Development and validation of an offline deep learning algorithm to detect vitreoretinal abnormalities on ocular ultrasound

Venkatesh Krishna Adithya, Prabu Baskaran[1], Aruna S[1], Arthi Mohankumar, Jean Pierre Hubschman[2], Aakriti Garg Shukla[3], Rengaraj Venkatesh

Purpose: We describe our offline deep learning algorithm (DLA) and validation of its diagnostic ability to identify vitreoretinal abnormalities (VRA) on ocular ultrasound (OUS). **Methods:** Enrolled participants underwent OUS. All images were classified as normal or abnormal by two masked vitreoretinal specialists (AS, AM). A data set of 4902 OUS images was collected, and 4740 images of satisfactory quality were used. Of this, 4319 were processed for further training and development of DLA, and 421 images were graded by vitreoretinal specialists (AS and AM) to obtain ground truth. The main outcome measures were sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and area under receiver operating characteristic (AUROC). **Results:** Our algorithm demonstrated high sensitivity and specificity in identifying VRA on OUS ([90.8%; 95% confidence interval (CI): 86.1–94.3%] and [97.1% (95% CI: 93.7–98.9%], respectively). PPV and NPV of the algorithm were also high ([97.0%; 95% CI: 93.7–98.9%] and [90.8%; 95% CI: 86.2–94.3%], respectively). The AUROC was high at 0.939, and the intergrader agreement was nearly perfect with Cohen's kappa of 0.938. The model demonstrated high sensitivity in predicting vitreous hemorrhage (100%), retinal detachment (97.4%), and choroidal detachment (100%). **Conclusion:** Our offline DLA software demonstrated reliable performance (high sensitivity, specificity, AUROC, PPV, NPV, and intergrader agreement) for predicting VRA on OUS. This might serve as an important tool for the ophthalmic technicians who are involved in community eye screening at rural settings where trained ophthalmologists are not available.

Key words: Artificial intelligence, deep learning, ophthalmic technicians, retina, ultrasound, vitreo retinal, vitreous

The demand for ophthalmic care continues to grow globally as the aging population expands,[1,2] inviting a revolution with the advent of artificial intelligence (AI). Chronic ophthalmic diseases require mass screening at the community level to allow for early detection and potentially better quality of life, posing an immense burden on the healthcare and human resources. The utility of imaging in ophthalmology makes it the ideal field to potentially benefit from AI applications.[3] This was evidenced by the first Food and Drug Administration-approved AI project, IDx-DR, which is an algorithm developed to screen diabetic retinopathy (DR). Further AI studies have focused on posterior segment images of DR, age-related macular degeneration, and glaucoma, the three leading causes of irreversible blindness worldwide.

In settings where the dense cataract or other media opacity prevents a view of the posterior segment, ocular ultrasound (OUS) is an important investigative tool. In India, the prevalence of mature cataracts is high at 7–10%, making OUS a commonly used modality.[4] The ophthalmic technicians who are routinely involved in community screening at rural settings are unable to assess the posterior segment. Furthermore, the lack of consistent internet connectivity in rural parts of India further limits teleophthalmic consult for image interpretation.

These obstacles to posterior segment evaluation pose a risk, especially in the setting of a comprehensive examination prior to cataract surgery. Visual prognosis cannot be adequately assessed prior to surgery, and the lack of complete information prior to surgery can potentially lead to medicolegal issues. While fundus photography is the subject of numerous AI studies,[5-8] OUS has not been evaluated by AI.

Our study aimed to develop an offline deep learning algorithm (DLA) for the ophthalmic technicians to identify multiple vitreoretinal abnormalities (VRA) on OUS as normal or abnormal. This might be a value addition for community eye screening done by ophthalmic technicians. Here, we report the development of our DLA strategy using two independent data sets, and its validation using 421 test images.

Methods

Our study was designed and carried out at two major tertiary eye care facilities in South India between January 2018 and December 2019. The study adhered to the tenets of Declaration of Helsinki and approved by the Institutional Review Board of

Glaucoma, Aravind Eye Hospital, Pondicherry, India, [1]Retina, Aravind Eye Hospital, Chennai, Tamil Nadu, India, [2]Retina, Jules Stein Eye Institute, UCLA, USA, [3]Glaucoma, Wills Eye Hospital, Philadelpia, USA

Correspondence to: Dr. Prabu Baskaran, Aravind Eye Hospital, Noombal, Poonamallee High Road, Chennai, Tamil Nadu - 600 077, India. E-mail: prabubaskaran@gmail.com

Aravind Eye Hospital. Patients were enrolled with informed written consent.

