ORIGINAL ARTICLE

Accuracy of the EyeArt artificial intelligence system in detecting referable diabetic retinopathy in patients with diabetes in Hong Kong

Sunny Chi Lik Au[1,2], *MBChB, MRCSEd, FCOphthHK, FHKAM (Ophthalmology); George Tsz Ho Shum*[3], *MBBS, MRCP(UK), FHKCP; Steffi Shing Yee Chong*[1,2], *MBBS, MRCSEd; Callie Ka Li Ko*[1,2], *MBBS, FCOphthHK, FHKAM (Ophthalmology)*

[1]*Department of Ophthalmology, Pamela Youde Nethersole Eastern Hospital, Hong Kong SAR, China*
[2]*Department of Ophthalmology, Tung Wah Eastern Hospital, Hong Kong SAR, China*
[3]*Department of Medicine, Pamela Youde Nethersole Eastern Hospital, Hong Kong SAR, China*

Correspondence and reprint requests:
Sunny Chi Lik Au, 9/F, MO Office, 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 evaluated the real-world performance in an Asian population with diabetes mellitus (DM) of the EyeArt system, which is the only United States Food and Drug Administration–approved system available in Hong Kong.

Methods: Fundus photographs taken from December 2020 to June 2021 during a diabetic retinopathy (DR) screening program in patients with DM who attended Pamela Youde Nethersole Eastern Hospital in Hong Kong were retrieved. Two vitreoretinal-trained ophthalmologists independently graded the fundus photographs for DR and concomitant pathologies. Discrepancies were resolved through discussion with a senior vitreoretinal ophthalmologist for a final decision. Fundus photographs (in jpg format) were then uploaded to the EyeArt AI system. The EyeArt system's sensitivity and specificity on detecting (1) referable DR (moderate or severe non-proliferative DR, proliferative DR, or diabetic macular edema [DME]) and (2) any DR or DME were calculated based on the ophthalmologists' diagnoses (the gold standard).

Results: 23 men and 79 women (93 Chinese, 3 Pakistanis, 2 Thais, and 4 Indonesians) aged 25 to 77 years with type 1 (n=8) or type 2 (n=94) DM were included for analysis. The mean duration of DM was 4.9 years; 56 (54.9%) patients were newly diagnosed with DM. 33 (32.4%) patients were taking ≥1 (up to 4) oral hypoglycemic agents; none were taking ≥5. 16 (15.7%) patients required both oral hypoglycemic agents and insulin for DM control; 31 (30.4%) patients injected insulin every day. As diagnosed by the ophthalmologists, 41 (20.1%) of 204 eyes had DR and/or DME, whereas 26 (25.5%) of 102 patients had referable DR. There was no discrepancy between the two ophthalmologists' gradings. On detecting referable DR, the EyeArt system had 96.2% sensitivity, 94.7% specificity, 86.2% positive predictive value, 98.6% negative predictive value, and 95.1% accuracy. On detecting any DR or DME, the EyeArt system had 92.7% sensitivity, 93.7% specificity, 79.2% positive predictive value, 98.0% negative predictive value, and 93.5% accuracy.

Conclusion: The EyeArt system is a reliable AI tool for DR screening in patients with DM in Hong Kong. Its usage in primary care and public healthcare programs should be considered.

Key words: Artificial intelligence; Diabetic retinopathy; Macular edema

Introduction

Artificial intelligence (AI) models intelligent behavior with minimal human intervention.[1] AI is increasingly used in clinical practice to improve healthcare services by expediting processes with greater accuracy.[2] Its applications include informatics approaches via machine learning and deep learning in clinical information management.[3] Image analysis is one such technology.[4]

In Hong Kong, diabetes mellitus (DM) is increasingly prevalent; DM and its related morbidities are important primary healthcare issues.[5] Diabetic retinopathy (DR) is the most common complication of DM.[6] Among patients with DM in Hong Kong, the prevalence of DR is 39%.[7] Screening for DR and timely treatment are essential to avoid vision loss, particularly for working-age patients with DM.[7]

Automated DR-detection algorithms have been developed to stratify patients' risks in terms of vision-threatening DR and diabetic macular edema (DME).[8,9] AI can facilitate the integration of community, primary, and specialist eye care services, optimize the flow of patients within healthcare networks, and improve the efficiency of DR management.[10] Nonetheless, applying AI in real-world clinical settings is challenging.

