



Screening Referable Diabetic Retinopathy Using a Semi-automated Deep Learning Algorithm Assisted Approach

Yueye Wang^{1†}, Danli Shi^{1†}, Zachary Tan^{2†}, Yong Niu³, Yu Jiang¹, Rulin Xiong¹, Guankai Peng⁴ and Mingguang He^{1,2,5,6*}

¹ State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China, ² Centre for Eye Research Australia, University of Melbourne, Melbourne, VIC, Australia, ³ Department of Ophthalmology, Guangzhou No. 11 People's Hospital, Guangzhou, China, ⁴ Guangzhou Vision Tech Medical Technology Co. Ltd., Guangzhou, China, ⁵ Department of Ophthalmology, Guangdong Eye Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, ⁶ Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:

Erping Long,
National Institutes of Health (NIH),
United States

Reviewed by:

Yih Chung Tham,
Singapore Eye Research Institute
(SERI), Singapore
Wei Han,
Zhejiang University, China
Hossam El-Din Moustafa,
Mansoura University, Egypt

*Correspondence:

Mingguang He
mingguang.he@unimelb.edu.au

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Ophthalmology,
a section of the journal
Frontiers in Medicine

Received: 14 July 2021

Accepted: 08 November 2021

Published: 25 November 2021

Citation:

Wang Y, Shi D, Tan Z, Niu Y, Jiang Y,
Xiong R, Peng G and He M (2021)
Screening Referable Diabetic
Retinopathy Using a Semi-automated
Deep Learning Algorithm Assisted
Approach. *Front. Med.* 8:740987.
doi: 10.3389/fmed.2021.740987

Purpose: To assess the accuracy and efficacy of a semi-automated deep learning algorithm (DLA) assisted approach to detect vision-threatening diabetic retinopathy (DR).

Methods: We developed a two-step semi-automated DLA-assisted approach to grade fundus photographs for vision-threatening referable DR. Study images were obtained from the Lingtong Cohort Study, and captured at participant enrollment in 2009–2010 (“baseline images”) and annual follow-up between 2011 and 2017. To begin, a validated DLA automatically graded baseline images for referable DR and classified them as positive, negative, or ungradable. Following, each positive image, all other available images from patients who had a positive image, and a 5% random sample of all negative images were selected and regraded by trained human graders. A reference standard diagnosis was assigned once all graders achieved consistent grading outcomes or with a senior ophthalmologist’s final diagnosis. The semi-automated DLA assisted approach combined initial DLA screening and subsequent human grading for images identified as high-risk. This approach was further validated within the follow-up image datasets and its time and economic costs evaluated against fully human grading.

Results: For evaluation of baseline images, a total of 33,115 images were included and automatically graded by the DLA. 2,604 images (480 positive results, 624 available other images from participants with a positive result, and 1500 random negative samples) were selected and regraded by graders. The DLA achieved an area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy of 0.953, 0.970, 0.879, and 88.6%, respectively. In further validation within the follow-up image datasets, a total of 88,363 images were graded using this semi-automated approach and human grading was performed on 8975 selected images. The DLA achieved an AUC, sensitivity, and specificity of 0.914, 0.852, 0.853, respectively. Compared against fully human grading, the semi-automated DLA-assisted approach achieved an estimated 75.6% time and 90.1% economic cost saving.

Conclusions: The DLA described in this study was able to achieve high accuracy, sensitivity, and specificity in grading fundus images for referable DR. Validated against long-term follow-up datasets, a semi-automated DLA-assisted approach was able to accurately identify suspect cases, and minimize misdiagnosis whilst balancing safety, time, and economic cost.

Keywords: diabetic retinopathy, artificial intelligence, screening, cost-saving analysis, deep learning

INTRODUCTION

Along with rapid increases in the prevalence of diabetes worldwide, growing numbers of patients are at risk of developing vision-threatening diabetic retinopathy (DR) that affects quality of life. DR has been reported in 103.1 million patients in 2020 and is estimated to affect 160.5 million in 2045 (1). Early detection and timely intervention are critical to prevent DR-related severe visual loss (2).

Regular screening is an effective strategy to identify DR patients amongst a large population, and those suitable for referral to ophthalmologists for treatment (3). However, annual screening as recommended by guidelines is poorly adhered to (4–6), especially in low- and middle- income regions (7). Given the uneven distribution of medical resources and retinal specialists in many countries, new methods and technologies are required to improve upon current screening strategies (8).

Automating the detection and classification of eye disease has received growing attention as a means to increase access to screening. Deep learning techniques are able to automatically capture and learn the most predictive features for classification from large training datasets (9). With no need to specify rules explicitly, deep learning algorithm (DLA)-based artificial intelligence AI systems have demonstrated promising accuracy and efficacy in detecting vision-threatening DR from fundus photographs (10–14), and may represent a cost-effective alternative to human grading. The implementation of these algorithms has the promise to enable more affordable, rapid, and consistent DR diagnosis.

Although now described extensively in the literature, the clinical application and translation of this AI technology remain limited. Researchers have made efforts to validate AI in clinical practice (15), revealing multiple socio-environmental factors that restrict the accuracy and adoption of this new technology (16). Clinical workflows present complex challenges including potential ethical, bias and generalizability limitations, and medicolegal implications when integrating these tools into existing clinical pathways (17). Further, the applicability and consistency of AI systems in detecting longitudinal change in single individuals has also not been evaluated.

