

Reliability of Graders and Comparison with an Automated Algorithm for Vertical Cup-Disc Ratio Grading in Fundus Photographs

Wei-han Tong,¹ MB BCH BAO, Maryanne Romero,¹ MD, Vivien Lim,¹ MMed (Ophth), Seng Chee Loon,¹ FRCSEd, Maya E Suwandono,² SpM, Yu Shuang,³ PhD, Xiao Di,³ PhD, Yogi Kanagasingam,³ PhD, Victor Koh,¹ MMed (Ophth)

Abstract

Introduction: We aimed to investigate the intergrader and intragrader reliability of human graders and an automated algorithm for vertical cup-disc ratio (CDR) grading in colour fundus photographs. **Materials and Methods:** Two-hundred fundus photographs were selected from a database of 3000 photographs of patients screened at a tertiary ophthalmology referral centre. The graders included glaucoma specialists (n=3), general ophthalmologists (n=2), optometrists (n=2), family physicians (n=2) and a novel automated algorithm (AA). In total, 2 rounds of CDR grading were held for each grader on 2 different dates, with the photographs presented in random order. The CDR values were graded as 0.1-1.0 or ungradable. The grading results of the 2 senior glaucoma specialists were used as the reference benchmarks for comparison. **Results:** The intraclass correlation coefficient values ranged from 0.37-0.74 and 0.47-0.97 for intergrader and intragrader reliability, respectively. There was no significant correlation between the human graders' level of reliability and their years of experience in grading CDR ($P = 0.91$). The area under the curve (AUC) value of the AA was 0.847 (comparable to AUC value of 0.876 for the glaucoma specialist). Bland Altman plots demonstrated that the AA's performance was at least comparable to a glaucoma specialist. **Conclusion:** The results suggest that AA is comparable to and may have more consistent performance than human graders in CDR grading of fundus photographs. This may have potential application as a screening tool to help detect asymptomatic glaucoma-suspect patients in the community.

Ann Acad Med Singapore 2019;48:282-9

Key words: Eyes, Glaucoma, Intraocular pressure

Introduction

Glaucoma is one of the world's leading causes of irreversible blindness, with certain types of glaucoma such as angle-closure glaucoma being more prevalent in East Asia.¹ The global prevalence is currently estimated to be 3.54% in those aged 40-80 years old.² People affected by glaucoma in 2013 was estimated to be 64.3 million.² This is expected to rise to 111.8 million by 2040.² This trend of an increasing eye disease burden is also expected to be seen in Singapore in future.^{3,4} As glaucoma is asymptomatic in the early stages, it is important for early detection and diagnosis so that blindness may be prevented with early treatment.⁵

Glaucoma is primarily diagnosed by ophthalmologists in tertiary eye care and commonly after incidental findings of high intraocular pressure or a raised vertical cup-disc ratio (CDR) detected on retinal photographs.

Apart from a simple colour photograph, there are other methods of glaucoma screening, such as the use of tonometry, the Van Herick method, optical coherence tomography (OCT), OCT-angiography, macula vessel density, ganglion cell inner plexiform layer thickness and automated perimetry tests.⁶⁻⁹ Studies have shown that the determination of the optic disc margin based on colour photographs does not always correlate with the Bruch's membrane opening,

¹Department of Ophthalmology, National University Hospital, Singapore

²Department of Glaucoma, KMN EyeCare, Indonesia

³Commonwealth Scientific and Industrial Research Organisation, Australia

Address for Correspondence: Dr Victor Koh, Department of Ophthalmology, National University Hospital, Level 7, National University Health System (NUHS) Tower Block, 1E Kent Ridge Road, Singapore 119228.

Email: victor_koh@nuhs.edu.sg

which should be the anatomically correct disc margin.^{10,11} This leads to a mismatch between the clinically perceived disc margin and the true Bruch's membrane opening, which can be detected by OCT.¹⁰ This can cause errors in the estimation of the optic disc size and therefore errors in CDR estimation. Hence, OCT had been proposed to be used in clinical assessment for better accuracy and consistency at measuring CDR.¹² Accurate measurement of CDR is important as variability in measurements may affect the risk estimation and subsequent management of patients who are glaucoma-suspects.¹³ However, many of the above approaches require expensive bulky equipment and are usually not practical for screening purposes. Primary care physicians often use colour fundus photographs as the main ocular screening modality as there is a large number of diabetic patients who require regular diabetic retinopathy screening. Most fundus photographs are either optic disc or macula centred images and the optic nerve head can be analysed for features of glaucomatous optic neuropathy, especially raised vertical CDR.

The optic nerve head appearance—in particular the vertical CDR—is usually determined subjectively by a grader who can include one of the following: ophthalmologists, optometrists, primary care physicians, trained graders or research staff within a grading centre. Studies have shown that this is prone to high intra and intergrader variability which could be worse in real-world clinical setting that is typically without standardised training.¹⁴ As such, the use of automated algorithms (AA) in the detection of vertical CDR on optic disc photos might have the potential to be used in mass screening settings.

