IMPACT OF MORPHOLOGICAL HETEROGENEITY ON ADHESION DYNAMICS OF RED BLOOD CELLS IN PATIENTS WITH SICKLE CELL DISEASE USING MACHINE LEARNING

September 1, 2019

Gundeep Singh
Major: Physics
Mentor: Prof. Michael Hinczewski, Department of Physics

0.1 Research Report

Red blood cells of sickle cell patients have a relatively stickier cell surface as compared to healthy patients. This enhanced stickiness is one of the main causes of crisis events in sickle cell patients. In this study, we have developed an Artificial Intelligence (AI) based computational tool as part of our overall goal of developing a quantitative diagnostic tool for measuring the severity of sickle cell disease by measuring the stickiness of sickle cells. Our AI algorithm provides a high-throughput functionality to identify and classify sickle cells based on cell structure.

0.1.1 Abstract

The morphological heterogeneity found in the red blood cells (RBCs) of Sickle Cell patients has been shown to be correlated to the biophysical property of dynamic adhesion of sickle cells to endothelial proteins. In this work, we have developed an Artifical Intelligence (AI) algorithm that predicts the deformabilty charateristic of individual sickle cell using morphological variation among the structural features of the cells. We have employed two deep learning-based networks that generate pixel-wise and object-wise segmentation and classification respectively. Our final result demonstrate that the combined capability of the two automated networks successfully matches manual annotation and quantification, while cutting down the required time from several hours to around 13 minutes.

0.1.2 Background

Sickle Cell Disease (SCD) is an inherited blood disorder affecting more than 14 million genetically predisposed individuals worldwide. The affected demographic is mostly endemic to or draws on ancestral lineage from parts of India and Africa. A key hindrance to current diagnoses and treatment of SCD is lack of access to economically and operationally light point of care (POC) screening and monitoring tools. Further, patients of SCD show abnormally high levels of red blood cell (RBC) adhesion to the membrane proteins expressed in the endothelial vessels. This is one of the main causes that leads to painful vaso-occlusive crisis in SCD patients. Comprehensively tackling SCD hence hinges

on better understanding of sickle cell RBC adhesive dynamics.

A key factor that governs the dynamics of sickle cell adhesion to endothelial proteins is the variant morphologies of individual RBCs in blood flow. Dr. Umut Gurkan's lab[1] studies the effect of dynamic deformability of RBCs on cellular adhesion levels as part of their larger goal of developing an alternative SCD diagnostic mechanism and technology. They have found a significant correlation between the two phenomena, suggesting an "interplay between dynamic deformability and increased adhesion of RBCs during vaso-occlusive events". [1]

In my previous research, I have studied the biophysical modelling of abnormal cell adhesion in patients with Sickle Cell Disease (SCD). However, the main focus of my previous projects was to explore the variation in cell adhesion strengths at an inter-patient scale. This coarser approach at the inter-patient scale fails to capture the intra-patient heterogeneity observed in the characteristics of red blood cells. A single patient's blood sample does not have a homogenous sickle cell population that demonstrate equal levels of adhesion to endothelial proteins. On the contrary, there exists an intrinsic stochasticity of mechanical, chemical, and biological pressures that determines the fate of each individual RBC's deformability modulus and expression levels of receptor proteins. [2] Thus, an assumption that adhesion dynamics of different patients can be determined and compared solely on the basis of an overall effective adhesion characteristic gives an incomplete description of abnormal adhesion dynamics in sickle cell patients and fails to account for finer cell-cell differences due to heterogeneity in cellular shape, size, rigidity, and sample-to-sample variations.

In this project we studied the impact of morphological variation on cellular adhesion dynamics of sickle red blood cells. The research project consisted of three main steps - experimental design and run, image processing and analysis, and theoretical modelling and experimental validation. A significant portion of my work during the summer of 2019 was towards the second step of the research, i.e. image processing and analysis, where we developed 2 convoluted neural networks (CNN) to identify and classify sickle cell images respectively. Convoluted Neural Network or CNN is one of the many deep learning algorithm

[1]CASE-Biomanufacturing and Microfabrication Laboratory, Department of Mechanical Engineering, CWRU

that has transformed the advances in the fields of medical image processing, image segmentation and computer-aided diagnosis.

