

# Automatic retinal image analysis methods using colour fundus images for screening glaucomatous optic neuropathy

Chuying Shi ,<sup>1,2</sup> Jack Lee,<sup>1</sup> Di Shi,<sup>3</sup> Gechun Wang,<sup>2,4</sup> Fei Yuan,<sup>2</sup> Benny Chung-Ying Zee <sup>1</sup>

**To cite:** Shi C, Lee J, Shi D, *et al.* Automatic retinal image analysis methods using colour fundus images for screening glaucomatous optic neuropathy. *BMJ Open Ophthalmology* 2024;**9**:e001594. doi:10.1136/bmjophth-2023-001594

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjophth-2023-001594>).

Received 27 November 2023  
Accepted 25 August 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Center for Clinical Research and Biostatistics, Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

<sup>2</sup>Department of Ophthalmology, Zhongshan Hospital, Fudan University, Shanghai, China

<sup>3</sup>Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, China

<sup>4</sup>Department of Ophthalmology, Affiliated Xiaoshan Hospital, Hangzhou Normal University, Hangzhou, Zhejiang, People's Republic of China

## Correspondence to

Dr Benny Chung-Ying Zee; [bzee@cuhk.edu.hk](mailto:bzee@cuhk.edu.hk)

## ABSTRACT

**Objectives** Train an automatic retinal image analysis (ARIA) method to screen glaucomatous optic neuropathy (GON) on non-mydratic retinal images labelled with the additional results of optical coherence tomography (OCT) and assess different models for the GON classification. **Methods** All the images were obtained from the hospital for training and 10-fold cross-validation. Two methods were used to improve the classification performance: (1) using images labelled with the additional results of OCT as the reference standard and (2) generating models using retinal features from the entire images, the region of interest (ROI) of the optic disc, and the ROI of the macula, and the combination of all the features.

**Results** Overall, we collected 1338 images with paired OCT scans. In 10-fold validation, ARIA achieved sensitivities of 92.2%, 92.7% and 85.7%, specificities of 88.8%, 86.7% and 80.2% and accuracies of 90.6%, 89.9% and 83.1% using the retinal features from the entire images, the ROI of the optic disc and the ROI of the macula, respectively. We found the model combining all the features has the best classification performance and obtained a sensitivity of 92.5%, a specificity of 92.1% and an accuracy of 92.4%, which is significantly different from other models ( $p < 0.001$ ).

**Conclusion** We used two methods to improve the classification performance and found the best model to detect glaucoma on colour fundus retinal images. It can become a cost-effective and relatively more accurate glaucoma screening tool than conventional methods.

## INTRODUCTION

Glaucoma is one of the leading causes of irreversible blindness worldwide. Globally, the number of adults (aged from 40 to 80) with glaucoma was estimated at around 64 million in 2013 and 76 million in 2020 and was expected to increase to 112 million in 2040.<sup>1</sup> Glaucoma is typically asymptomatic in the early stages. Without screening, many patients with glaucoma remain undiagnosed until the late stage, especially in developing countries. For example, in the Chinese population, more than 90% of patients with primary open-angle glaucoma were undiagnosed, and nearly 30% were eventually blind in at least one eye.<sup>2</sup> In a multiethnic Asian

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Many studies have proposed artificial intelligence approaches to diagnosing glaucoma based on fundus retinal images. The gaps are (1) the large-scale datasets they used relying on subjective human gradings of colour fundus retinal images to establish the 'ground truth' of labels to train and validate algorithms and (2) using the overall features of images or only optic nerve head features but neglecting the features on the macula area.

## WHAT THIS STUDY ADDS

⇒ We proposed a machine-learning method and used two methods to solve the problems based on optical coherence tomography (OCT) scans: (1) using images labelled with the additional results of OCT as the reference standard and (2) generating models using retinal features from the entire images, the region of interest (ROI) of the optic disc, and the ROI of the macula, and the combination of all the features. We found the best model to detect glaucoma on colour fundus retinal images, which showed a significant best performance ( $p < 0.001$ ).

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our proposed method has the potential to be a convenient, cost-effective and accurate glaucoma screening tool, which is especially useful in the areas where OCT examination is not available.

population of adults, more than two-thirds of glaucoma cases were undiagnosed.<sup>3</sup> Although glaucoma screening is a significant public health concern, the early detection and treatment of glaucoma can largely reduce visual loss. Therefore, glaucoma is an attractive screening disease because it is asymptomatic, prevalent and treatable.

