IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 24, 2021, accepted December 16, 2021, date of publication December 28, 2021,
date of current version January 7, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3139129

Classification of Diabetic Retinopathy Severity Based on GCA Attention Mechanism

BINHUA YANG[1], TONGYAN LI[1], HAIDI XIE[1], YULIN LIAO[1], AND YI-PING PHOEBE CHEN[2], (Senior Member, IEEE)
[1]College of Communication Engineering, Chengdu University of Information Technology, Chengdu 610225, China
[2]Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3086, Australia

Corresponding author: Tongyan Li (sunny_cs061@163.com)

ABSTRACT Diabetic retinopathy (DR) is one of the major complications caused by diabetes and can lead to severe vision loss or even complete blindness if not diagnosed and treated in a timely manner. In this paper, a new feature map global channel attention mechanism (GCA) is proposed to solve the problem of the early detection of DR. In the GCA module, an adaptive one-dimensional convolution kernel size algorithm based on the dimension of the feature map is proposed and a deep convolutional neural network model for DR color medical image severity diagnosis named GCA-EfficientNet (GENet) is designed. The training process uses transfer learning techniques with a cosine annealing learning rate adjustment strategy. The image regions of interest of GENet are visualized using a heat map. The final accuracy, precision, sensitivity and specificity of the DR dataset of the Kaggle competition reached 0.956, 0.956, 0.956, and 0.989, respectively. A large number of experiment results show that GENet based on the GCA attention mechanism can more effectively extract lesion features and classify the severity of DR.

INDEX TERMS Attention mechanism, convolutional neural network, deep learning, diabetic retinopathy, medical images.

I. INTRODUCTION

Diabetic retinopathy (DR) is one of the major complications of diabetes due to the retinal damage caused by the rupture of capillaries from high levels of sugar [1]. There are now 460 million people worldwide aged 20-79 years with diabetes and this number will exceed 700 million by 2045 [2], [3]. Due to the dramatic increase in the number of people with diabetes, the number of people with DR is expected to reach 191 million by 2030 [4]. Early-stage DR is less harmful, does not cause serious visual impairment and is clinically treatable [5], treatment in a timely manner can reduce the risk of visual impairment by approximately 57% [6]. Therefore, timely examination and treatment are the main measures to protect visual acuity.

2D color fundus images and 3D optical coherence tomography (OCT) images are the most common examination methods and the diagnostic basis for ophthalmic diseases [7]. 2D color fundus images are a common DR examination method with the advantages of time saving and low cost compared to OCT, additional lesions can be diagnosed through color images, such as macular edema, microaneurysm and optic disc edema, etc. Due to the increasing development of big data and computer technology in recent years, the application of deep learning in image processing and computer-aided diagnosis has become increasingly prevalent. The use of color fundus images to identify the severity of DR based on computer-aided diagnosis can not only improve the efficiency of doctors' diagnosis, but also save the cost of medical treatment and provide convenience for economically poor areas.

Deep convolutional neural networks (DCNNs) are widely used in computer vision and have achieved excellent results in many tasks. For DR severity diagnosis, Varun Gulshan *et al.* [8] used the Inception-v3-based deep convolutional neural network trained on EyePACS-1 and Messidor-2 datasets and achieved 90.3% sensitivity and 98.1% specificity on EyePAC-1 and Messidor-2. Wan *et al.* [9] performed transfer learning and hyperparameter finetuning on AlexNet, VggNet, GoogLeNet, ResNet on the Kaggle platform respectively, the best classification accuracy reached 95.68%. Gadekallu *et al.* [10] argued that previous studies lacked data preprocessing and dimensionality reduction, which led to poor results, proposed a feature extraction method combining the standardization of data by

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeeb Dey.

StandardScalar, principal component analysis and deep neural networks, the model was evaluated against the mainstream machine learning models available today.

When using DCNN for DR classification, because color fundus images are rich in detailed parts such as capillaries and are spread over the vast majority of the image, proper preprocessing of the original image is required, more importantly, the network model needs to consider the relevance of the original image at different locations and in different channels. This requires the network to introduce a proper attention mechanism, which makes the model adaptively enhance the perception of useful information. However, there is scant existing research on this issue. Runze Fan *et al.* [11] combined an attention model in the feature fusion stage of the DR classification model to adaptively update the weights of each feature block, Liu et al[12] combined a compact bilinear pooling model and an attention mechanism for DR fine-grained image classification. Based on the above background and previous studies, this paper proposes a DR severity classification model GCA-EfficientNet (GENet) using deep learning, which introduces a global channel attention mechanism for feature maps and fully takes into account the correlation between different dimensions of the feature map for DR severity-assisted diagnosis. The main contribution points of this paper are as follows.

(1) For the DR severity classification problem, a Global Channel Attention (GCA) mechanism is proposed to update the attention weights of the different channels of the feature map with the model training process.

(2) In the process of the GCA module parameter update, an adaptive one-dimensional convolution kernel size calculation method is proposed to adjust the size of the convolution kernel adaptively according to the dimensions of the feature maps of different feature extraction modules.

(3) Combining the GCA attention mechanism and EfficientNet, the GCA-EfficientNet (GENet) model is proposed, based on the transfer learning technique, the training of the model is accomplished, eventually the accuracy, precision, sensitivity and specificity reach 0.956, 0.956, 0.956, and 0.989 respectively on the DR dataset of the Kaggle competition.

II. RELATED WORKS

Diabetic retinopathy severity detection aims to help physicians make a timely diagnosis of early fundus disease and provide a rationale for further treatment based on the severity of DR by discriminating lesion features on color fundus images or OCT images through image processing techniques and computer technology. Early research on DR mainly used traditional machine learning techniques to identify lesion features. However, in recent years, with the rapid development of artificial intelligence technology and computer technology, an increasing number of scholars have used deep learning techniques for DR severity classification.

At the stage of DR detection using traditional machine learning techniques, researchers need to have some medical background and manual extract lesion features from the image dataset, whereupon the extracted lesion features are fed into a classification model to complete the detection of DR. Nguyen *et al.* [13] proposed a multilayer feedforward neural network with strong robustness for DR severity classification. For the early detection and classification of the main symptoms of DR, Zhang *et al.* [14] used a support vector machine (SVM) to classify preprocessed bright nonlesion areas, exudates and cotton wool spots. Zhang *et al.* [15] proposed a top-down strategy to detect fundus hemorrhage, proposed combined 2DPCA, and applied virtual SVM to achieve higher classification accuracy. To extract the feature vectors of the DR images, Soares *et al.* [16] used pixels and took a 2D Gabor wavelet transform at multiple scales, which were fed into a classification model to identify vascular and non-vascular. Nayak *et al.* [17] used image preprocessing, morphological processing and texture analysis techniques to detect lesion features and used them as inputs to an artificial neural network for the automatic detection of DR severity. An automatic system for analyzing DR lesions in the central field of retina is proposed by Barriga *et al.* [18], the system extracted features using amplitude and frequency modulation and used partial least squares (PLS) and a support vector machine (SVM) for classification. Priya *et al.* [19] compared the performance of a probabilistic neural network (PNN) and support vector machine (SVM) for DR binary classification, the SVM model achieved 97.608% accuracy which was better than the other models. Roychowdhury *et al.* [20] analyzed fundus images in different contexts and reduced the number of features used for lesion classification to generate DR severity classes using machine learning. In this paper [21], Srivastava *et al.* used a Frangi filter to extract features from the green channel of fundus images to train a SVM classifier, which predicted the severity of DR. Santhakumar *et al.* [22] divided the lesion features of fundus images into several rectangular patches, then passed the features of the patches into a support vector machine (SVM) for DR severity classification. This paper [23] attempts to detect red lesions from retinal fundus images, Srivastava *et al.* proposed a new filter with strong robustness to discriminate between vascular and red lesions, the lesion features were extracted using the corresponding filter for red lesions of different sizes, the experiment results show that this filter was helpful for the automatic detection of DR.