Data set

Patients who underwent OUS for documentation of posterior segment examination prior to anterior segment surgery (including cataract surgery) and those with known VRA were recruited to participate in this study. The OUS imaging for all the study patients was done by two fellowship-trained retina specialists (AM, AS) using standardized screening protocol. Though ultrasound imaging is a dynamic process, we have considered one representative OUS picture per eye as the study image for further processing. The horizontal axial scan position was preferred; if this scan was not representative of the overall clinical picture, the scan position which demonstrated the exact VRA was the designated study image. OUS images were retrieved directly from the ultrasound machine (Appasamy Associates, Gantec Corporation, India/ same machine was used at both centers) and processed further as described below.

A data set consisting of 4902 images was collected from 2486 patients and 4740 OUS images of satisfactory quality were used. The data set was further separated into a training set (4319 OUS images) and validation set (421 OUS images), which is shown in Fig. 1. The sample size for validation set was calculated based on previous literature work on AI related to DR. Images were placed in the validation set to ensure an equal number of normal and abnormal images. The training set images were graded by two fellowship-trained retina specialists (AM, AS) and reported as either normal or abnormal. Fig. 2a–c demonstrates the initial clinical grading required for DLA training. The final number of images, features assessed, and the data set used for training and validation are described in Table 1.

Data preprocessing

Automated preprocessing steps were applied to the original images to make these images suitable for training the DLA. These steps are described as follows. Images were downsized to 299 × 299 pixels and divided by 255 to scale the image pixels
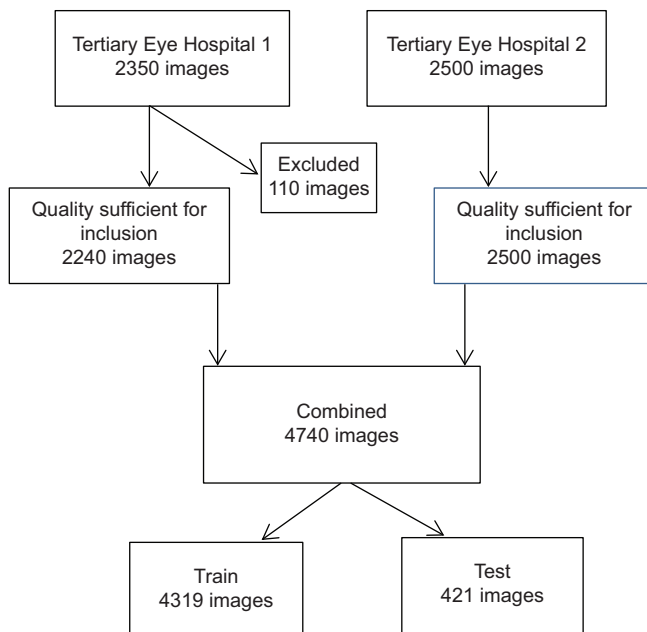
to a range of 0–1. Image augmentation including change in orientation (horizontal or vertical flip), rotation up to 45°, shifting width, height, sheer, and zoom in the range of -20 to + 20% were randomly applied during training. These steps increased the diversity of the data set, reducing the possibility of overfitting and making the DLA more robust.

Machine learning architecture

We used an Inception-ResNet-V2 architecture for classification.[9] Given our limited sample size, we used transfer learning to avoid overfitting. Transfer learning is the reuse of deep learning models that are pretrained in large data sets like the ImageNet[10] data set to reduce training time and maximize accuracy despite a small data set. We used Inception-ResNet-V2 that was pretrained on ImageNet and a classifier of fully connected layers was added to separate OUS images into normal and abnormal.

Python programming language (http://www.python.org/) and Keras library with Tensorflow backend (https://keras.io/) were used to train the deep learning model. All training was done using NVIDIA RTX 2080 graphics processing unit with NVIDIA cuda (version 10.0) and cudnn (version 7.6.5).

Input to the network consisted of 4319 training images. The model was trained with multiple epochs, each epoch with a batch size of 12, with a learning rate of 0.001 with an ADAM optimizer, which minimizes the loss function (error function). For classification, categorical cross-entropy was used. Training was stopped when the loss of the model and accuracy in the validation images decreased. Greatest accuracy of 94.6% was achieved in the 49th epoch.

