



Detecting glaucoma based on spectral domain optical coherence tomography imaging of peripapillary retinal nerve fiber layer: a comparison study between hand-crafted features and deep learning model

Ce Zheng^{1,2} · Xiaolin Xie² · Longtao Huang³ · Binyao Chen² · Jianling Yang² · Jiewei Lu³ · Tong Qiao¹ · Zhun Fan³ · Mingzhi Zhang²

Received: 28 August 2019 / Revised: 2 November 2019 / Accepted: 8 November 2019 / Published online: 7 December 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Purpose To develop a deep learning (DL) model for automated detection of glaucoma and to compare diagnostic capability against hand-craft features (HCFs) based on spectral domain optical coherence tomography (SD-OCT) peripapillary retinal nerve fiber layer (pRNFL) images.

Methods A DL model with pre-trained convolutional neural network (CNN) based was trained using a retrospective training set of 1501 pRNFL OCT images, which included 690 images from 153 glaucoma patients and 811 images from 394 normal subjects. The DL model was further tested in an independent test set of 50 images from 50 glaucoma patients and 52 images from 52 normal subjects. A customized software was used to extract and measure HCFs including pRNFL thickness in average and four different sectors. Area under the receiver operator characteristics (AROC) curves was calculated to compare the diagnostic capability between DL model and hand-crafted pRNFL parameters.

Results In this study, the DL model achieved an AROC of 0.99 [CI: 0.97 to 1.00] which was significantly larger than the AROC values of all other HCFs (AROCs 0.661 with 95% CI 0.549 to 0.772 for temporal sector, AROC 0.696 with 95% CI 0.549 to 0.799 for nasal sector, AROC 0.913 with 95% CI 0.855 to 0.970 for superior sector, AROC 0.938 with 95% CI 0.894 to 0.982 for inferior sector, and AROC 0.895 with 95% CI 0.832 to 0.957 for average).

Conclusion Our study demonstrated that DL models based on pre-trained CNN are capable of identifying glaucoma with high sensitivity and specificity based on SD-OCT pRNFL images.

Keywords Deep learning · Glaucoma · Spectral domain optical coherence tomography · Peripapillary retinal nerve fiber layer

✉ Mingzhi Zhang
zmz0754@126.com

Ce Zheng
zhengce@me.com

Xiaolin Xie
xiexl@jsiec.org

Longtao Huang
171thuang@stu.edu.cn

Binyao Chen
cby@jsiec.org

Jianling Yang
yjl@jsiec.org

Jiewei Lu
12jwlu1@stu.edu.cn

Tong Qiao
qiaojoel@163.com

Zhun Fan
zfan@stu.edu.cn

¹ Department of Ophthalmology, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai, China

² Joint Shantou International Eye Center of Shantou University and the Chinese University of Hong Kong, Shantou University Medical College, Shantou 515000, Guangdong, China

³ Department of Electronic Engineering, Shantou University, Shantou, Guangdong, China

Introduction

Glaucoma is an optic neuropathy characterized by a slow progressive degeneration of retinal ganglion cells (RGCs), which leads to irreversible visual field defects [1]. Structural alteration of the optic nerve head (ONH) and retinal nerve fiber layer (RNFL) is an early sign of glaucoma [2]. Newer ocular imaging methods, such as spectral domain optical coherence tomography (SD-OCT), allow for noninvasive quantitative imaging of these affected structures in order to facilitate the detection and monitoring of glaucoma [3]. As peripapillary RNFL (pRNFL) comprising the axons of all RGCs from the entire retina, pRNFL parameter measurement is currently the most commonly utilized OCT parameter for clinical glaucoma assessments [4]. Several studies have reported good diagnostic ability of pRNFL parameters in glaucoma using SD-OCT [5, 6]. These, as well as most of the previously studies with hand-crafted features (HCFs), have employed an object segmentation including hand-designed program, followed by utilizing domain knowledge to manually or semi-automatically extract features, like pRNFL thickness, and finally make diagnosis using statistical classifiers or shallow neural computational machine-learning classifiers designed specifically for each class of objects [3, 7].

The advantage of HCFs is that they can reflect expert knowledge by selecting and analyzing features which are suitable for a given problem, like glaucoma. However, hand-designed program and feature extraction may be difficult and time-consuming. By contrast to hand-craft features, deep learning (DL) algorithms learn the most predictive features directly from the images. The DL methods have been applied to various medical imaging modalities, such as fundus photography, lung radiography, computer tomography (CT), and MRI [8–11]. Moreover, some studies have report DL model significantly outperformed other machine learning methods for detection of glaucomatous visual field (VF) change [12]. To the best of our knowledge, there exists no study that compares diagnostic accuracies of DL model and HCFs of pRNFL parameters obtained with SD-OCT to differentiate normal eyes from eyes with glaucoma. Thus, the purpose of this study was to evaluate the efficacy of a DL model for detecting of glaucomatous pRNFL changes and to compare the DL model against HCFs based on SD-OCT pRNFL imaging.

Methods

Chinese subjects above the age of 20 years were prospectively recruited from Joint Shantou International Eye Center of Shantou University and the Chinese University of Hong Kong (JSIEC). All participants engaged in an informed consent process and signed a written consent document before study procedures were carried out. This study was conducted

according to the tenets of the Declaration of Helsinki and had the approval of the institutional review board.

