

# **AN AUTOMATED APPROACH for DIAGNOSIS of MALARIA DISEASE**

**Mohan Gowda K<sup>1</sup>, Seema Nagaraj<sup>2</sup>**

PG Student, Dept. of MCA, Bangalore Institute of Technology, Bengaluru, India<sup>1</sup>

Assistant Professor, Dept. of MCA, Bangalore Institute of Technology, Bengaluru, India<sup>2</sup>

**Abstract:** Malaria [1] is a dangerous disease, infection starts when an infected Anopheles mosquito injects plasmodium parasites into the skin of the human. However, patients who have been diagnosed with malaria can get early medical and clinical therapies to improve their possibility of surviving. Therefore, it's crucial to find malaria early on. In this research, we used a trained Convolutional Neural Network [6] to tackle the issue of detecting malaria through a web application. The thin blood smear image is given as input into the application, which then outputs the blood cell's class, which can be either benign (milder) or malignant (harmful), depending on the image. The model used in the web application had a validation loss of 0.3536 and an accuracy of 0.9588.

**Keywords:** Malaria, Thin Blood Smear, Web Application, Convolutional Neural Network (CNN).

## **I. INTRODUCTION**

A mosquito of the genus Anopheles is always responsible for transmitting human malaria. There are four main kinds of the complicated disease of malaria. Anopheles mosquitoes spread the disease, which is brought on by a protozoan of the species Plasmodium. It had a significant impact on human history and has resulted in illness, misery, and death. Only a patient's blood can provide a definitive diagnosis of malaria. A small amount of blood is applied to a glass slide to form a thin film, and the slide is then studied under a microscope. Malaria is confirmed if Plasmodium is discovered in any of the blood cells. Although this test is incredibly accurate, it is costly and requires specialist labs. Despite the fact that it has been studied for over a century, in 2019, there were a total of 409 thousand fatalities and over 229 million people were infected.

Methods that are revolutionary for the Malaria detection and treatment are constantly being improved. Traditional Machine Learning (ML) techniques [4] are used for classification, however they are comparably less accurate and require feature extraction and image pre-processing beforehand. The issue with these conventional ML methods is that the clinicians already know and use the attributes that were chosen as the foundation for detection. There may be additional characteristics that have not yet been identified by humans but on which malaria may be found.

Therefore, the creation of a quick, effective, reliable, and economical method for detecting malaria disease is urgently required. Our aim is to create a web application that, at its base, uses a CNN model to identify malaria in its early stages. In our web application, we used Transfer Learning [5] to construct the model. Our web application's primary goals are to: i) Using Deep Convolutional Neural Networks, accurately classify cell pictures and discover malaria. ii) To offer a user-friendly UI for entering the blood smear image into the online application.

## **II. RELATED WORK**

The use of machine learning to detect the malarial infection, that has become one of the most extensively studied areas in recent years. This issue has been addressed in a number of academic studies using image analysis.

### **1. Applications of Machine Learning**

B N Narayanan in [10] presented the effectiveness of different Deep Learning and Machine Learning techniques for identifying plasmodium on cell pictures using digital microscopy is examined in this paper. They used a dataset containing 27,558 cell pictures that was made accessible to the public and had an equal amount of parasitized and uninfected cells. For training and testing, the dataset is divided into 2 groups at random, each with 80 and 20 percent of the data. Depending on the categorization architecture used, they apply colour constancy and spatially resample all pictures to a specific size. They provide a quick Convolutional Neural Network(CNN) architecture for classifying cell

images. The effectiveness [2] of transfer learning algorithms constructed on well-known network topologies including AlexNet, ResNet, VGG-16, and DenseNet is also examined and analysed.

