Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5441366, 11 pages
https://doi.org/10.1155/2022/5441366

Research Article

Classification of Diabetic Retinopathy Based on Multiscale Hybrid Attention Mechanism and Residual Algorithm

Yue Miao and Siyuan Tang

Department of Computer Science and Technology, Baotou Medical College, Inner Mongolia University of Science and Technology, Baotou 014040, China

Correspondence should be addressed to Yue Miao; 102007036@btmc.edu.cn

Received 16 January 2022; Revised 4 May 2022; Accepted 9 May 2022; Published 2 June 2022

Academic Editor: Mu-Yen Chen

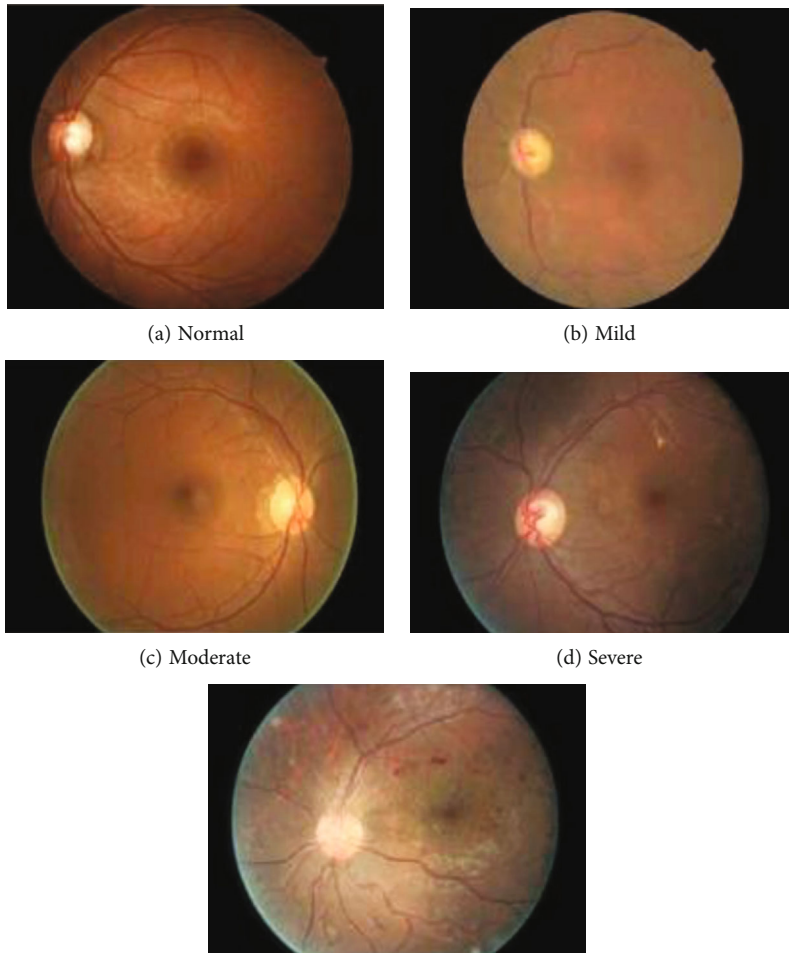
The key of classification diagnosis of diabetic retinopathy lies in the recognition of the features of small lesions, and it is difficult to extract the features of too small lesions by general extraction methods. In order to solve the problem that it is difficult to extract small focus, a hybrid attention mechanism combined with residual convolutional neural network model algorithm is proposed to improve the classification accuracy of diabetic retinopathy. Firstly, a multiscale deep learning network model with hybrid attention is designed, and then, the high-level features of images are extracted by using the network model; finally, after balancing different types of samples by sampling algorithm, the spatial attention and channel attention of the extracted features are enhanced; small-step learning strategy, loss function, and initial parameters are used to optimize the performance of the network model. The classifier based on multiscale hybrid attention network is used to judge the five classifications. Experimental results show that the proposed algorithm can learn more features of small targets and can effectively improve the classification performance of diabetic retina. An experimental test was performed on Kaggle's publicly available dataset of diabetic retinas, and the classification accuracy was 93.8%, compared to some existing classification models; the method proposed in this paper can achieve better classification results for diabetic retinopathy.

1. Introduction

With the improvement of economic level and the change of lifestyle, the incidence of diabetes is increasing year by year. Retinopathy is one of the complications of diabetes, mainly due to the long-term deterioration of small blood vessels in the retinal area. According to the degree of vascular lesions, it can be divided into two categories: nonhyperplasia and hyperplasia. Nonhyperplasia is the early stage of diabetic retinopathy (DR), which can be divided into symptomless, mild, moderate, and severe levels [1], as shown in Figure 1.

The clinical symptoms of nonproliferative DR lesions mainly include microaneurysms, hemorrhage, and soft and hard exudates. The development of the disease may enter the stage of pathological proliferation, mainly due to vascular obstruction resulting in retinal vascular hyperplasia.

In the early stage of the disease, patients cannot detect symptoms, such as symptoms, which has entered the serious stage and missed the best detection and treatment period. Therefore, early detection and intervention play a significant role in preventing vision loss or blindness caused by diabetes [2]. Regular screening can lead to early detection and treatment and slow the progression of the disease and prevent it from happening. Traditional screening mainly relies on ophthalmologists to manually grade and screen retinal images, and its screening intensity is far from meeting the needs of the present stage. The main reasons are as follows: first, there are fewer experienced ophthalmologists; second, manual screening time is long and the result feedback is slow; third, the number of diabetes in China is large. In view of the above situation, there is an urgent need for a high-precision DR automatic recognition and hierarchical diagnosis system to solve the current problems.



