



# Evaluation of an Artificial Intelligence System for Retinopathy of Prematurity Screening in Nepal and Mongolia

Emily Cole, MD, MPH,<sup>1,\*</sup> Nita G. Valikodath, MD, MS,<sup>1,\*</sup> Tala Al-Khaled, MD,<sup>1</sup> Sanyam Bajimaya, MBBS, MD,<sup>2</sup> Sagun KC, MSc,<sup>3</sup> Tsengelmaa Chuluumbat, MD,<sup>4</sup> Bayalag Munkhuu, MD,<sup>4</sup> Karyn E. Jonas, MSN, RN-BC,<sup>1</sup> Chimgee Chuluunkhuu, MD,<sup>5</sup> Leslie D. MacKeen, BSc,<sup>6,7</sup> Vivien Yap, MD,<sup>8</sup> Joelle Hallak, PhD,<sup>1</sup> Susan Ostmo, MSc,<sup>9</sup> Wei-Chi Wu, MD, PhD,<sup>10</sup> Aaron S. Coyner, PhD,<sup>9</sup> Praveer Singh, PhD,<sup>11</sup> Jayashree Kalpathy-Cramer, PhD,<sup>11</sup> Michael F. Chiang, MD,<sup>12</sup> J. Peter Campbell, MD, MPH,<sup>9</sup> R. V. Paul Chan, MD<sup>1</sup>

**Purpose:** To evaluate the performance of a deep learning (DL) algorithm for retinopathy of prematurity (ROP) screening in Nepal and Mongolia.

**Design:** Retrospective analysis of prospectively collected clinical data.

**Participants:** Clinical information and fundus images were obtained from infants in 2 ROP screening programs in Nepal and Mongolia.

**Methods:** Fundus images were obtained using the Forus 3nethra neo (Forus Health) in Nepal and the RetCam Portable (Natus Medical, Inc.) in Mongolia. The overall severity of ROP was determined from the medical record using the International Classification of ROP (ICROP). The presence of plus disease was determined independently in each image using a reference standard diagnosis. The Imaging and Informatics for ROP (i-ROP) DL algorithm was trained on images from the RetCam to classify plus disease and to assign a vascular severity score (VSS) from 1 through 9.

**Main Outcome Measures:** Area under the receiver operating characteristic curve and area under the precision-recall curve for the presence of plus disease or type 1 ROP and association between VSS and ICROP disease category.

**Results:** The prevalence of type 1 ROP was found to be higher in Mongolia (14.0%) than in Nepal (2.2%;  $P < 0.001$ ) in these data sets. In Mongolia (RetCam images), the area under the receiver operating characteristic curve for examination-level plus disease detection was 0.968, and the area under the precision-recall curve was 0.823. In Nepal (Forus images), these values were 0.999 and 0.993, respectively. The ROP VSS was associated with ICROP classification in both datasets ( $P < 0.001$ ). At the population level, the median VSS was found to be higher in Mongolia (2.7; interquartile range [IQR], 1.3–5.4) as compared with Nepal (1.9; IQR, 1.2–3.4;  $P < 0.001$ ).

**Conclusions:** These data provide preliminary evidence of the effectiveness of the i-ROP DL algorithm for ROP screening in neonatal populations in Nepal and Mongolia using multiple camera systems and are useful for consideration in future clinical implementation of artificial intelligence–based ROP screening in low- and middle-income countries. *Ophthalmology Science* 2022;2:100165 © 2022 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Retinopathy of prematurity (ROP) is a leading cause of preventable blindness in low- and middle-income countries (LMICs). Retinopathy of prematurity is characterized by abnormal vascular development resulting from premature birth, exposing the developing retina to relatively higher levels of oxygen as compared with the hypoxic environment in utero, which can lead to abnormal vessel formation that can cause devastating sequelae such as retinal detachment. Screening for ROP typically occurs via serial fundus examination by ophthalmologists during the neonatal period, often while infants are still in the neonatal intensive care unit. Retinopathy of prematurity is described using the International Classification of ROP (ICROP),

which describes zone (the extent of vessel growth), stage (the severity of findings), and plus disease (dilation and tortuosity of vessels portending a worse prognosis and requiring treatment). Worldwide, we are seeing rising ROP incidence, especially in LMICs, because of the growing number of neonatal intensive care units and the increased survival of preterm infants.<sup>1</sup> Infants with ROP in developing countries have been shown to have higher birth weights (BWs) and older gestational ages (GAs) compared with developed countries, which has led to more liberal screening guidelines and a higher population at risk in LMICs.<sup>2–4</sup> This increase in screening burden is especially challenging because fewer trained

ophthalmologists are available per capita than in higher-income countries.<sup>4</sup>

A number of telemedicine programs have been implemented in LMICs to address these challenges and to expand ROP screening.<sup>5–7</sup> These programs use remote grading of digital fundus images to diagnose and manage ROP. However, the high volume of ROP and shortage of trained ophthalmologists remain a challenge for telemedicine and ROP screening networks. Moreover, ROP diagnosis can be variable even among experts.<sup>8–10</sup> Theoretically, artificial intelligence (AI)-assisted screening programs reduce the human workload by providing automated diagnoses or preliminary readings of ROP images. Several proof-of-principle studies for the use of AI to detect clinically significant ROP have been published, including detection of both stage and plus disease.<sup>11–17</sup> Most of these studies used the RetCam (Natus Medical, Inc.), which limits widespread use, and were performed in high-income countries, where the clinical need may be less pressing.<sup>11–16</sup> To our knowledge, only 1 study evaluated AI for the detection of ROP stage in an LMIC using a different camera system, the Forus 3nethra neo (Forus Health).<sup>17</sup>

