Deep learning for retinopathy of prematurity screening

Daniel S W Ting,¹ Wei-Chi Wu,²,³ Cynthia Toth⁴

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).⁸⁹

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, Karnata 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 malignant melanoma¹⁵ and breast cancer lymph

<table>
<thead>
<tr>
<th>Table 1</th>
<th>International Classification of Retinopathy of Prematurity (ICROP)⁷</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stages</td>
<td>Retinal changes</td>
</tr>
<tr>
<td>1</td>
<td>Demarcation line</td>
</tr>
<tr>
<td>2</td>
<td>Ridge</td>
</tr>
<tr>
<td>3</td>
<td>Extraretinal fibrovascular proliferation</td>
</tr>
<tr>
<td>4</td>
<td>Partial retinal detachment</td>
</tr>
<tr>
<td>5</td>
<td>Total retinal detachment</td>
</tr>
<tr>
<td>Plus disease: dilated iris vessels, poor pupil dilation, venous dilation, arteriolar tortuosity and vitreous haze.</td>
<td></td>
</tr>
</tbody>
</table>

¹Duke-NUS Medical School, Singapore National Eye Centre, Singapore, Singapore
²Department of Ophthalmology, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan
³College of Medicine, Chang Gung University, Taoyuan, Taiwan
⁴Duke 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

Br J Ophthalmol Month 2018 Vol 0 No 0
node metastases. In ophthalmology, it has shown to have clinically acceptable diagnostic performance in detecting diabetic retinopathy (DR), glaucoma, age-related macular degeneration, and OCT. 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.

**DL TO DETECT PLUS DISEASE IN ROP SCREENING**

Brown et al described the use of DL system to detect plus diseases in ROP 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 0.99 for disease severity score, 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 13.6%.

**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, 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 semiautomatic 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, the AI system has been built and validated within the i-ROP setting, with many high-quality retinal images. Further studies are important to determine the diagnostic performance of this AI system in other...
CONCLUSION

In conclusion, DL system is a novel breakthrough in 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.

REFERENCES