Deep learning for retinopathy of prematurity screening

Daniel S W Ting,[1] Wei-Chi Wu,[2,3] Cynthia Toth[4]

INTRODUCTION

Retinopathy of prematurity (ROP) is a neurovascular disorder of retina, characterised by abnormal fibrovascular proliferation at the boundary of the vascularised and avascular peripheral retina. Globally, it is estimated that 19 million children are suffering from visual impairment.[1] Of those, ROP accounts for 6%–18% childhood blindness,[2] causing significant psychosocial impact on the child and the family.[3] According to the Early Treatment for Retinopathy of Prematurity (ETROP) trial,[4] early treatment has shown to be beneficial to improve the visual acuity of the high-risk patients with ROP, although 9% still eventually became blind. Thus, early screening with regular monitoring is extremely crucial.

WHO AND WHEN TO SCREEN?

The at-risk groups are babies who are born preterm or those with neonatal morbidity, for example, respiratory distress syndrome, infection and hyperglycaemia.[5] These groups of neonates usually require high oxygen demand due to the systemic issues. Oxygen regulation is important for normal retinal vascular development.

In the UK, the current guidelines recommend that babies born at <32 weeks or with birth weight <1.5 kg should be screened for ROP.[6] For the medically unstable infants who require high supplemental oxygen, the screening is recommended to be done earlier, although the screening criteria may vary slightly between different countries around the world.

WHAT TO SCREEN?

Based on the International Classification of ROP, ROP is divided into five stages (table 1) with or without plus diseases.[7] Early recognition of plus diseases is extremely crucial for initiation of treatment. The disease involvement is usually documented in terms of location and the extent of clock hours. For location, it is divided into zones 1–3 (figure 1). It is important to detect the patients with prethreshold (type 1 ROP and type 2 ROP) and aggressive posterior ROP (also known as 'rush disease' previously). Serial diagnostic examinations should be performed until each eye is considered no longer at risk for developing serious ROP (eg, full retinal vascularisation, postmenstrual age of 45 weeks, and no prethreshold disease, zone III retinal vascularisation without previous zone I or II ROP, or regression of ROP).[8 9]

HOW TO SCREEN?

Traditionally, ROP has been screened by either paediatric ophthalmologists or retinal specialists using 28 or 30 dioptre lens with indirect biomicroscopy examination followed by a detailed documentation. This screening method, however, requires the expertise of experienced paediatric ophthalmologists or retinal specialists. Many countries, unfortunately, do not have such expertise to screen for ROP, leading to inadequate ROP care delivery. With the advent of wide-field retinal imaging (eg, Retcam),[8] this provides the opportunity to perform tele-ROP screening in collaboration with a reading centre.[10] In the USA, the Imaging and Informatics in ROP (i-ROP) and Stanford University Network for Diagnosis of ROP are two major tele-ROP screening networks in the USA,[10] whereas in India, Karnataka Internet-assisted Diagnosis of ROP is a large telemedicine initiative to serve rural patients who otherwise have limited access to tertiary eye care services in big cities.[11]

WHAT IS MACHINE LEARNING AND DEEP LEARNING?

The term 'machine learning' has subsequently been coined by Arthur Samuel in 1959, stating, 'the computer should have the ability to learn without being explicitly programmed.'[12] Using machine learning, the algorithm can now learn and make predictions based on the data that have been fed into the training phase using either supervised or unsupervised manner. The machine learning technique has been widely adopted in applications such as computer vision and predictive analytics using complex mathematical models. For supervised learning, the computer was trained with labelled examples, also known as ground truth, whereas for unsupervised learning, no labelling is required for the algorithm to find its own structure in its input. With the advent of graphic processing unit with much improved processing power, deep learning (DL) is the most recent machine learning technique that uses multiple processing layers to learn representation of data with multiple levels of abstraction.[13] Using the backpropagation algorithm, DL is capable of discovering intricate structure in large data sets, then changing its internal parameters that are used to compute the representation in each layer from the previous one. It has been widely adopted in the image recognition, speech recognition and natural language processing.