The Imaging and Informatics in ROP (i-ROP) deep learning (DL) system was developed by the i-ROP Consortium and has demonstrated expert-level classification of plus disease.<sup>18</sup> In addition to classifying plus disease, the system has introduced the concept of a quantitative scale for ROP severity with a vascular severity score (VSS) that has been shown to correlate with the full ICROP classification and has potential use for monitoring disease progression and for screening in LMICs.<sup>18–23</sup> However, potential challenges to the performance of any AI system on images that differ from the training dataset exist, in this case because of more diverse disease phenotypes, higher disease prevalence, differences in demographics and fundus pigmentation, image quality, and different camera systems. Implementation of AI-based ROP screening is feasible only if effectiveness is demonstrated in the intended use population. The purpose of this study was to assess the diagnostic performance of the i-ROP DL system for detection of type 1, treatment-requiring ROP from screening programs in Mongolia (using the RetCam Portable) and Nepal (using the Forus 3nethra neo).

## Methods

This was a retrospective analysis of prospectively collected data in 2 separate ROP screening programs. The study complied with the Health Insurance Portability and Accountability Act, adhered to the tenets of the Declaration of Helsinki, and was approved by the institutional review board at the University of Illinois at Chicago. Written and verbal consent were obtained from patients' parents or guardians in Nepal, and a waiver of consent was obtained in Mongolia from the local institutional review board.

### Study Population in Nepal

The study population in Nepal included patients from an ROP screening program from 4 urban hospitals in Kathmandu, Nepal (Patan Hospital, Kanti Children's Hospital, Paropakar Maternity and Women's Hospital, and Tilganga Institute of Ophthalmology).

Data were collected from October 2016 through August 2018. Infants were screened if BW was less than 1700 g or GA was less than 36 weeks.

### Study Population in Mongolia

The study population in Mongolia included patients from an ROP screening program in a single national referral center (National Center for Maternal and Child Health) in Ulaanbaatar, Mongolia. Data were collected from December 2015 through January 2017. Screening guidelines, based on the previously published guidelines from India, included infants with GA of 36 weeks or less, BW of 2000 g or less, or both.<sup>24,25</sup>

### Clinical Data

Prospectively collected data included demographics such as BW, GA, and postmenstrual age, as well as zone, stage, and plus disease classification. Clinical data regarding zone (I, II, or III), stage (1–5), and plus disease classification (none, preplus, or plus) were determined by the local ophthalmologists at the time of the examination, and data was recorded in Research Electronic Data Capture software (Vanderbilt University) in the Mongolian dataset and iTeleGEN in the Nepali dataset. Each patient was assigned a unique identification. Using the ICROP and Early Treatment for Retinopathy of Prematurity guidelines, the results from each eye examination were assigned a category, including no, mild, type 2, or type 1 ROP (herein referred to as treatment-requiring [TR] ROP).<sup>26,27</sup> Treatment-requiring ROP includes infants with plus disease as well as zone I, stage 3 without plus disease.

### Exclusion Criteria

Patients without recorded BW or GA were excluded from this study. At the clinical examination level, visits were excluded if classification of ROP could not be determined after review of the medical record. Eyes with prior treatment were excluded from the study because the goal was to evaluate AI for secondary prevention (detection of incident type 1 ROP). Of the 373 and 321 babies in the Nepal and Mongolia datasets, respectively, 50 and 7 were excluded.

### Image Reference Standard Diagnosis

For evaluation of algorithm performance on plus disease classification, we used a reference standard diagnosis (RSD). The initial diagnosis was determined by the local screening physicians. All images were reviewed by a trained study coordinator (S.O.) who was masked to the initial diagnosis and assigned a diagnosis of normal, preplus, or plus disease. An ROP expert (J.P.C.) adjudicated any differences between the trained study coordinator and local physicians' diagnoses. An additional ROP expert (R.V.P.C.) adjudicated any differences in the event of a 3-way disagreement.

### Image Analysis

Images were obtained using the Forus 3nethra neo camera in Nepal and the RetCam Portable camera in Mongolia. Multiple image views of an eye (posterior pole-centered, superior, inferior, nasal, and temporal) were obtained during each clinical examination. To be used in this study, images were required to contain the optic disc because the i-ROP DL system has been validated only on said images. These images were identified using a previously trained DL algorithm. Briefly, a U-Net (implemented with Keras and TensorFlow) was trained on manual retinal vessel segmentations to segment and detect the optic disc from an entirely separate dataset

Table 1. Demographics and Clinical Severity of Retinopathy of Prematurity in Nepal and Mongolia

Variable	Nepal	Mongolia	P Value
Patient-level summarization			
No.	323	314	
Birthweight	1959.5 ± 557.1	1515.2 ± 384.2	< 0.001
Gestational age (wks)	33.3 ± 2.5	30.4 ± 2.1	< 0.001
ROP diagnosis			
None	301 (93.2)	202 (64.3)	< 0.001
Mild	12 (3.7)	47 (15.0)	< 0.001
Type 2	3 (0.9)	21 (6.7)	< 0.001
Treatment requiring	7 (2.2)	44 (14.0)	< 0.001
Plus disease diagnosis			
Normal	295 (91.3)	188 (59.9)	< 0.001
Preplus	21 (6.5)	76 (24.2)	< 0.001
Plus	7 (2.2)	50 (15.9)	< 0.001
Examination-level summarization			
No.	391	467	
ROP diagnosis			
None	360 (92.1)	250 (53.5)	< 0.001
Mild	18 (4.6)	96 (20.6)	< 0.001
Type 2	5 (1.3)	59 (12.6)	< 0.001
Treatment requiring	8 (2.0)	62 (13.3)	< 0.001
Plus disease diagnosis			
Normal	355 (90.8)	225 (48.2)	< 0.001
Preplus	28 (7.2)	186 (39.8)	< 0.001
Plus	8 (2.0)	56 (12.0)	< 0.001

ROP = retinopathy of prematurity.

