Malaria Parasite Detection using Various Machine Learning Algorithms and Image Processing

Yash Panchori[1], Nikita Agarwal[1], Aashiya Singhal[1], Sudhir Busa[2]

[1]Students of Electronics and Telecommunication Department,
[2]Assistant Professor of Electronics and Telecommunication Department,
Bharati Vidyapeeth (Deemed to be) University
Dhankawadi, Pune, India. 411043

***Abstract* —** *Malaria is mosquito-borne blood disease caused by protozoan parasites of the genus Plasmodium. The Conventional diagnostic tool for malaria is the examination of a stained blood cell of a patient in microscope which is time consuming and dependent on the experience of a pathologist. In this project, an improved image processing system along with different machine learning algorithms for detection of parasites is proposed. On implementation we found the accuracy of the model varying from 85% to 90% for different algorithms. This model has increased the efficiency of malaria parasite detection and minimizes the human intervention during the detection process.*

Keywords — *machine learning, image processing, malaria, Gaussian blur, classification, contour, regression*

I. INTRODUCTION

Malaria is caused by biting of female Anopheles mosquitoes which transmit the protozoan parasites of the genus plasmodium. According to the report of WHO on World Malaria in year 2015, an estimate 3.2 billion people in 95 countries and territories are at risk of being infected with malaria and developing disease. The traditional microscopic diagnostics is not standardized and depends heavily on the experience and skill of the microscopist. Some of the traditional methods for detection are light microscopy, rapid diagnostic test, flow cytometry and many more. It is common for microscopists in low-resource settings to work in isolation, with no rigorous system in place that can ensure the maintenance of their skills and thus the diagnostic quality.

Due to many drawbacks of the traditional method, we have proposed an automated detection method which will detect the presence/absence of parasites in the blood cell. The prerequisite for this model is the microscopic level image of blood cell. The technology used for development of model is machine learning and image processing. Image processing is the processing of RGB image into gray scale and further to black and white image to extract the necessary feature from the image. The image processing is done with the help of Gaussian Blur technique. Machine learning is an application of artificial intelligence (AI) that gives systems the ability to learn automatically and improve from experience without being explicitly programmed by programmers. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The feature used in the detection process of parasites is contour. Contours are the curve joining all the continuous points (along the boundary), having the same color or intensity. A non-infected blood image will only have one contour i.e. the boundary of cell image. On the other side, the infected blood cell image will have more than one contour due to the presence of parasites inside the cell which generate a contour during the image processing.

II. LITERATURE REVIEW

Mahdieh Poostchi et al. [4] wrote a survey paper on image analysis and machine learning methods used for detection and covered the basics, they mention that there is a high prospect for further development, especially using deep learning. In the paper, they discuss the possibility of thick film processing too. Adedeji Olugboja and Zenghui Wang [5] concluded that machine learning algorithms are a very powerful tool used in detection and classification of malaria parasites. Among the 6 algorithms used, they found out that Fine Gaussian SVM performed better in the classification tasks and the Subspace KNN had the best overall performance. Malihi L et al. [2] decided to take on a broader spectrum and used four types of parasites: falciparum, vivax, ovale, and malariae and five classifiers K-NN, 1-NN, NM, SVM and Fisher linear discriminator. Suman Kunwar et al. [3] presented an approach and developed an algorithm for automated malaria detection and quantification of malaria infection. They also developed a strategy to train using machine learning, for detection of malaria with other types of parasite and discussed the scope of increasing the predictive value with results. Lastly,

WHO website [1] provided us with sample images for training and testing to implement the algorithms and check accuracy.

III. IMAGE PROCESSING

The image processing is done with the help of OpenCV. OpenCV is a library of programming functions mainly aimed at real time computer vision. Digital image are the main data sources for this research. The images used in this research are taken from Lister Hill National center for Biomedical Communication part of National Library of Medicine. 27,203 image (both infected and non-infected) were used for development of model. These images are of highly defined and magnified. Image of RBC's that are taken in thin smear are further processed and analyzed for the Malaria Parasite detection. In this approach, instead of box filter consisting of equal filter coefficients, a Gaussian kernel is used. Gaussian filtering is highly effective in removing Gaussian noise from the image. It is achieved with the help of inbuilt function by specifying the width and height of the kernel which should be positive and odd along with the standard deviation in both directions. The blur RGB image is converted into gray scale image. Further on gray scale image a threshold is applied and black and white image is obtained. Using the inbuilt contour function all the contours are marked and top 5 contour areas are retrieved and saved in csv along with the status of the blood cell image which is used as dataset for building the machine learning model.

