Validation of computer-aided diagnosis of diabetic retinopathy from retinal photographs of diabetic patients from Telecamps

Sheila John, MBBS, DO, PhD\(^1\), Sangeetha Srinivasan, PhD\(^2\) and Sundaram Natarajan, MBBS, DO, DSc, FRCS (Glasgow)\(^3\)

**Affiliations:** \(^1\)Department of Teleophthalmology, Sankara Nethralaya, Chennai, India; \(^2\)Vision Research Foundation, Chennai, India; \(^3\)Aditya Jyot Eye Hospital, Mumbai, India

**Corresponding Author:** *Sheila John, Email: sheilajohn24@gmail.com*

**Keywords:** algorithm; diabetes; diabetic retinopathy; retinal photographs; screening.

**Section:** Original Research

**Objective:** To validate an algorithm previously developed by the Healthcare Technology Innovation Centre, IIT Madras, India, for screening of diabetic retinopathy (DR). We validated the algorithm using fundus images of diabetic patients from telecamps to examine the screening performance for DR.

**Design:** Photographs of patients with diabetes were examined using the automated algorithm for the detection of DR.

**Setting:** We conducted mobile teleophthalmology camps in two districts in Tamil Nadu, India, from January 2015 to May 2017.

**Participants:** A total of 939 eyes of 472 diabetic patients were examined at mobile teleophthalmology camps in Thiruvallur and Kanchipuram districts, Tamil Nadu, India. The fundus photographer obtained fundus images (40–45° posterior pole in each eye) for all patients using a non-mydriatic fundus camera.

**Main outcome measures:** Fundus images were evaluated for presence or absence of DR using a computer-assisted algorithm, by an ophthalmologist at a tertiary eye care center (reference standard), and by a fundus photographer.

**Results:** The algorithm demonstrated 85% sensitivity and 80% specificity in detecting DR compared to detection by an ophthalmologist. The area under the receiver operating characteristic curve was 0.69 (95% CI = 0.65–0.73). The algorithm identified 100%
of vision-threatening retinopathy just like the ophthalmologist. When compared to the photographer, the algorithm demonstrated 81% sensitivity and 78% specificity. The sensitivity of the photographer to detect DR was found to be 86% and specificity was 99% in detecting DR compared to detection by an ophthalmologist.

**Conclusions:** The algorithm can detect the presence or absence of DR in diabetic patients in real-life settings. There were many images labeled as ungradable by the algorithm because of physiological pupillary dilation, small pupils, and age-related cataractous changes. However, all findings of vision-threatening retinopathy could be detected with reasonable accuracy. The algorithm will help reduce the workload for human graders in remote areas.

Researchers predict that there will be 439 million diabetic individuals worldwide by 2030, and they will require annual retinal evaluation.\(^1\) In the diabetic population, less than 65% undergo annual retinal examination and in the rural population, it is only 10–20%.\(^2,3\) Sheppler et al. suggested that clinicians should explain the importance of annual eye examinations to all diabetic patients, and discuss the perceived misconceptions and barriers. The common barriers are transportation, lack of awareness, and cost.\(^4\)

The concentration of ophthalmologists and paramedics in urban settings, lack of infrastructure, as well as lack of adequately trained specialists are identified as reasons for the high magnitude of avoidable blindness. In remote and rural areas in India, the ophthalmologist to patient ratio is 0.9:100,000 – indicating an acute shortage of skilled professionals for screening.\(^5,6\) The majority of ophthalmologists are trained in cataract surgery, and only 7–8% are trained in the management of diabetic retinopathy (DR).\(^7\) A lack of resources for implementing large-scale DR screening, especially in a country like India, and overcoming geographic and economic constraints mean that automated DR screening might be advantageous.\(^8,9\) A large study across 86 centers in India also reported the need for resources to fill this gap.\(^9\)

Different models have been developed for DR screening and are implemented all over the world.\(^10\) The two methods of DR screening are ophthalmologist-based and ophthalmologist-led models.\(^11,12\) The former model is an outreach camp where the ophthalmologist examines patients at the campsite and refers those with vision-threatening DR (VTDR) to the hospital for treatment. In the latter model, paramedical staff visit the venue, acquire and then transfer the images to the ophthalmologist at the base hospital. In rural India, there are only 0.3 ophthalmologists per 100,000 population.\(^13\) Given the limited number of ophthalmologists available, the latter model is advantageous as a screening tool.\(^14\)