Data are presented as mean ± standard deviation or no. (%), unless otherwise indicated.

of retinal fundus images. It then was applied to images from these datasets.

Because eye examinations often included multiple images per eye, a patient-level examination VSS was assigned as follows. The i-ROP DL system was used to analyze each optic disc-containing image in the datasets to generate a VSS on a scale from 1 through 9 using methods published previously, based on the probability of the output of plus (score of 9), preplus (score of 5), or no plus (score of 1) disease.<sup>12,18,20</sup> The mean VSS of all images captured from an eye during an examination were calculated. To remove intereye correlations, a patient-level VSS was formed using the mean VSS of both eyes. Similarly, a patient-level plus disease diagnosis was formed using the median image-level prediction of an eye, then setting the patient-level diagnosis equal to the worse prediction between the two.

## Statistical Analysis

Plus disease classification performance of i-ROP DL was evaluated directly via confusion matrices and summary statistics (e.g., sensitivity, specificity, etc.). Treatment-requiring ROP classification performance was evaluated via confusion matrices by binarizing any patient-level diagnosis of preplus or plus disease into normal or not normal and comparing with infants with a binarized ROP diagnosis of TR ROP or not TR ROP. Finally, the continuous VSS was compared with binarized RSDs of plus disease and TR ROP using area under the receiver operating characteristic curve and area under the precision-recall curve. Although both are useful performance metrics, area under the precision-recall curve is better suited for highly imbalanced data (i.e., low plus disease and TR ROP prevalence).<sup>28</sup>

A *P* value of less than 0.05 was considered statistically significant. In each population, we compared the VSS with the ICROP disease category using analysis of variance. Stata MP software version 13 (StataCorp),<sup>29</sup> SAS software (SAS Institute, Inc.), and R software (R Foundation for Statistical Computing) were used for statistical analyses.

## Results

Table 1 displays the demographics of the two populations. The population of infants screened in Mongolia showed lower BW and GA compared with the population in Nepal (*P* < 0.001). Patient-level TR ROP was more prevalent in Mongolia than in Nepal, with an overall prevalence of 14.0% as compared with 2.2%, respectively (*P* < 0.001). This finding held true for plus disease as well, with prevalences of 15.9% and 2.2%, respectively.

### Diagnostic Performance for Plus Disease Classification

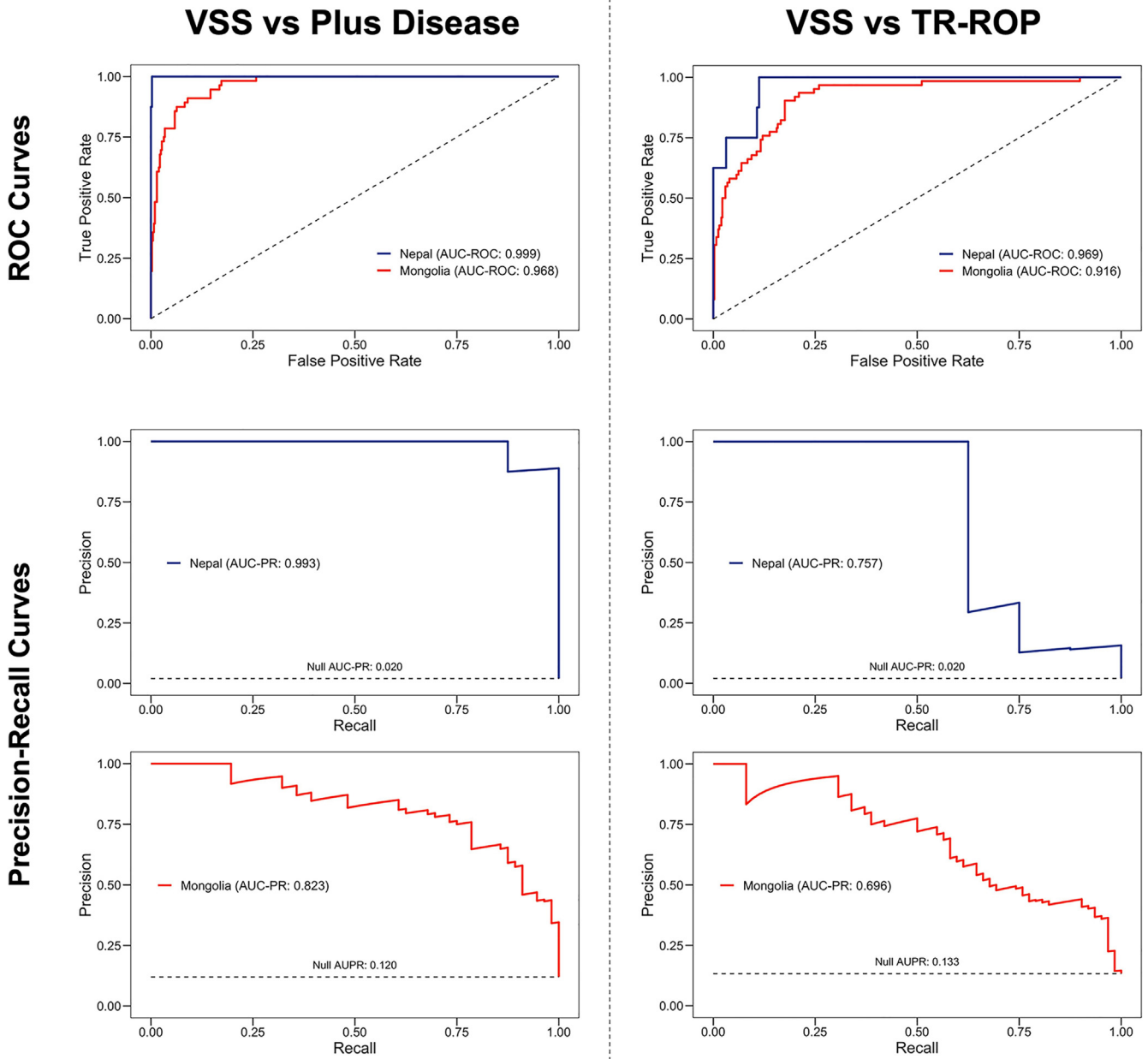
The area under the receiver operating characteristic curve for both plus disease and TR ROP by the VSS suggests high levels of detection (> 90%; Fig 1). However, the areas under the precision-recall curve suggested it may not be quite that high. This was confirmed by the confusion matrices for patient-level predictions as compared with patient-level RSDs (Table 2). Plus disease sensitivity was 75.0% and 89.3% for the Nepal and Mongolia datasets, respectively. However, when nonnormal plus disease findings were used to identify TR ROP, sensitivity was 100.0% and 96.8% in the Nepal and Mongolia datasets, respectively. Even with these high levels of sensitivity, specificity was moderately high at 64.5% and 54.3%.

