PEER REVIEWED PUBLICATIONS

Scientific Literature on offline AI

remidio

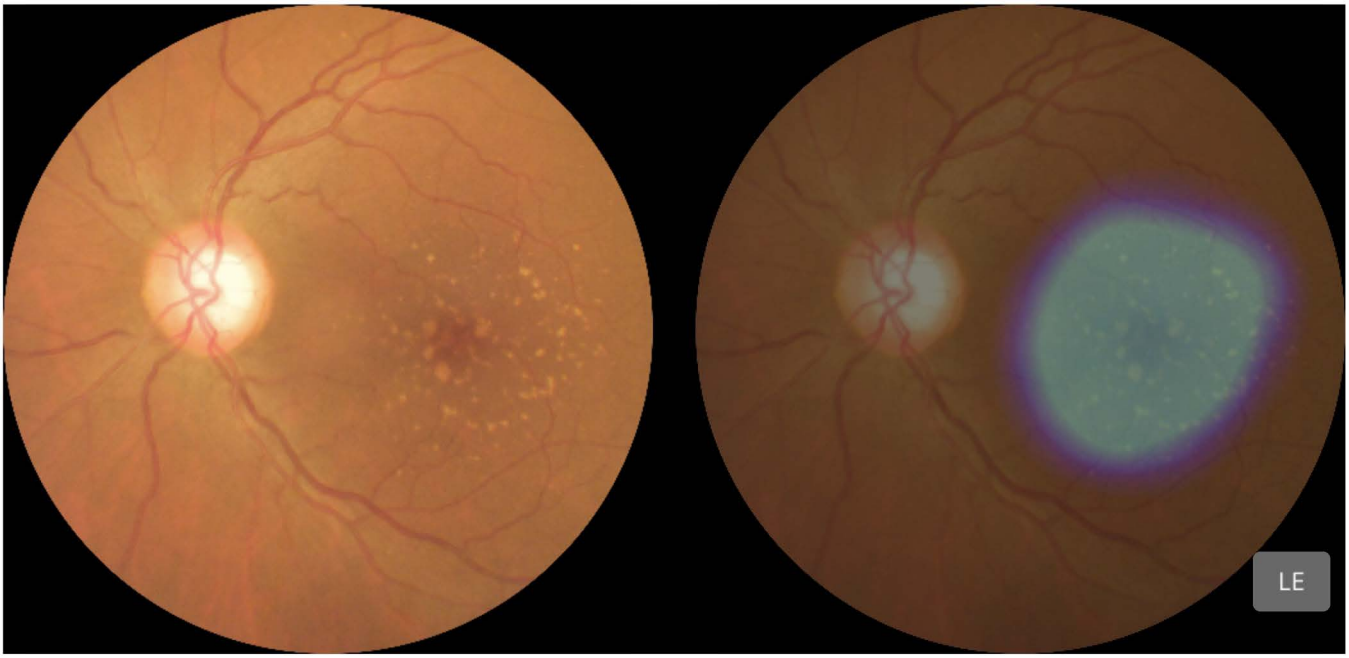
DR and AMD Report

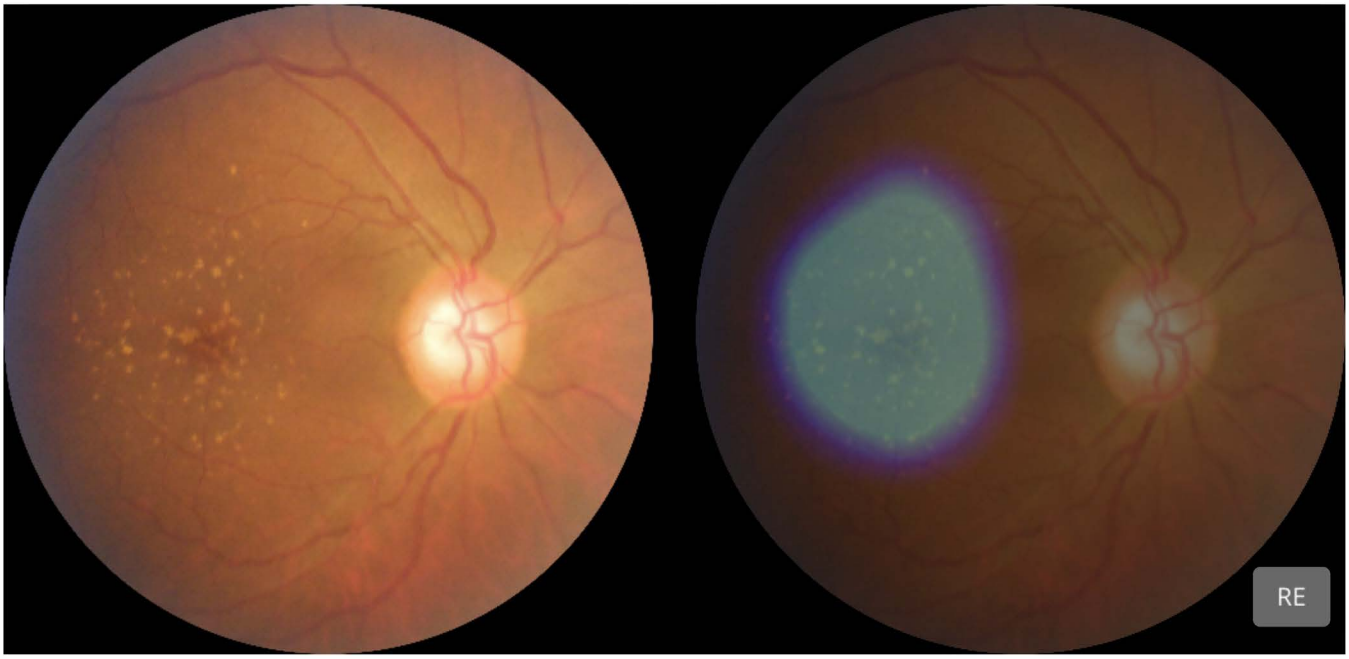
remidio

[redacted]

Result DR: **No signs of DR detected. Re-examine after 12 months.**

Result AMO: **Signs of AMO detected. Examples of lesions are highlighted.**





Medios AI is a physician assist software, not a replacement for an ophthalmologist's diagnosis. The results are only indicative of a high probability of Moderate NPDR, Intermediate AMD, or more severe disease. This report does not screen for any medical or vision conditions apart from DR or AMD. The images on this report are only thumbnails and must not be used for diagnostic purposes. Any heat maps shown are only indicative of some probable areas of abnormality.

Doctor's Signature

rem1d10 medios

Multi-ethnic, multi-site studies evaluating Medios DR AI
High sensitivity and specificity on FOP and multiple cameras

SETTING	STUDY	ETHNICITY	SAMPLE	STUDY DESIGN	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)
COMMUNITY	NATARAJAN ET AL., JAMA OPHTHALMOL, 2019 A	INDIAN	231	PROSPECTIVE	100.0% (78.2%-100.0%)	88.4% (83.2%-92.5%
	WROBLEWSKI ET AL. JDST, 2023 B	MEXICAN	248	RETROSPECTIVE	ANY DR: 94% (88%-97%)	ANY DR: 94% (88%-97%)
	POLYCLINICS, RAO ET AL, IJO, 2024C	ARMENIAN	478	RETROSPECTIVE	95.30% (91.9% 98.7%)	95.30% (91.9% 98.7%)
	MOBILE PRIMARY HEALTH CARE, DR SCREENING CLINIC, KEMP ET AL, BMJ OPEN OPHTHAL, 2023D	DOMINICAN	535	PROSPECTIVE	80.4% (75.5% - 86.3%)	91.5% (87.9% - 94.3%
	NHS COLLABORATION STUDY*	MULTI-ETHNIC	200K	RETROSPECTIVE (MULTIPLE CAMERAS)	98.9% (98.5 – 99.2%)	80.9% (80.7%-81%)
TERTIARY EYE HOSPITAL	RAO ET AL. CLINICAL OPHTHAL, 2022 E	INDIAN	135	RETROSPECTIVE (TOPCON CAMERA)	98.3% (96% -100%)	83.7% (73% - 94%)
	GRYZYBOWSKI ET AL. OPHT RES. 2023 F	POLISH	807	RETROSPECTIVE (TOPCON CAMERA)	95% (91 –98%)	80% (77 -83%)
DIABETES CLINICS	SOSALE ET AL (BMJ OPEN DIAB RES CARE, 2020 F	INDIAN	900	PROSPECTIVE	93% (91.3%-94.7%)	92.5% (90.8%-94.2%)

*TO BE PUBLISHED

A NATARAJAN ET AL. JAMA OPHTHALMOLOGY, 2019, PMIO: 31393538.
B WROBLEWSKI ET AL.JOURNAL OF DIABETES SCIENCE AND TECHNOLOGY, 2023, PMID: 37641576.
C RAO ET AL. INDIAN JOURNAL OF OPHTHALMOLOGY, 2014 (ACCEPTED FOR PUBLICATION).
D KEMP ET AL. BMJ OPEN OPHTHALMOLOGY, 2023, PMID: 38135351.
E RAO ET AL. CLINICAL OPHTHALMOLOGY, 2022, PMID: 36003071.
F SOSALE ET AL. BMJ OPEN DIABETES RESEARCH CARE. 2020, PMID: 32049632. G SOSALE ET AL. INDIAN JOURNAL OF OPHTHALMOLOGY. 2020. PMID: 31957735.

JAMA Ophthalmology | Original Investigation

Diagnostic Accuracy of Community-Based Diabetic Retinopathy Screening With an Offline Artificial Intelligence System on a Smartphone

Sundaram Natarajan, MD; Astha Jain, MD; Radhika Krishnan, MD; Ashwini Rogye, B Optom; Sobha Sivaprasad, FRCOphth

IMPORTANCE Offline automated analysis of retinal images on a smartphone may be a cost-effective and scalable method of screening for diabetic retinopathy; however, to our knowledge, assessment of such an artificial intelligence (AI) system is lacking.

OBJECTIVE To evaluate the performance of Medios AI (Remidio), a proprietary, offline, smartphone-based, automated system of analysis of retinal images, to detect referable diabetic retinopathy (RDR) in images taken by a minimally trained health care worker with Remidio Non-Mydriatic Fundus on Phone, a smartphone-based, nonmydriatic retinal camera. Referable diabetic retinopathy is defined as any retinopathy more severe than mild diabetic retinopathy, with or without diabetic macular edema.

DESIGN, SETTING, AND PARTICIPANTS This prospective, cross-sectional, population-based study took place from August 2018 to September 2018. Patients with diabetes mellitus who visited various dispensaries administered by the Municipal Corporation of Greater Mumbai in Mumbai, India, on a particular day were included.

INTERVENTIONS Three fields of the fundus (the posterior pole, nasal, and temporal fields) were photographed. The images were analyzed by an ophthalmologist and the AI system.

MAIN OUTCOMES AND MEASURES To evaluate the sensitivity and specificity of the offline automated analysis system in detecting referable diabetic retinopathy on images taken on the smartphone-based, nonmydriatic retinal imaging system by a health worker.

RESULTS Of 255 patients seen in the dispensaries, 231 patients (90.6%) consented to diabetic retinopathy screening. The major reasons for not participating were unwillingness to wait for screening and the blurring of vision that would occur after dilation. Images from 18 patients were deemed ungradable by the ophthalmologist and hence were excluded. In the remaining participants (110 female patients [51.6%] and 103 male patients [48.4%]; mean [SD] age, 53.1 [10.3] years), the sensitivity and specificity of the offline AI system in diagnosing referable diabetic retinopathy were 100.0% (95% CI, 78.2%-100.0%) and 88.4% (95% CI, 83.2%-92.5%), respectively, and in diagnosing any diabetic retinopathy were 85.2% (95% CI, 66.3%-95.8%) and 92.0% (95% CI, 97.1%-95.4%), respectively, compared with ophthalmologist grading using the same images.

CONCLUSIONS AND RELEVANCE These pilot study results show promise in the use of an offline AI system in community screening for referable diabetic retinopathy with a smartphone-based fundus camera. The use of AI would enable screening for referable diabetic retinopathy in remote areas where services of an ophthalmologist are unavailable. This study was done on patients with diabetes who were visiting a dispensary that provides curative services to the population at the primary level. A study with a larger sample size may be needed to extend the results to general population screening, however.

JAMA Ophthalmol. doi:10.1001/jamaophthalmol.2019.2923
Published online August 8, 2019.

Invited Commentary

Author Affiliations: Aditya Jyot Foundation for Twinkling Little Eyes, Mumbai, India (Natarajan, Jain, Krishnan, Rogye); Moorfields Eye Hospital, London, United Kingdom (Sivaprasad).

Corresponding Author: Astha Jain, MD, Aditya Jyot Foundation for Twinkling Little Eyes, Mumbai, 153, Major Parmeswaran Road, Wadala, Mumbai 400031, India (drradhika.ajftle@gmail.com).

Diabetic retinopathy (DR) is a major cause of preventable blindness in the working-age population in many countries of the world.[1] People with diabetes usually remain asymptomatic until an advanced stage of DR. Therefore, screening for sight-threatening complications is necessary to initiate timely treatment.[1,2] Prevalence of blindness attributable to DR has decreased in countries such as the United Kingdom because of effective population-based screening programs that use desktop retinal cameras to capture 1-field, 2-field, or 3-field mydriatic digital retinal photographs,[2,3] followed by primary or secondary human grading and an arbitration process.[1-3] These are costly, time-consuming, and complex screening programs that require considerable training in the use of cameras, as well as experienced retinal graders, limiting the relevance of these models in developing countries. There is a substantial unmet need for accurate and simple-to-use screening modalities that can be used globally to screen people for DR. Smartphone-based retinal imaging is emerging as a cost-effective way of screening for retinopathy in the community.[4,5] Similarly, automated analysis of retinal images captured using standard retinal cameras has the promise of being cost-effective and scalable within population-based DR screening programs.[6,7] Incorporating similar automated analysis into low-cost, smartphone-based devices has been shown to be acceptable for screening.[8]

To date, most automated algorithms use deep-learning and neural networks that require a processor-intensive environment for inferencing, resulting in images needing to be transferred to the cloud. However, there are many parts of the world where access to a stable internet connection is not assured. This study validates the performance of an offline automated analysis algorithm that runs directly off a smartphone. To our knowledge, this is the first study evaluating an offline artificial intelligence (AI) algorithm to detect DR using an affordable, easy-to-use, smartphone-based imaging device.

Key Points

Question To evaluate the performance of an offline, automated artificial intelligence system of analysis to detect referable diabetic retinopathy on images taken by a health worker on a smartphone-based, nonmydriatic retinal camera.

Finding In this cross-sectional study, fundus images from 213 study participants were subjected to offline, automated analysis. The sensitivity and specificity of the analysis to diagnose referable diabetic retinopathy were 100.0% and 88.4%, respectively, and the sensitivity and specificity for any diabetic retinopathy were 85.2% and 92.0%, respectively.

Meaning This study suggests these methods might be used to screen for referable diabetic retinopathy using offline artificial intelligence and a smartphone-based, nonmydriatic retinal imaging system.

Methodology

Fundus images were captured using the Remidio Non-Mydriatic Fundus on Phone (Remidio Innovative Solutions Pvt Ltd). The images so captured were subjected to automated analysis by the Medios AI (Remidio), a proprietary offline automated analysis of retinal images on a smartphone to detect referable diabetic retinopathy (RDR) on images taken by a health care worker on a smartphone-based, nonmydriatic retinal camera. These were also graded by a vitreoretinal resident physician and a vitreoretinal surgeon (A.J.) who were masked to the results from the AI system.

Institutional review board approval was obtained from the Aditya Jyot Eye Hospital Ethics Committee. Informed consent was obtained from all participants. The protocol adhered to the tenets of the Declaration of Helsinki. Both the offline automated analysis and the smartphone-based, nonmydriatic retinal imaging system are based on proprietary technologies. However, authors of the study have no financial interest in these technologies.

Capture of Retinal Images

This was a prospective, cross-sectional study of diagnostic accuracy. Patients with diabetes mellitus who were visiting the various dispensaries administered by the Municipal Corporation of Greater Mumbai in Mumbai, India, on a particular day were screened for DR using the portable, smartphone-based, nonmydriatic retinal imaging system. Preliminary data, such as age, sex, duration since diabetes onset, and postprandial blood glucose level, were collected. Patients' eyes were dilated using single drop of tropicamide eyedrops, 1%, which has previously been found to cause minimal risk of angle-closure glaucoma.[9] Fundus imaging was then done by a health care worker with no professional experience in the use of fundus cameras. An anterior segment photograph was first captured, followed by 3 fields of the fundus (namely, the posterior pole, including the disc and macula, and the nasal and temporal fields), as per the Early Treatment Diabetic Retinopathy Study protocol (**Figure 1**). The offline AI algorithm on the smartphone flags images of poor quality, prompting the operator to take additional pictures of the same retinal view until the images were deemed acceptable by the AI system.

Grading by Human Graders

The images were stored on a Health Insurance Portability and Accountability Act-compliant cloud server (Amazon Web Services) and graded by a vitreoretinal resident and a vitreoretinal surgeon (A.J.) at the Aditya Jyot Eye Hospital and Aditya Jyot Foundation for Twinkling Little Eyes in Mumbai, India, who were masked to the AI grading results. In case of a discrepancy between the grading of the resident and surgeon, the diagnosis of the surgeon was considered final. The grading of retinopathy was done according to the International Clinical DR severity scale.[10] The final diagnosis for each patient was determined by the stage of DR of the more affected eye per the International Clinical DR severity scale. Patients whose image of 1 or both eyes was considered ungradable were excluded from the AI analysis.

Grading by Offline AI System

The offline automated analysis application is integrated into the smartphone-based, nonmydriatic retinal imaging sys-

tem. It is a component of the camera control app and thus seamlessly integrates into the image acquisition workflow. It can be broadly divided into 2 core components. First, an algorithm checks the quality of the captured images. A second DR assessment mechanism generates a diagnosis by detecting DR lesions. This mechanism relies on 2 convolutional neural networks.

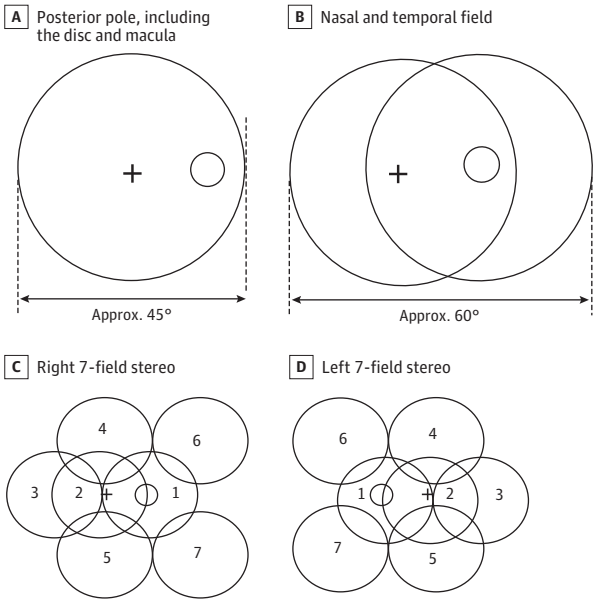
Captured images are first processed by a cropping algorithm. This removes the black border surrounding the circular field of view generated by the fundus camera. They are then down-sampled to a standardized image size. A first neural network assesses the quality of the image. This network is based on the MobileNet architecture. It has been trained with fundus images flagged as ungradable, as well as images of sufficient quality. The user is advised to recapture the image if the result of the neural network is negative.

Two neural networks have been trained separately to detect DR. They consist of binary classifiers based on the Inception-V3 (Google) architecture that separate healthy images from images with referable DR (defined as moderate nonproliferative DR and cases of greater severity). No images with mild nonproliferative DR have been used during training. The training set consisted of 34 278 images from the Eye Picture Archive Communication System (EyePACS) data set, 14 266 images taken with a Kowa VX-10α mydriatic camera at Diacon Hospital in Bangalore, India, and 4350 nonmydriatic images taken in screening camps by the smartphone-based, nonmydriatic retinal imaging system. The data set has been curated to contain as many referral cases as healthy ones. It has also been curated to contain images taken in a variety of conditions, including with nonmydriatic and low-cost cameras.

Three different data sets were used for internally validating networks and ensembling them. These are separate from the training data. Results on these data sets are shown in **Table 1.** Data set 1 consists of images taken with the mydriatic version of the smartphone-based, nonmydriatic retinal imaging system at Dr Mohan's Diabetes Specialties Center in Chennai, India. Data set 2 consists of images taken with the mydriatic mode (one of several modes available on this more recent device) of the smartphone-based, nonmydriatic retinal imaging system at Diacon Hospital in Bangalore, India. These institutions only provided images with their diagnosis and were not involved in computing the results.

One of the 2 networks has been trained directly on the captured images, while the other works on images that underwent an image-processing algorithm to boost their contrast. The contrast-enhancement algorithm has been empirically optimized to make DR lesions stand out in the input images. Both outputs of each network are then fed to a linear classifier that computes the final assessment of an image. This follows the ensemble learning paradigm. It improves the accuracy by combining several classifiers trained under different strategies. Common data-augmentation techniques, such as rotations, flipping, and zooming, were applied to both networks. Final referral recommendations are given on a patient level. A patient was considered to have referable disease if any image was flagged as referable by the algorithm.

Figure 1. Seven-Segment Early Treatment Diabetic Retinopathy Study Protocol



Class activation mapping[11] was also implemented. This gives a visual feedback to the physician by displaying the areas of the fundus image that have triggered a positive diagnosis. Examples of outputs are given in **Figure 2.**

The whole system has been implemented directly on the iPhone 6 (Apple) using high-performance image processing techniques based on CoreML version 2.0 (Apple) and Open Graphics Library ES 2.0 (Silicon Graphics), leveraging on the graphics processing unit of the device instead of an internet connection to a remote server. Both the image processing algorithms and the neural networks run in seconds on an iPhone 6 (Apple).

The same images of these patients were graded by the offline automated analysis algorithm to have either referable DR or no DR. The AI algorithm was run offline on the smartphone by the operator immediately after image acquisition. Adjudication of images that presented results that differed between the resident, surgeon, and AI system was handled by the vitreoretinal surgeon (A.J.) at Aditya Jyot Foundation for Twinkling Little Eyes.

The offline automated analysis is designed to binary-type only RDR and no DR. It does not grade the stages of DR, such mild nonproliferative DR, moderate nonproliferative DR, severe nonproliferative DR, and proliferative DR.

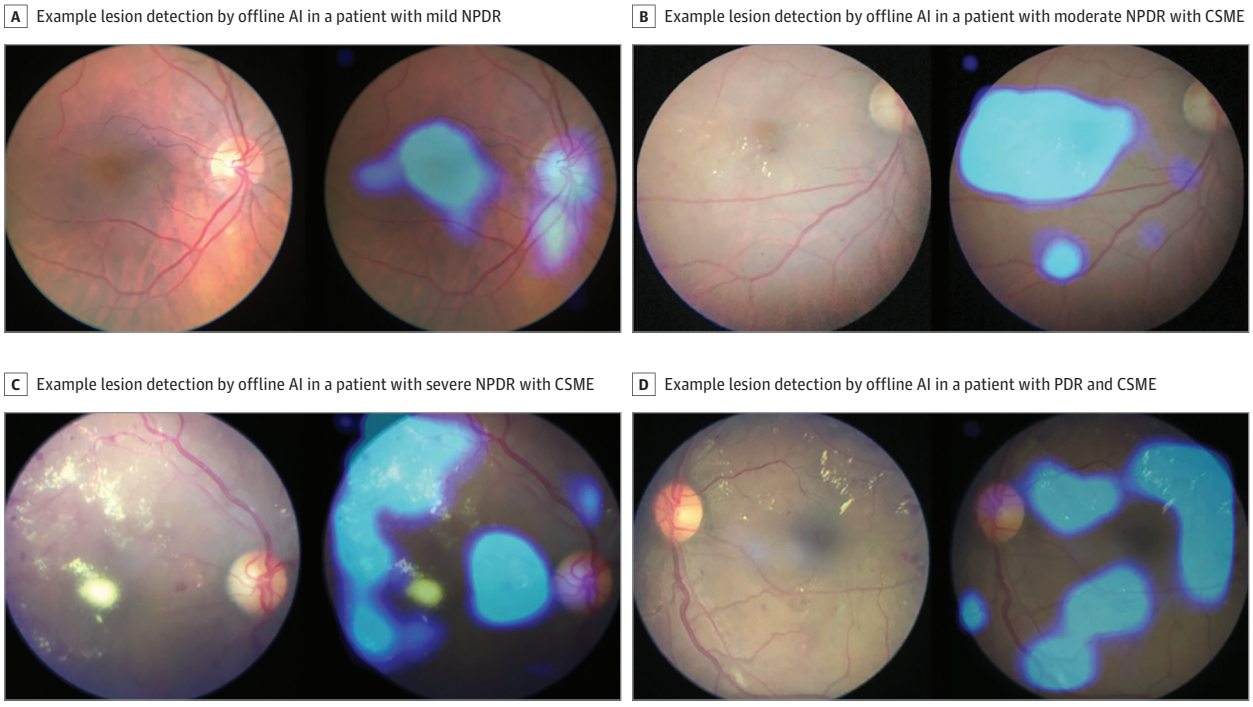
Statistical Analysis

Sensitivity and specificity statistics were computed for both any DR as well as RDR, assuming the ground truth to be the evaluation of the same patient images by the vitreoretinal resident and surgeon. The eye with the more severe retinopathy grading was considered the patient-level grading. The minimum number of patients needed to be screened in an opportunistic screening context, such as an outreach center in India, was calculated, assuming a margin of error of 7% for this

Table 1. Medios Artificial Intelligence Internal Validation: Performance Results on Data Sets

			Referable Diabetic Retinopathy		Any Diabetic Retinopathy	
Data Set	Images, No.	Patients, No.	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %
1	3038	301	95.9	81.3	86.2	99.1
2	1054	165	100.0	78.7	77.1	91.3

Figure 2. Class Activation Mapping in Mild, Moderate, and Severe Nonproliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR)



AI indicates artificial intelligence; CSME, clinically significant macular edema.

study on either side of the mean (given a previously published mandate from the US Food and Drug Administration of an end point of at least 86% diagnostic sensitivity of RDR).[12] At the 95% CI, and a population of 18.4 million in Mumbai (census of 2011), with a maximum DR prevalence of 20%,[13] this resulted in a minimum sample size of 200 patients.

Results

Of the 255 patients seen at the dispensary on the day the study was conducted, 231 consented for DR screening. The major reasons given by the 24 nonparticipating patients for refusal were the unwillingness to wait for screening and the blurring of vision that would occur after dilation. Images of one or both eyes from 18 patients were deemed ungradable by the ophthalmologist, and these individuals were therefore excluded. Hence, a total of 213 patients were analyzed for assessment of RDR using AI, with 1 or more ophthalmologists grading the same images regarded as the ground truth (defined as a direct observation serving as the gold standard). Among the 213 included patients, there were 110 female patients and 103 male patients. The mean (SD) age of the participants was 53.1 (10.3) years. The mean (SD) postprandial blood glucose level was 207.8 (74.7) mg/dL (to convert to millimoles per liter, multiply by 0.0555), and the mean (SD) duration since diabetic disease onset was 5.5 (4.75) years.

A total of 187 patients were diagnosed as having no DR by ophthalmologist grading. Of these, 172 patients were correctly diagnosed by AI, whereas 15 patients (8.0%) were incorrectly diagnosed as having RDR. Fifteen patients (8.0%) were identified as having RDR by ophthalmologist grading, and all 15 (100.0%) were correctly diagnosed by the AI. Among 12 individuals with cases of mild nonproliferative DR who were diagnosed by the ophthalmologists, 8 patients (67%) were diagnosed as having RDR by the AI, while 4 (33%) were diagnosed as not having DR. This gave a sensitivity and specificity of diagnosing RDR as 100% (95% CI, 78.2%-100%) and 88.4% (95% CI, 83.16%-92.53%), respectively; the same values for any DR were 85.2% (95% CI, 66.3%-95.8%) and 92.0% (95% CI, 97.1%-95.4%), respectively (**Table 2**). There was excellent intergrader agreement between the vitreoretinal resident and the vitreoretinal surgeon (eyewise grading: minimum $\kappa = 0.89$ [SE, 0.05]; clinically significant macular edema grading: minimum $\kappa = 0.77$ [SE, 0.06]).

Table 2. Medios Offline Artificial Intelligence Diagnoses vs Ophthalmologist Diagnoses

	Medios Artificial Intelligence					
	By Patient, After Excluding Poor-Quality Images		By Patient, With Poor-Quality Images Included		By Eye	
Diabetic Retinopathy Diagnosis	Referable Diabetic Retinopathy	No Diabetic Retinopathy	Referable Diabetic Retinopathy	No Diabetic Retinopathy	Referable Diabetic Retinopathy	No Diabetic Retinopathy
Ophthalmologist grading[a]						
None	15	172	28	159	35	317
Nonproliferative						
Mild	8	4	8	4	11	8
Moderate	12	0	12	0	17	0
Severe	2	0	2	0	4	0
Proliferative	1	0	1	0	2	0
Diagnostic accuracy, %						
Referable						
Sensitivity	100.0	NA	100.0	NA	100.0	NA
Specificity	88.4		81.9		87.6	
Any						
Sensitivity	85.2		85.2		81.0	
Specificity	92.0		85.0		90.1	

Abbreviation: NA, not applicable.

[a] Ground truth.

Since the image capture was done by a nonspecialist, there were images that did not meet the minimum image quality requirement of the AI. A separate analysis was performed wherein the health care worker was asked to use all images taken in every patient, including the images not meeting the minimum quality standards of the AI, in assessing the diagnostic accuracy of the AI (**Figure 3**). The sensitivity for detection of RDR continued to remain 100.0% (95% CI, 78.2%-100.0%), while the specificity dropped to 81.9% (95% CI, 75.9%-87.0%) as a result of an increase in the number of no DR cases graded as RDR by the AI, consequent to the inclusion of the poorer-quality images (Table 2). In fact, eyewise analysis of the same data set with the inclusion of poor-quality images showed sensitivity of detection of RDR of 100.0% (95% CI, 85.2%-100.0%), specificity of detection of RDR of 87.6% (95% CI, 83.8% to 90.8%), sensitivity of detection of any DR of 81.0% (95% CI, 65.9% to 91.4%), and specificity of detection of any DR of 90.1% (95% CI, 86.4% to 93.0%), as shown in Table 2.

Discussion

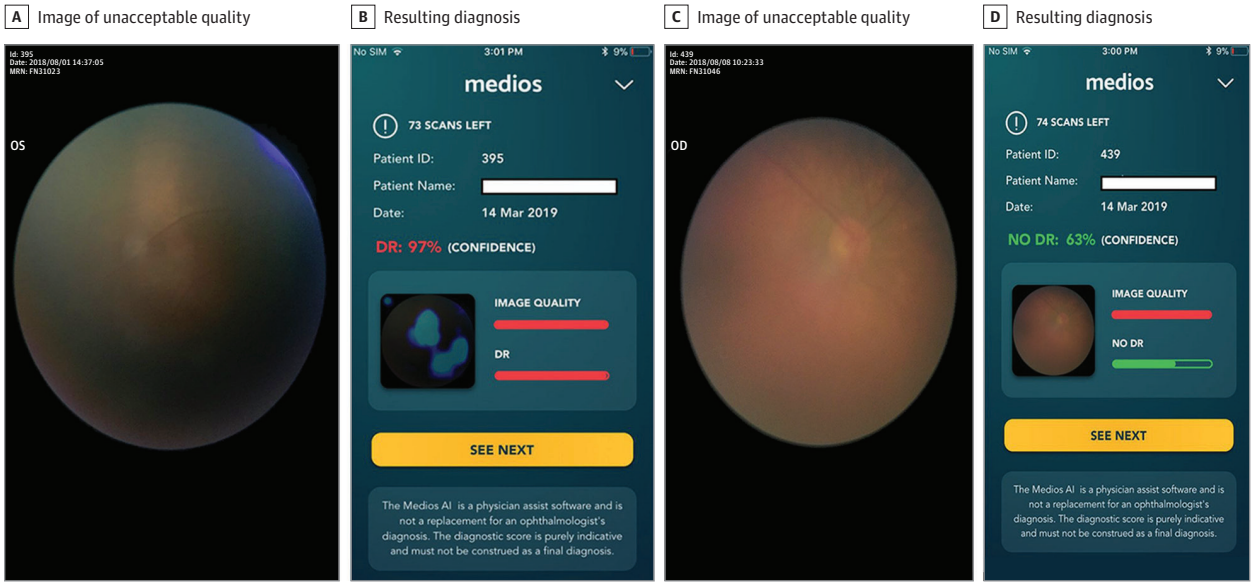
This study evaluates the diagnostic accuracy of an offline AI algorithm for detection of RDR on images taken from a smartphone-based portable camera. To our knowledge, this is the first study assessing an offline AI algorithm on a smartphone for detection of RDR.

In a developing country such as India, nearly 70% of the population resides in rural areas, and a ratio of only 1 ophthalmologist per 100 000 people[14] is available for the care of the entire population. This study shows how an offline AI algorithm can help address this lack of specialist access through automatic, instant grading of the retinal images, highlighting a possible solution for implementation of large-scale models for screening for RDR. Various software have been used previously for automated detection of DR.[6,7,12] The current study uses the offline automated analysis that gives results in real time.

The sensitivity for detection of RDR remained at 100.0% in both the eyewise analysis as well as the patientwise analysis, pointing to an inherent robustness of the offline algorithm in screening for RDR. Similar high sensitivity was found in the study by Gulshan et al[6] albeit in an in-clinic retrospective study, while in the EyePACS-1 data set, the sensitivity was 97.5% and the specificity was 93.4%. In the Messidor data set, the sensitivity was 96.1% and the specificity was 93.9% for detecting RDR. A study using the cloud-based software EyeArt (Eyenuk) showed a sensitivity and specificity of 95.8% and 80.2%, respectively, for detecting any DR and 99.1% and 80.4%, respectively, in detecting sight-threatening DR,[8] using the same smartphone-based retinal imaging system technology, albeit via an earlier model of the device offering mydriatic imaging alone. In a multiethnic study, Ting et al[7] showed a sensitivity and specificity of 90.5% and 91.6%, respectively, for RDR and 100% and 91.1%, respectively, for vision-threatening DR using a conventional, desktop fundus camera.

The specificity seen in this study for detecting RDR may have been slightly lower because many mild nonproliferative DR cases have been identified as RDR by the offline automated analysis. This is because the offline AI was purposely not trained on mild nonproliferative DR images, to ensure high specificity in no DR and RDR diagnoses. Retinal lesions other than DR, such as retinitis pigmentosa, drusen, and retinal pig-

Figure 3. Artificial Intelligence Analysis of Poor-Quality Images



ment epithelium changes, were overdiagnosed by the AI as RDR. Although incorrectly labeled as RDR, these cases likely would warrant a referral to an ophthalmologist.

The major advantage of the offline, automated analysis over the previously used deep-learning algorithms[6,7,12,15,16] may be that it can be used offline on a smartphone and would not require internet access for real-time transfer of images, which enables results to be given to patients immediately. The AI also provides a lesion detection map on the images (Figure 2), which guides the health worker and educates the patient.

Unlike previous retrospective studies that have assessed the performance of the AI algorithms in an in-clinic setting with images that are usually of excellent quality, this study involved the use of the smartphone-based, nonmydriatic retinal imaging system in the field by a health care worker who was trained for less than 2 weeks on how to use the device. Thus, not all the images were of excellent quality, which is representative of what would typically be expected in large-scale, opportunistic community screenings. Even when the offline AI system was subjected to images with quality deemed unacceptable by the AI (Figure 3), the sensitivity for RDR and any DR of the offline automated analysis remained unchanged, while the specificity for RDR and any DR decreased by 7% as a result of some cases of no DR being incorrectly graded as RDR. The superiority end point deemed by the FDA in the pivotal clinical trials evaluating the IDx AI algorithm was a sensitivity of 85% and a specificity of 82.5%.[12] The offline automated analysis algorithm provides a sensitivity and specificity to detect RDR of 100.0% and 88.4%, and this was above the defined thresholds.

Limitations

One limitation of this study is the small sample size of the test population on which the offline automated analysis was tested compared with other studies in literature.[6,7,12] Nevertheless, to estimate the specificity and sensitivity with a lower margin of error of less than 3% (at the 95% CI), a larger sample size of nearly 1050 patients will be needed, assuming a prevalence of up to 20% of DR in the Mumbai population.[13] Another drawback is that the current version of the offline AI does not permit grading of retinopathy according to the International Clinical DR severity scale or the National Health Service classification. Hence, this study is unable to assess the sensitivity and specificity of the offline AI for detection of individual grades of DR.

In this study, the analysis included imaging pupils as small as 3 mm. However, nonmydriatic imaging protocols in an Indian population have shown a large percentage of ungradable images, owing to the comorbidity of cataract and smaller mesopic pupil sizes.[17] This necessitated dilated retinal photography in screening for RDR, especially when implementing 3-segment digital retinal imaging protocols.[18]

Conclusions

The Municipal Corporation of Greater Mumbai, India, where the study was conducted, is the largest municipal corporation in the country, with 174 dispensaries and 210 health posts across the city. Of these, all 174 dispensaries run diabetes management services through monitoring and basic treatment. Of the 255 patients eligible for inclusion, 231 patients (90.6%) visiting the centers received examinations using a dilated imaging protocol, pointing to convenience, affordability, and access to instant reporting driving the demand for a proper DR screening with dilation. A validation of these findings on a larger data set that enables precise assessment of specificity and sensitivity with a standard error less than 3% is currently in progress.

ARTICLE INFORMATION

Accepted for Publication: May 26, 2019.

Published Online: August 8, 2019. doi:10.1001/jamaophthalmol.2019.2923

Author Contributions: Drs Natarajan and Jain had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Natarajan, Jain, Krishnan, Sivaprasad.
Acquisition, analysis, or interpretation of data: Rogye.
Drafting of the manuscript: Jain, Rogye.
Critical revision of the manuscript for important intellectual content: Natarajan, Krishnan, Sivaprasad.
Administrative, technical, or material support: Krishnan.
Supervision: Natarajan, Jain, Sivaprasad.

Conflict of Interest Disclosures: Dr Sivaprasad reported grants and personal fees from Novartis, Allergan, Boehringer Ingelheim, Roche, and Optos and grants, personal fees, and nonfinancial support from Bayer outside the submitted work. No other disclosures were reported.

Additional Contributions: We acknowledge Medios Technologies, Singapore, for providing the AI software for conducting the study and Florian M. Savoy, MSc, Medios Technologies, for authoring the paragraphs describing the technical design of the AI software in the methodology section. We acknowledge Anand Sivaraman, PhD, Remidio Innovative Solutions Pvt Ltd, for technical discussions on study design, training on the NM FOP 10 device, and inputs on statistical analysis. We acknowledge Diacon Hospital, Bangalore, India, and Dr Mohan's Diabetes Specialities Center, Chennai, India, for providing images for internal validation. The individuals named were not compensated.

REFERENCES

1. Squirrell DM, Talbot JF. Screening for diabetic retinopathy. *J R Soc Med*. 2003;96(6):273-276. doi:10.1177/014107680309600604

2. Bachmann MO, Nelson SJ. Impact of diabetic retinopathy screening on a British district population: case detection and blindness prevention in an evidence-based model. *J Epidemiol Community Health*. 1998;52(1):45-52. doi:10.1136/jech.52.1.45

3. Scanlon PH. The English national screening programme for diabetic retinopathy 2003-2016. *Acta Diabetol*. 2017;54(6):515-525. doi:10.1007/s00592-017-0974-1

4. Rajalakshmi R, Arulmalar S, Usha M, et al. Validation of smartphone based retinal photography for diabetic retinopathy screening. *PLoS One*. 2015;10(9):e0138285. doi:10.1371/journal.pone.0138285

5. Sengupta S, Sindal MD, Baskaran P, Pan U, Venkatesh R. Sensitivity and specificity of smartphone based retinal imaging for diabetic retinopathy: a comparative study. *Ophthalmol Retina*. 2019;3(2):146-153. doi:10.1016/j.oret.2018.09.016

6. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216

7. 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(22):2211-2223. doi:10.1001/jama.2017.18152

8. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye (Lond)*. 2018;32(6):1138-1144. doi:10.1038/s41433-018-0064-9

9. Pandit RJ, Taylor R. Mydriasis and glaucoma: exploding the myth, a systematic review. *Diabet Med*. 2000;17(10):693-699. doi:10.1046/j.1464-5491.2000.00368.x

10. Wilkinson CP, Ferris FL III, Klein RE, et al; Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110(9):1677-1682. doi:10.1016/S0161-6420(03)00475-5

11. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. Proceedings of the IEEE conference on computer vision and pattern recognition. http://cnnlocalization.csail.mit.edu/Zhou_Learning_Deep_Features_CVPR_2016_paper.pdf. Published 2016. Accessed July 8, 2019.

12. Abramoff M, Lavin PT, Birch M, Shah N, Folk J. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digital Medicine*. 2018;1(39):1-8. doi:10.1038/s41746-018-0040-6

13. Sunita M, Singh AK, Rogye A, et al. Prevalence of diabetic retinopathy in urban slums: the Aditya Jyot Diabetic Retinopathy in Urban Mumbai slums study, report 2. *Ophthalmic Epidemiol*. 2017;24(5):303-310. doi:10.1080/09286586.2017.1290258

14. Lundquist BM, Sharma N, Kewalramani K. Patient perceptions of eye disease and treatment in Bihar, India. *J Clinic Experiment Ophthalmol*. 2012;3:2. doi:10.4172/2155-9570.1000213

15. Walton OB IV, Garoon RB, Weng CY, et al. Evaluation of automated teleretinal screening program for diabetic retinopathy. *JAMA Ophthalmol*. 2016;134(2):204-209. doi:10.1001/jamaophthalmol.2015.5083

16. Kumar PNS, Deepak RU, Sathar A, Sahasranamam V, Kumar RR. Automated detection system for diabetic retinopathy using two field fundus photography. *Procedia Comput Sci*. 93:486-494.

17. Gupta V, Bansal R, Gupta A, Bhansali A. Sensitivity and specificity of nonmydriatic digital imaging in screening diabetic retinopathy in Indian eyes. *Indian J Ophthalmol*. 2014;62(8):851-856. doi:10.4103/0301-4738.141039

18. Vujosevic S, Benetti E, Massignan F, et al. Screening for diabetic retinopathy: 1 and 3 nonmydriatic 45-degree digital fundus photographs vs 7 standard early treatment diabetic retinopathy study fields. *Am J Ophthalmol*. 2009;148(1):111-118. doi:10.1016/j.ajo.2009.02.031

Original Article

Journal of Diabetes Science and Technology
1–7

DOI: 10.1177/19322968231194644
journals.sagepub.com/home/dst
Sage

Diabetic Retinopathy Screening Using Smartphone-Based Fundus Photography and Deep-Learning Artificial Intelligence in the Yucatan Peninsula: A Field Study

John J. Wroblewski, MD[1,2], Ermilo Sanchez-Buenfil, MD[3], Miguel Inciarte, MD[3], Jay Berdia, MD[2], Lewis Blake, PhD[4], Simon Wroblewski, BS[2], Alexandria Patti, BA[2], Gretchen Suter, BS[2], and George E. Sanborn, MD[5]

Abstract

Background: To compare the performance of Medios (offline) and EyeArt (online) artificial intelligence (AI) algorithms for detecting diabetic retinopathy (DR) on images captured using fundus-on-smartphone photography in a remote outreach field setting.

Methods: In June, 2019 in the Yucatan Peninsula, 248 patients, many of whom had chronic visual impairment, were screened for DR using two portable Remidio fundus-on-phone cameras, and 2130 images obtained were analyzed, retrospectively, by Medios and EyeArt. Screening performance metrics also were determined retrospectively using masked image analysis combined with clinical examination results as the reference standard.

Results: A total of 129 patients were determined to have some level of DR; 119 patients had no DR. Medios was capable of evaluating every patient with a sensitivity (95% confidence intervals [CIs]) of 94% (88%-97%) and specificity of 94% (88%-98%). Owing primarily to photographer error, EyeArt evaluated 156 patients with a sensitivity of 94% (86%-98%) and specificity of 86% (77%-93%). In a head-to-head comparison of 110 patients, the sensitivities of Medios and EyeArt were 99% (93%-100%) and 95% (87%-99%). The specificities for both were 88% (73%-97%).

Conclusions: Medios and EyeArt AI algorithms demonstrated high levels of sensitivity and specificity for detecting DR when applied in this real-world field setting. Both programs should be considered in remote, large-scale DR screening campaigns where immediate results are desirable, and in the case of EyeArt, online access is possible.

Keywords

artificial intelligence, diabetic retinopathy, fundus-on-phone camera, Mexico, rural health care, screening

Introduction

Approximately 463 million people live with diabetes worldwide, and about 700 million people are projected to have diabetes by 2045.[1] Individuals with diabetes are 25 times more likely to become blind than are those in the general population.[2] Thus, diabetic retinopathy (DR) is one of the leading causes of blindness worldwide, with sight-threatening DR affecting 28.5 million people.[3] In Mexico, increasing rates of obesity and a genetic predisposition for type 2 diabetes have led to the increasing prevalence (15.2%) of diabetes in adults.[4] In 2016, Mexico declared that diabetes had reached epidemic proportions and should be considered a major public health problem.[5] Furthermore, endocrine, nutritional, and metabolic diseases were the second leading causes of death in 2014.[6] Approximately

[1]Retina Care International, Hagerstown, MD, USA
[2]Cumberland Valley Retina Consultants, Hagerstown, MD, USA
[3]RetimediQ, Mérida, Mexico
[4]Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, USA
[5]Department of Ophthalmology, Virginia Commonwealth University, Richmond, VA, USA

Corresponding Author:
John J. Wroblewski, Retina Care International, 1150 Opal Ct., Hagerstown, MD 21740, USA.
Email: johnw@retinacare.net

one-third of Mexicans with diabetes have DR, with an alarming incidence of 38.9% in Chiapas.[7]

Early detection of DR can help prevent blindness in those with diabetes. However, early detection of DR is challenging because Mexico does not have a national DR screening program. Therefore, Retinacare International, a 501c3 US-based nonprofit organization, partnered with RetimediQ in 2005, a private ophthalmological clinic in Merida, Yucatan, to conduct annual and bi-annual DR screening campaigns throughout the state. To date, we have screened >6000 patients in seven cities and towns using dilated, indirect ophthalmoscopic examinations, often combined with high-magnification mydriatic funduscopic examinations.

Smartphone-based retinal imaging has emerged as an efficient, sensitive, specific, and cost-effective method for DR screening.[8-10] However, ophthalmologists or trained graders have been required to grade acquired images for the presence and severity of DR.[11,12] This grading requirement is not practical for screening in an outreach field setting. Thus, an automated image-grading system for detecting DR is needed to facilitate large-scale DR detection screening efforts and reduce health care provider burden.

Diabetic retinopathy detection can now be performed by computer-based analysis of fundus images using machine learning and artificial intelligence (AI).[13,14] Studies have demonstrated that deep-learning algorithms can accurately detect and grade DR in digital fundus images.[15-19] Others have investigated the feasibility of DR detection using smartphone-based fundus photography (Remidio fundus-on-phone [FOP] camera[18] combined with an offline (Remidio Medios)[20] or online version of AI software [Eyenuk EyeArt]).[21] These studies determined that both versions of AI software show high sensitivity and specificity for detecting DR in images acquired by smartphone-based fundus photography in tertiary care centers.[20,21]

This retrospective, noninterventional AI validation analysis compared the diagnostic accuracy of the offline Medios AI software and online EyeArt AI software for detecting DR on a single set of patient images acquired using the ultra-portable Remidio-FOP camera in an outreach field setting.

Methods

Patients

A total of 248 consecutive patients with a known history of diabetes were invited to participate in a DR screening campaign in the cities of Valladolid and Merida, Yucatan, Mexico, in June 2019. Each patient was assigned a unique identification number, and all data, including fundus images, were de-identified to ensure patient confidentiality. The study protocol (EXT-22-01) was approved by the Research Ethics Committee of the Association to Avoid Blindness in Mexico I.A.P (CONBIOETICA-09-CEI-006-20170306) and the Committee of the Association to Avoid Blindness in Mexico I.A.P. (COFEPRIS: 17 CI 09 003 142). Written informed consent was obtained from all patients.

Screening

Three graduate students (AP, SW, GES) who did not have professional experience in fundus photography acquired the images of dilated eyes using two portable Remidio-FOP cameras (Remidio Innovative Solutions Pvt Ltd, Karnataka, India). A minimum of three fundus fields (ie, posterior pole [disc and macula], nasal, temporal) were attempted to be captured for each eye using the portable devices mounted on a table stand.

The offline AI algorithm on the Remidio smartphone flagged images rated as poor quality and prompted the operator to take additional pictures of the same or near retinal view until the images were deemed acceptable by the AI system. A retinologist (ES-B, MI, JJW) performed indirect ophthalmoscopic examinations on all patients, which were often combined with a high-magnification funduscopic examination using the slit lamp. The presence of DR was determined based on the clinical information collected. A patient was positive for DR if any degree of DR was identified in at least one eye.

Image Analysis

After the DR screening campaign, image sets were fully analyzed by the Medios offline automated application integrated into the smartphone-based retinal imaging system. This application has two components: (1) an algorithm that checks the quality of the images and (2) a mechanism that generates an image-level diagnosis (or not) of DR. If an image is of marginal-to-poor quality, the image quality notification function flagged the image. However, this function could be overridden manually, allowing all patient image sets to be assessed. If at least one image in one eye was positive for any level of DR using the Medios AI software, the patient had a positive DR result.

All images captured by the two smartphones were evaluated further at a later date in masked fashion by two experienced retinologists (GES, JJW) for the presence or absence of DR. The DR status of the patients was determined using the images alone. When the two retinologists did not agree, a third, adjudicating retinologist (JB) evaluated the images to determine the presence or absence of DR.

The "ground truth" presence or absence of DR for each patient was determined using the photographic image determination combined with their clinical examination result. The "true" presence of DR was defined as any DR observed on clinical examination or DR missed on clinical examination but identified on one or more photographs by the two masked readers. The "true" absence of DR was defined as a negative clinical examination (ie, no degree of DR observed) combined with both masked readers not detecting any degree of DR in any photographs from either eye of the patient.

For the online analysis (after the campaign and masked review of images), all images (JPEGs) were uploaded to the Remidio confidential website and then sent to Eyenuk for analysis using an automated process with machine learning-enabled software, EyeArt v2.1. The architecture, data composition, and clinical validation studies for the Medios[18,20] and EyeArt algorithms[13,21] have been described. Each gradable patient was given a diagnosis (or not) of DR.

The Medios and EyeArt results were compared individually with the "true" presence or absence of DR (per the above criteria). A head-to-head comparison was performed in the limited number of patients for whom every image was deemed to be of good quality by Medios and whose image set was deemed to be gradable by EyeArt. This allowed for an "apples-to-apples" comparison, where all of the patient-level image criteria were met by both AI systems.

Statistical Analysis

The efficacy of the two software programs was evaluated using the following metrics: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR−), and Youden's *J* statistic (combines sensitivity and specificity into a single measure of each of the algorithm's performance). The closer *J* is to 1, the closer the software is to having no false positives and no false negatives. Each of these metrics provides unique insight into the performance of the diagnostic test.[22]

Apparent prevalence of DR was calculated as the number of patients who tested positive divided by the total number of patients tested. The Rogan-Gladen estimator[23] was used to approximate true prevalence. An exact 95% confidence interval (CI) was calculated using the Clopper-Pearson approach for the point estimates of apparent prevalence, true prevalence, sensitivity, specificity, PPV, NPV, and Youden's *J*.[24] For LR+ and LR−, 95% CIs were calculated using methods described by Simel et al.[25]

Results

Of the 248 patients, 212 were female (mean age 56.4 years [range: 5-80]) and 36 were male (mean age 55.9 years [range: 12-73]). Mean duration of diabetes was 14.7 years (range: 1-39) among female patients and 12.9 years (range: 5-21) among male patients.

A total of 2130 images were acquired. The "ground truth" results for the presence or absence of DR were: 129 patients were diagnosed as having some level of DR and 119 patients did not have any degree of DR. The two masked readers agreed on the presence or absence of DR for 92% (228/248) of patients; images for 20 patients required the third masked reader (JB) to adjudicate and determine the presence or absence of any DR.

Of the 248 patients, Medios recognized 82 patients who had at least one poor-quality image. However, Medios was able to analyze every image, thus yielding a DR determination on all 248 patients. EyeArt was able to evaluate 46 of these patients. Of the 129 patients with "true" DR, the Medios software identified 121 patients as having DR and 8 as not having DR. Table 1 summarizes the performance metrics of the Medios software based on these results.

EyeArt was able to evaluate and thus yield a DR determination in 156 of the 248 patients. The "true" results for the presence or absence of DR in this subset of patients determined that 82 patients had some level of DR; 74 patients did not have DR. EyeArt identified 87 patients as having DR and 69 patients as not having DR. Table 2 summarizes the

Table 1. Medios Analysis (N = 248).

Metric	Equation	Result	Point estimate	(95% CI)
Apparent prevalence (AP)	TP + FP / N	121 + 7 / 248	0.52	(0.45-0.58)
True prevalence	AP + (SP − 1) /SP + (SE − 1)	0.52 + (0.94 − 1) / 0.94 + (0.94 − 1)	0.52	(0.46-0.58)
Sensitivity (SE)	TP / TP + FN	121 / 121 + 8	0.94	(0.88-0.97)
Specificity (SP)	TN / TN + FP	112 / 112 + 7	0.94	(0.88-0.98)
PPV	TP / TP + FP	121 /121 + 7	0.95	(0.89-0.98)
NPV	TN / TN + FN	112 /112 + 8	0.93	(0.87-0.97)
LR+	SE / (1 − SP)	0.94 / 1 − 0.94	15.95	(7.76-32.76)
LR−	(1 − SE) / SP	(1 − 0.94) / 0.94	0.07	(0.03-0.13)
J statistic	SE + SP − 1	0.94 +0.94 − 1	0.879	(0.76-0.95)

Abbreviations: CI, confidence interval; TP, true positive; FP, false positive; SP, specificity (estimated probability that a patient without DR tests as not having DR); DR, diabetic retinopathy; SE, sensitivity (estimated probability that a patient with "true" DR tests as having DR); FN, false negative; TN, true negative; PPV, positive predictive value (% of positive tests that are TPs); NPV, negative predictive value (% of negative tests that are TNs); LR−, negative likelihood ratio (estimate of the probability that a patient who has DR is predicted as not having DR divided by the probability that a patient who does not have DR is predicted as not having DR); LR+, positive likelihood ratio (estimate of the probability that a patient with DR is predicted as having DR divided by the probability that a patient who does not have DR is predicted as having DR).

Table 2. EyeArt Analysis (N = 156).

Metric	Equation	Result	Point estimate	(95% CI)
Apparent prevalence (AP)	TP + FP / N	77 + 10 / 156	0.56	(0.48-0.64)
True prevalence	AP + (SP − 1) / SP + (SE − 1)	0.56 + (0.86 − 1) / 0.86 + (0.94 − 1)	0.53	(0.44-0.61)
Sensitivity (SE)	TP / TP + FN	77 / 77 + 5	0.94	(0.86-0.98)
Specificity (SP)	TN / TN + FP	64 / 64 + 10	0.86	(0.77-0.93)
PPV	TP / TP + FP	77 / 77 + 10	0.89	(0.80-0.94)
NPV	TN / TN + FN	64 / 64 + 5	0.93	(0.84-0.98)
LR+	SE / (1 − SP)	0.94 / (1 − 0.86)	6.95	(3.89-12.40)
LR−	(1 − SE) / SP	(1 − 0.94) / 0.86	0.07	(0.03-0.17)
J statistic	SE + SP − 1	0.94 + 0.86 − 1	0.804	(0.63-0.91)

Abbreviations: CI, confidence interval; TP, true positive; FP, false positive; SP, specificity (estimated probability that a patient without DR tests as not having DR); DR, diabetic retinopathy; SE, sensitivity (estimated probability that a patient with "true" DR tests as having DR); FN, false negative; TN, true negative; PPV, positive predictive value (% of positive tests that are TPs); NPV, negative predictive value (% of negative tests that are TNs); LR−, negative likelihood ratio (estimate of the probability that a patient who has DR is predicted as not having DR divided by the probability that a patient who does not have DR is predicted as not having DR); LR+, positive likelihood ratio (estimate of the probability that a patient with DR is predicted as having DR divided by the probability that a patient who does not have DR is predicted as having DR).

performance metrics of the EyeArt software based on these results.

A total of 110 patients had image sets that were deemed completely to be of good quality by Medios and gradable by EyeArt. In this head-to-head comparison, the "true" results for the presence or absence of DR were: 76 patients had some level of DR and 34 patients did not have DR. Table 3 summarizes the performance metrics of the head-to-head comparison of the Medios results and the EyeArt results.

Discussion

This is the first analysis to compare the head-to-head performance of two AI algorithms for detecting the presence of any degree of DR in a field setting using a portable fundus camera. Although a real-world, head-to-head analysis comparing multiple AI DR screening algorithms has been published, the analysis was performed on images obtained with a nonportable fundus camera in two strictly controlled Veterans' Affairs primary care clinical environments and did not include the Medios AI algorithm.[26] To our knowledge, our analysis is the first to use the Medios and EyeArt AI algorithms to analyze images from patients of Spanish-Mexican and Mayan descent in the Yucatan Peninsula, each having a distinct genetic phenotype.

The offline smartphone-based Medios AI algorithm was highly sensitive and specific for detecting the presence or absence of DR in binary fashion. These results are consistent with those from previous studies.[12,18,20,27] For the Medios AI analysis, the sensitivities and specificities of 0.94 and 0.94 for all 248 patients and 0.99 and 0.88 for the 110 patients in the head-to-head comparison were comparable with those detecting any DR in patients of Indian descent.[12,18]

For the cloud-based EyeArt AI algorithm analysis, the sensitivities and specificities of 0.94 and 0.86 for 156 patients with all gradable images and 0.95 and 0.88 in the head-to-head comparison of 110 patients were comparable with the results of previous reports for the presence of any DR and sight-threatening DR for English,[28-30] Indian,[21] and American patients.[31,32] Interestingly, both AIs had similar sensitivities and specificities despite being trained by different methodologies.[13,22,23] This is made more remarkable because DR phenotypes differ by region and ethnicity,[3] and the two AI algorithms used in this analysis were trained on different international DR data sets of patients of Indian descent[14,20,27] for Medios and of American and Northern Mexican descent[13,33] for EyeArt.

Because three inexperienced graduate students were tasked as camera operators, we confirmed that once they became familiar with the technology, the Remidio-FOP camera with accompanying Medios AI was easy to use, as previously described.[12,20] Interestingly, all 248 patients had images that Medios was able to analyze. This can be explained by its proprietary, two-step process: (1) an upstream image quality module with a per image notification function and (2) a downstream image automated analysis application module trained to recognize any sign of DR. Sensitivity increases at the potential expense of specificity. Although any image set could be deemed of marginal quality based on a single image of poor quality, the images could still be subjected to the evaluation module at the camera operator's discretion. Thus, Medios was able to evaluate all 248 patients, even though 82 patients had at least one poor-quality image. Only 67% of patients had a complete set of good-quality images. While there may be potential safety issues (eg, increased number of false negatives) with a strategy that allows for the analysis of images of marginal-to-poor quality, the overall benefit in an outreach field setting—where media opacities, patient compliance, and image acquisition speed have a greater impact—cannot be marginalized. Many of the 82 patients with

Table 3. Head-to-Head (Medios/EyeArt) Performance (N = 110).

Metric	Equation	Result	Point estimate	(95% CI)
Apparent prevalence (AP)	TP + FP / N	75 + 4 / 110	0.72	(0.62-0.80)
		74 + 4 / 110	0.69	(0.60-0.78)
True prevalence	AP + (SP − 1) / SP + (SE − 1)	0.72 + (0.88 − 1) / 0.88 + (0.99 − 1)	0.69	(0.60-0.78)
			0.69	(0.60-0.78)
		0.69 + (0.88 − 1) / 0.88 + (0.95 − 1)		
Sensitivity (SE)	TP / TP + FN	75 / 75 + 1	0.99	(0.93-1.00)
		74 / 74 + 4	0.95	(0.87-0.99)
Specificity (SP)	TN / TN + FP	30 / 30 + 4	0.88	(0.73-0.97)
		30 / 30 + 4	0.88	(0.73-0.97)
PPV	TP / TP + FP	75 / 75 + 4	0.95	(0.88-0.99)
		75 / 75 + 4	0.95	(0.87-0.99)
NPV	TN / TN + FN	30 / 30 + 1	0.97	(0.83-1.00)
		30 / 30 + 4	0.88	(0.73-0.97)
LR+	SE / (1 − SP)	0.99 / (1 − 0.88)	8.39	(3.34-21.07)
		0.95 / (1 − 0.88)	8.05	(3.20-20.25)
LR−	(1 − SE) / SP	(1 − 0.99) / 0.88	0.01	(0.00-0.10)
		(1 − 0.95) / 0.88	0.06	(0.02-0.16)
J statistic	SE + SP − 1	0.99 + 0.88 − 1	0.869	(0.65-0.97)
		0.95 + 0.88 − 1	0.830	(0.60-0.95)

Abbreviations: CI, confidence interval; TP, true positive; FP, false positive; SP, specificity (estimated probability that a patient without DR tests as not having DR); DR, diabetic retinopathy; SE, sensitivity (estimated probability that a patient with "true" DR tests as having DR); FN, false negative; TN, true negative; PPV, positive predictive value (% of positive tests that are TPs); NPV, negative predictive value (% of negative tests that are TNs); LR−, negative likelihood ratio (estimate of the probability that a patient who has DR is predicted as not having DR divided by the probability that a patient who does not have DR is predicted as not having DR); LR+, positive likelihood ratio (estimate of the probability that a patient with DR is predicted as having DR divided by the probability that a patient who does not have DR is predicted as having DR).

marginal-quality images had obvious DR stigmata identified by the two masked readers. However, portions of the image often were partially obscured by vitreous or preretinal hemorrhage, asteroid hyalosis, or cataract; thus, the images were deemed to be of marginal quality. The greater acquisition of fundus images, including those of marginal quality, can potentially improve the quality of the screening process by identifying more patients with sight-threatening DR and perhaps early DR.

EyeArt evaluated 63% (156/248) of the patients, primarily owing to camera operator error. During the screening campaign, the three camera operators were untrained and not familiar with EyeArt's mandatory image-specific and patient-specific capture criteria.[21] EyeArt processes are designed to produce patient-level results, not individual image-level results. Therefore, if any image was considered ungradable, the entire patient encounter was deemed ungradable because EyeArt does not skip ungradable images. Similar to Medios, EyeArt is trained to reduce the incidence of false negatives. However, EyeArt has a one-step AI algorithm that combines image quality/gradeability and image analysis for the presence of disease. As such, variations in image alignment, resolution, and exposure; not having a macula-centered image per eye; monocular status; or having >14 images per patient resulted in an ungradable encounter and a recommendation for referral to an eye specialist.

The high rate of ungradable images by EyeArt is in contrast to the results of a study conducted in primary care, general ophthalmology, and retina specialty centers with trained photographers using a nonportable tabletop Canon camera.[33] This study, which excluded participants with persistent visual impairment, demonstrated a 97.4% dilated eye gradable rate.[32] Another possible reason for EyeArt's high ungradable rate was the high incidence of media opacities observed. Incidence rates for these conditions tend to be higher in underserved outreach field settings. In addition, these imaging factors underscore the need for dilation in an outreach setting, as the incidence of ungradability has been shown to be substantially higher for nonmydriatic vs mydriatic images.[32,34]

In the head-to-head comparison—where all images for every patient were deemed of sufficient quality for grading by both AI algorithms—the sensitivity, specificity, and PPVs/NPVs of Medios and EyeArt were comparable and highly accurate. This very high degree of DR detection accuracy (ie, sensitivity) may be a function, at least in part, of the high DR prevalence (69%) in this cohort. These results are not typical in Western populations and are more representative of underserved populations. The high degree of sensitivity also could be linked to the screening of patients of Mexican-Spanish and Mayan descent who have a greater degree of background contrast from choroidal melanin and

fundus hyperpigmentation, which is often not seen in the blonde fundi of Western populations.[19] This also is in contrast to a study showing lower specificities when a significant portion of the screened population had mild nonproliferative retinopathy or no retinopathy.[29] Thus, provided that online access is available, either AI algorithm should be adequate at providing a relatively immediate and robust determination of the presence or absence of DR in a large-scale outreach campaign.

The primary limitation of this analysis is that the definition of "ground truth" DR was unconventionally determined by clinical examination findings combined with image analysis performed by two masked, fellowship-trained, vitreoretinal specialists, each with >30 years' experience. In previous reports evaluating both AI algorithms, three 45° and four wide-field tabletop fundus photographs evaluated by masked or expert readers were used to define "ground truth" DR.[12,32] However, these conventional methods were not feasible in a large-scale, real-world, five-day DR screening campaign in an outreach setting. Furthermore, by adding the clinical examination (including a 20-diopter indirect ophthalmoscopic evaluation) into the definition of "ground truth" DR, eyes with DR primarily limited to the pre-equatorial fundus were properly characterized. Although both AI algorithms were trained using only images of the posterior pole, the pattern recognition and deep-learning ability of each algorithm could potentially allow each to grade an image as having some degree of DR in the absence of any classic stigmata of DR visible to the naked eye on a photographic image. Pathology, including microvascular ischemia (in the absence of microaneurysm, intraretinal hemorrhage, or cotton wool spot), choriocapillary ischemic thinning, and primary neuronal cell loss may lead to retinal thinning, which is only recognizable by deep learning. As such, our definition of "ground truth," which accounts for these possibilities, potentially increased the accuracy of the results.

Other limitations included using untrained camera operators pressed for time and screening a smaller number of patients who had images deemed acceptable by both AI algorithms. This analysis was limited to detecting any degree of DR in a binary fashion and did not include grading the level of retinopathy if present. This was necessary, as the clinical examination reports did not segregate referable DR from mild nonproliferative DR. In addition, Medios is only trained to recognize the presence of moderate nonproliferative diabetic retinopathy or worse or no DR as a binary outcome. Nonetheless, detecting any level of DR in this underserved and likely poorly controlled population with diabetes takes on an added level of importance where rapid diabetic retinal disease progression and blindness from cataract formation and glaucoma are more probable.[35]

Conclusions

Our results demonstrate that both Medios and EyeArt AI-enabled algorithms can be effective in achieving high accuracy in an outreach field setting where portable fundus cameras are used and where medical professionals and other resources are limited. These robust findings suggest this same methodology of DR screening could be readily implemented in any office setting or location in the United States. The benefit of immediately determining, in an offline manner with Medios, the presence or absence of DR cannot be overstated. Both AIs should be considered equally viable options in large-scale DR screening campaigns where rapid results are needed and, in the case of EyeArt, online access is possible.

Abbreviations

AI, artificial intelligence; AP, apparent prevalence; CI, confidence interval; DR, diabetic retinopathy; FN, false negative; FOP, fundus on phone; FP, false positive; LR−, negative likelihood ratio; LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; SE, sensitivity; SP, specificity; TN, true negative; TP, true positive.