The heatmaps highlighting the regions in which abnormalities were detected using the DLA were generated using the Grad-CAM method for all true-positive images. This method calculates the gradients of the output of the DLA network with respect each pixel to identify pixels that have the greatest impact on the prediction. A senior retina specialist (PB) confirmed that heatmaps were effective in highlighting abnormal regions in OUS images.

Ground truth

The validation data set of 421 deidentified images were assessed separately by both vitreoretinal specialists (AM, AS) for ocular abnormalities shown in Table 1. Based on these structures, graders reported the OUS image as normal or abnormal. Discrepancy in reporting between the two graders was mutually agreed upon and adjudicated result was considered as the ground truth. Area under the receiver operating characteristic (AUROC), sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV) of the DLA were primary outcome measures.

Statistical plan

Data collection was performed in Microsoft Excel (Microsoft Corporation, Washington, US). The power of the study was set to 80%. $P < 0.05$ was considered to reflect statistical significance. The normality of the study was assessed using Kolmogorov–Smirnov test. The AUROC curve was obtained by plotting true-positive rate against false-positive rate. Intergrader agreement was calculated by Cohen's kappa. Data analysis was performed using Stata version 14, Statacorp, College Station, Texas.

Results

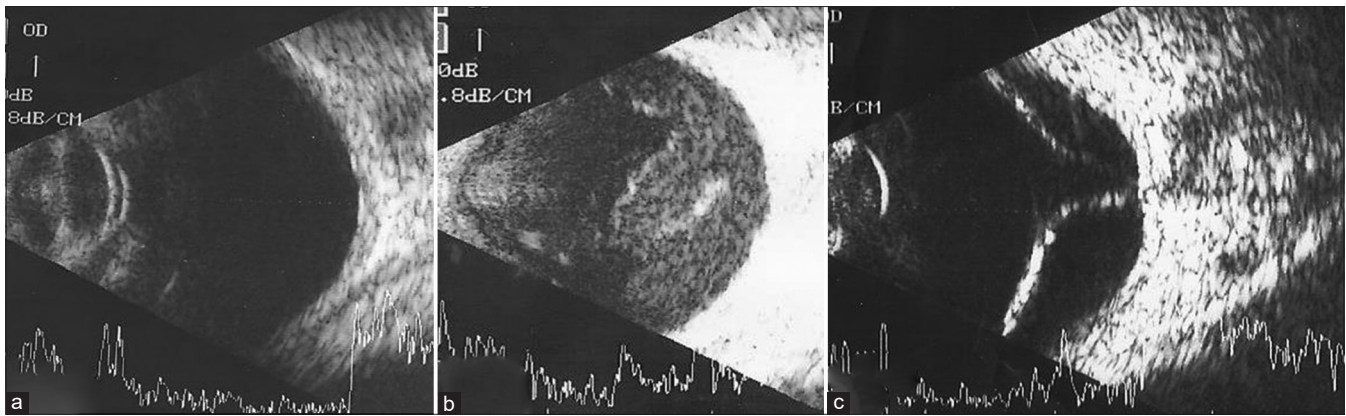As shown in the flowchart [Fig. 1], we collected 4850 OUS images prospectively and excluded 110 ungradable images.

Figure 1: Sample selection at Aravind Eye Hospital Pondicherry and Chennai

Figure 2: Ocular ultrasound scan image showing (a) normal structures, (b) multiple vitreous dot echoes, and attached retina (c) detached retina

