Prospective studies on artificial intelligence (AI)-based diabetic retinopathy screening

Onnisa Nanegrungsunk[1], Paisan Ruamviboonsuk[2], Andrzej Grzybowski[3,4]

[1]Retina Division, Department of Ophthalmology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand; [2]Department of Ophthalmology, College of Medicine, Rangsit University, Rajavithi Hospital, Bangkok, Thailand; [3]Department of Ophthalmology, University of Warmia and Mazury, Olsztyn, Poland; [4]Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland
Correspondence to: Onnisa Nanegrungsunk, MD. Retina Division, Department of Ophthalmology Faculty of Medicine, Chiang Mai University, 110 Inthawarorot Rd., Sri Phum, Muang, Chiang Mai 50200, Thailand. Email: onnisa.naneg@cmu.ac.th.
Comment on: Yang Y, Pan J, Yuan M, *et al.* Performance of the AIDRScreening system in detecting diabetic retinopathy in the fundus photographs of Chinese patients: a prospective, multicenter, clinical study. Ann Transl Med 2022;10:1088.

As one of the global health issues, 537 million people were affected with diabetes and 6.7 million deaths were caused by diabetes in 2021 (1). Currently, over 75% of adults with diabetes live in low- and middle-income countries which approximately only 5% of them received a full coverage of guideline-based comprehensive diabetes treatment (1,2). The estimated prevalence of global diabetes is on the rise to approximately 783 million in 2045. In addition, the relative increase in the prevalence of diabetes between 2021 and 2045 is expected to occur in middle-income countries (21.1%) compared to low-income (11.9%) and high-income (12.2%) countries (1). National or large-scale policies which advice on delivering comprehensive managements, such as medication, diet or exercise counselling, and screening for diabetes-related complications are urgent, as these could provide the effective long-term care for patients with diabetes in health systems.

In addition to mortality, diabetes is also one of the leading causes of morbidity, such as visual impairment and blindness. In 2020, diabetic retinopathy (DR), a diabetes-related ocular complication was responsible for blindness in 0.8 million patients globally (3). As increasing prevalence of diabetes is expected, a concomitant rising of DR prevalence and a larger burden for DR screening to reduce the risk of visual morbidity are also anticipated. Although the screening rate were increased with the advent of telemedicine, the limitations on human resources, equipment and cost of screening are still challenges, especially in low-income, middle-income or developing countries (4-6).

With adoption of artificial intelligence (AI) using machine learning (ML) or deep learning (DL) into automated DR detection algorithms, these disruptive technologies have been demonstrated high diagnostic performance, lower cost and shorter acquisition time in a number of studies (7,8). Although the comparison among these studies were difficult due to the differences in AI algorithms, study designs and the primary outcomes, the results of these studies still pointed toward the similar and positive directions of AI as a valuable tool for large-scale DR screening (6,7,9). The first AI algorithm that received approval from United States Food and Drug Administration (FDA) was IDx-DR, which was reported as having sensitivity of 87.2% [95% confidence interval (CI): 81.8–91.2%], specificity of 90.7% (95% CI: 88.3–92.7%) using the Early Treatment Diabetic Retinopathy Study photography as standard (10), and imageability rate (percentage of participants with a completed human grading and with a disease level AI system output) of 96.1% (95% CI: 94.6–97.3%) in detect more than mild non-proliferative diabetic retinopathy (NPDR) and diabetic macular edema (DME) (11). Among the current AI screening systems for DR, a head-to-head comparative study on diagnostic performance of these systems demonstrated the sensitivity, specificity, and area under the curve (AUC) in detecting

Page 2 of 5

Nanegrungsunk et al. Prospective studies on AI-based DR screening

Table 1 Summary of prospective studies on AI-based DR screening

Variables	United Kingdom	China	Australia	Thailand	China
Author	Heydon et al. (12)	Zhang et al. (13)	Scheetz et al. (14)	Ruamviboonsuk et al. (5)	Yang et al. (15)
Year	2021	2020	2021	2022	2022
Number of centers	3	155	5	9	3
Number of participants (or images)	>120,000 images	47,269	236	7,940	1,001
Number of image acquisition per eye	Two[a]	One[b]	One[b]	One[b]	Two[a]
Software	EyeArt (v2.1)	VoxelCloud Retina	Automated AI system	Google DL (v.2)	AIDRScreening system (v.1.0)
Primary outcome/condition	RDR	RDR	RDR	VTDR	RDR (but no DME)
Accuracy or AUC for RDR (95% CI)	–	75.8%[c]	0.92 (0.88–0.96)	94.8% (93.2–96.3%)	92.2% (90.3–93.8%)
Sensitivity for RDR (95% CI) (%)	95.7 (94.8–96.5)	83.3 (81.9–84.6)	96.9 (83.8–99.9)	98.4 (98.3–98.6)	86.7 (83.4–90.1)
Specificity for RDR (95% CI) (%)	–	92.5 (92.1–92.9)	87.7 (81.8–92.2)	93.4 (93.3–93.5)	96.1 (94.1–97.5)
Strengths	Systematic screening (all patients with diabetes are targets)	Largest sample size to date	Offline reading; real-time interpretation system; all of images were reviewed by specialists for reference	Real-time interpretation system; the results were also compared with over-readers specialists	–
Limitations	54% specificity for non-RDR	Fair accuracy	Excluded ungradable images	Exclude ungradable images	Excluded ungradable images; most eyes required pupil dilatation

