Validation and deployment of H&E image based model predicting total nucleic acid yield in multiple cancer types

Kshitij Ingale[1], Josh Och[1], Sun Hae Hong[1], Rohan P. Joshi[1], Kunal Nagpal[1], Yoni Muller[1], Kellen Malstrom[1], Ryan Jones[1]
[1]Tempus AI, Inc., Chicago, IL, USA

Disclosure Statement: All authors are current or former employees of Tempus AI, Inc. All current employees hold stock in the company.

INTRODUCTION

Next-generation sequencing (NGS) can be instrumental in enabling personalized treatment options for cancer patients. However, a primary reason for NGS testing failures is insufficient total nucleic acid (TNA) yield. In this work, we developed a model (p-Yield) to predict TNA yield score from routinely generated H&E slide images. This model was validated on a temporal test set, and is internally deployed to run on NGS samples to support lab workflow. We report robust performance across cancer types and procedure types in both validation and deployment data sets.

DESIGN

The p-Yield model was trained to predict TNA yield with NGS samples received between Jan - June 2023 using cell count (identified by a deep learning model[1]), sample age, tissue site, and procedure type. Predicted TNA yield was used to predict samples failed with DNA sequencing quantity-not-sufficient (QNS) status, and those yielding excess TNA (> 1000 ng). Binary metrics (NPV, PPV) used an operating point selected using a withheld tuning set. This model was validated on a temporally held-out set (July-Sept 2023) and evaluated on a post deployment set across multiple cancers and procedure types (July-Aug 2024).

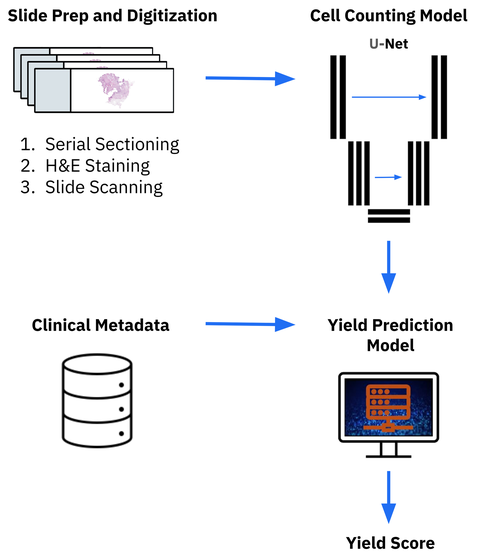

Figure 1. Model Workflow

SUMMARY

- The p-Yield model effectively predicted TNA yield and was robust across cancer types and procedure types
- Lab workflows can be enhanced by identifying slides with potentially low TNA yields and alerting clinicians to send additional samples for sequencing
- Cases with excess TNA yields are identified, enabling tissue conservation for sequencing and potential use in other biomarker assays

RESULTS

Receiver Operating Characteristic curves for labels stratified by clinical variables

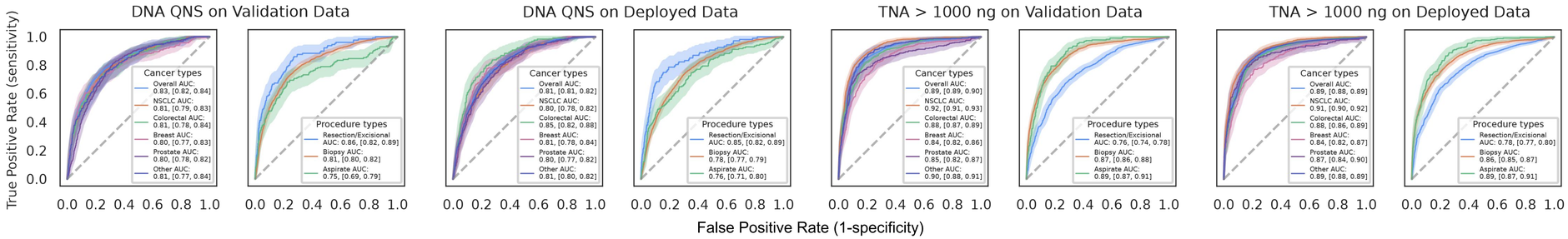

Table showing metrics based on operating point for DNA QNS and cases with excess TNA yield (TNA mass > 1000 ng) label

		DNA QNS						TNA mass > 1000 ng					
		Validation data			Deployed data			Validation data			Deployed data		
		Sensitivity	Specificity	Total Samples	Sensitivity	Specificity	Total Samples	Sensitivity	Specificity	Total Samples	Sensitivity	Specificity	Total Samples
Overall		0.39	0.94	13037	0.27	0.96	15769	0.73	0.89	13037	0.75	0.87	15769
Cancer types	Breast cancer	0.34	0.96	1807	0.23	0.96	1302	0.64	0.86	1807	0.64	0.87	1302
	Colorectal cancer	0.21	0.98	2680	0.18	0.97	1892	0.81	0.82	2680	0.78	0.83	1892
	NSCLC	0.33	0.95	4948	0.25	0.95	3431	0.71	0.92	4948	0.67	0.93	3431
	Prostate cancer	0.56	0.83	2002	0.57	0.84	1250	0.52	0.94	2002	0.51	0.96	1250
	Other	0.30	0.96	1600	0.22	0.97	7894	0.81	0.85	1600	0.80	0.84	7894
Procedure types	Aspirate	0.45	0.92	1152	0.38	0.93	1316	0.41	0.95	1152	0.44	0.96	1316
	Biopsy	0.40	0.93	8820	0.27	0.95	10173	0.43	0.96	8820	0.45	0.96	10173
	Large sample	0.21	0.98	2990	0.17	0.99	4062	0.95	0.25	2990	0.96	0.15	4062

ACKNOWLEDGMENTS: We thank Bolesław L. Osinski for technical support. We thank Dana DeSantis from the Tempus Science Communications team for poster development.
REFERENCES: [1] Osinski, B.L., et al. Artificial intelligence-augmented histopathologic review using image analysis to optimize DNA yield from formalin-fixed paraffin-embedded slides. Mod Pathol 35, 1791–1803 (2022)

Correspondence: kshitij.ingale@tempus.com

Conference-specific guidelines

USCAP 2025

* Draft due to SciComms: Feb. 28

* Legal Review Deadline: March 10

* Printing: Send to postersmith.com by March 18

* March 22-27, 2025: USCAP Conference