As the clinical implementation of AI is challenging and underexplored, we thus conducted this study to integrate this

technology into a research image dataset for the screening of vision-threatening referable DR. We tested a previously established DLA system in digital fundus photographs collected from the Lingtou Eye Cohort Study, which consisted of non-mydratic fundus images routinely captured at annual health screening within primary care centers for enrolled general Chinese participants. In addition, to achieve improved screening performance, we developed a novel semi-automated DLA-assisted approach that combined both AI and human grading procedures, and tested its performance within longitudinal datasets. Lastly, to evaluate cost-savings, time and economic outlays for this semi-automated approach were compared against fully human grading alone (**Supplementary Figure 1**).

MATERIALS AND METHODS

Study Design

The main aim of the study was to validate the screening performance of a previously established DLA against manual grading as a reference standard in a research image dataset. We then developed a novel semi-automated DLA-assisted workflow that integrated this DLA into a manual grading workflow and evaluated potential time and economic cost savings.

To test DLA performance in detecting vision-threatening referable DR (pre-proliferative DR or worse), a total of 33,115 fundus photographs captured at participant enrollment from the Lingtou Eye Cohort Study (“baseline images”) were included in this present study. The Lingtou Eye Cohort Study is an ongoing prospective cohort study that enrolled government employees attending the Guangzhou Government Servant Physical Check-up Center, Lingtou, China. At baseline recruitment in 2009–2010, a total of 4,939 participants were enrolled and all subsequently invited to take part in annual follow-up at primary care centers, including physical and ophthalmic examinations and health questionnaires. Detailed study methodology has been reported previously (18). Written informed consent was obtained from all participants.

Collection of Fundus Photographs

Non-mydratic standard digital fundus photographs were captured of each eye using a fundus camera (TRC-NW6S; Topcon, Tokyo, Japan) in two specified positions: centered on the optic disc (F1 image) and macular fovea (F2 image). Among 4,939 participants included at baseline, 39 did not receive fundus photography due to lack of cooperation, rejection of ophthalmic examination, amongst other reasons. Images from the remaining 4,900 participants were captured. During annual

Abbreviations: DLA, deep learning algorithm; DR, diabetic retinopathy; AUC, area under the receiver operating characteristic curve; CBIA, computer-based image analysis; AI, artificial intelligence; NHS, National Health Service; CI, confidence intervals; κ W, weighted kappa; USD, United States Dollars.

follow-up from 2011 to 2017, the numbers of participants with fundus photographs captured annually were 3,505, 3,325, 3,198, 3,104, 3,038, 2,873, and 2,579, respectively.

Original images included from baseline and follow-up were all jpeg files with a resolution of $2,000 \times 1,980$ pixels. Pre-processing was performed to normalize original images into an appropriate format for the DLA. Firstly, all images were clipped to 90% of its original size to remove the “boundary effect” and following, images were resized to a resolution of 299×299 pixels with red-green-blue (RGB) channels. Further, to eliminate potential noise in images, the local average color was subtracted and mapped to 50% gray.

Development of Semi-automated DLA-Assisted Approach

Using an Inception-v3 convolutional neural network, we previously trained an AI-based DLA for the automated detection of referable DR. The Inception-v3 neural network structure for the DLA was trained from scratch with a mini-batch gradient descent size of 32 and Adam optimizer of a 0.002 learning rate (19). After image normalization, the input of each image was transformed into a standard format of $299 \times 299 \times 3$, while outputs were warped into probability distributions (**Supplementary Figure 2**). A total of 71,043 retinal photographs (both F1 and F2 images) from different hospitals and clinics in China were used for training and internal validation. The formats of images used during development of the DLA were the same as those in this study. Detailed descriptions of this DLA have been reported in our previous study (19).

For these training and internal validation data, criteria used for grading DR originate from the National Health Service (NHS) (20). Images were categorized as R0 (no DR), R1 (background DR), R2 (pre-proliferative DR), and R3 (proliferative DR). Images categorized as R3 were further subcategorized as R3a for active disease, and R3s for stable disease. Images of poor quality or poor positioning were defined as ungradable. Vision-threatening referable DR was defined as either pre-proliferative or proliferative DR (R2, R3a, and R3s as per NHS guidelines). Detailed criteria for classification and fundus images for typical cases are shown in **Supplementary Table 1** and **Supplementary Figure 3**.

Performance of DLA grading was validated using images from the Lingtong Cohort and compared against trained human grader results as reference. Manual grading was carried out by three trained graders and two licensed ophthalmologists. All graders were masked to participant diagnoses and baseline characteristics to minimize bias and ensure comprehensive validation of the DLA's initial diagnosis. If grading outcomes were not consistent among graders, images would be reviewed by an assigned senior ophthalmologist, with diagnoses assigned at this step considered conclusive. A reference standard grading would be assigned once all graders achieved consistent grading outcomes or with a senior ophthalmologist's final diagnosis.

Utilizing the DLA system we previously established, we carried out a two-step semi-automated DLA-assisted approach to screen images sourced from the cohort study. Firstly, using

this DLA, all images captured at baseline and at follow-up were screened for referable DR and classified as positive, negative, or ungradable. Secondly, all images classified as positive by the DLA, and all other available images from the same patient (even if classified as negative or ungradable) would be selected and regraded by trained graders. In addition, images initially graded as negative by the DLA were randomly sampled for further trained grader review. Approximately 5% of images were randomly sampled and independently regraded by three trained graders. Finally, any images where there was disagreement between the DLA and human grader were manually reviewed, to investigate the causes that may underlie false DLA diagnoses.

