Evaluation of an Artificial Intelligence System for Detecting Diabetic Retinopathy in Community Healthcare Centres

Zia Saquib, PhD
BioScan AI Innovative Solutions

Amit Bhondve, MD
Diabease Diabetes and Thyroid Clinic, Ulwe, Navi Mumbai

Vishal Karande, MBBS
King Edward Memorial Medical College, Mumbai

Abstract:

In this paper we present a cross-sectional diagnostic study to evaluate the effectiveness of an Artificial Intelligence based System (Drishti AI) for detecting and grading diabetic retinopathy. The study spans Indian population across multiple centres in a community setting. Participants with a previous diagnosis of diabetes from three healthcare centres and two diabetic clinics were enrolled in the study. Single-field colour fundus photography was obtained and analysed by Drishti AI grading system and independently by two certified diabetic retinopathy graders. If there was a conflict between the two independent certified graders, then grading by the senior grader is accepted as the ground truth. Primary outcome measures include the sensitivity, specificity, positive predictive value, and negative predictive value with 95% confidence intervals (CIs) of the AI system in detecting DR and diabetic macular oedema (DME). These outcome measures are relevant in evaluating AI based systems in diagnostics.

Introduction:

Diabetic retinopathy (DR), if untreated, leads to progressive visual impairment and eventual blindness. Timely identification and referral to ophthalmologists could reduce blindness and disease complications. Those with poorly controlled diabetes should be screened for DR at least annually, and monitored by the treating physician; however, less than half of such patients receive screening. Screening currently requires referral to an eye specialist, and patients may not visit the specialist because of logistical barriers, cost of the visit, or lack of an eye specialist in their community. One method of improving access to DR screening is for primary care practices to obtain colour fundus images and send these to ophthalmologists or optometrists for reading. While such programs increase screening rates, there are logistical barriers, costs, scalability and time delays in having the images read by ophthalmologists or optometrists, given that there are only about 25,000 ophthalmologists in throughout the country (as on Jan 2023).

These limitations have driven interest in computer assessment of images through automated retinal image analysis systems (ARIASs). Such a system would decide in real time whether a patient needs referral and could potentially be much cheaper than having eye experts conduct screening. Several studies have used repositories of retinal images to test the performance of AI-based ARIAS grading systems in detecting DR and in April 2018 the US Food and Drug Administration approved an AI algorithm, developed by IDx/DR, used with Topcon Fundus camera (Topcon Medical) for DR identification.

Despite enthusiasm about the potential of AI-based grading systems, to our knowledge, there have been quite few evaluations of the performance of an AI system in a real-world multi-centre clinical setting within India. In this pilot study, we describe the performance of an AI system in a primary care practice.

Methods:

The study design and patient information and informed consent forms for study participants were part of a post-graduate thesis programme and supervised by a faculty member in the department of community medicine, the King Edward Memorial Hospital and Seth Gordhandas Sunderdas Medical College, Mumbai, India. Patients provided their written informed consent. We conducted the trial according to the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline.

AI-based grading system for DR:

Our AI system is based on deep-learning models for DR. It was developed and evaluated based on manually outlined pathologies using colour fundus images from several training data sets (altogether 92,000 images) including APTOS, Messidor and EyePACS datasets. The model was trained using the images from the 3 data sets. The deep learning model adopted a deep convolutional neural network architecture. We used the convolutional neural layers from a state-of-the-art deep-learning architecture as our base model and connected it to our customized top layers with several fully connected layers for the purpose of DR image classification. By applying transfer-learning techniques, the model training process includes the following steps: (1) manually classify the selected image data into 2 categories, DR disease and no DR disease; (2) divide the categorized image data into 2 parts, training data set (80% of the total) and test data set (20%) and keep the balance of the 2 categories in each data set; (3) normalize all the images and resize them to the dimension of 224 × 224 pixels; (4) load the pre-trained base model weights and use the training data set to train the top model initially; (5) use the training data set to retrain the whole model; and (6) monitor the accuracy and loss on the training data set and test data set and achieve the best model. The manual rule-based process adopted selection criteria results in 3 outcomes: (1) a binary identification of disease or no disease for clinically significant DR, (2) identification of specific pathologies (Micro-aneurysms, haemorrhages and exudates) related to DR, and (3) using regression technique calculate the severity of DR based on the International Clinical

Diabetic Retinopathy Disease (ICDRD) Severity Scale criteria. The AI system is designed to be compatible with most retinal imaging cameras but for the purpose of this study, the system was integrated with Intuvision Prime, a fundus camera commercially available for capturing non-mydriatic retinal images.

Severity level 2 and 3 were graded as more than mild DR (mTmDR) and severity level 4 (PDR or any DR with presence of DME) was graded as vision threatening DR (vTDR), consistent with the medically acceptable nomenclature.

An image quality control system was also developed that used deep-learning techniques to check the quality of the images. We classified selected images from the data sets into 2 classes: good/acceptable image quality for DR grading and unacceptable image quality for DR grading. Then we used the only good/acceptable quality images to train the convolutional neural network model. The trained model on the above dataset exhibited the following metrics:-

	mtmDR (95% CI)	vtDR (95% CI)
Sensitivity	92.85	94.85
Specificity	90.20	82.46
Accuracy	90.50	89.22

DR = Diabetic Retinopathy
Dataset: APTOS, Messidor and EyePACS with train-test split of 80% training, 20% test
mtmDR= more-than-mild DR, vtDR = vision threatening DR

Table 1:

Deployment in Primary Healthcare and Private Clinic:

We deployed the AI system sequentially for 4 weeks (between June 15, 2022, and July 15, 2022) at the Department of Community Medicine at King Edward Memorial Hospital, one private Diabetic clinic in Ulwe, one in Chennai, one Urban healthcare centre at Malad (Mumbai) and one community centre at Dadar (Mumbai).The study employed one diabetologist, one junior resident doctor and 2 healthcare workers. The tele-retinal screening software system included a colour fundus camera with image capturing software (Intuvision Prime), and the AI inference engine (Drishti AI) running on the AWS cloud computing infrastructure.

