Article

Deep Learning Algorithm for Automated Diagnosis of Retinopathy of Prematurity Plus Disease

Zachary Tan[1,2,3,*], Samantha Simkin[4,*], Connie Lai[5,6,*], and Shuan Dai[4,7,*]

[1] Save Sight Institute, The University of Sydney, Sydney, New South Wales, Australia
[2] St Vincent's Hospital Sydney, Sydney, New South Wales, Australia
[3] Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia
[4] Department of Ophthalmology, New Zealand National Eye Centre, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand
[5] Queen Mary Hospital, Hong Kong, China
[6] Department of Ophthalmology, The University of Hong Kong, Hong Kong, China
[7] Department of Ophthalmology, Queensland Children's Hospital, Brisbane, Queensland, Australia

Correspondence: Shuan Dai, Queensland Children's Hospital, Department of Ophthalmology, 501 Stanley St., South Brisbane, Brisbane, Queensland 4101, Australia. e-mail: shuandai@me.com

Received: 16 April 2019
Accepted: 30 September 2019
Published: 2 December 2019

Keywords: deep learning; artificial intelligence; retinopathy of prematurity

Citation: Tan Z, Simkin S, Lai C, Dai S. Deep learning algorithm for automated diagnosis of retinopathy of prematurity plus disease. Trans Vis Sci Tech. 2019;8(6):23, https://doi.org/10.1167/tvst.8.6.23

Purpose: This study describes the initial development of a deep learning algorithm, ROP.AI, to automatically diagnose retinopathy of prematurity (ROP) plus disease in fundal images.

Methods: ROP.AI was trained using 6974 fundal images from Australasian image databases. Each image was given a diagnosis as part of real-world routine ROP screening and classified as normal or plus disease. The algorithm was trained using 80% of the images and validated against the remaining 20% within a hold-out test set. Performance in diagnosing plus disease was evaluated against an external set of 90 images. Performance in detecting pre-plus disease was also tested. As a screening tool, the algorithm's operating point was optimized for sensitivity and negative predictive value, and its performance reevaluated.

Results: For plus disease diagnosis within the 20% hold-out test set, the algorithm achieved a 96.6% sensitivity, 98.0% specificity, and 97.3% ± 0.7% accuracy. Area under the receiver operating characteristic curve was 0.993. Within the independent test set, the algorithm achieved a 93.9% sensitivity, 80.7% specificity, and 95.8% negative predictive value. For detection of pre-plus and plus disease, the algorithm achieved 81.4% sensitivity, 80.7% specificity, and 80.7% negative predictive value. Following the identification of an optimized operating point, the algorithm diagnosed plus disease with a 97.0% sensitivity and 97.8% negative predictive value.

Conclusions: ROP.AI is a deep learning algorithm able to automatically diagnose ROP plus disease with high sensitivity and negative predictive value.

Translational Relevance: In the context of increasing global disease burden, future development may improve access to ROP diagnosis and care.

Introduction

Retinopathy of prematurity (ROP) is a vasoproliferative disorder affecting the retinae of premature infants with low birthweights. Although most ROP settles without significant visual sequelae, up to 5% of ROP will require timely treatment to avoid permanent visual loss. ROP remains one of the leading causes of preventable childhood blindness globally despite improvements in ophthalmic care.[1] Blindness from ROP is largely preventable with early case detection and timely treatment,[2,3] although these can be challenging.

In high-income countries, advances in neonatal care and screening protocols have resulted in ROP occurring mostly in infants born extremely premature. However, improving neonatal care in middle-income countries has led to the improved survival of premature infants without accompanying improvements in infrastructure for ROP case detection and treatment.[4,5] This has led to a global increase in disease burden, described as the "third-epidemic" of ROP.[4]

Access to ROP screening is limited due to the need for highly specialized personnel and the time-consuming nature of its clinical practice, as well as significant associated medicolegal risk.[6,7] Screening examinations have traditionally been carried out by a trained ophthalmologist using a binocular indirect ophthalmoscope. Modern modalities of ROP screening utilize digital retinal imaging and remote diagnosis via telemedicine by specialized clinicians, which have improved both access to screening and objectivity in diagnosis.[8–13] Common across the various screening modalities is the effort to identify the presence of treatment-requiring ROP. The Early Treatment for Retinopathy of Prematurity (ETROP) study redefined treatment-requiring and observational ROP as "type 1" and "type 2" ROP, respectively.[3,14] Type 1 ROP is defined as (1) any stage of ROP in zone I with plus disease, (2) stage 3 ROP in zone I without plus disease, or (3) stages 2 or 3 ROP in zone II with plus disease. Type 2 ROP is defined as stages 1 or 2 ROP in zone I without plus disease or stage 3 ROP in zone II without plus disease.

Type 1 ROP and indication for treatment is, thus, distinguished predominantly by the presence of plus disease. Plus disease was defined in the 1980s by an international consensus panel as arterial tortuosity and venous dilation of the posterior retinal vessels greater than or equal to that of a standard published fundal photograph.[15] More recently, a further intermediate level of pre-plus disease has been described, defined by a level of vascular dilation and tortuosity in between that of normal posterior pole vasculature but less so than plus disease.[16] Following the identification of type 1 ROP, treatment may be delivered via laser photocoagulation or intravitreal injection of anti-vascular endothelial growth factor agents.[14,17]

Significant interclinician subjectivity and regional variation in the diagnosis of plus disease is well documented and may lead to delayed treatment and poorer visual outcomes.[18–20] Diagnostic variation may be due to individual ophthalmologists evaluating differing features, focusing on wider fields of view than the standard photograph for diagnosis, or having different cutoff points for vascular abnormality required for determination of plus disease.[21] Several studies have demonstrated mild to moderate inter- and intraexpert agreement in the diagnosis of plus disease.[19,21,22] Geographical variation in diagnosis has also been established, with diagnosis rates of treatment-requiring ROP significantly lower in Australia and New Zealand (7.7% and 7.5%, respectively) than that of international counterparts in the United Kingdom (19.2%), Canada, and the United States (13.0%) in a recent multinational trial.[20]

In this global context of increasing disease burden, limited access to specialized case detection, and regional variabilities, several research groups have investigated the development of computer-based image analysis (CBIA) for the automated diagnosis of ROP plus disease.[23–26] Previously published CBIA tools have evaluated retinal image features selected a priori to reach a diagnostic conclusion. Various systems[27], including ROPtool,[24] Retinal Image multiscale Analysis,[28] and early iterations of the i-ROP tool, have been able to demonstrate >90% sensitivity for the detection of plus disease in two-level (normal versus plus disease) classification.

