ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Design of Intelligent Diagnosis System for Retinal Diseases Combined with Deep Learning

Huan Liu[1] *and Lei Wang*[1,*]

[1] The 967th Hospital of the Chinese People's Liberation Army Joint Logistic Support Force, Ophthalmology Department, Dalian, Liaoning, 116011, China

SUMMARY: *Scale-up screening of retina is important for prevention, early diagnosis and timely intervention and treatment of related diseases. In this paper, the construction of intelligent diagnosis system for retinal diseases is based on deep learning technology tools. The fundus image is pre-processed by performing region segmentation and extraction, image denoising and region feature enhancement. Based on the structure of retinal layers, the encoding of retinal boundaries is carried out through a dual boundary representation with the existence of complementary constraint relations, and multi-task consistency constraints are proposed correspondingly. Under the retinal layering of the dual-task boundary concern method, a fully convolutional network is introduced to extract a lesion feature and feed back the structural and spatial location information of this feature. The average performance of the system solution for boundary location segmentation of lesion images (9.80±7.26) is extremely similar to that of manual labeling data (8.08±5.66), and the average accuracy of segmentation of the four major lesion metrics is as high as 0.6773, which is capable of assisting in the localization and detection of lesion features with accurate and efficient retinal layer segmentation performance.*

KEYWORDS: *retinal disease diagnosis; fully convolutional network; double boundary representation; deep learning; region segmentation*

1 Introduction

The retina is important part of the eye as it captures light and transforms it into nerve impulses that travel to the brain and form visual images [1]. Nevertheless, there exist a wide range of factors that may cause a retinal disease, and they frequently cause vision problems and other symptoms and signs of eyesight [2]. It is difficult to detect these conditions early, which makes it necessary to diagnose retinal diseases quickly and accurately to be able to act in time and stop additional vision loss [3, 4]. Conventional diagnostic approaches depend on the manual analysis of medical images, which is tedious and highly subject to the knowledge and personal opinion of the physician [5, 6]. During the past several years, the accelerating progress of deep learning technologies and increasing demand of effective medical diagnostics have brought the creation of smart diagnostic systems of retinal diseases as the major theme of study in the medical sphere [7, 8].

Deep learning is a branch of artificial intelligence that uses artificial neural networks to replicate the architecture and processes of the human brain, allowing it to learn complex patterns in raw data, and to classify and predict [9, 10]. The creation of smart diagnostic

*wl18847502872@163.com
https://doi.org/10.65102/is2026477
Published: 30 April, 2026

systems to treat retinopathy with the help of deep learning is both a challenging task and one of paramount importance. With the choice of appropriate deep learning models, the improvement of the quality and size of data, the training of the system on a large set of fundus images, as well as the enhancement of system interpretability, it will be feasible to automatically identify and determine different retinal disorders and their severity [11-13]. It would enable healthcare workers to enhance the precision of diagnosis and the quality of care of patients, which could result in better patient treatment results. However, there are still issues, including the lack of available data, its quality, and standardized evaluation. Additional efforts are required to make deep learning models and algorithms more accurate and consistent to get a more precise and steady diagnostic system.

Various research articles have investigated the application of deep learning in the diagnosis of retinal diseases and have shown the efficacy of smart diagnostic systems. In another paper [14], a deep learning-based diagnostic system was developed to diagnose retinal diseases using fundus images and it was tested with a multi-retinal disease dataset. It performed better than both conventional diagnostic approaches, being more accurate. Another paper [15] dealt with retinal vein occlusion (RVO) where a deep learning model was created to recognize and diagnose RVO using fundus images. This system successfully detected the occurrence of RVO and categorized it depending on its occlusion location. Work in [16] has also described a deep learning-based automated diagnosis of different types of retinal diseases based on fundus images. It was noted that this approach may help the ophthalmologists make more accurate clinical decisions and save time and be efficient. The use of a deep learning-based healthcare system in the diagnosis of diabetic retinal diseases was discussed in [17]. The results indicated that the system had the ability of independently detecting and determining the severity of retinal diseases and that this could provide useful information to the medical practitioners. To overcome the difficulty of early diagnosis in retinal diseases, [18] suggested a deep learning-based automatic disease localization and segmentation method. Comparative analysis showed that this method is especially useful in diagnosing retinal diseases at their initial stage. An article in [19] studied the role of image-based deep learning algorithms in diagnosing retinal diseases and experiments conducted on retinal OCT images proved that the method was highly sensitive and accurate. Finally, the idea of deep learning models to diagnose retinal diseases was addressed in [20], highlighting that it can increase the diagnostic ability of ophthalmologists and improve the overall medical diagnosis of retinal conditions.

Moreover, research work [21] discusses different eye disorders including cataract, retinal disorders and glaucoma, and emphasizes the need to diagnose them properly as it can improve the quality of life of patients. The proposed solution is a deep learning solution, which has proven to be effective in terms of accuracy, recall, and precision. An analogous solution can be seen in [22], where an automated deep learning framework to diagnose diabetic retinopathy is presented. The system has been shown to be able to accurately identify and diagnose early retinal disorders in diabetic patients. In [23], the emphasis is made on enhancing early detection and care of retinal conditions, and a deep learning diagnostics tool known as DeepDR is created. As is evidenced by the experimental findings, this system improves the diagnostic accuracy at all the stages of retinal disease. Moreover, the importance of abnormal sign identification in retinal fundus photographs is highlighted as being crucial to the successful management of eye disorders in [24]. It investigates the capability of a diagnostic system of retinal diseases implemented using deep learning to confirm its diagnostic effectiveness. Other work [25] talks about the important contribution of artificial intelligence towards diagnosing common diseases and suggests the application of deep learning to diagnose and refer retinal diseases. As per the findings, such an approach

demonstrates diagnostic performance which is at least equal to experts. A joint algorithm in [26] combining deep learning is proposed to be used in the classification and diagnosis of retinal diseases. Ablation experiments and significance tests are used to test the effectiveness of the algorithm, and it has better recall, accuracy, and precision. Lastly, the literature [27] presents a deep learning algorithm, which can diagnose retinal diseases in diabetic patients. The analysis of 500 fundus images with specific disease indicators shows that the algorithm is highly accurate and is much more efficient in terms of diagnosing such conditions.

This paper uses the maximum inter-class variance approach in preprocessing of fundus images to divide the retinal region and determine the final region of interest. The median filtering is used to reduce noise and restricted contrast adaptive histogram equalization is used to increase the contrast of different regions of the fundus image. Also, a multi-task dual-boundary attention network, a custom dual-boundary representation, and loss function are proposed to come up with an automated segmentation of the retinal layers. In order to enhance retinopathy detection even more, a fully convolutional network is employed to effectively capture the features related to retinopathy, which helps to create a detection algorithm. Besides, the system also includes such operations as flattening and blood vessel extraction to establish an intelligent diagnosis system of retinal diseases.

2 Preprocessing of fundus images

To correctly extract lesions features and provide accurate diagnoses in a computer-aided diagnostic system, the preprocessing of the fundus image is necessary. Since the background of retinal images is very complex and there is an overlap of blood vessels with lesion areas, the effectiveness of preprocessing will also directly affect the precision of later feature extraction. It further influences the final results of the diagnostic process of the system. Therefore, the preprocessing techniques of the fundus images have the highest priority. They generally consist of three main stages: extracting the region of interest, denoising filtering, and improving the image quality.

