Assessing the Ability of Convolutional Neural Networks to Detect Glaucoma from OCT Probability Maps

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Thakoor, Kaveri A.[1]; Zheng, Qian[2]; Nan, Linyong[2]; Li, Xinhui[1]; Tsamis, Emmanouil[5]; Rajshekhar, Rashmi[5]; Dwivedi, Isht[2]; Drori, Iddo[2]; Sajda, Paul[1,3,4], Hood, Donald C.[5, 6]

Columbia University Departments of [1]Biomedical Engineering, [2]Computer Science, [3]Electrical Engineering, [4]Radiology, [5]Psychology, [6]Ophthalmology; New York, NY

Contact: k.thakoor@columbia.edu

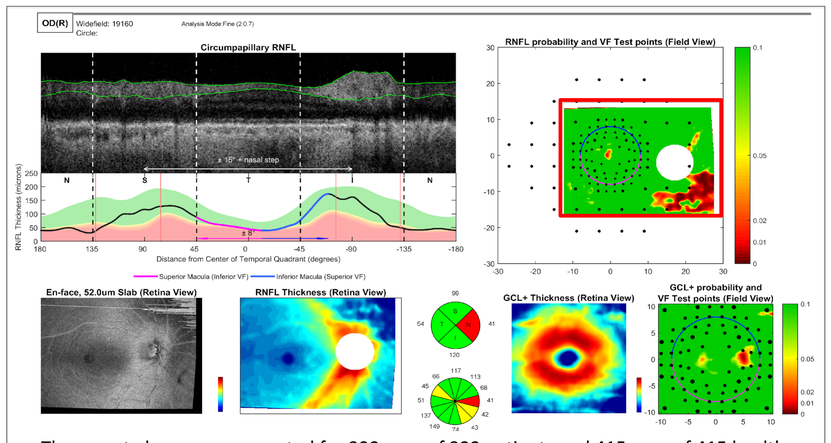
COLUMBIA | COLUMBIA UNIVERSITY DEPARTMENT OF OPHTHALMOLOGY

1464 - A0148

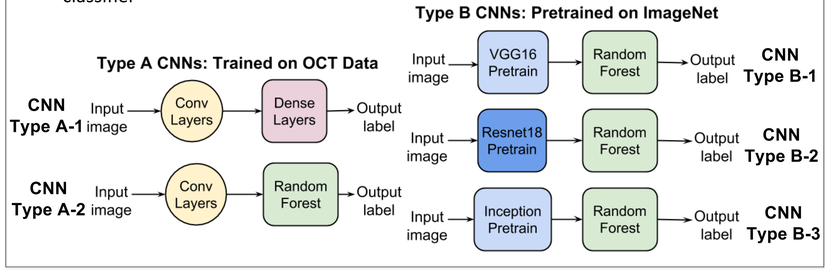
INTRODUCTION & PURPOSE

- Glaucoma is among the leading causes of irreversible blindness in the world
  - Over 50% of glaucoma cases go undetected due to lack of access to eye specialists/ophthalmologists for timely screening[1]
- We have developed multiple Convolutional Neural Network (CNN) architectures to:
  - Automate detection of glaucoma from Retinal Nerve Fiber Layer (RNFL) probability map images derived from OCT cube/volume scans
  - Evaluate eyes of glaucoma patients (**G**), suspects (**NG-S**), and healthy controls (**NG-H**)
  - Provide accuracy results, class activation map visualizations, and post-hoc analysis of false positives and false negatives

METHODS



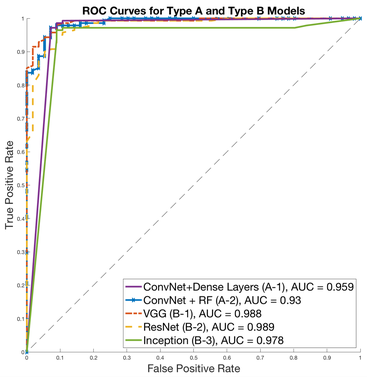
- The report above was generated for 322 eyes of 322 patients and 415 eyes of 415 healthy controls (NG-H) from wide-field OCT cube scans (Topcon).[2]
- Patients were early glaucoma or glaucoma suspects (mean deviation on 24-2 visual field better than -6 dB).
- Senior author (DCH) rated each patient eye on a scale between 0 (non-glaucomatous, **NG**, 0-49) and 100 (glaucomatous, **G**, 51-100) using report above.
- The RNFL probability maps (red rectangle) supplied the only input for all CNN models.
- The 192 **G** eyes and 545 **NG** eyes (415 NG-H and 130 NG-S) were divided into:
  - Training images: **395** (215 NG-H, 70 NG-S, and 110 G)
  - Validation images: **145** (100 NG-H, 18 NG-S, and 27 G)
  - Testing images: **197** (100 NG-H, 41 NG-S, and 56 G)
- Automated glaucoma detection was conducted with two CNN model types:
  - **CNN-A-Type:** without any natural image pretraining, (i.e., trained only on OCT data), followed by downstream classifiers (Random Forest or Dense Layers)
  - **CNN-B-Type:** pretrained on ImageNet[3], followed by a non-parametric Random Forest classifier

