AMERICAN ACADEMY OF OPHTHALMOLOGY®

Accuracy of Integrated Artificial Intelligence Grading Using Handheld Retinal Imaging in a Community Diabetic Eye Screening Program

Recivall P. Salongcay, MD, MPM,[1,2,3] *Lizzie Anne C. Aquino, BSc,*[2] *Glenn P. Alog, MD, MBA,*[2,3] *Kaye B. Locaylocay, MD, MBA,*[2,3] *Aileen V. Saunar, MD,*[2,3] *Tunde Peto, MD, PhD,*[1] *Paolo S. Silva, MD*[2,3,4,5]

Purpose: To evaluate mydriatic handheld retinal imaging performance assessed by point-of-care (POC) artificial intelligence (AI) as compared with retinal image graders at a centralized reading center (RC) in identifying diabetic retinopathy (DR) and diabetic macular edema (DME).

Design: Prospective, comparative study.

Subjects: Five thousand five hundred eighty-five eyes from 2793 adult patients with diabetes.

Methods: Point-of-care AI assessment of disc and macular handheld retinal images was compared with RC evaluation of validated 5-field handheld retinal images (disc, macula, superior, inferior, and temporal) in identifying referable DR (refDR; defined as moderate nonproliferative DR [NPDR], or worse, or any level of DME) and vision-threatening DR (vtDR; defined as severe NPDR or worse, or any level of center-involving DME [ciDME]). Reading center evaluation of the 5-field images followed the international DR/DME classification. Sensitivity (SN) and specificity (SP) for ungradable images, refDR, and vtDR were calculated.

Main Outcome Measures: Agreement for DR and DME; SN and SP for refDR, vtDR, and ungradable images.

Results: Diabetic retinopathy severity by RC evaluation: no DR, 67.3%; mild NPDR, 9.7%; moderate NPDR, 8.6%; severe NPDR, 4.8%; proliferative DR, 3.8%; and ungradable, 5.8%. Diabetic macular edema severity by RC evaluation was as follows: no DME (80.4%), non−ciDME (7.7%), ciDME (4.4%), and ungradable (7.5%). Referable DR was present in 25.3% and vtDR was present in 17.5% of eyes. Images were ungradable for DR or DME in 7.5% by RC evaluation and 15.4% by AI. There was substantial agreement between AI and RC for refDR ($\kappa = 0.66$) and moderate agreement for vtDR ($\kappa = 0.54$). The SN/SP of AI grading compared with RC evaluation was 0.86/0.86 for refDR and 0.92/0.80 for vtDR.

Conclusions: This study demonstrates that POC AI following a defined handheld retinal imaging protocol at the time of imaging has SN and SP for refDR that meets the current United States Food and Drug Administration thresholds of 85% and 82.5%, but not for vtDR. Integrating AI at the POC could substantially reduce centralized RC burden and speed information delivery to the patient, allowing more prompt eye care referral.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article.

Supplemental material available at www.ophthalmologyscience.org.

Diabetic retinopathy (DR) is one of the leading causes of preventable vision loss globally, especially among the working-age population.[1] The incidence of DR is rising yearly; hence, there is a constant need to increase DR screening capacity to prevent blindness among people with diabetes.[2] Teleophthalmology has enabled the systematic, widescale implementation of community-based DR screening programs (DRSPs).[3,4] With teleophthalmology DRSPs, screening units are placed at strategic community locations where patients with diabetes can have their retinal images taken, which are then analyzed remotely by trained graders at a centralized reading center (RC). The results are then sent back to the patient and screening sites, and include follow-up details and referral or treatment recommendations whenever appropriate. This process decreases the need for in-person eye examinations with a clinician and is a cost-effective approach to delivering a DRSP.[5]

The establishment of widescale community-based teleophthalmology programs is not straightforward. Significant hurdles remain in some settings due to resource limitations, geographic isolation, and high cost of screening equipment.[6,7] In recent years, the incorporation of handheld retinal imaging devices in DRSPs have gained acceptance, as they are significantly cheaper and allow broader accessibility and improved portability while maintaining acceptable levels of sensitivity (SN) and specificity (SP).[8–11] One of the main

https://doi.org/10.1016/j.xops.2023.100457
ISSN 2666-9145/23

challenges for DRSPs is how to ensure early identification and referral of eyes at risk of vision loss due to the sheer volume of images that need to be analyzed in a timely manner. The training and certification of image graders require considerable time and financial investment which can be prohibitive for many health care systems. The integration of artificial intelligence (AI) in the DRSP can potentially address this need. Artificial intelligence provides DR assessment at point-of-care (POC) and, therefore, decreases image-grading burden by identifying eyes with referable levels of DR automatically without any significant delay. One of the barriers to the deployment of AI in DRSP, however, is ensuring accuracy.

