Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity

Travis K Redd,[1] John Peter Campbell,[1] James M Brown,[2] Sang Jin Kim,[1,3] Susan Ostmo,[1] Robison Vernon Paul Chan,[4] Jennifer Dy,[5] Deniz Erdogmus,[5] Stratis Ioannidis,[5] Jayashree Kalpathy-Cramer,[2] Michael F Chiang,[1,6] for the Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium

For numbered affiliations see end of article.

Correspondence to
Michael F Chiang, Departments of Ophthalmology & Medical Informatics and Clinical Epidemiology, Casey Eye Institute, Oregon Health & Science University, Portland, OR 97239, USA; chiangm@ohsu.edu

Received 31 August 2018
Revised 3 October 2018
Accepted 17 October 2018
Published Online First 23 November 2018

ABSTRACT

Background Prior work has demonstrated the near-perfect accuracy of a deep learning retinal image analysis system for diagnosing plus disease in retinopathy of prematurity (ROP). Here we assess the screening potential of this scoring system by determining its ability to detect all components of ROP diagnosis.

Methods Clinical examination and fundus photography were performed at seven participating centres. A deep learning system was trained to detect plus disease, generating a quantitative assessment of retinal vascular abnormality (the i-ROP plus score) on a 1–9 scale. Overall ROP disease category was established using a consensus reference standard diagnosis combining clinical and image-based diagnosis. Experts then ranked ordered a second data set of 100 posterior images according to overall ROP severity.

Results 4861 examinations from 870 infants were analysed. 155 examinations (3%) had a reference standard diagnosis of type 1 ROP. The i-ROP deep learning (DL) vascular severity score had an area under the receiver operating curve of 0.960 for detecting type 1 ROP. Establishing a threshold i-ROP DL score of 3 conferred 94% sensitivity, 79% specificity, 13% positive predictive value and 99.7% negative predictive value for type 1 ROP. There was strong correlation between expert rank ordering of overall ROP severity and the i-ROP DL vascular severity score (Spearman correlation coefficient=0.93; p<0.0001).

Conclusion The i-ROP DL system accurately identifies diagnostic categories and overall disease severity in an automated fashion, after being trained only on posterior pole vascular morphology. These data provide proof of concept that a deep learning screening platform could improve objectivity of ROP diagnosis and accessibility of screening.

Linked

► http://dx.doi.org/10.1136/bjophthalmol-2018-313290

Check for updates

To cite: Redd TK, Campbell JP, Brown JM, *et al*. *Br J Ophthalmol* 2019;**103**:580–584.

INTRODUCTION

Retinopathy of prematurity (ROP) is a leading cause of vision loss in children worldwide. Unfortunately, the global burden of this disease remains inadequately addressed, partly due to lack of access to screening. Barriers to screening include the extensive time and training required to perform these specialised ophthalmic examinations, relatively low financial compensation and significant malpractice liability.[1 2] Compounding these issues, the demand for ROP screening examinations continues to increase as the incidence of disease rises worldwide, particularly in middle-income countries.[3]

Even for those infants who do undergo screening, accurate diagnosis of ROP is difficult and requires identification of three distinct examination parameters (zone, stage and plus disease) which are combined into a composite diagnostic category.[4] It has been shown that inter-examiner diagnostic variability is high, even among expert ROP clinicians.[1] This variability results in clinically significant differences in outcomes for premature infants.[5] Appropriate diagnosis and treatment of ROP reduces the risk of progression, emphasising the importance of accurate and timely diagnosis.[4] These factors have fostered interest in artificial intelligence technologies for ROP, which have the potential to improve access to screening and facilitate standardisation of ROP diagnosis.

