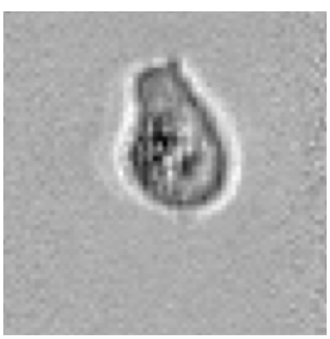
University of Tokyo Research Internship Program

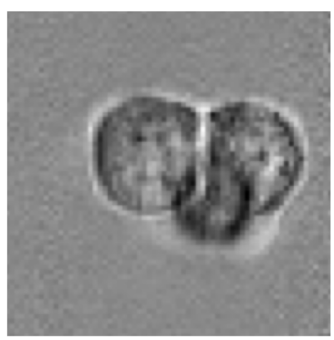
Research Report, Joseph Schull

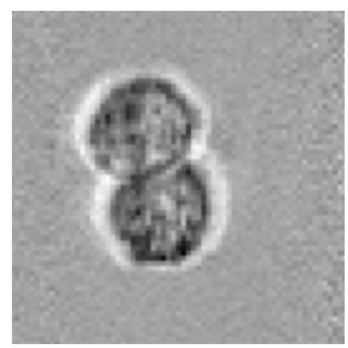
This past summer, I had the opportunity to work in the Goda Laboratory at the University of Tokyo. The Goda Lab works to develop tools for molecular imaging and spectroscopy; it is an interdisciplinary lab that integrates photonics, nanotechnology, microfluidics, and data science. I was working on the *Intelligent Platelet Morphometry* project, under the supervision of Yunjie Deng. It was an amazing experience, during which I was able to develop my data science and image analysis skills within a welcoming and supportive lab environment. This research report aims to provide an in depth overview of the work I conducted this summer, and the results of this research work.

Platelets are vital anucleate cells found in the bloodstream; their primary function is to prevent bleeding by instigating hemostatic responses that culminate in the formation of blood clots. Beyond their fundamental role in hemostasis, platelets play a role in various pathological processes, including atherosclerosis, inflammation, coronary artery disease, diabetes mellitus, tumor growth, cancer metastasis, COVID-19, and thrombosis. Thrombosis specifically is the leading cause of death and disability worldwide; in order to better understand and treat this condition, methods of imaging the formation of blood clots need to be improved. Clinical image-based testing methods, such as Ultrasound, Computed Tomography and Magnetic Resonance Imaging all lack sufficient spatial resolution, and in vitro methods of modeling and imaging thrombosis are limited in lacking disease relevance, or being too slow and labor-intensive. In order to tackle this problem, the Goda Lab partnered with clinicians at the University of Tokyo Hospital to devise an innovative tool that harnesses high-throughput optical imaging and deep learning to elevate platelet morphometry beyond conventional approaches. The goal of my research was to improve the image classification method the lab had in place for differentiating between images of single platelets and images of platelet aggregates.

Example Images:

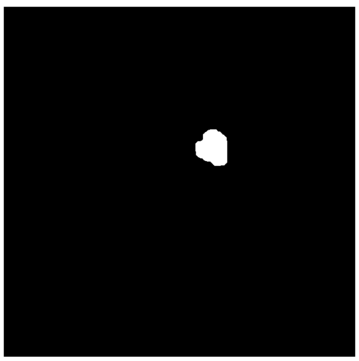






The old MATLAB script that was used to analyze these images used basic image processing techniques in order to create binary masks for images, and then single cell/cell aggregate classification would be made based upon the area and circularity of these binary masks.

However, the system was extremely inaccurate. In one 100 image sample, 75% of aggregate images were misclassified as being images of single cells; the aggregate misclassification rate varied between 50 and 75%. In order to understand the inaccuracy of this system, I first developed images of the binary masks that the old model was creating. Although certain masks looked accurate, many were clearly inaccurate.



Masks like these clearly capture no real aspects of cell shape, and instead are unintentionally capturing cell segments rather than the outlines of cells/cell aggregates. Given this inaccuracy, I decided that an entirely new method of image processing and classification had to be created. After a significant amount of trial and error, I decided upon the following method of image processing.

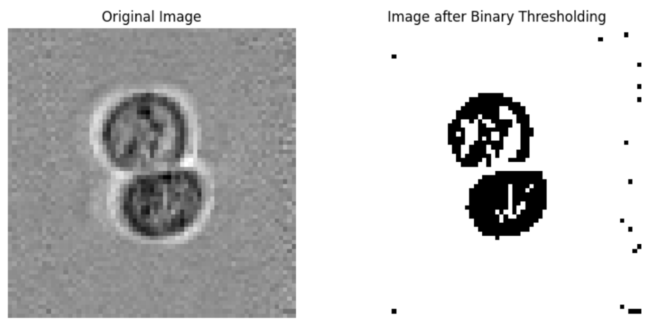
a) Binary Thresholding; this is an image processing technique that converts a grayscale image into a binary image by setting a threshold value. Pixels with intensity values below the threshold are assigned to black, while those above are assigned to white. In the context of cell images, binary thresholding helps to remove background noise whilst maintaining important image features, such as the silhouettes of cells.