More recently, fully automated techniques utilizing artificial intelligence (AI) deep learning technologies have also been validated in ROP by groups based in the United States[29] and China.[25] The performance of these systems has matured, achieving diagnostic performance comparable or exceeding those of human graders in the diagnosis of plus disease, attracting significant attention.[30]

This study describes the initial development of a deep learning algorithm, ROP.AI, that can automatically diagnose the presence of ROP plus disease. Fundal images for this algorithm were sourced from infants in New Zealand, which have markedly different demographic and clinical features to those internationally. Dissimilarities include ethnic heterogeneity, with infants of European (39.0%), Māori (the indigenous people of New Zealand; 23.9%), Pacific Peoples (18.0%), and Asian (including East and South Asians; 16.9%) backgrounds.[8] Furthermore, New Zealand ROP screening guidelines (<1250 g birth weight or <30 weeks gestational age)[31] are more restrictive than international guidelines[32,33] and infants screened were likely, on average, to be smaller and more premature.

Given these clinical differences and significant variations in the rate of plus disease diagnosis in Australia and New Zealand,[20] ROP.AI is the first algorithm trained using fundal images sourced from local image databases.

Methods

Ethics

This study was approved by the New Zealand Health and Disability Ethics Committee (approval number 14/NTA/183/AM08) and adhered to the tenets of the Declaration of Helsinki.

Data Sets

Deidentified fundal images were sourced from the Auckland Regional Telemedicine ROP (ART-ROP) image library, a database of images generated from routine ROP screening across four neonatal intensive care units in Auckland, New Zealand. Images used were captured from 2006–2015, inclusive. All fundal images were photographed using a commercially available camera at a standard field of view of 130° and resolution of 640 by 480 pixels (RetCam; Natus Medical Incorporated, Pleasanton, CA).

Diagnoses were supplied by the ART-ROP image library, which were given by author SD as part of real-world routine ROP screening and clinical care. Diagnoses were not added or altered, and images were classified as either normal retina or plus disease. Fundal images with pre-plus disease were not available in the supplied image library.

A total of 4926 deidentified fundal images were initially received, and all images were manually graded by author ZT for image quality. Criteria for image quality inclusion were that (1) the image was not grossly out of focus and (2) the image was not affected by blur. Following grading, 3487 fundal images were included for preprocessing and data augmentation. Preprocessing comprised of crop to remove image text annotations, and data augmentation with horizontal flips were carried out to double the number of images for the final training set.

ROP.AI Algorithm Development

Images were uploaded with the desktop Safari web browser (Apple Inc, Cupertino, CA) to a cloud-based deep learning platform (MedicMind, https://ai.medicmind.tech; MedicMind, Dunedin, New Zealand) utilizing TensorFlow's Inception-v3 (Alphabet Inc., Mountain View, CA) convolutional neural network (CNN).[34] CNNs are an advanced AI deep learning technology specialized for image recognition. CNNs operate by learning and applying a series of filters that emphasize image features that are relevant to the task at hand.

Graded classification was used, with normal and plus disease fundal images uploaded to separate bins. The algorithm was trained to diagnose only the presence of plus disease. The RMSProp optimizer was used, with a weight decay factor of 0.00004 and momentum of 0.9. The learning rate was exponentially decayed with a decay factor of 0.16, 30 epochs per decay, and an initial learning rate of 0.1. Binary cross-entropy loss was used for training. The batch size was 16. A softmax output layer was used to produce probabilistic outputs.

Training and validation of the CNN was performed by cloud-based graphics processing units hosted by Amazon Web Services (Amazon Inc., Seattle, WA).

Internal Validation

Images were randomly assigned in an 80:20 split to training and hold-out test sets. The ratio of normal to plus disease fundal images were, thus, expected to be comparable between image sets. The ROP.AI algorithm was trained on fundal images within the training set and evaluated on images within the hold-out test set. This internal validation assessed the algorithm's ability to diagnose plus disease on previously unseen images from the ART-ROP image library. Statistical performance for the classifier was measured by calculating sensitivity, specificity, accuracy, receiver operator characteristic (ROC) curve, and area under receiver operating characteristic curve (AUROC).

External Validation

The performance of the algorithm was subsequently evaluated against an external test set of 90 fundal images. These images were not included in the training and internal validation hold-out sets and were provided by author CL. Diagnoses were given as part of real-world routine ROP screening and clinical care and not subsequently added or altered. Of this external test set, 57 fundal images were of normal retina and 33 of plus disease.

Fundal images were uploaded to the ROP.AI algorithm. The statistical performance of the classifier in diagnosing plus disease was measured by calculating sensitivity, specificity, accuracy, positive predictive value, and negative predictive value.

Detection of Pre-plus Disease

ROP plus disease likely exists as a continuous spectrum of retinal vascular abnormality.[21,29] Thus, the performance of the algorithm in detecting intermediate and less severe pre-plus disease was evaluated. The statistical performance of the classifier was evaluated in an expanded external test set of 116 fundal images, which comprised of the initial external test set of 57 normal retina and 33 plus disease images, with an additional 26 pre-plus disease fundal images. Statistical performance was measured by calculating sensitivity, specificity, accuracy, positive predictive value, and negative predictive value.