### Retinopathy of Prematurity Vascular Severity Score and International Classification of Retinopathy of Prematurity Disease Category

Figure 2 displays the distribution of ROP vascular severity scores for all eye examinations in Nepal and Mongolia overall by ICROP category. At the population level, the median VSS was higher in Mongolia (2.7; interquartile range [IQR], 1.3–5.4) compared with Nepal (1.9; IQR, 1.2–3.4; *P* < 0.001). In both countries, the VSS was associated with overall disease category for no ROP, mild ROP, type 2 ROP, and TR ROP. In Mongolia, the median VSS for each category was 1.5 (IQR, 1.1–3.0), 3.6 (IQR, 1.6–5.5), 6.0 (IQR, 4.4–7.5), and 7.7 (IQR, 5.2–8.7), respectively (*P* < 0.001). In Nepal, the median VSS for each category was 1.5 (IQR, 1.1–3.0), 3.6 (IQR, 1.6–5.5), 6.0 (IQR, 4.4–7.5), and 7.7 (IQR, 5.2–8.7), respectively (*P* < 0.001).

## Discussion

In this retrospective external validation study, we evaluated the performance of the i-ROP DL system on images from Nepal and Mongolia obtained using different camera systems. The key findings of this study are (1) the system



**Figure 1.** Diagnostic accuracy of the Imaging and Informatics for Retinopathy of Prematurity DL algorithm in Nepal and Mongolia. Receiver operating characteristic (ROC) and precision-recall curves in Nepal (n = 391 examinations) and Mongolia (n = 467 examinations) for plus disease and treatment-requiring (TR) retinopathy of prematurity (ROP) diagnosis. AUC-PR = area under the precision-recall curve; AUC-ROC = area under the receiver operating characteristic curve; VSS = vascular severity score.

performed well on plus disease diagnosis in Nepal and Mongolia, despite being trained on data from North America; (2) the performance was as acceptable on an initial dataset of images from the Forus camera system compared with the RetCam, despite being trained on the RetCam; and (3) the VSS correlated well with overall ICROP severity and may be a useful epidemiologic and educational tool to compare assessment of disease severity across populations and to standardize assessment of disease severity.

Several previous studies have demonstrated the efficacy and effectiveness of the i-ROP DL system for ROP screening using the RetCam.<sup>18,20,23,30</sup> These data add further to the growing evidence of the effectiveness of this approach to ROP screening not only in North America but also in the LMIC setting, where disease epidemiologic features and phenotypes are very different. Performance can vary when AI algorithms are tested in populations (or cameras) different from the original training population.<sup>17,31</sup> Performance also can vary depending on the reference

Table 2. Confusion Matrix for Plus Disease Classification in Nepal and Mongolia

Examination-Level Prediction	Examination-Level Reference Standard Diagnosis					
	Nepal			Mongolia		
	Normal	Preplus	Plus	Normal	Preplus	Plus
Normal	245	2	0	193	28	1
Preplus	109	26	2	31	114	5
Plus	1	0	6	1	44	50

Comparison of the examination-level plus disease predictions output by Imaging and Informatics for Retinopathy of Prematurity deep learning algorithm with the reference standard diagnoses of plus disease.

standard diagnosis that is used. This is particularly true for ROP, where it is well established that experts often disagree on the diagnosis of plus disease.<sup>32–35</sup> That variability may be even more pronounced because ROP phenotypes often appear distinct in LMICs, especially when bigger babies demonstrate TR ROP, a distinction that is beyond the scope of this article but is critical to understanding for clinical implementation. The i-ROP DL algorithm initially was trained on a consensus reference standard diagnosis consisting of 3 independent gradings of the images and the ophthalmoscopic diagnosis in a North American population.<sup>20</sup> However, these results suggest that despite differences in fundus pigmentation, phenotypic appearance, and demographics, the performance in Nepal and Mongolia may be good enough for prospective clinical evaluation of AI-based ROP screening, including and beyond the North American population (Fig 1; Tables 2 and 3).