Artificial intelligence offers the opportunity to improve management of many medical conditions, particularly using a subset of techniques known as deep learning (DL).[6] DL is one method of training computer-based image analysis systems to automatically recognise and evaluate images and has been used successfully to diagnose a variety of ocular conditions, most notably diabetic retinopathy.[7–11] Several artificial intelligence systems have been developed to detect plus disease in ROP.[12–14]

The deep learning algorithm (DeepROP) developed by the Imaging & Informatics in ROP (i-ROP) research consortium has been incorporated into a system termed 'i-ROP DL'. This system has previously been shown to have very high accuracy for detecting plus disease from wide-angle posterior pole retinal images without the need for manual vessel segmentation and performs comparably or better than expert human examiners.[15] However, the system has been trained only to recognise plus disease. In this study, we investigate the overall clinical and public health applicability of this system by assessing its ability to identify broader diagnostic categories of ROP, as well as overall disease severity, from posterior pole images alone.

MATERIALS AND METHODS

Study population

This project was conducted as part of the multicentre i-ROP study. All data were collected

prospectively from seven participating institutions: Oregon Health and Science University, Weill Cornell Medical College, University of Miami, Columbia University, Children's Hospital Los Angeles, Cedars-Sinai Medical Centre and Asociación para Evitar la Ceguera en México. Subjects were infants who: (1) were examined in a participating neonatal intensive care unit between July 2011 and December 2016, (2) met published criteria for ROP screening examination and (3) had parents who provided informed consent. We excluded all examinations performed on eyes that had undergone prior treatment for ROP. This study was conducted in accordance with Health Insurance Portability and Accountability Act guidelines, prospectively obtained Institutional Review Board (IRB) approval from each institution and adhered to the tenets of the Declaration of Helsinki.

Disease classification and development of reference standard

Training a DL system requires a robust ground truth, or reference standard.[6] For this purpose, we utilised an overall reference standard diagnosis (RSD), an integration of ophthalmoscopic and image-based diagnoses from each ROP examination, using methods we have previously published.[16] Briefly, all infants underwent serial dilated ophthalmoscopic examinations by expert ROP clinicians according to current screening guidelines at each institution. Standard five-field retinal image sets were then obtained using a wide-angle camera (RetCam; Natus Medical Incorporated, Pleasanton, California, USA), and de-identified image sets were diagnosed by three independent experts. All ophthalmoscopic and image-based examination findings were documented according to the international classification of ROP (ICROP),[17] consisting of zone, stage and plus disease. Ophthalmoscopic and image-based diagnoses were integrated into a RSD for each image set, and diagnostic discrepancies were resolved by panel review.

The three key diagnostic parameters (zone, stage, plus disease) were then incorporated into one of the following overall disease categories based on Early Treatment for Retinopathy Of Prematurity specifications: (1) no ROP; (2) mild ROP, defined as ROP less than type 2 disease; (3) type 2 ROP (defined as zone I, stage 1 or 2, without plus disease or zone II, stage 3, without plus disease) and (4) type 1 or treatment-requiring ROP, defined as zone I, any stage, with plus disease; zone I, stage 3, without plus disease; or zone II, stage 2 or 3, with plus disease. Stages 4 and 5 ROP were excluded from this study in order to focus on identification of the onset of clinically significant disease. A category designated 'clinically significant ROP' was established to identify cases of ROP that would have warranted referral to a specialty centre. This category included type 1 ROP, type 2 ROP and pre-plus disease.

DL system development

The i-ROP DL system was developed based on the concept of convolutional neural networks using methods previously published.[15] Briefly, a reference standard diagnosis of plus disease was established by analysing 5511 wide-angle image sets from ROP screening in at-risk neonates. These images and their respective RSD for plus disease were then presented to the system in an iterative fashion. The system consisted of two consecutive neural networks, one trained for retinal vessel segmentation and the second trained to detect plus disease.[15] In the first network, images were reduced to 640×480 pixels and U-Net architecture was used to develop a 'vessel-ness' map by training on 200 manually labelled retinal images. The vessel-ness map was then used to create a circular mask of the original image. Images were then resized and cropped to 224×224 pixels. In the second network, training sets were augmented with geometric translations of the original images and then randomly sampled to achieve equal images from each class of plus disease severity in the training sets. Then, the Inception V1 (GoogLeNet) neural network architecture was used classify individual retinal images as normal, pre-plus, or plus, after pretraining on the ILSVRC ImageNet dataset. The softmax output layer was modified to perform three class prediction, and all network layers were subsequently fine tuned. The cross-entropy loss function was minimised by stochastic gradient descent (SGD) for 100 epochs, with a constant learning rate of 0.0001. A dropout rate of 0.4 was also used to mitigate overfitting. From this process, the system learnt to identify retinal image features deemed important to the diagnosis of plus disease, with near-perfect receiver operating curve characteristics.[15] Of note, information from the RSD regarding zone and stage was not provided to the system during this training. Five individual models were trained on different subsets of the overall data, and each applied to the remaining unseen data. This fivefold cross-validation method minimised bias in the output of the i-ROP DL system.