(a) Normal (b) Mild

(c) Moderate (d) Severe

(e) Proliferative

Figure 1: Examples of retinal images.

Early diabetic retinopathy is characterized by microaneurysms and exudation, but these symptoms are difficult to extract. This paper mainly solves the problem that small features are difficult to extract, which can be used for early screening and provide some help for doctors. The purpose of this paper is to extract enough small features from retinal fundus color images of diabetes mellitus by deep learning and try to apply them to the early diagnosis of diabetes mellitus, so as to assist doctors to make a reasonable diagnosis and treatment plan, improve the accuracy of classification of diabetic retinopathy, and reduce the burden on doctors.

2. Related Research

In the past few years, many researchers have made great progress in automatic diagnosis of DR using various algorithms, mainly using traditional machine learning algorithms. Basic operations include image preprocessing, feature extraction, and classification [3]. Feature information such as texture, color, and size of image is extracted manually, and the extracted features are input to support vector machine or random forest and other classifiers for classification and detection. Selecting the right and effective features requires expertise and adjustment of various parameters. These manually extracted features are limited and inaccurate, which will lead to wrong classification, thus affecting the classification performance of lesions and prone to misdiagnosis and missed diagnosis.

Sinthanayothin et al. [4] used Principal Component Analysis (PCA) and edge detection (EDT) to segment the blood vessels and remove background information such as optic disc. Hard exudates were detected by region growth. Sensitivity and specificity were 80.21% and 70.66%, respectively. After binarization of the image, Haloi et al. [5] used morphology to remove the blood vessel, extracted 22 features, and classified them by the support vector machine system. The hard exudates with different sizes could be recognized with a sensitivity of 96.54% and a specificity of 98.35%.

Zhang et al. [6] removed the complex background structure information such as vessels and optic disc, 27 features

were used, and the hard exudate was detected by random forest algorithm (RFA). The AUC is 0.935.

Quellec et al. [7] adopted the wavelet transform algorithm to detect the microhemangioma without removing the optic disc and blood vessel and used the template matching method for the fundus image; by detecting microhemangioma in color fundus image, green component fundus image, and vascular imager image, it is easy to mistake a small lesion for a microhemangioma. The sensitivity and specificity of the algorithm are 89.2% and 89.50%, respectively.

These algorithms are to detect a single disease; using the removal of background and other factors, using different algorithms to extract feature information, the extracted features are entered into a classifier such as a support vector machine or a random forest for classification and detection. How to select suitable and effective features depends on professional knowledge and the adjustment of various parameters. These manually extracted features are limited and inaccurate, which will lead to wrong classification, thus affecting the classification performance of lesions, prone to misdiagnosis and missed diagnosis.

In recent years, with the development of medical image processing and deep learning, deep learning technology has been applied to the detection and diagnosis of DR lesions. Deep convolutional neural network (CNN) can solve the problem of machine learning manual extraction of features which is not accurate and deep features cannot be extracted, while CNN realizes automatic extraction of deeper and more valuable features. Compared with traditional learning methods, its deep learning ability can approximate very complex functions, and its end-to-end characteristics, high accuracy, and robustness are favored by current researchers [8]. Automatic recognition and hierarchical diagnosis system based on deep learning can analyze image information more safely, accurately, efficiently, and noninvasively and can detect, locate, and classify diseases. Therefore, it is necessary to accelerate the application of deep learning in ophthalmic diagnosis, which can contribute to large-scale screening of DR patients, greatly improve clinical efficiency, and alleviate the relative shortage of medical resources [3].

In view of the above problems, under the influence of transfer learning [9], multiscale, attention, and weak supervision mechanisms [10–15], this study improved on the basis of CNN model and proposed a classification model based on residual double-attention mechanism [16], which can be a good solution to the small target difficult to extract the problem. Its main contributions are as follows. ① This paper uses a fusion of attention mechanism and inception module for the DR classification network model algorithm [17], which can strengthen the weight of small lesions and improve the accuracy of classification by modifying the loss function of model training. ② The residual mechanism is added [18]. The core of the residual mechanism uses cross-connection mode to avoid the loss of information transmitted in the layer and the disappearance of gradient, which can greatly accelerate the training of the deep neural network and improve the accuracy of the model.

3. DR Classification Method

The classification of DR lesions based on deep learning is mainly divided into three stages, as shown in Figure 2: image pretreatment stage, classification model training stage, and detection classification stage.

3.1. Image Preprocessing and Data Enhancement. RGB images collected from hospitals have many problems, so data preprocessing is needed before network training. Good and bad image quality has a great influence on the results of retinopathy classification. Prior to feature extraction, preprocessing is crucial to help identify lesions and distinguish the extent of actual lesions, thus improving the accuracy of DR lesion detection. Compared with the blue and red channels, the green channel has the most image information, the largest gap between the green channel and the background, the best contrast and the lowest noise. Green channel images are more conducive to image segmentation and classification, and the Contrast Limited Adaptive Histogram Equalization (CLAHE) method is adopted to enhance contrast. There is better detection of exudates and blood vessels. Illumination correction is for illumination irregularity to improve lumen and brightness of the image. Gaussian filtering and other denoising methods are used to smooth the image, and threshold method is used to delete meaningless black borders. However, network training requires a lot of data, and data enhancement is achieved by image mirroring, rotation, resizing, and clipping.