Automated analysis of biomedical images using machine learning provides a high-throughput platform that accurately provides biomedical quantification in a relatively small time period, thus, cutting down the need for subjective and time-consuming interpretation of biomedical data by expert biological researchers. Our experimental data consisted of whole channel bright field images of adhered sickle red blood cells (sRBCs) which hold a wealth of data that are needed to be comprehensively and reliably quantified. Single images can contain several thousands of adhered sRBCs. The quantification of these images can each take take up to several hours – not feasible for a targeted POC patient monitoring platform.

The technique of machine learning has already been applied to classify the different morphologies observed in sickle RBCs. [6] However, the focus of the classification in the algorithm published by Xu et al was to account for structural distinctions and not dynamic rigidity of cells in flow and thus can not be incorporated in real-time analyses to study how each category affects cellular adhesion dynamics. Thus, our CNN model for classification was based upon morphological differences that are indication of biophysical adhesion properties of the cells and were not based on characteristics related to cell growth and cycle. In the next section of Methodology, after describing the experimental setup, I will elaborate on how the two CNN algorithms were constructed for effectively identifying and classifying dynamic sickle cell mopohological-based adhesion.

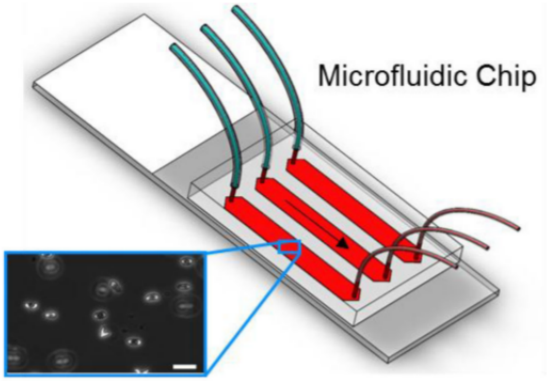


Figure 1: Experimental design showing microfluidic chip with patient blood flowing through one of the 3 channels of the chip. A snapshot of the sample frame is shown that is obtained and subsequently analyzed using AI-based image processing algorithm we have designed. Scale bar indicates 20 µm of length. [3]

0.1.3 Methodology

Our experimental setup consisted of an SCD monitoring platform [see Fig. 1] which ran clinical whole blood samples through protein functionalized microchannel in experiments designed to mimic conditions in microvasculature. As shown in the figure X., clinical whole blood samples of patients at University Hospitals, Cleveland, Ohio were flowed through microchannels and real-time videos of different patients were obtained that captured the dynamics of protein receptor-ligand interactions at a single-cell scale. The video frames obtained from these experiments were then cropped into smaller single-cell templates and were used to train the 2 convoluted neural networks.

The first CNN-based deep learning architecture is a pixel-wise segmentation network (as shown in Fig 2). The network looks for patterns in parameters based on shape, size, cellular morphology, and sample-to-sample variation to comprehensively segment sickle cell images. We utilized an encoder-decoder type semantic classifier network (see Fig 2) that relied on training images where each pixel had been annotated with one of the one of the five object categories our samples show up. (See [4] for more information on Semantic Segmentation) Training images were 32x32 pixel templates extracted from whole channel images and the complete set consisted of a total of 2867 cells (see Table 1 for breakdown). We started by attempting a binary classification of the adhered sickle cells into two morphologically heterogenous populations: deformable and non deformable cells (Fig 3). A separate out of focus category was created for images of non adhered cells outside the focal plane, a large number of which show a distinct halo- like diffraction pattern around the blurred cell image. Since the experiments are on whole blood samples, a large amount of background gunk, platelets, white blood cells etc show up. For the purposes of our sRBC monitoring platform, we want to be able to distinguish between these and our adhered sRBCs.

Label Category	No. of Object Templates
Background clutter	300
Deformable cells	1282
Non-deformable cells	917
Out of focus cells	368
Total	2867

Table 1: Training set for pixel-wise CNN network 1

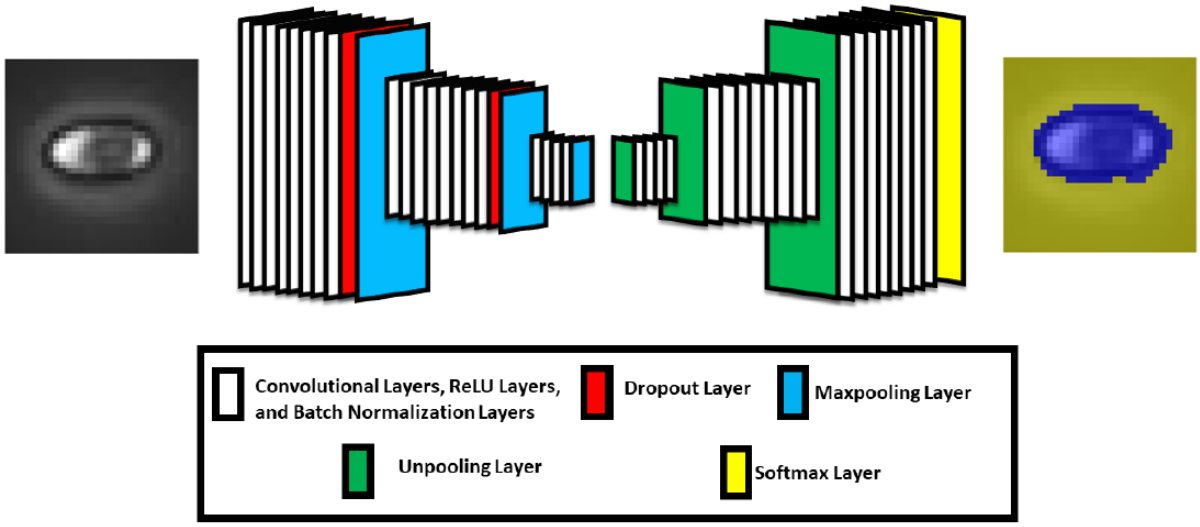


Figure 2: A schematic of pixel-wise CNN architecture. The network performs convolutions with trainable kernels to map out a feature space. For the ground truth preprocessing, each pixel is annotated based on one of the possible four classes: background, deformable, out of focus cells, and non-deformable.

Once we have a class for each of the pixels of the image, we then wrote another algorithm to move from pixel-wise to object-wise identification and classification of cells. We first identified all the objects where the majority of connected pixels belonged to either deformable or non-deformable classes and extracted those objects as 32X32 pixel template boxes. The second CNN network was then applied, which is a derived form of the popular image recognition architecture VGG16 [5]. The network was pretrained on object classification using 3363 templates (with 1121 templates for each of three classes - deformable, non-deformable and out-of-focus cells) and consisted of a deep network of convoluted layers followed by softmax and cross-entropy layers, required to assign labels to each of the single-cell templates. Finally, the quantification data of each of the classes were then shown for each of the whole-channel images.

0.1.4 Results

The performance of the two CNN networks - pixel-wise and object-wise identification and classification is shown in Fig 3 and 4. In Fig. 3, the network's performance is shown in the form of a confusion matrix, which shows the detailed breakdown of true and false positives and negatives (for more details,

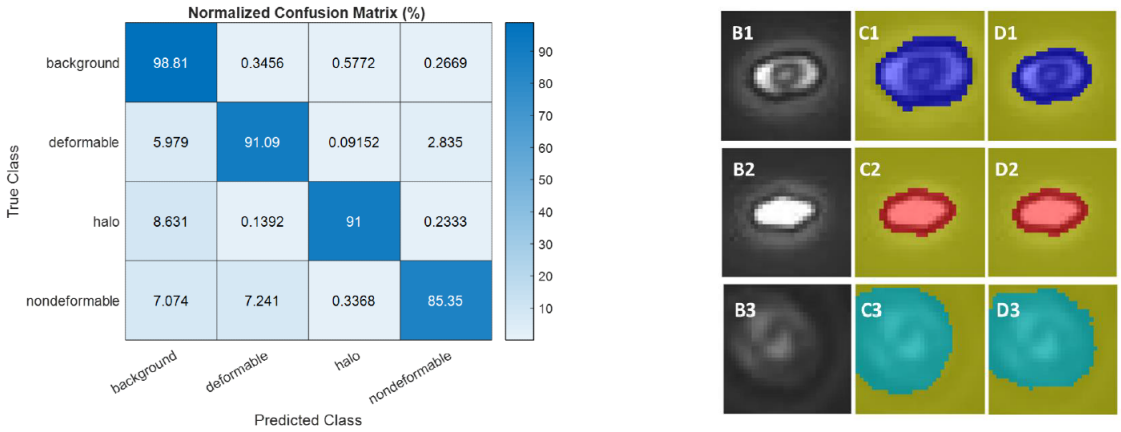
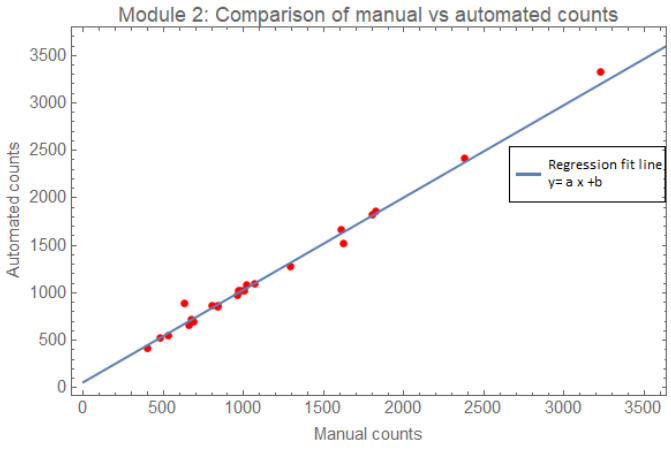


Figure 3: Fig.2a: (Left) The confusion matrix indicating network's overall classification accuracy on the test data. The diagonal elements indicate percentage of true positive detections. For eg. when presented with a deformable cell, it correctly classifies it as deformable 91.09% of the time. The off-diagonal elements show mislabeling percentages- an indicator of the network's 'confusion'. For eg. from the bottom row second column, we can see that when shown a test image of a deformable cell, 7.2% of the time the network incorrectly identifies it as a deformable cell. (Right) (B1-3) Three representative 32x32 pixel templates depicting a deformable, nondeformable, and out of focus cell in descending order (C1-3) The manual annotated ground truth images of background (yellow pixels), deformable (purple pixels), non-deformable (red pixels), and out of focus (blue pixels) (D1-3) The CNN segmented version of the templates.

see caption of Fig 3). The average accuracy of the pixel-wise classification is computed to be 92±6 %, i.e. on average the network correctly identified pixels (i.e. ground-truth annotation) 92±6 % of times. After passing the resulting identified templates into the second object-wise classification network, we estimated the overall accuracy by comparing the total count performance of manual and AI algorithm. AS it can be seen from Fig. 4, our network was able to replicate the job of a biomedical researcher. For each of the image, since each image consist thousands of cells, it can take up to several hours for researchers to quantify image data. However, the computational time of running both networks together is around 13 min for each image. Thus, we were successfully able to produce a high-throughput deep learning algorithm that identifies and classifies sickle-cell based on morphological variation, which is correlated to biophysical variation in cellular adhesion strength. However, we are yet to study the quantitative effect of how the heterogeneity of adhesion strength correlates with clinical parameters and patients' diagnostics. This constitutes our next step of the project as I describe in the next section.



	Estimate	Standard Error	t-Statistic	P-Value
a	0.973222	0.0161124	60.4022	7.61882×10^{-27}
b	60.42	28.8891	2.09145	0.0477265
AdjustedRSquared	0.997508			

Figure 4: Classifier performance matched against manual counts for total number of adhered cells in a set of whole channel images. A value of 1 for the regression parameter a would indicate exact agreement. At a=0.973 we are doing very well on the total cell counts.

0.1.5 Future Consideration

Analyzing the impact of morphological differences on cellular adhesion dynamics holds great potential for clinical diagnostic systems but has so far remained un-attempted mainly due to the sheer volume of labor involved in the manual analyses. In this study, we have proposed a machine learning based convoluted neural network approach to identify and categorize sickle RBCs on the basis of their morphological differences. In the next phase of our study, using the computational tool we have developed in this study we will study how the morphological differences correlate with the stickiness of the sickle red blood cells and whether we can use that information for devising a finer-scale severity measurement of sickle cell disease patients.

In the school year 2019-20, I will be continuing this work as part of my senior capstone project. The main focus would be on modelling the system using theoretical biophysics modelling and using the model fits to estimate the adhesion parameter of each of the categories. We hypothesize that the deformable cells would have less stickiness than the nondeformable cells, however, we would need the model parameter values to describe the comparison quantitatively. I am also planning to combine my senior project with another

Machine Learning based project, where I will be using AI to attempt to reduce the data required for creating Magnetic Resonance Imaging, the success of which would mean cutting down the time and cost of running of MRI machines.

0.2 Academic Publication and Reflection on Experience

0.2.1 Academic Publication

We are in the process of writing a paper for this study. The paper will primarily be a methodology paper where we would explain the CNN network architecture and performances with different test cases. We would also be highlighting our application algorithm in the paper as we are planning to publish the application publicly which can be used by biological researchers working at cellular scale systems to segment and classify 2D cell images.