Despite the large interest in AI for DR, to the best of our knowledge there are limited published prospective studies[12–14] and only 3 United States (US) Food and Drug Administration (FDA)-approved AI algorithms that are limited to 2 specific tabletop retinal cameras.[12–17] There is a need for more evidence that AI systems for DR evaluation perform as well as human graders in the clinical setting, especially on handheld retinal images in community settings with limited resources. In this study, we evaluated the performance of handheld retinal images assessed by AI at the time of imaging in identifying referable levels of DR, as compared with standard retinal image graders at a centralized RC in a community-based teleophthalmology DRSP.

Methods

Population and Sample

This was a prospective comparative study of AI assessment of referable DR (refDR) and vision-threatening DR (vtDR). A total of 5585 eyes from 2793 adult patients with diabetes from a community-based DRSP in Central Luzon, Philippines were included in the study. The inclusion criteria were as follows: (1) aged $\geq$ 18 years; (2) previously received a diagnosis of diabetes type 1 or 2; and (3) consented to undergo fundus photography. The exclusion criteria were any of the following: (1) the presence of media opacities such as dense cataracts or corneal scars that prevent sufficient view of the fundus; (2) contraindication to pupil dilation such as history of hypersensitivity to mydriatic eye drops; or (3) presence of any active eye infection or inflammation at the time of imaging visit. Patient enrollment and data collection took place from September 2021 to August 2022.

The study design complies with the ethical standards of the Declaration of Helsinki. The study protocol was approved by the institutional review board of The Medical City (reference number GVSOVS2021-116). All participants provided informed consent.

Image Acquisition and Analysis

All participants underwent mydriatic handheld fundus photography on the Aurora IQ camera (Optomed Ltd) using an imaging protocol previously described by our group.[11] Briefly, the participants underwent pupil dilation using 1 drop of tropicamide 0.5% + phenylephrine 0.5% eye drops, and 5-field (disc-centered, macula-centered, superior, inferior, and temporal to the macula) 50° retinal images were acquired using the Aurora handheld camera. All images were obtained by retinal imagers (L.A.C.A.) who underwent training and certification by the Gloucestershire Retinal Education Group (Gloucestershire Hospitals, National Health Service Foundation Trust, United Kingdom). Additional training on the use of the handheld camera and the AI system was provided by the device manufacturer. Before the start of the study, the retinal imagers had acquired $>$ 2500 retinal images with the handheld retinal camera in a clinical setting, ensuring proficiency in handheld image acquisition.

Reading center evaluation was performed independently by 3 masked trained graders who are board-certified ophthalmologists (G.P.A., K.B.L., and A.V.S.) at a centralized RC using high-resolution, high-definition, color-calibrated liquid-crystal display computer displays. The presence of diabetes-related retinal lesions such as hemorrhages or microaneurysms, venous beading, intraretinal microvascular abnormalities, new vessels, preretinal or vitreous hemorrhage, tractional membranes, and laser marks were assessed. Diabetic retinopathy severity was assessed using the international clinical classification for DR (no DR, mild nonproliferative DR [NPDR], moderate NPDR, severe NPDR, proliferative DR, or ungradable). Diabetic macular edema (DME) severity was assessed as no DME, DME, center-involving DME (ciDME [defined as hemorrhages or microaneurysms, or exudates within 200 μm from the center of the fovea]), or ungradable for DME. An image is considered ungradable if it is not possible to visualize retinal features or lesions in $\geq$ 50% of the image. Referable DR was defined as moderate NPDR or worse, any proliferative DR, any DME, or ungradable images, whereas vtDR was defined as severe NPDR or worse, any proliferative DR, ciDME, or ungradable images. Discrepancies were adjudicated by retina specialists experienced in image grading (R.P.S. or P.S.S.), and the adjudicated grade was considered the final assessment.

