

MALARIAL PARASITES DETECTION IN THE BLOOD CELL USING CONVOLUTIONAL NEURAL NETWORK

Deeksha Kotian¹, Brina Paula², Dhanush Hegde³, Darshith D N⁴, Annappa Swamy D R⁵

^{1,2,3,4} Computer Science & Engineering, MITE Moodabidri.

⁵ Associate Professor, Dept. of Computer Science & Engineering, MITE Moodabidri, Karnataka, India.

Abstract - Malaria is a life-threatening, infectious mosquito-borne disease caused by Plasmodium parasites. These parasites are transmitted by the bites of infected female Anopheles mosquitoes. It is a significant burden on our healthcare system and it is the major cause of death in many developing countries. Therefore, early testing is necessary to detect malaria and save lives. The standard diagnostic methods for malaria detection are Microscopy and Rapid Detection Test (RDT). Microscopy process requires a skilled microscopist which sometimes cannot be available in rural areas and it is impossible to manually detect the presence of parasites. The RDT may not be able to detect some infections with lower numbers of malaria parasites circulating in the patient's bloodstream. Therefore, there is a need for specialized technology that proves essential to combat this problem. This proposed system uses a deep learning model based on convolutional neural network (CNN) to classify single blood smears whether it is parasitized or non-parasitized. A variety of techniques were performed under CNN model such as activation layer, relu, maxpool, dropout, flatten, dense layer to optimize and improve the model accuracy. Our model deep-learning model predicts malaria parasites from images with an accuracy of 95.34%.

Key Words: CNN, Malaria detection, Deep Learning, Sequential Model, Blood cells.

1. INTRODUCTION

Malaria is a mosquito-borne disease caused by a plasmodium parasites. It is transmitted by the bite of infected mosquitoes. Worldwide, there is an estimated 300–500 million people who suffer from malaria each year, which results in 1.5–2.7 million deaths yearly. According to the World Health Organization (WHO), approximately 219 million cases were diagnosed with malaria resulting in 435,000 deaths globally in 2017. Malaria is a deadly disease which is more frequently found in rural areas where medical diagnosis and health care options are not easily accessible. For Malaria diagnosis, the RDT and microscopic diagnosis are the most used clinical methods. RDT is an effective and faster tool. Also, it does not require the presence of a trained

medical professional. But RDT has few drawbacks like susceptibility to damage by heat and humidity and higher cost compared to a light microscope.

The machine learning and deep learning approaches have proved to be successful in the diagnosis of a disease. Machine learning models require a lot of tuning, factor analysis, and feature engineering. The machine learning method is not scalable with more data provided. Machine learning requires feature engineering and feature training which is not a handy tool. Deep learning models are reliable and easily scalable with a higher accuracy rate. Convolutional Neural Networks (CNN) have proved to be really effective in a wide variety of computer vision tasks. Convolutional Neural Networks transform input image volume into an output volume holding a class label. The regular manual diagnosis of blood samples requires proper expertise in classifying and counting the infected and uninfected cells. This is a time-consuming task and accuracy will be very less.

To overcome the problem faced during the detection of malarial parasites in the blood cell, there is a need to develop a system which predicts the presence of malarial parasites in the blood cell efficiently.

2. LITERATURE REVIEW

Snehal Suryawanshi et al. [1] In the proposed system Poisson distribution using minimum error thresholding is used to detect the presence of malarial parasites in blood cells using the technique of image segmentation. The work presented in this paper is based on some extended techniques. Poisson distribution based minimum error thresholding algorithm automatically binarizes images, which is refined by morphological opening and hence foreground is being extracted. Then by a novel method, the seed points are detected combining multiscale Laplacian of Gaussian with gabor filtering. Then the features extracted are compared with a database to check whether the blood cell images are infected by malaria parasites.

Jigyasha Soni et al. [2] proposed another approach to differentiate between the simple RBC and malarial parasite affected blood cell. In order to maximize the productivity of the algorithm various approaches are

combined that take benefits of morphological operation and thresholding at appropriate positions.

