Robustness of Deep Learning Histogenomic Models to Tissue Area, Tumor Purity, and Scanner Type

Poster # 167

Qiyuan Hu, Kunal Nagpal, Rohan P. Joshi, Yoni Muller, Martin C. Stumpe, Nike Beaubier
Tempus AI, Inc., Chicago, IL

Disclosures: All authors are employees of and stockholders in Tempus AI, Inc., a for-profit company.

INTRODUCTION

Microsatellite instability-high (MSI-H) is a tumor-agnostic biomarker for immune checkpoint inhibitor (ICI) therapy. Previous studies have shown that AI-based imaging predictors can infer MSI status from hematoxylin and eosin (H&E) whole-slide images (WSIs). Deployment of these AI predictors requires the characterization of assay operating parameters and specimen requirements. Here, we evaluated the performance robustness of a model trained to predict MSI-H status from H&E WSIs of prostate cancer with respect to tissue area, tumor purity, and scanner type.

METHODS

H&E-stained WSIs of prostate biopsies and surgical resections containing carcinoma were split into a model training set (n=4015, MSI-H 1.8%) and a validation set enriched for MSI-H (n=1523, MSI-H 4.3%). A subset of 1000 slides was scanned on both Aperio GT450 and Philips UFS scanners.

Attention-based multiple instance learning models were trained to predict MSI-H status (labels determined via NGS) from H&E WSIs. Prediction scores were ensembled from 5-fold cross-validation. For tissue area simulations, model inference was run with various numbers of randomly sampled tiles from each slide. For tumor purity simulation, pathologists annotated tumor areas for 60 slides (equal parts of MSI-H and MSS), and different proportions of tumor tiles and stroma tiles were sampled to make up sets of 200 tiles to simulate various tumor percentages. The experiment was repeated 10 times with random samples.

	MSI-H, n = 138	**MSS**, n = 5400
Procedure Type		
Biopsy	18	578
Core needle biopsy	80	2910
Excisional biopsy	1	22
Incisional biopsy	0	1
Surgical resection	37	1669
Unknown	2	220
Gleason Score		
7	5	869
8	18	822
9	61	2076
10	33	368
Unknown	21	1265

SUMMARY

- Our histogenomic model is effective for tissue area down to 0.6 mm^2, is confident on samples with 50% and higher tumor purity, and is robust to scanner variation.
- These findings demonstrate the robustness of the model and quantify the limit of detection and quality control metrics for the model performance, establishing a foundation for future deployment of such histogenomic algorithms.

RESULTS

Model performance was robust down to a simulated tissue area of 0.6 mm^2, 0.01 percentile in our dataset and is much smaller than a typical core needle biopsy, and a simulated tumor percentage of 50%, 25 percentile in our dataset.

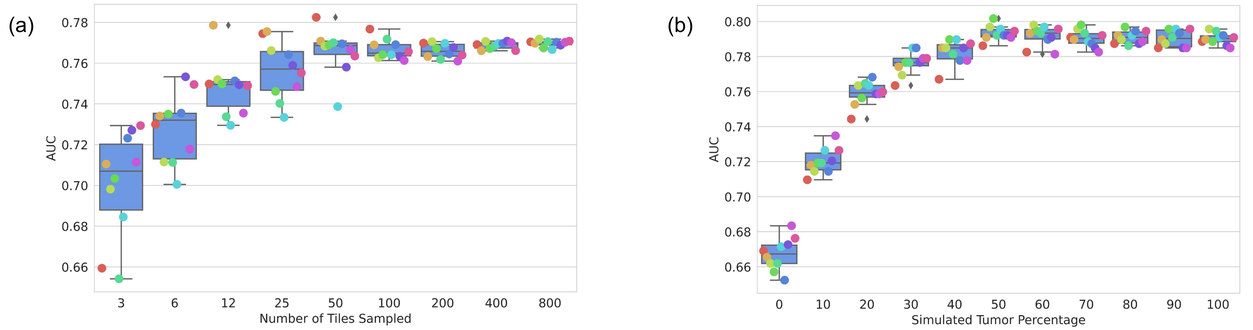

Figure 1. Area under the receiver operating characteristic curve (AUC) for predicting MSI status across the range of (a) tissue areas and (b) tumor percentages in simulation experiments.

Slides scanned by two different scanners yielded similar AUC, and prediction scores of these paired scans demonstrated high correlation.

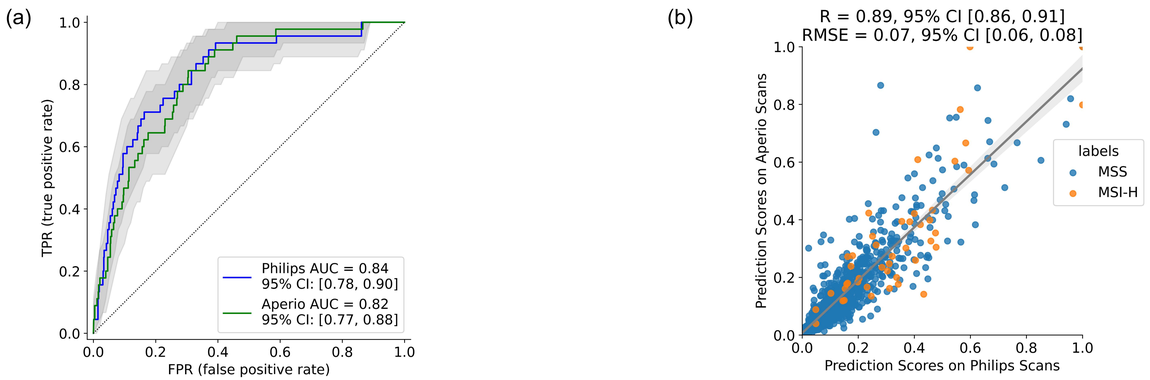

Figure 2. (a) Receiver operating characteristic (AUC) curves for predicting MSI status and (b) correlation between predictions scores from paired scans of the same slides scanned by two different scanners.

References:
1. Alam MR, et al. Recent Applications of Artificial Intelligence from Histopathologic Image-Based Prediction of Microsatellite Instability in Solid Cancers: A Systematic Review. Cancers. 2022
2. Hu Q, et al. Development and Validation of a Deep Learning-Based Microsatellite Instability Predictor from Prostate Cancer Whole-Slide Images. arXiv preprint arXiv:2310.08743. 2023 Oct 12.

Acknowledgments: We thank Dana DeSantis from the Scientific Communications team for poster development.

Correspondence: Isabelle.Hu@tempus.com