Point-of-care AI assessment of 2-field (disc-centered and macula-centered) images was provided by the Aurora camera's integrated AI software at the time of imaging (Aurora IQ, Optomed Ltd). The POC AI used in this study was SELENA Eyris deep learning system as reported by Ting et al[18] in JAMA. The deep learning system was trained on detecting DR using a data set of 76 370 images. The reported SN of the SELENA Eyris deep learning system for refDR was 90.5%, and SP was 91.6%. When analyzing the retinal images in this study, the images were analyzed without any preprocessing or postprocessing, regardless of image quality. The images acquired by the handheld retinal camera are transferred using wireless local-area network connectivity for cloud-based server to enable AI-based image analysis. Using specified severity thresholds of either refDR or vtDR, the AI software gives an automated score of 0 (eye does not meet the severity threshold), 1 (eye meets the severity threshold), or AI failure (ungradable images). A summary of the grading workflow of RC and POC AI is provided in Figure S1 (available at www.ophthalmologyscience.org).

Statistical Analysis

Baseline characteristics of participants were presented as means (standard deviation) or numerical values (percentage). The level of agreement between POC AI and RC image grading was assessed using kappa (κ). The strength of agreement was determined using the Landis and Koch interpretation of κ statistics (0.20: slight agreement; 0.21–0.40: fair agreement; 0.41–0.60: moderate agreement; 0.61–0.80: substantial agreement; 0.81–1.00: almost perfect agreement). Sensitivity and SP values for refDR, vtDR, and ungradable images were calculated. Sensitivity and SP performance thresholds of 85% and 82.5%, respectively, were used, following US FDA requirements.[12,15] Statistical analysis was performed using SAS software version 9.4 (SAS, Inc).

Results

Of the 2793 participants, 1798 (64.4%) were women; the mean (± standard deviation) age of participants was 59.1 (± 10.4) years, and the mean (± standard deviation) duration of diabetes was 7.1 (± 7.3) years. Type 2 diabetes was present in 95.5% of patients.

Diabetic retinopathy severity by RC evaluation was as follows: no DR (3758 eyes [67.3%]), mild NPDR (540 eyes [9.7%]), moderate NPDR (482 eyes [8.6%]), severe NPDR (271 eyes [4.8%]), proliferative DR (213 eyes [3.8%]), and ungradable (321 eyes [5.8%]). Furthermore, DME severity by RC evaluation was as follows: no DME (4490 eyes [80.4%]), non−ciDME (430 eyes [7.7%]), ciDME (246 eyes [4.4%]), and ungradable (419 eyes [7.5%]). Referable DR was present in 25.3% and vtDR was present in 17.5% of eyes. Images were ungradable for DR or DME in 15.4% of eyes by POC AI and in 7.5% by the RC. Exact agreement between RC and POC AI for ungradable images was 89%. The primary cause of the ungradable images was image quality in all cases. In 84 (1.5%) eyes, images were gradable by the POC AI but ungradable by the RC due to poor image quality preventing the assessment of DR severity in 27 (32.1%) eyes and DME in all 84 (100%) eyes. In 528 (9.4%) eyes, images were gradable by the RC but ungradable by the POC AI with 147 eyes (27.8%) having refDR. The baseline characteristics of participants and DR/DME severity assessment by RC evaluation are summarized in Table 1.

There was substantial agreement between POC AI assessment and RC evaluation for refDR ($\kappa = 0.66$; exact agreement = 86.0%), and moderate agreement for vtDR ($\kappa = 0.54$; exact agreement = 82.4%) and ungradable images ($\kappa = 0.47$; exact agreement = 89.0%). Sensitivity and SP for refDR were 0.86/0.86; vtDR, 0.92/0.80; and ungradable images, 0.80/0.90. Images were ungradable for DR or DME in 15.4% of eyes by POC AI. The US FDA SN and SP thresholds of 85% and 82.5% were met by POC AI for refDR.[15] Artificial intelligence achieved the vtDR thresholds for SN but not for SP. On the other hand, POC AI achieved ungradable thresholds for SP but not for SN. Table 2 summarizes the agreement rates, SN, and SP of POC AI against RC evaluation. When ungradable images are excluded from the analysis, the performance of the algorithm for both refDR and vtDR is improved and meets the FDA minimum thresholds for SN and SP (Table S3, available at www.ophthalmologyscience.org).

An assessment of the POC AI false-negative and false-positive images was performed after 3 months of patient enrollment. A total of 1477 (26.4%) eyes were evaluated with 32 false-negative eyes and 216 false-positive for refDR. Among the POC AI false-negative eyes on RC review, the human grader was more accurate in 29 eyes (90.6%) and the AI in 3 eyes (9.4%). When the human grader was more accurate, this was due to image quality in 26 eyes and field discrepancy in 3 eyes. In the subset of the first 40 (18.5%) eyes graded as false-positive for refDR by the POC AI, 39 (97.5%) eyes were confirmed by the RC to have no refDR present and 1 (2.5%) eye was assessed to have moderate NPDR. The POC AI did not provide an AI attention map. However, based on RC review of the images, the potential source of refDR finding by the POC AI is due to the following: image quality in 18 eyes, drusen or retinal pigment changes in 12 eyes, macular sheen in 5 eyes, myopic fundus changes in 2 eyes, epiretinal membrane in 1 eye, and chorioretinal scar in 1 eye.

