AMERICAN ACADEMY OF OPHTHALMOLOGY®

Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs

Zhixi Li, MD,[1,] Yifan He, BS,[2,*] Stuart Keel, PhD,[3,*] Wei Meng, BS,[2] Robert T. Chang, PhD,[4] Mingguang He, MD, PhD[1,3]*

Purpose: To assess the performance of a deep learning algorithm for detecting referable glaucomatous optic neuropathy (GON) based on color fundus photographs.

Design: A deep learning system for the classification of GON was developed for automated classification of GON on color fundus photographs.

Participants: We retrospectively included 48 116 fundus photographs for the development and validation of a deep learning algorithm.

Methods: This study recruited 21 trained ophthalmologists to classify the photographs. Referable GON was defined as vertical cup-to-disc ratio of 0.7 or more and other typical changes of GON. The reference standard was made until 3 graders achieved agreement. A separate validation dataset of 8000 fully gradable fundus photographs was used to assess the performance of this algorithm.

Main Outcome Measures: The area under receiver operator characteristic curve (AUC) with sensitivity and specificity was applied to evaluate the efficacy of the deep learning algorithm detecting referable GON.

Results: In the validation dataset, this deep learning system achieved an AUC of 0.986 with sensitivity of 95.6% and specificity of 92.0%. The most common reasons for false-negative grading (n = 87) were GON with coexisting eye conditions (n = 44 [50.6%]), including pathologic or high myopia (n = 37 [42.6%]), diabetic retinopathy (n = 4 [4.6%]), and age-related macular degeneration (n = 3 [3.4%]). The leading reason for false-positive results (n = 480) was having other eye conditions (n = 458 [95.4%]), mainly including physiologic cupping (n = 267 [55.6%]). Misclassification as false-positive results amidst a normal-appearing fundus occurred in only 22 eyes (4.6%).

Conclusions: A deep learning system can detect referable GON with high sensitivity and specificity. Coexistence of high or pathologic myopia is the most common cause resulting in false-negative results. Physiologic cupping and pathologic myopia were the most common reasons for false-positive results. *Ophthalmology 2018;125:1199-1206 © 2018 by the American Academy of Ophthalmology*

Glaucoma is a leading cause of irreversible blindness worldwide.[1–3] A recent global meta-analysis of 50 population-based studies reported the pooled glaucoma prevalence (age range, 40–80 years) to be 3.5%,[3] corresponding to an estimated 64.3 million individuals worldwide. As a result of population growth and ageing, this figure is expected to increase to 112 million by 2040.[3] Most vision loss resulting from glaucoma is avoidable through early detection and treatment strategies.[4–6] Despite this, approximately 85% of cases among the Singapore Chinese, the same rate for African American population of United States, and even an overall rate of 50% among the cases in the United States are undiagnosed.[7–12] High rates of undiagnosed disease can be attributed to chronic glaucoma often being asymptomatic until central visual acuity is affected in the advanced stages of disease. As glaucoma advances from the early to late stage, care costs increase by 4-fold, posing a significant financial burden wordwide.[13]

The assessment of the optic disc and retinal nerve fiber layer (RNFL) are the foundation of glaucoma diagnosis, although a dilated clinical fundus examination after mydriasis has been recommended for its advantage of offering a stereoscopic view of the optic disc.[14] However, monoscopic optic disc photographs offer some key advantages, including convenience and affordability. Furthermore, the Glaucomatous Optic Neuropathy Evaluation project demonstrated that subjective assessments of monoscopic optic disc photographs provide an equal diagnostic accuracy for glaucoma when compared with stereoscopic photographs.[15] Nevertheless, manual assessment of the optic disc is labor intensive and highly dependent on image interpretation by trained specialists. This significantly impacts the cost effectiveness of glaucoma screening,[16–18] and as a result, glaucoma screening strategies are not widely implemented in the general population.[18–20]

https://doi.org/10.1016/j.ophtha.2018.01.023
ISSN 0161-6420/18

Given the growing public health concern, improvements in screening methods for glaucoma are warranted. An emerging area of diagnostic imaging in ophthalmology involves the use of automated computer vision image interpretation with deep learning algorithms, which involves training the algorithm on large datasets of labelled images so that it can learn the features from the data itself rather than from predefined rules. Recent evidence suggests that deep learning algorithms can grade images with excellent diagnostic accuracy in identifying conditions such as diabetic retinopathy.[21,22] If this technology could be adopted to provide accurate assessment of glaucoma, there are significant potential benefits including an increased accessibility and affordability of glaucoma screening for specific and at-risk populations, thus improving access to care and decreasing the cost of glaucoma screening especially in remote and underserved communities. The purpose of this study was to evaluate the efficacy of a newly developed deep learning algorithm for the detection of referable glaucomatous optic neuropathy (GON) from monoscopic color fundus photographs.

Methods

In the current study, 70 000 fundus photographs were downloaded by random sampling from the online dataset LabelMe (Healgoo Ltd. *LabelMe dataset*; 2016. http://www.labelme.org. Accessed February 16, 2016.), which contains more than 200 000 color fundus photographs collected from various clinical settings in China. Subsequently, 48 116 images with a visible optic disc were selected for the labelling of GON.

