

A Study on Machine Learning-based Malaria Detection

Joymalya Biswas¹, Akash Singh², Adarsh Yadav³, Krishna Kunal⁴, Pankaj Sharma⁵

Abstract: *As deep learning and computer vision developed rapidly, the exact recognition of medical imagery was one of the major factors in medical diagnostics and decision-making. For this purpose, the Convolutional Neural Network (CNN) data-driven approach is proposed for detecting paludism parasites, which can automatically generate deep neural networks using evolutionary algorithms and optimized the structure of their network topology to detect a person being infected or not with a life-threatening disease like Malaria. Malaria continues to pose a significant threat to global health with some 200 million worldwide cases and over 400,000 deaths a year. Modern information technology, biomedical research, and political efforts play a major role in many efforts to combat the disease. One of the challenges to a promising decrease in mortality was in particular an inadequate diagnosis of malaria. This paper provides a description of these methods and examines the progress of the microscopic malaria detection field and we have chosen to use deep learning for the identification of malaria parasites. CNN is very helpful for this purpose. In particular, it has the advantage of automatically creating a network structure most commonly used for analyzing visual imagery. CNN is employed for this purpose as the quantity of pre-processing is less in relation to other image classification techniques. The CNN consists of various layers of mainly an input layer, a hidden layer, and an output layer. The middle layers of a feed-forward neural network are known as hidden layers as they are obscured by an activation function. The two main components used are TensorFlow and Keras. The input passed is in the shape of Tensor with a certain shape with its parameters and particularly VGG19 package of Keras that is a kind of CNN is employed in the classification and detection of data. Multiple Epochs are used in order for the algorithm works through the dataset multiple times.*

Keywords: CNN(Convolutional Neural Network), Malaria Detection, VGG(Visual Geometry Group), Epochs, TensorFlow, ANN(Artificial Neural Network).

Introduction

Malaria is a mosquito-borne disease. It is a communicable disease that can grow from person to person or other animals due to infectious bites of a female Anopheles mosquito. In a UNICEF/WHO report of date 2015, every year, nearly 3.2 billion people are at risk of getting infected by malaria, which is nearly one-half of the planet's total population. According to report of World Health Organization's 2019 World Malaria Report, nearly 228 million cases of malaria were reported worldwide in 2018, up from 251 million cases in 2010 and 231 million cases in 2017. Furthermore, in 2018, children under the age of five were the most exposed individuals, accounting for 67 percent (272,000) of all malaria-related deaths worldwide.

Malaria had been discovered to be the leading reason for death all over the world. In 2018, 94% of the death due to malaria were mentioned in the WHO African Region. The report claims that one of the main reasons for the children having or showing the signs and symptoms of fever but not receiving medical attention may be because of inadequate access to health-care providers, with an increased effort of availing malaria diagnosis and treatment across community level.

It has numerous signs such as fever, fatigue, vomiting, and headaches. If not treated swiftly and appropriately on time, it can cause yellow skin, diarrhea, profuse sweating, muscle pain and pain in joints, seizures, coma, and leading to death. Most of its symptoms are too generic making it challenging to detect.

In previously performed analysis, many systems are not standardized or rigorous enough to ensure an accurate diagnosis. As a result, they largely depend on the skills, expertise of the “microscopist,” and the environment where they work so results are not 100% accurate.

Such claims suggest that innovative ways are required in healthcare systems for malaria elimination. Malaria is a serious health issue, and for quicker results and diagnosis, a more computationally efficient mode is required. Machine learning is an attractive research area in computer engineering. Because of its, dynamic solution style and learning capability, this approach is applied in lots of real-world application areas. CNN is a deep neural network approach that is mainly a tool for categorising images, which exactly fits our research area on the current topic. Therefore, this paper, it is aimed to automate the diagnosis process of Malaria disease with the assistance of CNN.

In modern times with the increasing computation power of computers, various new techniques are approaching for better and quicker categorization of images. Various Deep learning approaches are present like CNN, RNN, Long Short-Term Memories, etc. CNN is mostly preferred for the classification of images.

Goal and Contribution

This research paper's goal is to train a model that can predict or detect malaria disease with high accuracy. This will help doctors with precise diagnoses and also add a layer of confirmation for them. Various layers of train and test datasets have become accustomed to get distinct output. This paper is schematizing in distinctive sections as follows:

Section 1 is ‘Introduction and Background Knowledge’;

Section 2 is ‘Related work and Design and Implementation’;

Section 3 is ‘Experimental Results’;

and finally, Section 4 is ‘Conclusion and References’.