Telemedicine helps with remote imaging of fundus photographs for VTDR which may be asymptomatic.\(^15,16\) Our eye hospital is a pioneer in mobile teleophthalmology practice in rural India.\(^17\) Camps were conducted through mobile teleophthalmology vans in rural villages where paramedical staff took non-mydriatic digital retinal images of patients with diabetes and transmitted those images via satellite or internet to the central telemedicine hub. The fundus images were then analyzed remotely by an ophthalmologist from the hospital.\(^17\) Telemedicine has been shown to be a cost-effective screening tool in the detection of DR, especially in rural and remote areas.\(^18,19\)

Teleophthalmology also addresses issues such as transportation, costs, concern over pupillary dilation, and adherence to recommended annual...
Accuracy of DR diagnosis in various telemedicine programs has been published. Severe visual loss is preventable in 90% of patients with DR by timely diagnosis and treatment.\textsuperscript{21,22}

A previously developed automated algorithm was validated in a vitreoretinal outpatient department at a tertiary eye care centre.\textsuperscript{23} In the current study, we validated the automated algorithm in a teleophthalmology setting.

**METHODS**

The study was initiated after the approval of our institutional ethics committee. Written informed consent was obtained from each patient; and the study was conducted in accordance with the Health Insurance Portability and Accountability Act and followed the tenets of the Declaration of Helsinki. Over the 2-year study period from January 2015 to May 2017, a team of well-trained and experienced optometrists and paramedical staff traveled in the outdoor, mobile teleophthalmology units to remote villages in Thiruvallur and Kanchipuram districts, Tamil Nadu, India, after obtaining permission from the head of the District Blindness Control Society. Figure 1 shows our teleophthalmology bus in a village.

Patients with type 2 diabetes, aged 35 years and above or those turning 35 years in the current calendar year, who were diagnosed by a medical practitioner or a diabetologist or diabetic by self-report were screened. Those with small or mitotic pupils, nystagmus, patients who have undergone ocular injections, or surgery for diabetic macular edema or proliferative DR, or...
any other for intraocular surgery (other than cataract surgery or laser for DR) were excluded from the study. After the campsite was set, patients were registered and then explained about diabetes and DR.

Patients underwent a comprehensive clinical examination by an optometrist including recording of case history, refraction on Snellen’s distance charts, muscle balance, cover test for distance and near, slit lamp examination, pupil reaction, and intraocular pressure measurement.

Fundus images were obtained for all diabetic patients using a non-mydriatic fundus camera (Topcon Retinal Fundus Camera TRC-NW8F, Topcon, Tokyo, Japan) by a non-certified but well-trained fundus photographer. The photographer had experience working in a DR population-based study for 5 years, acquired retinal photographs for over 5,000 patients, and he had additional experience with the fundus fluorescein angiography technique for 3 months in the retina department at the base hospital.

After 10 min of dark adaption, a single 45° digital fundus photograph centered on the macula was taken for both eyes. Under a fixed, predetermined imaging protocol, first, the right eye was photographed followed by 3 min of further dark adaption and then the left eye was photographed. If the quality of the images was found to be poor, reimaging instruction was given to the fundus photographer (Fig. 2). The room at the campsite was made dark with a dark cloth with the patient’s eyes closed, to help in physiological dilation.

THE PROCESS OF TELECONSULTATION

After the initial basic examination by the optometrist, patients requiring teleconsultation were identified. Any patient with loss of vision, diabetes or other systemic disease, trauma, ocular surgery, and any abnormal finding in the slit lamp or fundus image was referred to the ophthalmologist along with the patient electronic medical record (EMR) to the base hospital for evaluation by teleconsultation using internet connectivity (data card with laptop). All patient records were converted and stored in EMR format on the server at the base hospital.