Approval from the institutional review board of Zhongshan Ophthalmic Center was obtained (identifier, 2017KYPJ049), and this project was conducted according to the tenets of the Declaration of Helsinki. Because of the retrospective nature and fully anonymized usage of images in this study, the review board indicated that the informed consent was not required.

Fundus Photograph Grading, Quality Control, and Reference Standard

Figure 1A illustrates the process of image grading. Twenty-seven licensed ophthalmologists were invited to grade the images. During the training of ophthalmologists, 4 sets each with 100 images (30 had suspect GON or worse) were used for the test. The results of graders were compared with those of 3 senior glaucoma specialists (M.H., R.C., and X.L.), and participants passed the training until they achieved an unweighted κ value of 0.75 or more in any test set. Thus, 21 ophthalmologists qualified as graders to classify these images in an online grading system from September 2016 through March 2017.

Table 1 shows the details of the criteria of GON grading, which was classified into unlikely, suspect, and certain GON based on the criteria of previous population studies.[23–26] The field of the retinal fundus photograph and image quality also were included in the grading. Poor location was assigned to a photograph when the optic disc was not fully visible. Poor quality was used when vessels

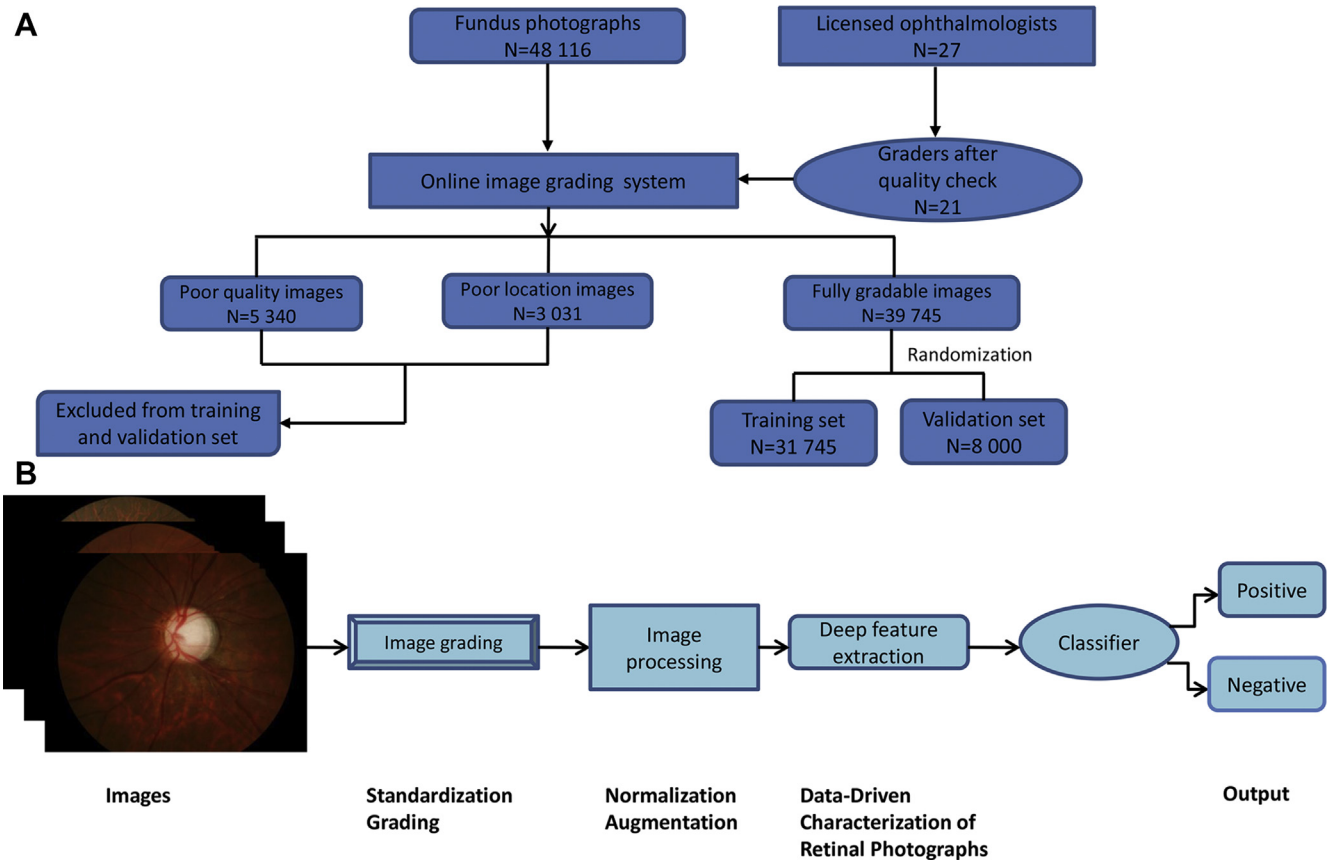

Figure 1. Diagrams showing an overview of the proposed automated detection for referable glaucomatous optic neuropathy using deep learning. **A**, Image grading workflow for our algorithm. **B**, Integration flow in the current study.

