

Malaria Detection Using Image Processing With Random Forest Algorithms

Khushbu Tamrakar¹, Akshat Khaskalam²

¹Dept of CSE

²Asstt. Prof, Dept of CSE

^{1,2}Takshshila Institute of Engineering & Technology,
Jabalpur, Madhya Pradesh, India.

Abstract- Malaria is a serious disease that can be fatal. An infected *Anopheles* mosquito bite is the most common way to contract the disease. Infected mosquitoes carry plasmodium parasites. Malaria is transferred when a mosquito bites an infected person and then bites a non-affected person. When parasites get mature, they travel to the liver and enter the blood-stream, where they begin to infect red blood cells after a few days. Therefore, there are several benefits of using an automated malaria detection process. It provides much higher accurate results compared to the manual process. It reduces the workload to serve more patients. Deep learning has that much potential to provide an accurate and faster result for parasites. Then it was tough for an inexperienced human to detect the disease correctly. But now researchers have got a tremendous result on this. Computer-aided diagnosis is the most powerful tool in this area. Hence in our research work we have tried to solve all these issues more accurately and precisely, hence developed a more accurate model for parasite detection as well as classification. The proposed model includes image processing and random forest method.

Keywords- Malaria prediction, machine learning, random forest, image processing, deep learning, early prediction, classification

I. INTRODUCTION

Malaria seems to be the most common disease in tropical and subtropical regions, where the parasites proliferate. According to the most recent World Malaria Report, estimated 241 million malaria cases in 2020, slightly higher compared to 227 million in 2019. It was estimated that Malaria killed 627,000 individuals in 2020. Children under age 5 are in the most dangerous position. In 2020, 80% of children worldwide died from malaria [1]. Plasmodium parasites are the cause of malaria. Humans can be infected by five different malaria parasites. These are *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. Malaria signs typically appear 10 to 4 weeks after the infection has occurred. It can take several months for symptoms to appear in some cases. Some common signs of malaria disease include moderate to severe shivering chills,

high fever, muscle pain, profuse sweating, headache, vomiting, bloody feces, abdominal pain, diarrhea, nausea, and convulsions [2]. Because of the loss of red blood cells, malaria can induce jaundice (yellow skin and eyes) and anemia. If not treated promptly, the infection can cause kidney failure, mental confusion, coma, convulsions, and death [3].

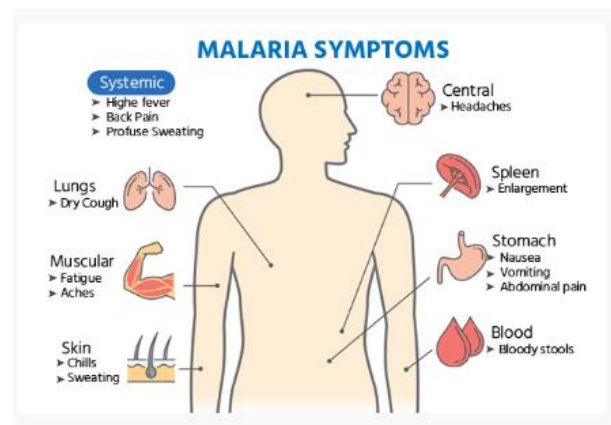


Figure 1: Some symptoms of Malaria.

Malaria detection framework that replicates the conventional gold standard diagnosis of malaria to identify malaria parasites in blood smears images. An automated diagnostic system can significantly improve the efficiency of the pathologists and also reduce the need for dedicated pathologists in rural villages [4][5]. In microscopy diagnosis, peripheral blood smear images are stained using a popular stain called Giemsa to perform efficient visual identification of malaria parasites. The staining procedure slightly colors the healthy RBCs but highlights parasites, platelets, artifacts and white blood cells (WBCs) with a deep purple color. To identify malaria infection, it is necessary to divide the purple-colored blood cell components into two categories using visual cues: parasite-infected blood cells and non-parasitic blood cell components such as platelets, artifacts and WBCs [6]. In undeveloped nations, disease diagnosis is a significant problem in rural areas for diagnosing malaria. Additionally, it is essential to detect diseases earlier and with greater precision because doing so may facilitate the delivery of medication to the patient who has been diagnosed[7]. False positives, on the other hand, could result in an increase in drug resistance and

