Clinical & Experimental Ophthalmology
RANZCO The Royal Australian and New Zealand College of Ophthalmologists
Clinical and Experimental Ophthalmology 2018; 46: 412–416 doi: 10.1111/ceo.13056

Original Article

Diabetic retinopathy screening using deep neural network

Nishanthan Ramachandran, Sheng Chiong Hong, Mary J Sime and Graham A Wilson
[1]Eye Department, Dunedin Hospital, Dunedin, New Zealand

Abstract

Importance: There is a burgeoning interest in the use of deep neural network in diabetic retinal screening.

Background: To determine whether a deep neural network could satisfactorily detect diabetic retinopathy that requires referral to an ophthalmologist from a local diabetic retinal screening programme and an international database.

Design: Retrospective audit.

Participants: Diabetic retinal photos from Otago database photographed during October 2016 (485 photos), and 1200 photos from Messidor international database.

Methods: Receiver operating characteristic curve to illustrate the ability of a deep neural network to identify referable diabetic retinopathy (moderate or worse diabetic retinopathy or exudates within one disc diameter of the fovea).

Main Outcome Measures: Area under the receiver operating characteristic curve, sensitivity and specificity.

Results: For detecting referable diabetic retinopathy, the deep neural network had an area under receiver operating characteristic curve of 0.901 (95% confidence interval 0.807–0.995), with 84.6% sensitivity and 79.7% specificity for Otago and 0.980 (95% confidence interval 0.973–0.986), with 96.0% sensitivity and 90.0% specificity for Messidor.

Conclusions and Relevance: This study has shown that a deep neural network can detect referable diabetic retinopathy with sensitivities and specificities close to or better than 80% from both an international and a domestic (New Zealand) database. We believe that deep neural networks can be integrated into community screening once they can successfully detect both diabetic retinopathy and diabetic macular oedema.

Key words: artificial intelligence, computer, diabetic retinopathy, neural network, screening.

Introduction

Diabetes mellitus poses a significant health burden globally, where 415 million (1 in 11) adults currently have diabetes mellitus, and this number is expected to increase to 642 million (1 in 10) adults by the year 2040.[1] The prevalence of diabetic retinopathy (DR) in individuals with diabetes mellitus is expected to be 35% and vision-threatening DR (severe non-proliferative DR to proliferative DR) to be 10% globally.[2] New Zealand (NZ) has similar proportions with more than 275,000 people living with diabetes; and of those, about 20–25% with DR.[3] There is good evidence that retinal screening and subsequent treatment reduces preventable blindness.[3]

Currently there is a nationwide DR screening programme provided by 26 centres in NZ[4]; however, this is neither linked nor co-ordinated. Grading can be undertaken by ophthalmologists, optometrists, non-ophthalmic medical practitioners and other allied health professionals, with secondary grading reserved for ophthalmologists and some optometrists.[4] Variation in grading centres have been described;[4] and this is not unexpected as there is likely to be inter-specialty and inter-reporter variability. Screening guidelines would typically expect a sensitivity and specificity to exceed 80%.[5]

■ **Correspondence:** Dr Nishanthan Ramachandran, Eye Department, Dunedin Hospital, 201 Great King Street, Dunedin 9016, New Zealand. Email: nishan.rama@gmail.com

Received 22 July 2017; accepted 31 August 2017.

Conflict of interest: None declared.

Funding sources: None declared.

There is evidence that artificial intelligence in the form of deep learning algorithm (deep neural network) is able to detect referable DR with sensitivity and specificity that well exceed typical expectations.[5] Gulshan *et al.* demonstrated that their deep neural network could detect referable DR to a sensitivity between 87% and 90% and specificity of 98%, with an area under receiver operating characteristic (ROC) curve of 0.99.[5,6] In Dunedin Hospital grading of DR photographs from Otago Diabetic Eye Monitoring Service (ODEMS) is mostly done by an ophthalmologist with help from an ophthalmic medical photographer.

