ORIGINAL RESEARCH

Towards a Device Agnostic AI for Diabetic Retinopathy Screening: An External Validation Study

Divya Parthasarathy Rao [ID][1], Manavi D Sindal[2], Sabyasachi Sengupta[3], Prabu Baskaran[4], Rengaraj Venkatesh[2], Anand Sivaraman[5], Florian M Savoy[6]

[1]Artificial Intelligence R&D, Remidio Innovative Solutions Inc, Glen Allen, VA, USA; [2]Vitreoretinal Services, Aravind Eye Hospitals and Postgraduate Institute of Ophthalmology, Pondicherry, India; [3]Department of Retina, Future Vision Eye Care and Research Center, Mumbai, India; [4]Vitreoretinal Services, Aravind Eye Hospitals and Postgraduate Institute of Ophthalmology, Chennai, India; [5]Artificial Intelligence R&D, Remidio Innovative Solutions Pvt Ltd, Bangalore, India; [6]Artificial Intelligence R&D, Medios Technologies, Singapore

Correspondence: Divya Parthasarathy Rao, Artificial Intelligence R&D, Remidio Innovative Solutions Inc, 11357 Nuckols Road, #102, Glen Allen, VA, 23059, USA, Tel +1 855 513-3335, Email drdivya@remidio.com

Purpose: To evaluate the performance of a validated Artificial Intelligence (AI) algorithm developed for a smartphone-based camera on images captured using a standard desktop fundus camera to screen for diabetic retinopathy (DR).

Participants: Subjects with established diabetes mellitus.

Methods: Images captured on a desktop fundus camera (Topcon TRC-50DX, Japan) for a previous study with 135 consecutive patients (233 eyes) with established diabetes mellitus, with or without DR were analysed by the AI algorithm. The performance of the AI algorithm to detect any DR, referable DR (RDR Ie, worse than mild non proliferative diabetic retinopathy (NPDR) and/or diabetic macular edema (DME)) and sight-threatening DR (STDR Ie, severe NPDR or worse and/or DME) were assessed based on comparisons against both image-based consensus grades by two fellowship trained vitreo-retina specialists and clinical examination.

Results: The sensitivity was 98.3% (95% CI 96%, 100%) and the specificity 83.7% (95% CI 73%, 94%) for RDR against image grading. The specificity for RDR decreased to 65.2% (95% CI 53.7%, 76.6%) and the sensitivity marginally increased to 100% (95% CI 100%, 100%) when compared against clinical examination. The sensitivity for detection of any DR when compared against image-based consensus grading and clinical exam were both 97.6% (95% CI 95%, 100%). The specificity for any DR detection was 90.9% (95% CI 82.3%, 99.4%) as compared against image grading and 88.9% (95% CI 79.7%, 98.1%) on clinical exam. The sensitivity for STDR was 99.0% (95% CI 96%, 100%) against image grading and 100% (95% CI 100%, 100%) as compared against clinical exam.

Conclusion: The AI algorithm could screen for RDR and any DR with robust performance on images captured on a desktop fundus camera when compared to image grading, despite being previously optimized for a smartphone-based camera.

Keywords: smartphone, Deep Learning, retina, imaging, screening

Introduction

Diabetes Mellitus (DM) is estimated to affect over 640 million people by 2040. The global prevalence of any form of DR among diabetics has increased to 34.6%, and 10.2% for sight-threatening DR (STDR), over the past decade.[1,2]

Artificial Intelligence (AI) methods based on Deep Learning (DL) have been at the forefront of DR screening programs. They particularly help in detecting DR in its early stages. AI-based DR screening algorithms have often been validated against consensus image grading from two or three field, two-dimensional fundus images with promising results.[3–7] However, stereoscopic clinical examination can provide significantly more macular details and inputs from the retinal periphery. This is especially important in diabetic macular edema (DME) and proliferative diabetic retinopathy (PDR) where neovascular changes can be missed at times by conventional fundus imaging techniques capturing posterior pole images. Evidence of AI performance to detect DR changes compared to clinical diagnosis is lacking in literature.[4]

It is established that the performance of the algorithm is closely tied to the fundus camera on which it has been trained and eventually deployed. Hence, the validation process entails ensuring optimum performance on the intended camera for

2659

use by regulatory authorities.[8,9] This, however, limits their utility across devices. There is limited literature on the performance of a DR algorithm on images obtained from different camera systems.