Although these methods can detect the severity of DR to some extent, machine learning methods based on the traditional approach require a large number of annotated features, this process consumes a lot of resources and time for feature annotation, it needs to segment the part of the lesion from the whole fundus image, which makes the whole annotation process more demanding in terms of medical background and inefficient. Moreover, it is easy to miss the lesion features in the fundus image during the annotation process. However, deep learning techniques do not require the manual annotation and segmentation of lesion features, for example, convolutional neural networks (CNNs) can extract lesion features from the entire fundus image without missing features compared to manual ones. In addition, when CNNs extract fundus image features, according to the different receptive field, it is easy to extract detailed features from convolutional kernels close to the network input such as the texture and shape of the image, while more semantic features are easily extracted for convolutional kernels close to the network output. Nowadays, an increasing number of researchers are applying deep learning techniques for DR severity detection. The adoption of CNN has made the DR diagnosis process simple and efficient, Pratt *et al.* [24] used CNN structures to extract disease features from fundus images and trained the model using data augmentation techniques to enable the extraction of complex lesion features. A method for deep visual feature (DVF) extraction based on scale-invariant color density and gradient location direction histogram was proposed by Abbas *et al.* [25], with the whole model having no pre-processing or post-processing stages, the extracted features were transformed and fed into a multilayer classification network to obtain the prediction results. Kanungo *et al.* [26] derived the impact of hyperparameters and the quality and quantity of training data on the model performance through a large number of comparative experiments. Due to the uninterpretable black-box nature of how CNNs make decisions internally based on image features, Quellec *et al.* [27] proposed a way to create heatmaps to show which pixels in an image play a role in prediction at the image level and applied it to DR screening. In this paper [28], Zhao *et al.* proposed a model combining an attention mechanism and a bilinear network for fine-grained classification to deal with small target lesion features in fundus images and proposed a new loss function for different classes of DR, the experiments showed that this model achieved excellent performance in DR severity classification. A contour detection image processing algorithm for vascular detection in fundus images based on Mamdani (Type-2) fuzzy rule was developed by Orujov *et al.* [29], this algorithm passing the green channel of the image through adaptive histogram equalization with restricted contrast and median filtering, then applying the Mamdani fuzzy rule to the gradient values of the image for edge detection. Das *et al.* [30] proposed a CNN-based DR detection and classification algorithm in which fundus images are preprocessed so that vascular branches can be extracted by a segmentation model, the segmented regions are corrected using a maximum principal curvature and adaptive histogram equalization. A residual convolutional block attention model (RCAM) was proposed by Fan *et al.* [31], the attention model is used in a multi-feature fusion technique with adaptive weights, which was combined with the MobileNetV3 network for DR severity classification. Liu *et al.* [12] considered DR severity classification as a fine-grained classification problem and proposed a compact bilinear pooling network model based on the attention mechanism for DR severity classification, which improved both the prediction accuracy and maintained the computational efficiency of the model. In this paper [32], Ramasamy *et al.* extracted and fused ophthalmic features from retinal images, which are based on texture gray level features. The diabetic retinopathy severity was classified using the sequential minimum optimization (SMO) classification method.

Different from previous approaches, this paper proposes a new attention mechanism and a flexible adaptive convolutional kernel sizing algorithm in the attention mechanism, which automatically adjusts the convolutional kernel size according to the size of the input feature matrix and fuses the local channel correlation to obtain the global channel correlation of the feature map and combines it with a deep convolutional neural network for DR severity classification. To determine the model's ability to detect lesion features at different stages, this paper uses a heatmap to visualize the area to which the model pays special attention.

III. METHOD

A. ATTENTION MECHANISM

The attention mechanism was proposed by Treisman *et al.* [33] to simulate a model of the human brain's attention, which can derive attention weights for different factors, emphasizing the impact of a particular factor on the model's results. The attention mechanism has been widely used in deep learning tasks such as sequence-to-sequence [34], image localization [35], image understanding [36], and lip translation [37]. The transformer structure proposed by the Google Machine Translation team [38], which discards the recursion and convolution structures and is based entirely on the simpler attention mechanism for processing feature sequences, achieved 28.4 BLEU in the WMT 2014 English-to-German translation task, which was 2 BLEU higher than the best result at the time.

The attention mechanism is used to adaptively adjust the higher-order abstract features extracted by the model for better performance and has been increasingly combined with computer vision in recent years. Hu *et al.* [39] proposed the ''Squeeze-and-Excitation''(SE). The SE structure assumes that the input feature map is $\boldsymbol{I} \in \mathbb{R}^{H\times W\times C}$, where H, W, and C denote the height, width, and number of channels of the feature map, respectively. The output feature matrix of the SE structure can be expressed as:

$$\boldsymbol{O} = \sigma(\boldsymbol{W_{ex}}ReLU(\boldsymbol{W_{sq}}(gap(\boldsymbol{I})))) \tag{1}$$

where $\boldsymbol{O} \in \mathbb{R}^{H\times W\times C}$, the gap denotes the global average pooling operation over the channel of the matrix, $\boldsymbol{W_{sq}}$ and $\boldsymbol{W_{ex}}$ are fully connected layers for downscaling and upscaling channel, $ReLU(\cdot)$ and $\sigma(\cdot)$ denote rectified liner unit and sigmoid activation function.

The SE structure is widely used in many classical network architectures, including MobileNet-v3 [40] and EfficientNet [41] due to its higher flexibility and obvious performance improvement of the network model, which has a large performance improvement for DCNN.

Although the inter-channel attention mechanism based on the SE structure considers the correlation between different channels of the feature matrix and gives the influence of channel weights on the model results, the fully connected layers used in both the SE stage make the number of parameters of the model increase drastically. Because the correlation between the channels of the feature map will yield more nonlinear information, which is beneficial to the performance of the network model [42], the bottleneck structure composed of two fully connected layers in the SE structure, although it reduces the number of parameters by dimensionality reduction to a certain extent, it means that some of the feature information extracted by the network is lost, which leads to some limitations in the model performance. SE-Var2 adopts a depth-wise separable convolution approach, which reduces the number of parameters by learning the weights of each channel independently, but does not take the correlation of the feature map channels into account, SE-Var3 adopts a fully connected layer mapping, which considers the correlation of the channels but significantly increases the number of parameters [42]. Although SE-Var3 makes some improvements to this problem, it never makes a better balance between complexity and channel correlation.

