IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 August 2023, accepted 4 September 2023, date of publication 7 September 2023,
date of current version 13 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312709

RESEARCH ARTICLE

Grading of Diabetic Retinopathy Images Based on Graph Neural Network

MEILING FENG[1], JINGYI WANG[1], KAI WEN[2], AND JING SUN[2]
[1]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300191, China
[2]Tianjin Eye Hospital, Tianjin Medical University, Tianjin 300070, China

Corresponding author: Meiling Feng (sdfml88@email.tjut.edu.cn)

This work was supported by the Scientific Research Project of the Tianjin Municipal Education Commission under Grant 2018ZD11.

ABSTRACT Diabetic Retinopathy (DR) has become one of the main reasons for the rise in the number of limited vision people worldwide, while high-definition color fundus images have brought great convenience to the diagnosis of DR. However, manual image reading is time-consuming and labor-intensive, and different doctors may make different diagnoses. At present, intelligent grading based on deep learning has become a hotspot in DR intelligent diagnosis. The existing DR intelligent classification model based on convolutional neural network has achieved good results, but the relationship between the deep features proposed by the network is not considered, while this relationship contains important classification information. In order to overcome the above-mentioned shortcoming of convolutional networks, this article draws on the powerful relationship capture capabilities of graph neural networks and proposes a new DR intelligent classification model. The model is composed of two cascaded networks. The convolutional neural network is used to extract the deep features of the DR image, and the graph neural network is used to further capture the relationship between the deep features of the convolutional network. Finally, the two network outputs are fused by adaptive weight, and give the grading result of the entire network. The proposed model is evaluated on the APTOS2019 and Messidor-2 datasets. Compared with other models, the grading accuracy and F1-score of the proposed model on APTOS2019 are improved by 1.1% and 1.3%, respectively. The grading accuracy and F1-score are improved by 1.4% and 1.8% on Messidor-2, respectively.

INDEX TERMS Diabetic retinopathy, grading, graph convolutional network, convolutional neural network, graph neural network.

I. INTRODUCTION

Diabetic retinopathy (DR) is a chronic disease caused by diabetes, which damages retinal microvessels, resulting in a series of typical lesions, which eventually lead to serious consequences such as visual impairment and even blindness [1]. At different stages of the disease, DR patients will have different pathological features, including: microaneurysms, hemorrhage, exudates (including hard exudates and soft exudates), and neovascularization, as shown in Fig. 1. Microaneurysms, abnormal retinal hemorrhage and exudation are the early pathological features of DR patients. If a patient is in this pathological state for a long time, some blood vessels will be blocked or closed, so that nutrients cannot reach the retina, thereby stimulating the growth of new blood vessels, causing extensive damage to the eye, and eventually leading to blindness [2].

In order to clearly define the severity of DR to help doctors in clinical diagnosis, a grading method is usually used in medicine. The current international commonly used five-class grading standard [3], in which the existence of each pathological feature determines the DR class, as shown in Fig. 2. According to the severity of the lesions, it can be divided into: Grade I (no disease), Grade II (mild non-proliferative stage), Grade III (moderate non-proliferative stage), Grade IV (severe non-proliferative stage), Grade V (proliferative stage).

According to statistics, nearly one-third of diabetic patients will develop DR, and nearly one-tenth of diabetic patients will develop severe vision-threatening DR. However, DR patients

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

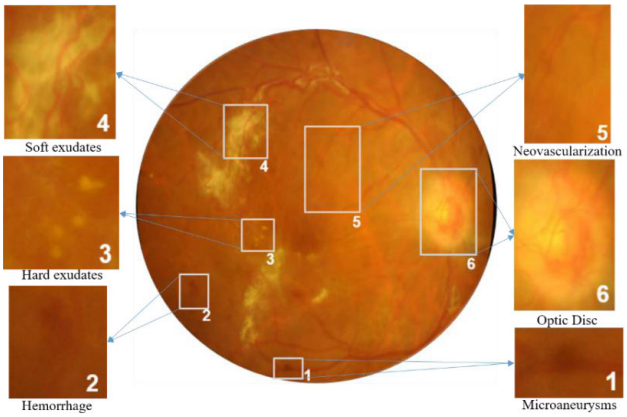


FIGURE 1. Example of typical lesions (1)-(5) and optic disc.

do not have any obvious symptoms in the early stage, until the lesions develop to a more serious class, the patient can perceive some symptoms of the eye. (for example: dark shadow, vision loss, etc.) Screening is very necessary [4]. With the rapid development of modern medical imaging technology, high-definition color fundus images have brought great convenience to the screening of DR. However, the differences in DR image features are small, and it is difficult to distinguish between adjacent class, and images of different class have their own characteristics. The pathological features of images of other class overlap to some extent. In clinical practice, only experienced doctors can grade the severity of DR images, and it takes 1-2 days for patients to get the results. At the same time, for the same DR Images, different doctors may make different diagnoses, and the same doctor may give different diagnosis results at different times, as shown in Fig. 3, this time-consuming and laborious diagnosis process may delay the treatment of the disease. Therefore, It is necessary to develop intelligent DR grading systems.

DR grading based on fundus images can be regarded as image classification, and each DR class can be regarded as each category of images. In recent years, medical image classification based on deep learning has achieved rich results and applications, such as: diagnosis of cervical cancer lesions, diagnosis of breast lesions, diagnosis of diabetic retinopathy, etc. At present, the DR intelligent grading model is mainly based on convolutional neural network (Convolutional Neural Network, CNN). The algorithm of middle retinopathy realizes intelligent grading of DR images. In 2016, Gulshan et al. [5] used convolutional neural network to learn the characteristics of DR images and established an algorithm for automatic detection of retinopathy, for DR images classifications. Pratte et al. [6] proposed a customized CNN model based on CNN architecture and data augmentation, which can identify complex features in classification tasks and intelligently grade DR images. Zhao et al. [7] added an attention mechanism to neural networks for improving DR image classification performance. Zhou et al. [8] believed that different classes are progressive, that is, there is a relationship between labels, and proposed a multi-cell multi-task learning network (M2CNN) to achieve DR image classification and label regression tasks. Later, Li et al. [9] considered the relationship between DR and diabetic macular edema (DME), learning DR and the relationship between DME to improve DR images classification performance.