Our second key finding is that the i-ROP DL algorithm demonstrated high diagnostic performance on a dataset of images collected using a non-RetCam camera system. Previous work demonstrated that algorithms could be developed for detection of ROP stage in Forus images<sup>17</sup>; however, this was one of the first evaluations of the i-ROP DL algorithm in a screening population. One of the key implementation barriers to regulatory approval of AI is the necessity of validation on each intended camera system. In some cases, this requires retraining of original algorithms.<sup>17</sup> In this case, it seems that the i-ROP DL

system performs well enough for prospective testing on the Forus without retraining. From a global health perspective, this is very important because the cost of the Forus camera is much lower than that of the RetCam, and it was developed for the LMIC setting.<sup>36,37</sup> Other low-cost approaches may be possible, such as smartphone-based ROP screening, that further aid the democratization of AI-based ROP screening to the regions where it is needed most, but it will be critical to evaluate each of these rigorously to ensure safe clinical implementation of AI.

The third key finding is that the vascular severity score derived from the i-ROP DL system correlated well with the ICROP classification in Nepal and Mongolia. This is consistent with prior reports in North America and highlights that there may be several potential indications for use of this technology besides AI-based screening.<sup>22,38</sup> The first is disease monitoring. Because ROP screening is an iterative process and change in vascular severity is associated with a change in either the stage or extent of disease,<sup>39</sup> analyzing the change in vascular severity may improve both the sensitivity and specificity of detecting disease progression. This quantitative framework could be used for objective longitudinal disease monitoring, identifying babies who are progressing toward TR ROP based on changes in severity, rather than cross-sectional evaluation alone. The second is risk prediction. Adding objective longitudinal data to existing risk models may improve the specificity of existing screening models, which are designed to be highly

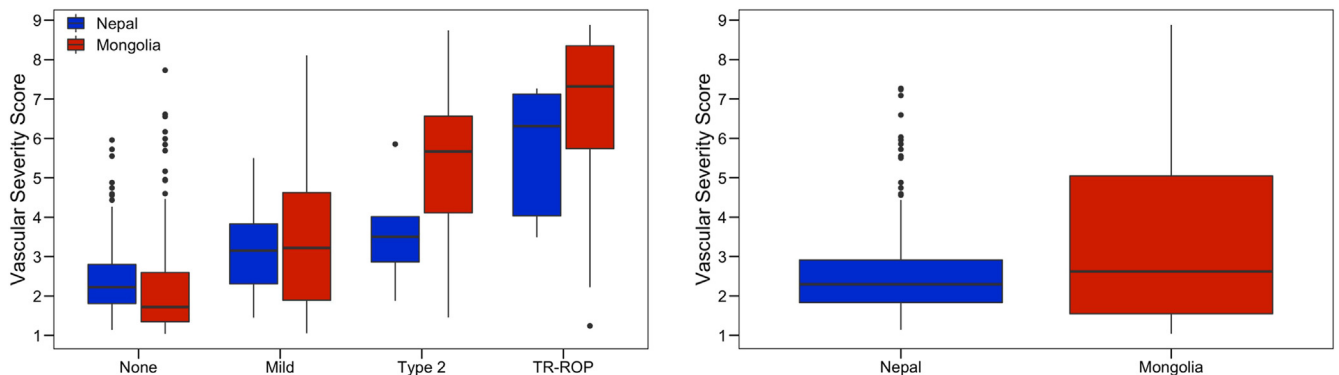


Figure 2. Box-and-whisker plots showing retinopathy of prematurity (ROP) vascular severity score (VSS) score in Nepal and Mongolia by disease category. The ROP VSS score was associated with disease category ( $P < 0.001$ ) in both countries and was higher in Mongolia compared with Nepal when associated with disease category ( $P < 0.001$ ). TR = treatment-requiring.

Table 3. Confusion Matrix for Type 1 Retinopathy of Prematurity Detection in Nepal and Mongolia

Examination-Level Prediction	Examination-Level Reference Standard Diagnosis			
	Nepal		Mongolia	
	Not Treatment-Requiring Retinopathy of Prematurity	Treatment-Requiring Retinopathy of Prematurity	Not Treatment-Requiring Retinopathy of Prematurity	Treatment-Requiring Retinopathy of Prematurity
Normal	247	0	220	2
Not normal	136	8	185	60

Comparison of nonnormal retinal vasculature patient-level predictions output by Imaging and Informatics for Retinopathy of Prematurity deep learning algorithm to binarized reference standard diagnoses of treatment-requiring retinopathy of prematurity versus not treatment-requiring retinopathy of prematurity.

