

Presentation 2021: Amgad Public Access TILs Nuclei Dataset

[Back to Presentations and Publications Main Page](#)

Presentation

NuCLS: A scalable crowdsourcing, deep learning approach and dataset for nucleus classification, localization and segmentation

- Presenter: **Mohamed Amgad, MD, MSc**
- Wednesday, March 10, 2021, at 2pm EST

Talk Summary

High-resolution mapping of cells and tissue structures provides a foundation for computational TILs assessment. Mohamed Amgad will be giving a public talk at the FDA describing their newest work, [NuCLS](#), and the public release of 220,000 annotations of cell nuclei in breast cancers.

Detailed Abstract

High-resolution mapping of cells and tissue structures provides a foundation for developing interpretable machine-learning models for computational pathology. Deep learning algorithms can provide accurate mappings given large numbers of labeled instances for training and validation. Generating adequate volume of quality labels has emerged as a critical barrier in computational pathology given the time and effort required from pathologists. In this paper we describe an approach for engaging crowds of medical students and pathologists that was used to produce a dataset of over 220,000 annotations of cell nuclei in breast cancers. We show how suggested annotations generated by a weak algorithm can improve the accuracy of annotations generated by non-experts and can yield useful data for training segmentation algorithms without laborious manual tracing. We systematically examine interrater agreement and describe modifications to the Mask.RCNN model to improve cell mapping. We also describe a technique we call Decision Tree Approximation of Learned Embeddings (DTALE) that leverages nucleus segmentations and morphologic features to improve the transparency of nucleus classification models. The annotation data produced in this study are freely available for algorithm development and benchmarking at: <https://sites.google.com/view/nucls/home>.