The ROYAL COLLEGE of
OPHTHALMOLOGISTS

ARTICLE

Artificial intelligence-based screening for diabetic retinopathy at community hospital

Jie He[1] · Tingyi Cao[1] · Feiping Xu[1] · Shasha Wang[1] · Haiqi Tao[2] · Tao Wu[2] · Liyan Sun[2] · Jili Chen[1]

Abstract

Objectives The purpose of this study is to assess the accuracy of artificial intelligence (AI)-based screening for diabetic retinopathy (DR) and to explore the feasibility of applying AI-based technique to community hospital for DR screening.

Methods Nonmydriatic fundus photos were taken for 889 diabetic patients who were screened in community hospital clinic. According to DR international classification standards, ophthalmologists and AI identified and classified these fundus photos. The sensitivity and specificity of AI automatic grading were evaluated according to ophthalmologists' grading.

Results DR was detected by ophthalmologists in 143 (16.1%) participants and by AI in 145 (16.3%) participants. Among them, there were 101 (11.4%) participants diagnosed with referable diabetic retinopathy (RDR) by ophthalmologists and 103 (11.6%) by AI. The sensitivity, specificity and area under the curve (AUC) of AI for detecting DR were 90.79% (95% CI 86.4–94.1), 98.5% (95% CI 97.8–99.0) and 0.946 (95% CI 0.935–0.956), respectively. For detecting RDR, the sensitivity, specificity and AUC of AI were 91.18% (95% CI 86.4–94.7), 98.79% (95% CI 98.1–99.3) and 0.950 (95% CI 0.939–0.960), respectively.

Conclusion AI has high sensitivity and specificity in detecting DR and RDR, so it is feasible to carry out AI-based DR screening in outpatient clinic of community hospital.

Introduction

Diabetic retinopathy (DR), a common retinal microvascular complication of diabetes mellitus (DM), is a leading cause of vision loss in the working-age population globally [1]. With the aging of the global population and the expansion of the DM epidemic, the prevalence of DR will continue to escalate [2, 3]. The pooled prevalence rate of DR in the Chinese general population was 1.14% from 1990 to 2017; for patients with DM, the pooled prevalence rate of DR was 18.45% [4]. It is worth noting that the onset of DR is relatively insidious, and there are few symptoms. When

diabetic patients first present to the hospital with fundus morbidities, they are always in intermediate or late stages DR. Early detection is necessary for a good DR prognosis. It will be very beneficial for all diabetic patients to undergo DR screening, especially in developing countries.

DR screening is usually performed by ophthalmologists through fundus examination or fundus photography. Although fundus cameras have become popular in primary hospitals, there is a lack of experienced ophthalmologists and the equipment is always underutilized. To assist community doctors in their work, the public health system has established telehealth services for DR, uploading fundus photos to higher-level hospitals and waiting for the results of manual interpretation [5]. However, the ophthalmologists of higher-level hospitals often cannot complete these interpretations in time due to the heavy workload. In addition, manual interpretation has certain subjectivity. The generation of automated techniques for DR diagnosis is undoubtedly critical to solving these problems [6]. Deep learning-based artificial intelligence (AI) grading of DR is fast and has achieved high validation accuracies. Raju et al. [7] reported a sensitivity of 80.28% and a specificity of 92.29% for automatic diagnosis of DR on the publicly

---

These authors contributed equally: Jie He, Tingyi Cao

✉ Jili Chen
corneachen@163.com

1 Department of Ophthalmology, Shanghai Shibei Hospital of Jing'an District, Shanghai, China

2 Community Health Service Center, PengPu Town, Jing'an District, Shanghai, China

available Kaggle dataset. Despite these studies, there is a lack of relevant data on the clinical use of this technology [8, 9].

In light of this, we applied deep learning-based AI grading of DR to community hospital clinics. The aim of this study is to assess the accuracy of AI-based techniques for DR screening and explore the feasibility of implementing AI-based techniques in community hospital clinic.

Methods

Participants

Diabetes patients who attended PengPu Town Community Hospital of Jing'an district between May 30, 2018 and July 18, 2018 were invited to participate in this study. All participants were 18 years of age or older and provided written informed consent. This study was approved by the ethical committee of Shibei hospital, Jing'an district, Shanghai (ChiCTR1800016785).