2.1 Region of interest extraction

The region of interest (ROI) in image processing is set as the focus of attention for image analysis, and in fundus images doctors mainly focus on the circular or oval retinal region. The region of interest of the fundus image is first extracted using threshold segmentation to remove the black background portion around the retinal region.

In the color fundus image, the green channel has the clearest vascular contour and edge area, which is commonly used for vascular extraction or lesion segmentation, and the red channel has the most obvious contrast with the surrounding black background area, so the red channel image is thresholded. By selecting an appropriate threshold value and comparing the size of each pixel's gray value with the threshold value in turn, the pixels in the image are classified into two categories so that the gray scale image of the red channel is converted into a binary image. The binary image data matrix contains only 0 (black) and 1 (white), i.e., the background region and the target region of interest. Let the original image be $f(x, y)$, the thresholded image be $g(x, y)$, and T be the selected threshold value, then we have equation (1):

$$g(x,y) = \begin{cases} 1, & f(x,y) > T \\ 0, & f(x,y) \le T \end{cases} \tag{1}$$

The thresholding method separates the region of interest and other background regions in the image better by selecting the best threshold value suitable for the image. Since different fundus images have different gray value distributions and fixed thresholds cannot be widely applied to different images, this paper adopts the maximum interclass variance method to adaptively calculate the optimal thresholds based on the fundus images to realize the extraction of regions of interest. The retinal region is initially segmented by the maximum interclass variance method, and the final region of interest is extracted by the maximum connectivity domain. In which the maximum inter-class variance method (OTSU) is realized as follows:

Assuming that the region of interest and other regions in an image are separable, the larger the value of the inter-class variance between the two means the larger the difference in the gray value of their connected regions. Assuming that the threshold of an image is thr, the proportion of the region of interest is ω_0, and its average gray value is μ_0; the proportion of the background region is ω_1, and its average gray value is μ_1, and the total gray value of the image is μ, and the value of the interclass variance of the objective function is $g(t)$, there is equation (2):

$$g(t) = \omega_0 \times (\mu_0 - \mu)^2 + \omega_1 \times (\mu_1 - \mu)^2 \tag{2}$$

For ease of calculation, take equation (3):

$$g(t) \approx \omega_0 \times \omega_1 \times (\mu_0 - \mu_1)^2 \tag{3}$$

The value that maximizes the interclass variance $g(t)$ is computed, at which point t is the threshold value that applies to the entire image. In this paper, the threshold is set to $0.3thr$ to segment all retinal regions as completely as possible.

2.2 Denoising Filter

In order to shield the interference factors introduced during image digitization and make its visual effect better, this paper adopts the median filtering method for image filtering, so as to reduce the interference brought by environmental factors to the fundus image to a certain extent.

Median filtering is a kind of nonlinear filtering realized by logic operation, which is usually used to shield irregular discrete interference factors. The original image is assumed to be $f(x, y)$, and the median filtered image is $g(x, y)$, with (x, y) as the center pixel point, the square, diamond and other small windows are selected according to the need, and all the pixels contained therein are arranged according to the grayscale value in ascending order, and the values of the original pixels are replaced by the intermediate values if the number of pixel points is odd; otherwise, the values of the two intermediate pixels are rounded to one-half of the values of the two intermediate pixels and the resulting values are rounded to one-half of the values of the two intermediate pixels. Otherwise, the value obtained by rounding one-half of the sum of the two intermediate pixel values is used as the gray value at (x, y). Let W be the window size of the template, then there is equation (4):

$$g(x, y) = med\left\{ f(x - i, y - j), (i, j \in W) \right\} \tag{4}$$

2.3 Image Enhancement

Image enhancement makes different lesion features more visible and suppresses unimportant areas and irrelevant information. Unlike the simple enhancement of ordinary images, the contrast enhancement of darker and brighter areas of the fundus image should be considered at the same time. And the image enhancement process should try to maintain the characteristics of different regions of the original fundus image without introducing new noise. In this paper, the restricted contrast adaptive histogram equalization is used for image enhancement.

Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is a combination of histogram equalization and adaptive histogram equalization algorithms. Histogram equalization increases image contrast by redistributing the pixel gray levels and making their probability distribution more uniform on the 0-255 gray scale. Nonetheless, this approach does not modify the various regions of the retina in an adaptive manner, hence restricting its utility to analyze the images of the retina. Conversely, adaptive histogram equalization subdivides the image into non-overlapping regions and performs histogram equalization to each region. Although this solution can make local information more prominent, it ignores the rest of the pixels in the image. Moreover, this method may also increase the amount of noise in each segmented block when used on noisy images, which will further degrade the quality of the image as a whole.

Contrast-limited adaptive histogram equalization limits the magnitude of local contrast enhancement by restricting the height of the local histogram, effectively suppressing the amplification of disturbing information while avoiding the over-enhancement of local regions. The principle of limiting contrast histogram equalization is shown in Fig. 1, which sets a threshold for limiting contrast to the height of the histogram of each region in the image, crops out the part that exceeds the threshold, and distributes it evenly to the gray values from 0 to 255.

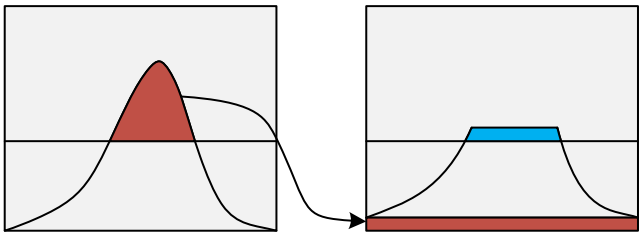


Figure 1: The principle of CLAHE

3 Intelligent diagnosis of retinal diseases based on fundus images

3.1 Retinal Layer Segmentation with Multitask Boundary Attention

3.1.1 Multi-tasking architecture

This paper has examined a number of existing image segmentation algorithms applied to the problem of retinal layer segmentation, and found that the process of boundary region segmentation often failed. This paper will present a multitask framework that includes boundary tasks to help solve the problem of segmenting boundaries, and increase the accuracy of retinal layer segmentation. There are three tasks involved in the multitask framework on retinal layer segmentation, one main segmentation task, and two boundary regression tasks. In

contrast to classical boundary representations, this method creates a pair of boundary representations which incorporate the spatial relationship between neighboring retinal layers. They are highly correlated such that the total of pixel values in similar positions in both boundary maps is always unity. On top of that, the paper presents a complementary loss term between predictions of these two boundary representations to enhance learning the boundaries. Moreover, to promote the mutual supervision among segmentation task and boundary tasks, a consistency loss function is created. This algorithm assigns image pixels into 13 classes: 11 retinal layers, cumulus and back ground. The specifics of the proposed retinal layer segmentation network are presented below.