Table 1. Baseline Characteristics and DR/DME Severity by RC Evaluation

	Value ± SD or (%)
Female sex	1798 (64.4)
Age, yrs	59.1 ± 10.4
Type 2 DM	2667 (95.5)
Duration of DM, yrs	7.1 ± 7.3
DR severity by RC evaluation	
No DR	3758 (67.3)
Mild NPDR	540 (9.7)
Moderate NPDR	482 (8.6)
Severe NPDR	271 (4.8)
PDR	213 (3.8)
Ungradable	321 (5.8)
DME severity by RC evaluation	
No DME	4490 (80.4)
Non-ciDME	430 (7.7)
ciDME	246 (4.4)
Ungradable	419 (7.5)
Referable DR	25.3%
Vision-threatening DR	17.5%
Ungradable for DR or DME	7.5%

ciDME = center-involving diabetic macular edema; DM = diabetes mellitus; DME = diabetic macular edema; DR = diabetic retinopathy; NPDR = nonproliferative diabetic retinopathy; PDR = proliferative diabetic retinopathy; RC = reading center; SD = standard deviation.

Discussion

In this study, POC AI assessment of DR severity using handheld retinal imaging meets the current thresholds for SN

Table 2. Agreement Rates and Measures of Performance of Point-of-Care Artificial Intelligence against Reading Center Evaluation

Threshold	Kappa	Exact Agreement	Sensitivity	Specificity	PPV	NPV
RefDR‡	0.66	86.0%	0.86*	0.86*	0.67	0.95
VtDR§	0.54	82.4%	0.92*	0.80†	0.50	0.98
Ungradable	0.47	89.0%	0.80†	0.90*	0.39	0.98

DR = diabetic retinopathy; NPV = negative predictive value; PPV = positive predictive value; refDR = referable diabetic retinopathy; vtDR = vision-threatening diabetic retinopathy.
*refDR or vtDR thresholds that met the 85% sensitivity and 82.5% specificity rates.
†refDR or vtDR thresholds that did not meet the 85% sensitivity and 82.5% specificity rates.
‡refDR is defined as moderate nonproliferative DR or worse, any diabetic macular edema or ungradable images.
§vtDR is defined as severe nonproliferative DR or worse, center-involving diabetic macular edema or ungradable images.

and SP for refDR of 82.5% and 85.0%, respectively. The use of POC AI and handheld imaging as a DR screening tool has the potential to decrease the burden on reading centers, especially in low-income settings or geographically isolated communities. Reliable AI assessment of DR at POC with real-time output can guide clinical decision-making and referral recommendations. The handheld retinal camera and imaging protocol used in this study have been validated against standard 7-field ETDRS photography.[11] The weighted κ level of agreement for referable DR, as compared with ETDRS standard 7-field photography, is 0.81, with a SN and SP of 0.84 and 0.97, respectively. No lid or lash artifacts were captured with the handheld retinal cameras. Macular image quality was the primary cause of RC ungradable images. For the POC AI, one of the potential reasons for the high ungradable rate is the difference in the camera type used to train the algorithm, as compared with the camera used in this study. The AI algorithm used in the present study trained on images acquired from tabletop retinal cameras as compared with a handheld retinal camera used in the previous study. Additionally, the data set used to train the camera did not include images from a Filipino population. This lack of representation in the AI training set may affect AI performance, particularly in underrepresented populations or with retinal cameras not typically used in high-resource settings such as the camera and population we evaluated in the Philippines.

