

# *Plasmodium vivax* thick blood image masks development

JORGE AUGUSTO BALSELLS ORELLANA      ERICK WILFREDO DÍAZ SABORIO

SOFIA MARIBEL SOBERANIS LOPEZ

CINDY CAROLINA ROCIO VILLALOBOS MORALES

HANS IVÓ ROCHE VILLAGRÁN

University of San Carlos de Guatemala

May 7, 2019

## I. ABSTRACT

Malaria is a disease transmitted by the bite of the female mosquito of the genus *Anopheles* infected by protozoa of the genus *Plasmodium*, which is obligate intracellular parasites. *P. vivax* is the sole cause of the disease in Guatemala and currently focuses on the departments of Escuintla, Alta Verapaz, Izabal and Suchitep  lquez. The signs and symptoms caused by *P. vivax* infections are fever and fatigue, this *Plasmodium* has the peculiarity of maintaining "dormant" forms (hypnozoites) in the liver, which gives the possibility of recurrence of the disease. In addition, it may present complications due to splenomegaly, rupture of the organ with concomitant internal bleeding (Pan American Health Organization, 2017).

In Guatemala, the techniques currently used for the diagnosis of *P. vivax* are the microscopic examination of the protozoan through two types of blood films: thick blood smear and peripheral smear (National Health Laboratory, 2017). However, this type of diagnosis has some disadvantages: trained and experienced microscopist personnel to identify and quantify the parasite in the thick drop (gold standard) and peripheral rub given that the parasite can be confused with platelets, precipitate, white blood cells or artifacts; high workload for staff depending on the area in which the disease is concentrated; high estimated time in observation of blood films, on average a trained person observes a thick blood film for 10 to 20 minutes, in endemic areas a health center can receive between 200 and 300 weekly blood films; Ocular fatigue, usually is the most influential factor in an incorrect diagnosis, this is due to long periods of time observing samples through the microscope.

This investigation objective, is to design a raw tick blood images, data processing pipeline that generates a clean dataset, that can be used for training models like convolutional neural networks (CNN) models. In order to train these models is required a good dataset, that have been quantitatively measured for an objective, in this case detect malaria. Our initial data exploration analysis on the images shows that, the images have different sizes and the image is not centered (**Supplementary Fig 1**). The code of this initial data exploration analysis can be found at [https://github.com/ErickDiaz/bioinformatic\\_thesis\\_project/blob/master/code/data-exploration/data\\_exploration.ipynb](https://github.com/ErickDiaz/bioinformatic_thesis_project/blob/master/code/data-exploration/data_exploration.ipynb).

## II. METHODS

This work pretends to find patterns in these biological data and detect the malaria parasite *P. vivax*, using the following techniques.

- Edge detection
- Sobel Filters
- Convolution Filters
- Histograms of image value distribution, and use the cumulative distribution function for finding the frequency of a given range of pixel intensity occurrence.

In order to quantify how good the filters are, we will use two cost functions that will measure how is our pipeline filtering the parasite. The cost function will measure the mask generated by the pipeline against manually generated masks by our team, using an open source tool called LabelImg. The goal of the data pipeline is to minimize the values of the cost functions.

- Mean Absolute error: Calculate the similarity, because we want to measure if there is a difference between two images after applying the masks.
- The intersection of the union(IOU): Measure the total number of shared pixels after applying the masks.

$$IOU = \frac{I_1 \cup I_2}{I_1 \cap I_2}$$

## III. RESOURCES

Resources		
Material	Finantial	Human
tick blood images	AWS, Instance c5.4xLarge: 0.68 USD/Hora	Jorge Balsells Orellana, Erick Diaz Saborio, Sofia Soberanis Lopez, Cindy Villalobos Morales, Hans IvÃş Roche VillagrÃan
Manual Bounding Boxes tooll : <a href="https://github.com/tzutalin/labelImg">https://github.com/tzutalin/labelImg</a>	\$ 0.00	
Python	\$ 0.00	
GitHub	\$ 0.00	

