



University of California
San Francisco

Limits of Fundus-Based Deep Learning for Alzheimer’s Disease Prediction: Anatomical Ablation and Clinical Benchmarking in UK Biobank

Eric Yang¹, Rohit Vashisht^{1,2}, PhD; Renaud La Joie^{1,3}, PhD; Thomas Hoffmann¹, PhD, MA; Peter Buto⁴; Willa Brenowitz⁵, PhD, MPH; Maria Glymour⁴, ScD, MS; Jacqueline Torres¹, PhD, MA, MPH; Yulin Hswen^{1,6}, ScD, MPH

¹Department of Epidemiology and Biostatistics, UCSF School of Medicine, San Francisco, CA; ²Bakar Computational Health Science Institute, San Francisco, CA; ³UCSF Weill Institute for Neurosciences, San Francisco, CA; ⁴Department of Epidemiology, Boston University School of Public Health, Boston, MA; ⁵Kaiser Permanente Center for Health Research, Portland, OR; ⁶Department of Epidemiology and Biostatistics, University of Maryland, College Park, MD



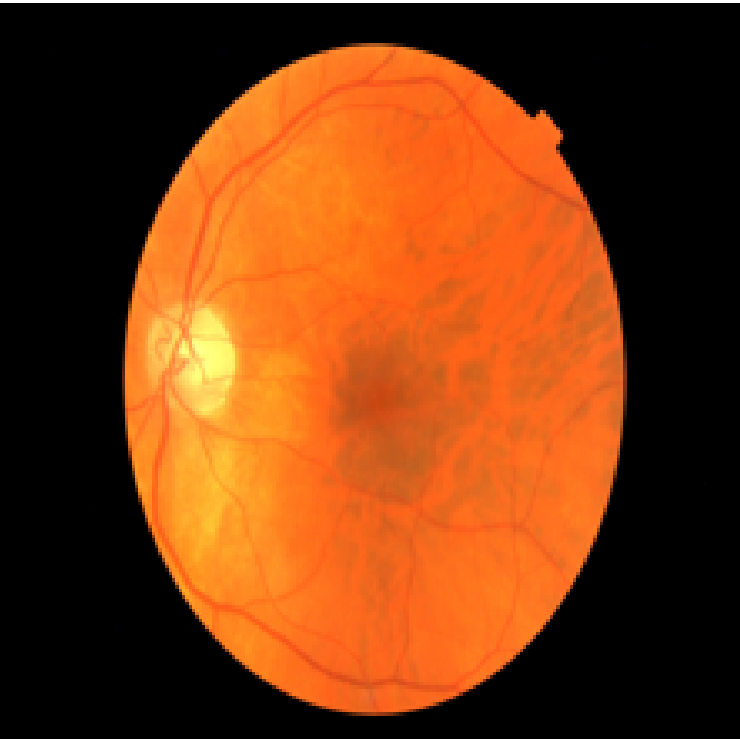
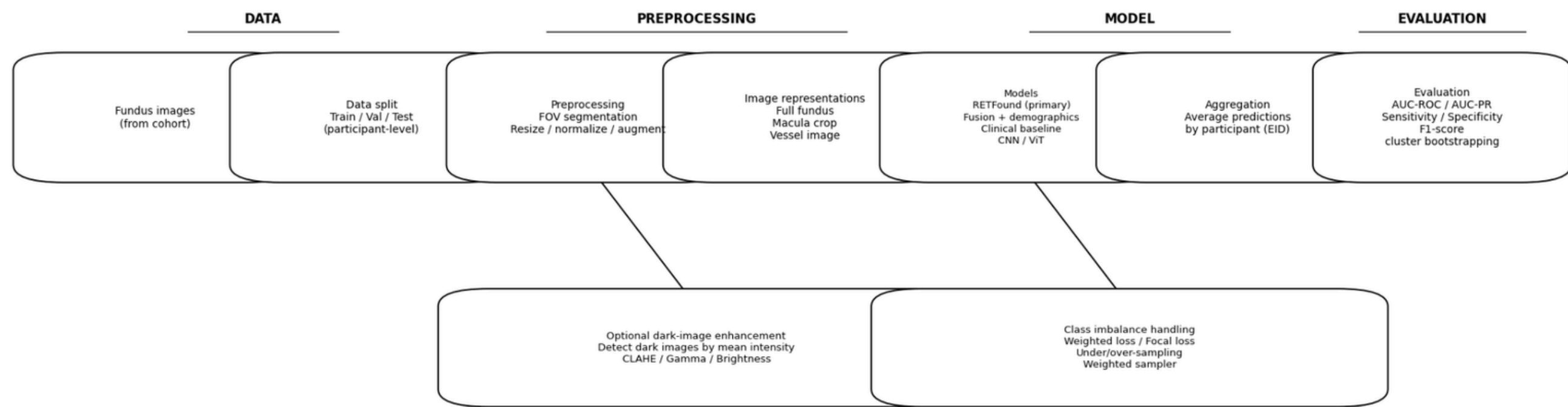
OBJECTIVE

