International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-11 Issue-2, December 2021

Malaria Parasite Detection in Thick Blood Smears using Deep Learning

K Venkata Shivaramakrishna Reddy, S Phani Kumar

Abstract: ***Malaria parasitized detection is very important to detect as there are so many deaths due to false detection of malaria in medical reports. So analysis has gained a lot of attention in recent years. Detection of malaria is important as fast as possible because detecting malaria is difficult in blood smears. Our idea is to build a transfer learning model and detect the thick blood smears whether the presence of malaria parasites in a drop of blood. The data consists of 5000 each infected and uninfected data obtained from the NIH website. In this paper, I propose to use three different types of neural networks for the performance evaluation of the malaria data by transfer learning using CNN, VGG19, and fine-tuned VGG19. Transfer learning model performed well among various other models by achieving a precision of 98 percent and an f-1 score of 96 percent.***

Keywords: ***VGG19, Malaria Parasitized, Deep Learning, Convolutional Neural Network, Transfer Learning, Image Augmentation, Python.***

I. INTRODUCTION

Malaria is a life-threatening disease that is caused by mosquito-borne infectious diseases it affects human beings and animals a lot. It is caused by an infected mosquito when it is bitten and usually symptoms begin after ten to fifteen days. Symptoms typically include fever, vomiting, tiredness, and headaches. In severe cases, it causes yellow skin, seizures, coma, or death. To detect the malaria parasite in our body examination of the blood and antigen detection tests. The parasite first grows in liver cells multiplies itself there, then in the red blood cells, and the red blood cells the parasite grows and kills the red blood cells. According to WHO, there are around 435,000 deaths in 2018 and 2019 it is increased to 487,000 as there is an increase in deaths it is a life-threatening disease. So, to detect malaria parasites we have microscopy examination of thin blood strain and thick blood strain. It's time-consuming and widely used for the detection of malaria in microscopy examination and examination required expertise 'parasitologist'. Due to the incorrect diagnostic of the malaria parasite which causes inappropriate treatment. In return of incorrect diagnosis leads to unnecessary use of antimalarial drugs and suffers from side effects.

Manuscript received on December 21, 2021.
Revised Manuscript received on December 26, 2021.
Manuscript published on December 30, 2021.
* Correspondence Author
K Venkata Shiva Rama Krishna Reddy*, Student, Department of Information Technology, Gitam University, Hyderabad (Telangana), India. E-mail. kvenkatshivareddy@gmail.com
S Phani Kumar, Department of Computer Science, Institution, Gitam University, Hyderabad (Telangana), India. E-mail. psingams@gitam.edu

So, therefore the development of automatic detection of malaria parasites is very important for patient treatment and management. It has two big advantages: fast detection of parasites will lead to the easy treatment of the parasite and second no need for expertise 'parasitologist' also time-consuming. Whereas there are two types of malaria parasite detection in blood one is thin blood smears and the other is thick blood smears. In thin blood smears both RBC's and WBC's are visible so automatic detection is difficult first we have to segment each of the RBC's and then WBC's and after that classify each segmented RBC as infected or uninfected. Whereas in the case of thick blood smears the nuclei of the RBC's are visible so parasites can detect them directly. The first step is typically challenged because WBC's nuclei and various non-parasites can lead to artifacts and cause false parasite detection.

VGG19

These are the pre-trained networks which train a million images database at a time and it's a type of convolutional neural network and has 19 layers deep. It has a kernel size of 3x3 and has a linear chain of layers. In 19 layers VGG 19 neural network consists of 16 convolutional layers, 5 Max Pool layers, 3 fully connected layers and 1 SoftMax layer. VGG 19 is a visual Geometry Group consisting of 2D Convolutional and Max Pooling layers. It was trained on the NVIDIA Titan Black GPU.

II. DATA AUGMENTATION

Data augmentation is a technique used to significantly increase the variety of data available for training models without actually acquiring any new data. It artificially creates new and varied types of training data. This is mostly implemented by deploying domain specific techniques. There are various data augmentation techniques such as cropping, horizontal/vertical flipping, rotating, etc. These are the main techniques used in training the neural networks. Image data augmentation in training neural networks does not use advanced augmentation techniques.