The efficient channel attention (ECA) mechanism proposed by Wang *et al.* [42] makes a trade-off between the performance and complexity of the model compared to the original SE structure and adopts an adaptive convolution kernel size adjustment method to effectively extract the correlation between different channels of the feature matrix. The ECA structure is similar to channel convolution [41], which helps to capture the intrinsic correlations between feature map channels and uses a one-dimensional convolution with an adaptive convolution kernel size in the channel attention mechanism, which greatly reduces the complexity of the model compared to the SE structure. In the ECA structure, assuming that the input feature matrix is $\boldsymbol{I} \in \mathbb{R}^{H\times W\times C}$, then

$$\boldsymbol{y} = f_k(gap(\boldsymbol{I})) \tag{2}$$

where *gap* denotes the global average pooling operation, f_k denotes a one-dimensional convolution of convolution kernel size k, with k positively correlated with the dimension C of the input feature matrix, $\boldsymbol{y} = [y_1, y_2, \ldots, y_C]$, and finally $\boldsymbol{y}$ is multiplied with the original matrix to obtain the model of the inter-channel attention mechanism.

$$\boldsymbol{O} = \sigma(\boldsymbol{y}) \otimes \boldsymbol{I} \tag{3}$$

where $\otimes$ denotes the multiplication in the channel dimension and the final output feature matrix $\boldsymbol{O} \in \mathbb{R}^{H\times W\times C}$.

Although the ECA structure can effectively control the parametric increase of the model and it is an efficient inter-channel attention mechanism, the one-dimensional convolution makes the weight of each channel derived from the final convolution operation only related by the fixed k channel features adjacent to it, ignoring the correlation between the global channel features of the feature map.

B. GCA STRUCTURE

In this paper, we propose a global channel attention (GCA) model for feature maps. The GCA structure effectively overcomes the problem where the ECA structure only considers local channel correlations, in addition, GCA takes the correlations of all feature map channels into account while maintaining the number of model parameters, which effectively improves the perceptual performance of the model for different channels. The GCA structure is shown in Fig. 1.

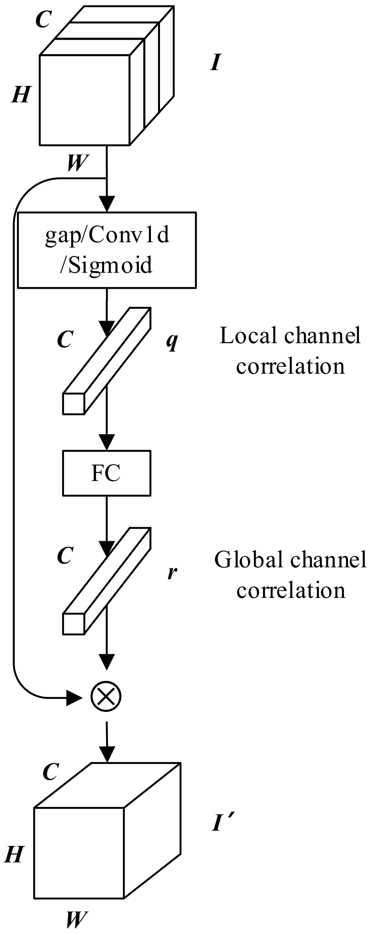

FIGURE 1. Schematic diagram of GCA structure. In the figure, $\otimes$ denotes the product of elements in the feature matrix channel dimension.

Assuming that the input feature matrix of the input GCA structure is $\boldsymbol{I} \in \mathbb{R}^{H\times W\times C}$, the features of each channel are obtained by a global average pooling operation.

$$\boldsymbol{p} = gap(\boldsymbol{I}) \tag{4}$$

For the extracted features $\boldsymbol{p}, \boldsymbol{p} \in \mathbb{R}^C$, an adaptive one-dimensional convolution operation is performed to obtain the local inter-channel correlation features $\boldsymbol{q}$ as follows:

$$\boldsymbol{q} = swish(\boldsymbol{W_1 p}), \quad \boldsymbol{q} \in \mathbb{R}^C \tag{5}$$

In the above equation, $\boldsymbol{W}_1 \in \mathbb{R}^{C\times C}$ is the parameter matrix of the one-dimensional convolution and $swish(\cdot)$ is the activation function defined as:

$$swish(x) = x \cdot \sigma(x) \tag{6}$$

where the parameter matrix $\boldsymbol{W_1} \in \mathbb{R}^{C\times C}$ can be expressed as shown at the bottom of the next page.

To determine the convolution kernel size of the 1D convolution in $\boldsymbol{W}_1$, this paper proposes an adaptive convolution kernel size design method based on the ECA structure [42], which makes the size of the convolution kernel adaptively adjusted with different feature maps. To obtain more local inter-channel correlations, the number of channels of the feature map and the convolution kernel size are positively correlated, $C \propto k$ so this paper proposes that the relationship between the feature map and the convolution kernel size is

$$C = g_{\gamma,\beta}(k) = e^{\gamma k+\beta} \tag{7}$$

where k denotes the convolutional kernel size, γ and β is the parameter of the linear mapping, which can be learned through the network training process, therefore, the relationship between the convolutional kernel size k and the number of feature map channels can be expressed as:

$$k = g^{-1}_{\gamma,\beta}(C) = \left|\frac{\ln(C) - \beta}{\gamma}\right| \tag{8}$$

As a result, local inter-channel correlation features with window length k are extracted, each feature can be expressed as:

$$q_i = swish(\sum_{j=1}^{k} p_i(k)w_{i,j}), \quad i \in (1, C) \tag{9}$$

where $p_i(k)$ denotes the k channel features adjacent to p_i.

After extracting the local inter-channel correlation feature $\boldsymbol{q} \in \mathbb{R}^C$, in order to extract the global channel correlation feature $\boldsymbol{r}$ from the local inter-channel correlation feature $\boldsymbol{q}$, a global linear operation is then performed on $\boldsymbol{q}$,

$$\boldsymbol{r} = \sigma(\boldsymbol{W}_2\boldsymbol{q}), \quad \boldsymbol{r} \in \mathbb{R}^C \tag{10}$$

where $\boldsymbol{W_2} \in \mathbb{R}^{C\times C}$ is the parameter matrix of the linear operation and the parameter matrix $\boldsymbol{W}_{\not\models}$ can be shown as:

$$\boldsymbol{W_2} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,C} \\ \vdots & \ddots & \vdots \\ w_{C,1} & \cdots & w_{C,C} \end{pmatrix} \tag{11}$$

Then, the global channel correlation $\boldsymbol{r} = [r_1, r_2, \cdots, r_C]$ is extracted based on the local inter-channel correlation $\boldsymbol{q}$,where r_i can be expressed as:

$$r_i = \sum_{j=1}^{C} w_{i,j}q_j \tag{12}$$

Finally, the global channel correlation features $\boldsymbol{r} \in \mathbb{R}^C$ are weighted with the input feature matrix $\boldsymbol{I} \in \mathbb{R}^{H\times W\times C}$ in the dimension of the channel to obtain the feature matrix $\boldsymbol{I}'$ of the global attention mechanism of the feature map, where $\boldsymbol{I}'$ can be expressed as:

$$\boldsymbol{I}' = \boldsymbol{I} \otimes \boldsymbol{r}, \quad \boldsymbol{I}' \in \mathbb{R}^{H\times W\times C} \tag{13}$$

where $\otimes$ denotes a multiplying weighting operation in the dimension of the feature map channel.