Examinations and inclusion criteria

All subjects were asked about their medical and ophthalmic history and underwent a standardized ophthalmic examination including: slit-lamp examination, Snellen visual acuity, intraocular pressure (IOP) measurement by Goldmann applanation tonometry, stereoscopic optic disc examination with a 78-D Volk lens, gonioscopy performed in the dark using a Goldmann 2-mirror lens, as well as VF testing (Humphrey Field Analyzer II-750i, Carl Zeiss Meditec).

Glaucomatous eyes were defined by the glaucomatous appearance of the optic nerve head on color fundus photographs, regardless of the presence or absence of glaucomatous VF defects. These characteristic changes in the optic nerve head included glaucomatous optic disc cupping, peripapillary atrophy, neuroretinal rim notching, wedge-shaped defects of RNFL adjacent to the edge of optic disc, and optic disc hemorrhage. Glaucomatous VF defects were defined as those with a cluster of three points with probabilities of < 5% on the pattern deviation map in at least 1 hemifield, including at least 1 point with a probability of < 1%, or a cluster of two points with a probability of < 1%, and/or a glaucoma hemifield test (GHT) result outside normal limits. Other inclusion criteria included (1) age > 40 years old and (2) no concomitant ocular disease, like retinal disease, diabetic mellitus, or significant senile cataract that could affect the results of SD-OCT examinations.

Inclusion criteria for the normal control group were (1) age > 20 years old; (2) best corrected visual acuity of 20/40 or better with spherical equivalent refractive errors between – 6.0 and 3.0 diopters; (3) intraocular pressure of 21 mmHg or less; (4) no history of glaucoma, retinal disease, diabetic mellitus, or significant senile cataract that could affect the results of SD-OCT examinations; (5) normal VF which was defined as a mean deviation (MD) and pattern standard deviation (PSD) within 95% confidence limits and a GHT result within normal limits.

SD-OCT imaging and hand-crafted features of pRNFL

SD-OCT imaging of pRNFL was carried out with Topcon 3D OCT-2000 (Topcon, Tokyo, Japan, software version: 8.11.003.04) by experienced operators (X.L, B.C, and J.Y). The SD-OCT uses a super luminescent diode laser with a center wavelength of 840 nm and a bandwidth of 50 nm as a light source. The acquisition rate of the SD-OCT is up to 20,000 A-scans per second. The transverse and axial resolutions were 20 and 5 μ m, respectively. A 3D scan disc protocol was used for pRNFL imaging in this study. Only images with a quality factor > 45 were used for analyses. RNFL imaging with 1024 points of resolution on a 3.46-mm circle diameter was exported and saved as .jpg format for quantitative analysis.

A customized software (Anterior Segment Analysis Program (ASAP)) was used to quantify pRNFL parameters, including thickness in four different sectors (nasal, temporal, superior and inferior) and on average. ASAP is a plug-in for image processing software (ImageJ version 1.38x; public domain software, <http://imagej.nih.gov/ij/>) [13]. The detail of this software had been reported by us and other research groups [14, 15]. After automatically delineating the boundary of pRNFL, the software then automatically calculated the pRNFL thickness, which include the overall average thickness in 360°, with 315° to 45° position designated temporal, 225° to 315° position inferior, 135° to 225° position nasal, and 45° to 135° position superior (Fig. 1). The inter-observer reproducibility of ASAP was assessed using a random subset of 20 images as evaluated by two examiners (X.L and C.Z) independently. The reproducibility of the ASAP was good to excellent, with an intra-class correlation coefficient that ranged from 0.792 to 0.979 for all the RNFL parameters. We used receiver operating characteristic (ROC) curves to describe the ability to discriminate glaucomatous from healthy eyes for each pRNFL parameter.

Datasets

Two datasets were included in this study. The first dataset, consisted of 1501 peripapillary RNFL OCT images obtained from 153 glaucoma patients with 690 images and 394 normal subjects with 811 images, was selected from a prospective study at JSIEC between September 2013 and August 2014. We had reported some of the results previously [16]. Briefly, the study was designed to investigate the profile and determinant of hand-crafted features in SD-OCT imaging with a semi-automatic software. This dataset was further randomly divided into two sets: a training set (80% with 552 images of glaucoma and 649 images of normal subjects) for the development of the algorithm and a development set (20% with 138 images of glaucoma and 162 images of normal subjects) for monitoring the performance of the trained model.

The other testing dataset, which included 50 images from 50 glaucoma patients and 52 images from 52 normal subjects, was used for evaluation of the algorithm and comparison of the algorithm with RNFL parameters. Subjects in the testing dataset were enrolled prospectively after the training dataset were established from the same center. The inclusion and exclusion criterion were identical for both the training dataset and testing dataset except that age was above 40 years old in normal subjects of testing dataset.

Development of Deep Learning Model

We used transfer learning with fine-tune technique to build the classifier in this study. To achieve this, a modified convolutional base of convolutional neural network (CNN)

model with weights pre-trained on ImageNet was used as DL model [17]. The pre-trained weights of DL model were further fine-tuned on our dataset for classification. The applied CNN was Inception-V3 with 159 layers (Google Inc) [18]. The DL model was implemented in Tensorflow framework (Google, version 1.10.0) with Keras API (version 2.2.4). All images were resized to 224 × 224 pixels as required by Keras API. Image pixel values were scaled to values in a range of 0 through 1. Data augmentation was performed to increase the amount and type of variation within the training dataset, including horizontal flipping, rotation of 20°, and sharpening and adjustments to saturation, brightness, contrast, and color balance. Training was then performed by a minibatch gradient descent of size 32 with an Adam optimizer learning rate of 0.001. Training was run for 100 epochs, as the loss of the model decreased and the accuracy of the validation set increased.

