



# Automated glaucoma screening method based on image segmentation and feature extraction

Fan Guo<sup>1</sup> · Weiqing Li<sup>1</sup> · Jin Tang<sup>1</sup> · Beiji Zou<sup>2</sup> · Zhun Fan<sup>3</sup>

Received: 17 February 2020 / Accepted: 25 July 2020  
© International Federation for Medical and Biological Engineering 2020

## Abstract

Glaucoma is a chronic disease that threatens eye health and can cause permanent blindness. Since there is no cure for glaucoma, early screening and detection are crucial for the prevention of glaucoma. Therefore, a novel method for automatic glaucoma screening that combines clinical measurement features with image-based features is proposed in this paper. To accurately extract clinical measurement features, an improved UNet++ neural network is proposed to segment the optic disc and optic cup based on region of interest (ROI) simultaneously. Some important clinical measurement features, such as optic cup to disc ratio, are extracted from the segmentation results. Then, the increasing field of view (IFOV) feature model is proposed to fully extract texture features, statistical features, and other hidden image-based features. Next, we select the best feature combination from all the features and use the adaptive synthetic sampling approach to alleviate the uneven distribution of training data. Finally, a gradient boosting decision tree (GBDT) classifier for glaucoma screening is trained. Experimental results based on the ORIGA dataset show that the proposed algorithm achieves excellent glaucoma screening performance with sensitivity of 0.894, accuracy of 0.843, and AUC of 0.901, which is superior to other existing methods.

**Keywords** Glaucoma screening · Neural network · Image segmentation · Feature extraction

## 1 Introduction

Glaucoma is a common chronic disease that threatens eye health, and it is the second leading cause of blindness worldwide [1]. According to the World Health Organization (WHO), approximately 65 million people around the globe are suffering from glaucoma [2]. Since the vision loss caused by glaucoma is irreversible and the symptoms are imperceptible in the early stages, glaucoma is considered to be “silent theft of sight” [3]. Although existing medical technology cannot cure glaucoma, early screening and corresponding treatment can help patients avoid vision loss and reduce the probability of blindness effectively.

One common clinical glaucoma detection technique is intraocular pressure (IOP) measurement. Increasing IOP is one of the symptoms of glaucoma; it can lead to optic nerve damage, visual field defects, and even blindness [4]. Therefore, IOP is considered as one of the important indicators of glaucoma. However, this method is inadequate because the IOP of some patients with glaucoma is normal [5]; thus, IOP measurement cannot detect those special cases. Another common method for screening glaucoma is optic nerve head (ONH) examination which relies on clinical ophthalmologists to screen glaucoma based on retinal images [6]. When screening glaucoma, ophthalmologists often enhance the retinal image manually and diagnose eye disease according to their own experience with domain skills. Since the diagnostic process is inefficient and time-consuming, the above two methods are not suitable for population screening. Thus, the design of an automatic glaucoma screening system is very helpful and necessary for early and mass detection.

On the other hand, with the development of computer and artificial intelligence, automatic glaucoma screening based on the processing of digital retinal images can achieve reliable accuracy and efficiency, and it is applicable for large-scale screening. Glaucoma is usually accompanied by pathological

✉ Jin Tang  
tjin@csu.edu.cn

<sup>1</sup> School of Automation, Central South University, Changsha 410083, China

<sup>2</sup> School of Computer Science and Engineering, Central South University, Changsha 410083, China

<sup>3</sup> Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Shantou University, Shantou 515063, China

phenomena, such as nerve retinal edge erosion [7, 8] and increased optic cup, and these pathological phenomena mainly occur in ONH. Thus, the analysis of ONH can obtain the basis for glaucoma screening as well.

Automatic glaucoma diagnosis methods based on fundus images consist of two main categories: clinical measurement analysis and image-based feature analysis. Clinical measurement analysis refers to the measurement of some glaucoma-associated geometrical features, such as the optic cup to disc ratio (CDR) [9], the diameter of the optic disc [10], and area of the optic cup. These characteristics are highly correlated with glaucoma screening, and CDR is the most important one which has been recognized by clinical ophthalmologists. In the color retinal image, the optic disc (OD) can be easily identified by the naked eye because it is usually displayed as a bright yellow oval area. However, the optic cup (OC) is relatively difficult to distinguish; it exists in the center of the OD and appears as the brightest oval or round region. Except for the OC area, the remaining marginal region in the OD is the neuroretinal rim. According to clinical experience and domain knowledge, a larger CDR implies a higher probability of suffering from glaucoma and vice versa, as shown in Fig. 1. Thus, many automatic glaucoma detection methods based on clinical features (e.g., CDR) have been proposed.

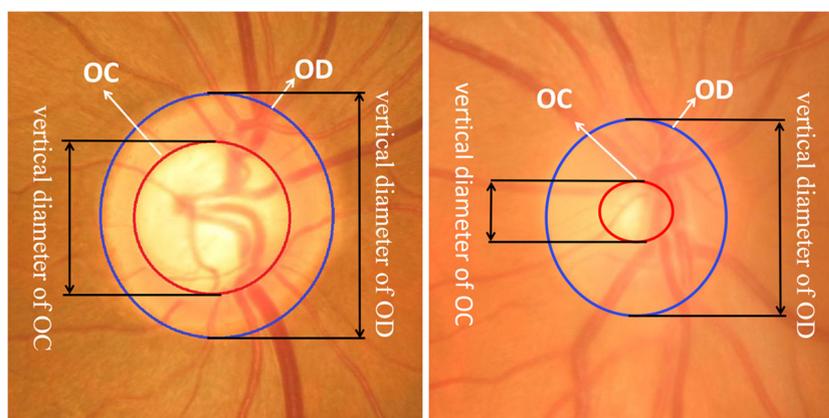
Segmenting the OD and OC from fundus image is paramount for these methods, and then glaucoma is identified with clinical measurements. For instance, Yin et al. [11] performed segmentation using circular Hough transformation and then measured CDR. Morphological operations [12] were employed by Nayak et al. [13] for the segmentation of OC and OD, and the CDR is measured to assess glaucoma risk. Cheng et al. [14] proposed a superpixel classifier for classifying each superpixel to distinguish the OD and OC from the fundus image and the CDR was obtained to screening glaucoma.

However, the measurement of clinical features is heavily dependent on the segmentation results, and the quality of fundus image can greatly affect the measurement results, so the

accuracy of glaucoma screening using only clinical measurement features has reached a bottleneck. Therefore, many image-based methods have been proposed for glaucoma detection. Noronha et al. [15] trained classifiers for automatic glaucoma detection by extracting HOS cumulants from the fundus images. Besides, linear discriminant analysis (LDA) was adopted for feature reduction to make the classifier achieve better performance. In [16], the Gabor transform was combined with a support vector machine (SVM) classifier for glaucoma screening. Dua et al. [17] employed two-dimensional discrete wavelet transform to extract hidden features from fundus images, and discussed the effectiveness of different feature selection methods for machine learning classifiers. Haleem et al. [18] designed a regional image features model (RIFM), which divides the OD area image into different parts and extracts multiple higher-order features of each part, and the contributions of each region to the glaucoma classification are compared. Besides, some important geometric features are also used to improve classification accuracy.

In recent studies, deep learning technique has been proved to be effective in the image segmentation [19] and classification, especially in biomedical image analysis. For OD or OC segmentation, an improved version of U-Net was proposed in [20]. Fu et al. [21] proposed a convolutional neural network (CNN), named M-Net, which realizes the simultaneous segmentation of OD and OC. For glaucoma screening, in [22], a six-layer CNN architecture was designed to obtain glaucoma diagnosis result from the input fundus image directly. Diaz-Pinto et al. [23] employed different CNN models, such as VGG series network, InceptionV3, and ResNet50, to perform glaucoma classification comparison experiments. Experiments adopted the transfer learning method based on ImageNet-trained weights, and the results indicated that the ImageNet-trained models can effectively improve the performance of glaucoma recognition. Although the CNN model can automatically extract the image features, these

**Fig. 1** Comparison of OD in glaucoma and normal fundus image. The OC is represented with the red contour and OD is shown with the blue contour in both images. The vertical diameter of OC in glaucoma image is significantly larger than that in normal images



(a) Glaucoma

(b) Normal

features only focus on the visual aspects of the image rather than the clinical information and lack clinical interpretability which is indispensable for glaucoma screening.

In this paper, we proposed a novel automatic glaucoma screening method combining both clinical measurements and image-based features. Different from the existing methods, increasing field of view (IFOV) feature model was proposed to extract hidden image-based features, such as texture features and Gabor transform features. Besides, some important clinical measurement features are also extracted after the segmentation of the OD and OC. The accuracy of glaucoma detection has been significantly improved. Our contributions include:

- 1) A neural network named CP-FD-UNet++ is proposed. The network is an improved UNet++ [24] and contains a nested U-Net architecture, mixed loss function, and the connection path at the encoder side. The different scales of information input and feature map sharing can be realized through the connection path, and it is beneficial for the performance of OD and OC simultaneous segmentation. The mixed loss function is the combination of focal loss [25] and dice coefficient, which can eliminate the impact of data imbalance.
- 2) We proposed the increasing field of view (IFOV) feature model to extract the image-based features. The visual field range of four different scales from the cup region to the global fundus image was defined; hidden visual features of gray-level co-occurrence matrix (GLCM) and Gabor transform are extracted in these regions. Both clinical measurement features and invisible features obtained from image transformation are taken for glaucoma detection.
- 3) We employed an adaptive synthetic sampling (ADASYN) algorithm [26] to lessen data imbalance problem. Finally, we found that gradient boosting decision tree (GBDT) classifier with ADASYN algorithm can achieve the best performance for glaucoma detection.

## 2 Methods

The framework of the proposed glaucoma screening method is shown in Fig. 2. We locate OD and crop the ROI from a fundus image first, and then OD and OC are segmented from the ROI simultaneously. The next step is feature extraction, and these extracted features will be used for supervised classification after selection and ranking.

### 2.1 Optic disc localization and ROI cropping

OD and OC only occupy a tiny part of the retinal image which will cause a heavy bias between the target and the

background. If OD and OC segmentation are based on the entire fundus image, the result will definitely be seriously affected. Generally, the resolution of the original retinal fundus image is very high, and inputting the original image directly into the network will cause the calculation to increase exponentially and seriously influence the segmentation efficiency. If the fundus image is input into the network after being scaled, the loss of pixel information caused by the scaling will also affect the accuracy of the segmentation result. Therefore, we segment OC and OD based on ROI to avoid the above problems. We positioned the OD and then cropped the ROI from the retinal image based on the localization results. A new OD localization method is proposed based on our past work [5], which includes:

- Step 1: Brightest region extraction
- Step 2: Vessels segmentation
- Step 3: Confidence calculation of the sliding window

In step 1, the morphological transformation (top-bottom-hat) is performed to improve the gray-level fundus image contrast, as shown in Fig. 3b. As the brightest area of the retinal image, OD occupies approximately 6.5% of the full image. Thus, a threshold is set based on 6.5% of the maximal pixel value of the improved gray-scale retinal image, and then the brightest region is extracted by threshold segmentation, as shown in Fig. 3c.

In step 2, the bottom-top-hat transformation is performed to segment the blood vessels directly in the green channel of fundus image. Before segmentation, the green channel needs to be processed by contrast limited adaptive histogram equalization (CLAHE) to obtain better segmentation results, as shown in Fig. 3d.