DL IN MEDICAL IMAGING ANALYSIS

DL has sparked tremendous interest in the medical field over the past 2 years, achieving a robust diagnostic performance in various medical conditions, including tuberculosis,[14 15] malignant melanoma[16] and breast cancer lymph

Table 1 International Classification of Retinopathy of Prematurity (ICROP)[7]

Stages	Retinal changes
1	Demarcation line
2	Ridge
3	Extraretinal fibrovascular proliferation
4	Partial retinal detachment
	A. Extrafoveal
	B. Foveal
5	Total retinal detachment

Plus disease: dilated iris vessels, poor pupil dilation, venous dilation, arteriolar tortuosity and vitreous haze.

[1]Duke-NUS Medical School, Singapore National Eye Centre, Singapore, Singapore
[2]Department of Ophthalmology, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan
[3]College of Medicine, Chang Gung University, Taoyuan, Taiwan
[4]Duke Eye Center, Duke University, Durham, North Carolina, USA

Correspondence to Dr Daniel S W Ting, Duke-NUS Medical School, Singapore National Eye Center, Singapore 168751, Singapore; daniel.ting.s.w@singhealth.com.sg

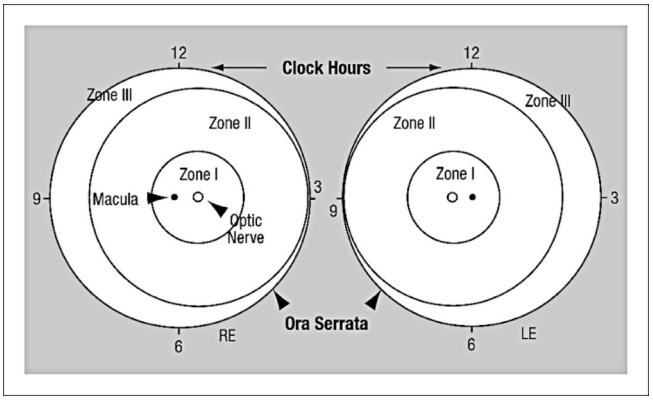

Figure 1 The location (clock hours) and zones according to the International Classification of Retinopathy of Prematurity (ICROP).7,34 Zone 1: radius from optic disc centre to twice the distance from the centre of optic disc to the macula. Zone 2: border of zone 1 to ora serrata (nasally) and equator (temporally). Zone 3: residual temporal crescent anterior to zone 2. The extent of disease is recorded as hours of the clock or as 30 sectors. RE, right eye; LE, left eye.

node metastases.[17] In ophthalmology, it has shown to have clinically acceptable diagnostic performance in detecting diabetic retinopathy (DR),[18–22] glaucoma,[18 23] age-related macular degeneration[18 24 25] from the fundus photographs[26] and OCT.[27 28] Furthermore, DL has also been applied to estimate refractive error and cardiovascular risk factors (eg, age, blood pressure, smoking status and body mass index) from retinal fundus photography.[29 30]

DL TO DETECT PLUS DISEASE IN ROP SCREENING

Brown *et al* described the use of DL system to detect plus diseases in ROP[31] using 5511 Retcam retinal photographs, U-net for vessel segmentation and pretrained Inception-V1 as the technical network. These retinal images were collected over a 5-year period from eight academic institutions for training. For the training set, the gold standards were: image-level diagnosis—three experts; patient-level diagnosis—one expert; prevalence of normal versus preplus versus plus disease: 82% vs 17% vs 3%, respectively. The trained algorithm was compared against the gold standards, as well as eight independent well-regarded experts in the field who had vast experience in managing ROP clinically. Using fivefold cross-validation, the DL system had an area under the curve of 0.94 and 0.98 for the diagnosis of normal (vs preplus disease or plus disease) and plus disease (vs normal or preplus disease). On a further evaluation using an independent sample of 100 retinal images, the algorithm continued to demonstrate robust sensitivity and specificity for detection of plus and preplus diseases (plus: 93% sensitivity/94% specificity; preplus or worse: 100% sensitivity/94% specificity).