The exponential increase in the prevalence of diabetes globally presents a corresponding increase in the prevalence of DR. According to projections by the International Diabetes Federation, approximately 537 million (or around 1 in 10) adults (aged 20−79 years) worldwide have diabetes in 2021.[19] This projection suggests that $>$ 1 billion eyes need to be screened for DR at least once annually, translating to around 3 million eyes daily. Even with the recent advances in technology such as low-cost, portable retinal imaging devices and teleophthalmology, the sheer number of images that need to be graded in a timely manner places an overwhelming burden on human graders. Traditionally, the responsibility of grading the fundus images lies with eye care clinicians who perform the task when time permits or once clinical responsibilities are accomplished. Oftentimes there is not enough time to grade all images captured for the day and the backlogs grow. Many DR screening programs have introduced certified and highly trained nonclinical graders to ease the burden of image analysis on clinicians.[20−22] In these settings, certified ophthalmic image graders have acted as force-multipliers to alleviate pressures on eye care clinicians. However, this is still not enough to cope with DR screening services' workload. The use of validated AI algorithms for DR evaluation is uniquely suited to address this need. Artificial intelligence can provide DR assessment at POC, identifying eyes with retinopathy automatically and without any delay; hence, only patients who are at risk of losing their sight will be referred for an in-person consult. This use relieves human graders and clinicians of the bulk of images that need to be evaluated, thereby reducing time delays, and the requirement for in-person consults is also minimized.

Before an AI platform can be incorporated into a DRSP, however, its SN and SP should be ensured to be at par with how human graders are performing. A study on an autonomous (i.e., does not require human review) AI-based grading of DR in primary care screening sites by Abramoff et al[12] exceeded the 85% SN and 82.5% SP end point targets set by the US FDA. Using the IDx-DR platform, they were able to achieve 87.2% SN and 90.7% SP in detecting the presence of more-than-mild DR. This pivotal trial led to the authorization of the first US FDA-approved AI system for autonomous use in DR screening.[15,23] Since then, other AI systems have also gained authorization from the US FDA.[16,17,24,25]

A prior study in Australia using a previously validated AI system in a clinical practice setting determined a 92% SP; however, SN was not generated due to low incidence of disease in the study cohort.[14] Additionally, the deployed AI system suffered from a high false-positive rate due to the low disease incidence and insufficient image quality. The authors recommended that images flagged by the AI system as having referable illness should undergo further review by an ophthalmologist before referral advice is given to the patient.[14] In another study incorporating the IDx-DR device in daily clinical workflow in an existing diabetes care system, SN/SP for refDR was 0.91/0.84 when using the EURODIAB classification and 0.68/0.86 when using the international scale, when compared with retina specialists.[26] The authors performed a post hoc analysis and found that overzealous adherence by the retina specialists to the international scale definitions led to misclassification of images and subsequent discrepancies in AI performance.[26] Discrepancies in AI accuracy measures between the 2 grading scales in this case suggest that differences in the reference standard and settings can impact the performance of the algorithm. In another study, a trained and validated AI system was deployed in the general ophthalmology clinics, a vitreoretinal clinic, and a teleophthalmology community screening unit among 2 hospital systems in India.[13] The SN/SP of the AI system for refDR was 88.9% to 92.1% and 92.2% to 95.2%, respectively. The results were consistent across the 2 hospitals, suggesting good reproducibility of the algorithm in prospective clinical settings.[13] Despite performing well in validation studies under controlled research settings, AI systems may encounter issues when implemented in clinical sites or screening programs. Hence, consistency of AI performance across multiple settings must be assured for model generalizability. These studies highlight the importance of external evaluation of AI systems in clinical settings before wide scale use in DRSPs.

In our study, POC AI achieved SN/SP values of 0.86/0.86 for refDR and 0.92/0.80 for vtDR, respectively, exceeding the US FDA threshold for refDR but not for vtDR.

The performance values of POC AI are comparable to the present, US FDA-approved AI systems (Table 4). All AI systems meet the SN and SP thresholds for refDR. However, POC AI suffers from a higher AI failure rate, and upon review this failure rate was mainly due to poorer image quality on handheld retinal cameras compared with tabletop cameras used by the 3 FDA-approved systems. Point-of-care AI was also primarily trained on images from tabletop cameras and not on handheld cameras which could have also contributed to this result. It will be imperative to

Table 4. Comparison with FDA-Approved AI Systems

Referable DR*	Point-of-Care AI†	IDx-DR	Eyenuk EyeArt†	AEYE-DS
Sensitivity	0.86 (0.92)	0.87	0.95–1.00 (0.89–1.00)	0.95
Specificity	0.86 (0.80)	0.90	0.87–0.92 (0.94–0.98)	0.89
Positive predictive value	0.67 (0.50)	0.73	0.46–0.65 (0.27–0.67)	0.54
Negative predictive value	0.95 (0.98)	0.96	0.99–1.00 (0.99–1.00)	0.99
Prevalence, %	32.4	23.8	10.5–12.2 (2.4–4.4)	12.3
AI failure rate, %	15.4	8.0	3.5	0.9