Cost-Saving Evaluation

We recorded the total working time and economic costs of grading for this semi-automated DLA-assisted approach and compared it to a fully human grading process. The cost-saving analysis were performed based on follow-up datasets with a total of 88,363 images.

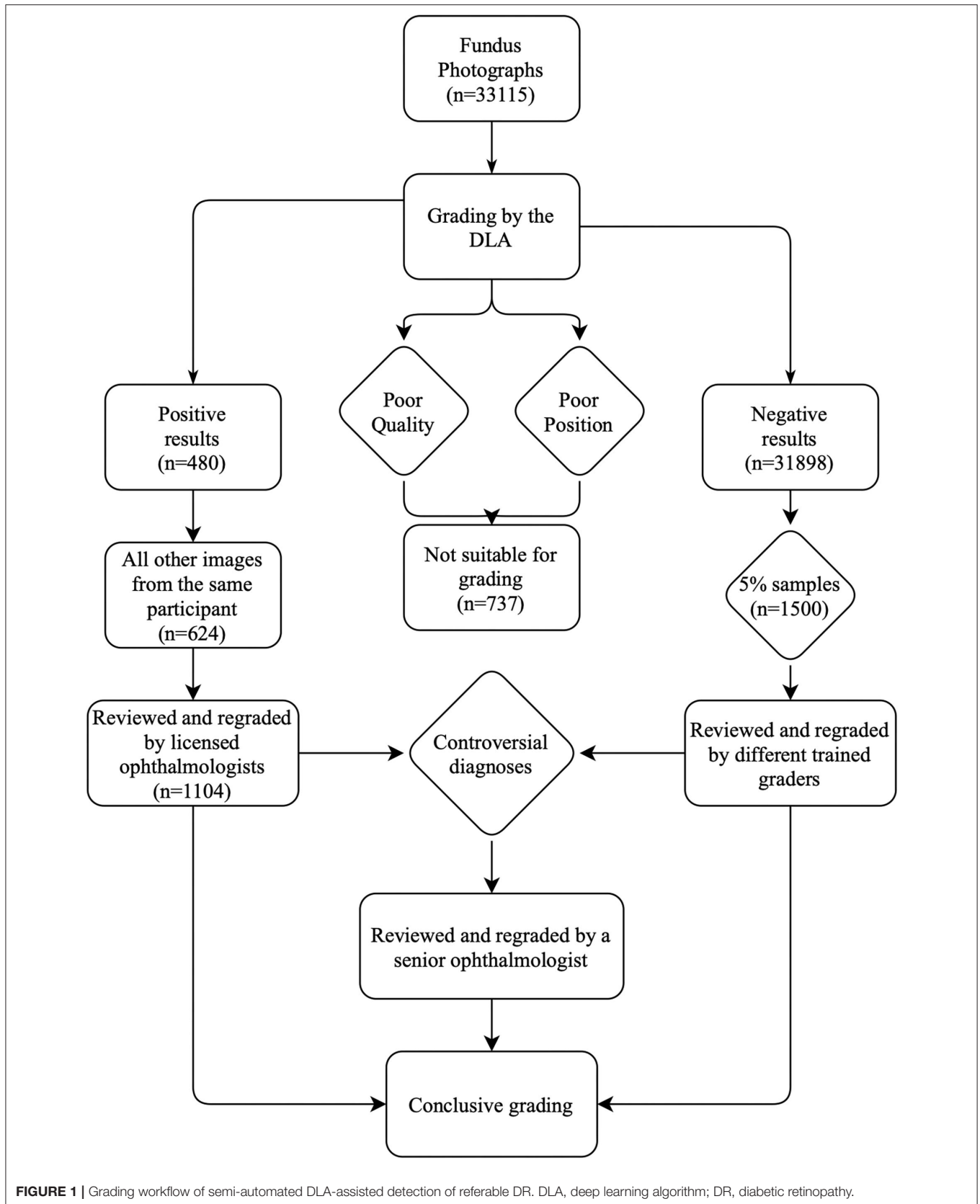
The total working time for the semi-automated approach was calculated as the sum of time required for the DLA to grade all images, and time required for subsequent human grader review of flagged images. As only flagged images within the follow-up datasets were reviewed by human graders, the time-cost human graders would require to grade the full dataset was calculated by multiplying the average time required per image across the total number of images.

In the calculation of economic costs of grading, only costs of human grading (i.e., payment to human graders as remuneration for their time) were considered in the cost-saving evaluation of the semi-automated and fully manual grading approaches. DLA costs were not considered as the marginal cost of DLA operation is minimal, and the cost of DLA development was outside the scope of this study. Time and economic costs for grading per image were evaluated and compared across the fully

TABLE 1 | Characteristics of baseline participants with fundus images.

Characteristics	Baseline (2009–2010)
No. of participants	4,900
Age (mean \pm sd, yrs)	59.1 \pm 8.78
Male (n, %)	2,846 (58.5%)
Smoke (n, %)	823 (16.9%)
Alcohol consumption (n, %)	2,545 (52.2%)
BMI (mean \pm sd, Kg/m ²)	24 \pm 3.00
SBP (mmHg)	129 \pm 18.0
DBP (mmHg)	75.0 \pm 11.3
Fast blood glucose (mmol/L)	5.60 \pm 1.17
Triglyceride (mmol/L)	1.75 \pm 1.43
Cholesterol (mmol/L)	5.45 \pm 0.95
Diabetes (n, %)	591 (12.0%)
Diabetes duration of diabetic patients (mean \pm sd, yrs)	6.44 \pm 6.01

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure.