This work will be utilized by researchers in Gurkan group to carry out automated quantification of their sickle cell experimental data. Thus, this work would also be credited in all of their future projects that employ the computational tool we have developed in this study.

0.2.2 Reflection on Experience

My work this summer was strongly aligned with my proposal. However, there is one big aspect that changed during the initial stages of my research. In my proposal, I wrote that my research would look at deformability as a dynamic property, which would thereby transcend the current literature that puts the cells into a binary category - deformable or non-deformable. However, developing an AI algorithm that computes a scalable deformability parameter just by looking at 2D cellular templates meant integrating unsupervised and supervised machine learning methods. This integration proved to be a lot more challenging than I initially anticipated, as it is still a recent advancement in computer vision and it's implementation is very difficult. Thus, in this work, we worked in developing a binary classification of cellular deformability as the first step before tackling the dynamic deformability problem.

I started my summer research project with very minimal background knowledge and experience of machine learning in both theory and implementation. Thus my first step was to teach myself the fundamental math of machine learning for which I enrolled in a free Coursera course on the subject taught by its co-creator Andrew Ng. I simultaneously wrote simple image processing programs on Matlab using different Git resources posted by various AI bloggers. After trying few fundamental examples, I wrote a network architecture and trained it with the data specific to our experiment. However, since we do not yet completely understand the arithmetic working of deep learning, it involves a lot of fine tuning and hit & trial tweaking, which at times felt completely contrasting to my physics background and scholarship. But once, we found the hyper-parameters that worked for our data, the results we got were astonishing and and unparalleled to any of our previous results we had derived using thresholding-based image processing methods.

Thus, in a very short period of time, I started to realize the potential of AI and the transformations it can bring, or has already brought, to both academic research ad medicinal applications. I had previously learned about the potential of machine learning in 'learning' the patterns in data without much user-input and knowledge, but experiencing this potential in my own research was a completely different thing. I am hopeful that learning and implementing deep learning in a research setting will prove to be a significant learning curve in my future career. Having the knowledge of implementing the state-of-the-art AI based technology in a biomedical setting would also make me a potential candidate for graduate school applications.

Another enlightening aspect of my research that was the interdisciplinary and collaborative nature of my project. This work would not have been possible without the collaboration of physicists, engineers and medical doctors. My involvement in this research project has made me realize how gratifying, social, and exciting a career in research could be. And thus my experience has further inspired me to pursue a career in academic research. It has also given me the tools to navigate scientific conversations in a collaborative environment, which will prove to be really helpful as I am planning to get a PhD in Biophysics, which is innately a collaborative and interdisciplinary field.

Finally, I would like to express my gratitude to the office of SOURCE for supporting my work in a number of different ways. The financial stipend

undoubtedly was a huge inspiration that further motivated to have a fruitful and productive summer research experience. The Lunch and Learn series also acted as constant source of inspiration, that not only allowed me to explore various different research projects, but also helped me to make connections with researchers outside of my academic discipline. The Tuesday seminars were also a great opportunity for me to understand the operations of many important outside-of-lab resources on campus. I particularly enjoyed the seminar on Intellectual Property as it is one of my interests to understand the working and nuances of patenting computational applications. Seminars on graduate and fellowship programs were also very relevant to my future career as I apply to graduate schools this year. Overall, the SOURCE program was a huge help in making my summer research into a multifaceted productive experience and I would highly recommend future undergraduate students to consider applying. One advice that I would give to them is not shying away from applying early in their research career and having an open-mind towards exploring different facets of their research.

Bibliography

[1] Alapan Y et al. *Dynamic deformability of sickle red blood cells in microphysiological flow*. Technology vol. 4,2 (2016): 71-79.

[2] Hierso R et al. *Variation of Red Blood Cell Rheology and Adhesiveness during Vaso-Occlusive Crises in Sickle Cell Patients*. Blood 128.22 (2016): 1311.

[3] Kim M et al. *Hypoxia-enhanced adhesion of red blood cells in microscale flow*. Microcirculation (New York, N.Y. : 1994) vol. 24,5 (2017): 10.1111/micc.12374.

[4] Long J et al. *Fully convolutional networks for semantic segmentation*. CoRR, abs/1411.4038, (2014).

[5] Simonyan K et al. *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556 (2014).

[6] Xu M et al. *A deep convolutional neural network for classification of red blood cells in sickle cell anemia*. PLOS Computational Biology 13:10 (2017).