International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2410286

Malaria Parasite Detection in Microscopic Blood Smear Images using Deep Learning Approach

Dr. M. Praneesh*[1], Sai Krishna P K[2], Febina. N[3], Ashwanth.V[4]

[1]Assistant Professor, PG & Research Dept of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India

[2, 3,4]PG Scholar, PG & Research Dept of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India

ARTICLEINFO

Article History:

Accepted: 10 April 2024

Published: 22 April 2024

Publication Issue

Volume 10, Issue 2

March-April-2024

Page Number

669-676

ABSTRACT

Malaria remains a significant global health concern, posing formidable challenges to healthcare systems. Conventional diagnostic methods rely on manual examination of blood smears under a microscope, a process prone to inefficiencies and subjectivity. Despite prior attempts to leverage Deep Learning algorithms for malaria diagnosis, practical performance has often fallen short. This paper presents a novel machine learning model centred on Convolutional Neural Networks (CNNs) designed to automate the classification and prediction of infected cells in thin blood smears on standard microscope slides. Through rigorous ten-fold cross-validation with 27,558 single-cell images. This paper reviews various image processing techniques employed for the detection of malaria infection in humans, presenting a comparative analysis of these methods

Keywords: Malaria, Automated diagnosis, Blood smears, Deep Learning

I. INTRODUCTION

Malaria originated in Africa, caused by the Plasmodium falciparum virus. Spread by mosquitoes, the disease has become global. While it thrives in warm climates, it cannot survive in cold temperatures. Dating back 40 million years, malaria affects animals and humans of all ages, leading to symptoms ranging from fever to coma and death. The virus targets and destroys red blood cells, impairing organ function. Diagnosis requires microscopic examination of blood samples.

Malaria is a dangerous illness caused by parasites transmitted through mosquito bites. It can be prevented and treated. In 2016, there were 216 million cases worldwide, with 445,000 deaths, mostly in Africa. Quick and accurate diagnosis is crucial for effective treatment. Current methods, like manual microscopy, take time and can be prone to mistakes. Our study suggests better ways to diagnose malaria using digital image processing on stained blood samples. These new methods are faster, more reliable, and reduce the chance of missing cases. We introduce simple techniques for spotting and identifying malaria parasites in blood smears. However, the success of

these methods depends on the quality of the microscope and images used, assuming the patient has a healthy blood count.

Malaria infection begins when an infected mosquito bites a person, injecting Plasmodium parasites called sporozoites. These parasites quickly travel to the liver and multiply for 7 to 10 days. They then transform into merozoites and enter the bloodstream, settling in lung capillaries. From there, they invade red blood cells, multiply further, and eventually burst the cells, continuing the cycle when passed to another person by a mosquito. When someone has malaria, they experience symptoms such as high fever, headache, nausea, vomiting, abdominal pain, or even coma. The body tries to fight the infection by triggering white blood cells to provide immunity .While various machine learning models exist to predict malaria, our study focuses on using a deep learning model for more accurate predictions. Detecting malaria parasites in a person's blood involves microscopic examination of blood samples spread on slides, either as thick or thin smears. These slides are stained to highlight blood cells, making it easier for pathologists to examine them. Thick smears help detect the presence of malaria parasites, while thin smears aid in identifying and quantifying parasite species. However, this process is time-consuming and requires highly skilled pathologists. In rural areas with limited medical facilities, getting a timely diagnosis can be challenging and may impact patient health, sometimes leading to fatal outcomes.

Malaria remains a significant global health concern, with efforts underway to eradicate it by 2040, led by organizations like the United Nations and the Bill & Melinda Gates Foundation. India, for example, aims to become malaria-free by 2030 through initiatives like the National Framework for Malaria Elimination. Achieving these goals requires robust and accessible diagnostic systems to reduce mortality rates and ultimately eliminate malaria.

Researchers have explored various techniques, including computer vision, machine learning, and deep learning, for detecting malaria parasites. While computer vision involves segmenting blood cells and extracting features for classification, recent advancements in deep learning have automated this process. Techniques such as Convolutional Neural Networks (CNNs), Transfer Learning, and Model Ensembles have shown promise in image classification tasks, achieving performance comparable to human intelligence.

This study aims to develop a deep learning system for timely and accurate malaria diagnosis using a discounted ensemble building technique called Snapshot Ensemble. This approach helps build strong learners while training only one deep learning model. The paper is structured into sections: a literature review of previous research on malaria classification methods, a description of the methods used in this study, an analysis of the results obtained a comparison with previous studies, and a conclusion summarizing the findings and suggesting avenues for future research.

II. REVIEW OF LITERATURE

Image processing and machine learning algorithms play a crucial role in detecting malaria parasites. Traditional machine learning algorithms rely on handcrafted features, which can be complex to extract and prone to errors, requiring expert domain knowledge. In contrast, deep learning alleviates this burden by automatically learning features from raw data, making it a preferred choice for malaria classification. Various image processing techniques are employed to enhance image quality and extract relevant features. These techniques include filtering, contrast stretching, segmentation, thresholding, and morphological operations. Filtering methods are commonly used to remove noise from whole slide images, ensuring clearer and more accurate results.

Peter et al. [10] proposed a model to detect malarial cells using a new genotypic signature. Inspired by this concept, we adopted the blood stain approach for our project. Raghuveer et al. [5] emphasized the importance of variability and artifacts in microscopic images of malarial cells, particularly using Leishman blood smears. Building on their work, we incorporated similar concepts into our paper.