As of 2022, the only United States Food and Drug Administration (FDA)–approved DR screening system available in Hong Kong was the EyeArt AI system (Eyenuk, Los Angeles [CA], USA). There are few extant studies examining the use of EyeArt in Asian populations.[11,12] Retinas in Asians are believed to differ from those in Caucasians, given the much higher prevalence of myopia among Asians.[13] This study aims to evaluate the accuracy of the EyeArt in detecting referable DR in patients with DM in Hong Kong.

Materials and methods

In April 2022, fundus photographs from patients with DM who attended Pamela Youde Nethersole Eastern Hospital in Hong Kong were retrieved after removing all patient identifiers. All the fundus photographs were taken by the non-mydriatic Auto Fundus Camera 330 (Nidek, Gamagori, Japan) between December 2020 and June 2021. Two photographs were taken for each eye: one centered on the macula and the other centered at the optic disc. Poor-quality photographs with the presence of cataracts or media opacities were not excluded to reflect real-world conditions. Eyes with retinal vascular diseases (such as retinal vein occlusion, age-related macular degeneration, and retinal artery microaneurysm) and eyes with missing fundus photographs were excluded.

Data on patient demographics (age, sex, ethnicity, and smoking history), diabetic status and control, other metabolic diseases, medication and medical history, and ophthalmic history were collected. Data were blinded from both the grading ophthalmologists and the AI operator. Two ophthalmologists, specialized in vitreoretinal diseases, independently graded the fundus photographs for DR and concomitant pathologies, such as cataracts and macular diseases. Discrepancies were resolved through discussion with a senior vitreoretinal ophthalmologist for a final decision.

Fundus photographs (in jpg format) of included patients were then uploaded to the EyeArt AI system (version 2.1.0) by a resident trainee. Items investigated included the photographs' gradeability, DR severity, presence of DME, and any ophthalmologist-referable recommendations (**Figure**). Gradings of the fundus photographs by both the ophthalmologists and EyeArt were based on the 2017 International Council of Ophthalmology Guidelines for Diabetic Eye Care recommendations.[14] The EyeArt system's sensitivity and specificity on detecting (1) referable DR (moderate or severe non-proliferative DR, proliferative DR, or DME) and (2) any DR or DME were calculated based on the ophthalmologists' diagnoses (the gold standard).

EyeArt's diagnostic sensitivity and specificity have been reported to be 96% and 88%, respectively.[11] In this study, sensitivity and specificity were calculated through the respective equations: $TP + FN = Z^2 \times [\text{sensitivity} \times (1 - \text{sensitivity}) / W^2]$ and $TN + FP = Z^2 \times [\text{specificity} \times (1 - \text{specificity}) / W^2]$, where Z is set to 1.96 for the normal distribution value, corresponding to the 95% confidence interval (CI). W, the maximum acceptable width of the 95% CI, is set to 10%,[15] and the expected sensitivity and specificity are based on estimates from the available information (ie, 95%). TP + FN is equal to 14.751744 and TN + FP is equal to 40.567296. The sample size required for sensitivity and specificity is [TP + FN / P] and [TN + FP / (1–P)], respectively,[15] in which P is the prevalence of DR in Chinese populations in Hong Kong. In a large-scale study of 174 532 patients, the estimated prevalence of DR was 39%.[7] Thus, 37 and 65 patients were required to determine sensitivity and specificity, respectively, with a total required sample size of 102. Statistical analyses were performed using SPSS (version 27, IBM, Armonk, [NY], USA).

Results

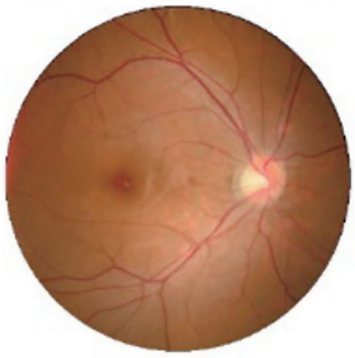
Of 112 patients' records retrieved, 10 were excluded owing to non-Asian patients (n=2), an incomplete set of fundus photographs (n=1), retinal vein occlusion (n=1), active age-related macular degeneration (n=5), or retinal artery microaneurysm (n=1). The remaining 23 men and 79 women (93 Chinese, 3 Pakistanis, 2 Thais, and 4 Indonesians) aged 25 to 77 years with type 1 (n=8) or type 2 (n=94) DM were included for analysis (**Table 1**). The mean duration of DM was 4.9 years; 56 (54.9%) patients were newly diagnosed with DM. 33 (32.4%) patients were taking ≥1 (up to 4) oral hypoglycemic agents; none were taking ≥5. The most commonly used oral hypoglycemic agent was metformin (biguanide) [n=30], followed by sodium-glucose transport

EyeArt Diabetic Retinopathy (DR) Exam Result Summary

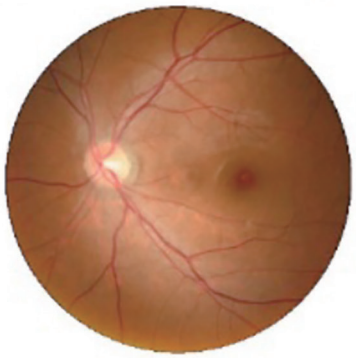
Negative for referable diabetic retinopathy.