3.1.2 Double boundary representation

The retinal segmentation task in the present paper divides the retinal OCT image into 11 layers of the retina, together with the effusion and background. Various attempts have been made to enhance the segmentation results by incorporating boundary representations to learn features at the boundaries via encoding. Nonetheless, many of these approaches are based on straightforward pixel distance contexts to determine the boundary coding, which complicates the process of capturing hierarchical relationships among neighboring retinal layers in OCT images. As one example, the background always remains over the first retinal layer. Taking the advantage of this relational property, we will introduce a double-boundary representation to focus more on delicate boundaries. It is calculated upon the smallest distance of a pixel to the border of a neighbor layer. The dual-boundary representation proposed in this paper has both higher and lower boundaries representation. Prior to defining these representations, two variables, BR_{upper}^{dis} and BR_{lower}^{dis} are introduced as given in Equations (5)-(8):

$$BR_{upper}^{c}(x)=\begin{cases}\inf_{y\in B_{upper}^{c}}\|x-y\|, & GT^{c}(x)=1\\ 0, & otherwise\end{cases} \tag{5}$$

$$BR_{upper}^{dis}=cat\left(BR_{upper}^{1},BR_{upper}^{2},\cdots,BR_{upper}^{C-1}\right) \tag{6}$$

$$BR_{lower}^{c}(x)=\begin{cases}\inf_{y\in B_{lower}^{c}}\|x-y\|, & GT^{c}(x)=1\\ 0, & otherwise\end{cases} \tag{7}$$

$$BR_{lower}^{dis}=cat\left(BR_{lower}^{1},BR_{lower}^{2},\cdots,BR_{lower}^{C-1}\right) \tag{8}$$

Here GT uses One-hot coding and $GT^{c}(x)$ refers to the xth pixel point of the cth channel of the segmentation mask. B_{upper}^{c} and B_{lower}^{c} refer to the upper and lower boundaries of the cth retinal layer, respectively. The $\|x-y\|$ represents the Euclidean distance between the point x and the point y. The function $\inf$ serves to obtain the smallest value from the set to get the minimum distance from the pixel to the upper and lower boundaries of the retinal layer where it is located. The cat function combines multiple one-channel data into multi-channel data. Based on BR_{upper}^{dis} and BR_{lower}^{dis}, the double-boundary representation in this paper is defined as Eqs. (9)-(10):

$$BR_{upper} = \frac{BR_{upper}^{dis}}{BR_{upper}^{dis} + BR_{lower}^{dis}} \tag{9}$$

$$BR_{lower} = \frac{BR_{lower}^{dis}}{BR_{upper}^{dis} + BR_{lower}^{dis}} \tag{10}$$

The equation above shows that the values of BR_{upper} and BR_{lower} are between 0 and 1, which makes the representation even stronger. Besides, the upper and lower boundary representations are interrelated, making the representation even more powerful. Also, this sum of two boundary representations may be converted to segmentation results, enabling reciprocal support with segmentation forecasts. The dual-boundary representation is used on the eleven retinal layers and the cumulus as well as the background giving rise to BR_{upper} and BR_{lower} forming $C-1$ channels.

3.1.3 Loss function

The network employs four different loss functions. The first of them is the segmentation loss, which is used on the segmentation task. The second one is the boundary loss, which is used in the boundary regression task. Also, boundary complementarity loss is added as an imposed mutual constraint on the two boundary representations. Lastly, multi-task consistency loss is used to guarantee that the boundary regression and segmentation tasks are well limited against one another.

(1) Segmentation loss

In contrast to the conventional cross-entropy loss, this method uses Dice loss in supervising segmentation predictions. Dice loss is used because it can deal with class imbalance better and delivers higher segmentation accuracy especially in small objects. Equation (11) gives a formal definition of Dice loss.

$$L_{segmentation} = 1 - \frac{1}{C}\sum_{c=0}^{C-1} \frac{2P^c G^{c+\tau}}{P^c + G^{c+\tau}} \tag{11}$$

Here GT^c denotes the c th channel of the true segmentation label and P^c denotes the c th channel of the segmentation prediction mask. C is equal to 13 and denotes the number of channels, corresponding to the number of categories segmented. τ is a constant close to zero to prevent computational anomalies.

(2) Boundary loss

Due to the fact that the algorithm contains two boundary regression problems, the boundary loss consists of two parts: a part related to the prediction of the upper boundary and the other to the prediction of the lower boundary. The definition of the boundary loss based on Mean Squared Error (MSE) is given by equations (12)-(14).

$$L_{upper} = \frac{1}{C-1}\sum_{c=1}^{C-1} \left\| PB_{upper}^c - BR_{upper}^c \right\|^2 \tag{12}$$

$$L_{lower} = \frac{1}{C-1}\sum_{c=1}^{C-1} \left\| PB_{lower}^c - BR_{lower}^c \right\|^2 \tag{13}$$

$$L_{boundary} = L_{upper} + L_{lower} \tag{14}$$

Here PB_{upper}^{c} and PB_{lower}^{c} are the c th channel where the upper boundary represents the prediction result and the lower boundary represents the prediction result, respectively, and BR_{upper}^{c} and BR_{lower}^{c} are the c th channel where the upper boundary represents the true label and the lower boundary represents the true label, respectively. The c starts from 1 instead of 0 because the background region is not involved in the computation in this process.

(3) Boundary Complementary Loss

The boundary representations equation means that, on every pixel, the sum of the upper and lower boundary representations will be always 1. This special feature is one of the crucial novelties of the algorithm. In order to improve the capability to capture boundaries, a boundary complementarity loss is defined, which is presented in equation (15).

$$L_{complementation} = \frac{1}{\left|PB_{add}\right|} \sum_{c=1}^{C-1} \sum_{x \in PB_{add}^{c}} d\left(PB_{add}^{c}(x)\right) \tag{15}$$

where there are equations (16)-(17):

$$PB_{add} = PB_{upper} + PB_{lower} \tag{16}$$

$$d\left(PB_{add}^{c}(x)\right) = \begin{cases} \left\| PB_{add}^{c}(x) - 1 \right\|^{2}, & GT^{c}(x) = 1 \\ 0, & otherwise \end{cases} \tag{17}$$

Here PB_{upper} and PB_{lower} are the prediction results of the upper and lower boundary representations, respectively, and PB_{add} is the result of doing addition operation for the double boundary representation. The $PB_{add}^{c}(x)$ denotes the x pixel of the c th channel of PB_{add} , and $|PB_{add}|$ refers to the number of pixel points of PB_{add} .

(4) Multi-task consistency loss

Since the sum of the two boundary predictions may be expressed as a result of the segmentation process, there is a certain high level of consistency between both the boundary representations and the segmentation. Thus, a multi-task regression loss is created to promote the mutual supervision of the boundary regression task and the segmentation task. The said loss function is given in equation (18) with the consistency loss representing the difference between the double-boundary representation prediction and the segmentation prediction.

$$L_{consistency} = \frac{1}{C-1} \sum_{c=1}^{C-1} \left\| PB_{add}^{c} - P^{c} \right\|^{2} \tag{18}$$

The consistency loss calculates the difference in consistency between the double-boundary representation predictions and the segmentation predictions.

