HM and EX diffuses on the retinal surface globally. It can be caught sight of Fig. 1 (d) and Fig. 1 (e), from NPDR to Proliferative Diabetic Retinopathy (PDR), an obvious variation is the emergence of neovascularization, which may bring about permanent retinal damage or even blindness. Clinically, different lesions in fundus images are the key to judge the severity of DR.

In recent years, convolutional neural network (CNN) has made great progress, and it is ubiquitous in computer vision filed. The convolution kernel in CNN can efficiently extract the features in the image, making it

[*] Corresponding author. School of Information Science and Technology, Linyi University, Linyi, China.
[**] Corresponding author.
E-mail addresses: jianmuweihk@163.com (M. Jian), lxg@bjut.edu.cn (X. Li).
[1] He contributes equally to this work and shares the first authorship.

https://doi.org/10.1016/j.compbiomed.2023.106631
Received 1 December 2022; Received in revised form 29 January 2023; Accepted 4 February 2023
Available online 7 February 2023

a significant breakthrough in target detection [4,5], semantic segmentation [6,7], image augmentation [40,43], and so on. Because of CNN's strong ability of advanced feature extraction and representation, they are also successfully applied in the field of intelligent medical treatment. For example, Yu et al. [41] proposed a dual branch network to segment nodules in thyroid ultrasound images. The network applied a convolution based soft erase module to expand the foreground response region while constraining the erroneous expansion of the foreground region by the enhancement of background features, with the aim of generating high-quality segmented outputs. Besides, Chen et al. [42] designed convolutional neural networks for iris segmentation and recognition simultaneously. The features extracted by the two networks complement each other in a series connection, thereby achieving the state-of-the-art segmentation and recognition performance on low-quality iris images. With regard to DR classification tasks, Li et al. [8] applied a pre-trained CNN based on transfer learning for DR grading. Specifically, they first load the pre-trained parameters into CNN, which as feature extractors to extract the lesions in the fundus image, and then use support vector machine (SVM) as a classifier to deal with the classification task. In Ref. [9], Chai et al. improved the residual network, which combined with SE block to grade DR according to fundus images and achieved good results.

Although the CNN based DR grading method has achieved impressive outcomes, it is clinical practice in reality is yet challenging owing to the complicacy of the mission. First of all, lots of lesions among fundus images are extraordinarily tiny, consisting of minority of pixels, which are prone to be ignored in the process of feature extraction, resulting in inaccurate classification results. Secondly, there is a visual similarity in the shape and color of the lesions in different types of fundus images, such as microaneurysms and bleeding spots, which is also apt to be confused, thus causing adverse effects on the classification performance. Thirdly, in contrasted to the dataset of billions of samples (e.g. ImageNet [10]), the number of samples in the public fundus image dataset is insufficient, which leads to the serious limitation of the DR classification task and difficulty in achieving satisfactory performance. Finally, among the limited fundus image samples, the number of samples between different categories is extremely imbalanced. The number of No DR samples account for the majority. Imbalanced data distribution will bring about the model focus the major categories.

In view of the above observations, we design a novel cascade network scheme called Triple-DRNet for Diabetic Retinopathy grading. The model is divided into three individual subnetworks: namely DR-Net, PDR-Net and NPDR-Net. The DR-Net is mainly used to distinguish whether there are lesions in the fundus image, that is, the binary classification of DR and No DR. The PDR-Net will further classify fundus images with pathological changes into PDR or NPDR. After that, the NPDR-Net further will discriminate the NPDR into distinct types of mild, moderate and severe. This cascade network model greatly alleviates the first problem found in the above observation. For the second issue, inspired by Ref. [11], we have carried out a multi-scale global pixel modeling on the feature map. This operation is capable of pouring more focus on the similar lesions among the image, and mitigate the problem that lesions are ignored in the feature extraction process to some extent. Meanwhile, we have applied attention gate [12] into our devised model, which can retain useful pathological information in the fundus image while suppressing irrelevant noises effectively. Finally, we introduced the CAB module designed by Ref. [13] in NPDR-Net, which can avoid the problem of category confusion during the DR grading with advantage.

The contributions of this article are summarized as below.

1) A triple-stage cascade network called Triple-DRNet for DR grading is proposed. This network can progressively divide DR/No DR, PDR/NPDR, and Mild/Moderate/Severe NPDR precisely, which effectively alleviates the problem that some categories are difficult to distinguish and the number of samples is imbalanced.
2) We designed two efficient attention modules in the subnetworks and their complementary functions can enable the model to better extract relevant lesions in fundus images as well as avoid missing of small lesions.
3) Extensive experiments have been carried out on APTOS 2019 Blindness Detection dataset. Experiments testify that our strategy has achieved satisfactory outputs and has culminated in state-of-the-art performance in five-class DR grading.

The remaining sections of this article are installed as follows. section 2 analyzes the convolutional neural network design and DR grading. The exploited DR grading Triple-DRNet is depicted in details in section 3. Intensive experiments are conducted to assess the behavior of the Triple-DRNet for DR grading in section 4. Finally, we summarized the Triple-DRNet and discussed the related future orientation in section 5.

2. Related work

Principally, we concisely reviewed the architecture of neural networks in deep learning and the latest research on DR grading.

2.1. CNN in deep learning

CNN is the mainstream network architecture in deep learning. Since the appearance of AlexNet [14], a large number of CNN frameworks have appeared in the domain of multimedia analysis, and how to build an efficient network has become a hotspot in the field. In 2014, a novel CNN architecture, GoogLeNet [15], was designed. GoogLeNet, also known as Inception, uses sparse connections instead of full connections and convolution operations, solving the problem of high computational complexity and easy overfitting of large networks. It is generally believed that there are three elements to improve the network performance: height, width and resolution. Among them, the network depth is the dominant factor to be considered. ResNet proposed by He et al. [16] introduced identity skip connections into the network, thus settling the drawbacks of gradient disappearance and gradient explosion, and raising the network depth factor to a heigh level. Soon after, Xie et al. [17] explored a new converged network ResNext, which combines the calculation mode of ResNet with that of Inception. Specifically, it integrates split transform merge in Inception into the shortcut branch, achieving higher accuracy with the same parameters as ResNet. Another improvement on ResNet comes from the work of Gao et al. [18], which takes the resolution dimension into account in the network architecture. Concretely, it divides the feature maps inside the Residual Block into multiple groups. In addition to the first group of feature maps, each group of feature maps is convolved and added to the previous group's feature maps to obtain multi-scale feature representation. In 2019, Tan