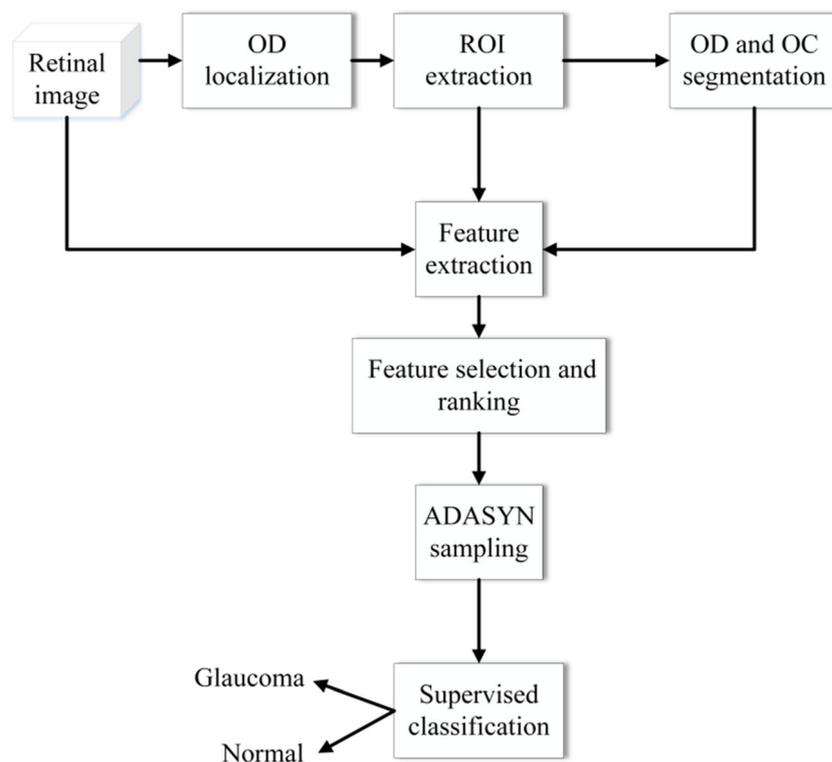
In step 3, the improved gray-level retinal image and the vessel image are combined to obtain a fusion image, as shown in Fig. 3e. The fusion image and the brightest region image were then scanned by a sliding window to calculate the average intensity score; the sliding window with the highest score is regarded as the OD location, as shown in Fig. 3f.

According to the OD localization result, a square area with width 2 times of the OD diameter is extracted as ROI finally. Figure 4 shows an illustrative example.

### 2.2 Automatic segmentation of OD and OC

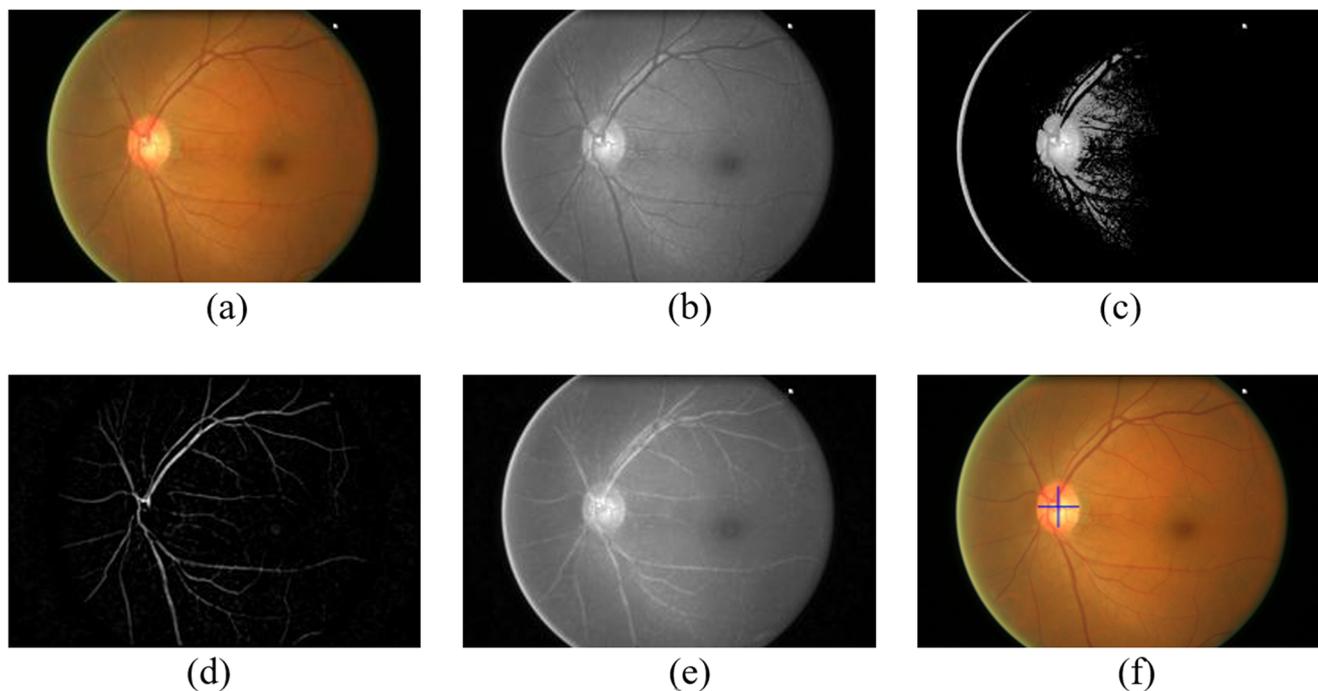
In [24], UNet++, as a powerful architecture, is adopted for medical image segmentation. Inspired by U-Net [27] and UNet++, we propose an improved UNet++ network structure, called CP-FD-UNet++, for the simultaneous segmentation of OD and OC based on ROI. It includes a nested U-Net architecture, mixed loss function, and the connective path at the encoder side. Figure 5 shows the proposed network architecture. Specifically, the network contains the

**Fig. 2** Framework of proposed glaucoma classification method



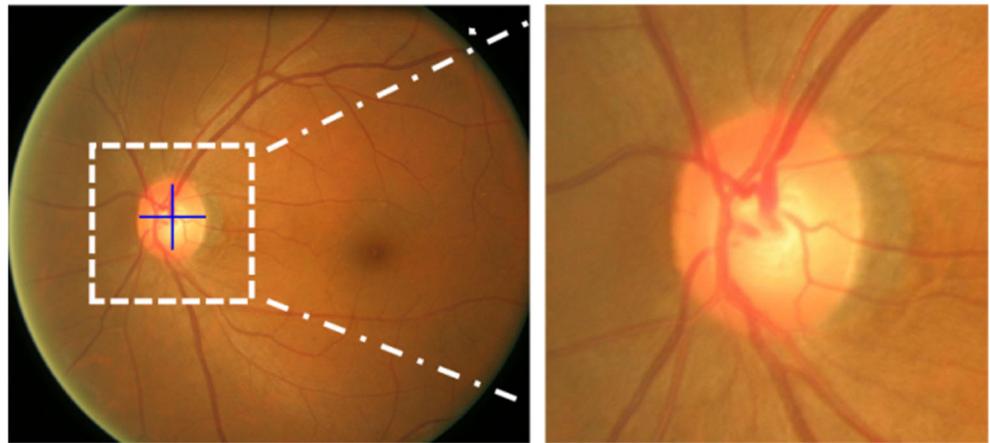
following: (i) a nested U-Net CNN network. Different from U-Net, the encoder and decoder of our proposed network are connected through nested dense convolutional blocks rather than skip connections. Thus, the semantic difference between the feature maps of the encoder and decoder can

be diminished [24]. (ii) The connection path at the encoder side. It can enable different layers of the encoder to obtain multilevel information input and feature maps sharing. (iii) Mixed loss function combines the focal loss with the dice coefficient.



**Fig. 3** Steps of OD localization. **a** Original fundus image. **b** Enhanced gray-scale retinal image. **c** Brightest area image. **d** Vessel image. **e** Fusion image. **f** OD localization result

Fig. 4 OD localization result and ROI



(i). A nested U-Net convolutional neural network U-shaped fully convolutional neural network possesses reliable capabilities and powerful effect in terms of semantic segmentation, especially in medical image segmentation. The main structure of the U-Net consists of a set of down-sampled encoder paths and up-sampled decoder paths, these paths employ convolution blocks to generate feature maps and automatically extract semantic features. In the original U-Net, the encoder and decoder paths are bridged by skip connections; the output feature maps of the encoder are copied and concatenated to the feature maps of the corresponding layer decoder. However, skip connections directly concatenate the shallow encoder feature maps to the deep decoder feature maps, which will cause a certain semantic gap. To solve this problem, skip connections can be changed to nested dense convolutions. According to [24], we insert 3, 2, and 1 convolution modules between the

first, the second, and the third layer encoder-decoder paths, respectively. Each convolution module consists of two consecutive  $3 \times 3$  convolution kernels. Similar to DenseNet [28], the input feature maps of each convolution module are composed of the output of all previous modules on the same layer encoder-decoder path and the upsampling output of the corresponding module on the next layer encoder-decoder path, as shown in Fig. 5. Thus, the features at all gradations can be integrated in a superimposed manner, which increases the reuse rate of features. Besides, the nested dense convolution can avoid the semantic gaps caused by skip connection between the encoder and the decoder, which enables the net to fully understand the abstract semantic features and construct the details of the semantic segmentation target more accurately. Therefore, the accuracy of segmentation results can be greatly improved.

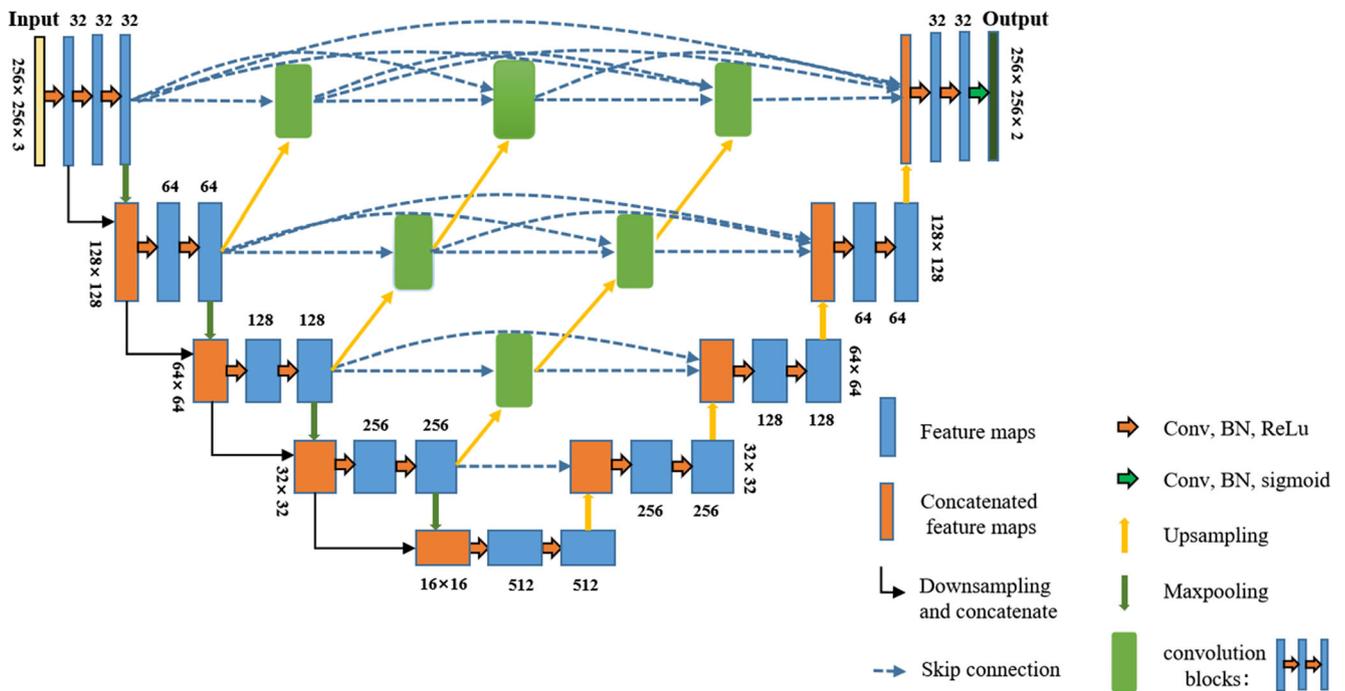


Fig. 5 The network architecture of CP-FD-UNet++

(ii). **Connection path** In view of the powerful feature extraction and reuse capabilities of nested U-Net, we propose a new feature connection method applied at the input of the encoder to make the best use of UNet++. This method concatenates the input feature maps of the previous encoder with its corresponding output feature maps, and the concatenated results are employed as the input of the next encoder. Therefore, it enables the encoder to obtain multi-scale information input and realizes feature map reuse.