Quantitative Severity Score

For each posterior pole image, the DL system produced a set of probabilities (*P*) that the exam represented normal vessels, pre-plus disease and plus disease. To reflect the continuous spectrum of disease, we generated a scaled score from this output to represent disease severity in a given examination according to the following formula: $\{[1\times P(normal)]+[5\times P(pre\text{-}plus)]+[9\times P(plus)]\}$.[15 18] The result was termed the 'i-ROP DL score', reflecting a quantitative measurement of the degree of vascular severity on a 1–9 scale.

Data analysis

The area under the receiver operating curve (AUROC) of the i-ROP DL score was determined for all diagnostic parameters and disease classifications of ROP. The AUROC quantifies the capability of a test to classify a binary outcome, with 0.5 representing random chance and 1.0 representing a perfect test.[19] Based on these curves, a hypothetical referral cut-off score was selected for detection of type 1 ROP.

An independent data set of 100 posterior pole photos (54 normal, 31 pre-plus and 15 plus) were excluded from the training data set and used for additional validation of the system.[15 20] These examinations underwent a series of pairwise comparisons by five independent experts asked to 'select the image that represents more severe disease'. These comparisons were combined into a consensus rank ordering of the entire data set according to overall severity of ROP, from 1 (least severe) to 100 (most severe), using the Elo algorithm.[18 20] The Spearman correlation coefficient was calculated for the association between the i-ROP DL score and the expert rank order of overall ROP severity for each image. Excel 2011 (Microsoft, Redmond, Washington, USA) was used for data management, and all statistical analysis was performed using Stata MP V.13.

RESULTS

A total of 4861 individual eye examinations from 870 infants were analysed. The mean±SD birth weight and gestational age were 901±304 g and 27±2 weeks, respectively. According to the RSD, 15 examinations (3%) demonstrated type 1 ROP and 912 examinations (19%) demonstrated clinically significant ROP. Specifically, 282 (6%) eye examinations demonstrated zone I,

Table 1 Performance of the retinopathy of prematurity deep learning (i-ROP DL) system for detecting various parameters and levels of ROP

	Frequency n (%)†	AUROC (95% CI)
Diagnostic parameter*		
Plus disease	128 (3)	0.989 (0.984 to 0.994)
Stage 3	299 (6)	0.880 (0.860 to 0.899)
Stage 3 without pre-plus or plus	68 (1)	0.672 (0.628 to 0.716)
Zone I	222 (5)	0.817 (0.786 to 0.849)
Zone I without pre-plus or plus	90 (2)	0.616 (0.565 to 0.667)
Disease category*		
Clinically significant ROP	912 (19)	0.914 (0.903 to 0.925)
Type 1 ROP	155 (3)	0.960 (0.941 to 0.978)
Type 2 ROP	300 (6)	0.867 (0.854 to 0.880)
Pre-plus disease	636 (13)	0.910 (0.900 to 0.920)

Results are displayed as area under the receiver operating curve (AUROC) compared with a reference standard diagnosis.
*According to reference standard diagnosis.
†Out of total of 4861 eye examinations from 870 infants.

4469 (92%) had zone II and 110 (2%) had zone III disease. A total of 2141 (44%) demonstrated stage 0, 1168 (24%) had stage 1, 1253 (26%) had stage 2 and 299 (6%) had stage 3 disease. With respect to plus disease, 4097 (84%) had no plus, 636 (13%) had pre-plus and 128 (3%) had plus disease.

