Predicting the attention trajectories of digital pathologists: Towards understanding the variability in cancer diagnosis

Greg Zelinsky, Psychology, Dimitrios Samaras, Computer Science, Kenneth Shroyer, Pathology, and Joel Saltz, Bioinformatics

A pathologist's interpretation of a tissue biopsy from breast or prostate or brain has life-impacting consequences, on par with a judge sentencing a person to prison. But unlike the legal system, where judgments obey intricate rules of law and jurors and opposing advocates exist for the sole purpose of getting the decision right, the process by which a pathologist arrives at an interpretation is highly idiosyncratic. No one knows how, or even, if, a pathologist's idiosyncratic viewing behavior while interpreting a slide biopsy affects their final diagnosis. What is known is that ~1.4 million women each year trust in pathologists to "get it right", but all too often they get it wrong. As one example, it is estimated that atypical hyperplasia and ductal carcinoma in situ are misclassified 50% of the time. This is an alarming error rate, given that each miss may mean that an aggressive form of cancer goes untreated and each false positive means that a person needlessly undergoes agonizing radiation or chemical treatment. We propose exploring the possibility that some of these errors may be linked to intra-pathologist and inter-pathologist variability in how slides are inspected during interpretation. Our broad goals will be to: rigorously characterize slide interpretation behavior and to quantify its variability, exploit recent advances in digital pathology and computer vision to identify the features used by pathologists in their interpretations, and apply this understanding to the creation of interventions that target this variability with the goal of reducing diagnostic errors. More specifically, our aims are as follows.

(1) Assemble a team of pathologists and a dataset of digital pathology images. Crucial to the success of this project is the assembly of two groups of participating pathologists. One group will be tasked with: selecting digitized whole-slide images of prostate and breast biopsies to be used as stimuli, determining ground truth interpretations, and labeling diagnostically-meaningful patches in these images for model training and model and behavioral evaluation. The second group will be tasked with interpreting this test set of images and rendering diagnoses, thereby producing the viewing behavior and diagnostic data that we will quantify and experimentally evaluate.

(2) Collect and analyze attention trajectories. Using the dataset created under Aim 1, in Aim 2 work we will collect attention trajectories (changes in magnification and x,y eye position in a viewport over time) from the testing group of pathologists as they interpret the digitized slides selected by the ground- truth group and make diagnoses. Similarity in attention trajectory will be computed between two pathologists viewing the same slide, and this will be done for all slides and all pair-wise combinations of pathologists. From this rich behavioral dataset we will be able to relate accuracy of diagnosis to variability in attention trajectory, measured both between pathologists interpreting the same slide(s) and across interpretations of different slides from the same pathologist(s).

(3) Use artificial intelligence to learn a pathologist's features of cancer. Aim 3 work will have deep neural networks learn the visual features that pathologists use while interpreting digital slides. One model will be trained using supervised learning based on the labels provided by the ground-truth group. Another model will be trained using attention-based learning, where the network learns features from the image regions that were fixated by the test group of pathologists during their interpretations. Predictions of diagnostic grade will be made by both models, and compared. To the extent that the attention-trained model better predicts a given pathologist's diagnosis, this would be computational support for a relationship between diagnosis and the attention trajectory taken by that pathologist during interpretation. It would also identify the features that pathologists use to guide their attention during slide interpretation.

(4) Explore AI-informed interventions for improving cancer classification success. Informed by the behavioral analysis and modeling work, two tools will be developed aimed at improving the accuracy of digital pathologists. These tools could be used both as interventions during pathology practice as well as

learning devices during medical school training. One tool will provide the user with potentially useful information that they may have missed during their interpretation, and the other tool will be aimed at reducing intra- and inter-pathologist variability by intervening when a deviant attention trajectory is identified. Interventions will consist of suggested “alternative views” and requests for “second looks”.