For the current study, we aimed to determine the intergrader and intragrader reliability between different professionals namely glaucoma specialists, general ophthalmologists, family physicians and optometrists. In addition, we aimed to validate and compare human grading with a novel AA in determining vertical CDR from colour fundus photographs.

Materials and Methods

This study was approved by the Domain Specific Review Board of the National Healthcare Group in Singapore. Fundus photographs were obtained from patients who came for routine eye screening at the ophthalmology department of a tertiary referral hospital. The photographs were taken using VISUCAM^{PRO} NM (Carl Zeiss Meditec Inc., United States of America). We selected 200 macular-centred fundus photographs for the grading process. The photographs were selected to show optic discs with different vertical CDR as well as poor quality images of optic discs. The photographs were grouped as vertical CDR 0.1-0.2, 0.3-0.4, 0.5-0.6, 0.7-0.8, 0.9-1.0 and ungradable images. The

sample photographs of each group are shown in Figure 1. There were 33, 33, 33, 34, 34 and 33 photographs in each group, respectively.

In this study, graders with different levels of experience in assessing vertical CDR on colour photographs were included. They included glaucoma specialists ($n = 3$), general ophthalmologists ($n = 2$), family physicians ($n = 2$) and optometrists ($n = 2$). Of the glaucoma specialists, there were 2 fellowship-trained glaucoma senior specialists, each with more than 10 years of experience, while another was a glaucoma specialist with 5 years of experience. One general ophthalmologist had 10 years of experience and the other had 5 years of experience. Both optometrists had at least 5 years of experience working in an ophthalmology department of a tertiary hospital. Both family physicians included in this study had basic knowledge of the optic nerve head anatomy but had graded less than 10 colour fundus photographs a year.

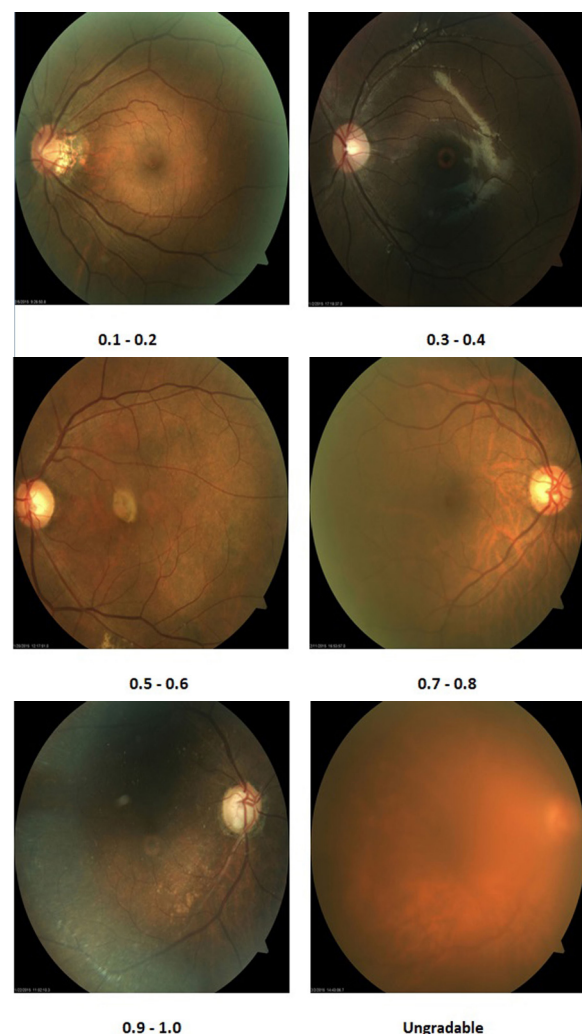


Fig. 1. Sample photographs of each vertical cup-disc ratio group.

The newly developed AA assessed in this study is currently an unvalidated system in large-scale clinical settings, particularly in Asia. It had been tested on various known datasets of mixed-ethnic groups, achieving comparable performance of what is state-of-the-art in the literature.¹⁵ The AA is mainly based on a deep learning model named U-Net, which is a U-shaped convolutional network originally developed for biomedical image segmentation. We developed a modified U-Net model for optic disc and cup segmentation. The proposed optic disc and cup segmentation procedure involves 2 main steps. A colour fundus image is firstly preprocessed with contrast enhancement, followed by resizing to the dimension of 512×512 pixels. Then, our single-label modified U-Net model is applied to roughly segment the optic disc region in the image as the region of interest (ROI). In step 2, the ROI is resized to the dimension of 512×512 pixels. Our multilabel U-Net model is employed to segment the optic disc and cup simultaneously. The predicted optic disc and cup from step 2 are further processed via a morphological postprocessing module for the finetuning of their boundaries. Finally, vertical CDR is calculated using the vertical diameter of the optic cup versus that of the optic disc.