Retinal images

Using an automatic nonmydriatic Topcon TRC-NW400 camera (Topcon, Tokyo, Japan), 45° colour retinal photographs were taken of each eye. All retinal images captured two fields, macula-centred and disc-centred, according to EURODIAB protocol [10]. AI equipment was installed and used in the community hospital. The fundus photographs of each participant were analysed by AI and transmitted to two ophthalmologists simultaneously. Participants who had unclear fundus photographs due to small pupils, cataracts or vitreous opacity were removed.

Human grading

All fundus photographs were graded independently by two ophthalmologists (retina specialists, Kappa $(\kappa) = 0.89954$) who were masked to each other and to AI device outputs. When the results between the two retina specialists were inconsistent, the third retina specialist would make a decision. The grading of retinopathy was evaluated according to the International Clinical Diabetic Retinopathy (ICDR) severity scale [11]. The classification of the five stages of DR are as follows: score 0, no apparent retinopathy, no abnormalities; score 1, mild nonproliferative DR (NPDR), microaneurysms only; score 2, moderate NPDR, more than just microaneurysms but less than severe NPDR; score 3, severe NPDR, one or more of the following: (i) more than 20 intraretinal haemorrhages in each of four quadrants, (ii) definite venous beading in two or more quadrants and (iii) prominent intraretinal microvascular abnormality in one or

more quadrants; score 4, proliferative DR (PDR), retinal neovascularization with or without vitreous/preretinal haemorrhage [8, 11]. Referable diabetic retinopathy (RDR) has been defined more than mild NPDR and/or macular oedema [12].

Automated grading

The AI device automated analysis and identified signs of retinal photographs with AI software (Airdoc, Beijing, China). Then, a DR screening report including referral recommendations was generated and delivered to the participant immediately. The core of the AI software is deep neural network. The deep neural network, a sequence of mathematical operations, is applied to certain inputs, such as images. The deep neural network, inception v4 architecture with input size 512×512 was used to build the classification network [13]. The network weights were pre-trained on the ImageNet dataset with 1.2 million images for 1000 categories classification and fine-tuned on the fundus image. This network was implemented by TensorFlow framework. In our experiment, the deep learning model is trained by using two Geforce GTX Titan X graphical processing units with CUDA version 9.0 cuDNN 7.0. The inputs of the network are the fundus images which are preprocessed by image mirror and image rotation. We augmented the training dataset via random rotating the image between $-15°$ and $+15°$. The output of the deep neural network is a vector, which indicates the category (five stages of DR) of the input image. Based on the values of the output vector from the heatmap, the AI software gives a DR stage prediction (Fig. 1).

Statistical analysis

Statistical analysis of the data was performed using the SPSS statistical package version 16.0. The performance of the AI algorithm was evaluated using the ophthalmologist grading as a reference standard. Kappa (κ) statistics were used to quantify and evaluate the degree of agreement between automated analysis and manual grading. The sensitivity and specificity of the AI algorithm for detecting DR were calculated.