3.2. Loss Functions Deal with the Problem of Unbalanced Dataset Classes. The dataset used in this study has the problem of category imbalance, which will affect the accuracy of the model. When the sample number of a certain category is small, the proportion of the loss value generated is also small, which does not conform to the characteristics of good performance of all classification categories in the multiclassification model. To solve this problem, the common strategy is resampling the dataset. The problem caused by resampling is oversampling or undersampling, which makes some data lack or reuse problems. Therefore, in this study, the improved loss function is mainly used to punish the classification learning model and modify the learning cost of samples to optimize the network model. On the basis of the original multiclassification focal loss function, regular terms are added to form a loss function suitable for this problem. Formula (1) is the multiclassification focal loss function.

$$L_{\mathrm{MCFL}} = -\sum_{i=1}^{n} \alpha y' (1-y)^{\gamma} \log\left(y'\right). \tag{1}$$

Formula (2) is the improved multiclassification DR-focal loss function

$$L_{\mathrm{DR\text{-}MCFL}} = -\sum_{i=1}^{n} \alpha y' (1-y)^{\gamma} \log\left(y'\right) + \lambda \left|\left|y' - y\right|\right|_2^2. \tag{2}$$

The regular term is added on the basis of the original loss function, which can solve the problem of the influence of

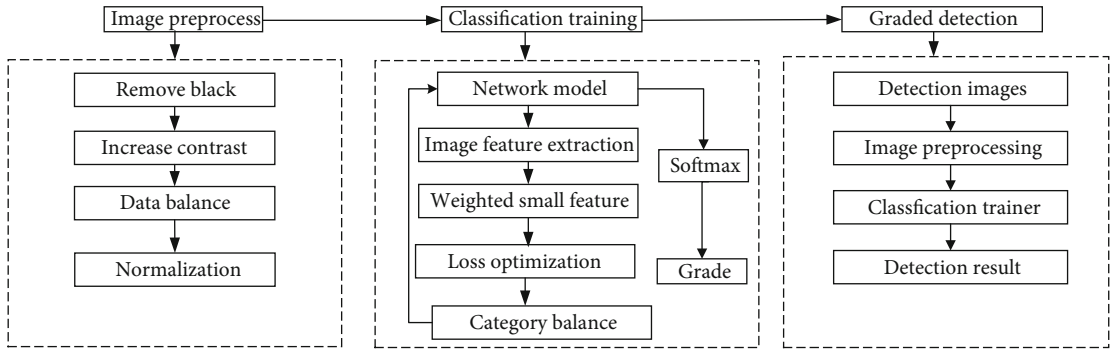


Figure 2: DR classification framework.

class imbalance on classification network learning. n is the total number of categories, i indicates a certain category, y' is the predicted result, y is the label value of the sample, α is the equilibrium factor, and λ is the coefficient of the weight of the regular term. The loss function can also accelerate the convergence of the network by adding a second norm as a regular term.

3.3. Related Principles and Mechanisms. DR lesions in the nonhyperplasia stage mainly show symptoms such as hemangioma, fundus hemorrhage, vitreous hemorrhage, and exudate, and the characteristic information of these symptoms is sparse. The traditional network model is not ideal for the detection of small lesions, and the accuracy and performance need to be further improved. In view of sparse lesion areas, attention mechanism can be used to better highlight the information of small lesion feature images, so as to extract richer features. In the network model, residual blocks were added and the information of the front output layer was used as the input of the back layer to prevent the loss of the features of the lesion region and further improve the detection ability of the features of small retinal lesions. Finally, multiclassification is carried out by the Softmax function to accurately achieve the classification of diabetic fundus images at five levels (healthy, mild, moderate, severe, and proliferation).

Based on the above reasons, this study added inception and attention modules on the basis of the basic deep learning network model and combined with the residual thought of ResNet [18]. The core of ResNet uses a cross-connection approach that avoids the loss of information and gradient disappearance problems transmitted in layers, which can greatly speed up training for deeper networks and improve the accuracy of the model. This model can extract more important features under the same amount of computation, so as to improve the training results and make more efficient use of computing resources.

3.3.1. Attention Mechanism. Human perception of the world does not process everything it sees, but rather makes sense of the world around it by capturing the parts that stand out. It is based on human's understanding of the world that this principle is applied to deep learning. Attention mechanism is widely used in natural image and natural language processing. There are two attention mechanisms, namely, spatial domain and channel. The channel mechanism focuses on the importance of the channel, while the spatial attention mechanism focuses on the importance of different positions on the same channel. If the two mechanisms are combined, the channel can be paid attention to, and the weight of different features on the channel can be given, and the problem of difficult extraction of small features can be solved by increasing the feature weight of small features, so as to improve the classification accuracy of the model [15].

3.3.2. Channel Attention Mechanism. For the natural image in the input convolutional neural network, there are two attributes, in which length and width are the scale space of the image, and the other attribute is the channel. The principle of channel attention mechanism is to firstly reduce channel dimension. After obtaining feature information through maximum pooling and average pooling, respectively, the two parts are splicing together to form a feature map by sharing multilayer perceptron, and then, the weight value is normalized to 0-1 by the sigmoid function. After the calculated weight matrix value is weighted by multiplying the original channel image, the importance of different channels can be finally learned [19], as shown in Figure 3.