After the above operations, the feature matrix $\boldsymbol{I}'$ of the global attention mechanism of the feature map is obtained. Unlike the SE structure and ECA structure, the GCA structure proposed in this paper overcomes both the loss of the feature map channel information caused by dimensionality reduction processing in the SE structure and the drawback of failing to consider the global channel correlation in the ECA structure, while the adaptive convolution kernel size adjustment method proposed in this paper can extract local channel correlation information at different scales according to the feature maps of different tasks.

The GCA structure proposed in this paper extracts the global channel correlation information of the feature map in two steps. The first step begins with the local inter-channel correlation derived by adaptive one-dimensional convolution with a small number of parameters; the second step integrates the local inter-channel correlation and extracts the global channel correlation. The two-step operation effectively avoids the huge number of parameters caused by two fully connected operations, so that the model does not suffer from overfitting problems and also extracts the global channel correlation features. In the later sections of this paper, a disease severity classification model based on the GCA structure will be proposed and trained based on transfer learning, and finally the performance of the model in this paper will be evaluated using the experiment results.

C. GENET STRUCTURE

The deep convolutional neural network model used in this paper is based on EfficientNet [41], which is based on the neural network architecture search technique (NAS) obtained by balancing the network width, depth and input image resolution, using a relatively small number of parameters but obtaining better performance, depending on the different resolutions of the input image, model width and depths. EfficientNet can be divided into eight models from EfiicientNet-B0 to EfficientNet-B7 [41]. EfficientNet-B7 exceeds the accuracy achieved by the best GPipe at that time, but with 8.4 times fewer number of parameters and 6.1 times faster computing speed [41]. However, because EfficientNet uses the same inverted residual structure MBConv as

$$\boldsymbol{W_1} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,k} & 0 & 0 & \cdots & 0 \\ 0 & w_{2,2} & \cdots & w_{2,k+1} & 0 & \cdots & 0 \\ 0 & 0 & \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & 0 & w_{C-1,C-k} & \cdots & w_{C-1,C-1} & 0 \\ 0 & 0 & \cdots & 0 & w_{C,C-K+1} & \cdots & w_{C,C} \end{pmatrix}$$

MobileNetv2 [43], where the inter-channel attention mechanism uses an SE structure consisting of two fully connected layers, the model has a larger number of parameters and loses some information due to the dimensionality reduction process in the SE structure.

In this paper, we improve on the MBConv convolutional structure and propose the GConv structure which integrated the GCA attention mechanism and the MBConv structure, as shown in Fig. 2. Finally, with reference to the EfficientNet-B0 model derived from NAS technology, the GCA-EfficientNet (GENet) based severity classification model used in this paper for DR is proposed, as shown in Fig. 3.

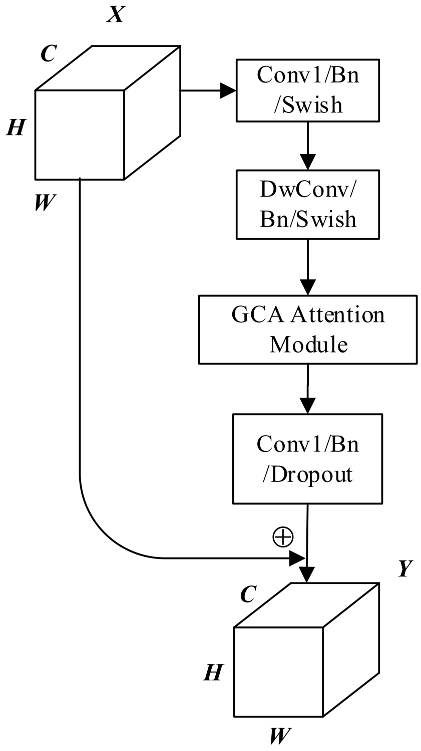

FIGURE 2. GConv structure diagram. The ⊕ in the figure indicates the summation by feature matrix elements, and this operation holds only when the input and output feature matrices have the same dimension.

D. VISUALIZATION

To solve the problem of invisibility inside the convolutional neural network model, which is like a "black box", this paper uses Grad-CAM [44] to visualize the attention region of the CNN for the input image in the form of a heat map. The gradient of the score y^c of any category c with respect to the feature map A^k of the convolution layer, i.e. $\partial y^c/\partial A^k_{ij}$, is first calculated, then a global average pooling operation is undertaken in the dimensions of height and width, the corresponding weight scores are calculated as follows:

$$\varphi_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{14}$$

The weight φ_k^c indicates the importance of the feature map for the prediction result, after filtering out the effect of negative values by $ReLU(\cdot)$, the final CNN visualization algorithm for DR severity detection is obtained as follows:

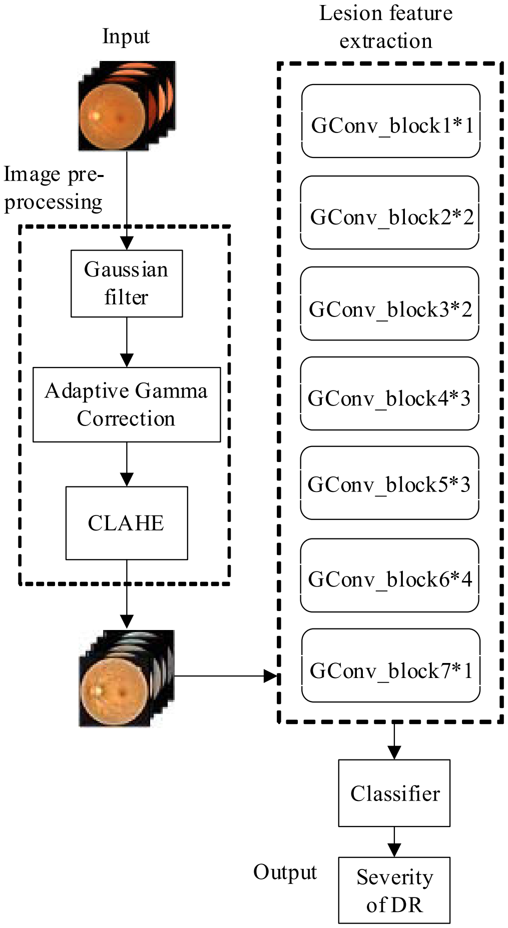

FIGURE 3. DR severity classification model based on GENet.

$$L^c_{Grad-CAM} = ReLU(\sum_k \varphi_k^c A^k) \tag{15}$$

IV. EXPERIMENT

A. DATASET INTRODUCTION

The dataset used in this experiment is a Kaggle competition dataset containing 35126 high-resolution color fundus images, which have been divided into 5 categories by professional clinicians according to the severity of DR, the number of samples in each category is shown in Table 1. Because the sample number of different categories in this dataset varies greatly, it will have a negative impact on the results of the model, which will be addressed by the preprocessing process in the next section.