III. TRANSFER LEARNING

It's a research problem while solving one problem knowledge gained is stored and focuses on applying on different related problems. Knowledge gained to recognize cars will apply while recognizing other vehicles. Types of neural network in transfer learning positive learning, negative learning and neutral learning.

CNN

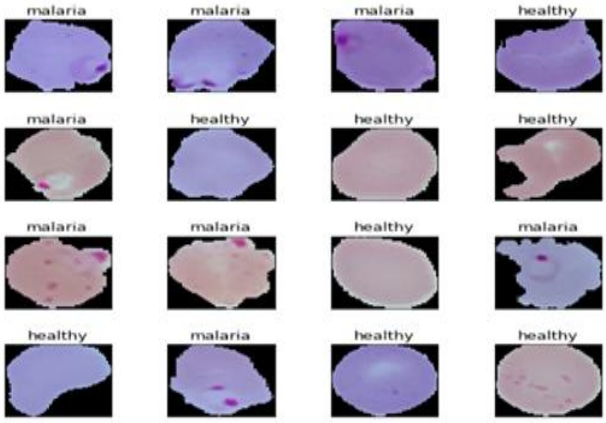
In deep learning, a CNN i.e convolutional neural network is a type of deep learning/deep neural networks, it is very avidly used to analyse visual imagery. People usually assume that Neural networks are mostly used for matrix multiplication but that is not really the case in reality when it comes to CNN because it uses the mathematical technique called convolution. There is no need for us to get into the mathematics of convolution because the CNN libraries are being used and they usually handle it for us.

IV. LITERATURE REVIEW

Several Malaria parasite detection solutions are proposed. There are approaches like threshold, morphological operation, genetic programming classifier, SVM classifier, and CNN classifier. These all approaches have fall-positive results. So, this is the motivation for researchers to use different image processing techniques. Kaewkamnerd on the HSV images he used an adaptive threshold to detect the parasite in thick blood smears from the background of WBC's and distinguish parasites from WBC's and distinguish parasites from WBC's according to their size. Evaluated on 20 images and achieved 60% accuracy. Chakrabortya used morphological segmentation with color images and performed an experiment on 75 images and the accuracy obtained was around 95%, the false-positive ratio of 10% as the dataset was very small, good accuracy was obtained. Elter used machine learning techniques based on the classification of the pre-selected parasite images using a support vector machine (SVM) for 256 images that have 97% sensitivity. But this process has complications because we must preselect the parasite images and then path level on each image. Yunda pre-segmented images are extracted for the texture features and then principal component analysis (PCA) after the neural network which shows the sensitivity of 76.45% on 110 parasite images. Quinn evaluated 2903 images and precision of 90% at a recall of 20% and used a randomized tree classifier for classification. Rosado evaluate 94 images based on the SVM classifier and parasite detection. In this model, it shows patch level accuracy of 80.5% and sensitivity of 91.8%, specificity of 93.5% for RBC's whereas in the case of WBC's detection achieves 98.2% and 72.1% of sensitivity & specificity this approach has evolved got performance in the patch level. The goal is to classify both parasite and false positive for a patient this classification has not good assured performance on an image level. Recently deep learning had the highest trend and superior performance in image processing of the big data and boosted in many areas including non-medical areas. So deep learning does not require any segmentation and helps in feature representation. Delhunt used classification and the neural network which are SVM and convolutional neural networks for localizing the parasite detection and used feature extraction on 143 patients' images which should have a specificity of 92% on patient level. Quinn proposed four-layered CNN models to extract from the RGB images and shows average precision of 97% on 182 images. The overview shows that on small datasets we do not achieve high accuracy on patch level performance is becoming qualitative and does not guarantee similar results our goal for using fine-tuned transfer learning method to increase the sensitivity and accuracy of the model.

