Effective Creation of Ground Truth Dataset for Malaria Diagnosis

Martha Shaka

The University of Dodoma, Tanzania
Goal

- Malaria is a life threatening disease transmitted by a bite of an infected female anopheles mosquito.

Who is at Risk?

- In 2017 nearly half of the world was at risk of malaria being the poorest and marginalized communities with the highest risk;

Death Burden

- 93% of the deaths caused by malaria was reported in Africa in 2017;
- 60% of the reported cases of death were children under 5 years;
- Malaria kills a child every 2 minutes!

Source: WHO malaria report 2018 showcasing malaria cases globally

The goal of this project is to reduce mortality rate related to malaria disease, particularly in marginalised communities.
Existing Strategies

- **Malaria is curable** and with a prompt diagnose and treatment can reduce death.

**Common diagnosis tools for malaria**

- **Microscope**
  - **Pros:** accurate.
  - **Cons:** takes time (15-30 mins), requires expert and labour in count of parasites.

- **RTD**
  - **Pros:** fast, portable, no need of expert, cheap.
  - **Cons:** can’t diagnose malaria at early stage, no quantitative analysis.

- Existing rapid diagnosis tests can’t identify the stage or count the number of parasites (quantitative diagnosis).
- But we can do so with object detection in Computer Vision.

**Why Quantitative Diagnosis matters?**

- Uncomplicated malaria patient receives oral antimalarial.
- Severe malaria patient receives parenteral antimalarial.
What has been done

- Several AI techniques have been used to solve challenges in the existing malaria diagnosis tools.
- In 2015, Sanchez Sanchez used Deep Convolutional Networks have been used for the detection of malaria parasites (S 2015).
- Mobile Malaria Lab (MOMALA)
DATA CHALLENGES

- one of the major drawbacks that hinder the development of AI-based applications for healthcare services in developing countries, such as Tanzania, is the lack of data for training, testing, and validation of such tools.
- In these countries, there is limited access to the available data from both government and non-governmental organizations.
- Moreover, even the little available data still lacks the necessary qualities in terms of pixels, labels that are required for the development of AI tools.
- To address this deficiency, we have collected and annotated 10,000 images of a stained blood smear to be used to develop different artificial intelligence tools for malaria diagnosis.
In the collection of this dataset, we first sought and were granted ethical clearance from the University of Dodoma and Benjamin Mkapa Hospital's research center.

We have collected 50 blood smear samples for patients confirmed with malaria and 50 samples for negative confirmed cases.

Each sample was stained by the lab technologist and 100 images were taken using iPhone 6S attached to a microscope. This led to having a total of 5000 images for the positive confirmed patients and 5000 imaged for the negative confirmed patient.
 IMAGE CAPTURING AND ANNOTATION

- Image captured using iphone 6s mounted on the olympus microscope
- Annotated with a green bounding box around the plasmodium
Three lab technologists have labelled and verified the dataset.

Challenge; in some occasion they had a different labelling and we had to find an average of the labelled data.
Create an annotation tool; Start using our data and train deep learning models in the development of the open-source annotation tool.
• Work with the AI4D team, we are looking for the best approach to follow when releasing our open-source dataset in the medical field.
• our overall aim is to develop a final product of our mobile application that will assist lab technologist in Tanzania and beyond in the onerous work of diagnosis malaria.
Reference