(1) The algorithm flow of the channel attention mechanism principle is as follows

(2) Feature input: assuming that the feature graph of the input is represented by $F(H, W, C)$, H is height, W is width, and C is channel

(3) Cycle processing: for $w = 1, 2, \cdots \cdots, W$, for $h = 1, 2, \cdots \cdots, H$, and for $c = 1, 2 \cdots \cdots, C$. Repeat the following operations to perform global average pooling and maximum pooling. Global pooling reduces dimension and maximum pooling extracts more influential channels, as shown in

$$F^{C}_{\text{chanel avg}} = \frac{\sum_{h=1}^{H}\sum_{w=1}^{W}\left(F^{C,H,W}\right)}{H \times W}, \tag{3}$$

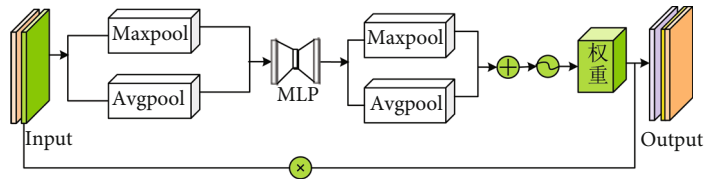


Figure 3: Channel attention mechanism diagram.

$$F^{C}_{\text{chanel max}} = \frac{\sum_{h=1}^{H}\sum_{w=1}^{W}\left(F^{C,H,W}\right)}{H \times W}. \tag{4}$$

(4) Generate channel attention weight value: after adding the maximum pooling layer and average pooling layer through multilayer perceptron operation containing two fully connected layers, the sigmoid function is used to calculate channel attention weight value, as shown in

$$M_c(F) = \text{sigmoid}\left(\text{MLP}\left(F^{C}_{\text{chanel avg}}\right) + \text{MLP}\left(F^{C}_{\text{chanel max}}\right)\right). \tag{5}$$

(5) Formula (6) is used to calculate the final channel attention graph F':

$$F' = F \times M_c(F). \tag{6}$$

3.3.3. Attention Mechanics in Space. In the aspect of space domain, it mainly deals with feature dimension. After the channel attention mechanism, the contribution degree of image features on each channel is also different; that is, the importance degree of each channel is different, and the importance of features on each channel is also differentiated. Through the spatial attention mechanism, features with high contribution on this channel can be found. The specific principle is that global average pooling and global maximum pooling are also used to obtain two feature graphs, and then, a convolution kernel is used to form a new feature graph after convolution. More critical and important feature information can be obtained by multiplying the weight value of the feature graph by the weight of the original image through the normalization of the sigmoid function. The attention mechanism can focus more attention on more important features, which is helpful to extract smaller and more difficult feature information [20], as shown in Figure 4.

3.4. Residual Principle of the Module. With the increase of the layers of the deep learning network model, the gradient explosion or gradient disappearance will occur. The main reason is that in the process of network backpropagation, the gradient value may be infinite or zero due to the nonlinear change, which makes the network model either in the state of training stagnation or you are in a state where the parameter value keeps increasing indefinitely. The overall training stability of the network will become very poor. In order to solve this problem, residual module is adopted in this research. The basic principle is to superimpose the output of shallow layer network on the output of deep layer network to protect the primitiveness and integrality of the characteristic information; learning the difference between its input and output simplifies the difficulty of network training. The strategy is illustrated in Figure 5.

Let X be the output of shallow layer, $H(x)$ is the deep output, $F(x)$ is the transformation represented by the middle two layers of the two, and then, the formula is $H(x) = x + F(x)$.

According to the above formula, when the output of the shallow network is superimposed on the output of deep network, when the network converges to the global optimal solution, the mapping of output layer reestablishes a new channel relational mapping of input to output, and the mapping of original layer is set to 0. After the characteristic information contained in shallow layer X was fully learned through the network, if the parameter adjustment of the back layer made the loss function tend to increase after the change of X, the loss function tended to be 0 through the residual connection channel, and x continued to be transported to the next hidden layer from the identity mapping. In the forward propagation of the network, the training speed of the shallow network layer is faster and easier than that of the layer network layer. Therefore, the training speed of the deep network layer can be accelerated by mapping the features learned at the shallow network layer to the corresponding positions of the deep network layer. In network backpropagation, gradient propagation is faster in the deep network layer due to residual connection branches, and gradient upward can be transmitted by an activation function with the help of residual connection paths. The introduction of residual connection will reduce the parameter values in the module layer and make the parameters in the network more sensitive to the loss function under reverse propagation. Therefore, the convergence of loss function in the network can be accelerated so that the training time of the network is shorter and the training efficiency is higher. Residual connection also regularizes the network.

3.5. Residual Double Attention Module Principles. The residual double attention module is mainly composed of residual, channel attention, and space attention [20]. Deep learning problems often encountered in the explosion problem are gradient disappear or gradient; the usual solution is to initialize the data standardization and batch, at the same time also can bring when depth deepening and the network

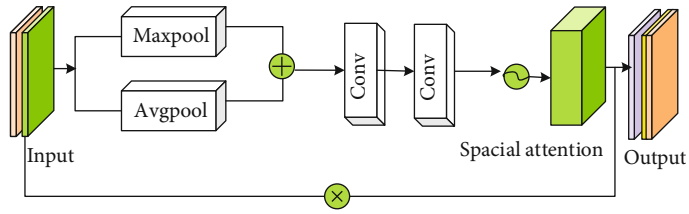


Figure 4: Spatial attention mechanism diagram.