B. PRE-PROCESSING AND DATA AUGMENTATION

The dataset used in this experiment is taken from fundus cameras in different environments, which introduce noise during data collection, there are negative impacts such as uneven light, so image preprocessing is necessary to reduce the impact of noise on the experiment results and improve the learning effect of the network model. Meanwhile, to solve

TABLE 1. Sample number of DR dataset.

Class	Severity of DR	Number
0	No DR	25810
1	Mild Non-Proliferative DR	2443
2	Moderate Non-Proliferative DR	5292
3	Severe Non-Proliferative DR	873
4	Proliferative DR	708

the uneven number of DR images in different classes, this paper makes the number of samples in each class basically the same by performing data augmentation techniques on negative samples.

1) IMAGE PRE-PROCESSING

The image pre-processing process comprises the following steps:

(A) Remove the black margin around the fundus image to reduce the impact of unnecessary information.

(B) Gaussian filtering, which inhibits Gaussian noise during image acquisition.

(C) Adaptive gamma correction improves the effect of uneven lighting during image collection, corrects images with too much gray or too little gray, as well as enhances contrast.

(D) Contrast-limited adaptive histogram equalization (CLAHE), converting RGB color space to Y-Cr-Cb color space, then adaptive histogram equalization in brightness to reduce the impact of uneven gray scale values in brightness.

After the above operations, the pre-processed DR images are obtained, the pairs of DR images before and after pre-processing for different categories are shown in Fig. 4. The pre-processed DR images can be better used for DCNN learning, the proposed DCNN model GENet can better capture the detailed features of the pre-processed images.

2) DATA AUGMENTATION

Although the pre-processed DR images can be directly trained for the network, it can be seen from Table 1 that the number of DR images in different categories varies greatly, which will adversely affect the result of the network model, so this paper applies data augmentation processing to the negative samples, rotating (90°, 180°, 270°), flipping horizontally and vertically, cropping at the four corners and the center of the negative sample images, ensuring the number of samples in each class is basically the same, solving the sample imbalance problem. A comparison of the sample numbers before and after data augmentation is shown in Table 2.

TABLE 2. Comparison of sample size before and after data augmentation.

Class	Severity of DR	Number	Data Augmentation
0	No DR	25810	25810
1	Mild Non-Proliferative DR	2443	24380
2	Moderate Non-Proliferative DR	5292	26440
3	Severe Non-Proliferative DR	873	26160
4	Proliferative DR	708	25488

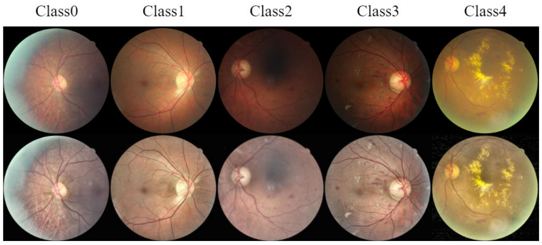

FIGURE 4. Comparison of DR image before and after pre-processing. The first row in the figure is the original image, the second row is the image after pre-processing.

C. EXPERIMENT ENVIRONMENT

The GENet proposed in this paper runs under PyTorch 1.7.0, Python 3.6 environment. It divides the dataset into training and validation sets according to 8 : 2, the image resolution is set as 224 × 224, the cross-entropy loss function is used, 100 epochs are learned on the training set, stochastic gradient descent with momentum is used as the model parameter optimizer, the initial learning rate is set to 0.01, the momentum is set to 0.9. In order to make the model finally converge, the cosine annealing learning rate adjustment strategy shown in Fig. 5 is used in this paper.

The lack of sufficient labeled data is a major challenge for medical image processing. When training DCNN models, a small training set is prone to overfitting. In addition, deep learning systems require much more training time and larger amount of data than traditional machine learning systems. In order to solve the above problems, in the model training stage, this paper adopts the transfer learning technique. Transfer learning is a deep learning training strategy that a pre-trained model with generalized features is reused in another task, and in the field of computer vision, specific low-level features such as edges, shapes, and textures can be shared between tasks. Therefore, the use of transfer learning techniques can fine tune the pre-trained model in downstream tasks, thus greatly saving training time and the amount of data, allowing the model to converge as soon as possible and avoiding the overfitting problem.

To ensure the model achieves the desired performance as soon as possible, the parameters of the same structure of GENet and EfficientNet-B0 were used to initialize the GENet network model using the transfer learning technique in the model training stage, the GCA modules with different structures were initialized using Kaiming initialization [45], the models used as comparisons are initialized with the official pre-trained models provided by PyTorch official. To analyze the effectiveness of GENet in DR disease detection, in this paper, GENet and classical DCNN networks are compared in the same environment, the experiment results are analyzed in the next section.

D. EXPERIMENT CONCLUSION AND ANALYSIS

To evaluate the DR classification performance of GENet rigorously, the model which was trained for 100 epochs was evaluated comprehensively for accuracy, precision,

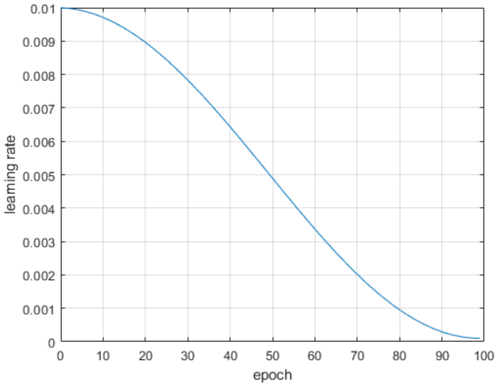

FIGURE 5. **Cosine annealing learning rate adjustment strategy.**

sensitivity and specificity on the validation set in this paper. The confusion matrix is defined in Table 3, confusion matrix for DR severity classification was drawn and the GENet model was compared with the classical DCNN.

TABLE 3. **Confusion matrix definition.**

	Label			
True Value Predict Value	Positive Positive	Negative Positive	Positive Negative	Negative Negative
Result	TP	FP	FN	TN

As shown in Table 3 accuracy, precision, sensitivity, and specificity are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$\Pr ecision = \frac{TP}{TP + FP} \tag{17}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{18}$$

$$Specificity = \frac{TN}{TN + FP} \tag{19}$$

Experiment 1: In this paper, the performance of the model trained for 100 epochs is evaluated on the validation set to obtain the detection of GENet for DR diseases of different severities, the confusion matrix obtained from the experiment is shown in Fig. 6.

It can be seen from Fig. 6 that GENet achieves an accuracy of 95.63% after 100 epochs of training, the classification performance for DR diseases of different severities is shown in Table 4.

TABLE 4. **Classification performance of GENet for DR diseases of different severities.**

Severity of DR	Precision	Sensitivity	Specificity
No DR	0.87	0.936	0.965
Mild Non-Proliferative DR	0.978	0.914	0.995
Moderate Non-Proliferative DR	0.948	0.931	0.987
Severe Non-Proliferative DR	0.998	0.999	0.999
Proliferative DR	0.996	1.0	0.999

It can be seen from Table 4 that GENet achieves excellent classification results for Severe Non-Proliferative DR and Proliferative DR, while slightly lower classification performance for No DR and Mild DR. We conjecture that due to the relatively small difference between the two types of images in the color fundus images, this problem will be overcome as the model is trained with more samples and for a longer period of time because the GENet model is not over-fitted.