To visualize the most predictive features for detecting glaucomatous pRNFL changes, a class activation map (CAM) was used to open the “black box” of DL model [19]. ACAM is a 2D grid of heatmap associated with a specific output class, computed for every location in any input image, indicating how important each location is with respect to the class considered.

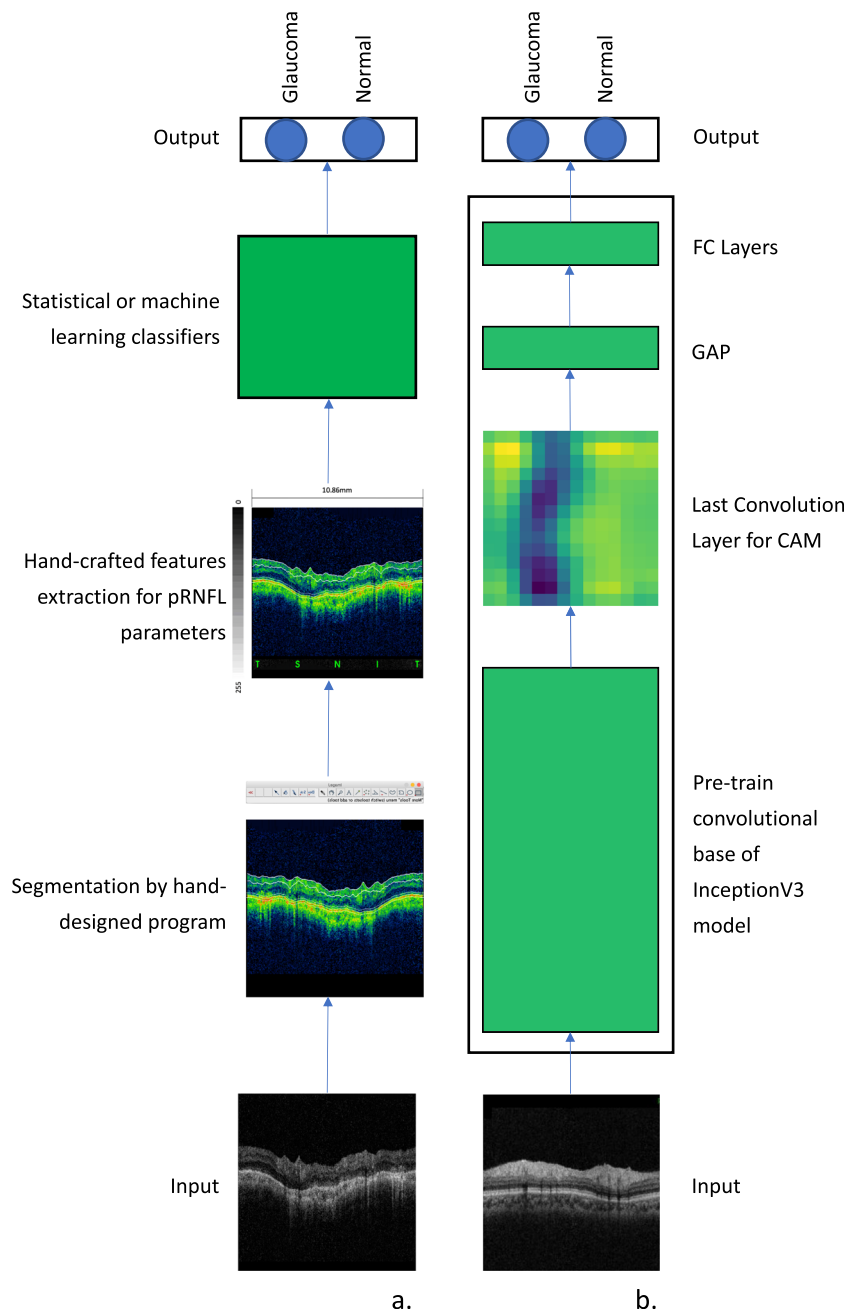
Statistical analysis

The Student *t* test or Mann-Whitney test was used to evaluate demographic and clinical differences between normal subjects and glaucoma patients. The mean and standard deviation of the following parameters were calculated: age, spherical equivalent, axial length, visual field (MD and PSD), image quality score, and pRNFL thickness in overall average and four different quadrants. The diagnostic accuracy of DL model and each pRNFL parameter to differentiate between normal and glaucomatous eyes was determined by computing the area under the curve (AUC) of ROC, sensitivity, specificity, and f1 score with 95% confidence intervals. Sensitivities at fixed specificities of 80% and 90% were determined for DL model and all the pRNFL parameters. Comparison of AUC was carried out using DeLong’s method [20]. All statistical analyses were carried out using commercial statistical software (IBM SPSS Statistics v. 17 for Windows; SPSS Inc. Chicago, IL) and statistical programming language Python (ver. 3.5.1, Python Software Foundation, Beaverton, USA).

