Reducing childhood blindness from retinopathy of prematurity using artificial intelligence

Background

Retinopathy of prematurity (ROP) is the leading cause of childhood blindness worldwide, yet most cases of blindness are preventable with appropriate primary, secondary, and tertiary prevention. [1,2] The incidence and severity of ROP can be markedly reduced through careful oxygen monitoring in the neonatal intensive care unit (NICU), yet many parts of the world have inadequate human and material resources (e.g. nurses and oxygen supply equipment) to ensure high-quality modern neonatal care. [3,4] As a result, ROP is epidemic in low- and middle-income countries (LMIC). [1] ROP screening, typically performed by ophthalmologists at the bedside, is time consuming, inefficient, and of variable quality as ROP diagnosis performed in this manner is subjective and many parts of the world, including the US, have too few trained examiners. Medical malpractice concerns further restrict the pool of clinicians willing to manage the disease.[5-7] Finally, it is well recognized that inter-observer diagnostic variability leads to real world differences in treatment for babies with ROP. [2,8-11] Previous work has identified that one reason for inter-observer disagreement is that severe ROP (plus disease) presents on a continuum and experts disagree, and are systematically biased, as to the level of disease that is clinically significant. [8-11] This bias also complicates interpretation of existing randomized clinical trial data.

It was this problem that led our team to develop machine learning methods for quantitative evaluation of plus disease in ROP. Initially using traditional machine learning methods, and more recently convolutional neural networks (CNNs), we have evaluated the performance of an automated classifier for detection of severe ROP. This paper reviews our initial attempts to apply this technology for improved secondary prevention (AI- assisted ROP screening, diagnosis, and improved risk prediction), and our preliminary analysis of AI for evaluation of primary prevention (NICU-level evaluation of ROP severity), and tertiary prevention (quantitative evaluation of treatment quality and disease response).

Methods

Using a dataset of 5511 images, and a reference standard diagnosis (plus disease vs. pre-plus disease vs. normal) determined by combining 3 independent image-based and 1 clinical (ophthalmoscopic) diagnosis, we trained a CNN for diagnosis of plus disease (severe ROP) using 5-fold cross validation and evaluating using area under the receiver operating characteristic curve (AUC).

Using the best-performing CNN, we performed the following experiments: 1) Comparison of DL performance to 8 world ROP experts on an independent test set of 100 images against the reference standard diagnosis. 2) Evaluation on real-world dataset from an Indian telemedicine program with 742 eye examinations. Using the output of the CNN, we developed a quantitative vascular severity score for plus disease ranging from 1-9, and performed the following additional experiments: 3) Comparison of the ROP quantitative vascular severity score to overall disease severity as determined by multiple experts. 4) Ability of the vascular severity score to monitor disease progression (and regression after treatment) over time. 5) Development of improved risk model using a quantitative ROP vascular severity score for early detection of severe ROP. 6) Evaluation of objective differences in disease severity and response to treatment for treatment-requiring ROP. 7) Population level analysis (NICU-level) of ROP severity as a function of established NICU quality measures. For this analysis, we used established WHO criteria to develop a quality scale from 0-5 with one point each for: nurse: patient ratio >1:5, oxygen blenders for every baby, pulse oximeters for every baby, posted oxygen target reminders, and having a neonatology trained physician.

Results
1) Using 5-fold cross validation, we found a mean AUC of 0.98 for the diagnosis of plus disease (range 0.973–0.993). On the independent test set of 100 images, the algorithm outperformed 7/8 ROP experts (weighted kappa 0.92 compared to reference standard diagnosis, expert range 0.73-0.93).

2) On the Indian dataset, compared to a single Indian physician telemedicine grader, the AUC was 0.88.

3) The ROP quantitative vascular severity score correlated significantly with overall reference standard diagnosis (Figure 1) suggesting that this technology may enable objective, quantitative disease classification in ROP. Retrospectively applied to the original database (5511 images), the severity score had an AUC of 0.96 for detection of TR-ROP.

4) Figure 2 demonstrates the change in SS over time, demonstrating the ability to identify babies progressing to severe treatment-requiring ROP.

5) Using this technology, we have developed a risk model that can identify progression to treatment-requiring ROP with an AUC 0.93.

6) Among 5 ophthalmologists with over 10 babies diagnosed with treatment-requiring ROP in our dataset, we found evidence both of intra-observer and inter-observer differences in severity. We further found that eyes that failed primary treatment had higher baseline severity score values at time of diagnosis, suggesting that they were either more aggressive, or treated too late.