However, since CNN does not consider the characteristics of DR images, the performance of CNN-based DR image classification is limited. In order to overcome the above shortcoming of CNN, Sakaguchi et al. [10] first proposed to use graph neural network for DR image classification to improve DR image classification performance.

Luo and Kamata [11] proposed a DR grading method using lesion-related graphs based on graph convolutional network and convolutional neural network, which demonstrated that there is a relationship between lesions in DR images. Liu et al. [12] considered the class correlation and proposed a grading model based on Graph Convolutional Network (GCN) to improve the performance of intelligent grading of DR images by learning the relationship between different class of fundus images. The above methods all use the existing CNN modules to capture the deep features of the fundus images, but they ignore the potential relationship among the deep features proposed by the convolutional network, which contains important class information.

In order to capture the relationship between the deep Features of the image, this paper draws on the powerful relationship capturing ability of the Graph Neural Network (GNN) and proposes a novel DR image classification method based on the Graph Neural Network, as shown in Fig. 4. The method obtains the deep features with stronger class information by learning DR image deep features correlation. Specifically, the model consists of two cascaded networks. First, CNN can learn the image deep features and obtain the high-class feature representation of the image, so as to obtain the score corresponding to each DR class. Secondly, a two-layer graph neural network can capture deep features correlation, and obtain new high-level feature representations, and finally obtain the DR image score at each class. These new high-class feature representations describe the relationship between deep features, which is a more accurate feature-level representation of fundus images with richer class-level feature information. Finally, the output of the two-way network is fused through an adaptive weight, and the paper sets the adaptive weight to be learnable and sum to 1, so as to control the contribution ratio of the two-way network to the final classification result.

In summary, the contributions of this paper include the following three aspects:

(1) For the first time, GCN learns the images deep features correlation for DR image classification, and can obtain deep features with richer class information, thereby improving the classification performance.

(2) This paper proposes a new mechanism that integrates the CNN branch network and the GCN branch network to learn the image deep features correlation, that is, the adaptive weight mechanism, which allows the network to adjust the proportion of the contribution of the branch network of GCN

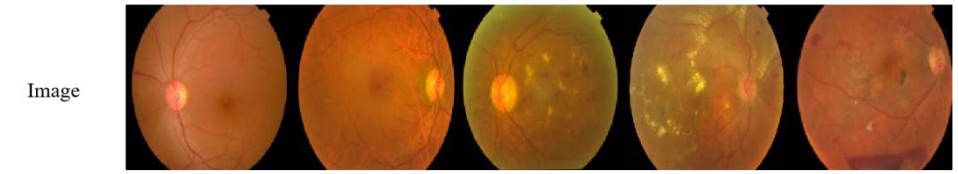
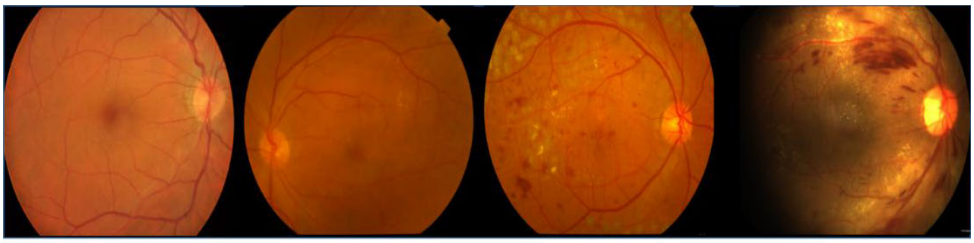


Image					
Lesion	None	Microaneurysms	Microaneurysms Hard exudates	Microaneurysms Hemorrhage Soft exudates	New blood vessel Massive hemorrhage
Grade	Grade I	Grade II	Grade III	Grade IV	Grade V
Severity	None	Mild NPDR	Moderate NPDR	Severe NPDR	PDR

FIGURE 2. International criterion on DR image grading.



None (1) Mild (2)	Mild (1) Moderate NPDR (2)	Severe NPDR (1) Moderate NPDR (2)	PDR (1) Severe NPDR (2)

FIGURE 3. Inconsistent labels from two physicians (i.e., 1 and 2) for the same DRfundus images. The small difference in features of the DR image causes this inconsistency and could limit thenetwork classification accuracy.

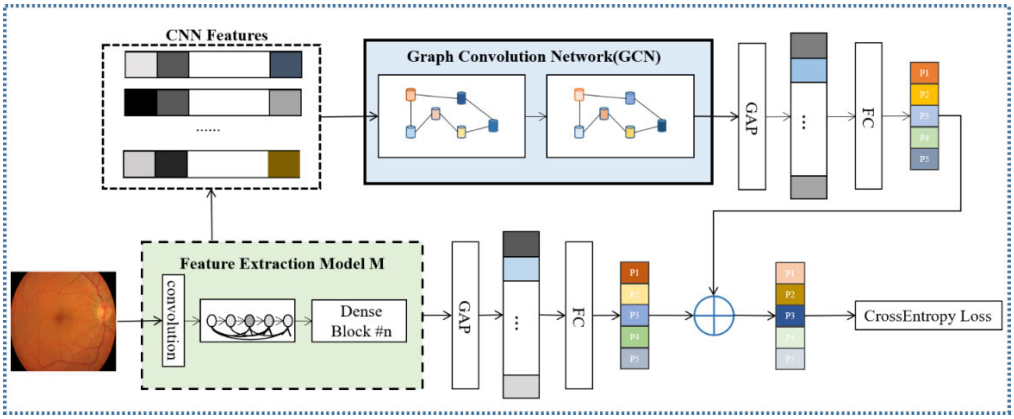


FIGURE 4. Grading of Diabetic Retinopathy Images Based on Graph Neural Network. (The original image classification network includes feature extractors and classifier in addition two layers of GCN to learn the deep features correlation of the image, the two networks respectively predict the classification results, finally our method sets an adaptive weight to control the contribution ratio of the two branches to the final result.)

learning the deep features correlation of the image to the final classification result according to different datasets.