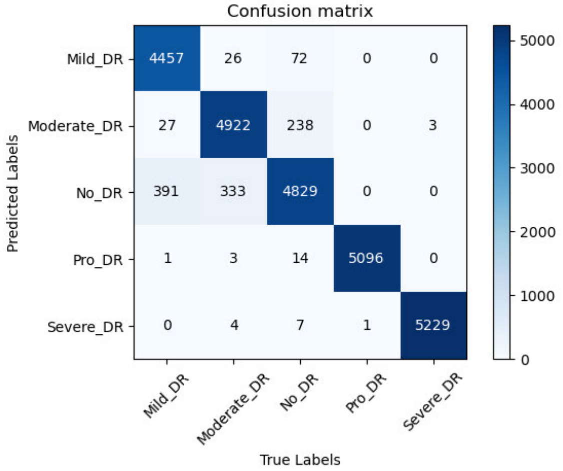

FIGURE 6. **The confusion matrix for DR diseases of different severities. The more concentrated samples are on the diagonal, indicating a better classification result.**

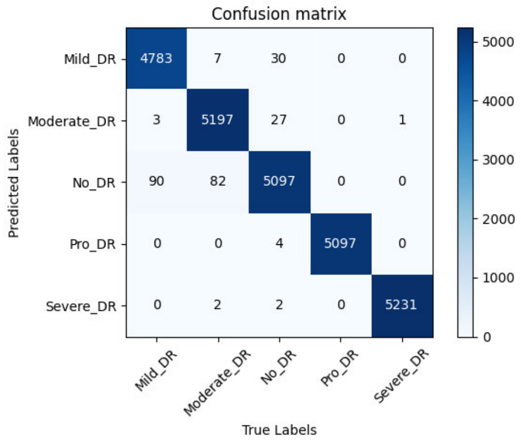

FIGURE 7. **The confusion matrix obtained after setting the training epoch to 150.**

In order to verify the above inference, we set the epoch of the above experiments to 150 and trained GENet with other hyperparameters unchanged. The confusion matrix derived after 150 epochs of training is shown in Fig. 7, and the detection performance for different DR severity is shown in Table 5. By comparing the confusion matrix with Fig. 6 and the DR detection performance with Table 5, it can be seen that our analysis is correct and GENet has significantly improved the detection effect for NO DR and Mild DR. Therefore, the GENet based on GCA attention mechanism proposed in this paper can effectively detect different DR severity.

Convolutional neural networks are like a “black box”, however, Grad-CAM [44] can visualize the region of interest of a CNN in the form of a heat map to understand intuitively how the CNN model derives its prediction results.

TABLE 5. The classification performance of GENet for different DR severity after setting epoch to 150.

Severity of DR	Precision	Sensitivity	Specificity
No DR	0.967	0.988	0.992
Mild Non-Proliferative DR	0.992	0.981	0.998
Moderate Non-Proliferative DR	0.994	0.983	0.998
Severe Non-Proliferative DR	0.999	1.0	1.0
Proliferative DR	0.999	1.0	1.0

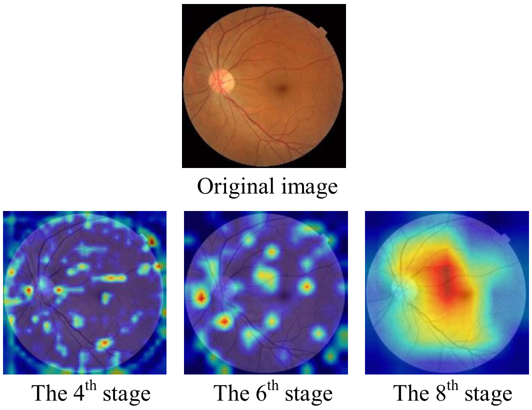
Original image
The 4th stage
The 6th stage
The 8th stage

FIGURE 8. Visualization of the GCA attention mechanism, heat map of stages 4, 6, and 8 in GENet.

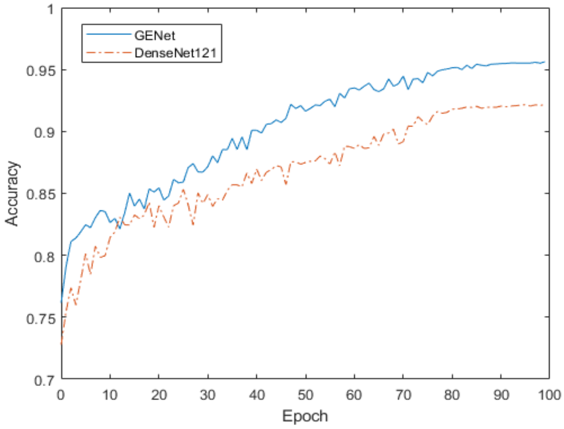

FIGURE 9. The graphs of GENet and DenseNet-121 are compared under the same experiment environment. The solid line in the figure indicates GENet and the dashed line indicates DenseNet-121. The accuracy of both models on the validation set keeps improving with training, but the accuracy of GENet always outperforms DenseNet-121.

The shallow layer of the convolutional neural network preserves the lower-order features such as contours, edges, and textures of the image, while the deep layer preserves the higher-order semantic features of the image. To have an intuitive understanding of the attention mechanism in GENet and how GENet makes predictions internally, blocks of different depths were selected for visualization in this paper. The heat map of DR images based on Grad-CAM is shown in Fig. 8, which demonstrates the different levels of attention of GENet to different regions of the fundus image after using the GCA attention mechanism. In the heat map, red indicates that the image features in the region have higher weights and blue indicates lower weights. From Fig. 8, it can be seen that the feature maps of different depths in the model contain different information about the lesions reflected in the DR images.

TABLE 6. Results of GENet compared with other DCNN models. Bold indicates better performance results.

Models	Accuracy	Precision	Sensitivity	Specificity
ResNet-50	0.860	0.863	0.860	0.965
DenseNet-121	0.921	0.924	0.921	0.980
GoogLeNet	0.9	0.901	0.9	0.975
GENet(Proposed)	**0.956**	**0.956**	**0.956**	**0.989**

Experiment 2: To verify the effectiveness of GENet in DR severity classification, in this paper, GENet and DenseNet-121 were compared on the same training and validation sets under the same experiment environment, the accuracy is shown in Fig. 9. From the figure, it can be seen that the accuracy of GENet proposed in this paper is better than DenseNet-121 in detecting DR diseases under the same experiment environment, no overfitting phenomenon occurs as the training process continues.

In this paper, we also conducted a comparison experiment between GENet and the classical DCNN model in the same experiment setting as above, analyzed the accuracy, precision, sensitivity, and specificity after 100 training epochs. The results are shown in Table 6. From Table 6, it can be seen that all evaluation indexes of the GENet network are better than the classical traditional CNN network, indicating that the DCNN model based on the GCA attention mechanism proposed in this paper can achieve better performance in the disease severity classification task.

