

Validation of Deep Convolutional Neural Network-based algorithm for detection of diabetic retinopathy – Artificial intelligence versus clinician for screening

Payal Shah, Divyansh K Mishra, Mahesh P Shanmugam, Bindiya Doshi, Hariprasad Jayaraj¹, Rajesh Ramanjulu

Purpose: Deep learning is a newer and advanced subfield in artificial intelligence (AI). The aim of our study is to validate a machine-based algorithm developed based on deep convolutional neural networks as a tool for screening to detect referable diabetic retinopathy (DR). **Methods:** An AI algorithm to detect DR was validated at our hospital using an internal dataset consisting of 1,533 macula-centered fundus images collected retrospectively and an external validation set using Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology (MESSIDOR) dataset. Images were graded by two retina specialists as any DR, prompt referral (moderate nonproliferative diabetic retinopathy (NPDR) or above or presence of macular edema) and sight-threatening DR/STDR (severe NPDR or above) and compared with AI results. Sensitivity, specificity, and area under curve (AUC) for both internal and external validation sets for any DR detection, prompt referral, and STDR were calculated. Interobserver agreement using kappa value was calculated for both the sets and two out of three agreements for DR grading was considered as ground truth to compare with AI results. **Results:** In the internal validation set, the overall sensitivity and specificity was 99.7% and 98.5% for Any DR detection and 98.9% and 94.84% for Prompt referral respectively. The AUC was 0.991 and 0.969 for any DR detection and prompt referral respectively. The agreement between two observers was 99.5% and 99.2% for any DR detection and prompt referral with a kappa value of 0.94 and 0.96, respectively. In the external validation set (MESSIDOR 1), the overall sensitivity and specificity was 90.4% and 91.0% for any DR detection and 94.7% and 97.4% for prompt referral, respectively. The AUC was .907 and .960 for any DR detection and prompt referral, respectively. The agreement between two observers was 98.5% and 97.8% for any DR detection and prompt referral with a kappa value of 0.971 and 0.980, respectively. **Conclusion:** With increasing diabetic population and growing demand supply gap in trained resources, AI is the future for early identification of DR and reducing blindness. This can revolutionize telescreening in ophthalmology, especially where people do not have access to specialized health care.

Key words: Deep convolutional neural networks, diabetic retinopathy screening, machine learning, validation of artificial intelligence

India has one of the highest prevalence of diabetes in the world with approximately 79 million people predicted to have diabetes by 2030.^[1] The prevalence of diabetic retinopathy (DR) in Indian diabetics has been reported to be around 18^[2]–27%.^[3] This emphasizes the need for proper diabetic screening to reduce the burden of DR-related blindness. However, paucity of trained retinal specialists in India limits effective screening of asymptomatic patients, thereby resulting in patients presenting late with advanced diabetic eye disease.

Fundus photograph-based DR screening in lieu of physical screening can be performed using manual grading of fundus images by trained graders or retina specialists. The technology of machine-based learning to detect DR has given a new horizon for DR screening and is improving rapidly. Its application in diagnosing referable DR patients would have a great impact

Department of Vitreoretina and Ocular Oncology, Sankara Eye Hospital, Bengaluru, Karnataka, India, ¹Lebcare Technologies Pte. Ltd, Singapore

Correspondence to: Dr. Payal Shah, Department of Vitreoretina and Ocular Oncology, Sankara Eye Hospital, Kundalahalli Gate, Bengaluru - 560 079, Karnataka, India. E-mail: n.payalshah@gmail.com

Received: 25-May-2019

Revision: 14-Aug-2019

Accepted: 10-Sep-2019

Published: 20-Jan-2020

in reducing the blindness burden. Several studies have shown that a 50° posterior pole fundus image alone can be used as a screening tool to identify DR.^[4,5]

The field of artificial intelligence (AI) has made tremendous strides since 1950s, with the emergence of machine learning by 1980s and deep learning by 2010. Machine learning is a subfield of AI, which does not require the program to perform a specific task. It recognizes the patterns and learns to predict automatically. Deep learning is a newer and advanced subfield in machine learning, which taps into neural networks and simulates the human brain in decision-making. It requires a huge database for training. A deep neural network consists of

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

Cite this article as: 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;68:398-405.

