

Malaria Detection using Machine Learning

Shivam Tiwari, Prashant Kumar, Minay Ghai, Priya Gupta, Virandra Pal

Computer Science and Engineering

Greater Noida Institute of Technology, Greater Noida, Uttar Pradesh, India

Dr. A K, T. University, Lucknow

Abstract: *Malaria is a major issue in the hot and mild regions, due to malaria thousands of people lose their lives every year. In 2021, it was found that about 274 million people were affected by malaria and approx 6,27,000 people lost their lives due to bites of Anopheles mosquitoes. This is caused by Plasmodium parasites, which have only a single cell. And most of the cases were found in Africa because Africa is less economically developed and there is a lack of medical facilities like the lack of a microscope for testing the blood smear and a lack of medicine. So it required a well-trained microscopic person who had to count the number of parasites present in the RBC in the blood sample. So there is always a chance of human error to overcome this problem, reduce human error, and increase the test speed. We can use machine learning that needs to be well-trained on whether the blood cells are parasitized or not. to discover that we should use a convolution neural network (CNN). We used various machine learning strategies like image, and map detection features. We will be going to embed our well-trained model in a Simple Board Computer(SBC), which is affordable and can be used in Developing Countries for detecting malaria.*

Keywords: CNN, Machine Learning, SBC, RBC, Parasitized, Auto-encoder

I. INTRODUCTION

Malaria is a medical condition of the person which is caused by a parasite called plasmodium, initially, In [1,2] this parasite is present inside the body of a mosquito which is the primary host of plasmodium, and when a mosquito bite us at that time these parasites enters the human body which is the secondary host of plasmodium, around 2-3 days are required for the parasite to successfully occupy the RBC, and then start to multiply itself inside the RBC, this belongs to the second step, after 2 days of this happen the parasite rupture the red blood cells and release more parasite into the human body, at the same time the red blood cell realizes 'hemozoin' and chillness in body, this belongs to the third step, now there parasite attack on the vital organs in the body which damage more, for repairing it human body use WBC, which means platelets eventually decrease, In [3][4] at this time body suffers with fever if this event is happening in the whole body then the person may die due to malaria, the better solution is it must have proper medication before second step of parasite started .

Our project fits perfectly in this situation; it detects malaria from the patient's body before the outbreak of the second stage. With the help of this, we help many patients, give proper medication on time, and cure them. In [1] Malaria is a worldwide problem where approx. half of the population lives under of risk of malaria in 90 countries and territories. In India things are not good, In 2017 number of cases are 7.47 million but now a day things become better around 3.22 million cases are registered in 2020, In [2,3,4] P. falciparum and P. vivax, two of the most common malaria parasites, are the main causes of malaria in India.

II. LITERATURE SURVEY

A parasite that frequently infects a specific species of mosquito that feeds on humans is the primary cause of the serious, occasionally fatal disease known as malaria. In [1,3] According to a WHO survey, there were 247 million cases of malaria worldwide and 619000 deaths were reported in India; cases were about 196000 and deaths are about 30000 in the year 2022. [2] From this information, it is very evident that malaria is a huge burden to our healthcare system. Although the malaria disease is curable and preventable if detected timely, in some parts of India as well as the World the severity of malaria disease is very high due unavailability of proper testing labs and high diagnosis time. Because the traditional method in which the sample of blood is taken and then tested is time-consuming, requires trained lab

technicians, and is expensive due to which many people infected with malaria go undetected. In today's world, the advancement in technologies like machine learning, deep learning, etc has brought great revolution not only in the technical world but also in the healthcare system.[1,2,3,4] It has been made possible to detect malaria by using some machine learning techniques which give results in real-time at a very low cost by analyzing the cellular image of blood. [2,3] According to research done till now deep learning algorithm for detecting malaria involves five basic steps: processing of the input image to remove unwanted noise, segmentation of the image, feature extraction of an image to identify infected and uninfected cells, classification to label infected and uninfected cells, then showing the output after checking for standard parameters of the parasite.

Firstly, the noise reduction operation on the input cellular image is performed by using techniques like Gaussian filter after this image segmentation process starts to get the required region of the image it is done using two techniques namely Otsu Segmentation and watershed segmentation after the completion of the above two processes features extraction on the image is performed (done is using CNN)[1,3] after this classification begins this is done using both K-Nearest Neighbor and convolution neural network the in the final stage the computed results (count of infected and uninfected RBCs) compared to the normal count of infected and uninfected RBCs. The output is shown to the user.

