

Rayhaan Rasheed, William Noone, Samuel Aboagye
DATS 6203 – Machine Learning II
Instructor: Amir Jafari
Spring 2019

Final Project Proposal

Our group decided to tackle the issue of detecting malaria in microscopic examinations of blood smears. Malaria is a deadly disease that is caused by parasites transmitted through the bite of the female Anopheles mosquito. According to the World Health Organization's most recent report on malaria, there were over 215 million cases around the world in 2017. In order to analyze blood smears, our group will be looking at images of stained blood slides from the U.S. National Library of Medicine (NLM). There are 27,000 images annotated as either normal or parasitemic. We will be using a custom convolutional neural network (CNN) built on the Caffe framework. Though, the model is being built from scratch, we are getting inspiration from the NLM's customized CNN built by Krishnan et. al in 2017. Their 13-layer CNN outperformed a pre-trained AlexNet and pre-trained VGGNet with accuracy over 95%. Another resource is the CNN built in 2016 by Liang et. al from the University of York. They created a 17-layer CNN with a Sigmoid layer attached before the Softmax output. This resulted in an amazing accuracy of 97.7%. Both architectures used a Stochastic Gradient Descent optimizer along with a Binary Cross-entropy loss function. Since the project is a binary classification problem, we will evaluate the model's performance using a Confusion Matrix, ROC Curve, and AUC. Our goal is to have all the labels encoded and paired with their respective images by Sunday, April 7th. All of the image resizing and processing will be completed by Thursday, April 11th, and the CNN architecture on Caffe will be completed by Wednesday, April 17th. Other deliverables, like the reports, will be completed by Monday, April 22nd.