

Artificial intelligence to detect referable diabetic retinopathy in a Chinese population in Hong Kong

Sunny Chi Lik Au^{1,2}, MBChB, MRCSEd, FCOphthHK; George Tsz Ho Shum³, MBBS, MRCP(UK); Steffi Shing Yee Chong^{1,2}, MBBS; Callie Ka Li Ko^{1,2}, MBBS, FCOphthHK, FHKAM (Ophthalmology)

¹Department of Ophthalmology, Tung Wah Eastern Hospital, Hong Kong SAR, China

²Department of Ophthalmology, Pamela Youde Nethersole Eastern Hospital, Hong Kong SAR, China

³Department of Medicine, Pamela Youde Nethersole Eastern Hospital, Hong Kong SAR, China

Correspondence and reprint requests:

Dr Sunny Chi Lik Au, 9/F, MO Office, Tung Wah Group of Hospitals Lo Ka Chow Memorial Ophthalmic Centre, Tung Wah Eastern Hospital, 19 Eastern Hospital Road, Causeway Bay, Hong Kong SAR, China. Email: kilihcua@gmail.com

Abstract

Objective: This study aimed to investigate the sensitivity and specificity of the Deep Fundus, an artificial intelligence program for diabetic retinopathy screening, in Chinese patients with diabetes.

Methods: Fundus photographs of Chinese patients with diabetes aged 50 to 90 years who attended Pamela Youde Nethersole Eastern Hospital, Hong Kong between December 2020 and June 2021 during the public diabetic retinopathy screening period were retrospectively retrieved in February 2022. Two ophthalmologists independently graded the fundus photographs for diabetic retinopathy (according to the recommendations in the 2017 International Council of Ophthalmology Guidelines for Diabetic Eye Care) and any concomitant eye pathology such as cataracts or macular diseases. On 27 February 2022, the photographs were uploaded to the Deep Fundus website. Fundus images were classified into referable or non-referable diabetic retinopathy in accordance with the abovementioned Guidelines. Sensitivity and specificity of the Deep Fundus were determined.

Results: Of 202 eyes screened, 96 eyes were included in the analysis. There was no discordance between the two ophthalmologists in the diabetic retinopathy gradings. 57 (59.4%) eyes were graded to have positive diabetic retinopathy findings. Using gradings of the two

ophthalmologists as the gold standard, sensitivity and specificity of the Deep Fundus were 62.3% and 90.7%, respectively. The positive and negative predictive values were 89.2% and 66.1%, respectively.

Conclusions: The sensitivity and specificity of the Deep Fundus for diabetic retinopathy screening in a Chinese population were 62.3% and 90.7%, respectively. Further real-world validation studies are needed to assess the artificial intelligence model, particularly among Chinese patients with cataracts, myopia, diabetic macular edema, or proliferative diabetic retinopathy.

Key words: Artificial intelligence; Diabetic retinopathy; East Asian people; Machine learning; Sensitivity and specificity

Introduction

Big data and machine learning are transforming healthcare. Artificial intelligence (AI) and deep learning have trained a wide range of computer models for medical triage and diagnostic testing.¹ AI has comparable performance to medical personnel in certain fields and may offer solutions to overburdened healthcare services resulting from an aging population and a shortage of healthcare personnel.^{2,3}

Among commercially available AI products for ophthalmology,^{4,5} analysis of fundus photographs⁶ and optical coherence tomography images⁷ remains the core technologies. Two AI systems for diabetic retinopathy

(DR) screening have been approved by the United States Food and Drug Administration for marketing.⁸ AI screening for age-related macular degeneration and glaucoma are developing as well.⁶

The International Diabetes Federation estimated that the global diabetic prevalence was 9.3% (equaling 463 million people) in 2019.⁹ The concept of DR screening began in 1989 with the St Vincent Declaration, which aimed to reduce diabetes mellitus (DM)-related blindness by one-third in 5 years.¹⁰ DR screening aims to detect important, treatable pathology in its early stages with a low-cost, efficient, and non-invasive diagnostic test.¹¹ In Hong Kong, DR screening is performed by primary care physicians, endocrinologists, optometrists, or specially-trained graders with non-stereoscopic digital color retinal fundus photographs.¹¹ Most AI models were validated in Caucasian populations, but the algorithm may differ in Chinese populations. For instance, Chinese populations have a higher prevalence of myopia than Caucasian populations. Therefore, this study aimed to investigate the sensitivity and specificity of the Deep Fundus, an AI program for DR screening, in Chinese patients with diabetes.

