CLASSIFICATION OF MALARIA CELL IMAGES WITH DEEP LEARNING :
Build an explainable deep learning model to classify malaria-infected
cells and deploy the model in a simple dashboard.

Abstract:
This page hosts a repository of segmented cells from the thin blood smear
slide images from the Malaria Screener research activity. To reduce the
burden for microscopists in resource-constrained regions and improve
diagnostic accuracy, researchers at the Lister Hill National Center for
Biomedical Communications (LHNCBC), part of National Library of Medicine
(NLM), have developed a mobile application that runs on a standard
Android smartphone attached to a conventional light microscope. 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. The de-identified
images and annotations are archived at NLM (IRB#12972). We applied a
level-set based algorithm to detect and segment the red blood cells. The
dataset contains a total of 27,558 cell images with equal instances of
parasitized and uninfected cells. An instance of how the patient-ID is
encoded into the cell name is shown herewith: "P1" denotes the patient-ID
for the cell labeled "C33P1thinF_IMG_20150619_114756a_cell_179.png". We
have also included the CSV files containing the Patient-ID to cell
mappings for the parasitized and uninfected classes. The CSV file for the
parasitized class contains 151 patient-ID entries. The slide images for
the parasitized patient-ID "C47P8thinOriginal" are read from two
different microscope models (Olympus and Motif). The CSV file for the
uninfected class contains 201 entries since the normal cells from the
infected patients' slides also make it to the normal cell category
(151+50 = 201).
The data appear along with the publication:Rajaraman S, Antani SK,
Poostchi M, Silamut K, Hossain MA, Maude, RJ, Jaeger S, Thoma GR. (2018)
Pre-trained convolutional neural networks as feature extractors toward
improved Malaria parasite detection in thin blood smear images.
PeerJ6:e4568 https://doi.org/10.7717/peerj.4568
An improvement in performance has been recently reported using deep
neural ensembles toward malaria parasite detection in thin-blood smear
images and is published in the PeerJ journal as cited herewith:Rajaraman
S, Jaeger S, Antani SK. (2019) Performance evaluation of deep neural
ensembles toward malaria parasite detection in thin-blood smear images.
PeerJ 7:e6977 https://doi.org/10.7717/peerj.6977
The datasets are available at cell_images.zip, the codes at
malaria_cell_classification_code.zip and the Patient-ID to cell mappings
for the parasitized and uninfected classes at
patientid_cellmapping_parasitized.csv and
patientid_cellmapping_uninfected.csv respectively.

DATA COLLECTION -
1.For this work, we collected data that is already available on the
internet. The dataset is collected from NIH (National Institute
of Health) which contains 27,588 images of parasitized and uninfected
cells. The number of images per class is equally distributed with
13,794 images per class. The difference between malaria-infected
and uninfected cells from the dataset, chosen randomly

2.After data collection, the next step was data preprocessing. In the preprocessing step, we convert all the images into 64*64 dimensions which is the standard input image size of most pre-trained CNN models for faster model convergence. Then the next step is data labelling. We represent the parasitized images as 1 and uninfected images as 0 for the data in this step.

3.The next step was data splitting. In this step, data is divided into two different parts which are known as data training and data validation. In this process, the data set's splitting ratio is (75, 25), i.e. 75% of data used for training and the remaining 25% of data used for validation.

DEEP LEARNING -

Deep learning techniques are now generally used for image classification and medical image analysis. It has been a proven method which increases the performance in any field. A convolutional neural network (CNN), a type of deep neural networks, is essentially considered for research in the computer vision field. The deep architecture of CNN is its main power. The convolutional layer in the CNN works as an automatic feature extract-tor that extracts hidden and important features. Extracted features are passed to a fully connected neural network which performs classification images by maximizing the probability scores

1.After data is splitting into training and validation then the CNN algorithm was chosen for classifying the data. We give the train data to the CNN algorithm by making use of the fit function and train the algorithm to detect the malaria disease. For the training process, we use the 20669 images of the taken data.

2. After training the next process is validation. In this process, we have to test the given data to whether the algorithm classifies the given data correctly or not and shows the predicted class label for the dataset.

ANALYSIS AND DISCUSSION -

Deep learning methods achieve impressive performance (even surpassing human-level performance) in several tasks over a wide spectrum of domains. However, in many domains like healthcare, reasoning the predictions of a model is vital. They are useful in evaluating the reliability and fairness of a model, and in effective data-driven decision making.

While being a generally black box approach, with modern techniques, deep learning methods can provide some explanations to their predictions. In the context of image classification, heat-maps on the image highlighting areas that highly influence the classification of the target can be generated (see examples below).

RESULTS -

We have adopted the following approach to assessing the proposed CNN model's performance for the classification of uninfected and parasitized cell images. The results based on the test plan that has been done. At the first trial, the results were nearly negative that the Accuracy of malaria detection was 96.0% and Loss was1.6. After the second test plan has been done, the results were positive: The Accuracy was 98.0%, and Loss was 0.9. the Accuracy was calculated by rule, as shown below:

$$ACCURACY = (TP+TN/TP+FP+FN+TN)$$

This network's approach is instead of layers to learn the underlying mapping; we allow the network to provide the residual

mapping. So, instead of say H(x), initial mapping, let the network
fit, F(x): = H(x) – x which gives H(x): = F(x) + x.

RECOMMENDATION –
THIS model achieved 98.0% of Accuracy and 0.9 of Loss.It was recommended
to develop our model by: deploying our best improved model to a mobile
application to facilitate a simple and fast detection of malaria parasite
in blood cell images.