We trained 4 members from our team as screening operators to use the Intuvision fundus camera and image capturing software along with AI grading system. All patients suspected with Type 1 or Type 2 diabetes mellitus associated with the primary care clinics were invited to participate in the study. Macula-cantered images were acquired for each patient and up to 2 images per eye were allowed depending on the image quality (confirmed by quality control software). After completing the imaging process, the system sent each patient's information and related images to the AI-based grading system. The AI based grading system provided a DR report of referable (with severity level depicted as coefficients) or non-referable to the diabetologist or primary care physician via an email/ and portal within a period of 60 seconds. Patients with moderate (mTmDR) or severe (vTDR) were referred to an

ophthalmologist or eye clinic immediately. All images were also sent to an archival storage to be referred to a retinal specialist grader for evaluation for tele-retinal system.

Statistical Analysis:

The binary classification (referable or non- referable) by the team of two certified DR graders was used as the reference or ground truth. If there was disagreement between the two certified graders then the image was sent to a retinal specialist doctor for adjudication. The grade provided by the ophthalmologist was treated as ground truth. These images with the ground truth so obtained were compared with the grading obtained from AI-based DR grading system. The sensitivity of the disease grading was derived as true-positive/(true-positive + false-negative), specificity was true-negative/(true-negative +false-positive), positive predictive value was true-positive/(true-positive + false-positive), and negative predictive value was true-negative/(true-negative + false-negative).

Results:

In this study, 445 individuals signed informed consent forms and 443 participants completed the study according to the protocol. Of the participants, 63.88% were female .According to the reference standard ICDRD, of the 848 eyes, 582 (68.63%) had no DR, 33 (3.89%) had mild NPDR, 121 (14.27%) had moderate NPDR, 68 (8.02%) had severe NPDR, and 44 (5.19%) had PDR. For Diabetic Macular Edema (DME), 56 (6.60%) eyes were diagnosed with CSME by the certified graders. The AI system successfully graded 440 individuals, with an image ability of 98.06% (8 eyes were judged as unrecognizable by the AI system due to the presence of signs such as vitreous haemorrhage that reduced the image clarity). Further analyses of the sensitivity, specificity, PPV, and NPV for the AI system to detect mtmDR, and vtDR are shown in **Table 2**.

	mtmDR (95% CI)	vtDR (95% CI)
Sensitivity	85.85	80.85
Specificity	80.20	98.46
Accuracy	80.50	98.22

DR= Diabetic Retinopathy, mtmDR= more-than-mild DR, vtDR= vision threatening DR

Table 2:

DME is one of the most frequent vision-threatening treatable complications of DR; the sensitivity and specificity for the AI system to detect DME are shown in **Table 3**.

Sensitivity	92.03
Specificity	90.12
Accuracy	90.65

DME= Diabetic Macular Edema

There were several factors that led to the false-positive results. Some retinal images had drusen that were similar in appearance to exudates. Other false positives were

driven by dirty lens reflections or uneven light exposure at the rim of images. The AI system incorrectly identified exudates that were sheen reflections around the optic disc, the papillomacular area, and the macula.

Discussion:

We evaluated the performance of an AI system that reads retinal images to identify DR in a real-world clinical setting. The system was successfully deployed and detected patients with moderate and severe DR requiring referral. Though there was a limited sample size, the AI system was effective in ruling out disease. However, the system had a moderate rate of false-positives. The specificity of the deployed system is similar to our prior validation using a database of retinopathy images and similar to other AI systems for reading retinopathy images. The moderate rate of false-positives was driven by the low incidence of disease. Prior validations of AI systems for identifying DR have used data from retinal image databases, and images were preselected such that the incidence of disease was much higher (roughly 1 of 3). The sensitivity, specificity and the AUC for the trained model with public domain images was 92.9%, 90.8% and 93.8% respectively. On average, when the disease incidence is lower, the positive predictive value will also be lower. This is consistent with other screening programs where false-positives are common, such as mammograms. The low incidence rate of DR we observed in our study is the norm in primary care; therefore, false-positives are likely to be an issue unless the specificity of our system or other systems is much higher.

Despite these limitations, we believe the AI system has potential for improving the efficiency of screening for DR in primary care. Roughly 92% of all patients were informally told at their primary care practice they had no DR and therefore no referral was needed. In this case, the number of patients that would have to be reviewed by an ophthalmologist was less than 10%. The ability to provide real-time eye screening at familiar primary care physician practices has many practical advantages, including comprehensive chronic disease management at a single location for patients with diabetes. There is also the potential for the AI system to be improved. Further training of the AI system to differentiate drusen, sheen reflections, and exudates is part of continuous training of the model.

Acknowledements

The authors are grateful to The Jio Institute, Ulwe, Navi Mumbai and in particular Ranjan Kumar, Soumya Shrivastava, Bharat Varyani, Aishwarya Sharma, Sheetal Jain, Ronak Dhedia and Somx Gupta for their help in the preparation of data for this paper and the accompanying field work. This study was part of a post graduate thesis and was supported by KEM Hospital, Urban Community Health Centre, Malvani, Mumbai.