2nd International Conference on Computational Sciences and Technologies, 17-19 Dec 2020 (INCCST 20), MUET Jamshoro

Malaria Disease Detection Using Machine Learning

Usha Kumari
Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro,Pakistan
kumariusha1638@gmail.com

Sanam Narejo
Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro,Pakistan
sanam.narejo@faculty.muet.edu

Mashal Afzal
Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro,Pakistan
mashalafzalmemon@gmail.com

Madeha Muzafar Memon
Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro,Pakistan
madeha.memon@gmail.com

***Abstract*—Malaria, which affects millions of lives per year, is a contagious disease. Standard Laboratory Malaria diagnosis includes an expert and thorough examination of the red blood cell (RBCs) that are healthy and sick. To prevent this form of error, this study aims to develop a computer-based approach for automatically detecting malaria parasite through Image Processing and Machine Learning (ML) techniques. By using different features and ML model, we have found the result of parasitic and non-parasitic images of malaria cell. The trained model gave the accuracy of 97.93%. Moreover, ROC curve shows the detecting rate of Support Vector classification model.**

Keywords— Support Vector Machine, Discrete Wavelength Transformation, Gray-level Co-occurrence Matrix, Histogram of Oriented Gradients, Local Binary Pattern

I. INTRODUCTION

Malaria, a life-threatening mosquito-borne disease, causes fever, vomiting, headaches, and exhaustion and can cause severe cases, lead to coma or even death. People and animals can be affected by this condition. This infection normally spreads by a female Anopheles mosquito. Malaria is caused by one cell microorganism of the Plasmodium genus, that is capable of infecting humans by five of their organisms. Amongst these species, P. falciparum is the most deadly; P. vivax, P. oval, P.knowlesi and P. malariae are among the other species [1].

In tropical and subtropical areas, malaria is prevalent, especially in Latin America, Sub-saharan Africa and Asia. Among 13 out of Bangladesh's 64 districts, the disease is considered endemic and presents a risk to about 14 million people. In 2016, around 731,000 deaths were confirmed worldwide due to malaria, 90% of the deaths were in Africa [2]. Microscopic exams of blood cells using blood video are usually identified with malaria. Approximately, 167 million blood films were screened using microscopy during 2010 for malaria, which is cheaper and reactionless-based diagnostic as compared to polymerase chain. While, it is commonly used, but still there are disadvantages of microscopic diagnosis discussed as follows. Becoming commonly related to suffering in low economies, malaria is not normal research equipment in most labs or medical facilities [3]. Furthermore, the diagnosis relies on the capacity of the person who tests the blood film and the parasite level. The monotonicity of the test greatly influences the quality of the test. Particularly, if there are a number of specimens in the lot. Global pathologist shortages typically have a major effect on developed countries' healthcare systems, and the case of malaria is not different [4].

Malaria is a life-threatening tropical disease with a worldwide impact. Over 200 million infections occur in a year among 100 countries and continuing malaria transmission. In 2015, 212 million cases of malaria were reported and 429,000 deaths from malaria were recorded by the World Health Organization, most of them were children in Africa [5]. Data on malaria infection in 87 countries have been reported by the World Health Organization 2016, 219 million malaria cases were reported in November 2017 and 217 million cases in 2016. It was also estimated that the number of malaria deaths in 2017 will be 4 35,000.

II. RELATED WORK

The diagnosis of malaria can be detected by observing a drop of the patient's blood under a microscope, which is opened on a slide as a "blood smear." Correctness, trust and speed are central to the management of diseases. According to the diagnosis of malaria, many experiments were undertaken to transfigure the manual interpretation in detail. In this regard, this section discusses a number of these reports.

In paper [6] the author established diagnostic methods in the literature which rely on and require instructions from trained operators. The documented approach prevents problems related to "rapid diagnostic" approaches, as they are species-specific and high per test expense, compared with other diagnostic techniques. This proposed process is entirely automated. The method has been developed especially for operating on low-resolution images and hence reduces the reliance on high-resolution microscopy that improves the practicality of the algorithm

In paper [7], the author proposed a study to use visual image or color photograph of microscope-based stem malaria blood to test the presence of RBC parasitaemia in a new technique to identify parasites in this area. The goal of this research is to automatically interpret the blood sample slides microscopically more precisely and quicker than ever.

In article [8], an improved Image Processing system along with different Machine Learning algorithm for detection of parasitic is proposed. On implementation, the accuracy of the model was found varying from 85% to 90% for various algorithms. The dataset utilized in this research is

ISBN-978-969-23372-1-2

taken from LISTER HILL NATIONAL center for biomedical medicine 27,203 image (both infected and uninfected) were used for the development of the model.

III. METHODOLGY

The lead of one phase works as the entry into the next phase sequentially. This means that every phase of development begins only when the preceding phase is complete. The waterfall system is a constant growth.

