# AIM-AHEAD

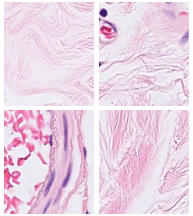
## 2023 Health Equity Data Challenge Lightning Talks

## University of Hawai‘i Team

Arianna Bunnell, Armin Soltan & Zain Jabbar

NIH National Institutes of Health
Office of Data Science Strategy

# Model Development



256* 256 Patches from Whole-Slide Biopsy Images

ResNet50 with ImageNet Weights

Demographic Balanced sampling

Pre-processing Technique for Fairness

Fine-tuned ResNet50

Stage Classification Head

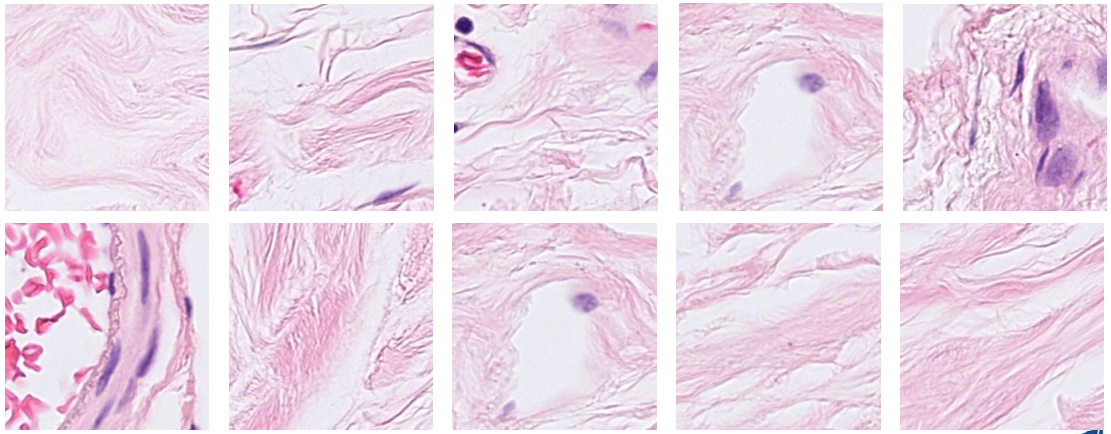
# Methodology and Model Used

- A ResNet50 was fine-tuned to take in a single (256 x 256) patch randomly selected from the whole-slide image to predict cancer stage.
- Balanced sampling was used during model training and validation to ensure every **group** in the dataset was represented equally.

White or Caucasian
Black or African American
American Indian or Alaska Native
Asian
Native Hawaiian or Pacific Islander
Other

# Model Evaluation

1. 10 probabilistic stage predictions generated per slide
2. Mean-pooling to a single slide-wide prediction
3. *n* slides mean-pooled per biopsy



# Results and Conclusions

- 0.68 Nightingale score on held-out, demographic-balanced test set

**One vs. Rest AUROC by Stage**

- Stage 0: 0.752
- Stage 1: 0.712
- Stage 2: 0.524
- Stage 3: 0.698
- Stage 4: 0.785

# Acknowledgement and References


[1] Cancer at a Glance 2014-2018, Hawai‘i Tumor Registry, 2022.

[2] Bifulco, C., Piening, B., Bower, T., Robicsek, A., Weerasinghe, R., Lee, S., Foster, N., Juergens, N., Risley, J., Nachimuthu, S., Haynes, K., & Obermeyer, Z. (2021). Identifying high-risk breast cancer using digital pathology images [Data set]. Nightingale Open Science. https://doi.org/10.48815/N5159B

[3] Mullainathan, S., & Obermeyer, Z. (2022). Solving medicine’s data bottleneck: Nightingale Open Science. Nature Medicine, 28(5), 897–899. https://doi.org/10.1038/s41591-022-01804-4



This research was, in part, funded by the National Institutes of Health (NIH) Agreement NO. 1OT2OD032581-01. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.

# AIM-AHEAD

Thank you

**Poster #90**

NIH National Institutes of Health
*Office of Data Science Strategy*