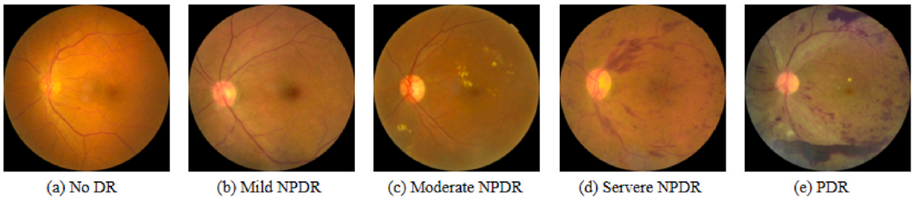
(a) No DR (b) Mild NPDR (c) Moderate NPDR (d) Servere NPDR (e) PDR

Fig. 1. Examples of fundus images at various stages of diabetes retinopathy. (a) no obvious lesions, and the severity of DR increases from (b) to (e).

et al. [19] believed that improving the height, width and resolution of the network alone would lead to the bottleneck of network performance. Therefore, they used the Neural Architecture Search technology to rationalize the above three dimensions, proposed EfficientNet, and achieved perfect performance. In the "2020s" era, Liu et al. [20] analyzed several techniques of the CNN architecture and compared with the design concept of Swin Transformer, providing a huge reference value for the design idea of the CNN architecture.

Throughout the development process of CNN, the proposal of the new architecture is always inseparable from two basic modules, namely, the residual module and the inception module. We believe that the design of the inception module will affect the actual computing efficiency, which is intolerable in medical applications. Meanwhile, the pathological changes in fundus images are very complex, so it is desirable to exploit a deeper network to extract features better. Therefore, in this study, we used ResNet as the backbone of the three cascade subnetworks. At the same time, inspired by the design concept of ConvNext, we replaced some ordinary Conv in ResNet with DWConv, and further reduced the number of parameters without affecting the accuracy of the model so as to promote the efficiency of the model. Finally, we refer to the scaling factor proposed in EfficientNet to balance the height, width and resolution of the devised model as optimal as possible.

2.2. DR grading

The purpose of DR grading is to achieve accurate and automatic staging of DR disease severity. Traditional DR classification methods require devising hand-crafted features, and using general classifiers or their different versions (e.g. support vector machine) to grade DR. For instance, Acharya et al. [21] employed morphological operation to extract visual features of blood vessels, microaneurysms, exudates and bleeding in fundus images, and input them into SVM for classification. Nayak et al. [22] utilized morphological operation and texture analysis to detect hard exudate area, vascular area, contrast and other features, and then these features were input into artificial neural network (ANN) for disease staging with an accuracy of 93%. In Ref. [23], Calleja et al. applied LBP to extract local features and fed them into ANN, SVM, and RF to detect DR. The results show that RF performs best in a dataset of 71 images. The accuracy achieved 97.46%. These methods have shown great potential. However, they rely on prior knowledge and still need to be improved under complex image conditions.

In recent years, deep learning algorithms achieved significant progress in the field of medical image analysis, providing strong support for DR grading. Ting et al. [24] used VGG to filter DR, and achieved advanced performance in 71896 images. Krause et al. [25] employed the Inception V4 to automatically detect diabetes retinopathy in fundus images and achieved excellent outcomes. Bellemo et al. [26] designed two different depth learning models VGG and ResNet to extract features, and integrated the scores of the two different models with the idea of ensemble learning to output the final results. In Ref. [37], Tariq et al. adopted deep transfer learning and combined the classification results of diverse convolutional neural networks for diagnosing Diabetic Retinopathy. The above research shows that the method based on deep learning is effective for DR grading. However, traditional depth-learning methods are difficult to capture small lesions, such as microaneurysms and bleeding spots. Considering this, Chai et al. [9] and He et al. [13] independently introduced attention mechanism to their model. Specifically, attention mechanism can highlight the pathological changes in fundus images that will affect grading, and suppress the background areas irrelevant to classification tasks, so as to ameliorate the performance of the framework. Later, Nneji et al. [38] proposed a weighted fusion framework based on the Inception V3 and VGG16 networks for feature extraction from low-quality fundus images and had achieved excellent Diabetic Retinopathy diacrisis. In Ref. [27], Shi et al. devised EfficientNet in their work, and for the first time adopted multi-stage migration learning method in DR grading via training and fine-tuning the network on multiple datasets, which alleviated the problem of insufficient datasets and reached impressive outcomes. In light of correlation between the left and right eyes of DR patients, Nirthika et al. [39] designed a siamese network based convolutional neural network architecture for DR grading, which can extract the features of patients' left and right eyes to enhance the grading performance of monocular DR. With the appearance of Vision Transformer (ViT) [28,29], Transformer-style model began to be applied to image processing tasks, and DR grading also ushered in a new framework. Sun et al. [30] employed a novel Transformer based model to process fundus images. This model is an encoder-decoder structure, wherein the encoder is utilized to model pixel relationships, and the decoder is a pathological filter. This model can perform DR grading and pathological discovery simultaneously.

The above-mentioned methods have made intensive investigation on the DR grading with deep learning networks. However, most of them have overlooked the special challenge in DR grading task, that is, the different categories may depend on different type of features. In reality, a hierarchical and cascade classification scheme may be more in line with the doctor's workflow.

According to the specific lesions in DR and the hierarchical characteristics of DR classification, we build a cascade Triple-DRNet, which alleviates the imbalance of class samples and effectively improves the performance of classification for confusing specific lesions. In addition, multi-scale global pixel modeling module and attention module are introduced to different stages of the network to further boost the classification performance.

3. Methodology

In this section, we will demonstrate the designed Triple-DRNet in detail, including: (1) The pre-processing module deals with initial images; (2) The three cascade subnets for hierarchical classification tasks; (3) Aggregation module conducts label anastomosing on the results of the triple subnetworks, and finally realizes five-class of DR gradation. Fig. 2 illustrates the workflow of the devised Triple-DRNet and its distinct subnetworks.