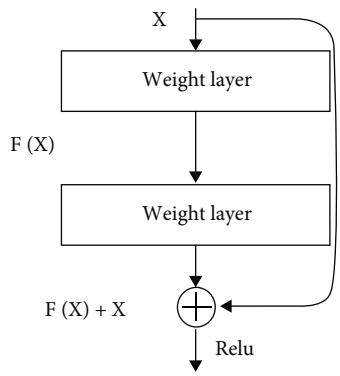


Figure 5: Residual schematic diagram.

performance problems, and the residual is mainly used to solve the problem of gradient and also can improve network performance and reduce the error rate; the specific formula is

$$H_{i,d}(x) = (1 + M_{i,c}(x)) \times F_{i,c}(x). \tag{7}$$

H is the feature of the output of residual attention, M is the feature of the attention mechanism, and F is the attention function. Different functions will extract different attention fields. F in Formula (8) stands for attention in the mixed domain.

$$F_{i,c}(x) = \frac{1}{1 + \exp(-x_{i,c})}. \tag{8}$$

F in Formula (9) stands for attention in the channel domain.

$$F_{i,c}(x) = \frac{x_{i,c}}{\|x_i\|}. \tag{9}$$

F in Formula (10) stands for attention in the spatial domain.

$$F_{i,c}(x) = \frac{1}{1 + \exp(-(x_{i,c} - \text{mean}_c)/\text{std}_c)}. \tag{10}$$

3.6. Model Structure. Through the combination of attention mechanism and residuals, the problem of difficult extraction of small features in focus was solved. In the experiment, using the dual attention mechanism of channel and space, not only the features on the important channel can be extracted but also the small features in different spaces can be focused. In addition, the residual parameter module is used in the model, which can not only reuse the low-order features of the image but also generate new high-order composite features continuously. There are two main modules in the model: inception module and attention module, as shown in Figure 6.

IRCSB has inception module and attention module [21], as shown in Figure 7. The three different scale convolutional layers used in the inception module can capture more locally diverse information, which is then fused, and finally, features are extracted by using a 1×1 convolutional layer for perception of different scales. The inception module main multiscale extraction feature, first through a 1×1 convolution extraction feature, the convolution of two 3×3s, a 3×3 convolution, a pool, and a 1×1 convolution are concatenated into the later attention module, thus paying attention to both the important information in the channel and the characteristic information in the space, to ensure the integrity of the information. The different scale convolutions used in the inception module can capture more local and diverse information and then fuse this information, combined with the attention mechanism to increase the weight of small features, which can extract smaller features. In the attention module, there are channel attention (CA) and space attention (SA); CA is mainly used to extract the importance of different channels, as shown in Figure 4. SA is mainly used to focus on areas with high-frequency information and calculate the importance of different areas. The structure of SA is shown in Figure 6.

4. Network Model Training and Testing

The research of this paper is to implement the training and testing of the whole model in Python programming language and PyTorch framework, as shown in Figure 8. The hardware environment of the experiment is as follows: CPU: Intel core i9-9980XE @ 3.00 GHz × 362; graphics card: NVIDIA GeForce RTX 2080 Ti × 4; memory: 128 GB.

The data used in the experiment came from the public dataset of Kaggle Competition [18–20]. The dataset is a large number of high-resolution retinal images taken under various imaging conditions, with a total of 35,126 images, with resolutions of 1440×960, 2240×1488, and 2304×1536, respectively. There are many problems in the image, such as artifacts, lack of focus, underexposure, or overexposure. In addition, there are also problems of data imbalance in different levels of images. The dataset was graded by the clinician on each DR lesion image, which was divided into five scales, with an integer 0-4 to represent the severity of the

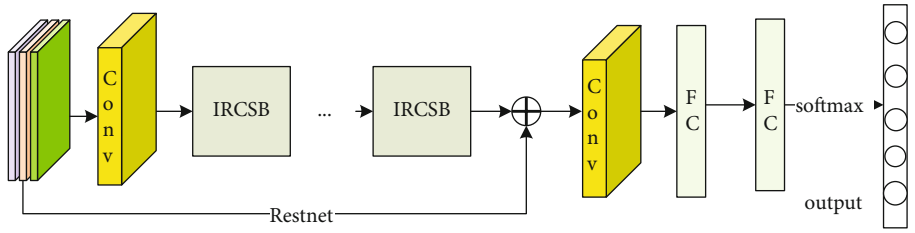


Figure 6: Model network structure chart.

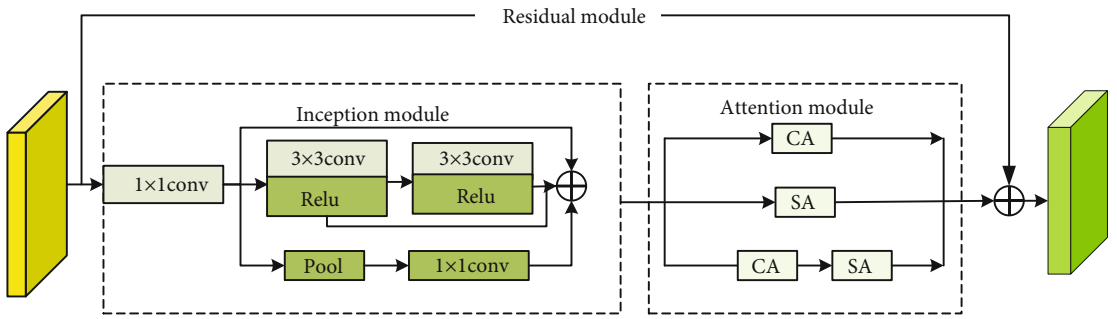


Figure 7: IRCSB network structure chart.