The i-ROP DL vascular severity score had AUROC 0.96 for detection of type 1 ROP and 0.91 for clinically significant ROP. For plus disease, the severity score had near-perfect AUROC (0.99 for two-level analysis, plus vs no plus) (table 1). For disease features that it was not trained to detect (zone 1 and stage 3), it was less effective: AUROC 0.82 and 0.88, respectively, but this lowered to 0.62 and 0.67 in cases when the vessels were normal (less than pre-plus disease).

Establishing a hypothetical i-ROP DL referral cut-off score of 3 for proof of concept, the i-ROP DL vascular severity would confer 94% sensitivity and 79% specificity for detection of type 1 ROP. For screening purposes, the negative predictive value (NPV) would be 99.7%, but positive predictive value (PPV) would only be 13% (figure 1). Using this cut-off in this study population, 10 cases (7%) of type 1 ROP would be missed. Nine (90%) of these cases had zone I, stage 3 disease without plus and 1 (10%) had zone I, stage 2 disease with plus according to the RSD (the i-ROP DL vascular severity score in this latter case was 2.93).

With respect to the independent data set of 100 rank ordered examinations, there was a very strong, statistically significant correlation between the expert rank ordering of overall disease severity and the i-ROP DL vascular severity score (Spearman correlation coefficient=0.93; p<0.0001). In this independent data set, the referral score of 3 would capture all cases of type 1 ROP and exclude 47 (94%) of cases with no or mild ROP (figure 2). This cut-off score would miss only one case of clinically significant ROP, which had an i-ROP DL score of 1.15 and was classified as no plus, zone I, stage 1 disease (type 2 ROP) by the RSD.

DISCUSSION

This study evaluates the performance of the i-ROP DL system for diagnosing ROP based on posterior pole fundus photographs. Key findings include: (1) despite only being trained to recognise plus disease, the i-ROP DL system accurately detects clinically significant ROP, with 94% sensitivity for type 1 ROP; (2) the i-ROP DL vascular severity score is strongly correlated with expert ranking of overall disease severity and (3) this system detects severe ROP based only on posterior pole vascular morphology, emphasising the collinearity of diagnostic parameters in ROP. These data provide proof of concept that a DL-based