The Medios AI (Medios Technologies, Remidio Innovative Solutions, Singapore) has been extensively validated when integrated on the Remidio smartphone-based fundus camera (Fundus on phone, FOP).[4,10,11] Though developed and trained on various desktop camera-based images, some architectural changes were made while optimizing the Medios AI for the Remidio FOP such that automated DR grading could be delivered offline on the smartphone itself, for eg, at a remote rural site with no internet.[4,10,11] The AI's ability to detect DR on desktop-camera-derived images after these optimizations has not been studied till date.

In this post-hoc analysis, we evaluated the performance of this AI algorithm on images obtained from a desktop fundus camera. This could add a unique capability of performing optimally on both low-cost and high-end cameras. Thus, it could potentially move the AI towards being device independent, expanding the use of the AI across different settings. Additionally, to the best of our knowledge, this is also the first study to compare the performance of an AI algorithm to both clinical examination and consensus image grading by retina specialists.

This AI algorithm gives a binary indication of referral for DR without staging disease. It has been trained to maximize the sensitivity for detecting referable DR (RDR) Ie, worse than mild non proliferative diabetic retinopathy (NPDR), excluding mild NPDR cases during the training process. While the algorithm was first intended for deployment on the Remidio FOP, images from a wide range of cameras were used during the training process. DR algorithms are often validated with datasets captured under similar conditions used for training. This can yield to higher accuracies than expected in real-world settings. The purpose of this study was to validate the performance of this AI as an independent external study on a different imaging system. Beyond performance, this study will also give insights on how the algorithm behaves for mild cases of DR when captured by a standard tabletop fundus camera.

Methods

This retrospective study was approved by the Institutional Ethics Committee at Aravind Eye Hospital and Postgraduate Institute of Ophthalmology, Pondicherry, a tertiary eye care center in south India. The study was performed according to the International Conference on Harmonisation Good Clinical Practice guidelines and fulfilled the tenets of the Declaration of Helsinki.

Study Population and Sample Size Calculation

Posthoc analysis was conducted on a dataset collected for an earlier study validating the smartphone-based camera (FOP, Remidio Innovative Solutions Pvt. Ltd., Bangalore, India) against a standard tabletop fundus camera (TRC-50DX, Topcon Corporation/Kabushiki-gaisha Topukon, Tokyo, Japan).[12]

The study methodology has been described in detail in an earlier publication.[12] In brief, two hundred consecutive diabetic subjects above 21 years of age meeting study criteria were enrolled in the study between April 2015 and January 2016 following a written informed consent. These included diabetic subjects with and without clinically gradable DR. Patients with significant corneal or lenticular pathology precluding fundus examination or those who had undergone prior laser treatment or vitreo-retinal surgeries were excluded from the study.

A sample size of 200 eyes was chosen in the earlier study to include adequate samples of each category of DR, namely no DR, mild to moderate NPDR, severe NPDR, and PDR. This sample estimate was found to be adequate for the present study too. The minimum required sample is 172 eyes to detect a sensitivity of 90% (and addressing a specificity of 80%) with a precision of 10%, incorporating 20% prevalence of referable diabetic retinopathy (RDR) and with a 95% confidence level.

Dilated Image Acquisition Protocol

An ophthalmic photographer used a standard Topcon tabletop fundus camera to capture mydriatic 45 degrees, three fields of view per eye – namely the posterior pole, nasal, and supero-temporal field images. All photographs were stored as JPEG files after removing all patient identifiers and assigning a randomly generated unique numerical identifier linked to the participant's study ID number.

Reference Standard for Comparison of the Performance of the AI

The reference standard for performance assessment of Medios AI consisted of – 1) The consensus image grading of two fellowship trained vitreo-retinal experts (MDS, PB) masked to the clinical grades, as well as each other's grades for all images, and 2) A clinical examination conducted by a single retina specialist (SS) for diagnosing the severity of DR using slit lamp biomicroscopy (+90D lens) and indirect ophthalmoscopy (+20D lens).

Two experts graded the level of DR based on the International Clinical Diabetic Retinopathy (ICDR) severity scale for each eye after examining images from the 3 fields of view.[13] The scale consists of No DR, Mild NPDR, Moderate NPDR, Severe NPDR, PDR and DME. Referable DR (RDR) was defined as moderate NPDR or worse disease and/or the presence of DME. Sight-threatening DR (STDR) was defined as severe NPDR or worse disease and/or the presence of DME. DME was defined as presence of surrogate markers of macular edema such as presence of hard exudates within 1 disc diameter of the center of the fovea. Additionally, all the misclassified false-positive images detected as RDR by the AI were provided to the two expert graders for an adjudicated grading. They also graded the quality of images as "excellent", "acceptable" and "ungradable" as described elsewhere.[12]

The image diagnosis of each doctor was then converted to the following categories as shown in Table 1. The clinical diagnosis of DR was based on the ICDR severity scale as well.