(5) Total loss function

During network training, the total loss function combines the four loss functions mentioned above, as expressed in equation (19).

$$L_{total} = \omega 1 L_{segmentation} + \omega 2 L_{boundary} + \omega 3 L_{complementation} + \omega 4 L_{consistency} \quad (19)$$

Here $\omega 1$, $\omega 2$, $\omega 3$ and $\omega 4$ are used to balance the loss weights for importance between different losses. Experiment based attempts were finally set to $\omega 1 = 10$ and $\omega 2 = \omega 3 = \omega 4 = 1$ in the algorithm implementation.

3.2 Retinopathy detection algorithm based on fully convolutional network

3.2.1 Overview of the methodology

Aiming at the traditional retinal lesion detection algorithms that are difficult to cope with the problems of coexistence of multiple lesion structures, huge differences in lesion structures, and the interference of retinal blood vessels, this paper designs a retinal lesion detection algorithm based on a fully convolutional network, which consists of three functional modules, namely, convolution module, local contrast enhancement module, and anti-convolution module. First, the convolutional module can extract rich multilevel and multiscale features from the original input retinal OCT images, so as to effectively represent various retinal lesion structures; in addition to this, in order to improve the network's ability to detect various retinal lesion structures, a local contrast enhancement module is further designed, which is able to locally contrast-enhance the features of the retinal lesion structures; finally, an Finally, the inverse convolution module is used to gradually map the high-level abstract semantic features back to the spatial resolution of the original input OCT image, so as to recover the detailed spatial location information of retinal lesion structures.

3.2.2 Retinopathy detection network

All the functional modules in the retinopathy detection network (LDN) proposed in this paper can form a 4 × 5 grid structure (except for the Global module and the Score module), in which the features processed in each column of the grid have the same spatial dimensions, while the spatial dimensions of the features processed in each row vary (e.g., 224 × 224, 112 × 112, ..., 7×7). The structure of each row in the grid and its features are described in detail below.

The first row of the grid consists of five convolutional modules (Conv-1 to Conv-5), which are derived from all the convolutional and pooling layers of the VGG16 network, and each module contains two or three convolutional layers as well as a pooling layer. The convolutional layers are used to extract multilayer and multiscale features from the input OCT image, and the pooling layer is able to reduce the spatial resolution of the feature map to 1/2 of the original. For example, the Conv-1 module consists of two convolutional layers with convolutional kernel size 3×3, 64 feature maps, and a 2×2 maximum pooling layer. After these 5 convolutional modules, the original input OCT image is downsampled 32 times and the spatial resolution is gradually reduced from 224×224 to 7×7. The rightmost part of the first row is the Global module, which uses three convolutional layers with convolutional kernel size 3×3 and number of feature maps 128 to extract the global features X_G of the input retinal OCT image.

The second row of the grid also contains five convolutional modules (Conv-6 to Conv-10), which are connected to the five convolutional modules Conv-1, Conv-2, ..., Conv-5 in the first row of the grid, each of which contains a convolutional layer with a convolutional kernel size of 3×3 and a number of feature maps of 128 . These 5 convolutional modules in the second row of the mesh can reduce the number of convolved feature maps in the first row of the mesh, thus greatly reducing the number of parameters of the entire LDN network. These 5

convolutional modules are able to extract local convolutional features $X_i\left(i=1,2,\cdots,5\right)$ from different layers respectively, which are used as inputs to the third row of the grid.

The third row of the grid consists of five local contrast enhancement modules (Contrast-1 to Contrast-5), which can effectively enhance the local contrast information of the lesion structures, thus improving the network's ability to detect various lesion structures. In general, there is a significant contrast difference between retinal lesion structures and their surrounding non-lesion regions (e.g., cystic lesions often appear as dark black hollow structures and fibrous scar lesions often appear as long, white, striped structures with a large luminance of the RPE layer, relative to the surrounding normal retinal tissue), which means that there is also a significant feature difference between the features of the lesion region in the CNN network and those of its surrounding region. Therefore, a local contrast enhancement module is employed for each local convolutional feature X_i separately to perform local contrast enhancement on the features of the lesion structure. Specifically, each local contrast enhancement module $Contrast-i$ is implemented as in Eq. (20) by subtracting its 3 × 3 average pooled features from its local convolutional feature X_i:

$$C_i = X_i - AvgPool\left(X_i\right) \tag{20}$$

where $C_i\left(i=1,2,\cdots,5\right)$ represents the local contrast features output by the local contrast enhancement module, and $AvgPool\left(\cdot\right)$ represents the 3×3 average pooling operation.

The fourth row of the grid contains a series of deconvolution modules (Deconv-2 to Deconv-5) that are capable of gradually increasing the spatial resolution of the feature map from 7×7 (lower right) to 112×112 (lower left) to recover the spatial location information of the lesion structure. The inverse convolution module takes as input the local convolution feature X_i, the local contrast feature C_i, and the up-sampling feature U_i, and it splices the three features in series in the channel dimensions and inputs the spliced feature maps into an inverse convolutional layer which is capable of doubling the spatial resolution of the input feature map, a process that can be expressed as equation (21):

$$U_i = Deconv\left(X_i, C_i, U_{i+1}\right) \tag{21}$$

where U_i is the up-sampled feature output from this inverse convolution module, U_{i+1} is the up-sampled feature from the previous inverse convolution module, and $Deconv\left(\cdot\right)$ represents a convolution kernel size of 5×5 with a step size of 2 for the inverse convolutional layer with a step size of 2. In the lower left corner of the mesh is a Local module, which uses a convolutional layer with a convolutional kernel size of 1×1 and 640 feature maps to extract the final local features X_L as in equation (22):

$$X_L = Conv\left(X_1, C_1, U_2\right) \tag{22}$$

The rightmost side of the network has a Score module, which consists of two convolutional layers with convolutional kernel size 1×1 and the number of feature maps is 2. These two convolutional layers deal with the global features X_G and local features X_L extracted from the LDN network, which respectively map the global features X_G and local

feature X_L into two category scores, respectively.

The two category scores are first added and then combined, then the resultant category scores are processed through the Softmax layer. It produces the probability of a pixel being assigned to one of the categories, thereby coming up with a lesion probability map. On this map, every spatial position value shows the probability that the respective pixel of the input OCT image is an element of a lesion structure. For a fixed retinal OCT image, the lesion probability map computed by the LDN network has a spatial size of half the resolution of the original input image. This map is upscaled with bilinear interpolation to make it the same spatial resolution as the input image. Loss of the LDN network is the cross-entropy loss, which is given by Equation (23).

$$Loss = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{m}\sum_{c\in\{0,1\}} I\left(g_i = c\right)\log p\left(c\left|p_i\right.\right) \tag{23}$$

where N is the total number of training images, m is the total number of pixels per image, g_i is the true labeling of pixel p_i, and $p\left(c\left|p_i\right.\right)$ is the predicted probability that pixel p_i belongs to category c.

4 Construction and validation of an intelligent diagnostic system for retinal diseases

This chapter optimizes the preprocessing of fundus images through retinal flattening and blood vessel extraction, highlights the hierarchical features of fundus images, reduces layered interference factors, and assists in improving the efficiency of fundus segmentation images. Combining this preprocessing method with the intelligent diagnosis method of retinal diseases based on fundus images, an intelligent diagnosis system of retinal diseases under deep learning is constructed.