To evaluate retinal fundus imaging for predicting Alzheimer’s Disease and Related Dementias (ADRD) using a pretrained retinal foundation model (RETFound), including assessment of model performance from baseline models, retinal region contributions, and added value beyond demographic risk factors.

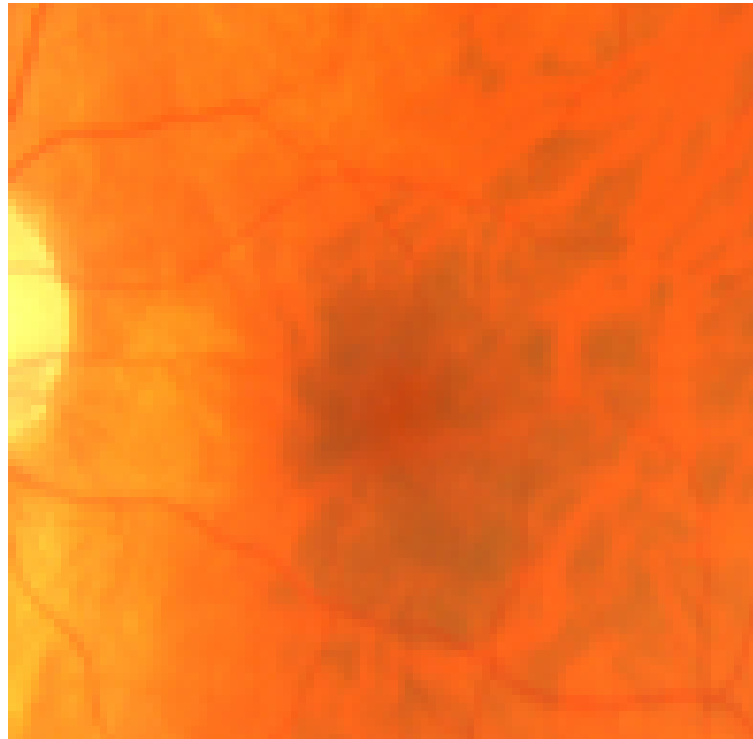
INTRODUCTION

- The retina reflects brain neurobiology through neuromodulator release while retinal ganglion cell-inner plexiform layer (GC-IPL) and retinal nerve fiber layer (RNFL) thinning contributes to blood vessel inflammation and **impaired exchange of nutrients, and wastes linking retinal changes to systemic neurodegeneration and cognitive decline**.
- Instead of using costly and labor intensive traditional MRI scans for Alzheimer’s Disease diagnosis, fundus images and OCT scans in the eye can potentially provide a **cheap and efficient** alternative for providing accessible and equal care among people and communities.
- The retina is a non-invasive biomarker for ADRD**. Optical coherence tomography (OCT) studies show thin RNFL and GC-IPL are associated with higher dementia risk and poorer cognition in large cohorts (e.g. UK Biobank), establishing retinal structure as a marker of cognitive decline.
- Advances in machine learning (ML) have shifted ADRD research from traditional models to deep learning. Neural networks extract complex image features, and models using retinal and vascular biomarkers achieve strong performance (average AUC to ~0.85)
- RETFound is a pretrained retinal foundation model trained on 1.6M+ retinal images**. Its self-supervised framework enables efficient fine-tuning. We apply RETFound to UK Biobank fundus images and clinical variables to evaluate ADRD prediction against baseline models, trained using nature images, and custom sigmoidal multimodal fusion model.
- OCT Logistic Regression (LR) acts as clinical baseline using age, sex and human derived GC-IPL and RNFL numerical values from fundus images. Human derived variables are very labor intensive and are not ideal for mass data.