Table 1: Assessed ocular abnormalities for training and validation data sets

Total no. of OUS images		4850	
Total no. of gradable images		4740 (97.73%)	
Parameters	**Features looked at for labeling normal vs abnormal**	**Training data set (4319 eyes)**	**Validation data set (421 eyes)**
Lens status	Normal – no lens echo noted	2049 (47.4%)	206 (48.9%)
	Normal – lens echo noted	1661 (38.4%)	148 (35.1%)
	Normal – IOL reverberations noted	583 (13. %)	61 (14.4%)
	Abnormal – Subluxated/dislocated lens/IOL	26 (0.6%)	6 (1.4%)
Vitreous dot echoes	Normal – no dot echoes	706 (16.3%)	61 (14%)
	Normal – mild dot echoes	2381 (55.1%)	219 (52%)
	Abnormal – moderate echoes	633 (14.6%)	81 (19.2%)
	Abnormal – plenty of dot echoes	588 (13.6%)	60 (14.2%)
Vitreous clump echoes	Normal – no clump echoes	3830 (88.6%)	354 (84%)
	Abnormal – clump echoes present	489 (11.3%)	67 (15.9%)
Vitreous membranous echoes	Normal – no membranous echoes	3589 (83.1%)	326 (77.4%)
	Abnormal – single membranous echo present	479 (11.0%)	59 (14%)
	Abnormal – two membranous echo present	220 (5.0%)	35 (8.3%)
	Abnormal – multiple membranous echoes present	31 (0.7%)	1 (0.2%)
Attachment of membranous echo to disc	Normal – not attached	3870 (89.6%)	361 (85.7%)
	Normal – point attachment (IPVD)	223 (5.1%)	21 (4.9%)
	Abnormal – broad attachment (RD)	224 (5.1%)	39 (9.2%)
Retina	Normal – attached	3952 (91.5%)	383 (90.1%)
	Abnormal – detached	242 (5.6%)	38 (9.9%)
Choroid	Normal – no choroidal detachment	4257 (98.5%)	416 (98.8%)
	Abnormal - choroidal detachment present	61 (1.4%)	5 (1.1%)
Retina choroid sclera	Normal	4131 (95.6%)	401 (95.2%)
	Abnormal – thickened	77 (1.7%)	9 (2.1%)
ST fluid	Normal – absent	4290 (99.3%)	416 (98.8%)
	Abnormal – present	28 (0.6%)	5 (1.1%)
T sign	Normal – absent	4293 (99.4%)	416 (98.8%)
	Abnormal – present	25 (0.5%)	5 (1.1%)
Final diagnosis	Normal	2535 (58.6%)	204 (48.4%)
	Abnormal (if any one of the above parameters was found to be abnormal, then the final diagnosis was termed abnormal)	1784 (41.3%)	217 (51.5%)

IPVD – incomplete posterior vitreous detachment, RD – retinal detachment, and ST – subtenon

The 4740 images were split into training set (4319 images) and validation set (421 images), which are detailed in Table 1.

Of the 421 images tested for accuracy, the DLA software reported 203 abnormal and 218 normal OUS images. Findings are displayed in Table 2. The sensitivity and specificity of the

algorithm to identify a posterior segment abnormality were 90.8% (95% CI: 86.1–94.3%) and 97.1% (95% CI: 93.7–98.9%), respectively. The PPV and NPV of the algorithm were high at 97.0% (95% CI: 93.7–98.9%) and 90.8% (95% CI: 86.2–94.3%), respectively. The AUROC was high at 0.939 [Fig. 3]. The agreement between the two masked clinical graders was 96.9% (Cohen's kappa 0.9382, $P < 0.001$). Analyses performed on false-negative and true-positive images are shown in Table 3, which demonstrate the range of abnormalities found on OUS images and the performance of the model in detecting these abnormalities.

To visualize how the DLA makes abnormal predictions from OUS images, heatmaps were generated to indicate the region of abnormality. Fig. 4 presents examples of VRA activation maps, which are accompanied by the corresponding original image. Heatmaps effectively highlighted regions of the abnormality.

Discussion

Our DLA with Inception-Resnet-V2 network detected VRA using OUS images with high sensitivity and specificity and

Table 2: Contingency table comparing the results between our DLA and ground truth

Our DLA (diagnostic test) results	Clinical grading (ground truth) results		Total
	Abnormal	Normal	
Abnormal	197	6	203
Normal	20	198	218
Total	217	204	421

Table 3: Analysis of false-negative and true-positive results

Clinical condition	Number of images tested	Identified (%)	Missed (%)
Vitreous hemorrhage	137	137 (100)	0
Retinal detachment	38	37 (97.4)	1 (2.6)
Silicone oil filled globe	18	17 (94.5)	1 (5.5)
Choroidal detachment	5	5 (100)	0
Choroidal thickening	2	0	2 (100)

Area under ROC curve = 0.9392
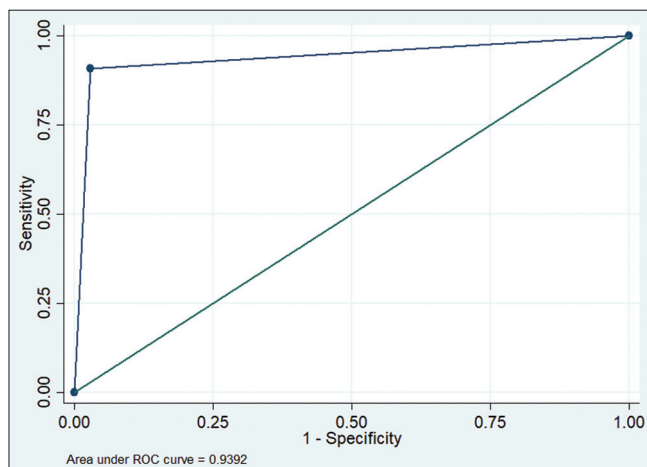
**Figure 3:** Area under the receiver operating curve for the deep learning algorithm