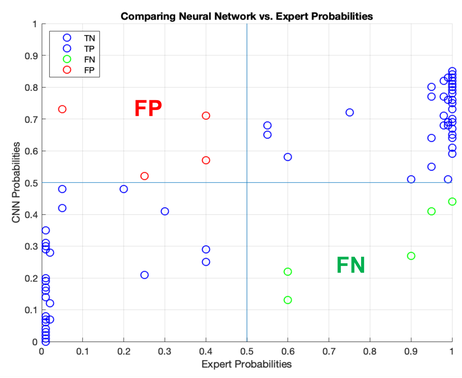


RESULTS: ACCURACY RATES, ROC CURVES, POST-HOC ANALYSIS

Training	Models	Accuracy (%)	FN-G (N = 56)	FP-NG-S (N = 41)	FP-NG-H (N = 100)
OCT Trained Only (A-1)	Conv Layers + Dense Layers	**95.7 ± 0.57**	5 (8.9%)	3 (7.3%)	1 (1.0%)
OCT Trained Only (A-2)	Conv Layers + Random Forest	94.0 ± 1.6	5 (8.9%)	4 (9.8%)	2 (2.0%)
ImageNet Pre-trained (B-1)	VGG16 PT + Random Forest	**95.0 ± 0.42**	7 (13%)	3 (7.3%)	0 (0%)
ImageNet Pretrained (B-2)	ResNet18 PT + Random Forest	94.8 ± 0.42	4 (7.1%)	6 (15%)	1 (1.0%)
ImageNet Pretrained (B-3)	InceptionNet PT + Random Forest	94.2 ± 0.91	6 (11%)	4 (9.8%)	1 (1.0%)

- All models exhibited high accuracy performance and high AUC scores (see ROC curve, below left)
- CNN A-1 had highest accuracy with 4 false positives (FP) and 5 false negatives (FN)
- Correlation between human expert rating probabilities and model probabilities was high, R value of 0.87 (see scatterplot, below right)

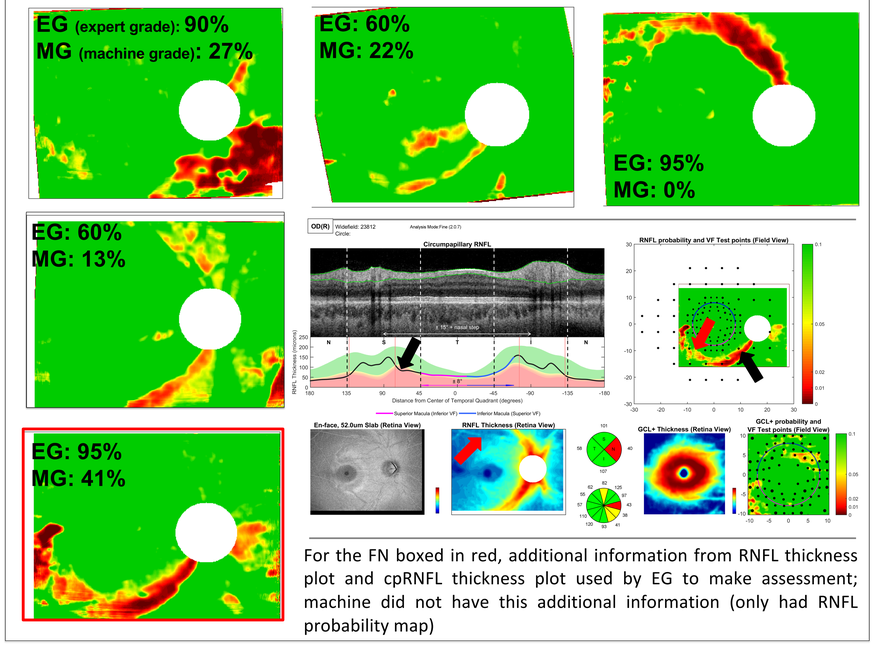




False Negatives: Expert Grader Decision Based on Additional Information

Post-Hoc Analysis

- In 4/5 FNs, patterns resembled 1 or more of those found in Healthy Controls/True Positives (e.g. blood vessel artifacts appear at similar regions as arcuates in patients; angle of arcuates more acute than that of blood vessels – can be easily confused)
- In all 5 cases, expert grading (**EG**) was made using additional information from full report (examples of additional information below)