Results

Demographics of the training and testing sets are given in Table 1. In testing set, mean age was 53.57 and 50.70 years for normal [range 40–76 years] and glaucoma [range 24–72 years] subjects respectively. In training set, mean age was

Fig. 1 Overview diagram of classic machine learning with hand-crafted features vs. deep learning. **a** A classic machine learning with hand-crafted features (HCFs) involves an object segmentation using hand-designed program, follows by utilizing domain knowledge to manually or semi-automatically extract features, and finally makes diagnosis using statistical classifiers or machine-learning classifiers to make a prediction. **b** A deep learning approach learns the most predictive features by a series of hidden layers which extracts increasingly abstract features from the image directly



41.84 and 49.54 years for normal [range 18–80 years] and glaucoma [range 21–76 years] subjects respectively. Comparison of HCFs, including pRNFL thickness on average and in four different sectors, is shown in Table 2. The average and different sectors of pRNFL thickness were significantly smaller in the glaucoma group compared with the normal group in both training and testing datasets (all with $p < 0.001$) except for nasal sector.

The training curve for the DL model is shown in Fig. 2. We trained each DL model for 100 epochs (iterations through the entire dataset) until the absence of further improvement in both accuracy and loss function. As shown in Fig. 2, the loss

function value and accuracy rate of both training and validation dataset changed dramatically at the first five epochs and then stabilized after that, showing that our DL models reach good convergence.

Using the pRNFL thickness in four different sectors, the AUC obtained in the independent test set varied between 0.661 and 0.938 (the lowest AUC value of 0.661 with 95% CI 0.549 to 0.772 for pRNFL thickness in temporal sector, the highest AUC value of 0.938 with 95% CI 0.894 to 0.982 for pRNFL thickness in inferior sector respectively), as shown in Table 3 and Fig. 3. The DL model achieved an AUC value of 0.990 [95% CI 0.974 to 1.000], which was significantly larger

Table 1 Demographics of the training and testing datasets

| | Testing set | | | Training set | | | | |
|--------------------------|-------------------------|---------------------------|------------|--------------------------|----------------------------|------------|-----------------------|------------------------|
| | Normal (<i>n</i> = 52) | Glaucoma (<i>n</i> = 50) | <i>p</i> * | Normal (<i>n</i> = 394) | Glaucoma (<i>n</i> = 153) | <i>p</i> * | <i>p</i> ⁺ | <i>p</i> ⁺⁺ |
| Age, y (mean ± sd) | 53.57 ± 10.06 | 50.70 ± 15.01 | 0.262 | 41.84 ± 16.55 | 49.54 ± 16.05 | < 0.001 | 0.652 | < 0.001 |
| Gender (male/female) | 20/32 | 42/8 | < 0.001 | 191/203 | 101/54 | < 0.001 | 0.005 | 0.173 |
| Eye (OD/OS) | 27/25 | 26/24 | 0.984 | 196/198 | 81/72 | 0.503 | 0.908 | 0.769 |
| Spherical equivalent (D) | 0.14 ± 0.29 | − 0.90 ± 2.43 | 0.006 | − 1.07 ± 1.96 | − 1.36 ± 3.26 | 0.298 | 0.353 | < 0.001 |
| Axial length (mm) | 23.05 ± 0.84 | 23.88 ± 1.22 | 0.030 | 23.88 ± 1.15 | 24.15 ± 1.60 | 0.150 | 0.315 | < 0.001 |
| Visual field MD(dB) | − 1.04 ± 1.46 | − 12.63 ± 10.84 | < 0.001 | − 1.00 ± 1.32 | − 12.90 ± 10.93 | < 0.001 | 0.879 | 0.818 |
| Visual field PSD(dB) | 1.58 ± 0.47 | 6.31 ± 4.21 | < 0.001 | 1.50 ± 0.44 | 6.54 ± 4.00 | < 0.001 | 0.716 | 0.164 |
| Image quality score | 57.81 ± 4.49 | 53.60 ± 3.83 | < 0.001 | 55.24 ± 4.21 | 51.95 ± 6.19 | < 0.001 | 0.077 | < 0.001 |

*p** comparison between glaucomatous and normative eyes in the training or testing dataset, *p*⁺ comparison between glaucomatous eyes in the training dataset and those in the testing dataset, *p*⁺⁺ comparison between normative eyes in the training dataset and those in the testing database

than the AUC values of all other HCFs (*p* < 0.001, DeLong's method with Holm's method for multiple comparisons). CAMs were generated to visualize the regions on SD-OCT imaging that were most important for the glaucoma prediction. Figure 4 shows two examples of the strongest activations within the DL model, after being assigned the positive for glaucoma. For normal pRNFL imaging, no area is activated. Most of the pRNFL area has been activated by CNN in glaucoma. This shows that the network is focusing on the part of the image where the glaucomatous changes are present.

Discussion

In this study, we had shown that a DL model was effective at distinguishing glaucoma from normal SD-OCT imaging of pRNFL, and its accuracy was higher when compared with pRNFL parameters which were extracted by a hand-crafted

software. Our results demonstrated the possibility of applying DL to assist screening and diagnosis of glaucoma in SD-OCT images.