Historically, glaucoma screening's main barriers have been cost-effectiveness and the lack of an appropriate test.<sup>4</sup> The non-mydratic retinal images are low-cost, portable, quick and straightforward to operate and interpret. In addition, artificial intelligence (AI) can learn good features from images. It may

outperform glaucoma specialists in detecting the disease based on imaging.<sup>5</sup>

Many studies have proposed AI approaches to diagnosing glaucoma based on fundus retinal images. However, the large-scale datasets they used relied on subjective human gradings of colour fundus retinal images to establish the ‘ground truth’ of labels to train and validate algorithms, which only focused on the structure of glaucomatous changes in optic disc photography.<sup>6–9</sup> Nevertheless, relying on retinal images is insufficient to detect glaucoma.<sup>10–11</sup> Ophthalmologists’ subjective interpretation of glaucoma on fundus images is often poor-to-modest intra and interobserver agreement<sup>12–14</sup> as well as poor reproducibility<sup>13</sup> in interpreting images for glaucoma. Essentially, the algorithm cannot outperform the reference standard for the training. Therefore, although their algorithms perform well, their algorithms were limited to subjective gradings used as the reference and the performance may not reflect the true ability of glaucoma detection.

Optical coherence tomography (OCT) is a non-contact, non-invasive and objective structural imaging device for cross-sectional and three-dimensional (3D) viewing of the macula and optic nerve head (ONH). It is a powerful and common tool to diagnose glaucoma, especially mild to moderate glaucoma, by obtaining more objective high-resolution images for the structural assessment of glaucomatous damage<sup>11</sup> and enabling ophthalmologists to achieve good diagnostic accuracy in clinical settings<sup>15–17</sup> with three structures for the assessment: (1) the ONH parameters (includes disc area, rim area, average cup-to-disc ratio (C/D), vertical C/D and cup volume), (2) the retinal nerve fibre layer (RNFL) thickness and (3) the ganglion cell-inner plexiform layer (GCIPL) thickness.<sup>18</sup> Current AI approaches use the overall features of images or only ONH features by segmentation of the optic disc. Few of them used features on the macula area. However, GCIPL thicknesses, a critical characteristic of the macula, also help detect glaucomatous eyes and even showed a higher area under the curve (AUC) and a higher sensitivity than RNFL for detecting glaucoma in the early stages.<sup>19</sup> In this sight, features generated from the macular area can be included in developing algorithms for glaucoma detection.

Although OCT and images can provide different information about optic nerve status, both are useful adjuncts to glaucoma detection.<sup>10</sup> Unfortunately, due to its expensive and inconvenient features, OCT can be used in the clinical setting but is unlikely to be used in a general medical office or a community screening.<sup>20</sup> Moreover, one study found that glaucoma screening by combining OCT with fundus photography showed a 25% higher sensitivity than using fundus photography alone.<sup>21</sup> Hence, the classification performance of AI algorithms should be more accurate and objective if we combine fundus images and OCT measurements as the reference standard rather than the grader’s subjective assessment merely based on fundus images.

Therefore, in this study, we aim to not only train the automated retinal image analysis (ARIA) method to detect glaucomatous optic neuropathy (GON) on non-mydratric retinal images automatically but also improve the performance of GON detection by two methods: (1) using images labelled with the additional results of OCT as the reference standard and (2) combining all the features generated from the entire images, the region of interest (ROI) of the optic disc and the ROI of the macula.

## METHOD

### Datasets

#### Data collection

The datasets for this study were collected from the Department of Ophthalmology, Zhongshan Hospital of Fudan University.

All the images used for training and testing were obtained by a non-mydratric retinal camera (TOPCON/TRC-NW100) at the Department of Ophthalmology, Zhongshan Hospital of Fudan University from 31 October 2012 to 30 September 2021. Besides images, we also collected patients’ age, sex, previous medical history and OCT reports. Each image was paired with the corresponding OCT, the date of which is closest to the date of this image. The inclusion criteria were (1) age equal to or older than 18 years old and (2) paired images and OCT scans. The exclusion criteria were (1) other ocular or systemic diseases that may affect the optic nerve, (2) ungradable colour fundus retinal images, (3) unreliable OCT scans, (4) the interval between the dates of images and the date of corresponding OCT more than 3 months and (5) missing data on fundus images or OCT optic disc scans. Poor image quality was defined as optic disc and vessels invisible on the surface. Unreliable OCT reports were defined as inaccurate data resulting from apparent eye movements, involuntary blinking or saccade or those with a signal strength index <6 were excluded, as recommended by the manufacturer. We used all the eligible images of dataset 1 for the overall features generated from all the images to detect GON.

