Deep Learning for Image Quality Assessment of Fundus Images in Retinopathy of Prematurity

Aaron S. Coyner[1], Ryan Swan[1], James M. Brown[3], Jayashree Kalpathy-Cramer[3], Sang Jin Kim[2,4], J. Peter Campbell[2],Karyn E. Jonas[5], Susan Ostmo[2], R.V. Paul Chan[6], Michael F. Chiang[1,2]

[1]Medical Informatics, [2]Ophthalmology, Oregon Health & Science University, Portland, OR United States.
[3]MGH/Harvard Medical School, Boston, MA, United States.
[4]Ophthalmology, Sungkyunkwan University School of Medicine, Seoul, Korea.
[5]University of Illinois at Chicago, Chicago, IL, United States.
[6]Ophthalmology, Illinois Eye and Ear Infirmary, Chicago, IL, United States.

PURPOSE

To implement a prescreening method for retinal fundus images that (A) can asses whether or not a captured image is that of a retina and (B) can predict the quality of the image for diagnosis of Retinopathy of Prematurity (ROP).

INTRODUCTION

- ROP is a leading cause of preventable childhood blindness.[1]
- Roughly 90% of all infants with ROP require no treatment, as the disease improves without leaving permanent damage.[1]
- Treatment-requiring ROP is tightly associated with the presence of plus disease, which is characterized by highly tortuous and dilated retinal blood vessels.[1-4]
- Accurate image-based ophthalmic diagnosis of plus disease relies upon the clarity and quality of retinal fundus images.[2-4]
- Lower-quality images, when used in emerging techniques such as telemedicine and computer-based image analysis, can lead to higher rates of disease misclassification.[2-4]
- Automated retinal image quality assessment is a crucial preprocessing step for images that will ultimately be evaluated via computer-based image analysis or by an ophthalmologist.

METHODS

DATA SETS

Data Set	Number of Images	Source
i-ROP Image Set	6,043	Collected during routine ROP examinations
MicroImageNet	20,000	ImageNet/Kaggle

DATA SET DISTRIBUTIONS

Classifier	Training	Validation	Test
Retinal Image Classifier	5,802	288	19,959
Retinal Image Quality Classifier	2,770	200	3,073

CLASSIFIER DETAILS

Property	Value
Classifier	Deep Convolutional Neural Network
Architecture	Pre-trained VGG19
Image Input Size	64 x 64 x 3 (Retinal Image Classifier) 256 x 256 x 3 (Retinal Image Quality Classifier)
Optimizer	Stochastic Gradient Descent
Training Epochs	Early stopping; maximum 100 epochs allowed

RESULTS

RETINAL IMAGE CLASSIFIER

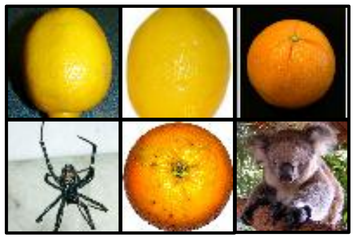

Figure 1: The six images that were incorrectly classified as retinal fundus images by the retinal image classifier.

Table 1: Summary statistics for the retinal image classifier. A total of six out of 19,959 images were misclassified as retinal images. No retinal images were misclassified as non-retinal images.

Statistic	Value
Accuracy	99.97%
Area Under Receiver Operating Characteristic Curve (AUC, AUROC)	0.999

RETINAL IMAGE QUALITY CLASSIFIER

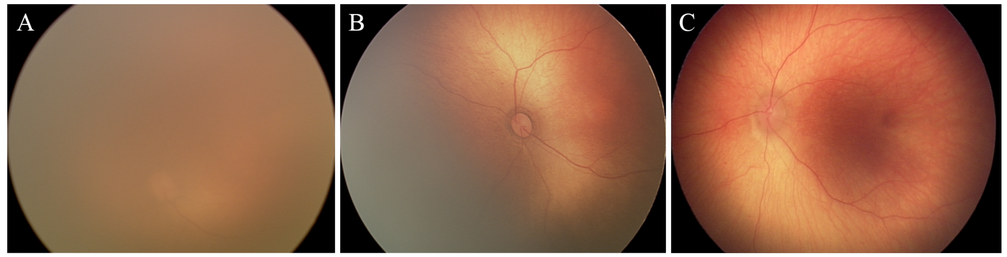

Figure 2: Example retinal fundus images: (A) Not acceptable, (B) Possibly acceptable, and (C) Acceptable for diagnosis of ROP.