3.1. Preprocessing module

Before pre-processing, we organize the labels of the dataset into hierarchal two-category labels and follow three-category labels, and then divide the database into training sets and testing sets with a proportion of 8:2. After that, we separate the original training set and original testing set into three specific training sets T_1, T_2, T_3, and three specific testing sets V_1, V_2, V_3 corresponding to the networks at different stages. In pre-processing phase, we refer to the method designed by Graham [31] to rescale the image so that each image has the same radius. Then, Gaussian blur is applied to the image, subtracted the local average color and mapped the local average to 50% gray. Finally, the fundus part of the image is cut to 90% of the original image size to eliminate boundary effects. This pre-processing scheme is used to eliminate the differences between images caused by different lighting conditions, camera resolutions, etc.

3.2. Triple-DRNet

The Triple-DRNet consists of three cascade subnets, including DR-Net, PDR-Net and NPDR-Net. DR-Net uses T_1 and V_1 as training set and testing set, respectively. The label includes DR/No DR, and this distinct network determines whether DR exists in the fundus image through binary classification. The PDR-Net employs T_2 and V_2 as training set and testing set, accordingly, which will further differentiate the lesions into PDR or NPDR in the images with DR. In the end, the NPDR-Net utilizes T_3 and V_3 to further classify NPDR into mild/moderate/severe types based on the grading results of the previous stage.

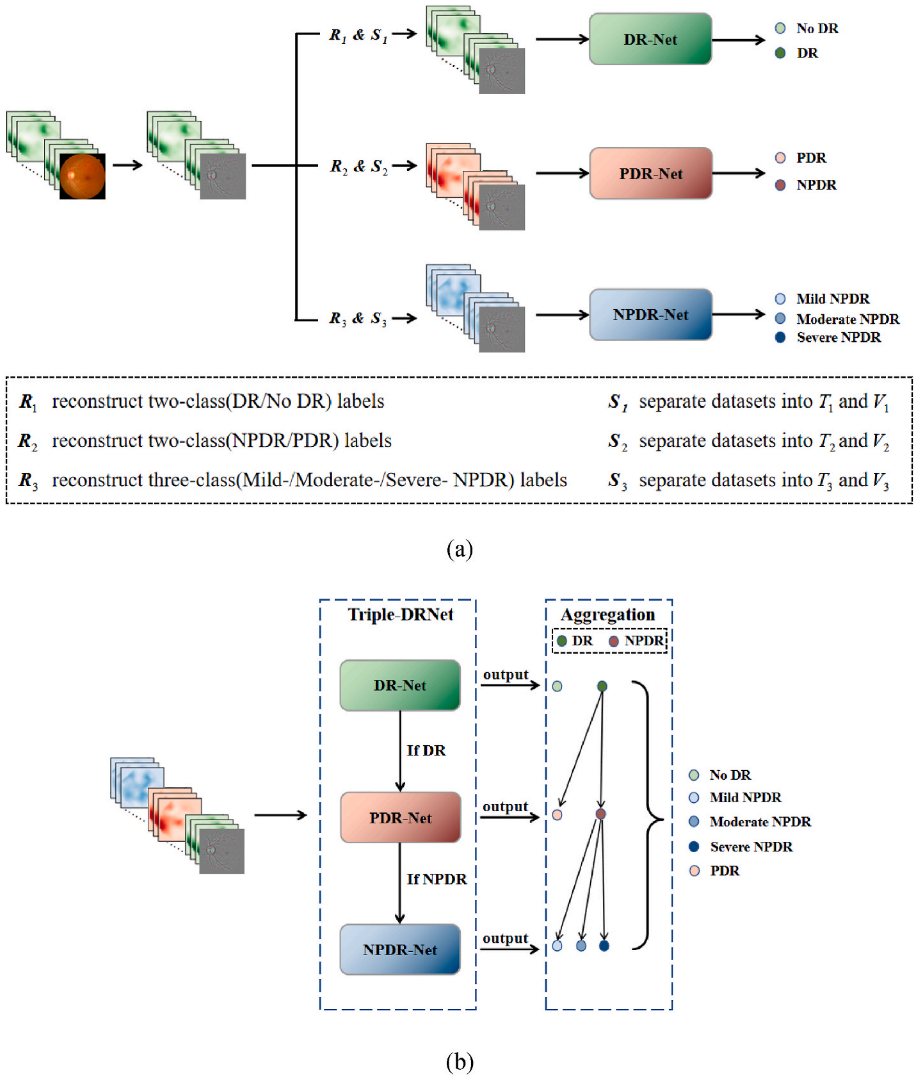
R_1 & S_1
R_2 & S_2
R_3 & S_3
DR-Net
PDR-Net
NPDR-Net
No DR
DR
PDR
NPDR
Mild NPDR
Moderate NPDR
Severe NPDR
R_1 reconstruct two-class(DR/No DR) labels
R_2 reconstruct two-class(NPDR/PDR) labels
R_3 reconstruct three-class(Mild-/Moderate-/Severe- NPDR) labels
S_1 separate datasets into T_1 and V_1
S_2 separate datasets into T_2 and V_2
S_3 separate datasets into T_3 and V_3
(a)
Triple-DRNet
DR-Net
If DR
PDR-Net
If NPDR
NPDR-Net
output
output
output
Aggregation
DR
NPDR
No DR
Mild NPDR
Moderate NPDR
Severe NPDR
PDR
(b)

Fig. 2. Diagram of our Triple-DRNet. In (a), the training and testing phases of the subnetworks are shown, and each subnetwork performs different DR grading tasks. For clarity, the workflow of Triple-DRNet is displayed in (b), and the trained subnetworks are integrated to perform five-class DR grading.

3.2.1. DR-Net module

Based on ResNet [16], we built our DR-Net, which contains a stem, four stages, three down-sample layers and a separate convolution layer, as shown in Fig. 3. Among them, stage 1 and stage 2 adopt residual block, which are recorded as B_r, B_r with three convolution layers. The input image gets the feature map after feature extraction as below:

$$F = f_{1\times1}(\sigma(f_{3\times3}(\sigma(f_{1\times1}(I))))), \tag{1}$$

where σ represents ReLu activation function, $f_{1\times1}$ and $f_{3\times3}$ denote convolution layers using the kernel size of 1×1 and 3×3, respectively. I is the original input image of B_r block. To reduce the number of parameters of the subnet, we have improved B_r in stage 3 and stage 4. Specifically, we follow the improvement method mentioned in Section 2.1 and adopt DWConv instead of ordinary Conv. The improved B_r is recorded as B_r'. After feature extraction of the original image I', the feature map F' is calculated as follows:

$$F' = f'_{1\times1}\left(\sigma\left(f'_{1\times1}\left(\sigma\left(f'_{3\times3}(I')\right)\right)\right)\right), \tag{2}$$

Meanwhile, we built a separate down-sample layer after each stage; it can reduce the scale of the feature maps to enable the subsequent convolution operation with a larger receptive field, so as to further extract high-level features. The specific operations are described as follows:

$$F_l' = BN(f_{2\times2}(F_l)), \tag{3}$$

where $F_l' \in \mathbb{R}^{C'\times H'\times W'}$, $(C' = 2C, H' = H/2)$, $(W' = W/2)$ and $F_l \in \mathbb{R}^{C\times H\times W}$, and BN denotes batch normalization.