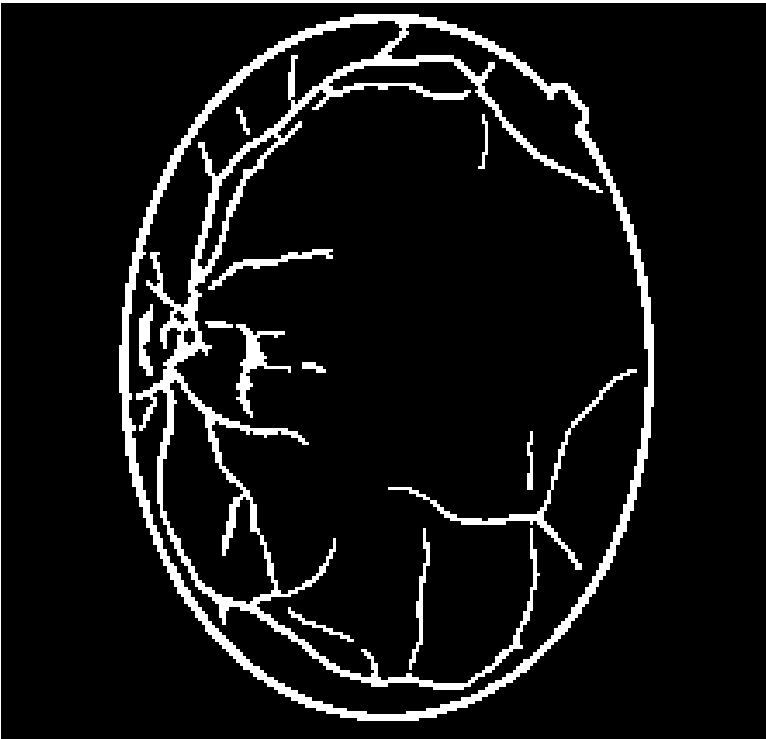
METHODS



Masked Retina



Macula Cutoff

