Rayhaan Rasheed
DATS 6203
Machine Learning II

<u>Individual Report</u>

Introduction

The objective of this project was to classify malaria in images of stained red blood cells. Malaria is a deadly parasitic disease that infects millions of people every year. Most of these victims come from underdeveloped countries like India and those in Africa. Previous publications have ran CNNs on the RGB images using Keras, but there have not been any publications that use different color spaces or deep learning frameworks. Our group decided to create 3 different CNNs to classify RGB and Grayscale blood cell images. Also, we decided to build the CNNs using Caffe as the framework.

Individual Work

It was my decision to use this dataset due to my background in Biomedical Engineering, specifically in Medical Imaging and Image Analysis. It was my job to do all the background research on what other groups have done in regards to malarial cell classification using Neural Networks. WIth a very strong understanding of this domain, I led the group by developing the design of the project and all the image analysis code.

The image analysis algorithm was inspired by the feature enhancement and segmentation algorithms I have worked with in the past. During my time researching, I could not find any publication that looked at converting the image from RGB (3 channels) to Grayscale (1 channel). The goal of my code was to establish a pipeline that would take in the images from my GCP and run a variety of transformations was to eliminate any background noise as well as play with other color spaces.

Firstly, I took in all the images and resized them to be 100x100 that way there is consistency in the data, afterwords, I used the ndimage library in SciPy to call upon the *rgb2grey* function. This created another set of images that were the grayscale equivalent. Now that grayscale is established, a gaussian filter is run to eliminate background noise and smooth out the image. Finally, a threshold mask is built to only keep pixels with an intensity of less than or equal to 0.45 out of 1 (See Figure 1). The

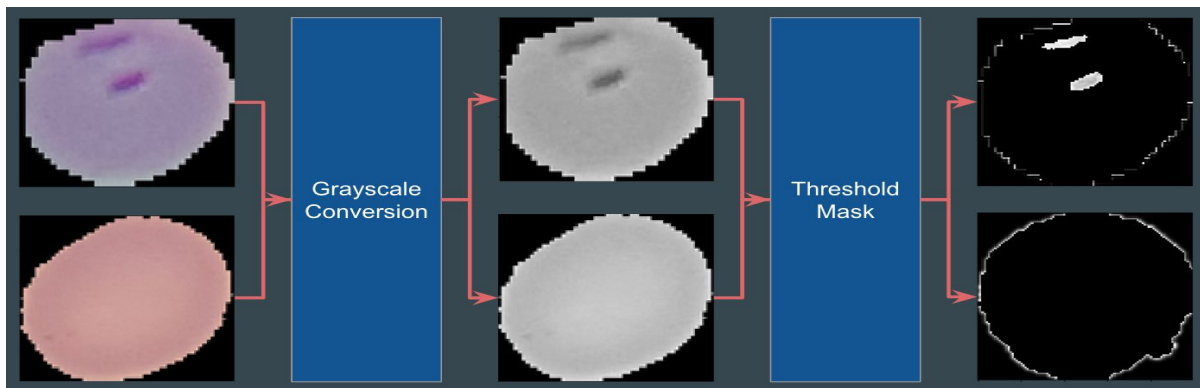resultant images and the unprocessed images were sent to the LMDB file to be run on the CNN.

Figure 1: Image Analysis Pipeline

There were two sources that I used to design the CNN that was built in Caffe by my partner. The first inspiration was from Liang et al 2016 who created a CNN of repeating Conv-ReLU-Pool layers then ended with a couple fully connected layers and a Sigmoid. The second inspiration was Krishnan et al. 2017 who worked off of the previous CNN model but left out the sigmoid. The figures below reflects my research on these two groups because it incorporates certain characteristics but modifies other to create our own CNN.