(3) The experimental results on the APTOS2019 and Messidor-2 datasets show that the superiority of our method.

II. RELATED WORK

This paper builds a network main model based on CNN. The proposed module for learning the relationship between deep features of images needs to rely on the powerful relationship capture ability of GCN. This section briefly describes the theoretical basis of our model.

A. CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks have become the dominant method for image classification. Although convolutional neural networks have been proposed 20 years ago,

improvements in computer hardware and network structures have made the training of deep CNNs possible. The early LeNet5 [13] includes five layers, VGG19 [14] includes 19 layers, and in recent years DenseNet and ResNet have more than 100 layers. A convolutional neural network is an artificial neural network, usually a convolutional neural network includes a convolutional layer, a pooling layer, nonlinear activation layer and a fully connected layer, as shown in Fig. 5. Convolution is the basic convolutional layer, pooled layer, fully connected layer. Enhance the original signal characteristics through operation of CNN, and its main features are: through the convolution operation, the original signal features are enhanced and noise is reduced. The reason for using a pooling layer is that subsampling the image can reduce the amount of computation while keeping the image rotation invariant according to the principle of image local correlation. The activation function makes the model introduce nonlinear factors, so that the neural network can complete nonlinear mapping. The role of the fully connected layer is to implement classification. The feature space output by CNN is used as the input of the fully connected layer, and the fully connected layer is used to complete the mapping from the input to the label set, that is, classification.

B. GRAPH NEURAL NETWORK (GNN)

Although deep learning can effectively capture the hidden patterns of data in Euclidean space, a large number of learning tasks today need to deal with graph data, which contains two parts of information: attribute information and structural information. The attribute information describes the inherent properties of the objects in the graph; the structure information describes the properties of the association between the objects. The structure generated by this association is not only of great help to the characterization of the nodes in the graph data, but also to the characterization of the whole graph. also plays a key role. These graph data contain rich inter-element relationships. The performance of traditional deep learning methods in processing non-Euclidean spatial data is unsatisfactory. Therefore, deep learning and graph theory are combined to produce a graph neural network for processing graph data. Abilityto capture dependencies within graph data (non-Euclidean data) through message passing between graph nodes [15]. At present, graph neural networks are divided into five categories, namely: Graph Convolutional Networks, Graph Attention Networks [16], Variational Graph Auto-encoders [17], Graph Generation Networks (Graph Generative Networks) [18] and Graph Spatial-temporal Networks [19].

C. GRAPH CONVOLUTION NEURAL NETWORK (GCN)

Graph Convolutional Networks (GCN) [20], [21] generalize convolution operations from grid data to graph data and enable end-to-end learning on graph data. GCN methods are divided into two categories, namely: spectral-based and spatial. Spectral-based methods define graph convolution by introducing filters from a graph signal processing perspective; spatial-based methods characterize graph convolution as aggregating feature information from nearby neighbors.

The main idea of GCN is to convolve the neighbor nodes of each node in the graph structure, and effectively combine the inherent structural information of the graph with the features of the nodes during the learning process, and finally generate a more advanced node feature representation. Given a graph $G = (V, E)$ with n nodes, nodes $v_i \epsilon V$, edges $(v_i, v_j) \epsilon E$ and an adjacency matrix $A \epsilon R^{n \times n}$ describing the relationship between nodes, the purpose of GCN is to pass neural The network model $f(X, A)$ encodes the graph, where $X \in R^{N \times D}$, D represents the dimension of each node feature. AX is the sum of the features of all neighbor nodes, so GCN can capture the inherent structural information of the graph. A multilayer GCN updates node features according to the following propagation rules:

$$H^{(l+1)} = f_{relu}(\hat{A} H^{(l)} W^{(l)}) \tag{1}$$

Among them, $H^l \epsilon R^{N \times d_l}$ represents the feature description of the node in Layer l, d_l represents the feature dimension of the node at layer l, H^{l+1} represents the node feature updated after a layer of GCN, and is used as The input of the next layer of GCN node features, $\hat{A} \in R^{N \times N}$ represents the regularized form of the adjacency matrix A, and the regularization matrix A^ needs to be obtained through the original adjacency matrix A and degree matrix δ. δ is a diagonal matrix, and δ can be obtained by the following formula:

$$\delta_{ij} \overset{\text{def}}{=} \begin{cases} \deg(v_i) & if\ i = j \\ 0 & otherwise \end{cases} \tag{2}$$

f_{relu} is an activation function, and $W^{l+1} \epsilon R^{d_l \times d_{l+1}}$ is a trainable weight matrix. Therefore, after two layers of GCN, node feature representation with richer semantic information can be obtained.

III. MATERIAL AND METHODS

In 2020, Wang et al. [22] clustered and composed the latent features of Covid-19 images, and fused the deep features of two layers of GCN and CNN to improve the classification performance. Inspired by this, this paper proposes a DR image grading model based on GCN and CNN based on the characteristics of DR images, as shown in Fig. 4: Use CNN to extract the deep features of the image, use two layers of GCN to learn the relationship between the deep features of the image, and finally, the two-way network completes fusion through adaptive weights. Therefore, the network can obtain high-level feature representation with richer class information, so that it has better advantages in classification performance. Each module of the model will be described in detail below.