Graders reviewed and graded all 200 fundus photographs independently. Each photograph was graded with vertical CDR to 1 decimal place or as “ungradable”. “Ungradable” was defined as a poor quality photograph including unfocused image of the optic disc and obstruction over the optic disc that precluded grading. Two rounds of grading were conducted for each human grader using the same 200 photographs on 2 different dates. During the second round of grading, the photographs were presented back to them at random, in a different order compared to the first round of grading.

Statistical analysis was carried out using SPSS Statistics (Version 25). The gradings by the 2 senior glaucoma specialists were set as “gold standards” in the study to serve as the 2 reference points for comparison to other graders. Analysis was carried out to assess intragrader and intergrader reliability. This was measured by the intraclass correlation coefficient (ICC). The ICC can be categorised into different levels of reliability. Values <0.5 indicate poor reliability, 0.5 – 0.75 indicate moderate reliability, 0.75 – 0.9 indicate good reliability and >0.90 indicate excellent reliability.¹⁶ Statistical significance was set at $P < 0.05$. Correlation between levels of reliability and years of experience were assessed using logistic regression. Correlation was also assessed using Bland Altman plots.

In addition, receiver operating characteristic (ROC) curves were plotted. ROC serves to assess characteristics of the AA via area under the curve (AUC). Outcomes in the

ROC curve were dichotomised into binary outcomes based on the senior glaucoma specialists’ gradings of “Normal CDR” and “Referable CDR” (with CDR values of <0.6 being classified as “Normal CDR”). The cut-off value of 0.6 was chosen because once $CDR \geq 0.6$, the probability of abnormality increases dramatically.¹⁷ Ungradable images were also classified under “Referable CDR”. This division was to determine the functional capability of the AA (if adapted in future for community eye screening) to classify patients correctly into the 2 categories—healthy individuals or individuals who are potential glaucoma-suspects that need further specialist assessment.

Results

The results of intergrader reliability are shown in Table 1 with senior glaucoma specialist 1 as the reference. Table 2 shows the intergrader reliability with senior glaucoma specialist 2 as the reference. Both senior glaucoma specialists achieved moderate intergrader reliability. ICC values showed a “moderate” level of reliability for most of our graders compared to the reference grader. One grader (optometrist) exhibited a “poor” level of intergrader reliability (ICC 0.37). Interestingly, there was no significant correlation between the graders’ level of intergrader reliability and their experience in CDR grading (Table 3). The family physician group—who had minimal grading experience—was able to display a similar level of reliability to that of the glaucoma specialists. On the other hand, for the optometrists—despite having similar experience—there was a large disparity in their levels of reliability.

The intragrader reliability are shown in Table 4. Once again, there was no relationship between grader experience

Table 1. Intergrader Reliability in Vertical Cup-Disc Ratio Grading (Senior Glaucoma Specialist 1 As Reference)

	Intergrader Reliability			
	Intraclass Correlation Coefficient	95% Confidence Interval	P Value	Level of Reliability
Senior glaucoma specialist 2	0.69	0.46 – 0.80	<0.001	Moderate
Glaucoma specialist	0.69	0.61 – 0.76	<0.001	Moderate
General ophthalmologist 1	0.70	0.62 – 0.76	<0.001	Moderate
General ophthalmologist 2	0.71	0.62 – 0.77	<0.001	Moderate
Optometrist 1	0.74	0.67 – 0.80	<0.001	Moderate
Optometrist 2	0.37	0.19 – 0.52	<0.001	Poor
Family physician 1	0.61	0.51 – 0.69	<0.001	Moderate
Family physician 2	0.67	0.58 – 0.74	<0.001	Moderate
Automated algorithm	0.52	0.39 – 0.63	<0.001	Moderate

Table 2. Intergrader Reliability in Vertical Cup-Disc Ratio Grading (Senior Glaucoma Specialist 2 As Reference)

	Intergrader Reliability			
	Intraclass Correlation Coefficient	95% Confidence Interval	P Value	Level of Reliability
Senior glaucoma specialist 1	0.69	0.46 – 0.80	<0.001	Moderate
Glaucoma specialist	0.71	0.60 – 0.80	<0.001	Moderate
General ophthalmologist 1	0.72	0.41 – 0.64	<0.001	Moderate
General ophthalmologist 2	0.49	0.16 – 0.69	<0.001	Poor
Optometrist 1	0.73	0.54 – 0.83	<0.001	Moderate
Optometrist 2	0.37	0.21 – 0.45	<0.001	Poor
Family physician 1	0.66	0.49 – 0.77	<0.001	Moderate
Family physician 2	0.55	0.19 – 0.74	<0.001	Moderate
Automated algorithm	0.65	0.55 – 0.72	<0.001	Moderate

Table 3. Correlation Between Grader Experience and Reliability Levels (Senior Glaucoma Specialist 1 As Reference)