Results

At the PengPu town Community Clinic, 889 diabetic patients agreed to participate in DR screening, including 418 men and 471 women. The average age of the participants was 68.46 ± 7.168 years old. A total of 3556 retinal images were obtained and graded in this study. According to the ICDR classification scale, 149 participants had DR in
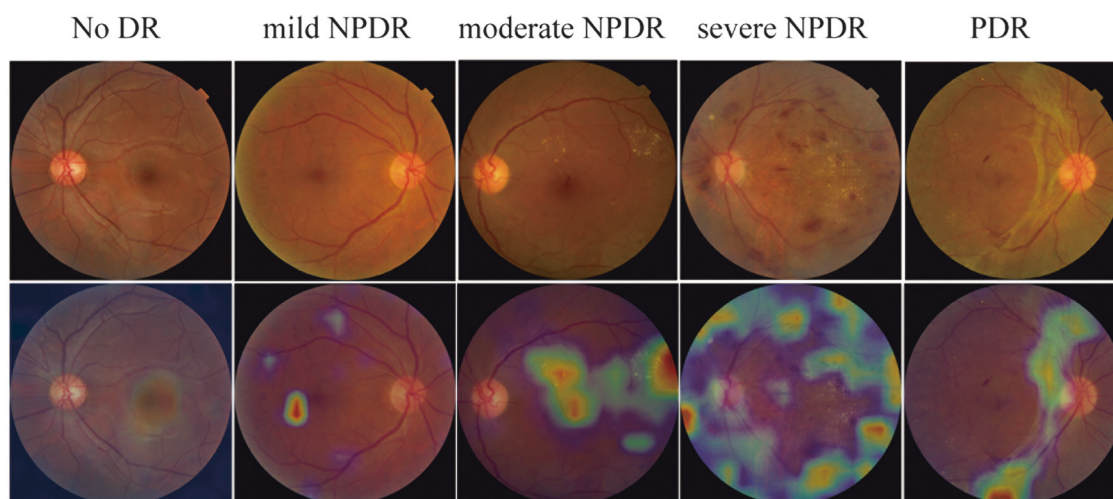
No DR   mild NPDR   moderate NPDR   severe NPDR   PDR

Fig. 1 Examples of the heatmap generated by AI on different severity levels of DR
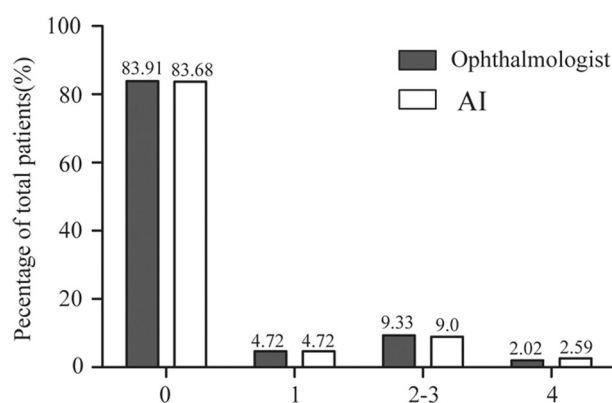
Fig. 2 Comparison of diabetic retinopathy (DR) grading between ophthalmologist and AI

Fig. 3 Venn diagram showed the overlap comparison of RDR between human and automated grading

at least one eye that was detected by the ophthalmologist or AI. Ophthalmologists detected DR in 143 (16.1%) participants, while AI detected DR in 145 (16.3%) participants. The proportion of different DR grades between the two interpretation modes is shown in Fig. 2. Most of the participants' fundus photographs reveal no DR. Even among participants with DR, most were scored as moderate NPDR (score 2) or severe NPDR (score 3). RDR was diagnosed in 101 (11.4%) participants based on manual grading and in 103 (11.6%) participants using the deep learning algorithm. The matched diagnosis of RDR between ophthalmologist and AI grading was observed in 91 participants (Fig. 3).

With ophthalmologist grading as the reference standard, the sensitivity and specificity of AI for detecting score 0, score 1, score 2–3, score 4, any DR and RDR are shown in Table 1. For detecting any DR, the sensitivity and specificity were 90.79% (95%, CI 86.4–94.1) and 98.5% (95%, CI 97.8–99.0), respectively. For detecting RDR, AI showed 91.18% (95%, CI 86.4–94.7) sensitivity and 98.79% (95%,
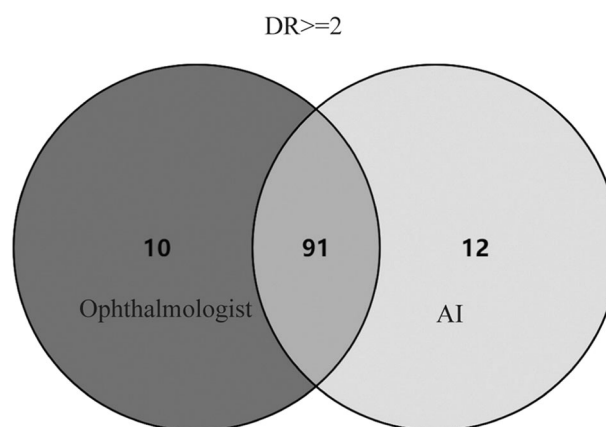
CI 98.1–99.3) specificity. The area under the curve (AUC) was 0.946 (95%, CI 0.935–0.956) when testing the ability of AI to detect any DR; for the detection of RDR, the AUC was 0.950 (95%, CI 0.939–0.960) (Table 1).