A. GRAPH CONVOLUTIONAL NETWORK MODULE FOR DR CLASSIFICATION

1) RESNET

The emergence of the ResNet [23] model was a milestone in CNN history. Because the traditional convolutional neural

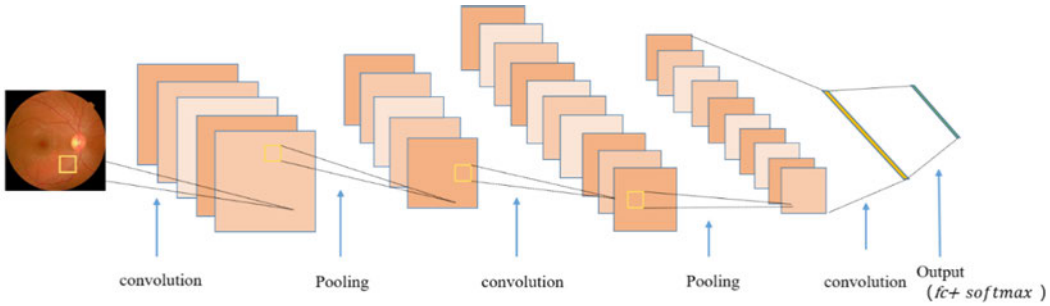


FIGURE 5. Convolutional neural network structure diagram, convolutional neural network structure includes: convolutional layers and reduce noise; Use pooling layers to subsample images to reduce the amount of computation.

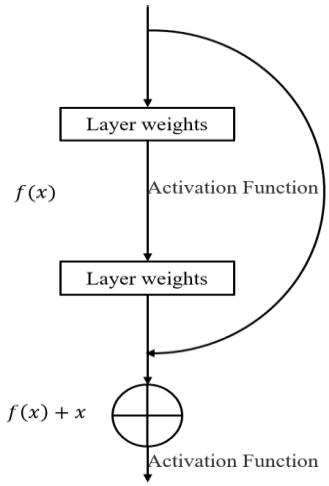


FIGURE 6. The residual learning module greatly protects the integrity of the information by "bypassing" the input information to the output.

network or fully connected network, in the information transmission, there will be more or less information loss, loss and other issues, but also lead to gradient disappearance or gradient explosion, making it difficult to train a very deep network. However, ResNet (Residual Neural Networks) proposes a kind of residual learning that solves the problem of deep neural network degradation to some extent, as shown in Fig. 6, the block composed of $F(x) + x$ is called the residual block. A number of similar residual blocks are connected in series to form ResNet, as shown in Fig. 7, each layer and the previous layer (usually 2 to 3 layers) are connected by elemental level addition, short circuit. This facilitates the back propagation of gradients during training, so that deeper CNN networks can be trained.

2) DENSENET

Compared with ResNet which is dense connections, DenseNet [24] has the advantage that DenseNet directly connects feature maps from different layers, which can achieve feature reuse and improve efficiency. As a result, DenseNet has become a popular network today. DenseNet consists primarily of Dense Block and Transition, as shown in Fig. 8. Inside Dense Block, the feature map of the layers is the same size and can be connected on the channel dimension, that is, it proposes a more aggressive dense connection mechanism

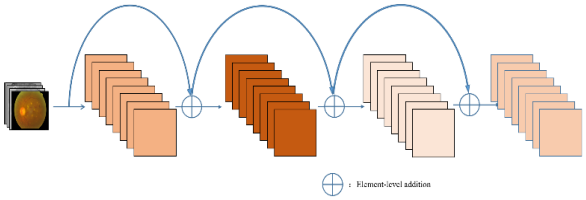


FIGURE 7. ResNet block diagram, each layer is connected to a short circuit in front of a layer, the connection method is element-level addition.

that connects all the layers to each other, and each layer will be connected with all the previous layers in the channel dimension as input to the next layer, as shown in Fig. 9. The Transition layer mainly connects two adjacent Dense Blocks and reduces the size of the feature map, which can compress the model. In the experiment, the image was resize to a fixed size of 256 × 256, so that several feature maps of 7*7 size are available after the CNN (ResNet, DenseNet) extracts the features.

B. GRAPH CONVOLUTIONAL NETWORK MODULE FOR DR CLASSIFICATION

In order to improve the performance of DR classification, this paper proposes a method to learn the relationship between DR images deep features using GCN. The basic CNN model is used to obtain the deep features of the DR image, but the CNN model does not combine the characteristics of the DR image itself, and fails to consider the relationship between the deep features of the DR image, so the obtained high-level semantic feature representation is not accurate enough. Using GCN on top of CNN has better results than just using CNN.

In this paper, for each DR image I, the DenseNet121 model pre-trained on ImageNet [25] is used as the base CNN model in the experiment, and 1024 feature maps are obtained. First, the global average pooling layer (GAP) is used to obtain the high-level feature representation of the image $I_C \epsilon R^D$, where D represents the dimension of the vector is 1024. Secondly, the obtained 1024 feature maps are respectively expanded into vectors, and 1024 vectors $X \epsilon R^{1024 \times d}$ are obtained, where d represents the expanded vector dimension of each feature map. Then the graph structure G is constructed based on X and the corresponding adjacency matrix $A \epsilon R^{1024 \times 1024}$ is

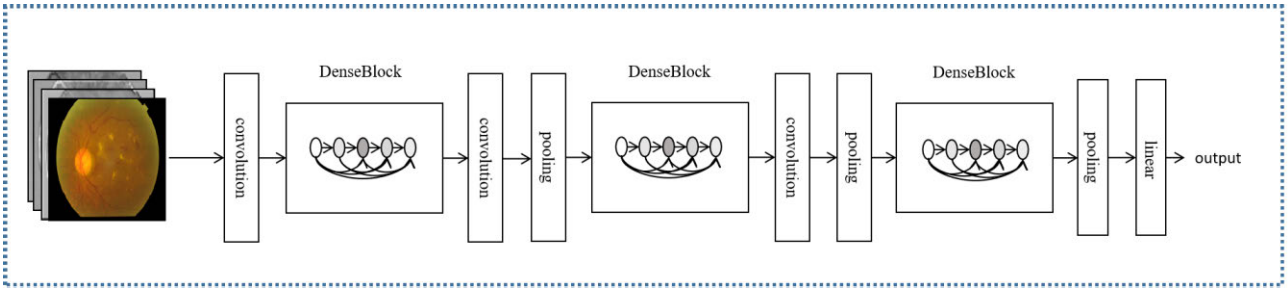


FIGURE 8. A deep DenseNet with three dense blocks, The layers between two adjacent blocks are referred to as transition layers and changefeature-map sizes via convolution and pooling.

