12-23-2021

Machine Learning Applications with Qupath in Histopathology

Bianca Lang
Washington University in St. Louis

Independent Study:

Machine Learning Applications with

Qupath in Histopathology

Lake Lab

Fall 2021

Bianca Lang

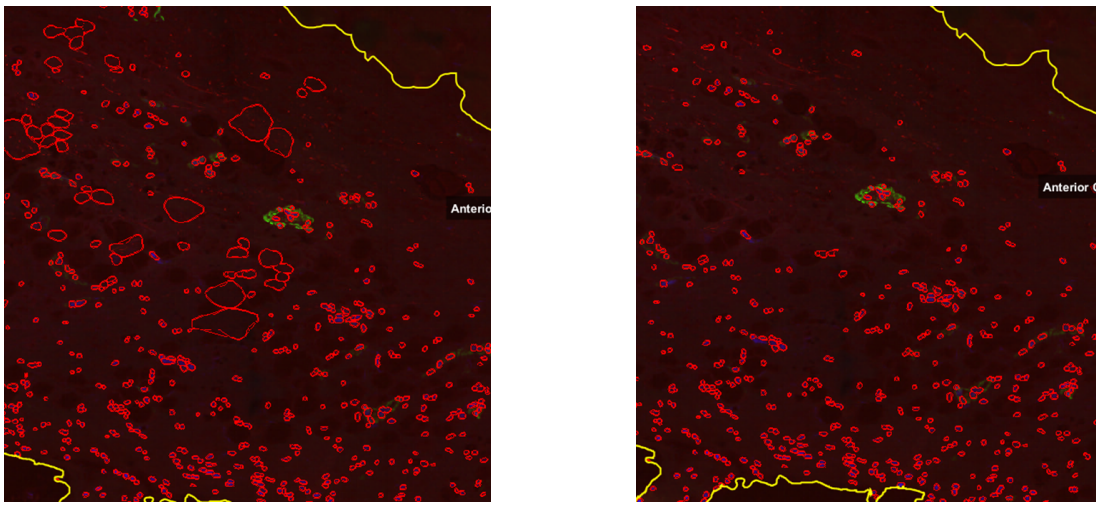
December 22, 2021

The Lake Lab studies the debilitating condition of post traumatic joint contracture (PTJC), characterized by loss of elbow function/motion. PTJC develops because an acute elbow trauma leads to damage and irreversible changes in the periarticular soft tissues, such as the joint capsule and cartilage. Current therapies are often unsuccessful in preventing or reversing PTJC due to the lack of knowledge on the tissue pathogenesis. Further research on the pathogenesis is necessary to create preventative treatment strategies. The Lake lab utilizes an animal injury model to help understand the cellular mechanisms and healing response of the elbow tissue, in turn allowing for the development of various tissue-targeted treatments for elbow joint dislocation. Currently the lab uses histopathology to study cellular mechanisms of disease in the animal model, but histopathology has several limitations due to the invasive, manual nature and the necessity for a pathologist's evaluation. As a result of these limitations, new approaches are now warranted. Digital histopathology, when paired with machine learning techniques, show promise of discovering new information about elbow soft tissues. Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that imitates the way that humans think and learn.

My primary objective for this semester was to help develop machine learning techniques that help further our understanding of the cellular mechanism and the healing response. These techniques assist in quantifying the cartilage and capsule features. I began by familiarizing myself with QuPath, a bioimage analysis software used for digital pathology applications. The application of QuPath is instrumental in the categorization of different features of the DH slides. The intention with the application of machine learning is to reduce time spent identifying different parts of the cell and focus

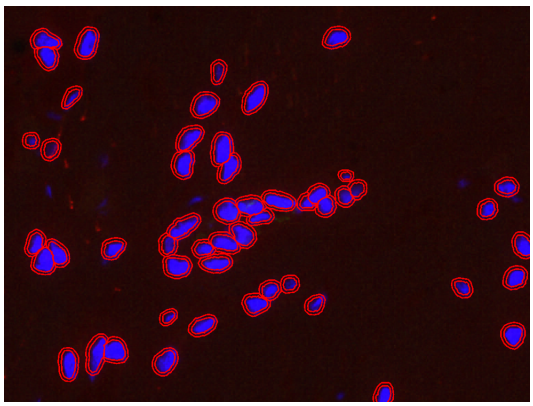
on data analysis. While QuPath has several built in features for analyzing H&E slides, it also has the capacity to add-on stardist, allowing the user to create customized scripts. StarDist is a deep-learning-based method of nucleus detection that can be applied in Qupath. Using these models and applications allows us to reduce time consumption, since we will no longer be individually tracing each nuclei, and allows us to create broader datasets through the parameter outputs of qupath.

To take advantage of stardist's features, I learned how to code in groovy. I also familiarized myself with existing applications of histopathology in qupath through published research papers and online trouble shooting qupath forums. I was then able to adapt code taken from published work of H&E samples in other projects, to better create scripts in groovy that could help us identify nuclei within the different slides. It provided the ability to adjust the thresholds of certain detection parameters (ex: intensity threshold, constraining the cell expansion, intensity values, the percentile normalization, shape measurements, etc.) My script also better suited the needs of the projects since it was able to overcome certain irregularities that the built-in script did not detect.



False positive removal above nuclei of 50 micrometers in the anterior humerus articular cartilage

For example, I was able to write a script that removes false positives above 50 micrometers in the detection of a nucleus. Another issue that was overcome was the overlap of nuclei, it was able to detect false overlap and separate each individual nuclei. By running this script on all the slides, I was able to remove the need for much of the individual mechanical work of tracing the cell features. Under the supervision of Mike, we can then use the output data sets created by QuPath and analyze to see the cellular difference between injured and noninjured models.



Differentiation and Separation of neighboring nucleus/cells

Other responsibilities I took upon within the lab include the lab work of MRI isolation and sample preparation. I was better able to understand chemical procedures, and helped prep slides before they were put in the image nanozoomer. One procedure I have assisted was the staining of cells. In addition to helping develop an algorithm for nuclei identification in the Machine Learning process, I focused on tracing in qupath. I primarily traced the anterior humerus articular cartilage and anterior humerus bone within a set of H&E images. Throughout the semester, I also read scientific literature to learn the background knowledge necessary to understand the purpose of the research.

I read about stardist, qupath, and learned about machine learning applications and other relevant biology topics such as the different stains applied to the cells for analysis. I synthesized relevant information from the readings to have a better understanding of my role and the overall goal of the lake lab. Throughout this semester, I effectively wrote an algorithm to identify specific DH and non-invasive image features used to quantify cell and tissue health. Ultimately, this along with my other contributions in the lab aided in predicting the disease state of post-traumatic elbow conditions.

.