AI-Based Software Architecture

The Medios AI consists of an ensemble of two convolutional neural networks (based on the Inception-V3 architecture). They classify colour fundus images for the presence RDR. The detailed software architecture has been published previously.[4,10] The training set consisted of 52,894 images of which 34,278 images originated from the Eye Picture Archive Communication System tele-medicine program (EyePACS LLC, Santa Cruz, California).[14] This dataset contained images from multiple ethnicities and desktop-based cameras. Additionally, 14,266 mydriatic images were taken with a Kowa VX-10α (Kowa American Corporation, CA, USA) at a Tertiary Diabetes Center, India and 4350 non-mydriatic images were taken in screening camps in India using the Remidio FOP NM10. The dataset was curated to contain as many referral cases as healthy ones.

The AI algorithm was initially intended for deployment on the Remidio FOP. Therefore, the final models were selected based on their performance on an internal test dataset consisting of only Remidio FOP images. The AI has been optimized for the sensitivity of RDR and specificity of any DR to minimize under-detection of referable cases. In other words, it reduces false negatives from a screening perspective. While this leads to a small proportion of mild NPDR being flagged as RDR, it makes the chances of missing an RDR lower.

Automated Image Analysis

Image captured on the Topcon TRC-50DX (Topcon, Japan) were de-identified and uploaded on a secure Virtual Machine to be analyzed by Medios AI software. Each patient received an automated image quality analysis followed by an automated DR analysis. The AI DR analysis output, ie, No RDR, or RDR, as well as the image quality analysis results were noted. The DR results of patients with images deemed ungradable by the AI were included in the analysis if they received a consensus grading by the experts. The quality check AI presents results as ungradable vs gradable. The last

Table 1 Image Diagnosis of Each Doctor and the Corresponding Severity

Severity	Image Diagnosis
Ungradable	DR and/or DME images ungradable
Healthy	No DR and no DME
Sight threatening diabetic retinopathy (STDR)	Severe NPDR and Proliferative DR and/ OR DME
Referable diabetic retinopathy (RDR)	Moderate NPDR and more severe and/ OR DME
Any diabetic retinopathy (any DR)	Any grade of DR and/or DME

Abbreviations: DR, Diabetic Retinopathy; DME, Diabetic Macular Edema, NPDR, Non-Proliferative Diabetic Retinopathy.

step of the AI algorithm consists of thresholding a probability value where 0 is ungradable and 1 is gradable. A threshold of 0.2 is used when deploying the model on the Remidio FOP.

Outcome Measures

The primary outcome measures were the sensitivity, specificity and predictive values (performance metrics) of the AI in detecting RDR when compared to the image grading provided by the specialists.

The secondary measures included assessment of the sensitivity, specificity, predictive values of the AI for any DR, sensitivity in detecting STDR against image grading as well as intergrader reliability for diagnosis. Additionally, the same performance metrics of the AI in detecting any DR, RDR and STDR compared to the diagnosis based on clinical examination were measured.

Statistical Analysis

A 2*2 confusion matrix was used to compute the sensitivity, specificity and Kappa to detect any stage of DR, RDR and STDR by the AI. Additional metrics included the positive predictive value (PPV) and the negative predictive value (NPV). Wilson's 95% confidence Intervals (CI) were calculated for sensitivity, specificity, NPV, and PPV. A weighted kappa statistic was used to determine the interobserver agreement (including the AI as a grader) to the consensus image grading. Kappa of 0–0.20 was considered as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement.[15] All data were stored in Microsoft Excel and were analyzed using pandas (1.1.0), numpy (1.19.5) and scikit-learn (0.23.1) libraries in python 3.7.7.

Results

The study involved analysis of images of 233 eyes from a study cohort of 135 participants aged above 21 years. Subjects had a mean age of 54.1±8.3 years and 65% were men. The average duration of diabetes was 10.7 years (median, 10 years; interquartile range, 8–15 years). As per the clinical examination, 55 eyes (23%) had no DR, 70 eyes (30%) had mild to moderate NPDR, 46 eyes (20%) had severe NPDR, and 62 eyes (27%) had PDR. Forty-four eyes (19%) had DME. The image diagnosis of each doctor was first classified as any DR, RDR, STDR, healthy and ungradable categories. Consensus amongst doctors was then computed. A total of 170 eyes were included in the final analysis. Refer to Figure 1 for illustration.

Comparing the AI Results Against Image Grading

Comparing AI results against image grades (consensus grading followed by adjudicated grading of misclassified false positive images by the AI) by two independent vitreo-retina specialists on images deemed gradable, there was a high sensitivity and specificity for RDR as well as any DR, and the sensitivity for STDR was nearly 100% (Tables 2 and 3).

There were 8 false-positive cases (4.7%) when comparing AI to image grading for RDR of which 4 were mild NPDR and 4 were no DR. There were 2 referable cases missed by the AI, 1 of them was RDR, and 1 was STDR. Kappa agreement between AI and image grading for RDR was 0.85. Figure 2 shows examples of a true positive, a true negative, a false-positive and a false-negative subject. Figure 3 shows the retinal photographs taken from both the cameras highlighting how factors such as field of view and image quality compare between both systems.

Comparing the AI Results Against Clinical Assessment

Comparing AI results against clinical examination, there was 100% sensitivity to detect RDR and STDR with a high sensitivity to detect any DR as well. The specificity for RDR was moderate and any DR was high (Tables 4 and 5). There were 23 (13.5%) false-positive cases when comparing AI to clinical assessment for RDR, 18 being mild DR and 5 cases of no DR. There were no missed cases of RDR or STDR. Kappa agreement between AI and clinical grading for RDR was 0.70.
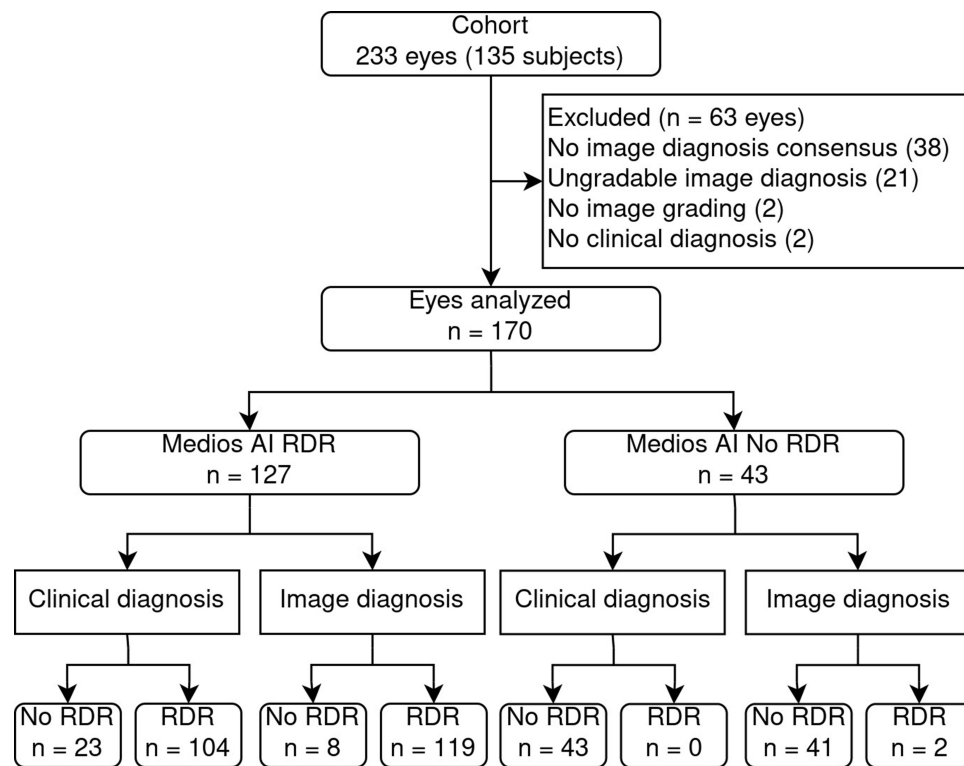
Figure 1 STARD flowchart: AI output for RDR against clinical assessment and image-based grading.

Intergrader Reliability

The intergrader reliability (weighted kappa, ĸ) for detecting RDR was assessed against the consensus (including AI as a grader). Kappa of the AI was 0.81, and that of the clinical ground truth was 0.72 and that of the two graders were 0.89 and 0.86.

Table 2 Confusion Matrix: AI vs Consensus Image Grading

N= 170	**Consensus Image Grading (N, %)**			
AI	No DR	Any DR	RDR	STDR
No RDR	40 (23.5%)	1 (0.58%)	1 (0.58%)	1 (0.58%)
RDR	4 (2.35%)	4 (2.35%)	24 (14.11%)	95 (55.88)

Abbreviations: AI, Artificial Intelligence; DR, Diabetic Retinopathy; RDR, Referable Diabetic Retinopathy; STDR, Sight-threatening Diabetic Retinopathy; N, number of eyes.