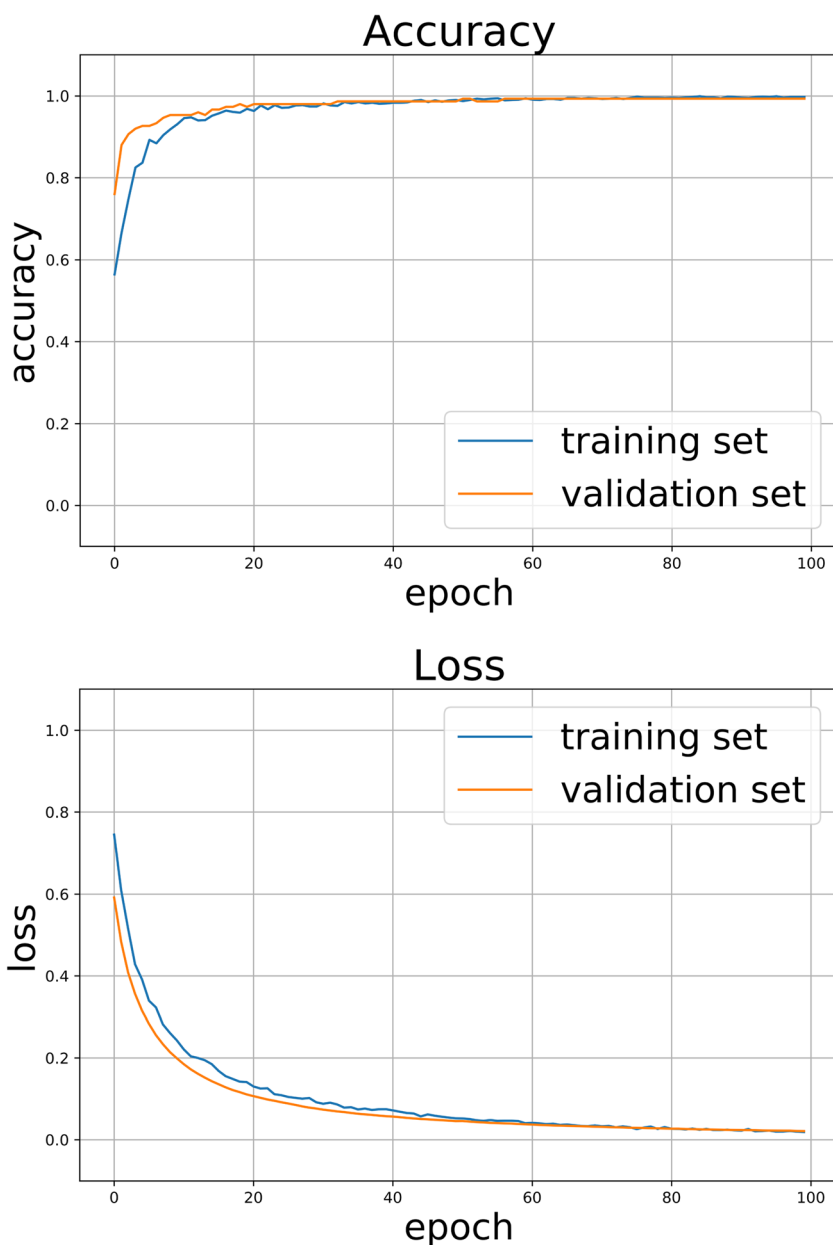
SD-OCT is widely used for glaucoma detection, and peripapillary RNFL analysis represents the scanning protocol most used for routine patient management. Both manual and semi-automatic methods have been reported to diagnose glaucoma based on SD-OCT images. By using build-in software, Harsha et al. reported AUC between 0.792 and 0.884 for pRNFL parameters, which was comparable to the current study [21]. Using RTVue-100 system, Seong et al. report a higher AUC value from 0.921 to 0.976 for pRNFL parameters in patients with different glaucoma stages [22]. Although there are differences in the dataset and reference standards compared with the previous studies, the present study extends this body of work by using DL model with high sensitivity and specificity. Our study had a unique difference when compared with the previous studies. Most previous studies used HCFs,

Table 2 Comparison of pRNFL parameters between training and testing set

| | Training set | | | Testing set | | | | |
|-----------------|-------------------------|---------------------------|------------|--------------------------|----------------------------|------------|-----------------------|------------------------|
| | Normal (<i>n</i> = 52) | Glaucoma (<i>n</i> = 50) | <i>p</i> * | Normal (<i>n</i> = 394) | Glaucoma (<i>n</i> = 153) | <i>p</i> * | <i>p</i> ⁺ | <i>p</i> ⁺⁺ |
| Parameters | Mean (95% CI) | Mean (95% CI) | | Mean (95% CI) | Mean (95% CI) | | | |
| pRNFL thickness | | | | | | | | |
| Average | 108.69 ± 10.09 | 80.74 ± 18.64 | < 0.001 | 106.97 ± 14.44 | 78.10 ± 20.39 | < 0.001 | 0.539 | 0.115 |
| Inferior sector | 139.10 ± 15.71 | 90.76 ± 25.65 | < 0.001 | 133.73 ± 13.82 | 89.87 ± 28.55 | < 0.001 | 0.305 | 0.362 |
| Superior sector | 129.54 ± 13.49 | 88.26 ± 24.01 | < 0.001 | 128.72 ± 13.67 | 90.82 ± 25.44 | < 0.001 | 0.228 | 0.665 |
| Nasal sector | 84.94 ± 14.85 | 73.44 ± 17.02 | 0.389 | 78.87 ± 15.00 | 68.93 ± 15.14 | 0.778 | 0.278 | 0.805 |
| Temporal sector | 80.75 ± 10.76 | 70.14 ± 18.64 | < 0.001 | 86.00 ± 14.43 | 66.35 ± 17.63 | < 0.001 | 0.57 | 0.023 |

*p** comparison between glaucomatous and normative eyes in the training or testing dataset, *p*⁺ comparison between glaucomatous eyes in the training dataset and those in the testing dataset, *p*⁺⁺ comparison between normative eyes in the training dataset and those in the testing database

Fig. 2 The training curve for the deep learning model. The loss function value and accuracy rate of both training and validation dataset changed dramatically at the first 5 epochs and then stabilized after that, showing that our DL models reach good convergence



which are labor intensive and need to be specified by experts. The supervised DL technologies adopted in this study allow training entirely end-to-end. This would help

to aid in the implementation of large-scale screening programs by providing quick and objective SD-OCT image assessment.