The images of eyes in datasets 2, 3 and 4 are all derived from dataset 1. Dataset 2 segmented the optic disc and generated glaucomatous ONH features based on OCT optic disc scans. Therefore, for better optic disc localisation and ONH feature generation, we included images of good quality on the optic disc and paired OCT optic disc scan with a signal strength index  $\geq 7$ . Dataset 3 segmented the macula and generated glaucomatous macular features based on OCT macular scans. Similarly, we included images of good quality on the macula and paired the OCT macular scan with a signal strength index  $\geq 7$ . Dataset 4 was used to compare the GON classification performance of different models. Thus, only the images of overlapped eyes in datasets 1, 2 and 3 were included, and dataset 4 was composed. Online supplemental figure 1 describes the process of image selection and dataset building.

### Labelling for the primary dataset

All the images in the primary datasets were labelled as one of three classifications:

1. Images without GON.
2. Images with probable GON.
3. Images with definite GON.

Images were subjected to two ophthalmologists (CS and GW) with 3 years of clinical experience in glaucoma assessed GON based on pairs of gradable images and the information on reliable OCT optic disc reports and labelled each image as ‘no GON’, ‘probable GON’ or ‘definite GON’ based on the definition of classification<sup>7</sup> showing in online supplemental table 1. Probable GON and definite GON were combined as a GON group for binary classification in training and testing. If there were a discordance between ophthalmologists, another specialist (FY) with glaucoma experience over ten years would make a final decision.

### Colour fundus retinal images

We used a Non-Mydriatic Retinal Camera (TOPCON TRC-NW100 Non-Mydriatic Retinal Camera, Tokyo, Japan), which can obtain high-resolution colour images of the retina with small pupils. The colour touch-screen display monitor provides total control of image acquisition and display of the acquired images.

### OCT scanning

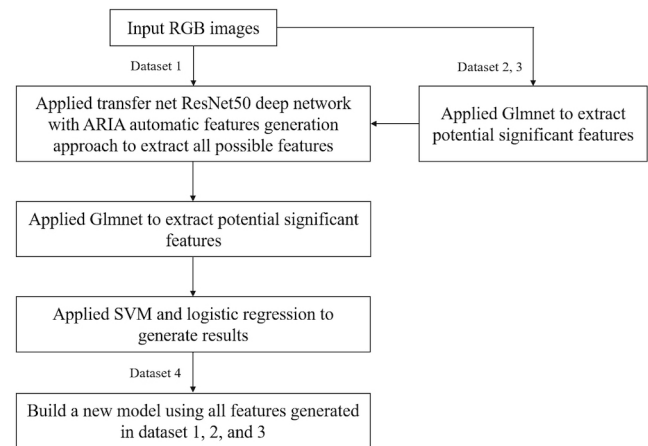
Cirrus HD-OCT (Carl Zeiss Meditec 5000, Dublin, California, USA) optic disc scan automatically locates a circle of 3.46mm diameter evenly around the centre of the optic disc. It generates a 200×200 Optic Disc Cube of data through a 6mm square grid by acquiring 200 horizontal scan lines, each composed of 200 A-scans. The system also calculates the average RNFL thickness for each A-scan in pixels 30 µm square throughout the cube outside the optic disc.

Cirrus HD-OCT macular scan includes the Ganglion Cell OU Analysis measures the thicknesses for the sum of the ganglion cell layer and IPL (GCL+IPL) using data from the 512×128 Macular Cube. This macular scan generates a cube of data through a 6mm square grid by acquiring 128 horizontal scan lines, each composed of 512 A-scans. The thickness map shows thickness measurements of the GCL+IPL in the 6mm by 6mm cube, which contains an elliptical annulus centred on the fovea.

### Development of the automatic retinal image analysis method

#### ARIA method

The presented analysis method incorporates the ARIA features generation approach we have developed using Matlab computer software to acquire and analyse retinal images. The ARIA method has been reported (US Patent 8787638 B2 <https://patents.google.com/patent/US20120257164A1/en> (<https://patents.google.com/patent/US20120257164A1/en>)).<sup>22</sup> Approaches including Haralick texture features analysis, fractal analysis and a set of modified pretrained deep networks (ie, modified



**Figure 1** Flow chart of the presented method in GON detection. Glnet, generalised linear model via penalised maximum likelihood; GON, glaucomatous optic neuropathy; ResNet50, Residual network that is 50 layers deep; SVM, support vector machine.

transfer network resnet-50) were used to create the highly related pixels of features. Training of the ARIA entailed exposure to multiple labelled retinal images (with and without each condition). After the training, the ARIA could automatically classify different categories.