DLA grading, semi-automated DLA-assisted grading, and fully human grading approaches.

Statistical Analysis

To assess the validity of classification by the DLA; sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC) with 95% confidence intervals (CI) were calculated and compared to human grading results. A

weighted kappa (κ_w) statistic (with Cicchetti–Allison weighting) was used to compare agreement between the DLA and human grading.

Statistical analyses were performed using standard statistical software (R, v.4.0.4; Stata, v.15.1).

RESULTS

Assessment of DLA and Human Grading

A total of 33,115 fundus photographs were captured at participant enrollment for the Lingtou Eye Cohort Study (“baseline images”) and screened for referable DR. Baseline clinical characteristics of included participants were shown in **Table 1**. Among these images, 737 (2.23%) were classified as ungradable due to poor quality or poor position, which may result from extremely small pupils or eye movement during image capture. A total of 32,378 images were conclusively graded by the DLA. For further human grading, images graded as positive by the DLA ($n = 480$, 1.45%), other images from participants with an image graded as positive by the DLA ($n = 624$, 1.88%), and randomly selected negative samples ($n = 1500$, 4.5%), were reviewed (**Figure 1**). After manually grading

TABLE 2 | Confusion matrix of DLA grading results for baseline images.

No. of fundus images		Reference (human grading)		Total
		Positive	Negative	
DLA grading	Positive	(TP) 191	(FP) 289	480
	Negative	(FN) 7	(TN) 2,117	
Total		198	2,124	2,604

DLA, deep learning algorithm; TP, true-positive; FP, false-positive; FN, false-negative; TN, true-negative.

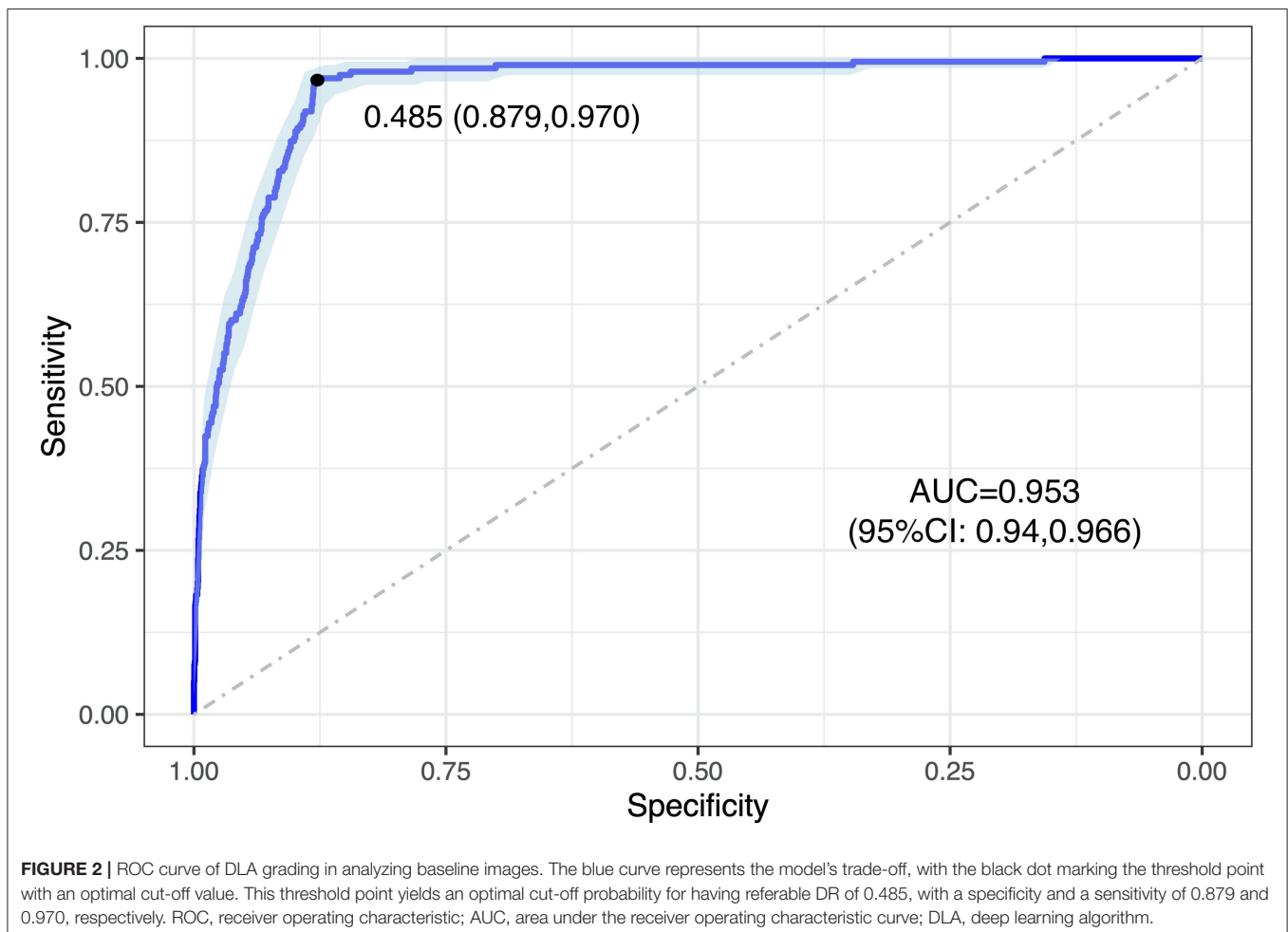


TABLE 3 | Features and numbers of DLA-classified false-positive and false-negative referable DR cases.