Discussion

This study assessed the accuracy of AI-based techniques for DR screening. We found that it was feasible to use an AI-based DR screening model in Chinese community clinic.

In our study, the prevalence of DR in DM patients was 16%, similar to the report by Song et al. They conducted a META analysis of 31 studies and found that the prevalence of DR in Chinese DM patients was 18.45% [4]. In India, the prevalence of DR in patients with type 2 diabetes was 17.6% [14]. In the US diabetes population, the incidence of any DR was 33.2% [15]. And Yau et al. estimated that the

Table 1 Sensitivity, specificity and AUC of AI for detection of any varying degrees of DR with ophthalmologist grading as reference standard

Stage	Sensitivity, %	(95% CI)	Specificity, %	(95% CI)	AUC	(95% CI)	p
0	98.70	98.0–99.2	92.28	88.2–95.3	0.955	0.944–0.964	<0.0001
1	83.33	70.7–92.1	99.54	99.1–99.8	0.914	0.900–0.927	<0.0001
2–3	91.38	86.2–95.1	99.00	98.4–99.4	0.952	0.941–0.961	<0.0001
4	80.36	67.6–89.8	99.42	98.9–99.7	0.899	0.884–0.913	<0.0001
Any DR	90.79	86.4–94.1	98.50	97.8–99.0	0.946	0.935–0.956	<0.0001
RDR	91.18	86.4–94.7	98.79	98.1–99.3	0.950	0.939–0.960	<0.0001

Score 0 = no apparent retinopathy, score 1 = mild nonproliferative DR (NPDR), score 2 = moderate NPDR, score 3 = severe NPDR, score 4 = proliferative DR (PDR)

Any DR diabetic retinopathy (any stage), *RDR* referable diabetic retinopathy

prevalence of any DR was 34.6% by summarizing 35 population-based study data from around the world [1]. Therefore, the difference in DR prevalence reported in different regions may be related to the research methods, the demographic characteristics, and the DR identification and classification.

The screening method for DR is constantly improving, but it still fails to meet the demand of the explosive growth of DM patients [16]. Verma et al. [17] attempted to reduce the workload of ophthalmologists in tertiary hospitals with the short-term training of nonophthalmologists and optometrists to examine the fundus under direct ophthalmoscopy. Later, the computerized "disease/no disease" scoring system was slightly better than the manual system [18]. In recent years, the emergence of deep learning algorithms has allowed for the quick and accurate identification of diabetic macular oedema (DMO) and the grading of DR [6]. This system can help not only in the early screening of DR but also in the long-term follow-up of DM patients [9].

Our statistical analysis showed that the sensitivity of DR screening using AI was high, reaching 90.79%, and the sensitivity of RDR screening using AI was 91.18%. Abràmoff et al. [19] showed the performance of a deep learning enhanced algorithm for the automated detection of RDR. The sensitivity of the algorithm was 96.8%, the specificity was 87.0% and the AUC was 0.980. Gargeya et al. [20] also reported a 94% sensitivity, a 98% specificity and an AUC of 0.970. However, our sensitivity is similar to the data reported by Ting et al. [21], with a sensitivity of 90.5%, a specificity of 91.6% and an AUC of 0.936. Because their research is closer to that of the clinic, it is conducted in community and clinical multiracial populations with diabetes. Moreover, the specificity of RDR in our study was 98.79%, which is slightly higher than that of previously reported cases. In our study, the detection sensitivity of proliferative DR was 80.36% and perhaps the relatively lower sensitivity in this group was relevant to the small sample size. Therefore, our system has the advantages in detecting "any retinopathy" and "no

retinopathy". The initial screening results will be further confirmed by ophthalmologists. Given the relatively low incidence of DR, it will effectively reduce the workload of ophthalmologists.