an unnecessary burden on the economy [8]. As a result, a new method of diagnosis must be developed. The K-means clustering algorithm is used in this pixel-based method to identify segments of malaria parasite tissue. The machine learning algorithms received sufficient training data in [9]. Images taken with a standard microscope are used to identify the parasites in a blood smear. The clustering of the various parasite species and life stages was the focus of a few additional studies [10]. The motivation behind the work is to early diagnosis of malaria is key to successful treatment and death prevention, hence leading to reduced malaria transmission, and hence reducing the death rate due to malaria.

II. RELATED WORK

In Ref. [11], the author proposed a conventional image processing method. Red blood cells are segmented using the Otsu binarization technique. He extracted statistical features manually from the images. To cluster image values, K-means clustering is applied. Support vector machine (SVM) is used as a classifier to classify the disease.

In Ref. [12], the author showed a variety of approaches for malaria diagnosis. Image processing, cell segmentation, feature extraction, and classification can all be done by a variety of techniques. Preprocessing methods he mentioned mean filter, Laplacian filter, wiener filter, median filter, adaptive histogram equalization, contrast enhancement, etc. Segmentation techniques are morphological operation, Hough transformation, k-means clustering, Fuzzy segmentation, etc. Features that are mentioned in his paper are a color feature, texture feature, and morphologic feature. Nowadays, the most often used approach is the convolutional neural network (CNN). It has the potential to extract features from images automatically.

In Ref. [13], the author proposed a fast Convolutional Neural Network of six convolutional layers, one fully connected layer, and one classification layer. He also evaluated different transfer learning techniques such as Alex Net, VGG16, ResNet50, DenseNet121. Finally, he provided an overall accuracy of 96.7% for all models using the test dataset. Transfer learning is a way of using pre-trained networks. He found an accuracy of 95.9% using the transfer learning model.

In Ref. [14], the author developed a shallow CNN model that performs similarly to the VGG16 and ResNet50. In deep learning, high computational complexity and high computational cost is a big issue. His proposed custom CNN has lower computational complexity and lower computational run time. Evaluation metrics used in his research are accuracy,

sensitivity, specificity, F1 score, and MCC (Mathew's Correlation coefficient). He found an accuracy of 95.32% in his research.

In Ref. [15], the author proposed a three-stage pipeline that is segmentation, crop & mask, and classification. In the segmentation stage, a segmentation neural network (SNN) has been built for red blood cell (RBC) segmentation. He got an overall 93.72% accuracy from his custom CNN architecture.

In Ref. [16], the author introduced a new approach to classify malaria disease that is CNN-SVM and CNN-KNN. In this approach, CNN is employed as a feature extractor and SVM/KNN is applied as a classifier. In addition, he demonstrated a new autoencoder technique. Autoencoder is a form of neural network that is trained to regenerate output from its input.

In Ref. [17], the author proposed a custom CNN architecture. Bilateral filter & data augmentation techniques were applied as data preprocessing steps. The preprocessed input images were then fed into a custom-built CNN model. He found an accuracy of 96.82%. In Ref. [18], the author introduced a unique CNN architecture called Attentive Dense Circular Net (ADCN) in their research. ADCN was inspired by ResNet and DenseNet. There are three parts to the ADCN architecture. In the first part, there is a custom CNN model with two dense blocks and one attention module. The dense block has six dense conv blocks that consists of two convolutional layers. Every dense conv block is connected with another dense conv block. Finally, an attention module is split into two parts. One is a 1x1 convolutional layer, while the other is an attention branch for extracting the attention feature map. To ensure that the branch's output size is consistent, the attention branch uses downsampling and upsampling techniques.