sensitive. Other published algorithms and ROP prediction models exist to identify TR ROP or to alleviate the screening burden on local ophthalmologists.<sup>40–42</sup> These models factor in BW, GA, and weight gain; however, none of these models have been validated in the LMIC setting, where the implications of BW are different.<sup>43–45</sup> The third is epidemiologic evaluation. Previous work demonstrated that the VSS was higher in neonatal care units that did not have oxygen blenders or pulse oxygenation monitoring in India.<sup>30</sup> These results demonstrate differences in overall severity between the populations in Nepal and Mongolia. Future work may demonstrate the usefulness of this tool for epidemiologic monitoring not only across geographic borders but over time. The fourth is quantitative diagnosis of ROP. Objective assessment of vascular severity in ROP may provide a tool to standardize education, research, and clinical care in ROP, which has been limited by the subjectivity of diagnosis and interobserver differences.<sup>32</sup>

This study has several limitations that should be considered. First, a number of challenges exist in data collection in the LMIC setting. Inconsistencies may exist in the recording of clinical data that can lead to some records being unusable for this analysis and in data cleaning between the United States–based and internationally based teams who prepared the datasets. Despite this, we do not believe that this led to systematic bias or otherwise affected the key findings in our study. However, a prospective, longitudinal evaluation of the VSS in ROP disease progression is needed to characterize better the clinical usefulness of the VSS and its potential role as an adjunct to clinical diagnosis. In Mongolia, we used a secure, web-based ROP database

called Research Electronic Data Capture. In Nepal, data management software called iTeleGEN was implemented. iTeleGEN enabled the systematic input of data, and further studies are needed to assess its performance. Second, the precise operating point in actual clinical settings for ROP screening remains to be determined. This study examined the correlation between the i-ROP DL algorithm and RSD for plus disease as well as the VSS and ICROP category but did not use findings from simulated ROP screenings. Third, we did not assess systematically the impact of image quality on the results, although this will be a key component of clinical implementation. Despite this, the algorithm demonstrated acceptable performance and might mimic clinical settings more accurately, where it would be difficult to filter every image for quality.

In conclusion, this was one of the first studies to evaluate the performance of the i-ROP DL system for ROP screening in Nepal and Mongolia and with Forus images. Implementation of AI-based ROP screening seems feasible; however, many questions remain about how to accomplish this safely and effectively. A low-cost, AI-assisted ROP screening program theoretically could reduce the human workload, could feed into integrated risk models with serial VSS evaluation, and could identify all cases of TR ROP with significantly lower human resources compared with current screening programs, which rely on ophthalmoscopic screening or telemedicine. Although the road from successful demonstration of AI diagnostic accuracy in an article to clinical implementation at the bedside seems to be long, these data suggest that we may be on our way to clinical use of AI-based ROP screening.

## Footnotes and Disclosures

Originally received: March 29, 2022.

Final revision: April 19, 2022.

Accepted: April 19, 2022.

Available online: April 25, 2022. Manuscript no. XOPS-D-22-00061

<sup>1</sup> Department of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois Chicago, Chicago, Illinois.

<sup>2</sup> Tilganga Institute of Ophthalmology, Kathmandu, Nepal.

<sup>3</sup> Helen Keller International, Kathmandu, Nepal.

<sup>4</sup> National Center for Maternal and Child Health, Ulaanbaatar, Mongolia.

<sup>5</sup> Orbis International, New York, New York.

<sup>6</sup> The Hospital for Sick Children, Toronto, Canada.

<sup>7</sup> Phoenix Technology Group, Pleasanton, California.

<sup>8</sup> Department of Pediatrics, Weill Cornell Medical College, New York, New York.

<sup>9</sup> Department of Ophthalmology, Casey Eye Institute, Oregon Health & Science University, Portland, Oregon.

<sup>10</sup> Chang Gung Memorial Hospital, Taoyuan, Taiwan, and Chang Gung University, College of Medicine, Taoyuan, Taiwan.

<sup>11</sup> Harvard Medical School, Boston, Massachusetts.

<sup>12</sup> National Eye Institute, National Institutes of Health, Bethesda, Maryland.

Presented in part at: Association for Research in Vision and Ophthalmology, May 2020, Baltimore, Maryland.

\*Both authors contributed equally as first authors.

#### Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form. The author(s) have made the following disclosure(s): L.D.M.: Employee – Phoenix Technology Group, LLC

J.K.-C.: Financial support – Genentech

M.F.C.: Consultant – Novartis; Financial support – Genentech; Equity owner – Intelereina

J.P.C.: Financial support – Genentech; Equity owner – Boston AI Labs, Inc. (this potential conflict of interest has been reviewed and managed by OHSU) Founder – Siloam Vision, LLC

R.V.P.C.: Scientific Advisory Board – Phoenix Technology Group, LLC; Consultant – Alcon; Financial support – Genentech, Regeneron; Founder – Siloam Vision, LLC

The Imaging and Informatics for Retinopathy of Prematurity deep learning algorithm has been licensed by OHSU, Massachusetts General Hospital, University of Illinois at Chicago, and Northeastern University to Boston AI Labs, Inc., and may result in royalties to the universities, Dr Kalpathy-Cramer, Dr Campbell, Dr Coyner, and Dr Chan in the future.

Supported by the National Institutes of Health, Bethesda, Maryland (grant nos. R01EY19474, K12EY027720, P30 EY001792, and P30EY10572); the National Science Foundation, Arlington, Virginia (grant nos.: SCH-1622679, SCH-1622542, and SCH-1622536); the VitreoRetinal Surgery Foundation (NV); unrestricted departmental funding; Research to Prevent Blindness, Inc., New York, New York (career development award [J.P.C.]); the Ulverscroft Foundation (UK); the United States Agency for International Development Child Blindness Program, Washington, DC; and the Cless Family Foundation, Northbrook, Illinois.