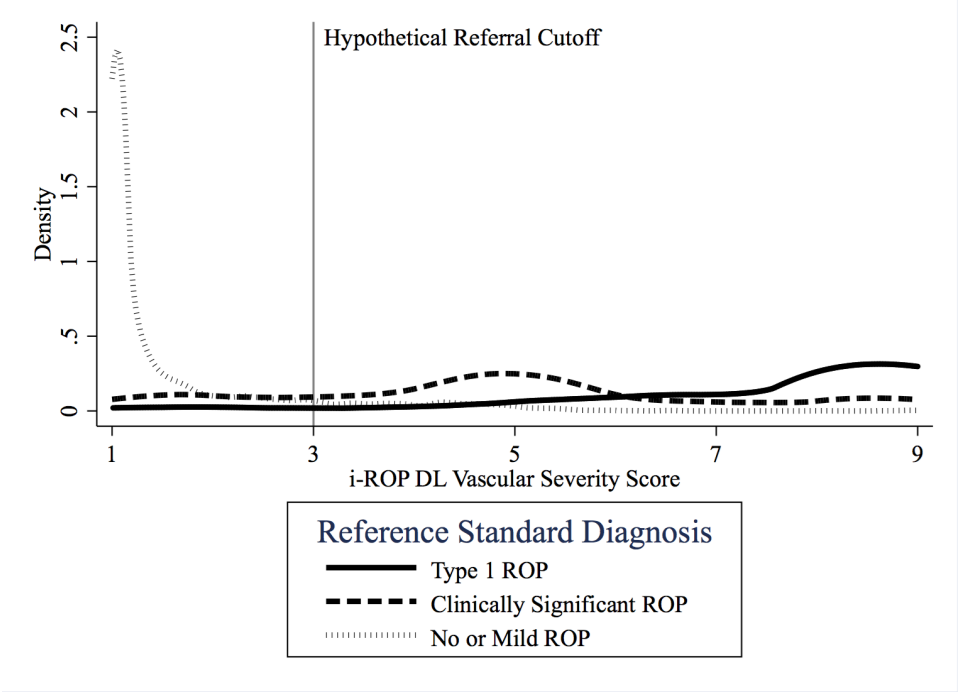

Figure 1 Distributions of Imaging & Informatics retinopathy of prematurity deep learning (i-ROP DL) vascular severity score in eye examinations with different reference standard diagnoses. Data are shown for 4861 eye examinations. In this data set, a hypothetical referral cut-off score of '3' would effectively exclude 89% of examinations with no or mild ROP, while capturing 94% of examinations with type 1 ROP.

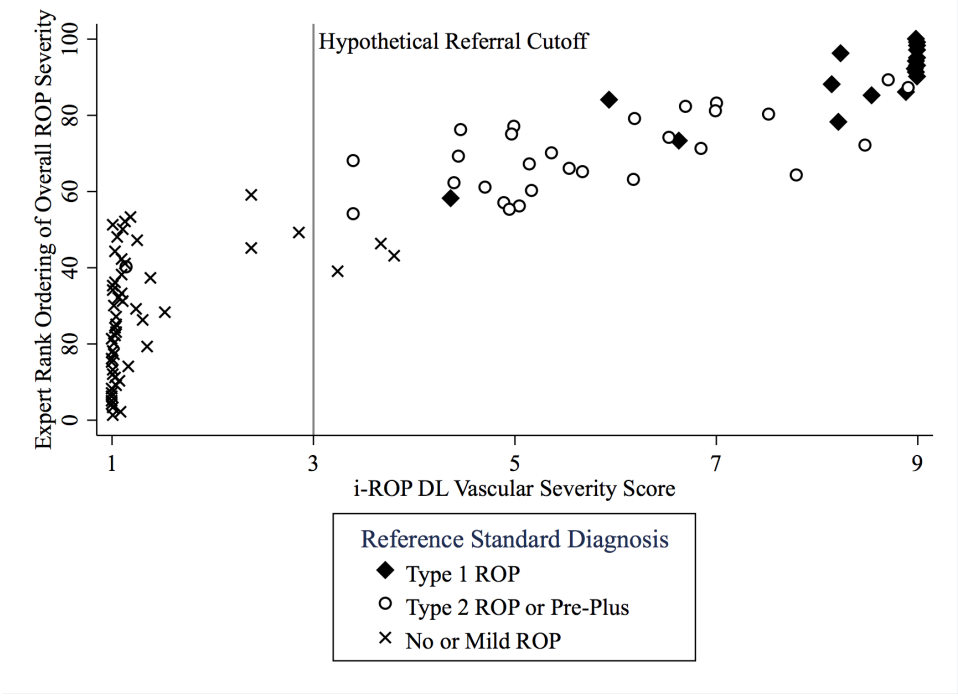

Figure 2 Association between Imaging & Informatics retinopathy of prematurity deep learning (i-ROP DL) vascular severity score and ordered ranking of overall ROP disease severity of 100 images by five experts. In this data set, a hypothetical referral cut-off score of '3' would effectively exclude 94% of cases of no or mild ROP, while capturing 100% of cases of type 1 ROP.

screening platform could be deployed to improve the objectivity of ROP diagnosis and improve access to screening.

The first key finding is that the i-ROP DL system has high accuracy for detecting clinically significant ROP (table 1). We have previously shown that this system has very high accuracy for plus disease in ROP.[15] These results extend those findings beyond plus disease and demonstrate that despite only being trained to recognise plus disease, the system has high accuracy for broader diagnostic categories of ROP, particularly severe disease. A hypothetical referral score of '3' has a sensitivity of 94% for type 1 ROP (figure 1). More importantly, the negative predictive value is 99.7%, meaning that a posterior pole image scoring less than this threshold value would have only a 3 in 1000 chance of being type 1 ROP. The sensitivity and NPV are the most important parameters in a screening test for a disease such as ROP, where underdiagnosis has critical implications.[21]

The second key finding is that the i-ROP DL vascular severity score correlates with the continuum of disease severity as determined by expert graders (figure 2).[18] We have previously demonstrated that ROP phenotypes appear to run a continuum from mild to severe and that experts agree on relative disease severity better than they agree on zone, stage, plus or overall category.[18 20] This study reaffirms the concept of a continuous spectrum of disease and provides an automated and accurate method of measuring this continuum. This has implications for disease screening as suggested in figure 1, as well as for following disease progression over time.

The third key finding is that the severity score generated by the i-ROP DL system detects severe ROP based only on posterior pole vascular morphology. The fact that it achieves this without being trained to detect zone or stage suggests that severe ROP rarely occurs in the absence of detectable changes in the posterior vasculature, which is supported by the literature.[4] In this population, a cut-off score of 3 missed 10 cases (7%) of type 1 ROP, all of which would have been detected if the system were additionally trained to detect zone I and stage 3 disease.

It is conceivable that future iterations of this system could provide an automated point-of-care screening test to identify patients with clinically significant ROP who require full ophthalmoscopic evaluation. To be successful, the sensitivity will need to be higher (so that no cases of type 1 disease are missed) and we will need to reconceptualise current ROP screening models relying on full ICROP classification. If all patients with ROP in need of urgent intervention could be identified, the rest could be rescreened in a defined time period (eg, 1–2 weeks) for objective automated signs of disease progression. There is preliminary evidence that the i-ROP DL system can accurately identify this progression.[22] The Food and Drug Administration recently approved the first DL-based system for health screening to detect referable diabetic retinopathy.[23] Utilising the i-ROP DL system in a similar fashion might greatly reduce the number of ophthalmoscopic exams needed to adequately screen for ROP, addressing the high and growing demand for screening worldwide.[3]

This study has several limitations. First, the definition of clinically significant ROP is a composite of type 2 ROP, type 1 ROP and pre-plus disease. The justification for including pre-plus in this category was that the concept of pre-plus was not introduced until 2005,[17] after the definition of type 2 ROP was established.[4] Pre-plus disease has since been shown to be a strong independent risk factor for progression of disease[24] and thus should necessitate referral for specialist examination. Second, an inherent problem in the evaluation of any ROP screening system is the interexpert variability in the diagnosis of ROP. For this study, we utilised a consensus RSD both for system training and evaluation. Third, the accuracy of any artificial intelligence system is dependent on the quality of data it presents. In this study, images of inadequate quality for clinical diagnosis were excluded. We

are developing software to automatically determine if images are of sufficient quality.[25]

This study provides proof of concept for the use of artificial intelligence in autonomous or assistive ROP screening in the future. The i-ROP DL system demonstrates high sensitivity for detection of clinically significant ROP based only on posterior pole photography, strengthens the case for the development and validation of a continuous ROP severity score and may have important applications for ROP care in resource-limited settings.

Author affiliations
[1]Department of Ophthalmology, Casey Eye Institute, Oregon Health & Science University, Portland, Oregon, USA
[2]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, Maryland, USA
[3]Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[4]Department of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois at Chicago, Chicago, Illinois, USA
[5]Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts, USA
[6]Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon, USA

Contributors JB, JD, DE, SI and JKC developed the artificial intelligence system evaluated in this study. TR, JPC and MFC conceived the study design and drafted the manuscript. SO and JPC gathered, cleaned and organised the data. TR performed all data analysis. JB, SJK, RVPC and JKC performed critical revision of the manuscript.

Funding Supported by grants R01EY19474, K12 EY027720, P30EY10572, and P30EY001792 from the National Institutes of Health (Bethesda, Maryland, USA) by grants SCH-1622679, SCH-1622542, and SCH-1622536 from the National Science Foundation (Arlington, Virginia, USA), and by unrestricted departmental funding from Research to Prevent Blindness (New York, New York , USA).