DL TO DETERMINE ROP SEVERITY LEVEL

Redd *et al* further expanded on the same DL system to quantify the different ROP severity levels on a 1–9 scale, using a consensus reference standard diagnosis of clinical and image-based diagnosis, with the experts ranking the second data set of 100 posterior images.[32] According to overall ROP severity, the authors divide ROP into four levels based on ETROP classifications: (1) no ROP; (2) mild ROP; (3) type 2 ROP; and (4) type 1 or treatment-requiring ROP. The authors excluded stage 4 and 5 ROP to focus on the onset of clinically significant disease. The DL system was evaluated for detection of type 1 ROP, and 'clinically significant ROP', defined as type 1 ROP, type 2 ROP and preplus disease, cases that warrant an urgent referral to a specialty centre. Using the formula {[1×P(normal)]+[5×P(preplus)]+[9×P (plus)]},[33] they also develop a disease severity score, termed 'i-ROP DL score' to quantify the disease severity on a 1–9 scale. In their study, the artificial (AI) system achieved an area under the receiver operating characteristic curve (AUC) of 0.96 and 0.91 for detection of type 1 ROP and clinically significantly ROP, respectively, with an AUC of 0.99 for disease severity score to differentiate plus from no plus disease. Using a hypothetical cut-off i-ROP DL referral score of 3, the AI system had 94% sensitivity and 79% specificity to detect type 1 ROP, with a negative predictive value of 99.7% and positive predictive value of 13%.

HOW TO CLINICALLY TRANSLATE THE TECHNOLOGY?

Given the above-mentioned performance, it is important to consider a few screening models for clinical translation in ROP screening. Combining the results of the two papers,[26 32] the group showed robust performance in detecting plus diseases, type 1 ROP, clinically significant ROP and generating an effective disease severity score. What would be the appropriate screening models for ROP within and outside i-ROP network? First, the DL system can be integrated as a semiautomated fashion with DL system as the first stage grading, followed by manual grading by the human graders. With an appropriate setting of an operating threshold aiming at an extremely high sensitivity (eg, between 95% and 99%), the DL system can help filter off the normal and mild ROP cases. For those retinal images 'deemed' to be referable (either with plus diseases, type 1 ROP, clinically significant ROP or ungradable), these retinal images can be manually graded by the human at the second stage to avoid over-referrals of the unnecessary cases to the tertiary settings. Outside i-ROP setting, this algorithm could be used as a stand-alone software to help screen for ROP, though the preset sensitivity may need to be different from the semiautomated model. Without the presence of human graders, the DL system presensitivity and specificity will need to be realistic to ensure all appropriate cases are referred urgently (high sensitivity), but at the same time avoid causing over-referrals (decreased specificity). The screening model can be either cloud based or hospital based by uploading these retinal images onto an online server or grading website, or onto a desktop or laptop using an application programming interface.

POTENTIAL CHALLENGES OF AI APPLICATION IN MEDICINE

Although DL has generated an explosion of interest in many healthcare domains recently, many clinical, technical and regulatory challenges still remain unanswered. First, the generalisability of the DL system for other ethnic groups or retinal cameras is critical to increase the uptake of the technologies. Hence, it is always important to validate a DL system using many external data sets. For this particular AI system described by Redd *et al*,[26 32] the AI system has been built and validated within the i-ROP setting, with many good-quality retinal images. Further studies are important to determine the diagnostic performance of this AI system in other

ethnicities and countries with varying retinal photography skills. Second, lack of explainability of the algorithm due to the 'black-box' nature, though many studies have tried using heat map, statistical and mathematical models to address this issue. Third, given that this technology is still relatively new, the health authority and legislative bodies have yet to set the rules for medicolegal, regulatory and financial rebate policies. ROP is a complex disease that requires multidisciplinary clinical care teams and whether the use of DL system would generate an unnecessary anxiety among the parents is still yet to be surveyed.

CONCLUSION

In conclusion, DL system is a novel breakthrough machine learning technology that has revolutionised the automation of medical imaging analysis. There is a huge potential for this technology to help with millions of premature babies who may suffer from this potentially blinding disease. In order to increase the success of clinical translation of this technology, a multidisciplinary collaborative effort from different stakeholders, including the clinical and technical teams, patients, family members, hospital and health authorities, is paramount. Future research will be of great value to clinically validate this algorithm prospective manner in a real-world setting.