## **2. Machine Learning Models**

A Molina in [12] discussed, Methods: The identification of 2852 colour and texture characteristics came from using the watershed and histogram thresholding techniques[12] for segmentation of 15 660 erythrocyte images from 87 smears. The training and assessment portions of the dataset were separated. A training set was used to create the entire system, and different categorization algorithms were contrasted to see which one provided the most precise recognition. The evaluation set was then used to evaluate the recognition system, which involved 2 steps: analysing complete smears to determine whether malaria infected cells was present and to classify each individual cell picture to test the system's recognition skills.

## **3. Sequential Classification**

H I Okagbue in [14] presented 6 models were classified using only the 15 symptoms, their expected accuracy decreased. Adaboost nonetheless had the best results, 96.60% of precision, 98% of accuracy and 1.80% of error rate. Once more, performance of Logistic Regression was weakest. The current investigation found a significant relationship between the outcome and both sex and age. The Adaboost approach can be used to develop online or mobile-based decision support systems or fast diagnostic tools for diagnosing malaria. The current study, may reduce the rate of misdiagnosis, reduce the number of malaria-related deaths, and improve the general health of people living in malaria-endemic regions.

## **4. Random Forest**

Y W Lee in [15] discussed, Methods: To construct datasets, patient data was extracted from PubMed abstracts published between 1956 and 2019. One dataset exclusively contained information on parasitic infections, and the other had data on other ailments. The six machine learning models we tested were the Random Forest (RF), CatBoost, Support Vector Machine (SVM), AdaBoost, Gradient Boosting (GB), and Multilayered Perceptron. Additionally, a synthetic minority oversampling strategy (SMOTE) was used to solve the issue of imbalance in data.

Results: Whether or not SMOTE was included, random forest was the best model found for the dataset on only parasite diseases. It was determined that GB was the best size for the full dataset. SMOTE, on the other hand, led to the best performance of RF. The most important factor found was nationality while malaria prediction given the asymmetry of the data. The symptom was the important factor in the balanced data with SMOTE.

## **5. Health and Technology**

M Poostchi in [13] discusses recent developments in machine learning and image analysis for the detection of malaria and gives an overview of these methods. According to the imaging, picture preprocessing, segmentation of cell and parasite detection feature computation, and automatic cell categorization approaches used, the literature is divided into groups. The various techniques are included in the table that are relevant literature noted for both thick and thin blood smear pictures. They also covered most recent advancement in smartphone technology and deep learning malaria detection in future.

# **III. METHODOLOGY**

## **A. Data Set:**

We created a system in the first module to obtain the input from the dataset for training and testing. We used the data from Malaria Cell Images Dataset, from Kaggle. Dataset contains both infected and uninfected images, total of 27558 thin blood smear images make up the dataset.

## **B. Overview of the Proposed Method:**

Malaria Disease Detection application's architecture is shown below. It contains mostly the two building components of machine learning and image processing. Thin blood smear image is inputted to the web application for image processing, subsequently scale to an acceptable size as part of pre-processing stage. The CNN is trained using the dataset Malaria Cell Images to develop model, and later saved for eventual deployment in machine learning block. This model is then given the pre-processed image and classifies it, predicting whether it is "Infected" or "Uninfected."