#### Feature generation

The flow chart presents a detailed analysis method, as shown in figure 1. To generate exact overall features on the images, we applied transfer net ResNet-50 deep network with retinal images in dataset 1 as input and features generated at the layer of ‘fc1000\_softmax’ as output based on pixels associated with GON.<sup>23</sup> Then, the same ARIA automatic features generation approach was used to extract features associated with GON, written in Matlab.<sup>22</sup> Finally, we also applied the Glnet approach to select the most important subsets of features highly associated with GON.<sup>24</sup>

#### Automatic segmentation

In datasets 2 and 3, to generate exact optic disc features and macular features, we first localised and segmented the optic disc and macula by setting the ROI as the rectangular area surrounding the optic disc or macula, respectively (online supplemental figure 2). The intensity-based method was used for ROI localisation. To consistently detect the optic disc and the macula, we corrected background brightness by normalising the intensity of various images beforehand. The optic disc shows the highest intensity with the retinal pattern while the fovea (centre of the macula) shows lower intensity in fundus images. The probability density function (PDF) peak was found to be the highest intensity value in the histogram of the optic disc. Then, the pixels were set to the optic disc region. Accordingly, we could locate the macular region with the lowest intensity values by computing the intensity values within the 99% CI of the

PDF from the centre of mass. After segmentation, the process of feature extraction in datasets 2 and 3 was the same as in dataset 1.

### Model building

With the features generated in datasets 1, 2 and 3, we used machine-learning methods and support vector machine (SVM) to generate models A, B and C, respectively. In dataset 4, we used all features generated in datasets 1, 2 and 3 to build a new model D using SVM. Since dataset 4 is composed of overlapping images in datasets 1, 2 and 3, the comparison of classification performances of four models was conducted in dataset 4 using SVM. To avoid the overfitting problem, we applied 10-fold cross-validation to generate more robust results.

Similarly, we used logistic regression models to further verify the performance improvement by combining all features. First, logistic regression models 1, 2 and 3 were generated with the features generated in datasets 1, 2 and 3, respectively, were generated. Then, in dataset 4, we also generated a new model 4 using all features extracted in models 1, 2 and 3. Finally, the classification performance of new model 4 was compared with the logistic regression models 1, 2 and 3 in dataset 4.

### Statistical analysis

For patients' demographic data, the analysis of variance was used to compare continuous data, and the  $\chi^2$  test was used to compare categorical data. The performance of models was evaluated by calculating sensitivity, specificity, accuracy and the AUC of the receiver operator characteristic (ROC) analysis curve. All the features were included in the logistic regression analysis. The backward elimination method selected the statistically significant variables with a level of significance <10% ( $p < 0.1$ ) for prediction models. The DeLong and Associates method,

which compares areas under the ROC curve (AUROC)<sup>25</sup> between algorithms and manual gradings, was used to compare the classification performance of different models. P values <0.05 were considered statistically significant. All analyses were performed using the software SPSS V.20, SAS Studio and Matlab 2020a.

### RESULTS

We collected 1338 images with paired OCT optic disc scans and built dataset 1 (table 1 and online supplemental figure 3). There were 544 patients (190 without GON, 174 with probable GON and 180 with definite GON). Age is significantly different in the three groups of patients ( $p < 0.001$ ), whereas sex has no significant difference ( $p = 0.051$ ). In datasets 2, 3 and 4, there are 1131 images with paired OCT optic disc scans, 1021 images with paired OCT macula scans and 955 images with paired OCT optic disc scans and macular scans, respectively (online supplemental table 2).

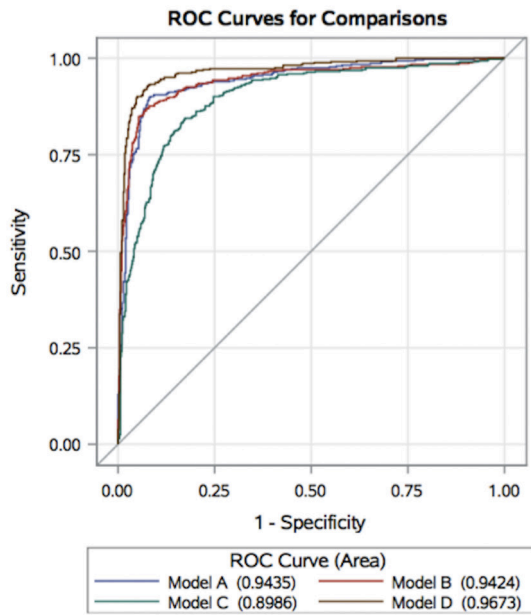
In 10-fold cross-validation (online supplemental table 3), the SVM model A in dataset 1 achieved the best performance with a sensitivity of 92.4%, a specificity of 85.1% and an accuracy of 89.2%, followed by SVM model B in dataset 2 with a sensitivity of 93.8%, a specificity of 82.9% and an accuracy of 88.9%. The SVM model C in dataset 3 achieved the worst performance, with a sensitivity of 82.2%, a specificity of 72.8% and an accuracy of 78.0%.