V. CONCLUSION

Diabetic retinopathy is one of the major complications of diabetes mellitus, failure to diagnose and treat it in time can lead to severe eye vision loss or even complete blindness. However, diabetic retinopathy can be prevented through routine screening and effective treatment, thus avoiding the occurrence of irreversible blindness. With the continuous development of machine learning and artificial intelligence technologies, an increasing number of machine learning techniques are used in the medical field to assist doctors in routine diagnosis and treatment.

Therefore, this paper proposes a global channel attention mechanism for feature maps, named the GCA attention mechanism. Furthermore, a deep convolutional neural network model GENet, in which the GCA attention mechanism and EfficientNet are integrated, is proposed for the early detection of diabetic retinopathy. In the disease feature extraction stage, for the network model to fully consider the correlation between feature map channels, this paper proposes an adaptive convolutional kernel size adjustment algorithm for extracting local channel correlation, which makes GENet adaptively adjust the convolutional kernel size in different tasks, so that the network model is enough to achieve better performance. The training process uses transfer learning techniques and cosine annealing algorithms to ensure that the model eventually converges as quickly as possible. The final GENet model achieves 0.956 accuracy, 0.956 precision, 0.956 sensitivity and 0.989 specificity on the DR validation set, demonstrating that the deep convolutional neural network model based on the GCA attention mechanism proposed in this paper is effective in the classification of the severity of DR. In future work, we will combine the GCA attention mechanism with more deep learning models to improve the performance of the model to detect small differences between categories, so that GENet can be used in more scenarios.

REFERENCES

[1] C. P. Wilkinson, F. L. Ferris, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, and J. T. Verdaguer, ''Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales,'' *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, Sep. 2003, doi: 10.1016/S0161-6420(03)00475-5.

[2] P. Saeedi, P. Salpea, S. Karuranga, I. Petersohn, B. Malanda, E. W. Gregg, N. Unwin, S. H. Wild, and R. Williams, ''Mortality attributable to diabetes in 20–79 years old adults, 2019 estimates: Results from the international diabetes federation diabetes atlas, 9th edition,'' *Diabetes Res. Clin. Pract.*, vol. 162, Apr. 2020, Art. no. 108086, doi: 10.1016/j.diabres.2020.108086.

[3] C. Sabanayagam, R. Banu, M. L. Chee, R. Lee, Y. X. Wang, G. Tan, J. B. Jonas, L. Lamoureux, C.-Y. Cheng, B. E. K. Klein, P. Mitchell, and R. Klein, ''Incidence and progression of diabetic retinopathy: A systematic review,'' *Lancet Diabetes Endocrinol.*, vol. 7, no. 2, pp. 140–149, 2019, doi: 10.1016/s2213-8587(18)30128-1.

[4] D. S. Ting, G. C. Cheung, and T. Y. Wong, ''Diabetic retinopathy: Global prevalence, major risk factors, screening practices and public health challenges: A review,'' *Clin. Exp. Ophthalmol.*, vol. 44, no. 4, pp. 260–277, May 2016, doi: 10.1111/ceo.12696.

[5] N. Cho, J. E. Shaw, S. Karuranga, Y. Huang, J. D. da Rocha Fernandes, A. W. Ohlrogge, and B. Malanda, ''IDF diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045,'' *Diabetes Res. Clin. Pract.*, vol. 138, pp. 271–281, Apr. 2018, doi: 10.1016/j.diabres.2018.02.023.

[6] Early Treatment Diabetic Retinopathy Study Research Group, ''Grading diabetic retinopathy from stereoscopic color fundus photographs—An extension of the modified airlie house classification: ETDRS report number 10,'' *Ophthalmology*, vol. 127, no. 4S, pp. S99–S119, Apr. 2020, doi: 10.1016/S0161-6420(13)38012-9.

[7] M. D. Abràmoff, M. K. Garvin, and M. Sonka, ''Retinal imaging and image analysis,'' *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010, doi: 10.1109/RBME.2010.2084567.

[8] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, and S. Venugopalan, ''Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,'' *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016, doi: 10.1001/jama.2016.17216.

[9] S. Wan, Y. Liang, and Y. Zhang, ''Deep convolutional neural networks for diabetic retinopathy detection by image classification,'' *Comput. Electr. Eng.*, vol. 72, pp. 274–282, Nov. 2018, doi: 10.1016/j.compeleceng.2018.07.042.

[10] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, I.-H. Ra, and M. Alazab, ''Early detection of diabetic retinopathy using PCA-firefly based deep learning model,'' *Electronics*, vol. 9, no. 2, p. 274, Feb. 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/2/274

[11] Y. Zhang, C. P. Huynh, and K. N. Ngan, ''Feature fusion with predictive weighting for spectral image classification and segmentation,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6792–6807, Sep. 2019, doi: 10.1109/TGRS.2019.2908679.

[12] P. Liu, X. Yang, B. Jin, and Q. Zhou, ''Diabetic retinal grading using attention-based bilinear convolutional neural network and complement cross entropy,'' *Entropy*, vol. 23, no. 7, p. 816, Jun. 2021, doi: 10.3390/e23070816.

[13] H. T. Nguyen, M. Butler, A. Roychoudhry, A. G. Shannon, J. Flack, and P. Mitchell, ''Classification of diabetic retinopathy using neural networks,'' in *Proc. 18th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 4, Nov. 1996, pp. 1548–1549, doi: 10.1109/IEMBS.1996.647546.

[14] Z. Xiaohui and O. Chutatape, ''Detection and classification of bright lesions in color fundus images,'' in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 1, Oct. 2004, pp. 139–142, doi: 10.1109/ICIP.2004.1418709.

[15] X. Zhang and O. Chutatape, ''A SVM approach for detection of hemorrhages in background diabetic retinopathy,'' in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 4, Aug. 2005, pp. 2435–2440, doi: 10.1109/IJCNN.2005.1556284.

[16] J. V. Soares, J. J. Leandro, R. M. Cesar Junior, H. F. Jelinek, and M. J. Cree, ''Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification,'' *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006, doi: 10.1109/TMI.2006.879967.

[17] J. Nayak, P. S. Bhat, R. Acharya, C. M. Lim, and M. Kagathi, ''Automated identification of diabetic retinopathy stages using digital fundus images,'' *J. Med. Syst.*, vol. 32, no. 2, pp. 107–115, Apr. 2008, doi: 10.1007/s10916-007-9113-9.

[18] E. S. Barriga, V. Murray, C. Agurto, M. Pattichis, W. Bauman, G. Zamora, and P. Soliz, ''Automatic system for diabetic retinopathy screening based on AM-FM, partial least squares, and support vector machines,'' in *Proc. IEEE Int. Symp. Biomed. Imag.: Nano Macro*, Apr. 2010, pp. 1349–1352, doi: 10.1109/ISBI.2010.5490247.

[19] R. P. R. Priya and P. Aruna, ''SVM and neural network based diagnosis of diabetic retinopathy,'' *Int. J. Comput. Appl.*, vol. 41, no. 1, pp. 6–12, Mar. 2012, doi: 10.5120/5503-7503.

[20] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, ''DREAM: Diabetic retinopathy analysis using machine learning,'' *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1717–1728, Sep. 2014, doi: 10.1109/JBHI.2013.2294635.

