Plasmodium Detection using Machine Learning

Neetha K S, Ram Rushendranath G, Sri Vardhan R, Mani Kumar R, Pramodh K

***Abstract**: Plasmodium is one of India's biggest public health problems. Early prediction of a malaria epidemic is that the secret to malaria morbidity management, mortality as well as reducing the risk of malaria transmission in the community will benefit politicians, health care providers, medical officers, health ministry and other health organizations to better target medical resources to areas of greatest need. In this project, we acquire data sets from hospital databases, which have the information about the causes of malaria, and the images of cells infected with malaria. We then analyze these data sets and feed them to our machine-learning model. Here we are using contour detection and random forest algorithms for training the model and predicting the output*

Keywords: Plasmodium, Miasma, Parasites, Malaria Detection, Image acquisition, RBC count, Stained object detection.

I. INTRODUCTION

The WHO (World Health Organization) survey predicted that malaria could occur in almost 33 hundred million cases. Blood-borne disease, miasma is caused by Plasmodium, red blood cells infected with parasites, and a specific type of mosquito called Anopheles transmits it. A person suffering from malaria will exhibit many clinical manifestations from very mild to severe cases, which could even lead to the person's death. Malaria detection using a microscope is both a time-consuming and difficult process. This traditional method requires the considerable expertise of a microscopes or technician in a laboratory. Experienced microscopic parasites potentially play an important role in the diagnosis of miasma. According to research carried out in and, it is estimated that out of 300-450 million acute miasma diseases 1-3 million are almost lethal. Diagnosis is very complicated in areas most affected by malaria and medications are provided based on symptoms only.   Diagnosing sickness is a significant issue in creating   nations like Uganda, where just 50% of the country's   Well-being focuses have magnifying instruments and only one-fourth of them have prepared lab experts for jungle fever conclusion. Likewise, identifying illnesses at the most punctual with better exactness is significant, as it might help in giving the drug to the analyzed patient at a beginning period. Besides, bogus negatives could bring about the casualty, and bogus positives could cause an increase in superfluous financial weight and medication obstruction. Thus, there is a need to build up an alternate technique for determination.

For the determination of image converter and   computer vision strategies. As of late, Khan and his devoted group proposed another computer vision strategy dependent on the way to deal with distinguish the Parasite from the light This is a pixel-based methodology, which utilizes K-means clustering algorithm for the segment identification of malaria parasite tissue. The parasites present in blood smear are recognized utilizing pictures shot through standard magnifying instrument. Not many different examinations looked considerably further at grouping the various species and the various phases of the parasites life cycle. Picture preparing strategies are as yet being rehearsed on the grounds that, we would prefer not to clear out human specialist's demonstrative procedure totally, however in a specific degree for definite judgment dependent on the blood smear. This procedure will improve the adequacy of lab specialists by serving to triage their fixation and furthermore actualize the intestinal sickness finding over a remote system association.

This paper is based on computerized detection of Plasmodium by discovering and group the infected red blood cells from healthy ones into images of cheap-quality bloodstain. We use classical machine algorithms as well as deep learning algorithms because conventional algorithms fail to process these low-quality images. Thus, our system can detect malaria without any human intervention or at least the system can serve as a helpful medium for technicians to reduce their work   and   also   possibly   increase   the   diagnostic accuracy.[1]

Neeta K S*, Dept. Of CSE, GITAM University, Bengaluru, India. Email: neetasrinath@gitam.edu

Ram Rushendranath Gulla, Dept. Of CSE, GITAM University, Bengaluru, India. Email: ramrushendhranath.gulla@gmail.com

Sri Vardhan Rachamallu, Dept Of CSE, GITAM University, Bengaluru, India. Email: srivardhan.venkat@gmail.com

Mani Kumar Rankireddy, Dept Of CSE, GITAM University, Bengaluru, India. Email: manikumar4580@gmail.com

Pramodh K, Dept Of CSE, GITAM University, Bengaluru, India. Email: pramodhkri30@gmail.com

II. LITERATURE SURVEY

Miasma is caused by the Plasmodium parasite which attacks red blood cells (RBC) and is transfused by mosquitoes. The magnitude of Miasma varies from moderate to extremely serious, which eventually leads to human death. Neural networks have been used in analyzing the possibility of RBC's and parasites in the blood smear . Using the Bayesian pixel classifier, whose purpose is staining pixels,

is trained with the learned features in the weighted KNN (K-Nearest Neighbours) algorithm which is an instance based learning algorithm. In Smear preparation it is important to identify multi-class parasites in terms of their life cycle stage and their forms in Basic thresholding using a histogram-based approach to recognize the presence of Plasmodium in blood smears as differences in these might cause variations, as predominant as imaging conditions.