Model 1 (Figure 2) consists of two Conv-ReLU-Pool layers and two fully connected layers at the end. This was our initial attempt at a customized CNN. I also wanted to compare this model to other configurations, so the group added more layers and produced Model 2 (Figure 3).
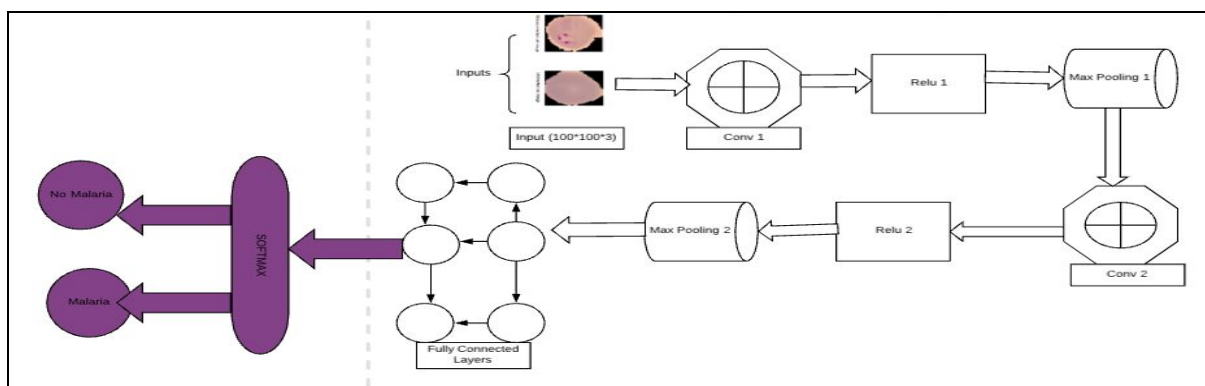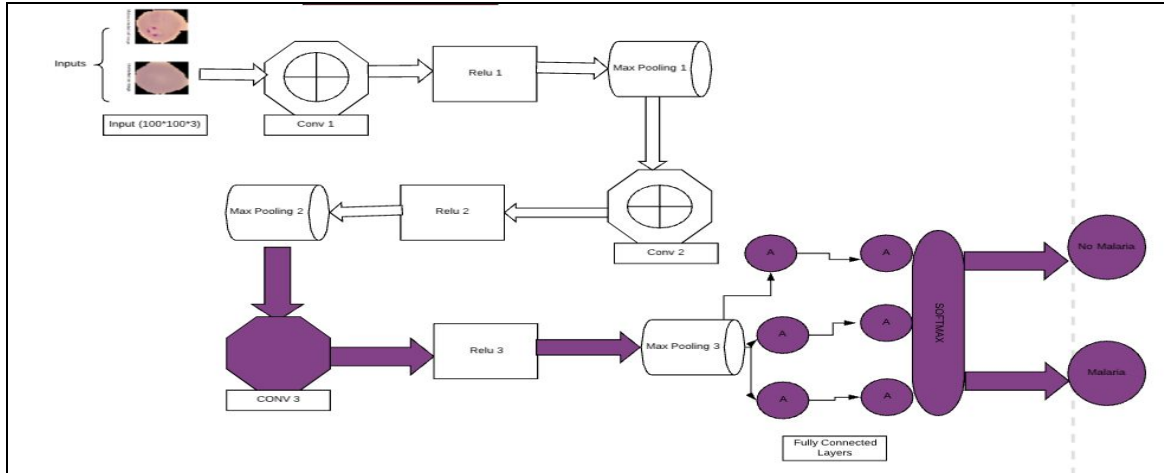

Figure 2: Two-Conv CNN

Flgure 3: Three-Conv CNN

Results

Our initial hypothesis was that the processed images would result in a better accuracy and faster computation time. After testing both unprocessed and processed images in all the various model configurations, the unprocessed images proved better. Unprocessed images on the first model produced an accuracy of 91.7%, and the processed images gave us an accuracy of 87.1%. Next, running the images on the second model gave an accuracy of 92% for the unprocessed images and an accuracy of 81.5% for the processed images.

Summary and Conclusions

Surprisingly, image processing techniques had a negative effect upon CNN predictive performance. This could be improved upon by modifying the parameters of the image filtering and masking steps. It is a time intensive process to convert image sets from JPG to LMDB format suitable for the Caffe framework. Additional challenges were the prohibitive with respect to time and available resources. This study underscores the importance of image processing on image classification tasks. The team recognizes that the study is not exhaustive as countless specifications (dropout, learning rate, optimizers, etc) could be altered in the pursuit of performance tuning. More specialized image analysis approaches are needed to take into account the specific patterns and color spaces in these blood cell images. Proper classification of malarial cells is vital for classification of specific strains. Using the proposed classification technique as an a priori for identifying various strains of malaria, unique treatment can be synthesized to help save lives.

Calculate the percentage from the internet

(20 - 10)/(20 + 84) * 100 = 9.6% of my code is used from the internet

References

[1] Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15 (June 2014), 1929–1958.

[2] Nasir AA, Mashor M, Mohamed Z. Segmentation based approach for detection of malaria parasites using moving k-means clustering. In: EMBS conference on biomedical engineering and sciences (IECBES). IEEE; 2012:653–8.

[3] Poostchi, M., Silamut, K., Maude, R. J., Jaeger, S., Thoma, G. Image analysis and machine learning for detecting malaria. Journal of Translational Research 2017:194-33

[4] Sampathila, N., Shet, N., Basu, A. Computational approach for diagnosis of malaria through classification of malaria parasite from microscopic image of blood smear. Journal of Biomedical Research 2018: 29-970

[5] LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature 521:436–444

[6] Rajaraman et al. (2018), Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ 6:e4568; DOI 10.7717/peerj.4568

[7] Krishnan S., Antani, S., Jaeger, S., Visualizing Deep Learning Activations for Improved Malaria Cell Classification.

[8] Liang et al. (2016), CNN-Based Image Analysis for Malaria Diagnosis. In: EMBS conference on biomedical engineering and sciences (IECBES). IEEE