Operating Point Optimization

ROP.AI returns a probability value following evaluation of a fundal image. Values between 0 and 1.00 are returned. By default, an operating point threshold of 0.50 is used, with values above this point indicative of plus disease.

As the algorithm can have multiple operating points, its performance can be tuned to match the requirements for specific clinical settings.[35] Given the consequences of missing a ROP diagnosis, high sensitivity and negative predictive value are required in a screening setting. Thus, the algorithm was subsequently tested against its normal retina/plus disease external validation set at 0.01 operating point intervals between 0.01 and 1.00. An operating point optimized for sensitivity and negative predictive value was subsequently identified.

Results

In total, we obtained 4926 deidentified fundal images from the ART-ROP image database (Fig. 1). Of the 4926 supplied fundal images, 3792 (77.0%) were of normal retina and 1134 (23.0%) plus disease. As images were deidentified, the unique number of infants is unknown, although are likely to have been obtained from >300 total unique imaging sessions.

Following manual grading, there were 1439 images that did not meet initial image quality criteria and were excluded from the ROP.AI training set. A total of 3487 fundal images were subsequently identified as suitable for training. Following image preprocessing to remove image text annotations and data augmentation with horizontal flips, 6974 fundal images were included in the ROP.AI algorithm training set.

Figure 2 displays the ROC curve for the ROP.AI algorithm tested against its internal validation set ($n = 1395$, 20% of the provided data). The AUROC for the diagnosis of plus disease was 0.993.

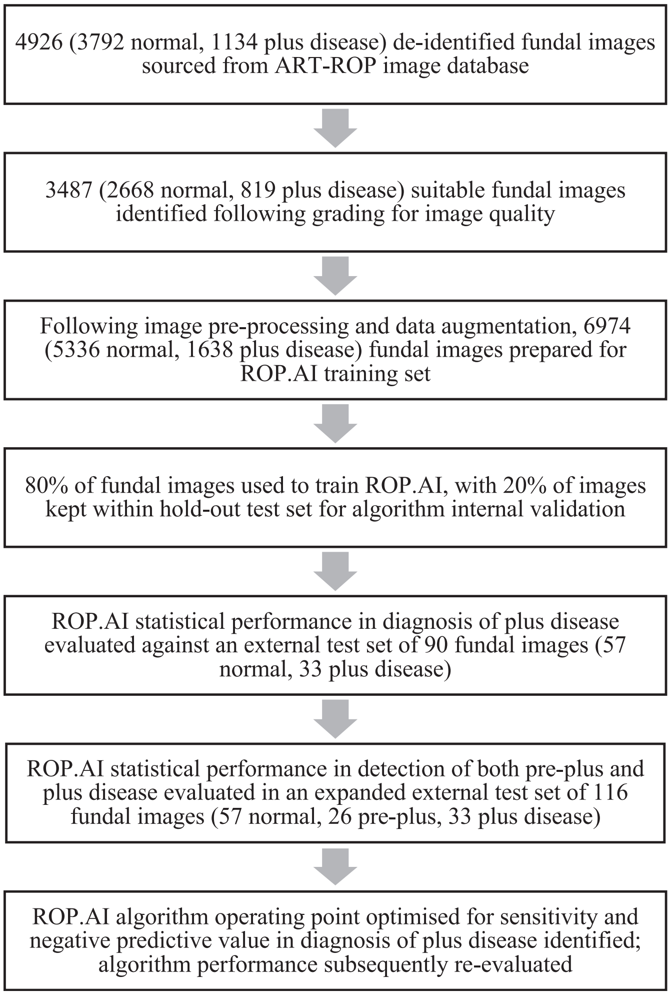
4926 (3792 normal, 1134 plus disease) de-identified fundal images sourced from ART-ROP image database
3487 (2668 normal, 819 plus disease) suitable fundal images identified following grading for image quality
Following image pre-processing and data augmentation, 6974 (5336 normal, 1638 plus disease) fundal images prepared for ROP.AI training set
80% of fundal images used to train ROP.AI, with 20% of images kept within hold-out test set for algorithm internal validation
ROP.AI statistical performance in diagnosis of plus disease evaluated against an external test set of 90 fundal images (57 normal, 33 plus disease)
ROP.AI statistical performance in detection of both pre-plus and plus disease evaluated in an expanded external test set of 116 fundal images (57 normal, 26 pre-plus, 33 plus disease)
ROP.AI algorithm operating point optimised for sensitivity and negative predictive value in diagnosis of plus disease identified; algorithm performance subsequently re-evaluated

Figure 1. Development of initial training set and algorithm validation process.

On internal validation, the sensitivity and specificity of the ROP.AI algorithm were 96.6% and 98.0% respectively, providing an overall accuracy of 97.3% (96.6%–98.0%, 95% confidence interval).

External Validation

The ROP.AI algorithm was evaluated against an external test set of 90 images, which it had not been trained with or encountered previously. Figure 3 displays the ROC curve for this validation. The AUROC for the diagnosis of plus disease was 0.977.

Using a default operating point threshold of 0.50, the algorithm provided a sensitivity and specificity of 93.9% and 80.7%, respectively. Overall accuracy, positive predictive value, and negative predictive value of 85.6%, 73.8% and 95.8% was achieved.