(iii). **Mixed loss function** The OD and OC only occupy a little portion of the fundus image or ROI, which leads to a serious imbalance between the segmentation target and the background. Besides, the OC is more difficult to distinguish compared with OD; it may make the network biased towards OD and cause the distortion of OC. To eliminate the impact of data imbalance, we propose to combine focal loss [25] with the dice coefficient as the loss function. Focal loss can effectively improve the accuracy of hard-to-distinguish samples, and the combination with dice coefficient can reduce the impact of unevenly distributed data. The mixed (FD) loss function is represented as:

$$L_{FD} = \lambda_1 (L_{focal}^{disc} + L_{focal}^{cup}) + \lambda_2 (L_{dice}^{disc} + L_{dice}^{cup}) \quad (1)$$

$$L_{focal} = - \sum_{i=1}^N \alpha \cdot (q^i)^\gamma \cdot \log p^i, \quad L_{dice} = - \sum_{i=1}^N \frac{2|p^i \cdot q^i|}{|p^i|^2 + |q^i|^2} \quad (2)$$

where  $L_{focal}^{disc}$  and  $L_{focal}^{cup}$  represent the focal loss of OD and OC respectively and  $L_{dice}^{disc}$  and  $L_{dice}^{cup}$  are the dice coefficient loss.  $p^i$  refers to the probability that pixel  $i$  is predicted as the target label value, and  $q^i$  represents the ground truth label.  $\lambda_1$  and  $\lambda_2$  in Eq. (1) are weigh parameters to balance the contribution of two different loss functions.  $\alpha$  in Eq. (2) is the balance coefficient used to balance the unevenness of the samples.  $\gamma$  in Eq. (2) can reduce the loss of easy-to-classify samples so that the model will pay more attention to the difficult and misclassified samples. Since the values of dice coefficient loss and focal loss are in the same order of magnitude, and the two loss functions are equally important in OC and OD segmentation, both  $\lambda_1$  and  $\lambda_2$  are set to 0.5. According to [25], the best results can be obtained by setting  $\alpha$  and  $\gamma$  to 0.25 and 2, respectively.

Figure 6 shows the simultaneous segmentation result using the proposed model. The CP-FD-UNet++ model can automatically extract semantic features from the input fundus image to segment OC and OD. Experimental results show that the segmentation accuracy of the model is close to the level of clinical ophthalmologists, while the operation time of the model is far less than that consumed by doctors in manual annotation. Thus, the model has both excellent accuracy and real-time performance. Besides, the segmentation network also has a low implementation cost, which makes the model have the

potential for clinical application. Moreover, the proposed model can make full use of the extracted semantic features due to its unique network structure, which enables the network to perform OC and OD segmentation under different imaging conditions and individual differences. Therefore, the CP-FD-UNet++ model has excellent clinical reliability and can be effectively applied to OC and OD segmentation under clinical conditions.

### 2.3 Feature extraction

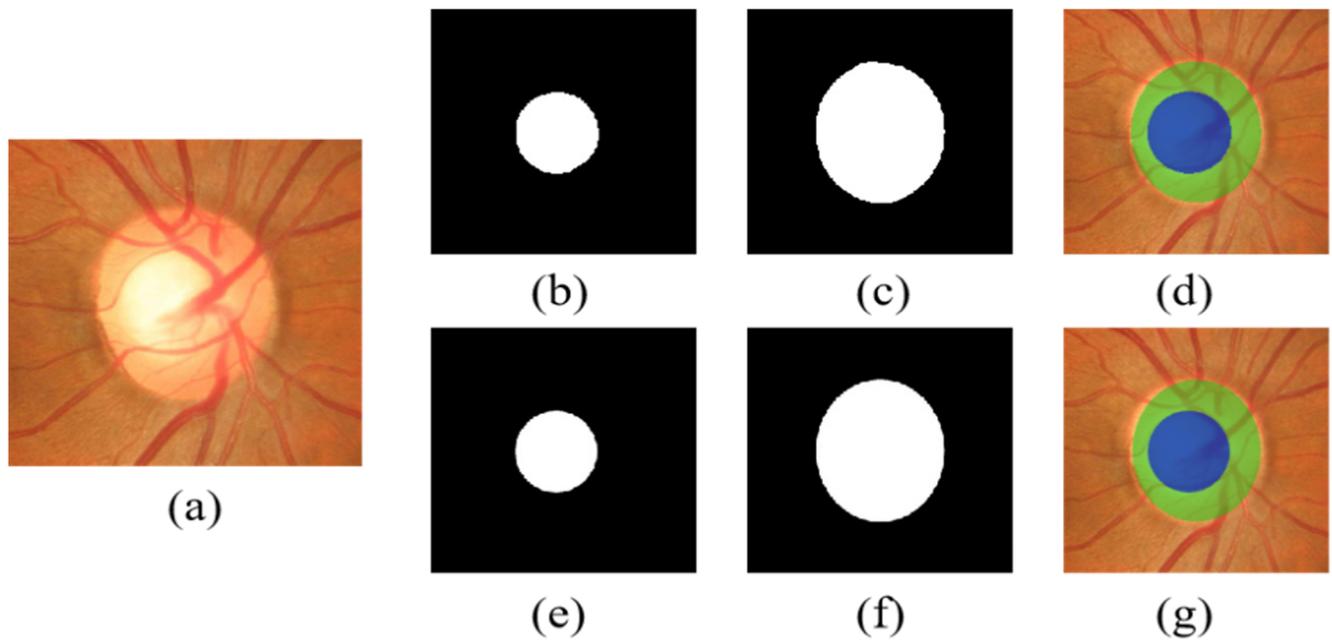
In this paper, we propose a feature extraction method combining both clinical measurement features which inspired by domain knowledge and image-based features for automatic glaucoma screening. Different from the existing methods, increasing field of view (IFOV) feature model was proposed to extract the image-based features. In this method, the hidden features of the different receptive fields are considered; thus, the extracted features have wider coverage and can give more reliable and accurate glaucoma screening results. Figure 7 shows the feature extraction flowchart.

### 2.4 Clinical measurements

Neuroretinal rim (NRR) (see the green area in Fig. 8b) refers to the region between OC and OD; it is determined to consist of four quadrants, i.e., inferior (I), superior (S), nasal (N), and temporal (T). The ISNT rule is a common basis for judging whether the NRR shape is normal in the clinical diagnosis of glaucoma. It specifies that the thickness value of the I quadrant is the largest while the T quadrant is the thinnest. Besides, the thickness of the S quadrant is greater than that of the N quadrant. The CDR and ISNT rules are the indispensable basis for clinical diagnosis of glaucoma. The CDR value of the normal eye is found to be 0.3 to 0.5 [29] and it will increase with the aggravation in neuroretinal degeneration. Therefore, the rising CDR value suggests an increased risk of glaucoma. Besides, healthy eye structure conforms to the INST rule; otherwise, glaucoma may occur.

Based on the above domain knowledge, the following geometric features were selected as our clinical measurement features: the vertical CDR and horizontal CDR, OC area, OD area, NRR area, and area ratio of NRR to OD. Figure 8 shows the clinical measurement features.

In accordance with the ISNT rule, the region of NRR which fall in the ISNT four quadrants are respectively named as inferior NRR (INR) (see Fig. 8c), superior NRR (SNR) (see Fig. 8d), nasal NRR (NNR) (see Fig. 8e), and temporal NRR (TNR) (see Fig. 8f). The area, thickness, and area ratio to OD of the four regions were calculated as the distinguishing characteristics of glaucoma. Besides, the area and thickness ratios between these areas are also important for the inspection of the ISNT rule. Figure 8 shows an illustrative example.



**Fig. 6** OD and OC segmentation result. **a** Original ROI image. **b** OC segment result of our network. **c** OD segment result of our network. **d** Our network’s visualized fusion result. **e** OC segment result of ground truth. **f** OD segment result of ground truth. **g** Ground truth’s visualized fusion results

### 2.5 IFOV feature model

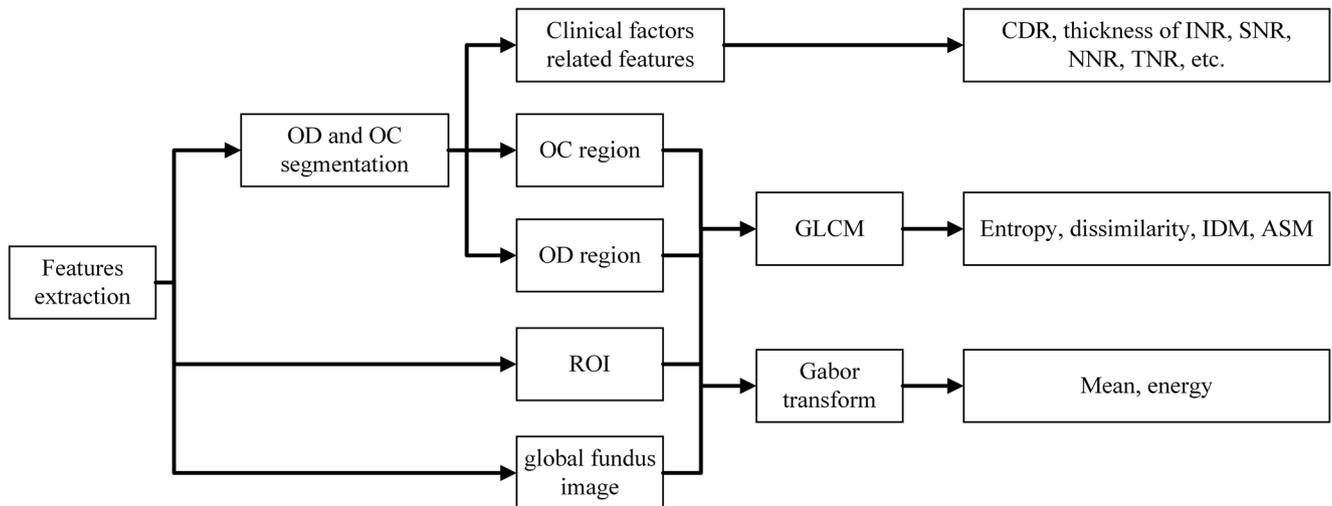
We propose the increasing field of view (IFOV) feature model to extract the image-based features. In the IFOV model, there are four visual fields with different scales, ranging from small to large: OC region, OD region, ROI, and global fundus image; various hidden visual features are extracted from Gabor transform and GLCM that applied to these regions. As can be seen in Fig. 9, the field of vision is increasing, according to the order in which the ophthalmologist observes the lesion. Thus, image features at different levels can be extracted. The extracted features possess wide coverage and representative ability, providing more evidence for glaucoma detection.

#### 2.5.1 GLCM feature

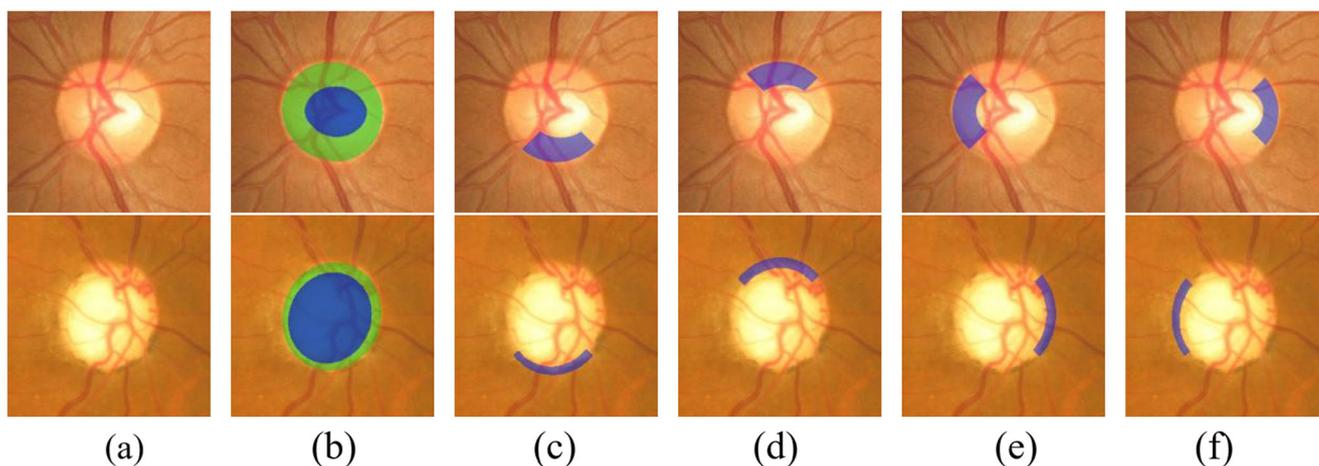
GLCM is a second-order statistical measure method used to analyze image texture features. It describes the appearance frequency of different gray-scale value combinations in an image and measures the relationship between neighborhood pixels in four angles ( $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ ). For the image of size  $m \times n$ , the  $C_d(i, j)$  for GLCM can be written as

$$C_d(i, j) = \begin{cases} (p, q), (p + \Delta x, q + \Delta y) : \\ I(p, q) = i, I(p + \Delta x, q + \Delta y) = j \end{cases} \quad (3)$$

where  $(p, q), (p + \Delta x, q + \Delta x) \in m \times n$ .  $\Delta x$  and  $\Delta y$  are offset