[21] R. Srivastava, D. W. K. Wong, L. Duan, J. Liu, and T. Y. Wong, ''Red lesion detection in retinal fundus images using frangi-based filters,'' in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 5663–5666, doi: 10.1109/EMBC.2015.7319677.

[22] S. R. M. Tandur, E. R. Rajkumar, G. K S. G. Haritz, and K. T. Rajamani, ''Machine learning algorithm for retinal image analysis,'' in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2016, pp. 1236–1240, doi: 10.1109/TENCON.2016.7848208.

[23] R. Srivastava, L. Duan, D. W. K. Wong, J. Liu, and T. Y. Wong, ''Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels,'' *Comput. Methods Programs Biomed.*, vol. 138, pp. 83–91, Jan. 2017, doi: 10.1016/j.cmpb.2016.10.017.

[24] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, ''Convolutional neural networks for diabetic retinopathy,'' *Proc. Comput. Sci.*, vol. 90, pp. 200–205, Jan. 2016, doi: 10.1016/j.procs.2016.07.014.

[25] Q. Abbas, I. Fondon, A. Sarmiento, S. Jiménez, and P. Alemany, ''Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features,'' *Med. Biol. Eng. Comput.*, vol. 55, no. 11, pp. 1959–1974, Nov. 2017, doi: 10.1007/s11517-017-1638-6.

[26] Y. S. Kanungo, B. Srinivasan, and S. Choudhary, ''Detecting diabetic retinopathy using deep learning,'' in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2017, pp. 801–804, doi: 10.1109/RTEICT.2017.8256708.

[27] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, ''Deep image mining for diabetic retinopathy screening,'' *Med. Image Anal.*, vol. 39, pp. 178–193, Jul. 2017, doi: 10.1016/j.media.2017.04.012.

[28] Z. Zhao, K. Zhang, X. Hao, J. Tian, M. C. Heng Chua, L. Chen, and X. Xu, ''BiRA-Net: Bilinear attention net for diabetic retinopathy grading,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1385–1389, doi: 10.1109/ICIP.2019.8803074.

[29] F. Orujov, R. Maskeliūnas, R. Damaševičius, and W. Wei, ''Fuzzy based image edge detection algorithm for blood vessel detection in retinal images,'' *Appl. Soft Comput.*, vol. 94, Sep. 2020, Art. no. 106452, doi: 10.1016/j.asoc.2020.106452.

[30] S. Das, K. Kharbanda, S. M. R. Raman, and D. D. Edwin, ''Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy,'' *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102600, doi: 10.1016/j.bspc.2021.102600.

[31] R. Fan, Y. Liu, and R. Zhang, ''Multi-scale feature fusion with adaptive weighting for diabetic retinopathy severity classification,'' *Electronics*, vol. 10, no. 12, p. 1369, Jun. 2021, doi: 10.3390/electronics10121369.

[32] L. K. Ramasamy, S. G. Padinjappurathu, S. Kadry, and R. Damaševičius, "Detection of diabetic retinopathy using a fusion of textural and ridgelet features of retinal images and sequential minimal optimization classifier," *PeerJ Comput. Sci.*, vol. 7, p. e456, May 2021, doi: 10.7717/peerj-cs.456.

[33] A. Treisman, "Features and objects in visual processing," *Sci. Amer.*, vol. 255, no. 5, pp. 114–125, Nov. 1986, doi: 10.1038/scientificamerican1186-114B.

[34] T. Bluche, "Joint line segmentation and transcription for end-to-end handwritten paragraph recognition," in *Proc. NIPS*, 2016, pp. 838–846.

[35] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," presented at the 28th Int. Conf. Neural Inf. Process. Syst., Montreal, QC, Canada, 2015.

[36] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," 2015, *arXiv:1502.03044*.

[37] J. S. Chung, A. Senior, O. Vinyals, and A. Zisserman, "Lip reading sentences in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3444–3453, doi: 10.1109/CVPR.2017.367.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," presented at the 31st Int. Conf. Neural Inf. Process. Syst., Long Beach, CA, USA, 2017.

[39] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[40] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," 2019, *arXiv:1905.02244*.

[41] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.

[42] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539, doi: 10.1109/CVPR42600.2020.01155.

[43] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," 2018, *arXiv:1801.04381*.

[44] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," 2015, *arXiv:1502.01852*.

BINHUA YANG received the bachelor's degree in electronic science and technology from the Chengdu University of Information Technology, in 2019, where he is currently pursuing the master's degree.

He has participated in many mathematical modeling competitions and won awards, received several academic scholarships, and has several software copyrights. His research interests include deep learning and computer vision.

TONGYAN LI received the Ph.D. degree in communication and information system from the University of Electronic Science and Technology of China, in 2010.

Since 2010, she has been a Graduate Student Tutor with the Department of Communication Engineering, Chengdu University of Information Technology. She has been involved in a number of projects, like the "863" major project, National Natural Science Fund Project, Fund Project of Sichuan Provincial Department of Education and Found Project of Science, and Technology Department in Sichuan Province. As the first author, she has more than 20 academic papers published in academic journals and conferences, of which two papers were indexed by SCI, and more than 20 papers were indexed by EI. Her research interests include data mining, recommendation systems, big data processing, machine learning, and artificial intelligence methods.

HAIDI XIE received the bachelor's degree from the Chengdu University of Information Technology, where she is currently pursuing the degree majoring in electronic information with the School of Communication Engineering.

The completed projects include a personalized recommendation system, and the design and implementation of a weather data platform for the Internet of Things based on the rest architecture. A book on recommender systems is currently being written. Her research interests include data mining, recommendation systems and big data processing, and she is currently researching recommendation algorithms.

YULIN LIAO is currently pursuing the degree with the Chengdu University of Information Technology, majoring in communication engineering and electronic information. In 2016, she studied with the School of Communication Engineering, Chengdu University of Information Technology. During the undergraduate period, she participated in the recommendation system and other projects and won provincial awards. She won the third prize in the 15th Wuyi Mathematical Modeling Contest. In the study and work and won the "school-level excellent stem," scholarship. Her current research interests include artificial intelligence and intelligent information processing.

YI-PING PHOEBE CHEN (Senior Member, IEEE) has been a Professor and the Chair with the Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia, since April 2010. She has been working in many emerging areas, such as bioinformatics, multimedia, artificial intelligence, scientific visualization, pattern recognition, health informatics, data mining, deep learning, and databases. She has published over 240 research papers, many of them appeared in top journals and conferences, such as *Artificial Intelligence*, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Biomedical Engineering, *Pattern Recognition*, IEEE Transactions on Information Technology in Biomedicine, IEEE Transactions on Knowledge and Data Engineering, *Information Systems*, *Bioinformatics*, *Nature Machine Intelligence*, *Molecular Systems Biology*, *Aging Cell*, *Nucleic Acids Research*, *SIGMOD*, and *ACM MM*.

She is the Steering Committee Chair of Asia-Pacific Bioinformatics Conference (a Founder) and international conference on multimedia modeling. She has been on the program committees of over 100 international conferences, including top ranking conferences, such as ICDE, ICPR, ISMB, and CIKM.