Right Eye: No apparent signs of DR [0.0] detected

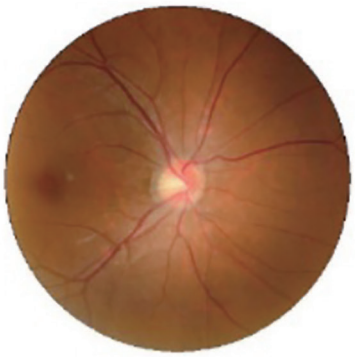
Left Eye: No apparent signs of DR [0.0] detected



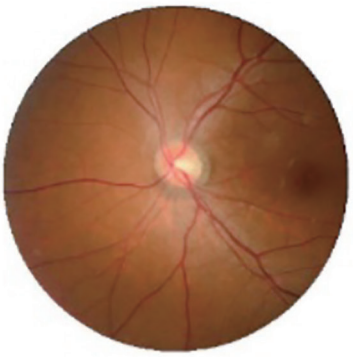
Macula Centered, Right Eye



Macula Centered, Left Eye



ONH Centered, Right Eye



ONH Centered, Left Eye

* Do not use the above thumbnail images for diagnostic purposes.

Notes

A negative result for referable diabetic retinopathy indicates a low risk for moderate non-proliferative DR, severe non-proliferative DR, proliferative DR, and markers for DME.

Recommended scheduling of next diabetic eye screening, per national guidelines and/or ICO Guidelines for Diabetic Eye Care (2017).

Figure. Example report of the EyeArt system

protein 2 inhibitors (n=8), dipeptidyl peptidase IV inhibitors (n=7), sulfonylureas (n=7), thiazolidinediones (n=4), and glucagon-like peptide 1 receptor agonists (n=3). 16 (15.7%) patients required both oral hypoglycemic agents and insulin for DM control; 31 (30.4%) patients injected insulin every day. The mean hemoglobin A1c was 7.0%. Seven (6.9%) patients were active smokers, and 17 (16.7%) patients were ex-smokers.

As diagnosed by the ophthalmologists, 41 (20.1%) of 204 eyes had DR and/or DME, and five (2.5%) eyes were ungradable owing to media opacity (**Table 2**), whereas 26 (25.5%) of 102 patients had referable DR. There was no discrepancy between the two ophthalmologists' gradings. Five (2.5%) eyes had laser marks on the retina, which reflects a history of panretinal photocoagulation treatment for proliferative DR. There were no cases of vitreous, pre-retinal, or subhyaloid hemorrhages.

On detecting referable DR, the EyeArt system had 96.2% (95% CI=80.4%-99.9%) sensitivity, 94.7% (95% CI=87.1%-98.6%) specificity, 86.2% (95% CI=70.6%-94.2%) positive predictive value, 98.6% (95% CI=91.3%-99.8%) negative predictive value, and 95.1% (95% CI=88.9%-98.4%) accuracy (**Table 3**). On detecting any DR or DME, the EyeArt system had 92.7% (95% CI=80.1%-98.5%) sensitivity, 93.7% (95% CI=88.7%-96.9%) specificity, 79.2% (95% CI=67.5%-87.5%) positive predictive value, 98.0% (95% CI=94.3%-99.3%) negative predictive value, and 93.5% (95% CI=89.1%-96.5%) accuracy (**Table 4**).

