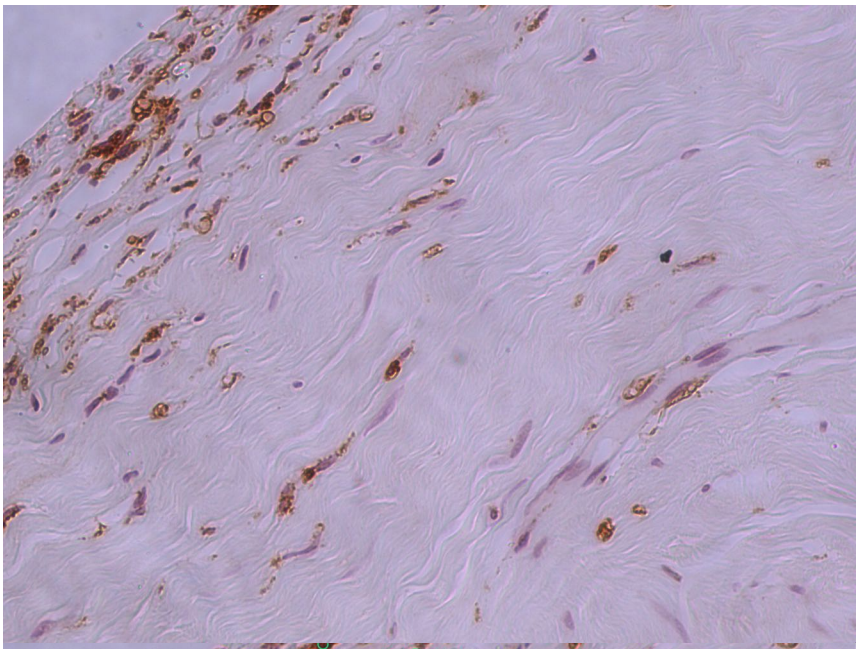
(Working) Project Title:
Detection of the Number of Stained Cells on Histology Slides Pertinent to Capsular Contracture on Histology Slides