The human graders – the general ophthalmologist and fundus photographer – read all digital fundus images. All photographs were anonymized and coded with an identification number and uploaded to a secure database. Information on the patient’s age, sex, and duration of diabetes was shared, and other details of demographic data and medical records of the patients were withheld from the readers. The reader was asked to read the images in a given order, masked to other patient details. A reader was not allowed to contact others concerning his or her reading. The retinal photographs were stored as JPEG images and viewed in a darkened room. All digital fundus images were run through the automated system and were also read by the human graders. The readers used the same computer and monitor for the grading, and were allowed to magnify and move the images, but not modify brightness or contrast. Readers were allowed to label images as ‘gradable’ based on their clinical judgment.

Over the 2-year study period from January 2015 to May 2017, patients with poor quality fundus images were referred to the base hospital for further evaluation. Patients who were referred to base hospital were noted in a special register and issued a patient identity card containing their name, village name, reason for referral, and contact number. These patients were followed up 1 week later to ensure they reported to the
hospital. The referred patients were provided treatment free of cost at the base hospital.

RESULTS
Demographics
We enrolled 939 eyes of 472 patients to test the accuracy of the algorithm to detect DR in the teleophthalmology setting. The mean age of participants was 54.5 ± 10.9 years (median = 54 years, interquartile range [IQR] = 47–61 years, range = 34–83) and 619 (66%) were men. The mean duration of diabetes in this cohort was 6.9 ± 6.2 years (median = 5 years, IQR = 2–10 years, range = 0.5–25 years).

The algorithm successfully graded 478 out of 939 possible images (51%), and diagnosed the presence of DR in 262 images. The mean image gradeability score was 0.11 ± 0.0 (median = 0.103, IQR = 0.03–0.17). The gradeability score was 0.18 ± 0.06 in eyes with gradable images compared to 0.03 ± 0.03 for those with ungradable images ($P < 0.001$, Wilcoxon test). The overall DR score was 0.37 ± 0.28 (median = 0.34, IQR = 0.12–0.57). Eyes with DR had a mean score of 0.73 ± 0.13, and those without DR had a DR score of 0.22 ± 0.17 ($P < 0.001$, Wilcoxon test).

Algorithm vs. ophthalmologist
Compared to the ophthalmologist (reference standard), 461 (49%) images were ungradable by the algorithm. Overall, the ophthalmologist found only 42 (4.4%) images to be ungradable compared to 461 (49%) images by the algorithm. There was only slight agreement in terms of image gradeability between the ophthalmologist and algorithm, Kappa = 0.019 (95% CI = -0.008–0.046). The sensitivity and specificity were both 80% for all images (gradable and ungradable combined).

Algorithm vs. ophthalmologist in only gradable images
The sensitivity of the algorithm to detect DR was found to be 85% and specificity was 80% compared to the ophthalmologist. The area under

<table>
<thead>
<tr>
<th>All images</th>
<th>Value (%)</th>
<th>95% CI</th>
<th>Only gradable images</th>
<th>Value (%)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Algorithm vs. Ophthalmologist</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>80.17</td>
<td>71.75 to 87.00%</td>
<td>Sensitivity</td>
<td>84.85</td>
<td>73.90 to 92.49%</td>
</tr>
<tr>
<td>Specificity</td>
<td>79.26</td>
<td>76.24 to 82.05%</td>
<td>Specificity</td>
<td>80.00</td>
<td>75.71 to 83.83%</td>
</tr>
<tr>
<td>Positive predictive value (PPV)</td>
<td>36.47</td>
<td>30.55 to 42.70%</td>
<td>PPV</td>
<td>41.48</td>
<td>33.07 to 50.27%</td>
</tr>
<tr>
<td>Negative predictive value (NPV)</td>
<td>96.42</td>
<td>94.67 to 97.72%</td>
<td>NPV</td>
<td>96.93</td>
<td>94.43 to 98.52%</td>
</tr>
<tr>
<td><strong>Algorithm vs. Photographer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>76.34</td>
<td>66.40 to 84.54%</td>
<td>Sensitivity</td>
<td>81.63</td>
<td>67.98 to 91.24%</td>
</tr>
<tr>
<td>Specificity</td>
<td>78.14</td>
<td>75.11 to 80.97%</td>
<td>Specificity</td>
<td>78.22</td>
<td>73.87 to 82.15%</td>
</tr>
<tr>
<td>PPV</td>
<td>28.98</td>
<td>23.38 to 35.10%</td>
<td>PPV</td>
<td>31.25</td>
<td>23.35 to 40.04%</td>
</tr>
<tr>
<td>NPV</td>
<td>96.58</td>
<td>94.87 to 97.85%</td>
<td>NPV</td>
<td>97.23</td>
<td>94.81 to 98.73%</td>
</tr>
</tbody>
</table>
the receiver operating characteristic curve was 0.69 (95% CI = 0.65–0.73) (Table 1). The algorithm detected all photographs with VTDR (100%) as identified by the ophthalmologist.

