**Malaria Detection – Milestone 1**

1. **Problem Definition**
   1. The context - Why is this problem important to solve?
      1. Malaria is a significant health burden with that is a vector-borne tropical diseases with 212 million cases in 2015 [1, 2]. A majority of deaths from malaria are account of a single *Plasmodium species P. falciparum*, that resulted in approximately 445,000 deaths in 2016 in the sub-Saharan desert, although the number of deaths has decreased by 29% since 2010 [1, 2]. An additional 4 or 5 *Plasmodium* species contributed to the worldwide malaria burden, even though 120 different species are capable of infecting mammals, birds and reptiles [1].
   2. The objectives - What is the intended goal?
      1. The goal is to use computer vision to detect the presence of malaria from the images of red blood cells that have been labeled for the presence of malaria. It is known that host’s red blood cells (RBC) infected with *P. falciparum* cluster in smaller blood vessels and avoid being cleared by the body by residing in the spleen [1]. Meanwhile, these isolated RBC damage the cell lining and obstruct the system of smaller vessels. Infected blood cells can bind to uninfected blood cells, which further worsens the damage.
      2. As a consequence to this damage to the vascular system, symptoms of severe malaria include acute lung injury, kidney injury, tubular necrosis and acidosis, and cerebral malaria. Among those who develop cerebral malaria, 10-20% of those who are treated die, whereas 50% of pregnant women with cerebral malaria die [1].
      3. Malaria has still not been eradicated in certain regions of the world, so it will be critical to identify whether computer vision will more successfully detect abnormalities in the RBCs.
   3. The key questions - What are the key questions that need to be answered?

Microscopic examination of blood films is currently used to screen for malaria, where the test screens first look for presence of parasitic infection, where as few as 10-50 parasites per microliter can be detected [2]. Species-level detection is important for treatment decisions, but it is also important to improve detection at an earlier stage, to the point of preceding clinical detection [2].

* 1. The problem formulation - What is it that we are trying to solve using data science? CDC’s telediagnostic service currently identifies 70% of cases last reported between 2005 and 2010 [2]. Due to challenges in detecting malaria, deep learning may improve detection of malaria.

1. **Data Exploration – Anaconda stopped working during this submission…attempted to run this analysis on AWS struggling to access unzipped files in S3. Need to uninstall and reinstall anaconda**
   * Data Description - What is the background of this data? What does it contain?
   * Observations & Insights - What are some key patterns in the data? What does it mean for the problem formulation? Are there any data treatments or pre-processing required?
2. **Proposed approach**
   1. Potential techniques - What different techniques should be explored?
      * Possible techniques to explore include performing a convolutional neural network to capture the spatial structure
      * Alternating the number of hidden layers, the type of activation function used, use of forward and back propagation, and utilization of batch normalization through ridge or lasso regression to reduce the potential for overfitting

* 1. Overall solution design - What is the potential solution design?
     + Convolutional neural network with several hidden layers and an activation function that permits non-linear fitting
     + Other techniques for regularization will include data augmentation, and performing drop-out
     + Transformations will improve performance
  2. Measures of success - What are the key measures of success to compare potential techniques?
     + We will utilize similar performance measures as utilized in the supervised learning training modules, including measures of the confusion matrix, including recall, f1, precision, and accuracy.

References

1. Barber, B.E., et al., *World Malaria Report: time to acknowledge Plasmodium knowlesi malaria.* Malaria journal, 2017. **16**(1): p. 1-3.

2. Mathison, B.A. and B.S. Pritt, *Update on Malaria Diagnostics and Test Utilization.* J Clin Microbiol, 2017. **55**(7): p. 2009-2017.