Acknowledgments

The authors acknowledge Medios Technologies, Singapore, for providing an oral description of the technical design of their software and assistance with image transfers. The authors thank Eyenuk, Los Angeles, CA, for providing a description of EyeArt's functionality and technical support in the submission of acquired images to EyeArt. Linda Goldstein, PhD, CMPP, provided medical writing assistance funded by Retinacare International.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. None of the authors have any conflicts of interest to disclose. No author has a relationship of any kind with Remidio, Medios Technologies, or Eyenuk. The two Remidio-FOP cameras and smartphones were purchased by and are the property of Retina Care International. Retina Care International did not receive any funding of any kind for this study. Medios Technologies and Eyenuk provided the AI for use in the study and had no role in the study design, funding, execution, data collection and analysis, or publication.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

John J. Wroblewski https://orcid.org/0000-0002-0600-8666
George E. Sanborn https://orcid.org/0000-0002-0444-6988

References

1. International Diabetes Federation. IDF diabetes atlas. https://diabetesatlas.org/en/. Accessed August 11, 2021.
2. Kahn HA, Hiller R. Blindness caused by diabetic retinopathy. *Am J Ophthalmol*. 1974;78(1):58-67.
3. Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045:

systematic review and meta-analysis. *Ophthalmology*. 2021;128(11):1580-1591.
4. Villalpando S, de la Cruz V, Rojas R, et al. Prevalence and distribution of type 2 diabetes mellitus in Mexican adult population: a probabilistic survey. *Salud Publica Mex*. 2010;52(suppl 1):S19-S26.
5. Mendoza-Herrera K, Quezada AD, Pedroza-Tobias A, Hernandez-Alcaraz C, Fromow-Guerra J, Barquera S. A diabetic retinopathy screening tool for low-income adults in Mexico. *Prev Chronic Dis*. 2017;14:E95.
6. Pan American Health Organization. *Health in the Americas+, 2017 Edition. Summary: Regional Outlook and Country Profiles*. World Health Organization; 2017. https://iris.paho.org/handle/10665.2/34321.
7. Polack S, Yorston D, López-Ramos A, et al. Rapid assessment of avoidable blindness and diabetic retinopathy in Chiapas, Mexico. *Ophthalmology*. 2012;119(5):1033-1040.
8. Rajalakshmi R, Arulmalar S, Usha M, et al. Validation of smartphone based retinal photography for diabetic retinopathy screening. *PLoS ONE*. 2015;10(9):e0138285.
9. Russo A, Morescalchi F, Costagliola C, Delcassi L, Semeraro F. Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy. *Am J Ophthalmol*. 2015;159(2):360-364.e1.
10. Ryan ME, Rajalakshmi R, Prathiba V, et al. Comparison Among Methods of Retinopathy Assessment (CAMRA) study: smartphone, nonmydriatic, and mydriatic photography. *Ophthalmology*. 2015;122(10):2038-2043.
11. Malerbi FK, Andrade RE, Morales PH, et al. Diabetic retinopathy screening using artificial intelligence and handheld smartphone-based retinal camera. *J Diabetes Sci Technol*. 2022;16(3):716-723.
12. Sengupta S, Sindal MD, Baskaran P, Pan U, Venkatesh R. Sensitivity and specificity of smartphone-based retinal imaging for diabetic retinopathy: a comparative study. *Ophthalmol Retina*. 2019;3(2):146-153.
13. Bhaskaranand M, Ramachandra C, Bhat S, et al. Automated diabetic retinopathy screening and monitoring using retinal fundus image analysis. *J Diabetes Sci Technol*. 2016;10(2):254-261.
14. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.
15. 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;1:39.
16. 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(9):987-993.
17. Lim JI, Regillo CD, Sadda SR, et al. Artificial intelligence detection of diabetic retinopathy: subgroup comparison of the EyeArt system with ophthalmologists' dilated examinations. *Ophthalmol Sci*. 2023;3(1):100228.
18. 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.
19. Raman R, Srinivasan S, Virmani S, Sivaprasad S, Rao C, Rajalakshmi R. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. *Eye*. 2019;33(1):97-109.
20. Sosale B, Sosale AR, 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.
21. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*. 2018;32(6):1138-1144.
22. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35.
23. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol*. 1978;107(1):71-76.
24. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26:404-413.
25. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol*. 1991;44(8):763-770.
26. Lee AY, Yanagihara RT, Lee CS, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. *Diabetes Care*. 2021;44(5):1168-1175.
27. 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(1):e000892.
28. Bhaskaranand M, Ramachandra C, Bhat S, et al. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technol Ther*. 2019;21(11):635-643.
29. Heydon P, Egan C, Bolter L, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol*. 2021;105(5):723-728.
30. Olvera-Barrios A, Heeren TF, Balaskas K, et al. Diagnostic accuracy of diabetic retinopathy grading by an artificial intelligence-enabled algorithm compared with a human standard for wide-field true-colour confocal scanning and standard digital retinal images. *Br J Ophthalmol*. 2021;105(2):265-270.
31. Mokhashi N, Grachevskaya J, Cheng L, et al. A comparison of artificial intelligence and human diabetic retinal image interpretation in an urban health system. *J Diabetes Sci Technol*. 2022;16(4):1003-1007.
32. 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.
33. Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *J Diabetes Sci Technol*. 2009;3(3):509-516.
34. Piyasena M, Yip JLY, MacLeod D, Kim M, Gudlavalleti VSM. Diagnostic test accuracy of diabetic retinopathy screening by physician graders using a hand-held non-mydriatic retinal camera at a tertiary level medical clinic. *BMC Ophthalmol*. 2019;19(1):89.
35. Shah P, Mishra DK, Shanmugam MP, Doshi B, Jayaraj H, Ramanjulu R. Validation of Deep Convolutional Neural Network-based algorithm for detection of diabetic retinopathy—artificial intelligence versus clinician for screening. *Indian J Ophthalmol*. 2020;68(2):398-405.

Original Article

Evaluation of an AI algorithm trained on an ethnically diverse dataset to screen a previously unseen population for diabetic retinopathy

Divya P Rao[1], Florian M Savoy[2], Anand Sivaraman[3], Sreetama Dutt[3], Marianne Shahsuvaryan[4,5], Nairuhi Jrbashyan[6], Narine Hambardzumyan[5], Nune Yeghiazaryan[5], Taraprasad Das[7]

Purpose: This study aimed to determine the generalizability of an artificial intelligence (AI) algorithm trained on an ethnically diverse dataset to screen for referable diabetic retinopathy (RDR) in the Armenian population unseen during AI development. **Methods:** This study comprised 550 patients with diabetes mellitus visiting the polyclinics of Armenia over 10 months requiring diabetic retinopathy (DR) screening. The Medios AI-DR algorithm was developed using a robust, diverse, ethnically balanced dataset with no inherent bias and deployed offline on a smartphone-based fundus camera. The algorithm here analyzed the retinal images captured using the target device for the presence of RDR (i.e., moderate non-proliferative diabetic retinopathy (NPDR) and/or clinically significant diabetic macular edema (CSDME) or more severe disease) and sight-threatening DR (STDR, i.e., severe NPDR and/or CSDME or more severe disease). The results compared the AI output to a consensus or majority image grading of three expert graders according to the International Clinical Diabetic Retinopathy severity scale. **Results:** On 478 subjects included in the analysis, the algorithm achieved a high classification sensitivity of 95.30% (95% CI: 91.9%–98.7%) and a specificity of 83.89% (95% CI: 79.9%–87.9%) for the detection of RDR. The sensitivity for STDR detection was 100%. **Conclusion:** The study proved that Medios AI-DR algorithm yields good accuracy in screening for RDR in the Armenian population. In our literature search, this is the only smartphone-based, offline AI model validated in different populations.

Access this article online

Website: https://journals.lww.com/ijo

DOI: 10.4103/IJO.IJO_2151_23

Quick Response Code:

Key words: Artificial Intelligence, deep learning, diabetic retinopathy, eye screening

Diabetic retinopathy (DR), a microvascular complication of diabetes mellitus (DM), is one of the leading causes of preventable blindness. It is estimated that 642 million people would be living with diabetes by 2040 worldwide.[1] The global prevalence of DR among people with diabetes is 34.6%, and it is 10.2% for sight-threatening diabetic retinopathy (STDR).[2] Over the past decade, the number of people with diabetes has increased.[1] Such high numbers not only pose a great economic burden but also create an ever-increasing demand for accessible eye care. Artificial intelligence (AI) can help bridge the otherwise widening gap between ophthalmologists and patients.

Current deep learning (DL)-based AI algorithms have shown performances approaching that of clinicians in detecting DR.[3–7] This encourages the deployment of such systems to reduce the burden on ophthalmologists. There is a requirement across all geographies to tackle the global problem of preventable blindness. It is thus important to focus on unbiased and robust AI systems, which work equally well across ethnicities and populations.

Patient attributes, such as race/ethnicity, can introduce biases in AI systems, and these biases pose significant challenges in the development of AI-based models for DR screening. It is crucial for effective solutions deployed in different geographies to demonstrate consistent accuracy across diverse populations.[4,8,9] It should be noted that ethnic groups with darker skin tend to have higher melanin content within their uveal melanocytes, resulting in darker fundus pigmentation.[10,11] Consequently, although DR lesions are similar across all ethnic groups, the background color of the fundus can make these lesions more or less distinct. Thus, fundus pigmentation may also impact the interpretation of AI systems.[10]

We hypothesize that an AI algorithm can generalize across different populations. This should hold true even if the population is not represented in the training set. This assumption relies on an ethnically balanced and diverse

[1]AL& ML, Remidio Innovative Solutions, Inc, Glen Allen, USA, [2]AI&ML, Medios Technologies Pte Ltd, Remidio Innovative Solutions, Singapore, [3]AI&ML, Remidio Innovative Solutions Pvt Ltd, Bengaluru, India, [4]Ophthalmology, Yerevan State Medical University, [5]Armenian Eyecare Project, [6]Dept of Economics and Management, Yerevan State University, Armenia, [7]Vitreoretinal Services, Kallam Anji Reddy Campus, LV Prasad Eye Institute, Hyderabad, India

Correspondence to: Dr. Divya P Rao, Remidio Innovative Solutions, Inc and Medios Technologies Pte Ltd are Wholly Owned Subsidiaries of Remidio Innovative Solutions Pvt Ltd, Bengaluru, India. E-mail: drdivya@remidio.com

Received: 11-Aug-2023 Revision: 22-Dec-2023
Accepted: 02-Feb-2024 Published: 29-Jul-2024

Cite this article as: Rao DP, Savoy FM, Sivaraman A, Dutt S, Shahsuvaryan M, Jrbashyan N, *et al*. Evaluation of an AI algorithm trained on an ethnically diverse dataset to screen a previously unseen population for diabetic retinopathy. Indian J Ophthalmol 2024;72:1162-7.

dataset. The Medios AI-DR is an offline, smartphone-based algorithm trained on a diverse dataset that did not include Armenian eyes. We report here the accuracy of this algorithm on an Armenian population for the detection of referable diabetic retinopathy (RDR). The scope of the DR screening program in Armenia falls under the aegis of the Armenian Eye Care Project (AECP) team in cooperation with the Armenian Ministry of Health and the World Diabetes Foundation with an aim to end preventable blindness due to DR.[12] The AECP project is part of a larger initiative led by this team to ensure accessible eye care for all Armenians.

Methods

This retrospective study was approved (IRB approval no.: N4-3/2020) by the institutional ethics committee. It was performed according to the tenets of the Declaration of Helsinki.

Study Population Sampling and Imaging Protocol: Patients with established diabetes attending the polyclinics of Armenia for DR screening were included in this study. The images were captured using the Fundus-on-Phone (FOP NM-10, Remidio Innovative Solutions), a smartphone-based fundus camera with a field of view of 40°, and collected over a 10-month period between July 2019 and April 2020. The study population included 550 consecutive subjects. Subjects with no established ground truth diagnosis were excluded.

Retinal image acquisition: The images were captured by non-medically trained operators in real-world conditions. The imaging protocol consisted of acquiring one disc and one macula-centered image per eye of each patient without dilation. This was performed during routine DR screening. Repeat images were captured when required to ensure sufficient image quality.

Image Grading: The consensus or majority image grading of three expert graders, two fellowship-trained ophthalmologists, and one certified optometrist formed the reference standard for assessing the AI algorithm. One grader, a fellowship-trained ophthalmologist, provided image-based diagnosis during the screening program. In total, 53 patients had no consensus after grading by the three graders. They were presented to two senior retina specialists, whose adjudicated diagnosis was deemed final. All graders were masked to the Medios AI-DR output and to each other's grading. The de-identified images were graded for the stage of DR and the presence of diabetic maculopathy. The International Clinical Diabetic Retinopathy classification was used to grade images. The images were graded as either no DR, mild non-proliferative DR (mild NPDR), moderate non-proliferate DR (moderate NPDR), severe non-proliferate DR (severe NPDR), proliferative DR (PDR), or ungradable. The graders also looked for hard exudates within 1 disc diameter of the fovea center. This was used as a surrogate marker for the presence of clinically significant diabetic macular edema (CSDME), which is considered a standard guideline in a screening context in the absence of stereo imaging.[13–15] An image was deemed ungradable if a reliable diagnosis of DR was not possible. This could happen in two distinct scenarios: 1) if major vessels could not be clearly identified, or 2) due to blurring, artifacts, under/over-exposure, or glare spanning half or more of the image. The patient-level diagnosis was inferred by the DR stage of the more affected eye. The consensus grading for each patient formed the final diagnosis. RDR was defined as moderate NPDR or higher severity and/or the presence of CSDME. STDR was defined as severe NPDR or higher severity and/or the presence of CSDME.

AI-based software architecture: The Medios AI-DR consists of an ensemble of two convolutional neural networks based on the Inception-V3 architecture. It classifies color fundus images for the presence of RDR. The detailed software architecture has been previously published.[6] In brief, the training set consisted of 52,894 images as follows: 34,278 images were obtained from the Eye Picture Archive Communication System telemedicine program (EyePACS LLC, Santa Cruz, California). In total, 14,266 mydriatic images were captured using a Kowa VX-10α (Kowa American Corporation, CA, USA) at a tertiary diabetes center in India, and 4350 non-mydriatic images were taken in mass screening camps in India by using the Remidio FOPNM-10. The dataset was curated to contain as many referral cases as healthy ones.

Automated image analysis: The analysis was performed on the FOP NM10 with offline Medios AI-DR that does not require internet to provide a report. The de-identified images were loaded through Remidio's secure cloud connectivity system. The Medios AI-DR was manually run as per standard protocol. Following an automated analysis of image quality, each patient underwent an automated analysis for DR. The AI output No RDR or RDR, as well as the image quality analysis results, were noted. The "proceed anyway" override option was used for images that failed the AI quality check but received a consensus grading by the experts.

Outcome measure: The primary outcome measures were the sensitivity, specificity, false positives, false negatives, and predictive values (positive and negative) of the Medios AI-DR algorithm for detecting RDR on images captured using FOPNM-10 on this population. The secondary outcome measures were the sensitivity, specificity, and predictive values (positive and negative) of the algorithm for detecting any grade of DR. In addition, the sensitivity for detecting STDR was reported.

Statistical analysis: Considering a sensitivity of at least 80% with a precision of 10% and a prevalence of RDR of 20%, the required sample size was 308 patients for a 95% confidence interval.[16] We, however, looked at a larger sample size of 550 patients imaged over 10 months.

Next, 2 * 2 confusion matrices were used to compute the sensitivity (true positive [TP] rate) and specificity (true negative [TN] rate) to detect RDR, any stage of DR, and STDR. The positive predictive value (PPV) and the negative predictive value (NPV) were evaluated. Furthermore, 95% confidence intervals (CI) were calculated for sensitivity, specificity, NPV, and PPV. The false positive (FP) rate was calculated as FP/FP + TN. The false negative rate was calculated as FN/FN + TP. The kappa statistic was used to determine the agreement between each expert grader to the consensus grading. A kappa value of above 0.8 was considered as high agreement, between 0.5 and 0.79 as moderate, and below 0.5 as poor agreement. Data were analyzed using the pandas (1.1.0), NumPy (1.19.5), and scikit-learn (0.23.1) libraries in Python (ver 3.7.7).

Results

The mean age in this study cohort of 550 subjects was 61.6 ± 9.94 years (range: 12–83 years). Among them,

63.45% (n = 349) were females. In the final analysis, 478 subjects were included after excluding duplicate entries (n = 6) and patients deemed ungradable (n = 66) [Fig. 1].

Based on consensus image diagnosis by experts, any DR was present in 159 (33.26%) patients, RDR was seen in 149 patients (31.17%), and STDR in 62 patients (12.97%). The intergrader agreement (quadratic weighted kappa) between the individual certified experts and the majority diagnosis was moderately good (0.56–0.72).

In total, 478 subjects were fed to the AI to generate outputs for the presence or absence of RDR. The sensitivity for RDR was 95.30 % (95% CI: 91.9%–98.7%) and specificity was 83.89% (95% CI: 79.9%–87.90%). The NPV, that is, the probability of a subject with a negative screening test by the AI to truly not have RDR, was high (97.53%, 95% CI: 95.7%–99.3%). The key performance metrics for Medios AI-DR on the Armenian population are listed in Tables 1A and 1B.

The AI made a false diagnosis of RDR in 53 subjects. Four subjects had a consensus grading of mild NPDR, and the other 49 subjects had a consensus grading of no DR by the graders. Thus, the false positive rate was 16.11%. Among the 53 subjects, 17 were graded with AMD by at least one of the graders, while another nine were graded as having another pathology.

Table 1A: Confusion matrix to evaluate Medios AI- DR performance for RDR

n=478	Image grading RDR positive	Image grading RDR negative
AI RDR positive	142 (29.7%)	53 (11.1%)
AI RDR negative	7 (1.5%)	276 (57.7%)

All the seven subjects falsely diagnosed as RDR-negative cases by the AI were moderate NPDR cases. The false-negative rate was 4.70%. No cases of STDR were missed (100% sensitivity). Fig. 2 shows examples of subjects diagnosed correctly and incorrectly by the AI for RDR.

Discussion

This study showed the clinical effectiveness of the Medios AI-DR algorithm in detecting RDR in an Armenian population, reproducing its results from previous validation studies on the Indian population [Table 2]. The study highlighted the generalizability of this offline system in an ethnic population unseen during training. The study revealed excellent sensitivity (95.30%) and acceptable specificity (83.89%) for RDR.

The current study demonstrated a 100% sensitivity for sight-threatening DR. This subgroup entails patients at immediate risk of blindness if left untreated. Hence, this result supports the use of the Medios AI-DR as an aid in the early diagnosis of RDR. Thus, the specialists could focus on the treatment of patients with sight-threatening diseases rather than on screening. This is particularly valuable in regions with low density of ophthalmologists. This conclusion is further supported by a low false negative rate of 4.70% (7 cases of moderate NPDR missed), a parameter critical to denote the safety of the AI system. DR is understood to progress slowly in its early stages. The risk of missing such cases is mitigated by a recommendation for follow-up screenings every year.

The study showed 53 false positives (49 with no DR and 4 with mild NPDR). Furthermore, among the 53 subjects, 17 were graded with AMD by at least one of the graders, while another nine were graded as having another pathology. The accuracy is comparable to our findings on Indian cohorts.[6,7,17]

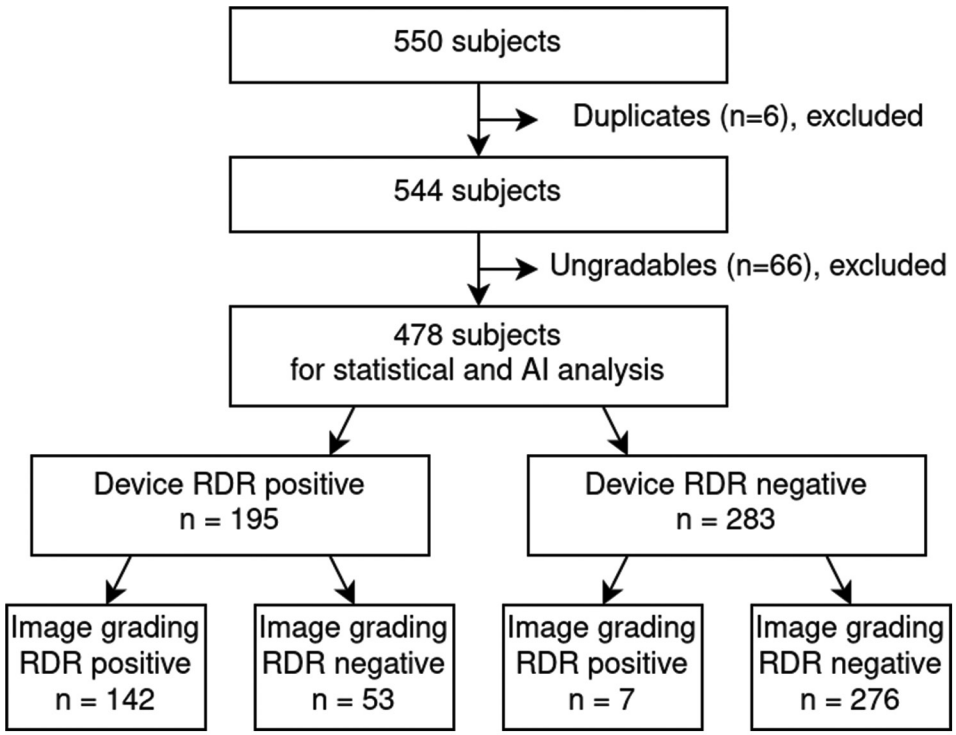


Figure 1: STARD Diagram for Medios AI-DR output for RDR

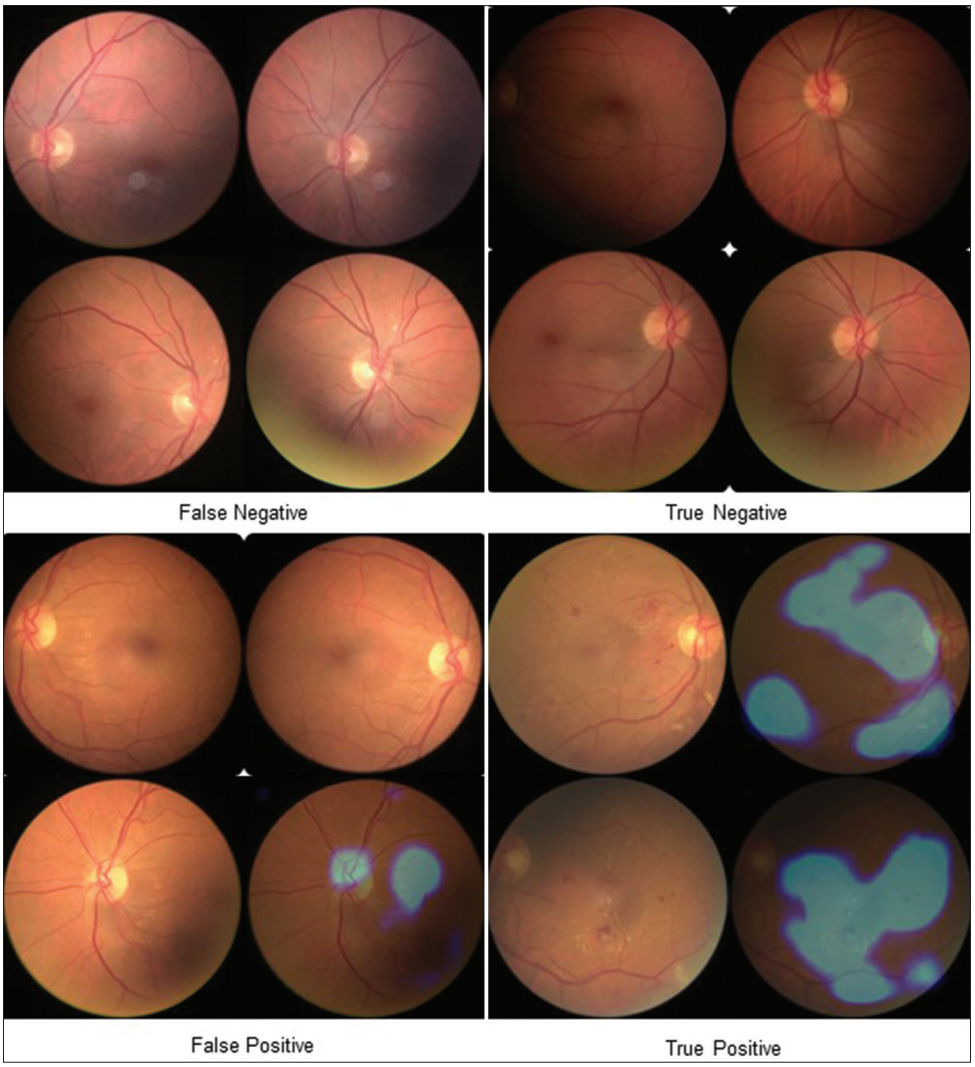


Figure 2: Examples of False Negative, True Negative, False Positive and True Positive patient images analyzed for RDR using Medios AI- DR. Activation maps are shown for images with positive AI diagnosis

Table 1B: Performance of Medios AI-DR on the Armenian population

	RDR	Any DR	STDR
Sensitivity (95% CI)	95.30% (91.9%–98.7%)	91.82% (87.6%–96.1%)	100.00 (1.0–1.0)
Specificity (95% CI)	83.89 (79.9%–87.9%)	84.64% (80.7%–88.6%)	NA
PPV (95% CI)	72.82% (66.6%–79.1%)	74.87% (68.8%–81.0%)	NA
NPV (95% CI)	97.53% (95.7%–99.3%)	95.41% (93.0%–97.8%)	NA

Table 2: Performance of Medios AI-DR in screening for RDR in previous studies

Study	Ethnicity	Sensitivity	Specificity
Natarajan *et al*.[6]	Indian	100.0% (95%CI: 78.2%–100.0%)	88.4% (95% CI: 83.2%–92.5%)
Sosale *et al*.[7] (SMART Study)	Indian	93% (95% CI: 91.3%–94.7%)	92.5% (95% CI: 90.8%–94.2%)
Sosale *et al*.[17]	Indian	98.84% (95% CI: 97.62%–100%)	86.73% (95% CI: 82.87%–90.59%)
Current Study	Armenian	95.30% (91.9%–98.7%)	83.89% (79.9%–87.9%)

We hypothesize that retinal pigmentation did not affect the system's performance. The large and diverse dataset used during development ensured representation across dark to lightly pigmented retinas. Here, the misdiagnosis is rather explainable by the presence of other pathologies being picked up by the AI (26/53, 49%) as referable diseases. We are currently in the process of deploying other disease-specific models such as AMD, which will address this in the near future. Graders,

however, only reported the status of DR. The presence of another retinal disease warranting a referral to a specialist and triggering a positive report by the AI thus cannot be ruled out. The false positive rate is within acceptable limits. The FDA-mandated superiority endpoints of 85% sensitivity and 82.5% specificity for an autonomous RDR AI were surpassed in this study.[14,15]

Previous studies have reported that a lack of ethnic diversity in training data impacted the performance of AI-based systems across ethnicities.[10] Varying concentrations of melanin in different ethnic groups affect the pigmentation of uveal melanocytes as well.[11] A different contrast between fundus and DR lesions due to a varying color of fundi may affect the performance of AI algorithms. AI systems should thus be trained on data from various ethnic groups to reduce bias toward a particular group. This is of particular importance when considering the deployment of systems across many geographies.

The Medios AI-DR had previously been validated in cohorts visiting primary and tertiary care centers in India only. Most studies relied on real-world images captured by minimally trained operators. They showed a sensitivity of 93%–100% and a specificity of 86.7%–92.5% [Table 2].[6,7,17] Though these studies demonstrated the consistency of the algorithm in different clinical settings and the community, these had not assessed its consistency across different ethnicities. The current study complements the earlier one and demonstrates generalizability to a new population, with good sensitivity and specificity. There are some reports of the use of ethnically diverse training datasets on the AI-based detection of RDR. Bellemo *et al*.[9] reported a sensitivity of 92.25% and a specificity of 89.04% in detecting RDR in an African cohort by using an AI algorithm trained with images predominantly from Chinese, Malay, and Indian populations. The system showed consistency and generalizability in detecting RDR in patients with dark fundi of the African population.[9] We are unsure if the good sensitivity and specificity were related to the AI training dataset that also included the Indian eyes. Both Indian and African groups rank higher in melanin synthesis and have nearly similar dark-colored fundi. Similarly, Ting *et al*.[4] reported a sensitivity of 90.5% and a specificity of 91.6% in identifying RDR in a multi-ethnic population. Again, these investigators had included images from Indian, Chinese, and Malay populations. The validation set further comprised African-American, Caucasian, and Hispanic populations with additional multi-ethnic validation datasets. All these studies indicate a common pattern: a sufficiently diverse training set can lead to generalizability beyond the ethnic groups included during training. While this hypothesis is reinforced in this study, we evaluated the AI performance in a different low-resource population that has significant accessibility issues for DR screening. To our knowledge, there is no peer-reviewed published evidence (MEDLINE literature search) on the performance of AI on images captured in a real-world diabetic screening program in Armenia. In addition, on the technical front, the Medios AI-DR is a lightweight ensemble architecture that utilizes a low processor environment of a smartphone-based fundus camera, allowing it to be deployed on the edge on the device. This could overcome barriers of internet connectivity and cloud-based inferencing in limited resource settings.

Strengths: First, the Medios AI-DR algorithm for RDR has been developed with over 50,000 images, which also included 34,278 images from the Kaggle-EyePACS dataset. The Kaggle-EyePACS dataset includes populations of indigenous American, African, European, Asian, and Indian subcontinent descent by design. It does not have any inherent issues with diversity or bias.[18,19] This could partly explain the absence of bias in this algorithm when used on both dark and lightly pigmented fundi. The results of this study further strengthened this statement. Second, this study truly captured real-world data where images were captured by minimally trained operators. Thus, the AI was subject to test on images of par quality unlike the pristine images obtained in the clinic where the performance will be far better.

Limitations: This is a retrospective study with inherent limitations attributable to any retrospective study. A notable limitation was the absence of a "live" AI quality check and feedback to the operator to recapture images. This implied that the study did not follow the exact protocol of two sufficient quality images per eye (one disc and one macula centered) in all patients, as required for optimum AI performance results. Instead, the study used all images captured (up to 7 images per patient) from multiple capturing attempts giving the worst-case scenario results. Inclusion of all available images might have also adversely affected the sensitivity and specificity. In addition, an accurate assessment of the image quality algorithm was not possible. A prospective trial would have been ideal to assess the best performance of the algorithm. The DR AI algorithm and image grading at least relied on two 40° fields of view per eye – one macula and one disc-centered image (covering approximately 60° field of view), potentially overlooking a few instances of severe DR extending beyond the captured area, particularly affecting the nasal region or extending beyond major blood vessel arcades. However, utilizing non-mydriatic single or two-field fundus photography for DR screening aligns with the acceptable practices recommended by the International Council of Ophthalmology as well as the American Academy of Ophthalmology Guidelines.[20,21]

Conclusion

There is evidence that a training set biased against a specific ethnic group does not generalize well beyond that group.[10] However, others and we have shown that an AI algorithm trained with an ethnically diverse dataset overcomes this deficiency.[4,9] In addition, we feel that additional investigations are needed to evaluate variations of fundus pigmentation across ethnicities and understand the similarities and differences to develop robust AI solutions by using retinal images.

Acknowledgement

The authors thank the AECP personnel and patients who contributed to this study. The authors acknowledge Dr. Usha Sharma, Consultant Retina Specialist, and Sheetal Panicker, Optometrist for providing image grading assistance.

Financial support and sponsorship: Nil.

Conflicts of interest: Divya Parthasarathy Rao, Florian M. Savoy, Anand Sivaraman, Sreetama Dutt are employees of Remidio Innovative Solutions Pvt. Ltd. All the other authors declare no known financial interests.

References

1. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, *et al*. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017;128:40-50.
2. Thomas RL, Halim S, Gurudas S, Sivaprasad S, Owens DR. IDF Diabetes Atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018. Diabetes Res Clin Pract 2019;157:107840.
3. 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;3:294-304.
4. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, *et al*. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017;318:2211-23.
5. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond) 2018;32:1138-44.
6. 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;137:1182-8.
7. Sosale B, Aravind SR, Murthy H, Narayana S, Sharma U, Gowda SGV, *et al*. Simple, mobile-based artificial intelligence algorithm in the detection of diabetic retinopathy (SMART) study. BMJ Open Diabetes Res Care 2020;8:e000892.
8. Yip MYT, Lim G, Lim ZW, Nguyen QD, Chong CCY, Yu M, *et al*. Technical and imaging factors influencing performance of deep learning systems for diabetic retinopathy. NPJ Digit Med 2020;3:40.
9. Bellemo V, Lim ZW, Lim G, Nguyen QD, Xie Y, Yip MYT, *et al*. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: A clinical validation study. Lancet Digit Health 2019;1:e35-44.
10. Burlina P, Joshi N, Paul W, Pacheco KD, Bressler NM. Addressing artificial intelligence bias in retinal diagnostics. Transl Vis Sci Technol 2021;10:1-13.
11. Hu DN, Savage HE, Roberts JE. Uveal melanocytes, ocular pigment epithelium, and Müller cells in culture: *In vitro* toxicology. Int J Toxicol 2002;21:465-72.
12. Armenian EyeCare Project Diabetes Program - Armenian EyeCare Project. Available from: https://eyecareproject.com/diabetes-program/. [Last accessed on 2022 Mar 08].
13. Diabetic eye screening: Professional guidance - GOV.UK. Available from: https://www.gov.uk/government/collections/diabetic-eye-screening-commission-and-provide. [Last accessed on 2022 Mar 08].
14. 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;1:39.
15. Ipp E, Liljenquist D, Bode B, Shah VN, Silverstein S, Regillo CD, *et al*. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. JAMA Netw Open 2021;4:e2134254.
16. Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003-2016. Acta Diabetol 2017;54:515-25.
17. Sosale B, Sosale AR, Murthy H, Sengupta S, Naveenam M. Medios- An offline, smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. Indian J Ophthalmol 2020;68:391-5.
18. Cuadros J, Bresnick G. EyePACS: An adaptable telemedicine system for diabetic retinopathy screening. J Diabetes Sci Technol 2009;3:509-16.
19. 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;14:2021-35.
20. Williams GA, Scott IU, Haller JA, Maguire AM, Marcus D, McDonald HR. Single-field fundus photography for diabetic retinopathy screening: A report by the American Academy of Ophthalmology. Ophthalmology 2004;111:1055-62.
21. Wong TY, Sun J, Kawasaki R, Ruamviboonsuk P, Gupta N, Lansingh VC, *et al*. Guidelines on Diabetic Eye Care: The International Council of ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings. Ophthalmology 2018;125:1608-22.