	Years of Experience	Inter-grader Reliability	P Value	Intra-grader Reliability	P Value
Glaucoma specialist	5	Moderate	0.91	Moderate	0.57
General ophthalmologist 1	10	Moderate	0.91	Moderate	0.57
General ophthalmologist 2	5	Moderate	0.91	Good	0.57
Optometrist 1	5	Moderate	0.91	Moderate	0.57
Optometrist 2	5	Poor	0.91	Good	0.57
Family physician 1	0	Moderate	0.91	Poor	0.57
Family physician 2	0	Moderate	0.91	Moderate	0.57

Table 4. Intragrader Reliability in Vertical Cup-Disc Ratio Grading

	Intergrader Reliability			
	Intraclass Correlation Coefficient	95% Confidence Interval	P Value	Level of Reliability
Senior glaucoma specialist 1	0.70	0.61 – 0.76	<0.001	Moderate
Senior glaucoma specialist 2	0.97	0.96 – 0.98	<0.001	Excellent
Glaucoma specialist	0.71	0.63 – 0.77	<0.001	Moderate
General ophthalmologist 1	0.87	0.83 – 0.90	<0.001	Good
General ophthalmologist 2	0.64	0.55 – 0.71	<0.001	Moderate
Optometrist 1	0.87	0.83 – 0.90	<0.001	Good
Optometrist 2	0.47	0.36 – 0.57	<0.001	Poor
Family physician 1	0.52	0.41 – 0.61	<0.001	Moderate
Family physician 2	0.62	0.52 – 0.70	<0.001	Moderate

and level of intragrader reliability (Table 3). Senior glaucoma specialist 2, who served as one of the reference standards, achieved an “excellent” level of intragrader reliability. One of the optometrists and a general ophthalmologist achieved a “good” level of intragrader reliability, which was better than the glaucoma specialist.

The ROC curves were plotted for the AA and senior glaucoma specialist 1 (Fig. 2) and for senior glaucoma specialist 2 (Fig. 3). In comparison to the reference standard set by senior glaucoma specialist 1, AUC for the AA was 0.829 (comparable to the AUC value of 0.823 of the glaucoma specialist). Comparing to senior glaucoma specialist 2, the AUC for AA was 0.847, while AUC of the glaucoma specialist was 0.876. These results suggest that the AA grader had discriminative power that was comparable to that of a glaucoma specialist.¹⁸

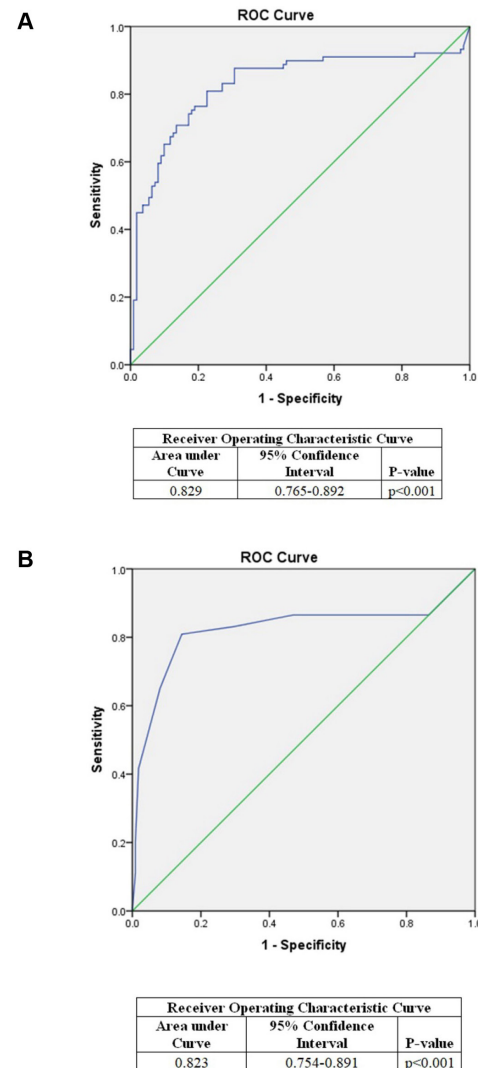


Fig. 2. Receiver operating characteristic curves. A: Automated algorithm (senior glaucoma specialist 1). B: Glaucoma specialist (senior glaucoma specialist 1).

