Deep Learning Identifies Microsatellite Instability in H&E Whole Slide Images from Prostate, Esophageal and Gastric Cancers and Generalizes across Cancer Types

Poster: 165

Rohan Joshi, MD, PhD [1], Andrew Kruger, PhD [1], Elle Moore [1], Ryan Jones, MD [1], Martin Stumpe, PhD[1]

[1]Tempus Labs, Inc., Chicago, IL // **Disclosure:** All authors are employees of Tempus Labs, a for-profit company // **Correspondence:** rohan.joshi@tempus.com

INTRODUCTION

Defective mismatch repair (dMMR) proteins and high microsatellite instability (MSI-H) are associated with a positive response to checkpoint inhibitor therapy in colorectal and non-colorectal cancers. Because of the low prevalence in certain non-colorectal cancers, testing for dMMR and/or MSI-H is not routinely performed, particularly at the time of initial biopsy. Here, we tested the ability to predict MSI status from H&E whole slide images (WSIs) in prostate cancer and to generalize model predictions to gastric and esophageal cancers.

METHODS

WSIs and MSI labels (MSI-H or microsatellite stable [MSS]; obtained via next-generation sequencing) were collected from primary and metastatic colorectal, endometrial, and prostate cancer specimens (Figure 1).

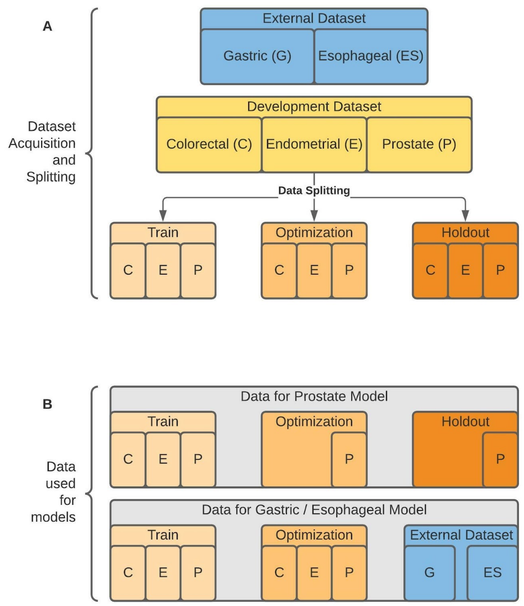

Figure 1. An attention-based convolutional neural network was trained to predict MSI status for each WSI in the training set. Hyperparameters and operating points were selected using the optimization set, targeting prostate cancer prediction (prostate cancer model) or simultaneous prostate, endometrial, and colorectal cancer prediction (gastric/esophageal cancer model). Results are reported on a fully-independent holdout set (prostate cancer model) or independently collected datasets (gastric and esophageal cancer model) to assess generalizability.

SUMMARY

Computational modeling can predict MSI status from H&E images, demonstrating the ability to enrich for patients who are likely to have actionable findings.

Our modeling framework detects MSI in prostate cancer (despite its low prevalence) and generalizes to gastric and esophageal cancers while not being trained on these types.

RESULTS

Illustration of deep learning pipeline for predicting MSI status from WSIs

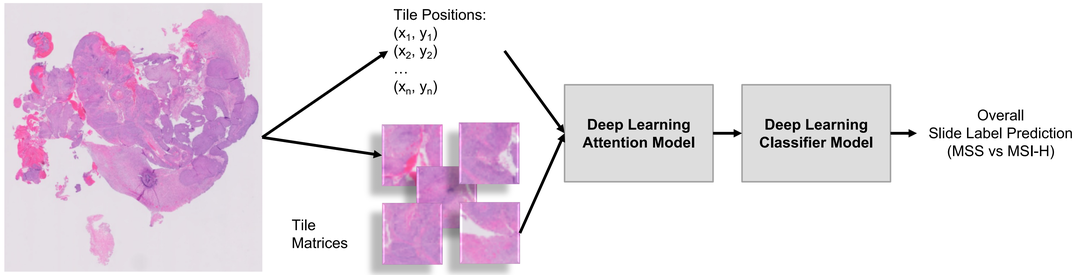

Figure 2. Whole slide images (WSIs) are broken up into 20x magnification tiles. Tile data and positions are recorded and passed to attention and classifier deep learning modules to create a prediction. During the training process, the weights of the deep learning models are iteratively updated until the area under the receiver operating characteristic curve (ROC-AUC) for the validation set no longer improves.