Table 1. Patient characteristics (n=102)

Characteristic	Value*
Age, y	44.2±13.8 (25-77)
Body mass index, kg/m^2	25.7±4.8 (15.0-42.4)
Duration of diabetes mellitus, y	4.9±8.1 (0-35)
Hemoglobin A1c, %	7.0±2.9 (4.5-19.0)
No. of oral hypoglycemic agents taken	0.6±1.1 (0-4)
Triglycerides, mmol/L	1.38±0.77 (0.51-4.38)
High-density lipoprotein, mmol/L	1.19±0.33 (0.40-1.83)
Calculated low-density lipoprotein, mmol/L	2.30±0.67 (1.13-3.71)

** Data are presented as mean ± standard deviation (range)*

Table 2. Diabetic retinopathy gradings of 204 eyes as determined by ophthalmologists

Diabetic retinopathy	No. of eyes	Referable
Not apparent	158	No
Mild non-proliferative	7	No
Moderate non-proliferative	20	Yes
Severe non-proliferative	5	Yes
Proliferative	9	Yes
Diabetic macular edema	4	Yes
Ungradable	5	Yes

Table 3. Accuracy of the EyeArt system in detecting referable diabetic retinopathy, compared with the gold standard by ophthalmologists

No. of patients (n=102) with referable diabetic retinopathy*		
EyeArt	Ophthalmologists	
	Yes	No
Yes	25	4
No	1	72

** Five eyes that were ungradable were included to simulate real-world conditions.*

Table 4. Accuracy of the EyeArt system in detecting any diabetic retinopathy or diabetic macular edema, compared with the gold standard by ophthalmologists

No. of eyes (n=199) with any diabetic retinopathy or diabetic macular edema*		
EyeArt	Ophthalmologists	
	Yes	No
Yes	38	10
No	3	148

** Five eyes were ungradable*

Discussion

The EyeArt system exhibits high sensitivity and specificity on detecting DME and DR, including referable DR, in patients with DM in Hong Kong. It exceeds the benchmark required for graders in primary healthcare settings (sensitivity of ≥95% and specificity of ≥85%).[7] In our sample, sensitivity is greater as determined by the EyeArt than that based on our retropsective review of patients' records (96.2% vs 69.2%).

Of the handful of commercially available AI systems that have been FDA-approved for DR screening, only IDx-DR (Digital Diagnostics, Coralville [IA], USA) and EyeArt are supported by searchable scientific publications.[9] IDx-DR is not yet available in Hong Kong. EyeArt was similarly unavailable before our study in 2022, but it has become the only FDA-approved AI system for DR screening currently in use in Hong Kong. There is no previous study of EyeArt in a purely Asian population. To the best of our knowledge, the present study is the first to evaluate the real-world performance of an FDA-approved AI system for DR screening in an Asian population.

We used real-world data from a 7-month public screening service to reflect the reality of implementing this diagnostic AI tool in daily clinical practice. The high proportion (more than half) of newly diagnosed DM patients might be related to our hospital setting,[16] as these DM patients might be at the stage of oral hypoglycemic agents and insulin titration and have not yet stepped down to family medicine or general outpatient clinics for regular medication refills. However, this phenomenon could also result from better patient compliance with DR screening soon after diagnosis.

Additionally, the mean patient age of 44.2 years in our study is also younger than that of 60 years in previous studies.[16,17] This might explain why our patients have a lower prevalence of DR than previously published data.[7] Our relatively younger DM population also accounts for the lower percentage (2.5%) of ungradable photographs owing to the presence of cataracts.

Our study reported a higher proportion of referrable DR than a study in large-scale primary care settings did (25.5% vs 9.8%).[7] This could be because our study was conducted in a tertiary hospital that comprises pediatricians, renal dialysis services, orthopedic surgeons, ophthalmologists, intensive care units, and even the sole hyperbaric oxygen service for diabetic foot ulcers to manage all kinds of diabetic complications.[18] Difficult and complicated DM cases tend to stay in hospital settings rather than family medicine or general outpatient clinics, where fewer oral hypoglycemic agents and insulins are available.

AI DR screening systems have advantages over human graders. First, AI systems can grade photographs non-stop 24 hours a day, 7 days a week.[19] This is important to tackle the growing demand for public DR screening services that comes with the rise in DM population; increased service

provision means shorter waiting times.[20] Second, training graders takes time,[21] and the high turnover of personnel is a big issue.[22] Third, human graders perform inconsistently owing to mood changes or time constraints; they can be absent from work during the COVID-19 outbreak, for example.

Despite the advantages of AI DR screening systems, the risks associated with cybersecurity and dependence on overseas private commercial products should not be overlooked.[23-25] The present study is limited by the lack of stereoscopic fundus photographs or optical coherence tomography for DME diagnosis. Human grading based purely on two-dimensional fundus photographs may give a high false positive rate of DME.[26,27] Therefore, EyeArt's genuine diagnostic accuracy of DME requires further validation in future studies.

Conclusion

The EyeArt system is a reliable AI tool for DR screening in patients with DM in Hong Kong. Its usage in primary care and public healthcare programs should be considered.