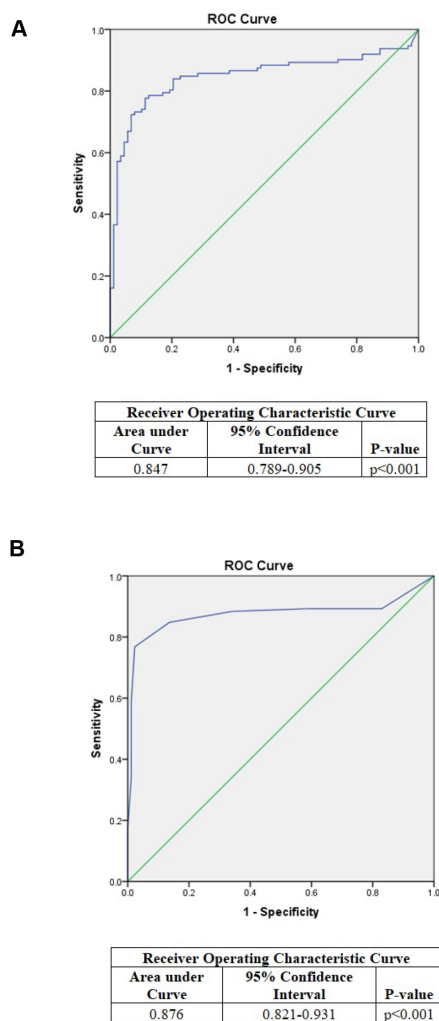


Fig. 3. Receiver operating characteristic curves. A: Automated algorithm (senior glaucoma specialist 2). B: Glaucoma specialist (senior glaucoma specialist 2).

In addition, Bland Altman plots were performed. The plots compared the results of the AA against the 2 senior glaucoma specialists (Fig. 4). There appears to be more proportional bias in Figure 4A, which may be related to the poorer intragrader reliability of senior glaucoma specialist 1 as compared to senior glaucoma specialist 2. Hence, for subsequent plots, senior glaucoma specialist 2 was used as the reference grader in comparison to other graders (Fig. 5). Based on the plots, the AA appears to have reliability comparable to that of a glaucoma specialist.

Discussion

Fundus photography remains one of the most common and simplest methods for eye screening, especially for clinics located in the community where family medicine is practised. In such settings, sophisticated and expensive ophthalmology equipment are not practical or cost-effective. Hence, fundus photography remains a useful modality for

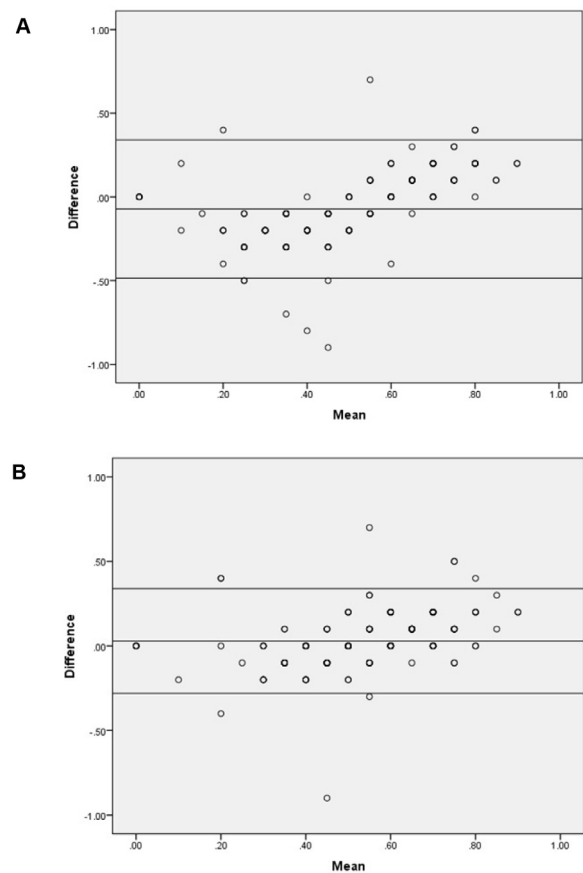


Fig. 4. Bland Altman plots of the automated algorithms. A: Against senior glaucoma specialist 1. B: Against senior glaucoma specialist 2.

the screening of ocular diseases and is the focus in the development of AAs for mass screening. Portable fundus cameras are also an option for use in tele-ophthalmology diagnosis of glaucoma which can potentially benefit patients with mobility issues and who face difficulties visiting eye clinics due to the long distance from their homes.^{19,20} They may also benefit people living in more rural areas with minimal access to tertiary eye care. Use of the same imaging modality for subsequent fundus photographs will also allow monitoring of CDR and for the detection of any changes.²¹

Developing an AA system to determine CDR accurately does not come without challenges. Many methods have been previously proposed for the processes of optic disc extraction, optic cup extraction and peripapillary atrophy (PPA) localisation.²² Different methods exhibit different levels of accuracy.²² Various new AA systems are being developed for possible future use in glaucoma screening.²³⁻²⁵ Each system employs different techniques for glaucoma detection on colour fundus photographs. Some used OCT imaging for CDR calculation while another used a combination of structural and non-structural features for glaucoma detection.^{23,25} Shibata et al developed