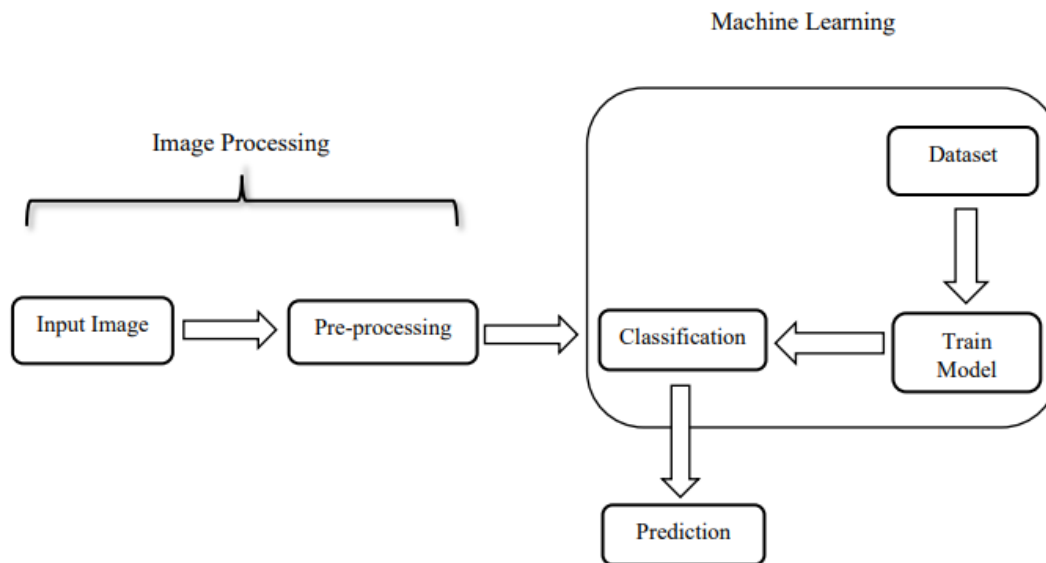


Fig. 1 System Architecture

## Image Pre-Processing:

Additionally, for datasets that are unbalanced, it is required to augment more pictures to avoid for their lower contributions to a trained model. Training set is subjected to image pre-processing techniques like Resize, Random Rotation transformations, Colorjitter, and Random Horizontal Flip. The testset is left unchanged because it shouldn't be changed in any way.

## Training Model:

Resnet's main principle is to introduce a "identity shortcut connection" where more than one layer will be skipped. The 1st step in Resnet before entering common layer behavior block, which contains batch normalization, convolution and ReLU operation. The following process will be batch normalization, which is an element-wise process and does not alter the volume size of the data. Transformation functions will be used to convert the images that are given as input, later it will be given to neural network model, in our case we used Resnet. For comparison, we created a custom Resnet9 architecture and a Resnet34 pre-trained model. Ten epochs were used to train the Resnet9 model. The training approach for the Resnet34 model consisted of two steps: 5 epochs were used to train the fully connected layer of the model, and ten epochs were used to train the whole model. Then, using PyTorch functions, the state of the learnt weights are saved for deployment in the application.

## IV. RESULTS

Two convolutional neural network models have been trained. By defining the kernel size, padding and out features, in features to be used in each layer, we created a unique architecture for Resnet9. The other is the pre-trained Resnet34, which can be found in the machine learning library for PyTorch torchvision models. For an increasing number of epochs, Resnet34's validation accuracy fluctuation was plotted in Fig. 2. The graph demonstrates that a model's accuracy is 0.3594 in the first epoch and at the conclusion of ten epochs it was 0.9588.

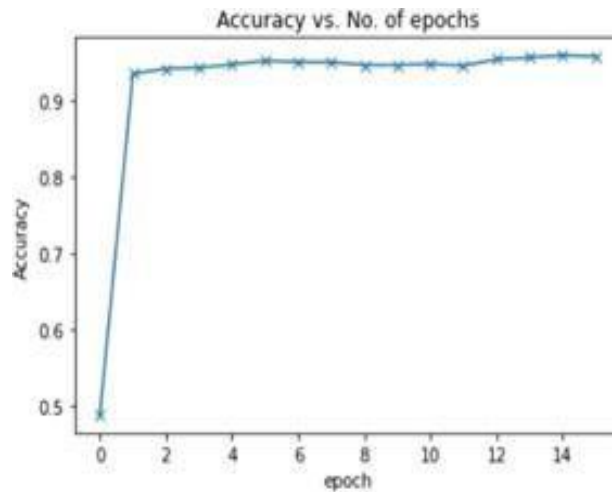


Fig. 2 Resnet34 model's graph showing epochs vs. accuracy.