**Fig. 7** Flowchart of the feature extraction



**Fig. 8** Clinical measurements feature illustration. The upper row is an example of non-glaucoma, and the lower row is glaucoma. **a** Original ROI image. **b** Visualized fusion result of OC and OD segmentation. **c** INR. **d** SNR. **e** NNR. **f** TNR

values which are ranged from 1 to 6 to evaluate pixel adjacency at different scales. The probability of pixel adjacency  $P_d(i, j)$  is defined as

$$P_d(i, j) = \frac{C_d(i, j)}{\sum C_d(i, j)} \tag{4}$$

In our work, the GLCM features, i.e., entropy, dissimilarity, inverse different moment (IDM), and angular second moment (ASM) are selected for glaucoma classification. These features are mathematically formulated as

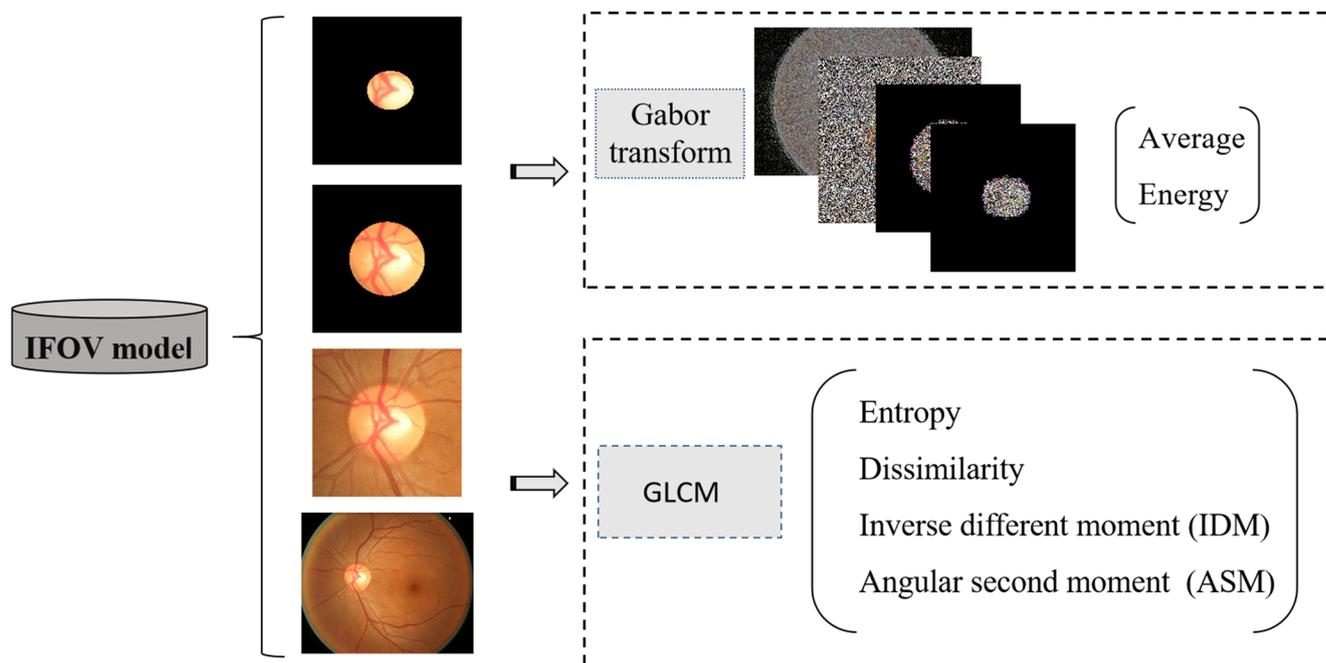
$$\text{entropy} = - \sum_{i=1}^m \sum_{j=1}^n C_d(i, j) \log C_d(i, j) \tag{5}$$

$$\text{dissimilarity} = \sum_{i=1}^m \sum_{j=1}^n C_d(i, j) * |i-j| \tag{6}$$

$$\text{IDM} = \sum_{i=1}^m \sum_{j=1}^n \frac{(C_d(i, j))^2}{1 + (i-j)^2} \tag{7}$$

$$\text{ASM} = \sum_{i=1}^m \sum_{j=1}^n (C_d(i, j))^2 \tag{8}$$

The above features are extracted from the four receptive fields, i.e., OC region, OD region, ROI, and global fundus image. By adjusting the angle and offset value, a total of 384 GLCM features can be obtained per fundus image.



**Fig. 9** The illustration of the IFOV feature model

### 2.5.2 Gabor feature

Gabor transform refers to the convolution operation of Gabor filter kernel and the image to obtain multi-scale and multi-direction amplitude domain maps. Thus, it can extract structural information of a specific frequency and direction [30]. Gabor filters can achieve local optimization in both the spatial and frequency domains, and they can describe local information efficiently. The Gabor filter is represented as

$$\text{gabor}_{f,\sigma,\theta,\gamma}(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) * \exp(i2\pi fx) \quad (9)$$

$$x' = x\cos\theta + y\sin\theta \quad y' = y\cos\theta - x\sin\theta \quad (10)$$

In order to extract comprehensive feature information, we construct a complete set of Gabor filters by changing the Gabor kernel parameters. These parameters include two different frequencies, four different orientations ( $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ ), five different  $\gamma$  values (0.25, 0.5, 1, 2, 4), and four different Gabor scales ( $\sigma = 1, 3, 5, 7$ ). The two frequencies ( $f = 1.1781, 2.3562$ ) are given by [16]. The average value and energy of the Gabor transform coefficients are extracted for glaucoma classification. Thus, a total of 1280 Gabor features can be obtained for a fundus image, and the number of features of different types can be seen in Table 1.

### 2.6 Feature selection and ranking

In our method, a total of 1688 original features were extracted, such redundant features may lead to low computational efficiency and depreciation of classifier performance. Therefore, the dimension reduction and selection of features are crucial. To obtain the best classification performance based on the fewest features, multiple feature selection methods are used to determine the best feature combination in this work.

Step 1. We employ the variance analysis to remove features with extremely low variance since the values of these features are approximately the same or basically constant, and they contribute little to the classification of glaucoma. We set the variance threshold of

discarded features to 0.001, which can filter out most low-variance features and improve computational efficiency without reducing performance.

Step 2. Pearson correlation coefficient is employed to measure the correlation between two features. In our work, we define two features with an absolute value of the Pearson correlation coefficient greater than 0.9 as extremely correlated features. The redundant features will be deleted since these features have little contribution to the classification, affect computing efficiency, and even reduce classification accuracy. The heatmap of the feature's correlation matrix is shown in Fig. 10.

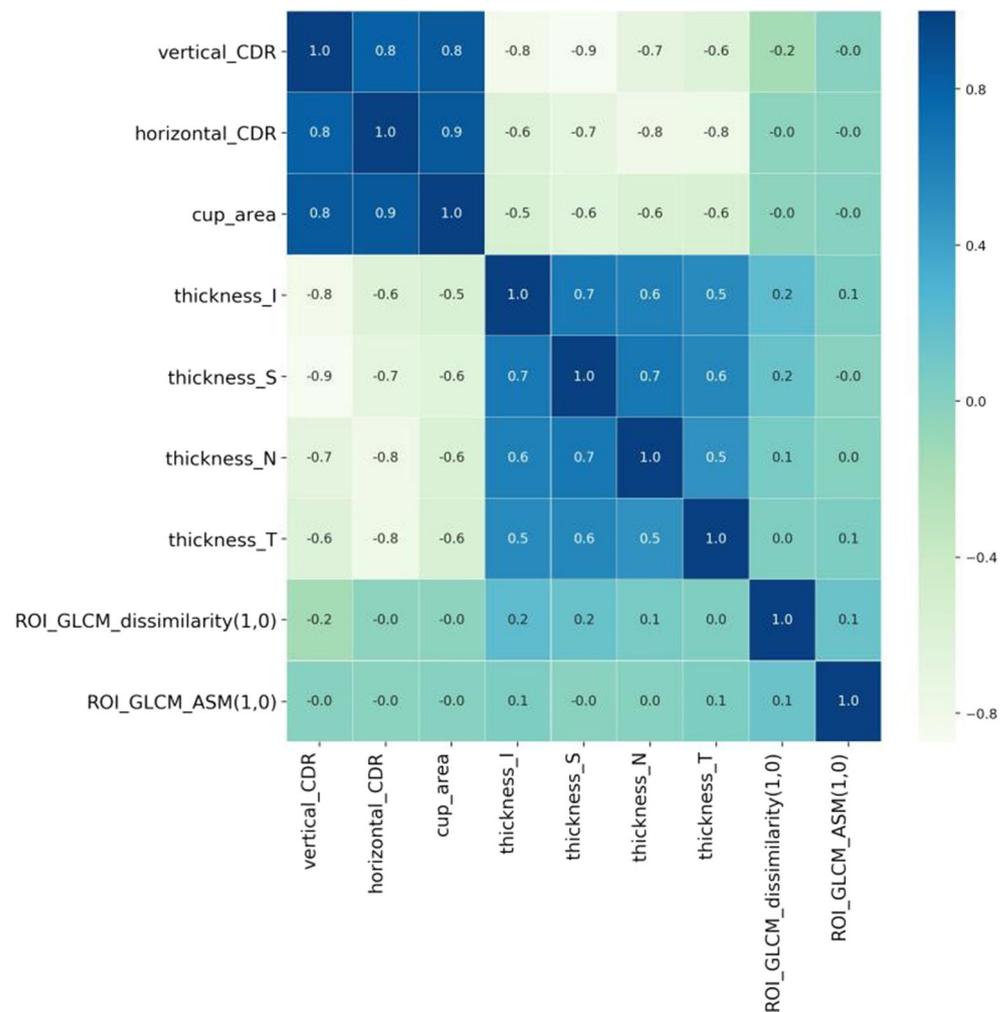
Step 3. After variance and correlation analysis, the feature redundancy still exists and has a great impact on the construction of the glaucoma classifier. To achieve the best classification accuracy with the least number of features, it is necessary to sort the entire feature set and select important feature subsets. Thus, we compare the three different candidate feature ranking methods, i.e., chi-square ( $[\chi^2]$ ) [31], mutual information [32, 33], and GBDT [34, 35]. In the chi-square method, the importance of a feature is estimated by computing the value of its  $\chi^2$  statistic. Mutual information is used to measure the degree of interdependence between two variables and calculate content information shared between them. The importance of features is measured by mutual information. GBDT is an integrated model of decision trees, which can be effectively applied to tasks, such as regression, classification, and feature ranking. When GBDT is used for feature ranking, it evaluates the importance of features based on the reduction of impurities after node splitting and then ranks the features according to the importance. Mean square error (MSE) is adopted for measuring the impurity of the dataset.

The combination of features based on different ranking methods will greatly affect the performance of the classifier. To find the best feature ranking method and feature combination, we use different methods to rank all candidate features and select some of the top-ranked features to train the classifier. By continuously increasing the number of top-ranked features, we can compare the accuracy of the classifiers to determine the best feature ranking method and feature combination.