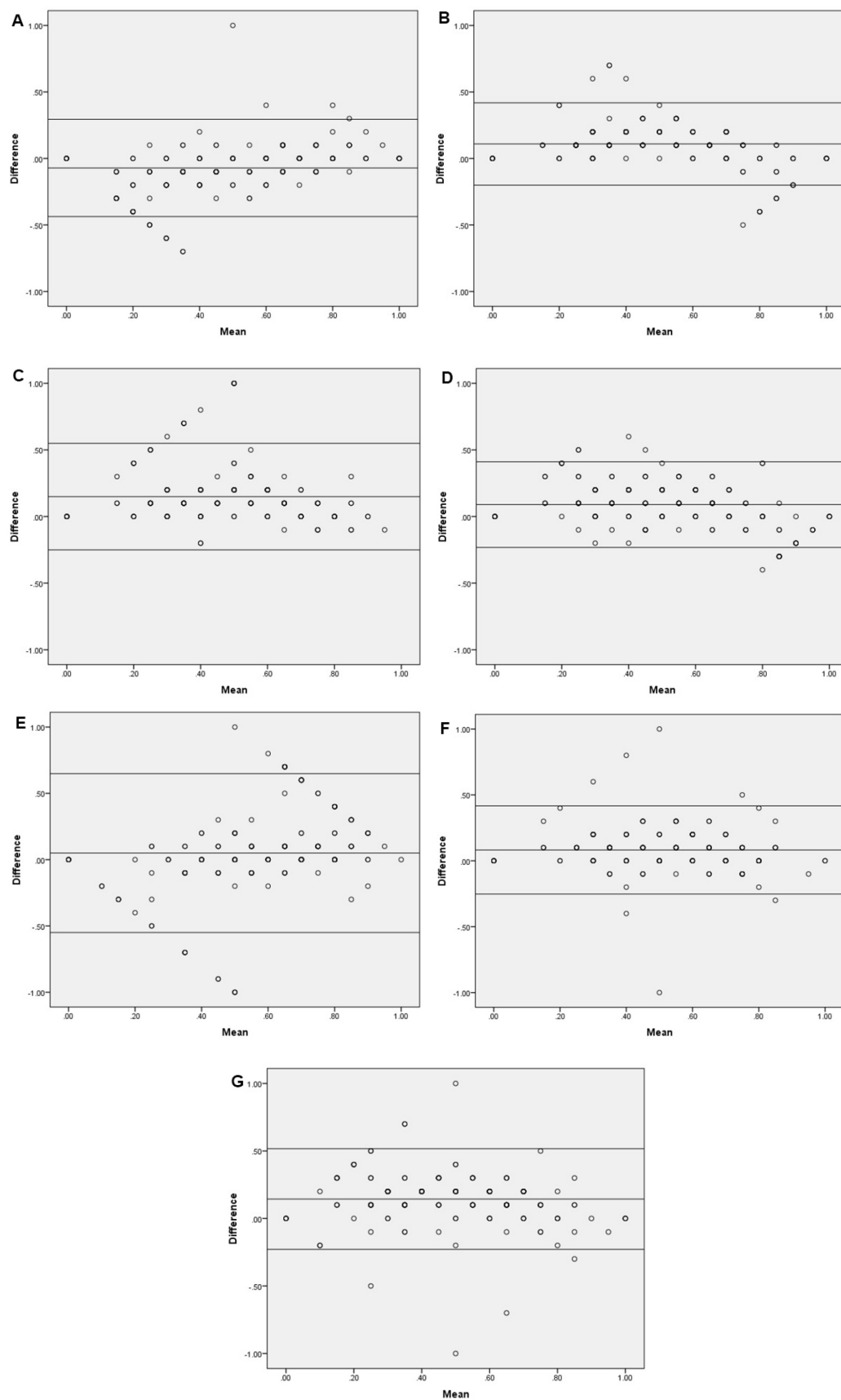


Fig. 5. Bland Altman plots. A: Glaucoma specialist against senior glaucoma specialist 2. B: General ophthalmologist 1 against senior glaucoma specialist 2. C: General ophthalmologist 2 against senior glaucoma specialist 2. D: Optometrist 1 against senior glaucoma specialist 2. E: Optometrist 2 against senior glaucoma specialist 2. F: Family physician 1 against senior glaucoma specialist 2. G: Family physician 2 against senior glaucoma specialist 2.

an AA that showed higher diagnostic performance to their ophthalmology residents.²⁶ Christopher et al also evaluated various deep learning architectures to assess their performance in detecting varying degrees of glaucomatous optic neuropathy in fundus photographs.²⁷ Their best performing model achieved an AUC of 0.91 for healthy eyes, 0.89 in eyes with mild glaucoma and 0.97 in moderate to severe glaucomatous eyes.²⁷ These 2 systems aimed to detect various glaucomatous changes, all at the same time based on retinal images alone and not just to calculate CDR. The changes included rim notching, laminar dot sign, disc edge haemorrhages and retinal nerve fibre layer defects.²⁶ In a previous study that is similar to ours, Muramatsu et al proposed a method of automated CDR determination on stereo fundus images via segmentation of the cup region using a depth map.²⁸ AUC achieved by the AA system in that study was 0.90 compared to an ophthalmologist's assessment.²⁸ In another recent study, Li et al presented a deep learning algorithm that achieved an AUC of 0.986 in detecting glaucomatous optic neuropathy on fundus photographs.²⁹