Fig. 3 displays the fluctuation in loss for the Resnet34 model's validation and train sets with increasing no. of epochs.

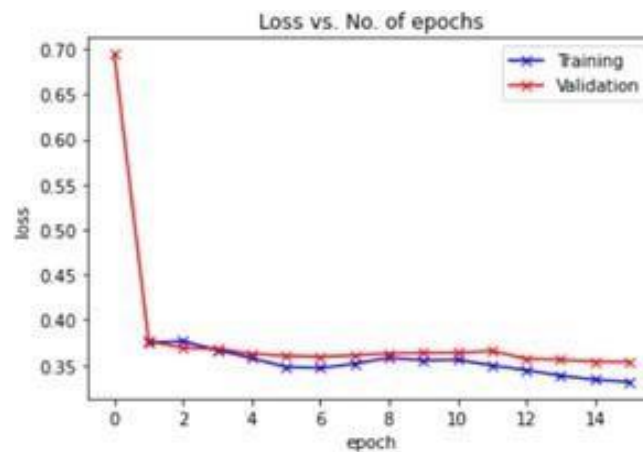


Fig. 3 Loss vs the no. of epochs for the validation and train sets of the Resnet34.

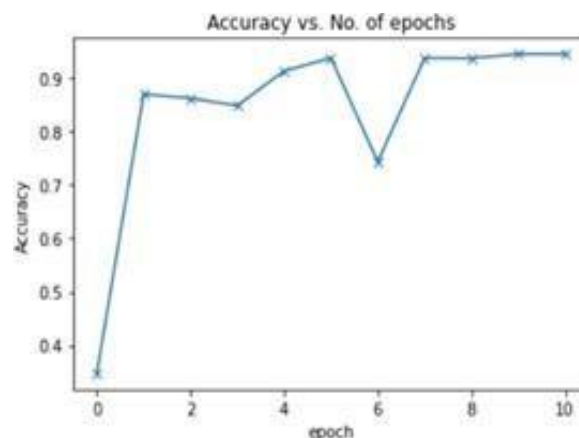


Fig. 4 Accuracy v/s no. of epochs of Resnet9

The findings from training the customised Resnet9 model over 10 epochs are very promising. The accuracy value at the end of the tenth epoch was 0.9444, which is lower than the accuracy value for the Resnet34 model that has been trained beforehand, which was 0.9588. Graphs drawn for accuracy v/s no. of epochs and loss v/s no. of epochs are shown in Fig. 4 and Fig. 5, respectively.

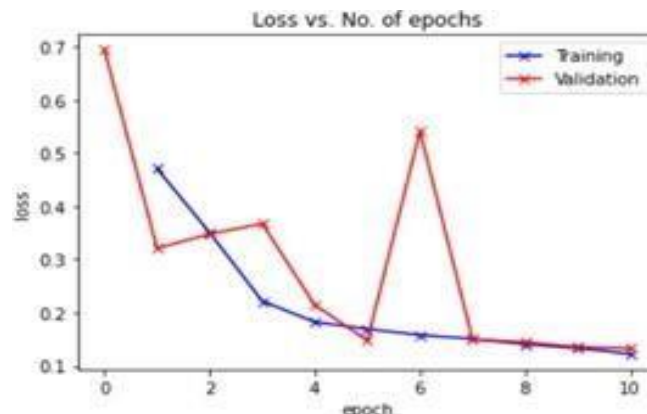


Fig. 5 Loss v/s no. of epochs for validation and train sets of Resnet9

We employed the pre-trained Resnet34 model in the application because to its superior performance after studying and comparing the results produced for the custom model Resnet9.

## V. CONCLUSION

A fully automated application for detecting malaria disease is presented by the system. The application demonstrated demonstrates how convolutional neural networks effectively classify blood smear cells. The Deployed model is tested for accuracy by being compared to a number of other pretrained models and their actual accuracy over the validation set, with the best one being chosen for deployment. Because of the logically designed Graphical User Interface (GUI), the end user finds it easier to operate the application.