This study is much better than the previous one because without needing any laboratory instruments and proper lighting conditions, it can be worked

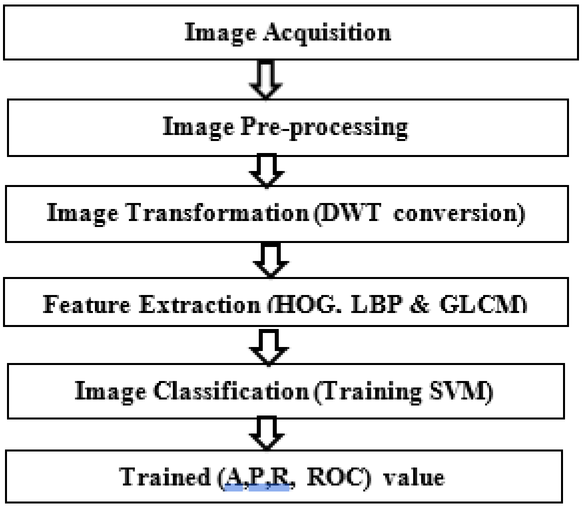


Fig. 1. Flowchart of Malaria cell

In this section, considering limitation from previous studies, an approach is designed and discussed for automatic Malaria Detection.

For this purpose, we have followed the Waterfall methodology. The Waterfall Model is the earliest SDLC technique of software design [10]. In Waterfall method, the entire software design procedure is divided into various technique where progress is seen as continually flowing downward (such as waterfall) in its layout. The software design method is thus also called the "linear sequential life cycle model" [12]

Furthermore, the whole implementation is performed using MATLAB 2019b software to perform various steps which are shown in Fig. 1.

A. Data Acquisition

For the very first step of this study, the dataset is gather images from the “KAGGLE” website [11]. The optimal images from the dataset are then converted into a numerical form which can be later manipulated on a computer. To train the system for distinguishing Parasitic or infected RBC cell from non-parasitic or uninfected ones, a proper dataset with a useful number of records is required. This dataset contains two portions. One is infected and other is uninfected and the total images in the dataset for Malaria cell are 27,558.

Fig. 2. Actual Image ,Fig,2(b) Resized Image

Fig. 2. Malaria cell. (a) shows original image of malaria cell from dataset. (b) shows resized image for the reference

B. Image Pre-processing

Image pre-processing is an elementary step to convert an image into digital form. In Image pre-processing, firstly it takes an image as input then all the suitable preprocessing operations are performed such as image sharpening, image filtering, and image enhancement [13]. In this research, we have used multiple techniques like image enhancement, feature extraction, image classification for preprocessing, enhancing, compressing, and reconstructing Malaria cell image.

C. Image Transformation using DWT Conversion

By using the image transformation [15], we can easily detect the infected area or pointed area. This technique is used to transfer the image original into Grayscale.

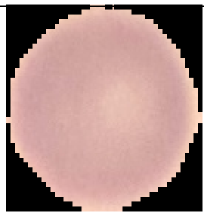
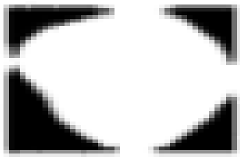
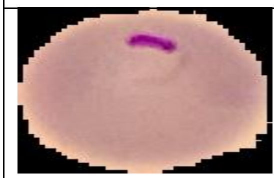
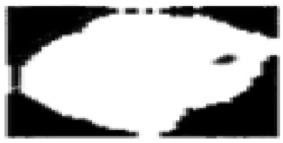
	
(a) Original Image	(b) Grayscale Image
	
(c) Original Image	(d) Grayscale Image

Fig. 3. Infected and non infected cell a)show the original image of non infected image.(b) show the conversion of non- infected cell.(c)show the original image of infected cell (d)show the conversion of infected cell

Changing the size of an object is called transformation and image transformation is a method to convert an image from one empire to another empire. The db4 wavelet divided into four parts that are i) Low-Low ii) Low-High iii) High-Low iv) High-High. Mainly two wavelets low-low and high-high are used because only Low-Low and High-High have total information about wavelet.

D. Feature Extraction

In this step, multiple features are extracted using different techniques to differentiate between infected and non-infected images. For this purpose, we have used GCLM [14], LBP [16] and HOG [17] techniques.

1) Gray Level Co-occurrence Matrix (GLCM)

By using GLCM we first convert RGB image to HSV. Afterwards, it is necessary, to scale HSV matrix to value between o and 64. This does that co-occurrence matrix be computational possible. After this, we can calculate the co-occurrence matrix, for H, S AND V matrices. Thus, we will have three co-occurrence matrix, and we can now determine the parameters (entropy, variance and RMS), for each one of the matrices. It is necessary to establish the correlation between parameters, for determining which of this are relevant.