In dataset 4, we built SVM model D by using all the features generated in datasets 1, 2 and 3 (figure 2). It achieved the best performance with a sensitivity of 92.5%, a specificity of 92.1%, an accuracy of 92.4% and AUC of 0.967. In the comparison of the classification performance of SVM models in dataset 4 (table 2 and online supplemental figure 4), SVM model D also showed statistically better classification performance than models A

**Table 1** Demographic data of dataset 1 for training and testing (n=1338)

	Control	Probable GON	Definite GON
	Mean±SD		
<b>Patients</b>			
Number (n)	190	174	180
Age (year)	48.3±15.9	52.4±18.2	62.9±15.4
Sex (female (%))	106 (55.8)	83 (47.7)	78 (43.3)
<b>Images</b>			
Number (n)	577	416	345
RNFL thickness (µm)	97.70±8.73	87.82±12.01	70.20±16.77
Rim area (mm <sup>2</sup> )	1.33±0.24	1.14±0.30	0.80±0.25
Disc area (mm <sup>2</sup> )	2.01±0.39	2.12±0.52	2.14±0.49
Average C/D	0.55±0.15	0.63±0.17	0.78±0.09
Vertical C/D	0.51±0.15	0.60±0.17	0.78±0.09
Cup volume (mm <sup>3</sup> )	0.22±0.16	0.37±0.27	0.55±0.27

C/D, cup-to-disc; GON, glaucomatous optic neuropathy; RNFL, retinal nerve fibre layer.



**Figure 2** The comparison of classification performance of SVM model A (entire image), model B (ROI of ONH), model C (ROI of macula) and model D (combination) in dataset 4. ONH, optic nerve head; ROC, receiver operator characteristic; ROI, region of interest; SVM, support vector machine

( $p=0.0001$ ), B ( $p=0.0002$ ) and C ( $p<0.0001$ ). In **figure 2**, the orange line (model D) also shows the best AUC curve in classifying images with GON and no GON.

Classification performances of logistic regressions in terms of confusion matrix, sensitivities, specificities, accuracies and AUROC values are displayed in online supplemental table 4. In dataset 1, model 1 achieved a sensitivity of 90.3%, a specificity of 89.4%, an accuracy of 89.9% and an AUC of 0.945. In dataset 2, model 2 achieved a sensitivity of 93.9%, a specificity of 85.5%, an accuracy of 90.2% and an AUC of 0.950. In dataset 3, model 3 achieved a sensitivity of 82.8%, a specificity of 79.1%, an accuracy of 81.1% and an AUC of 0.889.

**Table 2** The classification performance of support vector machine models using 10-fold cross-validation in dataset 4

	Se	Sp	Acc	AUC	P value
Model A	92.2	88.8	90.6	0.943	<b>0.0001</b>
Model B	92.7	86.7	89.9	0.942	<b>0.0002</b>
Model C	85.7	80.2	83.1	0.899	<b>&lt;0.0001</b>
Model D	92.5	92.1	92.4	0.967	–

Model A: entire image; model B: region of interest (ROI) of optic nerve head; model C: ROI of macula; mode D: combine all features generated in model A, B and C.

P values are derived from the DeLong and Associates method comparing areas under the ROC curve.

Se: sensitivity; Sp: specificity; Acc: accuracy.

AUC, area under the curve.

A comparison of the classification performance of logistic regression models 1, 2, 3 and 4 in dataset 4 is shown in **table 3** and online supplemental figure 5. Model 4, which used all the features generated from datasets 1, 2 and 3, achieved the highest performance with a sensitivity of 92.8 %, a specificity of 93.5%, an accuracy of 93.2% and an AUC of 0.981. Model 4 also performs significantly better classification than models 1, 2 and 3 ( $p<0.0001$ ). In online supplemental figure 5, the orange line (model 4) also shows the best AUC curve in classifying images with GON and no GON.

## DISCUSSION

In this study, we developed and tested the ARIA method using colour fundus retinal images as the input for automated GON identification. We also proposed two methods to improve GON detection accuracy based on the characteristics of glaucoma compared with other AI approaches. The proposed algorithm performs well in detecting probable GON and definite GON on images with the additional information of OCT scans. Therefore, our presented method has the potential to be an accurate, convenient and cost-effective screening tool and may be comparable to OCT examination for GON detection.