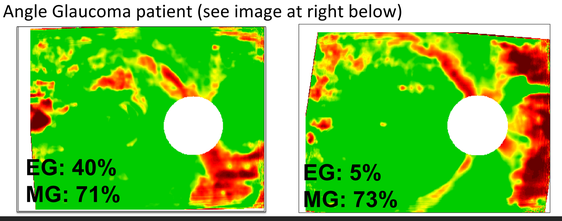
For the FN boxed in red, additional information from RNFL thickness plot and cpRNFL thickness plot used by EG to make assessment; machine did not have this additional information (only had RNFL probability map)

FALSE POSTIVE/FALSE NEGATIVE POST-HOC ANALYSIS

False Positives: Machine May Be Right (2 out of 4 FP)

Post-Hoc Analysis:

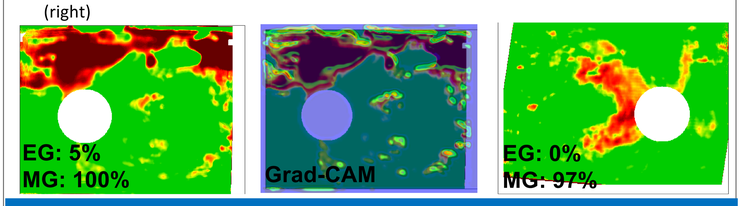
- For 1st FP, rating by Expert Grader indicated uncertainty (40%). This eye could be actually a *true positive* based on other information (family history of ocular hypertension; see image at left below).
- For the 2nd FP, machine may also be correct, as this is the fellow eye of a Juvenile Open Angle Glaucoma patient (see image at right below)



False Positives: Poor Scan Quality, Non-Glaucoma Artifacts (2 out of 4 FP)

Post-Hoc Analysis:

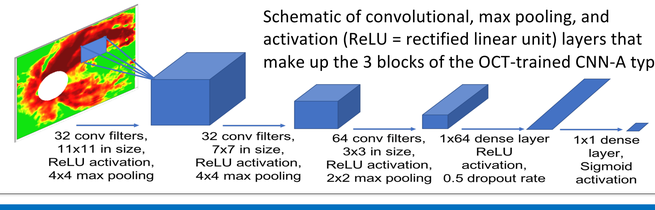
- For 3rd FP, the edges of artifacts due to poor scan (left) were mistaken for an arcuate (see Grad-CAM[4] at center)
- For last FP, artifact due to anatomical variation was recognized by EG but not by MG (right)



CONCLUSIONS & FUTURE DIRECTIONS

- Developed purely OCT-trained (Type A) as well as transfer-learning based (Type B) CNN architectures; all achieved high accuracy and high AUC-score detection of glaucoma from OCT probability map images.
- Post-hoc analysis of false positives and false negatives, aided by Grad-CAM[4] visualizations, shows strong correlation between human expert and machine performance; FP and FN may be reduced with multimodal input data
- This work is a step towards enabling automated eye disease detection especially in situations when access to vision experts may not be possible.

APPENDIX



Schematic of convolutional, max pooling, and activation (ReLU = rectified linear unit) layers that make up the 3 blocks of the OCT-trained CNN-A types.

REFERENCES, SUPPORT, & ACKNOWLEDGMENTS

References: [1] Rudnicka, A. R., Mt-Isa, S., Owen, C. G., Cook, D. G., and Ashby, D. *Invest. Ophthalmol. Vis. Sci.*, 2006.
[2] Z. Wu, D.S. Weng, R. Rajshekhar, A. Thenappan, R. Ritch, D.C. Hood. *Translational Vision Science & Technology*, 2018.
[3] O. Russakovsky, J. Deng, J. Krause, *et al.*, *International Journal of Computer Vision*, 2015.
[4] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, *International Conference on Computer Vision*, 2017.

Acknowledgments: Thanks to all members of the Hood Lab (Columbia) and the LIINC Lab (Columbia); thank you to Dr. C. Gustavo De Moraes for his suggestions and insights; special thanks to Sol Labruna for help with data pre-processing.
Support: This work was supported by NIH Grant RO1-EY02115 awarded to D.C.H. The National Science Foundation Graduate Research Fellowship awarded to K.A.T. is also gratefully acknowledged. **CR:** D. C. Hood: Topcon, Inc.; F (Financial Support), C (Consultant), R (Recipient)