Localized Tumor Purity Prediction from H&E (Hematoxylin & Eosin) Stained Whole Slide Images using Cell Type Proportions from Cell-type Deconvolution Methods

Neil Daniel, Edward Kim, Alim A. Oraz, Neha Ranjith - EDIT AI & Machine Learning Program
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth Health

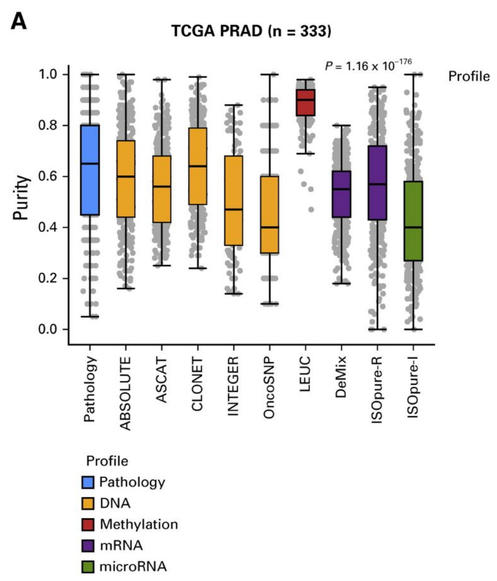
ABSTRACT

- **Novel Approach:** Developed a CNN using cell-type deconvolution data from spatial transcriptomics to predict localized tumor purity from H&E stained images, offering a cost-effective and accurate alternative to traditional methods
- **Key Results:** Achieved a mean absolute error (MAE) of 0.1016 on unseen H&E images, demonstrating strong generalizability that outperforms competing computational approaches
- **Clinical Impact:** Enhances precision in tumor assessment, aiding in personalized cancer treatment and improving the interpretability of genomic data
- **Future Directions:** Aiming to extend predictions to the whole slide image level and refine the model for broader clinical applications in computational pathology, including the integration of a cell detection framework

INTRODUCTION

- **Complex Tumor Microenvironment:** Tumor tissues contain diverse cell types, including stromal, epithelial, and immune cells, complicating the interpretation of molecular profiles and influencing treatment decisions
- **Critical Role of Tumor Purity:** Tumor purity, the proportion of cancerous cells in a tumor, is essential for interpreting genomic data, determining tumor mutational burden (TMB), and guiding patient-specific cancer therapies
- **Challenges with Current Methods:** Traditional tumor purity estimation methods, such as manual H&E slide evaluation and genomic approaches, are limited by inefficiencies, subjectivity, and high costs, leading to a demand for more precise and accessible alternatives
- **Need for Innovative Solutions:** Existing computational approaches, including multiple instance learning (MIL) and deep learning-based spatial purity mapping, offer potential but fall short in providing strongly accurate and localized purity estimates
- **Research Objective:** This study introduces a novel method using spatial transcriptomics and cell-type deconvolution to train a CNN that predicts localized tumor purity from H&E stained images, aiming to improve accuracy, generalizability, and accessibility in tumor assessment

- **Fig. 1** (Haiden et al., 2020) illustrates the large discrepancies between different purity estimation methods
- This complexity challenges molecular profiling and underscores the need for advanced methods to accurately assess tumor purity
- Our research aims to address this challenge, enhancing the precision of tumor purity estimation for better-informed cancer treatment decisions