Materials and methods

Fundus photographs of patients with diabetes who attended a public DR screening at Pamela Youde Nethersole Eastern Hospital, Hong Kong, between December 2020 and June 2021 were retrospectively retrieved in February 2022, with patient identifiers removed. All fundus photographs were taken by the Non-Mydriatic Auto Fundus Camera AFC-330 (Nidek, Gamagori, Japan). Two photographs were taken per eye: one centered on the macula and the other centered on the optic disc.¹² Photographs of all sorts of quality were included to reflect the real-world situation of dense cataracts or different media opacities. Patients aged <50 years and >90 years were excluded to comply with the age range defined by the Deep Fundus.¹³ Non-Chinese patients were excluded, as were patients with branch retinal vein occlusion, age-related macular degeneration, retinal macroaneurysm, or missing fundus photographs.

Two ophthalmologists specializing in vitreoretinal diseases who were blinded from patient data independently graded the fundus photographs for DR (according to the recommendations in the 2017 International Council of Ophthalmology Guidelines for Diabetic Eye Care)¹⁴ and any concomitant eye pathology such as cataracts or macular diseases.

The Deep Fundus is accessible at www.deepfundus.com and is trained in the Eyetelligence (also available online),^{13,15} which has been approved as a Class I medical device for clinical use in the European Union, United Kingdom, Australia, and New Zealand.¹⁶⁻¹⁹ On 27 February 2022, the photographs were uploaded to the Deep Fundus website. Items investigated included valid fundus image, laterality, referable DR, DR severity scale, diabetic macular edema,

and cataracts. Only eyes with valid fundus image and laterality were included in the sensitivity and specificity analysis. Fundus images were classified into referable or non-referable DR in accordance with the recommendations in the 2017 International Council of Ophthalmology Guidelines for Diabetic Eye Care.¹⁴

The sample size needed to determine the model's accuracy was calculated based on its sensitivity and specificity.^{20,21} We searched PubMed, MEDLINE, Embase, and Scopus but found no previous publication on the performance of Deep Fundus on DR screening; therefore, we assumed the sensitivity and specificity to be 95% as claimed by the Deep Fundus. We first calculated the true positive (TP) + false negative (FN) for sensitivity and the true negative (TN) + false positive (FP) as follows:

$$TP+FN=Z^2 \times [\text{sensitivity} \times (1-\text{sensitivity})/W^2]$$

$$TN+FP=Z^2 \times [\text{specificity} \times (1-\text{specificity})/W^2]$$

where Z is set to 1.96 for the normal distribution value, corresponding to the 95% confidence interval (CI), and W, the maximum acceptable width of the 95% CI, is set to 10%.²⁰ The sample size calculations for sensitivity and specificity were $TP+FN/P$ and $TN+FP/(1-P)$, respectively, in which P is the prevalence of DR in the Chinese population in Hong Kong. In a previous study involving 174 532 patients, the estimated prevalence of DR was 39%.¹¹ Thus, 47 and 30 eyes were required for determining sensitivity and specificity, respectively, with a total of 77. Statistical analyses were performed using SPSS (Windows version 27.0; IBM Corp, Armonk [NY], USA).

Results

Of 202 eyes screened during the public DR screening period, 106 eyes were excluded owing to missing fundus photographs (n=1), non-Chinese ethnicity (n=14), or age <50 years (n=91), whereas 96 eyes in patients aged 50 to 77 (mean, 58.3±7.8) years were included in the analysis. The male-to-female ratio was 1.29:1. All eyes were graded correctly on laterality by the Deep Fundus.

There was no discordance between the two ophthalmologists in the DR gradings. 12 (12.5%) eyes were determined to be myopic according to the grading system by the meta-analysis for pathologic myopia study group.²² 57 (59.4%) eyes were graded to have positive DR findings (**Table 1**). Five (5.2%) eyes had a history of pan-retinal photocoagulation laser treatment. No eye had vitreous, pre-retinal, or subhyaloid hemorrhages. One (1.0%) eye had tractional retinal detachment. Nine eyes with proliferative DR had neovascularization at disc or elsewhere. Fundus photograph quality was suboptimal in 26 (27.1%) eyes owing to cataracts. When macular details and the entire posterior pole of the fundus were obscured, efforts were made to grade the periphery. Nine (9.4%) eyes had macular diseases other than diabetic macular edema including dry age-related macular degeneration, myopic maculopathy, choroidal naevus, chorioretinal scars, and retinal artery macroaneurysms.

