**User-story**

When a patient is diagnosed/expected to be affected with cancer, the primary thing which is done irrespective of demographics of patient and the type of cancer is a biopsy of the area that is expected to have been affected. This step is recommended by a doctor ***(oncologist).***

As a part of this procedure, the clinician takes the tissue sample extracted from the patient and performs a H&E staining. On completion of this staining when this slide is placed under an imaging device, one can see different components of interest highlighted in different colors on the resultant whole -slide image ***(WSI)***. Most commonly they are trying to check the concentration and the spatial distribution of immune cells and their types. This is an extremely important analysis in determining the spread and stage of cancer.

Essentially, if this image had only a few components then the job of clinician would have been easy, and no need of further computational intervention would be needed.

Although, in cases of cancer, the clinician needs to manually calculate the number of certain features on the image, and this requires a high amount of precision and accuracy on clinician’s end. Also, the interpretation of image varies from clinician to clinician based on factors like imaging equipment, hours the clinician has been working for, expertise, and many other factors. This level of uncertainness has been a matter of concern for a very long time and thereby a tool aiding clinician and giving them a clearer perspective is the goal of this project.

The current computational work would essentially do the following things:

* Take an input WSI from the imaging equipment and/or the user (clinician).
* Make equisized patches of this WSI and run an unsupervised clustering algorithm on each of these patches. Post clustering overlay the original image with a clustered image and help clinician point out exactly where the different components lie.
* Based on clustering in 2., decide the cluster which contains immune cells and further segregate them into different types ***(Lymphocytes and macrophages)***.
* Automate the entire process and/or streamline the process such that it is highly intuitive for the clinician to use, and they can explore various features as and when required.

**Use-Cases**

* Generate clusters from an individual patch determine every component separately.
* Overlay a single, multiple or all clusters at once on a patch depending upon the analysis that clinician needs to perform.
* Automated process of an input WSI and multiple output overlayed patches.
* Automating components of analysis which are constant and enabling enough interventional capability for the user.
* Ability to classify different components within a particular subcluster of interest ***(if any exists)***

**Design and Components**

* **Language:** Python
* **Primary Libraries:** Numpy, Pandas, Matplotlib, OpenCV, Sci-kit learn.

Components required to achieve the Use-Cases:

**Component:** Using a clustering algorithm to generate clusters:

**Progress:** Currently, we are working towards choosing a clustering algorithm based on the type of patches and results we are obtaining from them. We have done a substantial amount of work using KMeans clustering. Next, we will be observing the performance of Spectral Clustering as well. Based on the prior research and the current insights, we believe ***KMeans and Spectral Clustering*** are the top two candidates for the task at hand.

**Component:** Pipeline Automation ***(WSI to overlayed images)***

**Progress:** We are working a code that would initially eliminate background form a WSI and create patches only for the tissues. The prior part of generating patches automatically is successfully completed.

**Component:** Capability for user inputs

**Progress:** Prior steps must be successfully completed to investigate this.

**Component:** Exploring a cluster of interest further.

**Progress:** We must be able to robustly generate clusters for any WSI that is an input. To do this the first two components must be looked at.

**Component:** Tests *(unit-tests)*

**Progress:** Some more thinking needs to be done before deciding the tests.