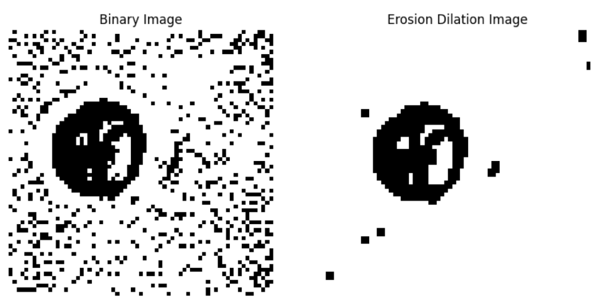
b) Morphological closing; this processes images by using dilation followed by erosion to close gaps and connect nearby regions in a binary image. It is particularly useful for smoothing and enhancing objects in binary images. Closing is achieved by first dilating the image, which expands the boundaries of white regions, and then eroding the dilated image, which shrinks those boundaries.

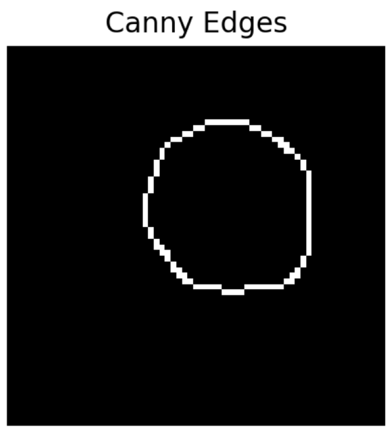
c) Distance transformation: an image processing technique that assigns each pixel in a binary image a value corresponding to its distance from the nearest background pixel (usually a black pixel). It quantifies the proximity of each object pixel to the background and assigns each pixel a new intensity based on this proximity. Various distance metrics, such as Euclidean distance or Manhattan distance, are used to produce a grayscale image, known as a distance map, where brighter values represent closer distances and darker values represent farther distances.

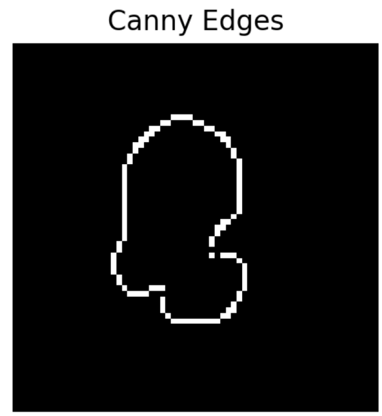
d) Canny edge detection: a technique which identifies edges and boundaries within an image by detecting rapid changes in pixel intensity. It involves smoothing the image, finding gradient magnitudes and orientations, applying non-maximum suppression to thin the edges, and utilizing hysteresis to trace and connect edges.

Following this sequence of steps, I applied a technique called perimeter contouring (which helps to identify outer contours of shapes in an image) to produce the new binary masks of the images. The following photos represent this sequence of steps.

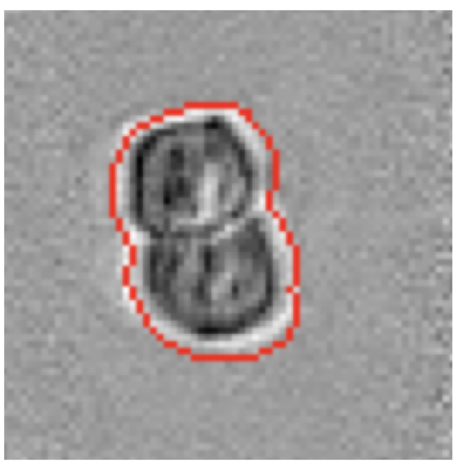


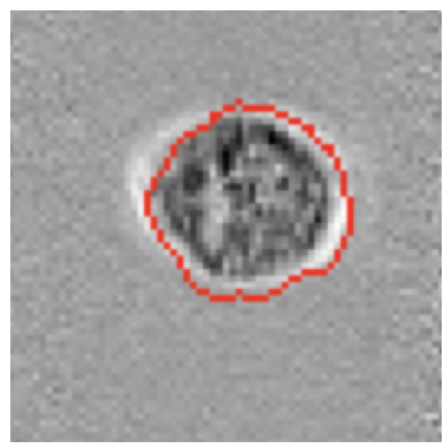






The following image displays the final contour upon original cell images.





In order to classify cells, I experimented with a variety of different distinguishing thresholds, and ultimately decided upon a decision tree for making classifications. My model ultimately had a 12% aggregate misclassification rate (compared to the 50-75%

misclassification rate of the old model), and an 88.2% overall model accuracy. I was satisfied with the improvement in accuracy that my model achieved.

I would like to thank the organizers of the UTRIP program, and everyone in the Goda Lab for contributing to such an enjoyable, fulfilling research experience!