BACKGROUND KNOWLEDGE

Machine and learning words have a history as old as humanity. It is quite similar to how human beings learn they are provided with pieces of data based on image or writing and some tutor or teacher teaches it to them. This is Machine learning terms is known as the training data. The Student or the model both learn and gets inside information and try to derive some conclusion based on that. Then comes the test where the student has to prove his ability or how much he has learned from it. This is Machine learning terms is known as test data. The percentage accuracy determines how good your data is working.

A. Deep Learning

It is a subgroup of the machine learning created with the structure of artificial neural networks. It uses various layers to extract higher results from the given raw input. It is used in various applications like speech recognition, machine vision, NLP, medical image association, drug details, social media, biometrics, etc. It uses ANN for computation and has various features.

B. Artificial Neural Network (ANN)

It collects weighted values and sends them to a threshold function. If the value exceeds the threshold function, it is fired to the subsequent layer. The error is calculated between the actual value and the estimated the neural network's worth and the coefficients(weights) are updated with the backpropagation algorithm.

C. Convolutional Neural Network (CNN)

In contrast to ANN, CNN perform a convolution process instead of weighing in the processing phase. Perceptron essentially works like a kernel unit. It aims to locate the appropriate weights and ascertain the properties of image.

D. TensorFlow

It is an open-source software library used for Machine Learning. It is used for large number of tasks but used basically for deep neural tasks and image classification. It is based on dataflow and differential programming.

E. Keras

It is a neural networks API. It is very user-friendly as it supports a large number of deployment options with backend technologies like TensorFlow, CNTK, etc. It requires a strong set of graphical processing to work fluently.

F. Confusion Matrix

It is an efficiency metric for machine learning classification problems with two or more groups as output. There are four different variations of expected and real values in this table.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

It's perfect for assessing Memory, Precision, Specificity, Consistency, and in particular, the AUC-ROC Curve.

Related Work :

To diagnose malaria using machine learning several algorithms are proposed. Generally, information from RBCs is collected and further implementations are applied on top of that to get a better understanding of the blood smear.

- 1) The first and most important requirement is a clear image of RBC.
- 2) Amount of workload can be widely decreased with the assistance of this technique.
- 3) Overall costs can be lowered heavily.

Firstly the clear image is taken and certain pre-processing algorithms are applied to normalize the image for different kinds of blood straining or any kind of light effects. Blood parasites that are present are discovered in the next step to get a deeper knowledge of the cell. Segmenting the protein and getting the cell clusters to get the count of cells. Then in the final step by typical visuals, the model is distributing the cells as infected or uninfected. To discriminate

between infected and uninfected cells we use CNN. Using ML ensures that we do not require continuous human efforts or hand-crafted features. However, a huge dataset for training is provided for better understanding.

Related to this there are various approaches to solve this problem and various kinds of CNN present such as VGG, Lenet, AlexNet, ResNet, etc but for the model VGG 19 in accordance with Keras to give the best accuracy close to 90%. Hence Tensorflow in accordance with Keras is very important.

Design and Implementation

As discussed earlier we used image-based classification for malaria detection and its examination. These are present within the deep learning framework and hence quite successfully used. For classification of image CNN is used and likewise VGG 19 for classification. The network is constantly updated with actual values and the model is optimized and updated.

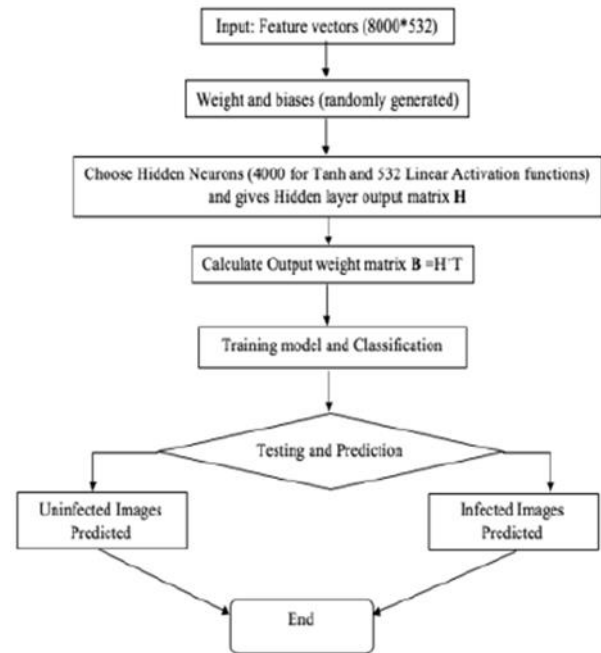


Fig 2. Steps followed

Our model name model_1. model_1 is created using VGG. After creating model the set of possibilities are 20,074,562. Out of these, set of trainable parameters are 50,178 and non-trainable parameters are 20,024,384.