achieved an AUROC of 0.939. Our approach avoided numerous example images for model convergence by fine-tuning the weights of the Inception Resnet V2 model, which was pretrained on ImageNet data set. To our knowledge, our study is the first to describe the use of AI for OUS interpretation. Our results suggest the competence of our algorithm in providing cost-effective and objective diagnostics for VRA, reducing the need to depend on vitreoretinal specialists. Additionally, the ability of OUS to detect VRA in the setting of media opacities make it especially useful as compared to most of the modalities currently studied by AI (including ultrawide field fundus images and optical coherence tomography images). Moreover, as this algorithm can perform offline, it may be especially suitable for use in developing countries, where internet connectivity may not be consistently available.

Although prior AI applications of OUS are not available in the literature, our algorithm delivered results that were similar to other modalities of posterior segment evaluation. Li et al.[8] introduced a deep learning system for predicting retinal detachment from ultrawidefield fundus camera fundus images and found a sensitivity of 96.1%, specificity of 99.6%, and an AUROC of 0.989 (95% CI: 0.978–0.996). Additionally, Ohsugi et al.[11] developed DLA software with 98% sensitivity and 97% specificity for detecting retinal detachment based on ultrawidefield fundus photographs. These studies used fundus photographs and were limited to the analysis of retinal detachment, while our model utilized OUS and studied multiple VRA types. In spite of this, our model had a comparable sensitivity to the Li et al. model (97.4% vs. 99.6%, respectively) and Ohsugi et al. model (97.4% vs. 97.6%, respectively) in detecting abnormalities in the setting of retinal detachment.

While analyzing the true-positive and false-negative images, the model performed effectively and predicted abnormalities in images with vitreous hemorrhage, retinal detachment, silicone oil in the posterior segment, and choroidal detachment, with sensitivities of 100, 97.4, 94.4, and 100%, respectively [Table 3]. On the other hand, the model performed poorly in detecting choroidal thickening; however, this is probably due to the limited number of images of this diagnosis in the training set.

This study has several limitations. First, the data set was limited by the novelty of this study and the lack of open source OUS data sets. Our own collection of OUS images included a greater proportion of normal cases than abnormal cases in the training set, leading to underexposure to some of the less common VRA. Second, although we used data sets from two distinct hospitals, all images were from individuals of South Asian descent, potentially limiting its applicability elsewhere. External validation of this data set across various ethnicities and geographic locations would test its broader applicability. Third, the aim of our model is not to identify the exact abnormality on OUS. The current version can only separate the abnormal OUS images from the normal OUS images. Future directions include developing separate algorithms for detecting most common OUS abnormalities in the future, potentially allowing for efficient triaging of cases that require specialty care.