METHODS

- **Data Sourcing:** Single-cell gene expression data were sourced from the c295 colon cancer atlas (371,223 cells) and Dartmouth Hitchcock Medical Center (10 tumors: 8 colon, 2 breast), focusing on the most variable genes for each dataset
- **Label Transfer:** Conducted label transfer from the c295 atlas to the Dartmouth single-cell data using a semi-supervised SCANVI model, providing cell type annotations necessary for accurate cell-type deconvolution
- **Cell-Type Deconvolution:** Applied the cell2location model to estimate cell type proportions within visium spots, enabling localized tumor purity calculation by dividing tumor epithelial cells by total epithelial cells
- **Model Architecture and Training:** Developed a CNN with 5 convolutional layers, batch normalization, max pooling, and dropout layers, using ReLU activation. The model was trained with the Adam optimizer (learning rate: 10^-3) on 256x256 pixel H&E image patches, following an 80-20 train-test split with iterative patient selection to manage memory constraints
- **Normalization and Tumor Purity Calculation:** Normalized cell type abundances to ensure consistency across visium spots, then calculated tumor purity using the proportion of tumor epithelial cells, providing ground truth for CNN training
- **Model Evaluation:** Assessed the CNN's performance on the validation set using mean absolute error (MAE) and mean squared error (MSE) metrics, achieving strong generalizability on unseen WSIs and suggesting real-world utility
- **Limitations and Future Work:** Addressed memory constraints by loading and training image patches in batches, with future work focusing on extending predictions to the whole slide image level and integrating a cell detection framework for broader applicability

RESULTS

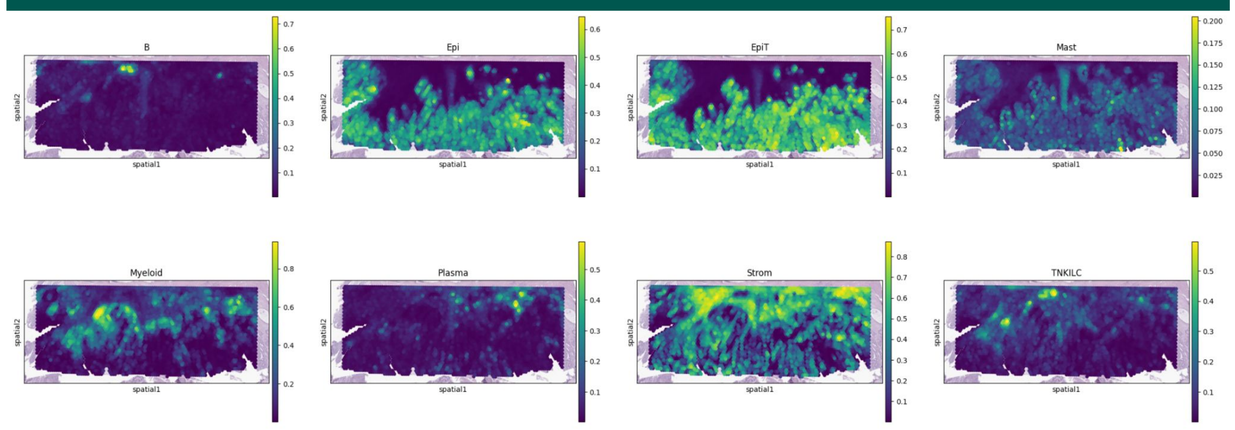


Fig. 2: Proportion heatmaps for each cell type on a selected Whole Slide Image (WSI)

- **Model Accuracy:** Achieved a mean absolute error (MAE) of 0.1016, demonstrating strong generalizability in predicting localized tumor purity from H&E stained images
- **Comparison with Existing Methods:** Improved on traditional methods by offering localized purity estimates, enhancing tumor heterogeneity understanding
- **Clinical Impact:** This method offers a cost-effective way to assess tumor purity, potentially improving personalized treatment decisions in clinical practice
- **Future Directions:** Further work will extend predictions to whole slide images