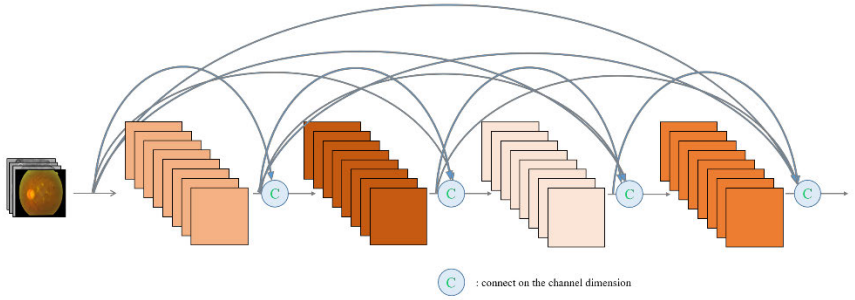


FIGURE 9. DenseBlock, each layer takes all preceding feature-maps as input.

defined as follows:

$$A_{ij} = \begin{cases} 1, & if\ X_i \in KNN(X_j)\ or\ X_j \in KNN(X_i) \\ 0, & otherwise \end{cases} \tag{3}$$

where $KNN(X_i)$ denotes the k nearnest neighbors of X_i based on Cosine similarity, as illustrated in Fig. 10, assuming k = 4, then the three nearest neighbors of node i and node j are $KNN(X_i) = (1, 2, 3, j)$, $KNN(X_j) = (4, 5, 6, 7)$. Therefore, $X_j \in KNN(X_i)$ and $X_i \notin KNN(X_j)$ can be obtained. Then, the node feature $X = H^{(0)}$ and the regularized adjacency matrix $\hat{A}$^ are used as the input of the two-layer GCN through the two-layer GCN, as shown in Fig. 11, there are:

$$H^{(1)} = f_{relu}(\hat{A}XW^{(0)}) \tag{4}$$

$$H^{(2)} = f_{relu}(\hat{A}XW^{(1)}) \tag{5}$$

$W^{(0)} \in R^{d_0 \times d_1}$, $W^{(1)} \in R^{d_1 \times d_2}$ are a trainable weight matrix, and the feature $H^{(1)} \in R^{1024 \times d_1}$ of the node is obtained after the first layer of GCN, after the second layer of GCN, the final node feature representation $H^{(2)} \in R^{1024 \times d_2}$ is obtained. ($d0$,$d1$,$d2$) are hyperparameters will be set in the experiment. $H^{(2)}$ through the GAP layer as well, after that a more accurate high-level feature representation $I_G \in R^{1024}$ that takes into account the relationship between the deep features of the image can be obtained; I_C and I_G are respectively passed through the fully connected layer:

$$R_I = (W^{(0)})^T I_C + b \tag{6}$$

$$S_I = (W^{(1)})^T I_G + b \tag{7}$$

$W^{(0)}, W^{(1)} \in R^{1024 \times C}$, where C is the number of classes and b is the bias. We can get the scores of the DR image at each level $R_I \in R^C$, $S_I \in R^C$.

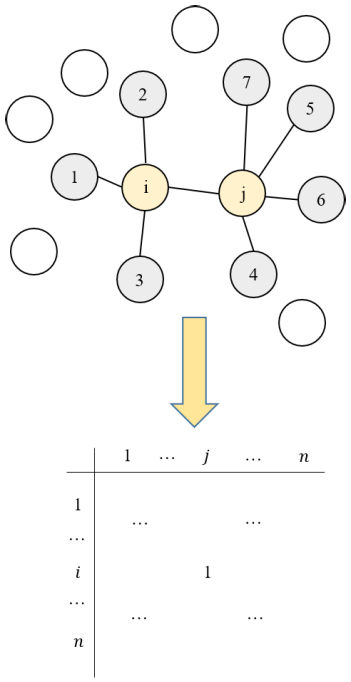


FIGURE 10. Illustration of KNN-based adjacency matrix.

C. FUSION MODULE OF TWO-WAY NETWORK

The output R_I, S_I of the two-way network is fused through adaptive weights w(w initialized to 0.5) to obtain the final score O_I.

$$O_I = (1 - w) \times R_I + w \times S_I \tag{8}$$

finally the entire network is trained end-to-end with the cross-entropy loss function [26]. Therefore, the classification

result of each image combines the classification result of CNN and the classification result after GCN captures corresponding deep features.

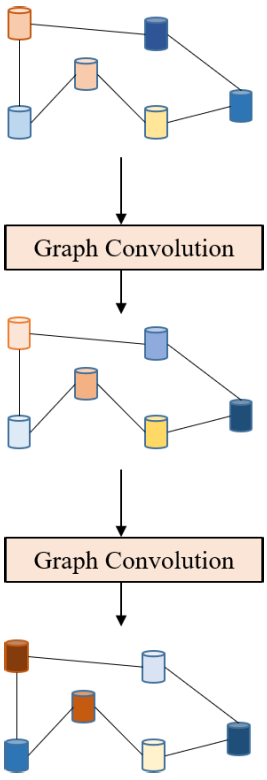


FIGURE 11. A two-layer GCN. (Different color cylinders mean different node characteristics).

IV. EXPERIMENTS

A. DATASET

To evaluate the performance of our method for DR image classification, APTOS 2019 Blindness Detection (APTOS2019) [26] and Messidor-2 [27] are used in this paper. The APTOS2019 dataset provided by the 2019 kaggle competition APTOS contains 3662 fundus images. Messidor-2 provided by Messidor project partners contains 1749 fundus images.

1) APTOS2019 DATASET

APTOS2019 was obtained by the Aravind Eye Hospital in India to screen the disease of the rural population. The full name is Asia Pacific Tele-Ophthalmology Society (APTOS) Symposium. The training and testing images were collected by multiple cameras at different times in multiple different clinics, for a total of 3662 training samples and 1928 testing samples. According to the severity of the lesions, the pictures are marked into five classes, among which the training samples have labels and the test samples have no labels. Therefore, this paper only performs five classifications on 3662 training samples (see Table 1 for data distribution details) to evaluate the performance of the model.

TABLE 1. Data distribution of the APTOS2019 dataset.