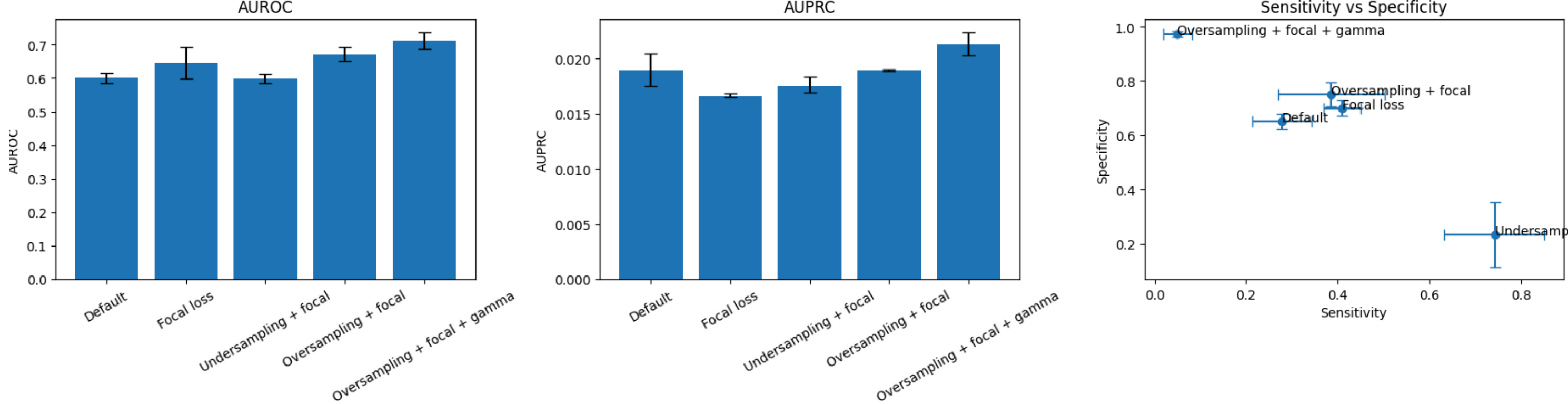
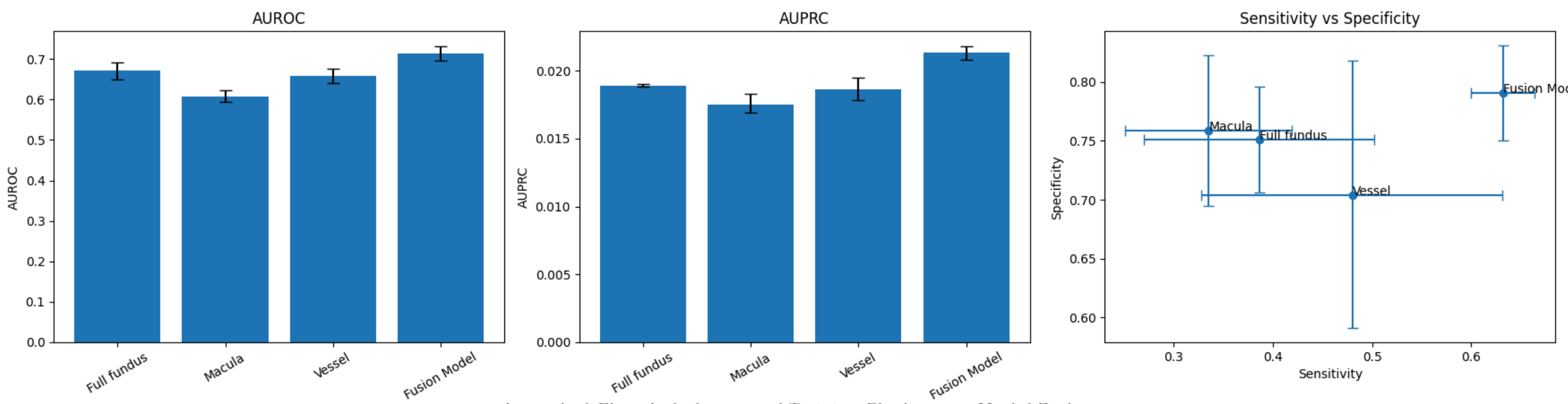