Features	No.	Proportion
False-positive images		
Myopic retinopathy	102	35.29%
Normal fundus with artifacts	80	27.68%
Opaque refracting media	32	11.07%
Age-related macular degeneration	28	9.69%
Retinal pigment changes	17	5.88%
Retinal atrophy	9	3.11%
Retinal vessel occlusion	8	2.77%
Background DR	2	0.69%
Others	11	3.81%
Total	289	100.00%
False-negative images		
Intraretinal microvascular abnormalities	3	42.86%
Blurred peripheral retina exudation	2	28.57%
Diabetic macular exudation	1	14.29%
Questionable new vessels	1	14.29%
Total	7	100.00%

DLA, deep learning algorithm; DR, diabetic retinopathy.

by trained human graders, 191 images were classified as true-positive (TP) and 2,117 as true-negative (TN), while 289 were classified as false-positive (FP) and 7 as false-negative (FN) (Table 2).

As a result, 88.6% [(TP + TN)/total graded images] of the baseline images were correctly graded by the DLA compared to human grading. Agreement between the DLA and human grading was acceptable with a κ W of 0.511 (95% CI: 0.465–0.557). Compared to human grading, the AUC, sensitivity, and specificity of the DLA were 0.953 (95% CI: 0.94–0.966), 0.970, and 0.879 respectively, with a cutoff point of 0.485 (Figure 2).

The most common cause of false-positive cases ($n = 289$) was myopic retinopathy [$n = 102$ (35.29%)], associated with diffuse or patchy choroidal atrophy, while intraretinal microvascular abnormalities [$n = 3$ (42.86%)] were most commonly responsible for the false-negative cases ($n = 7$). Features and examples of typical cases classified by the DLA are shown in Table 3 and Figure 3.

Semi-automated DLA-Assisted Grading Approach Testing

To test the reproducibility and reliability of the semi-automated DLA-assisted grading approach, we applied this protocol to fundus images captured at annual Lingtou Eye Cohort Study follow-ups from 2011 to 2017. A total of 88,363 fundus photographs from these 7 years were obtained and graded. The percentage of ungradable images due to poor quality or positioning was on average 4.72% across the annual datasets. DLA performance across these years is shown in Table 3. As per the semi-automated study protocol, 8,975 images with either positive DLA grading or other images from participants with an image graded as positive by the DLA were selected and re-graded by human graders. Compared against human grading,

the AUC, sensitivity, and specificity of the DLA across the different follow-up years were 0.914 (0.87–0.939), 0.852 (0.718–0.915), 0.853 (0.741–0.918) respectively, as shown in Table 4 and Supplementary Figure 4.

For the fully automated DLA grading process, all 88,363 follow-up images were classified within 84.8 h, with an average grading time of 0.058 min per image. In the human grading process, three graders and two ophthalmologists combined required 55 h on the whole to grade a total of 8,975 flagged images, with an average grading time of 0.368 min per image. For this, human graders were paid a total of \$1544 United States Dollars (USD) (\$0.172 per image). The hypothesized time and economic cost that would be required for human graders to grade all follow-up images was calculated as ~543 h and \$15,202 USD.

The total time required for semi-automated DLA-assisted analysis of the follow-up datasets was 139 h (0.09 min per image), and the total cost ~\$1544 USD (\$0.017 per image). This represented an estimated 75.6% time and 90.1% cost saving compared to entirely human grading of the same dataset. Comparison of the total and mean time and cost of the different grading protocols (fully automated DLA grading, human grading, and semi-automated DLA-assisted grading) are shown in Table 5.

DISCUSSION

Although various AI-based algorithms have been developed for DR detection, implementation within clinical practice remains nascent and under investigation. In this study, we developed a semi-automated DLA-assisted approach to diagnose referable DR and validated it across 121,478 images collected longitudinally from a cohort study. This semi-automated DLA-assisted approach presented advantages in time and economic savings for grading, with combined DLA and human grading enabling accurate and efficient diagnoses.

With rapid advances in the development of AI technology for detecting DR, some studies have investigated the clinical application of AI-based grading and diagnosis (21, 22). Within limited samples, the performance of AI tools has been impressive. He et al. conducted a study of AI-based screening for DR amongst 889 diabetic patients at a Chinese community hospital, achieving high sensitivity (90.8%) and specificity (98.5%) in the detection of DR (23). Real-world performance may however be affected in larger populations. According to a recent multicenter study, 7 DR screening algorithms had significant variability in performance on real-world clinical data, with sensitivities varying from 50.98 to 85.90% (24). Shah et al. evaluated the performance of an AI algorithm in the detection of referable DR in a screening program with 2,680 diabetic patients, achieving 100% sensitivity and 82% specificity (25). These results however may not be ideal in a primary care screening setting given relatively low expected positive predictive value, which may result in individuals who are healthy or have background DR mistakenly flagged for referral (26). DR screening guidelines issued by Diabetes UK recommend a minimum of 95% specificity, which remains a challenging threshold for AI grading to achieve in large real-world settings (27). Within our 8-year longitudinal cohort study dataset, our DLA was not able to achieve the