Algorithm vs. fundus photographer
Out of the 461 eyes that were deemed ungradable by the algorithm, 123 (27%) were reported to have DR. The gradeability and diagnosis of DR (presence or absence) were analyzed as two separate entities (Figure 2).

Compared to the fundus photographer, 49% of images were ungradable by the algorithm. Overall, the fundus photographer found only 50 images to be ungradable compared to 461 images by the algorithm, with only a slight agreement in terms of image gradeability between the fundus photographer and the algorithm, Kappa = 0.002 (95% CI = -0.027–0.031). The algorithm demonstrated 76% sensitivity and 78% specificity for all images, and 81% sensitivity and 78% specificity in gradable images. The sensitivity of the photographer to detect DR was found to be 86%, and specificity was 99% in detecting DR compared to the ophthalmologist.

DISCUSSION
The algorithm was assessed for the presence or absence of DR, and the gradeability of the images. There was no difference in image gradeability of the algorithm based on the VTDR status of the eye as graded by the ophthalmologist. Out of the 461 (41%) eyes that were deemed ungradable by the algorithm, 123 (27%) were reported to have DR. The presence or absence of DR and the gradeability of the images were examined as two separate streams

Figure 2—Fundus photography at the campsite.
independent of each other, thus presenting the possibility of an eye with DR being ungradable.

Compared to the fundus photographer, 461 (49%) images were ungradable by the algorithm. The sensitivity of the algorithm to detect DR was found to be 82% and specificity was found to be 78% in detecting DR compared to fundus photographer.

There was no difference in image gradeability of the algorithm based on the VTDR status of the eye as graded by the photographer. The sensitivity of the photographer to detect DR was found to be 86% and the specificity was 99% in detecting DR compared to ophthalmologist.

In our study, the algorithm demonstrated 76–84% sensitivity and 78–80% specificity in the automated detection of DR, which is acceptable. However, in our study, the algorithm interprets many images as ungradable. The area under the curve is less than 70% even if only gradable images are considered. Some of the fundus images were found to be of sub-optimal quality, as an outcome of being captured by inducing physiological dilation of pupils by exposing patients to darkness, and by avoiding use of any pharmacological mydriatic.

Additionally, Indian eyes have a darker iris and smaller basal pupillary diameter, greater incidence of cataract, and probably more VTDR with vitreous hemorrhage, which can also negatively influence the performance of the algorithm. Overall, we believe that the differences in proportion of mydriatic images, cameras used to acquire images, controlled settings versus outreach settings, and greater proportion of cataract played a role in the inferior performance of our algorithm compared to the deep machine-learning algorithm reported by Gulshan et al. 24,25 A high rate of ungradable images is identified in non-mydriatic versus

\[Figure\ 3\—(A)\ and\ (B)\ show\ a\ patient\ with\ VTDR\ correctly\ identified\ by\ the\ algorithm.\ (C)\ and\ (D)\ show\ patient\ images\ with\ cataract\ ungradable\ by\ the\ algorithm.\]
mydriatic fundus imaging. Ungradable images must be included in the statistical analysis as positive findings. Readers were required to label images as gradable or ungradable based on their clinical judgment, and poor-quality fundus images were referred to the base hospital for further evaluation, investigation, and treatment.