Using morphological agents, the conflicting RBCs were segregated. Anomalies are identified by examining photographs of blood cells, where the true picture is binarized using a fluid measuring method and then cells are classified in the film. In addition, these labeled cells are classified into platelets, leukocytes, and erythrocytes using an architecture called a hierarchical neural network using certain attributes such as colour, size, and characteristics. The algorithm consists of four stages: to detect the edges, to link the edges, to split the clumps and then to detect the parasite. Pre-processing is an efficient equalization of the histogram used in this algorithm. A color segmentation technique is used to separate the pixels into erythrocytes, parasites, and backgrounds, which are based on classically supervised models. Supervised classification algorithms such as Support Vector Machines, K-Nearest Neighbour, and Naive Bayes were tested using different color models, including RGB (Red Green Blue), uniform RGB, HSV (Hue Saturation Value), and YCbCr, respectively. In recent years, there has been much development of new methodologies for malaria diagnosis, including rapid antigen detection, fluorescent microscopy detection, and method PCR(Polymer Chain Reaction) detecting specific nucleic acid sequences. Given that, the process of diagnosing light microscopy is the most commonly employed procedure. Edges are observed using the matrices of scattered banded filters, and identified using X-ray signal detection in the tumor. Microscopy may discriminate between species groups, measure parasitism and analyze the parasite's multiple asexual stages and. Nevertheless, this procedure requires qualified technicians and it is a time-consuming method and the actual precision of the diagnosis relies on the microscopist's experience and skills, as well as the amount of time spent in learning each slide.

III. METHODOLOGY

In this paper we are doing the survey regarding the machine learning algorithms as well as deep learning algorithms and among them Random Forest algorithms having more accuracy besides performance.Since this algorithm overcomes the overfitting. So the performance increases to 97.50 and it is more accurate than CNN which is a deep learning algorithm and many other algorithms.In Random forest we propose five approaches. In the first approach, we collect the data. In the second approach - we separate images for training and testing. In the third approach we convert training images into gray scale and apply contour detection. In the fourth approach - we apply random forest machine learning algorithm to the training images. Finally we test the model with testing data and check its accuracy. The image data-set used in this paper is captured using oil immersion objective lens from 1000 individuals using 1000x magnification. The images which were out-of-focus, poor quality images and images which were difficult to label the parasite were eliminated. Lastly the data set contains images

of blood smear with boundary boxes of miasma parasites. Then, each image is split into overlapping patches and the patches are labelled as 0 or 1 using the bounding box. Training data-set uses 80% of the labelled data blood smear images, which contains patches where it is annotated as Parasites, and the remaining 20% blood smear images of the data is used for testing purposes. It includes patches as well as non-parasite images annotated as Parasites.

That picture identified as either a 0 or a 1 was broken into patches that overlapped. There are some pixels in each picture patch layer. Using this dataset of picture patches labeled, we address the challenge of defining malaria as a binary pattern. In the classification task the raw formatted pixel data in image patches will not be directly useful. Instead, we use a description that will not be influenced by the speed of localization, rotation, and continuous offsets. In the Plasmodium detection problem the structure of artifacts in the input patches is the main concern. We need to scale the images if they are collected with different sizes and require a representation invariant to representation of the expression, duration, and rotation.

An essential step in the development of the predictive malaria diagnosis system is function engineering. Next, we need to consider a description of the evidence resulting in good performance on plasmodium identification, and then have a general representation of the forms found in the photos comprising blood smear without artifacts such as leukocytes or the different hemoparasites, so that the same method can be used in the future to classify the other related problems.Generally, color information can also be very useful, though it isn't informative when using blood films, which are treated with the field's stain. Therefore, statistical representations of the shapes are used for this task. In general, we have to transform the color patches to gray patches for extraction of the function. Here two types of features are being used: one is derived from connected components, and the other is derived by calculating the moment's patches, thresholded at multiple levels.