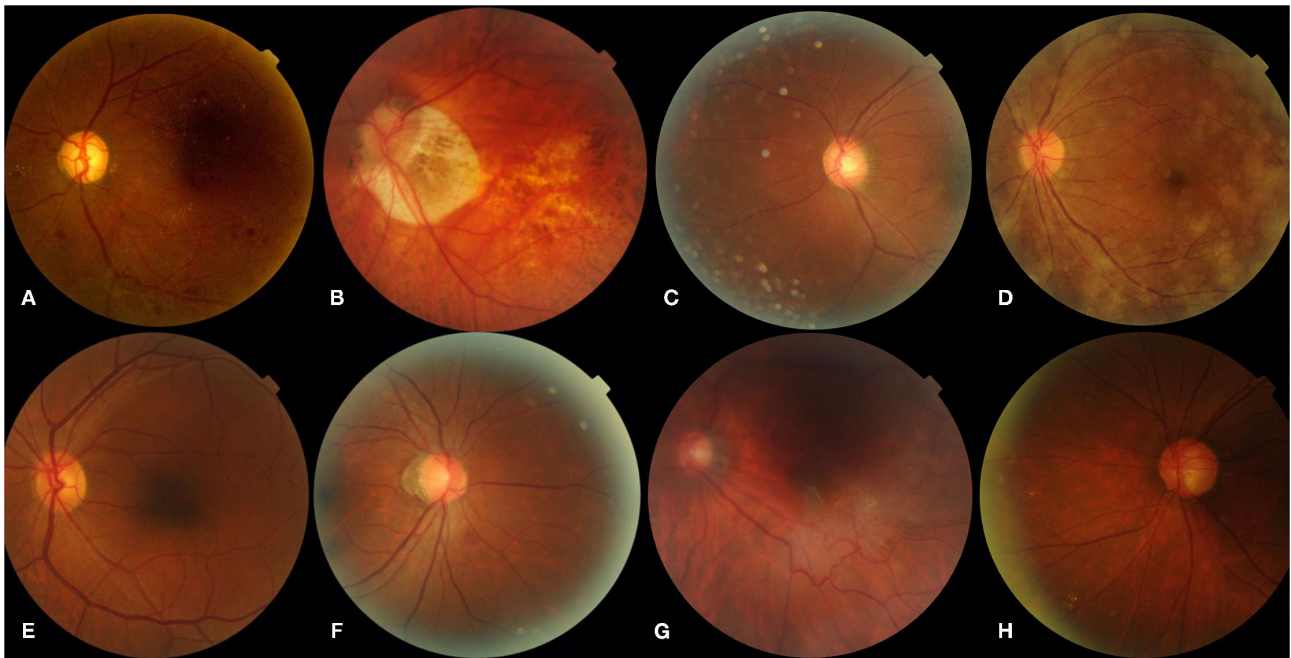


FIGURE 3 | Representative sample of typical images classified by the DLA. **(A)** Represents a true-positive image, **(B–D)** represent typical false-positive images, **(E)** represents a true-negative image, **(F–H)** represent typical false-negative images. **(A)** Pre-proliferative DR (R2) with microaneurysms, multiple blot hemorrhages, hard exudates; **(B)** myopic retinopathy; **(E)** normal fundus with artifacts; **(D)** opaque refracting media; **(A,E)** normal fundus; **(F)** intra-retinal microvascular abnormalities; **(G)** questionable new vessels; **(H)** blurred peripheral retina exudates. DLA, deep learning algorithm.

TABLE 4 | DLA grading results across the baseline and 2011–2017 follow-up cohorts.

Grading for referable DR	Baseline	2011	2012	2013	2014	2015	2016	2017	Total in follow-up
No. of participants	4,900	3,505	3,325	3,198	3,104	3,038	2,873	2,579	-
No. of included images	33,115	13,927	13,590	13,089	12,929	12,303	11,749	10,776	88,363
Positive images (n, %)	480 (1.45%)	186 (1.34%)	185 (1.36%)	169 (1.29%)	182 (1.41%)	199 (1.62%)	196 (1.67%)	150 (1.39%)	1267 (1.43%)
Negative images (n, %)	31,898 (96.3%)	13,363 (96.0%)	12,992 (95.6%)	12,378 (94.6%)	11,834 (91.5%)	11,316 (92.0%)	10,998 (93.6%)	10,094 (93.7%)	82,975 (93.9%)
Ungradable images (n, %)	737 (2.23%)	378 (2.71%)	413 (3.04%)	542 (4.14%)	913 (7.06%)	788 (6.40%)	555 (4.72%)	532 (4.94%)	4121 (4.72%)
No. of images graded by human	2,604	1,628	1,419	1,355	1,333	1,179	1,099	962	8,975
Positive images by human grading	198	57	62	55	80	76	91	71	492
AUC	0.953	0.870	0.910	0.919	0.939	0.921	0.937	0.899	0.914
Sensitivity	0.970	0.903	0.855	0.873	0.887	0.816	0.915	0.718	0.852
Specificity	0.879	0.741	0.883	0.835	0.867	0.900	0.824	0.918	0.853
Accuracy of the DLA	0.886	0.921	0.913	0.916	0.923	0.896	0.904	0.918	0.913

DLA, deep learning algorithm; DR, diabetic retinopathy; AUC, area under the receiver operating characteristic curve.

recommended 95% specificity standard. In light of the challenges of using AI to completely replace human grading, a more practical solution may be the integration of AI into existing DR screening workflows. The semi-automated DLA-assisted approach described in this present study may be more feasible in achieving the requisite accuracy.

To implement new AI diagnostic strategies within clinical practice, three different general models have been previously

proposed: triage, replacement, and add-on (28). All these models have been trialed for DR screening in previous attempts. Tufail et al. trialed a replacement and add-on model for DR screening, where automated retinal image analysis served as either an alternative for human grading, or as a filter prior to manual grading (29). Within a real-world screening environment, both the replacement and add-on filter strategies achieved acceptable accuracy at a lower overall cost of grading compared to fully

TABLE 5 | Comparison of time and cost for grading follow-up images between different grading procedures.