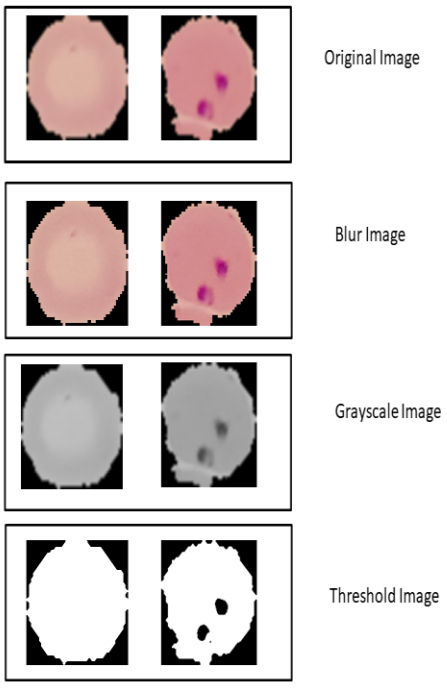

Fig 1. Image Processing

IV. BUILDING MACHINE LEARNING MODEL

Data is generated using the image processing procedure discussed in above chapter. The data consists of five maximum areas of the contour along with the status. The data is visualized using the pandas library. Later on the data is cleaned by removing the null value, infinity value generated due to some error. The dataset is divided into training and testing data at random. The ratio of training and testing data is 4:1.The training data is used to train the model whereas testing data is to evaluate the model. The classification algorithms used are Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Support Vector Machine and Bagging Classifier.

A. Logistic Regression

It is the regression analysis conducted when the dependent variable is binary i.e. present or absent. It is used to describe data and to explain relation between the dependant variable and the rest of independent variable.

B. Decision Tree Classifier

This approach is used for the classification purpose as well as regression purpose. In this method a set of training examples is broken down into smaller and smaller subsets while at the same time an associated decision tree gets incrementally developed. At the end of the learning process, a decision tree covering the training set is returned.

C. Random Forest Classifier

This approach consists of a large number of individual decision trees that operates as an ensemble. Each tree in forest results out a class prediction and the class with the most votes becomes our model's prediction.

D. Support Vector Machine

Support Vector Machine is a supervised learning algorithm in which the dataset is plotted using the independent variables and the dependent variable. The points plotted are classified into two groups. A support vector machine outputs a map of the sorted data with the margins between the two as far apart as possible.

E. Bagging Classifier

It is also known as bootstrap aggregating. It is a machine learning ensemble meta-algorithm which was designed to improve the stability and accuracy of machine learning algorithm. Avoidance of overfitting problem and reduction in variance can be done with the help of bagging classifier.

V. RESULTS AND DISCUSSION

The end goal of the project was to create an automated decision support system using machine learning algorithms to identify malaria parasites in blood smear images. Images were processed and 5 contour areas

were obtained for evaluation. The table shows the different algorithms used and their accuracies.

We observed that in SVM, parameters – regularization and gamma were varied and an output was obtained ranging from 84% to 87%. When kernel was changed, the accuracy varied from 10% to 87%. Highest accuracy was obtained using random forest classifier.

Table I. ACCURACIES

Algorithm	Accuracy(%)
Logistic Regression	89.74
Decision Tree Classifier	86.00
Random Forest Classifier	90.05
Support Vector Machine	87.24
Bagging Classifier	86.44

VI. CONCLUSION

Overall, the algorithms that showed the best output are random forest and logistic regression. The least accuracy was seen when Decision Tree Classifier algorithm was used. Logistic Regression shows best results for binary classification for a condition i.e. - yes or no.

ACKNOWLEDGEMENT

The team members would like to acknowledge the efforts of our project guide, Prof. S.S.Bussa. His continuous efforts have propelled us to achieve our ideas. We would like to thank our college: Bharati Vidyapeeth University, College of Engineering, for providing us with the infrastructure to perform the project.

REFERENCES

[1] WHO, "fact sheet: World malaria Report 2016," in World Health Organization, world Health Organisation, 2016. [Online]. Availablehttp://www.who.int/malaria/media/world-malaria-report-2016/en.

[2] Malihi L, Ansari-Asl K, Behbahani A. Malaria parasite detection in giemsa-stained blood cell images. In: 8th Iranian conferenceon machine vision and image processing (MVIP). IEEE.

[3] https://www.researchgate.net/publication/322819026_Malaria_Detection_Using_Image_Processing_and_Machine_Learning

[4] Mahdieh Poostchi, Kamolrat Silamut, Richard Maude, Stefan Jaeger, George Thoma, Image analysis and machine learning for detecting malaria, Translational Research (2018), TRSL 1210 S1931-5244(17)30333-X.

[5] Malaria Parasite Detection Using Different Machine Learning Classifier -Adedeji Olugboja, Zenghui Wang - Machine College of Science, Engineering and Technology, University of South Africa, Florida- 978-5386-0408-3/17/$31.00 ©2017 IEEE