Because of these labelled image spots, the diagnosis of miasma may be presented as a classification issue, i.e. classification of either 0 (another disease) or 1 (malaria).We use several Machine Learning algorithms such as CNN ,AdaBoost, Random Forest, Decision Tree, KNN to detect malaria. Random Forest was very good in detecting malaria with accuracy of 0.975. The outcome is assessed in the current work focused on the existence of parasite at patch level and not on increasing the patient's whole picture stage. The individual is considered infected; if a clear patch in the picture sample occurs at least. Because the photos we have for our studies are from malaria-infected individuals, per-patient sensitivity and specificity outcomes cannot be given. This device should be used as a help system, such that technicians will conveniently make the decision.This results in processing the images taken from the microscope for making the expert's attention to focus only on the objects within those images that are more likely of the containing Plasmodium. To this purpose, a particular limit of greater sensitivity is chosen. They use different classification standards for getting different fake positives and negatives.
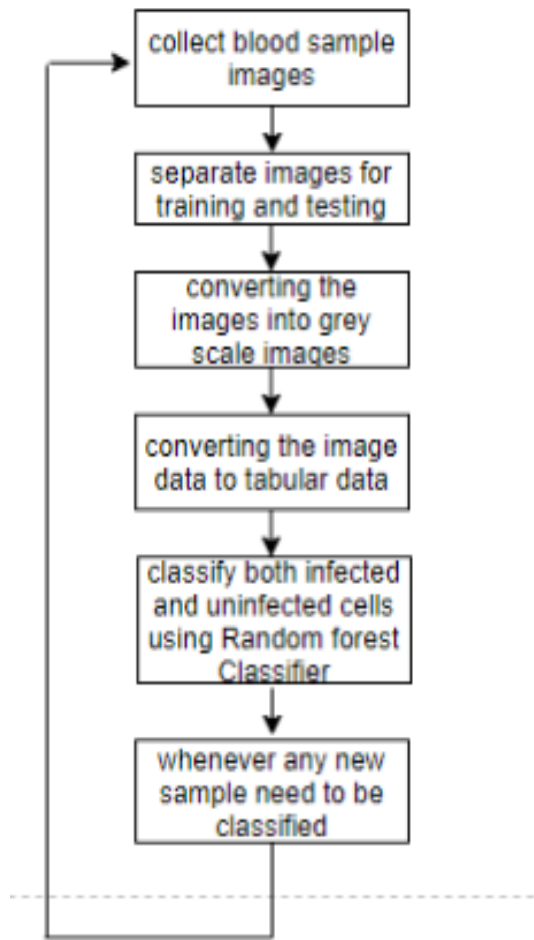
Figure 1: Proposed Framework for applying machine learning algorithms and deep learning algorithms for miasma detection using microscopic images

IV. RESULT ANALYSIS

Table 1: Performance Measure of the proposed work for malaria detection using classical machine learning algorithms and deep learning algorithms.

Algorithm	Accuracy
Ada Boost	0.962
Decision Tree	0.946
KNN	0.940
CNN	0.952
Naive Bayes	0.858
ANN	0.948
SVM	0.928
Linear Regression	0.943
Extra Trees	0.956
Random Forest	0.975

V. CONCLUSION

We have proposed a Malaria parasite detection method using shallow machine learning algorithms and deep learning algorithms. This method of detecting the malaria parasite can be very useful to health workers in countries where there is less number of trained laboratory experts and lack of resources. In the present work, we divided the image into patches and analysed based on the presence or absence of miasmal parasites. To accomplish this, we have used various classical machine learning algorithms and deep learning algorithms such as Ada Boost, Decision Tree, KNN,CNN,SVM, Random Forest, etc.

The accuracy of our model assists the laboratory technicians in decision making, by focusing on the parts of images prone to Plasmodium. Automated Malaria diagnosis also aids for data collection. Furthermore, our extraction of features and classification framework may be sufficiently general for other diagnostic tests like hemoparasites, worm infestations, or tuberculosis