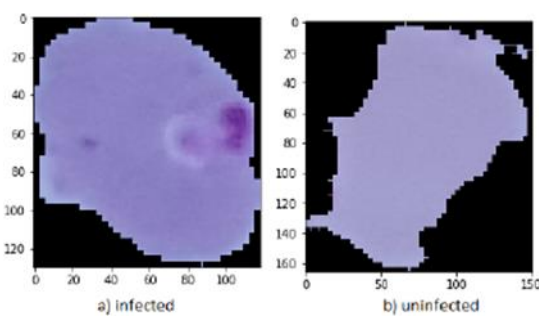


Fig 1. Infected and Uninfected samples

As shown in the above pictures there are two varieties of cell images are portrayed. All the cell images used in the model are gathered from various medical organizations and then divided into two categories. The first image being an infected one and the second one being uninfected. Preprocessing of the photo to reduce their sizes to (224, 224, 3) are done.

```

Model: "model_1"
-----
Layer (type)                Output Shape                Param #
-----
input_1 (InputLayer)        (None, 224, 224, 3)        0
block1_conv1 (Conv2D)        (None, 224, 224, 64)       1792
block1_conv2 (Conv2D)        (None, 224, 224, 64)       36928
block1_pool (MaxPooling2D)   (None, 112, 112, 64)       0
block2_conv1 (Conv2D)        (None, 112, 112, 128)     73856
block2_conv2 (Conv2D)        (None, 112, 112, 128)     147584
block2_pool (MaxPooling2D)   (None, 56, 56, 128)       0
block3_conv1 (Conv2D)        (None, 56, 56, 256)       295168
block3_conv2 (Conv2D)        (None, 56, 56, 256)       590080
block3_conv3 (Conv2D)        (None, 56, 56, 256)       590080
block3_conv4 (Conv2D)        (None, 56, 56, 256)       590080
block3_pool (MaxPooling2D)   (None, 28, 28, 256)       0
block4_conv1 (Conv2D)        (None, 28, 28, 512)       1180160
block4_conv2 (Conv2D)        (None, 28, 28, 512)       2359808
block4_conv3 (Conv2D)        (None, 28, 28, 512)       2359808
block4_conv4 (Conv2D)        (None, 28, 28, 512)       2359808
block4_pool (MaxPooling2D)   (None, 14, 14, 512)       0
block5_conv1 (Conv2D)        (None, 14, 14, 512)       2359808
block5_conv2 (Conv2D)        (None, 14, 14, 512)       2359808
block5_conv3 (Conv2D)        (None, 14, 14, 512)       2359808
block5_conv4 (Conv2D)        (None, 14, 14, 512)       2359808
block5_pool (MaxPooling2D)   (None, 7, 7, 512)         0
flatten_1 (Flatten)         (None, 25088)              0
dense_1 (Dense)              (None, 2)                   50178
-----
Total params: 20,074,562
Trainable params: 50,178
Non-trainable params: 20,024,384
  
```


Another important task is fitting the images into a model using ImageDataGenerator and passing the images in various classes. In our model we have used 5 epochs further to improve accuracy more layers can be added easily. Furthermore, creating a confusion matrix is another important term. Confusion matrix is basically a table that determines the performance of a classifier on a particular test dataset.

There are three main steps to create a confusion matrix

- 1) We need a test dataset with expected outcome values.
- 2) Prediction for each row of a dataset is done.
- 3) From the prediction count and expected outcomes: check how many predictions are right for each class of a dataset.

EXPERIMENTAL RESULTS

The main target is to recognize persons who are infected by checking their blood smears or by searching for dark spots in a cell with the aid of CNN techniques. It is very crucial to test the training of network to get accurate results. So, in this section we will discuss the system and specifications used and results obtained by it. used quite easily. All elements used are clearly mentioned in Table 1.

TABLE I

EXPERIMENTAL ENVIROMENT

Experimental Enviroment	Specifications
GPU	AMD R5 M330
Operating System	Windows 10
Development Enviroment	Jupyter Notebook
Model Package	KERAS 2.3.1
Backend	TensorFlow 2.1.0

Training and testing of neural networks require a lot of time. But with increase in the graphical power and GPU helps in having multiple layers of image processing and helps in preparing the data better and improve

accuracy. In the end the neural networks works very well with TPU(Tensor Processing Units) which is specially designed for deep learning approaches. In our current model we developed an artificial neural network(ANN) on the GPU and we used AMD R5 M330 for that. Furthermore, the processor present was i7 7500U with a base clock speed of 2.7GHz and 8 GB ram. The system worked on Windows 10 operating system. . The experimental environment was developed with KERAS version 2.3.1 and TensorFlow backend (2.1.0) . KERAS, a python deep learning library, is used along with VGG19(A type of CNN) for image processing along with TensorFlow for backend. Additionally we used the following components as they were readily available and easy to use. Also, it's a library which can be used quite extensively for image processing and reading or augmentation of image data is quite easy. Jupyter notebook is used for coding as it is readily available with Anaconda and can be used quite easily. All elements used are clearly mentioned in Table 1.