V. IMPLEMENTATION

We developed a python script to obtain the required parasites by applying relevant data which is image processing data. And code has written in python 3.8 system configuration is intel i7 processor, 1Tb hard disk, 16 Gb ram & windows 11 operating system. So, these are the steps involved in this malaria parasite detection. Data is image data having parasitized and uninfected datasets and reading all the data as the input dataset. Hence required dependencies for loading the python dataset and examining the image dataset. Pandas, Os, Glob, Cv2, Threading, Matplotlib, Label encoder, TensorFlow, Datetime, Keras, Model-evaluation Utils Concurrent, NumPy. Install all these dependencies in windows for processing the data. And official dataset contains images of different malaria parasites which contains data of 13780 images of each of the parasites. After the data is loaded we have to train, validate and test the data for predicting the outcome by using different types of neural networks. We have to evaluate the data by performing the different model performances. After comparing the model we have to predict the best algorithm for malaria parasite detection. We have to split the data into train and test by splitting the data into equal partitions after that load the image and resize the image dimensions. Sample images of the healthy and malaria parasitized images.



To generate the predictions of our dataset we use three different neural networks one is the basic convolutional neural network and the second one is using transfer learning pre-trained CNN which is VGG(visual geometry group) and the last one is fine-tuned and pre-trained CNN with image augmentations To generate the predictions of our own images we used three different types of neural networks. Here in our first basic CNN model without changing the code of the model, train our dataset like that only to repeat two more models one is frozen pre-trained CNN and the other one is fine-tuned pre-trained CNN with data augmentation, and the model is frozen for performance evaluation of each algorithm.

Retrieval Number: 100.1/ijeat.B32931211221
DOI: 10.35940/ijeat.B3293.1211221
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Dataset is loaded and trained for 25 epochs and contains a batch size of 64 and the learning rate will be 0.01 approximately it will take 1 to 2 days as the GPU performance required faster training because the malaria image dataset has very huge and sometimes it switches from TPU to GPU to decrease the load on the GPU. The performance evaluation of each model is stored in a 2D graph between accuracy and loss for train and validation sets. A predicted confusion matrix is stored and evaluation is done there to predict which model performance is good. After all that the models are predicted their output we go for performance evaluation, this is called inference mode.

VI. EXPERIMENTAL RESULTS

The evaluation of the dataset is carried on the Jupyter notebook and runtime as TPU and sometimes runtime changes to GPU runtime to decrease the workload on the data. The data consists of 13780 parasitized and uninfected data in different directories. The data is loaded and data is split into train and test data. After that loading the dimensions and decreasing the image dimensions and resize the image.

```
Model Performance metrics:
------------------------------

Model Classification report:
------------------------------
              precision    recall  f1-score   support

     healthy       0.94      0.98      0.96      4075
     malaria       0.98      0.94      0.96      4193

   micro avg       0.96      0.96      0.96      8268
   macro avg       0.96      0.96      0.96      8268
weighted avg       0.96      0.96      0.96      8268

Prediction Confusion Matrix:
------------------------------
                 Predicted:
                   healthy malaria
Actual: healthy       4004      71
        malaria        260    3933
```

Here the sample images are loaded so that healthy and malaria data can be distinguished. Configure the batch size and epochs for the models which we are performing.

First CNN model we build the basic architecture model and after that training, the model basic CNN performance is shown in a graph between train accuracy and validation accuracy and another graph between train loss and validation loss. Then the model is saved for performance evaluation.

```
Model Performance metrics:
------------------------------

Model Classification report:
------------------------------
              precision    recall  f1-score   support

     healthy       0.93      0.95      0.94      4075
     malaria       0.95      0.93      0.94      4193

   micro avg       0.94      0.94      0.94      8268
   macro avg       0.94      0.94      0.94      8268
weighted avg       0.94      0.94      0.94      8268

Prediction Confusion Matrix:
------------------------------
                 Predicted:
                   healthy malaria
Actual: healthy       3871     204
        malaria        312    3881
```

Same steps are repeated for other two models in the next model frozen pre trained CNN for transfer learning.

```
Model Performance metrics:
------------------------------

Model Classification report:
------------------------------
              precision    recall  f1-score   support

     healthy       0.95      0.95      0.95      4075
     malaria       0.95      0.95      0.95      4193

   micro avg       0.95      0.95      0.95      8268
   macro avg       0.95      0.95      0.95      8268
weighted avg       0.95      0.95      0.95      8268

Prediction Confusion Matrix:
------------------------------
                 Predicted:
                   healthy malaria
Actual: healthy       3884     191
        malaria        225    3968
```

After that image augmentation for the next fine tuned pre-trained CNN model. All the models performance evaluation is stored and in the end data is performed and stored the performance evaluation.