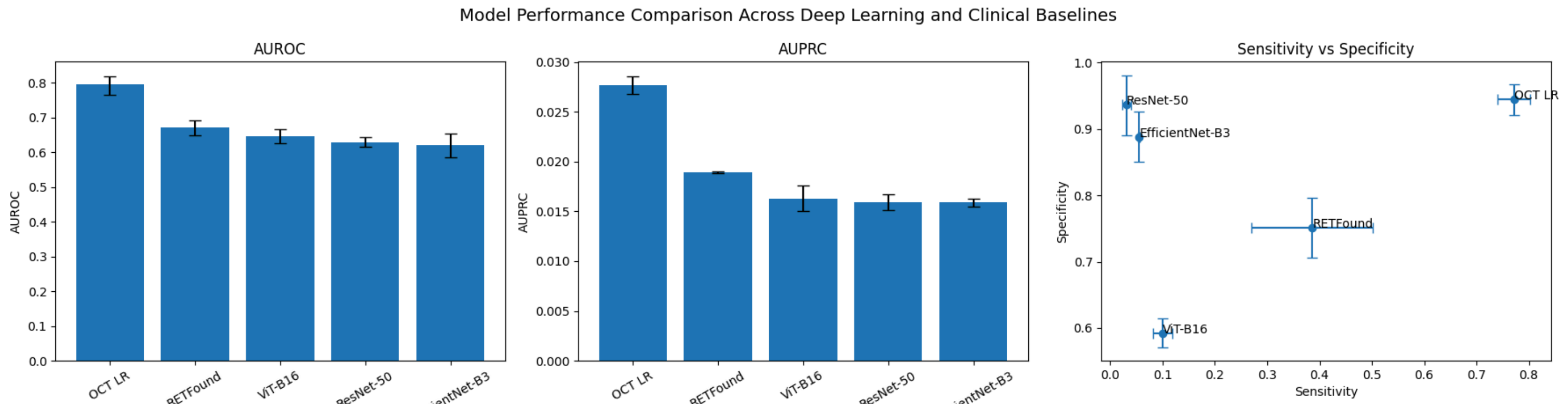
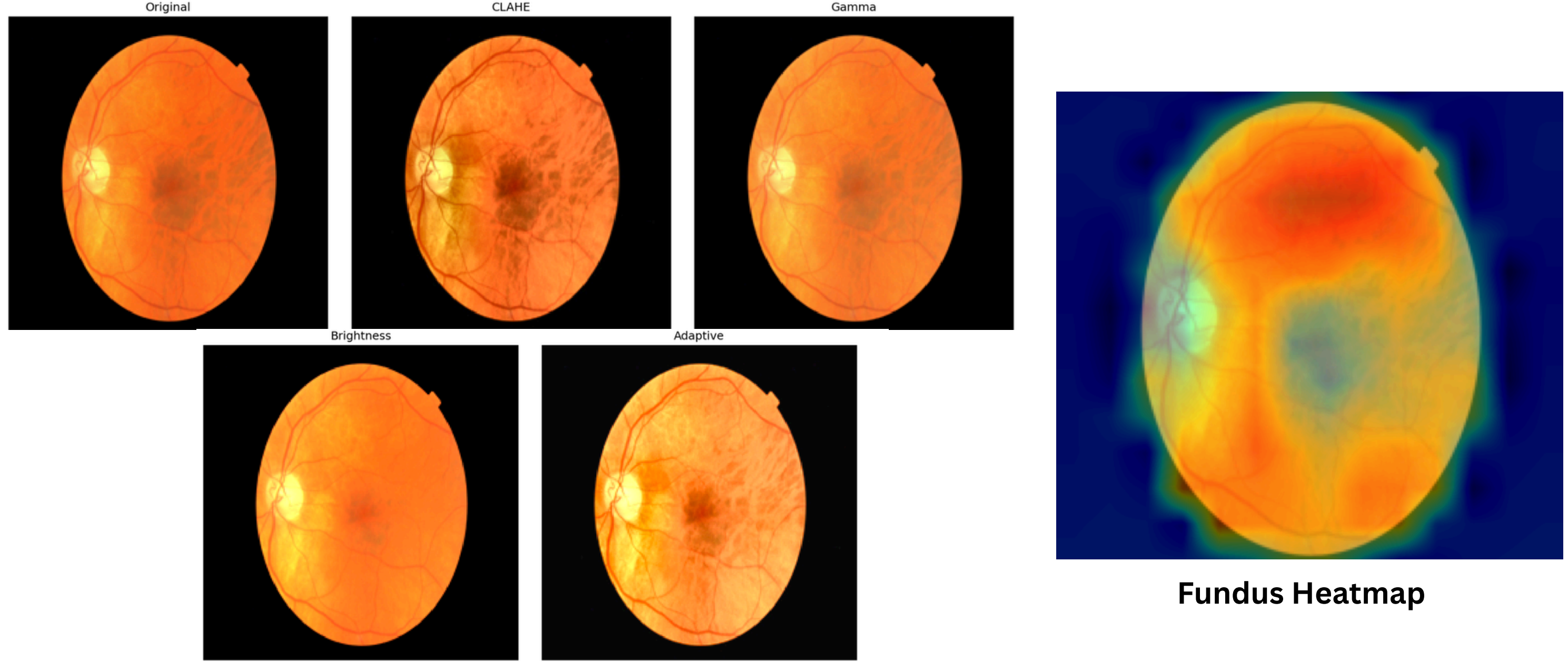