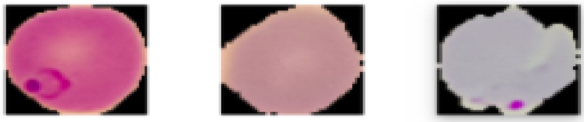


Fig. 4. Segmented cells after transformation to gray-scale a) Original b) Green channel c) Saturation (from HSV)

2) Histogram of Oriented Gradients (HOG)

The HOG is a function descriptor used for object detection purposes in computer vision and image processing. In localized portions of an image, the technique counts the occurrence of gradient orientation.

3) . Local Binary Pattern (LBP)

For any given pixel, LBP may be a local definition of the picture supporting the neighborhood. The neighborhood of a pixel is given within a radius of R within the form of P number of neighbors. It's a strong descriptor that detects within the image all the possible edges.

E. Image Classification using SVM

Support Vector Machine transforms linear issue into higher dimensional feature space, thus providing linear classification for sensitive, non-linear problems detection of the infected and non-infected parasite, without increasing algorithm complexity. It takes labelled dataset, and outputs an optimal hyperplane which is a line dividing the plane into two portions with each class on either side, thereby categorizing new records. Firstly, it takes a linearly distributed dataset and adds several samples in the anyone of the categories so that, the samples contained on either side of hyperplane exhibit a significant gap in between representing two prominent categories. When a new entry is encountered, it is checked against the margin (distance between hyperplane/decision boundary and data point) of each data point. It is then classified into its respective category.

F. Trained (A,P,R, ROC) value

Precision-Recall Curve is a useful performance measure to evaluate the success of a machine learning model. It represents the skill of the model at predicting true class . Sensitivity validation results, also favour Support vector Machine and the accuracy achieved in Precision-Recall was also higher.

IV. RESULTS

Using the test results achieved from the application of pre-defined Machine learning algorithms and analyzing the factors that play a part in making the system perform better, we have implemented those Machine learning algorithms on the described dataset, which comparatively gave reliable results among all. Furthermore, this model has provided more accurate distinction among parasitic and non-parasitic malaria cell has been deployed.

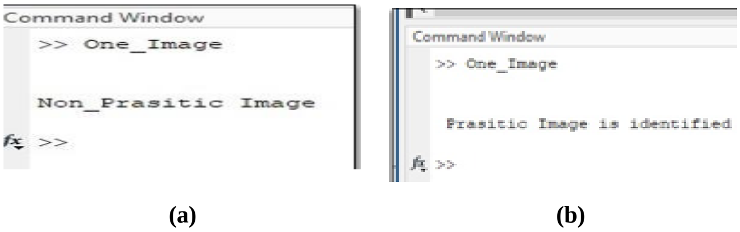


Fig. 5. show the result of non-parasitic image (b)show the result of parasitic image

A. Comparative Analysis Discussion

After visualizing the performance of each of the algorithms, they have been comparatively analyzed against one another.

TABLE I. TABLE 1: FACTORS WITH THEIR PERCENTAGE RESULTS

Factors	Percentage
Accuracy	97.9314
Recall_1	97.8963
Sensitivity	97.9666
Specificity	97.8963
Precision	97.8955
FPR	-96.8963
F1_Score	9.7931e+03
Error	2.0686

From the above results, it can be seen that the performance of Support Vector Machine is commendably better in terms of accuracy and classification report and the number of errors made is comparatively lesser.

By using different Feature and using ML model we have found out the result of parasitic and non-parasitic images of Malaria cell with the accuracy of 97.93%.

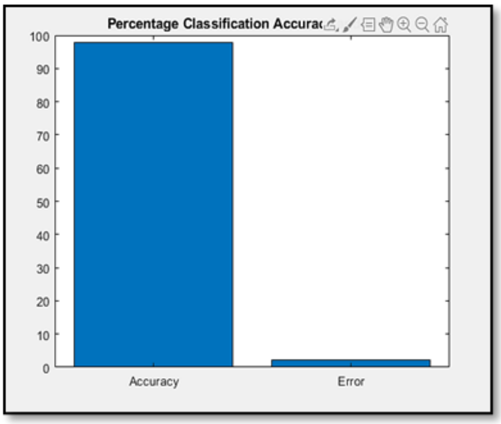


Fig. 6. Percentage classification Accuracy and Error

B. Receiver Operating Characteristic Curve (ROC Curve)

The purpose of Roc is to show the detecting rate of Support Vector classification model. In the below diagram, the horizontal line shows the result of the false-positive rate of the curve whereas the vertical line shows the detecting rate of the curve. This curve shows the Accuracy of our model is 0.99.