To learn discriminative features, we added a single convolution layer after stage 4. Finally, through the global average pooling operation, the fully connected layer and softmax are activated to produce the outputs of the two-class classification:

$$P = SoftMax(FC(GAP(f_{1\times1}(F')))), \tag{4}$$

where $\in \{0, 1\}$, 0 and 1 represent No DR and DR, respectively. *GAP* denotes global average pooling operation, and *FC* is the fully connected layer.

3.2.2. PDR-Net module

Since an important lesion feature of PDR is the appearance of neovessels in the image, in the PDR-Net module, we affiliate the vessel segmentation map to the model as medical priori knowledge for effective classification. Specifically, because the green channel in the fundus image has the highest contrast, we extract the green channel separately for subsequent operations. Next, we perform adaptive histogram

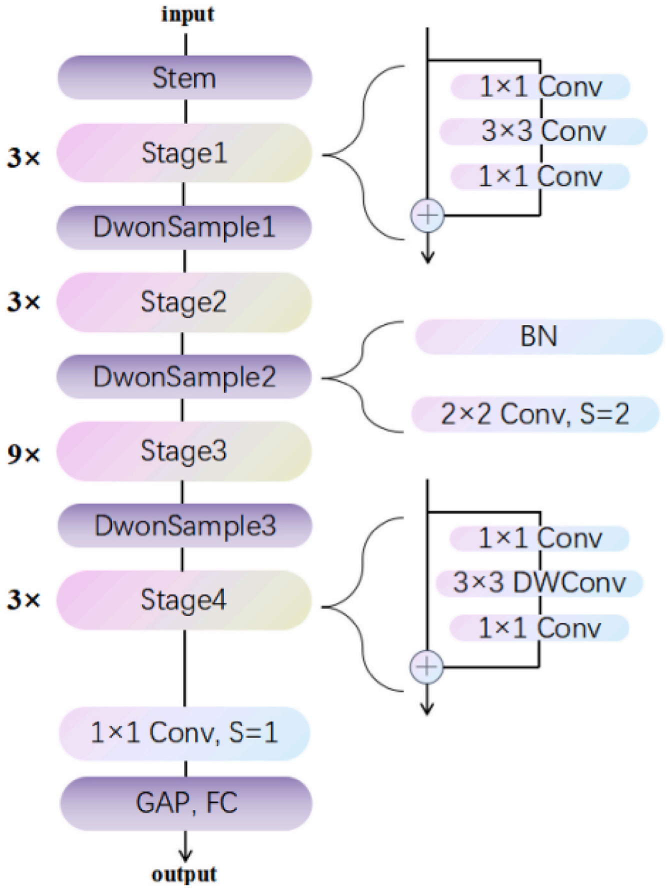

Fig. 3. The structure of the DR-Net module.

equalization on the image to further strengthen the image's contrast. After contrast enhancement to obtain image I_e, we sequentially performed opening operation and closing operation with kernels 5, 11, 23 to produce image I_{oc}:

$$I_{oc} = C_{23\times23}(O_{23\times23}(C_{11\times11}(O_{11\times11}(C_{5\times5}(O_{5\times5}(I_e)))))), \tag{5}$$

where O and C refer to the opening operation and closing operation, respectively. Subtracting I_e from I_{oc} yields a segmentation map, then activating with sigmoid function yields I_s to be augmented to the network as medical prior knowledge:

$$I_s = \text{Sigmoid}(Enhance(I_{oc} - I_e)), \tag{6}$$

We follow the backbone in the DR-Net module, meanwhile, in order to make the model better at extracting feature related to the lesion region of interest and suppressing lesion independent background information, we redesigned a novel attention module inspired by attention gate [10] and swelled it to B_r and $B_r^{'}$, as shown in Fig. 4. In the case of B_r, the two inputs of attention module are the original I (low-level feature map) of B_r and the output F (high-level feature map), respectively. First, I and F are fused by two different 1 × 1 convolutions and element-wise summation, and then ReLU activation is adopted to obtain the fused feature map Q:

$$Q = \sigma\left(f_{1\times1}(I) \oplus f_{1\times1}^{'}(F)\right), \tag{7}$$

where $\oplus$ denotes element-wise summation, and σ represents ReLU activation function.

After acquiring the fused feature map Q, we performed average pooling and maximum pooling operations along the channel dimensions to engender two corresponding feature maps Q_{avg} and Q_{max} with spatial information:

$$g(Q) = \begin{cases} Q_{avg}, g = AvgPool \\ Q_{max}, g = MaxPool \end{cases}. \tag{8}$$

Then feature maps Q_{avg} and Q_{max} are concatenated and reduced dimensions through a convolution layer with the kernel size of 7 × 7, and the final activation map Q_{att} is obtained by Sigmoid activation:

$$Q_{att} = Sigmoid\left(f_{7\times7}\left(\left[Q_{avg}; Q_{max}\right]\right)\right). \tag{9}$$

By element-wise multiplication, the output F_{att} can be derived with the attention:

$$F_{att} = Mul(F, Q_{att}), \tag{10}$$

where $Mul(\bullet)$ denotes element-wise multiplication.