We proposed an innovative method and found a statistically significant improvement in performance after combining all the features, compared with those models using features generated from entire images, ROI of the optic disc and ROI of the macula separately using machine-learning methods and traditional logistic regression methods. In the traditional logistic regression models, the model combining all the features has a better sensitivity (92.8% vs 91.7%), a better specificity (93.5% vs 89.4%) and a better accuracy (93.2% vs 90.5%), compared with the model that used features generated from entire images with the second good performance. In the machine-learning models in 10-fold cross-validation, the model combining all the features has a better sensitivity (92.5% vs 92.2%), a better specificity (92.1% vs 88.8%) and a better accuracy (92.4% vs 90.6%) compared with the model that used features generated from entire images with the second good performance.

Many studies used all the images or the ROI of the optic disc area when training algorithms to detect GON because the mainstay of detection of glaucoma focuses on the ONH and retinal RNFL outside ONH.<sup>26–29</sup> Our results are consistent with this. The models (1 and 2 or A and B) include the ROI features of the optic disc and have relatively better performance than the models (3 or C) that did not include them.

However, in recent decades, with the development of OCT, GCIPL in the macular area could be quantitatively assessed. Glaucoma was also found tending to preferentially influence the innermost retinal layers of the macula, including the GCL, which contains retinal ganglion cell (RGC) bodies and IPL, which contains RGC dendrites, bipolar cell axons and amacrine processes.<sup>30</sup> OCT

**Table 3** Comparison of classification performance of logistic regression models 1, 2, 3 and 4 in dataset 4 (n=955)

	Label: 0	Label: 1	Total	Se	Sp	Acc	AUC	P value
Model 1				91.7	89.4	90.5	0.953	<0.0001
Est: 0	408	54	462					
Est: 1	37	456	493					
Model 2				93.7	85.6	89.9	0.949	<0.0001
Est: 0	381	32	413					
Est: 1	64	478	542					
Model 3				79.1	81.4	80.3	0.883	<0.0001
Est: 0	352	95	447					
Est: 1	93	415	508					
Model 4				92.8	93.5	93.2	0.981	–
Est: 0	413	33	446					
Est: 1	32	477	509					
Total	445	510	955					

P values are derived from the Delong and Associates method comparing areas under the ROC curve.

Se: sensitivity; Sp: specificity; Acc: accuracy.

AUC, area under the curve; Est, estimation.

imaging of the ONH, RNFL and macula was a routine of the clinical examinations for glaucoma.<sup>18</sup> Moreover, another study found combining structural measurements of ganglion cell complex, RNFL and ONH variables from OCT improved the accuracy of glaucoma diagnosis.<sup>31</sup> Therefore, in our study, the improvement of model 4 and model D could be well explained by the clinical characteristics of glaucoma and further confirmed that ONH parameters, RNFL and GCIPL are complementary in glaucoma diagnosis, and the combination of features can statistically increase the diagnostic accuracy.

The classification ability of our algorithms is comparable with other studies with sensitivity, specificity and accuracy >90% in logistic regression models and SVM classifiers. We achieved the highest AUC of 0.967 in the SVM and 0.981 in the logistic regression model, whereas many studies' AUCs are slightly over 0.90. Although a few studies also achieved AUC>0.96,<sup>6</sup> our algorithm should have relatively better accuracy because their algorithms were limited to the reference they used, the manual assessment based on images.

Our algorithm also showed comparable classification ability compared with three studies (AUC 0.996, 0.970 and 0.939) that used additional examination results as the reference for labelling.<sup>32–34</sup> However, two studies achieving an AUC of 0.996 and an AUC of 0.970 only classified clinically diagnosed glaucoma and healthy control but did not mention the inclusion of probable glaucoma.<sup>32–34</sup> In our study, demographic information showed that probable GON is at the intermediate level between no GON and definite GON, which means the classification between probable GON and no GON is more complex than the classification between definite GON and no GON. Additionally, the category of probable GON is essential and meaningful in the setting of

glaucoma screening. It refers to a patient with clinical findings or a constellation of risk factors that indicate a significant likelihood of converting to glaucoma, which may vary from person to person and increase with the increment of their risk factors.<sup>10</sup> Therefore, these patients should be detected and referred to ophthalmologists for treatment determination (including the decision of treatment commencement, setting of target intraocular pressure and therapies) and the following monitoring with different visit periods varying among patients.<sup>10</sup>