Although the given created models have shown successful results for detecting malaria this is still not user-friendly (complex) and cannot be used everywhere Unless it is embedded in a system that is specifically designed to execute this model and can be made available in the market in the form of small testing kits in all the pharmacy kits which will be able to provide the malaria diagnosis results in real-time at low cost and also will eliminate the need of lab technicians or labs

III. METHODOLOGY

3.1. DEEP LEARNING FOR DETECTION:

[1,2,4] A parasitized divided into the uninfected and parasitized image. Nowadays machine learning is an important part of life and is used widely across many fields. but we have seen that deep learning is used nowadays due to many features like image classification and video recognition and speech recognition these are trending

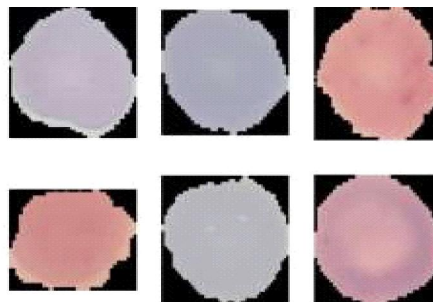


fig.1 (Un-infected)

Defining the above first images as we see in the first image there is (un-infected blood cell from malaria parasite) and it is clearly observe by the modal on the bases of spots on the blood cell.

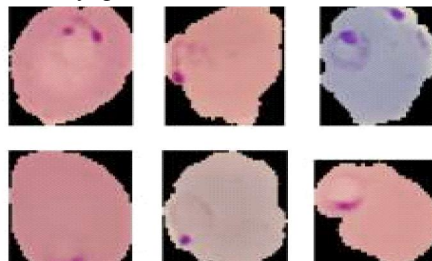


fig.2 (infected)

Defining the above second image which clearly indicates an infected blood cell and modal can observe it on the bases of spots which indicate malaria parasite already attack on the cell and it is infected by malaria parasite.

[2] This project will use a Convolution neural network (CNN), which is a type of deep learning. this is the most famous algorithm used for research in the field of computer vision because the CNN model

3.2 DATA SPLITTING

We have divided our dataset into two parts: the training set and the testing dataset. the ratio can be able to extract hidden information from the model. In this model, we trained our model to classify the image of blood swears. The architecture of the CNN model is described below.

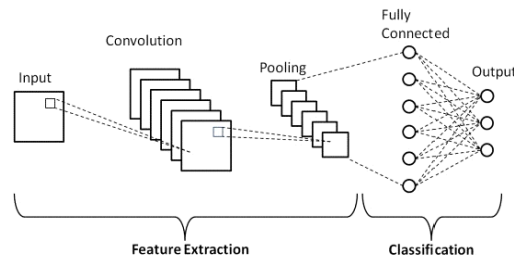


Fig.3 CNN ARCHITECTURE

The above image shows how the cnn modal works on some particular condition.CNNs use convolutional layers to extract features from images, and pooling layers to reduce their size. Fully connected layers classify the resulting features into output classes.

between the training and testing dataset is 80:20 and we are using the machine learning Scikit-learn library which is an open-source library containing a lot of functionality and tools for training and testing the model.

3.3 MODEL PRE-PROCESSING:

[2,4] Supervised learning behavior and performance of the model wholly depend on the that we feed our model so it is necessary to feed the correct dataset. and pre-processing knowledge is important in any experiment. this model works by manually correcting the image according to the input. images patches are resized and map it to among zero to one that varies to the crystal rectifier to obtain a quicker.

IV. DATA SETS

4.1 DATASETS COLLECTION:

We have used the datasets from Kaggle convergence results. This will enhance the model's performance and efficiency.

4.2 MODEL DESIGN:

The auto-encoder has unique features that make websites. and we have used the 27558 images for the training of our model. And 15832 images were used for the validation, and for testing the model. 27558 images are equally input images and reduce the size of the image dimensionality and again produce the desired output. It has unique features that take input and produce the output equally by using the feed forwarding networks. Components of the auto-encoder.

the above flow chart shows how the model used from starting like first we import important libraries to system and then load dataset to train model with configuring dataset and then load that model to user interaction platform system

The workflow is shown in the diagrams connected and forms the forwarding mesh. And code strongly feed acts as a single layer that acts as its own dimension , Encoder and decoder are completely

4.3 ENCODER

Encoder means converting the datasets into the required format. In machine learning, we convert the image into the text or sequence of words into multidimensional vectors that have many hidden layers. The encoder is built by the Recurrent Neural network.