In Ref. [19], the author proposed three different techniques that are custom CNN network, transfer learning as VGG16, and CNN-SVM[20]. Then he used the ensemble learning method to reduce prediction variance and generalization error. The results from each of the three networks are combined, and the final prediction is calculated using a weighted average ensemble.

In the paper [21], author's compare the machine learning algorithms namely KNN, Decision Tree, Logistic regression and Random forest and implementing transfer learning with deep learning models VGG19, modified Resnet50 to improve the accuracy achieved with machine learning models thus proposing the best model for predicting

malaria only by observing by blood cell image rather than doing any staining of blood, thus reducing any expert requirement.

In study [22] utilizes a color-based segmentation to detect the parasite in three color spaces, i.e. RGB, YCbCr, and Lab. Furthermore, image-based object counting is introduced to count the number of parasites. The best results are achieved by color-based segmentation in RGB color space. In total, there are 27,558 images consisting of infected (13,779) and uninfected (13,779) blood smears. The threshold values are set to 105–240, 10–80, and 75–175 for R, G, and B channels. Overall the accuracy, sensitivity, and specificity achieved using this scheme is 94.75%, 96.02%, and 93.47%.

III. PROPOSED SYSTEM

Malaria is a severe infectious disease transferred through the bite of an infected mosquito caused by a blood parasite of the genus Plasmodium. In the proposed work, our approach primarily focuses on image processing techniques to process and performing feature extraction, as well as machine learning techniques for the final classification of feature space. The proposed system have computed color features using a novel Bins Approach and further, these images are classified into parasitized and uninfected cells by applying machine learning classifiers such as Linear SVM & Random Forest Algorithm (RF). The performance of the algorithms is evaluated and compared using different parameters like accuracy, precision, recall, and F1-score.

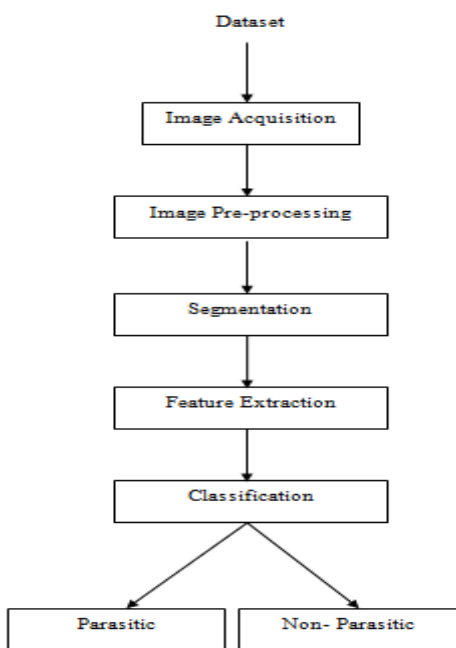


Fig 1: Proposed architecture

Following are important consideration of the proposed system:

- The proposed system uses the dataset provided by ‘National Institutes of Health’ consisting of 27558 RGB cell images in .png format.
- The input RGB cell image is first processed to remove noise from the image.
- The pre-processed image is then segmented to extract the Region of interest(ROI) and we get the segmented image.
- Input to the feature extraction stage is the segmented image and output will be the feature vectors.

Next stage is classification stage where the input will be the feature vectors and output is the classified label as parasitic and non-parasitic.

IV. RESULT

The dataset used in our study contains Giemsa stained image of thin blood smears obtained from a malaria screening campaign in 50 healthy patients and 150 Plasmodium parasites infected patients. It is taken from the National Institute of Health (NIH). Images in a dataset are manually annotated by slide reader experts of Mahidol Oxford Tropical Medicine Research Unit Bangkok, Thailand, and collected at the National Library of Medicine (NLM). The dataset contains 27558 images with the equal occurrence of infected and uninfected red blood cell images as shown below:

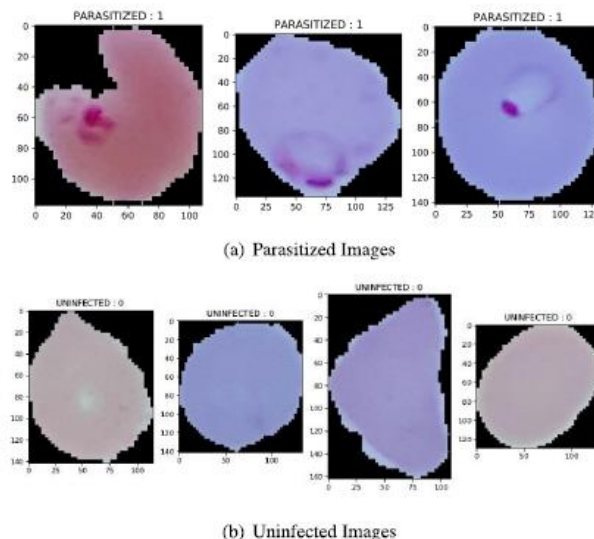


Fig 2: Sample dataset image

Now we are going to compare the results of all the classifiers in the tables below. Table 6.1 below represents the comparisons of accuracy for base classifiers and proposed classifiers. From the results it is clear that the proposed

Random forest classifiers performed well as compared to SVM and Neural Network for malaria parasite classification.

Table 1 Comparison of Existing and Proposed Model.

Algorithm Implemented	Accuracy in (%)
Support Vector Machine	93.1
CNN	94.0
Proposed Classifier	96.3

Graphical chart is shown below:

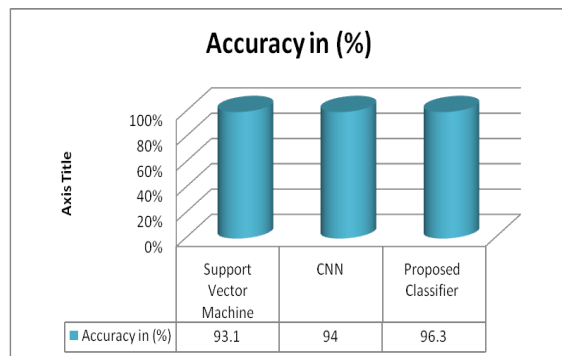


Fig 3: Comparison on accuracy

V. CONCLUSION

This thesis proposed a model that will play a major role in assisting the individuals as well as doctors to make a decision on the best treatment to be given on their hospitals located depending on their location and environment. This system considers various parameters and factors, to predict accurate disease based on symptoms and images. The proposed model provides an advantage by using the accuracies of proposed Random Forest algorithms and gives a resulting accuracy which is higher than the accuracy of the existing techniques.