In addition, we draw lessons from the work of Wang et al. [11] and integrate features of different levels with full consideration of pixel remote dependencies. The constructed module is illustrated in Fig. 5. Concretely, we first feed the output $F_{s3} \in \mathbb{R}^{C_3\times H_3\times W_3}$ of stage 3 into two convolution layers with the kernel size of 2 × 2, and also lessen the amount of channels to obtain two sets of characteristics $F_q \in \mathbb{R}^{C_3^{'}\times S}$ and $F_k \in \mathbb{R}^{S\times C_3^{'}}$, where $C_3^{'} = C_3/2$, $S = H_3W_3/4$. Through multiplication operation, the correlation matrix between each pixel is obtained, and the weight matrix $V \in \mathbb{R}^{S\times S}$ can be derived through softmax activation:

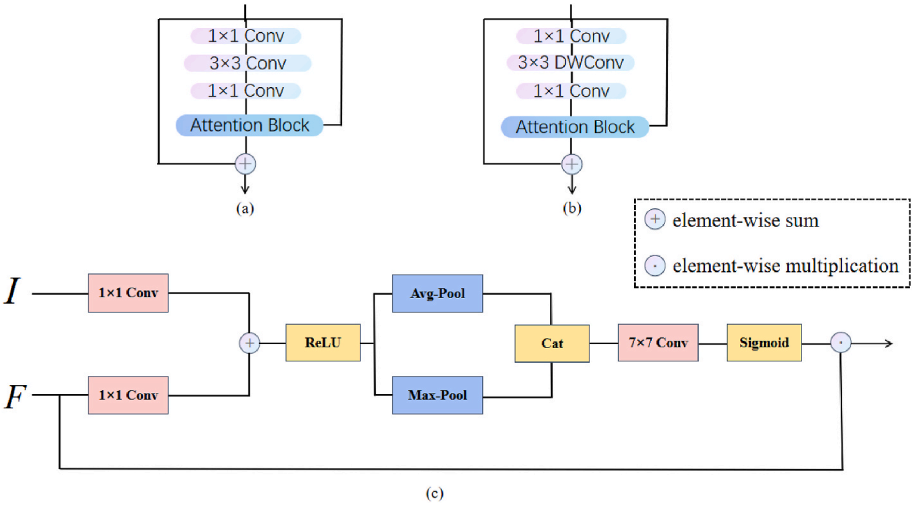

Fig. 4. The structure of the PDR-Net module. The added attention block is shown in (a) and (b), while (c) is the overall structure of the attention block.

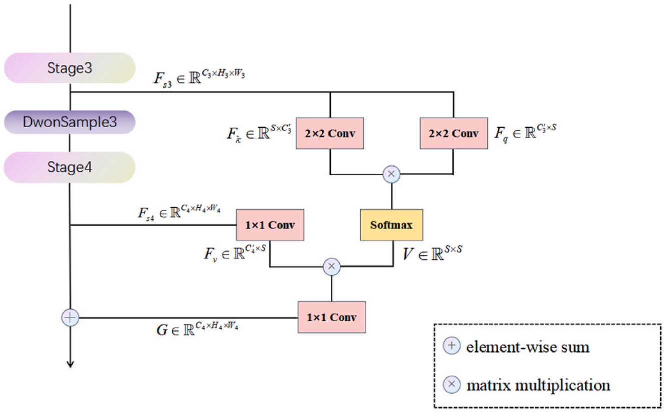
Stage3
$F_{s3} \in \mathbb{R}^{C_3 \times H_3 \times W_3}$
DwonSample3
Stage4
$F_k \in \mathbb{R}^{S \times C_3'}$
2×2 Conv
2×2 Conv
$F_q \in \mathbb{R}^{C_3' \times S}$
$F_{s4} \in \mathbb{R}^{C_4 \times H_4 \times W_4}$
1×1 Conv
Softmax
$F_v \in \mathbb{R}^{C_4' \times S}$
$V \in \mathbb{R}^{S \times S}$
1×1 Conv
$G \in \mathbb{R}^{C_4 \times H_4 \times W_4}$
⊕ element-wise sum
⊗ matrix multiplication

Fig. 5. Global pixel modeling structure with different level features.

$$V = SoftMax\left(F_q \otimes F_k\right), \tag{11}$$

where ⊗ denotes matrix multiplication.

Next, we reduce the dimension of the output $F_{s4} \in \mathbb{R}^{C_4 \times H_4 \times W_4}$ of stage 4 through a 1 × 1 convolution layer to attain $F_v \in \mathbb{R}^{C_4' \times S}$, where $C_4' = C_4/2$ and $H_4W_4 = S = H_3W_3/4$. Through multiplication of F_v and V, we gain the feature map $G \in \mathbb{R}^{C_4' \times S}$ with different levels of features and global information. Finally, we raise the dimension to generate $G \in \mathbb{R}^{C_4 \times H_4 \times W_4}$, and fuse it with F_{s4} to produce the final output. The formula is expressed as follows:

$$output = F_{s4} \oplus f_{1\times 1}(G), \tag{12}$$

where $G = F_v \otimes V$.

The structure of the classification part is the same as DR-Net module, and finally a binary classification result $P_2 \in \{0, 1\}$ is fulfilled, where 0 and 1 represent NPDR and PDR, respectively.

3.2.3. NPDR-Net module

The NPDR-Net module aims to further classify the grade NPDR outputted by the cascade PDR-Net module into three severity grades, involving mild, moderate and severe NPDR. For exploring the detailed variation between categories, we specially adjust the input size of the model to 528 × 528 pixels in the experiments.

Follow the backbones in DR-Net, we introduce the CAB block proposed by Ref. [13] after Stage 4. CAB block is a kind of attention module based on category, which allocates certain amount of feature channels for individual DR severity grade to ensure that individual DR grade has the identical feature channel, which is conducive to avoiding channel deviation and thus preventing category confusion during classification.

The structure of the classification part is as same as in the DR-Net module. Finally, a three-class classification result $P_3 \in \{0, 1, 2\}$ is achieved. Among them, 0, 1, 2 represent mild, moderate and severe NPDR, respectively.

3.3. Aggregation module

All testing set V_{test} is separated to three sub-sets, where $V_{test} \in V_1 \cup V_2 \cup V_3$. Two-class gradation results of DR and No DR are obtained through DR-Net module, DR is further divided into PDR and NPDR in PDR-Net module, while NPDR is further divided into three-class classification results of mild, moderate and severe NPDR in the corresponding NPDR-Net module. Finally, five-class gradation results are accomplished in the aggregation module to realize the five categories of DR grading.

4. Experimental settings

Here, we provided the composition of the database and the selection of evaluation indicators.

4.1. Dataset

A publicly accessible Kaggle competition dataset, APTOS 2019 Blindness Detection dataset [32], is employed to train and test the proposed framework. The APTOS 2019 competition aims to exploit a pragmatic DR detection system, so that more potential patients can be treated timely and accurately. This database has two parts: training set and testing set. The training set includes 3662 fundus images, all of which are jointly labeled by multiple professional doctors. Overall, there are only a few blurred and uneven lighting images. We removed these images, and the remaining 3545 images are used for the experiment. They are marked as five levels (i.e. No DR, Mild NPDR, Moderate NPDR, Severe NPDR, PDR) in database. The tangible division of data samples is list in Table 1.