**HUMAN SUBJECTS:** Human subjects were included in this study. The human ethics committees at the University of Illinois at Chicago approved the study. All research complied with the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and adhered to the tenets of the Declaration of Helsinki. Written and verbal consent were obtained from

patients' parents or guardians in Nepal, and a waiver of consent was obtained in Mongolia from the local institutional review board.

No animal subjects were included in this study.

#### Author Contributions:

Conception and design: Cole, Valikodath, Al-Khaled, Jonas, Hallak, Coyner, Singh, Kalpathy-Cramer, Chiang, Campbell, Chan

Analysis and interpretation: Cole, Valikodath, Al-Khaled, Bajimaya, Sagun, Chuluunbat, Munkhuu, Jonas, Chuluunkhuu, MacKeen, Yap, Hallak, Ostmo, Wu, Coyner, Singh, Kalpathy-Cramer, Chiang, Campbell, Chan

Data collection: Cole, Valikodath, Al-Khaled, Bajimaya, Sagun, Chuluunbat, Munkhuu, Jonas, Chuluunkhuu, MacKeen, Yap, Hallak, Ostmo, Wu, Coyner, Singh, Kalpathy-Cramer, Chiang, Campbell, Chan

Obtained funding: N/A

Overall responsibility: Cole, Valikodath, Al-Khaled, Bajimaya, Sagun, Chuluunbat, Munkhuu, Jonas, Chuluunkhuu, MacKeen, Yap, Hallak, Ostmo, Wu, Coyner, Singh, Kalpathy-Cramer, Chiang, Campbell, Chan

#### Abbreviations and Acronyms:

**BW** = birth weight; **DL** = deep learning; **GA** = gestational age; **ICROP** = International Classification of Retinopathy of Prematurity; **IQR** = interquartile range; **i-ROP** = Imaging and Informatics for Retinopathy of Prematurity; **LMIC** = low- and middle-income country; **ROP** = retinopathy of prematurity; **RSD** = reference standard diagnosis; **TR** = treatment-requiring; **VSS** = vascular severity score.

#### Keywords:

Artificial intelligence, Deep learning, Mongolia, Nepal, Retinopathy of prematurity.

#### Correspondence:

R. V. Paul Chan, MD, MSc, MBA, Department of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois at Chicago, 1905 West Taylor Street, Chicago, IL 60612. E-mail: [rvcphan@uic.edu](mailto:rvcphan@uic.edu)

## References

- Quinn GE. Retinopathy of prematurity blindness worldwide: phenotypes in the third epidemic. *Eye Brain*. 2016;8:31–36.
- 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(2):77–82.
- 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(5):e518–e525.
- Dogra MR, Katoch D. Clinical features and characteristics of retinopathy of prematurity in developing countries. *Ann Eye Sci*. 2018;3(1):1–7.
- Shah PK, Ramya A, Narendran V. Telemedicine for ROP. *Asia Pac J Ophthalmol (Phila)*. 2018;7(1):52–55.
- Skalet AH, Quinn GE, Ying G-S, et al. Telemedicine screening for retinopathy of prematurity in developing countries using digital retinal images: a feasibility project. *J AAPOS*. 2008;12(3):252–258.
- Ossandón D, Zanolli M, López JP, et al. [Telemedicine correlation in retinopathy of prematurity between experts and non-expert observers]. *Arch Soc Esp Ophthalmol*. 2015;90(1):9–13.
- Hewing NJ, Kaufman DR, Chan RVP, Chiang MF. Plus disease in retinopathy of prematurity: qualitative analysis of diagnostic process by experts. *JAMA Ophthalmol*. 2013;131(8):1026–1032.
- Chiang MF, Jiang L, Gelman R, et al. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. *Arch Ophthalmol*. 2007;125(7):875–880.
- Campbell JP, Ryan MC, Lore E, et al. Diagnostic discrepancies in retinopathy of prematurity classification. *Ophthalmology*. 2016;123(8):1795–1801.
- Ding A, Chen Q, Cao Y, Liu B. Retinopathy of prematurity stage diagnosis using object segmentation and convolutional neural networks [published online April 3, 2020]. *arXiv*; 2020, 200401582 [cs, eess]. Available at: <http://arxiv.org/abs/2004.01582>. Accessed 24.04.20.
- 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(1):11–19. quiz 10, 20.
- Abbey AM, Besirli CG, Musch DC, et al. Evaluation of screening for retinopathy of prematurity by ROPTool or a lay reader. *Ophthalmology*. 2016;123(2):385–390.
- Worrall DE, Wilson CM, Brostow GJ. Automated retinopathy of prematurity case detection with convolutional neural networks. In: Carneiro G, Mateus D, Peter L, et al., eds. *Deep Learning and Data Labeling for Medical Applications*.