4.1 Retinal Planarization and Vessel Extraction in OCT Fundus Images

An OCT scanner was used as the fundus image scanning instrument in this paper, and the fundus image preprocessing method proposed in Chapter 2 was used to design the experimental OCT images for retinal flattening and blood vessel extraction, and to narrow down the layered work area.

The gradient value changes in three different dimensions of the retina under OCT images are shown in Fig. 2, in which the two peak points with the largest gradient in each column and not adjacent to each other are selected as the upper and lower baseline positions in the column, respectively. The endpoints of the upper baseline are used as the reference for moving each column in the vertical direction to complete the upper level flattening, and then the endpoints of the lower baseline are used as the reference for stretching and transforming all the pixels between the upper and lower baselines on the result of the upper level flattening. The transformation relationship of each column is composed into a transformation matrix, which can be used to find the edge position in the original image by inverse transformation according to the position of the boundary points after the layering is finished.

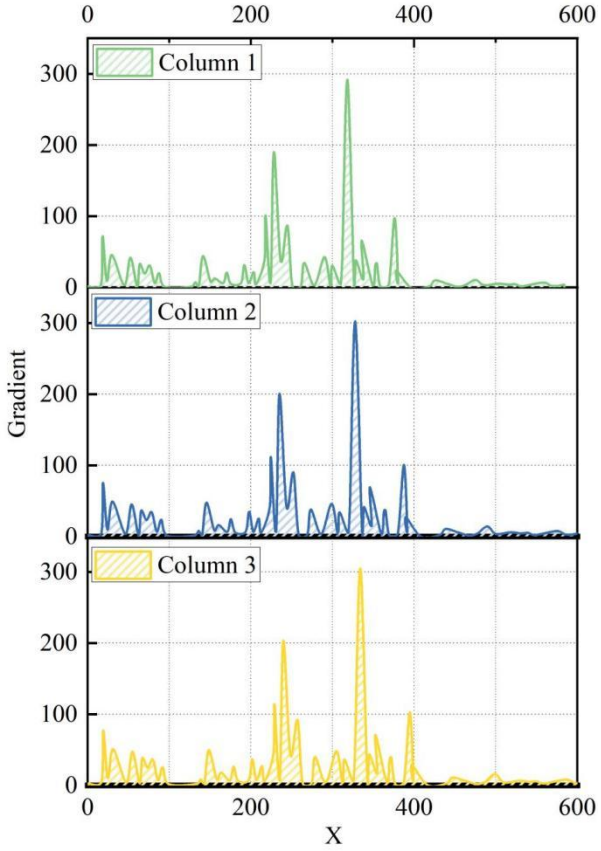


Figure 2: The gradient values of the retinal image columns change

Since blood vessels appear as dark rod-shaped regions on OCT images, the RPE and OS layers, which are better flattened, have larger gray values and are all crossed by blood vessels, and the contrast between the two is obvious. Therefore, in this paper, the extraction of blood vessels was performed in the RPE and OS layers. The grayscale projection statistics corresponding to the sampling results of the above flattened OCT fundus images are shown in Fig. 3. Combined with the sampled images, it can be seen that the peak points after grayscale projection in Fig. 3 correspond to the edge positions of retinal blood vessels.

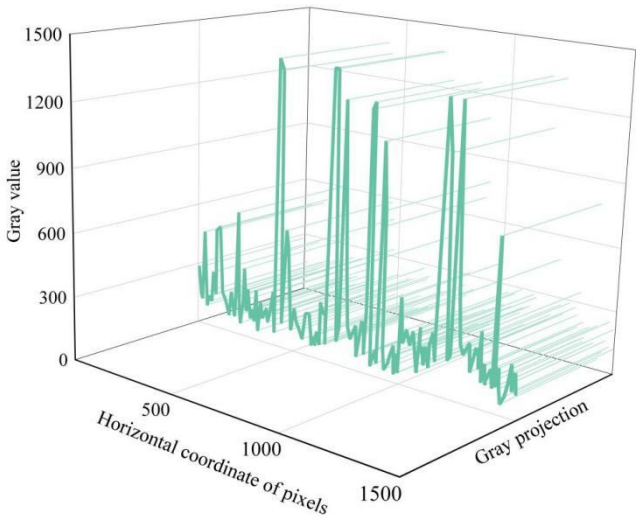


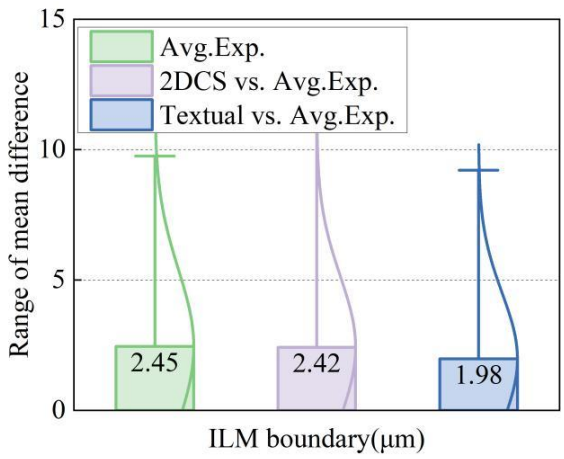
Figure 3: Grayscale projection in the sampling interval

4.2 Effect of retinal layer segmentation for different eye images

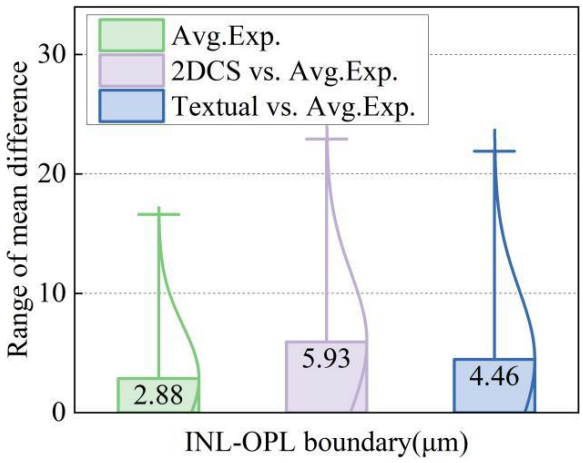
4.2.1 Segmentation of the normal eye

2DGS, 3DGS and MS3DGS were selected as the control retinal layer segmentation algorithms, and combined with the retinal layer image segmentation algorithms in this paper to compare the segmentation effects of OS-PRE boundaries, ONL-IS boundaries and BM boundaries of fundus images. Two clinical ophthalmologists were also invited to manually mark the retinal tissue layer boundaries of OCT images (Exp. 1, Exp. 2), and the manual marking results of the two doctors were averaged to form a standard (Avg. Exp.).