**Table 1** Number of different features

Type	Clinical features	GLCM features	Gabor features	Total
Number	24	384	1280	1688

**Fig. 10** Heatmap of feature correlation matrix



### 2.7 Adaptive synthetic sampling approach

An imbalanced dataset refers to the extremely uneven sample size of each category in the dataset. The samples of the majority class occupy too much proportion in the total samples, and the minority class samples are ignored. Therefore, the trained classifier has a bias for the majority class, which will cause an obvious degradation in classification performance. In this paper, the dataset used in the experiment has the problem of uneven sample distribution. For example, the ratio of normal and glaucoma fundus images in the ORIGA [36] dataset is 482:168. In order to reduce the imbalance of the sample distribution, we employ the ADASYN [26] algorithm. This algorithm generates new feature samples adaptively for minority class based on the features extracted from minority ethnic samples, minimizing the feature sample differences between normal and glaucoma classes as much as possible. The ADASYN algorithm for glaucoma classification problem is described in algorithm 1.

---

**Algorithm 1** ADASYN algorithm for glaucoma classification

**Input:**  $n$  is the total number of fundus image samples,  $n_s$  is the number of minority class i.e. glaucoma samples,  $n_l$  is the number of majority class i.e. normal samples.

---

**Algorithm:**

- 1) Calculate the imbalance index:
 
$$d = n_s / n_l \tag{11}$$
- 2) Get the number of feature samples to be synthesized:
 
$$Q = (n_l \cdot n_s) \tag{12}$$
- 3) KNN algorithm is employ to determine the nearest  $K$  neighbor features for each sample in the minority class, calculate the parameter  $r_i$ :
 
$$r_i = \Delta_i / K, \quad i = 1, \dots, n_s \tag{13}$$
 where  $\Delta_i$  is the number of feature samples belonging to the majority class among  $K$  nearest neighbors.
- 4) Calculate  $r'_i$  for each minority sample:
 
$$r'_i = r_i / \sum_{j=1}^{n_s} r_j \tag{14}$$
- 5) Calculate the feature sample quantity that need to be produced for each minority sample:
 
$$g_i = r'_i \times Q \tag{15}$$
- 6) For each feature sample in the minority class produce  $g_i$  synthetic feature examples according to the following steps:
  - Do the **Loop** from 1 to  $g_i$  :
  - Select one feature sample  $x_{ij}$  from the generated nearest  $K$  neighbors for each minority sample  $x_i$  to synthesize feature example  $x_{si}$  :

$$x_{si} = x_i + (x_{ij} - x_i) \times 0.5 \tag{16}$$

**End Loop**

---

Experimental results show that the imbalance of feature samples between normal and glaucoma is greatly relieved and the classifier achieves better performance.

## 2.8 Classifier setting

We compare SVM, random forest (RF), and GBDT classifiers to select the best method for the automated glaucoma diagnosis based on the selected features.

SVM is a widely used supervised learning tool. It can realize classification by projecting features into space to find the best classification hyperplane. Kernel functions can be employed to map feature points which cannot be distinguished linearly to a higher dimensional space where these feature points can be linearly separated. According to [5], we use radial basis kernel function for glaucoma detection. The SVM model has two important hyperparameters:  $C$  and gamma.  $C$  is the penalty coefficient, and gamma affects the number of support vectors. From the experiment of adjusting parameters, we find that setting  $C$  to 150 and gamma to 0.001 can obtain the best glaucoma classifier performance.

As an integrated learning method, Random forest (RF) is also a variant of the bagging algorithm. It can randomly establish a large number of decision trees which are not related to each other, the input samples are classified according to the decision-making situation of each decision tree in the forest. RF has the ability to analyze the complex interactive features and can maintain robustness in the presence of noisy data. The number of decision trees is an important parameter in the RF model. If the decision trees are insufficient, the fitting ability of the RF model will be weakened. However, too many decision trees will increase the calculation and lead to overfitting. Therefore, we keep the number of decision trees to be moderate, namely 50.

Unlike random forests, GBDT [34] is an integrated learning algorithm that combines decision trees with boosting method. It adopts the gradient descent method to optimize the loss function in the function space. Besides, the regression idea is possessed by the GBDT algorithm, which makes it capable of processing complex data. The determination of the decision tree number in the GBDT model is similar to that of the RF model, so we choose 50 decision trees as a moderate choice.

## 3 Experiments and results

The proposed glaucoma detection method is evaluated in three aspects: OD and OC segmentation, feature ranking and selection, and glaucoma classification performance assessment. Three datasets are used to evaluate these parts: ORIGA [36], RIGA [37], and DRISHTI-GS1 [38].

The ORIGA [36] dataset consists of 650 color fundus images and their glaucoma diagnosis results, and the segmentation labels of OD and OC are also provided. Thus, the ORIGA dataset is used to assess the performance of OD and OC segmentation and glaucoma detection.

The RIGA [37] provides 750 color fundus images gathered from three resources, including the MESSIDOR dataset [39], Magrabi Eye Center [37], and Bin Rushed Ophthalmic Center [37]. Besides, the RIGA dataset also provides the boundaries of OD and OC labeled by six experienced clinical ophthalmologists. Thus, the RIGA dataset is adopted to verify the performance of OD and OC segmentation.

To further evaluate the accuracy of our proposed glaucoma detection method, the comparison experiment with other state-of-art methods was conducted based on the DRISHTI-GS1 [38] dataset which contains 101 color fundus images.

### 3.1 Evaluation metrics

For OD and OC segmentation performance, we compute overlap coefficient  $S$ , the balanced accuracy  $Acc$  and absolute CDR error as evaluation measurements. The overlap coefficient  $S$  denotes the overlap level of two regions which can be defined as:

$$S(A, B) = \frac{\text{Area}(A \cap B)}{\text{Area}(A \cup B)} \quad (17)$$

where  $S \in [0, 1]$ , and higher  $S$  value implies a better segmentation performance.  $A$  refers to the ground truth mask, and  $B$  represents the predicted mask.  $\text{Area}(\cdot)$  denotes the region area.  $\cap$  refers to the intersection and,  $\cup$  refers to the union operation.

**Table 2** Segmentation evaluation results on ORIGA dataset

Method	Overlap coefficient of OD ( $S_{OD}$ )	Balanced accuracy of OD ( $Acc_{OD}$ )	Overlap coefficient of OC ( $S_{OC}$ )	Balanced accuracy of OC ( $Acc_{OC}$ )	Absolute CDR error ( $\delta_{CDR}$ )
Supersixel [14]	0.898	0.964	0.736	0.918	0.077
DeepCDR [21]	0.929	0.983	0.770	0.930	0.071
U-Net [27]	0.925	0.965	0.756	0.914	0.090
UNet++ [24]	0.926	0.973	0.777	0.928	0.070
CP-UNet++	0.947	0.976	0.805	0.928	0.062
FD-UNet++	0.948	0.971	0.802	0.935	0.062
Proposed	0.960	0.976	0.813	0.943	0.061

**Table 3** Segmentation evaluation results on RIGA dataset

Dataset	Method	Overlap coefficient of OD ( $S_{OD}$ )	Balanced accuracy of OD ( $Acc_{OD}$ )	Overlap coefficient of OC ( $S_{OC}$ )	Balanced accuracy of OC ( $Acc_{OC}$ )	Absolute CDR error ( $\delta_{CDR}$ )
BinRushed [37]	DeepCDR [21]	0.845	0.923	0.727	0.936	0.060
	U_Net [27]	0.940	0.967	0.771	0.929	0.062
	UNet++ [24]	0.944	0.970	0.786	0.933	0.057
	Proposed	0.957	0.972	0.829	0.947	0.046
Magrabia [37]	DeepCDR [21]	0.882	0.942	0.709	0.906	0.063
	U_Net [27]	0.942	0.965	0.784	0.939	0.060
	UNet++ [24]	0.935	0.960	0.770	0.956	0.052
	Proposed	0.937	0.965	0.830	0.959	0.047
MESSIDOR [39]	DeepCDR [21]	0.826	0.913	0.624	0.931	0.074
	U_Net [27]	0.921	0.969	0.782	0.933	0.060
	UNet++ [24]	0.910	0.970	0.804	0.925	0.058
	Proposed	0.938	0.978	0.822	0.957	0.030

The balanced accuracy Acc which consists of sensitivity and specificity is also adopted. They are defined as:

$$Acc = \frac{Sensitivity + Specificity}{2}$$

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

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

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Extracting clinical measurement features, especially CDR, is the purpose of OD and OC segmentation. The quality of the segmentation directly affects the accuracy of glaucoma screening. Therefore, we adopted the absolute CDR error to measure the performance of segmentation, which is defined as

$$\delta_{CDR} = |CDR_A - CDR_B| \tag{19}$$

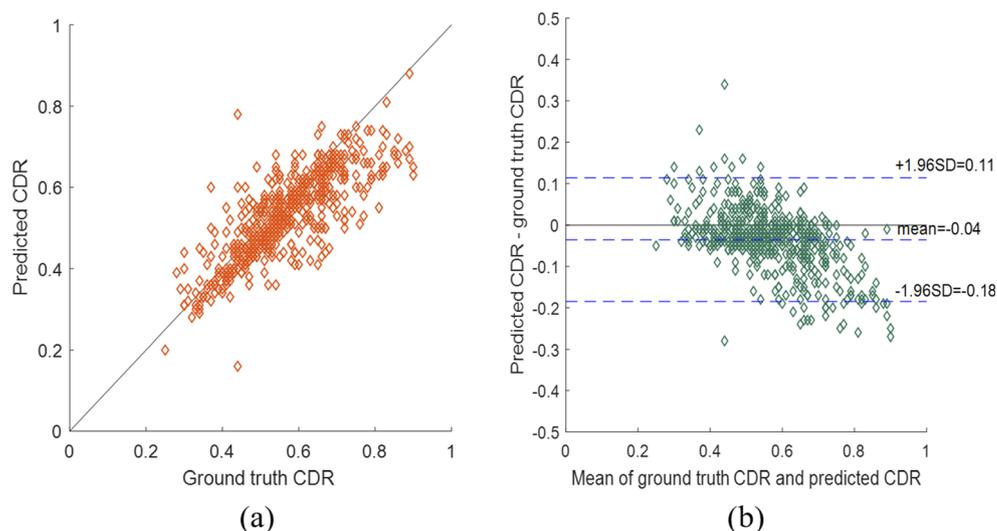
where  $CDR_A$  refers to the ground truth CDR, and  $CDR_B$  is the CDR calculated by our proposed CP-FD-UNet++ model.

### 3.2 OD and OC segmentation evaluation

We have conducted an experimental evaluation for OD and OC segmentation from three aspects:

- 1 Segmentation performance based on quantitative assessment
- 2 Quantitative assessment based on different classification algorithms
- 3 Quantitative assessment based on different datasets

**Fig. 11** Comparison of ground truth CDR and predicted CDR. **a** Scatter plot of the CDR on the two datasets. **b** Bland-Altman diagram



We compare the proposed segmentation method with other state-of-the-art methods: superpixel method [14], U-Net [27], UNet ++ [24], and deepCDR [21]. To confirm the effectiveness of our proposed connection path and FD loss designs for OD and OC segmentation, we combine them with UNet++ separately and conduct relevant comparative experiments. Besides, two different datasets, i.e., ORIGA dataset [36] and RIGA dataset [37] are used to evaluate the robustness of the segmentation method. We randomly select 50% samples from different datasets as the training set and the remaining 50% as the test set. The ADAM method is adopted to optimize the network, and cosine annealing method is used to adjust the learning rate for it can help the model jump out of the local best point and obtain better performance. Besides, horizontal flip, random rotation, and color jitter are used to augment the training image, which can greatly improve the performance of the model and avoid overfitting.

Table 2 and Table 3 show the segmentation results of the ORIGA dataset [36] and the RIGA dataset [37], respectively. All evaluation results are the average of the test set samples. Results on both tables illustrate that our proposed method can achieve the best results for most indexes in different datasets. Besides, the OC segmentation performance of our model is greatly superior to several other methods, which indicates that the CP-FD-UNet++ model can ease the impact of data imbalance effectively, and the model focus on the samples that difficult to distinguish. Table 2 shows that two improved versions of U-Net++, i.e., CP-UNet++ and FD-UNet++ based on connection path and FD loss proposed by our work, can contribute to the improvement of segmentation performance.