- Lecture Notes in Computer Science*. New York, NY: Springer International Publishing; 2016:68–76.
15. Rabinowitz MP, Grunwald JE, Karp KA, et al. Progression to severe retinopathy predicted by retinal vessel diameter between 31 and 34 weeks of postconception age. *Arch Ophthalmol*. 2007;125(11):1495–1500.
  16. Wilson CM, Cocker KD, Moseley MJ, et al. Computerized analysis of retinal vessel width and tortuosity in premature infants. *Invest Ophthalmol Vis Sci*. 2008;49(8):3577–3585.
  17. Chen JS, Coyner AS, Ostmo S, et al. Deep learning for the diagnosis of stage in retinopathy of prematurity: accuracy and generalizability across populations and cameras. *Ophthalmol Retina*. 2021;5(10):1027–1035.
  18. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol*. 2019;103(5):580–584.
  19. Coyner AS, Swan R, Campbell JP, et al. Automated fundus image quality assessment in retinopathy of prematurity using deep convolutional neural networks. *Ophthalmol Retina*. 2019;3(5):444–450.
  20. 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(7):803–810.
  21. 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 [published online July 3, 2019]. *JAMA Ophthalmol*. 2019;137(9):1029–1036.
  22. 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(9):1022–1028.
  23. Greenwald MF, Danford ID, Shahrawat M, et al. Evaluation of artificial intelligence-based telemedicine screening for retinopathy of prematurity. *J AAPOS*. 2020;24(3):160–162.
  24. Rashtriya Bal Swasthya Karyakram. *Revised ROP guidelines*; June 2017. Available at: [https://nhm.gov.in/images/pdf/programmes/RBSK/Resource\\_Documents/Revised\\_ROP\\_Guidelines-Web\\_Optimized.pdf](https://nhm.gov.in/images/pdf/programmes/RBSK/Resource_Documents/Revised_ROP_Guidelines-Web_Optimized.pdf). Accessed 05.01.20.
  25. Online Neonatal Orientation Programme in India. NNF clinical practice guidelines: retinopathy of prematurity. <https://www.ontop-in.org/ontop-pen/Week-12-13/ROP%20NNF%20Guidelines%20.pdf>; 2022. Accessed 30.05.20.
  26. 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(12):1684–1694.
  27. International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol*. 2005;123(7):991–999.
  28. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10(3):e0118432.
  29. Cook J. *PRTAB: Stata module to compute precision-recall curves*. Boston College Department of Economics; 2019. <https://ideas.repec.org/c/boc/bocode/s458554.html>. Accessed 09.02.21.
  30. Campbell JP, Singh P, Redd TK, et al. Applications of artificial intelligence for retinopathy of prematurity screening. *Pediatrics*. 2021;147(3):e2020016618. <https://doi.org/10.1542/peds.2020-016618>.
  31. Chang K, Beers AL, Brink L, et al. Multi-institutional assessment and crowdsourcing evaluation of deep learning for automated classification of breast density. *J Am Coll Radiol*. 2020;17(12):1653–1662.
  32. 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)*. 2018;32(1):74–80.
  33. Reynolds JD, Hardy RJ, Kennedy KA, et al. Lack of efficacy of light reduction in preventing retinopathy of prematurity. Light Reduction in Retinopathy of Prematurity (LIGHT-ROP) Cooperative Group. *N Engl J Med*. 1998;338(22):1572–1576.
  34. Choi RY, Brown JM, Kalpathy-Cramer J, et al. Real-world variability in plus disease identified using a deep learning-based retinopathy of prematurity severity scale [published online May 4, 2020]. *Ophthalmol Retina*. 2020;4:1016–1021.
  35. Quinn GE; e-ROP Cooperative Group. Telemedicine approaches to evaluating acute-phase retinopathy of prematurity: study design. *Ophthalmic Epidemiol*. 2014;21(4):256–267.
  36. Vinekar A, Rao SV, Murthy S, et al. A novel, low-cost, wide-field, infant retinal camera, “neo”: technical and safety report for the use on premature infants. *Transl Vis Sci Technol*. 2019;8(2):2.
  37. Vinekar A, Bhende P. Innovations in technology and service delivery to improve retinopathy of prematurity care. *Community Eye Health*. 2018;31(101):S20–S22.
  38. Bellsmith KN, Brown J, Kim SJ, et al. Aggressive posterior retinopathy of prematurity: clinical and quantitative imaging features in a large North American cohort. *Ophthalmology*. 2020;127(8):1105–1112.
  39. Campbell JP, Kim SJ, Brown JM, et al. Evaluation of a deep learning-derived quantitative retinopathy of prematurity severity scale. *Ophthalmology*. 2021;128(7):1070–1076.
  40. Binenbaum G, Ying G, Quinn GE, et al. A clinical prediction model to stratify retinopathy of prematurity risk using postnatal weight gain. *Pediatrics*. 2011;127(3):e607–e614.
  41. Ying G-S, Quinn GE, Wade KC, et al. Predictors for the development of referral-warranted retinopathy of prematurity in the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) study. *JAMA Ophthalmol*. 2015;133(3):304–311.
  42. van Sorge AJ, Schalij-Delfos NE, Kerkhoff FT, et al. Reduction in screening for retinopathy of prematurity through risk factor adjusted inclusion criteria. *Br J Ophthalmol*. 2013;97(9):1143–1147.
  43. Alizadeh Y, Zarkesh M, Moghadam RS, et al. Incidence and risk factors for retinopathy of prematurity in north of Iran. *J Ophthalmic Vis Res*. 2015;10(4):424–428.
  44. Karkhaneh R, Mousavi SZ, Riazi-Esfahani M, et al. Incidence and risk factors of retinopathy of prematurity in a tertiary eye hospital in Tehran. *Br J Ophthalmol*. 2008;92(11):1446–1449.
  45. Ebrahim M, Ahmad RS, Mohammad M. Incidence and risk factors of retinopathy of prematurity in Babol, North of Iran. *Ophthalmic Epidemiol*. 2010;17(3):166–170.