Table 3 The diagnostic performance of DL model as compared with hand-crafted features of pRNFL imaging

| | Accuracy (95% CI) | <i>p</i> | Sensitivity at 80% specificity (%) | Sensitivity at 90% specificity (%) |
|----------------------|------------------------|----------|------------------------------------|------------------------------------|
| Deep learning models | 0.990 (0.974 to 1.000) | | 0.981 | 0.981 |
| pRNFL thickness | | | | |
| Average | 0.895 (0.832 to 0.957) | < 0.001 | 0.865 | 0.577 |
| Inferior sector | 0.938 (0.894 to 0.982) | < 0.001 | 0.971 | 0.730 |
| Superior sector | 0.913 (0.855 to 0.970) | < 0.001 | 0.885 | 0.620 |
| Nasal sector | 0.696 (0.594 to 0.799) | < 0.001 | 0.346 | 0.250 |
| Temporal sector | 0.661 (0.549 to 0.772) | < 0.001 | 0.135 | 0.077 |

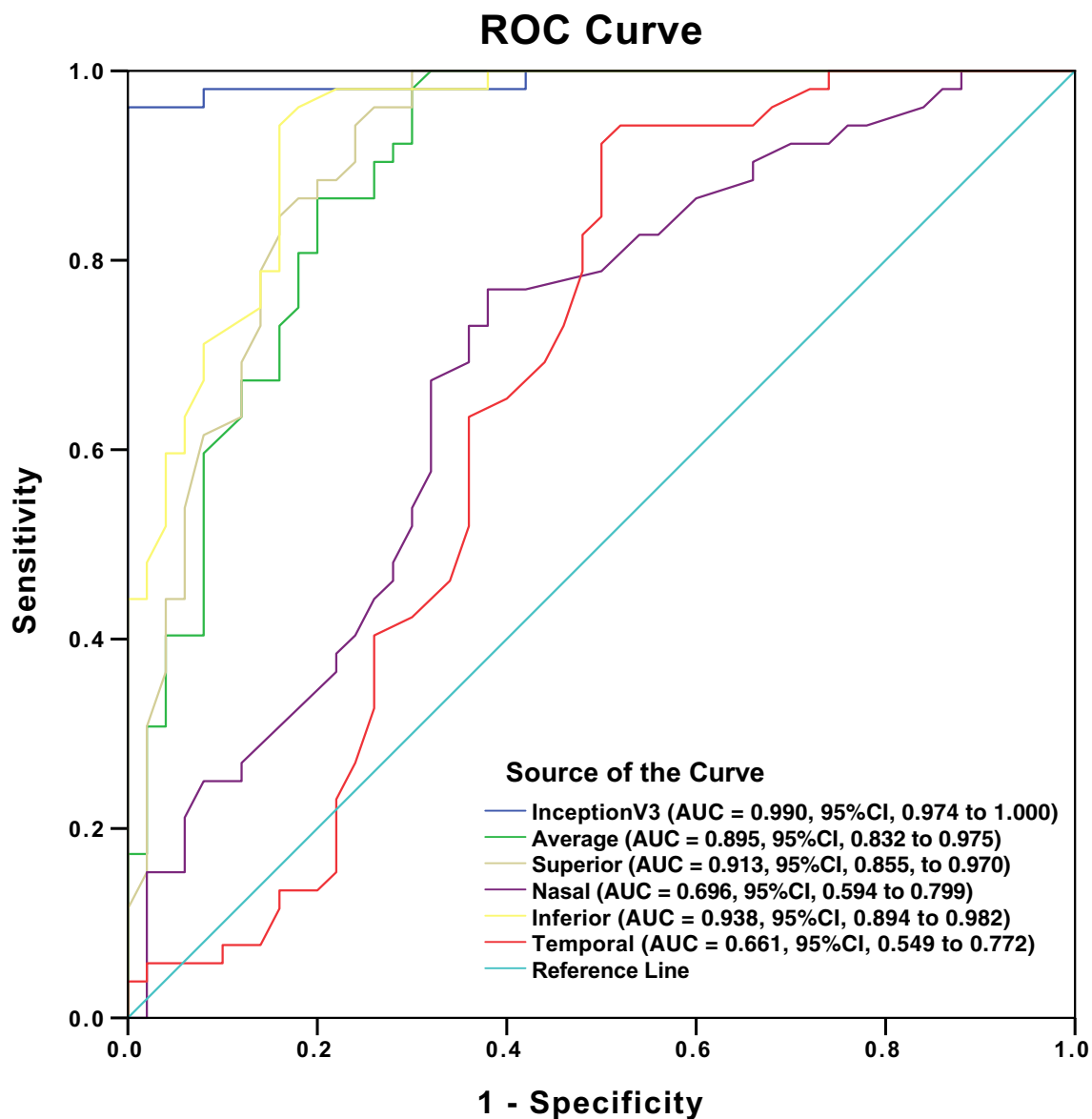


Fig. 3 Receiver operating character (ROC) curves in the testing set

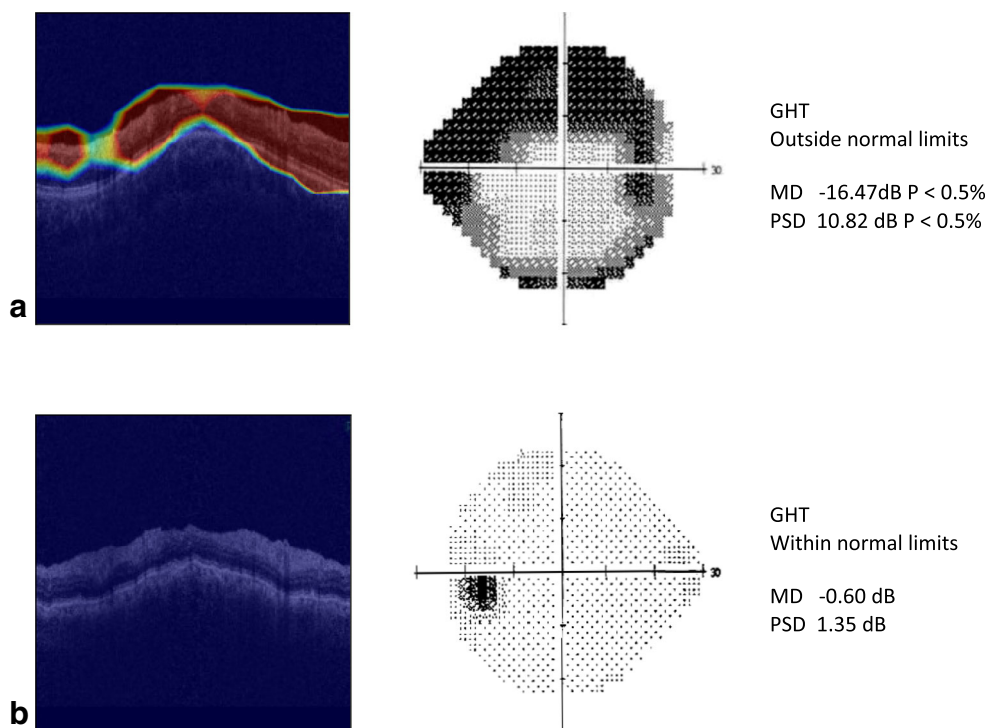
The application of DL is usually limited due to the lack of large training sets, like several tens of thousands of examples to train DL model effectively. Given a finite amount of OCT images can be prepared in the clinical setting, several authors suggested to use transfer learning to improve the performance of DL model. Transfer learning has proven to be a highly effective technique and been widely used in multiple medical domains with limited data [7, 23]. Rather than training a completely blank network, we adapted DL model base pre-trained on the ImageNet dataset which involved more than 14 million images with 1000 classes. After running pre-trained CNN base over our dataset, we then extended the models by adding dense layers on top and running the whole thing end-to-end on the input data. These techniques yield

a more accurate model in much less time with relatively small dataset.

Deep learning models have often been functionally so-called black boxes because it is difficult to determine how DL makes its predictions [24]. CAMs are tools that can help aid visualization of a CNN. It is interesting to notice that features outside RNFL region also contribute to DL model predictions. Previous studies had reported several other glaucomatous structure defects, such as the ganglion cell layer (GCL) and peripapillary microvascular changes [25, 26]. It is possible that CAMs will prove to be even more useful to reveal the structure–function relation in glaucoma.

Our study has several limitations. First, our DL models are developed based on Topcon model. It had been reported that pRNFL thickness values obtained by