REFERENCES

1. Víctor Uc-Cetina, Carlos Brito-Loeza,and Hugo Ruiz-Piña " Chagas Parasite Detection in Blood Images Using AdaBoost " 2015.
2. Sajana Tiruveedhula , Manda rama narasingarao "Machine learning techniques for malaria disease diagnosis " 2017.
3. Md. Jaffar Sadiq , Dr. V.V.S.S.S. Balaram " Decision Tree based binary Classification using Feature Selection and Optimization for Malaria Infected Erythrocyte Detection " 2017.
4. Sai Prashanth G, Naren Babu R, ArunPriya J, Vinayakumar R, Sowmya V, Soman K P " Performance Comparison of Machine Learning Algorithms For Malaria Detection Using Microscopic Images " 2019.
5. Mahdieh Poostchi, Kamolrat Silamut, Richard James Maude, Stefan Jaeger " Image analysis and machine learning for detecting malaria " 2018.
6. Akshay Nanoti, Sparsh Jain, Chetan Gupta, Garima Vyas " Detection of malaria parasite species and life cycle stages using microscopic images of thin blood smear " 2016.
7. Zhaohui Liang, Andrew Powell, Ilker Ersoy, Mahdieh Poostchi " CNN-based image analysis for malaria diagnosis " 2016.
8. W. David Pan, Yuhang Dong and Dongsheng Wu " Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks" 2017.
9. Sajana Tiruveedhula, Manda rama narasingarao "Classification of Imbalanced Malaria Disease Using Naïve Bayesian Algorithm " 2018.
10. T.F. Michael Raj and S. Prasanna " Implementation of ML Using Naïve Bayes Algorithm for Identifying Disease-Treatment Relation in Bio-Science Text " 2013.
11. Purnima Pandit , A. Anand " Artificial Neural Networks for Detection of Malaria in RBCs " 2016.
12. S Sakthivel , R Thenmozhi " Automatic Malaria Detection Technique Using ANN Classifiers " 2018.
13. Saeid Afkhami , Hassan Rashidi Heram-Abadi "Detection of Malarial Parasite in Blood Images by two classification Methods: Support Vector Machine (SVM) and Artificial Neural Network (ANN) " 2017.
14. Adedeji Olugboja , Zenghui Wang " Malaria parasite detection using different machine learning classifier " 2017 .
15. Emmanuel P. Mwanga, Elihaika G. Minja, Emmanuel Mrimi, Mario González Jiménez, Johnson K. Swai, Said Abbasi, Halfan S. Ngowo, Doreen J. Siria, Salum Mapua, Caleb Stica, Marta F. Maia, Ally Olotu, Maggy T. Sikulu-Lord, Francesco Baldini, Heather M. Ferguson, Klaas Wynne, Prashanth Selvaraj, Simon A. Babayan & Fredros O. Okumu " Detection of malaria parasites in dried human blood spots using mid-infrared spectroscopy and logistic regression analysis " 2019.
16. Aimon Rahman , Hasib Zunair , M Sohel Rahman , Jesia Quader Yuki , Sabyasachi Biswas , Md Ashraful Alam , Nabila Binte Alam , M.R.C. Mahdy " Improving Malaria Parasite Detection from Red Blood Cell using Deep Convolutional Neural Networks " 2019.

AUTHORS PROFILE

Mrs. Neeta K S, working as Assistant Professor in the Department of Computer Science and Engineering, GITAM School of Technology, Bengaluru, India. Currently pursuing a PhD in the area of Machine Learning and AI. She also published around two papers in reputed journals and conferences.

Ram Rushendranath Gulla, pursuing final year B.Tech at GITAM University, Bengaluru. His areas of interest are Machine Learning, Robotic Process Automation. He has certifications on Big Data Analytics and Robotic Process Automation. He has done a project on TNP Interface. Also certified by DELL EMC for Data Analysis certification. He completed his schooling at Loyola Public School and higher education at Sri Chaitanya Educational Institution.

Rachamallu Venkata Sri Vardhan Reddy, pursuing B. Tech Final Year at GITAM, Bengaluru. He has been certified by DELL EMC for his course completion in Big Data Analytics and also by Automation Anywhere for RPA. His areas of interest are Machine learning, and Data analysis. He completed many projects during his course time, some of them are Mapping Bot and Report Card Generation. He completed his schooling Keshava Reddy Educational Institution and higher education at Narayana Educational Institution.

Rankireddy Siva Surya Naga Mani Kumar pursuing B. Tech Final Year at GITAM, Bengaluru. He has been certified by DELL EMC for his course completion in Big Data Analytics and also by Automation Anywhere for RPA. His areas of interest are Machine learning, and Data analysis. He completed many projects during his course time, some of them are Chat Bot and Report Card Generation. He completed his schooling at Narayana Techno School and higher education at Siddartha Junior College.

Pramodh K, pursuing B. Tech Final Year at GITAM, Bengaluru. His areas of interest are Big Data, and IoT. He completed his schooling at Kendra Vidyalaya and higher education at Narayana PU College.