REFERENCES

- [1] Malaria, World Health organization (WHO), Retrieved from, <https://www.who.int/news-room/fact-sheets/detail/malaria>, 2021. (Accessed 25 January 2022).
- [2] D. Burke, Malaria: causes, symptoms, and diagnosis. Health line, Retrieved from, <https://www.healthline.com/health/malaria#symptoms>, 2019. (Accessed 25 January 2022).
- [3] CDC - Malaria - About Malaria - FAQs, Global Health, Division of Parasitic Diseases and Malaria, 2021. <https://www.cdc.gov/malaria/about/faqs.html>.
- [4] Blood Smear - Understand the Test, Retrieved from, <https://www.testing.com/tests/blood-smear>, 2021. (Accessed 26 January 2022).
- [5] D. K. Das, M. Ghosh, M. Pal, A. K. Maiti, and C. Chakraborty, "Machine learning approach for automated screening of malaria parasite using light microscopic images," *Micron*, vol. 45, pp. 97–106, Sep. 2013.
- [6] F. B. Tek, A. G. Dempster, and I. Kale, "Computer vision for microscopy diagnosis of malaria," *Malaria J.*, vol. 8, no. 1, p. 153, 2009.
- [7] WHO, Malaria Microscopy Quality Assurance Manual. Geneva, Switzerland: World Health Organization, 2016.
- [8] Anggraini, D., Nugroho, A.S., Pratama, C., Rozi, I.E., Pragesjvara, V. and Gunawan, M., 2011, December. Automated status identification of microscopic images obtained from malaria thin blood smears using Bayes decision: a study case in Plasmodium falciparum. In 2011 International Conference on Advanced Computer Science and Information Systems (pp. 347-352), IEEE.
- [9] Di Ruberto, C., Dempster, A., Khan, S. and Jarra, B., 2002. Analysis of infected blood cell images using morphological operators. *Image and vision computing*, 20(2), pp.133-146.
- [10] Díaz, G., Gonzalez, F. and Romero, E., 2007, November. Automatic clump splitting for cell quantification in microscopical images. In *Iberoamerican Congress on Pattern Recognition* (pp. 763-772). Springer, Berlin, Heidelberg.
- [11] Sunarko, Djuniadi, M. Bottema, N. Iksan, K.A.N. Hudaya, M.S. Hanif, Red blood cell classification on thin blood smear images for malaria diagnosis, *J. Phys. Conf.* 1444 (1) (2020), <https://doi.org/10.1088/1742-6596/1444/1/012036>.
- [12] M. Poostchi, K. Silamut, R.J. Maude, S. Jaeger, G. Thoma, Image analysis and machine learning for detecting malaria, *Transl. Res.* 194 (2018) 36–55, <https://doi.org/10.1016/j.trsl.2017.12.004>. Mosby Inc.
- [13] B.N. Narayanan, R. Ali, R.C. Hardie, Performance analysis of machine learning and deep learning architectures for malaria detection on cell images, in: *Applications of Machine Learning*, 2019, September 6, <https://doi.org/10.1117/12.2524681>.
- [14] S. Sarkar, R. Sharma, K. Shah, Malaria Detection from RBC Images Using Shallow Convolutional Neural Networks, 2020 arXiv Preprint arXiv: 2010.11521.
- [15] M. Delgado-Ortet, A. Molina, S. Alferez, J. Rodellar, A. Merino, A deep learning approach for segmentation of red blood cell images and malaria detection, *Entropy* 22 (6) (2020) 657, <https://doi.org/10.3390/e22060657>.
- [16] K.M.F. Fuhad, J.F. Tuba, M.R.A. Sarker, S. Momen, N. Mohammed, T. Rahman, "Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application", *Diagnostics* 10 (5) (2020) 329, <https://doi.org/10.3390/diagnostics10050329>.

- [17] Maqsood, M.S. Farid, M.H. Khan, M. Grzegorzec, Deep malaria parasite detection in thin blood smear microscopic images, *Appl. Sci.* 11 (5) (2021) 2284, <https://doi.org/10.3390/app11052284>.
- [18] Q. Quan, J. Wang, L. Liu, An effective convolutional neural network for classifying red blood cells in malaria diseases, *Interdiscipl. Sci. Comput. Life Sci.* 12 (2) (2020) 217–225, <https://doi.org/10.1007/s12539-020-00367-7>.
- [19] Rahman, H. Zunair, M. Sohel Rahman, J. Quader Yuki, S. Biswas, M. Ashraful Alam, N. Binte Alam, M.R.C. Mahdy, Improving Malaria Parasite Detection from Red Blood Cell Using Deep Convolutional Neural Networks, 2019 arXiv Preprint arXiv:1907.10418.
- [20] Mosabbir Bhuiyan, Md Saiful Islam, “A new ensemble learning approach to detect malaria from microscopic red blood cell images”, *Sensors International*, Volume 4, 2023, 100209, ISSN 2666-3511, <https://doi.org/10.1016/j.sintl.2022.100209>.
- [21] R. Mishra, S. S. Saranya and M. Shafahad, "Analysis of different Machine Learning and Deep Learning Techniques for Malaria Parasite Detection," 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2022, pp. 1-7, doi: 10.1109/ICERECT56837.2022.10059648.
- [22] W. Setiawan, A. Faisal, N. Resfita and Y. A. Rahman, "Detection of Malaria Parasites using Thresholding in RGB, YCbCr and Lab Color Spaces," 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2021, pp. 70-75, doi: 10.1109/iSemantic52711.2021.9573224.