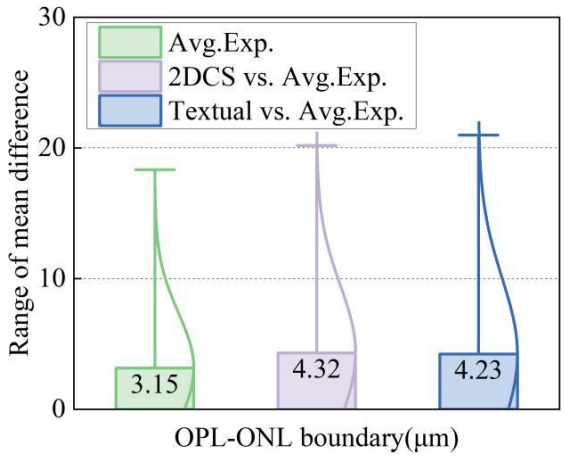
Three eyes with a total of 316 images were randomly selected from the dataset containing 25 normal eyes for quantitative analysis, and the mean errors of the boundary locations of a total of three sets of OCT images were plotted in Fig. 4(a)-(f), where the data labels are the average of the algorithm segmentation results. The recognition effect of this paper's algorithm on a total of five retinal boundary combinations, ILM, IML-OPL, OPL-ONL (4.23), OS-PRE and BM, is the closest to the average manual labeling among the four algorithms, with the mean difference of the mean values being, in order, 1.98, 4.46, 4.23, 2.83, and 2.40, which preliminarily verifies the retinal layer under the multi-task boundary network segmentation validity.



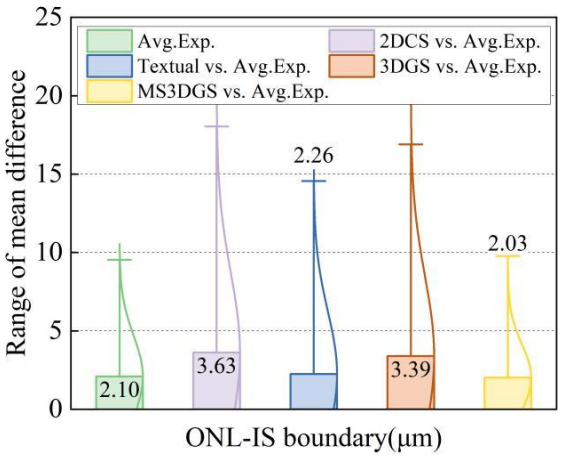
(a) ILM



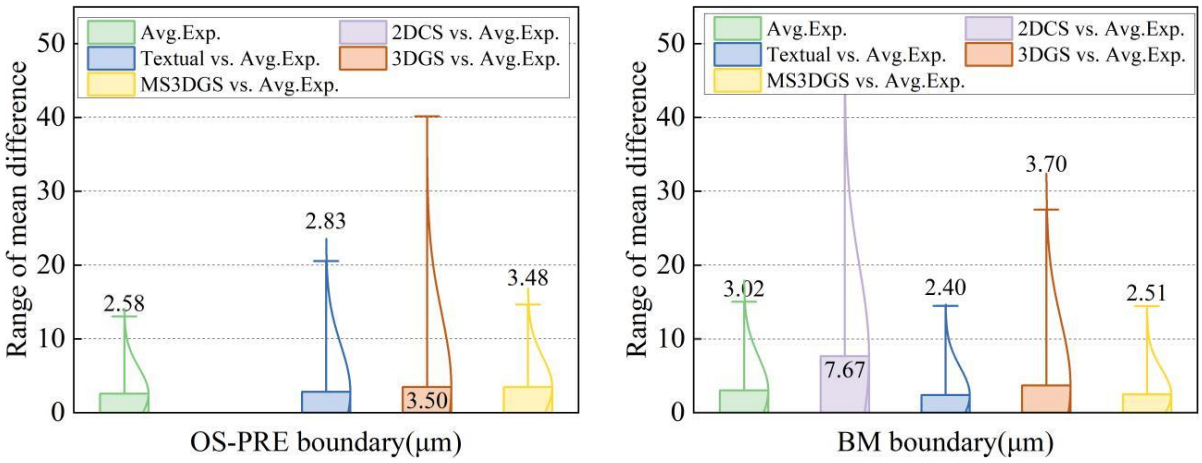
(b) IML-OPL



(c) OPL-ONL



(d) ONL-IS



(e) OS-PRE (f) BM

Figure 4: Mean error of the boundary of the retinal tissue layer

The retinal thickness differences are further listed in Table 1, and only the comparison between this paper's algorithm and the 2DGS method is listed because the 3D method only segments the IS-OS boundary and RPE layer of the lesion image. The algorithm in this paper is within 1.50 of the mean manual labeling results of the two physicians and 0.250 of the mean manual labeling results, which is better than the 2DGS method.

Table 1: The differences in retinal thickness

Retinal Layer	Textual vS.Exp.1	Textual vs.Exp.2	Textual vs.Avg.Exp.	2DGS vs.Avg.Exp.	Exp.1 vs.Exp.2
RNFL-INL	1.23±0.75	3.08±0.04	2.12±0.18	1.58±0.61	1.87±0.72
OPL	5.08±1.60	6.04±1.28	5.57±0.72	2.77±4.00	0.91±0.28
ONL	7.02±1.25	4.77±1.31	5.68±0.76	7.61±2.42	1.75±0.02
IS-OS	1.39±0.98	1.25±0.17	1.28±0.08	NULL	0.06±0.85
RPE	0.12±2.91	1.52±2.21	0.82±1.83	NULL	1.33±0.62
Total retina	1.31±0.56	0.82±0.22	0.25±0.27	12.89±0.92	2.16±0.32

4.2.2 Segmentation results of lesion images

From the dataset containing the classical 20 lesion images, a set of images with both lesions was selected for quantitative evaluation. The above four methods are still used to compare with manual segmentation, and the results of the absolute error of the boundary position are shown in Table 2.For the lesion images, the errors of manual segmentation and automatic segmentation methods are relatively large, mainly because the existence of lesions makes the corresponding retinal tissue layers deformed, and even some retinal tissue layers can not be identified with the naked eye, which causes some difficulties in the layer segmentation, so the segmentation results are relatively large compared with the normal eyes. Therefore, the segmentation results have a larger error than normal eyes. However, compared with other methods, the algorithm in this paper has the lowest absolute error in boundary position (9.80±7.26), which is most similar to the absolute error in manual labeling (8.08±5.66). Overall, the algorithm in this paper is able to achieve optimal segmentation results for almost all boundaries, regardless of normal eye or lesion images.

Table 2: Comparison of absolute errors in the boundary positions of lesion images

Retinal Layer	ONL-IS	OS-RPE	BM	Mean
Exp.1 Vs Exp.2	7.81±6.15	8.31±6.27	8.11±4.57	8.08±5.66
2DGS vs Exp.1	10.46±5.46	11.71±7.67	9.17±5.68	10.45±6.27
Textual Vs Exp.2	11.21±10.08	11.01±7.61	7.17±4.10	9.80±7.26
3DGS Vs Avg.Exp.	10.41±5.46	11.50±7.54	8.09±4.76	10.00±5.92
MS3DGS Vs Avg.Exp.	10.43±7.66	12.36±8.72	7.57±4.63	10.12±7.00

4.3 Performance evaluation of retinopathy detection algorithms

In the experiments of this section, a dataset of 60 lesion images was chosen, including four types of lesions: hard exudates (EX), soft exudates (SE), microaneurysms (MA) and hemorrhages (HE). To identify the various lesions, four similar contrast networks Resnet18-P, l-seg1631, HEDNet, and SD-P were used. In order to overcome the issue of segmentation of various lesions, the issue was solved as a collection of binary segmentation problems. Such an approach has enabled distinguishing the representations of different lesions and reducing the coupling between the models. All the networks had the same hyperparameters with the exception of the networks themselves. The outcomes of the quantitative comparison of the retinal lesion detection algorithm using the full-volume machine network suggested in this paper and the four comparison networks (based on the number of parameters, computational power (FLOPs) and the segmentation index AUC per lesion) are given in Table 3.