The orange diamond highlights the performance of

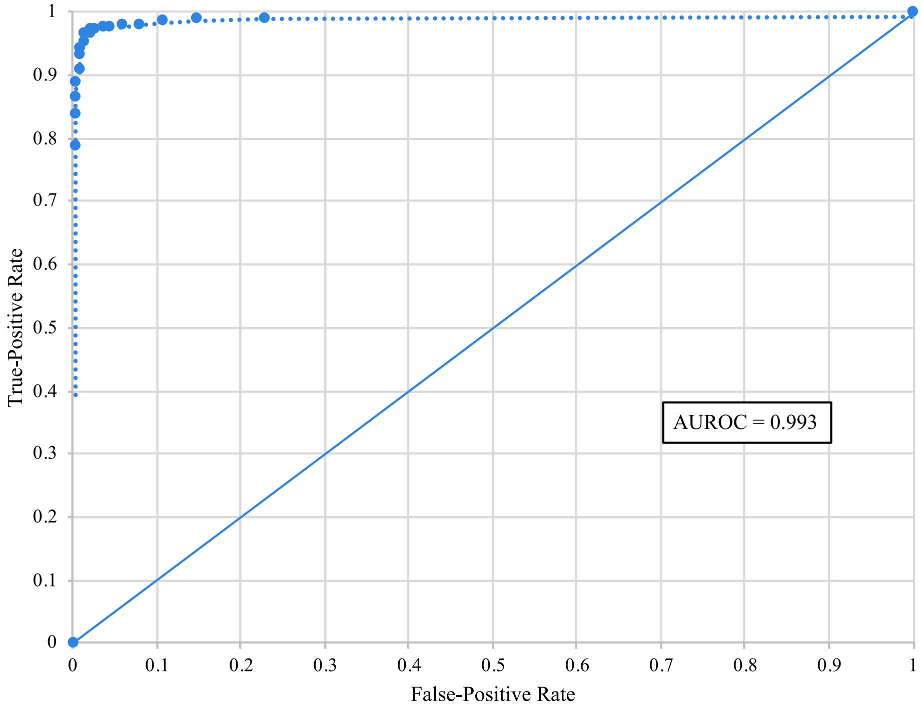

Figure 2. Receiver operator characteristic (ROC) curve for algorithm diagnosis of retinopathy of prematurity plus disease.

the algorithm at the optimized operating point identified for high sensitivity.

Detection of Pre-plus Disease

Despite being trained on normal retina and plus disease fundal images only, the performance of the ROP.AI algorithm in detecting both pre-plus and plus disease was evaluated in an expanded external test set of 57 normal, 26 pre-plus, and 33 plus disease images. Using a default operating point threshold of 0.50, the algorithm provided a sensitivity and specificity of 81.4% and 80.7%, respectively. Overall accuracy, positive predictive value, and negative predictive value of 81.0%, 81.4%, and 80.7% was achieved.

The average outputs produced by the algorithm for normal, pre-plus, and plus disease images were 0.23, 0.65, and 0.93, respectively. The distribution of these probability outputs is illustrated in the violin plot in Figure 4.

The violin plot shows the distribution of probability outputs produced by the algorithm for normal, pre-plus, and plus disease fundal images. The 25th, 50th, and 75th percentile outputs for normal, pre-plus, and plus disease fundal images are 0.002, 0.088, and 0.317; 0.387, 0.760, and 0.879; and 0.963, 1.00, and 1.00, respectively. The operating point optimized for high sensitivity is shown with the horizontal line at 0.38.

Operating Point Optimization

The algorithm was subsequently tested against its normal retina/plus disease external validation set at 0.01 operating point intervals between 0.01 and 1.00 (Fig. 5). Given the consequences of missing a ROP diagnosis, high sensitivity and negative predictive value are required in a screening setting. An optimized operating point of 0.38, which maximized these values, was subsequently identified.

Performance metrics at the optimized operating point of 0.38 are highlighted.

With an optimized operating point of 0.38, the sensitivity and specificity values for diagnosing plus disease were 97.0% and 78.9%, respectively. Overall accuracy, positive predictive value, and negative predictive value of 85.6%, 72.7%, and 97.8% was achieved.

Discussion

This study describes the initial development of a deep learning algorithm, ROP.AI, trained to automatically diagnose ROP plus disease. Our results have shown that a deep learning algorithm can successfully diagnose this form of treatment-requiring ROP with high accuracy. Over the last decade, digital retinal imaging to screen for ROP has been implemented in