The advantage of this study is that AI equipment is studied in clinical scenarios. Our results and previous reports [8, 9] indicate that AI-based DR screening for outpatients seems to be feasible. Firstly, a community hospital clinic can provide an ideal environment for capturing individuals with DR in DM patients who may not routinely participate in an eye examination. What's more, the patient's pupils do not need to be dilated when taking a picture of the fundus, which is more easily accepted by the patient. Finally, the report is quickly released on the spot, and the doctor in community hospitals can recommend referral to the upper level hospital ophthalmologist according to the condition.

Undoubtedly, there are also some limitations of this study. The sample size is relatively small. For fundus photographs in which the pupil is small or the refractive interstitial is turbid, the quality of some image is poor [16]. For AI, it is difficult to accurately diagnose macular oedema based solely on the fundus photos. The classification of DMO will be an important improvement in the AI system in the future.

In summary, AI has a high sensitivity and specificity for identifying DR in Chinese community clinic. This approach seems to be feasible. Further research is needed to assess compliance with referral rates and the effectiveness of DR vision improvement.

Summary

What was known before

- Although deep learning-based artificial intelligence (AI) grading of DR is fast and has achieved high validation accuracies, there is a lack of relevant data on the clinical use of this technology.

What this study adds

- This research is a supplement to the data on the clinical use of this technology.

Compliance with ethical standards

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

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

References

1. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35:556–64.
2. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010;87:4–14.
3. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract. 2014;103:137–49.
4. Song P, Yu J, Chan KY, Theodoratou E, Rudan I. Prevalence, risk factors and burden of diabetic retinopathy in China: a systematic review and meta-analysis. J Glob Health. 2018;8:010803.
5. Peng J, Zou H, Wang W, Fu J, Shen B, Bai X, et al. Implementation and first-year screening results of an ocular telehealth system for diabetic retinopathy in China. BMC Health Serv Res. 2011;11:250.
6. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, 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. Raju M, Pagidimarri V, Barreto R, Kadam A, Kasivajjala V, Aswath A. Development of a deep learning algorithm for automatic diagnosis of diabetic retinopathy. Stud Health Technol Inf. 2017;245:559–63.
8. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye. 2018;32:1138–44.
9. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep. 2018;8:4330.
10. Aldington SJ, Kohner EM, Meuer S, Klein R, Sjolie AK. Methodology for retinal photography and assessment of diabetic retinopathy: the EURODIAB IDDM complications study. Diabetologia. 1995;38:437–44.
11. Wilkinson CP, Ferris FL 3rd, Klein RE, Lee PP, Agardh CD, Davis M, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003;110:1677–82.
12. Abramoff MD, Folk JC, Han DP, Walker JD, Williams DF, Russell SR, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol. 2013;131:351–7.
13. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. 2016;arXiv preprint arXiv:1602.07261.
14. Rema M, Premkumar S, Anitha B, Deepa R, Pradeepa R, Mohan V. Prevalence of diabetic retinopathy in urban India: the Chennai Urban Rural Epidemiology Study (CURES) eye study, I. Investig Ophthalmol Vis Sci. 2005;46:2328–33.
15. Wong TY, Klein R, Islam FM, Cotch MF, Folsom AR, Klein BE, et al. Diabetic retinopathy in a multi-ethnic cohort in the United States. Am J Ophthalmol. 2006;141:446–55.
16. Sinclair SH. Diabetic retinopathy: the unmet needs for screening and a review of potential solutions. Expert Rev Med Devices. 2006;3:301–13.
17. Verma L, Prakash G, Tewari HK, Gupta SK, Murthy GV, Sharma N. Screening for diabetic retinopathy by non-ophthalmologists: an effective public health tool. Acta Ophthalmol Scand. 2003;81:373–7.
18. Philip S, Fleming AD, Goatman KA, Fonseca S, McNamee P, Scotland GS, et al. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. Br J Ophthalmol. 2007;91:1512–7.
19. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investig Ophthalmol Vis Sci. 2016;57:5200–6.
20. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology. 2017;124:962–9.
21. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318:2211–23.