Table 3 Performance of AI Against Image Grading

	RDR	**Any DR**	**STDR**
Sensitivity (95% CI)	98.3% (96.1%, 100%)	97.6% (95%, 100%)	99.0% (96.9%, 100%)
Specificity (95% CI)	83.7% (73.3%, 94%)	90.9% (82.4%, 99.4%)	NA
PPV (95% CI)	93.7% (89.5%, 97.9%)	96.9% (93.8%, 99.9%)	NA
NPV (95% CI)	95.3% (89.1%, 100%)	93% (85.4%, 100%)	NA

Abbreviations: AI, Artificial Intelligence; DR, Diabetic Retinopathy; RDR, Referable Diabetic Retinopathy; STDR, Sight-threatening Diabetic Retinopathy; PPV, Positive Predictive Value, NPV, Negative Predictive Value.
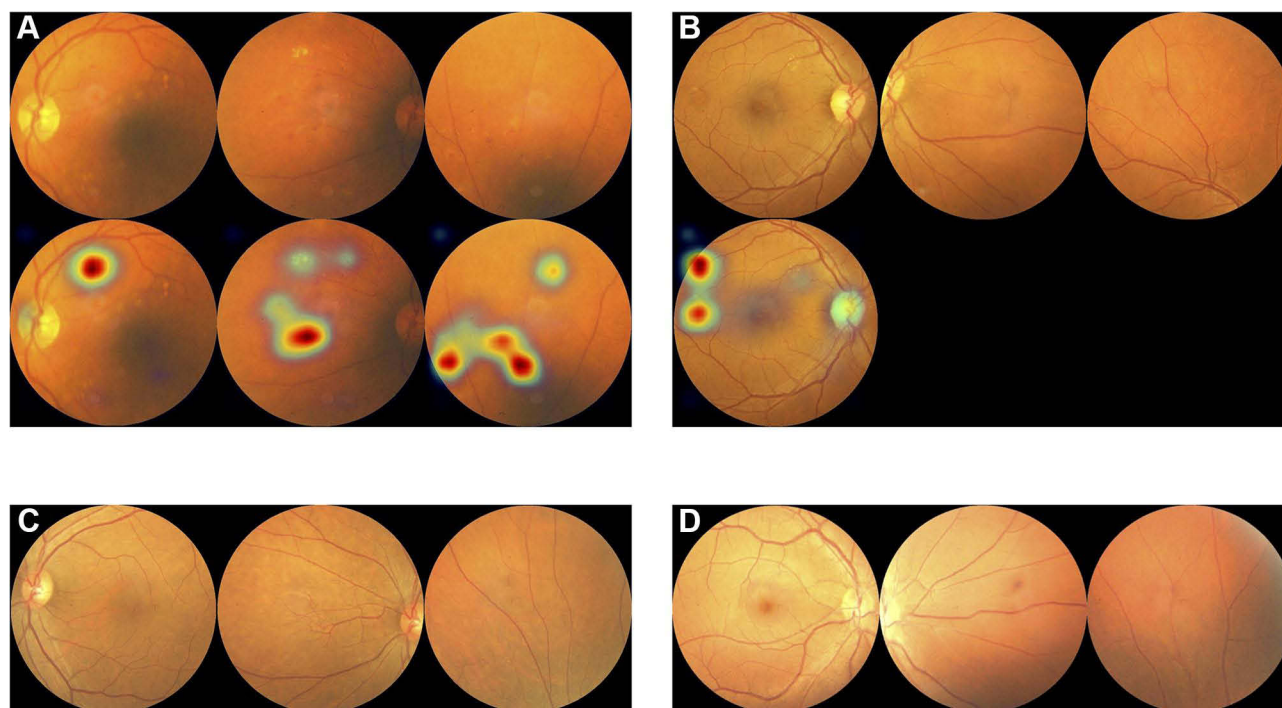
Figure 2 Images of true positive (**A**), false positive (**B**), false negative (**C**) and true negative (**D**) subject with activation maps for image triggering positive diagnosis.
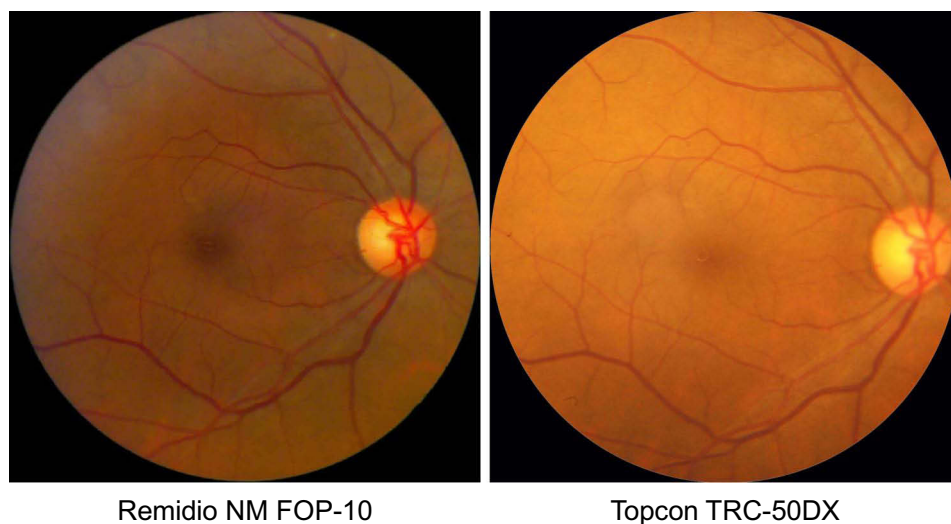

