Detecting Malaria using Deep Learning

Hassan Mahmoud Abdelkhaleq Sarwat

July - 2019

Domain Background

Malaria is a mosquito-borne blood disease caused by the Plasmodium parasites. It remains a major burden on global health, affecting 200 million people worldwide and killing 400,000 annually. One of the obstacles facing mortality reduction is malaria diagnosis.

An individual diagnosis is an arduous task involving a lot of time and requires the availability of an expert. In this project we try to automate the process of malaria diagnosis using deep learning, shortening the time required and completely removing the necessity of having an expert to detect the virus.

Problem Statement

This is a binary classification task that we will solve using supervised learning, to be exact we will use convolutional neural networks, as they have proven so far to be among the best at image classification. Given an image of a cell, we'd like to identify whether it is parasitized or uninfected.

Dataset and Inputs

Giemsa-stained thin blood smear slides from 150 P. falciparum-infected and 50 healthy patients were collected and photographed at Chittagong Medical College Hospital, Bangladesh. The smartphone's built-in camera acquired images of slides for each microscopic field of view. The images were manually annotated by an

expert slide reader at the Mahidol-Oxford Tropical Medicine Research Unit in Bangkok, Thailand

We download the data from the U.S. National library of medicine. It contains a total of 27,558 images with equal instances of parasitized and uninfected cells which we will use to train our deep learning model.

Solution Statement

We will use a convolutional neural network trained on our dataset so that we can identify whether a cell is parasitized or uninfected.

Benchmark Model

The benchmark model we will be using is the Resnet-50, it can be seen that of all the pretrained models used in [3] the Resnet-50 had the highest accuracy therefore we will set it as our benchmark

Evaluation Metrics

This project will have 2 evaluation metrics, first of all how accurate our model is and recall. Since this is a medical task a false positive is not really as dangerous as a false negative.

Project Design

First part of our project will be reading and resizing the images so they can fit into our model. After that we will try out different pretrained models while changing the last layer and hyper tuning parameters then observe how the model performs. We will try VGG-19 and InceptionV3 and InceptionResNet as those have

not been included in previous research. Our dataset will be divided into training, validation, and testing, 75%, 10%, and 15% respectively.

References

1- https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria
2- https://ceb.nlm.nih.gov/repositories/malaria-datasets/
3- https://peerj.com/articles/4568/