The aim of this study was to assess whether a deep neural network could satisfactorily select patients enrolled in the ODEMS, who need to be seen by an ophthalmologist in clinic due to the severity of their DR or diabetic macular oedema (DME). For comparison to international data, we also assessed the deep neural network's ability to satisfactorily grade referable DR and DME from the publicly available international Messidor database.[7]

METHODS

This retrospective study took place in April 2017 and looked at diabetic retinal screening photos from 1 October 2016 to 31 October 2016 from ODEMS as well as the 1200 diabetic retinal photos from Messidor.

The ODEMS database yielded 294 patients who were called for diabetic screening during this time. All diabetic retinal photos were photographed using 'Canon CR-2 Plus Digital Non-Mydriatic Retinal Camera (Canon Inc., Melville, New York, USA). The field of view was 45° and the photos were of 18 megapixel resolution (5184 × 3456 pixels).[8] The trained ophthalmic medical photographers took at least two posterior pole photos (a macula centred and a macula off-centred temporally by one disc diameter) and one nasal photo for each eye. Three quarters of photos were captured with pupil dilation and one quarter without dilation.

All ODEMS photos were assessed initially by a primary grader, an accredited ophthalmic medical photographer as per the New Zealand Ministry of Health (NZ MoH) guideline, followed by an ophthalmologist if grading by a secondary grader was required.[3,4] Each eye was assigned a DR and DME grade by the assessor as per the NZ MoH guideline.[3] Messidor is a French database of posterior pole retinal photos that has been publicly disseminated since the year 2008.[7] The images were captured by three ophthalmology departments in three sets of 400 images of the following resolution: 1440 × 960 pixels, 2240 × 1488 pixels and 2304 × 1536 pixels for each respective set.[9] A total of 800 images were photographed with pupil dilation and 400 were photographed without dilation.[9] Photographs were taken using 'colour video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinograph with a 45° field of view'.[9] Microaneurysms in images were denoted by two specialists in each ophthalmology department.[9] Referable DR in this study was defined as moderate DR (DR more severe than that seen in Early Treatment Diabetic Retinopathy Study 2A photo)[3,10] or moderate DME (presence of exudates within one disc diameter of the fovea).[3]

The study used a third-party deep neural network software called Visiona Intelligent Diabetic Retinopathy Screening Platform (Visiona Medtech International Limited, Hong Kong), abbreviated to Visiona, to assess DR grades of the photos from ODEMS and Messidor. Deep neural network is an advanced form of artificial intelligence, which is able to program itself by learning from a large training set to perform a specific task.[6,11] Visiona has been trained by a training set consisting of more than a 100 000 DR images of posterior pole of the fundus. In the training set, more than 30% of the images were graded as being of grade severe enough to be referable to an ophthalmologist. Each retinal image in the training set was graded by more than one experienced grader (ophthalmologist, optometrist or trained retinal image grader). The Visiona software used for this study provided a DR score 0.00–4.00, where a score of 1.50 or higher would indicate at least moderate DR severity and hence referable to an ophthalmologist in clinic. However, Visiona did not specifically grade DME.

The investigator (NR) collected and selected best quality posterior pole photos that were graded during October 2016 from ODEMS and uploaded these onto Visiona, along with posterior pole Messidor photos for grading. Statistical analysis was carried out using 'IBM SPSS statistics 24'. ROC curve was plotted to test DR score from Visiona against referable criteria for both ODEMS and Messidor. ROC curve is a method for selecting the optimal cut-off value for a test to maximize sensitivity and specificity of finding the abnormal condition.[12] Ethics approval was sought from 'Health Research South' before the commencement of the study (Ethics committee reference number HD17/002).

RESULTS

A total of 294 patients (a potential of 588 eyes) were registered to be seen by ODEMS in Dunedin Hospital for diabetic retinal photography between 1st and 31st October. Out of this, a total of 485 eyes were photographed and graded for both DR and DME. Fifty of these eyes (10.3%) were graded solely by the primary grader and 435 were graded by the

Table 1. Summary of Visiona's performance on Otago Diabetic Eye Monitoring Service (ODEMS) and Messidor for detecting referable diabetic retinopathy (DR)

Database	Area under ROC curve (95% CI)	DR score cut-off value	Sensitivity (%)	Specificity (%)
ODEMS	0.901 (0.807–0.995)	0.55	84.6	79.7
Messidor	0.980 (0.973–0.986)	1.89	96.0	90.0

CI, confidence interval; ROC, receiver operating characteristic.

primary grader and the ophthalmologist. A total of 103 potential eyes were not used for this study either due to patient non-attendance or poor quality of photos that were ungradable. The ability to satisfactorily grade photos as part of ODEMS was determined by the ophthalmologist.

Out of the 485 eyes, 13 eyes (2.7%) were graded by ODEMS as being referable for ophthalmology input in clinic (that is at least of moderate DR or moderate DME). As per the NZ MoH guideline, all referable fundal photos were also graded by the ophthalmologist.[4] The Messidor database contained 297 of 1200 images (24.8%) with referable DR or DME.

Visiona's ability to detect referable DR is summarized in Table 1 with ROC curve areas and best combination of sensitivities and specificities at respective DR score cut-off values. Figure 1 shows ROC curves for Visiona's performance on ODEMS and Messidor.

DISCUSSION

With the global burden of diabetes mellitus and hence DR projected to increase, there will be a growing demand for DR screening services in a world with limited resources. Our results have shown that a deep neural network can successfully detect referable DR with high areas under the ROC curves for both a domestic (ODEMS) and international database (Messidor). The area under the ROC curve with ODEMS images was 0.901 (95% confidence interval [CI], 0.807–0.995), while the area under the ROC curve for the Messidor was 0.980 (95% CI 0.973–0.986). The wider CI for the ODEMS is explained by the smaller proportion of referable DR and the smaller sample size compared with Messidor. There was no statistically significant difference at $P < 0.05$ between the areas under the two ROC curves.

The best combination of sensitivities and specificities with ODEMS (84.6% sensitivity, 79.7% specificity) and Messidor (96.0% sensitivity, 90.0% specificity) were at DR score cut-off values 0.55 and 1.89, respectively. Considering that a cut-off value of 1.50 or more is graded by Visiona as referable DR, there is a suggestion that ODEMS may be overcalling the DR grades. The impression from the ODEMS ophthalmologist is that there are artefacts on the ODEMS camera which may result in overcalling DR and DME grades.

Other groups have recently published studies using artificial intelligence for detecting referable DR. Abràmoff *et al.* used an algorithm to detect referable DR achieving sensitivity of 97% and specificity of 59%,[13] Solanki *et al.* used advanced machine learning with 94% sensitivity and 72% specificity to detect referable DR,[14] and Gulshan *et al.* used deep neural network with 87–90% sensitivity and 98% specificity for detecting referable DR.[5,6] Walton *et al.*, also used deep neural network to detect sight threatening diabetic eye disease with 66% sensitivity and 73% specificity.[15] The results of our study are comparable to the above studies, especially as Abràmoff *et al.*,[13] Solanki *et al.*[14] and Gulshan *et al.*[6] used Messidor 2 database for testing, an extension of the Messidor database that we used in this study.

As well as being comparable to other published studies, other strengths of this study is the lack of bias and confounding. There was also a large number of images from two different datasets, with generalizability to a domestic and an international setting. The main limitation is Visiona's inability to specifically detect DME. However, our definition of referable DR assigned photos with either significant (moderate grade or worse) DR or DME as abnormal and hence our ROC curve analysed Visiona's ability to detect both significant DR and DME through Visiona's DR score. Four of 13 referable eyes from ODEMS had significant DME without significant DR and 43 of 297 referable eyes from Messidor had significant DME without significant DR.

Our exclusion of ungradable photos due to poor quality is consistent with other studies,[6,15] and allowed us to produce comparable results. While ODEMS uses at least three fundal photos to grade each eye, Visiona used one photo per eye and this is consistent with other studies.[6,13–15] Moreover Lin *et al.* showed highly significant agreement (kappa = 0.97) for DR detection by human assessors between single monochromatic digital photograph and seven standard colour photos,[16] and hence we do not believe this to be a significant limitation. We also assumed each eye to be independent of one another like other studies[6,13,14]; however, this is not strictly true with a pair of eyes from the same person.

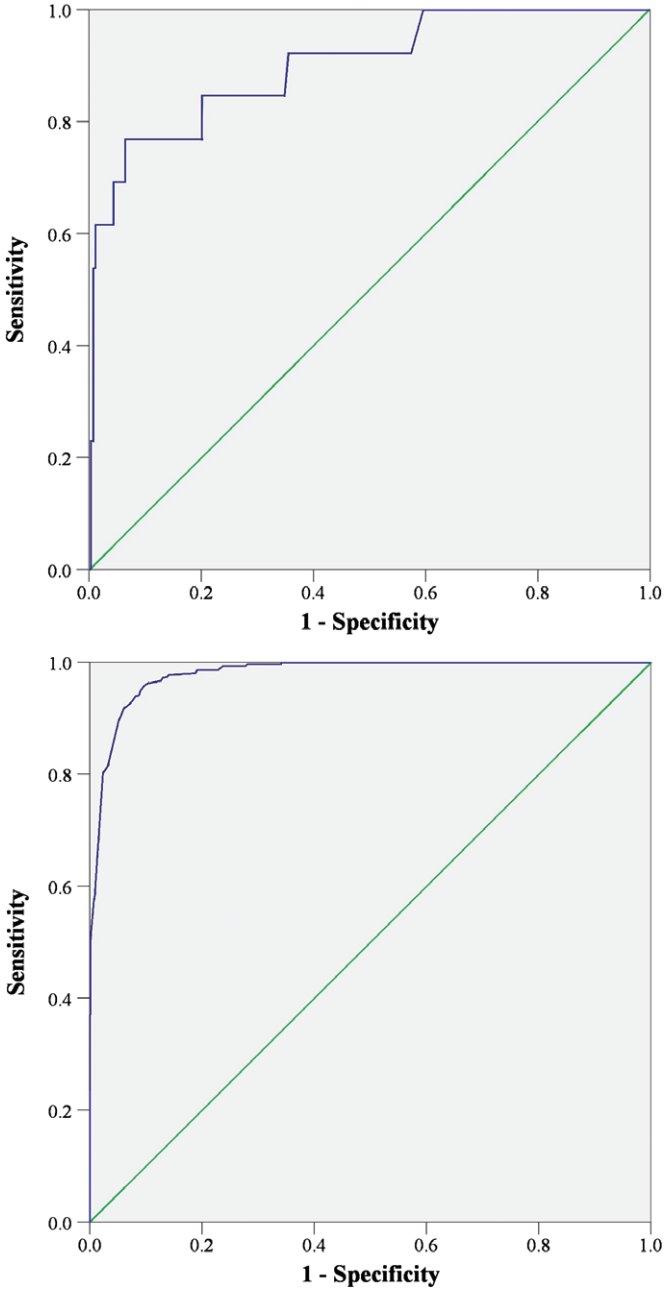

Figure 1. ROC curve for Visiona's performance on Otago Diabetic Eye Monitoring Service (ODEMS) (top) and Messidor (bottom). Blue line represents line corresponding to respective receiver operating characteristic curve. Green line represents reference line corresponding to no discriminatory power.

Visiona is also unable to identify incidental findings such as glaucoma, age related macular degeneration, retinal vein occlusion and choroidal tumours, which would otherwise be identified by a human assessor.