AI = artificial intelligence; DR = diabetic retinopathy; FDA = Food and Drug Administration.
These are presented as a reference only and do not represent comparative performance. It is important to note that each of the DR grading programs referenced in the table were evaluated using different image sets and grading methods.
*Moderate nonproliferative DR or worse or any level of diabetic macular edema. Based on United States FDA submission.
†Values in () are for vision-threatening retinopathy.

address this high failure rate in future versions of the software since this will affect the efficiency and acceptability of POC AI. Many authors have noted a tendency for algorithms to underperform when deployed to populations which are distinct from the ones on which they were trained, leading to concerns that this could potentially propagate health care disparities.[27–31] Since an AI system works well if it is applied on images similar to those on which it was developed, the POC AI algorithm may also need to undergo proper optimization using a training set that is comparable to the intended population. As a caveat, it is important to note that comparing these AI systems is not straightforward. The difference in the performance measures among the various AI platforms may be due to several factors, including difference in the cameras, image sets and grading methods used, dissimilar clinical settings, variation in algorithm threshold, or a combination of these factors. These are presented here merely as a reference and generally do not represent comparative performance.

The prevalence of refDR in the POC AI study population was considerably higher than in other studies due to a significant first-pass effect that is observed during the first round of screening in newly implemented DRSP.[32] Since most patients in the study never had any prior DR screening visits, a higher level of previously undetected prevalent disease was noted in the initial year of screening. This initial large demand on specialized eye care services will need to be considered in the planning of the DRSP.[33] Our results also suggest that currently, when using POC AI, only refDR (and not vtDR) may be applied as the threshold for screening when using handheld retinal imaging. With the increased accessibility provided by AI integration into systematic DR screening programs, the repetitive nature of the retinal examinations may act as a fail-safe to potentially identify disease that may have been missed. Although the current threshold for vtDR was not met with POC AI for handheld imaging, advances in algorithm development, computational methods, and the increasing availability of handheld imaging data sets may substantially improve AI performance in the future.

The strengths of this study include the prospective study design, large sample size, enrollment of patients from an active existing community-based DRSP, use of a standardized imaging protocol by trained imagers, and centralized RC evaluation by certified graders (board-certified ophthalmologists/retina specialists). Additionally, this use marked the first clinical use of POC AI in the Philippines, emphasizing the potential to effectively address health care disparities in underserved regions where access to care is limited. A limitation of POC AI was that it only grades the disc- and macula-centered images taken using the handheld camera. This limitation may also explain, in part, the higher ungradable rate of POC AI compared with RC evaluation as there were more fields (and therefore more areas of the retina) available for assessment in RC. Despite this, we deemed it was essential to emulate clinical conditions in this comparative study. Since the patients were recruited in an active, community-based DRSP that uses a validated 5-field handheld retinal imaging protocol that has been shown to perform favorably compared with standard ETDRS 7-field photography (weighted κ, 0.75),[11] any POC AI that will be deployed in the DRSP should perform well enough compared with the RC evaluation of 5-field images. Another limitation is that this study focused only on DR and was not designed to evaluate other retinal lesions that may be present in people with diabetes and that may require referral for specialist care. Other issues surrounding AI for DR screening, such as legal and regulatory approvals and user acceptability, are beyond the scope of this study and must be addressed in further investigations. Future work on POC AI should focus on the reduction of the ungradable rate and development of systems trained on handheld retinal images.[34]

This study demonstrated that POC AI at the time of imaging, following a defined retinal imaging protocol, using handheld fundus cameras, has SN and SP for refDR that meets the current US FDA thresholds. Integrating AI at the POC in a community-based DRSP could substantially reduce centralized RC burden and speed information delivery to the patient, allowing more prompt eye care referrals. Looking ahead, health systems globally should start to explore how AI can be integrated into their existing DRSP to help cope with the current and expected rise in demand for DR screening services. Teleophthalmology and DRSP should strive to take hold of the significant advances

in AI and retinal imaging and build on what can be sustainably used. Paramount to these is maintaining the quality of care and establishing standards to ensure patient safety and outcomes are improved. Artificial intelligence for DR screening can help ensure that every person with diabetes gets screened at appropriate intervals, and that those with DR are properly identified and offered timely treatment with the ultimate goal of saving sight.