Our study had several strengths. First, we reviewed both the colour fundus retinal images and corresponding OCT scans for ground truth labelling. It would be more accurate and consistent to identify GON, especially probable GON, than the manual assessment based on fundus images. Second, from the ophthalmologist's perspective, we considered and included the structure changes of GON not only in the optic disc area (ONH parameters and RNFL) but also in the macular area (GCIPL) in training and found the model with the combined features significantly outperforming models with features in respective areas. Third, we included probable GON (individuals who demonstrate features suggestive of glaucoma but do not yet completely fulfil the diagnostic criteria) in the training and testing, which poses more challenges to the classification because the distinguishment between probable GON and control is more complex and less distinguishable compared with the distinguishment between definite GON and control. Thus, our algorithms can detect patients at an early stage of glaucoma as a screening tool. Fourth, all the images in our studies were non-mydratic photography, which is more convenient, feasible and tolerable than mydratic images. Although non-mydratic retinal images pose more challenges in training due to the low pixel

and relative blurriness, our results still showed excellent performance, facilitating the glaucoma screening in telemedicine. Few studies mentioned using non-mydriatic images to assess glaucoma. However, teleglaucoma, as a practical, convenient and beneficial solution for patients and the healthcare system, is more commonly performed without dilating agents and uses non-mydriatic images.

However, there were still several limitations. First, the sample size of training is relatively small, but the demographic data of datasets showed that the images contain a variety of GON and health control, which guaranteed the robustness to be further applied in the clinic. Second, we did not include an external validation dataset. We considered combining our methods with the OCT information and segmentation methods to improve the classification performance. Then, we can optimise our algorithms and test them in a final validation before applying them in real-world settings.

## CONCLUSION

ARIA method for probable and definite GON detection with the additional information of OCT scans achieved a good discriminative performance. Our proposed machine-learning methods combining all the features of entire images, ROI of the macula and ROI of the optic disc, was proved to have a significant improvement in GON detection using the reference with the information of images, OCT optic disc scans and OCT macular scans. Therefore, it has the potential to be a convenient, cost-effective and accurate glaucoma screening tool, which is especially useful in the area where OCT examination is not available.

**Contributors** CS was involved in planning the work, collecting and organising data, performing the analysis, interpreting the results, drafting the manuscript and designing the figures. JL performed the analysis. DS, GW and FY did the data collection and data organisation. BC-YZ supervised this study and is the guarantor. All authors discussed the results and commented on the manuscript.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** This study was approved by the Ethics Committee of Zhongshan Hospital (reference number: B2021-799R).

**Provenance and peer review** Part of a Topic Collection; Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Chuying Shi <http://orcid.org/0000-0002-8010-8585>