Access this article online

Website:

www.ijo.in

DOI:

10.4103/ijo.IJO_966_19

Quick Response Code:



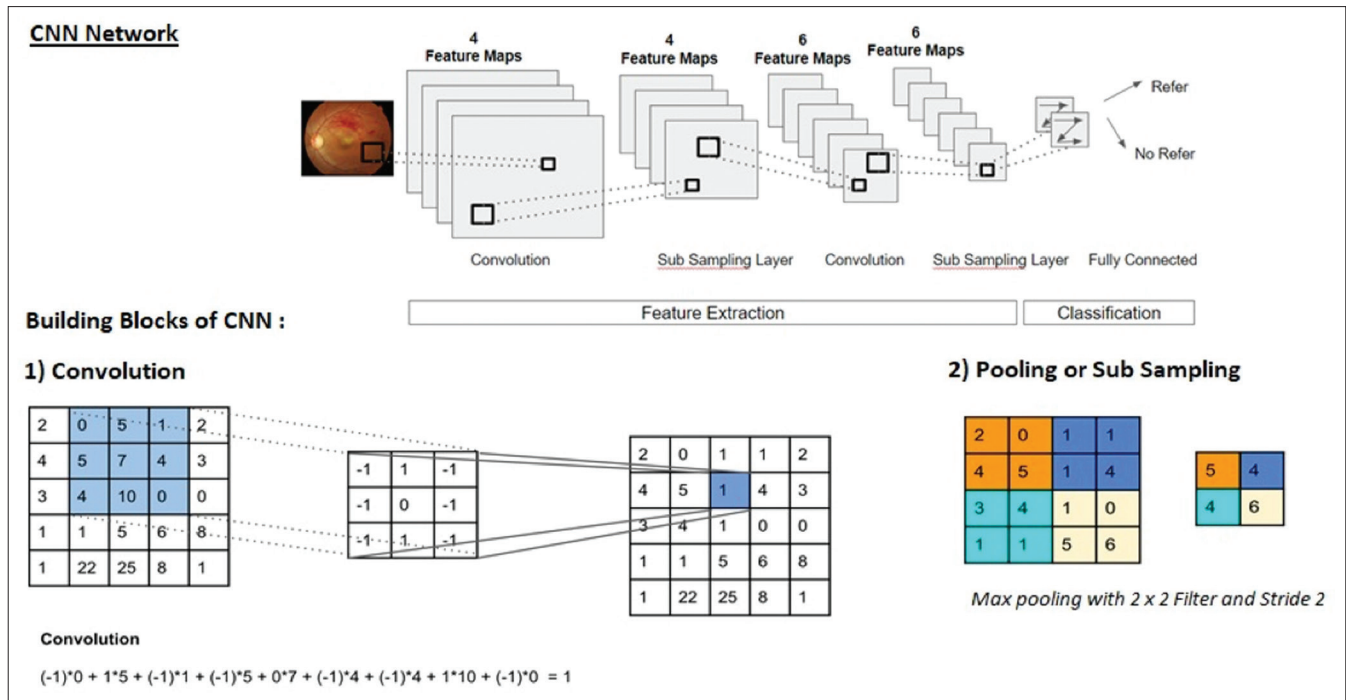


Figure 1: Typical convolutional network showing the sequence of transformations – convolution and pooling

convolutional layers and pooling layers as shown in Fig. 1. To describe in detail about convolution and pooling is beyond the scope of this article and suggested further reading can be found in the article by Escontrela.^[6]

Recent review articles on AI based on DCCN in DR by Li et al.^[7] and Ting et al.^[8] have also concluded that AI shows a great promise in ocular disease screening, both efficiency-wise and affordability-wise.

The aim of our study is to validate a machine learning algorithm developed based on deep convolutional neural networks (DCNNs) as a tool to detect referable DR to enable accessible, affordable, and accurate screening in the fight against avoidable blindness. We herein present our experience in validating such algorithm to screen fundus images and analyze its efficiency in detecting referable DR.

Methods

Approval for the study was obtained from the institutional review board. It was a retrospective exploratory non-interventional study carried out between April 2017 and August 2018. We conducted the trial according to the Standards for Reporting of Diagnostic Accuracy reporting guidelines.

Training of AI algorithm

This algorithm underwent a series of training modules, i.e. V1 (first version), V2 (second version), and V3 (third version). The training was started with 80,000 fundus images of diabetic patients in V1 and gradually, the training was improved by feeding more images in V2 (96,500) targeted to detect microaneurysms and hard exudates. The final version V3 used 112,489 images.

The four stage Convolutional Neural Network (CNN) classifiers were trained and optimized to

- Stage 1: Detect retinal photographs from nonretinal images

- Stage 2: Detect generic quality distortion for automated image quality assessment
- Stage 3: Detect DR stage
- Stage 4: Detect and annotate the lesions based upon pixel density in the fundus images - microaneurysms, hard exudates, cotton wool spots, superficial and deep hemorrhages, neovascularizations and fibrovascular proliferations.

The technique of *Data Augmentation*^[9] was used to avoid overfitting and making the model more robust. The training dataset consisted of 112,489 deidentified fundus images sourced from various hospitals. The images were derived from various fundus cameras of both mydriatic and nonmydriatic models including Topcon TRC NW8, Topcon TRC NW400 (TOPCON Medical Systems, Japan), Bosch Hand-Held Fundus camera (Robert Bosch Eng. And Business Pvt Solutions, Germany), Intucam Prime, Intucam 45 (Intuvision Labs, India) Trinetra Classic (FORUS Health, India), and Remidio (REMIDIO Innovative Solutions, India). About 72% of the training images were from mydriatic and remaining 28% were from nonmydriatic cameras.

Image quality assessment

Digital fundus photography is a common procedure in ophthalmology and provides critical diagnostic information of retinal pathologies, such as DR, glaucoma, age-related macular degeneration, and vascular abnormalities. An algorithm able to automatically assess the quality of the fundus image is an important preprocessing step for reliable lesion detection for an AI-based screening system. Due to factors like level of operator expertise, type of equipment used and patient conditions, the acquired retinal image might not have the minimum quality that would facilitate feature extraction, leading to incorrect analytics.

Traditional Image Quality Assessment Algorithm relies on handcrafted features^[10,11] that are based on either generic or structural quality parameters, such as global histogram features, textural features, vessel density, local nonreference perceptual sharpness etc., metric which does not generalize well on different dataset, camera, and field of view (FOV) show bias. "Image Quality Assessment" algorithm is based on CNN, which uses a custom 17-layer CNN network to train the data. The training dataset consisted of 103,578 good quality (gradable) retina images and 8911 ungradable retina images.

Grading of DR

Images were graded using the international clinical DR severity scale (ICDRS)^[12] (for both training and validation) as:

No DR.

NPDR further classified as

- Mild NPDR: Microaneurysms only
- Moderate NPDR: More than just microaneurysms and less than severe DR
- Severe NPDR: Presence of Intraretinal microvascular abnormalities (IRMAs) or venous beading in 2+ quadrants, or > 20 hemorrhages in each quadrant.

Proliferative diabetic retinopathy (PDR).

PDR: Neovascularization or preretinal/vitreous hemorrhage.

Macular edema was defined as presence of hard exudates with one disc diameter of the macula.

Referable DR (RDR)/prompt referral was defined as moderate NPDR or above and/or presence of macular edema.

Sight threatening DR (STDR) was defined as severe NPDR and PDR or above.

Validation of the algorithm:

The accuracy of the algorithm in detecting referable DR was validated using two datasets; an internal validation set with a retrospective collection of fundus images sourced from patients visiting our hospital and an external validation set – MESSIDOR 1. AI results were verified for consistency by running the two image sets for results on three separate occasions. The results seen were identical and there were no variations in the AI results for any image over the three test runs.

Data collection method for validation

Internal validation set

Deidentified fundus images were sourced from our medical records (Imaged between January 2015 and June 2018). The dataset consisted of 1533 independent images and for the purpose of this study was not assigned patient wise. These images did not overlap with the dataset used in training the algorithm. All images were captured using Topcon TRC-50DX and Topcon DRI OCT-TRITON camera with 50° field, posterior pole-centered images. This included both dilated and undilated images. Images with media haze, over/under exposure were excluded from the validation study.

All 1533 images were reviewed by two retina specialists (Observer A and Observer B) and disagreements were further reviewed by a third grader. Agreement between at least two graders was considered as ground truth (GT).

External validation set

MESSIDOR 1 online database consists of 1200 fundus images acquired using Topcon TRC NW6 nonmydriatic camera with a 45° field of view, of which 800 were acquired with pupil dilation and 400 without dilation.

All 1200 images were reviewed by our two retina specialists and disagreements were mutually reviewed to arrive at a consensus grading. ICDR classification scale was used for grading.