Acknowledgments

The authors thank Lloyd Paul Aiello, MD, PhD and Jerry D. Cavallerano, OD, PhD (Beetham Eye Institute, Joslin Diabetes Center, Department of Ophthalmology, Harvard Medical School) for their valuable contributions in the review of the scientific content of the manuscript. The authors also thank Mrs Lilibeth Manganip (Philippine Eye Research Institute, University of the Philippines, Manila) for administrative support during the conduct of the study.

Footnotes and Disclosures

Originally received: July 11, 2023.
Final revision: November 8, 2023.
Accepted: December 11, 2023.
Available online: December 15, 2023. Manuscript no. XOPS-D-23-00169R1.

[1] Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom.

[2] Philippine Eye Research Institute, University of the Philippines, Manila, Philippines.

[3] Eye and Vision Institute, The Medical City, Metro Manila, Philippines.

[4] Beetham Eye Institute, Joslin Diabetes Center, Boston, Massachusetts.

[5] Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts.

Disclosures:

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s):

T.P.: Research support − MRC Newton Fund Research Grant; Financial support − Optomed, Optos; Personal fees − Novartis, Bayer, Roche, Heidelberg, Optos.

P.S.S.: Research Support − Newton-Agham Grant from the Department of Science and Technology, Republic of the Philippines and the Medical Research Council, United Kingdom; Financial support − Optomed, Hillrom, Optos; Personal Fees − Novartis, Bayer, Roche, Optomed, Optos; Equipment − Optomed, Hillrom.

The other authors have no proprietary or commercial interest in any materials discussed in this article.

Supported by the UK Medical Research Council (58 Victoria Embankment, London EC4Y 0DS, UK) and the Philippine Council for Health Research and Development (Saliksik Building, DOST Compound, Gen. Santos Ave., Bicutan, Taguig City 1631, Metro Manila, Philippines) through the Newton-Agham Grant. Project Reference: MR/R025630/1. The funding agencies have no role in the design or conduct of the study.

Presented in part as a free paper during the Association for Research in Vision and Ophthalmology (ARVO) 2022 Annual Meeting on May 2, 2022, at Denver, Colorado, and as a free paper during the European Association for Diabetic Eye Complications (EAsDEC) 33rd Meeting on June 3, 2023, at Coimbra, Portugal.

HUMAN SUBJECTS: Human subjects were included in this study. The study design complies with the ethical standards of the Declaration of Helsinki. The study protocol was approved by the institutional review board of The Medical City (reference number GVSOVS2021-116). All participants provided informed consent.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Salongcay, Peto, Silva

Data collection: Salongcay, Aquino, Alog, Locaylocay, Saunar, Peto, Silva

Analysis and interpretation: Salongcay, Aquino, Silva

Obtained funding: Silva, Peto

Overall responsibility: Salongcay, Silva

Abbreviations and Acronyms:
AI = artificial intelligence; **ciDME** = center-involving diabetic macular edema; **DME** = diabetic macular edema; **DR** = diabetic retinopathy; **DRSP** = diabetic retinopathy screening program; **FDA** = Food and Drug Administration; **κ** = kappa; **NPDR** = nonproliferative diabetic retinopathy; **POC** = point-of-care; **RC** = reading center; **refDR** = referable diabetic retinopathy; **SN** = sensitivity; **SP** = specificity; **US** = United States; **vtDR** = vision-threatening diabetic retinopathy.

Keywords:
Artificial intelligence, Diabetic retinopathy, Handheld devices, Retinal imaging, Screening.

Correspondence:
Paolo S. Silva, MD, Beetham Eye Institute, Joslin Diabetes Center, 1 Joslin Place, Boston, MA 02215. E-mail: paoloantonio.silva@joslin.harvard.edu.

References

1. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990−2020: a systematic review and meta-analysis. *Lancet Glob Health*. 2017;5:e1221−e1234.
2. Burton MJ, Ramke J, Marques AP, et al. The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. *Lancet Glob Health*. 2021;9:e489−e551.
3. Salongcay RP, Silva PS. The role of teleophthalmology in the management of diabetic retinopathy. *Asia Pac J Ophthalmol (Phila)*. 2018;7:17−21.
4. Silva P, Cavallerano J, Paz-Pacheco E, Aiello LP. Diabetic retinopathy in Southeast Asia: a call for ocular telehealth programs. *J ASEAN Fed Endocr S*. 2012;27.
5. Avidor D, Loewenstein A, Waisbourd M, Nutman A. Cost-effectiveness of diabetic retinopathy screening programs using telemedicine: a systematic review. *Cost Eff Resour Alloc*. 2020;18:16.
6. Vujosevic S, Aldington SJ, Silva P, et al. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol*. 2020;8:337−347.
7. Piyasena MMPN, Murthy GVS, Yip JLY, et al. Systematic review on barriers and enablers for access to diabetic retinopathy screening services in different income settings. *PLOS ONE*. 2019;14:e0198979.
8. DeBuc DC. The role of retinal imaging and portable screening devices in tele-ophthalmology applications for diabetic retinopathy management. *Curr Diab Rep*. 2016;16:132.