Retinal Vessels

CONCLUSION

RETFound improved fundus-based ADRD prediction over traditional deep learning methods by 0.08 in AUROC, but performance remained limited by class imbalance and subtle retinal signal, with traditional OCT-derived biomarkers still outperforming all image-based models.

RESULTS



RESOURCES

1. Hao, J., Kwapong, W.R., Shen, T. et al. Early detection of dementia through retinal imaging and trustworthy AI. *npj Digit. Med.* 7, 294 (2024). <https://doi.org/10.1038/s41746-024-01292-5>

2. Ko, F., Muthy, Z.A., Gallacher, J. et al. Association of Retinal Nerve Fiber Layer Thinning With Current and Future Cognitive Decline: A Study Using Optical Coherence Tomography. *JAMA Neurol.* 2018;75(10):1198-1205. doi:10.1001/jamaneurol.2018.1578

3. Muttu, U., Colijn, J.M., Ikram, M.A. et al. Association of Retinal Neurodegeneration on Optical Coherence Tomography With Dementia: A Population-Based Study. *JAMA Neurol.* 2018;75(10):1256-1263. doi:10.1001/jamaneurol.2018.1563

4. Rong, H., Jianhao, Xiong, Gail, Li, Yidan, Zhu, Zongyuan, Ge, Yanjun, Ma, Meng, Fu, Chenglong, Li, Bin, Wang, Li, Dong, Xin, Zhao, Zhiqiang, Ma, Jili, Chen, Xinxiao, Gao, Chao, He, Zhaohui, Wang, Wenbin, Wei, Fei, Wang, Xiangyang, Gao, Yuzhong, Chen, Qiang, Zeng, Wuxiang, Xie, Development and validation of a deep learning algorithm based on fundus photographs for estimating the CAIDE dementia risk score, Age and Ageing, Volume 51, Issue 12, December 2022, afac282, <https://doi.org/10.1093/ageing/afac282>

5. Sara Ahsadi, Kenneth A Wilson, Boris Babenko, Cory Y McLean, Drew Bryant, Orion Pritchard, Ajay Kumar, Enrique M Carrera, Ricardo Lamy, Jay M Stewart, Avinash Varadarajan, Marc Berndt, Pankaj Kapahi, Ali Bashir (2023) Longitudinal fundus imaging and its genome-wide association analysis provide evidence for a human retinal aging clock eLife 12:e82364 <https://doi.org/10.1101/2023.04.11.531111>

6. Wisely, C. E., Wang, D., Henao, R., Grewal, D. S., Thompson, A. C., Robbins, C. B., Yoon, S. P., Soundararajan, S., Polascik, B. W., Burke, J. R., Liu, A., Carin, L., & Fekrat, S. (2020). Convolutional neural network to identify symptomatic Alzheimer's disease using multimodal retinal imaging. *Investigative Ophthalmology & Visual Science*, 61(14), 34.

7. Zhou, Y., Chia, M.A., Wagner, S.K. et al. A foundation model for generalizable disease detection from retinal images. *Nature* 622, 156-163 (2023). <https://doi.org/10.1038/s41586-023-06555-x>

DISCUSSION

- RETFound performance:** RETFound outperformed ImageNet models, supporting the value of retinal-specific pretraining, though overall performance remained moderate.
- Vessel as a retinal signal:** Ablation analysis showed predictive features are not localized; vessel images performed comparably to full fundus, while macula images underperformed, supporting physiology behind retina and dementia
- Impact of class imbalance:** Performance was sensitive to sampling strategies. Oversampling with focal loss provided balanced results, while other methods shifted sensitivity–specificity trade-offs, affecting clinical usability
- Comparison to clinical features:** OCT-derived biomarkers (e.g., GC-IPL, RNFL) outperformed image-based models, indicating fundus imaging may serve as a complementary rather than standalone tool
- Clinical relevance:** Despite low AUPRC due to rare disease prevalence, models showed enrichment over baseline, suggesting potential utility in screening settings. However, achieving clinically acceptable sensitivity and specificity remains challenging

LIMITATIONS

- Class imbalance:** Low ADRD prevalence led to imbalance despite mitigation, contributing to low AUPRC and prediction difficulty
- Modest performance:** AUROC was moderate, suggesting limited predictive signal from fundus images alone and possible label limitations
- Preprocessing & generalizability:** Image representations used heuristic methods rather than precise segmentation. Models were trained only on UK Biobank
- Limited multimodal data:** Only demographics were included; additional expert-guided features may improve performance and interpretability



eric.yang2@ucsf.edu