Table 2: Summary statistics for the retinal image quality classifier.

Statistic	Value
Accuracy	89.1%
Area Under Receiver Operating Characteristic Curve (AUROC)	0.964
Area Under Precision-Recall Curve (AUPR)	0.966
Spearman Rank Correlation of CNN to Expert Consensus Ranking	0.891

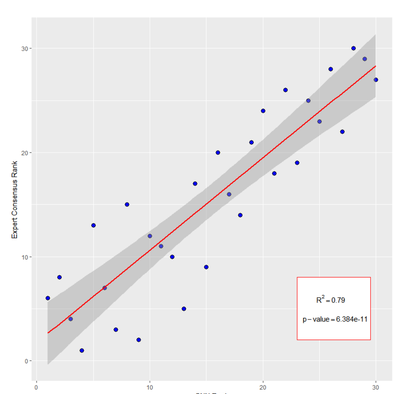

Figure 3: A smaller image set of 30 retinal fundus images of varying quality were ranked in terms of quality by six independent experts by performing pairwise comparisons. A consensus rank of the images was formed. The retinal image quality CNN scores each image from 0 to 1. It then bins images into "Acceptable for Diagnosis" and "Not Acceptable for Diagnosis" based on the raw score. The raw score was used to assess the ability of the CNN to rank images as compared to human expert graders. The CNN ranks image quality with a high degree of similarity to human graders.

CONCLUSIONS

RETINAL IMAGE CLASSIFIER

- The retinal image classifier can distinguish between retinal fundus images and non-retinal fundus images with a very high degree of accuracy.
- Qualitatively, it is easy to see why the misclassified images were incorrectly labeled as retinal fundus images.

RETINAL IMAGE QUALITY CLASSIFIER

- The retinal image quality classifier can distinguish between images of acceptable quality and those of possibly acceptable (or worse) quality with a fairly high degree of accuracy.
- The AUROC and AUPR are nearing 1, indicating that the CNN can reliably discriminate between images of acceptable quality and images of possibly acceptable quality or worse.
- Expert grader ranks were highly correlated with one another (correlation coefficient [CC] 0.89-0.94).
- Individual expert grader ranks were highly correlated with the consensus rank (CC 0.94 – 0.98).
- The CNN rank was highly correlated with the consensus rank (CC 0.89) and individual expert ranks (CC 0.84- 0.92).

ACKNOWLEDGEMENTS

Supported by NIH grants 4T15LM007088-25, NIH R01EY19474, NIH P30EY10572, and NIH T32EY23211, by NSF grant SCH-1622679, and by unrestricted departmental funding from Research to Prevent Blindness.

DISCLOSURES

Michael F. Chiang is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems, a Consultant for Novartis, and an initial member of Inteleretina. R. V. Paul Chan is on the Scientific Advisory Board for Visunex Medical Systems, and a Consultant for Genentech.

REFERENCES

1. Facts About Retinopathy of Prematurity (ROP) | National Eye Institute [Internet]. Nei.nih.gov. 2018 [cited 20 April 2018]. Available from: https://nei.nih.gov/health/rop/rop
2. Biten H, Redd T, Moleta C, Campbell J, Ostmo S, Jonas K et al. Diagnostic Accuracy of Ophthalmoscopy vs Telemedicine in Examinations for Retinopathy of Prematurity. JAMA Ophthalmology. 2018.
3. Chiang M, Wang L, Busuioc M, Du Y, Chan P, Kane S et al. Telemedical Retinopathy of Prematurity Diagnosis. Archives of Ophthalmology. 2007;125(11):1531.
4. Richter G, Williams S, Starren J, Flynn J, Chiang M. Telemedicine for Retinopathy of Prematurity Diagnosis: Evaluation and Challenges. Survey of Ophthalmology. 2009;54(6):671-685.