Remidio NM FOP-10   Topcon TRC-50DX

Figure 3 Retinal image photographs from Remidio FOP and Topcon camera.

Accuracy of the Image Quality by the AI

At a threshold of 0.2 image quality (original version of the image quality model deployed on Remidio FOP), the sensitivity for detecting ungradable images was 100% with a sensitivity of 82.8% for gradable images. At a threshold of 0.5, the sensitivity for detecting ungradable images dropped to 96.15% with a sensitivity of 89.0% for gradable images.

Table 4 Confusion Matrix- AI Vs Clinical Exam

N=170	Clinical Grading (N, %)			
AI	No DR	Any DR	RDR	STDR
No RDR	40 (23.5%)	3 (1.76%)	0 (0%)	0 (0%)
RDR	5 (2.94%)	18 (10.58%)	19 (11.17%)	85 (50%)

Abbreviations: AI, Artificial Intelligence; DR, Diabetic Retinopathy; RDR, Referable Diabetic Retinopathy; STDR, Sight-threatening Diabetic Retinopathy; N, number of eyes.

Table 5 Performance of AI Against Clinical Exam

	RDR	Any DR	STDR
Sensitivity (95% CI)	100.0% (100%, 100%)	97.6% (94.9%, 100%)	100% (100%, 100%)
Specificity (95% CI)	65.2% (53.7%, 76.6%)	88.9% (79.7%, 98.1%)	NA
PPV (95% CI)	81.9% (75.2%, 88.6%)	96.1% (92.7%, 99.4%)	NA
NPV (95% CI)	100.0% (100%, 100%)	93.0% (85.4%, 100%)	NA

Abbreviations: AI, Artificial Intelligence; DR, Diabetic Retinopathy; RDR, Referable Diabetic Retinopathy; STDR, Sight-threatening Diabetic Retinopathy; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Discussion

In this study, we found that the AI performance to detect RDR and any stage of DR on images captured with a conventional desktop fundus camera (Topcon) was high when compared to image grading of vitreo-retina specialists.

The Medios AI-DR algorithm was trained on a diverse dataset, despite being architecturally modified to function optimally on the smartphone-based Remidio FOP. During development, images of varying image quality from high-end tabletop systems on top of the original target device were utilized. We hypothesize this to have contributed to the encouraging results obtained in this study, which was not established in any previous study thus far. This is a step towards a device-agnostic algorithm, a much-needed approach in locations where validated fundus cameras are already part of the DR screening programmes. This also establishes that diversity in dataset is key to developing an AI algorithm that is more generalizable across different camera systems. The dataset encompassed a variety of cameras and capturing conditions, while restricting images to a certain field of view (30 to 45 degrees). We hypothesise that the neural network is able to generalize to the differences of colour tint and spatial resolution resulting from using different cameras. This would have, however, not happened if the field of view changed in more drastic ways. This scenario would likely require adaptations to the image pre-processing steps, the neural network architecture and the training dataset. In this study, the field of view (45 degrees) was within the fields of view presented during training, and thus DR lesions have similar relative sizes across different images. Additionally, the smartphone-based fundus camera has been validated against standard desktop systems for image quality.[12,16] This has specifically shown that DR grading by experts is comparable on the systems.