Table 1: Distribution of DR staging between two retina specialists for internal and external validation datasets

Observer B	Internal validation set					Total	Weighted Kappa	0.953		
	Observer A								Standard error	0.005
	No DR	Mild NPDR	Mod NPDR	Severe NPDR	PDR					
No DR	121	4	0	0	0	125 (8.2%)				
Mild NPDR	8	16	6	0	0	30 (2.0%)				
Mod NPDR	1	3	468	21	0	493 (32.2%)				
Severe NPDR	0	0	11	130	37	178 (11.6%)				
PDR	0	0	0	3	704	707 (46.1%)				
Total	130	23	485	154	741	1533				
Observer B	External Validation Set					Total	Weighted Kappa	0.961		
	Observer A								Standard error	0.006
	No DR	Mild NPDR	Mod NPDR	Severe NPDR	PDR					
No DR	538	10	0	0	0	548 (45.7%)				
Mild NPDR	6	214	8	2	0	230 (19.2%)				
Mod NPDR	1	0	329	5	2	337 (28.1%)				
Severe NPDR	0	0	10	50	4	64 (5.3%)				
PDR	0	0	0	0	21	21 (1.8%)				
Total	545	224	347	57	27	1200				

The percentage values in brackets indicate the proportion of cases under each stage of DR

Interobserver agreement

Distribution of DR staging between two retina specialists for the internal and external validation datasets is shown in Table 1.

Internal validation set - κ value of 0.95 indicates “Very Good” strength of agreement between the two graders. A total of 94 images were graded differently and hence were reviewed by a third grader to arrive at the GT.

External validation set - Of the 1200 images, the two observers agreed with 1152 image grades (96% agreement) with a κ value of 0.96 showing “Very Good” strength of agreement. And 48 variance images were reviewed between the two graders to arrive at the consensus grading (GT).

Statistical analysis

Statistical analysis was done using Microsoft excel 2016 and MedCalc Statistical Software version 18.11.6 (MedCalc Software bvba, Belgium, 2019). Sensitivity and specificity for identification of any DR, identification of referable/prompt DR and sight threatening DR was calculated. AUC was calculated for the same.

Cohen’s κ was run to determine the agreement between the two graders and between GT and AI results on the presence and severity of DR in the validation dataset. Classification of Cohen’s kappa (κ) was used as per guidelines from Altman (1999):^[13] <0.20 poor, 0.21--0.40 fair, 0.41--0.60 moderate, 0.61--0.80 good, and 0.81--1.00 very good.

Results

Performance of the AI in detecting “any DR,” “Referable DR,” and “Sight Threatening DR” was evaluated in comparison with the GT. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, ROC, and Cohen’s κ was derived. The performance of the AI was also compared with that of the two primary graders to understand if the AI can perform at par, below par, or better than the human experts.

Internal validation set

Of the total 1533 images, 1399 images had some form of DR and 134 had no DR. The distribution of the grades in comparison with GT and the AI output is shown in Table 2.

κ value of 0.69 indicates “Good” strength of agreement in DR severity staging between GT and AI. We noticed the tendency of AI to classify moderate NPDR as severe and PDR as severe, thereby the lower κ value as compared to the strength of agreement between the two retinal experts.

AI shows high sensitivity and specificity for any DR detection and referable DR detection as shown in Table 3. For STDR though the sensitivity is high the specificity is 56.38%. We have observed a tendency of the AI to overclassify moderate NPDR as severe, thereby the lower specificity. However, of the 898 STDR images, AI has referred 897 images thereby demonstrating high accuracy in screening settings.

κ value shows very good agreement between GT and AI results for Any DR detection and referable DR detection. For STDR the agreement is moderate. However, AI identified and considered all images with STDR as referable. It is also notable that AI has shown high sensitivity of 97.7% in detecting STDR, whereas the specificity is 56.4% - this was mostly due to classification of moderate NPDR as severe and hence as STDR.

The AUC for any DR detection was **0.991** and for prompt referral was **0.969** as shown in Fig. 2.

External validation set

Of the 1200 images, 546 had no DR and 654 had some form of DR of which, 432 had referable DR while 84 had sight threatening DR.

Comparison of consensus grade and AI derived grades is given in Table 4.

κ value of 0.82 shows very good strength of agreement in DR severity grading. However, 49 cases without DR were classified as mild NPDR and 62 cases with mild NPDR were classified as no DR by the AI, indicating that the algorithm requires further training and validation in detecting subtle changes such as microaneurysms and thereby the accurate identification of mild NPDR, though this may not be clinically relevant as no specific treatment is indicated for mild NPDR.

As observed in the internal validation set, AI has demonstrated high levels of accuracy in detecting both referable and sight threatening DR for external validation set as well as seen in Table 3.

Table 2: Distribution of DR staging between Ground Truth grading and AI results for the internal validation dataset

AI_Grade	Consensus grade (GT) for internal validation set					Total
	No DR	Mild NPDR	Mod NPDR	Severe NPDR	PDR	
No DR	132	3	1	0	0	136 (8.9%)
Mild NPDR	1	11	12	0	1	25 (1.6%)
Mod NPDR	1	7	189	11	10	218 (14.2%)
Severe NPDR	0	0	273	131	269	673 (43.9%)
PDR	0	0	4	9	468	481 (31.4%)
	134 (8.7%)	21 (1.4%)	479 (31.2%)	151 (9.8%)	748 (48.8%)	1533
Weighted Kappa^a						0.688
Standard error						0.013
95% CI						0.663-0.713

^aLinear weights. The percentage values in brackets indicate the proportion of cases under each stage of DR

κ value for STDR shows moderate agreement with a sensitivity of 91.67% and specificity of 92.92%. Of the 84 images with STDR, AI has classified 77 as STDR and the remaining 7 as moderate NPDR, i.e. AI has identified and referred all instances of STDR.

The AUC for any DR detection was .907 and for prompt referral was .960 as shown in Fig. 3.

Discussion

From our study, we can observe that this AI algorithm showed a very good sensitivity and specificity for detection of any DR, i.e., 99.7% and 98.5% and for prompt referral, i.e. 98.9% and 94.84%, respectively, comparable to the existing algorithms, thus suitable as a very good screening tool. AI has been shown to have a sensitivity close to 90% to detect referable DR in the studies so far performed. The sensitivity and specificity of AI improves as the machine sees more and more images. One of the greatest advantages of the algorithm used in our study is that it has been developed using fundus images from multiethnic populations, thus diverse disease presentations

were used for training. The large training data sets with wide spectrum of clinical findings help the algorithm to better understand the varied presentation of the disease. The second advantage is that the validation set images in our study were retrieved directly from clinic rather than online datasets as in several other studies.^[14-16] Third, the fundus images in the training set have been retrieved from a wide variety of fundus cameras including smart phone-based fundus cameras, thus helping to train AI in a very robust way with a heterogeneous set of images. Fourth, unlike in other studies^[16] where the training set itself has been used for validation, we have used a distinct set for validation so as to reduce the training bias in the output results. Fifth, the algorithm under study differs from others, as it not only identifies referable DR but also annotates the lesions; and thus, lesion identification helps the ophthalmologist understand what the AI is comprehending and also to assess progression of the disease on follow-up. It also helps in making the patient understand the severity of the DR for better understanding of the eye problem, better counseling, and in turn better compliance for the management

Table 3: Performance of AI in comparison to Ground Truth for internal and external validation sets

	Internal validation set					
	Any DR detection		Referable DR detection		Sight threatening DR detection	
Sensitivity	99.71%	99.27%-99.92%	98.98%	98.30%-99.44%	97.55%	96.32%-98.46%
Specificity	98.50%	94.71%-99.82%	94.84%	90.08%-97.75%	56.31%	52.35%-60.21%
AUC	0.991	0.985-0.995	0.969	0.959-0.977	0.77	0.747-0.790
PPV	99.86%	99.44%-99.96%	99.42%	98.86%-99.70%	76.00%	74.34%-77.58%
NPV	97.06%	92.53%-98.87%	91.30%	86.16%-94.65%	94.20%	91.43%-96.10%
k	0.975	0.956-0.995	0.922	0.890-0.954	0.572	0.532-0.612
	External validation set					
	Any DR detection		Referable DR detection		Sight threatening DR detection	
Sensitivity	90.37%	87.84%-92.52%	94.68%	92.12%-96.60%	91.67%	83.58%-96.58%
Specificity	91.03%	88.31%-93.29%	97.40%	96.01%-98.40%	92.92%	91.26%-94.36%
AUC	0.907	0.889-0.923	0.96	0.948-0.971	0.923	0.906-0.937
PPV	92.34%	90.22%-94.04%	95.34%	92.99%-96.93%	49.36%	43.84%-54.90%
NPV	88.75%	86.17%-90.90%	97.02%	95.62%-97.98%	99.33%	98.65%-99.67%
k	0.812	0.779-8.845	0.922	0.899-0.945	0.606	0.531-0.680

Any DR=Stage 1, 2, 3, 4; Referable DR=Stage 2, 3, 4; Sight threatening DR=Stage 3 and 4

Table 4: Distribution of DR staging between Ground Truth grading and AI results for the external validation dataset

AI Grade	Consensus grade (GT) for external validation set					Total
	No DR	Mild NPDR	Mod NPDR	Severe NPDR	PDR	
No DR	497	62	1	0	0	560 (46.7%)
Mild NPDR	48	141	22	0	0	211 (17.6%)
Mod NPDR	1	19	246	7	0	273 (22.8%)
Severe NPDR	0	0	78	52	8	138 (11.5%)
PDR	0	0	1	1	16	18 (1.5%)
Total	546 (45.5%)	222 (18.5%)	348 (29.0%)	60 (5.0%)	24 (2.0%)	1200

Weighted Kappa^a

0.82173

Standard error

0.01013

95% CI

0.80187-0.84159

^aLinear weights. The percentage values in brackets indicate the proportion of cases under each stage of DR

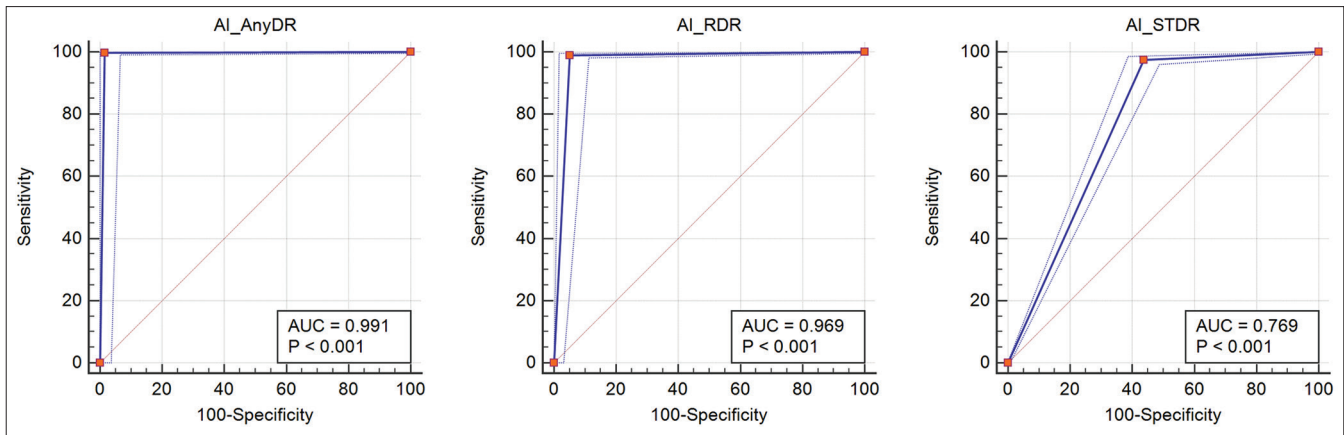


Figure 2: AUC of internal validation set

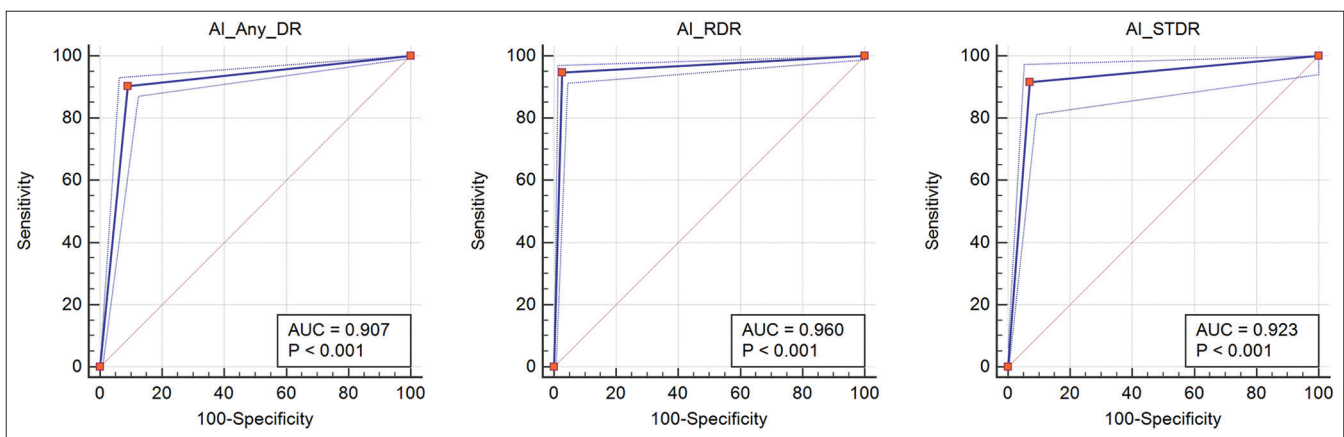


Figure 3: AUC of external validation set

and follow-up. When used in screening, it will aid in accurate identification of the lesions by the nonophthalmologists and thus the referrals can also be clarified as immediate or urgent or emergent depending upon the stage of DR. An example of the output with lesion identification is shown in Fig. 4.