Table 3: The quantitative comparison results of five models

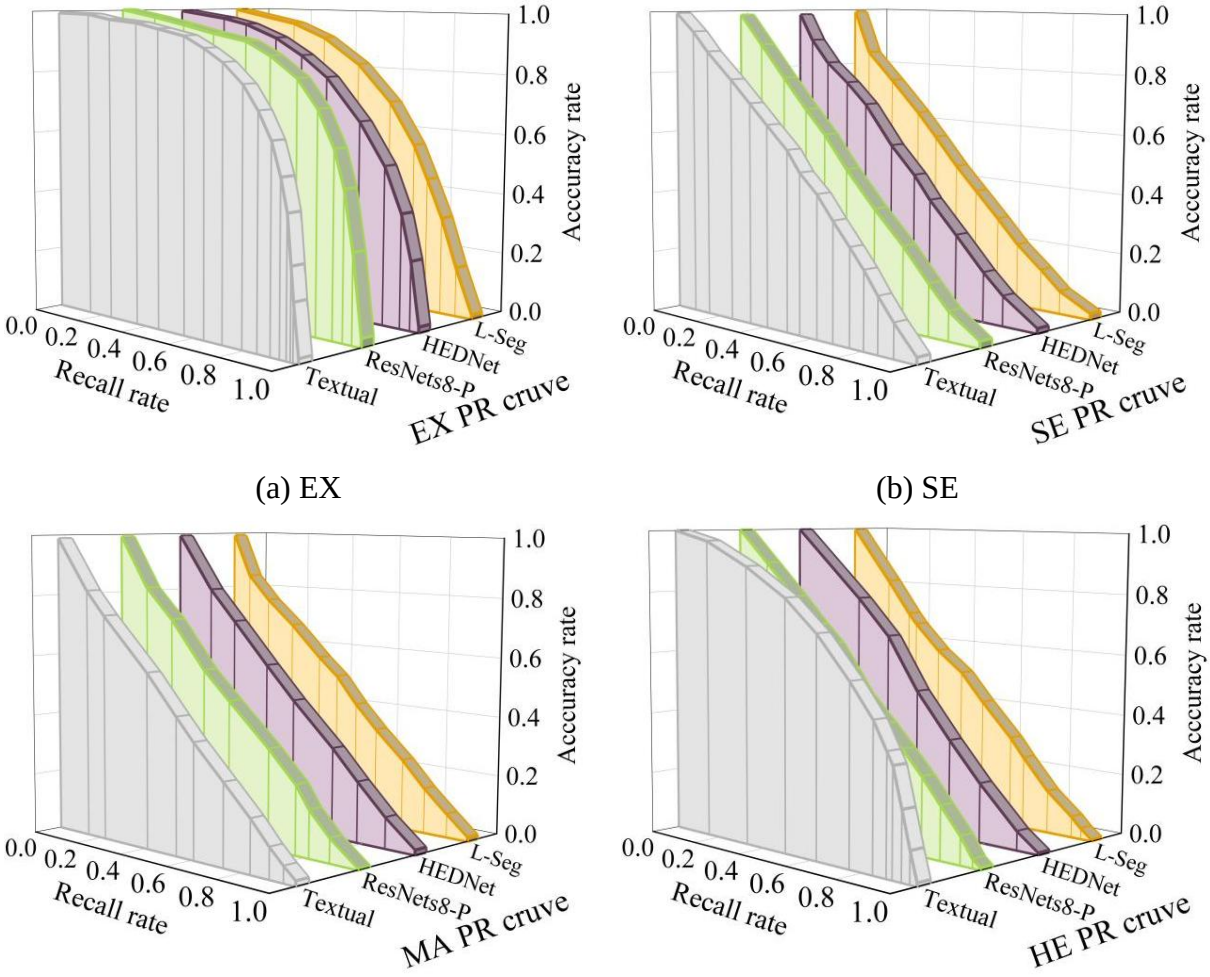
Network	ResNet18-P	L-Seg1631	HEDNet	SD-P	Textual
Parameter quantity(M)	14.5	14.7	14.5	14.7	14.3
FLOPs(G)	81.9	27.8	76.4	81.8	20.5
EX	0.9133	0.9316	0.8685	0.8392	0.9555
SE	0.4148	0.4087	0.5728	0.5105	0.5963
MA	0.4586	0.4533	0.4798	0.4476	0.4947
HE	0.4166	0.5929	0.5566	0.4796	0.6925
mAUC	0.5314	0.5261	0.5936	0.5289	0.6773

In terms of the number of parameters and computational power, the suggested algorithm performs better than the other algorithms, with the values of 14.3M parameters and 20.5G FLOPs. In case of the four types of lesion segmentation, all algorithms had the highest performance of detecting hard exudates (EX) of more than 0.8000 AUC. The proposed algorithm and the ResNet18-P network both scored over 0.9000, and the most optimal was the proposed one (0.9555). In case of the remaining three lesion types, i.e., soft exudates (SE), microaneurysms (MA), and hemorrhages (HE), the segmentation results of these lesions fell within the range of 0.4000 to 0.7000. The proposed algorithm once again proved to have the highest level of performance with AUC scores of 0.4947, 0.6825, and 0.6773. Considering the average accuracy (mAUC) of all the four categories, the performance of the four algorithms ranked as follows; full convolutional network (0.6773) > HEDNet (0.5936) > ResNet18-P (0.5314) > SD-P (0.5289) > L-Seg1631 (0.5261). These findings show that the full convolutional network-based retinal detection algorithm provided the best overall performance in the multi-lesion segmentation task.

In the experiments of this section, a dataset of 60 lesion images was chosen, including four types of lesions: hard exudates (EX), soft exudates (SE), microaneurysms (MA) and hemorrhages (HE). To identify the various lesions, four similar contrast networks Resnet18-P,

l-seg1631, HEDNet, and SD-P were used. In order to overcome the issue of segmentation of various lesions, the issue was solved as a collection of binary segmentation problems. Such an approach has enabled distinguishing the representations of different lesions and reducing the coupling between the models. All the networks had the same hyperparameters with the exception of the networks themselves. The outcomes of the quantitative comparison of the retinal lesion detection algorithm using the full-volume machine network suggested in this paper and the four comparison networks (based on the number of parameters, computational power (FLOPs) and the segmentation index AUC per lesion) are given in Table 3.

The PR curve performance of the four algorithms in the segmentation process of the four metrics is shown in Fig. 5(a)-(d), whose curve data are the segmentation data of 60 image instances after thresholding. It can be seen that the PR curve performance of the four algorithms is consistent with the segmentation performance trend of their corresponding indexes, and the hard exudation (EX) index is also the best. And among the four indexes, the curve of this paper's algorithm surrounds the largest area, which again verifies the lesion segmentation and detection effect of the lesion detection algorithm based on the full convolutional network.