4.2. Evaluation metrics

For the three distinct subnetworks, we applied accuracy (ACC) to measure their performance. For the five categories of DR gradation, we executed ACC and quadratic weighted Kappa metrics (QWK) [35] to evaluate the superiority of the devised models, and compared with other typical methods. ACC and QWK are defined as below:

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \tag{13}$$

$$QWK = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}, \text{w.r.t } w_{i,j} = \frac{(i-j)^2}{(N-1)^2}, \tag{14}$$

where TP and TN respectively represent the number of correctly classified positive samples and correctly classified negative samples; FP and FN denote the number of negative samples wrongly classified as positive and positive samples wrongly classified as negative, respectively. $O_{i,j}$ is the observed probabilities while $E_{i,j}$ is the expected probabilities, N denotes total number of classes and i and j represent a certain class, respectively.

5. Experimental results

First, we introduced the details of the experiment implementation. Then, through a series of substantial experiments, we verified the effectiveness of the design of Triple-DRNet and its different constituent subnetworks, and compared it with other advanced network models. Finally, we conducted a series of ablation studies to prove the validity of the proposed model.

5.1. Implementation details

The computer platform is equipped with an Intel XEON Platinum 8176 2.1 GHz 28C/56T CPU, and NVIDIA RTX 3090, 24G GPU. All procedures are implemented based on Python 3.8 and PyTorch 1.10.0. The CUDA version applied is 11.3.

In the training period, the three subnets were trained separately. DR-Net is trained for 300 epochs with the AdamW optimizer and the Focal loss [33]. PDR-Net and NPDR-Net have the same training settings as DR-Net except that AdamW is replaced by Adan [34]. Batch size is adapted to 16 for all procedures. The base learning rate is turn to 1e-3 when iteration reduction ceases diminishing.

5.2. Performance of Triple-DRNet and its subnetworks

The advantage of the exploited Triple-DRNet is that it can specifically classify the classes with special lesions, so as to improve the performance of classification. It can be seen from Table 2 that the three

Table 1
The tangible division of data samples in APTOS 2019 Blindness Detection dataset.

Models		Images	No DR		DR		
DR-Net	Train	2749	1354		1395		
	Test	796	451		345		
		Images	NPDR		PDR		
PDR-Net	Train	1395	1173		222		
	Test	345	290		55		
		Images	Mild NPDR		Moderate NPDR	Severe NPDR	
NPDR-Net	Train	1173	278		750	145	
	Test	290	70		180	40	
		Images	No DR	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Triple-DRNet	Test	796	451	70	180	40	55

Table 2
Performance of Triple-DRNet and its subnetworks on APTOS database.

Metric	DR-Net	PDR-Net	NPDR-Net	Triple-DRNet
Accuracy	0.9899	0.8840	0.8020	0.9208

AUC	0.9901 (DR or not)	0.9403 (PDR or NPDR)	0.9480(Mild or not)
			0.9150(Moderate or not)
			0.9354(Severe or not)
				/

# of parameters	19.1 M	21.1 M	21.9 M	61.1 M

cascade subnetworks have achieved promising results in the APTOS database, of which DR-Net has reached 98.99% accuracy; PDR-Net and NPDR-Net have achieved 88.40% and 80.20% accuracy respectively. And the final classification accuracy of the five-class has culminated 92.08%. In addition, we have shown the AUC of each category, which are 0.9901, 0.9480, 0.9150, 0.9354 and 0.9403 respectively.

5.3. Comparison experiments

For the five categories of DR gradation, the Triple-DRNet is evaluated between these representative models, including ResNet-50 [16], ConvNext [20], EfficientNet [19], ResNext [17], Inception-V4 [25], Simple-method [36], SE-ResNet [9] and CBANet [13].

As shown in Table 3, our devised network achieves the best results in terms of ACC and QWK metrics. ResNet and EfficientNet have both excellent feature extraction capabilities, but due to the particularity of DR grading, they cannot accurately judge specific classes. ConvNext is the other advanced CNNs model in recent years and its performance once exceeded that of Transformer-style model. However, own to its use of DWConv to decrease the amount of parameters, the feature reprsentation ability is weakened, resulting in poor performance in DR grading. SE-ResNet [9] and CBANet [13] introduced various attention mechanisms to improve the performance of the backbone, but in essence, they still had some limitations to identify specific lesions in each class. The designed Triple-DRNet contains three cascade subnetworks, and its merit is that each individual subnetwork can focus on distinguishing specific classes, eliminating other interfering factors and reducing the difficulty of overall classification.

Table 3
Comparative results of the Triple-DRNet and other SOTA methods based on the APTOS 2019 Dataset.

Models	Metric	
	ACC	QWK
ResNet-50 [16]	0.8631	0.9075
ConvNext [20]	0.8379	0.8727
EfficientNet [19]	0.8819	0.9081
ResNext [17]	0.8681	0.9218
Inception-V4 [25]	0.7626	0.7880
Simple-method [36]	0.8480	0.9013
SE-ResNet [9]	0.8178	0.8620
CBANet [13]	0.8869	0.9282
Triple-DRNet (ours)	**0.9208**	**0.9362**

Moreover, we introduce different attention modules in the subnetwork, which strengthens the capacity of each subnetwork during five-class DR grading. Fig. 6 displays the detailed results of DR grading of the top three models with the highest accuracy in the comparison experiment, which provided the results of experiments comparing the Triple-DRNet to other top three models in the form of confusion matrix. According to Fig. 6, we can note that compared with the results of the other three models, Triple-DRNet maintains the high accuracy of No DR and Moderate NPDR classification, and the Moderate NPDR has a certain improvement in accuracy. For Mild NPDR and PDR, our model's accuracy has been significantly improved, and the misdiagnosis of Mild NPDR and PDR as Moderate NPDR is evidently reduced. For severe NPDR, Triple-DRNet alleviates the diagnosis of Severe NPDR as Moderate NPDR. In general, our proposed model achieves the state-of-the-art DR grading outputs.

5.4. Ablation study

We performed ablation experiments on two subnetworks, PDR-Net and NPDR-Net, to assess the availability of the designed network and diverse attention mechanisms, as shown in Table 4.