	Accuracy	F1 Score:	Precision:	Recall
Basic CNN	0.9497	0.9497	0.9497	0.9497
VGG-19 Frozen	0.9376	0.9376	0.9379	0.9376
VGG-19 Fine-tuned	0.9600	0.9600	0.9610	0.9600

VII. CONCLUSION

In this paper we have performance evaluation of malaria parasitized for three different types of models which evaluates each model's performance. In the first model which is common CNN model where the accuracy was 94.97% and in the VGG19 frozen network we got the accuracy of 93.76% so in the third model VGG19 is fine tuned and data augmentation is done so that accuracy and performance should be increased the accuracy we achieved is 96.00%. for this huge data model has performed well as compared to other models and for less data the model performance and evaluation will be promising good results. So in the end the model can detect malaria in a parasite without any losses. In this research interesting challenges are obtained for detecting malaria parasitized detection.

REFERENCES

1. Ashwini Awchite et.al., "A Survey on Detection of Malarial Parasites in Blood Using Image Processing", International Journal Of Innovative Research in Computer and Communication Engineering, vol.1 No.1 pp. 1096-1100, October 2011.
2. Ms.Deepali Ghate, Mrs. Chaya Jadhav, Dr. N Usha Rani " Automatic detection of parasite from blood images", International Journal of Advanced Computer Technology(IJACT), vol 4, Number 1,2011.
3. Pallavi T. Suradkar, "Detection of malarial parasites in a blood smear using image processing", International Journal Of Engineering and Innovative Technology,vol 2, April-2013.
4. Korenromp, E., Miller, J., Nahlen, B., Wardlaw, T., Young, M.: World Malaria Report, Technical Report, World Health organization, geneva (2005).
5. Tek, F.B., Dempster, A.G., Kale, I.: Malaria parasite detection in peripheral blood images. In: Proc. Br. Mach. Vis. Conf., Edinburgh, UK (2006).

6. World Health Organization, Malaria Report 2018.
7. WHO, "Fact sheet: World Malaria Report 2016," in World Health Organization, World Health Organization, 2016. [Online].
8. Ahirwar A., Pattnaik S., Acharya B., Advanced Image Analysis Based System for Automatic Detection and Classification of Malarial Parasite in Blood Images. International Journal of Information Technology and Knowledge Management.
9. CDC, "Frequently asked questions (FAQs)," CDC, 2016. [Online]. Available: https://www.cdc.gov/malaria/about/faqs.html.
10. M. Sheikh Hosseini, H.Rabbani, M.Zekri, A.Talebi,“ ”, WSEAS transaction on biology and biomedicine, Issue Automatic diagnosis of malaria based on complete circle sear [6] ch algorithm ”. RMSjournal of Microscopy.
11. Kishor Roy, Shayla Sharmin, Rahma Bintey Mufiz Mukta, Anik Sen, using image processing” February 2018 [3] Pallavi T Suradkar, [4] Aimi Salihah Abdul “ “Detection of malaria parasite in Giems a blood sample, International Journal of Computer Science & Information Technology (IJCSIT)

AUTHORS PROFILE

K Venkata Shiva Rama Krishna Reddy, M-Tech Data Science, Efficiently diagnosing breast cancer detection from hypsographical images by combining FCN and Bi LSTM model. Research Areas Machine Learning, Data analytics and Image Processing.

S Phani Kumar, M.Tech, PhD, Research Areas: Software Engineering, Safety Critical Systems, Machine Learing, Wireless networks; And Published so many papers like Genetic Algorithms with Feasible Operators for Solving Job Shop Scheduling Problem. Hybrid shape-based Image Retrieval Performance improvement using Parallel Processing. International Journal of Scientific and Technology Research. Also International Conference NGCT 2017, International Conference on Inventive Systems and Control, International Conference on Advanced Material Technologies (ICAMT). Attended as guest lectures Pipelining Hazards, Fuzzy Systems E&ICT Academy & NIT Warangal.