In another work, Naveed Abbas et al. [3] proposed an approach to identify the different disorders like Malaria, Anaemia and Leukaemia by observing different features of blood cells like size, shape, color and internal features like nucleus its size and color (WBC).

Gautham Shekhar et al. [4] proposes a deep learning model based on convolutional neural network. The accuracy of the model are compared by combining the three types of CNN model- Basic CNN, VGG-19 Frozen CNN, and VGG-19 Fine Tuned CNN. The model with a higher rate of accuracy is acquired.

Abhishek Kanojiya et al. [5] The proposed system developed a model that can detect multiple cell images in thin blood smear and predict them as either infected or uninfected by effectively using image processing techniques along with machine learning approaches. The Machine Learning approach used in this model is Support Vector Machine algorithm.

Aliyu Abubakar et al. [6] The proposed model uses machine-learning models to detect the presence of malaria in blood cells by evaluating and testing the model which uses six different features—VGG16, VGG19, ResNet50, ResNet101, DenseNet121, and DenseNet201 models. Later Decision Tree, Support Vector Machine, Naïve Bayes, and K-Nearest Neighbor classifiers were trained using these six features. The System compared the accuracy of each classifier with features using ROC curve.

Jung Yoon et al.[7] developed an automated microscopic system for malarial detection and also parasitemia measurement. The proposed system in this paper was comprised of components such as microscope, fluorescent dye, plastic chip and an image analysis program. The analytical performance concerning the precision, linearity and limit of detection are evaluated and then it was compared to the result that obtained by conventional microscopic Peripheral Blood smear examination and flow cytometry. This System is proven to be produced the result of high precision in measurement of density of malaria parasites and also high degree of linearity.

Razzak M. I [8] proposed a method to overcome the problems faced in manual diagnosis. The malaria classification is carried out using different set of features which are forward to the ANN. Three feature set is used and has been found that Feature matrix III provided promising results.

Preedan.W et al. [9] proposed a model in which Giemsa-stained thin blood cell images are used to predict the presence of malarial parasite. Using SVM binary

classifier, Statistical features are computed for each cell.

F.Boray et al.[10] introduced a framework to detect and identify the malarial parasites in Giemsa stained thin blood cells. By identifying various life cycle stages and species the malarial parasites are detected in the blood cells. The model operate at 0- 1% parasitemia without any false detection and also with less than 10 false detections with the rate of 0.01%.

Nicholas E.Ross et al.[11] proposed a system to automate the malarial diagnosis using image processing algorithms. The system was developed particularly for the identify the malarial parasites present in thin blood cell and differentiate the different species of malaria. Morphological and selection techniques are used to identify the red blood cells and possible parasites present in the blood cell. Then the back propagation algorithm was used to diagnoses the species. It has been found that

11 out of 15 sample species were correctly determined.

K.M.Khatri et al.[12] evaluated and tested a method to detect the malaria accurate and faster. The system was based on image processing. Total of 30 blood samples were taken as a dataset. Result produced by this system was compared with manual analysis.

3. METHODOLOGY

The Proposed Work in this paper relates to the detection of malarial parasites in the blood cell. The problem starts with identifying whether the blood sample is infected with malarial parasites or not.

The dataset is collected from the National Institution of Health's (NIH) website which consists of 27,557 images of both parasitized and non-parasitized blood cells. The datasets contain 13,778 infected images and 13,779 uninfected images.

OpenCV is used for image processing where the images provided are being processed. Then the median filter is applied to the images where the noise in the images are removed. Median filter is applied to each image in parasitized and uninfected folder respectively.

After the image is being processed, it starts with classifying, training, and validation of the dataset. For splitting the dataset into training and validation sets, the model uses 20% of the validation split. After the splitting the training dataset has 22,048 images belonging to two classes parasitized and uninfected and 5,510 images belonging to the validation set. The model used in this paper is the Convolutional Neural Network .