7) Figure 3 demonstrates the vascular severity score as a function of NICU quality for 9 hospitals in a telemedicine network in South India. We found that overall ROP severity decreased with increasing NICU quality, controlling for other underlying demographic risk factors. This suggests that AI may be used for evaluation of primary prevention of ROP and may be a surrogate for NICU quality in LMIC in the future.
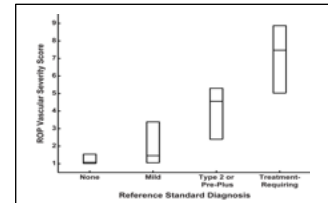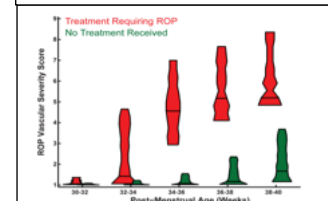

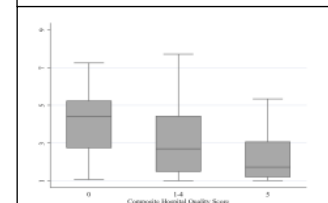Fig.1. Severity score compared to RSD.

Fig. 2. Severity score over time.

Fig. 3. Severity score by NICU quality from 0 (low) to 5 (high).

Implications for improving value of care
AI has been shown to be able to accurately diagnose retinal disease in photographs in multiple diseases, most notably diabetic retinopathy (DR). Like DR, blindness from ROP is nearly always preventable with optimal primary prevention (glucose/oxygen control), timely and accurate secondary prevention (in person or telemedical eye exams), and effective tertiary prevention (laser or intravitreal pharmacologic treatment). Unlike DR, in which patients may be asymptomatic until late in the disease stage and not present for screening, patients at risk for ROP exist in defined (captive) population in the NICU. Thus, in theory the implementation gap to deliver this technology (AI–assisted secondary prevention) to the entire at-risk population in the world is a more solvable problem than for DR, and the focus of our current translational efforts. Moreover, our data suggests a potential role for AI in the assessment of NICU quality, which may improve primary prevention of ROP. Finally, AI-based quantitative diagnosis and monitoring of disease may lead to improved tertiary prevention with more consistent anatomic and visual outcomes following treatment. ***We believe that in the next ten years, this technology will:***
1) Become standard of care for ROP screening
2) Lead to increased utilization of ROP telemedicine in NICUs worldwide.
3) Improve risk modeling and detection of severe ROP, enabling early diagnosis and treatment.
4) Encourage malpractice insurance companies to provide objective documentation of disease and reduce adverse outcomes from ROP
5) Be part of an objective assessment of NICU quality in LMIC, where heterogeneity in primary prevention is part of an ongoing epidemic of ROP.

Please consider for either Topic 1 or Topic 4.

References

1. Gilbert C. Retinopathy of prematurity: A global perspective of the epidemics, population of babies at risk and implications for control. *Early Human Development*. 2008;84(2):77-82. doi:10.1016/j.earlhumdev.2007.11.009.

2. Norman M, Hellström A, Hallberg B, et al. Prevalence of Severe Visual Disability Among Preterm Children With Retinopathy of Prematurity and Association With Adherence to Best Practice Guidelines. *JAMA Netw Open*. 2019;2(1):e186801. doi:10.1001/jamanetworkopen.2018.6801.

3. Mora JS, Waite C, Gilbert CE, Breidenstein B, Sloper JJ. A worldwide survey of retinopathy of prematurity screening. *British Journal of Ophthalmology*. 2017;102(1):9-13. doi:10.1136/bjophthalmol-2017-310709.

4. Gilbert C, Blencowe H. Retinopathy of prematurity: it is time to take action. *Community Eye Health*. 2017;30(99):45-48.

5. Engelhard SB, Collins M, Shah C, Sim AJ, Reddy AK. Malpractice Litigation in Pediatric Ophthalmology. *JAMA Ophthalmol*. 2016;134(11):1230-1235. doi:10.1001/jamaophthalmol.2016.3190.

6. Mills MD. Retinopathy of prematurity malpractice claims. *Arch Ophthalmol*. 2009;127(6):803-804. doi:10.1001/archophthalmol.2009.117.

7. Moshfeghi DM. Top five legal pitfalls in retinopathy of prematurity. *Current Opinion in Ophthalmology*. 2018;29(3):206-209. doi:10.1097/ICU.0000000000000471.

8. Campbell JP, Ryan MC, Lore E, et al. Diagnostic Discrepancies in Retinopathy of Prematurity Classification. *Ophthalmology*. 2016;123(8):1795-1801. doi:10.1016/j.ophtha.2016.04.035.

9. 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;0(0):2345-2351. doi:10.1016/j.ophtha.2016.07.020.

10. 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(11):2338-2344. doi:10.1016/j.ophtha.2016.07.026.

11. 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 (Lond)*. 2017;123:1-7. doi:10.1038/eye.2017.150.