Cost-saving analysis	Fully human grading		Fully automatic DLA grading (<i>n</i> = 88,363)	Semi-automated DLA-assisted grading (<i>n</i> = 88,363)
	For flagged images (<i>n</i> = 8975)	Hypothesized for whole dataset (<i>n</i> = 88,363)		
Time for grading (h)	55	543	84.8	139
Mean time of grading per image (min)	0.368	0.368	0.058	0.094
Total cost of grading (\$)	1,544	15,202	–	1,544
Mean cost of grading per image (\$)	0.172	0.172	–	0.017

DLA, deep learning algorithm.

human grading. In triage models, AI tools are likely to be adopted as diagnostic decision-making supports, that generate immediate reports for qualified clinicians to review. One triage model that adopted AI in endocrinology and primary care settings within Australia reduced the workload for telemedicine graders by over 50% (17).

The semi-automated DLA-assisted approach described in this present study may be adopted within a triage model, and potentially reduce the workload of human graders. For instance, images in which the DLA has deemed to be highly likely to be negative may not require further review, enabling human graders to focus their review on images in which the pre-grading probability of DR is significantly higher. We were able to validate the accuracy of this semi-automated DLA-assisted approach within a general population with a significantly lower prevalence of DR (2), compared to previous validation studies that were performed in diabetic populations with a significantly higher existing prevalence of DR (30). This is an important contribution in validating the adoption of AI-based DR screening for general populations, including within primary care, where there is a very low pre-test probability of patients having referable DR. In addition, our findings validate the performance of a semi-automated DLA-assisted approach, where DLA graded negative images can safely not require further human grader review.

Despite requiring additional human grader review for some images, our semi-automated DLA-assisted approach achieved substantial cost savings compared to a fully human grading approach. Previous studies including the Singapore Epidemiology of Eye Diseases Study evaluated the cost-saving of two AI-based DR screening models, comparing a fully automated approach and a semi-automated approach (in which the AI served as a triage filter), against fully human grading (31). That study found the semi-automated approach to be the least expensive, achieving an estimated 20% saving in current annual screening costs if applied across the health system. Similarly, the semi-automated DLA-assisted approach described in this study may represent a significant cost saving compared to a fully human grading approach, particularly when applied to screening large general populations in primary care settings. These time and economic savings may be increasingly important going forward, as the requirement for grading and diabetic retinopathy screening increases with the growing population prevalence of diabetes. In this present study, we have shown that a semi-automated DLA approach may address some of these challenges.

The findings of this study must be considered in the context of its limitations. Firstly, the semi-automated DLA-assisted approach described in this study was only assessed within the Lingtuo Cohort Study datasets, which is representative only for the primary care setting within this region. Further application in other settings requires caution. Secondly, ~5% of images were ungradable due to poor quality or poor positioning, which have may resulted in some positive DR cases being missed. Established standards for fundus photography should be strictly followed to minimize potential ungradable images. Thirdly, fundus images used in this study were two-field, non-stereoscopic images, which were easy and quick to capture in primary care clinics, but may have reduced DR detection compared to gold standard seven-field stereoscopic images. Additional studies may be required to better understand and optimize the means in which these diagnostic strategies are integrated into existing clinical practice workflows, optimizing for accuracy, cost, and accessibility.

In conclusion, the DLA described in this study was able to achieve high accuracy, sensitivity, and specificity in detecting vision-threatening referable DR. A semi-automated DLA-assisted approach that integrates initial automated DLA diagnosis and human grading for high-risk cases, was able to accurately identify suspect DR and avoid unnecessary review. Validated across multiple datasets, this approach minimized misdiagnosis whilst balancing safety, time, and cost. This approach has potential for adoption in increasing DR screening efficiency, enabling clinicians to meet the increasing eye care demands of rapidly growing global diabetes prevalence.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study was approved by the Ethics Committee of the Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants.

AUTHOR CONTRIBUTIONS

MH: study conception, design, and supervision. YW, DS, YN, YJ, and RX: acquisition, analysis, or interpretation. YW and ZT: drafting of the manuscript. MH, DS, ZT, and GP: critical revision of the manuscript for important intellectual content. YW, YJ, and RX: statistical analysis. YN and MH: administrative, technical, or material support. All authors contributed to the article and approved the submitted version.

FUNDING

The present work was supported by the Fundamental Research Funds of the State Key Laboratory of Ophthalmology, Project of Investigation on Health Status of Employees in Financial

Industry in Guangzhou, China (Z012014075), and Science and Technology Program of Guangzhou, China (202002020049). MH receives support from the University of Melbourne at Research Accelerator Program and the CERA Foundation. The Centre for Eye Research Australia receives Operational Infrastructure Support from the Victorian State Government. The sponsor or funding organization had no role in the design or conduct of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.740987/full#supplementary-material>