Table 1. The Classification for Glaucomatous Optic Neuropathy in the Online System

Classification	Presence of Clinical Features
Unlikely	Does not meet any of the following criteria
Suspect	Any criterion of the following: $0.7 \leq$ VCDR < 0.9 0.05 DD < rim width $\leq$ 0.1 DD RNFL defect Disc hemorrhage
Certain	Any criterion of the following: VCDR $\geq$0.9 Rim width $\leq$0.05 DD or localized notches RNFL defect corresponds to narrowing of rim or localized notches
Poor quality	Any criterion of the following: Vessels within 1 DD of the optic disc margin cannot be identified $\geq$50% of the area is obscured
Poor location	Only part of optic disc is visible in the image

DD = disc diameter; RNFL = retinal nerve fiber layer; VCDR = vertical cup-to-disc ratio.

within 1 disc diameter of the optic disc margin could not be identified or 50% or more of the area was obscured. Retinal photographs of poor quality and poor location were regarded as ungradable for referable GON and were excluded from the training and validation dataset. Referable GON was defined as suspect or certain GON.

In the process of online grading, each image was assigned randomly to a single ophthalmologist for initial grading and was assigned sequentially to other individual graders until 3 consistent grading outcomes were achieved. This specific grading outcome was considered as the conclusive grading for a given image.

Image Processing and Development of the Deep Learning Algorithm

Figure 1B displays the deep learning pipeline. The training dataset was used to train the deep learning network, and the validation dataset was used for evaluation of the algorithm. Because a large number and variation of images were collected, we performed several preprocessing steps to normalize the images for variation in our database. Image pixel values were scaled to values in a range of 0 through 1 and then downsized to a 299 × 299 matrix. Local space average color[27] was subtracted to solve the issue of color constancy, which indicates a human observer could see the color of an object with consistency, even if the light reflected by an object changed with the illumination type. Data augmentation, a method of image transformations within a dataset to enlarge image heterogeneity, but to retain prognostic features in the image itself, was performed to enhance the dataset, including random horizontal shifts of 0 to 3 pixels and random rotations of 90°, 180°, and 270°. We adopted Inception-v3 architecture,[28] which is a convolutional neural network composed of 11 inception modules. A diagram of the deep learning architecture is shown in Figure 2. A minibatch gradient descent of size 32 was used for training, with an Adam optimizer learning rate of 0.002 for better convergence.[29] The exact model version of this study was 20170811. Using the ground truth graded by consensus from the ophthalmologists, we adopted 2 levels detection of referable GON for this deep learning system: no referable GON and referable GON (comprising suspected and certain GON).

Characteristics of Misclassification by the Deep Learning Algorithm

To characterize misclassified photographs, an experienced ophthalmologist (Z.L.) reviewed them and categorized false-negative images according to the 3 most commonly seen features arbitrarily developed by a consensus meeting: (1) eyes with coexisting eye conditions, such as pathologic or high myopia, diabetic retinopathy, and age-related macular degeneration; (2) eyes without other eye conditions; and (3) others (only presence of RNFL defect alone or optic disc hemorrhage). Similarly, the false-positive photographs were classified into the following arbitrary features: (1) eyes with other eye conditions, such as physiologic large cupping, pathologic or high myopia, nonglaucomatous optic atrophy, congenital optic disc abnormalities, and other retinal diseases; and (2) eyes with a normal fundus.

Statistical Analyses

The accuracy of each grader was calculated as a proportion of the grading results that matched the conclusive grading outcome over the total number of images graded by that individual. We used the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals to evaluate the performance of this algorithm. Accuracy, sensitivity, and specificity of the system also were evaluated. All statistical analyses were performed using Stata software version 14.0 (Stata Corp., College Station, TX).

Results

A total of 48 116 fundus photographs were graded for glaucoma by program-trained ophthalmologists, with each fundus

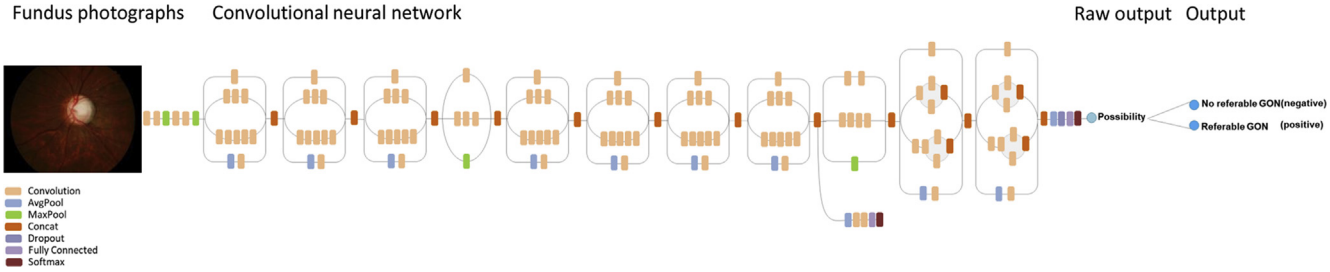

Figure 2. Diagram showing a deep convolutional neural network. This deep learning network contains 22 layers, of which had 11 inception modules. GON = glaucomatous optic neuropathy.