Moderate NPDR is the cut off for detecting RDR as per the International Council of Ophthalmology guidelines for screening of DR and the AAO preferred practice patterns.[2,17] Accordingly, this was also the threshold used for the AI to trigger referral. When the AI results were compared to the clinical grades for RDR, the sensitivity was 100% and specificity was 65.2%, respectively. The specificity was lower than that reported in previous validation studies (86.73–92.5%) on the smartphone-based system.[4,10,11] On further analyzing the low specificity, we found that there were eighteen cases of mild NPDR and five cases of no DR on clinical assessment that were detected by the AI algorithm as RDR. Interestingly, when the consensus image grading of the same mild NPDR patients were cross verified, fifteen were graded as RDR, with two of them graded as STDR.

On analyzing the five no DR cases on clinical exam that were picked as RDR positive by the AI, two had a consensus of any DR, with one of them being RDR too on image grading. We re-examined the class activation maps on these five

subjects and found that three subjects had other lesions – drusens (in two) and Pigment Epithelial Detachment (in one) that triggered the AI to give a positive result.

The analysis of the spuriously low specificity of the RDR algorithm against clinical exam also showed that the kappa for clinical exam (Cohen's kappa 0.72) was lower, compared to the agreement obtained by experts during image grading (Cohen's kappa 0.89 and 0.86). The variation found was higher in milder stages of disease. Literature indicates a wide range of interobserver and grader reliability, ranging from 0.22 to 0.91.[18] We found that the consensus image diagnosis from two experts was more consistent and reliable than a single observer clinical evaluation. This further justifies Krause et al's interpretation where they found that majority decision to have a higher sensitivity than any single grader.[18] Most of the images (15/18 eyes) that were graded as mild NPDR on clinical exam were graded as moderate NPDR or more severe disease on image grading. Thus, specificity went up considerably to 83.7% on image grading with multiple graders. Well known clinical trials like the ACCORD and FIND have also found image grading to be superior to clinical grading to detect early to moderate changes in DR over time.[19] This is also backed by regulatory authorities like FDA who advocate for image grading by multiple certified graders on a consensus or adjudication basis. While clinical assessment provides an opportunity to examine the entire retina, three field imaging with multiple graders provided sufficient information for reliable screening to detect RDR.[10,11]

While we found the results to be comparable to our previous studies using the smartphone-based camera, the modest increase in sensitivity[4] and decrease in specificity[4,10,11] is possibly due to minor variations expected in image sharpness. The decrease in specificity is primarily due to an overcall of mild NPDR cases. A desktop camera like Topcon has better sharpness with mild lesions being more prominent and hence more likely to be picked up by the AI.

The Medios AI system consists of two components: an AI for image quality analysis and an AI for referable DR. This allows the operators to get live automated feedback at the time of image capture. It enables the user to understand whether the image captured is of sufficient quality or needs a recapture. This image quality algorithm has been optimized for use on the Remidio FOP-NM10 device. In this study, we assessed the performance of the AI quality check on the images captured with Topcon camera using the same 0.2 threshold that was used on the original version of the system (deployed on Remidio FOP). The sensitivity of detecting ungradable images was 100% and the sensitivity to detect gradable images was 82.8%. We found that an improved performance can be achieved on the Topcon system by setting the threshold at 0.5. The sensitivity for detecting gradable images improved to 89.0% with a sensitivity drop to 96.15% for detecting ungradable images. Given the minor differences in the sharpness of their imaging system, the threshold of the algorithm will require to be varied prior to deployment on a new camera system.

The strengths of this study are post-hoc analysis on a dataset with good representation of all stages of disease, simultaneous comparison of the AI performance to two reference standards (image grading and clinical assessment) as well as an assessment of the image quality algorithm.

This study has some limitations. First, the AI has been tested with images with similar fields of view. The performance of the AI models when deployed on images with a significantly different field of view needs to be assessed. Second, the images were analyzed using the same AI model as deployed offline on the Remidio FOP, but on a Cloud Virtual Machine. The performance of the system after a future offline integration of the models on a Topcon Fundus camera system will require further study.

Conclusion

To the best of our knowledge, this study is the first of its kind to compare an AI-based screening algorithm for DR to both clinical examination and consensus image-based grading. This study adds to the growing evidence on image-based grading being more consistent and reliable for screening DR than a clinical exam. The AI which had previously been validated only on a smartphone-based fundus camera showed a high sensitivity and specificity in screening for RDR and any stage of DR on images captured on a standard desktop camera. This indicates that this algorithm can be used on both a high-end desktop fundus camera like Topcon and a smartphone-based system to screen for DR given the diversity in training dataset. Thus, it is a positive move towards a device-agnostic application of the AI for expanding the use in screening for DR in different settings. Further studies need to be conducted to assess the efficiency of the system on images from other cameras. This may provide a big boost in reducing the huge economic burden posed by DR globally.