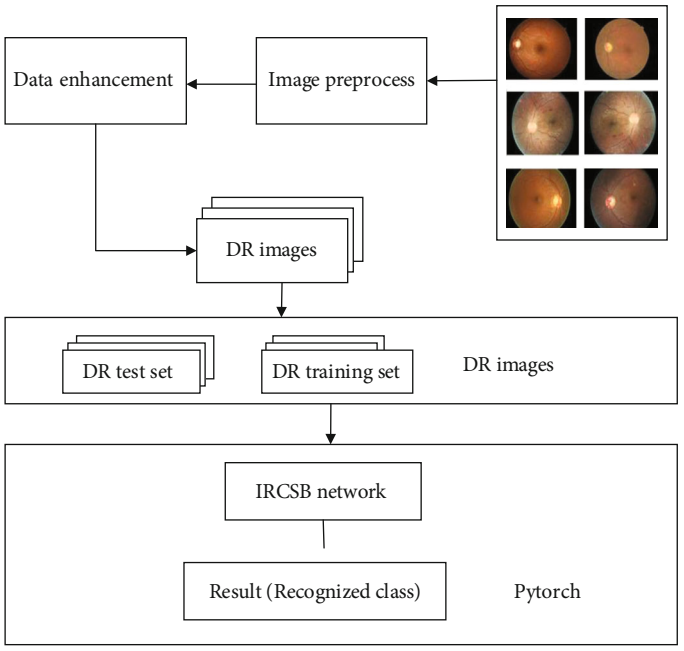


Figure 8: DR diagnosis technique using IRCSB network structure.

lesion, namely, asymptomatic, mild, moderate, severe, and proliferative. There were 35,126 images in the test dataset, including 25,810 without DR symptoms, 2443 mild NPDR, 5292 moderate NPDR, 873 severe NPDR, and 708 proliferation maps. There is a serious imbalance between different categories, and image enhancement is used for categories with relatively little data. The details are shown in Table 1.

4.1. Image Preprocessing. After histogram equalization, remove black borders from all images. Because images have different tones and lighting, color and lighting should be balanced to improve the robustness of the model. Otherwise, feature extraction of lesions will be affected in the later stage. Image denoising is to remove the noise in the image without blurring the edges, as shown Figure 9. Median filter is used to remove noise and retain some image features such as discontinuity, edge, or line.

4.2. Image Cutting Processing. The common lesions in eye images mainly include microhemangioma, hard exudate,

Table 1: Results of datasets.

Lesion level	Quantity (sheet)
Normal	25,810
Slight	2443
Average	5292
Serious	873
Proliferation	708
Total	35,126

hemorrhagic spots, cotton patch, and neovascularization. These small lesion areas are difficult to extract features. For the features of small lesion areas in the eyes are not easy to extract, it is necessary to do some cutting processing on the image. Firstly, the original high-resolution image was scaled to a suitable size (480×480) to extract the global features of the image. Then, the original image is scaled to get a subimage with a size of 1000×1000, which is cut into four parts. Then, the four images are scaled to 480×480. Local features of the image can be extracted by using these four small images. Finally, five 480×480 images will be obtained as the input data of the network. The network can extract both global information and local small features so that the convolutional network can fully extract more useful image features. In the experiment, in order to speed up the training of convolutional network, data were normalized, as follows:

$$X = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \tag{11}$$

5. Experimental Results and Analysis

5.1. Experimental Parameter Settings. The SGD optimizer is adopted, the initial learning rate $\mathrm{lr} = 2e - 3$, the learning rate is adjusted dynamically according to the change of loss, and the $l1$ regularization is added. Different activation functions such as Sigmoid, Tanh, ReLU, and LReLU are used to test the accuracy of the network. Finally, LReLU activation function is selected to accelerate the convergence of the network, epochs = 100, batch size = 32, and the learning strategy is step. Four activation functions, Sigmoid, Tanh, ReLU, and LReLU, are used to compare the experimental results, which are shown in Table 2.