The correspondence of ground truth CDR and the predicted CDR in all datasets is shown in the scatter plot (see Fig. 11a). Besides, we use the Bland-Altman method to further measure the consistency of the ground truth CDR and the predicted CDR, as shown in Fig. 11b. The Bland-Altman method uses

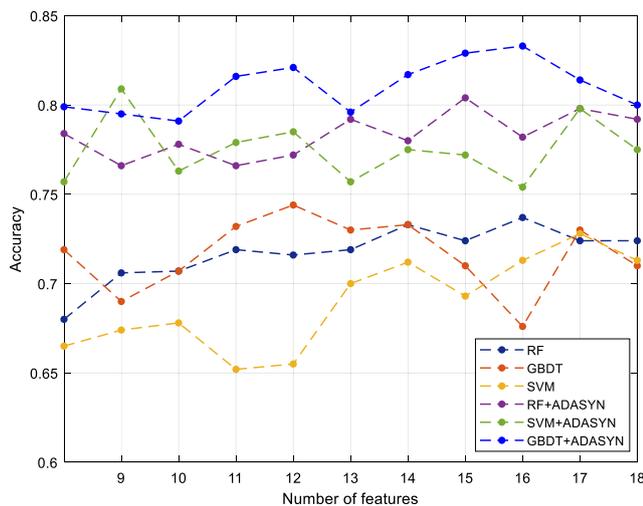


Fig. 12 The performance results obtained using chi-square ranking method

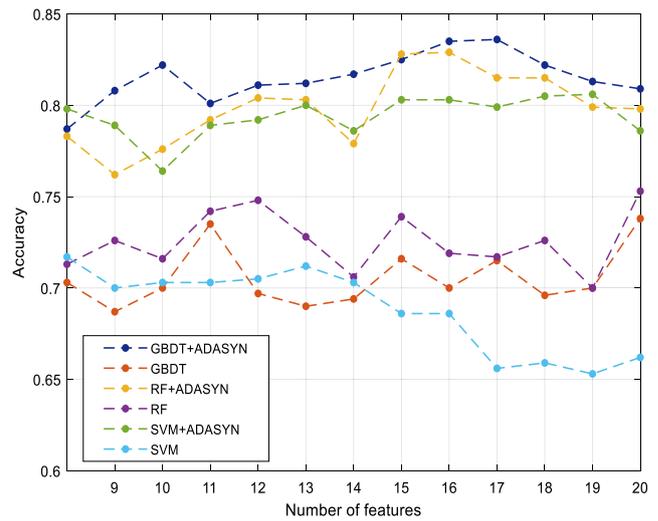


Fig. 13 The performance results obtained using mutual information ranking method

the mean of the data as the horizontal axis and the difference as the vertical axis to make a scatter plot. It calculates the 95% distribution range of the difference as the limits of agreement (LoA) and then compares the distribution of scattered points within LoA. Specifically, the lower limit of the LoA is the mean difference minus 1.96 times standard deviation (SD) of the difference, and the upper limit is the mean difference plus 1.96 times SD of the difference. It can be seen from the results that 4.98% of the difference points exceed the LoA, and 95.02% of the samples are within the consistency limit, indicating that the CDR calculated based on our proposed model is very close to ophthalmologist’s prediction.

### 3.3 Feature ranking and selection evaluation

To obtain the best feature combination for glaucoma classifier, we conduct comparison experiments based on different

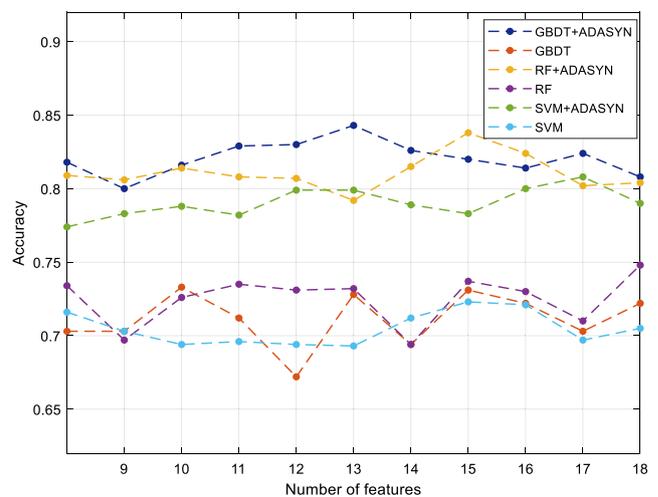


Fig. 14 The performance results obtained using GBDT ranking method

**Table 4** The performance result comparison of different ranking methods

Ranking method	Best classifier	Number of the most important features	Balance accuracy
GBDT	GBDT+ADASYN	13	0.843
Mutual Information	GBDT+ADASYN	17	0.835
Chi2	GBDT+ADASYN	16	0.833

feature ranking methods by adjusting the number of top-ranked features. Balance accuracy of the glaucoma classifiers is considered to be the assessment criteria of the combination of different feature ranking methods and the most important feature quantities. In our case, the classifiers are the SVM, RF, and GBDT and their combination with data balancing algorithm ADASYN. Comparative experiments are performed using the ORIGA dataset [36]. The performance results obtained by chi-square, mutual information, and GBDT ranking methods are shown in Fig. 12, Fig. 13, and Fig. 14.

The average comparison results of different ranking methods and feature numbers are shown in Table 4. The results show that no matter which feature ranking method is chosen, the best classification accuracy can be obtained by using the GBDT+ADASYN classifier. Besides, when the GBDT+ADASYN classifier is combined with the GBDT ranking method, the best classification performance can be achieved with the minimum number of the most important features. Therefore, the 13 most significant features by using the GBDT ranking method were selected, and their importance ranking is shown in Fig. 15.

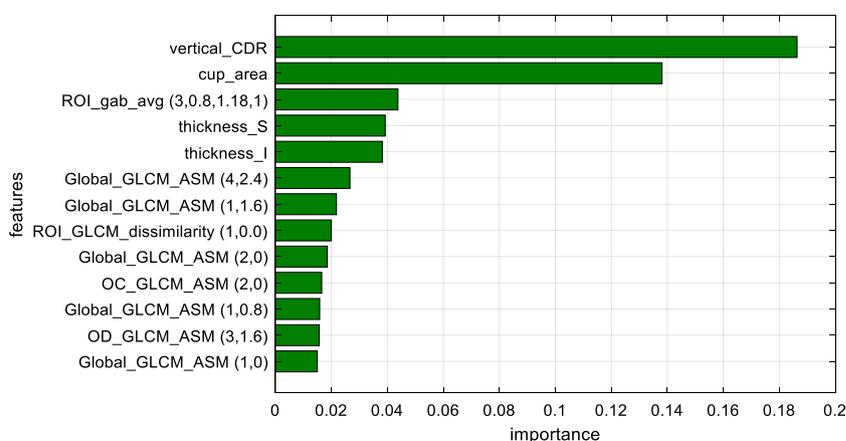
A total of 1688 original features are reduced to 13 without affecting the accuracy of classification, which also greatly improves the efficiency of the classification system. From Fig. 15, we can clearly see that CDR achieves the highest importance score, which indicates that it is indispensable for the glaucoma screening. Some clinical measurement features related to Neuroretinal rim are also important to the classification. GLCM and Gabor features are considered to contribute greatly to the screening of glaucoma because they can

discover hidden information related to glaucoma diagnosis. Besides, the OC region, OD region, ROI, and global images all contribute to the extraction of unstructured features, indicating that the proposed IFOV feature model plays an important role in the glaucoma classification.

### 3.4 Glaucoma classification evaluation

Since the balance accuracy of the classifier has been verified in the feature selection step, to comprehensively evaluate the classification performance of different classifiers, we also adopt three other evaluation criteria: sensitivity, specificity, and area under the ROC curve (AUC). Quantitative assessment is performed on ORIGA [36] and DRISHTI-GS1 [38] datasets. The glaucoma samples are defined as positive while the normal samples are defined as negative. Then, sensitivity and specificity are calculated according to Eq. 18. We conduct experiments on different datasets separately and randomly select 50% from each dataset as the training set, and the remaining 50% as the testing set.

The average quantitative assessment results for glaucoma screening of various classifiers are shown in Table 5, and the ROC curves are shown in Fig. 16. From Table 5, we can see that all these classifiers without ADASYN can obtain high specificity but low sensitivity. That is because the distribution of training samples is extremely uneven which can cause severe predictive bias of the classifiers and leads to many glaucoma samples being misdiagnosed as normal. Thus, the data balancing strategy is important for glaucoma screening performance. As a result, GBDT with the ADASYN method obtains

**Fig. 15** Ranking of feature importance. For more details about the features, please see Table 9 in Appendix

the highest sensitivity, Acc, and AUC. Besides, it also achieves the best specificity among the methods that use data balancing strategies and attains the best performance.

### 3.5 Discussion

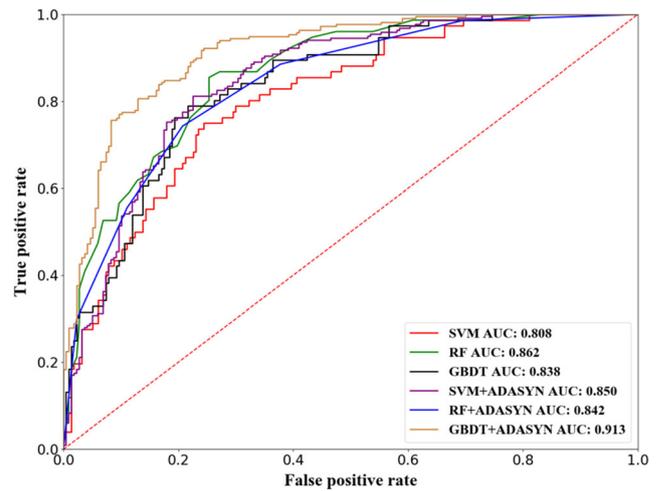
#### 3.5.1 Comparison with different algorithms

In order to test the superiority of the proposed glaucoma screening method, some of the latest glaucoma screening methods—deepCDR [21] and wavelet-based method (Wavelet) [40]—were adopted to compare against GBDT+ADASYN. For the deepCDR method [21], a deep learning architecture, named M-Net, is proposed to solve the OD and OC segmentation, and the glaucoma screening is conducted based on the calculated CDR value. Wavelet-based method extracts statistical moments or entropy as features using the wavelet transform, and SVM is employed for glaucoma classification. Besides, we also compared the methods using Gabor transform [16] and GLCM [41], separately. These methods diagnose glaucoma by hidden statistical information without clinical measurement features. Table 6 shows the average comparison results of the classification performance on the ORIGA dataset, and the ROC curves of different methods are shown in Fig. 17.

From the result, we can deduce that the statistical features extracted based on a single transform cannot achieve satisfactory Acc and specificity, so the features obtained by these methods cannot effectively describe the characteristics of glaucoma and lack certain discrimination. DeepCDR method using CDR can achieve satisfactory performance with 0.791 Acc, 0.865 AUC, and the highest specificity, which outperforms wavelet-based, Gabor, and GLCM-based methods. For the proposed GBDT+ADASYN method, clinical measurement features, hidden statistics features, and texture features are adopted for glaucoma detection, and it obtains the best overall performance, which has a 6.6% improvement on Acc, a 14.6% improvement on sensitivity, and a 5.5% improvement on AUC than deepCDR. Figure 18 compares the ranking results of the evaluation indicators of different methods. Although the specificity of the proposed method is

**Table 5** Classification results of different classifiers on ORIGA dataset

Method	Sensitivity	Specificity	Acc	AUC
RF	0.526	0.908	0.732	0.862
SVM	0.487	0.880	0.693	0.808
GBDT	0.539	0.903	0.735	0.838
RF+ADASYN	0.844	0.774	0.809	0.842
SVM+ADASYN	0.849	0.747	0.797	0.850
GBDT+ADASYN	0.894	0.793	0.843	0.913



**Fig. 16** ROC curves of different classifiers on ORIGA dataset

not the highest, it is very close to deepCDR and greatly superior to other algorithms. Therefore, our proposed method is much superior to state-of-the-art methods and obtains the most reliable performance using the ORIGA dataset.

Besides, we perform comparative experiments based on another dataset DRISHTI-GS1 [38] to further verify the robustness of the proposed glaucoma screening method. Table 7 shows the average classification performance comparison of different algorithms on DRISHTI-GS1. From the table, we can see that the proposed method can obtain the highest specificity, Acc, and AUC, which has 32.9%, 11.6%, and 12.2% improvement than deepCDR. We also find that although deepCDR achieves the highest sensitivity, the specificity was extremely low, indicating that deepCDR does not work well on DRISHTI-GS1 datasets; thus, the robustness of DeepCDR is unsatisfactory.