[a], two CFPs were capture in each eye, one centred on the optic disc and the other on the macula; [b], a single, macula-centered, non-stereoscopic, CFP was captured in each eye; [c], Youden index. AI, artificial intelligence; AUC, area under the ROC curve; CFP, color fundus photograph; CI, confidence interval; DL, deep learning; DME, diabetic macular edema; DR, diabetic retinopathy; NPDR, non-proliferative diabetic retinopathy; RDR, referable diabetic retinopathy (moderate NPDR or worse and/or DME); ROC, receiver operator characteristic; v, version; VTDR, vision-threatening diabetic retinopathy (severe NPDR or worse and/or DME).

referable DR (RDR: moderate NPDR or worse and/or DME) ranged from 87.2% to 100%, 73.3% to 98.5%, and 0.89 to 0.99, respectively (6).

Most of the AI research in DR and Ophthalmology were restrained with a limited number of prospective studies concerning the usability, deployment and performance of AI in the real world. The first few prospective studies of DL for DR screening were conducted in the United Kingdom, China, Australia and Thailand (*Table 1*) (5,12-14). The sensitivity and specificity of these studies were ranged from 83.3% to 98.4% and 87.7% to 96.1% for detecting RDR or sight-threatening DR, respectively.

In China, the world's largest population of patients with diabetes, 13% of adults aged ≥45 years were estimated of having diabetes (16). With an anticipation of increasing prevalence and unmet need for DR screening, China's

prevention first policy ("Healthy China 2030 Plan") was launched (17). The DR screening program was implemented in this policy as well as the development of AI algorithm for the screening in China. A previously published study by Zhang et al. prospectively validated AI-enabled DR screening in 155 diabetes centers across China (15,805 randomly selected participants) using a single, non-mydriatic, macula-centered color fundus photograph (CFP) per eye. They reported 83.3% (95% CI: 81.9–84.6%) sensitivity and 92.5% (95% CI: 92.1–92.9%) specificity to detect RDR including DME using diagnoses by expert panel as standard. The performance of this DL system for classification of five severity stages of DR was comparable to the interobserver variability of the expert specialists (concordance, 83.0% *vs.* 84.3%) (13).

Another study from Yang *et al.*, recently published in

Annals of Translational Medicine, was also conducted in order to prospectively evaluate the performance of another DL algorithm, AIDRScreening system (version 1.0), for DR screening in a Chinese population, including 1,001 participants. In the detection of RDR (not including DME), this study showed sensitivity of 86.7% (95% CI: 83.4–90.1%), specificity of 96.1% (95% CI: 94.1–97.5%) and a false-positive rate of 3.9%, compared to manual human grading. The positive predictive value (PPV) was 94.0% (95% CI: 91.1–96.2%) and negative predictive value (NPV) was 91.1% (95% CI: 88.5–93.3%). The diagnostic accuracy gain rate of the system was 16.6% higher than that of the human graders whereas the diagnostic duration was 37.3% lower (15). The relatively high sensitivity in this study implied that RDR patients would have a high possibility of being detected by DL while the high specificity implied less unnecessary detection of referrals for non-RDR patients. This study, again, demonstrated a potential of high performance of AI for DR screening and a feasibility of the AI system to be deployed into the current DR screening programs in various clinical centers.

However, several points need to be addressed in the study from Yang *et al.* First, the algorithm in this study was designed to detect only RDR which was moderate NPDR or worse, not DME. Thus, patients with mild NPDR with DME might have been missed by the detection. Although analysis on the CFP images is not an ideal method for DME detection compared to analysis on optical coherence tomography images, the previous study on another DL system for CFP demonstrated the capability in prospective detection of DME with 94.7% (95% CI: 93.2–96.1%) accuracy, 89.9% (95% CI: 84.8–94.2%) sensitivity and 95.6% (95% CI: 94.2–96.8%) specificity using adjudication of three retinal specialists on grading CFP as standard (5). Thus, the use of AI system in DME detection from CFP is still promising and feasible. Moreover, the setting of moderate NPDR as a referral threshold may cause referral of patients with moderate NPDR without DME. This group of patients may not require treatment at referral centers and have poor compliance on referrals.

Second, only gradable and qualified CFP images were included in the analysis. These situations, like in other prospective studies on DL for DR, may not reflect the real clinical setting in which a certain proportion of the images is ungradable or not qualified. Third, the DL system in this study requires two CFP images per eye, one centered at the optic disc and another one centered at the macula, for analysis. This means the duration for screening each patient, including CFP capture and analysis, may be longer than the DL systems that require an analysis on a single CFP image per eye. The deployment of these systems may be more practical in countries where there is a huge population of patients with diabetes like China or India. Some prospective studies on AI for DR screening using a single CFP image per eye have demonstrated very high performances in detecting patients with RDR. The summary of recently published prospective studies on AI for DR screening including screening strategies and diagnostic parameters of the studies is shown in *Table 1*.

Diagnostic performances of DL for DR screening, however, may not be significant issues to be addressed in prospective validation studies since the results from most of these studies confirmed the high performances, although the performances may not be as robust as those reported in retrospective validation studies. The other important outcomes to be measured would be those associated with deployment, such as the proportion of patients successfully screened, the proportion of patients successfully referred, the downtime of the DL system, or the burden on healthcare personnel to be alleviated by DL. These outcomes may not be easy to measure as the diagnostic outcomes.

Another important area on further studies of DL for DR screening may rest on economic evaluation. There have been only a few published studies on DL in terms of cost-effectiveness compared with human graders for the screening. This evaluation in both healthcare provider and societal perspectives in the context of the countries where the DL system are being deployed is important to advocate policy makers for decision making and resource preparation (18).

After the Covid-19 pandemic, the existing healthcare system has been disrupted and the new models of heath care are implemented globally; these include reducing unnecessary clinic visit and duration, reducing unnecessary care or referral, reducing unnecessary work of health personnel, and increasing remote health care via telemedicine (6,19). These new models of health care further support the utilization of telemedicine in combining with AI in ophthalmology, particularly in DR, and other specialties in medicine.

Acknowledgments

Page 4 of 5

Nanegrungsunk et al. Prospective studies on AI-based DR screening

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Annals of Translational Medicine*. The article did not undergo external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-2022-71/coif). ON received supports for medical writing (for other study) from Novartis, payment for lecture from Allergan, attending the educational meetings from Bayer and Novartis, and participating as an advisory board in Roche. PR received grants from Roche and Novartis; consulting fees from Bayer, Novartis and Roche; payment as a speaker from Bayer, Novartis, Roche and Topcon. AG received grants from Alcon, Bausch&Lomb, Zeiss, Teleon, J&J, CooperVision and Hoya; lecture fees from Thea, Polpharma and Viatris; member of advisory boards of Nevakar, GoCheckKids and Thea. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119.
2. Flood D, Seiglie JA, Dunn M, et al. The state of diabetes treatment coverage in 55 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data in 680 102 adults. Lancet Healthy Longev 2021;2:e340-51.
3. GBD 2019 Blindness and Vision Impairment Collaborators; Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health 2021;9:e144-60. Erratum in: Lancet Glob Health. 2021 Apr;9(4):e408.
4. Jani PD, Forbes L, Choudhury A, et al. Evaluation of Diabetic Retinal Screening and Factors for Ophthalmology Referral in a Telemedicine Network. JAMA Ophthalmol 2017;135:706-14.
5. Ruamviboonsuk P, Tiwari R, Sayres R, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. Lancet Digit Health 2022;4:e235-44.
6. Li JO, Liu H, Ting DSJ, et al. Digital technology, telemedicine and artificial intelligence in ophthalmology: A global perspective. Prog Retin Eye Res 2021;82:100900.
7. Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Lond) 2020;34:451-60.
8. 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.
9. Ting DSW, Cheung CY, Lim G, 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.
10. Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. Ophthalmology 1991;98:786-806.
11. Abràmoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018;1:39.
12. Heydon P, Egan C, Bolter L, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. Br J Ophthalmol 2021;105:723-8.
13. Zhang Y, Shi J, Peng Y, et al. Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study. BMJ Open Diabetes

Res Care 2020;8:e001596.

14. Scheetz J, Koca D, McGuinness M, et al. Real-world artificial intelligence-based opportunistic screening for diabetic retinopathy in endocrinology and indigenous healthcare settings in Australia. Sci Rep 2021;11:15808.

15. Yang Y, Pan J, Yuan M, et al. Performance of the AIDRScreening system in detecting diabetic retinopathy in the fundus photographs of Chinese patients: a prospective, multicenter, clinical study. Ann Transl Med 2022;10:1088.

16. Shi G, Zhu N, Qiu L, et al. Impact of the 2020 China Diabetes Society Guideline on the Prevalence of Diabetes

Mellitus and Eligibility for Antidiabetic Treatment in China. Int J Gen Med 2021;14:6639-45.

17. Wang W. Interpretation of the Diabetes Prevention and Control Action of the Healthy China Initiative 2019-2030. China CDC Wkly 2020;2:143-5.

18. Ruamviboonsuk P, Chantra S, Seresirikachorn K, et al. Economic Evaluations of Artificial Intelligence in Ophthalmology. Asia Pac J Ophthalmol (Phila) 2021;10:307-16.

19. Saleem SM, Pasquale LR, Sidoti PA, et al. Virtual Ophthalmology: Telemedicine in a COVID-19 Era. Am J Ophthalmol 2020;216:237-42.