Dataset overview

Characteristic	MSI-H	MSS
Overall	441	5779
Development		
Colorectal	231	3132
Endometrial	126	430
Prostate	43	1322
External		
Gastric	26	320
Esophageal	15	575

Table 1. A subset of Tempus data was allocated across training, optimization, and holdout/external sets. Cancer types in which MSI-H is more prevalent (colorectal and endometrial) were leveraged in training to improve prediction for cancers with lower MSI-H prevalence (prostate) or fewer data (gastric, esophageal). "Development" data were used in both models whereas "External" data were used in the gastric/esophageal model.

Deep learning models discriminate MSI-H from MSS in Prostate, Esophageal, and Gastric Cancer

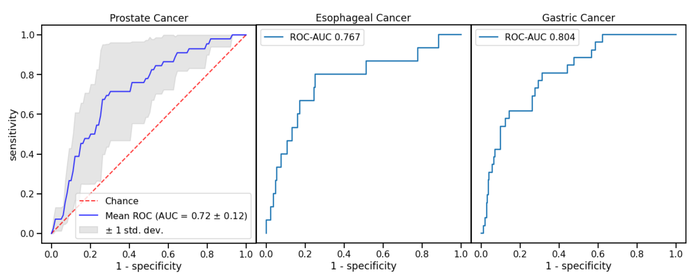

Figure 3. Receiver operating characteristics demonstrate the predictive accuracy of our models. For prostate cancer, there were enough cases available to perform training and holdout testing with cross-validation. The models distinguish MSI-H and MSS in prostate cancer despite low prevalence, and esophageal and gastric cancer despite not being trained on these cancer types.

Acknowledgements: We thank Elle Moore, Adam Hockenberry, Andrew Kruger, and other members Tempus for critical review and assistance with data visualization and presentation.

Demonstrating the predictive accuracy of trained models in a binary classification task

Prostate Cancer	True Status		Total
	MSS	MSI-H	
Predicted MSS	193	0	193
Predicted MSI-H	81	8	89
Total	274	8	
Summary Metrics	Sensitivity = 100% (95% CI: 63–100%) Specificity = 70% (95% CI: 65–76%) Positive Likelihood Ratio = 3.4 (95% CI: 2.8–4.0)		

Esophageal Cancer	True Status		Total
	MSS	MSI-H	
Predicted MSS	423	3	426
Predicted MSI-H	152	12	164
Total	575	15	
Summary Metrics	Sensitivity = 80% (95% CI: 52–96%) Specificity = 74% (95% CI: 70–77%) Positive Likelihood Ratio = 3.0 (95% CI: 2.3–4.0)		

Gastric Cancer	True Status		Total
	MSS	MSI-H	
Predicted MSS	185	5	190
Predicted MSI-H	135	21	156
Total	320	26	
Summary Metrics	Sensitivity = 81% (95% CI: 61–93%) Specificity = 58% (95% CI: 52–63%) Positive Likelihood Ratio = 1.9 (95% CI: 1.5–2.4)		

Table 2. Models were trained and the operating points were determined based on the optimization set, with a target of high sensitivity for a screening-type test. Here, confusion matrices show model predictions and operating point application to the holdout/external sets. Using estimated real-world prevalences of 5%, 2%, and 20% of MSI-H (in prostate, esophageal, and gastric cancers, respectively), we expect that 15%, 6%, and 32% (respectively) of patients would have detectable MSI-H status on follow-up NGS testing after a positive model result.