Figure 19 shows the ROC curves of different methods, and Fig. 20 shows the ranking results of different methods according to the evaluation indicators. These results show that the proposed glaucoma detection method has excellent sensitivity, specificity, Acc, and AUC using the DRISHTI-GS1 dataset. Therefore, the proposed method can be effectively applied to different datasets with

**Table 6** Performance measures of different methods using ORIGA dataset

Method	Sensitivity	Specificity	Acc	AUC
DeepCDR [21]	0.780	0.803	0.791	0.865
Wavelet [40]	0.882	0.668	0.774	0.834
Gabor [16]	0.788	0.735	0.746	0.800
GLCM [41]	0.794	0.714	0.754	0.771
Proposed	0.894	0.793	0.843	0.913

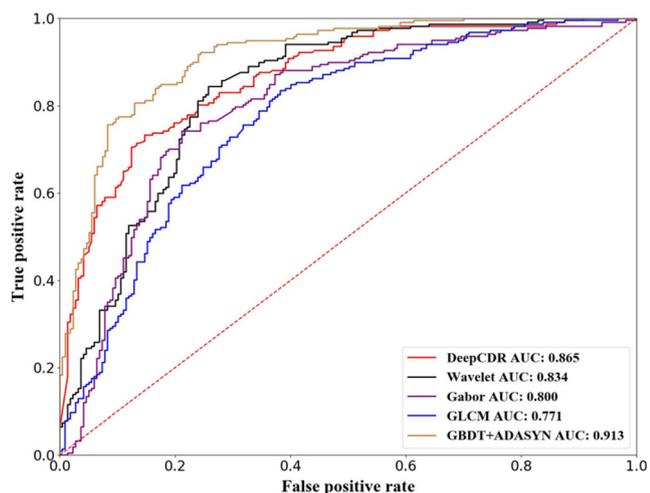


Fig. 17 ROC curves of different methods using ORIGA dataset

excellent robustness and its comprehensive performance is superior to the state-of-the-art algorithm.

It can be seen from the experimental results that glaucoma screening using clinically measured features can achieve good results, but its accuracy is limited to a certain extent, and it is difficult to meet the needs of clinical diagnosis. The reason is that the method ignores

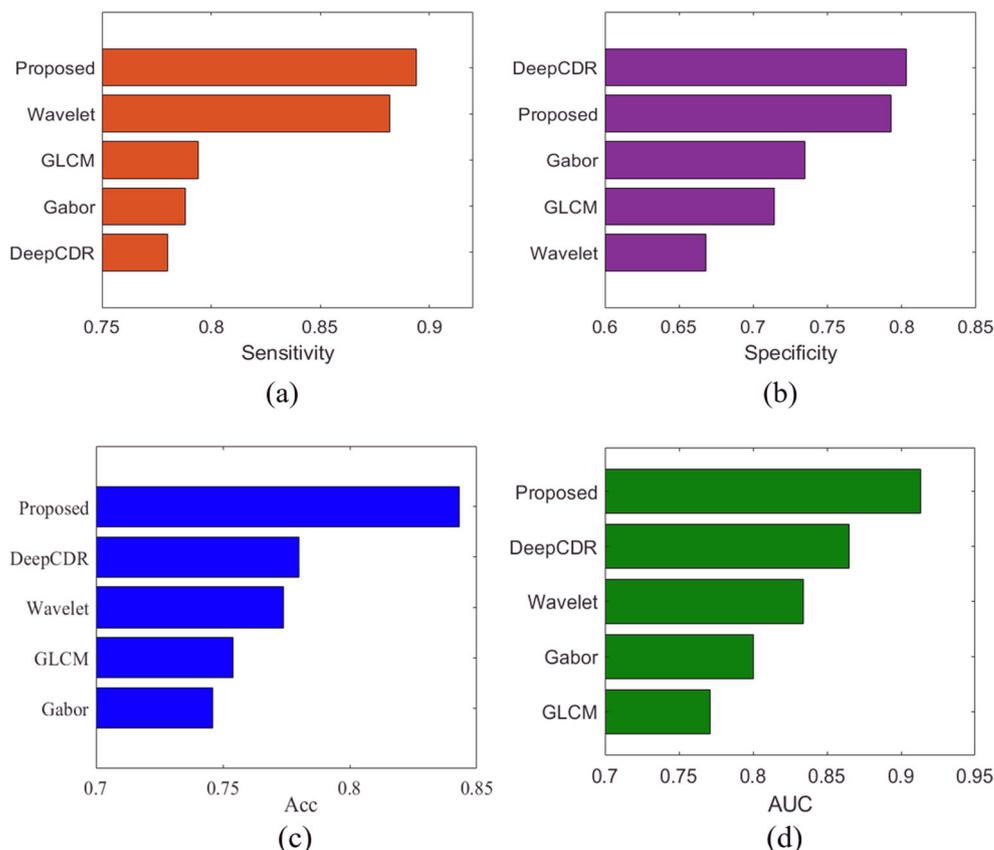
Table 7 Performance measures of different methods using DRISHTI-GS1 dataset

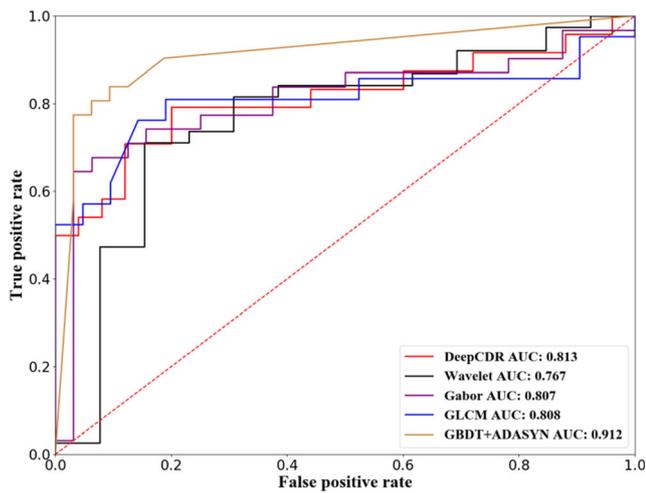
Method	Sensitivity	Specificity	Acc	AUC
DeepCDR [21]	0.938	0.581	0.750	0.813
Wavelet [40]	0.876	0.648	0.760	0.767
Gabor [16]	0.939	0.588	0.781	0.807
GLCM [41]	0.844	0.643	0.743	0.808
Proposed	0.904	0.772	0.837	0.912

other valuable hidden evidence in the fundus image, namely texture and statistical features. The glaucoma detection method based on texture or statistical characteristics can achieve excellent sensitivity, but it cannot be applied to clinical diagnosis because of its low specificity. Therefore, in our method, the above two different kinds of features are combined.

In terms of extracting image-based features, we proposed the IFOV feature model to extract GLCM and Gabor transform features from different receptive fields and fully quantify hidden evidence related to glaucoma. In addition, we combine clinical measurement features related to CDR, NRR, and

Fig. 18 Ranking results of different methods using ORIGA dataset. a Sensitivity ranking result. b Specificity ranking result. c Acc ranking result. d AUC ranking result





**Fig. 19** ROC curves of different methods using DRISHTI-GS1 dataset

ISNT rules with image-based features to enhance the performance of the classifier.

Furthermore, we use the ADASYN method to solve the problem of data imbalance, which also greatly improves the specificity and sensitivity of our classification model. Experimental results show that our proposed method has the best glaucoma screening performance compared with the methods that use clinical

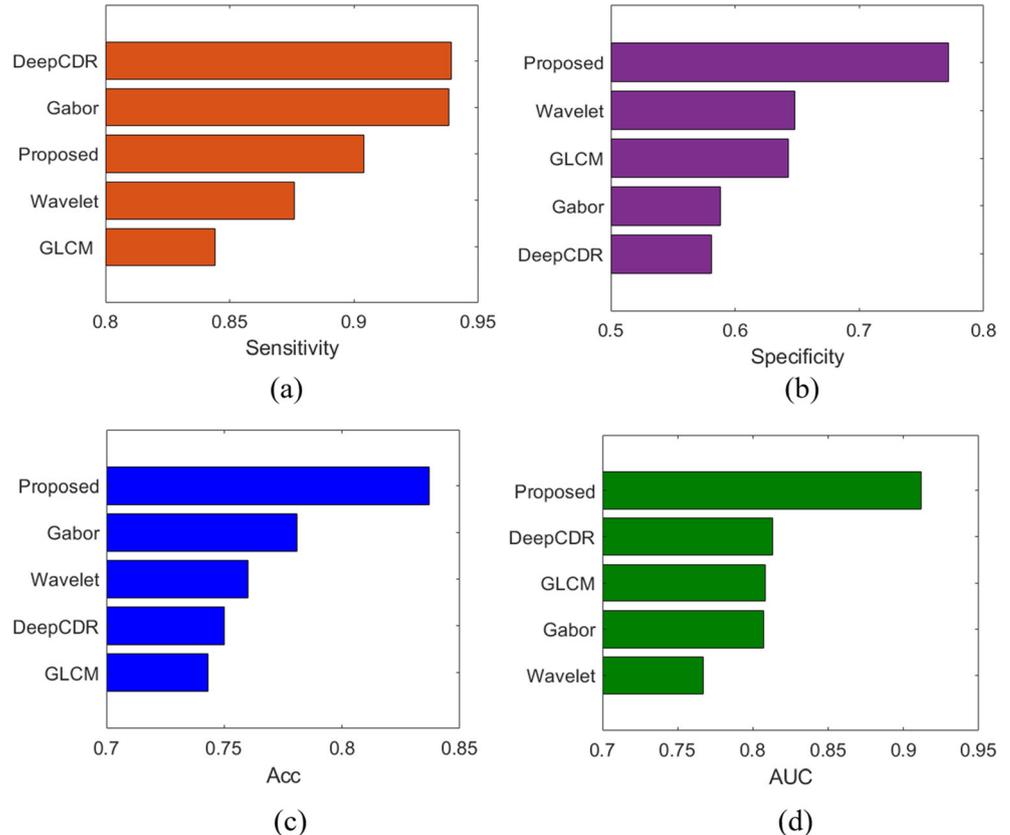
measurement features, such as CDR alone or the methods with a single transform feature model.

### 3.5.2 Running time

The OC and OD simultaneous segmentation task is implemented on the NVIDIA RTX 2080 Ti GPU, and the training phase using the ORIGA dataset takes about 1 h. Besides, the algorithm needs to extract a large number of raw features for analysis and comparison during the feature selection stage, so this process is a little time-consuming. However, the processes of the model training and feature selection can be done offline; their calculation time will not affect the clinical use of the system. Therefore, we compare the time required of the complete test phase including OC and OD segmentation, extraction of selected image features, and glaucoma detection. Specifically, the calculation time of OC and OD segmentation is 0.42 s per image, and the calculation time of the selected feature extraction and glaucoma screening stage is 0.19 s per image. The average test time of a single fundus image of different algorithms is shown in Table 8.

Since the proposed algorithm extracts multiple types of features from different receptive fields, the time taken for feature extraction is slightly longer than other algorithms. But the proposed glaucoma detection method greatly

**Fig. 20** Ranking results of different methods using DRISHTI-GS1 dataset. **a** Sensitivity ranking result. **b** Specificity ranking result. **c** Acc ranking result. **d** AUC ranking result



**Table 8** Comparison of time required to execute different methods

Method	Time	System
DeepCDR [21]	0.50s	NVIDIA Titan X GPU
Wavelet [40]	0.48 s	Intel Xeon E5 2.5 GHz
Gabor [16]	0.50s	NVIDIA RTX 2080 Ti GPU
GLCM [41]	0.55 s	
Proposed	0.61 s	

surpasses other algorithms in AUC, accuracy, sensitivity, and specificity. Besides, obtaining accurate glaucoma screening result in 0.61 s is much more efficient than clinical ophthalmologists, making it acceptable for clinical diagnosis.

## 4 Conclusions

A new automatic glaucoma screening algorithm is proposed in this paper. Both clinically measured features and image-based features are extracted to fully explore the evidence of glaucoma.