Significant strengths of our study include its novel evaluation of OUS, a common, cost-effective, and useful diagnostic test, and our DLA's high accuracy in OUS image analysis. While additional work on this algorithm may improve its accuracy, we believe the present report provides a valuable start to AI analysis of OUS images. As a screening tool, this DLA may help in primary health care centers in rural regions in India and globally to enable telemedicine. Moreover, one of the major practical difficulties that occur with any AI module is the need for consistent internet connectivity.[12] We have
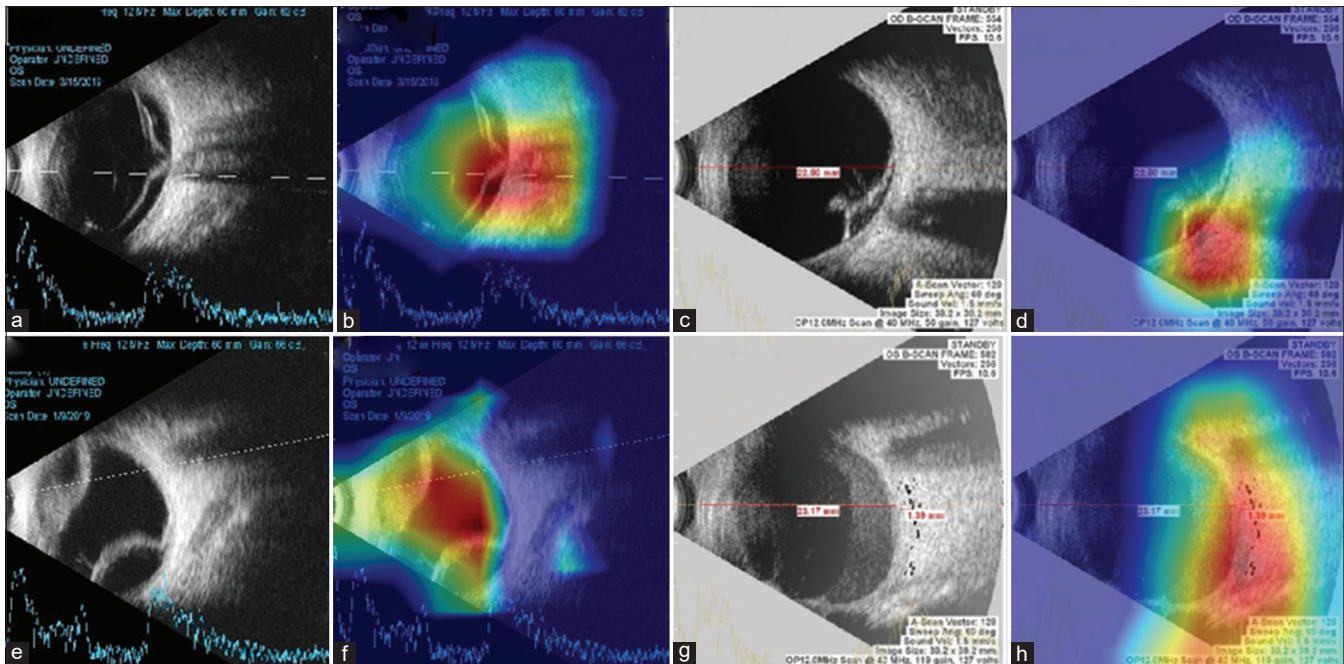
Figure 4: Heat maps highlighting regions with abnormalities detected using the DLA. (a) OUS of rhegmatogenous retinal detachment. (b) Heat map of DLA identifying the site of vitreoretinal abnormalities. (c) OUS of tractional retinal detachment. (d) Heatmap of DLA identifying the abnormal tenting of retina. (e) OUS of choroidal detachment. (f) Heatmap of DLA identifying the site of vitreoretinal abnormalities. (g) OUS of vitreous hemorrhage. (h) Heatmap of DLA identifying the site of vitreoretinal abnormalities

designed our algorithm in such a way that it can work without internet connectivity (offline), allowing its applicability in remote regions.

Conclusion

The DLA achieved high sensitivity (90.8% [95% CI: 86.1–94.3%]) and specificity (97.1% [95% CI: 93.7–98.9%]) to identify VRA on OUS. We believe that our algorithm may be useful for ophthalmic technicians and may improve the eye care standards in rural areas, where there is a lack of trained ophthalmologists and internet access.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

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. Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng CY, *et al.* Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis. Lancet Glob Health 2014;2:e106-16.

3. Faes L, Liu X, Wagner SK, Fu DJ, Balaskas K, Sim D, *et al.* A clinician's guide to artificial intelligence: How to critically appraise machine learning studies. Transl Vis Sci Technol 2020;9:7.

4. Singh S, Pardhan S, Kulothungan V, Swaminathan G, Ravichandran J, Ganesan S, *et al.* The prevalence and risk factors for cataract in rural and urban India. Indian J Ophthalmol 2019;67:477.

5. Kim KE, Kim JM, Song JE, Kee C, Han JC. Development and validation of a deep learning system for diagnosing glaucoma using optical coherence tomography. J Clin Med 2020;9:2167.

6. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks CME Quiz at jamanetwork.com/learning. JAMA Ophthalmol 2017;135:1170-6.

7. Mukkamala SK, Patel A, Dorairaj S, McGlynn R, Sidoti PA, Weinreb RN, *et al.* Ocular decompression retinopathy: A review. Surv Ophthalmol 2013;58:505-12.

8. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, *et al.* Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. Commun Biol 2020;3:15.

9. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. 31st AAAI Conf Artif Intell AAAI 2017, 2017. p. 4278-84.

10. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, *et al.* ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115:211-52.

11. Ohsugi H, Tabuchi H, Enno H, Ishitobi N. Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Sci Rep 2017;7:9425.

12. 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.