5.2. Evaluation Index. For multiclassification problems, the evaluation criteria of sensitivity, specificity, and precision, three indicators, were used to evaluate the experimental results. True positive rate (TPR) refers to the probability of being correctly predicted in actual positive samples, and true negative rate (TNR) refers to the probability of being correctly predicted in actual negative samples, also known as recall rate. Precision is the probability of being correctly predicted in positive samples of predicted results. The AUC is the area under the curve; it is a kind of performance index to measure the classification quality.

$$\text{Accuracy(ACC)} : \mathrm{ACC} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})}, \tag{12}$$

the proportion of the number of samples correctly classified by the model to the total number of samples.

$$\text{True positive rate(TPR)} : \mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}, \tag{13}$$

the percentage of positive samples that are correctly classified among all positive samples.

$$\text{True negative rate(TNR)} : \mathrm{TNR} = \frac{\mathrm{TN}}{(\mathrm{FP} + \mathrm{TN})}, \tag{14}$$

the percentage of negative class samples that are correctly classified among all negative class samples.

$$\text{Precision(precision)} : P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{15}$$

the proportion of positive samples in positive examples determined by the classifier.

$$F1 \text{ score} : F1 = \frac{(2 \times P \times \mathrm{TNR})}{P + \mathrm{TNR}},$$

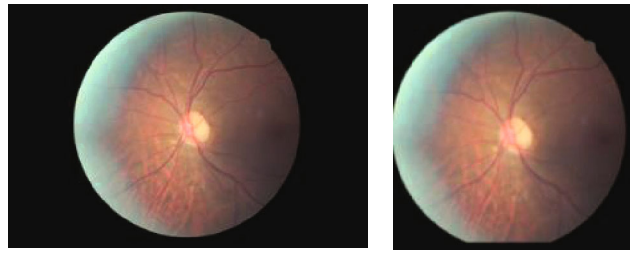
$$\text{AUC(area under the curve)} : \mathrm{AUC} = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i)(y_i + y_{i+1}). \tag{16}$$

The vertical axis of the receiver operating characteristic curve is the true positive rate (TPR), the horizontal axis is the false positive rate (FPR), and the AUC is the area under the curve; it is a kind of performance index to measure the classification quality.

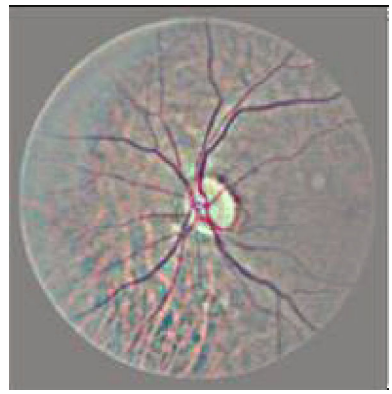
5.3. Experimental Results. When training the network, the learning rate has a certain impact on the convergence of the model. The learning rate is set to 0.002, 0.05, 0.1, and 0.5, respectively. After continuous test and comparison, it is determined that the initial learning rate is 0.002, and the total number of iterations tends to converge when it is about 1400, which not only ensures the convergence of the loss function but also makes the classification accuracy reach the highest, and the accuracy reaches 93.8%.

5.4. Experimental Comparison. This paper compares the classification performance of multiscale mixed attention model with several traditional models, other deep learning models, and approximate structure models and further verifies the effectiveness of the classification model proposed in this paper.

5.4.1. Compared with Machine Learning Algorithm. Machine learning algorithm method is useful for industry [22, 23] and medical field. Power tools can be diagnosed using the developed method. In the study of recognition of diabetic retinopathy using BP neural network algorithm, the main factors that affect the classification result of neural network are the number of hidden layers and the selection of excitation function. The BP neural network has 8 layers; it consists of an



(a) Original image (b) Remove black frame



(c) Image enhance

Figure 9: DR image preprocess result.

Table 2: Experimental results of different activation functions.

Activation function	Accuracy (%)	AUC
Sigmoid	89.5	0.87
Tanh	90.3	0.88
ReLU	92.8	0.91
LReLU	93.0	0.93

Table 3: Results of different network models on datasets.

Evaluation results	BP	SVM	Proposed
SE	0.85	0.86	0.934
SP	0.84	0.85	0.962
ACC	0.87	0.88	0.938

input layer, six hidden layers, and an output layer. The full connection between layers is used, and three different excitation functions, hyperbolic tangent function, sigmoid function, and quasilinear function, are used to test the results. Finally, the sigmoid function was chosen, and its prediction accuracy was up to 87%.

In SVM, kernel function is the final classification effect. Kernel function mainly includes polynomial kernel, radial basis kernel, and sigmoid kernel. These functions were tested, respectively, repeated 100 times, the final selection of radial basis function kernel, the error is the smallest, and the accuracy is 88%.

Table 4: Results of different network models on datasets.

Model	SP	SE	AUC	ACC
AlexNet	89.07%	79.01%	0.7988	81.76%
LeNet	86.32%	82.45%	0.88.52	87.02%
GoogleNet	90.34%	88.46%	0.9188	92.01%
Proposed	96.2%	93.40%	0.9205	93.80%

This experiment is compared with the traditional machine learning algorithm classifier, and the experimental results are shown in Table 3. The results show that the accuracy rate, true positive rate, and false positive rate of classification have been improved by using the network model of this scheme, and the misclassification rate has been effectively reduced, further confirming the feasibility of this scheme.

5.4.2. Comparison with Other Deep Learning Algorithms. This paper makes an experimental comparison with LeNet, AlexNet, and GoogleNet. LeNet contains two convolution layers and two full connection layers, with a total of 60,000 learning parameters. The accuracy rate on this dataset is 87.02%. In the AlexNet network model, there are 5 convolution pooling layers, 3 full connection layers, and 1000 neurons in the output layer. The first convolution uses a larger core size of 11×11 with a step size of 4. The core size of the subsequent convolution layer is relatively small, 5×5 or 3×3, with a step size of 1. The accuracy on this dataset is 81.76%. GoogleNet won the first place in the 2014

ImageNet challenge. Through the concept multiscale structure, it not only expands the depth and width but also improves the utilization of computing resources. This experiment compares with the three network structures of LeNet, AlexNet, and GoogleNet. The comparative experiment shows that the accuracy of the multiscale hybrid attention network proposed in this paper is higher than that of the other three network structures, and the convergence speed is much faster than that of the other two network structures. The overall effect is the best. The results are shown in Table 4.