Table 1. Diabetic retinopathy (DR) characteristics of eyes according to the recommendations in the 2017 International Council of Ophthalmology Guidelines for Diabetic Eye Care (n=96)

Classification	No. of patients	Referable
No apparent DR	39	No
Mild non-proliferative DR	4	No
Moderate non-proliferative DR	33	Yes
Severe non-proliferative DR	3	Yes
Proliferative DR	9	Yes
Diabetic macular edema	9	Yes
Ungradable (treated proliferative DR or unclear fundus photo)	8	Yes

Table 2. Referable diabetic retinopathy (DR) grading according to the Deep Fundus and ophthalmologists

Referable DR grading by the Deep Fundus	Referable DR grading by ophthalmologists	
	Yes	No
Yes	33	4
No	20	39

Numbers of referable DR according to the Deep Fundus and the two ophthalmologists are shown in **Table 2**. Using gradings of the two ophthalmologists as the gold standard, sensitivity and specificity of the Deep Fundus were 62.3% (95% CI=47.9%-75.2%) and 90.7% (95% CI=77.9%-97.4%), respectively. The positive and negative predictive values were 89.2% (95% CI=76.0%-95.6%) and 66.1% (95% CI=57.7%-73.6%), respectively. **Figure 1** shows fundus photographs of true positive and true negative grading, whereas **Figure 2** shows fundus photographs of false positive and false negative grading.

Discussion

The Deep Fundus has a high specificity but a low sensitivity for detecting referable DR in Chinese patients with diabetes aged 50 to 90 years. To the best of our knowledge, this study is the first to investigate the clinical application of the Deep Fundus in Chinese populations. We used real-world data from a 7-month public screening service. The number of eyes studied (n=96) was well above the minimum required sample size. All diagnostic results from the Deep Fundus were generated on the same day (27 February 2022); this may eliminate potential bias from the ongoing update or learning of the AI model. The Deep Fundus is only valid for patients aged 50 to 90 years; 45.0% (n=91) of eyes in patients aged <50 years were excluded; some of these patients had type 1 DM and/or gestational DM.

Disease prevalence affects the sample size calculation for determining the sensitivity and specificity of a diagnostic test.^{20,21} We used the DR prevalence of 39% in primary care settings¹¹ to calculate the sample size, but the present study had a higher prevalence of DR at 59.4%. This is expected among patients in ophthalmology and internal medicine departments of a tertiary-level hospital. Consequently, DM cases were more complicated and severe than those typically encountered in primary care settings. Using the prevalence of 59.4% in the sample size calculation, the numbers of eyes required for sensitivity and specificity would be 31 and 45, respectively, and the minimum sample size would be 76. Therefore, our study is powered to demonstrate the real-world sensitivity and specificity of the Deep Fundus.

DM is a risk factor for the development of cataracts, and the prevalence of cataracts in DM patients ranges from 20% to 66%.^{23,24} In our study, 27.1% of our DM patients had a visually significant cataract affecting the quality of their fundus photographs. However, the sample size of patients with cataracts was too small for subgroup analysis. Possible confounding factors include the variable severity of cataracts and the retrospective study design, which could introduce

