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;■:1–8 © 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]

Published by Elsevier Inc.
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

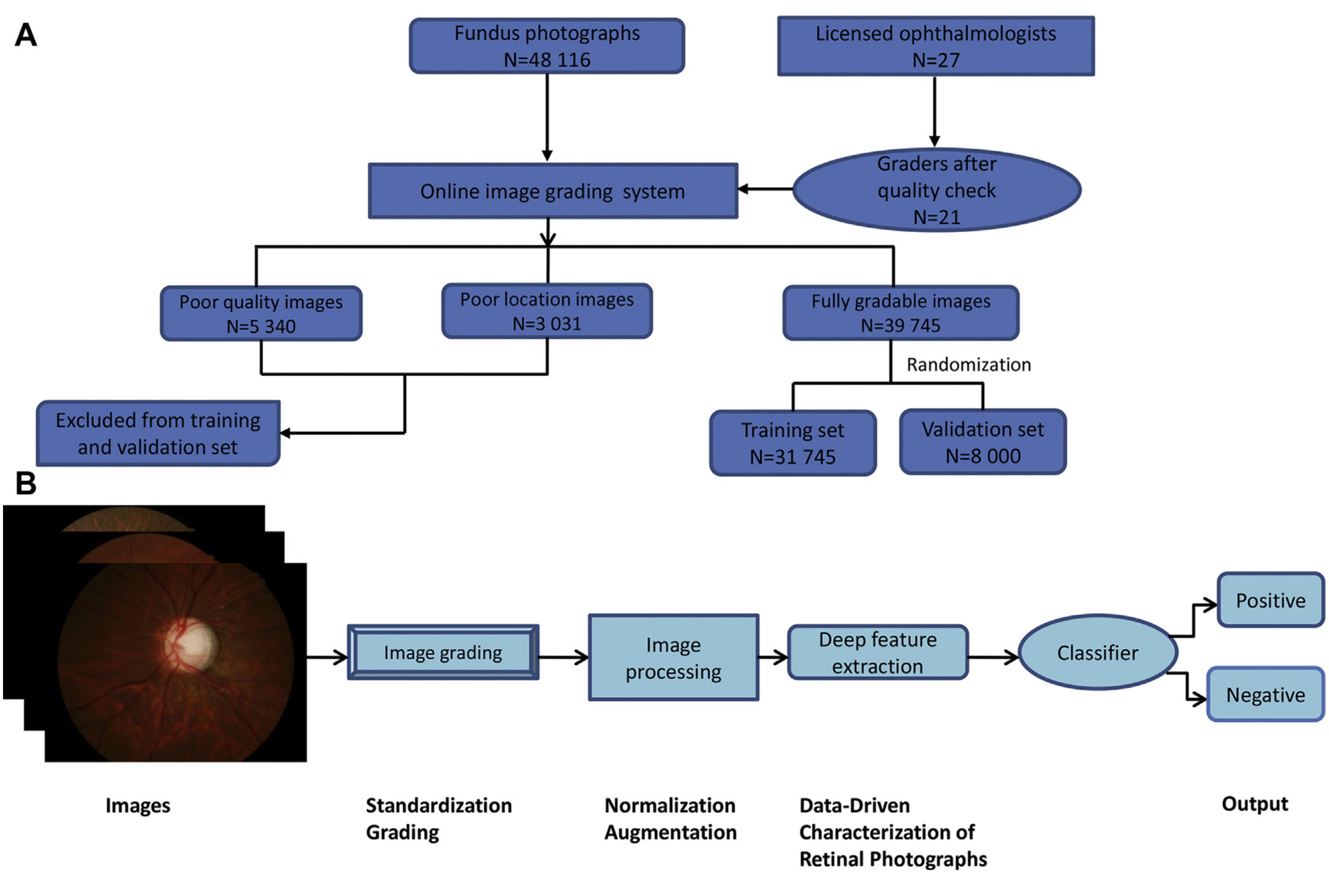

Figure 1. Diagrams showing an overview of the proposed automated detection for referable glaucomatous optic neuropathy using deep learning. **A**, Image grading workflow for our algorithm. **B**, Integration flow in the current study.