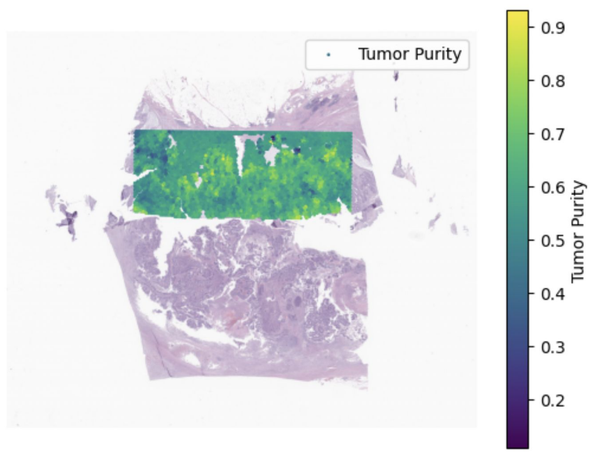


Fig. 3: Spatial Purity Map generated for a WSI

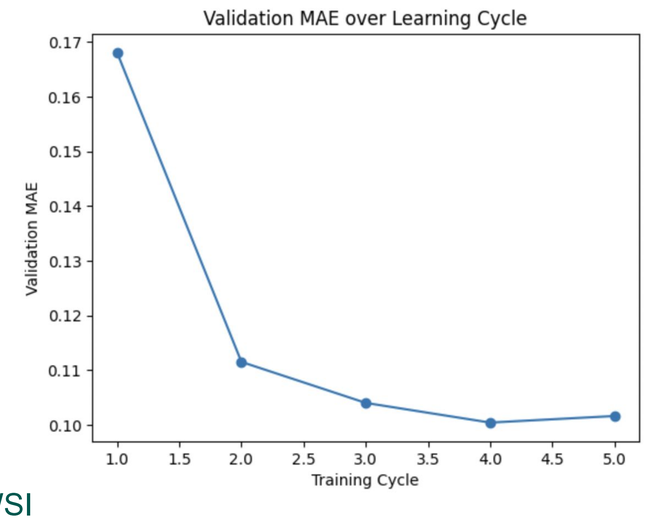


Fig. 4: Mean Absolute Error On Validation Set

CONCLUSION

- **Innovative Approach:** Developed a CNN-based method leveraging spatial transcriptomics to predict localized tumor purity from H&E stained images, advancing computational pathology
- **Clinical Impact:** The model's potential for precise tumor purity estimation supports more personalized and precise cancer treatments
- **Strong Generalizability:** Achieved a MAE of 0.1016, with strong validation results, though further refinements are needed for broader applicability
- **Comparison to Existing Methods:** Offers localized purity estimates, improving on traditional methods and providing more detailed insights into tumor heterogeneity
- **Future Directions:** Next steps include extending predictions to whole slide images and integrating a cell detection framework to improve accuracy and clinical utility
- **Data and Code Availability:** Data and code available on request subject to privacy/ethical restrictions

Acknowledgements: EDIT, DPLM, Pathology Shared Resource, DCC@DHMC

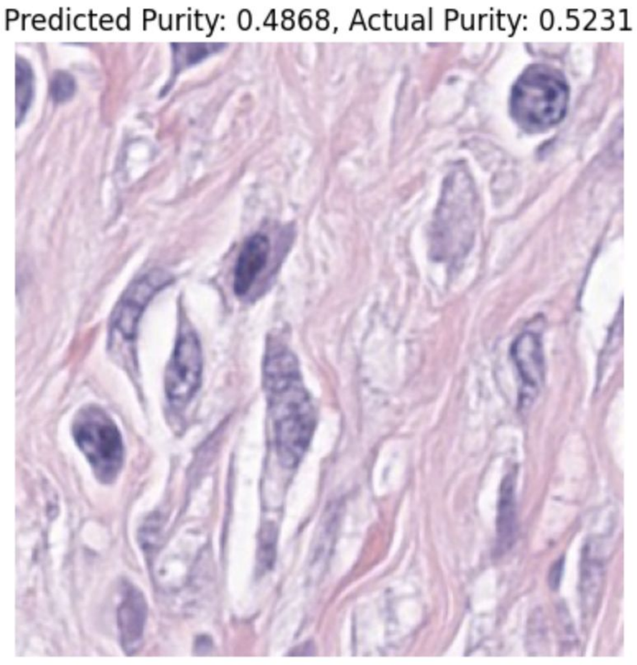


Fig. 5: Predicted and Actual Purity on Single H&E Image Patch From Validation Set

Manuscript available at QR code:

edit.