Table 2. Proportion of Glaucomatous Optic Neuropathy in the Training and Validation Dataset

	Training Set	Validation Set
Unlikely	23 433 (73.8)	6033 (75.4)
Suspect	2190 (6.9)	430 (5.4)
Certain	6122 (19.3)	1537 (19.2)
Total	31 745 (100)	8000 (100)

Data are number (%) unless otherwise indicated.

photograph graded between 3 and 9 times. The median quantity of images per ophthalmologist graded was 3337 (range, 407−34 621). Twelve graders graded more than 3000 images. The median accuracy of the ophthalmologist graders (n = 15) for referable glaucoma was 82.1% (range, 76.0%−86.5%). A total of 5340 images were graded as poor quality and 3031 were labeled as poor location, and these images were excluded from the training and validation dataset. Finally, among 39 745 fully gradable fundus photographs, 8000 images were selected randomly using a simple random sampling and treated as the validation set, and the remaining 31 745 images were used as the training set. The proportion of labels in the training and validation datasets is described in Table 2. The AUC of the deep learning algorithm was 0.986 (95% confidence interval, 0.984−0.988) for referable GON. An accuracy of 92.9% with a sensitivity of 95.6% and a specificity of 92.0% was obtained for the classification of referable GON (Fig 3).

The reasons for false-negative findings are provided in Table 3. The most common reason for false-negative classification (n = 87) was undetected GON with other eye conditions (n = 44 [50.6%]), mainly including pathologic or high myopia (n = 37 [42.6%]; Fig 4A). Table 4 shows the clinical features of the false-positive photographs of this deep learning algorithm. Among 480 false-positive images, most of these images (n = 458 [95.4%]) showed an abnormal fundus. A misclassification of physiologic large cupping was the most frequent reason (n = 267 [55.6%]) that these images showed borderline vertical cup-to-disc ratio with healthy RNFL (Fig 5A). A normal fundus was found in 22 images (4.6%; Fig 5F).

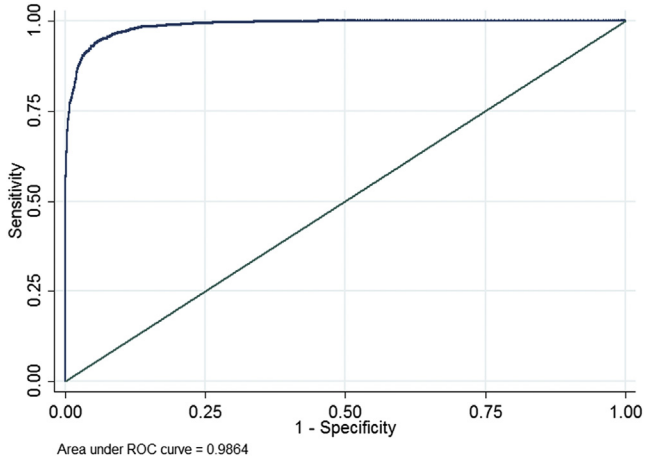

Figure 3. Graph showing an area under receiver operating characteristic (ROC) curve of 0.986 (95% confidence interval, 0.984−0.988) derived from the validation dataset that was obtained using the deep learning algorithm.

Discussion

In this study, the efficacy of the deep learning method in identifying referable GON based on 48 116 fundus photographs was investigated. This deep learning algorithm showed a robust performance (AUC, 0.986; sensitivity, 95.6%; and specificity, 92.0%) for the detection of referable GON. Recently, several reports of automated methods for the evaluation of glaucoma have been published.[30−34] Singh et al[30] developed a technique using wavelet feature extraction techniques from segmented optic discs on 44 fundus photographs and validated it in 19 images, achieving an accuracy of 94.7%. Issac et al[32] described an accuracy rate of 94.1% with 100.0% sensitivity by an adaptive threshold-based image processing method on 67 fundus photographs. On a total of 2252 fundus images, Chakrabarty et al[31] reported an AUC of 0.792 with a sensitivity of 71.6% and a specificity of 71.7% using a feature extraction technique. Xiangyu et al[33] and Annan et al[34] developed a deep learning algorithm with combined feature extraction and reported AUCs of 0.887 and 0.838, respectively, based on a small sample of fundus images (n < 1250).

Our study had a number of unique differences when compared with the previous studies. First, most previous studies[30,31,33,34] used technology on feature extraction, which inevitably would introduce errors in the localization and segmentation that would lead to misalignment and misclassification.[34,35] The supervised deep learning technologies adopted in this study avoid such problems through learning predictive features directly from the global labeled images. Second, our training and validation sets were substantially larger than those of previous reports (all the above previous datasets are smaller than 2500). An insufficient sample size for the training and validation datasets compromises the accuracy to detect glaucoma on a large scale. Finally, the large dataset in this study consisted of data collected from a variety of clinical settings, and therefore is more representative of the real world.

