Malaria Diagnosis Using Deep Learning Methods

Abstract:

Diagnosing diseases requires effort, time, and people specialized in the field, and the possibility of human error in diagnosis is very high. Therefore, the need for an automated system for diagnosis has increased. One of the most dangerous diseases worlds around is malaria, which could infect humans with parasites that are transmitted through the bites of infected female Anopheles mosquitoes. Where the estimated deaths in 2019 reached about 409000 cases. Based on the purpose of early and automated diagnosing of malaria this project build. In this project I Implement the images classification using the convolution neural network CNN with two different training strategies, the first model is the pre-trained CNN model (MobileNetV2), and the second model is CNN with training over the dataset from scratch the second model achieves an accuracy of 96%. the result shows that we can take the advantage of using deep learning as diagnosing tool to help those in the medical field for easy and accurate diagnosis the malaria disease.

The Dataset: The used dataset is microscopic cells images (27,558 cell images), labelled to parasitized(abnormal) and uninfected cells(normal), this dataset was collected by developing a mobile application that runs on a standard Android smartphone attached to a conventional light microscope. The dataset is uploaded in Kaggle[1]. Figure 1 below shows a sample of the dataset. The dataset is a balanced dataset where the number of image samples of the parasitized class is the same as the number of uninfected classes which equal 13779 images.

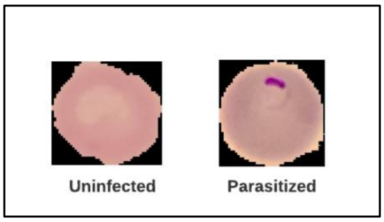


Figure 1:Dataset Sample

The Design: This project will aim to diagnose malaria based on classification the cells using the deep learning method (CNN). The first step is preparing the dataset by reading the images file and dividing it into 3 parts (training, valid, test). After that using ImageDataGenerator() to apply image rescaling and image resizing. Then start to build two CNN models the first one is implement using transfer learning and the other one is trained from scratch. After that evaluate the models to define the best performing model.

Algorithms:

To diagnose malaria automatically we implement two deep learning models based on CNN.

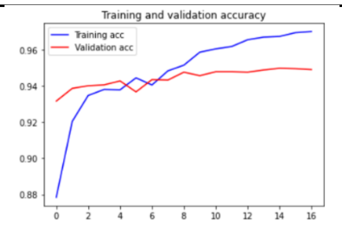
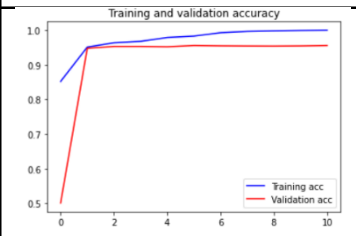
The first model is the pre-trained CNN model (MobileNetV2) where the process includes multiple steps. First, upload the pre-trained model from Keras application with defining the parameters of the model as follow: include_top = False(it means leave out the last layer of the model because I want to add my own output layer) and the parameter weights = 'imagenet'(which is pre-training on ImageNet) and make all the layers in the base model (non_trainable).

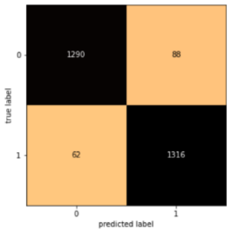
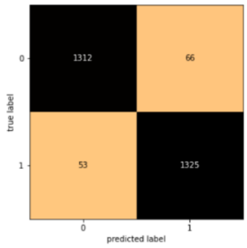
The second model is CNN which trained from scratch, constructed with different layers: (Input + Conv 1 + ReLU + Max Pooling+ BatchNormalization), (Conv 2 + ReLU + Max Pooling+ BatchNormalization), (Conv 3 + ReLU + Conv4+ ReLU+ Max Pooling+ BatchNormalization),(Conv 5 + ReLU + Max Pooling+ BatchNormalization), (Fully Connected + ReLU), (Output).

For model evaluation I use accuracy and classification report and confusion matrix.

Result:

The evaluation of models shows that the MobileNetV2 obtain an accuracy of 95% in the testing set and the CNN which trained from scratch achieves better accuracy with 96% accuracy. the table below represents the accuracy in training and validation sets in both models also the confusion matrix represents the performance for each model.

Performance	MobileNetV2	CNN
Accuracy in training and validation sets		

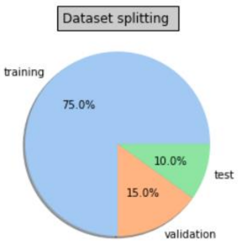
Confusion matrix		
Accuracy	95%	96%

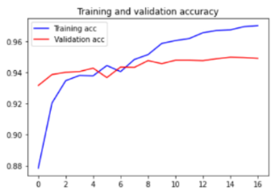
Tools:

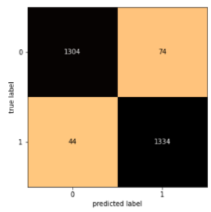
- Preprocessing: os , numpy, pandas, split-folders
- The model: keras, sklearn, tenseorflow
- Visualization: matplotlib, seaborn, slide show

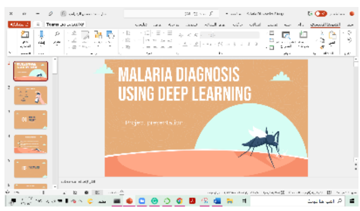
Communication:

Used multiple kinds of plots and graphs to present the dataset splitting and the accuracy and loss in the training process and the confusion matrix for illustrating the performance of the models, finally use the PowerPoint slide show to show the project presentation.









References:

[1] “Malaria Cell Images Dataset | Kaggle.” https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria (accessed Nov. 27, 2021).