REFERENCES

- Teo ZL, Tham YC, Yu M, Chee ML, Rim TH, Cheung N, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*. (2021) 128:1580–91. doi: 10.1016/j.ophtha.2021.04.027
- Jampol LM, Glassman AR, Sun J. Evaluation and care of patients with diabetic retinopathy. *N Engl J Med*. (2020) 382:1629–37. doi: 10.1056/NEJMra1909637
- Scanlon PH. Screening intervals for diabetic retinopathy and implications for care. *Curr Diab Rep*. (2017) 17:96. doi: 10.1007/s11892-017-0928-6
- Rathi S, Tsui E, Mehta N, Zahid S, Schuman JS. The current state of teleophthalmology in the United States. *Ophthalmology*. (2017) 124:1729–34. doi: 10.1016/j.ophtha.2017.05.026
- Ophthalmology AAO. *Diabetic Retinopathy PPP*. (2019). Available online at: <https://www.aao.org/preferred-practice-pattern/diabetic-retinopathy-ppp> (accessed June 15, 2021).
- Schoenfeld ER, Greene JM, Wu SY, Leske MC. Patterns of adherence to diabetes vision care guidelines: baseline findings from the diabetic retinopathy awareness program. *Ophthalmology*. (2001) 108:563–71. doi: 10.1016/S0161-6420(00)00600-X
- Beagley J, Guariguata L, Weil C, Motala AA. Global estimates of undiagnosed diabetes in adults. *Diabetes Res Clin Pract*. (2014) 103:150–60. doi: 10.1016/j.diabres.2013.11.001
- Teo ZL, Tham YC, Yu M, Cheng CY, Wong TY, Sabanayagam C. Do we have enough ophthalmologists to manage vision-threatening diabetic retinopathy? A global perspective. *Eye*. (2020) 34:1255–61. doi: 10.1038/s41433-020-0776-5
- Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T, et al. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol*. (2020) 8:337–47. doi: 10.1016/S2213-8587(19)30411-5
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. (2016) 316:2402–10. doi: 10.1001/jama.2016.17216
- De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. (2018) 24:1342–50. doi: 10.1038/s41591-018-0107-6
- 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. doi: 10.1016/j.ophtha.2018.01.023
- Milea D, Najjar RP, Zubo J, Ting D, Vasseneix C, Xu X, et al. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med*. (2020) 382:1687–95. doi: 10.1056/NEJMoa1917130
- Tan Z, Simkin S, Lai C, Dai S. Deep learning algorithm for automated diagnosis of retinopathy of prematurity plus disease. *Transl Vis Sci Technol*. (2019) 8:23. doi: 10.1167/tvst.8.6.23
- Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. (2018) 1:39. doi: 10.1038/s41746-018-0040-6
- Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI: Association for Computing Machinery. (2020). p. 1–12.
- He M, Li Z, Liu C, Shi D, Tan Z. Deployment of artificial intelligence in real-world practice: opportunity and challenge. *Asia Pac J Ophthalmol*. (2020) 9:299–307. doi: 10.1097/APO.0000000000000301
- Hu Y, Niu Y, Wang D, Wang Y, Holden BA, He M. The association of longitudinal trend of fasting plasma glucose with retinal microvasculature in people without established diabetes. *Invest Ophthalmol Vis Sci*. (2015) 56:842–8. doi: 10.1167/iovs.14-15943
- Li Z, Keel S, Liu C, He Y, Meng W, Scheetz J, 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. doi: 10.2337/dc18-0147
- England PH. *NHS Diabetic Eye Screening Program: Grading Definitions for Referable Disease*. (2016). Available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/582710/Gradingdefinitionsforreferable2017new110117.pdf (accessed May 2, 2015).
- Gunasekaran DV, Ting DSW, Tan GSW, Wong TY. Artificial intelligence for diabetic retinopathy screening, prediction and management. *Curr Opin Ophthalmol*. (2020) 31:357–65. doi: 10.1097/ICU.0000000000000693
- Scheetz J, Koca D, McGuinness M, Holloway E, Tan Z, Zhu Z, 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. doi: 10.1038/s41598-021-94178-5
- He J, Cao T, Xu F, Wang S, Tao H, Wu T, et al. Artificial intelligence-based screening for diabetic retinopathy at community hospital. *Eye*. (2020) 34:572–6. doi: 10.1038/s41433-019-0562-4
- Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee YE, 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. doi: 10.2337/dc20-1877

25. Shah A, Clarida W, Amelon R, Hernaez-Ortega MC, Navea A, Morales-Olivas J, et al. Validation of automated screening for referable diabetic retinopathy with an autonomous diagnostic artificial intelligence system in a spanish population. *J Diabetes Sci Technol*. 15:655–63. doi: 10.1177/1932296820906212
26. Cuadros J. The real-world impact of artificial intelligence on diabetic retinopathy screening in primary care. *J Diabetes Sci Technol*. 15:664–65. doi: 10.1177/1932296820914287
27. Scanlon PH. The english national screening programme for diabetic retinopathy 2003-2016. *Acta Diabetol*. (2017) 54:515–25. doi: 10.1007/s00592-017-0974-1
28. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*. (2006) 332:1089–92. doi: 10.1136/bmj.332.7549.1089
29. Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*. (2017) 124:343–51. doi: 10.1016/j.ophtha.2016.11.014
30. Leasher JL, Bourne RR, Flaxman SR, Jonas JB, Keeffe J, Naidoo K, et al. Global estimates on the number of people blind or visually impaired by diabetic retinopathy: a meta-analysis from 1990 to 2010. *Diabetes Care*. (2016) 39:1643–9. doi: 10.2337/dc.15-2171
31. Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekaran DV, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health*. (2020) 2:e240–9. doi: 10.1016/S2589-7500(20)30060-1

Conflict of Interest: GP is employed by Guangzhou Vision Tech Medical Technology Co. Ltd., Guangzhou, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Shi, Tan, Niu, Jiang, Xiong, Peng and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.