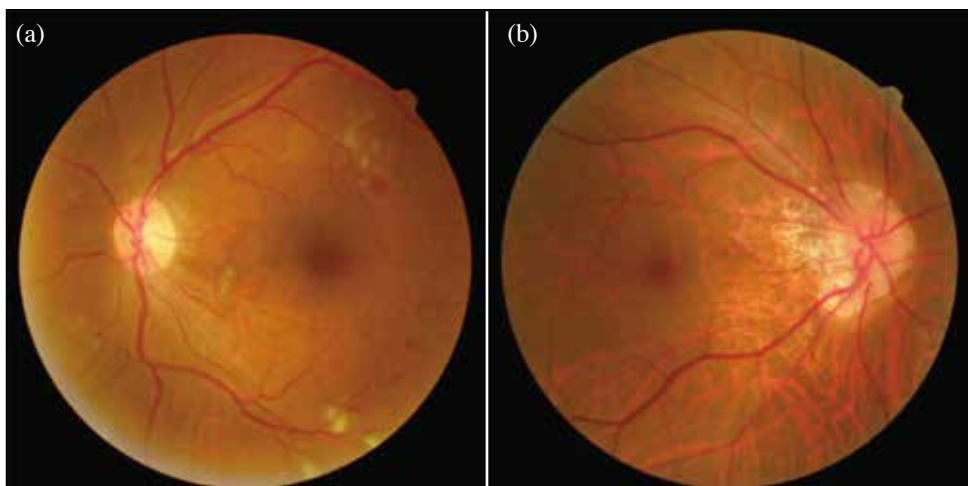


Figure 1. Fundus photographs showing (a) true positive and (b) true negative referable diabetic retinopathy grading by the Deep Fundus, which is graded as myopia by ophthalmologists.

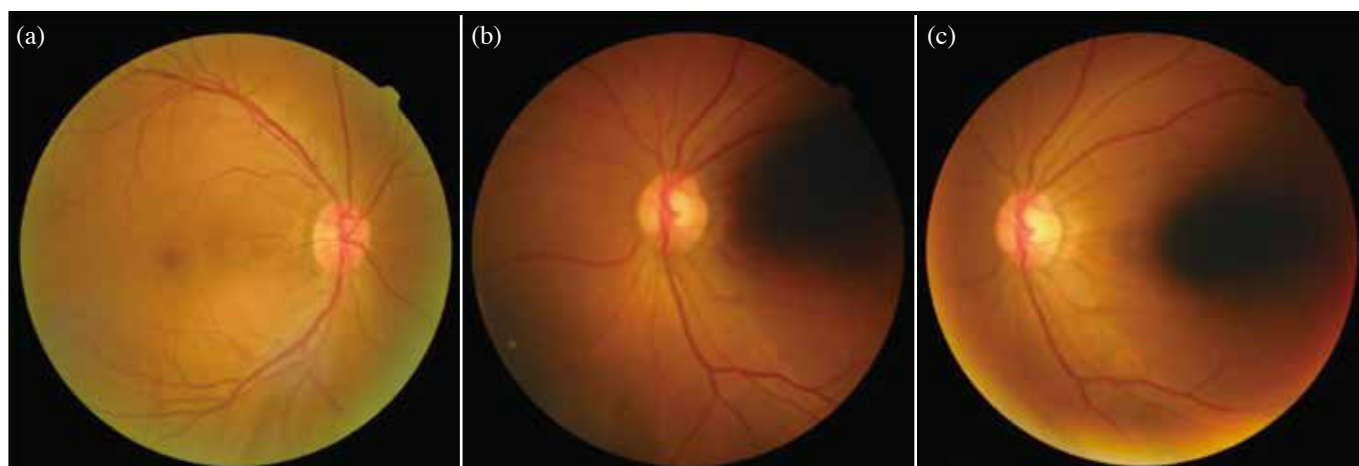


Figure 2. (a) A fundus photograph showing false positive referable diabetic retinopathy grading by the Deep Fundus. Only microaneurysms are seen, with no macular edema. This is graded as mild non-proliferative diabetic retinopathy by ophthalmologists but as moderate non-proliferative diabetic retinopathy by the Deep Fundus. (b and c) Fundus photographs showing a false negative referable diabetic retinopathy grading by the Deep Fundus. Cataracts obscure the macula; this patient should be referred according to the ophthalmologists.

selection bias. The quality of the fundus images can be a confounding factor for the Deep Fundus performance. We included photographs of varying quality to reflect the real-world situation of cataracts or different causes of media opacities. However, the prevalence and severity of media opacities differ across different populations with different demographics.²⁵ This variation would affect the sensitivity and specificity of the Deep Fundus in clinical practice.

The United States Food and Drug Administration has approved two AI systems for DR screening: the IDx-DR (Digital Diagnostics, Coralville [IA], USA) and EyeArt (Eyenuk, Woodland Hills [CA], USA), both of which have sensitivity and specificity approaching 90% as demonstrated in large-scale studies.²⁶⁻²⁸ Other AI systems to assess DR are also available.²⁹ Retmarker DR (Retmarker, Coimbra, Portugal) has been certified as a Class IIa medical device in Europe.³⁰ In China, AI-based software for DR includes Airdoc (Beijing, China).³¹ When adopting AI systems in clinical settings in Hong Kong, the algorithms should ideally be trained by photographs of Asian/Chinese eyes. The low sensitivity (62.3%) of the Deep Fundus suggests a considerable number of false negative cases of referable DR, which might progress rapidly to more severe grades of DR that are not timely detected by annual screening. Therefore, we suggest manually reviewing all negative cases graded by Deep Fundus to avoid missing the golden period of referral to ophthalmologists for monitoring. Nonetheless, if there are AI models with higher sensitivity, a semi-automated model can be considered in Hong Kong, in which retinal photographs initially analyzed by an AI model to filter out non-referable DR cases, followed by grading by ophthalmologists. Such models have been used in the UK and Singapore.³²

Refraction status can change after corneal refractive or cataract surgery. Some of our patients have pseudophakia, and the axial length measurement was not available. Thus, fundus photographs were used to determine myopia. According to the meta-analysis for pathologic myopia study group,²² visible signs of myopia on fundus photographs include a tessellated fundus, diffuse chorioretinal atrophy, patchy chorioretinal atrophy, macular atrophy, lacquer cracks, myopic choroidal neovascularization, and Fuchs spots. Some of these may mimic DR. For example, myopic choroidal neovascularization or lacquer cracks can be misinterpreted as DR hemorrhages, given that all present as red dots. Similarly, chorioretinal atrophy can be misinterpreted as exudate given that both are white in color. However, the prevalence of myopia in our patients was 12.5%, which is lower than the 41.1% reported in other studies in Asia.^{33,34} Subgroup analysis of myopic patients was not possible because of the small sample size (n=12).

Conclusion

The sensitivity and specificity of the Deep Fundus for DR screening in a Chinese population were 62.3% and 90.7%, respectively. Further real-world validation studies are needed to assess the AI model, particularly among Chinese patients with cataracts, myopia, diabetic macular edema, or proliferative DR.

Contributors

SCLA designed the study; all authors acquired the data; SCLA and GTHS analyzed the data; SCLA drafted the manuscript; and SCLA critically revised the manuscript for important intellectual content. All authors had full access to

the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

Conflicts of interest

As an editor of the journal, SCLA was not involved in the peer review process. Other authors have disclosed no conflicts of interest.

Funding/support

This study received no specific grant from any funding

agency in the public, commercial, or not-for-profit sectors.

Data availability

All data generated or analyzed during the present study are available from the corresponding author on reasonable request.

Ethics approval

The study was approved by Hong Kong East Cluster Research Ethics Committee (Reference: HKEC-REC-2022-019). The patients were treated in accordance with the tenets of the Declaration of Helsinki.