The Convolutional Neural Network(CNN) are artificial neural network designed to process pixel data and used for image classification and recognition. A CNN consists of an

input layer, a hidden layer and an output layer. In CNN the hidden layer has layers that perform convolutions. The model uses TensorFlow, Keras, Python libraries. Keras is the python library for neural network. TensorFlow is a machine learning platform that can be used to develop training models using Python. The proposed system implements a sequential CNN model. As a sequential model, CNNs are appropriate for a stack of layers that is made up of exactly one input tensor and one output tensor. The model uses two Convolutional Layers which are applied on image to extract features from it, two Max Pooling Layers which is a type of Pooling used to reduce the size of image, three Dropout Layers as they prevent overfitting on the training data, one Flatten Layer which converts multi-dimensional matrix to single dimensional matrix and two Dense Layers which helps to classify image based on output from previous layers.

The input to the model is a 68×68×3 pixel resolution. The convolutional layers uses 3×3 filters. The first convolutional layer has 16 filters and the second convolutional layer has 64 filters. The convolution layers use the ReLU activation function which is a nonlinear function which helps to prevent the exponential growth in the computation required to operate the neural network. Max-pooling layer with a pooling window of 2×2 follows the convolutional layer. A dropout of 0.2 is added to increase the learning of the model. The output is sent to the next 64 filter convolution layer, the max pooling layer and a 0.3 dropout layer. The pooled feature map generated is transformed into a one-dimensional vector using the Flatten layer as this vector will now be fed into an artificial neural network for classification. The one-dimensional vector image is sent to the dense layer with 64 units which represents the output size of the layer. A dropout of

0.5 is added. Finally, the activation of sigmoid is used with one-unit dense layer for calculating the accuracy.

The printing of the model summary is used to identify the parameters of the model. There are a total of 931,457 trainable parameters and zero non-trainable parameters that are acquired. The training of the model takes place with details of generator, steps per epoch, epochs, validation data, and validation steps. Fitting of the model is done with 6 epochs to decrease the loss percentage. At the 6th epoch, the readings are loss: 0.1357, accuracy: 0.9564, validation loss: 0.1810 - validation accuracy: 0.9414. The graph of accuracy vs epoch for the model is used to determine the overfitting and underfitting of data and the loss vs epoch graph gives us a snapshot of the training process and the direction in which the network learns. So, once the model is trained, it is saved in the extension of malaria.h5. The h5 file is a data file saved in hierarchical data format. It is used to store scientific data in multi-dimensional array.

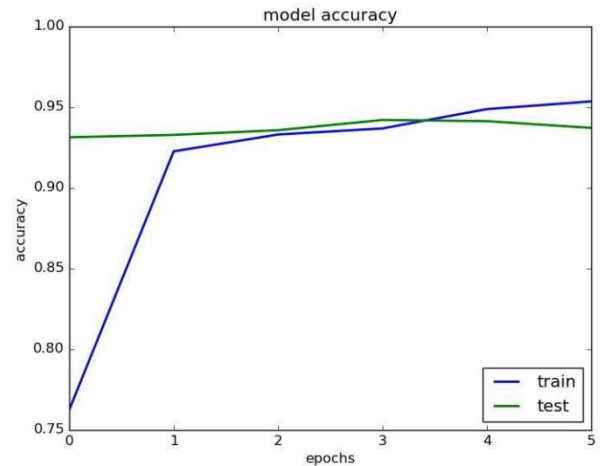


Fig 1: - The model accuracy vs epoch

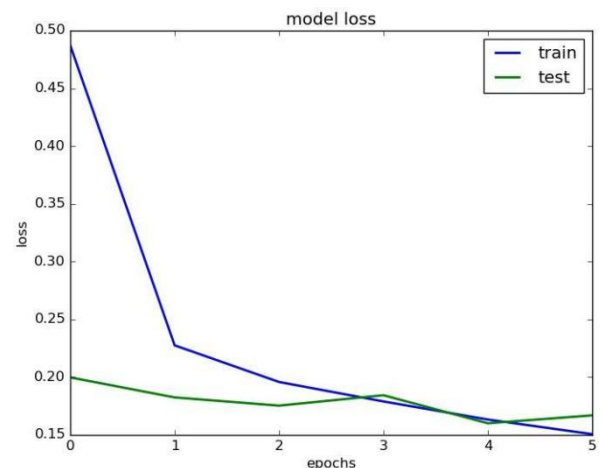


Fig 2: - The model loss vs epoch

The Fig. 1 and Fig.2 shows the model accuracy Vs epoch graph and the model loss Vs epoch graph of the CNN model which portrays that the training and testing accuracy line meets each other which shows an accurate result from the model. After conducting multiple iterations to make the results as accurate as possible, the final accuracy rate of the proposed model is obtained as 95.64%.

4.CONCLUSION

Traditional method for malaria detection requires trained personnel and consumes more time hence convolutional neural network model is used to predict malarial parasite in the blood cell. The proposed model gives higher accuracy with less time for computation. Various disease prediction along with health care application can be the new evolution and digitization.

REFERENCES

- [1] Suryawanshi, Ms Snehal, and V. Dixit. "Improved technique for detection of malaria parasites within the blood cell images." *Int J Sci Eng Res* 4 (2013): 373-375
- [2] Soni, Jigyasha, Nipun Mishra, and Chandrashekhar Kamargaonkar. "AUTOMATIC DIFFERENTIATION BETWEEN RBC AND MALARIAL PARASITES BASED ON MORPHOLOGY WITH FIRST ORDER FEATURES USING IMAGE PROCESSING." *International Journal of Advances & Technology* 1, no. 5 (2011): 290. *International Journal of Advances in Engineering*
- [3] Abbas, Naveed, and Dzulkipli Mohamad. "Microscopic RGB color images enhancement for blood cells segmentation in YCBCR color space for k-means clustering." *J Theor Appl Inf Technol* 55, no. 1 (2013): 117-125.
- [4] Shekar, G., Revathy, S. and Goud, E.K., 2020, June. Malaria detection using deep learning. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)* (pp.746-750). IEEE.
- [5] Gomes, C., Kanojiya, A., Yadav, A. and Motwani, K., 2020. Malaria Detection using Image Processing and Machine Learning.
- [6] Abubakar, A., Ajuji, M. and Yahya, I.U., 2021. DeepFMD: Computational Analysis for Malaria Detection in Blood-Smear Images Using Deep-Learning Features. *Applied System Innovation*, 4(4), p.82.
- [7] Yoon, J., Jang, W.S., Nam, J., Mihn, D.C. and Lim, C.S., 2021. An automated microscopic malaria parasite detection system using digital image analysis. *Diagnostics*, 11(3), p.527.
- [8] Razzak, M.I., 2015. Automatic detection and classification of malarial parasite. *International Journal of Biometrics and Bioinformatics (IJBB)*, 9(1), pp.1-12.
- [9] Preedan, W., Phothisonthai, M., Senavongse, W. and Tantisatirapong, S., 2016, February. Automated detection of plasmodium falciparum from Giemsa-stained thin blood films. In *2016 8th international conference on knowledge and smart technology (KST)* (pp. 215-218). IEEE.
- [10] Tek, F.B., Dempster, A.G. and Kale, I., 2010. Parasite detection and identification for automated thin blood film malaria diagnosis. *Computer vision and image understanding*, 114(1), pp.21-32.
- [11] Ross, N.E., Pritchard, C.J., Rubin, D.M. and Duse, A.G., 2006. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Medical and Biological Engineering and Computing*, 44(5), pp.427-436.
- [12] Khatri, K.M., Ratnaparkhe, V.R., Agrawal, S.S. and Bhalchandra, A.S., 2013. Image processing approach for malaria parasite identification. In *International Journal of Computer Applications, National Conference on Growth of Technologies in Electronics, Telecom and Computers-India's Perception*.