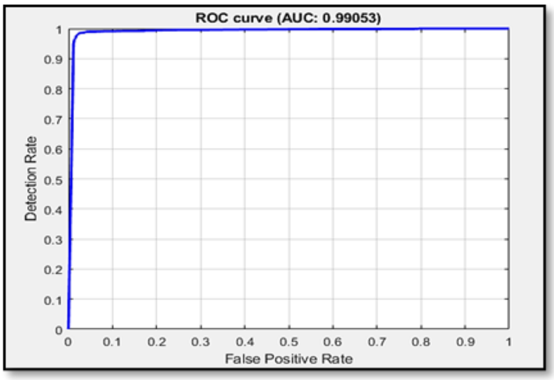


Fig. 7. show the Roc curve

V. CONCLUSION

The main objectives of our project was to propose an automatic method for the detection of Parasitic and Non Parasitic cell by using image processing and Support Vector Machine algorithm. From the observation that has been made, we concluded the Malaria cell, Parasitic and Non-Parasitic recognition and processing is precise, accurate, and the proposed work is user-friendly to use. Moreover, our technique provides better display results using these techniques.

Machine learning, which is becoming a desirable approach for solving most of the complex real-world problems, is being chosen for detecting Malaria disease as well. Thus, our project has made the use of Machine learning models to serve the cause.

REFERENCES

[1] Caraballo, H., & King, K. (2014). Emergency department management of mosquito-borne illness: malaria, dengue, and West Nile virus. Emergency medicine practice, 16(5), 1-23.

[2] Rahman, A., Zunair, H., Rahman, M. S., Yuki, J. Q., Biswas, S., Alam, M. A., & Mahdy, M. R. C. (2019). Improving malaria parasite detection from red blood cell using deep convolutional neural networks. arXiv preprint arXiv:1907.10418.

[3] Wilson, M. L. (2012). Malaria rapid diagnostic tests. Clinical infectious diseases, 54(11), 1637-1641.

[4] Chaity, A. Z. (2017, December 13 Published). Bangladeshis flock to Indian, Thai hospitals in huge numbers. Dhaka Tribune, Retrieved from: https://www.dhakatribune.com/feature/health-wellness/2017/11/30/doctor-trustbangladesh

[5] Nadjm, B., & Behrens, R. H. (2012). Malaria:: An update for physicians. Infectious Disease Clinics, 26(2), 243-259.

[6] Purwar, Y., Shah, S. L., Clarke, G., Almugairi, A., & Muehlenbachs, A. (2011). Automated and unsupervised detection of malarial parasites in microscopic images. Malaria journal, 10(1), 364.

[7] Raviraja, S., Bajpai, G., & Sharma, S. K. (2007). Analysis of detecting the malarial parasite-infected blood images using the statistical-based approach. In 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006 (pp. 502-505). Springer, Berlin, Heidelberg.

[8] Panchori Y., Agarwal N., Singhal A., and S. Busa. (2020). Malaria Parasite Detection using Various Machine Learning Algorithms and Image Processing. 7(2), 68–70.

[9] Mishra, V. K., Kumar, S., & Shukla, N. (2017). Image acquisition and techniques to perform image acquisition. SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, 9(01), 21-24.

[10] Chari, K., & Agrawal, M. (2018). Impact of incorrect and new requirements on waterfall software project outcomes. Empirical Software Engineering, 23(1), 165-185.

[11] Arunava. (2018, December 05). Malaria Cell Images Dataset. Retrieved December 11, 2020, from https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria

[12] Balaji, S., & Murugaiyan, M. S. (2012). Waterfall vs. V-Model vs. Agile: A comparative study on SDLC. International Journal of Information Technology and Business Management, 2(1), 26-30.

[13] Kunwar, S. (2018). Malaria Detection Using Image Processing and Machine Learning. http://arxiv.org/abs/1801.10031

[14] Habe, T. T. (2017). Feature Extraction of Malaria. 4(12), 997–1001.

[15] Mohideen, S. K., Perumal, S. A., & Sathik, M. M. (2008). Image De-noising using Discrete Wavelet transform. International Journal of Computer Science and Network Security, 8(1), 8–11.

[16] SONG, K.-C., YAN, Y.-H., CHEN, W.-H., & ZHANG, X. (2013). Research and Perspective on Local Binary Pattern. Acta Automatica Sinica, 39(6), 730–744. https://doi.org/10.1016/s1874-1029(13)60051-8

[17] Surasak, T., Takahiro, I., Cheng, C. H., Wang, C. E., & Sheng, P. Y. (2018). Histogram of oriented gradients for human detection in video. Proceedings of 2018 5th International Conference on Business and Industrial Research: Smart Technology for Next Generation of Information, Engineering, Business and Social Science, ICBIR 2018, 172–176. https://doi.org/10.1109/ICBIR.2018.8391187