References

1. Marchiori C, Dykeman D, Girardi I, et al. Artificial intelligence decision support for medical triage. *AMIA Annu Symp Proc* 2021;2020:793-802.
2. Baker A, Perov Y, Middleton K, et al. A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. *Front Artif Intell* 2020;3:543405. [Crossref](#)
3. Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. *Int J Environ Res Public Health* 2021;18:271. [Crossref](#)
4. Bali J, Bali O. Artificial intelligence in ophthalmology and healthcare: an updated review of the techniques in use. *Indian J Ophthalmol* 2021;69:8-13. [Crossref](#)
5. Goldhagen BE, Al-Kharsan H. Diving deep into deep learning: an update on artificial intelligence in retina. *Curr Ophthalmol Rep* 2020;8:121-8. [Crossref](#)
6. Zapata MA, Royo-Fibla D, Font O, et al. Artificial intelligence to identify retinal fundus images, quality validation, laterality evaluation, macular degeneration, and suspected glaucoma. *Clin Ophthalmol* 2020;14:419-29. [Crossref](#)
7. Kapoor R, Whigham BT, Al-Aswad LA. Artificial intelligence and optical coherence tomography imaging. *Asia Pac J Ophthalmol (Phila)* 2019;8:187-94.
8. Lim G, Bellemo V, Xie Y, Lee XQ, Yip MYT, Ting DSW. Different fundus imaging modalities and technical factors in AI screening for diabetic retinopathy: a review. *Eye Vis (Lond)* 2020;7:21. [Crossref](#)
9. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract* 2019;157:107843. [Crossref](#)
10. Diabetes care and research in Europe: the Saint Vincent declaration. *Diabet Med* 1990;7:360. [Crossref](#)
11. Gangwani RA, Lian JX, McGhee SM, Wong D, Li KK. Diabetic retinopathy screening: global and local perspective. *Hong Kong Med J* 2016;22:486-95. [Crossref](#)
12. Gangwani R, Lai WW, Sum R, et al. The incidental findings of age-related macular degeneration during diabetic retinopathy screening. *Graefes Arch Clin Exp Ophthalmol* 2014;252:723-9. [Crossref](#)
13. Deep Fundus. Available from: <https://www.deepfundus.com/>
14. Wong TY, Sun J, Kawasaki R, et al. Guidelines on diabetic eye care: the International Council of Ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings. *Ophthalmology* 2018;125:1608-22. [Crossref](#)
15. Eyetelligence. Available from: <https://eyetelligence.ai/us/>.
16. Kiburg KV, Turner A, He M. Telemedicine and delivery of ophthalmic care in rural and remote communities: drawing from Australian experience. *Clin Exp Ophthalmol* 2022;50:793-800. [Crossref](#)
17. Keel S, Li Z, Scheetz J, et al. Development and validation of a deep-learning algorithm for the detection of neovascular age-related macular degeneration from colour fundus photographs. *Clin Exp Ophthalmol* 2019;47:1009-18. [Crossref](#)
18. 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. [Crossref](#)
19. Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care* 2018;41:2509-16. [Crossref](#)
20. Negida A, Fahim NK, Negida Y. Sample size calculation guide - Part 4: how to calculate the sample size for a diagnostic test accuracy study based on sensitivity, specificity, and the area under the ROC curve. *Adv J Emerg Med* 2019;3:e33.
21. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996;3:895-900. [Crossref](#)
22. Ohno-Matsui K, Kawasaki R, Jonas JB, et al. International photographic classification and grading system for myopic maculopathy. *Am J Ophthalmol* 2015;159:877-83.e7. [Crossref](#)
23. Pék A, Szabó D, Sándor GL, et al. Relationship between diabetes mellitus and cataract in Hungary. *Int J Ophthalmol* 2020;13:788-93. [Crossref](#)
24. Becker C, Schneider C, Aballéa S, et al. Cataract in patients with diabetes mellitus-incidence rates in the UK and risk factors. *Eye (Lond)* 2018;32:1028-35. [Crossref](#)
25. Lee AY, Yanagihara RT, Lee CS, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care* 2021;44:1168-75. [Crossref](#)
26. van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol* 2018;96:63-8. [Crossref](#)

27. Bhaskaranand M, Ramachandra C, Bhat S, et al. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technol Ther* 2019;21:635-43. [Crossref](#)
28. Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referable and vision-threatening diabetic retinopathy. *JAMA Netw Open* 2021;4:e2134254. [Crossref](#)
29. Li S, Zhao R, Zou H. Artificial intelligence for diabetic retinopathy. *Chin Med J (Engl)* 2021;135:253-60. [Crossref](#)
30. Tufail A, Kapetanakis VV, Salas-Vega S, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess* 2016;20:1-72. [Crossref](#)
31. Gu C, Wang Y, Jiang Y, et al. Application of artificial intelligence system for screening multiple fundus diseases in Chinese primary healthcare settings: a real-world, multicentre and cross-sectional study of 4795 cases. *Br J Ophthalmol* 2023;322940. [Crossref](#)
32. Cheung CY, Tang F, Ting DSW, Tan GSW, Wong TY. Artificial intelligence in diabetic eye disease screening. *Asia Pac J Ophthalmol (Phila)* 2019;8:158-64.
33. Mak CY, Yam JC, Chen LJ, Lee SM, Young AL. Epidemiology of myopia and prevention of myopia progression in children in East Asia: a review. *Hong Kong Med J* 2018;24:602-9. [Crossref](#)
34. Van Newkirk MR. The Hong Kong vision study: a pilot assessment of visual impairment in adults. *Trans Am Ophthalmol Soc* 1997;95:715-49.