Grade	Number
Grade 0	1802
Grade 1	370
Grade 2	999
Grade 3	193
Grade 4	295
Total	3662

TABLE 2. Data distribution of the messidor-2 dataset.

Grade	Number
Grade 0	1017
Grade 1	270
Grade 2	347
Grade 3	75
Grade 4	35
Total	1748

TABLE 3. Four combinations of prediction and true categories.

True Category	Prediction Category	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

TABLE 4. Hyperparameter settings on the APTOS2019 dataset.

Hyperparameter	Value
learning rate	0.01
k	3
d1	64
d2	64

2) MESSIDOR-2 DATASET

The Messidor-2 dataset is a publicly available dataset that has been used by many organizations as a benchmark for performance testing of diabetic retinopathy image detection algorithms. The researchers graded retinal images from this dataset using the corresponding international grading reference standards, and images were annotated into five classes. The Messidor-2 dataset contains 1748 retinal color images of 874 subjects (see Table 2 for details of the data distribution).

Subjects were imaged using a color video 3CCD camera on a Topcon TRC NW6 non-dispersive fundus camera with a 45-degree field of view, centered on the fovea of the retina, at 1440 × 960, 2240 × 1488, 2304 × 1536 pixels. This paper

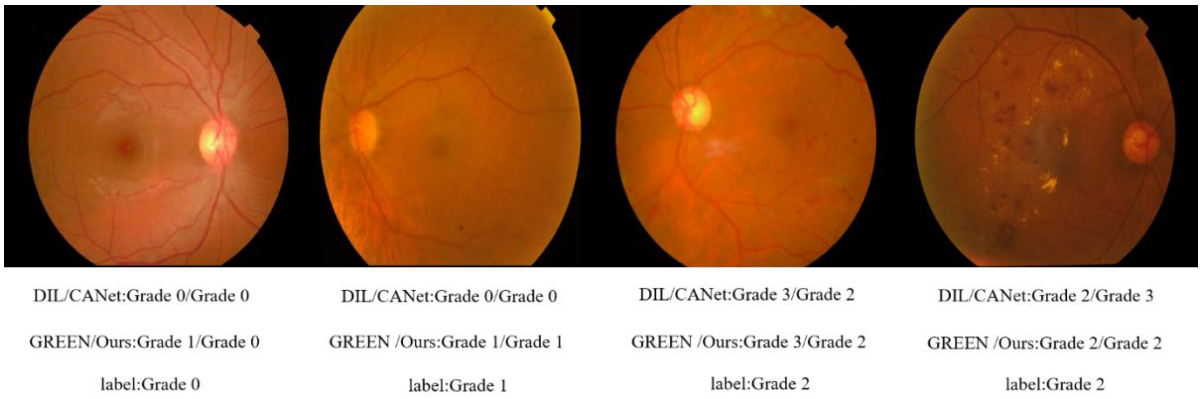


FIGURE 12. Visual evaluations on the APTOS2019 dataset. Our method is compared to DIL, CANet, GREEN, and ground truth(label). The results show our method is effective to perform DR grading.

TABLE 5. Comparison with state-of-the-art on the APTOS2019 dataset.

Method	Backbone	Evaluation Metrics		
		Acc	kappa	F1 score
DLI[30]	ResNet50	0.825	0.895	0.803
CANet[9]	ResNet50	0.832	0.900	0.813
GREEN[12]	ResNet50	0.844	0.908	0.836
Ours	ResNet50	0.848	0.909	0.843

TABLE 6. Hyperparameter settings on the messidor-2 dataset.

Hyperparameter	Value
Learning rate	0.005
k	5
d_1	64
d_2	64

evaluates the performance of our model by five classifications on the entire Messidor-2 dataset.

B. EXPERIMENTAL SETTINGS

In this paper, DenseNet121 and ResNet50 pre-trained on ImageNet are used as the backbone network for DR image classification. After the DR image is extracted from the backbone network, several 7×7 feature maps are obtained. The number of neighbor nodes k is selected from{3, 4, 5, 6, 7, 8, 9, 10}. The output dimension of the first layer GCN varies from {32, 64, 128}, and the second layer GCN output dimension varies from {32, 64, 128}. The network is based on the Pytorch framework, using SGD as the optimizer, the initial learning rate is adjusted to {0.0001, 0.0005, 0.001, 0.005, 0.01, 005}, the size of the Batchsize is determinedas 64, and the network is stopped at the 64th round. All the experiments are performed on a computer with an 3.60GHz Intel Xeon Gold 5122 CPU and a GPU of NVIDIA GTX 2080Ti.

C. EVALUATION INDICATORS

Similar to the standard setup in the literature [12], 5-fold cross-validation is performed to maximize the use of samples; The dataset was randomly divided into five equal parts, with four of them as the training set and the remaining one as the test set, for a total of five training and testing, using the accuracy rate Accuracy (Acc), Kappa value, and F1 score as evaluation indicators [28].

The combination of predicted and true classes for each sample, as shown in Table 3:

TP (True Positive): Both the prediction and true classes are positive.

TN (True Negative): Both the prediction and true classes are negative.

FP (False Positive): Predicted positive and true classes negative.

FN (False Negative): Predicted negative and true classes positive.

Accuracy represents the ratio of the number of correctly classified samples to the total number of samples for a given test data set, calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

Precision represents the percentage of accurate predictions that are positive, calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

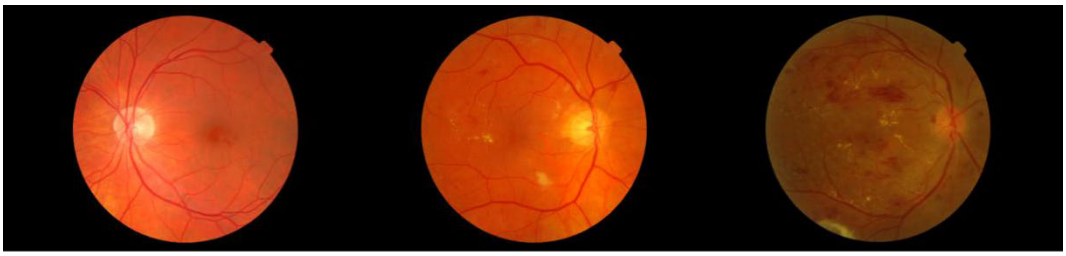


FIGURE 13. Visual evaluations on the Messidor-2 dataset. Our method is compared to DIL, CANet, GREEN, and ground truth(label). The results show our method is effective to perform DR grading.