9. Sengupta S, Sindal MD, Baskaran P, et al. Sensitivity and specificity of smartphone-based retinal imaging for diabetic retinopathy: a comparative study. *Ophthalmol Retina*. 2019;3:146−153.
10. Xiao B, Liao Q, Li Y, et al. Validation of handheld fundus camera with mydriasis for retinal imaging of diabetic retinopathy screening in China: a prospective comparison study. *BMJ Open*. 2020;10:e040196.
11. Salongcay RP, Aquino LAC, Salva CMG, et al. Comparison of handheld retinal imaging with ETDRS 7-standard field photography for diabetic retinopathy and diabetic macular edema. *Ophthalmol Retina*. 2022;6:548−556.
12. 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.
13. Gulshan V, Rajan RP, Widner K, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol*. 2019;137:987−993.
14. Kanagasingam Y, Xiao D, Vignarajan J, et al. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. *JAMA Netw Open*. 2018;1:e182665−e.
15. US-FDA. *De Novo Classification Request for IDx-DR. 11 April 2018*. Maryland, USA: US Food and Drug Administration; 2018.
16. US-FDA. *Aeye-DS - 510(k) Premarket Notification*, 10 November 2022. Maryland, USA: US Food and Drug Administration; 2022.
17. US-FDA. *Eyenuk's EyeArt 510(k) Summary*, 03 August 2020. Maryland, USA: US Food and Drug Administration; 2020: K200667.
18. 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;318:2211−2223.
19. Federation ID. *IDF Diabetes Atlas*. 10th ed. International Diabetes Federation; 2021.
20. Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003-2016. *Acta Diabetol*. 2017;54:515−525.
21. Cavallerano AA, Cavallerano JD, Katalinic P, et al. A telemedicine program for diabetic retinopathy in a Veterans Affairs Medical Center—the Joslin Vision Network Eye Health Care Model. *Am J Ophthalmol*. 2005;139:597−604.
22. Curran K, Congdon N, Hoang TT, et al. Impact of targeted diabetic retinopathy training for graders in Vietnam and the implications for future diabetic retinopathy screening programmes: a diagnostic test accuracy study. *BMJ Open*. 2022;12:e059205.
23. US-FDA. *FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems*. Maryland, USA: US Food and Drug Administration; 2018. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye. Accessed January 8, 2024.
24. US-FDA. *Aeye-DS Device 510(k) Number K221183. 10 November 2022*. Maryland, USA: US Food and Drug Administration; 2022.
25. US-FDA. *K200667 - 510(k) Premarket Notification. 03 August 2020*. Maryland, USA: US Food and Drug Administration; 2020.
26. van der Heijden AA, Abramoff MD, Verbraak F, et al. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol*. 2018;96:63−68.
27. Campbell JP, Mathenge C, Cherwek H, et al. Artificial intelligence to reduce ocular health disparities: moving from concept to implementation. *Transl Vis Sci Technol*. 2021;10:19.
28. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72:100759.
29. Ibrahim H, Liu X, Zariffa N, et al. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health*. 2021;3:e260−e265.
30. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447−453.
31. Chia MA, Hersch F, Sayres R, et al. Validation of a deep learning system for the detection of diabetic retinopathy in Indigenous Australians. *Br J Ophthalmol*. 2023:bjo−2022.
32. 31st Meeting of the European Association for the Study of Diabetes Eye Complications Study Group (EASDec) Odense, Denmark, 28th − 30th October 2021. *Eur J Ophthal*. 2021;31(2_suppl):1−41.
33. World Health Organization. Regional Office for E. Diabetic Retinopathy Screening: a Short Guide: Increase Effectiveness, Maximize Benefits and Minimize Harm. World Health Organization. Regional Office for Europe; 2020.
34. Jacoba CMP, Doan D, Salongcay RP, et al. Performance of automated machine learning for diabetic retinopathy image classification from multi-field handheld retinal images. *Ophthalmol Retina*. 2023;7:703−712.