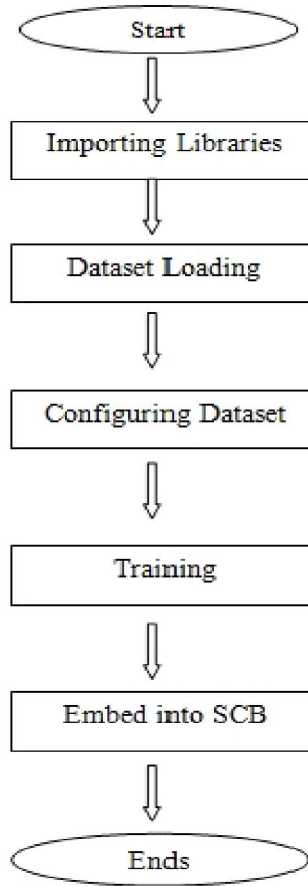


Fig. 4 flow chart of model design

4.4 DECODER:

Decoding means converting the coded data into the original form; the role of the decoder is to convert the multi-dimensional vector into the desired output or original and required image.

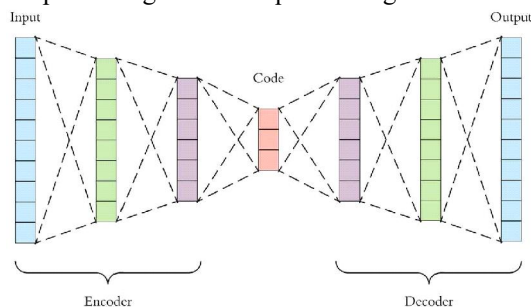


Fig.5 (Auto-encoder)

Auto-encoder is a type of neural network used for unsupervised learning that compresses and decompresses data to reconstruct the input, allowing it to learn efficient representations of the data

V. MODEL TRAINING:

In this process, we have trained the model by performing the following operation.

In the above steps we successfully complete the image processing part now we train our modal so that it can predict the malaria in the image of blood of the patient

This step include the code part which contain some instruction which identify the spots in the blood images than we train you modal with 49,000 images some of them are infected and some of them are un-infected we divided both classes on the ratio of 70% and 30% where 70% is training data and else is test data , to get more accurate we use more data to train our modal .

After that we collect all the similar images and divide them into two levels of infected and uninfected images with the help of SVM(support vector machine). This can be done by creating a line with the help of the nearest two different dataset and drawing a line to separate the data.

VI. RESULT

Taking the image of the blood of the patient with the help electronic microscope there are several process which help to take the image of blood , like with the help of electronics microscope , in every hospitals there is electronic microscope which take images of blood samples and then segmentation operation perform on the image for feature detection and loading it on the modal and performing image processing to identify the individual blood cell and then identify the spots on the blood cell .

Using SVM (Support vector machine) we are trying to make a group of infected cells and uninfected cells from the features of an image which contains a parasite in the blood cell or not then count accordingly.

With the use of CNN, it learns to extract meaningful features from the input blood image by applying a series of convolutional filters and pooling operations. This allows it to classify images with high accuracy. and also we can predict the output on different parameters such us (gender, age, WBC count, and so on) which help the modal to predict the output .

Than we calculate the result to the patient in the form of a report such as blood count, platelet count , and at last predict the malaria of the patient if the report of patient contains malaria then we refer that patient to meet nearest doctor

After using the training and testing datasets we can clearly tell which cell is infected or which is not because our trained model has about 95.70% accuracy which is good enough to predict the result.

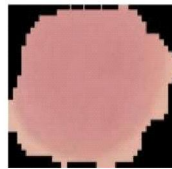


fig. 6 Uninfected cell

In the above image we clearly see that there is not any spot which means there is no parasite attack on the red blood cell which mean the cell is infected

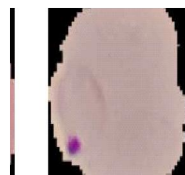


fig. 7 Infected cell

In the above image we clearly see that there is presence of spot which means there is parasite attack on the red blood cell which mean the cell is infected

VII. CONCLUSION:

Malaria is one of the major problems in a developing country where the temperature is hot and mild. It also causes death. Starting symptoms of malaria are headache, pains, vomiting, fevers, body aches, etc. We made a system for Cell Detection in Digital form and developed the trained model with an accuracy of approx 95.70% which is much more accurate to predict any result. We have used CNN which is a very powerful technique for predicting the results. That means that it's a better option for that country where there is not good connectivity. This is a vast area of research and anyone who is interested in machine learning and can do the research work further.