TABLE 7. Comparison with state-of-the-art on the messidor-2 dataset.

Method	Backbone	Evaluation Metrics		
		Acc	kappa	F1 score
Base	ResNet50	0.652	0.585	0.456
Base	DenseNet121	0.655	0.627	0.627
GREEN[12]	ResNet50	0.668	0.621	0.630
GREEN[12]	DenseNet121	0.666	0.650	0.642
Ours	ResNet50	0.671	0.663	0.655
Ours	DenseNet121	0.680	0.673	0.660

TABLE 8. Experimental results in the APTOS2019 dataset.

Method	Backbone	Evaluation Metrics		
		Acc	kappa	F1 score
Base	DenseNet121	0.828	0.738	0.824
GREEN[12]	DenseNet121	0.838	0.889	0.826
Ours	DenseNet121	0.855	0.906	0.849

Recall represents the proportion of samples that are actually positive that are judged positive, calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

F1-score combines accuracy with sensitivity, calculated as follows:

$$F1 - score = 2\frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

The Kappa coefficient is an indicator used for consistency testing and can also be used to measure the effectiveness of classification. For the classification problem, the consistency is that whether the predicted result of the model and the actual classification result are consistent, calculated as follows:

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \tag{13}$$

Thereinto:

p_0: Observation consistency, percentage of consistent results of the two tests;

p_c: Expected consistency, the probability that both test results are expected to be the same.

D. RESULTS

1) RESULTS ON APTOS2019 DATASET

With DenseNet121 as the backbone network, the final settings of hyperparameters on the APTOS 2019 dataset are shown in Table 4:

We compare our method with the most typical networks for grading DR images including DLI [29], CANet [9], GREEN. DLI uses randomly augmented noise. CANet considers the relationship between DR and DME. GREEN considers DR image classes correlation for DR grading improvement. For the fairness of the experiment, our uses ResNet50 as the feature extraction backbone network, and DenseNet121 as the backbone network in the ablation experiment. Table 5 lists the experimental results on the APTOS2019 dataset. It can be seen from Table 5 that under the same backbone network, all evaluation indicators of the model in our paper are better

than other methods. In addition, this paper also verifies the superiority of our method when DenseNet121 is used as the backbone network. Fig. 12 shows the visual classification results, which can be seen from Fig. 12: Generally speaking, DLI, CANet and GREEN networks cannot accurately classify DR images, because they do not consider the relationship between the deep features of the image. Therefore, the obtained high-level semantic feature representation is not accurate enough, which affects the final classification performance. In contrast, our method learns deep features correlation of the image through two layers of GCN, and performs auxiliary adjustment on the classification results of the original network for significant DR grading improvement. Moreover, the experiments have verified that our method has achieved good results in grading DR images at different stages.

2) RESULTS ON MESSIDOR-2 DATASET

With DenseNet121 as the backbone network, the final settings of hyperparameters on the Messidor-2 dataset are shown in Table 6:

We compare our method with the most popular network GREEN [12]. Table 7 lists the experimental results on the Messidor-2 dataset. When using the backbone network for direct classification,DenseNet121 performs betterthan.

We compare our method with the most popular network GREEN [12]. Table 7 lists the experimental results on the Messidor-2 dataset. When using the backbone network for direct classification, DenseNet121 performs better than ResNet50 in all evaluation indicators. Under the same backbone network, compared with GREEN, our method has improved in all three indicators. Fig. 13 shows the visualization results. Our method is compared with the GREEN network and the real labels on the Messidor-2 dataset. The experimental results show that our method can achieve better results when grading different classes of DR images. The main reason is that our method considers the deep features correlation of the image.

E. ABLATION EXPERIMENT

One of the contributions of our model is to use GCN to capture the deep features correlation of DR images. In order to verify this contribution, ablation experiments are analyzed. Table 8 shows the experimental results of using DenseNet121 as the backbone network on the APTOS2019 dataset.

F. ADAPTIVE WEIGHTw

The second contribution of our model is to construct an adaptive weight w, that fuse the results of CNN as the original classification network with the results of learning the deep features correlation of the image through two layers of GCN(CNN as the contribution of the original classification network to the final result is w, and the contribution ratio of GCN learning the deep features correlation of the image to the final result is $1 - w$), so as to control the contribution ratio of the two-way network branch to the final classification result. To verify its effectiveness, the weight w thatbe achievable after training on the APTOS2019 dataset is 0.7365. The weight w achievable on the Messidor-2 dataset is 0.7138. From this conclusion that the original classification network has a relatively large contribution to the final result, and the network branch of GCN learning the deep features correlation of the image plays an auxiliary adjustment role in the final result, which can further improve the original classification network performance, as expected.

V. CONCLUSION

This paper proposes an intelligent grading method for DR images based on graph neural network. This method uses a CNN to learn the deep features of the image, and uses a two-layer GCN to learn the deep features correlation of the image. Therefore, our method can obtain the high-level semantic feature representation of DR images with more class information, thereby improving the performance of classification. Compared with the backbone network, the experimental results on the APTOS2019 dataset and Messidor-2 dataset are significantly improved, which shows the effectiveness of our method. Since DR Patients may also suffer from cataract, glaucoma and other diseases, and these diseases may also cause poor fundus image quality, it is a question worthy of consideration whether this will interfere with the intelligent diagnosis of DR. We considered training fundus images with cataract, glaucoma and other diseases separately in this model, and then making adjustments after checking the classification effect.