Overfitting, whereby the algorithm performs well on the training dataset but generalizes inappropriately to unseen cases, may occur in machine learning, including deep learning. It is more common when the training sample size is small. The network in this study was built with a dropout technique in the third to last layer. Dropout is a method to drop units randomly with their connections during the training process, and it has been proven to be an effective tool to reduce overfitting.[36] Overall, the system in this study used a large dataset, dropout, and model regularization method to help reduce the possibility of this issue.[36,37] The evaluation of the training curve could be used to assess the overfitting possibility. The training curve (Fig 6) may suggest that this model is reasonably well fitted. If overfitting occurs, the loss of validation data increases, while training loss steadily decreases. In contrast, best-fitted curves indicate that

Table 3. The Proportion of Reasons for False-Negative Classification by the Deep Learning Algorithm

Reason	No.	Proportion (%)
With coexisting eye conditions	44	50.6
Pathologic or high myopia	37	42.6
Diabetic retinopathy	4	4.6
AMD	3	3.4
Without other eye conditions	33	37.9
Others (RNFL defect alone or optic disc hemorrhage only present)	10	11.5
Total	87	100

AMD = age-related macular degeneration; RNFL = retinal nerve fiber layer.

the validation loss has its global minimum. Although our results were insightful, certainly a further study also is expected to validate this algorithm using a larger-scale dataset.

The deep learning system strives to differentiate the images into those with or without referable GON. When looking into the reasons for false-negative classification, approximately half of all false-negative cases were a result of confounding optic disc features secondary to high myopia or pathologic myopia (42.6%). These optic discs usually are characterized by shallow cups, optic disc tilting, torsion, and peripapillary atrophy, which are also difficult for ophthalmologists or human graders to identify glaucoma even when visual field data is available.[38] The vast majority of remaining false-negative cases (37.9%) were a result of true misclassification, where a GON image without coexisting eye disease or secondary disc change was misclassified as normal. An ideal screening program should minimize the number of false-negative results. We expect more studies to explore how this happened and to identify strategies to minimize errors.

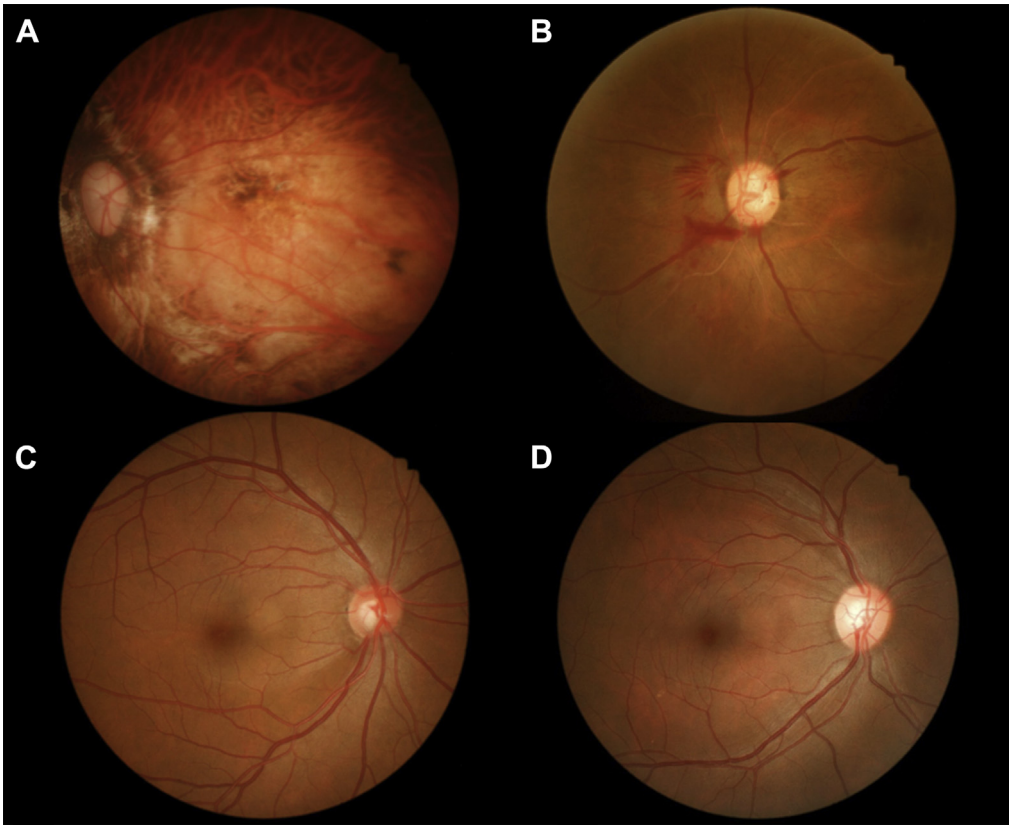

Figure 4. Fundus photographs showing typical false-negative cases. **A**, Glaucomatous optic neuropathy in pathologic myopia. **B**, Glaucomatous optic neuropathy accompanied with diabetic retinopathy. **C**, Glaucomatous optic neuropathy without other eye conditions. **D**, Retinal nerve fiber layer defect alone.

Table 4. The Proportion of Reasons for False-Positive Results by the Deep Learning Algorithm

Reasons	No.	Proportion (%)
With other eye conditions	458	95.4
Physiologic large cupping	267	55.6
Pathologic or high myopia	90	18.8
Nonglaucomatous optic atrophy	58	12.1
Congenital optic disc abnormalities	3	0.6
Others retinal diseases	40	8.3
Normal fundus	22	4.6
Total	480	100

In any screening program, false-positive results create unnecessary referral and burden to the healthcare system. In the present study, nearly all the false-positive cases (95.4%) were in fact a result of abnormalities of the optic disc or retina not related to GON. More than half of the eyes with false-positive results showed physiologic large cupping that also required further investigation by eye care professionals. Therefore, the increased workload resulting from false-positive classification seems to be reasonable given that most such eyes would benefit from further clinical investigation.