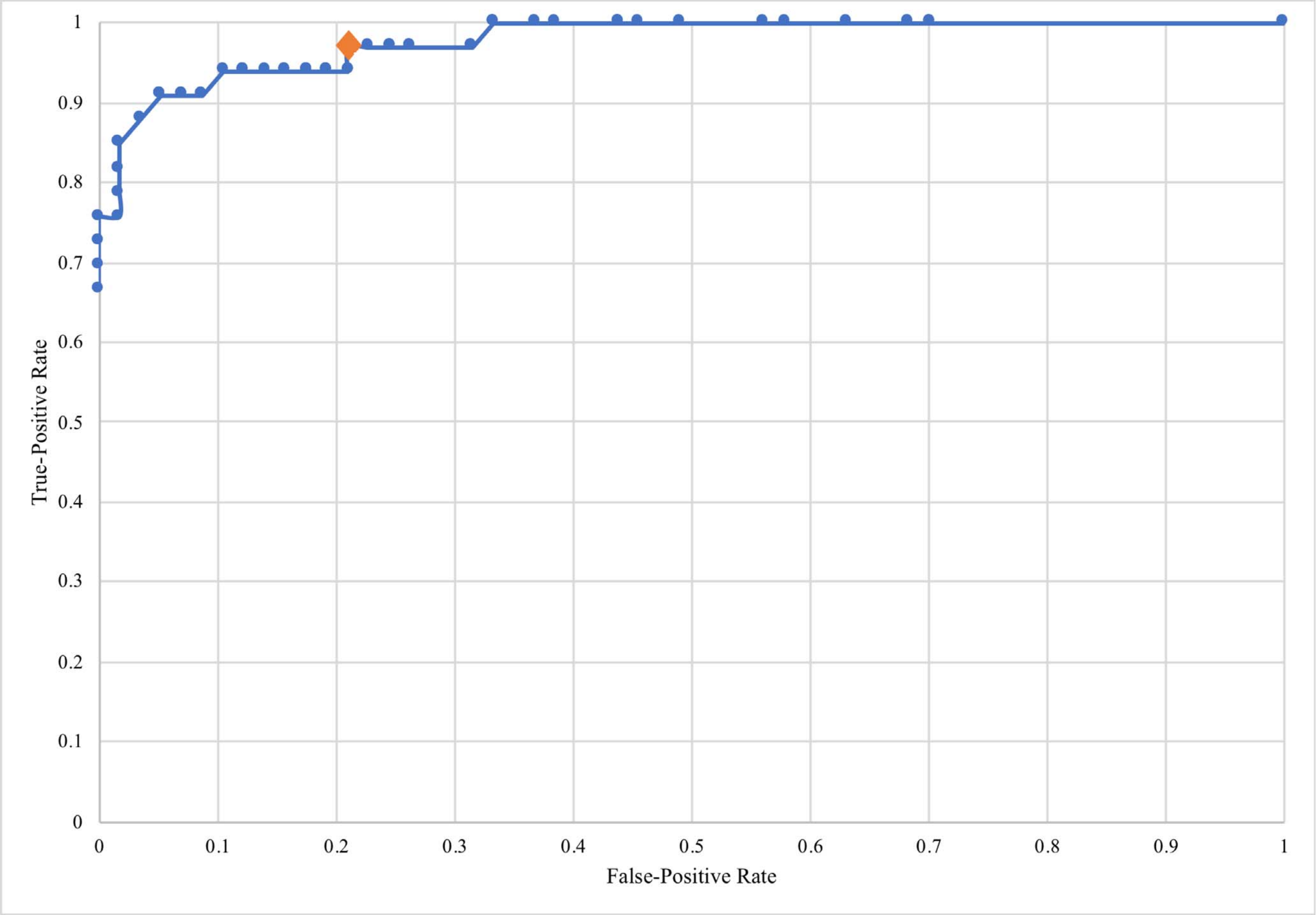

Figure 3. Receiver operator characteristic (ROC) curve for algorithm performance in the external validation set.

numerous Australian and New Zealand centers, with large fundal image databases now available. Given the significant differences in infant demographic features and well-documented regional variations in the threshold for plus disease diagnosis,[20] ROP.AI is the first algorithm trained using fundal images sourced from local databases. The performance of this algorithm was subsequently validated against images from an external test set obtained from a separate geography, in Hong Kong.

Early CBIA tools developed for the diagnosis of ROP required the manual identification of the optic disc and segmentation of key vessel segments.[24,26] Our algorithm is a refinement on these earlier tools, uses advanced deep learning techniques, and is the first fully automated tool for ROP diagnosis from Australasia.

A key finding is that this algorithm is able to diagnose ROP plus disease with high accuracy. Following operating point optimization, the algorithm had a 97.0% sensitivity and 97.8% negative predictive value in the diagnosis of plus disease (Fig. 5). This statistical performance is comparable with those recently reported by international groups.[25,29] In the clinical context of screening, high sensitivity and negative predictive value are critical to avoid missed and underdiagnosed ROP. This is particularly acute in ROP where delayed case detection and clinical intervention may have severe visual impairment and medicolegal sequelae.

Furthermore, despite being initially trained to diagnose plus disease only, the algorithm has shown promise also in the detection of pre-plus disease, an intermediate, less severe disease state. Following evaluation of a fundal image, ROP.AI generates a continuous number between 0 and 1.00, reflecting the probability of plus disease present. Average probability values returned for fundal images diagnosed clinically as normal, pre-plus, and plus disease were 0.23, 0.65, and 0.93, respectively.

Although the number of pre-plus images available for evaluation was relatively small, this may support

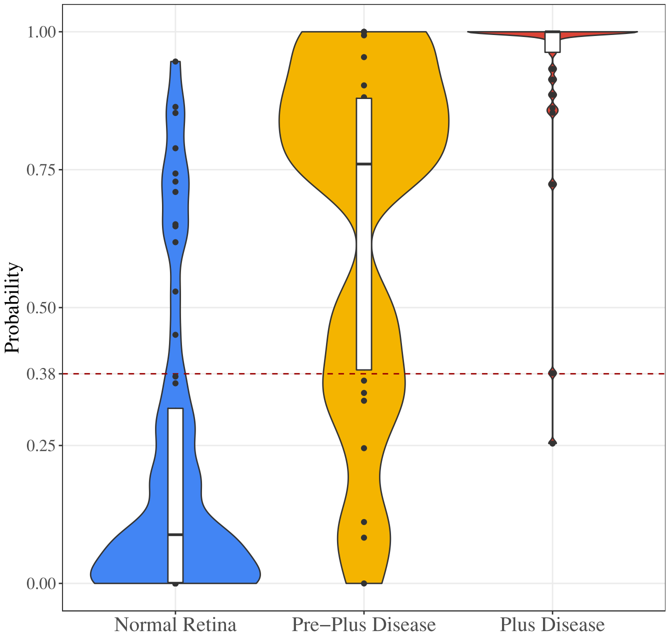

Figure 4. Violin plot for algorithm performance in normal, pre-plus, and plus disease fundal images.

reports that plus disease likely exists as a continuous spectrum of retinal vascular abnormality (Fig. 4),[29,36] a hypothesis also suggested in recent ROP deep learning literature.[29,37,38] With significant further development and validation, the algorithm may be able to assist in the objective quantification of ROP disease severity and the monitoring of disease progression.

Limitations

This initial study has several key limitations. The ROP.AI algorithm was trained using fundal images from only a single institution and clinician-provided diagnoses. CNNs use "supervised learning" where performance is only as robust as the quality of the input data. Given known interclinician variations in the threshold of plus disease diagnosis,[20] this algorithm likely reflects the diagnostic preferences of the single treating clinician and institution. This may result in overfitting in internal validation and may restrict the overall applicability of the algorithm in its current stage of development. Despite this limitation, the algorithm was still able to perform with high accuracy in external validation against a test set of