Our study differs from those described as our main focus is for the AA to accurately determine only the vertical CDR from colour fundus photographs. The aim is for the AA to serve as an initial screening tool to detect glaucoma-suspects and when found, the patients would still require prompt specialist assessment for diagnosis, monitoring and treatment. It is not meant to be a substitute for a full ocular assessment by an ophthalmologist. The method of detecting raised CDR is also simple and easier to comprehend, even for non-medical personnel. If any doubts arise when it is implemented in clinical practice, a trained human grader can help to verify the CDR determined by the AA. In contrast, verifying other signs of glaucomatous change on the optic nerve—such as rim notching and laminar dot sign—may not be as easy for the untrained grader. Unlike Muranatsu et al, we did not use stereo images in our study.²⁸ Our aim is also for the AA to function in the community without advanced equipment such as stereo fundus photography. In our validation study, the AA's AUC is very similar to that of a glaucoma specialist, suggesting that it can detect referable CDR accurately. With additional training, the AA system might be used as a good first-line screening tool in large groups of people in the community, followed by subsequent assessment by doctors when there is suspicion of glaucoma. It is difficult to compare AAs in various studies because they were evaluated by different methods that used different public or private databases.

There is a significant amount of variability between observers in determining clinical vertical CDR.¹⁴ Our study also showed the inherent intergrader and intragrader variations in CDR assessment among different individuals.

This is expected as the method of manual grading by looking at the physical appearance of the optic nerve is highly subjective. The reasons for this may be due to differences in prior training in grading colour fundus photographs, lack of stereoscopic views, variations in optic disc morphology and lack of familiarity in CDR grading.^{30,31} Even between glaucoma specialists in our study, there is considerable intergrader variability.³² This suggests that a more standardised automated screening tool might be useful in improving consistency. First, it will minimise intergrader variations during colour fundus photograph grading of vertical CDR. Second, an AA will be faster in screening a large number of optic disc photographs.

This study has various limitations. First, the number of fundus photographs was small and comprised only Asian eyes. In future, larger numbers of fundus photographs are needed to obtain results that will be representative of screening large numbers of people. The AA has yet to be validated on large-scale clinical datasets. In addition, before it can be applied in real life, it should ideally be evaluated under clinical conditions where a large number of patients with different demographics and ocular issues (e.g., presence of cataracts) are assessed. However, it is interesting to note that the current AA is also able to identify “ungradable” optic disc images and this would be useful in a clinical setting. Second, this study was done with assessment made in reference to CDR grading by either 1 of the 2 designated “gold standard” graders who were senior glaucoma specialists. Identification of true glaucoma cases is inadequate and accurate CDR measurements are prone to grader variability. It will be better if fundus photographs are interpreted in conjunction with other ocular investigations, such as OCT scans of the retinal nerve fibre layers and visual field tests. This will enable the identification of true glaucoma patients and consistent CDR measurements, thereby creating an accurate reference for comparison to other graders.

In conclusion, we showed that there is significant inter and intragrader variability among human graders for vertical CDR in fundus photographs. The use of an AA is comparable to a glaucoma specialist and can potentially play an important role in mass screening of glaucoma-suspects.

Acknowledgements

The authors would like to thank the following individuals for their assistance in the study: Family physicians Drs Yap Juin Shiong and Tong Shanhan and optometrists Ms Lor Keli and Ms Jacqueline Liew for their assistance with cup-disc ratio grading. The authors would also like to thank Dr Shen Liang (senior biostatistician, Biostatistics Unit, National University of Singapore) for her help in statistical analysis.

REFERENCES

- Ang LP. Current understanding of the treatment and outcome of acute primary angle-closure glaucoma: an Asian perspective. *Ann Acad Med Singapore* 2008;37:210-5.
- Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* 2014;121:2081-90.
- Ansah JP, Koh V, de Korne DF, Bayer S, Pan C, Thiyagarajan J, et al. Projection of eye disease burden in Singapore. *Ann Acad Med Singapore* 2018;47:13-28.
- See JL, Wong TY, Yeo KT. Trends in the pattern of blindness and major ocular diseases in Singapore and Asia. *Ann Acad Med Singapore* 1998;27:540-6.
- Lim AS, Khoo CY, Ang BC, Tan J, Heng LK. Eye diseases in the elderly in Singapore. *Ann Acad Med Singapore* 1987;16:46-53.
- Muhammad H, Fuchs TJ, De Cuir N, De Moraes CG, Blumberg DM, Liebmann JM, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma* 2017;26:1086-94.
- Rolle T, Dallorto L, Tavassoli M, Nuzzi R. Diagnostic ability and discriminant values of OCT-angiography parameters in early glaucoma diagnosis. *Ophthalmic Res* 2019;61:143-52.
- Silva FR, Vidotti VG, Cremasco F, Dias M, Gomi ES, Costa VP. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using spectral domain OCT and standard automated perimetry. *Arq Bras Oftalmol* 2013;76:170-4.
- Hoh ST. Evaluating the optic nerve and retinal nerve fibre layer: the roles of Heidelberg retina tomography, scanning laser polarimetry and optical coherence tomography. *Ann Acad Med Singapore* 2007;36:194-202.
- Reis ASC, Sharpe GP, Yang H, Nicoleta MT, Burgoyne CF, Chauhan BC. Optic disc margin anatomy in patients with glaucoma and normal controls with spectral domain optical coherence tomography. *Ophthalmology* 2012;119:738-47.
- Amini N, Miraftabi A, Henry S, Chung N, Nowroozizadeh S, Caprioli J, et al. The relationship of the clinical disc margin and Bruch's membrane opening in normal and glaucoma subjects. *Invest Ophthalmol Vis Sci* 2016;57:1468-75.
- Chauhan BC, Burgoyne CF. From clinical examination of the optic disc to clinical assessment of the optic nerve head: a paradigm change. *Am J Ophthalmol* 2013;156:218-27.e2.
- Chan PP, Chiu V, Wong MO. Variability of vertical cup to disc ratio measurement and the effects of glaucoma 5-year risk estimation in untreated ocular hypertensive eyes. *Br J Ophthalmol* 2019;103:361-8.
- Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc* 1976;74:532-72.
- Yu S, Xiao D, Frost S, Kanagasingam Y. Robust optic disc and cup segmentation with deep learning for glaucoma detection. *Comput Med Imaging Graph* 2019;74:61-71.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-63.
- Quigley HA. Open-angle glaucoma. *N Engl J Med* 1993;328:1097-106.
- Šimundić AM. Measures of diagnostic accuracy: basic definitions. *EJIFCC* 2009;19:203-11.
- Yogesana K, Constable IJ, Barry CJ, Eikelboom RH, Morgan W, Taylor-Kearney ML, et al. Evaluation of a portable fundus camera for use in the teleophthalmologic diagnosis of glaucoma. *J Glaucoma* 1999;8:297-301.
- Miller SE, Thapa S, Robin AL, Niziol LM, Ramulu PY, Woodward MA, et al. Glaucoma screening in Nepal: cup-to-disc estimate with standard mydriatic fundus camera compared to portable nonmydriatic camera. *Am J Ophthalmol* 2017;182:99-106.
- Mwanza JC, Grover DS, Budenz DL, Herndon LW, Nolan W, Whiteside-de Vos J, et al. A comparison of cup-to-disc ratio estimates by fundus biomicroscopy and stereoscopic optic disc photography in the Tema Eye Survey. *Eye (Lond)* 2017;31:1184-90.
- Haleem MS, Han L, van Hemert J, Li B. Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: a review. *Comput Med Imaging Graph* 2013;37:581-96.
- Salam AA, Khalil T, Akram MU, Jameel A, Basit I. Automated detection of glaucoma using structural and non structural features. *Springerplus* 2016;5:1519.
- Fernandez-Granero MA, Sarmiento A, Sanchez-Morillo D, Jiménez S, Alemany P, Fondón I. Automatic CDR estimation for early glaucoma diagnosis. *J Healthc Eng* 2017;2017:5953621.
- Miri MS, Abramoff MD, Lee K, Niemeijer M, Wang JK, Kwon YH, et al. Multimodal segmentation of optic disc and cup from SD-OCT and color fundus photographs using a machine-learning graph-based approach. *IEEE Trans Med Imaging* 2015;34:1854-66.
- Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci Rep* 2018;8:14665.
- Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep* 2018;8:16685.
- Muramatsu C, Nakagawa T, Sawada A, Hatanaka Y, Yamamoto T, Fujita H. Automated determination of cup-to-disc ratio for classification of glaucomatous and normal eyes on stereo retinal fundus images. *J Biomed Opt* 2011;16:096009.
- Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 2018;125:1199-206.
- Hanson S, Krishnan SK, Phillips J. Observer experience and cup: disc ratio assessment. *Optom Vis Sci* 2001;78:701-5.
- Teitelbaum BA, Haefs R, Connor D. Interobserver variability in the estimation of the cup/disk ratio among observers of differing educational background. *Optometry* 2001;72:729-32.
- Kwon YH, Adix M, Zimmerman MB, Piette S, Greenlee EC, Alward WL, et al. Variance owing to observer, repeat imaging, and fundus camera type on cup-to-disc ratio estimates by stereo planimetry. *J Glaucoma* 2009;18:305-10.