Fig 3. Epochs and accuracy

```
Epoch 1/5
13/13 [=====] - 217s 17s/step - loss: 0.3136 - accuracy: 0.8894 - val_loss: 0.3729 - val_accuracy: 0.7687
Epoch 2/5
13/13 [=====] - 216s 17s/step - loss: 0.2733 - accuracy: 0.9038 - val_loss: 0.2791 - val_accuracy: 0.8134
Epoch 3/5
13/13 [=====] - 216s 17s/step - loss: 0.2660 - accuracy: 0.8894 - val_loss: 0.2567 - val_accuracy: 0.8507
Epoch 4/5
13/13 [=====] - 216s 17s/step - loss: 0.3179 - accuracy: 0.8486 - val_loss: 0.3157 - val_accuracy: 0.8507
Epoch 5/5
13/13 [=====] - 216s 17s/step - loss: 0.2692 - accuracy: 0.8942 - val_loss: 0.1510 - val_accuracy: 0.8209
```

Fig 4. Loss Score of the System

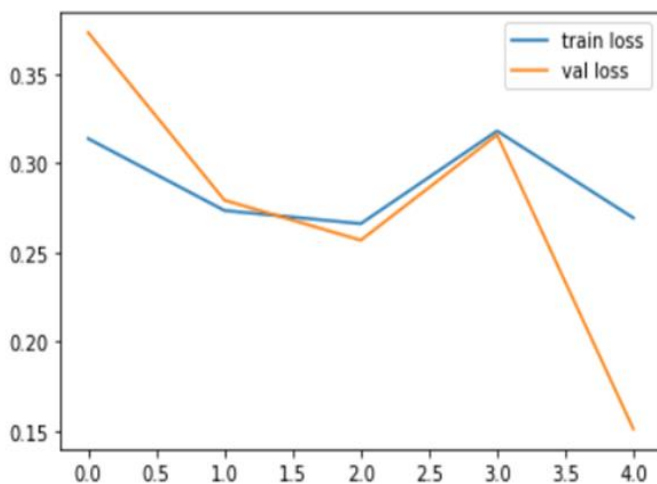
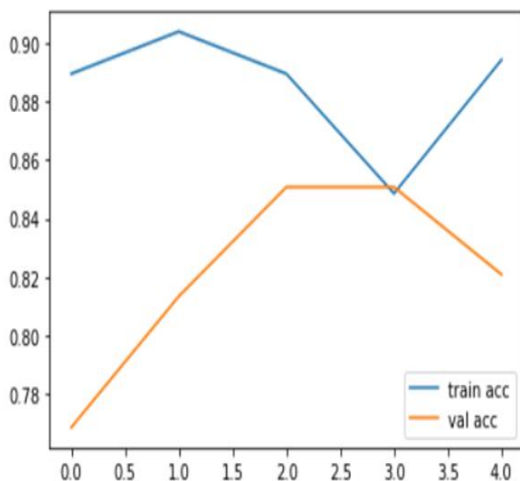


Fig 5. Accuracy Score of the System



CONCLUSIONS AND FUTURE WORKS

Malaria is one of the most deadly diseases, especially emerging in hot and humid climate near the equator and tropical regions. This include mostly countries of Africa, southern Asia, South America and so on. Detection and proper diagnosis of this disease is very important where doctors use microscope to check for pathogen in blood cells. But in countryside areas and poor countries cannot afford doctors all the time. So in that case our model could be very helpful for lot of people around the globe. In recent time heavy growth in computerized systems and algorithms are widely used mostly in the field of medicine.

In the future we aim to train the data over a larger dataset and add more feature values to it. Also installing more powerful graphical units(GPUs) and optimizing algorithm can make this faster and better. Also parallel computing devices can be done with this purpose. Finally, increasing accuracy and adding additional layers of epochs will significantly aid in the development of a better model.