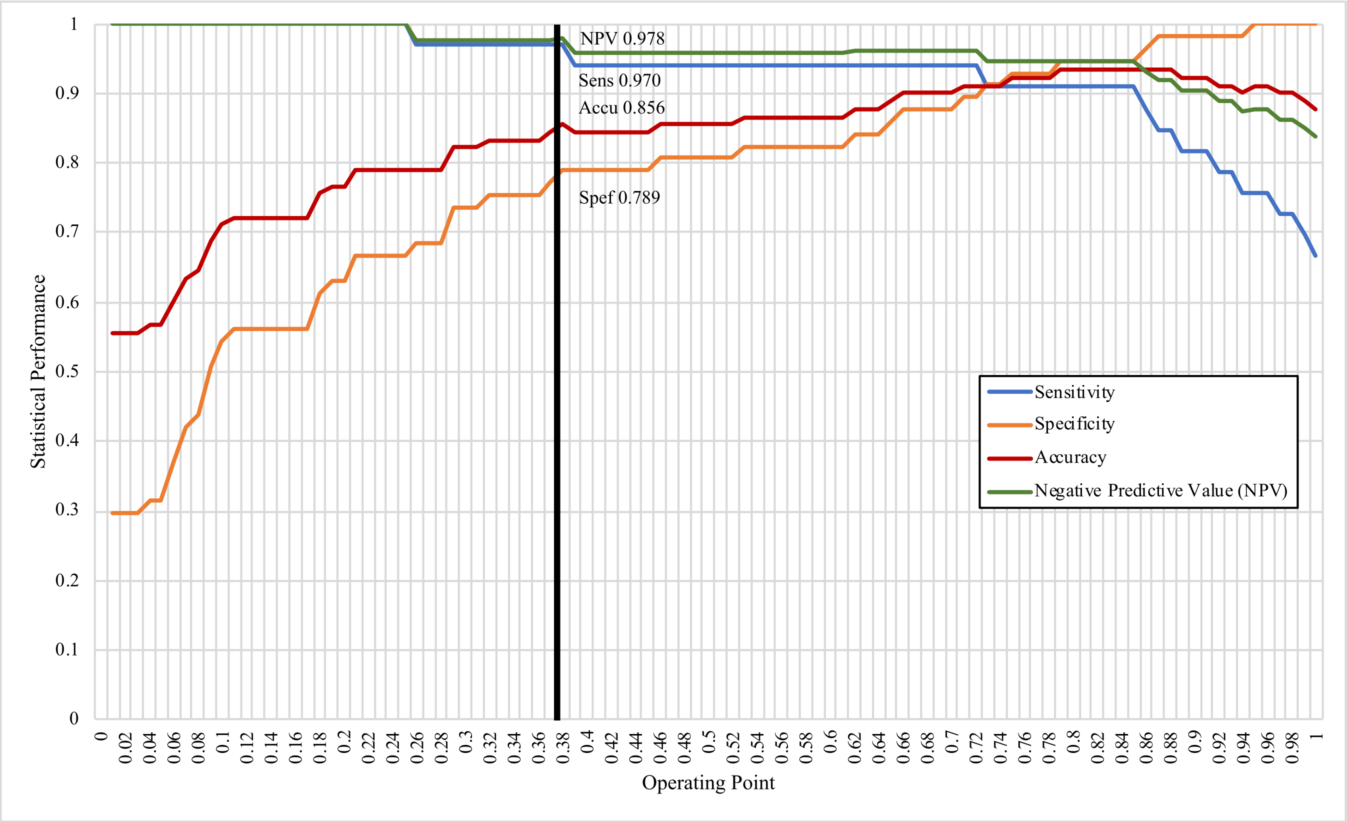

Figure 5. Statistical performance of the algorithm at operating point intervals between 0.01 and 1.00.

fundal images provided by an independent clinician and institution. Further training with images from multiple clinicians and institutions and the development of a training image set with consensus expert diagnoses may mitigate the impact of individual clinician diagnostic preferences in the future.

Furthermore, a significant proportion of fundal images initially sourced from the ART-ROP database were of insufficient image quality (1439 of 4926 initial images; Fig. 1). Images were predominantly excluded due to motion blur or other artifacts. These images reflect the challenging nature of the real-world capture of fundal images in ROP screening. Potential future systems for automated ROP diagnosis will likely require a layer to assess for fundal image quality before evaluation for diagnosis.

The algorithm was trained with fundal images from Australia and New Zealand and tested against an external test set obtained from a separate region, in Hong Kong. Although our fundal images are sourced from a network with a significant proportion of Asian infants,[8] ethnic compositions were likely to significantly differ between the training and external test sets, with differences in retinal pigmentation potentially adversely affecting the accuracy of these results.

Our algorithm was developed on a postdata augmentation training set of 6974 fundal images, including 5336 normal and 1638 plus disease images. The relatively low number of plus disease images may result in algorithm overfitting, resulting in erroneously high statistical performance reported on internal validation (Fig. 2). The size of this training image set, however, is consistent with those reported by other groups.[25,29,36] The total number of plus disease images available to train deep learning algorithms in ROP may be limited due to the relatively low incidence of plus disease in centers where digital retinal imaging is currently available and where neonatal care is sufficiently progressed to largely limit the development of plus disease.[39]

Furthermore, the ratio of plus disease to normal fundal images in the image sets included in our study is much higher than the real-world prevalence of plus disease, which may be as low as $<5\%$.[20] This may skew the reported sensitivity and negative predictive values, with the algorithm potentially not performing as effectively in a real-world setting. If the training data set of the size used in this study reflected real-world prevalence of plus disease, there would likely have been insufficient images to have trained the algorithm effectively. Potential future development to mitigate the limitations posed by this class imbalance may include attaining more fundal images, using techniques including synthetic oversampling, and oversampling through augmentation.

Lastly, the algorithm is currently only trained to detect plus disease changes. Plus disease remains only one component of determining treatment-requiring type 1 ROP, which also includes consideration of disease stage and zone. Previous groups, however, have demonstrated that severe ROP that results in changes in stage and zone rarely occurs in isolation of vascular abnormalities observed in pre-plus or plus disease.[14,36]

Clinical Applications

As recently discussed by Ting et al,[30] deep learning has significant clinical potential in improving the reliability of and access to ROP screening. Deep learning algorithms, including ROP.AI, may in the future be adopted into existing models of care that use fundus camera systems for digital retinal imaging. In high-demand regions, including middle-income countries, these may serve as triage tools to identify cases that demand further clinician review. Furthermore, particularly given known interclinician and regional diagnostic differences, these algorithms may have potential in assisting clinicians to objectively quantify and monitor the severity of ROP disease.