The key to glaucoma diagnosis is to recognize morphologic changes of optic disc and RNFL as a result of progressive retinal ganglion cell loss.[39] However, functional abnormalities are presented first in some patients. In most cases, a single isolated observation cannot always confirm the diagnosis of glaucoma even in clinical practice. Therefore, in the future, to improve the accuracy of the deep learning algorithm further, one should ensure that the clinical labeling (classification) is as close as possible to the so-called ground truth. To improve the accuracy of glaucoma classification, more real-world clinical data, such as visual field or longitudinal changes, or even a history of glaucoma surgery, should be included in the clinical labeling of the images.

The strengths of this study include recruiting more than 20 validated ophthalmologists, multiple interpretations of each image, a large image dataset, and a high accuracy of classification. Limitations of this study also should be considered. First, although we used a large set of fundus images, these images were collected only from hospitals in China. Further research is warranted to investigate the performance of the algorithm among other ethnic groups in different settings. In addition, as is the nature of supervised deep learning, the classification of glaucoma was from the global image labeling, rather than direct definitions of local features. It is unclear what exact features were seen and learned by the network. Further work on the visualization of deep learning networks may help patients, physicians, and health care providers understand better the process of its learning.[40] Overall, the current study demonstrated that deep learning can be applied to create an algorithm that is capable of identifying referable GON with high sensitivity and specificity in a large dataset. Further studies are required to explore the usefulness of this algorithm deployed in different population settings and for different ophthalmic conditions.

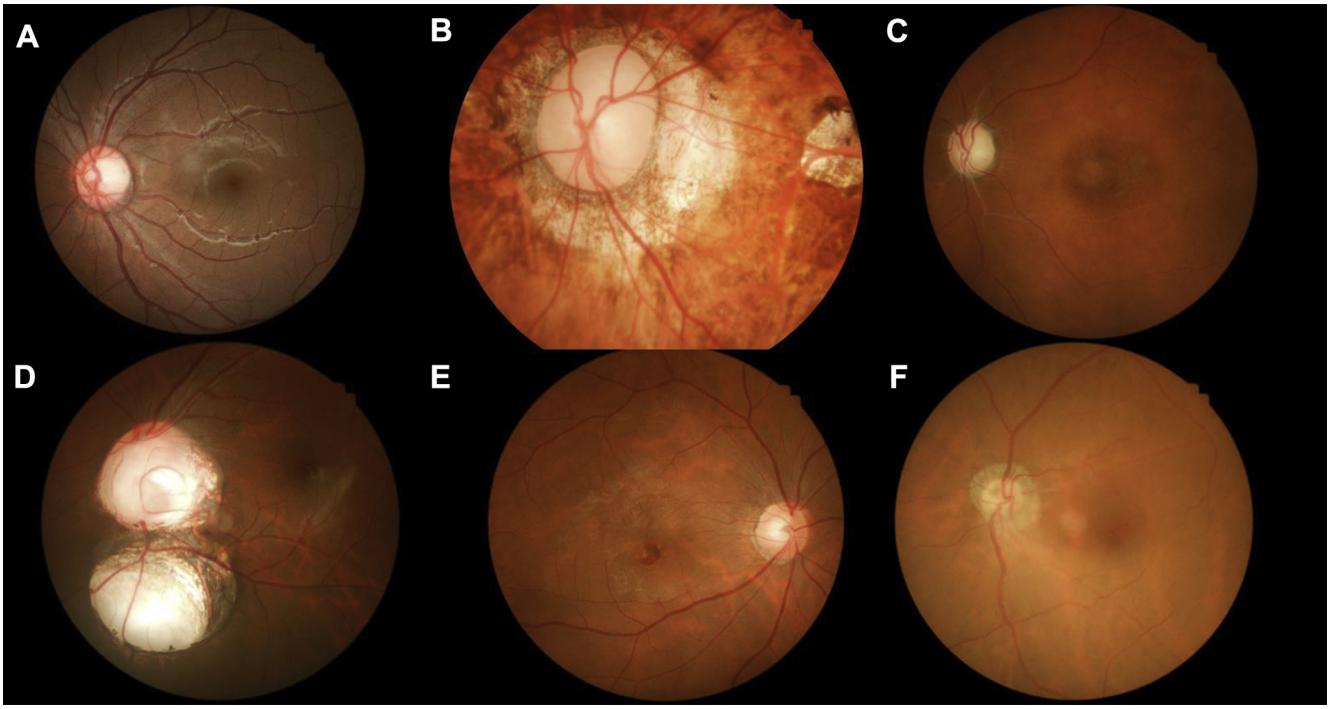

Figure 5. Fundus photographs showing typical false-positive cases. **A**, Physiologic large cupping. **B**, Pathologic myopia. **C**, Nonglaucomatous optic atrophy. **D**, Congenital optic disc and sclera coloboma. **E**, Macular hole (other retinal conditions). **F**, Normal fundus.