Fig. 4 Classic activation map (CAM) for glaucomatous and normal peripapillary retinal nerve fiber layer images. **a** A glaucomatous pRNFL defect with severe visual field defect (MD = -16.47). Notice the highlight regions in the pRNFL area which is activated by CNN. This shows that the network is focusing on the part of the image where the glaucomatous changes are present. **b** A normal pRNFL image without any activation area



Topcon OCT are significantly higher than those measured by Cirrus OCT in both normal and glaucomatous eyes [27]. Second, the macular data of SD-OCT was not available in the training set in current study. Some systematic review has reported pRNFL parameters are still preferable to macular parameters for diagnosing manifest glaucoma, although the differences are small. Third, previous study had reported the effects of peripapillary atrophy on the diagnostic ability of SD-OCT parameters. Further studies are needed to explore the effect of peripapillary atrophy on DL model developed in this study [28]. Fourth, all SD-OCT images were collected from the same center and external validation test is needed to confirm the diagnostic performance in different centers or OCT devices in the future. Also, most of the glaucoma cases were quite severe and this made classification easier in the current study. Lastly, we used images of Chinese eyes only, so our results may not be applicable to other populations.

In spite of these limitations, DL model developed in this study has achieved higher sensitivity and specificity compared to traditional HCFs. DL has a high diagnostic capability and may help screening and diagnosis of glaucoma in clinic. Further studies with larger dataset are needed to evaluate whether DL model also predict other clinical outcomes, including pRNFL imaging from different eye center, OCT devices, different size or appearance of optic disc, or different glaucoma stages.

Author's contributions CZ, MZZ, and ZF contributed towards the conception and design, drafted the manuscript, and approved the final version. CZ, JWL, and LTH developed the artificial intelligence system evaluated in this study. XLX, JLY, TQ and BYC gathered, cleaned, and organized the data.

Funding information This study was funded by the National Natural Science Foundation of China (81371010) and Clinical Research Funds of Shantou University Medical College (2014).

Compliance with ethical standards

This study was conducted according to the tenets of the Declaration of Helsinki and had the approval of the institutional review board.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments.