In settings with limited access to specialized clinician screening or specialized fundus cameras, deep learning algorithms may have the potential to improve accessibility to ROP screening. The use of digital retinal imaging captured via existing or new camera systems, coupled with sufficiently validated algorithms, may allow for the point-of-care diagnosis (Fig. 6) of treatment-requiring ROP in settings without previous access to case detection.

Numerous evidence generation and regulatory milestones remain prior to realizing this potential novel model of care.[30] The efficacy of deep learning algorithms for ROP diagnosis in real-world prospective settings remains to be established, and there continue to be significant ethical and intelligibility concerns impeding the uptake of AI systems in healthcare.[40] Ophthalmology, however, appears to be a leader in the adoption of deep learning systems, with algorithms for the diagnosis of diabetic retinopathy having already received regulatory approval for routine clinical use.[40,41]

Future Directions

Future research goals are in development to address the limitations highlighted in this initial

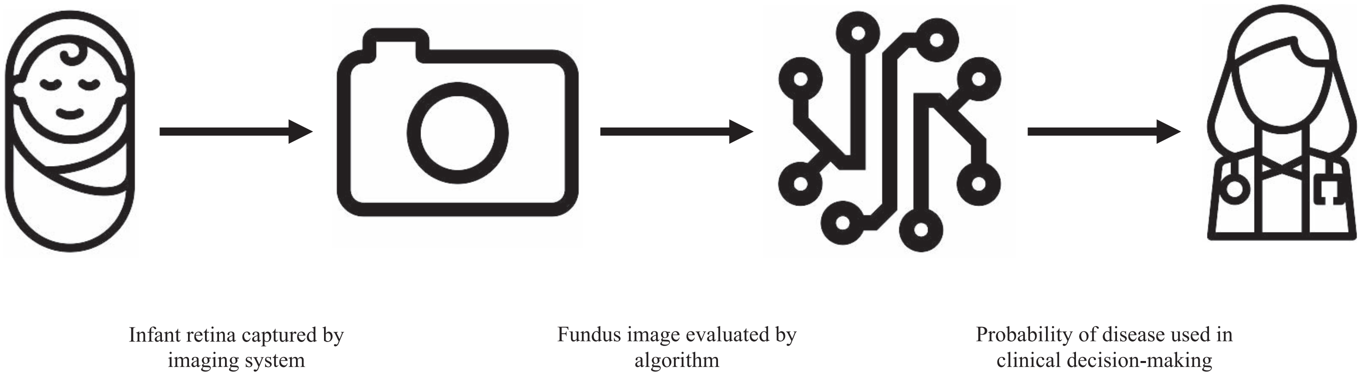
Infant retina captured by imaging system
Fundus image evaluated by algorithm
Probability of disease used in clinical decision-making

Figure 6. Potential novel retinopathy of prematurity screening model of care.

study. A multicenter local consortium that combines fundal images from ROP screening centers across Australia and New Zealand is under consideration to enable significantly expanded training and testing data sets. Data review and consensus diagnoses by multiple clinicians may reduce bias introduced by individual clinician diagnostic preferences. Greater numbers of fundal images and formal training with pre-plus and plus disease images may mitigate algorithm overfitting and improve the performance of the algorithm.

Furthermore, to improve the generalizability of the deep learning algorithm, technical variations in the fundal imaging process (e.g., different camera models, lenses, and optical aberrations) must be accounted for. Advanced deep learning frameworks utilizing segmentation maps, which decouple technical variations in optical coherence tomography imaging, have recently been successfully described and may potentially also be applied to fundal imaging.[42] A consideration of similar frameworks may be required to enable novel models of care that utilize varying fundal image capture systems.

Lastly, given the current discussions and ongoing review of ROP diagnostic criteria, consideration should also be given to the ontology of data used to train and test future algorithms. Pre-plus and plus disease remain discrete diagnostic categories that may not capture the full nuance present in ROP's disease processes. Novel diagnostic categories, aided in output by an algorithm, may in the future better reflect patient status and prognosis.

Conclusion

This study describes the initial development of a deep learning algorithm, ROP.AI, trained to automatically detect ROP. ROP.AI is the first algorithm trained using fundal images sourced from local Australasian image databases and shows high performance in the detection of plus disease. In the context of ROP's third epidemic, future development may allow for novel models of ROP screening and improve access to care for this leading cause of childhood blindness.

Acknowledgments

Special thanks to ART-ROP for access to deidentified fundal images for ROP.AI algorithm training.

Special thanks to Daniel Tian, Felipe Flores; Martin Seneviratne; Clark Stevenson; Hong Sheng Chiong, MedicMind; and Glenn Linde, MedicMind for their technical support and time in advising for the training of the ROP.AI algorithm.

Disclosure: **Z. Tan**, None; **S. Simkin**, None; **C. Lai**, None; **S. Dai**, None

*All authors contributed equally to this work.

References

1. Blencowe H, Vos T, Lee ACC, et al. Estimates of neonatal morbidities and disabilities at regional and global levels for 2010: introduction, methods overview, and relevant findings from the Global Burden of Disease study. *Pediatr Res*. 2013;74:4–16.
2. Multicenter trial of cryotherapy for retinopathy of prematurity: preliminary results. *Pediatrics*. 1988;81:697–706.
3. 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; 121:1684–1694.

