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 colours on the resultant whole -slide image *(WSI)*. In this procedure, the clinician goes through small patches from this **WSI** and looks for important biomarkers to analyse the spread and severity of cancer. 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.
- 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

- Automatically generate small *equisized* patches from the WSI generated by the clinician.
- Generate clusters from an individual patch to 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, Openslide, Matplotlib, Opencv-python (cv2), Sci-kit learn (sklearn), skimage, Unittest.

- • **Design requirements to achieve the *use-cases*:**

Component: Automatically generate small patches from the H&E-stained images.
Design: Used a combination of NumPy, Open slide, and skimage to generate patches.

Component: Developing a clustering algorithm to generate clusters:
Design: Utilized *k-means clustering* to generate centroid values for **RGB** *values.* Further developed a super-patch to account for deviations in representations because of varying contrasts.

*Component: Pipeline Automation **(WSI to overlayed images):***
Design: Developed a code for synthetically staining H&E patches using the clustering algorithm.

Component: Tests:
Design: Used *unittests* to build tests for the functions.