REFERENCES

- [1] Malaria parasite detection using different machine learning classifier Adedeji Olugboja;Zenghui Wang 2017 International Conference on Machine Learning and Cybernetics (ICMLC).
- [2] Malaria Disease Prediction Based on Machine Learning Octave Iradukunda;Haiying Che;Josiane Uwineza;Jean Yves Bayingana;Muhammad S Bin-Imam;Ibrahim Niyonzima 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP) .
- [3] Malaria Cell Detection Using Evolutionary Convolutional Deep Networks Bin Qin;Yufan Wu;Zhili Wang;Haocheng Zheng 2019 Computing, Communications and IoT Applications (ComComAp).
- [4] The self- upgrading mobile application for the automatic malaria detection Isabel H.J. Song;Wanzin Yazar;Austin Tsang 2020 10th Annual Computing and Communication Workshop and Conference (CCWC).
- [5] Convolutional Neural Network Based Malaria Parasite Infection Detection Using Thin Microscopic Blood Smear Samples Sidharth S. Prakash;Binsu C. Kovoov;K. Visakha 2020 Second International Conference on

Inventive Research in Computing Applications (ICIRCA).

[6] Deep Learning for Smartphone-Based Malaria Parasite Detection in Thick Blood Smears

Feng Yang;Mahdieh Poostchi;Hang Yu;Zhou Zhou;Kamolrat Silamut;Jian Yu;Richard J. Maude;Stefan Jaeger;Sameer Antani IEEE Journal of Biomedical and Health Informatics.

[7] Effective Performance of Bins Approach for Classification of Malaria Parasite using Machine Learning Hrishikesh Telang;Kavita Sonawane 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA).

[8] Unsupervised Deep Learning CAD Scheme for the Detection of Malaria in Blood Smear Microscopic Images Priyadarshini Adyasha Pattanaik;Mohit Mittal;Mohammad Zubair Khan.

[9] Malaria Disease Recognition through Adaptive Deep Learning Models of Convolutional Neural Network Sammy V. Militante 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences.

[10] Malaria Parasite Detection From Peripheral Blood Smear Images Using Deep Belief Networks Dhanya Bibin;Madhu S. Nair;P. Punitha IEEE Access.

[11] Malaria Detection using Deep Learning Gautham Shekar;S. Revathy;Ediga Karthick Goud 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184).

[12] CIDMP: Completely Interpretable Detection of Malaria Parasite in Red Blood Cells using Lower-dimensional Feature Space Anik

Khan;Kishor Datta Gupta;Deepak Venugopal;Nirman Kumar 2020 International Joint Conference on Neural Networks (IJCNN).

[13] Deep Learning Based Approach For Malaria Detection in Blood Cell Images Amogh Manoj Joshi;Ananta Kumar Das;Subhasish Dhal 2020 IEEE REGION 10 CONFERENCE (TENCON)

[14] An in silico Approach to Detect Efficient Malaria Drug Targets to Combat the Malaria Resistance Problem Segun Fatumo;Ezekiel Adebisi;Gunnar Schramm;Roland Eils;Rainer Konig 2009 International Association of Computer Science and Information Technology - Spring Conference.

[14] Smart diagnostic algorithms for automated detection of childhood pneumonia in resource-constrained settings Elina Naydenova;Althanasios Tsanas;Climent Casals-Pascual;Maarten De Vos 2015 IEEE Global Humanitarian Technology Conference (GHTC).

[15] Mobile Microscopic Device to Detect Parasitological Cell-Related Diseases Using Machine Learning Aishah Khan;Ashraf Khalil;Hassan Hajjdiab 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW).

[16] Detecting and Segmenting White Blood Cells in Microscopy Images of Thin Blood Smears Golnaz Moallem;Mahdieh Poostchi;Hang Yu;Kamolrat Silamut;Nila Palaniappan;Sameer Antani;Md Amir Hossain;Richard J Maude;Stefan Jaeger;George Thoma 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR).

[17] Extreme Phenotype Sampling Improves LASSO and Random Forest Marker Selection for Complex Traits Cai John;Wellington Muchero;Scott Emrich 2020 IEEE International

Conference on Bioinformatics and Biomedicine (BIBM).

[18] Extreme Phenotype Sampling Improves LASSO and Random Forest Marker Selection for Complex Traits Cai John; Wellington Muchero; Scott Emrich 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

[19] WHO. Malaria microscopy quality assurance manual-version 2. World Health Organization; 2016.

[20] Performance of Convolutional Neural Network in Detecting Plasmodium Parasites Hanung Adi Nugroho; Julisa Bana Abraham; Aina Hubby Azzira; Eka Legya Frannita; Rizki Nurfauzi; Faza Maula Azif; E. Elsa Herdiana Murhandarwati 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET).