Multi-pathology classification of retinal OCT scans

Ali Salehi, PhD[1]; Gary C. Lee, PhD[1]; Hugang Ren, PhD[1]; Sophia Yu, BS [1]; Rishi Singh, MD [2]; Katherine Talcott, MD [2]; Alline Melo, MD[2]; Tyler Greenlee, DO[2]; Eric Chen, MD[2]; Thais Conti, MD[2]; Niranchana Manivannan, PhD[1]

[1]Carl Zeiss Meditec Inc., Dublin, CA, USA; [2]Cleveland Clinic Cole Eye Institute, Cleveland, OH, USA

Poster #1086 - C0180

PURPOSE

- Optical coherence tomography (OCT) scans are commonly used for eye disease diagnosis. Each OCT volume scan consists of hundreds of individual B-scans.
- In a typical OCT review workflow, clinicians need to examine each B-scan individually.
- We developed an efficient multi-label Convolutional Neural Network (CNN) that flags the presence of eight different retinal abnormalities in individual B-scans.

METHODS

- 76,544 B-scans (598 eyes of 598 subjects; 128 B-scans from each eye) were captured by CIRRUS™ HD-OCT 4000 and 5000 (ZEISS, Dublin, CA), labeled by two experts, and were adjudicated by a third expert [1].
- Multi-head CNN was trained to perform eight-label binary classification on each input image.
- Architecture of the network and abnormality labels are shown in Figure 1.
- We trained the entire network end-to-end with a pre-trained model (ResNet-50 trained on ImageNet).
- Later, by freezing the shared representation sub-net, we fine-tuned each classification head separately to maximize the sensitivity/specificity for each abnormality.

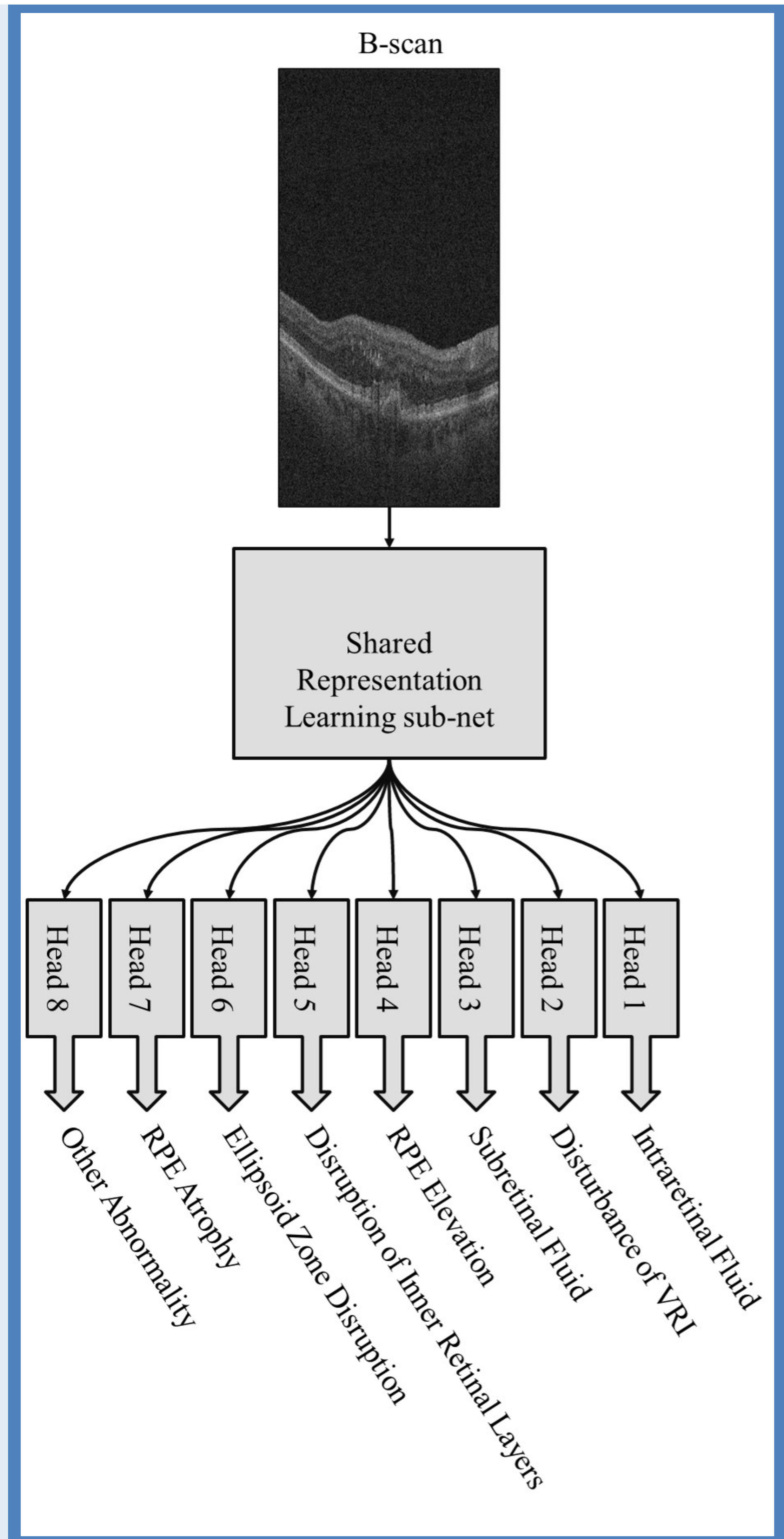

Figure 1: Architecture of the multi-head classification CNN

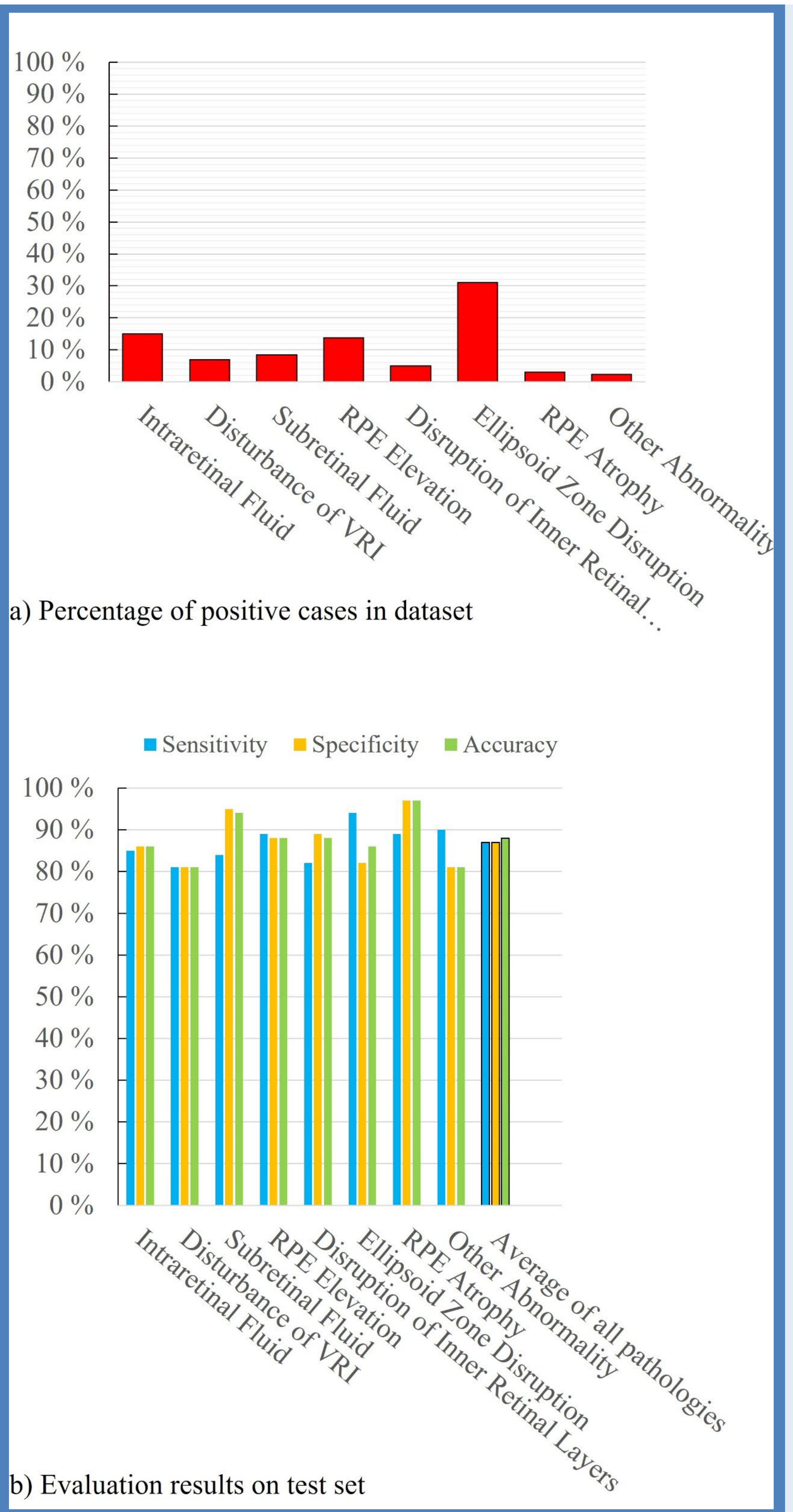

Figure 2: Distribution of the data and evaluation results of the model

Disclosures: AS (E), GL (E), RH (E), YS (E), MN (E), KT (F): Carl Zeiss Meditec, Inc.;
RS (C): Gyroscope, Novartis, Alcon, Bausch and Lomb, Genentech, Regeneron;
RS (F): Apellis, Graybug, Aerie; KT (C): Roche, Genentech

Email: ali.salehi@zeiss.com

- Data were split into train, validation, and test sets with 80%/10%/10% split at the patient level.
- Data was highly imbalanced (shown in Figure 2a), affecting the model's performance.
- Macro soft F-score with a beta of 2 was used as a loss function to address imbalance in training and increase sensitivity and specificity simultaneously.

RESULTS

- Despite the highly unbalanced and limited training data, our method achieved mean sensitivity of 86.8%, specificity of 87.4%, and a mean accuracy of 87.6%.
- Figure 2b shows the detailed accuracy metrics for each abnormality in the hold-out test set.

CONCLUSIONS

By designing a multi-head CNN and using a custom loss function, we were able to train a single model to perform eight-label classification on OCT B-scan images to improve the accuracy and efficiency of clinical workflow.

REFERENCES

[1] Yu, Sophia, et al. "Performance validation of B-scan of interest algorithm on normative dataset." *Investigative Ophthalmology & Visual Science* 61.9 (2020): PB0085.