## IV. RESULTS AND EXPECTED IMPACT

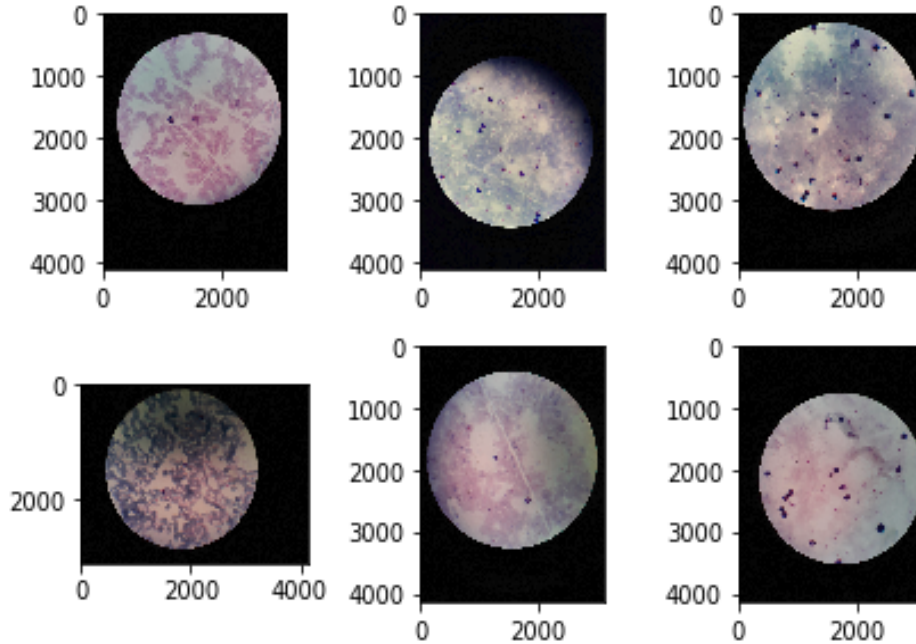
- Create a raw tick blood image processing pipeline that generates a database of treated images, classified by positive and negative samples. For each positive sample a set of masks that will indicate the location of the parasites. The purpose will be its use to experiment with different models of neural networks, in order to detect positive samples and the count of parasites in the sample.
- Contribute to the scientific advance of bioinformatics in Guatemala by sharing the dataset of malaria blood samples in Kaggle. This website is a community of data scientists, in which users can publish datasets, code of artificial intelligence models called kernels and discussion forums. Once a kernel is publish it allows users to fork this allows collaborative programming, also vote for a kernel and add comments.

## V. REFERENCES

- Clinically applicable deep learning for diagnosis and referral in retinal (Pearse A. Keane, 2018)
- Deep learning enabled multi-wavelength spatial coherence microscope for the classification of malaria-infected stages with limited labelled data size (Neeru Singla and Vishal Srivastava, 2019)
- Purwar, Y., Shah, S. Clarke, G. & Muehlenbachs A. (2011). Automated and unsupervised detection of malarial parasites in microscopic images. *Malaria Journal* 10 (364).
- Organizaci3n panamericana de la Salud. (2017). Marco para la eliminaci3n de la malaria. Washington: Organizaci3n Panamericana de la Salud.
- Laboratorio Nacional de Salud. (2017). Manual de normas y procedimientos de laboratorio para el diagn3stico de malaria. Guatemala: Ministerio de Salud P3blica y Asistencia Social, Direcci3n de Regulaci3n, Vigilancia y Control de la Salud, Laboratorio Nacional de Salud.
- Lopez, E. (2019). Ops/oms Guatemala - Guatemala Avanza En La Lucha Contra La Malaria. Recuperado de: [https://www.paho.org/gut/index.php?option=com\\_content&view=article&id=1220%3Aguatemala-avanza-en-la-lucha-contr-la-malaria&Itemid=441](https://www.paho.org/gut/index.php?option=com_content&view=article&id=1220%3Aguatemala-avanza-en-la-lucha-contr-la-malaria&Itemid=441)

## VI. SUPPLEMENTARY INFORMATION

### i. Supplementary Figures



**Figure 1:** Example of images in the raw dataset.