Disclosure

Divya Parthasarathy Rao, Anand Sivaraman and Florian M Savoy are Employees of Remidio Innovative Solutions. Medios Technologies, Singapore, where the AI has been developed, and Remidio Innovative Solutions Inc. USA, are wholly owned subsidiaries of Remidio Innovative Solutions Pvt Ltd, India. Dr Sabyasachi Sengupta reports personal fees from Novartis, India, Bayer, Intas, Allergan, outside the submitted work. The authors report no other conflicts of interest in this work.

References

1. Huemer J, Wagner SK, Sim DA. The evolution of diabetic retinopathy screening programmes: a chronology of retinal photography from 35 mm slides to artificial intelligence. *Clin Ophthalmol*. 2020;14:2021–2035. doi:10.2147/OPTH.S261629
2. International Council of Ophthalmology. Guidelines for diabetic eye care. Available from: https://www.urmc.rochester.edu/MediaLibraries/URMCMedia/eye-institute/images/ICOPH.pdf. Accessed March 28, 2022.
3. Rosses APO, Ben ÂJ, Souzade CF, et al. Diagnostic performance of retinal digital photography for diabetic retinopathy screening in primary care. *Fam Pract*. 2017;34(5):546–551. doi:10.1093/fampra/cmx020
4. Sosale B, Aravind SR, Murthy H, et al. Simple, Mobile-based Artificial Intelligence Algorithm in the detection of Diabetic Retinopathy (SMART) study. *BMJ Open Diabetes Res Care*. 2020;8:e000892. doi:10.1136/bmjdrc-2019-000892
5. Nielsen KB, Lautrup ML, Andersen JKH, Savarimuthu TR, Grauslund J. Deep learning–based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance. *Ophthalmol Retina*. 2019;3(4):294–304. doi:10.1016/j.oret.2018.10.014
6. Wang S, Zhang Y, Lei S, et al. Performance of deep neural network-based artificial intelligence method in diabetic retinopathy screening: a systematic review and meta-analysis of diagnostic test accuracy. *Eur J Endocrinol*. 2020;183(1):41–49. doi:10.1530/EJE-19-0968
7. Tufail A, Rudisill C, Egan C, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*. 2017;124(3):343–351. doi:10.1016/j.ophtha.2016.11.014
8. Abràmoff 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;3:1–8.
9. Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. *JAMA Netw Open*. 2021;4(11):e2134254. doi:10.1001/jamanetworkopen.2021.34254
10. Natarajan S, Jain A, Krishnan R, Rogye A, Sivaprasad S. Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone. *JAMA Ophthalmol*. 2019;137(10):1182–1188. doi:10.1001/jamaophthalmol.2019.2923
11. Sosale B, Sosale A, Murthy H, Sengupta S, Naveenam M. Medios– an offline, smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. *Indian J Ophthalmol*. 2020;68(2):391–395. doi:10.4103/ijo.IJO_1203_19
12. Sengupta S, Sindal MD, Baskaran P, Pan U, Venkatesh R. Sensitivity and specificity of smartphone-based retinal imaging for diabetic retinopathy. *Ophthalmol Retina*. 2019;3(2):146–153. doi:10.1016/j.oret.2018.09.016
13. Hansen MB, Abràmoff MD, Folk JC, Mathenge W, Bastawrous A, Peto T. Results of automated retinal image analysis for detection of diabetic retinopathy from the Nakuru Study, Kenya. *PLoS One*. 2015;10(10):e0139148. doi:10.1371/journal.pone.0139148
14. Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *J Diabetes Sci Technol Online*. 2009;3 (3):509–516. doi:10.1177/193229680900300315
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310
16. Prathiba V, Rajalakshmi R, Arulmalar S, et al. Accuracy of the smartphone-based nonmydriatic retinal camera in the detection of sight-threatening diabetic retinopathy. *Indian J Ophthalmol*. 2020;68(13):S42–6. doi:10.4103/ijo.IJO_1937_19
17. Flaxel CJ, Adelman RA, Bailey ST, et al. Diabetic retinopathy preferred practice pattern®. *Ophthalmology*. 2020;127(1):66–145.
18. Krause J, Gulshan V, Rahimy E, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 2018;125(8):1264–1272. doi:10.1016/j.ophtha.2018.01.034
19. Gangaputra S, Lovato JF, Hubbard L, et al. Comparison of standardized clinical classification with fundus photograph grading for the assessment of diabetic retinopathy and diabetic macular edema severity. *Retina*. 2013;33(7):1393–1399. doi:10.1097/IAE.0b013e318286c952