VIII. FUTURE SCOPE

The best part of our project is we take image of blood which contain blood cells and many other things like white blood cells , platelets, substances, proteins , sugar and fat particles.with the help of there things we are not limited to test only malaria but other tests will also be perform like fatigue, chronic infection, anemia, Thalassemia, Lymphoma, Leukemia etc all these reports will also be prepared by machine learning which is fast and more accrue than human

REFERENCES

- [1]. john A. Quinn, alfred Andama, Ian Munabi,"Automate Blood Smear Analysis for Mobile Malaria Diagnosis." , <https://www.researchgate.net/publication/266146329>
- [2]. Rudrasinh Ravalji, Nilay Kumar Shah,"Malaria Disease Detection using Machine Learning", International Research Journal of Engineering and Technology, Volume:07 Issue: 12 , Dec 2020, IRJET
- [3]. Priya Singh, Priyanshu Singh, Rahul Sharma, "Malaria Detection System Using Microscopic Blood Smear Image ." International Research Journal of Engineering and Technology, Volume:09 Issue: 05 , may 2022, IRJET
- [4]. D Ghate, C. Jadhav, and N. U. Rani, A., and S. Busa., "AUTOMATIC DETECTION OF MALARIA PARASITE FROM BLOOD IMAGES,". [Online]. Available: <http://ijact.org/volume4issue1/IJ0410050.pdf>. Accessed: Jan. 25, 2020
- [5]. Ahirwar A., Pattnaik S., Acharya B., Advanced Image Analysis Based System Learning Algorithms and Image Processing. 7(2), 68–70.
- [6]. Mishra, V. K., Kumar, S., & Shukla, N. (2017) "Automatic Detection and Classification of Malarial Parasite in Blood Images." International Journal of Information Technology and Knowledge Management. Jan-June 2012, Volume 5, No 1, pp. 59-64
- [7]. Chari, K., & A. An Impact of incorrect on waterfall software Empirical Software Engineering Agrawal, M. (2018). and new requirements project outcomes.
- [8]. Centers For Disease Control and Prevention , website for information about malaria, <https://www.cdc.gov/malaria/about/faqs.html>
- [9]. WHO Malaria microscopy quality assurance manual- version-2 World Health Organization
- [10]. World Health Organization. Malaria: fact sheet No. WHO-EM/MAC/035/E. World Health Organization , Regional Office for the eastern Mediterranean,
- [11]. Wilson, Michael L, "Malaria rapid diagnostic tests," Clinical infectious diseases 54 no.
- [12]. Gollin, douglas, and Christian zimmermann, "Malaria disease impacts and long-run income differences.".
- [13]. Hinton G, Vinyals O , Dean J, Distilling the knowledge in a neural network.
- [14]. Bengio Y. Rmsprop and equilibrated adaptive learning rates for nonconvex optimization
- [15]. O.A Koita,O.K. Doumbo, A Ouattara ,L.k. Tall, A Konare , M. Diakite , M. Dialle, I. Sagara , G.L. Masinde, S.N. Doumbo , et al False- negative rapid diagnostic tests for malaria and deletion of the histidinerich repart region of the hrp2 gerie.
- [16]. V.V. Makkapati and R.M. rao. Segmentationof malaria parasites in peripheral blood smear images . IEEE International Conference on Acoustics, Speech and signal Processing
- [17]. L.M. Milne, M.s. Kyi, P.L. chiodini, and D.C. warhurst . Accuracy of routine laboratory diagnosis of malaria in the United Kingdom. Journal of clinical pathology
- [18]. M.G. Mubangizi, C. Ikae, A. Spilipoulou, and J.A. Quinn Coupling spatiotemporal disease modaling with diagnosis .In Proceeding of the International Conference of Artificial Intelligence
- [19]. C.K.Murray, R.A.Gasser, A.J.Miller. pdate on rapid diagnostic testing foe malaria. Clinical Microbiology Reviews
- [20]. World Health Organization. Malaria light microscopy : creating a culture of quality . In report of WHO SEARO/ WPRO workshop on quality assurance for malaria microscopy, Geneva
- [21]. A.Roca-Feltrer,I.Carneiro, A Schellenberg, and Joanna R.M. Estimates of the burden of malaria morbidity in Africa in children under the age of 5 years. Tropical medicine and International Health
- [22]. C. Ronse, Set-theoretical algebraic approaches to connectivity in continuous or digital spaces.

- [23]. F.B. tek , A.G. Dempster, and I kale. Malaria parasite detection in peripheral blood images. British Machine Vision Conference
- [24]. F.B. tek , A.G. Dempster, and I kale. Computer vision for microscopy diagnosis of malaria.
- [25]. F.B. tek , A.G. Dempster, and I kale. Parasite detection and identification for automated thin blood film malaria diagnosis. Computer Vision and Image Understanding.