6. Conclusions

Due to the increasing number of diabetic population and cases of retinopathy, there is an increasing demand for automated DR diagnostic systems. However, there are still some problems in the direct application of these DR systems in clinical practice, so it is urgent to develop a more reliable and practical automatic DR diagnostic grading system to help clinicians do auxiliary examinations [24]. We propose an automatic DR classification system based on attention and residual parameter mechanism. Combined with inception multiscale module feature extraction, the problem of small feature extraction is solved. Finally, it achieves the effect of 5 classification with an accuracy of 93.8%, which is greatly improved compared with the traditional machine classification algorithm. The results show that the algorithm used in the model is superior to the other two algorithms in the recognition of diabetic retinopathy. The contribution of this study is that the diabetic retina can be early screened; it can not only reduce misdiagnosis caused by human factors but also greatly shorten the time of diabetic retinopathy diagnosis, which is of great clinical significance in preventing visual loss and treatment.

The shortage of this study is the use of less external datasets to verify, which needs to use multicenter data to verify the results of this model, so there is a certain gap in clinical application. What we can do on the basis of this study is to use more general data to verify the results of this model. In the future, we can also use multimodal data to realize the study, use multimodel integration or fusion method to train the model, and further improve the classification accuracy.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Consent

Informed consent was obtained from all individual participants included in the study.

Conflicts of Interest

All authors declare that they have no conflict of interest.

Authors' Contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was supported by No. 2021MS06010. This work was also supported by No. NJZY21068.

Acknowledgments

The authors acknowledge the help from the university colleagues.

References

[1] R. Ma and Z. Lu, "The value of ophthalmological imaging in the early diagnosis and treatment of diabetic retinopathy," *Imaging Research and Medical Application*, vol. 1, no. 12, 2017.

[2] L. Yin and H. Peng, "Progress in the treatment of diabetic retinopathy," *Modern Medicine and Health*, vol. 33, no. 1, pp. 80–83, 2017.

[3] S. K. Somasundaram and P. Alli, "A machine learning ensemble classifier for early prediction of diabetic retinopathy," *Journal of Medical Systems*, vol. 41, no. 12, p. 201, 2017.

[4] C. Sinthanayothin, V. Kongbunkiat, S. Phoojaruenchanachai, and A. Singalavanija, "Automated screening system for diabetic retinopathy," in *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*, pp. 915–920, Rome, Italy, 2003.

[5] M. Haloi, S. Dandapat, and R. Sinha, "A Gaussian scalespace approach for exudates detection, classification and severity prediction," *Computer Science*, vol. 56, no. 1, pp. 3–6, 2015.

[6] X. Zhang, G. Thibault, E. Decencière et al., "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Medical Image Analysis*, vol. 18, no. 7, pp. 1026–1043, 2014.

[7] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1230–1241, 2008.

[8] F. Meng, W. Yin, and J. He, "Detection of bleeding points in fundus images based on deep learning," *Journal of Shandong University (Science Edition)*, vol. 55, no. 9, pp. 62–71, 2020.

[9] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Shanghai, China, 2017.

[10] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, p. 3361, MIT Press, 1995.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[12] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, and G. Yang, "Hierarchical retinal blood vessel segmentation based on feature and ensemble learning," *Neurocomputing*, vol. 149, pp. 708–717, 2015.

[13] P. Ding, Q. Li, Z. Zhang, and F. Li, "Deep neural network classification method for diabetic retina images," *Computer Applications*, vol. 37, no. 3, pp. 699–704, 2017.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, Computer Science, http://arxiv.org/abs/1409.1556v6.

[15] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *IEEE 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, 2015.

[16] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," 2017, Thirty-First AAAI Conference on Artificial Intelligence, http://arxiv.org/abs/1602.07261.

[17] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," *Computers and Electrical Engineering*, vol. 72, pp. 274–282, 2018.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision & Pattern Recognition*, IEEE Computer Society, 2016.

[19] K. Xu, J. Ba, R. Kiros et al., "Show, attends and tells: neural image caption generation with visual attention," in *Proceedings of the 32nd International Conference on Machine Learning. PMLR 37*, pp. 2048–2057, New York, 2015.

[20] S. Woo, J. Park, J. Y. Lee, and I. Kweon, "Clam: convolutional block attention module," *Proceedings of the European conference on computer vision (ECCV)*, vol. 63, pp. 3–19, 2018.

[21] Y. Miao, S. Tang, P. Du, and Z. Li, "Research on deep learning in the detection and classification of diabetic retinopathy," in *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, pp. 107–113, Fuzhou, China, 2021.

[22] A. Glowacz, "Ventilation diagnosis of angle grinder using thermal imaging," *Sensors*, vol. 21, no. 8, p. 2853, 2021.

[23] A. Glowacz, "Thermographic fault diagnosis of ventilation in BLDC motors," *Sensors*, vol. 21, no. 21, p. 7245, 2021.

[24] M. Chen and D. Gong, "Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on Tensor Flow framework with feature selection," *Journal of Investigative Medicine*, vol. 67, Suppl 1, p. A3, 2019.