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

References

1. Quigley HA (2011) Glaucoma. *Lancet* 377:1367–1377
2. Weinreb RN, Aung T, Medeiros FA (2014) The pathophysiology and treatment of glaucoma: a review. *JAMA* 311:1901–1911

3. Sung KR, Kim JS, Wollstein G et al (2011) Imaging of the retinal nerve fibre layer with spectral domain optical coherence tomography for glaucoma diagnosis. *Br J Ophthalmol* 95:909–914. <https://doi.org/10.1136/bjo.2010.186924>
4. Medeiros FA, Zangwill LM, Bowd C et al (2005) Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. *Am J Ophthalmol* 139:44–55
5. Leung CK, Cheung CY, Weinreb RN et al (2009) Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology* 116:1257–1263
6. Knight OJ, Chang RT, Feuer WJ et al (2009) Comparison of retinal nerve fiber layer measurements using time domain and spectral domain optical coherent tomography. *Ophthalmology* 116:1271–1277
7. Asaoka R, Murata H, Hirasawa K et al (2018) Using deep learning and transform learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol*. <https://doi.org/10.1016/j.ajo.2018.10.007> (1879–1891 (Electronic))
8. Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316(22):2402
9. Shen W, Zhou M, Yang F et al (2015) Multi-scale convolutional neural networks for lung nodule classification. *Info Process Med Imaging* 24:588–599
10. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284(2):574–582. <https://doi.org/10.1148/radiol.2017162326>
11. Le MH, Chen J, Wang L et al (2017) Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks. *Phys Med Biol* 62(16):6491–6514. <https://doi.org/10.1088/1361-6560/aa7731>
12. Asaoka R, Murata H, Iwase A et al (2016) Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology* 123(9):1974–1980. <https://doi.org/10.1016/j.ophtha.2016.05.029>
13. Rasband WS. ImageJ, US. National Institutes of Health, Bethesda, Maryland, USA. <http://imagej.nih.gov/ij/>. Accessed 2 Oct 2019
14. Shabana N, Aquino MC, See J et al (2012) Quantitative evaluation of anterior chamber parameters using anterior segment optical coherence tomography in primary angle closure mechanisms. *Clin Exp Ophthalmol* 40(8):792–801. <https://doi.org/10.1111/j.1442-9071.2012.02805.x>
15. Zheng C, de Leon JM, Cheung CY et al (2016) Determinants of pupil diameters and pupil dynamics in an adult Chinese population. *Graefes Arch Clin Exp Ophthalmol* 254(5):929–936. <https://doi.org/10.1007/s00417-016-3272-7>
16. Chen B, Gao E, Chen H et al (2016) Profile and determinants of retinal optical intensity in normal eyes with spectral domain optical coherence tomography. *PLoS One* 11(2):1–16. <https://doi.org/10.1371/journal.pone.0148183>
17. Russakovsky O, Deng J, Su H et al (2015) ImageNet large scale visual recognition challenge. *Int J Comput Vis* 115.3(2015):211–252. <https://doi.org/10.1007/s11263-015-0816-y>
18. Szegedy C, Vanhoucke V, Ioffe S, et al. (2016) Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
19. Zhou B, Khosla A, Lapedriza A, et al. (2015) Learning deep features for discriminative localization. CVPR'16 (arXiv:1512.04150, 2015)
20. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3):837–845
21. Rao HL, Zangwill LM, Weinreb RN et al (2010) Comparison of different spectral domain optical coherence tomography scanning areas for glaucoma diagnosis. *Ophthalmology* 117(9):1692–1699.e1. <https://doi.org/10.1016/j.ophtha.2010.01.031>
22. Seong M, Sung KR, Choi EH et al (2010) Macular and peripapillary retinal nerve fiber layer measurements by spectral domain optical coherence tomography in normal-tension glaucoma. *Investig Ophthalmol Vis Sci* 51(3):1446–1452. <https://doi.org/10.1167/iovs.09-4258>
23. Wang X, Peng Y, Lu L, et al. (2017) ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3462–3471. <https://doi.org/10.1109/CVPR.2017.369>
24. Yosinski J, Clune J, Nguyen A, et al. (2015) Understanding neural networks through deep visualization. arXiv preprint. <https://arxiv.org/abs/1506.06579>. Published June 22, 2015. Accessed August 12, 2019
25. Vermeer KA, van der Schoot J, Lemij HG et al (2012) RPE-normalized RNFL attenuation coefficient maps derived from volumetric OCT imaging for glaucoma assessment. *Investig Ophthalmol Vis Sci* 53(10):6102–6108. <https://doi.org/10.1167/iovs.12-9933>
26. Xu H, Zhai R, Zong Y et al (2018) Comparison of retinal microvascular changes in eyes with high-tension glaucoma or normal-tension glaucoma: a quantitative optical coherence tomography angiographic study. *Graefes Arch Clin Exp Ophthalmol* 256(6):1179–1186
27. Rolle T, Briamonte C, Curto D et al (2011) Ganglion cell complex and retinal nerve fiber layer measured by fourier-domain optical coherence tomography for early detection of structural damage in patients with preperimetric glaucoma. *Clin Ophthalmol* 5:961–969. <https://doi.org/10.2147/OPHTH.S20249>
28. Kim S Y , Park H Y L , Park C K . (2012) The effects of peripapillary atrophy on the diagnostic ability of stratus and cirrus oct in the analysis of optic nerve head parameters and disc size [J]. *Invest Ophthalmol Vis Sci* 53(8)

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