To accurately extract clinical measurement features, we propose a network that was called CP-FD-UNet++ to segment

**Table 9** The description of the selected features

Feature name	Feature description
vertical_CDR	Vertical diameter ratio of OC-to-OD
cup_area	Area of OC
ROI_gab_avg(3,0.8,1.18,1)	Energy of the Gabor transform coefficients based on ROI (the Gabor kernels: $\sigma = 3$ , $\theta = 45^\circ$ , $f = 1.1781\text{HZ}$ , $\gamma = 1$ )
thickness_S	Thickness of superior region
thickness_I	Thickness of inferior region
Global_GLCM_ASM(4,2.4)	Angular second moment of the GLCM based on global image (offset = 4, $\theta = 135^\circ$ )
Global_GLCM_ASM(1,1.6)	Angular second moment of the GLCM based on global image (offset = 1, $\theta = 90^\circ$ )
ROI_GLCM_dissimilarity(1,0.0)	Dissimilarity of the GLCM based on ROI (offset = 1, $\theta = 0^\circ$ )
Global_GLCM_ASM(2,0)	Angular second moment of the GLCM based on global image (offset = 2, $\theta = 0^\circ$ )
OC_GLCM_ASM(2,0)	Angular second moment of the GLCM based on OC region (offset = 2, $\theta = 0^\circ$ )
Global_GLCM_ASM(1,0.8)	Angular second moment of the GLCM based on global image (offset = 1, $\theta = 45^\circ$ )
OD_GLCM_ASM(3,1.6)	Angular second moment of the GLCM based on OD region (offset = 3, $\theta = 90^\circ$ )
Global_GLCM_ASM(1,0)	Angular second moment of the GLCM based on global image (offset = 1, $\theta = 0^\circ$ )

the OD and OC simultaneously based on ROI. The network uses a nested U-Net structure, and the input feature maps of the encoders are cascaded to achieve multi-scale information input and feature map sharing. The mixed loss function which combines focal loss with dice coefficient is adopted to weaken the imbalance between the target and the background. Based on the accurate OD and OC segmentation results, clinically measured parameters, such as CDR and NRR, can be obtained. Besides, we propose the IFOV model to fully extract hidden visual features. There are four visual fields with different scales in the IFOV model: OC area, OD area, ROI, and global fundus image. Various texture and statistical features are extracted from Gabor transform and GLCM that applied to these areas. All features are used to train a GBDT classifier after being ranked and selected. The combination of clinical measurement features and hidden texture and statistical features can obtain abundant evidence for glaucoma discrimination and greatly improve the accuracy of glaucoma screening classifier. Experimental results on the ORIGA and DRISHTI-GS1 datasets show that compared with other existing methods, our glaucoma screening method can generally achieve better results in accuracy, sensitivity, specificity, and AUC. Thus, the proposed method can be regarded as a suitable and reliable method for automatic glaucoma screening in a variety of clinical settings.

**Funding information** This work was supported by the National Natural Science Foundation of China under Grant 61502537, the Natural Science Foundation of Hunan Province of China under Grant 2018JJ3681, the Independent Exploration and Innovation Project for Graduate Students of Central South University under Grant 2020zzts567, and the Open Project Fund of the Key Laboratory of Digital Signal and Image Processing of Guangdong Province under Grant 2018GDDSIPL-01.

## Appendix

The description of the features after ranking and selected in Fig. 15 can be seen in Table 9

## References

1. Tham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y (2014) Global prevalence of glaucoma and projections of glaucoma burden through. *Ophthalmology* 121:2081–2090
2. Bourne RRA (2006) Worldwide glaucoma through the looking glass. *Br J Ophthalmol* 90:253–254
3. Shen SY, Wong TY, Foster PJ, Loo JL, Rosman M, Loon SC, Wong WL, Saw SM, Aung T (2008) The prevalence and types of glaucoma in Malay people: the Singapore Malay Eye Study. *Invest Ophthalmol Vis Sci* 49:3846–3851
4. Raghavendra U, Gudigar A, Bhandary SV, Raol TN, Ciaccio EJ, Acharya UR (2019) A two-layer sparse autoencoder for glaucoma identification with fundus images. *J Med Syst* 43:299

5. Zhao X, Guo F, Mai YX, Tang J, Duan XC, Zou BJ, Jiang LZ (2019) Glaucoma screening pipeline based on clinical measurements and hidden features. *IET Image Process* 13:2213–2223
6. Garway-Heath DF, Hitchings RA (1998) Quantitative evaluation of the optic nerve head in early glaucoma. *Brit J Ophthalmol* 82:352–361
7. Budenz DL, Barton K, Whiteside-de Vos J, Schiffman J, Bandi J, Nolan W, Herndon L, Kim H, Hay-Smith G, Tielsch JM (2014) Prevalence of glaucoma in an urban west african population. *JAMA Ophthalmol* 131:651–658
8. Derick R, Pasquale L, Pease M, Quigley H (1994) A clinical study of peripapillary crescents of the optic disc in chronic experimental glaucoma in monkey eyes. *Arch Ophthalmol* 112:846–850
9. Jonas JB, Bergua A, Schmitz-Valckenberg P et al (2000) Ranking of optic disc variables for detection of glaucomatous optic nerve damage. *Invest Ophthalmol Vis Sci* 41:1764–1773
10. Hancox OD, Michael D (1999) Optic disc size, an important consideration in the glaucoma evaluation. *Clin Eye Vis Care* 11:59–62
11. Yin F, Liu J, Wong DW, Tan NM et al (2012) Automated segmentation of optic cup and disc in fundus images for glaucoma diagnosis', In: *The 25th international symposium on computer-based medical systems (CBMS)*
12. Smola A, Vishwanathan S (2008) *Introduction to machine learning*. Cambridge University Press
13. Nayak J, Acharya R, Bhat P, Shetty N, Lim T-C (2009) Automated diagnosis of glaucoma using digital fundus images. *J Med Syst* 33: 337–346
14. Cheng J, Liu J, Xu Y, Yin F, Wong DWK, Tan NM, Tao D, Cheng CY, Aung T, Wong TY (2013) Superpixel classification based optic cup and disc segmentation for glaucoma screening. *IEEE Trans Med Imaging* 32:1019–1032
15. Noronhaam KP, Acharya UR, Nayaka KP, Martis RJ, Bhandary SV (2014) Automated classification of glaucoma stages using higher order cumulant features. *Biomed Signal Proces Control* 10: 174–183
16. Acharyaa UR, Ng EYK, Eugenea LWJ, Noronha KP, Min LC, Nayak KP, Bhandarye SV (2015) Decision support system for the glaucoma using Gabor transformation. *Biomed Signal Proces Control* 15:18–26
17. Dua S, Acharya UR, Sree SV (2012) Wavelet-based energy features for glaucomatous image classification. *IEEE Trans Inf Technol Biomed* 16:80–87
18. Haleem M, Han L, Hemert J, Fleming A, Pasquale LR, Silva PS, Song BJ, Aiello LP (2016) Regional image features model for automatic classification between normal and glaucoma in fundus and scanning laser ophthalmoscopy (SLO) images. *J Med Syst* 40:132
19. Long J, Shelhamer E, and Darrell T (2015) Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intel* 39:640 651
20. Sevastopolsky A Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural. In: *The 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*
21. Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao XC (2018) Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans Med Imaging* 37:1597–1605
22. Chen XY, Xu YW, Wong DWK, Wong WY, Liu J (2015) Glaucoma detection based on deep convolutional neural network. In: *The 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*
23. Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A (2019) CNNs for automatic glaucoma assessment using fundus images: an extensive validation. *Biomed Eng Online* 18:29
24. Zhou ZW, Siddiquee MMR, Tajbakhsh N, Liang J (2018) UNet++: a nested U-Net architecture for medical image segmentation. In: *The 4th Deep Learning in Medical Image Analysis (DLMIA) Workshop*
25. Lin T-Y, Goyal P, Girshick R, He KM, and Dollár P (2017) Focal loss– focal loss for dense object detection. In: *The IEEE International Conference on Computer Vision (ICCV)*
26. He HB, Bai Y, Garcia EA, and Li S (2008) ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: *IEEE International Joint Conference on Neural Networks (IJCNN)*
27. Ronneberger O, Fischer P, and Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: *Proceedings of International Conference on Medical image computing and computer-assisted intervention*
28. Huang G, Liu Z, Van Der Maaten L, and Weinberger KQ (2017) Densely connected convolutional networks. In: *Proceedings of CVPR*
29. Vanajakshi B, Sujatha B, Srirama K (2010) A study on implementation of advanced morphological operations. *IJCSNS Int J Comput Sci Netw Secur* 10:6–9
30. Daugman JG (1988) Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans Acoust Speech Sig Process* 36:1169–1179
31. Liu H, and Setiono R (1995) Chi2: Feature selection and discretization of numeric attributes. In: *Proc. IEEE 7th Int. Conf. Tools with Artif. Intell*
32. Kwak N, Choi C-H (2002) Input feature selection for classification problems. *IEEE T Neural Netw* 13:143–159
33. Vergara JR, Estevez PA (2014) A review of feature selection methods based on mutual information. *Neural Comput Applic* 24: 175–186
34. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232
35. Rao H, Shi X, Rodrigue AK, Feng J, Xia Y, Elhoseny M, Yuan X, Gu L (2019) Feature selection based on artificial bee colony and gradient boosting decision tree. *Appl Soft Comput* 74:634–642
36. Zhang Z, Yin FS, Liu J, Wong WK (2010) Origa-light: An online retinal fundus image database for glaucoma analysis and research. In: *Proceedings of Engineering in Medicine and Biology Society (EMBC)*, pp. 3065–3068
37. Almazroa A, Alodhayb S, Osman E et al (2018) Retinal fundus images for glaucoma analysis: the RIGA dataset. In: *Proceeding of SPIE Medical Imaging, Houston*, pp.1–8
38. Sivaswamy J, Krishnadas S, Chakravarty A et al (2015) A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomed Imaging Data Papers* 2: 1–7
39. Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, Gain P, Ordonez R, Massin P, Erginay A, Charton B, Klein JC (2014) Feedback on a publicly distributed image database: the Messidor database. *Image Anal Stereol* 33:231–234
40. Koh JE, Acharya UR, Hagiwara Y et al (2017) Diagnosis of retinal health in digital fundus images using continuous wavelet transform (CWT) and entropies. *Comput Biol Med* 84:89–97
41. Nugroho HA, Oktoeberza KZW, Adji TB, Najamuddin F (2015) Detection of exudates on color fundus images using texture-based feature extraction. *Int J Technol* 6:121

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Fan Guo** received the B.S. degree in computer science and technology and the M.S. and Ph.D. degrees in computer application technology from Central South University (CSU), Changsha, China, in 2005, 2008, and 2012, respectively. She is currently an Associate Professor with the School of Automation, CSU. Her main research interests include image processing, computer vision, and pattern recognition.



**Beiji Zou** received a B.S. degree from Zhejiang University, Hangzhou, China, in 1982, and the M.S. and Ph.D. degrees in computer science and technology from Tsinghua University in 1984 and Hunan University in 2001, respectively. He is currently a Professor at Central South University. Until now, he has published more than 100 papers in journals. His research focus is on computer graphics, image processing, and pattern recognition.



**Weiqing Li** is currently pursuing a M.S. degree in the Intelligent Information Processing and Systems Laboratory, School of Automation, Central South University, Changsha, China. His research interests include deep learning, machine learning, and digital image processing.



**Zhun Fan** received the B.S. and M.S. degrees from the Huazhong University of Science and Technology, China, in 1995 and 2000, respectively, and the Ph.D. degree from Michigan State University, USA, in 2004. He is currently a Professor with the Department of Electronic and Information Engineering, Shantou University, where he has also served as the Dean of the Key Laboratory of Digital Signal and Image Processing of Guangdong Province. His re-

search focus is on computer vision, machine learning, and artificial intelligence.



**Jin Tang** received the B.S. degree and the M.S. degree from Peking University, Beijing, China, in 1987 and 1990, respectively, and the Ph.D. degree in pattern recognition and intelligence system from Central South University (CSU), Hunan, China, in 2002. He is currently a professor in the School of Automation, CSU, Changsha. His current research interests are focused on image processing, pattern recognition, and computer vision.