Benny Chung-Ying Zee <http://orcid.org/0000-0002-7238-845X>

## REFERENCES

- 1 Tham Y-C, Li X, Wong TY, *et al*. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* 2014;121:2081–90.
- 2 Song W, Shan L, Cheng F, *et al*. Prevalence of glaucoma in a rural northern china adult population: a population-based survey in kailu county, inner mongolia. *Ophthalmology* 2011;118:1982–8.
- 3 Chua J, Baskaran M, Ong PG, *et al*. Prevalence, Risk Factors, and Visual Features of Undiagnosed Glaucoma: The Singapore Epidemiology of Eye Diseases Study. *JAMA Ophthalmol* 2015;133:938–46.
- 4 Tan NYQ, Friedman DS, Stalmans I, *et al*. Glaucoma screening: where are we and where do we need to go? *Curr Opin Ophthalmol* 2020;31:91–100.
- 5 Girard MJA, Schmetterer L. Artificial intelligence and deep learning in glaucoma: Current state and future prospects. *Prog Brain Res* 2020;257:37–64.
- 6 Li Z, He Y, Keel S, *et al*. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology* 2018;125:1199–206.
- 7 Liu H, Li L, Wormstone IM, *et al*. Development and Validation of a Deep Learning System to Detect Glaucomatous Optic Neuropathy Using Fundus Photographs. *JAMA Ophthalmol* 2019;137:1353–60.
- 8 Ting DSW, Cheung CY-L, Lim G, *et al*. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA* 2017;318:2211–23.
- 9 Phene S, Dunn RC, Hammel N, *et al*. Deep Learning and Glaucoma Specialists: The Relative Importance of Optic Disc Features to Predict Glaucoma Referral in Fundus Photographs. *Ophthalmology* 2019;126:1627–39.
- 10 Prum BE, Lim MC, Mansberger SL, *et al*. Primary Open-Angle Glaucoma Suspect Preferred Practice Pattern® Guidelines. *Ophthalmology* 2016;123:112–51.
- 11 Bae HW, Lee KH, Lee N, *et al*. Visual Fields and OCT Role in Diagnosis of Glaucoma. *Optom Vis Sci* 2014;91:1312–9.
- 12 Jampel HD, Friedman D, Quigley H, *et al*. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. *Am J Ophthalmol* 2009;147:39–44.
- 13 Abrams LS, Scott IU, Spaeth GL, *et al*. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology* 1994;101:1662–7.
- 14 Rossetto JD, Melo LAS Jr, Campos MS, *et al*. Agreement on the evaluation of glaucomatous optic nerve head findings by ophthalmology residents and a glaucoma specialist. *Clin Ophthalmol* 2017;11:1281–4.
- 15 Bussell II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol* 2014;98 Suppl 2:ii15–9.
- 16 Lu AT-H, Wang M, Varma R, *et al*. Combining nerve fiber layer parameters to optimize glaucoma diagnosis with optical coherence tomography. *Ophthalmology* 2008;115:1352–7.
- 17 Kumar S, Giubilato A, Morgan W, *et al*. Glaucoma screening: analysis of conventional and telemedicine-friendly devices. *Clinical Exper Ophthalmology* 2007;35:237–43.
- 18 Chen TC, Hoguet A, Junk AK, *et al*. Spectral-Domain OCT: Helping the Clinician Diagnose Glaucoma: A Report by the American Academy of Ophthalmology. *Ophthalmology* 2018;125:1817–27.
- 19 Yadav VK, Rana J, Singh A, *et al*. Evaluation of ganglion cell-inner plexiform layer thickness in the diagnosis of pre-perimetric glaucoma and comparison to retinal nerve fiber layers. *Indian J Ophthalmol* 2024;72:357–62.
- 20 Quigley HA. Identifying Glaucoma in Primary Care Offices. *JAMA Ophthalmol* 2022;140:663–4.
- 21 Watanabe T, Hiratsuka Y, Kita Y, *et al*. Combining Optical Coherence Tomography and Fundus Photography to Improve Glaucoma Screening. *Diagnostics (Basel)* 2022;12:1100.
- 22 Zee CY, Lee JW, Li Q. Method and device for retinal image analysis. US8787638 B2. 2014. Available: <https://patents.google.com/patent/US20120257164A1/en>



- 23 Kaiming H, Xiangyu Z, Shaoqing R, *et al.* *Deep Residual Learning for Image Recognition*. 2016:770–8.
- 24 Hastie T. The elements of statistical learning: data mining, inference, and prediction. Second edition ed. In: Tibshirani R and Friedman JH ed. New York: Springer, 2009.
- 25 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
- 26 Chauhan BC, O’Leary N, AlMobarak FA, *et al.* Enhanced detection of open-angle glaucoma with an anatomically accurate optical coherence tomography-derived neuroretinal rim parameter. *Ophthalmology* 2013;120:535–43.
- 27 Skaat A, De Moraes CG, Bowd C, *et al.* African Descent and Glaucoma Evaluation Study (ADAGES): Racial Differences in Optic Disc Hemorrhage and Beta-Zone Parapapillary Atrophy. *Ophthalmology* 2016;123:1476–83.
- 28 Zhang X, Parrish RK, Greenfield DS, *et al.* Predictive Factors for the Rate of Visual Field Progression in the Advanced Imaging for Glaucoma Study. *Am J Ophthalmol* 2019;202:62–71.
- 29 Yu M, Lin C, Weinreb RN, *et al.* Risk of Visual Field Progression in Glaucoma Patients with Progressive Retinal Nerve Fiber Layer Thinning: A 5-Year Prospective Study. *Ophthalmology* 2016;123:1201–10.
- 30 Geevarghese A, Wollstein G, Ishikawa H, *et al.* Optical Coherence Tomography and Glaucoma. *Annu Rev Vis Sci* 2021;7:693–726.
- 31 Loewen NA, Zhang X, Tan O, *et al.* Combining measurements from three anatomical areas for glaucoma diagnosis using Fourier-domain optical coherence tomography. *Br J Ophthalmol* 2015;99:1224–9.
- 32 Liu S, Graham SL, Schulz A, *et al.* A Deep Learning-Based Algorithm Identifies Glaucomatous Discs Using Monoscopic Fundus Photographs. *Ophthalmol Glaucoma* 2018;1:15–22.
- 33 Lee J, Kim Y, Kim JH, *et al.* Screening Glaucoma With Red-free Fundus Photography Using Deep Learning Classifier and Polar Transformation. *J Glaucoma* 2019;28:258–64.
- 34 Hemelings R, Elen B, Barbosa-Breda J, *et al.* Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. *Acta Ophthalmol* 2020;98:e94–100.