Screening guidelines typically recommend values of >80% sensitivity and specificity.[5] Visiona was able to successfully detect referable DR with accuracy close to (ODEMS: 84.6% sensitivity, 79.7% specificity) and well above recommended guidelines (Messidor: 96.0% sensitivity, 90.0% specificity). The major limitation was Visiona's inability to specifically detect DME. It is currently possible to use artificial intelligence to interpret optical coherence tomography scans to detect macular pathology.[17,18] We believe that deep neural networks could be integrated into community screening once they can successfully detect both DR and DME. The use of artificial intelligence in a nationwide screening programme would eliminate inter-grader and intra-grader variability, while at the same time liberating valuable health resources to be allocated elsewhere.

Acknowledgements

The authors extend thanks to Lyn Fuller for her advice on ODEMS and Andrew Gray for his help with statistics.

References

1. International Diabetes Federation. *IDF Diabetes Atlas*, 7th edn. Brussels: International Diabetes Federation, 2015. Accessed Jul 2017. Available from: http://www.diabetesatlas.org
2. Zheng Y, He M, Congdon N. The worldwide epidemic of diabetic retinopathy. *Indian J Ophthalmol* 2012; **60**: 428–31.
3. Ministry of Health. *Diabetic Retinal Screening, Grading, Monitoring and Referral Guidance*. Wellington: Ministry of Health, 2016 Accessed July 2017. Available from: http://www.health.govt.nz/publication/diabetic-retinal-screening-grading-monitoring-and-referral-guidance
4. Hutchins E, Coppell KJ, Morris A, Sanderson G. Diabetic retinopathy screening in New Zealand requires improvement: results from a multi-centre audit. *Aust N Z J Public Health* 2012; **36**: 257–62.
5. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *JAMA* 2016; **316**: 2366–7.
6. 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**: 2402–10.
7. Decencière E, Zhang X, Cazuguel G *et al*. Feedback on a publicly distributed image database: the Messidor database. *Image Anal Stereol* 2014; **33**: 231–4.
8. Canon. *CR-2 PLUS AF Digital Non-Mydriatic Retinal Camera*. Melville, NY: Canon, 2015. Accessed Jul 2017. Available from: https://www.usa.canon.com/internet/portal/us/home/products/details/eyecare/digital-non-mydriatic-retinal-cameras/cr-2-plus-af/
9. Messidor Program Partners. *Methods to Evaluate Segmentation and Indexing Techniques in the Field of Retinal Ophthalmology (MESSIDOR)*. France: Messidor Program Partners, 2016. Accessed Jul 2017. Available from: http://www.adcis.net/en/Download-Third-Party/Messidor.html
10. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic

color fundus photographs – an extension of the modified Airlie House classification: ETDRS report number 10. *Ophthalmology* 1991; **98**: 786–806.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44.
12. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr* 2007; **96**: 644–7.
13. Abràmoff MD, Folk JC, Han DP *et al*. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol* 2013; **131**: 351–7.
14. Solanki K, Ramachandra C, Bhat S, Bhaskaranand M, Nittala MG, Sadda SR. EyeArt: automated, high-throughput, image analysis for diabetic retinopathy screening. *Invest Ophthalmol Vis Sci* 2015; **56**: 1429.
15. Walton OB, Garoon RB, Weng CY *et al*. Evaluation of automated teleretinal screening program for diabetic retinopathy. *JAMA Ophthalmol* 2016; **134**: 204–9.
16. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol* 2002; **134**: 204–13.
17. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmol Retina* 2017; **1**: 322–7.
18. Khalid S, Akram MU, Hassan T, Nasim A, Jameel A. Fully automated robust system to detect retinal edema, central serous Chorioretinopathy, and age related macular degeneration from optical coherence tomography images. *Biomed Res Int* 2017; **2017**: 7148245. https://doi.org/10.1155/2017/7148245