Computer Science Defense

Computational discovery of interpretable histopathologic prognostic biomarkers in invasive carcinomas of the breast

Mohamed Amgad
Emory University

Abstract: While microscopic examination of tumor resections and biopsies has been a cornerstone in breast cancer grading for decades, it suffers from considerable inter-rater variability due to perceptual limitations and high clinical caseloads. Computational analysis of whole-slide image scans using convolutional neural networks (CNN) can help address this challenge. Unfortunately, CNNs can be difficult to interpret, which motivates our adoption of an approach called concept bottlenecking, where models first detect various tissue structures then use them to make their prediction. Concept bottleneck models require a large set of manual annotation data to train. Unfortunately, manual delineation of histopathologic structures is very demanding and impractical given pathologists' time constraints. This dissertation describes contributions that fall under the themes of scalable data collection, deep learning-based tissue detection, and the discovery of novel histopathologic biomarkers and associations. ¡br¿¡br¿
First, we examine crowdsourcing approaches that engage medical students to collect manual annotation data. Our results show that a structured, collaborative approach with pathologist supervision is scalable; the resultant publicly-released BCSS and NuCLS datasets contain 20,000 and 200,000 annotations of tissue regions and nuclei, respectively. We show that medical students produce accurate annotations for predominant, visually distinctive structures and that algorithmic suggestions help scale and improve the accuracy of annotations.¡br¿¡br¿
Second, we describe a set of CNN modeling approaches for the accurate delineation of histopathologic structures. We describe various improvements to enhance the performance of nucleus detection CNN models and introduce a technique called Decision Tree Approximation of Learned Embeddings, which helps explain CNN nucleus classifications without compromising prediction accuracy. Additionally, we offer consensus recommendations from the International Immuno-Oncology Working Group surrounding the computational detection of tumor-infiltrating lymphocytes, a critical emerging biomarker. Following these recommendations, we develop and validate a multi-scale CNN model that jointly detects tissue regions and nuclei, employing pre-defined biological constraints to improve accuracy.¡br¿¡br¿
Finally, we describe the development of a morphologic signature based on quantitative features extracted from computationally-delineated histopathologic regions and cells. This morphologic signature relies partly on a set of stromal features not captured by clinical guidelines for breast cancer grading, and has a stronger independent prognostic value.

Tuesday, November 30, 2021, 1:00 pm
https://northwestern.zoom.us/j/95813522155

Computer Science
Emory University