PROJECT REPORT

ON

MALARIA CLASSIFICATION

Akshay Sasheendran

akshaysasheendran.98@gmail.com

ABSTRACT

Malaria poses a global health problem every day, as it affects millions of lives all over the world. Traditional diagnoses are time-consuming and prone to error. Thus, automated identification and classification of malaria type and stage of advancement can offer patients with a faster and more accurate diagnosis. Our application is able to classify between malaria and non-malaria infected cell. The model is trained using a resnet50 model which uses keras. It has achieved 94% validation accuracy. At the end an application using streamlit to classify a given image is deployed.

INTRODUCTION

Malaria is a disease caused by parasites that are transmitted to people through the bites of infected mosquitoes. It is a leading cause of illness and death, particularly in sub-Saharan Africa. One way to diagnose malaria is by examining a blood sample under a microscope to identify the presence of malaria parasites. This can be a time-consuming and labor-intensive process, and there is a need for more efficient and accurate methods of diagnosis.

Based on the simulation findings, it was proved that these deep-learning algorithms were capable of reaching extraordinarily high accuracy in pattern recognition. One way to do this is to use a convolutional neural network (CNN), such as a ResNet model, to classify the images. The ResNet model is a type of CNN that has been successful in a variety of image classification tasks and may be well-suited for the task of malaria cell classification.

ResNet Neural Network

Residual Network (ResNet) is a classic neural network. Training very deep neural networks was an issue until ResNet was introduced due to the vanishing gradient problem. ResNet's skip connection technique helped solve the vanishing gradient problem and help to get better results using deep neural networks like ResNet50.
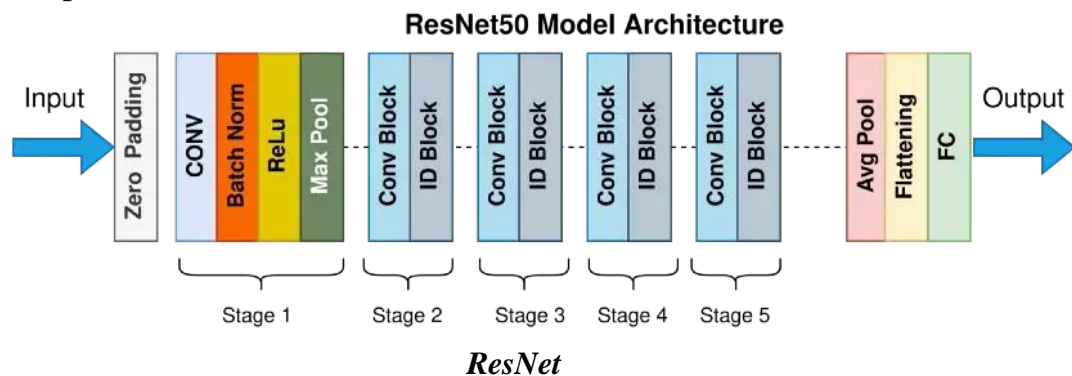
ResNet

And our resnet50 model is specified with include top= False, parameter which determines not to include the fully connected layers at the top of the network and weights='imagenet' are often used as a starting point for fine-tuning on a new dataset, here the weights are the pretrained weights on the ImageNet dataset. Also, input shape specifies the shape of the input to the network. The metrics used to measure the dataset were Accuracy and Loss. These metrics were measured for both training and validation data.

DATA PREPARATION

The dataset is taken from the National Institute of Health's (NIH) official website. The dataset contains 27,557 images: 13,778 infected images, and 13,779 non-infected images

Dataset images (a) parasitized (b) uninfected.

Finally, the dataset was divided into three parts: a training set, a verification set, and a testing set for training, verification, and testing.

☐Defining number and characteristics of classes.

• Parasitized- the one with infected cell image

• Uninfected – the one with non infected cell image

☐Seperating into train, validation and test sets.

• 80% - images in training set

• 10% - images in validation set

• 10% - images in test set

IMPLEMENTATION

Here, we decided to resize all the images to [112,112,3] 112 x 112 dimension and RGB with three channels. It is obvious now that choosing this fixed size for images is a matter of trade-off between computational efficiency and accuracy.

And resnet50 model is specified with include top= False, parameter determines not to include the fully connected layers at the top of the network and weights='imagenet' are often used as a starting point for fine-tuning on a new dataset, here the weights are the pretrained weights on the ImageNet dataset. Also, input shape specifies the shape of the input to the network. And finally our model is fitted and trained with five epochs. After

each epoch the error function gradually decreases to obtain the minimum value.

Here, it can be observed that loss is substantially being reduced as the number of batches is increased for processing. The same can be observed for training and validation loss on the fine-tuned ResNet model. Also, for each epoch we can see that graph accuracy metric is improving as it tend to upward and flattens at the end. In deploying the streamlit the saved model is called and the test model is converted to array dimensions to predict the classifier model suggesting parasitized and uninfected cell.
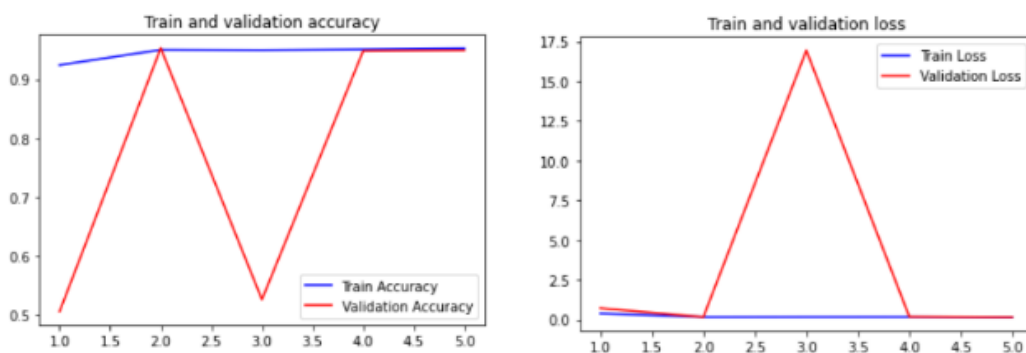
Training-validation accuracy and loss graphs of ResNet model.

CONCLUSION

The project shows promising results for malaria cell classification with the ResNet model with an accuracy of 94%. In context, the results obtained with ResNet architecture is promising for the classification of malaria parasites in the field of medical imaging field.

REFERENCE

1. Sureshan, S. (2021, October 12). *Malaria detection model: Deep learning based malaria detection*. Analytics Vidhya. Retrieved January 5, 2023, from https://www.analyticsvidhya.com/blog/2021/10/deep-learning-based-malaria-detection-model-for-beginners/
2. Dwivedi, P. (2019, March 27). *Understanding and coding a ResNet in Keras*. Medium. Retrieved January 5, 2023, from https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33
3. Team, K. (no date) *Keras Documentation: Resnet and RESNETV2*, *Keras*. Available at: https://keras.io/api/applications/resnet/ (Accessed: January 5, 2023).