VI. FUTURE PLANS

In the future, it is hoped to further improve the classification performance of DR images, and at the same time extend this method to the diagnosis of other diseases of the eye. The following are simple steps listed to achieve the goal:

(i) Apply this DR classification method to other larger datasets;

(ii) Experiment with other deep model frameworks as backbone networks;

(iii) Extend our DR grading method to glaucoma and cataract grading.

REFERENCES

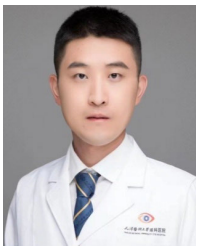
[1] Ophthalmology Society of Chinese Medical Association, "Chinese clinical guidelines for the diagnosis and treatment of diabetic retinopathy," *Chin. J. Ophthalmol.*, vol. 50, no. 11, 2014, pp. 851–865.
[2] F. Jia-Wei, Z. Ru-Ru, L. Meng, H. Jia-Wen, K. Xiao-Yang, and C. Wen-Jun, "Applications of deep learning techniques for diabetic retinal diagnosis," *J. Automatica Sinica*, vol. 47, no. 5, pp. 985–1004, 2021.
[3] (Oct. 2002). *International Clinical Diabetic Retinopathy Disease Severity Scale*. [Online]. Available: http://www.icoph.org/dynamic/attachments/resources/diabetic-retinopathy-detail
[4] U. R. Acharya, E. Y. K. Ng, J.-H. Tan, S. V. Sree, and K.-H. Ng, "An integrated index for the identification of diabetic retinopathy stages using texture parameters," *J. Med. Syst.*, vol. 36, no. 3, pp. 2011–2020, Jun. 2012.
[5] V. Gulshan, L. Peng, M., Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, and R. Kim, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016.
[6] 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.

[7] 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)*, Taiwan, Sep. 2019, pp. 1385–1389.
[8] K. Zhou, Z. Gu, W. Liu, W. Luo, J. Cheng, S. Gao, and J. Liu, "Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, Jul. 2018, pp. 2724–2727.
[9] X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "CANet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1483–1493, May 2020.
[10] A. Sakaguchi, R. Wu, and S.-I. Kamata, "Fundus image classification for diabetic retinopathy using disease severity grading," in *Proc. 9th Int. Conf. Biomed. Eng. Technol.*, Swoul, Republic Korea, Mar. 2019, pp. 190–196.
[11] D. Luo and S.-I. Kamata, "Diabetic retinopathy grading based on lesion correlation graph," in *Proc. Joint 9th Int. Conf. Informat., Electron. Vis. (ICIEV), 4th Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, Kitakyushu. Japan, Aug. 2020, pp. 1–7.
[12] S. Liu, L. Gong, K. Ma, and Y. Zheng, "GREEN: A graph REsidual RE-ranking network for grading diabetic retinopathy," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Lima, Peru, 2020, pp. 585–594.
[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[15] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.
[16] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent.*, Vancouver, BC, Canada, 2018, pp. 1–12.
[17] T. N. Kipf and M. Welling, "Variational graph auto-encoders," 2016, *arXiv:1611.07308*.
[18] H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo, "GraphGAN: Graph representation learning with generative adversarial nets," *IEEE Trans. Knowl. Data Eng.*, 2017, doi: 10.1109/TKDE.2019.2961882.
[19] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2018, vol. 32, no. 1, pp. 1–9.
[20] N. I. Kajla, M. M. S. Missen, and M. M. Luqman, "Additive angular margin loss in deep graph neural network classifier for learning graph edit distance," *IEEE Access*, vol. 8, pp. 201752–201761, 2020.
[21] N. I. Kajla, M. M. S. Missen, M. M. Luqman, and M. Coustaty, "Graph neural networks using local descriptions in attributed graphs: An application to symbol recognition and hand written character recognition," *IEEE Access*, vol. 9, pp. 99103–99111, 2021.
[22] S.-H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y.-D. Zhang, "Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network," *Inf. Fusion*, vol. 67, pp. 208–229, Mar. 2021.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.
[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, Florida, Jun. 2009, pp. 248–255.
[26] (2019). *APTOS 2019 Blindness Detection*. [Online]. Available: https://www.kaggle.kaggle.com/c/aptos2019blindnessdetection/data
[27] E. Decenciere, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, and A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed image database: The messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.
[28] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*.
[29] A. Rakhlin, "Diabetic retinopathy detection through integration of deep learning classification framework," *BioRxiv*, 2017, Art. no. 225508, doi: 10.1101/225508.

MEILING FENG received the bachelor's degree in computer science from Ludong University, in 2000, and the master's degree in computer science from Tianjin Normal University, in 2003. She is currently an Associate Professor with the School of Computer Science and Engineering, Tianjin University of Technology, China. Her current research interests include machine learning and data mining.

JINGYI WANG received the bachelor's degree in engineering from the Hebei University of Science and Technology, in 2020, and the master's degree from the Tianjin University of Technology, in 2023.

She conducted research with the Computer Vision Laboratory, Tianjin University of Technology, while studying for the master's degree. Her research interests include computer vision, deep learning, and medical image processing.



KAI WEN received the master's degree from Tianjin Medical University, China, in 2018, where he is currently pursuing the Ph.D. degree in ophthalmology.

He is currently a Researcher with a passion for the intersection of medicine and computer science. Collaborating with the School of Computer Science and Engineering, Tianjin University of Technology, he is also conducting research on lens visualization in ophthalmology. In addition, he is also involved in the basic medical research of ophthalmology. His research interests include artificial intelligence for identification of highly myopic fundus lesions and prediction of visual acuity after cataract surgery assisted by artificial intelligence.

JING SUN was selected into the Tianjin University Excellent Young Teacher Funding Program, in 2008. She was a Postdoctoral Fellow with the University of Hamburg, Germany, from 2019 to 2010. Her research interests include the combination of eye disease diagnosis and artificial intelligence, application of artificial intelligence in the calculation formula of intraocular lens power, and pathogenesis of high myopia.

She is currently a member of the Ophthalmology Branch of China Association for the promotion of medical and health care and a member of the American Academy of Ophthalmology. She is an Expert of the National College Health Education Teaching Steering Committee.