As we can observe from Table 4, all the distinct module devised in the subnetworks can improve their performance progressively. In PDR-Net module, the addition of a blood vessel segmental map can make the model capture neovessels effectively, which is conducive to a specific lesion. The applied attention module can further weaken the useless information in the background while highlighting the distinguishing features. In addition, global pixel modeling can sufficiently avoid missing small lesions in the fundus image and enhancing the model's accuracy. Additionally, in the NPDR-Net module, the implemented CBA procedure can distinguish the different characteristics of diverse classes and frustrate the confusion of classes in the DR grading process.

6. Conclusion and discussion

In this article, we design a novel classification model, called Triple-DRNet, to implement DR grading. Based on the division-and-rule strategy, the Triple-DRNet comprises three cascade subnetworks, which partitions the five categories (i.e. No DR, Mild-, Moderate-, Severe-NPDR and PDR) classification so that each individual subnetwork can focus on classifying a specific class. Moreover, we also apply different attention modules in the cascade subnetwork to ultimately improve the

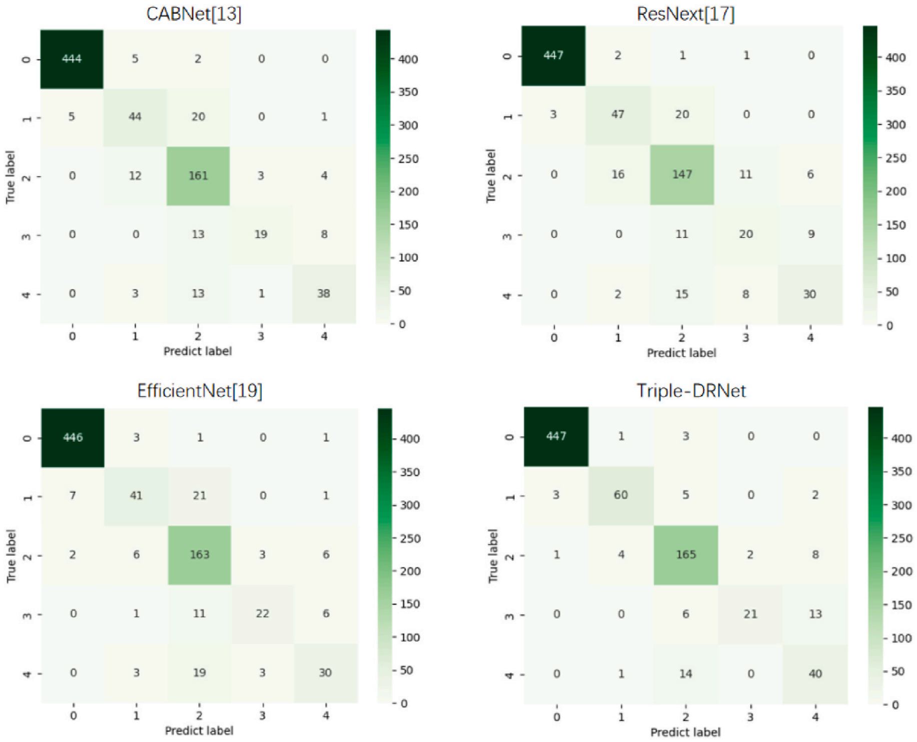

Fig. 6. Confusion matrices comparing the Triple-DRNet with the top three models. From left to right, the first row is CBANet and ResNext, and the second line is EfficientNet and Triple-DRNet, correspondingly.

Table 4
Ablation study on the impact of various modules on the performance of subnetworks.

Models	Accuracy
PDR-Net	0.8629
PDR-Net + vessels	0.8775
PDR-Net + vessels + Attention	0.8823
PDR-Net + vessels + Attention + Global	**0.8840**
NPDR-Net	0.7771
NPDR-Net + CBA Block	**0.8020**

performance of Triple-DRNet in the five categories DR gradation. In the light of the APTOS 2019 Blindness Detection dataset, our developed model is superior to other mainstream models according to both ACC and QWK, which verifies the validity of the exploited framework.

In the experiments, we find that the performance of NPDR-Net in the subnetwork is still has vast promotion space, and the parameters of the whole network can be further compressed, so as to optimize the model lightweight to apply in real clinical applications. In addition, since each subnetwork is trained independently, it is impossible to fuse the features extracted from different subnetworks during the training process, which has certain limitations on improving the performance of the devised model. In the future, solving the limitations of feature fusion and exploring the balance between lightweight vehicles and excellent performance are our main research contents.

Data availability statements

All data generated or analyzed during this study are included in this published article [32].

Compliance with ethical standards

- **Disclosure of potential conflicts of interest**

The authors declare that they have no competing interests.

- **Research involving Human Participants and/or Animals**

The authors declared that they have no involving Human Participants and/or Animals to this work.

Declaration of competing interest

The authors declared that they have no conflicts of interest to this work.

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgment

This work was supported by National Natural Science Foundation of China (NSFC) (61976123, 61601427); Taishan Young Scholars Program of Shandong Province; and Key Development Program for Basic Research of Shandong Province (ZR2020ZD44).

References

[1] N.H. Cho, J.E. Shaw, S. Karuranga, Y. Huang, J.D. da Rocha Fernandes, A. W. Ohlrogge, B. Malanda, IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045, Diabetes Res. Clin. Pract. 138 (2018) 271–281.

[2] C.P. Wilkinson, F.L. Ferris III, R.E. Klein, P.P. Lee, C.D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, Proposed international clinical Diabetic Retinopathy and diabetic macular edema disease severity scales, Ophthalmology 110 (9) (2003) 1677–1682.

[3] N. Tsiknakis, D. Theodoropoulos, G. Manikis, E. Ktistakis, O. Boutsora, A. Berto, K. Marias, Deep learning for Diabetic Retinopathy detection and classification based on fundus images: a review, Comput. Biol. Med. 135 (2021), 104599.

[4] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2016) 1137–1149.

[5] T. Melo, A.M. Mendonça, A. Campilho, Microaneurysm detection in color eye fundus images for Diabetic Retinopathy screening, Comput. Biol. Med. 126 (2020), 103995.

[6] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

[7] Z. Han, M. Jian, G. Wang, ConvUNeXt: an efficient convolution neural network for medical image segmentation, Knowl. Base Syst. 253 (2022), 109512.

