AI-Driven Single-Cell Tumor Classification in Pancreatic Cancer for Enhanced Histologic Diagnosis

Ava Hsieh[1*], Le Zhou[2,3], Adel Kardosh[4], Shaun Goodyear[4] , Zheng Xia[2,3,4]
[1]Partnership for Scientific Inquiry (PSI) Program, Knight Cancer Institute, Oregon Health and Science University, Portland, OR
[2]Department of Biomedical Engineering; [3]Computational Biology Program; [4]Knight Cancer Institute, Oregon Health & Science University, Portland, OR
[*]Presenting author

Background

- **Single-cell image classification in hematoxylin and eosin (H&E)-stained histopathology is a powerful technique for profiling tumor heterogeneity and guiding clinical decision-making.**
- **Adenocarcinoma** cells originate from glandular epithelial tissues and are a hallmark of malignant progression in a variety of cancers.
- Adenocarcinomas represent a critical diagnostic marker due to their potential to spread and disrupt organ function.
- **Squamous** epithelial cells play a protective role in surface tissues but are also prone to malignant transformation, often leading to squamous cell carcinoma.
- Accurate identification of these two cell types provides valuable insight into tumor origin, grade, and aggressiveness.

Objective

In this study, we explore the use of machine learning to improve the accuracy and efficiency of cell segmentation and classification, focusing on distinguishing adenocarcinoma cells and squamous epithelial cells.

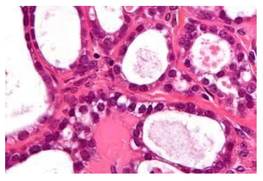


Fig. 1 Adenocarcinoma Cell

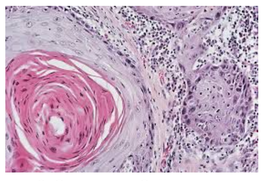


Fig. 2 Squamous Cell

Methodology

Fig. 1 Vision Transformer model

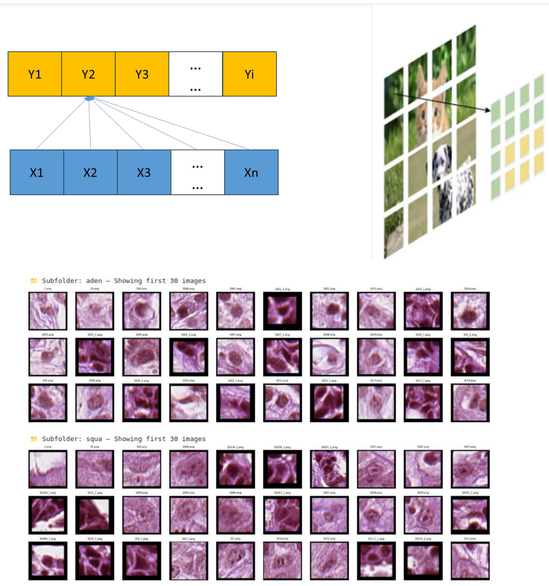


Fig. 2 Training data preparation with adenocarcinoma and squamous cells

- To classify the images, we used a ViT (vision transformer)
- To create training data, we manually annotated cell regions using QuPath, then applied both its built-in cell detection algorithm and the deep learning-based InstaSeg for segmentation.
- QuPath identifies cell nuclei by detecting local intensity maxima and estimates full-cell boundaries using a watershed transform, simulating topographic flooding to isolate adjacent cell regions with high precision.

Results

1. **Table representing the class, recall, precision, and F-1 score of each type of cell**

Fold	Fold 1		Fold 2		Fold 3		Fold 4		Fold 5	
Class	aden	squa	aden	squa	aden	squa	aden	squa	aden	squa
Recall	0.7875	0.8281	0.7156	0.85	0.7875	0.8469	0.8469	0.7781	0.8438	0.7156
Precision	0.8208	0.7958	0.8267	0.7493	0.8372	0.7994	0.7924	0.8356	0.7479	0.8208
F1-Score	0.8038	0.8116	0.7672	0.7965	0.8116	0.8225	0.8187	0.8058	0.793	0.7646

Figure 1. Table describing the model's scores on a basis of multiple factors; the type of cell, the accuracy that it finds cells, and accuracy that it correctly finds cells.

Dataset 1

- Extracted 12,191 squamous cells; 2,319 adenocarcinoma cells
- Randomly selected 2,000 squamous and 2,000 adenocarcinoma for training
- The rest 300 squamous and randomly selected 300 adenocarcinoma used for testing

300 squamous (accuracy 85%)

	Count
Aden	77
Squa	223

300 adenocarcinoma (accuracy 74%)

	Count
Aden	256
Squa	44

Dataset 2

- 6,000 adenocarcinoma and 6,000 squamous for training
- 15,699 adenocarcinoma and 29,991 squamous for testing

15,699 adenocarcinoma (accuracy 93%)

	Count
Aden	14,610
Squa	1,089

29,991 squamous (accuracy 86%)

	Count
Aden	4,069
Squa	25,822

Conclusion

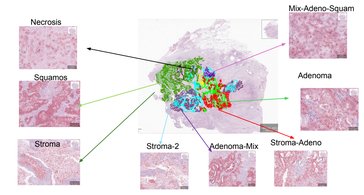
Of the adenocarcinoma cells, 14,610 were correctly classified, yielding a 93.1% accuracy. For the squamous cells, 25,822 were correctly identified, with an accuracy of 86.1%. Misclassifications primarily occurred between the two cell types, reflecting the morphological overlap in certain regions.

These findings highlight the potential of automated single-cell classification to support pathologists in tumor characterization.

Future Steps

Future work will focus on expanding the dataset to include additional cell types (e.g., stromal, lymphocytic) and further improving model generalization through multimodal learning and transformer-based architectures.

New annotations with more types of data



Acknowledgements

Partnership for Scientific Inquiry
Building tomorrow's researchers today