The initial breakthrough study using AI for DR was by Gulshan *et al.*^[14] from the Google AI Healthcare group in 2016. The most robust study till date by Ting *et al.*,^[17] where the deep learning system (DLS) was developed and validated using images from various multiethnic populations along with their Singapore national DR screening program, showed a sensitivity and specificity of 90.5% and 91.6% respectively for referable DR in primary validation set. Li *et al.*^[18] have also developed a multiethnic-based DLS which showed a sensitivity and specificity of 97% and 91.4% for sight threatening referable DR. They also highlighted the characteristics of false positives and false negatives when using AI in their study. The DLS developed by Abramoff *et al.*,^[19] has received U.S. FDA approval in April 2018 for diagnosis of DR, which has a sensitivity and specificity of 87.2% and 90.7%, respectively. A brief summary of various studies and their results are shown in Table 5.^[14-17,20,21]

In the present scenario, every diabetic patient needs to be referred to a retina specialist for diagnosis and treatment of DR. If an automated software is able to identify sight threatening DR as precisely as a clinician, it would greatly reduce the burden



Figure 4: Fundus image with lesion annotations by the AI algorithm

of screening on vitreoretinal (VR) surgeons. It would also help to reduce the financial burden on the patient. Kanagasingam *et al.* have also tried AI-based algorithms to detect DR in primary health care setting.^[22]

A study by Keel *et al.*^[23] to evaluate the feasibility and patient acceptability of AI-based DR screening in endocrinology clinics showed 96% of the patients were satisfied with the automated

Table 5: Brief summary of previous studies and their comparison with our results

	Dataset	Sensitivity	Specificity
Any DR			
Pratt <i>et al.</i> ^[20]	5000 images	30%	95%
Gargeya and Leng ^[15]	Messidor 2 (1748 images)	93%	87%
Abramoff <i>et al.</i> ^[21]	819 images	87.2%	90.7%
Our study (India)	1533 images	99.7%	98.5%
	Messidor 1 (1200 images)	90.37%	91.03%
Referable DR			
Gulshan <i>et al.</i> ^[14]	EyePACS-1 (9963)	90.3%	98.1%
	Messidor 2 (1748)	97.5%	93.4%
		87%	98.5%
		96.1%	93.9%
Ting <i>et al.</i> ^[17]	Multiple sets of images	90.5%	91.6%
Ramachandran <i>et al.</i> ^[16]	ODEMS (382)	84.6%	79.7%
	Messidor 1 (1200)	96%	90%
Our study (India)	1533 images	98.98%	94.84%
	Messidor 1 (1200)	94.68%	97.40%

DR screening model and 78% reported that they preferred automated over manual screening with the mean assessment time of only 6.9 minutes. The algorithm in the present study takes an average of 5 s per image for detecting DR and marking lesions, which is as good as a clinician.

Shortcomings of AI in the current scenario are as follows:

1. False positives if the image has artefacts due to poor image capturing;
2. Difficulty in distinction between mild and moderate NPDR in the absence of hard or soft exudates, since presence of dot hemorrhages alone can mimic microaneurysms. This would not have a great clinical significance as such patients would not require any treatment although they fall into moderate NPDR group according to ICDRS
As mentioned by Krause *et al.*^[24] in their study of grader variability in machine learning models, most common discrepancies were due to missing microaneurysm, artifacts, and misclassified hemorrhages.
3. Sharply delineated/hard drusen/pigments can sometimes give false positive results.
4. False negatives due to inability to identify cystoid macular edema in the absence of hard exudates;
5. False negatives in inactive peripheral lasered PDR cases.

Limitations of our study

Although we have only used images obtained from the Topcon fundus camera in the internal validation set, we were able to obtain similar results with the external validation set (MESSIDOR 1).

Also, we are conducting an ongoing DR screening study where we are trying to validate the software on different fundus cameras in a real-world outreach screening program. The sensitivity to identify referable DR from its pilot study also showed promising results (unpublished data).

