Breast Cancer Stage Classification from Whole Slide Histology Images using a Pre-trained Framework

Arianna Bunnell[1,2], Armin Soltan[1], Zain Jabbar[1]
University of Hawai'i at Mānoa, Honolulu, HI, USA[1], University of Hawai'i Cancer Center, Honolulu, HI, USA[2]

BACKGROUND:

- Breast cancer is staged between 0 and IV. Stage 0 represents a collection of abnormal cells which may develop into cancer, in Stages I - III cancer is present and represent increasing tumor size and spread, and finally, Stage IV indicates that the cancer has spread to parts of the body distant from the original site.
- Breast cancer in advanced stages (Stages III and IV) is more complex to treat and patient diagnosed at this stage have poorer survival outcomes.
- The advanced-stage breast cancer rate in Hawai'i is 22%, while in the U.S. more generally is is around 10% [1]. Advanced-stage breast cancer rates are highest among Native Hawaiian women, who suffer from the highest mortality from breast cancer in Hawai'i: 24.9% [1].
- Whole slide histology imaging is performed after breast biopsy to diagnose breast cancer. Stage is typically clinically evaluated through imaging and lymph node evaluation.
- Providing breast cancer stage through whole slide histology imaging has the potential to reduce burden on the patient for travel for additional imaging services and increase time to treatment, improving patient outcomes.

RESEARCH QUESTION:

This study aims to build an AI model which can take in a whole slide image and provide a breast cancer stage prediction consistently, regardless of patient's demographic characteristics.

METHODOLOGY:

- Data were sourced from Nightingale [2,3].
- A ResNet50 was fine-tuned to take in a single patch randomly selected from the whole-slide image to predict cancer stage.
- Balanced sampling was used during model training and validation to ensure every racial/ethnic group in the dataset was represented equally.

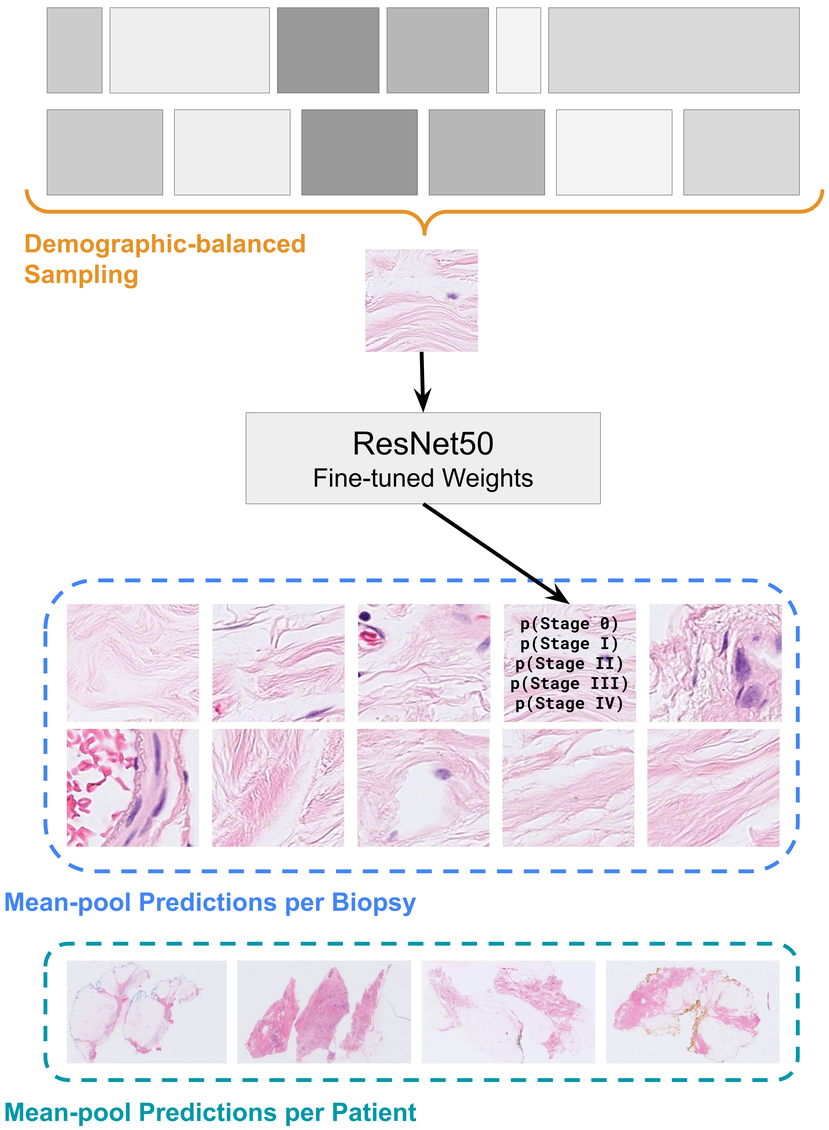

Figure 1: Model design for the final model submitted for the Nightingale competition. Model was trained on patches to predict cancer stage. During test time, 10 patches are fed through the model and mean pooled to come to a final model prediction.

MODEL EVALUATION:

- The dataset has been split randomly at the patient level with 75% of the data made available and a 25% held-out test set.
- During testing, 10 probabilistic stage predictions were generated per biopsy and mean-pooled to come to a single class probability for each stage, per biopsy. *n* biopsies were then mean-pooled per patient to come to a single stage prediction.

RESULTS:

- 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

CONCLUSION:

Cancer stage can be classified by whole-slide histology imaging on a test set balanced by demographic background, particularly in early- and advanced-stages. Future work will involve further fine-tuning of the method and subgroup analysis of performance.

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.