**Automatic Interpretation of Digital Pathology Images Using Deep Learning**

**Overview**

**Prepared by PI Iyad Obeid, PhD  
12/19/2022**

In Year 1 of this two-year project, the team has focused its efforts in two main areas. The first area is the creation, annotation, and curation of digital pathology slides. This has been done in collaboration between the lead team at the College of Engineering and partners at the Larry Katz School of Medicine (LKSOM) and the Fox Chase Cancer Center (FCCC). During 2022, the team digitized approximately 25,000 slides from Temple Hospital as well as another 15,000 from Fox Chase. Following the process outlined in the project proposal, these have been scanned using high resolution, multi-level imaging and stored in our HIPAA-compliant petabyte fileserver. The process has the approval of Temple IRB. Slides sourced from Temple Hospital are covered by IRB Protocol #24943, while those from FCCC were ruled to be “not human research.”

The process of curating slides is on-going. As outlined in our proposal, we have been training a cohort of undergraduate students to identify abnormal tissue specimens for our database. Those students receive hands on training on this process from pathologists on the LKSOM team. The process of training students and performing quality control on their work is slow and challenging, but already we have labeled about 1000 slides and expect to do substantially more in 2023. To date, the data science team in CoE has employed 18 undergraduate and graduate students.

The second main area of effort has been preparing and submitting proposals for federal funding. This was a key requirement of the Catalyst Program funding. In doing so, we have sought to leverage the data and team-building that have resulted from this project. We received a consultation from McAllister & Quinn in which we discussed candidate federal programs that would be a good fit for our work.

In August 2022, we submitted “Deep Learning for Whole Slide Segmentation of Prostate Cancer Needle Biopsies” to the Congressionally Directed Medical Research Program. In this proposal (3 years, $1.5M), we seek to create image processing software for analyzing slides of prostate biopsy tissue. As stated in the proposal, “Tissue samples from prostate biopsies are usually examined under a microscope by a pathologist. The process is time consuming and error prone because the pathologist must examine every millimeter of every slide in order not to miss anything important. Instead of spending time identifying, describing, and staging possibly cancerous cells, pathologists spend the bulk of their time simply looking for those cells. This effectively dilutes the pathologist’s expertise, which is best spent carefully studying suspicious cells for signs of cancer.”

In November 2022, we submitted “SCH: Unsupervised Deep Learning for Segmentation of Digital Pathology Images” to the National Science Foundation. Instead of a specific focus area (such as prostate cancer or breast cancer) the focus of this proposal is “agnostic” machine learning technology that can identify abnormal tissue in any domain. This is a significant and important technical challenge, since most machine learning in this area is specific to a single application and not transferable to other domains. If funded (4 years, $1.2M) this program will allow us to make significant developments in this field and will likely translate into commercial opportunities.

In addition to these main thrusts, this project has also sought to create new techniques for using artificial intelligence to link important information in textual physician reports to pathology images. Our strategy for innovating in this space was to use eye tracking to measure what the physician looks at when reading a report and use that to train software to finding important text automatically.

The primary work in this area during 2022 has been to better define the technical requirements for the eye tracking technology. The College of Science and Technology team has performed a market study on high end eye tracking (Tobii) as well as free webcam-based alternatives and is currently evaluating several candidate technologies. This market study has included interactions with representatives from various manufacturers and university labs. The process of gathering training data and building a prototype AI is expected to commence in 2023.