Contributors DSWT, WCW and CT have all contributed to the drafting, critical review and final approval for this editorial letter.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests DSWT is the coinventor of a deep learning system for retinal diseases.

Patient consent Not required.

Provenance and peer review Commissioned; internally peer reviewed.

To cite Ting DSW, Wu W-C, Toth C. *Br J Ophthalmol* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/bjophthalmol-2018-313290

Linked

► https://doi.org/10.1136/bjophthalmol-2018-313156

Br J Ophthalmol 2018;**0**:1–3.
doi:10.1136/bjophthalmol-2018-313290

REFERENCES

1 Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. *Br J Ophthalmol* 2012;96:614–8.
2 Fleck BW, Dangata Y. Causes of visual handicap in the Royal Blind School, Edinburgh, 1991-2. *Br J Ophthalmol* 1994;78:421.
3 Blencowe H, Vos T, Lee AC, *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(Suppl 1):4–16.
4 Early Treatment for Retinopathy of Prematurity Cooperative Group, Good WV, Hardy RJ, *et al*. Final visual acuity results in the early treatment for retinopathy of prematurity study. *Arch Ophthalmol* 2010;128:663–71.
5 Lee J, Dammann O. Perinatal infection, inflammation, and retinopathy of prematurity. *Semin Fetal Neonatal Med* 2012;17:26–9.
6 Fierson WM, American Academy of Pediatrics Section on Ophthalmology, American Academy of Ophthalmology, *et al*. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics* 2013;131:189–95.
7 International Committee for the Classification of Retinopathy of Prematurity. The international classification of retinopathy of prematurityInternational Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol* 2005;123:991–9.
8 Kemper AR, Wallace DK, Quinn GE. Systematic review of digital imaging screening strategies for retinopathy of prematurity. *Pediatrics* 2008;122:825–30.
9 Section on Ophthalmology. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics* 2006;117:572–6.
10 Fijalkowski N, Zheng LL, Henderson MT, *et al*. Stanford University Network for Diagnosis of Retinopathy of Prematurity (SUNDROP): four-years of screening with telemedicine. *Curr Eye Res* 2013;38:283–91.
11 Vinekar A, Gilbert C, Dogra M, *et al*. The KIDROP model of combining strategies for providing retinopathy of prematurity screening in underserved areas in India using wide-field imaging, tele-medicine, non-physician graders and smart phone reporting. *Indian J Ophthalmol* 2014;62:41–9.
12 Samuel AL. *Some studies in machine learning using the game of checkers. Computer Games*. New York: Springer, 1988: 335–65.
13 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
14 Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;284:574–82.
15 Ting DSW, Yi PH, Hui F. Clinical applicability of deep learning system in detecting tuberculosis with chest radiography. *Radiology* 2018;286:729–31.
16 Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
17 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al*. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199–210.
18 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.
19 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.
20 Lee CS, Tyring AJ, Deruyter NP, *et al*. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express* 2017;8:3440–8.
21 Abràmoff MD, Lou Y, Erginay A, *et al*. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci* 2016;57:5200–6.
22 Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017;124:962–9.
23 Li Z, He Y, Keel S, *et al*. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 2018;125:1199–206.
24 Burlina PM, Joshi N, Pekala M, *et al*. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol* 2017;135:1170–6.
25 Grassmann F, Mengelkamp J, Brandl C, *et al*. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology* 2018;125:1410–20.
26 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.
27 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.
28 Kermany DS, Goldbaum M, Cai W, *et al*. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018;172:1122–31.
29 Poplin R, Varadarajan AV, Blumer K, *et al*. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018;2:158–64.
30 Varadarajan AV, Poplin R, Blumer K, *et al*. Deep learning for predicting refractive error from retinal fundus images. *Invest Ophthalmol Vis Sci* 2018;59:2861–8.
31 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.
32 Redd TK, Campbell JD, Brown JC. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol* 2018:bjophthalmol-2018-313156.
33 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.