Studies have developed and tested automated algorithms in the past using new retinal images and existing public datasets such as EyePACS (Santa Cruz, CA) and Messidor (Saint-Contest, France), and the performances of these algorithms are higher than in our study.24–27 First, we believe that the results reported by Gulshan et al. were after 5–7 years of beta testing of the product involving 128,175 images in the developmental phase.25 Our algorithm is still in the beta testing phase and has already yielded about 80% sensitivity, though specificity is lower. Another major difference between our results and those reported by Gulshan et al. is that more than 40% of the images from the EyePACS and Messidor datasets were acquired after pupillary dilatation. In contrast, in our study, more than 90% of our images were non-mydriatic retinal images.25

In another study by Tufail et al., sensitivity and specificity of four different automated image analysis software programs were studied on 102,856 images in the UK.27 Since all of these were based on image analysis and not deep machine learning, it may be more appropriate to compare their results with ours. In the Tufail et al. study, the authors found a much higher sensitivity and specificity (>90%) using the EyeArt (Eyenuk Inc., Woodland Hills, CA) and Retmarker (Coimbra, Portugal). The gradeability reporting also shows superior results compared to our outcomes. We also found nearly 80% sensitivity and specificity in the mobile camps using non-mydriatic tabletop cameras. Major differences in study designs could have contributed to the differences in results as well. First, the UK study used only mydriatic images, whereas we used non-mydriatic images in all of our settings. Indian eyes are known to have smaller pupils in scotopic conditions, limiting the image quality and thereby compromising the software’s assessment capabilities. Second, Tufail et al. used images obtained inside ophthalmology clinic settings and images were obtained by trained technicians. We acquired images in outreach camps where lighting is not entirely under our control, but where technicians are trained fundus photographers. Lastly, all the differences mentioned above in comparing our study with the algorithm reported by Gulshan et al. are applicable in this case as well.

In the outreach camps, there is no dark room and, therefore, only a makeshift dark environment can be arranged for imaging. In addition, electricity shortages are common; hundreds of people are screened on a single morning, and many patients above the age of 50 years have media haziness because of cataract. New ways to structure workflow, from image acquisition to image analysis will be needed. To improve retinal imaging and to decrease the number of ungradable images, there is a need for a fundus camera with the following features: lightweight, non-mydriatic, inexpensive, portable, with no alignment issues, easy user-interface and image transfer features, and high sensitivity and specificity in detecting VTDR.

In a recent landmark paper on the current state of teleophthalmology in the United States, Rath et al. described applications of teleophthalmology in many diseases including DR.28 They mention the upcoming role of automated DR screening using various algorithms to ease the human burden on manual DR screening. They concluded by saying that although the findings are...
encouraging, further work remains to improve the clinical validity of these algorithms. The authors also stated that given the increasing prevalence of DR, the emergence of automated screening serves as a promising tool to address this public health issue.

In our experience in telecamps, we also found that trained fundus photographers were able to detect the presence/absence of DR and identify VTDR with satisfactory agreement with ophthalmologists. This is encouraging because we can consider reporting from photographers in the outreach camps without having to transfer images to the base hospital for DR detection. This will enable screening in very remote areas without internet connectivity.

CONCLUSIONS
Our novel software showed acceptable sensitivity and specificity in teleophthalmology settings, though improved results would be beneficial in improving predictive value and reducing unnecessarily excessive referrals. The main areas that require additional work are the reduction in ungradable images, and improving the agreement of gradeability and DR status with human graders.

ACKNOWLEDGEMENTS
We acknowledge the contributions of Ms Keerthi Ram and Dr Mohanasankar Sivaprakasam, Healthcare Technology Innovation Centre, IIT Madras, Chennai, India.

Contributor contributions: SJ – involved in conceptualization of the study; SJ, SS, SN – involved in data analysis; SJ, SS – involved in the drafting of the manuscript; SJ, SS, SN – involved in interpretation of results and in critical review of the manuscript.

Ethics approval: Approval was provided by the ethics committee of Vision Research Foundation, Sankara Nethralaya, Chennai, India. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Conflict of interest and funding: The authors declare that there are no conflicts of interest. The authors have not received any funding or benefits from industry or elsewhere to conduct this study.

REFERENCES
7. Resnikoff S, Felfch W, Gauthier TM, Spivey B. The number of ophthalmologists in practice and training worldwide: a


Copyright Ownership: This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, adapt, enhance this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0.