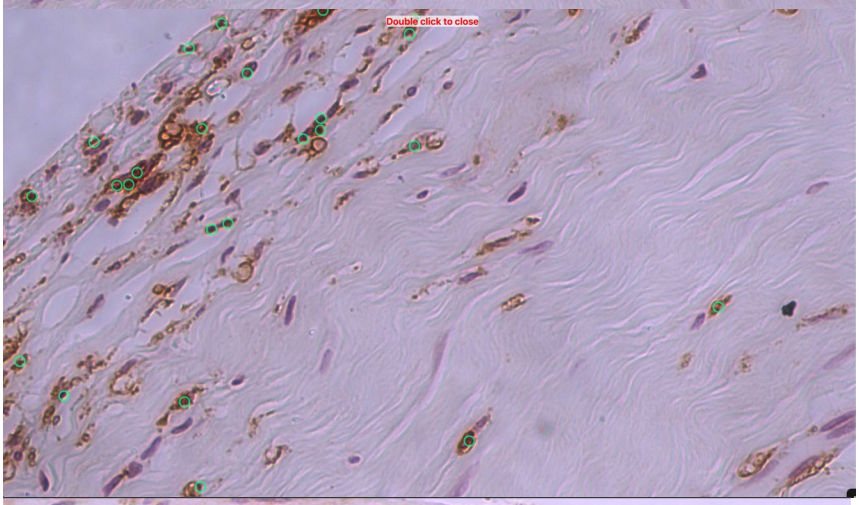
Team member: Fardad Jabbary

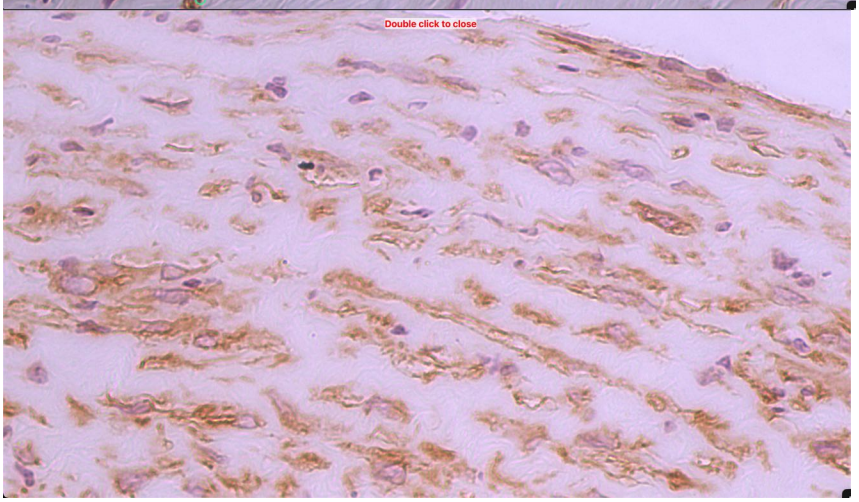
Capsular contracture is a pathology that could develop after implantation of a foreign capsule (e.g. a breast implant) into the body. Its exact causes are unknown and there is active research on understanding the disease better. There are some cells of interest thought to be contributive towards this pathology. In the process of doing further research on this issue, researchers around the world must manually count the number of stained cells on these slides – an arduous task, taking invaluable time away from highly skilled researchers.
Given the inherent diversity in the stained slides due to multitude of factors (e.g. amount of time in seconds stain left on the slide), as well as other factors such as where the pathology specimen was cut (hence out of a 3D volume of a cell, there may be only parts of the cell present in a slide without its nucleus) as well as others, this is an untrivial task, however, if it can be done with a certain amount of accuracy, a highly valuable one.

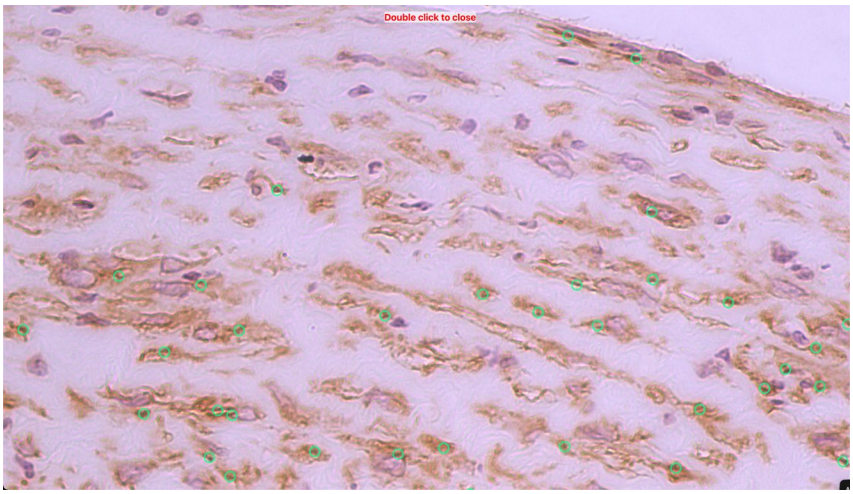
1. **Problem Statement**
   Estimating the number of stained cells of interest on a histology slide within a 10% margin of error of the actual count.

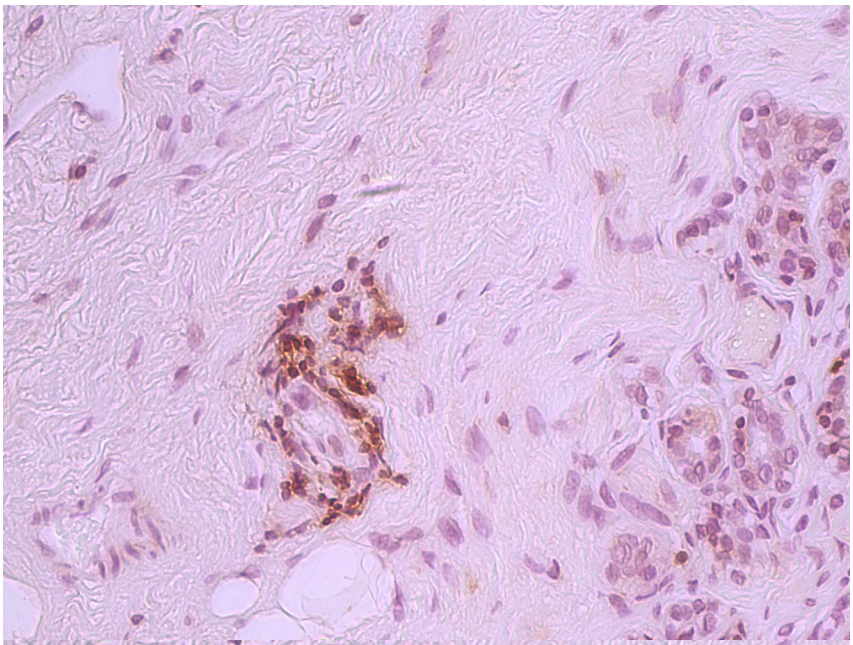
2. **Data Preprocessing**
   The original dimension of the images is 2048×1536. They otherwise have the same width and height. Some images are 'darker', or so they appear, because of the length of time the staining agents were left on them.
   This, however, cannot and must not be corrected as the real test data will also be variable. So the model must be trained on a diverse training dataset to be robust enough to handle real test cases too.
   Having said that, images that are deemed to be 'too bad' by a human expert to be used in the current methodology (counting by hand) are excluded from the dataset (i.e. these images would not be used in the current research process either, so their exclusion from this project will not introduce any bias).
   Some examples of the diversity are shown below:

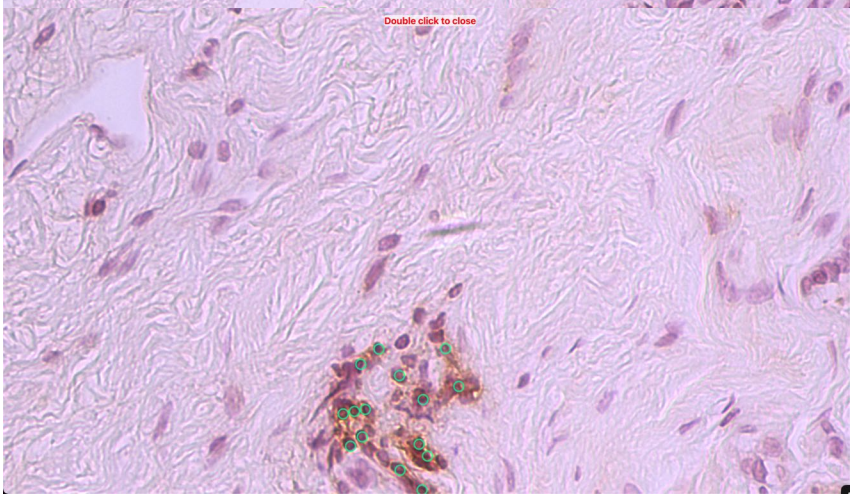



Double click to close


Double click to close


Double click to close




Double click to close

3. **Machine learning model**
   A DL model will suit this project the best. The two specific types of models in mind:
   - CNNs (specifically U-nets)
   - ViT

   Work on training the actual model has not begun yet (since they are upcoming in the course). The focus for the moment is to the data preprocessing.
4. **Preliminary Results**
   Currently the annotation step of the project is ongoing. Once enough data is annotated, then work on training and testing the model can begin. So prelim results are currently pending that step.

5. **Next Steps**
   Next step is to finalise enough annotations for models in order to be able to try basic models on them. Currently there are some CNN models which are pre-trained for histology classification, however, they are typically trained for detection of nuclei on a slide with a lot of homogeneity in the slides (both of these conditions are not necessarily true in our case: there is large heterogeneity in the nature of the slides and as well some nuclei may be completely absent from a slide yet if some staining occurs, it will need to be counted).