

# CIS 545 Project Proposal

Eric Oh and Mo Huang

April 8, 2018

For the final project, we propose to develop an algorithm to automatically detect nuclei of cells. Disease pathologists often examine tissue slides to diagnose diseases. Location and inspection of the nuclei in these tissues is a particularly important task as characteristics such as size, shape, and chromatin pattern can be important factors to distinguish different types of cells and understand disease progression. However, for large scale studies, manual inspection can be infeasible as they can involve thousands of slides. Therefore, automatic detection of nuclei can provide a tool for researchers to obtain insights quickly.

The dataset comes from Kaggle and contains 670 nuclei microscopy images. Associated with each image, we have the actual image file and the segmented masks of each nucleus. We anticipate a good amount of working with the data to be able to run algorithms on it. The images were obtained under differing magnification and imaging modalities, thus it might be necessary to standardize them across all images before analysis. In particular, to deal with different color images, it might be necessary to make all of the images greyscale. Additionally, we want to obtain separate masks for each nuclei; however, some cells are combined into a single mask making it difficult to segment the nuclei. There are standard techniques such as “mask erosion” that can separate out differences between the cells that we plan to apply. Lastly, we may apply some data augmentation procedures that have become standard for neural networks.

Overall, the objective of this project is to get familiar with classical image techniques and some advanced machine learning algorithms. Neural networks have achieved state of the art results on computer vision problems such as these, and as such, we will ultimately implement a CNN and possibly more recent neural network advances (such as the U-Net). We will also compare this to classical image techniques which do not take advantage of the training data.