[8] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, T. Wang, Convolutional Neural Networks Based Transfer Learning for Diabetic Retinopathy Fundus Image Classification, 2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI). IEEE., 2017, pp. 1–11.

[9] R. Chai, D. Chen, X. Ma, S. Liu, Y. Wang, Y. Wang, Diabetic Retinopathy Diagnosis Based on Transfer Learning and Improved Residual Network, 2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS). IEEE, 2022, pp. 941–946.

[10] J. Deng, W. Dong, R. Socher, L. Li, K. Li, F. Li, Imagenet: A Large-Scale Hierarchical Image Database, 2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009, pp. 248–255.

[11] X. Wang, R. Girshick, A. Gupta, K. He, Non-local Neural Networks, Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7794–7803.

[12] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: learning to leverage salient regions in medical images, Med. Image Anal. 53 (2019) 197–207.

[13] A. He, T. Li, N. Li, K. Wang, H. Fu, CABNet: category attention block for imbalanced Diabetic Retinopathy grading, IEEE Trans. Med. Imag. 40 (1) (2020) 143–153.

[14] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[15] C. Szegedy W. Liu, Y. Jia, Y. Sermanet, P. Reed, S. Anguelov, A. Rabinovich, Going Deeper with Convolutions, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[16] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[17] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated Residual Transformations for Deep Neural Networks, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1492–1500.

[18] S. Gao, M. Cheng, K. Zhao, X. Zhang, M. Yang, P. Torr, Res2net: a new multi-scale backbone architecture, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2) (2017) 652–662.

[19] M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, Int. Conf. Machine Learn. (2019) 6105–6114.

[20] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A Convnet for the 2020s, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.

[21] U.R. Acharya, C.M. Lim, E.Y.K. Ng, C. Chee, T. Tamura, Computer-based detection of diabetes retinopathy stages using digital fundus images, Proc. IME H J. Eng. Med. 223 (5) (2009) 545–553.

[22] J. Nayak, P.S. Bhat, U.R. Acharya, C.M. Lim, M. Kagathi, Automated identification of Diabetic Retinopathy stages using digital fundus images, J. Med. Syst. 32 (2) (2008) 107–115.

[23] J.D.L. Calleja, L. Tecuapetla, A. Medina, E. Bárcenas, A.B. Urbina Nájera, LBP and machine learning for Diabetic Retinopathy detection, in: International Conference on Intelligent Data Engineering and Automated Learning, 2014, pp. 110–117.

[24] D.S.W. Ting, C.Y.L. Cheung, G. Lim, G.S.W. Tan, N.D. Quang, et al., Development and validation of a deep learning system for Diabetic Retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes, JAMA, J. Am. Med. Assoc. 318 (22) (2017) 2211–2223.

[25] J. Krause, V. Gulshan, E. Rahimy, P. Karth, K. Widner, G.S. Corrado, L. Peng, D. R. Webster, Grader variability and the importance of reference standards for evaluating machine learning models for Diabetic Retinopathy, Ophthalmology 125 (8) (2018) 1264–1272.

[26] V. Bellemo, Z.W. Lim, G. Lim, Q.D. Nguyen, Y. Xie, et al., Artificial intelligence using deep learning to screen for referable and vision-threatening Diabetic Retinopathy in Africa: a clinical validation study, Lancet Digital Health 1 (1) (2019) e35–e44.

[27] L. Shi, J. Zhang, Few-shot Learning Based on Multi-Stage Transfer and Class-Balanced Loss for Diabetic Retinopathy Grading, 2021 arXiv preprint arXiv: 2109.11806.

[28] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, et al., An image is worth 16x16 words: transformers for image recognition at scale, ICLR (2021).

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017) 5998–6008.

[30] R. Sun, Y. Li, T. Zhang, Z. Mao, F. Wu, Y. Zhang, Lesion-aware transformers for diabetic retinopathy grading, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10938–10947.

[31] B. Graham, Kaggle Diabetic Retinopathy Detection Competition Report, University of Warwick., 2015, pp. 24–26.

[32] Asia Pacific Tele-Ophthalmology Society, APTOS 2019 Blindness Detection Dataset, 2019.

[33] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal Loss for Dense Object Detection, In Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.

[34] X. Xie, P. Zhou, H. Li, Z. Lin, S. Yan, Adan: Adaptive Nesterov Momentum Algorithm for Faster Optimizing Deep Models, 2022 arXiv preprint arXiv: 2208.06677.

[35] H. Li, X. Dong, W. Shen, F. Ge, H. Li, Resampling-based cost loss attention network for explainable imbalanced Diabetic Retinopathy grading, Comput. Biol. Med. 149 (2022), 105970.

[36] A. Sugeno, Y. Ishikawa, T. Ohshima, R. Muramatsu, Simple methods for the lesion detection and severity grading of Diabetic Retinopathy by image processing and transfer learning, Comput. Biol. Med. 137 (2021), 104795.

[37] H. Tariq, M. Rashid, A. Javed, E. Zafar, S.S. Alotaibi, M.Y.I. Zia, Performance analysis of deep-neural-network-based automatic diagnosis of diabetic retinopathy, Sensors 22 (1) (2021) 205.

[38] G.U. Nneji, J. Cai, J. Deng, H.N. Monday, M.A. Hossin, S. Nahar, Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans, Diagnostics 12 (2) (2022) 540.

[39] R. Nirthika, S. Manivannan, A. Ramanan, Siamese network based fine grained classification for Diabetic Retinopathy grading, Biomed. Signal Process Control 78 (2022), 103874.

[40] M. Wang, H. Chen, Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis, Appl. Soft Comput. 88 (2020), 105946.

[41] M. Yu, M. Han, X. Li, X. Wei, H. Jiang, H. Chen, R. Yu, Adaptive soft erasure with edge self-attention for weakly supervised semantic segmentation: thyroid ultrasound image case study, Comput. Biol. Med. 144 (2022), 105347.

[42] Y. Chen, H. Gan, H. Chen, Y. Zeng, L. Xu, A.A. Heidari, X. Zhu, Y. Liu, Accurate iris segmentation and recognition using an end-to-end unified framework based on MADNet and DSANet, Neurocomputing 517 (2023) 264–278.

[43] Y. Chen, X.H. Yang, Z. Wei, A.A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, Q. Guan, Generative adversarial networks in medical image augmentation: a review, Comput. Biol. Med. 144 (2022), 105382.