4. Gilbert C. Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. *Early Hum Dev*. 2008;84:77–82.
5. Gilbert C, Fielder A, Gordillo L, et al. Characteristics of infants with severe retinopathy of prematurity in countries with low, moderate, and high levels of development: implications for screening programs. *Pediatrics*. 2005;115:e518–e525.
6. Demorest BH. Retinopathy of prematurity requires diligent follow-up care. *Surv Ophthalmol*. 1996;41:175–178.
7. Sekeroglu MA, Hekimoglu E, Sekeroglu HT, Arslan U. Retinopathy of prematurity: a nationwide survey to evaluate current practices and preferences of ophthalmologists. *Eur J Ophthalmol*. 2013;23:546–552.
8. Simkin SK, Misra SL, Han JV, McGhee CN, Dai S. Auckland regional telemedicine retinopathy of prematurity screening network: a ten year review. *Clin Exp Ophthalmol*. 2019;1–9, https://doi.org/10.1111/ceo.13593. Published July 17, 2019.
9. Dai S, Chow K, Vincent A. Efficacy of wide-field digital retinal imaging for retinopathy of prematurity screening. *Clin Exp Ophthalmol*. 2011;39: 23–29.
10. Murakami Y, Jain A, Silva RA, Lad EM, Gandhi J, Moshfeghi DM. Stanford University Network for Diagnosis of Retinopathy of Prematurity (SUNDROP): 12-month experience with telemedicine screening. *Br J Ophthalmol*. 2008;92:1456–1460.
11. Shah PK, Ramya A, Narendran V. Telemedicine for ROP. *Asia Pac J Ophthalmol (Phila)*. 2018;7: 52–55.
12. Richter GM, Williams SL, Starren J, Flynn JT, Chiang MF. Telemedicine for retinopathy of prematurity diagnosis: evaluation and challenges. *Surv Ophthalmol*. 2009;54:671–685.
13. Ells AL, Holmes JM, Astle WF, et al. Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology*. 2003;110:2113–2117.
14. Good WV, Early Treatment for Retinopathy of Prematurity Cooperative G. Final results of the Early Treatment for Retinopathy of Prematurity (ETROP) randomized trial. *Trans Am Ophthalmol Soc*. 2004;102:233–250.
15. An international classification of retinopathy of prematurity. The Committee for the Classification of Retinopathy of Prematurity. *Arch Ophthalmol*. 1984;102:1130–1134.
16. International Committee for the Classification of Retinopathy of P. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol*. 2005;123:991–999.
17. Mintz-Hittner HA, Kennedy KA, Chuang AZ. Efficacy of intravitreal bevacizumab for stage 3+ retinopathy of prematurity. *N Engl J Med*. 2011; 364:603–615.
18. Wallace DK, Quinn GE, Freedman SF, Chiang MF. Agreement among pediatric ophthalmologists in diagnosing plus and pre-plus disease in retinopathy of prematurity. *J AAPOS*. 2008;12: 352–356.
19. Gschliesser A, Stifter E, Neumayer T, et al. Inter-expert and intra-expert agreement on the diagnosis and treatment of retinopathy of prematurity. *Am J Ophthalmol*. 2015;160:553–560.e553.
20. 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 (London, England)*. 2018;32:74–80.
21. 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;123:2338–2344.
22. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. *JAMA Ophthalmol*. 2007;125:875–880.
23. Campbell J, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis. *JAMA Ophthalmol*. 2016;134: 651–657.
24. Abbey AM, Besirli CG, Musch DC, et al. Evaluation of screening for retinopathy of prematurity by ROPtool or a lay reader. *Ophthalmology*. 2016;123:385–390.
25. Wang J, Ju R, Chen Y, et al. Automated retinopathy of prematurity screening using deep neural networks. *EbioMedicine*. 2018;35:361–368.
26. Heneghan C, Flynn J, O'Keefe M, Cahill M. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Med Image Anal*. 2002;6: 407–429.
27. Wittenberg LA, Jonsson NJ, Chan RVP, Chiang MF. Computer-based image analysis for plus

disease diagnosis in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus*. 2012;49:11–20.

28. Gelman R, Martinez-Perez ME, Vanderveen DK, Moskowitz A, Fulton AB. Diagnosis of plus disease in retinopathy of prematurity using Retinal Image multiScale Analysis. *Invest Ophthalmol Vis Sci*. 2005;46:4734–4738.
29. Brown JM, Campbell J, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136:803–810.
30. Ting DSW, Wu W-C, Toth C. Deep learning for retinopathy of prematurity screening. *Br J Ophthalmol*. 2018; bjophthalmol-2018-313290.
31. Darlow BA, Clemett RS. Retinopathy of prematurity: screening and optimal use of the ophthalmologist's time. *Aust N Z J Ophthalmol*. 1990;18: 41–46.
32. Section on Ophthalmology, American Academy of Pediatrics; American Academy of Ophthalmology; American Association for Pediatrics Ophthalmology and Strabismus. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*. 2006;117:572–576.
33. Vartanian RJ, Besirli CG, Barks JD, Andrews CA, Musch DC. Trends in the screening and treatment of retinopathy of prematurity. *Pediatrics*. 2017;139:e20161978.
34. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016:2818–2826.
35. 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;316:2402–2410.
36. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J ophthalmology*. 2018;103:580–584.
37. Gupta K, Campbell JP, Taylor S, et al. A quantitative severity scale for retinopathy of prematurity using deep learning to monitor disease regression after treatment. *JAMA Ophthalmology*. 2019;137:1029–1036.
38. Taylor S, Brown JM, Gupta K, et al. Monitoring disease progression with a quantitative severity scale for retinopathy of prematurity using deep learning. *JAMA Ophthalmol*. 2019;137:1022–1028.
39. Tan Z, Chong C, Darlow B, Dai S. Visual impairment due to retinopathy of prematurity (ROP) in New Zealand: a 22-year review. *Br J Ophthalmol*. 2015;99:801–806.
40. Tan Z, Scheetz J, He M. Artificial intelligence in ophthalmology: accuracy, challenges, and clinical application. *Asia Pac J Ophthalmol (Phila)*. 2019; 8:197–199.
41. Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Retrieved from https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye. Published April 11, 2018.
42. 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: 1342–1350.