Contributors

SCLA designed the study, acquired and analyzed the data, and drafted the manuscript. GTHS acquired and analyzed the data. SSYC acquired the data. CKLK acquired the data and 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

All authors have disclosed no conflicts of interest.

Funding/support

Eyenuk provided free trials of the EyeArt system. Skyview Optical Company Limited provided hardware and software support.

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 the Hong Kong East Cluster Research Ethics Committee (reference: HKEC-REC-2022-019). All research protocols were conducted in accordance with the Declaration of Helsinki.

References

1. *Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017;69S:S36-S40.*
2. *Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artificial Intelligence in Healthcare 2020:25-60.*
3. *Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6:94-8.*
4. *Fasterholdt I, Naghavi-Behzad M, Rasmussen BSB, et al. Value assessment of artificial intelligence in medical imaging: a scoping review. BMC Med Imaging 2022;22:187.*
5. *Wong MCS, Huang J, Kong APS. Diabetes screening revisited: issues related to implementation. Hong Kong Med J 2020;26:283-5.*
6. *Sinclair SH, Schwartz SS. Diabetic retinopathy: an underdiagnosed and undertreated inflammatory, neuro-vascular complication of diabetes. Front Endocrinol (Lausanne) 2019;10:843.*
7. *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.*
8. *Abramoff MD, Niemeijer M, Russell SR. Automated detection of diabetic retinopathy: barriers to translation into clinical practice. Expert Rev Med Devices 2010;7:287-96.*
9. *Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Lond) 2020;34:451-60.*
10. *Gunasekeran DV, Ting DSW, Tan GSW, Wong TY. Artificial intelligence for diabetic retinopathy screening, prediction and management. Curr Opin Ophthalmol 2020;31:357-65.*
11. *Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. JAMA Netw Open 2021;4:e2134254.*
12. *Lim JI, Regillo CD, Sadda SR, et al. Artificial intelligence detection of diabetic retinopathy: subgroup comparison of the EyeArt System with ophthalmologists' dilated examinations. Ophthalmol Sci 2022;3:100228.*
13. *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.*
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.*
15. *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.*
16. *Fung MM, Yap MK, Cheng KK. Community-based diabetic retinopathy screening in Hong Kong: ocular findings. Clin Exp Optom 2011;94:63-6.*
17. *Fu MSS, Lai LKP, Chan P, Chow K, Luk MMH, Chao DVK. The prevalence and associated factors of diabetic retinopathy*

in Chinese hypertensive patients newly diagnosed with type 2 diabetes mellitus: a cross sectional study in 3 primary care clinics in Hong Kong. Hong Kong Pract 2017;39:67-76.

18. *Zhang Z, Zhang W, Xu Y, Liu D. Efficacy of hyperbaric oxygen therapy for diabetic foot ulcers: an updated systematic review and meta-analysis. Asian J Surg 2022;45:68-78.*
19. *Raumviboonsuk P, Krause J, Chotcomwongse P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. NPJ Digit Med 2019;2:25.*
20. *Sun J, Lin Q, Zhao P, et al. Reducing waiting time and raising outpatient satisfaction in a Chinese public tertiary general hospital: an interrupted time series study. BMC Public Health 2017;17:668.*
21. *Islam FMA. Accuracy and reliability of retinal photo grading for diabetic retinopathy: remote graders from a developing country and standard retinal photo grader in Australia. PLoS One 2017;12:e0179310.*
22. *Guo H, Zhao Y, Niu T, Tsui KL. Hong Kong Hospital Authority resource efficiency evaluation: via a novel DEA-Malmquist model and Tobit regression model. PLoS One 2017;12:e0184211.*
23. *Niki O, Saira G, Arvind S, Mike D. Cyber-attacks are a permanent and substantial threat to health systems: education must reflect that. Digit Health 2022;8:20552076221104665.*
24. *Kiener M. Artificial intelligence in medicine and the disclosure of risks. AI Soc 2021;36:705-13.*
25. *Radanliev P, De Roure D. Advancing the cybersecurity of the healthcare system with self-optimising and self-adaptative artificial intelligence (part 2). Health Technol (Berl) 2022;12:923-9.*
26. *Wong IYH, Wong RLM, Chan JCH, Kawasaki R, Chong V. Incorporating optical coherence tomography macula scans enhances cost-effectiveness of fundus photography-based screening for diabetic macular edema. Diabetes Care 2020;43:2959-66.*
27. *Wong RL, Tsang CW, Wong DS, et al. Are we making good use of our public resources? The false-positive rate of screening by fundus photography for diabetic macular oedema. Hong Kong Med J 2017;23:356-64.*