BMJ Open Ophthalmology

Real-world evaluation of smartphone-based artificial intelligence to screen for diabetic retinopathy in Dominica: a clinical validation study

Oliver Kemp,[1] Covadonga Bascaran ,[1] Edyta Cartwright,[2] Lauren McQuillan,[2] Nanda Matthew,[3] Hazel Shillingford-Ricketts,[3] Marcia Zondervan,[1] Allen Foster,[1] Matthew Burton[1,4]

To cite: Kemp O, Bascaran C, Cartwright E, *et al*. Real-world evaluation of smartphone-based artificial intelligence to screen for diabetic retinopathy in Dominica: a clinical validation study. *BMJ Open Ophthalmology* 2023;**8**:e001491. doi:10.1136/bmjophth-2023-001491

OK and CB are joint first authors.

Received 11 September 2023
Accepted 10 December 2023

[1]London School of Hygiene and Tropical Medicine, London, UK
[2]University Hospitals Sussex NHS Foundation Trust, Worthing, UK
[3]Dominica China Friendship Hospital, Roseau, Dominica
[4]Moorfields Eye Hospital NHS Foundation Trust, London, UK

Correspondence to
Dr Covadonga Bascaran; covadonga.bascaran@lshtm.ac.uk

ABSTRACT

Objective Several artificial intelligence (AI) systems for diabetic retinopathy screening have been validated but there is limited evidence on their performance in real-world settings. This study aimed to assess the performance of an AI software deployed within the diabetic retinopathy screening programme in Dominica.

Methods and analysis We conducted a prospective, cross-sectional clinical validation study. Patients with diabetes aged 18 years and above attending the diabetic retinopathy screening in primary care facilities in Dominica from 5 June to 3 July 2021 were enrolled.
Grading was done at the point of care by the field grader, followed by counselling and referral to the eye clinic. Images were then graded by an AI system. Sensitivity, specificity with 95% CIs and area under the curve (AUC) were calculated for comparing the AI to field grader as gold standard.

Results A total of 587 participants were screened. The AI had a sensitivity and specificity for detecting referable diabetic retinopathy of 77.5% and 91.5% compared with the grader, for all participants, including ungradable images. The AUC was 0.8455. Excluding 52 participants deemed ungradable by the grader, the AI had a sensitivity and specificity of 81.4% and 91.5%, with an AUC of 0.9648.

Conclusion This study provides evidence that AI has the potential to be deployed to assist a diabetic screening programme in a middle-income real-world setting and perform with reasonable accuracy compared with a specialist grader.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Many diabetic retinopathy (DR) algorithms have been shown to perform with high accuracy when compared with human grading, but limited evidence has been published on real-world validation of artificial intelligence (AI) for DR.

WHAT THIS STUDY ADDS

⇒ The study reports on the performance of AI for DR when deployed in real-world conditions in an existing DR programme in a middle-income setting.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ At national level in Dominica, this study will inform policy and practice in service delivery for DR services. Globally, this study builds on the evidence in application of AI in real-world settings.

INTRODUCTION

Diabetic retinopathy (DR) is the most common microvascular complication of diabetes mellitus. It is a major cause of vision impairment and blindness.[1] Retinal screening and referral for treatment for those identified having DR can prevent vision loss.[2–5] For this reason, many countries are introducing DR screening and treatment programmes.[6–8]

A recent systematic review of DR screening found that in low-income and middle-income countries (LMIC), common barriers include limited skilled human resources and lack of access to eye facilities.[9] Use of artificial intelligence (AI) for grading of retinal images could help to reduce the time spent by ophthalmic specialists reviewing images.[10 11] AI in DR screening can allow quick assessment of a large number of images and communication of the decision to refer, or not, to the patients at the point of care, and in the last few years these technologies have started to be validated.[12–14] As the quality of smartphone cameras improves, there has been investment and research into their use as portable retinal cameras, offering a lower cost and transportable option in low resource and rural settings.[15]

Four recent meta-analyses reported sensitivities for AI to grade DR between 87% and 97%.[16–19] Most studies reported AI systems which used datasets from high-quality images taken with state-of-the-art retinal cameras in eye clinic settings. Some studies, including a

large-scale real-world use of AI in Thailand, have assessed community screening in field settings, reporting sensitivities between 84% and 91% for referable DR and 91% for vision threatening DR.[20–22]

The prevalence of diabetes in the adult population in Dominica is estimated to be 17.7%.[23] Dominica has been screening for DR since 2005, but its programme coverage is limited with approximately 1500 of the estimated 7000 adults living with diabetes being screened each year. There are two employed ophthalmic technicians in the public sector in Dominica certified to grade retinal images, but their time to screen DR is limited by other clinical duties. There are two retinal cameras, one fixed (Centervue DRS) in a hospital in Roseau, the capital, and a smartphone camera (Remidio) used in a mobile clinic that visits rural districts. The ophthalmology services in Dominica are equipped to deliver treatment to patients with vision threatening DR.

AI-assisted grading in the mobile clinic could help overcome human resources constraints and increase DR screening coverage. There is an AI software application that can be used offline with the smartphone-based 'Fundus on Phone' retinal camera currently used in Dominica.[24] Studies in India using this AI software and camera have reported a sensitivity of 83% to detect any DR, and a sensitivity of 93% to detect 'referable' DR.[25–27]

This study aimed to evaluate the diagnostic accuracy of Medios AI software for the diagnosis of referable diabetic retinopathy (RDR) using mydriatic retinal images when deployed and integrated in a real-world DR screening programme in a Caribbean population in Dominica.

MATERIALS AND METHODS

Study design

This prospective, cross-sectional clinical validation study was conducted to assess the performance of an AI software application in identifying referable DR, compared with a human grader (reference standard). The technology we tested was Medios DR AI software (NM App V.2.0, Mediostech, Singapore) hereafter referred to as 'AI system', incorporated into a Non-Mydriatic Fundus on Phone Camera, Model FOPNM-10, (Remidio Innovative Solutions, Bangalore, India). This AI system is Conformitè Europëenne marked and was chosen as it was compatible with the camera routinely used in the mobile programme.

The reference standard was the image grading performed in the field by the senior Dominican screener–grader, holder of a Certificate of Higher Education in DR Screening, Gloucester Retinal Education Group, University of Gloucestershire, UK (hereafter referred to as field grader).[28] The grading by the field grader was compared with remote grading by senior graders in the English National Screening Programme, and the interobserver reliability kappa coefficient was calculated.[29]

Participants and setting

A consecutive sample of patients with diabetes over the age of 18 years attending the mobile DR screening clinic in Dominica from 5 June to 3 July 2021 was enrolled in the study. Screening was conducted in primary care health facilities in four health districts. Informed consent was obtained from all participants. There was no change to normal practice in the screening programme clinical pathway.[30]

Image acquisition and grading

Following the local protocol, the pupils of patients were dilated (tropicamide 0.5% and phenylephrine HCL 5%). A minimum of one image centred on the optic disc and one image centred on the macula were taken of each eye using the hand-held camera by the field grader. The field grader performed DR grading and decided to refer or not based on the grading. Patients received the usual standard of care, which includes counselling on diabetes control and referral to the eye clinic.

Although the AI system can work offline and therefore potentially provide a point of care decision, in this validation, study AI grading was deferred to the end of the study to ensure that any AI output did not influence grading and clinical decisions about referral.[27]

Analysis

RDR was defined as moderate non-proliferative diabetic retinopathy or worse, or diabetic macular oedema, or ungradable image in either eye. Sensitivity, specificity with 95% CIs and area under the curve (AUC) were calculated for RDR comparing the AI system to field grader as gold standard. Vision-threatening diabetic retinopathy (VTDR) was defined as the presence of proliferative diabetic retinopathy and/or diabetic macular oedema in either eye. Data were collected using electronic tablets and later converted into Excel and analysed using Excel and Stata X software.

AI and human grading

The AI system is based on convolutional neural networks and its functionality has been described in detail elsewhere.[27] The AI provides a binary output of 'signs of DR detected' or 'signs of DR not detected' with a threshold of 'moderate non-proliferate DR' and above, according to the International Classification of Diabetic Retinopathy (ICDR).[31]

The field grader has been trained on, and uses, the English Grading System for DR.[6] This system does not correspond directly with the ICDR. The lower grade of DR, referred to as R1 in the English system is equivalent to both 'mild and moderate non-proliferative DR' in the ICDR. To allow comparability in the study, we asked the field grader to record retinal DR features in all mild and moderate cases and subsequently classified images accordingly.

Ungradable images

We defined ungradable images as those reported as such by the field grader. The AI system does not report an

ungradable category, rather it performs a quality assessment for each image and notifies the user if the image is low quality and prompts a recapture of the image.[27] This gives the technician the chance to retake the image until the AI quality threshold is achieved. This functionality was not used in the study, as we did not use the AI in the field to avoid introducing bias with the field grader. As the AI system actually produces a grade output for every image, regardless of the quality, we obtained AI grades for all images in this study, but in the analysis excluded AI reports for patients which the field grader reported as both eyes being ungradable.

Sample size

Based on previous validation studies, we assumed that the AI system would have an estimated sensitivity of 93% and a specificity of 89% for detecting moderate non-proliferative DR or worse, the threshold used in our definition of referable DR.[25–27] We also estimated that 3 in every 10 patients screened in the programme require referral to the diabetic eye clinic based on previous Dominica data; this is consistent with the expected prevalence of DR in people with diabetes.[32] Our sample calculations, with a margin of error of 5%, gave for sensitivity sN=333 and for specificity spN=461. We took the largest estimate and added 46 participants to account for an estimated 10% ungradable cases leading to a total minimum sample of n=507.[33]

Table 1 Participant characteristics (n=587)

Age (years)	Mean (SD)	64 (12.3)
	Range	26–94
Gender	Women	426 (72.6%)
	Men	161 (27.4%)
Ethnicity	Black Caribbean	570 (97.1%)
	Carib	17 (2.9%)
Years lived with diabetes*	Mean (SD)	12 (8.8)
	Range	0–49
Methods of diabetes control	Diet and exercise only	5 (0.9%)
	Tablet medication	517 (88.1%)
	Insulin	100 (17.0%)
	Insulin and tablet	54 (9.2%)
Type of diabetes	Type 1	6 (1.0%)
	Type 2	581 (99.0%)
Field grader DR grading	RDR	267 (45.4%)
	VTDR	111 (18.9%)
	One eye ungradable	20 (3.4%)
	Both eyes ungradable	52 (8.8%)

*n=549, some missing data for years lived with diabetes.
DR, diabetic retinopathy; RDR, referable diabetic retinopathy; VTDR, vision-threatening diabetic retinopathy.

RESULTS

Our study included 587 participants, with a mean age of 64 years (range 26–94); 426 (72.6%) were women (table 1). The predominant ethnicity was black Caribbean (570, 97.1%). A total of 2327 images were obtained from these 587 participants. The field grader classified 72 participants in the study as having ungradable images in at least one eye (72/587, 12.2%), of which 52 had ungradable images in both eyes (52/587, 8.8%). The interobserver agreement between the field and remote image graders for detecting any DR was K=0.69 (good agreement 0.61–0.80).

The prevalence of RDR (moderate non-proliferative diabetic retinopathy or worse or diabetic macular oedema), including all participants (n=587), was 45.4% (95% CI, 41.5% to 49.5%) by the field grader and 39.8% (95% CI, 35.9% to 43.8%) by the AI system. The prevalence of RDR in the sample, excluding the ungradable participants (n=535), was 40.1% (95% CI, 36.0% to 44.3%) by the field grader and 37.7%% (95% CI, 33.6% to 41.9%) by the AI system.

For all participants, including ungradable images, the AI system had a sensitivity of 77.5% and specificity of 91.5% compared with the field grader for detecting RDR. The AUC was 0.84 (table 2).

Excluding the 52 participants deemed ungradable by the field grader resulted in the AI system having a sensitivity of 81.4% and a specificity of 91.5%, with an AUC of 0.96, for detecting RDR (table 3).

The analysis comparing the remote graders with the AI, excluding 65 participants deemed ungradable by the remote graders resulted in a sensitivity, specificity of 83.7% and 83.7% and AUC of 0.86 (table 4).

The prevalence of VTDR, (proliferative diabetic retinopathy and/or diabetic macular oedema) by the field grader in the entire sample was 18.9% (95% CI 15.7% to 22.1%) and excluding ungradable participants (n=52) it was 20.7% (95% CI 17.3% to 24.2%). In the sample excluding ungradable participants, the AI system had a sensitivity of 89.2% (95% CI 82.8% to 95.2%) for detecting the presence of VTDR (which it classified as 'signs of DR detected'). The specificity of detecting VTDR could not be calculated as the AI system only gives a binary output for DR. There were 12 participants identified as having VTDR by the field grader, but not identified by the AI system. None of the 12 had proliferative diabetic retinopathy, all were graded as having diabetic maculopathy by the field grader. On further scrutiny of these 12 images, 7 had other macular pathology, which resulted in the field grader referring. If these were excluded from the analysis, the sensitivity of the AI increases to 95.2% (95% CI, 90.7% to 99.3%).

DISCUSSION

A good screening test for diabetic retinopathy should ideally have a sensitivity higher than 80% and a specificity higher than 95%.[6 34] Our study demonstrated a sensitivity and specificity for the AI system of 77.5% and 91.5%

Table 2 Grading comparison between AI system and field grader, including ungradable participants

Field grader					
AI system		Not referable	Referable	Total	
	Not referable	293	60	353	
	Referable	27	207	234	
	Total	320	267	587	
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC
Referable or not	77.5% (72.0% to 82.3%)	91.5% (87.9% to 94.3%)	88.4% (84.1% to 91.7%)	83.0% (82.0% to 87.9%)	0.84

AI, artificial intelligence; AUC, area under the curve; NPV, Negative Predictive Value; PPV, Positive Predictive Value.

when ungradable participants were included, and 80.4% and 91.5% when participants deemed ungradable by the field grader were excluded.

The analysis excluding ungradable participants probably gives the more reliable indication of the actual performance of the AI algorithm compared with the field grader. The AI system when used in the field prompts for a repeat image if the quality is low. To avoid bias, we could not use this feature during the study and therefore we run the AI in all images irrespective of quality.

At programme level however, it is important to consider all ungradable images as by definition those patients will need to be examined by an ophthalmologist and may have corneal pathology or cataract which results in poor retinal images.

The prevalence of DR (moderate non-proliferative diabetic retinopathy or worse or diabetic macular oedema) among our study participants was 40.1% (field grader) and 37.7% (AI system). This is similar to the estimated prevalence of DR for North America and the Caribbean region of 38.1%.[32] The regional estimates indicate 7.8% of people with diabetes have VTDR and are therefore at risk of vision loss if not treated. In our study participants, the prevalence of VTDR was 20.7%, significantly higher than the current regional estimates. The mean years living with diabetes in the study sample is quite high (12 years) and this may differ from the population-based studies included in regional estimates. Another explanation is that the higher prevalence found may indicate late diagnosis or poor diabetes control. Also, the prevalence of obesity and hypertension in Dominica is high, possibly compounding the higher progression to VTDR of our study population.[23]

This study was conducted in a real-world outreach mobile programme. The sensitivity values are below those previously reported in the literature for Medios AI (93%–100%).[25–27] A recent review of AI software used for DR screening found sensitivities ranging from 86% to 100% for detecting 'referable DR', with most of these using the same definition for referable DR as our study.[10] It is important to point out that, although the study was not powered to detect VTDR, there were 12 cases where the grader classified patients as VTDR, due to suspected maculopathy, that were not identified by the AI system, giving a sensitivity for VTDR of 89%. This reflects the fact that field graders in real-world programmes make decisions on referral of other pathology that they find while screening. In this case, seven participants had non-DR macular signs that prompted referral which the AI is not trained to pick up. An adequately powered large scale field validation of AI in Thailand achieved a sensitivity for identifying VTDR of 91.4% and reported that most of the discrepancies were related to the grading of diabetic maculopathy.[22] When we remove the seven referrals with non-DR macular changes from the analysis, the sensitivity of the AI for VTDR increases to 95.2%.

Table 3 Grading comparison between AI system and field grader, excluding ungradable participants (n=52)

Field grader					
AI system		Not referable	Referable	Total	
	Not referable	293	40	333	
	Referable	27	175	202	
	Total	320	315	535	
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC
Referable or not	80.4% (75.5% to 86.3%)	91.5% (87.9% to 94.3%)	86.6% (81.7% to 90.3%)	87.9% (84.6% to 90.6%)	0.96

AI, artificial intelligence; AUC, area under the curve; NPV, Negative Predictive Value; PPV, Positive Predictive Value.

Table 4 Grading comparison between AI system and remote grader, excluding ungradable participants (n=64)

Remote grader					
AI system		Not referable	Referable	Total	
	Not referable	324	22	346	
	Referable	63	113	176	
	Total	387	135	522	
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC
Referable or not	83.7% (75.6% to 90.4%)	83.7% (75.6% to 90.4%)	93.6% (88% to 98%)	64.2% (54.5% to 73.5%)	0.86

AI, artificial intelligence; AUC, area under the curve; NPV, Negative Predictive Value; PPV, Positive Predictive Value.

The balance of sensitivity and specificity is very relevant at programme level. A low specificity would imply too many patients being unnecessarily referred to the eye clinic, overloading the services. The specificity of the AI system in our study was quite high, which suggests the appropriateness of the referrals made. The programme guidelines in Dominica have a low threshold for referral, with mild forms of DR being referred to the eye clinic. This is because there is no robust system for annual recall of diabetic patients for an eye examination. Referring less severe cases of DR gives an opportunity for patient education about diabetes and hypertension control and ensures the patients are registered a in the eye clinic which facilitates regular review. The threshold for referral varies from country to country and is determined by local guidelines for DR management.[35–38] With the current programme referral thresholds, the AI system resulted in a postive predictive value (PPV) of 88.4% and 85.4% (including and excluding ungradable images in the analysis).

Our study had a women-to-men ratio of 3.5:1. Although it is reported that women are more likely to have diabetes than men in Caribbean populations, the WHO STEPwise approach to surveillance survey (STEPS) data for Dominica in 2008 showed a higher prevalence of diabetes in men.[23 39] It is plausible that this has changed in the last decade in Dominica. An alternative explanation is that women may be accessing diabetes services more than men and are therefore overrepresented in the DR screening programme. If this is the case, it will be important to explore the reasons for the lower uptake of screening by men and implement strategies to improve it.

This study reports the performance of an AI system fully integrated in a functioning DR screening programme in an LMIC. It provides evidence that an AI system with off-line capabilities has the potential to be deployed in a mobile community DR screening programme and perform with reasonable accuracy compared with a trained specialist grader. In order to leverage the contribution of AI technology to improve DR screening coverage and address the specialised human resource constraints, it is recommended as a next step to research the performance of the smartphone camera and AI system in the hands of trained community nurses.

Acknowledgements Remidio/Medios: Remidio provided technical support under the framework of their existing goods and services relationship with Dominica ophthalmology services. Remidio/Medios had no role in study design, data collection, data analysis, data interpretation or writing or reviewing the report. Open Solutions for Health, a local company in Dominica, provided support building a software for the DR screening programme in which we could embed the study data collection.

Contributors CB and OK are responsible for the overall content of the article. CB, HS-R and MZ conceived the study idea. OK and CB designed the study. OK, HS-R, NM, EC and LM contributed to data collection. CB and OK conducted data analysis. CB wrote the manuscript. All authors reviewed and commented on the manuscript. CB is guarantor of the work.

Funding The Dominica MoH and department of ophthalmology provided funding support for the project. Consumables and staff incentives were funded by the VISION 2020 LINKS Programme. LSHTM provided funding for travel and field work. MB is supported by the Wellcome Trust (207472/Z/17/Z).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval This study involves human participants. The study was approved by the London School of Hygiene & Tropical Medicine Ethics Committee and the Dominica Ministry of Health Ethics Committee. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iD

Covadonga Bascaran http://orcid.org/0000-0002-5662-3325

REFERENCES

1 Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of Avoidable blindness in relation to VISION 2020: the right to sight: an analysis for the global burden of disease study. *Lancet Glob Health* 2021;9:e144–60.
2 Photocoagulation treatment of proliferative diabetic retinopathy. clinical application of diabetic retinopathy study (DRS) findings, DRS report number 8. The diabetic retinopathy study research group. *Ophthalmology* 1981;88:583–600.

3 Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness Certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. *BMJ Open* 2014;4:e004015.
4 Early Photocoagulation for diabetic retinopathy. ETDRS report number 9. early treatment diabetic retinopathy study research group. *Ophthalmology* 1991;98(5 Suppl):766–85. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=2062512 [Accessed 6 Nov 2023].
5 Bäcklund LB, Algvere PV, Rosenqvist U. New blindness in diabetes reduced by more than one-third in Stockholm County. *Diabet Med* 1997;14:732–40.
6 Scanlon PH. The English national screening programme for diabetic retinopathy 2003-2016. *Acta Diabetol* 2017;54:515–25.
7 Astbury N, BurgessP FA, *et al*. *Takling diabetic retinopathy globally through the VISION 2020*. LINKS Diabetic Retinopathy Network.Eye News.2017;23(5),
8 Kristinsson JK, Hauksdóttir H, Stefánsson E, *et al*. Active prevention in diabetic eye disease. A 4-year follow-up. *Acta Ophthalmol Scand* 1997;75:249–54.
9 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.
10 Grzybowski A, Brona P, Lim G, *et al*. Correction to: artificial intelligence for diabetic retinopathy screening: a review. *Eye (Lond)* 2020;34:604.
11 Bastawrous A, Hennig BD. The global inverse care law: a distorted map of blindness. *Br J Ophthalmol* 2012;96:1357–8.
12 Tufail A, Kapetanakis VV, Salas-Vega S, *et al*. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess* 2016;20:1–72.
13 Heydon P, Egan C, Bolter L, *et al*. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol* 2021;105:723–8.
14 Bhaskaranand M, Ramachandra C, Bhat S, *et al*. The value of automated diabetic retinopathy screening with the Eyeart system: A study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technol Ther* 2019;21:635–43.
15 Rajalakshmi R, Prathiba V, Arulmalar S, *et al*. Review of retinal cameras for global coverage of diabetic retinopathy screening. *Eye (Lond)* 2021;35:162–72.
16 Wewetzer L, Held LA, Steinhäuser J. Diagnostic performance of deep-learning-based screening methods for diabetic retinopathy in primary care-A meta-analysis. *PLoS One* 2021;16:e0255034.
17 Wu J-H, Liu TYA, Hsu W-T, *et al*. Performance and limitation of machine learning Algorithms for diabetic retinopathy screening: meta-analysis. *J Med Internet Res* 2021;23:e23863.
18 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;183:41–9.
19 Islam MM, Yang H-C, Poly TN, *et al*. Deep learning Algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. *Computer Methods and Programs in Biomedicine* 2020;191:105320.
20 Barriga ES, Dewi ER, Baldivieso O, *et al*. Using a Handheld retinal camera and artificial intelligence for diabetic retinopathy screening in Bolivia. *Investigative Ophthalmology & Visual Science* 2020;61:1645.
21 Ming S, Xie K, Lei X, *et al*. Evaluation of a novel artificial intelligence-based screening system for diabetic retinopathy in community of China: a real-world study. *Int Ophthalmol* 2021;41:1291–9.
22 Ruamviboonsuk P, Tiwari R, Sayres R, *et al*. Real-time diabetic retinopathy screening by deep learning in a Multisite national screening programme: a prospective Interventional cohort study. *Lancet Digit Health* 2022;4:e235–44.
23 Ricketts P. Dominica STEPS Survey, . 2008Available: https://cdn.who.int/media/docs/default-source/ncds/ncd-surveillance/data-reporting/dominica/steps/dominica_2008_steps_factsheet.pdf?sfvrsn=b1ee05e4_5&download=true [Accessed 6 Nov 2023].
24 Remidio. Fundus on Phone FOP NM-10 - Company Advert, . 2021Available: https://www.remidio.com/products/fop [Accessed 6 Nov 2023].
25 Sosale B, Aravind SR, Murthy H, *et al*. Mobile-based artificial intelligence algorithm in the detection of diabetic retinopathy (SMART) study. *BMJ Open Diabetes Res Care* 2020;8:e000892.
26 Natarajan S, Jain A, Krishnan R, *et al*. Diagnostic accuracy of community-based diabetic retinopathy screening with an Offline artificial intelligence system on a Smartphone. *JAMA Ophthalmol* 2019;137:1182–8.
27 Sosale B, Sosale AR, Murthy H, *et al*. Medios- an Offline, Smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. *Indian J Ophthalmol* 2020;68:391–5.
28 GREG. Certificate of Higher Educaiton in Diabetic Retinopathy Screening, Available: https://www.gregcourses.com/certificate-of-higher-education-in-diabetic-retinopathy-screening [Accessed 6 Nov 2023].
29 Altman DG. Practical statistics for medical research. In: *Practical Statistics for Medical Research*. Chapman & Hall, 1991.
30 Matthew N. Running a mobile diabetes screening service in Dominica. *Community Eye Health* 2020;33:51–2.
31 Wilkinson CP, Ferris FL III, Klein RE, *et al*. Proposed International clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677–82.
32 Teo ZL, Tham Y-C, Yu M, *et al*. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology* 2021;128:1580–91.
33 Buderer NM. Statistical methodology: I. incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996;3:895–900.
34 Vujosevic S, Aldington SJ, Silva P, *et al*. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol* 2020;8:337–47.
35 Republic of Kenya Ministry of Health. Guidelines for the screening and managment of diabetic retinopathy screening in Kenya, . 2017Available: https://osk.or.ke/downloads/ [Accessed 6 Nov 2023].
36 Goverment of the Republic of Malawi. Ministry of Health. National Guidelines for Diabetic Eye Care, . 2021
37 Healthcare Improvement Scotland. Diabetic retinopathy screening standards. 2016.
38 Public health England. Diabetic eye screening programme: standards, . 2019
39 Sobers-Grannum N, Murphy MM, Nielsen A, *et al*. Female gender is a social determinant of diabetes in the Caribbean: a systematic review and meta-analysis. *PLoS ONE* 2015;10:e0126799.

Clinical Ophthalmology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

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

Divya Parthasarathy Rao[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

Received: 21 April 2022
Accepted: 11 July 2022
Published: 17 August 2022

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.

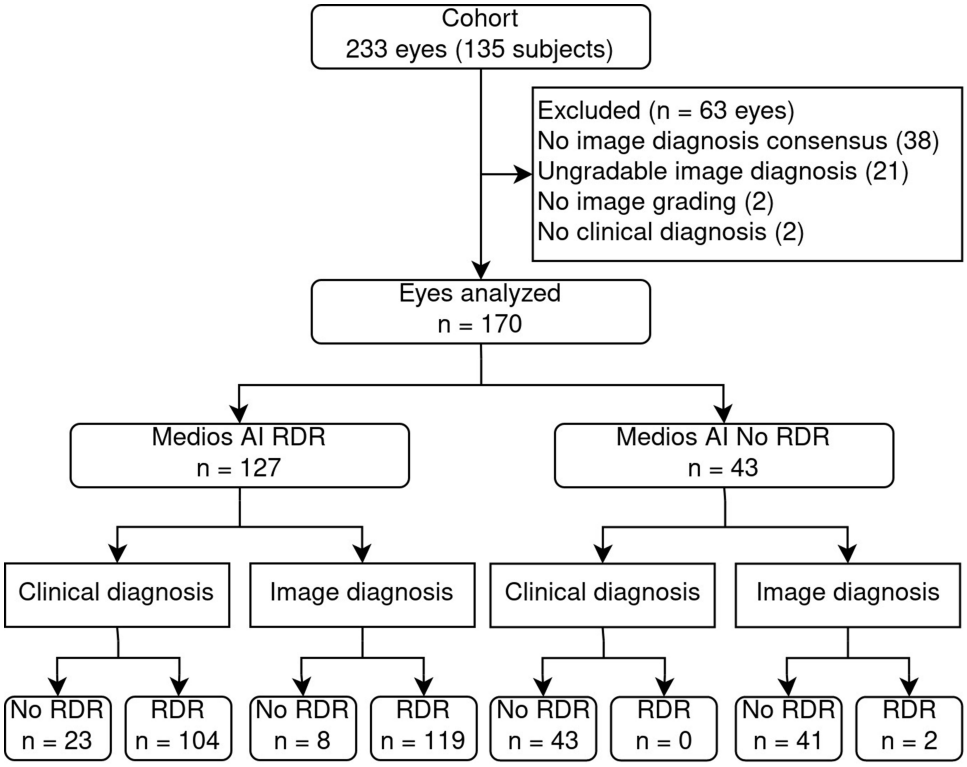


Figure 1 STARD flowchart: AI output for RDR against clinical assessment and image-based grading.

Intergrader Reliability

The intergrader reliability (weighted kappa, ķ) 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.

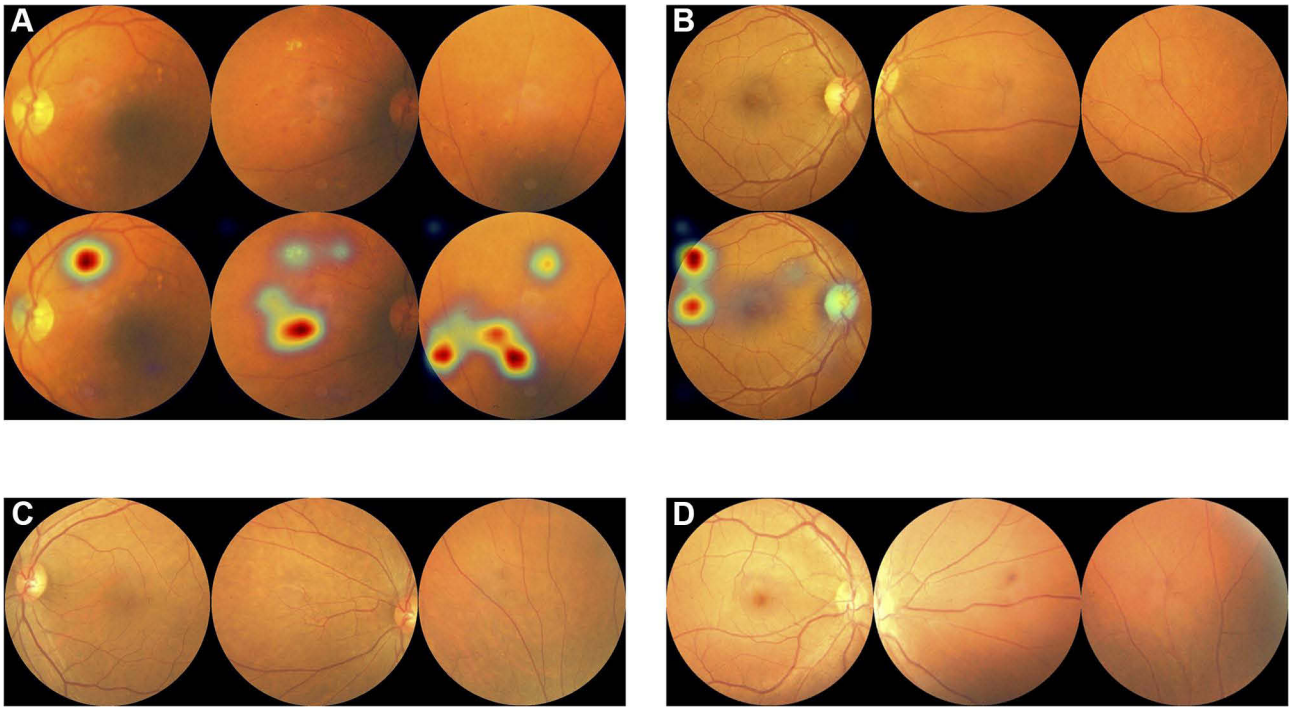


Figure 2 Images of true positive (**A**), false positive (**B**), false negative (**C**) and true negative (**D**) subject with activation maps for image triggering positive diagnosis.

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

Ophthalmic Research

Ophthalmic Res , DOI: 10.1159/000534098
Received: February 17, 2023
Accepted: September 8, 2023
Published online: September 27, 2023

Diagnostic accuracy of Automated Diabetic Retinopathy Image Assessment Softwares: IDx-DR and MediosAI

Grzybowski A, Rao DP, Brona P, Negiloni K, Krzywicki T, Savoy FM

ISSN: 0030-3747 (Print), eISSN: 1423-0259 (Online)
https://www.karger.com/ORE
Ophthalmic Research

Research Article

Diagnostic accuracy of Automated Diabetic Retinopathy Image Assessment Softwares: IDx-DR and MediosAI

Andrzej Grzybowski[1], Divya Parthasarathy Rao[2], Piotr Brona[3], Kalpa Negiloni[4], Tomasz Krzywicki[5], Florian M Savoy[6]
[1]Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland
[2]Department of AI R & D, Remidio Inc, Glen Allen, Virginia, USA
[3]Department of Ophthalmology, Poznan City Hospital, Szwajcarska 3, Poznan 60-285, Poland
[4]Department of Clinical Research, Remidio Innovative Solutions Pvt Ltd, Bangalore, India
[5]Department of Mathematical Methods of Informatics, University of Warmia and Mazury, Olsztyn, Poland
[6]Department of AI R & D, Medios Technologies, Remidio Innovative Solutions, Singapore

Short Title: Performance of two DR screening AI softwares

Corresponding author:
Divya Parthasarathy Rao
Department of AI R & D, Remidio Innovative Solutions Inc.,
11357 Nuckols Rd, #102, Glen Allen, Virginia, 23059, USA
Tel +1 (855) 513-3335
Email drdivya@remidio.com

Number of Tables: 2
Number of Figures: 1
Word count: 3018
Keywords: Diabetic Retinopathy, AI, Medios, IDx-DR

Abstract

Introduction: Numerous studies have demonstrated the use of Artificial Intelligence for early detection of referable diabetic retinopathy (RDR). A direct comparison of these multiple Automated DR Image Assessment Softwares(ARIA) is however challenging. We retrospectively compared the performance of two modern ARIAs, IDx-DR and Medios AI.

Methods: In this retrospective-comparative study, retinal images with sufficient image quality were run on both ARIAs. They were captured in 811 consecutive patients with Diabetes visiting diabetic clinics in Poland. For each patient, four non-mydriatic images, 45-degree field of view i.e two sets of one optic disc and one macula-centered image using Topcon NW400 were captured. Images were manually graded for severity of DR as no DR, any DR (mild NPDR or more severe disease), RDR (moderate NPDR or more severe disease and/or clinically significant diabetic macular edema (CSDME)) or sight-threatening DR (severe NPDR or more severe disease and/or CSDME) by certified graders. The ARIAs output was compared to manual consensus image grading (reference standard).

Results: On 807 patients, based on consensus grading, there was no evidence of DR in 543 patients (67). Any DR was seen in 264 (33%) patients, of which 174 (22%) were referable DR, and 41 (5%) sight-threatening DR. The sensitivity

of detecting RDR against reference standard grading was 95% (95%CI 91, 98%) and the specificity was 80% (95%CI 77, 83%) for Medios AI. They were 99% (95%CI 96, 100%) and 68% (95%CI 64, 72%) for IDx-DR respectively.
Conclusion: Both the ARIAs achieved satisfactory accuracy, with few false negatives. Although false-positive results generate additional costs and workload, missed cases raise the most concern whenever automated screening is debated.

Introduction

Diabetes is a global epidemic and one of the world's fastest-growing diseases. The number of patients with diabetic retinopathy (DR) and sight-threatening DR is also expected to rise. There are only a few established nationwide DR screening programmes and overall DR screening services remain inadequate in most of the developing world and even some developed countries [1]. This is further compounded by the increasing resources needed for the implementation and maintenance of comprehensive DR screening programs [2].

One of the proposed solutions to this global issue is the use of automated diabetic retinopathy image assessment software (ARIA) to grade fundus images instead or alongside human graders. There are multiple ARIAs currently available with many more being developed worldwide [1]. Although there is an abundance of studies looking into the performance of a single ARIA, studies comparing multiple ARIAs are currently rare, as direct comparison is often difficult [3]. Based on previous studies it is clear that the performance of even state-of-the-art algorithms may vary considerably [3,4]. We set out to analyze the performance of two modern ARIAs, IDx-DR and MediosAI.

Materials and Methods

Study Design: In this retrospective comparative study, the performance of two different ARIAs in screening for DR were compared to human graders (reference standard). The screening for DR was conducted and retinal images were obtained from diabetic clinics in Poznan, Poland between March 2020 and April 2021. The Institutional Review Board (Ophthalmology 21, Foundation for the Advancement of Ophthalmology) waived the need for IRB approval and written informed consent from participants for this retrospective study. The study was in adherence to the tenets of the Declaration of Helsinki. All the extracted images were anonymized, and no change in the clinical pathway was anticipated.

The primary outcome of the study was to assess the sensitivity and specificity of ARIAS in detecting referable Diabetic Retinopathy (RDR). The secondary outcomes were to assess the positive & negative predictive values of ARIAS to detect RDR and to assess the sensitivity of ARIAS in detecting sight-threatening diabetic retinopathy.

Sample size: Using an alpha error of 0.05, a precision rate of 10% (two sided), an estimated sensitivity of 85%, and an estimated incidence of RDR (International Clinical Diabetic Retinopathy (ICDR) -Moderate Non-proliferative DR (NPDR) and/or presence of clinically significant diabetic macular edema (CSDME)) to be 7%, the sample size calculated was 700 participants. Given these assumptions and expecting that 10% of subjects may be qualified as insufficient quality, a sample size of 800 subjects was chosen.

Inclusion & Exclusion criteria: The retinal images of subjects with established diabetes mellitus that were captured at the time of DR screening were included. Those that did not have at least one disc and one macula-centred image of sufficient quality were excluded from the study. Additionally, subjects who received treatment for DR (lasers or intraocular injections) were excluded.

Retinal Image Acquisition: The screening process involved undilated fundus images captured using a Topcon camera Nw-400 by trained operators who followed a specific imaging protocol. For each patient, a total of four images (45 degrees field of view each) were captured. They included one image centred on the optic disc and one centred on the macula for each eye. Additional images were taken to ensure sufficient quality. Retinal images were obtained from 811 consecutive patients with established diabetes mellitus who underwent screening for DR. Images deemed of sufficient quality graded by the IDx-DR AI software were selected. A total of 3200 sufficient quality images from 811 patients were used for the study.

Reference standard grading: The patients with images of sufficient quality were split into two sub-datasets. 362 patients were graded by three Polish retina specialists and 491 by three certified graders in India. All the graders are masked to the output of the AI and to each other's grading. Images were graded for severity of diabetic retinopathy based on International Clinical Diabetic Retinopathy (ICDR) severity classification as no DR, mild NPDR, moderate NPDR, severe NPDR and Proliferative Diabetic retinopathy (PDR). Macular edema was determined by the presence of surrogate markers like hard exudates. If hard exudates were found within 1 DD of the fovea, macular edema was determined as significant and labelled as clinically significant diabetic macular edema (CSDME) present. Image grading was done on a per eye basis. The final diagnosis for each patient was determined by the stage of DR of the more affected eye. Consensus image grading was regarded as the final reference standard based on Polish and Indian graders for the comparison of both AI systems. All the analysis was performed at the patient level.

Definitions: Referable diabetic retinopathy was defined as moderate NPDR and more severe disease (moderate NPDR, severe NPDR, PDR) and/or the presence of CSDME. Sight-threatening diabetic retinopathy (STDR) was defined as severe NPDR and more severe disease (severe NPDR, PDR), and/or the presence of CSDME

Artificial Intelligence (AI) Analysis using automated grading systems: We used two different Automated Diabetic Retinopathy Image Assessment Softwares (ARIAs) i.e., Medios AI for DR (Medios Technologies, Remidio Innovative Solutions, Singapore) and IDx-DR (Digital Diagnostics, Iowa, USA). The retinal images were run on both the ARIAs to screen for DR. Both the systems processed images deemed as sufficient quality by the IDx-DR system. IDx-DR results were recorded during live screening and all images captured for the patient were analyzed on a per patient basis. Two images per eye that passed the AI quality check were submitted to the AI for DR analysis. For Medios AI analysis, anonymized images for each patient were securely transferred to a cloud platform and the images were analyzed on an automated script version of the AI on a server instead of a manual analysis through the standard iPhone app deployment.

Both the AI systems are based on Convolutional Neural Networks (CNN), with the Medios system being based on the Inception-V3 architecture. Detailed description of the model is provided in the literature [5]. In brief, the Medios AI algorithm evaluated two possible outputs: “no signs of DR detected” (non-referable DR), “signs of DR detected” (referable DR). Report was generated on a per patient basis. The IDx-DR system also has an image quality and a diagnostic algorithm. The IDx-DR system outputs the stage of DR and generates a per patient report.

Statistical analysis:

All data was stored in Microsoft Excel sheets and Apache Parquet files and was analyzed using R and Python programming languages along with Numpy, Pandas, Scikit learn and Scipy libraries. The diagnosis of the AI using Medios and IDx-DR AI systems were tabulated against the consensus image diagnosis (reference standard) by constructing 2×2 tables. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) with 95% CIs were calculated. Inter-rater agreement for Polish and Indian graders was measured by calculating the kappa statistic.

Results

The study included the images of 811 patients. Image quality analysis was evaluated as part of the clinical workflow using the IDx-DR AI system. Four patients deemed ungradable by the graders were excluded. In total, 807 patients were included for further analysis. An additional two patients were removed from STDR analysis as they did not have a consensus for a STDR diagnosis despite being labelled as RDR by consensus. Figure 1 presents the STARD diagram of retinal image selection in the study.

Grades from the Polish and Indian graders were converted to No DR, any DR, RDR and STDR before computing consensus. Based on consensus grading, there was no evidence of DR in 543 patients (67%). Any DR was seen in 264 (33%), of which 174 (22%) were referable DR, and 41 (5%) sight-threatening DR. The inter-rater agreement (Cohen's kappa) for Poland graders was 0.679 (Ophthalmologist 1), 0.904 (Ophthalmologist 2) and 0.848 (Ophthalmologist 3). For the Indian graders, kappa was 0.632 (Ophthalmologist 1), 0.916 (Ophthalmologist 2) and 0.87 (Ophthalmologist 3).

IDx-DR AI system gives an output at a stage level. 567 patients (567/807; 70.3%) were flagged positive for DR (mild, moderate or threaten). 374 of them were also categorized as RDR (46.3%, moderate or threaten) and 189 as STDR (23.4%, threaten). Out of 174 patients with ground truth labeled as RDR, the IDx-DR system detected 172. This translates to a sensitivity of 99% (95% CI 96%, 100%). The Medios AI gives a binary output. It detected the presence of Referable Diabetic Retinopathy in 291 patients (36.1%). It correctly identified 166 of the 174 patients with a ground truth diagnosis of RDR. This translates to a sensitivity of 95% (95% CI 91%, 98%). The diagnostic ability of both the AI systems including sensitivity, specificity, positive and negative predictive values are tabulated in Table 1 and 2.

Discussion

With the rising burden of DR, the importance of early detection and screening cannot be overstated. To address this glaring need, advanced technologies like AI software have emerged as promising tools for DR screening. These AI systems are meticulously developed and optimized using diverse datasets. Before implementing such AI software in real-world scenarios, it is crucial to conduct comparisons among different solutions available. In our evaluation, we examined the performance of two automated DR Image Assessment Softwares: Medios AI and IDx DR. The results exhibited comparable performance in terms of sensitivity, with Medios AI achieving 95% and IDx-DR achieving 99% in identifying referable DR respectively. These findings underscore the potential of these software solutions in facilitating early detection and screening of DR.

Overall, the prevalence of DR in the sample analyzed was 33% for any DR and 22% for RDR, significantly higher than commonly reported in other studies. Scottish National Diabetic Retinopathy Screening Programme reported rates of RDR between 4.3% and 7%, large primary-care-based screening in California reported RDR rate of 8.2% and a hospital-based study in Ethiopia found any DR rate of 18.9% [6,7,8]. It is also much higher than previous estimates for DR prevalence in Poland [9]. This is likely a side-effect of the original screening set-up. The screening is based around diabetic clinics and diabetes medical centres, therefore selecting for a higher-risk population with other diabetic complications or difficult to control disease. A similarly high prevalence of DR was found in a study of 297 patients attending a tertiary center for diabetes care in India with DR prevalence of 40.8% [5].

Only patients who initially had images of sufficient quality for IDx-DR during the initial screening were included in this study. The IDx-DR image quality assistant process was used as part of the original screening program and is not evaluated herein. Out of 811 patients deemed gradable by IDx-DR only 4 (less than 0.5%) were excluded by the manual graders indicating that overall, the images selected for this study have good image quality.

The accuracy measures for Medios AI are in line with previously published studies. Natarajan et al reported accuracy of the Medios AI offline, smartphone-based version, with sensitivity and specificity pairs of 100% and 88.4% for RDR and 85.2% and 92.0% for any DR [10]. In the aforementioned study based in a tertiary diabetes center, Medios AI achieved 98.8% and 86.7% sensitivity and specificity for any DR [5]. In another India based study of 900 prospectively included patients Medios AI achieved 83.3%, 95.5% sensitivity and specificity for any DR and 93% and 92.5% respectively for RDR [11].

Crucially, all of the above-mentioned studies were done using images gathered with the Remidio FOP mobile smartphone-based camera in contrast to using a stationary, full-size automatic fundus camera for this study. Previous studies describing Medios AI were smartphone-based, with the algorithm app being run on a smartphone, also used to take the fundus pictures. This is the first study outside of India to investigate using Medios AI with images from a stationary fundus camera in a real-world screening scenario. Images captured with different cameras may differ in resolution, level of detail, contrast, noise and other parameters that may influence the accuracy of an algorithm. It is unclear whether the software or human graders may benefit from higher resolution images and provide a more robust golden-standard, and if so, how significant the difference is. For this study the previous smartphone-based results obtained by Medios AI seem to translate into comparable accuracy when using dedicated stationary fundus camera. These results are similar to another study where Medios AI was evaluated on Topcon images in an Indian population. This demonstrates generalizability of the model performance on a desktop system. This device agnostic

approach is particularly useful in screening programs that have already invested in camera systems and would want to move towards an AI-based approach without having to replace expensive cameras. IDx-DR exceeded the sensitivity measures of Medios AI at the cost of lower specificity. We have previously reported sensitivity and specificity of IDx-DR of 94% and 95% when compared to a single reader [3,12]. In this study IDx-DR retained excellent sensitivity at 99%, with a significantly lower specificity. This was more pronounced for any DR, with IDx-DR specificity of only 44%. This may be explained in part by the fact that IDx-DR has been specifically marketed for the detection of more than mild DR, and the specificity for detection of RDR is much higher at 68% with a 99% sensitivity. For comparison, in the pivotal trial that led to IDx-DR receiving FDA approval, where IDx-DR was compared against a diagnosis based on a 7-field ETDRS study with stereoscopic images and OCT, it achieved 87% sensitivity and 90% specificity for more than mild DR [13]. Both systems over-referred mild cases (False positives included 55% milds, 45% no DR by Medios AI; 38% mild, 62% no DR cases by IDx-DR). Another possibility of lower specificity could be referral of patients with similar lesions and concurrent pathologies that were not evaluated as part of this grading. Overall, both systems achieved satisfactory accuracy, particularly when patient safety is concerned with excellent negative predictive values, meaning very few patients receiving a false-negative result. Although false-positive results generate additional costs and workload due to the increase in referrals, it is the patients with missed disease that raise the most concern whenever automated screening is debated.

Many of the studies regarding automated analysis of DR from fundus images are sponsored or even performed directly by the respective software's owner company, which raises questions regarding bias. As previously mentioned Medios AI does not offer a dedicated desktop application at this point, therefore it was necessary to submit the images to Remidio, owner of the Medios AI algorithm, for a remote analysis on their system. As the authors of this study collaborated remotely, we could not directly oversee or verify the Medios AI output on site. Upon reviewing of the study methodology, we considered this to be a source of potential bias and asked Remidio for a way to independently verify some of the software's results. We submitted the subset of images assessed by Polish graders, through a dedicated API (application programming interface) provided by Remidio with live results. Images were anonymized without changing the image content. The results were in line with those previously submitted by Remidio for all but 3 patients.

Out of those three patients, for whom the initial MediosAI output differed from the verification, all three decisions changed from no RDR to RDR. For two of those patients the new MediosAI result now matched the grader decision of RDR, for the remaining patient the new MediosAI now disagreed with the grader consensus. All three of those patients had very subtle retinal signs. This is a study looking into MediosAI outside of the smartphone application which involves a custom-made deployment for the study. The discrepancies are likely due to challenges surrounding the implementation of the algorithm on different hardware or inconsistencies in image compression parameters between the version of the images submitted to Remidio for the first analysis and the version of the images sent for verification through the API.

Inter-grader and intra-grader variability is known amongst DR graders. Previous studies have shown that inter-grader kappa scores typically range from 0.40 to 0.65 in DR grading [13-19]. The kappa values for both Polish (0.68-0.90) and Indian graders (0.63-0.91) in the study showed similar variability but were well within the limits showing overall good agreement. This ensured reliability while having the data split and graded by both groups separately. The possible reasons for variability amongst graders could be identification and differentiation of subtle DR features (retinal hemorrhages, microaneurysms, hard exudates, new vessels, intraretinal microvascular abnormalities, neovascularization, and surrogate markers of macular edema), variation in image quality due to artifacts, brightness or contrast of images. This has been found in other studies and in other fields of medical imaging as well [13-15]. Gold standard grading in the current study was done based on a majority decision by the graders. As the individual grades were converted to a binary decision for each grader before computing consensus grading there was a majority decision for any DR and RDR for each of the patients. The reliability of the human grading could be improved with an adjudication process for patients without a full consensus [14].

This study included only patients with good quality, non-mydriatic images, which may not be representative of the whole screening cohort. Using a non-mydriatic protocol may underrepresent patients with smaller pupils or media opacities, particularly the elderly. In a previous study about DR grader reliability, based on the same screening programme in Poland from which images for this study were taken, out of 495 patients only 335 were deemed as sufficient quality by IDx-DR and all three human graders [15]. How many of those low-quality screening encounters

could we image and diagnose after mydriasis remains to be seen, as is the comparative performance of both systems in those patients.

In conclusion, our study compared the performance of two AI screening software, Medios AI and IDx DR, in detecting RDR. Both software systems demonstrated robust performance, with high accuracy and sensitivity, highlighting their potential as reliable tools for screening DR in real-world settings. Continued research and validation in larger and diverse patient populations will be essential to strengthen the evidence base and ensure the widespread adoption of these AI screening tools. Our study underscores the promise of these AI systems for DR screening, facilitating early detection and timely intervention for improved patient outcomes.

Statement of Ethics: The Institutional Review Board (Ophthalmology 21, Foundation for the Advancement of Ophthalmology, Application No. 2/2022) waived the need for IRB approval and written informed consent from participants for this retrospective study. The study was in adherence to the tenets of the Declaration of Helsinki.

Conflict of Interest: AG has Grants/Contracts from Alcon, Bausch & Lomb, Zeiss, Hoya, Thea, Viatris, Teleon, J&J, Cooper Vision, Essilor and Polpharma. AG has consulting fees / honoraria from Thea, Polpharma, Viatris and stock with GoCheckKids. DRP, FMS and KN are employees of Remidio Innovative Solutions. Remidio Innovative Solutions, Inc, USA and Medios Technologies are wholly owned subsidiary of Remidio Innovative Solutions Pvt Ltd, India. FMS has patents (mentioned in ICMJE) and stock (ESOP and stock, Remidio Innovative Solutions Pvt Ltd). Other authors declare no financial disclosures.

Funding/Support: There was no funding for this study

Author contribution

AG – Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing

DRP – Conceptualization; Formal analysis;; Investigation; Methodology; Project administration; Resources; Supervision; Visualization; Validation; Writing - review & editing

PB – Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Roles/Writing - original draft; Writing - review & editing

KN –Formal analysis; Writing - original draft; Writing - review & editing

TK - Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing

FMS - Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing - review & editing

Data Availability Statement

The data that support the findings of this study is not publicly available due to ethical reasons and are available on request from the Dr Andrzej Grzybowski (email: ae.grzybowski@gmail.com).

References

1. Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GSW, Abramoff M, Ting DSW. Artificial intelligence for diabetic retinopathy screening: a review. Eye 2020;34(3):451-460.
2. Ting DS, Cheung GC, Wong TY. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. Clin Exp Ophthalmol. 2016;44(4):260-77.
3. Grzybowski A, Brona P. Analysis and Comparison of Two Artificial Intelligence Diabetic Retinopathy Screening Algorithms in a Pilot Study: IDx-DR and Retinalyze. J Clin Med. 2021;10(11):2352.
4. Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee YE, el al. Multicenter, Head-to-Head, Real-World Validation Study of Seven Automated Artificial Intelligence Diabetic Retinopathy Screening Systems. Diabetes Care. 2021;44(5):1168-1175.
5. Sosale B, Sosale AR, 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.
6. Looker HC, Nyangoma SO, Cromie DT, Olson JA, Leese GP, Black MW, et al. Scottish Diabetes Research Network Epidemiology Group; Scottish Diabetic Retinopathy Collaborative. Rates of referable eye disease in the Scottish National Diabetic Retinopathy Screening Programme. Br J Ophthalmol. 2014;98(6):790-5.

7. Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. J Diabetes Sci Technol. 2009;3(3):509-16.
8. Tilahun M, Gobena T, Dereje D, Welde M, Yideg G. Prevalence of Diabetic Retinopathy and Its Associated Factors among Diabetic Patients at Debre Markos Referral Hospital, Northwest Ethiopia, 2019: Hospital-Based Cross-Sectional Study. Diabetes Metab Syndr Obes. 2020;13:2179-2187.
9. Kozioł M, Nowak MS, Udziela M, Piątkiewicz P, Grabska-Liberek I, Szaflik JP. First nation-wide study of diabetic retinopathy in Poland in the years 2013-2017. Acta Diabetol. 2020;57(10):1255-1264.
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.
11. Sosale B, Aravind SR, Murthy H, Narayana S, Sharma U, Gowda SGV, Naveenam M. Simple, Mobile-based Artificial Intelligence Algorithm in the detection of Diabetic Retinopathy (SMART) study. BMJ Open Diabetes Res Care. 2020;8(1):e000892.
12. Grzybowski A, Brona P. A pilot study of autonomous artificial intelligence-based diabetic retinopathy screening in Poland. Acta Ophthalmol. 2019;97(8):e1149-e1150.
13. 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;1:39.
14. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, et al. Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. Ophthalmology. 2018;125(8):1264-1272.
15. Grzybowski A, Brona P, Krzywicki T, Gaca-Wysocka M, Berlińska A, Święch A. Variability of Grading DR Screening Images among Non-Trained Retina Specialists. J Clin Med. 2022;11(11):3125.
16. Guan MY, Gulshan V, Dai AM, Hinton GE. Who said what: Modeling individual labelers improves classification. In: Thirty-Second AAAI Conference on Artificial Intelligence. 2018. arXiv:1703.08774v2.
17. Sedova A, Hajdu D, Datlinger F, Steiner I, Neschi M, Aschauer J, Gerendas BS, Schmidt-Erfurth U, Pollreisz A. Comparison of early diabetic retinopathy staging in asymptomatic patients between autonomous AI-based screening and human-graded ultra-widefield colour fundus images. Eye (Lond). 2022 Mar;36(3):510-516.
18. Gangaputra S, Lovato JF, Hubbard L, Davis MD, Esser BA, Ambrosius WT, Chew EY, Greven C, Perdue LH, Wong WT, Condren A, Wilkinson CP, Agrón E, Adler S, Danis RP; ACCORD Eye Research Group. Comparison of standardized clinical classification with fundus photograph grading for the assessment of diabetic retinopathy and diabetic macular edema severity. Retina. 2013 Jul-Aug;33(7):1393-9.
19. Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, Campana BJ, Phene S, Hemarat K, Tadarati M, Silpa-Acha S. Deep Learning vs. Human Graders for Classifying Severity Levels of Diabetic Retinopathy in a Real-World Nationwide Screening Program. arXiv preprint arXiv:1810.08290. 2018 Oct 18.

Figure Legends:

Fig. 1. STARD flow diagram showing the patient breakdown for the Referable Diabetic Retinopathy (RDR) analysis

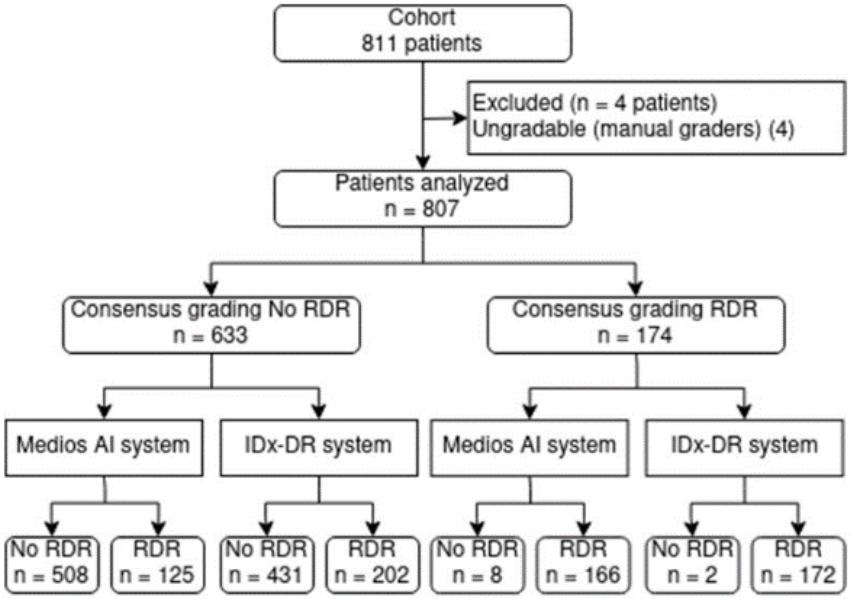


Table 1a) Performance of Medios and IDx DR with ground truth cutoff at any DR (n=264)

		Automated DR Image Assessment Softwares				
		Medios AI		IDx DR		
		Positive	Negative	Positive	Negative	TOTAL
Reference Standard (Consensus grading)	Positive (Mild NPDR and above and/or CSDME)	235 (29%)	29 (4%)	262 (32%)	2 (0.2%)	264
	Negative (No DR)	56 (7%)	487 (60%)	305 (38%)	238 (29%)	543
	TOTAL	291	516	567	240	807

Table 1b) Performance of Medios and IDx DR with ground truth cutoff at RDR (n=174)

		Automated DR Image Assessment Softwares				
		Medios AI		IDx DR		
		Positive	Negative	Positive	Negative	TOTAL
Reference Standard (Consensus grading)	Positive (Moderate NPDR and above and/or CSDME)	166 (21%)	8 (1%)	172 (21%)	2 (0.2%)	174
	Negative (No DR and mild NPDR)	125 (15%)	508 (63%)	202 (25%)	431 (53%)	633
	TOTAL	291	516	374	433	807

Table 1c) Performance of Medios and IDx DR with ground truth cutoff at STDR (n=41)

		Automated DR Image Assessment Softwares				
		Medios AI		IDx DR		
		Positive	Negative	Positive	Negative	TOTAL
Reference Standard (Consensus grading)	Positive (Severe, PDR or CSDME)	40 (5%)	1 (0.1%)	39 (5%)	2 (0.2%)	41

*

DR – Diabetic Retinopathy, NPDR- Non-Proliferative Diabetic Retinopathy, PDR – Proliferative Diabetic Retinopathy, CSDME – Clinically significant Diabetic Macular Edema, RDR - Referable Diabetic Retinopathy, STDR – Sight-threatening Diabetic Retinopathy

Table 2: Performance Analysis of Artificial Intelligence System Compared to Reference Standard

Values in % (95% CI)	For any DR		For referable DR		For sight-threatening DR	
Variables	Medios AI	IDx-DR	Medios AI	IDx-DR AI	Medios AI	IDx-DR AI
Sensitivity	89 (85, 93)	99 (97, 100)	95 (91, 98)	99 (96, 100)	98 (87, 100)	95 (83, 99)
Specificity	90 (87,92)	44 (40, 48)	80 (77, 83)	68 (64, 72)	NA	80 (77, 83)
PPV	81 (76, 85)	46 (42, 50)	57 (51, 63)	46 (41, 51)	NA	21 (15, 27)
NPV	94 (92, 96)	99 (97, 100)	98 (97, 99)	100 (98, 100)	NA	100 (99, 100)

BMJ Open Diabetes Research & Care

Simple, Mobile-based Artificial Intelligence Algorithm in the detection of Diabetic Retinopathy (SMART) study

Bhavana Sosale ,[1] Sosale Ramachandra Aravind,[1] Hemanth Murthy,[2] Srikanth Narayana,[3] Usha Sharma,[3] Sahana G V Gowda,[3] Muralidhar Naveenam[2]

To cite: Sosale B, Aravind SR, Murthy H, *et al*. Simple, Mobile-based Artificial Intelligence Algorithm in the detection of Diabetic Retinopathy (SMART) study. *BMJ Open Diab Res Care* 2020;**8**:e000892. doi:10.1136/bmjdrc-2019-000892

This study was presented at the Annual Conference of the American Diabetes Association on 8 June 2019, San Francisco, USA.

Received 10 September 2019
Revised 6 December 2019
Accepted 4 January 2020

[1]Diabetology, Diacon Hospital, Bangalore, India
[2]Ophthalmology, Retina Institute of Karnataka, Bangalore, India
[3]Ophthalmology, Diacon Hospital, Bangalore, India

Correspondence to
Dr Bhavana Sosale;
bhavanasosale@gmail.com

ABSTRACT

Introduction The aim of this study is to evaluate the performance of the offline smart phone-based Medios artificial intelligence (AI) algorithm in the diagnosis of diabetic retinopathy (DR) using non-mydriatic (NM) retinal images.

Methods This cross-sectional study prospectively enrolled 922 individuals with diabetes mellitus. NM retinal images (disc and macula centered) from each eye were captured using the Remidio NM fundus-on-phone (FOP) camera. The images were run offline and the diagnosis of the AI was recorded (DR present or absent). The diagnosis of the AI was compared with the image diagnosis of five retina specialists (majority diagnosis considered as ground truth).

Results Analysis included images from 900 individuals (252 had DR). For any DR, the sensitivity and specificity of the AI algorithm was found to be 83.3% (95% CI 80.9% to 85.7%) and 95.5% (95% CI 94.1% to 96.8%). The sensitivity and specificity of the AI algorithm in detecting referable DR (RDR) was 93% (95% CI 91.3% to 94.7%) and 92.5% (95% CI 90.8% to 94.2%).

Conclusion The Medios AI has a high sensitivity and specificity in the detection of RDR using NM retinal images.

Significance of this study

What is already known about this subject?

- Artificial intelligence (AI) algorithms can aid in the diagnosis of diabetic retinopathy.
- These algorithms work with images taken from expensive table top fundus cameras.
- Non-availability of a fundus camera and need for high-speed internet access are limitations to their use in practice.

What are the new findings?

- This study evaluated the performance a new AI algorithm that works offline on a smart phone fundus camera.
- The novel 'offline' Medios AI algorithm had a high sensitivity for the diagnosis of referable diabetic retinopathy and sight threatening diabetic retinopathy on non-mydriatic (NM) images captured with the Remidio fundus on phone camera.

How might these results change the focus of research or clinical practice?

- The Medios AI and the portable Remidio NM fundus-on-phone camera together are a complete integrated solution for diabetic retinopathy detection and can make screening accessible and scalable in countries with limited resources.

INTRODUCTION

Diabetic retinopathy (DR) is the most common cause of preventable blindness. India has close to 73 million individuals with diabetes.[1–3] Screening and early diagnosis of DR results in early referral to the specialist, and initiation of measures to improve glycemic control and reduce progression.[4–6]

Lack of awareness, limited access to ophthalmologists, need for expensive equipment and socioeconomic barriers are challenges to screening.[1] Although tele-ophthalmology makes screening more accessible, it is not free from challenges like the need for pupil dilatation, size and cost of fundus cameras, network connectivity issues, intergrader variability and access to ophthalmologists or trained readers.

Artificial intelligence (AI) is a potential scalable alternative in DR screening. It helps to reduce the manual burden on ophthalmologists and overcome the barriers with tele-ophthalmology. Recent advances in machine learning and convolutional neural networks has made it possible to analyze large amounts of data, recognize patterns and generate reports. AI algorithms developed for DR screening (eg, Google AI, EyeArt and IDx-DR) work on cloud-based platforms.[7–10] The captured images are uploaded online and the algorithm provides an output within an acceptable turn over time. In low-income and middle-income countries, limited internet access or reduced bandwidth limits the use of these solutions. In addition, most cameras integrated with AI software are the traditional expensive, large fundus cameras which require the operator to capture a dilated retinal image.

The AI algorithm by Medios Technologies, Singapore is to our knowledge the first offline software for DR screening integrated with the smart phone-based fundus camera, the Remidio non-mydriatic (NM) fundus-on-phone (FOP).[11] Studies evaluating the performance of this algorithm are limited. This study aims to evaluate the performance of an offline AI algorithm—Medios in DR screening using NM retinal images taken from the smartphone-based Remidio NM FOP retinal camera.

AIMS AND OBJECTIVES

Primary aim

To evaluate the performance of the AI algorithm in detecting any grade of DR using NM retinal images captured from patients with diabetes mellitus.

Secondary aims

To evaluate the performance of the AI algorithm in detecting referable diabetic retinopathy (RDR). RDR is defined as presence of disease greater than moderate non-proliferative DR or the presence of diabetic macular edema (DME). In addition, the ability of the algorithm to correctly identify all cases identified as STDR (severe NPDR or more severe disease or the presence of DME) by image diagnosis was also evaluated.

METHODS

The study was carried out as per the tenets of the Declaration of Helsinki (NCT03572699). Informed consent was provided was all participants enrolled in the study.

This study prospectively enrolled patients attending the outpatient department of Diacon Hospital, a university recognized, tertiary center for diabetes care and research, Bangalore, India between July and November 2018. All subjects, above the age of 18 years, with diabetes mellitus were invited to enroll for the study. Eyes with significant media opacity such as corneal opacity or cataract that precluded retinal imaging were excluded and those with known retinal vascular (artery or vein) occlusion were excluded. Enrollment continued until gradable retinal images were obtained from 900 patients. All consenting individuals meeting the inclusion criteria were screened for DR as part of routine care.

Retinal image acquisition

Undilated retinal images were captured using the smartphone based 'Remidio FOP camera' (Remidio Innovative Solutions, Bangalore, India) by a trained technician. Two images (ie, disc centered (nasal field) and macula centered (posterior pole)) were captured from each eye of each patient. The technician was trained to recognize the characteristics of an excellent image and was urged to capture more than one image per field of view if required to obtain excellent images. Two additional attempts were allowed to capture the image if the image was of poor quality (eg, an out-of-focus image, or in those with a small pupil).

Image grading by retina specialist

The de-identified (ie, anonymized) images with the subject ID were uploaded online from the FOP to an Amazon Web Services (AWS) hosted cloud service provided by the manufacturer. The images were accessed from the cloud by five retina specialists, that is, three fellowship-trained vitreoretinal surgeons and two medical retina specialists. The retina specialists individually graded the set of four retinal photographs from every eye using the International Clinical Diabetic Retinopathy Classification Severity Score.[12] Images were graded as no DR, mild non-proliferative DR (mild NPDR), moderate non-proliferate DR (moderate NPDR), severe non-proliferate DR (severe NPDR) and proliferate DR (PDR). The images with DR were then evaluated for DME. The diagnosis of diabetic macular edema (DME) was graded as no DME, mild DME, moderate DME and severe DME. The eye with the more severe stage of retinopathy was considered as the final diagnosis for that patient, in cases where each eye had a different stage of disease severity. Patients whose images were considered as ungradable by the retina specialists were excluded from the final analysis. The majority diagnosis of the five graders was considered as the final image diagnosis. The patient-wise diagnosis obtained from the retina specialists were considered as gold standard for comparison. Each retina specialist was blinded to the diagnosis of the others and to the diagnosis of the AI.

Image analysis using AI-based offline software

The images captured from the subjects were run offline on the iPhone6 using the Medios AI and the diagnosis was recorded in binary as DR present or absent.

Description of the AI software

The AI diagnosis system developed by Medios Technologies is based on Convolutional Neural Networks. It consists of a first neural network for image quality assessment and two other distinct neural networks that detect DR lesions. A final per-patient DR diagnosis is computed from the outputs of both DR neural networks and applied on all images of that patient.

Image processing is applied before feeding the images to the neural networks. The images are cropped by removing the black border surrounding the circular field of view typical of retinal images. They are resized to a common 512×512 pixels resolution. The neural network responsible for quality assessment is based on a MobileNet architecture. It consists of a binary classifier trained with images deemed as ungradable as well as with images deemed of sufficient quality. If the output is negative, a message prompts the user to recapture the image.

The other two neural networks are based on an Inception-V3 architecture and have been trained to separate healthy images from images with referable DR (moderate NPDR and above). The final output is a binary recommendation of referral to an ophthalmologist. No mild NPDR images have been used during training of the AI. The system has thus been engineered to maximize the

sensitivity for referable DR and the specificity for any DR. Both networks independently analyze the images. One uses images that have been preprocessed by a contrast enhancement image processing algorithm, while the other does not. A linear classifier merges outputs of both networks into a final per-image prediction. A patient is deemed as a referable case if the prediction for one or more images is positive.

A comprehensive dataset consisting of images taken in a variety of conditions has been used for training, with a proportion of it taken using NM and/or low-cost cameras. These include 4350 NM images taken during screening camps with the Remidio FOP, and 14 266 images captured with a KOWA vx-10 mydriatic camera and 34 278 images come from the EyePACS dataset. Half of the training set contained DR cases, and the other half healthy ones.

Neural networks traditionally run on computationally powerful servers to which the end user connects and sends images. In this case, the neural network is deployed directly on the phone, leveraging smartphone technologies to make full usage of the inbuilt hardware. The whole AI diagnosis pipeline runs offline on the iPhone of the Remidio NM FOP. 'Offline' refers to the computational unit on which AI inference is performed. Thanks to leveraging on the high-performance capabilities of the smartphone with Core Machine Learning platforms and Open Graphics Library, image processing is done directly on the Graphics Processing Unit instead of relying on a connection to a server on the internet. There is no degradation in performance of the algorithm as a result of deploying it offline versus online. This is because, the offline mode is primarily a method of deployment that uses the smartphone to run the same algorithm as it would have on a cloud server. With newer updates, continuously trained models can be deployed through the app store, which will enable the model to get the best inferencing convenience, re-training and continuous deployment using the iPhone as a platform.

The interface and the report also provide a visual representation of the areas of the retinal images that are responsible for a positive diagnosis. This is based on a deep learning technique called class activation mapping.

Two distinct datasets have been used for internal validation and fine-tuning of the linear classifier. Both datasets had not been used for training and consist of images taken in the mydriatic mode of the camera. One dataset was captured at Dr Mohan's Diabetes Specialities Center in Chennai, while the other was captured at Diacon Hospital in Bangalore. These results were computed independently of the institutions who provided the data. The datasets consisted of 3038 and 1054 images, respectively. The images used for training and internal validation of the AI do not overlap with those captured for the SMART study.

Outcome measures

The primary aim was to determine the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the AI algorithm in detecting all DR compared with the gold standard diagnosis by retina specialists. The secondary aims were to determine the sensitivity, specificity, PPV and NPV of the algorithm in the diagnosis of RDR. RDR was defined as moderate NPDR or more severe disease or the presence of DME. The ability of the algorithm to correctly identify all cases identified as sight threatening DR (STDR) by image diagnosis was also evaluated. STDR was defined as severe NPDR or more severe disease or the presence of DME.

Statistical analysis

The Food and Drug Administration (FDA) mandated superiority cut-offs (for AI algorithms for DR screening) for sensitivity and specificity were 85% and 82.5%.[7] The sample size required for a sensitivity of 85%, given a sensitivity of 75% under the null hypothesis using a one-sided test, 0.025 alpha and 90% power was 171 individuals with RDR. The sample size required for a specificity of 82.5% given a specificity of 75% under the null hypothesis using a one-sided test, 0.025 alpha and 90% power was 682 individuals with no RDR (combined no DR and mild NPDR). The minimum sample required was 853 and we planned to continue enrollment until gradable images could be obtained from 900 individuals.

All data were stored in Microsoft Excel and was analyzed using StataCorp V.14.2. The diagnosis of the AI was tabulated against the image diagnosis (reference standard) by constructing 2×2 tables. The sensitivity, specificity, PPV and NPV with 95% CIs were calculated, and area under the curve (AUC) plotted for all DR and RDR. In individuals diagnosed with STDR, the sensitivity (ability of the AI to correctly identify those with disease) was measured. Intergrader agreement was measured by calculating the kappa statistic.

RESULTS

The study enrolled 922 patients and the analysis included images from 900 patients (figure 1). Based on the image diagnosis, there was no evidence of DR in 648 participants (72%). Mild NPDR was seen in 51 (5.67%), moderate NPDR in 163 (18.11%), severe NPDR in 3 (0.33%) and

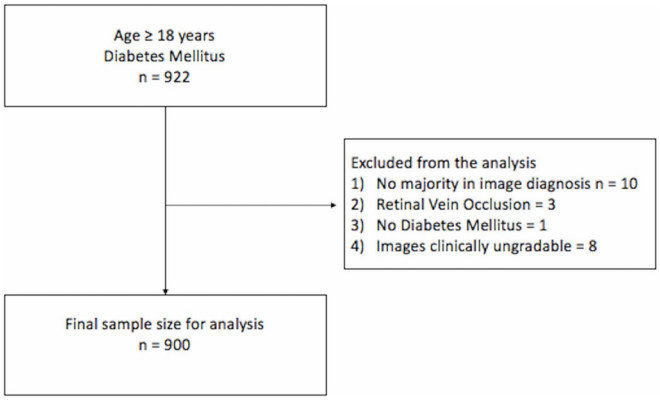


Figure 1 Flow chart depicting study enrollment.

Table 1 Performance of the Medios AI

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC
All DR	83.3% (80.9% to 85.7%)	95.5% (94.1% to 96.8%)	87.8% (85.7% to 90%)	93.6% (92% to 95.2%)	0.9
RDR	93% (91.3% to 94.7%)	92.5% (90.8% to 94.2%)	78.2% (75.5% to 80.9%)	97.8% (96.9% to 98.8%)	0.88

AI, artificial intelligence; AUC, area under the curve; DR, diabetic retinopathy; NPV, negative predictive value; PPV, positive predictive value; RDR, referable diabetic retinopathy.

PDR in 35 (3.89%). Mild DME was present in 12 (4.76%), moderate DME in 32 (12.69%) and severe DME in 3 (1.19%) individuals with DR with different grades of non-proliferative or proliferative DR. The intergrader agreement (quadratic weighted kappa) between the individual ophthalmologists and the majority diagnosis was between 0.79 and 0.91. Common causes of differences in diagnosis between retina specialists and the majority diagnosis were missed single microaneurysms (MA), and differentiating dot hemorrhages from MA.

The AI classified 239 (26.5%) of images as DR and 661 (73.4) % as no DR. The performance of the AI in detecting all DR and RDR is summarized in table 1. The AUC for all DR and RDR are shown in figure 2. An example of the output from the Medios AI algorithm with an image diagnosis of RDR is shown in figure 3.

The AI was able to correctly diagnose 76/80 cases graded as STDR as having signs of retinopathy. Sensitivity for STDR was 95.2% (95% CI 88.2% to 98.6%). The three PDRs missed by the AI were postlaser images with no active changes visible. When these three images were excluded, the AI correctly identified 76/77 cases of STDR as having signs of retinopathy. The sensitivity for STDR was found to be 98.7% (95% CI 92.9% to 99.7%).

The kappa that is, agreement between the AI and the opthalmologists' diagnosis was found to be 0.8.

DISCUSSION

To our knowledge, this is the first study to evaluate the performance of an AI algorithm for DR screening using NM images captured from a portable smartphone-based fundus camera. The analysis from this large study showed that the Medios AI has a high sensitivity in the detection of RDR and STDR.

The use of AI algorithms as fully automated screening solutions for DR diagnosis is on the rise. The only one to have made it past FDA's cut is IDx-DR on the basis of a clinical study conducted with mydriatic retinal images obtained from 900 individuals. In this study, the sensitivity and specificity of the IDx-DR system in identifying RDR was 87% and 90%, meeting the FDA superiority sensitivity and specificity cut-offs of 85% and 82.5%, respectively. Despite its accuracy, it is not recommended for evaluating rapidly progressive DR.[7] Limitations include the need for integration with expensive traditional fundus cameras. The Iowa Detection Program, a clinical study conducted to evaluate IDx-DR V.X2.1, from mydriatic retinal photographs, showed a sensitivity of 96.8% and specificity of 87.0%.[8]

The performance of the Google AI was studied on the EyePACS-1 and Messidor 2 datasets. In the EYE-PACS

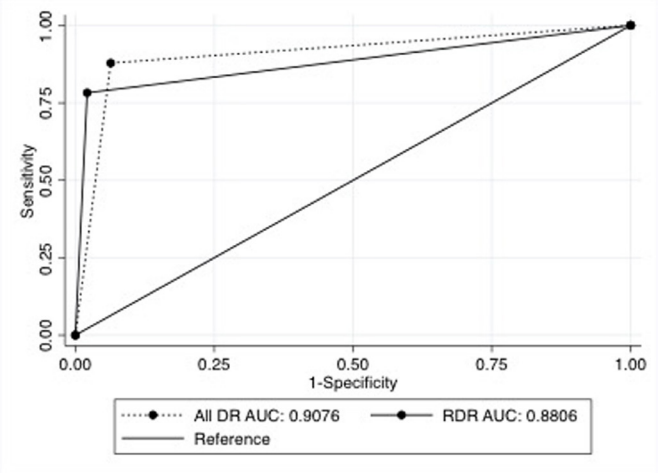


Figure 2 Area under the curve (AUC) of the Medios artificial intelligence algorithm for all diabetic retinopathy (all DR) and referable diabetic retinopathy (RDR).

Figure 3 Example of the output of the Medios artificial intelligence algorithm in an individual with a diagnosis of referable diabetic retinopathy.

dataset, the sensitivity and specificity of the algorithm for RDR was 90.1% (95% CI 87.2% to 92.6%) and 98.2% (95% CI 97.8% to 98.5%). In the Messidor 2 dataset, the sensitivity and specificity was 86.6% (95% CI 80.5% to 90.7%) and 98.4% (95% CI 97.5% to 99%) for the detection of RDR.[10] In a prospective study, the Google AI was validated across two sites in India. The sensitivity and specificity of the algorithm for the detection of RDR at Aravind Eye Hospital was 88.9% (95% CI 85.8% to 91.5%) and 92.2% (95% CI 90.3% to 93.8%); and 92.1% (95% CI 90.1% to 93.8%) and 95.2% (95% CI 94.2% to 96.1%) at Shankara Nethralaya.[13]

EyeArt (Eyenuk, Woodland Hills, California, USA) using dilated retinal images of 296 patients captured by the Remidio NMFOP was validated by Rajalaksmi *et al*. The authors reported a sensitivity of 95.8% and specificity of 80.2% for detecting any DR and a sensitivity of 99.1% and a specificity of 80.4% for detecting STDR.[14] In a recent retrospective study, Bhaskaranand *et al* reported a sensitivity and specificity of 91% using EyeArt on 101 710 individuals.[9] Another study by Ting *et al* with multiple retinal images taken with conventional fundus cameras from multiethnic cohorts of people with diabetes, reported a sensitivity and specificity for identifying RDR of 90.5% and 91.6%.[15]

Most of these AI algorithms require high-speed computational power and internet access for immediate reporting, in addition to the need for expensive desktop fundus cameras. This sets the Medios AI apart from other AI solutions, in being the first offline end-to-end solution integrated on the smart phone camera.

The Remidio FOP is an FDA510k cleared medical device validated in head-to-head studies against Topcon TRC 50DX and Zeiss FF450. It is the only smartphone-based device shown to have a high sensitivity and specificity in detection of all grades of DR, in non-mydriatic imaging.[11 16] The results seen with the Medios AI using images captured from the Remidio FOP meet the FDA superiority cut-offs and are comparable to results observed with other AI algorithms, such as Google AI, EyeArt or IDx-DR for the detection of RDR.[7 9 13] In a previous study by Rajalakshmi *et al* published in *Eye*, the Eyenuk AI algorithm, EyeArt was found to have very high sensitivity and specificity for detection of RDR and STDR when used on the Remidio FOP camera images (despite EyeArt not having been earlier trained on the Remidio FOP images).[14]

A recent study by Natarajan *et al* evaluated the performance of the Medios AI using dilated retinal images captured using the Remidio FOP from 231 individuals with diabetes. The images were captured by a healthcare worker in a primary healthcare community screening camp. The authors reported that the sensitivity and specificity of the AI in the diagnosis of RDR as 100% and 88.4%, and for any DR as 85.2% and 92%.[17]

There are a few differences between the study done by Natarajan *et al* and our study.[17] Natarajan *et al* evaluated the AI's performance using mydriatic images taken during community screening by healthcare workers. A sensitivity analysis was performed to assess the AI's performance using both good quality images, and images that did not meet the minimum quality standards of the AI. In this analysis, the sensitivity of the AI for RDR remained unchanged, while the specificity dropped to 81.9%. The increase in false positive outputs were attributed to image quality. This did not translate to a concern regarding patient safety as all individuals with RDR were detected by the AI. In contrast, our study used NM images captured in a clinic setting by a trained camera technician on a larger number of individuals. Both studies demonstrate a high sensitivity and specificity for the detection of RDR. The ease of use of the device by a community healthcare worker, and results observed with both mydriatic and NM images support the use of smart phone fundus imaging and AI-based reporting for DR screening.

Cloud-based AI algorithms require internet access for real time reporting. In countries like India, where mass screening is the need of the hour, access to continuous electricity and internet is a constraint. The FOP with inbuilt offline AI can address these operational challenges in rural and urban areas in the low-income and middle-income countries with limited resources. The offline mode of AI is advantageous in the context of clinical work flow and ground deployment constraints, to ensure that DR screening can move forward without interruptions.

In this study, we observed that the AI was unable to identify laser marks as 'DR' in those who had 'no active DR changes' post pan-retinal photocoagulation. It is worthwhile to note that other studies exclude individuals who have undergone laser treatment and hence it is not possible to ascertain if other AI algorithms also behave similarly.[7 13 18 19] Considering the practical application of these AI algorithms to be in primary care and screening (and not in tertiary hospitals visited by those with DR postlaser treatment), this finding in no way should undermine the robustness of this algorithm's performance.

Images from eight individuals were considered clinically ungradable (figure 1). The AI algorithm had flagged six of these images as poor quality, but did provide an output for these images. Since there was no ground truth for comparison with, these images were excluded from the analysis. However, the ability to identify signs of disease pathology invisible to the human eye in less than ideal conditions—a trait of deep learning algorithms—deserves merit. In order to report if AI algorithms perform better than a clinician on poor quality/hard to grade images, it may be necessary to analyze this with a larger pool of ungradable images. Hence, at this point in time, in a real-life situation, during screening, it is necessary to use caution, and refer cases where the AI quality check flags the image as poor quality.

Limitation of our study was that it only included NM images. Hence, screening in elderly or in those with a small pupil (<3 mm) can be a challenge. Dilatation with a drop of 1% tropicamide solution may be necessary in

these cases. The strengths of the study include prospective validation of the AI in a large sample against the diagnosis of five retinal specialists. Future studies that assess the performance of the AI compared with the adjudicated diagnosis, the clinical diagnosis and studies that evaluate integration of the AI into the clinical workflow are needed. We acknowledge that the AI in its current version works only integrated with the FOP and has the ability to only provide a diagnosis of referral versus no referral. Even though the algorithm is currently unable to give an output of STDR directly, we believe that every end-user should be aware of the ability of the AI to correctly identify those with STDR (at the highest risk of vision loss) and be aware of the rates and reasons of a missed diagnosis in its current version. The AI is currently being trained to grade DR, provide a diagnosis of DME and STDR in the future versions (which will continue to be deployed offline on the iPhone), to assist in triaging and immediate referral.

Our study is the first in validating the use of Medios AI in a large clinical setting using NM images. Our results show that the AI has a high sensitivity and specificity in the detection of RDR. This is the only AI system that works offline and produces real time reports on a smart phone. Multiple large-scale studies that validate the algorithm are necessary. If results are reproducible in both the mydriatic and NM setting, the Medios AI has the potential to be the scalable solution to make DR screening accessible at the primary care level.

Acknowledgements The authors would like to thank Medios Technologies and Remidio Innovative Solutions for providing the NM FOP 10 camera and the AI software for conducting this study. The authors would like to thank Florian M. Savoy for the description of the technical design of the AI software. The authors would also like to thank Mr Satish, Mrs Roopa and Mrs Rekha for help with imaging and data collection.

Contributors This manuscript has been read and approved by all the authors, the requirements for authorship have been met and each author believes that the manuscript represents honest work. BS, SRA and MN contributed to the study concept, design, intellectual content, literate search, clinical study, data acquisition, manuscript preparation, editing and review. BS conducted the statistical analysis. HM, SN, US and SGVG contributed to the clinical study, data acquisition, manuscript editing and review. All authors contributed to the final version of the manuscript and are responsible for the integrity of the work.

Funding The study was funded by Diacon Hospital and did not receive any external funding.

Competing interests BS and SRA are related to one of the founders of Medios Technologies, a subsidiary of Remidio Innovative Solutions. The authors have received no consulting fees or remuneration of any form from Medios or Remidio. The study design, conduct and analysis was conducted independently by Diacon Hospital. The study was funded by Diacon Hospital. The authors acknowledge Medios and Remidio for providing the fundus camera and the AI algorithm for use during the study. The other authors have no conflict of interest.

Patient consent for publication Not required.

Ethics approval This study has been approved by the Diacon Hospital Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Addtional data are available on reasonable request. All data relevant to the study are included in the article.

ORCID iD

Bhavana Sosale http://orcid.org/0000-0002-3658-1252

REFERENCES

1 Graham-Rowe E, Lorencatto F, Lawrenson JG, *et al*. Barriers to and enablers of diabetic retinopathy screening attendance: a systematic review of published and grey literature. *Diabet Med* 2018;35:1308–19.
2 Gadkari SS. Diabetic retinopathy screening: telemedicine, the way to go! Indian. *J Ophthalmol* 2018;66:187–8.
3 IDF Diabetes Atlas - Across The Globe 2017. Available: http://www.diabetesatlas.org/across-the-globe.html [Accessed 1Aug 2019].
4 American Diabetes Association. 11. Microvascular Complications and Foot Care: *Standards of Medical Care in Diabetes-2019*. *Diabetes Care* 2019;42:S124–38.
5 Solomon SD, Chew E, Duh EJ, *et al*. Diabetic retinopathy: a position statement by the American diabetes association. *Diabetes Care* 2017;40:412–8.
6 Bajaj S. RSSDI clinical practice recommendations for the management of type 2 diabetes mellitus 2017. *Int J Diabetes Dev Ctries* 2018;38:1–115.
7 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.
8 Abràmoff MD, Lou Y, Erginay A, *et al*. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci* 2016;57:5200–6.
9 Bhaskaranand M, Ramachandra C, Bhat S, *et al*. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technol Ther* 2019;21:635–43.
10 Gulshan V, Peng L, Coram M, *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus Photographs. *JAMA* 2016;316:2402–10.
11 Rajalakshmi R, Arulmalar S, Usha M, *et al*. Validation of smartphone based retinal photography for diabetic retinopathy screening. *PLoS One* 2015;10:e0138285.
12 Wilkinson CP, Ferris FL, Klein RE, *et al*. Proposed International clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677–82.
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.
14 Rajalakshmi R, Subashini R, Anjana RM, *et al*. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye* 2018;32:1138–44.
15 Ting DSW, Cheung CY-L, Lim G, *et al*. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318:2211–23.
16 Sengupta S, Sindal MD, Baskaran P, *et al*. Sensitivity and specificity of smartphone-based retinal imaging for diabetic retinopathy. *Ophthalmology Retina* 2019;3:146–53.
17 Natarajan S, Jain A, Krishnan R, *et al*. Diagnostic accuracy of community-based diabetic retinopathy screening with an Offline artificial intelligence system on a smartphone. *JAMA Ophthalmol* 2019;137:1182.
18 Tufail A, Rudisill C, Egan C, *et al*. Automated diabetic retinopathy image assessment software. *Ophthalmology* 2017;124:343–51.
19 Scanlon PH. The English national screening programme for diabetic retinopathy 2003–2016. *Acta Diabetol* 2017;54:515–25.

Agreement of a Novel Artificial Intelligence Software With Optical Coherence Tomography and Manual Grading of the Optic Disc in Glaucoma

Sujani Shroff, MD, Divya P. Rao, MD,† Florian M. Savoy, MS,‡*
S. Shruthi, MD, Chao-Kai Hsu, MS,‡ Zia S. Pradhan, MD,**
P. Jayasree, V., MD, Anand Sivaraman, PhD,§ Sabyasachi Sengupta, MD,‖*
Rohit Shetty, MD, PhD, and Harsha L. Rao, MD, PhD¶*

Précis: The offline artificial intelligence (AI) on a smartphone-based fundus camera shows good agreement and correlation with the vertical cup-to-disc ratio (vCDR) from the spectral-domain optical coherence tomography (SD-OCT) and manual grading by experts.

Purpose: The purpose of this study is to assess the agreement of vCDR measured by a new AI software from optic disc images obtained using a validated smartphone-based imaging device, with SD-OCT vCDR measurements, and manual grading by experts on a stereoscopic fundus camera.

Methods: In a prospective, cross-sectional study, participants above 18 years (Glaucoma and normal) underwent a dilated fundus evaluation, followed by optic disc imaging including a 42-degree monoscopic disc-centered image (Remidio NM-FOP-10), a 30-degree stereoscopic disc-centered image (Kowa nonmyd WX-3D desktop fundus camera), and disc analysis (Cirrus SD-OCT). Remidio FOP images were analyzed for vCDR using the new AI software, and Kowa stereoscopic images were manually graded by 3 fellowship-trained glaucoma specialists.

Results: We included 473 eyes of 244 participants. The vCDR values from the new AI software showed strong agreement with SD-OCT measurements [95% limits of agreement (LoA) = −0.13 to 0.16]. The agreement with SD-OCT was marginally better in eyes with higher vCDR (95% LoA = −0.15 to 0.12 for vCDR > 0.8). Interclass correlation coefficient was 0.90 (95% CI, 0.88–0.91). The vCDR values from AI software showed a good correlation with the manual segmentation by experts (interclass correlation coefficient = 0.89, 95% CI, 0.87–0.91) on stereoscopic images (95% LoA = −0.18 to 0.11) with agreement better for eyes with vCDR > 0.8 (LoA = −0.12 to 0.08). Conclusions: The new AI software vCDR measurements had an excellent agreement and correlation with the SD-OCT and manual grading. The ability of the Medios AI to work offline, without requiring cloud-based inferencing, is an added advantage.

Key Words: artificial intelligence, deep learning, glaucoma, glaucoma screening, glaucoma diagnostic imaging

(*J Glaucoma* 2022;00:000–000)

Glaucoma is one of the leading causes of irreversible blindness with more than 70 million cases worldwide. Projections suggest an even higher prevalence in the future.[1] Early diagnosis of glaucoma is of paramount importance to preserve as much visual function as possible. Tele-glaucoma, with digital optic disc images captured and transmitted to glaucoma specialists, has recently received a lot of interest.[2] These programs aim at facilitating early diagnosis and periodic monitoring of patients with glaucoma. Some of these models have also been found to be cost-effective with significant savings per quality-adjusted life years gained.[3]

Although the diagnosis of glaucoma is best established with multimodal testing including visual fields and optical coherence tomography (OCT), optic disc evaluation remains the cornerstone for clinical diagnosis. Most often the appearance of the optic disc arouses an initial suspicion and further testing confirms the presence of glaucoma. Therefore, optic disc photography has been at the heart of tele-glaucoma efforts globally. In addition, smartphone-based disc imaging has widened the reach of tele-glaucoma to regions where access to eye care has been very difficult.[4–6]

Technology is being increasingly mobilized for home-based diagnostics, not only for glaucoma but for other diseases such as diabetic retinopathy (DR).[6–8] This labor-intensive process of manual grading is overburdening physicians. Artificial intelligence (AI)-based software has therefore been deployed for screening images. Patients with the probable disease are automatically filtered and referred to physicians. This eases the specialist's load.[9–11] Yet, this workflow has limitations. Images may be acquired with different devices. They then need to be transmitted to a server for AI interpretation. The report is then transmitted back to the point of care and disseminated. Logistic difficulties such as a high-speed internet connection (5G or Wi-Fi), continuous electricity supply and tabletop imaging devices are all implementation challenges preventing the wide use of such cloud-based AI tele-glaucoma systems.

In this study, a portable, handheld, smartphone-based imaging device was used for acquiring fundus images. The device image quality has previously been validated in

Received for publication June 26, 2022; accepted November 19, 2022.
From the *Department of Glaucoma, Narayana Nethralaya, Rajajinagar; ¶Department of Glaucoma, Narayana Nethralaya, Bannerghatta Road; §Remidio Innovative Solution Pvt Ltd, Bengaluru, Karnataka, India; ‖Future Vision Eye Care and Research Centre, Mumbai, Maharashtra, India; †Remidio Innovative Solution Inc., Glen Allen, VA; and ‡Medios Technologies, Remidio Innovative Solutions Pvt Ltd, Singapore.
D.P.R., F.M.S., C.K.H., and A.S. are employees of Remidio Innovative Solutions. H.L.R is a consultant for Santen, Allergan, and Pfizer. The remaining authors declare no conflict of interest.
Reprints: Divya P. Rao, MD, Remidio Innovative Solution Inc., 11357 Nuckols Road #102, Glen Allen, VA 23059 (e-mail: drdivya@remidio.com).
Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.glaucomajournal.com.

DOI: 10.1097/IJG.0000000000002147

comparison to tabletop devices for the detection of DR.[7,8,12] The device is also equipped with an offline AI-based software capable of effectively detecting DR.[13,14] To the best of our knowledge, there is no other AI-based glaucoma detection software, which is integrated into a smartphone, and which can work in offline automatically. The purpose of this study was to determine the agreement between the vertical cup-to-disc ratio (vCDR) measurements acquired from this offline AI software and the values obtained from manual grading by specialists on stereoscopic images and from SD-OCT, in eyes with and without glaucomatous disc changes.

METHODS

This was a cross-sectional study conducted at Narayana Nethralaya, a tertiary eye Care center in Bengaluru, South India, between July 2021 and February 2022. The methodology adhered to the tenets of the Declaration of Helsinki and was approved by the Institute's Ethics Committee (EC Ref No: C/2021/02/02). Written informed consent was obtained from all participants. The study included all consecutive, consenting patients above 18 years of age attending the glaucoma clinic with varying degrees of glaucomatous disc damage. Patients without glaucoma were recruited from the general ophthalmology clinics and acted as controls. Participants with acute or sudden vision loss, and those deemed to have narrow angles on gonioscopy and could not be safely dilated were excluded. Similarly, patients with coexisting ocular pathologies and those with significant media opacity (eg, advanced cataracts) precluding adequate view of the disc were also excluded. In addition, participants having any condition that, in the opinion of the investigator, would preclude participation in the study (eg, uncontrolled intraocular pressure, active eye infection, <3 months post glaucoma filtering surgery, <1-month postcataract surgery, unstable medical status including blood pressure or glycemic control, photosensitivity, etc.) were also excluded.

Clinical Assessment

After recording the history and demographics, all participants underwent a complete ophthalmic evaluation including the best-corrected visual acuity, slit-lamp examination, intraocular pressure by Goldmann Applanation Tonometer, gonioscopy using a 4-mirror goniolens, and dilated fundus examination. A dilated slit-lamp evaluation was done to assess cataract status using The Lens Opacities Classification System III (LOCS III).[15] Any nuclear sclerosis/opalescence grade 3 (NS/NO3) and/or cortical cataract C4 and/or posterior subcapsular cataract P4 was categorized as "advanced immature" cataract and excluded from the study. A dilated fundus evaluation was performed. It included vCDR measurement in increments of 0.05 and identification of other typical features of glaucomatous optic disc viz. neuroretinal rim thinning, notching, splinter hemorrhages, retinal nerve fiber layer defects, and beta zone peripapillary atrophy. After dilated fundus evaluation, all patients underwent the following imaging protocol:

Imaging Protocol

All participants underwent all 3 imaging modalities by an ophthalmic photographer. A single 42-degree disc-centered image per eye was captured on the fundus on phone (FOP NM-10) device (Remidio Innovative Solutions Pvt. Ltd.). All acquired images were subjected to evaluation by the Medios AI-Glaucoma software (Medios Technologies, Remidio Innovative Solutions Pvt. Ltd) for image quality and vCDR. The image quality assessment is based on the visualization of the optic disc, surrounding nerve fiber layer and third-order vessels. If the image was of insufficient quality, the operator was alerted to take another image of better quality. The operator made a maximum of 2 attempts to get an image of sufficient quality. After image acquisition, the estimated vCDR value provided by the software was recorded for analysis. Figure 1 shows the workflow of the Medios Artificial Intelligence software.

All patients also had a single 30-degree disc-centered stereoscopic image taken using the standard tabletop fundus camera (Kowa NM WX-3D stereoscopic camera). After this, patients underwent imaging of the optic disc using an SD-OCT device (Zeiss Cirrus SD-OCT). The optic nerve head and retinal nerve fiber layer was imaged using the optic disc cube scan with the vCDR being recorded for this comparative analysis. All images 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.

Interpretation by Glaucoma Specialists

All stereoscopic images acquired using the Kowa device were further evaluated by 3 fellowship-trained glaucoma specialists (S.S., S.S., and J.P.V.). The disc and cup outlines were manually delineated to obtain the semi-automated vCDR measurement using the proprietary Kowa software installed on the device user interface. The inner margin of the scleral rim was outlined for the disc margin, and the bend of optic vessels was outlined for the cup margin. The glaucoma specialists were masked to the clinical examination details and findings from other imaging devices.

Automated vCDR Analysis on Remidio FOP Using Medios AI-Glaucoma Tool

The Medios AI-Glaucoma is a proprietary, fully automated, deep learning–based tool that runs on monoscopic fundus images captured using Remidio FOP. It provides an automated segmentation of the optic disc and cup with a vCDR measurement.

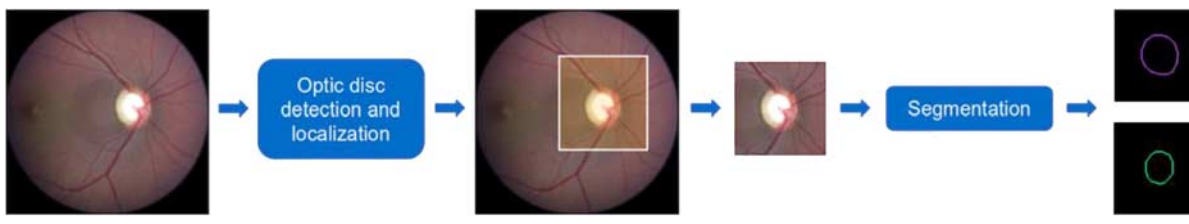


FIGURE 1. The artificial intelligence software workflow. Figure 1 can be viewed in color online at www.glaucomajournal.com.

The AI system consists of 2 deep neural networks. First, an assistive network detects the presence and the center coordinates of the optic disc. A region of interest image is then cropped around the disc. This cropped image is then fed to the main segmentation neural network, which returns the outline of the disc and the cup. The cup-to-disc ratio is finally computed with these outlines.

The main segmentation network consists of FPN (Feature Pyramid Network) architecture with a Resnet-50 backbone. It outputs a segmentation mask with 3 classes (background, cup, and disc). It was fine tuned from imagenet with a combination of disc and focal losses. Extensive data augmentation was used. This consisted of shifting, scaling, rotation, flipping, sharpening, and blurring, as well as slight modifications of brightness, contrast, hue, saturation, and value.

The network was trained using 4483 images in the train set (3700 images from South Asian population and 783 images from Caucasian population) and 560 images in the validation set. This included both dilated and undilated images. The train set had a maximum proportion of images coming from the target device, Remidio FOP (1862 images from Remdio FOP, 2621 images from 3 different standard desktop fundus cameras) and was also validated on the same device on which it was intended to be deployed. The ground truth for training was obtained by manual segmentation of the optic disc and cup by 5 fellowship-trained glaucoma specialists.

Precision Study

A repeatability (precision) substudy was conducted on 37 eyes of 33 patients with varying vCDR values. Each subject's eye in the substudy underwent the imaging protocol 3 times, imaged by 2 different operators.

Outcome Measures

The primary outcome measure was the agreement between AI software vCDR value and SD-OCT–based vCDR along with manual grading by specialists on stereoscopic images. Secondary outcomes were a correlation analysis and a pairwise assessment of the vCDR between the AI and SD-OCT, as well as manual grading. A pairwise analysis for glaucoma risk was performed in subcategories based on manual vCDR expert grading (low vCDR < 0.6, moderate vCDR ≥ 0.6 to ≤ 0.8, and high vCDR > 0.8).

Sample Size

Assuming 90% power, a precision error of 5% and the 95% limits of agreement (Bland-Altman LoA) of 0.2 in vCDR between the AI software and SD-OCT or manual grading by experts,[16] the minimum required sample size was calculated to be 465 eyes.

Statistical Analysis

All continuous variables were expressed as mean with SD or median with interquartile range. Group differences between continuous variables were analyzed using the Student *t* test for parametric or the Wilcoxon rank sum test for nonparametric distributions. Categorical variables were expressed as proportions (n, %) and group differences were analyzed using the χ^2 test.

The vCDR obtained from the AI software was compared with the other 2 modalities. The 95% limits of agreement between the vCDR from the AI software and other modalities were assessed using the Bland-Altman analysis and plotted. Interclass correlation coefficients (ICC—2-way mixed effects, absolute agreement, and multiple measurements) were also calculated and presented along with 95% CIs to understand correlations between diagnostic modalities. ICC values < 0.5 were considered as poor correlation, values between 0.5 and <0.75 were considered as moderate correlation, values between 0.75 and <0.90 were considered good correlation and values ≥ 0.90 were considered excellent correlation.[17] Correlations between vCDR from AI software and the other modalities were assessed using Pearson correlation coefficient and plotted using scatter plots with locally weighted smoothening (LOWESS) curves.

All data were entered into Microsoft Excel and analyzed using STATA 12.1 (Stata Corp). All *P*-values <0.05 were considered statistically significant.

RESULTS

We recruited 290 participants (580 eyes) for the study. Of 580 eyes, we excluded 25 eyes because of other causes of optic neuropathy, 20 eyes because of missing images from any of the modalities. In addition, 43 eyes where AI quality check failed and 22 eyes where image quality was deemed insufficient by the experts were excluded. This encompassed all eyes that failed the AI quality check as the AI has been designed to give a reliable output on an image with minimum image quality. We also excluded all eyes, which failed image quality on at least 2 modalities (AI quality, fundus image quality for grading on FOP and stereo images, SD-OCT images with signal strength <6).

We thus included 473 eyes of 244 participants in this study. Of these, the manual-based vCDR was <0.6 in 152 eyes (32%), between 0.6 and 0.8 in 273 eyes (58%) and > 0.8 in the remaining 48 eyes (10%). Table 1 shows a comparison of agreement between vCDR from the AI and the other modalities.

Comparison of vCDR Between Medios AI and SD-OCT

The vCDR from the AI software was delivered in <10 seconds. It showed strong agreement with SD-OCT measurements; and the upper and lower 95% LoA of vCDR were well within 0.2 units of the OCT. The agreement with SD-OCT got stronger with advancing vCDR (Fig. 2A). It was marginally better for eyes with vCDR > 0.8 compared to those with vCDR < 0.6 (Table 2). The overall mean absolute error was 0.02 and the AI software showed excellent correlation with the vCDR from the SD-OCT (95% CI of ICC = 0.88–0.91). A Pearson correlation analysis also showed good correlation between the AI and SD-OCT

TABLE 1. Overall Comparison of Vertical Cup-to-Disc Ratio From the Artificial Intelligence With the 2 Other Modalities

Comparison with	MAE	LL of agreement*	UL of agreement*	ICC† (95% CI)
SD-OCT	0.02	−0.13	0.16	0.90 (0.88–0.91)
Manual grading	0.03	−0.18	0.11	0.89 (0.87–0.91)

*Using Bland-Altman analysis.
†Two-way mixed effects, absolute agreement, multiple measurements.
ICC indicates interclass correlation coefficient; LL, lower limit; MAE, mean absolute error; SD-OCT, spectral-domain optical coherence tomography; UL, upper limit.

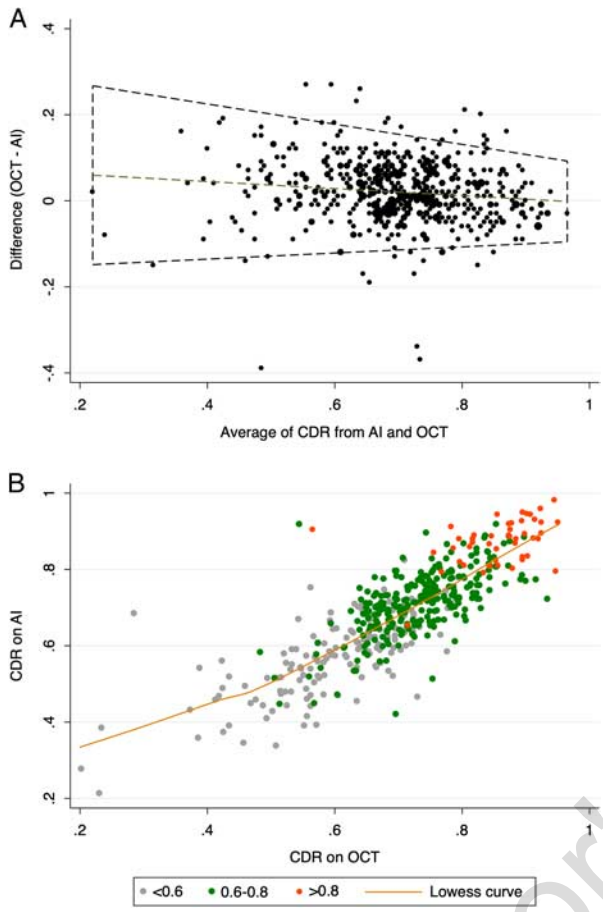


FIGURE 2. A, Bland-Altman plot showing trends in agreement with 95% CIs between CDR obtained from the OCT and AI. B, A 2-way scatter plot with a locally weighted smoothening curve showing correlation between vertical CDR obtained from the OCT and AI stratified by 3 groups based on manual grading of the vCDR. AI, artificial intelligence; CDR, cup-to-disc ratio; OCT, optical coherence tomography. Figure 2 can be viewed in color online at www.glaucomajournal.com.

(Pearson correlation coefficient = 0.82). Figure 2B depicts a 2-way scatter plot with a LOWESS curve showing correlation between vCDR obtained from the OCT and AI stratified by 3 groups based on manual grading of the vCDR.

The eye with the best agreement between AI and SD-OCT and with worst agreement is shown as in Figures 3A and B.

Comparison of vCDR Between the Medios AI and Manual Grading

On Bland-Altmann plot (Fig. 4A), the lower LoA of vCDR was within −0.18 units and the upper limit was 0.11. Similar to SD-OCT, the agreement of the vCDR between AI and manual grading was higher for eyes with vCDR > 0.8 (Table 2). The vCDR from the AI software also showed a good correlation with the vCDR from the manual grading (95% CI of ICC = 0.87–0.91). Pearson correlation analysis also revealed a good correlation between the manual vCDR grading and AI (Pearson correlation coefficient = 0.82). Figure 4B depicts a 2-way scatter plot with a LOWESS curve showing correlation between vCDR obtained from manual grading and AI stratified by 3 groups based on manual grading of the vCDR.

TABLE 2. Comparison of Vertical Cup-to-Disc Ratio From the AI With the 2 Other Modalities With Respect to Glaucoma Risk

Comparison with	MAE	LL of agreement*	UL of agreement*
For CDR <0.6 (n = 152)			
SD-OCT	0.03	−0.13	0.18
Manual grading	−0.05	−0.21	0.11
For CDR from 0.6 to 0.8 (n = 273)			
SD-OCT	0.02	−0.11	0.15
Manual grading	−0.03	−0.16	0.10
From CDR >0.8 (n = 48)			
SD-OCT	−0.015	−0.15	0.12
Manual grading	−0.02	−0.12	0.08

*Using Bland-Altman analysis.
LL indicates lower limit; MAE, mean absolute error; SD-OCT, spectral-domain optical coherence tomography; UL, upper limit.

The difference in correlation between AI and SD-OCT when compared with AI and manual grading was not statistically significant ($P = 0.24$).

Precision Study

Repeatability on 37 eyes of 33 patients imaged 3 times generated a total of 111 images of good quality. The vCDR measurements were found to have an excellent correlation, ICC was 0.98 (95% CI, 98.1–99.4). The mean coefficient of variation was 2.87. The vCDR output on 31 (84%) eyes were within 0.05 on all attempts, irrespective of the operator or repeat. Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/IJG/A672 presents the breakdown of the results by vCDR categories.

DISCUSSION

The vCDR reported by the new AI software was found to have strong agreement with vCDR measurements from SD-OCT and manual grading. The limits of agreement for vCDR between the AI and SD-OCT, as well as between AI and manual grading were well below the 0.2 units threshold, an acceptable limit even between 2 manual graders.[16] The agreement of the AI with both the modalities improved with advancing vCDR.

The Medios AI-Glaucoma is a fully automated software integrated on the same smartphone used to acquire images via the Remidio NM-FOP device. It uses deep learning technology and is trained based on ground truth provided by glaucoma specialists on a diverse dataset of several thousand monoscopic images obtained using the Remidio FOP device primarily, along with several other desktop cameras. The algorithm was trained with images from different ethnicities because there have been reports of racial and ethnic disparities in the appearance of optic nerve head, RNFL thickness, and neuroretinal rim characteristics.[18] Manual grading was the cornerstone for the AI development process. This is demonstrated by the excellent agreement between the manual and AI vCDR values. Although clinical assessment by a glaucoma specialist is standard clinical practice, it comes with an inherent limitation of subjectivity.[16,19] Hence, we also compared the AI to SD-OCT, an objective, repeatable, and well-established tool for optic disc analysis. The acceptable differences obtained against manual grading on stereo images and SD-OCT can be attributed to several factors. Different

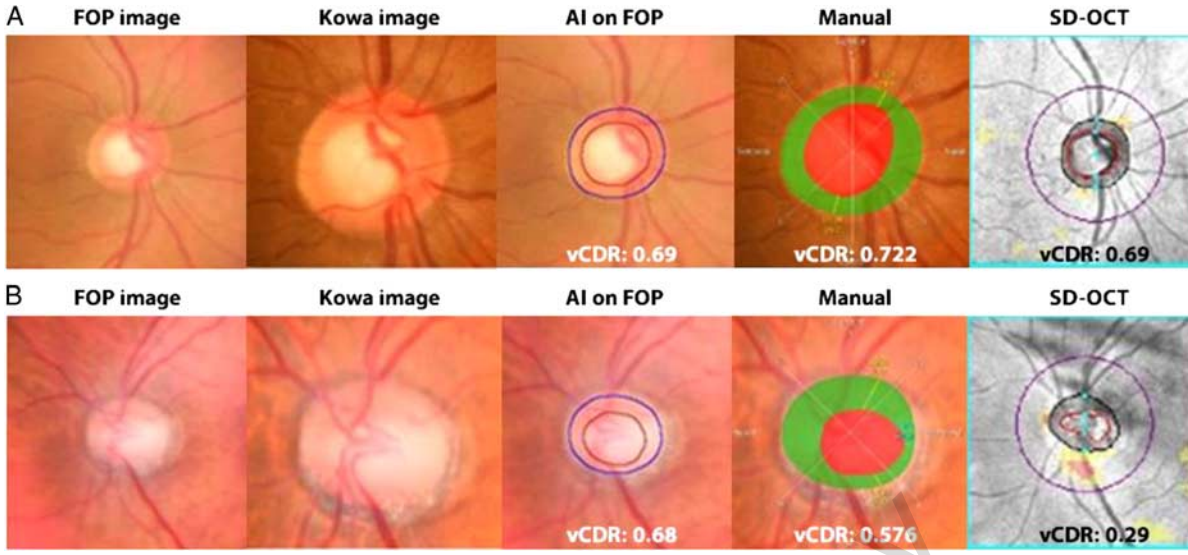


FIGURE 3. A, Reference images for comparison between AI, manual grading and SD-OCT (example for best mean absolute error between AI and SD-OCT). B, Reference images for comparison between AI, manual grading, and SD-OCT (example for worst mean absolute error between AI and SD-OCT). AI, artificial intelligence; FOP, fundus on phone; SD-OCT, spectral-domain optical coherence tomography; vCDR, vertical cup-to-disc ratio. Figure 3 can be viewed in color online at www.glaucomajournal.com.

modalities use different anatomic landmarks to measure vCDR, and the precision in identifying landmarks differs between SD-OCT, stereoscopic disc imaging, and monoscopic images. In another similar study, albeit with a much smaller sample size of 50 eyes (28 eyes with glaucoma), Varshney et al[20] also found excellent correlation (ICC = 0.86) between the vCDR obtained from the AI software and a swept source OCT device.

The repeatability of the automated measurements is a key metric to assess the utility of the tool. This is especially important in the context of screening by health care workers where more than 1 imaging attempt may be necessary to obtain an image of sufficient quality. The correlation was found to be excellent even on multiple attempts. It was well within the acceptable variability seen on SD-OCT.[21] The minor variability is explainable by the slightly different visibility of the optic disc landmarks in each image.

A multitude of software is available for automated detection of the vCDR from glaucomatous and normal eyes.[9,11,22–26] However, most of these either use a high-end desktop camera or require an online upload of images to the cloud, followed by a cloud-based AI analysis and an online report generation. To the best of our knowledge, this new AI is the only software that works offline on a smartphone fundus camera and can produce the vCDR values instantly without the need for an active internet connection. The obvious advantages of such software, in addition to its accuracy in computing vCDR, make it an excellent proposition for tele-glaucoma or AI-based screening and evaluation, especially in remote areas without any access to health care.

The AI system performed well when compared with manual grading of stereo images from a high-end desktop fundus camera. This shows that monoscopic images captured on a smartphone-based fundus camera is sufficient to make a reliable segmentation of the optic disc and cup and to estimate vCDR. To better understand how the tool might perform in the real world particularly on images of not ideal quality, we also performed a subanalysis on the exclusions (n = 65). As expected, the mean absolute error between the AI and the other 2 modalities was higher than that reported in the main analysis, but within 0.1 (vs. SD-OCT 0.08, vs. manual grading 0.1). The upper and lower 95% LoA of vCDR were within 0.21 units of the SD-OCT but higher when compared against manual grading (LLA and ULA against SD-OCT −0.21 and 0.16, respectively, against manual grading −0.19 and 0.27, respectively). This shows better agreement between the 2 objective modalities. The lower agreement with manual grading on this subset could possibly be explained by difficulty in manual human segmentation on those images with poorer image quality. This further demonstrates the robustness of the AI model.

Unlike several other systems which only output a measurement value, this system displays a visible segmentation of the disc and cup delineating the neuroretinal rim.[22–27] This offers more potential for a clinical decision assist tool to the physician. Furthermore, this tool is 1 component of a more comprehensive automated screening system for Referable Glaucoma using Remidio FOP images. The vCDR model complements another classification model. This other model reports the presence of referable glaucoma based on structural changes such as neuroretinal rim abnormalities, retinal nerve fiber layer defects, disc hemorrhages, and peripapillary atrophy. It displays areas of abnormality based on Class Activation Maps. The utility of this entire system is now evaluated in a prospective validation study that has just concluded. The tool has been developed for screening. It is not a replacement for a specialist who provides definitive services. It is meant to identify the potential undetected cases in the community and move them to the referral care pathway for confirmatory diagnosis and further management. This system caters to the requirements of remote screening and is affordable. The portable, nonmydriatic fundus camera (Remidio NM-FOP-10) is a fraction of the cost of a desktop system with a per scan cost of the Referable Glaucoma tool of under 3$ in a developing country like India.

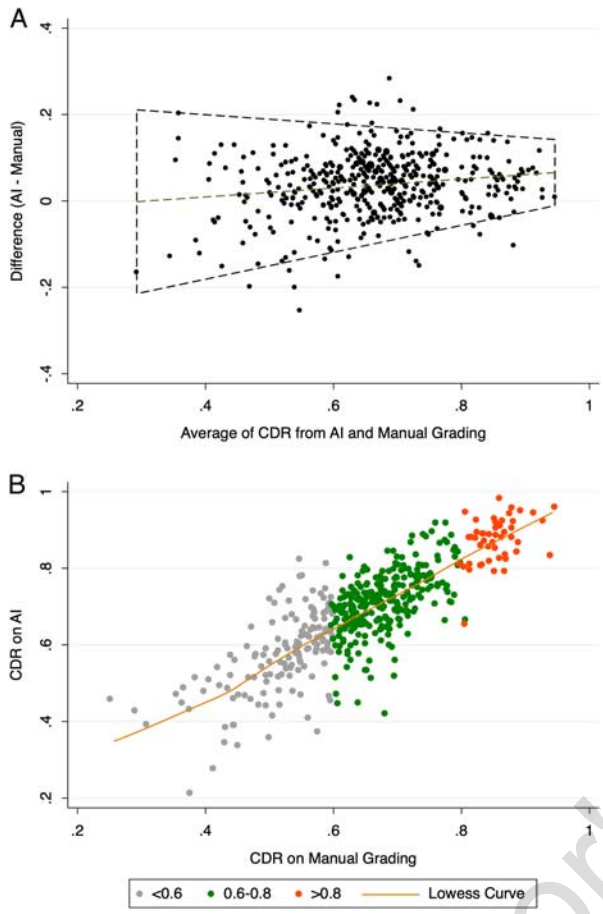


FIGURE 4. A: Bland-Altman plot showing trends in agreement with 95% CIs between CDR obtained from the Manual grading and AI. B, A 2-way scatter plot with a locally weighted smoothening curve showing correlation between vertical CDR obtained from manual grading and AI stratified by 3 groups based on manual grading of the vertical CDR. AI, artificial intelligence; CDR, cup-to-disc ratio. Figure 4 can be viewed in color online at www.glaucomajournal.com.

This study has 2 key advantages. The sample size was relatively large with a wide distribution of eyes across the spectrum of vCDR. Second, the comparison was made with 2 modalities commonly relied upon by glaucoma specialists in the clinic. One limitation of this study is its restriction to a homogenous South Asian population. Further validation will be essential in other populations to provide evidence to generalizability. As the first step, we validated the tool in a tertiary glaucoma center because of availability of relevant equipment and specialist resources. As expected, a larger proportion of cases was in the 0.6–0.8 vCDR range, having been referred to a glaucoma specialist to provide a definitive diagnosis of those with suspicious looking discs with higher vCDR. Hence, the distribution of vCDR is different from what one would expect in a population setting. The next step is a community-based validation on true distribution of vCDR on undilated images that would provide further evidence on the utility of this nonmydriatic system.

In conclusion, the vCDR from this new, offline AI software was found to have an excellent agreement and good correlation with the vCDR from the SD-OCT and manual grading by experts on stereo images. However, real-world data are required to see whether the vCDR given by this model when used in remote locations matches vCDR values obtained by clinical examinations of the same patients in clinics. Furthermore, the upcoming results of a larger validation study will provide evidence on the comprehensive glaucoma AI software which this tool is part of. It will show whether glaucoma can be accurately detected from various optic disc features in addition to this vCDR tool.

REFERENCES

1. Allison K, Patel D, Alabi O. Epidemiology of glaucoma: the past, present, and predictions for the future. *Cureus*. 2020;12:e11686.
2. Thomas S-M, Jeyaraman MM, Jeyaraman M, et al. The effectiveness of teleglaucoma versus in-patient examination for glaucoma screening: a systematic review and meta-analysis. *PLoS One*. 2014;9:e113779.
3. Thomas S, Hodge W, Malvankar-Mehta M. The cost-effectiveness analysis of teleglaucoma screening device. *PLoS One*. 2015;10:e0137913.
4. Idriss BR, Tran TM, Atwine D, et al. Smartphone-based ophthalmic imaging compared with spectral-domain optical coherence tomography assessment of vertical cup-to-disc ratio among adults in Southwestern Uganda. *J Glaucoma*. 2021;30:e90–e98.
5. Stratton S, Luna J, Roh S, et al. Smartphone-based fundus photography for remote glaucoma assessment in a low-resource setting. *Invest Ophthalmol Vis Sci*. 2021;62:1616.
6. Wintergerst MWM, Jansen LG, Holz FG, et al. Smartphone-based fundus imaging-where are we now? *Asia Pac J Ophthalmol (Phila)*. 2020;9:308–314.
7. 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.
8. Sivaraman A, Nagarajan S, Vadivel S, et al. A novel, smartphone-based, teleophthalmology-enabled, widefield fundus imaging device with an autocapture algorithm. *Transl Vis Sci Technol*. 2021;10:21.
9. Mayro EL, Wang M, Elze T, et al. The impact of artificial intelligence in the diagnosis and management of glaucoma. *Eye (Lond)*. 2020;34:1–11.
10. Li J-PO, Liu H, Ting DSJ, et al. Digital technology, telemedicine and artificial intelligence in ophthalmology: a global perspective. *ProgRetin Eye Res*. 2021;82:100900.
11. Asaoka R, Tanito M, Shibata N, et al. Validation of a deep learning model to screen for glaucoma using images from different fundus cameras and data augmentation. *Ophthalmol Glaucoma*. 2019;2:224–231.
12. Rajalakshmi R, Arulmalar S, Usha M, et al. Validation of smartphone based retinal photography for diabetic retinopathy screening. *PLoS One*. 2015;10:e0138285.
13. Jain A, Krishnan R, Rogye A, et al. Use of offline artificial intelligence in a smartphone-based fundus camera for community screening of diabetic retinopathy. *Indian J Ophthalmol*. 2021;69:3150–3154.
14. Sosale B, Sosale AR, Murthy H, et al. Medios—an offline, smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. *Indian J Ophthalmol*. 2020;68:391–395.
15. Chylack LT, Wolfe JK, Singer DM, et al. The lens opacities classification system III. The longitudinal study of cataract study group. *Arch Ophthalmol*. 1993;111:831–836.
16. Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology*. 1992;99:215–221.
17. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.

18. Poon LY, Antar H, Tsikata E, et al. Effects of age, race, and ethnicity on the optic nerve and peripapillary region using spectral-domain OCT 3D volume scans. *Transl Vis Sci Technol*. 2018;7:12.
19. Tielsch JM, Katz J, Quigley HA, et al. Intraobserver and interobserver agreement in measurement of optic disc characteristics. *Ophthalmology*. 1988;95:350–356.
20. Varshney T, Parthasarathy DR, Gupta V. Artificial intelligence integrated smartphone fundus camera for screening the glaucomatous optic disc. *Indian J Ophthalmol*. 2021;69: 3787–3789.
21. Satue M, Gavin A, Orduna E, et al. Reproducibility and reliability of retinal and optic disc measurements obtained with swept-source optical coherence tomography in a healthy population. *Jpn J Ophthalmol*. 2019;63:165–171.
22. Gonzalez-Hernandez M, Gonzalez-Hernandez D, Perez-Barbudo D, et al. Fully automated colorimetric analysis of the optic nerve aided by deep learning and its association with perimetry and oct for the study of glaucoma. *J Clin Med*. 2021; 10:3231.
23. Hatanaka Y, Noudo A, Muramatsu C, et al. Automatic measurement of cup to disc ratio based on line profile analysis in retinal images. *AnnuIntConf IEEE Eng Med Biol Soc*. 2011; 2011:3387–3390.
24. MacCormick IJC, Williams BM, Zheng Y, et al. Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile. *PLoS One*. 2019;14:e0209409.
25. Muramatsu C, Nakagawa T, Sawada A, et al. Automated segmentation of optic disc region on retinal fundus photographs: comparison of contour modeling and pixel classification methods. *Comput Methods Programs Biomed*. 2011;101: 23–32.
26. Snyder BM, Nam SM, Khunsongkiet P, et al. Accuracy of computer-assisted vertical cup-to-disk ratio grading for glaucoma screening. *PLoS One*. 2019;14:e0220362.
27. Anton A, Nolivos K, Pazos M, et al. Diagnostic accuracy and detection rate of glaucoma screening with optic disk photos, optical coherence tomography images, and telemedicine. *J Clin Med*. 2021;11:216.

The ROYAL COLLEGE of OPHTHALMOLOGISTS

ARTICLE OPEN

Check for updates

Evaluation of an offline, artificial intelligence system for referable glaucoma screening using a smartphone-based fundus camera: a prospective study

Divya Parthasarathy Rao[1] ✉, Sujani Shroff[2], Florian M. Savoy[3], Shruthi S[2], Chao-Kai Hsu[3], Kalpa Negiloni[4], Zia Sultan Pradhan[2], Jayasree P V[2], Anand Sivaraman[4] and Harsha L. Rao[2]

BACKGROUND/OBJECTIVES: An affordable and scalable screening model is critical for undetected glaucoma. The study evaluated the performance of an offline, smartphone-based AI system for the detection of referable glaucoma against two benchmarks: specialist diagnosis following full glaucoma workup and consensus image grading.

SUBJECTS/METHODS: This prospective study (tertiary glaucoma centre, India) included 243 subjects with varying severity of glaucoma and control group without glaucoma. Disc-centred images were captured using a validated smartphone-based fundus camera analysed by the AI system and graded by specialists. Diagnostic ability of the AI in detecting referable Glaucoma (Confirmed glaucoma) and no referable Glaucoma (Suspects and No glaucoma) when compared to a final diagnosis (comprehensive glaucoma workup) and majority grading (image grading) by Glaucoma specialists (pre-defined criteria) were evaluated.

RESULTS: The AI system demonstrated a sensitivity and specificity of 93.7% (95% CI: 87.6–96.9%) and 85.6% (95% CI:78.6–90.6%), respectively, in the detection of referable glaucoma when compared against final diagnosis following full glaucoma workup. True negative rate in definite non-glaucoma cases was 94.7% (95% CI: 87.2–97.9%). Amongst the false negatives were 4 early and 3 moderate glaucoma. When the same set of images provided to the AI was also provided to the specialists for image grading, specialists detected 60% (67/111) of true glaucoma cases versus a detection rate of 94% (104/111) by the AI.

CONCLUSION: The AI tool showed robust performance when compared against a stringent benchmark. It had modest over-referral of normal subjects despite being challenged with fundus images alone. The next step involves a population-level assessment.

Eye (2024) 38:1104–1111; https://doi.org/10.1038/s41433-023-02826-z

INTRODUCTION

Glaucoma is a leading cause of global irreversible blindness. The prevalence is projected to increase from 76 million in 2020 to 111.8 million in 2040 [1]. Undetected glaucoma raises the risk of blindness and as the disease advances to late stages, the treatment and care cost significantly increase, posing a financial burden. This necessitates timely diagnosis and treatment [2, 3].

Glaucoma is a progressive degeneration of the optic nerve, with loss of retinal ganglion cells, thinning of the retinal nerve fibre layer, and progressive excavation of the optic disc [4]. Manual assessment of the optic nerve head (ONH), a crucial component of glaucoma diagnosis is labour-intensive and dependent on trained specialists. Fundus photography along with technology like Artificial Intelligence (AI) can help overcome this challenge.

AI helps triaging patients and ensuring emergent cases are referred appropriately to ophthalmologists [5, 6]. Global research for the development of an automated tool for glaucoma screening using fundus images has been promising [7, 8]. However, to the best of our knowledge, this is the first study validating an offline AI system in a prospective clinical study. Additionally, algorithms have typically been developed for bulky, expensive desktop fundus camera systems. This poses several challenges to widespread adoption. Requirements for stable internet connectivity for reporting and continuous power supply are barriers to accessibility in remote areas. To overcome these challenges, a novel AI for referable Glaucoma has been integrated offline on a validated smartphone-based, portable fundus camera. It can run in seconds without the need for internet or cloud-based inferencing [9]. The purpose of this study is to evaluate the performance of this novel system in detecting referable glaucoma on monoscopic fundus images.

MATERIALS AND METHODS

A prospective, cross-sectional study was conducted at Narayana Nethralaya, a tertiary eye care centre, in South India between July 2021

[1]Remidio Innovative Solutions Inc, Glen Allen, VA, USA. [2]Narayana Nethralaya Eye Hospital, Glaucoma Services, Bangalore, India. [3]Medios Technologies Pte Ltd, Singapore, Singapore. [4]Remidio Innovative Solutions Pvt Ltd, Bengaluru, India. Remidio Innovative Solutions, Inc and Medios Technologies Pte Ltd are wholly owned subsidiaries of Remidio Innovative Solutions Pvt Ltd. ✉email: drdivya@remidio.com

Received: 11 October 2022 Revised: 27 October 2023 Accepted: 1 November 2023
Published online: 13 December 2023

and February 2022. The study adhered to the tenets of the Declaration of Helsinki and was approved by the Institute's Ethics Committee (EC Ref No: C/2021/02/02). The study included consecutive patients visiting the clinic and written informed consent was obtained from all participants. The performance of the novel AI system (Medios AI-Glaucoma, Medios Technologies, Remidio Innovative Solutions, Singapore) was evaluated. The AI is integrated on a portable, smartphone-based fundus camera (Remidio NM-FOP 10, Remidio Innovative Solutions Pvt Ltd, Bengaluru, India). The AI system was compared against two benchmarks: standard of care i.e., final diagnosis provided by Glaucoma specialists following a thorough glaucoma evaluation as well as against the majority image grading diagnosis by three glaucoma specialists.

The study included consecutive, consenting patients above 18 years of age attending the glaucoma clinic with varying degrees of glaucomatous optic disc damage. In the control group, patients without glaucoma were recruited from the general ophthalmology clinics. Normal subjects were those who either walked into the general clinic for a routine evaluation or those who were referred from other hospitals or other departments of the same hospital for a glaucoma workup. The details of the exclusion criteria are presented in Supplementary Methods Section 1.

Clinical evaluation

After recording the history and demographics, all participants underwent a complete ophthalmic evaluation including best corrected visual acuity (BCVA), slit lamp examination, intraocular pressure (IOP) by Goldmann Applanation Tonometer and gonioscopy using a 4-mirror goniolens. A dilated fundus evaluation included vertical cup-to-disc ratio (vCDR) measurement in increments of 0.05, and identification of other typical features of glaucomatous optic disc viz. neuroretinal rim thinning, notching, splinter haemorrhages, retinal nerve fibre layer defects and beta zone peripapillary atrophy. Following this, all patients underwent the imaging protocol described below by Optometrists with 1 year of experience.

Imaging protocol. A single 42-degree disc-centred image per eye was captured on the fundus on phone non-mydriatic (FOP NM-10) device (Remidio Innovative Solutions Pvt. Ltd, Bangalore, India). All acquired images were subjected to evaluation by the inbuilt image quality algorithm. The image quality assessment is based on the visualization of the optic disc, surrounding nerve fibre layer and 3rd-order vessels. If the image was of insufficient quality, the operator was alerted to take another image. The operator made a maximum of 2 attempts to get an image of sufficient quality.

Patients also underwent a single 30-degree disc-centred stereoscopic image captured on a standard tabletop fundus camera (Kowa NM WX-3D stereoscopic camera, Kowa, Japan). Following this, they underwent imaging of the optic disc using an SD-OCT device (Zeiss Cirrus SD-OCT, Dublin, CA). The optic nerve head and retinal nerve fibre layer were imaged using the optic disc cube scan.

Visual field examination (Humphrey visual field 24-2 or 10-2 programme) was performed in all new cases to establish the diagnosis of glaucoma and in confirmed cases if it was beyond 1 year since the last reliable fields.

All images were stored as JPEG files after removing patient identifiers and assigning a randomly generated unique numerical identifier linked to the participant's study ID number.

Final diagnosis

The glaucoma specialists (SS, SS, JVP) corroborated all the test results for a final diagnosis and categorized each eye into normal, glaucoma suspects, or glaucoma based on a predefined criteria [10] (Supplementary Methods Section 2). The worse eye diagnosis constituted the patient-level diagnosis. This was used as a reference standard against the binary output of the AI for referable glaucoma.

'Referable glaucoma' referred to those with glaucoma and 'No referable glaucoma' included glaucoma suspects and normal.

Fundus image quality control and grading

All the images captured using the Kowa stereoscopic camera and the FOP-NM 10 device were evaluated by three fellowship-trained glaucoma specialists (SS, SS, JVP). They were masked to the clinical examination details, investigational reports as well as each other's grading. The graders initially evaluated the quality of the images as excellent, acceptable, or insufficient based on the criteria mentioned in Supplementary Methods Section 3. Excellent and acceptable grades qualified as sufficient image quality. A predefined criterion from previous population studies was used by the specialists for making a provisional diagnosis (unlikely glaucoma, disc suspects or likely glaucoma) of glaucoma as mentioned in Supplementary Methods Section 4 [11–14]. Glaucoma severity was determined based on visual field MD as per Hodapp-Parish and Anderson criteria. Mean Deviation (MD) less than −6 dB was early, −6 to −12 dB was moderate and worse than −12 dB was defined as severe disease [15].

'Referable' glaucoma referred to those with likely glaucoma and 'No referable glaucoma' included disc suspects and unlikely glaucoma.

Automated referable glaucoma AI detection system

The AI system consists of two main components: a cup and disc segmentation model and a binary classification model. The segmentation model has been described and externally validated in a prospective study [16]. The classification model segregates images with glaucoma from suspects and normal eyes. It has been trained using 6674 images. 1813 (27.2%) were glaucoma, 1142 (17.1%) were suspects and 3719 (55.7%) were normal eyes. 4373 images (65.5%) were captured using the Remidio FOP (target deployment device), and 2301 (34.5%) using desktop fundus cameras. 5082 images (76.1%) were captured on a South Asian population, and 1592 (23.9%) on a Caucasian population. The model uses a ResNet-50 architecture and was pre-trained on the ImageNet dataset. Additionally, the datasets were carefully curated during development such that there was no overlap of patient data during training and testing. Two other assistive AI models were trained. The first is a quality check which outputs an indication of sufficient image quality for a reliable glaucoma diagnosis. The second is a disc localization model. It detects the location of the centre of the disc in the retinal image. The disc coordinates are used to crop a region of interest around the disc. This is a pre-processing step for the two main AI models (segmentation and classification algorithms). Supplementary flowchart summarizes the different elements of the AI system. This study was conducted following AI development and internal testing.

The images of all the participants were analysed using the AI tool. The AI graded the images as Referrable or No Referable Glaucoma. Referrable glaucoma included those with likely glaucoma requiring immediate referral and no referable glaucoma included disc suspects and no glaucoma. The AI also categorizes images with high VCDR (vCDR 0.7–0.85) and no other glaucomatous disc changes as 'high VCDR (disc suspect)' with a non-urgent referral to the ophthalmologist.