(a) EX (b) SE

(c) MA (d) HE

Figure 5: P-R curve of different lesions with different methods

5 Conclusion

The proposed intelligent diagnostic system of retinal diseases has been developed on the basis of deep learning, and works with preprocessed fundus images. The main components of the system are retinal layer segmentation and lesion detection. To segment the retinal layers, a boundary-sensitive multitasking image segmentation network is used, which has a segmentation performance comparable to manual labeling. When the difference between manual labelling and segmented values of retinal thickness in normal eyes are compared, the deviation between them is less than 1.50 with an average of manual labelling error of less than 0.250. In terms of lesion boundary segmentation, the absolute error is (9.807.26) which is the best result of other algorithms like this. The algorithm to detect retinal lesions uses only 14.3 million parameters and only 20.5G of computational power based on a fully convolutional network. It attains a segmentation rate of at most 0.9555 and an average of 0.6773 on the four large types of lesions, which satisfies the conditions to be considered as intelligent diagnosis of retinal diseases.

References

[1] Ptito, M., Bleau, M., & Bouskila, J. (2021). The retina: a window into the brain. Cells, 10(12), 3269.

[2] Duh, E. J., Sun, J. K., & Stitt, A. W. (2017). Diabetic retinopathy: current understanding, mechanisms, and treatment strategies. JCI insight, 2(14), e93751.

[3] Birtel, J., Yusuf, I. H., Priglinger, C., Rudolph, G., & Issa, P. C. (2021). Diagnosis of inherited retinal diseases. Klinische Monatsblätter für Augenheilkunde, 238(03), 249-259.

[4] Rusciano, D., & Marsili, S. (2025). Editorial to the Special Issue “Retinopathies: A Challenge for Early Diagnosis, Innovative Treatments, and Reliable Follow-Up”. Medicina, 61(4), 662.

[5] Matuszewski, W., Bandurska-Stankiewicz, E., Modzelewski, R., Kamińska, U., & Stefanowicz-Rutkowska, M. (2017). Diagnosis and treatment of diabetic retinopathy—historical overview. Clinical Diabetology, 6(5), 182-188.

[6] Jiménez-Báez, M. V., Márquez-González, H., Bárcenas-Contreras, R., Morales-Montoya, C., & Espinosa-García, L. F. (2015). Early diagnosis of diabetic retinopathy in primary care. Colombia Médica, 46(1), 14-18.

[7] Shankar, K., Perumal, E., & Elhoseny, M. (2021). An IoT-Cloud Based Intelligent Computer-Aided Diagnosis of Diabetic Retinopathy Stage Classification Using Deep Learning Approach. Computers, Materials & Continua, 66(3).

[8] Sedik, A., El-Shafai, W., El-Hag, N. A., El-Banby, G. M., & Abd El-Samie, F. E. (2025). Retinal disorder diagnosis based on hybrid deep learning models. Multimedia Tools and Applications, 1-25.

[9] Bartlett, P. L., Montanari, A., & Rakhlin, A. (2021). Deep learning: a statistical viewpoint. Acta numerica, 30, 87-201.

[10] Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. World Information Technology and Engineering Journal, 10(07), 3897-3904.

[11] Zhang, K., Liu, X., Liu, F., He, L., Zhang, L., Yang, Y., ... & Lin, H. (2018). An interpretable and expandable deep learning diagnostic system for multiple ocular diseases: qualitative study. Journal of medical Internet research, 20(11), e11144.

[12] Ranjan, S., & Choubey, D. K. (2024). Machine Learning, Deep Learning and IoT in Smart Healthcare for Retinal Diseases Management. In Machine Learning and IoT Applications for Health Informatics (pp. 216-238). CRC Press.

[13] Li, F., Chen, H., Liu, Z., Zhang, X. D., Jiang, M. S., Wu, Z. Z., & Zhou, K. Q. (2019). Deep learning-based automated detection of retinal diseases using optical coherence tomography images. Biomedical optics express, 10(12), 6204-6226.

[14] Vaiyapuri, T., Srinivasan, S., Sikkandar, M. Y., Balaji, T. S., Kadry, S., & Meqdad, M. N. (2022). Intelligent Deep Learning Based Multi-Retinal Disease Diagnosis and Classification Framework. Computers, Materials & Continua, 73(3).

[15] Xu, W., Yan, Z., Chen, N., Luo, Y., Ji, Y., Wang, M., & Zhang, Z. (2022). Development and application of an intelligent diagnosis system for retinal vein occlusion based on deep learning. Disease Markers, 2022(1), 4988256.

[16] Hussein, S. A., Farouk, A. A., & Saeid, M. M. (2025). Intelligent retinal disease detection using deep learning. Scientific Reports, 15(1), 43282.

[17] Jaiswal, A. K., Tiwari, P., Kumar, S., Al-Rakhami, M. S., Alrashoud, M., & Ghoneim, A. (2021). Deep learning-based smart IoT health system for blindness detection using retina images. IEEE Access, 9, 70606-70615.

[18] Nazir, T., Irtaza, A., Javed, A., Malik, H., Hussain, D., & Naqvi, R. A. (2020). Retinal image analysis for diabetes-based eye disease detection using deep learning. Applied Sciences, 10(18), 6185.

[19] Li, F., Chen, H., Liu, Z., Zhang, X., & Wu, Z. (2019). Fully automated detection of retinal disorders by image-based deep learning. Graefe's Archive for Clinical and Experimental Ophthalmology, 257(3), 495-505.

[20] Kim, K. M., Heo, T. Y., Kim, A., Kim, J., Han, K. J., Yun, J., & Min, J. K. (2021). Development of a fundus image-based deep learning diagnostic tool for various retinal diseases. Journal of Personalized Medicine, 11(5), 321.

[21] ARSLAN, G., & Erdaş, Ç. B. (2023). Detection of cataract, diabetic retinopathy and glaucoma eye diseases with deep learning approach. Intelligent Methods In Engineering Sciences, 2(2), 42-47.

[22] Mansour, R. F. (2018). Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. Biomedical engineering letters, 8(1), 41-57.

[23] Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., ... & Jia, W. (2021). A deep learning

system for detecting diabetic retinopathy across the disease spectrum. Nature communications, 12(1), 3242.

[24] Son, J., Shin, J. Y., Kong, S. T., Park, J., Kwon, G., Kim, H. D., ... & Park, S. J. (2023). An interpretable and interactive deep learning algorithm for a clinically applicable retinal fundus diagnosis system by modelling finding-disease relationship. Scientific Reports, 13(1), 5934.

[25] De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., ... & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine, 24(9), 1342-1350.

[26] Guo, B., Wang, D., Zhang, R., Hou, J., Liu, W., Wu, Y., ... & Zhang, L. (2025). WaveAttention-ResNet: a deep learning-based intelligent diagnostic model for the auxiliary diagnosis of multiple retinal diseases. Frontiers in Radiology, 5, 1608052.

[27] Wang, X. N., Dai, L., Li, S. T., Kong, H. Y., Sheng, B., & Wu, Q. (2020). Automatic grading system for diabetic retinopathy diagnosis using deep learning artificial intelligence software. Current Eye Research, 45(12), 1550-1555.