The Malaria Disease Detection application is mainly targeted at the medical industry concerned with the diagnosis of malaria. It enables one to quickly identify and counter malaria conditions since the report generated in the application is fast and reliable, thereby reducing the risk. It also helps practitioners new to the field, perfect their diagnostic skills as it provides more accurate results.

It also helps researchers to develop deep learning models and enhance the accuracy of the model even further. It serves as a base application which could be developed further to identify and diagnose other diseases based on the blood smear conditions as there are plenty of other disease types left that may affect the human system.

## REFERENCES

- [1]. E. A. Ashley, A. P. Phyto, and C. J. Woodrow, "Malaria," *The Lancet* 391, no. 10130, 2018, pp. 1608-1621.
- [2]. R. Sivaramakrishnan, S. Jaeger, and S. K. Antani, "Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images," *Peer J* 7, 2019, pp. e6977.
- [3]. J. Jackovich, and R. Richards, "Machine learning with AWS," Packt, 2018.
- [4]. S. Ardabili, A. Mosavi, and A. R. Varkonyi-Koczy, "Advances in machine learning springer reviewing hybrid and ensemble methods," *International Conference on Global Research and Education*, Springer, Cham, 2019, pp. 215-227.
- [5]. A. Mosavi, S. Ardabili, and A. R. Varkonyi-Koczy, "List of deep learning models," *International Conference on Global Research and Education*, Springer, Cham, 2019, pp. 202-214.
- [6]. K. O'Shea, and R. Nash, "An introduction to convolutional neural networks," *arXivpreprint arXiv:1511.08458*, 2015.
- [7]. S. Ardabili, A. Mosavi, M. Dehghani, and A. R. Varkonyi-Koczy, "Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review," *International Conference on Global Research and Education*, Springer, Cham, 2019, pp. 52-62.
- [8]. R. Claywell, L. Nadai, I. Felde, S. Ardabili, and A. Mosavi, "Adaptive Neuro-Fuzzy Inference System and a Multilayer Perceptron Model Trained with Grey Wolf Optimizer for Predicting Solar Diffuse Fraction," *Entropy* 22, no. 11, 2020, p.1192.
- [9]. S. Ardabili, A. Mosavi, and A. R. Varkonyi-Koczy, "Systematic review of deep learning and machine learning models in biofuels research," *International Conference on Global Research and Education*, Springer, Cham, 2019, pp. 19-32.

- [10]. B. N. Narayanan, R. Ali, and R. C. Hardie, "Performance analysis of machine learning and deep learning architectures for malaria detection on cell images," *Applications of Machine Learning*, vol. 11139, International Society for Optics and Photonics, 2019, pp. 111390W.
- [11]. B. Choubin, M. Abdolshahnejad, E. Moradi, X. Querol, A. Mosavi, S. Shamshirband, and P. Ghamisi, "Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain," *Science of The Total Environment* 701, 2020, p.134474.
- [12]. A. Molina, S. Alf  rez, L. Bold  , A. Acevedo, J. Rodellar, and A. Merino, "Sequential classification system for recognition of malaria infection using peripheral blood cell images," *Journal of clinical pathology* 73, no. 10, 2020, pp.665-670.
- [13]. M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, "Image analysis and machine learning for detecting malaria," *Translational Research*, 194, 2018, pp. 36-55.
- [14]. H. I. Okagbue, P. E. Oguntunde, E. C. Obasi, P. I. Adamu, and A. A. Opanuga, "Diagnosing malaria from some symptoms: a machine learning approach and public health implications," *Health and Technology* 11, no. 1, 2021, pp.23.
- [15]. Y. W. Lee, J. W. Choi, and E. H. Shin, "Machine learning model for predicting malaria using clinical information," *Computers in Biology and Medicine* 129, 2021, p.104151.