Competing interests MFC is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, California, USA), a Consultant for Novartis (Basel, Switzerland) and an initial member of Inteleretina, LLC (Honolulu, Hawaii, USA). RVPC is a Scientific Advisory Board member for Visunex Medical Systems (Fremont, California, USA) and a Consultant for Alcon (Fort Worth, Texas, USA), Allergan (Irvine, California, USA) and Bausch and Lomb (St. Louis, Missouri, USA). JPC is a consultant to Allergan (Irvine, California, USA).

Patient consent Patient/gaurdian consent obtained.

Ethics approval Oregon Health and Science University Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Chiang MF, Jiang L, Gelman R, *et al*. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. *Arch Ophthalmol* 2007;125:875–80.
2 Braverman RS, Enzenauer RW. Socioeconomics of retinopathy of prematurity in-hospital care. *Arch Ophthalmol* 2010;128:1055–8.
3 Sommer A, Taylor HR, Ravilla TD, *et al*. Challenges of ophthalmic care in the developing world. *JAMA Ophthalmol* 2014;132:640–4.
4 Early Treatment For Retinopathy Of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. *Arch Ophthalmol* 2003;121:1684–96.
5 Fleck BW, Williams C, Juszczak E, *et al*. An international comparison of retinopathy of prematurity grading performance within the Benefits of Oxygen Saturation Targeting II trials. *Eye* 2018;32:74–80.
6 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
7 Gulshan V, Peng L, Coram M, *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402–10.
8 Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017;124:962–9.
9 Abràmoff MD, Lavin PT, Birch M, *et al*. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39.
10 Ting DSW, Cheung CY, Lim G, *et al*. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318:2211–23.
11 De Fauw J, Ledsam JR, Romera-Paredes B, *et al*. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50.
12 Wittenberg LA, Jonsson NJ, Chan RV, *et al*. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus* 2012;49:11–19.
13 Heneghan C, Flynn J, O'Keefe M, *et al*. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Med Image Anal* 2002;6:407–29.
14 Wallace DK, Zhao Z, Freedman SF. A pilot study using "ROPtool" to quantify plus disease in retinopathy of prematurity. *J Aapos* 2007;11:381–7.
15 Brown JM, Campbell JP, Beers A, *et al*. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol* 2018;136:803–10.
16 Ryan MC, Ostmo S, Jonas K, *et al*. Development and evaluation of reference standards for image-based telemedicine diagnosis and clinical research studies in ophthalmology. *AMIA Annu Symp Proc* 2014;2014:1902–10.
17 International Committee for the Classification of Retinopathy of Prematurity. The international classification of retinopathy of prematurity revisited. *Arch Ophthalmol* 2005;123:991–9.
18 Campbell JP, Kalpathy-Cramer J, Erdogmus D, *et al*. Plus disease in retinopathy of prematurity: a continuous spectrum of vascular abnormality as a basis of diagnostic variability. *Ophthalmology* 2016;123:2338–44.
19 Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561–77.
20 Kalpathy-Cramer J, Campbell JP, Erdogmus D, *et al*. Plus disease in retinopathy of prematurity: improving diagnosis by ranking disease severity and using quantitative image analysis. *Ophthalmology* 2016;123:2345–51.
21 Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. *Inhal Toxicol* 2014;26:811–28.
22 Brown JM, Kalpathy-Cramer J, Campbell JP. Fully automated disease severity assessment and treatment monitoring in retinopathy of prematurity using deep learning. *Proc SPIE Med Imaging 2018 Imaging Informatics Heal Res Appl*;10579.
23 News Release FDA, 2018. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Available from: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm [Accessed Aug 2018].
24 Wallace DK, Freedman SF, Hartnett ME, *et al*. Predictive value of pre-plus disease in retinopathy of prematurity. *Arch Ophthalmol* 2011;129:591–6.
25 Coyner A, Swan R, Brown JM. Deep learning for image quality assessment of fundus images in retinopathy of prematurity. *AMIA Annu Symp Proc*. In press.