Conclusion

The advantage of AI would be in its ability to detect DR without the need for a trained retina specialist, remote screening, and ability to rapidly screen large numbers. This can revolutionize telescreening in ophthalmology, especially where people have poor access to specialized health care. With the increasing diabetic population and the growing demand supply gap in trained resources for disease screening, AI can be the future for early identification of DR and thereby reducing the blindness burden. Integrated into fundus cameras, an optometrist or a trained technician can screen diabetic patients much earlier using portable nonmydriatic cameras. Easier accessibility of these software and their integration with more portable fundus camera devices, which can be operated by a technician alone, will give a huge scope for even a nontrained health professional in screening DR. This would especially benefit in developing countries where penetration of specialized ophthalmic care is poor.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004;27:1047-53.
2. Raman R, Rani PK, Reddi Racheppalle S, Gnanamoorthy P, Uthra S, Kumaramanickavel G, *et al.* Prevalence of diabetic retinopathy in India: Sankara Nethralaya diabetic retinopathy epidemiology and molecular genetics study report 2. *Ophthalmology* 2009;116:311-8.
3. Shah S, Das AK, Kumar A, Unnikrishnan AG, Kalra S, Baruah MP, *et al.* Baseline characteristics of the Indian cohort from the IMPROVE study: A multinational, observational study of biphasic insulin aspart 30 treatment for type 2 diabetes. *Adv Ther* 2009;26:325-35.
4. Bawankar P, Shanbhag N, K SS, Dhawan B, Palsule A, Kumar D, *et al.* Sensitivity and specificity of automated analysis of single-field non-mydratic fundus photographs by Bosch DR Algorithm-Comparison with mydratic fundus photography (ETDRS) for screening in undiagnosed diabetic retinopathy. *PLoS One* 2017;12:e0189854.
5. Srihatrai P, Hlowchitsieng T. The diagnostic accuracy of single- and five-field fundus photography in diabetic retinopathy screening by primary care physicians. *Indian J Ophthalmol* 2018;66:94-7.
6. Escontrela A. Convolutional Neural Networks from the ground up. Towards data science, June 2017. Available from: <https://towardsdatascience.com/convolutional-neural-networks-from-the-ground-up-c67bb41454e1>.
7. Li Z, Keel S, Liu C, He M. Can artificial intelligence make screening faster, more accurate, and more accessible? *Asia Pac J Ophthalmol (Phila)* 2018;7:436-41.
8. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, *et al.* Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103:167-75.
9. Wang J, Perez L. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv: 1712.04621v1, Dec 17.
10. Yu H, Agurto C, Barriga S, Nemeth S, Soliz P, Zamora G, *et al.* Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening. *IEEE Southwest Symposium*

- on Image Analysis and Interpretation 2012:125-8. doi: 10.1109/SSIAI.2012.6202469.
11. Tennakoon R, Mahapatra D, Roy P, Sedai S, Garnavi R. Imagequality classification for DR Screening using Convolutional neural networks. In: Chen X, Garvin, Liu J, Trucco E, Xu Y, editors. Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop, OMIA 2016, 2016. p. 113-120. Available from: <https://doi.org/10.17077/omia.1054>.
 12. Wilkinson CP, Ferris FL 3rd, Klein RE, Lee PP, Agardh CD, Davis M, *et al.* proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677-82.
 13. Altman DG. Practical Statistics for Medical Research. Boca Raton: Chapman and Hall/CRC; 1991.
 14. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;304:649-56.
 15. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017;124:962-9.
 16. Ramachandran N, Hong SC, Sime MJ, Wilson GA. Diabetic retinopathy screening using deep neural network. *Clin Exp Ophthalmol.* 2018;46:412-6.
 17. 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.
 18. Li Z, Keel S, Liu C, He Y, Meng W, Scheetz J, *et al.* An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care* 2018;41:2509-16.
 19. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. U.S. Food and Drug Administration Web site. Available from: <https://www.idf.org/e-library/epidemiology-research/diabetes-atlas/13-diabetes-atlas-seventh-edition.html>. Updated December 4, 2018. [Last accessed on 2018 Dec 06].
 20. Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. *Procedia Comput Sci* 2016;90:200-5.
 21. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39.
 22. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney ML, Mehrotra A. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. *JAMA Netw Open* 2018;1:e182665.
 23. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, Maclsaac RJ, *et al.* Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: A pilot study. *Sci Rep* 2018;8:4330.
 24. 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;125:1264-72.