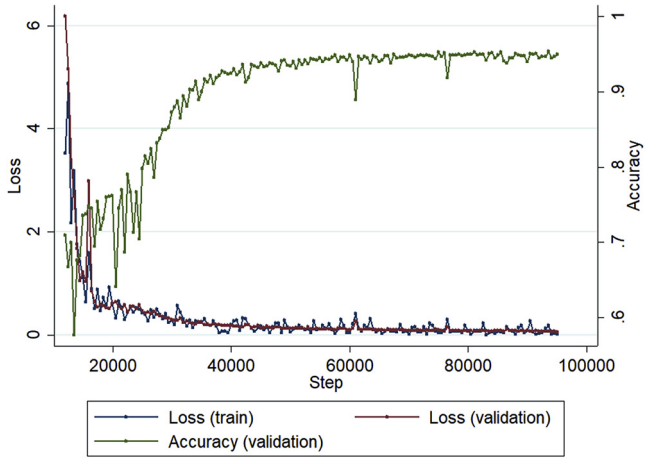

Figure 6. Training curve for the deep leaning algorithm. The green line shows that the accuracy over the training course increased over time in the validation set. The blue and red line show that the loss of training and validation dataset decreased over time. The loss means the fit between a prediction of the algorithm and the given manual label. When the loss decreased over the training course with accuracy increases and the loss of validation increased, but that of training remained stable, appreciable overfitting is indicated.

Acknowledgments

The authors thank Zehong Zhou, International Department, Affiliated High School of South China Normal University, for his contributions to the data cleaning, integration, and convolutional neural network construction.

References

1. Stevens GA, White RA, Flaxman SR, et al. Global prevalence of vision impairment and blindness: magnitude and temporal trends, 1990−2010. *Ophthalmology*. 2013;120:2377−2384.
2. Bourne RR, Stevens GA, White RA, et al. Causes of vision loss worldwide, 1990−2010: a systematic analysis. *Lancet Glob Health*. 2013;1:e339−e349.
3. Tham YC, 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;121:2081−2090.
4. Hood DC, Raza AS, de Moraes CG, et al. Glaucomatous damage of the macula. *Prog Retin Eye Res*. 2013;32:1−21.
5. Tatham AJ, Weinreb RN, Medeiros FA. Strategies for improving early detection of glaucoma: the combined structure-function index. *Clin Ophthalmol*. 2014;8:611−621.
6. Tatham AJ, Medeiros FA, Zangwill LM, Weinreb RN. Strategies to improve early diagnosis in glaucoma. *Prog Brain Res*. 2015;221:103−133.
7. Baskaran M, Foo RC, Cheng CY, et al. The prevalence and types of glaucoma in an urban Chinese population: the Singapore Chinese Eye Study. *JAMA Ophthalmol*. 2015;133:874−880.
8. Varma R, Ying-Lai M, Francis BA, et al. Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye Study. *Ophthalmology*. 2004;111:1439−1448.
9. Quigley HA, West SK, Rodriguez J, et al. The prevalence of glaucoma in a population-based study of Hispanic subjects: Proyecto VER. *Arch Ophthalmol*. 2001;119:1819−1826.
10. Gupta P, Zhao D, Guallar E, et al. Prevalence of glaucoma in the United States: the 2005−2008 National Health and Nutrition Examination Survey. *Invest Ophthalmol Vis Sci*. 2016;57:2577−2585.
11. Munoz B, West SK, Rubin GS, et al. Causes of blindness and visual impairment in a population of older Americans: the Salisbury Eye Evaluation Study. *Arch Ophthalmol*. 2000;118:819−825.
12. Shaikh Y, Yu F, Coleman AL. Burden of undetected and untreated glaucoma in the United States. *Am J Ophthalmol*. 2014;158:1121−1129.
13. Lee PP, Walt JG, Doyle JJ, et al. A multicenter, retrospective pilot study of resource use and costs associated with severity of disease in glaucoma. *Arch Ophthalmol*. 2006;124:12−19.
14. Kirwan JF, Gouws P, Linnell AE, et al. Pharmacological mydriasis and optic disc examination. *Br J Ophthalmol*. 2000;84:894−898.
15. Chan HH, Ong DN, Kong YX, et al. Glaucomatous optic neuropathy evaluation (GONE) project: the effect of monoscopic versus stereoscopic viewing conditions on optic nerve evaluation. *Am J Ophthalmol*. 2014;157:936−944.
16. Moyer VA. Screening for glaucoma: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2013;159:484−489.
17. Fleming C, Whitlock EP, Beil T, et al. Screening for primary open-angle glaucoma in the primary care setting: an update for the US preventive services task force. *Ann Fam Med*. 2005;3:167−170.
18. Miller SE, Thapa S, Robin AL, et al. Glaucoma screening in Nepal: cup to disc estimate with standard mydriatic fundus camera compared to portable non-mydriatic camera. *Am J Ophthalmol*. 2017;182:99−106.
19. Pizzi LT, Waisbourd M, Hark L, et al. Costs of a community-based glaucoma detection programme: analysis of the Philadelphia Glaucoma Detection and Treatment Project. *Br J Ophthalmol*. 2018;102(2):225−232.
20. Zhao D, Guallar E, Gajwani P, et al. Optimizing glaucoma screening in high-risk population: design and 1-Year findings of the Screening to Prevent (SToP) Glaucoma Study. *Am J Ophthalmol*. 2017;180:18−28.
21. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124:962−969.
22. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402−2410.
23. Iwase A, Suzuki Y, Araie M, et al. The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. *Ophthalmology*. 2004;111:1641−1648.
24. He M, Foster PJ, Ge J, et al. Prevalence and clinical characteristics of glaucoma in adult Chinese: a population-based study in Liwan District, Guangzhou. *Invest Ophthalmol Vis Sci*. 2006;47:2782−2788.
25. Topouzis F, Wilson MR, Harris A, et al. Prevalence of open-angle glaucoma in Greece: the Thessaloniki Eye Study. *Am J Ophthalmol*. 2007;144:511−519.
26. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol*. 2002;86:238−242.
27. Ebner M. Color constancy based on local space average color. *Machine Vision and Applications*. 2009;11:283−301.
28. Szegedy C, Vanhouke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. http://arxiv.org/pdf/1512.00567v3.pdf; 2015. Accessed May 10, 2017.
29. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial

networks. https://arxiv.org/pdf/1511.06434.pdf; 2015. Accessed June 12, 2017.
30. Singh A, Dutta MK, ParthaSarathi M, et al. Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image. *Comput Methods Programs Biomed*. 2016;124:108–120.
31. Chakrabarty L, Joshi GD, Chakravarty A, et al. Automated detection of glaucoma from topographic features of the optic nerve head in color fundus photographs. *J Glaucoma*. 2016;25:590–597.
32. Issac A, Partha SM, Dutta MK. An adaptive threshold based image processing technique for improved glaucoma detection and classification. *Comput Methods Programs Biomed*. 2015;122:229–244.
33. Xiangyu C, Yanwu X, Damon WKW, et al. Glaucoma detection based on deep convolutional neural network. *Conf Proc IEEE Eng Med Biol Soc*. 2015:715–718.
34. Annan L, Jun C, Damon WKW, Jiang L. Integrating holistic and local deep features for glaucoma classification. *Conf Proc IEEE Eng Med Biol Soc*. 2016:1328–1331.
35. Yang M, Zhang L, Shiu SC, Zhang D. Robust kernel representation with statistical local features for face recognition. *IEEE Trans Neural Netw Learn Syst*. 2013;24:900–912.
36. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–1958.
37. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012:1–9.
38. Ding X, Chang RT, Guo X, et al. Visual field defect classification in the Zhongshan Ophthalmic Center-Brien Holden Vision Institute High Myopia Registry Study. *Br J Ophthalmol*. 2016;100:1697–1702.
39. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311:1901–1911.
40. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *JAMA*. 2016;316:2366–2367.

Footnotes and Financial Disclosures

Originally received: November 24, 2017.
Final revision: January 10, 2018.
Accepted: January 18, 2018.
Available online: March 2, 2018. Manuscript no. 2017-2689.

[1] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China.

[2] Guangzhou Healgoo Interactive Medical Technology Co. Ltd., Guangzhou, China.

[3] Centre for Eye Research Australia; Departments of Ophthalmology and Surgery, University of Melbourne, Melbourne, Australia.

[4] Department of Ophthalmology, Byers Eye Institute, Stanford University, Palo Alto, California.

*These authors contributed equally as first authors.

Financial Disclosure(s):
The author(s) have made the following disclosure(s): W.M.: Patent – Using deep learning models to process color fundus images

M.H.: Patent – Using deep learning models to process color fundus images

Supported in part by the Fundamental Research Funds of the State Key Laboratory in Ophthalmology, National Natural Science Foundation of China (grant no.: 81420108008); the Science and Technology Planning Project of Guangdong Province (grant no.: 2013B20400003); the University of Melbourne at Research Accelerator Program of Australia (M.H.); the CERA Foundation of Australia (M.H.); Victorian State Government of Australia (Operational Infrastructure Support to the Centre for Eye Research Australia); and Research to Prevent Blindness, Inc., New York (to Stanford University Eye Department). The sponsors or funding organizations had no role in the design or conduct of this research.

HUMAN SUBJECTS: Human subjects were included in this study. The institutional review board of Zhongshan Ophthalmic Center approved the study and determined that informed consent was not required because of retrospective nature and fully anonymized usage of images in this study. The study was performed in accordance with the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Li, Meng, M.He

Analysis and interpretation: Li, Keel, Chang

Data collection: Li, Y.He, Keel, Meng

Obtained funding: None

Overall responsibility: Li, Y.He, Keel, Chang, M.He

Abbreviations and Acronyms:
AMD = age related macular degeneration; **AUC** = area under receiver operator characteristic curve; **DD** = disc diameter; **GON** = glaucomatous optic neuropathy; **RNFL** = retinal nerve fiber layer; **VCDR** = vertical cup to disc ratio.

Correspondence:
Mingguang He, MD, PhD, Stata Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, People's Republic of China. E-mail: mingguang.he@unimelb.edu.au.