The primary outcome measure was the diagnostic ability of AI in detecting referable Glaucoma when compared to a final diagnosis made by a glaucoma specialist following a complete glaucoma evaluation. The secondary outcome measures were (1) diagnostic ability of the AI when compared against a majority image grading diagnosis provided by glaucoma specialists (2) comparing the image quality and diagnostic accuracy in the detection of referable glaucoma using monoscopic and stereoscopic fundus camera images and (3) repeatability analysis of the AI output.

Sample size calculation

The minimum required sample calculated to detect the sensitivity of 80% (and addressing a specificity of 80%) with a precision of 10% was 154 patients. This incorporates a 40% prevalence of referable Glaucoma and a 95% confidence level. A sample size of 200 patients was also sufficient to measure rate of discordance in referrable glaucoma between the AI software and glaucoma specialist from the true rate of discordance by ≤8% assuming a true discordance rate ranging between 10 and 50%, and sensitivity of at least 80%. We aimed for at least 250 patients for the current study assuming a 25% attrition due to incomplete tests, dropouts and quality/reliability issues from various devices.

Statistical analysis

A patient-level analysis included the diagnosis of the worse eye for the presence of referable glaucoma. A 2*2 confusion matrix was used to compute the sensitivity and specificity of the AI. Additional metrics included the likelihood ratios (LR) and accuracy along with Wilson's 95% Confidence Intervals (CI). A weighted kappa statistic (pairwise) was used to determine the interobserver agreement. Kappa of 0–0.20 was considered as slight agreement, 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 [17].

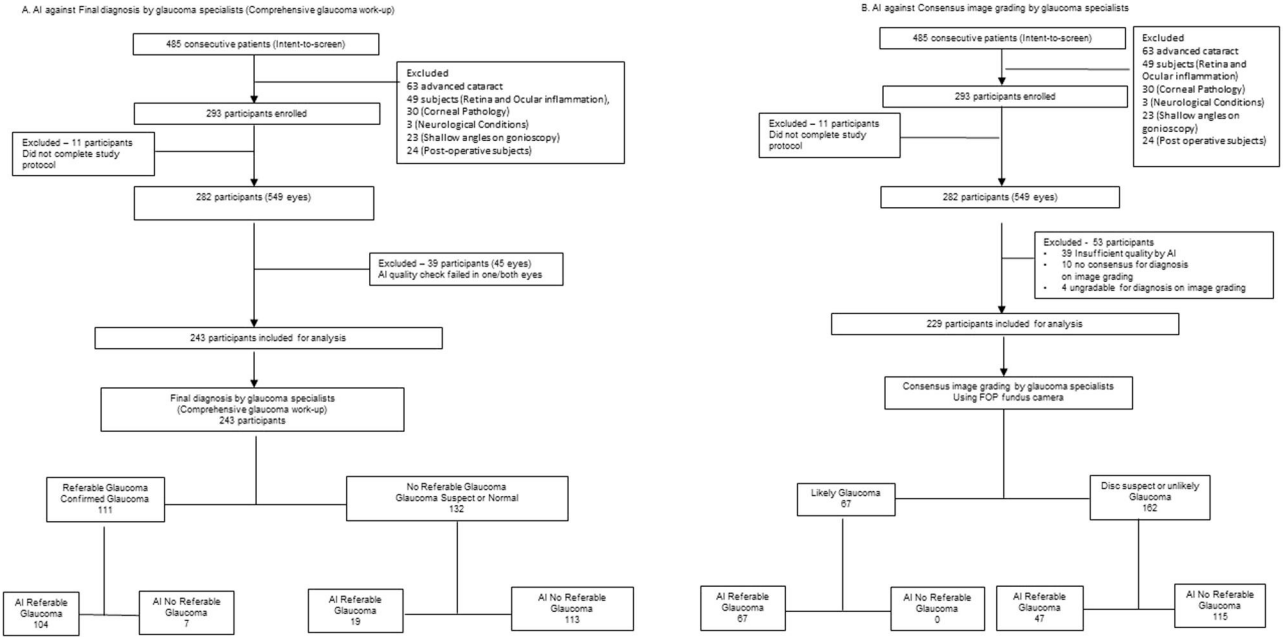


Fig. 1 Flow diagram for participant disposition in medios automated referable glaucoma detection artificial intelligence system study.

Image quality of the monoscopic and stereoscopic images was assessed on a majority grading basis as a proportion of sufficient (excellent and acceptable images) and insufficient quality images for a reliable glaucoma diagnosis. Additionally, the AI image quality algorithm was evaluated by image-ability, defined as the percentage of images determined as sufficient quality by the AI within the subset of images deemed sufficient by the graders [18]. All data was stored in Microsoft Excel and was analysed using Python 3.7, as well as the NumPy 1.21 and SciPy 1.7 libraries.

RESULTS

A total of 485 consecutive patients were screened and 293 participants were recruited. The mean age was 59 ± 12 years (range, 21, 83), 92% were greater than 40 years and 49% ($n = 144$) were female. There were 242 eyes with early to moderate cataract and 143 pseudophakia included in the study. 11 subjects were excluded as they did not complete the study protocol. Of the 282 participants (549 eyes), 39 were excluded (45 eyes) due to failed AI image quality in one or both eyes (image capture technology failure). 243 participants were included in the final analysis (Fig. 1).

Comparison of AI output against final diagnosis following a comprehensive glaucoma workup

Following a thorough glaucoma evaluation of 243 subjects, 111 subjects (45.67%), were diagnosed to have glaucoma, 56 (23.05%) were glaucoma suspects and 76 (31.28%) were normal. The AI system accurately detected glaucoma in 104 out of the 111 subjects. The sensitivity and specificity were 93.7% (95% CI: 87.6–96.9%) and 85.6% (95% CI: 78.6 – 90.6%), respectively in the detection of referable glaucoma. The true negative rate in definite non-glaucoma cases (i.e., the proportion of patients being normal on thorough glaucoma evaluation which have been correctly identified as no glaucoma by the AI) was 94.7% (95% CI: 87.2–97.9%). There were 7 (6.3%) false negative glaucoma cases (three diagnosed as disc suspect and four as normal by AI). On a closer evaluation, 4 were found to be early, 3 were found to be moderate glaucoma and none with advanced glaucoma. There were 19 (14.4%) false positive cases that included 15 diagnosed as disc suspects and 4 determined to be normal by the specialists. The performance of the AI system is summarized in Table 1. Representative outputs of correctly (True Negative and True Positive) and incorrectly (False Negative and False Positive) identified images by the algorithm along with class activations maps for the positive images are presented in Fig. 2.

Comparison of monoscopic images (FOP NM-10) vs stereoscopic images (Kowa) for image quality and agreement for glaucoma diagnosis

282 participants had a total of 549 images (15 one-eyed subjects), which were graded by three blinded, glaucoma specialists. Of these, 45 images failed AI quality check and 504 images (from 275 participants) were of sufficient quality. (Supplementary Table 1). 493/504 (97.8%) images on the FOP and 496/503 (98.6%) images on the Kowa were deemed to be of sufficient quality for a reliable glaucoma grading by the graders. Table 2 describes the details of image quality analysis between the two systems. The three specialists had consensus on 95.8 to 96.7% of the images on both systems for making a diagnosis. A pair-wise kappa analysis was between 0.72–0.74 on the FOP and 0.70–0.79 on the Kowa (Table 2).

Evaluation of the image quality AI on the FOP: 56 out of 549 FOP images received an insufficient image quality label by either the AI or the image graders or had no consensus. The graders identified 23 images as ungradable, and 4 had no consensus. Thus, 522 images were deemed to have sufficient quality by the graders. Amongst them, an additional 29 (5.6%) received an insufficient image quality from the AI. Thus, image-ability, was high at 94.4% (493/522). Supplementary Table 1 provides a summary of the results.

Comparison of AI against image grading by Glaucoma specialists on FOP NM-10 Fundus camera

Of 282 subjects, 229 were included for analysis of AI performance against image grading on FOP (Fig. 1). The specialists detected 60% (67/111) of true glaucoma cases by grading just fundus images versus a detection rate of 94% (104/111) by the AI. Table 3 details the performance of the algorithm against image grading.

Repeatability

A repeatability analysis was performed on a subset of 32 eyes. This included 15 eyes with a final diagnosis of glaucoma and 17 eyes with a final diagnosis of no glaucoma randomly chosen. Each

Table 1. Referable Glaucoma AI performance when compared against final diagnosis following comprehensive glaucoma evaluation.

			Glaucoma specialist diagnosis ($n = 243$)		
			Confirmed Glaucoma	**Glaucoma Suspects**	**Normal**
(a) Confusion matrix—AI system versus final diagnosis by Glaucoma specialists					
AI Diagnosis	Referable Glaucoma		104 (43%)	15 (6%)	4 (2%)
	No Referable Glaucoma	Disc Suspect	3 (1%)	19 (8%)	18 (7%)
		No Glaucoma	4 (2%)	22 (9%)	54 (22%)
		Total	111	56	76
(b) Confusion matrix—AI system versus final diagnosis based on Glaucoma severity (HAP criteria [15]) by the specialists ($N = 111$ confirmed glaucoma)					
			Glaucoma severity diagnosis by specialists		
			Early	Moderate	Advanced
AI Diagnosis	Referable Glaucoma		26	22	56
	No Referable Glaucoma	Disc Suspect	2	1	
		No Glaucoma	2	2	
(c) AI performance in the detection of Referable Glaucoma (Final diagnosis)					
Sensitivity			93.7% (95% CI: 87.6–96.9%)		
Specificity			85.6% (95% CI: 78.6–90.6%)		
Accuracy			89.3% (95% CI: 84.7–92.9%)		
Positive likelihood ratio			6.51 (95% CI: 4.28–9.90)		
Negative likelihood ratio			0.07 (95% CI: 0.04–0.15)		
Recall- No glaucoma			94.7% (95% CI: 87.2–97.9%)		

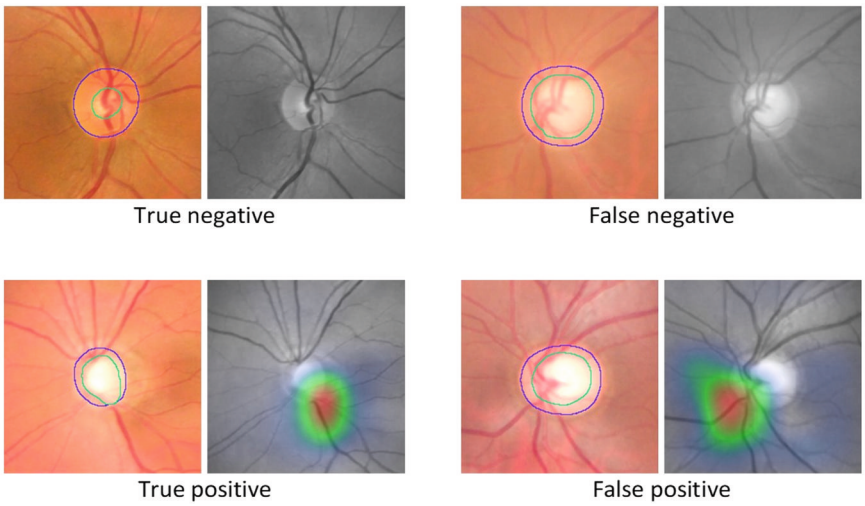


Fig. 2 Representative outputs of the AI system along with Class Activation Maps (CAMs) for the positive cases.

eye was imaged three times, with all three resulting images being fed to the AI independently. For 30/32 eyes, the output of the AI was identical amongst all three runs. The two cases with disagreements consisted of one glaucoma and one normal case. The repeatability was thus 93.75%.

DISCUSSION

An alarming trend shows more than 90% of glaucoma in the community being undetected in developing nations. Additionally, more than 50% have advanced disease and nearly 20% are blind at the time of diagnosis [19–21]. Compounding this problem is an acute shortage of glaucoma specialists. Studies in developing countries have shown that Glaucoma screening can be cost-effective [22, 23]. This necessitates a tool that leverages technologies like AI to address the inequities in screening making it effective and labour-sparing in at least the high-risk populations. Adding to the challenge is the absence of objective, standardized criteria that is universally agreed upon for diagnosing suspicious discs. This leads to subjectivity in not only the diagnosis but also the management of glaucoma suspects and early disease. We aimed to develop a novel, affordable screening tool using fundus images that can accurately identify those well-established glaucoma cases who are undetected in the community. They would benefit from immediate referral and management or would otherwise go blind. Due to the low prevalence of the disease, the algorithm was developed with the idea of maximizing the sensitivity for those with established glaucoma while maintaining a high specificity to avoid an over-referral or alarm amongst normal subjects.

Generally, structural changes in the optic nerve head (ONH) like neuroretinal rim abnormalities and enlargement of ONH excavation precede functional loss detectable on visual field assessment [4]. Hence, these morphological changes are considered early

Table 2. Comparison of monoscopic images (FOP NM-10) vs stereoscopic images (Kowa) for image quality and agreement for glaucoma diagnosis.

		Image grading by specialists	
		Monoscopic images (FOP NM-10) $N = 504$ images	**Stereoscopic images (Kowa) $N = 503$ images**
Quality of fundus images	Excellent	372 (73.8%)	413 (82.1%)
	Acceptable	121 (24.0%)	83 (16.5%)
	Total sufficient quality	493 (97.8%)	496 (98.6%)
	Insufficient	8 (1.6%)	3 (0.6%)
	No consensus	3 (0.6%)	4 (0.8%)
Consensus amongst graders on diagnosis (Patient level)	Yes	229 (95.8%)	233 (96.7%)
	No	10 (4.2%)	8 (3.3%)
Inter-grader agreement (Cohens kappa, Glaucoma diagnosis)	Ophthalmologist 1 and 2	0.72	0.70
	Ophthalmologist 1 and 3	0.74	0.76
	Ophthalmologist 2 and 3	0.73	0.79

biomarkers for glaucomatous optic neuropathy (GON). Fundus cameras capturing monoscopic colour images, red-free images or stereo images of the optic disc and RNFL have been widely used to detect structural changes and monitor glaucoma [24]. Stereoscopic imaging has better visualization of ONH morphology due to depth perception. However, these systems are large, unwieldy and expensive. In the current study, while the proportion of excellent quality images on the traditional desktop stereo camera was higher (82.1% Kowa vs 73.8% on FOP), the overall sufficient quality images for a reliable glaucoma diagnosis between the monoscopic (97.8% sufficient quality) and stereoscopic fundus camera (98.6% sufficient quality) were similar. While the specialists identified a marginally higher number of likely glaucoma cases on the stereoscopic camera (33% on Kowa vs 29% on FOP), the AI performance on the smartphone camera was unaffected when compared against imaging grading on either device. The AI correctly detected all the glaucoma cases identified by the specialists on either device (Sensitivity of AI 100% against both for image-based grading). This shows that the monoscopic fundus camera integrated with the robust AI has the potential for Glaucoma screening. It has significant public health relevance as it is easier to capture images on a portable fundus camera that is a fraction of the cost of a high-end expensive stereo fundus camera. This highlights the potential application of the AI system in a population-based setting to be used either independently or along with teleophthalmology as a clinical assist tool.

To present the accuracy of the AI system in referable glaucoma detection, we compared the AI system against two benchmarks: final diagnosis following a thorough glaucoma evaluation (standard of care) and image grading by glaucoma specialists on the same set of patients. This provides a better understanding of the reliability of image grading for glaucoma diagnosis. The AI system had a sensitivity and specificity of 93.7% and 85.6%, respectively, in comparison against standard of care. The 7 false negative cases were early (4) and moderate (3) glaucoma cases with no advanced case being missed. False positives (19 cases, 14.4%) included both disc suspects and normal cases being flagged as glaucoma by the AI. While the specificity seems relatively low, it is essential to recognize that the false positives were primarily disc suspects (15/19 cases) who would require a glaucoma workup and periodic yearly monitoring while not requiring urgent attention. This could also be attributable to a larger proportion of suspicious discs being evaluated in a tertiary centre. Interestingly, only 4 out of 76 normal subjects were considered referable glaucoma. Hence, the true negative rate in the definite non-glaucoma cases, or in other words, accurately identifying those without glaucoma was 94.7% (72/76; 95% CI: 87.2–97.9%). This is critical in a disease like glaucoma where minimal over-referral of normal subjects is pivotal to preventing overburdening of an already stretched health care system. On a closer evaluation, three of these subjects had a higher-than-average vCDR. It must be noted that at the population level, the prevalence of disease is low and hence the distribution of those with no glaucoma will be significantly higher. Hence, population-level specificity is to be evaluated in a subsequent study. Direct comparison to other global research groups is challenging due to differences in disease definitions, comparison standards, models utilized and the population in which the algorithm was validated. However, our model performed on par with other groups despite having a more difficult benchmark of comparison. Supplementary Table 2 summarizes various glaucoma detection studies using AI and Deep Learning on fundus photographs [25–33]. In the future, to improve the accuracy of the deep learning algorithm and further reduce the false negatives, more data coming from early-moderate cases along with corresponding OCT information during development will be useful.

The AI had a sensitivity of 100% for referable Glaucoma when compared against the consensus image grading of three glaucoma specialists. Inspecting the specificity of 71% (47 false positives) against image grading, we observed that 55% (26 cases) of false positives were graded as disc suspects and 21 as unlikely glaucoma by the specialists. Interestingly, 18 among these 26 cases and 10 out of 21, respectively, were diagnosed as having glaucoma on full evaluation contributing to the apparently low specificity on image grading. Overall, the specialists detected 60% (67/111) of true glaucoma cases by grading just fundus images versus a detection rate of 94% (104/111) by the AI on the same images. We hypothesize that the algorithm may have learnt, during the development phase, to identify subtle structural changes on fundus images that may not be very evident to the human eye. It shows great promise as a screening tool. However, it is important to address that this AI system cannot replace an ophthalmologist in decision-making on the final diagnosis for glaucoma. The gold standard still remains an ophthalmologist's diagnosis based on history, detailed clinical exam along with interpretation of multi-modal testing (structural and functional assessment) while excluding other causes of optic neuropathy.

Most AI algorithms require fast internet connectivity and high computational power for reporting [25, 30]. Additionally, they are developed to work on high-end, costly tabletop fundus cameras limiting their utility in resource-constrained settings [18, 34]. The current AI system utilizes lightweight deep neural network architectures that are deployed on a low-cost, smartphone-based fundus camera without compromising on efficiency or accuracy, which is a key highlight. This makes the implementation

Table 3. Referable Glaucoma AI performance when compared against image grading using FOP and Kowa fundus camera images.

			Image grading using FOP fundus camera ($n = 229$)			Image grading using Kowa fundus Camera ($n = 233$)		
			Likely Glaucoma	Disc Suspect	Unlikely Glaucoma	Likely Glaucoma	Disc Suspect	Unlikely Glaucoma
AI Diagnosis	Referable Glaucoma		67 (29%)	26 (11%)	21 (9%)	77 (33%)	24 (10%)	17 (7%)
	No Referable Glaucoma	Disc Suspect	0	13 (6%)	25 (11%)	0	12 (5%)	25 (11%)
		No Glaucoma	0	6 (3%)	71 (31%)	0	5 (2%)	73 (31%)
(b) AI performance in the detection of Referable Glaucoma (consensus image grading)								
			Image grading using FOP fundus camera			Image grading using Kowa camera		
Sensitivity			100 % (95% CI: 94.6–100%)			100% (95% CI: 95.2–100%)		
Specificity			71.0% (95% CI: 63.6–77.4%)			73.7% (95% CI: 66.3–80.0%)		
Accuracy			79.5% (95% CI: 73.7–84.5%)			82.4% (95% CI: 76.9–87.1%)		
Positive likelihood ratio			3.45 (95% CI: 2.71–4.39)			3.80 (95% CI: 2.93–4.95)		
Negative likelihood ratio			0.00			0.00		
Recall- no glaucoma			82.1 (95% CI: 74.1–88.0%)			85.2% (95% CI: 77.6–90.6%)		

of screening programmes in the outreach practical. To the best of our knowledge, it is the first offline, on-the-edge software for screening eye conditions such as Diabetic Retinopathy and Glaucoma that can give a report within a few seconds without the need for internet connectivity [35–37]. The portable design of this device with its embedded AI system makes it user-friendly and can be used by minimally trained health workers [38, 39].

Strengths of the study: This is the first prospective study evaluating an offline AI for screening referable glaucoma using smartphone-based monoscopic fundus images and showing promising performance. Additionally, the accuracy has been determined against two benchmarks: comprehensive evaluation and image grading by glaucoma specialists. The diagnostic criteria for both evaluations were standardized to lower the chance of subjective assessment. A stringent assessment against the gold standard despite the AI being presented with fundus images allows for a robust evaluation of the AI system. Adequate sample size with a good distribution of disease spectrum from no glaucoma to suspects to confirmed glaucoma ensured a thorough evaluation.

Limitations of the study: The performance of the AI has been evaluated in a South Asian population. To understand the generalizability of the model across geographies, a multi-ethnic validation will be essential. The purpose of this study was to evaluate the performance of this novel algorithm in a tertiary glaucoma centre (controlled setting) given the necessity to establish a robust ground truth with a comprehensive glaucoma work-up requiring several diagnostic modalities (clinical, structural and functional). Expectedly, the number of glaucoma and suspect cases was higher. While the results are promising, further evaluation in a real-world community setting is essential to understand whether the results can be extrapolated to a population setting with true disease prevalence, which is currently underway.

In conclusion, the novel AI integrated on a portable fundus camera can have a significant impact in screening for referable glaucoma. It can enable healthcare workers in low resource environments to screen and help break barriers to eyecare access. While this tool shows promising results, it is essential to start working towards strengthening the existing healthcare system to take on the additional burden of patients being moved into the referral care pathway. This will ensure that improved patient outcome is ultimately achieved.

SUMMARY

What was known before

- Currently, available tools are not ideal for glaucoma screening.
- Global research has found promising utility in using AI algorithms on fundus images for screening. However, they have typically been developed for bulky, expensive desktop fundus cameras with cloud-based inferencing that pose several challenges for widespread adoption.
- There is also a lacuna in terms of a prospective clinical study to validate these solutions against a gold standard diagnosis of glaucoma.

What this study adds

- A novel, offline AI deployed on a portable, affordable and validated smartphone-based fundus camera shows a robust performance in detecting referable glaucoma in a prospective clinical study.
- Comparison against gold standard diagnosis demonstrates

the true potential of the solution to triage undetected glaucoma cases to the referral care pathway.
- It holds promise for a scalable solution as it provides instant reports and overcomes several barriers associated with current technology for screening in the community.

DATA AVAILABILITY

The data can be shared upon reasonable request to the corresponding author

REFERENCES

1. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014;121:2081–90.
2. Zhang Y, Jin G, Fan M, Lin Y, Wen X, Li Z, et al. Time trends and heterogeneity in the disease burden of glaucoma, 1990–2017: a global analysis. J Glob Health. 2019;9:020436.
3. Delgado MF, Abdelrahman AM, Terahi M, Miro Quesada Woll JJ, Gil-Carrasco F, Cook C, et al. Management of glaucoma in developing countries: challenges and opportunities for improvement. Clinicoecon Outcomes Res. 2019;11:591–604.
4. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. JAMA. 2014;311:1901–11.
5. Gunasekeran DV, Wong TY. Artificial intelligence in ophthalmology in 2020: a technology on the cusp for translation and implementation. Asia Pac J Ophthalmol (Philos). 2020;9:61–6.
6. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. Clin Exp Ophthalmol. 2019;47:128–39.
7. Mayro EL, Wang M, Elze T, Pasquale LR. The impact of artificial intelligence in the diagnosis and management of glaucoma. Eye. 2020;34:1–11.
8. Mursch-Edlmayr AS, Ng WS, Diniz-Filho A, Sousa DC, Arnold L, Schlenker MB, et al. Artificial intelligence algorithms to diagnose glaucoma and detect glaucoma progression: translation to clinical practice. Transl Vis Sci Technol. 2020;9:55.
9. Varshney T, Parthasarathy DR, Gupta V. Artificial intelligence integrated smartphone fundus camera for screening the glaucomatous optic disc. Indian J Ophthalmol. 2021;69:3787–9.
10. Mariottoni EB, Jammal AA, Berchuck SI, Shigueoka LS, Tavares IM, Medeiros FA. An objective structural and functional reference standard in glaucoma. Sci Rep. 2021;11:1752.
11. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol. 2002;86:238–42.
12. Iwase A, Suzuki Y, Araie M, Yamamoto T, Abe H, Shirato S, et al. The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. Ophthalmology. 2004;111:1641–8.
13. He M, Foster PJ, Ge J, Huang W, Zheng Y, Friedman DS, et al. Prevalence and clinical characteristics of glaucoma in adult Chinese: a population-based study in Liwan District, Guangzhou. Invest Ophthalmol Vis Sci. 2006;47:2782–8.
14. Topouzis F, Wilson MR, Harris A, Anastasopoulos E, Yu F, Mavroudis L, et al. Prevalence of open-angle glaucoma in Greece: the Thessaloniki Eye Study. Am J Ophthalmol. 2007;144:511–9.
15. Hodapp E, Parrish RK II, Anderson DR. Clinical decisions in glaucoma. St Louis: The CV Mosby Co; 1993. pp. 52–61.
16. Shroff S, Rao DP, Savoy FM, Shruthi S, Hsu CK, Pradhan ZS, et al. Agreement of a novel artificial intelligence software with optical coherence tomography and manual grading of the optic disc in glaucoma. J Glaucoma. 2023;32:280–286.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
18. 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;1:39.
19. Dandona L, Dandona R, Srinivas M, Mandal P, John RK, McCarty CA, et al. Open-angle glaucoma in an urban population in southern India: the Andhra Pradesh eye disease study. Ophthalmology. 2000;107:1702–9.
20. Ramakrishnan R, Nirmalan PK, Krishnadas R, Thulasiraj RD, Tielsch JM, Katz J, et al. Glaucoma in a rural population of southern India: the Aravind comprehensive eye survey. Ophthalmology. 2003;110:1484–90.
21. Vijaya L, George R, Paul PG, Baskaran M, Arvind H, Raju P, et al. Prevalence of open-angle glaucoma in a rural south Indian population. Invest Ophthalmol Vis Sci. 2005;46:4461–7.
22. Tang J, Liang Y, O'Neill C, Kee F, Jiang J, Congdon N. Cost-effectiveness and cost-utility of population-based glaucoma screening in China: a decision-analytic Markov model. Lancet Glob Health. 2019;7:e968–78.
23. John D, Parikh R. Cost-effectiveness and cost-utility of community screening for glaucoma in urban India. Public Health. 2017;148:37–48.
24. Shabbir A, Rasheed A, Shehraz H, Saleem A, Zafar B, Sajid M, et al. Detection of glaucoma using retinal fundus images: a comprehensive review. Math Biosci Eng. 2021;18:2033–76.
25. Phan S, Satoh SI, Yoda Y, Kashiwagi K, Oshika T. Evaluation of deep convolutional neural networks for glaucoma detection. Jpn J Ophthalmol. 2019;63:276–83.
26. Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA Ophthalmol. 2019;137:1353–60.
27. Liu S, Graham SL, Schulz A, Kalloniatis M, Zangerl B, Cai W, et al. A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs. Ophthalmol Glaucoma. 2018;1:15–22.
28. Liu S, Graham SL, Schulz A, Kalloniatis M, Zangerl B, Cai W, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci Rep. 2018;8:16685.
29. Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep. 2018;8:1–9.
30. 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.
31. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318:2211–23.
32. Chakrabarty L, Joshi GD, Chakravarty A, Raman GV, Krishnadas SR, Sivaswamy J. Automated detection of glaucoma from topographic features of the optic nerve head in color fundus photographs. J Glaucoma. 2016;25:590–7.
33. Issac A, Partha Sarathi M, Dutta MK. An adaptive threshold-based image processing technique for improved glaucoma detection and classification. Comput Methods Prog Biomed. 2015;122:229–44.
34. Ipp E, Liljenquist D, Bode B, Shah VN, Silverstein S, Regillo CD, et al. EyeArt Study Group. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. JAMA Netw Open. 2021;4:e2134254.
35. Sosale B, Aravind SR, Murthy H, Narayana S, Sharma U, Gowda SGV, et al. Simple, mobile-based artificial intelligence algorithm in the detection of diabetic retinopathy (SMART) study. BMJ Open Diabetes Res Care. 2020;8:e000892.
36. Sivaraman A, Nagarajan S, Vadivel S, Dutt S, Tiwari P, Narayana S, et al. A novel, smartphone-based, teleophthalmology-enabled, widefield fundus imaging device with an autocapture algorithm. Transl Vis Sci Technol. 2021;10:21.
37. Prathiba V, Rajalakshmi R, Arulmalar S, Usha M, Subhashini R, Gilbert CE, et al. Accuracy of the smartphone-based nonmydriatic retinal camera in the detection of sight-threatening diabetic retinopathy. Indian J Ophthalmol. 2020;68:S42–6.
38. Sivaprasad S, Netuveli G, Wittenberg R, Khobragade R, Sadanandan R, Gopal B, et al. Nayanamritham Project Collaborators. Complex interventions to implement a diabetic retinopathy care pathway in the public health system in Kerala: the Nayanamritham study protocol. BMJ Open. 2021;11:e040577. https://doi.org/10.1136/bmjopen-2020-040577.
39. 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;137:1182–8. https://doi.org/10.1001/jamaophthalmol.2019.2923.

AUTHOR CONTRIBUTIONS

DRP, FMS, AS and HLR: Study conception and design, analysis and interpretation of data, manuscript drafting and revision. SS (1): Study conception and design, acquisition, analysis and interpretation of data, manuscript drafting and revision. KN: analysis and interpretation of data, manuscript drafting and revision. SS (2), ZSP and JPV: acquisition of data, analysis and interpretation of data, manuscript drafting and revision. All authors have read and approved the final manuscript.

FUNDING

The study was funded in part by Remidio Innovative Solutions, Pvt Ltd.

COMPETING INTERESTS

DPR reported being an employee of Remidio Innovative Solutions Inc. FMS reported being an employee of Medios Technologies and has a financial interest in Remidio Innovative Solutions, Pvt Ltd. C-KH reported being an employee Medios Technologies. KN reported being an employee of Remidio Innovative Solutions. AS

reported being an employee of and has a financial interest in Remidio Innovative Solutions. HLR reported being a consultant for Santen, Allergan and Pfizer outside of the submitted work. No other disclosures were reported.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41433-023-02826-z.

Correspondence and requests for materials should be addressed to Divya Parthasarathy Rao.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Glaucoma Screening Report

PATIENT DETAILS

[redacted]

SCREENING RESULT

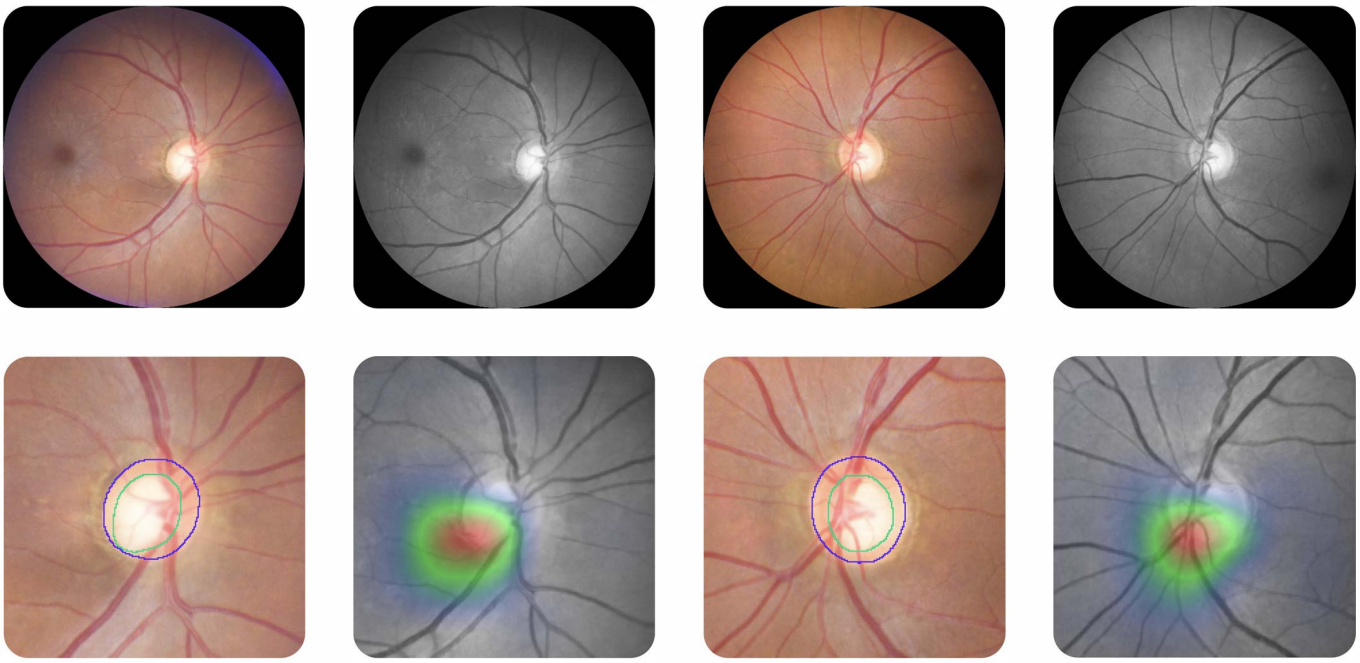
Right Eye	Left Eye
VCDR - 0.78 (Borderline High)	VCDR - 0.71 (Borderline High)

Referable Glacuoma - Refer immediately for further glaucoma evaluation

FUNDUS IMAGES

Right Eye

Left Eye



Medios AI is a physician assist software, not a replacement for an ophthalmologist's diagnosis. The results are only indicative of a high probability of Glaucoma/Suspicion of Glaucoma. This report does not screen for any medical or vision conditions apart from glaucoma. The images on this report are only thumbnails and must not be used for diagnostic purposes. Any heat maps shown are only indicative of some probable areas of abnormality.

Doctor's Signature

rem1d10 medios

Remidio Innovative Solutions Pvt Ltd
www.remidio.com