Similarly, Ratnaparkhe et al. [4] demonstrated image processing techniques using OpenCV and contour detection, which we applied to identify attributes of blood cells in our project. Zhaoui et al. [7] introduced the concept of convolutional neural networks (CNNs) for classifying infected blood cells, inspiring us to build our project around scratch CNNs and other CNN architectures.

Di Ruberto et al. [11] employed a (5x5) median filter for smoothing and noise removal, while Díaz et al. [12] used a low-pass filter to eliminate noisy components from slide images. Anggraini et al. [13] and May et al. [14] applied median filtering techniques for noise removal, with May et al. [14] also suggesting the use of a Weiner filter for Gaussian noise. Savkare and Narote [15] recommended Laplacian filtering for image smoothing and edge enhancement, and Arco et al. [16] implemented a Gaussian low-pass filter.

In terms of contrast enhancement, Anggraini et al. [13] utilized contrast stretching to boost parasite object properties, while May et al. [14] employed histogram stretching for contrast enhancement. Savkare and Narote [15] utilized Histogram Equalization and Partial Contrast Stretching, and Arco et al. [16] used Adaptive Histogram Equalization for localized contrast corrections. Among these techniques, partial contrast stretching and adaptive histogram equalization were found to perform better on pathological images.

Weihong et al. [1] introduced the VGG-16 model based on Visual Geometry Graphics (VGG), which we extended to the VGG-19 model for our project. Zhuocheng et al. [8] showcased automatic blood cell classification using LeNet, AlexNet, and GoogleNet, informing our development of three CNNs.

Ross et al. [3] introduced backpropagation feedforward neural networks with a higher learning rate than basic CNNs, influencing our neural network design. Gavet et al. [2] utilized Time Series Classification algorithms and Recurrence Plots to build a unified network with CNNs, guiding our approach to understanding these techniques.

Dallet et al. [6] developed real-time malarial cell identification using mobile phones, inspiring our exploration of pulse and calorie monitoring. Gopalakrishna et al. [9] focused on artificial microscopic slides of Plasmodium falciparum cells, aiding our understanding of malarial cell morphology. Furthermore, Vadavalli et al. [11] and Subashini et al. [12] explored deep learning and image processing techniques, while Madhukeerthana et al. [13] and Revathy et al. [14] evaluated medical image processing and machine learning models for disease prediction, respectively.

Di Ruberto et al. [11] employed a (5x5) median filter for smoothing and noise removal, while Díaz et al. [12] used a low-pass filter to eliminate noisy components from slide images. Anggraini et al. [13] and May et al. [14] applied median filtering to remove noise, with May et al. [14] also suggesting the use of a Weiner filter for Gaussian noise. Savkare and Narote [15] recommended Laplacian filtering for image smoothening and edge enhancement, and Arco et al. [16] implemented a Gaussian low-pass filter.

In terms of contrast enhancement, Anggraini et al. [13] utilized contrast stretching to boost parasite object properties, while May et al. [14] employed histogram stretching for contrast enhancement. Savkare and Narote [15] utilized Histogram Equalization and Partial

Contrast Stretching, and Arco et al. [16] used Adaptive Histogram Equalization for localized contrast corrections.

III.MATERIALS AND METHODS

A. Dataset Description

The dataset utilized in this study is sourced from the National Library of Medicine, a division of the National Institute of Health [39]. It comprises an archive of red blood cells segmented from Giemsa stained slides. These samples were collected from 150 infected and 50 uninfected individuals participating in the Malaria screener research activity at the CMC hospital in Bangladesh. The dataset consists of 27,558 erythrocyte images, evenly distributed with 13,779 instances from each category. Sample data from the parasitized and non-parasitized categories are illustrated in Figure 1 shows a few data samples from the parasitized and non-parasitized category

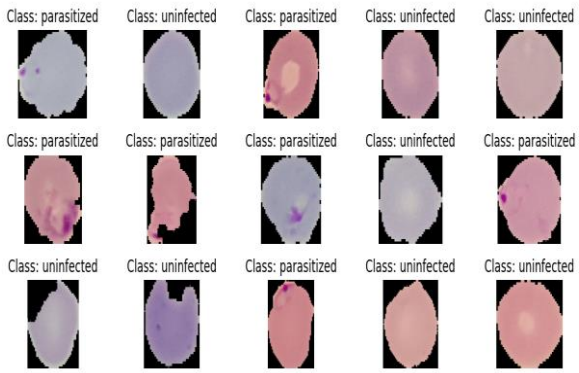

Figure 1 - Sample Data

Table 1. Distribution of the images in the dataset.

S.no	Number of Images	Class Name
1	13779	Parasitized
2	13779	Uninfected

B. Data Analysis

Biomedical images exhibit significant diversity, even for the same pathological condition, owing to variations in lighting conditions, marker stains, image extraction processes, and dimensions. Image pre-processing is essential to standardize the images and remove any irrelevant noise for analysis. In this study, all images are resized to dimensions of (135, 135) and undergo normalization and standardization processes, centering pixel values on the mean to facilitate faster convergence during training. This ensures a simple, accurate, and robust classification system

B. Data Split

Given the substantial training data requirements of deep learning systems to capture underlying image patterns and representations, the dataset is split in a ratio of 75:15:10. The training set receives 75% of the data, while the test set and validation set receive 15% and 10% of the data, respectively.

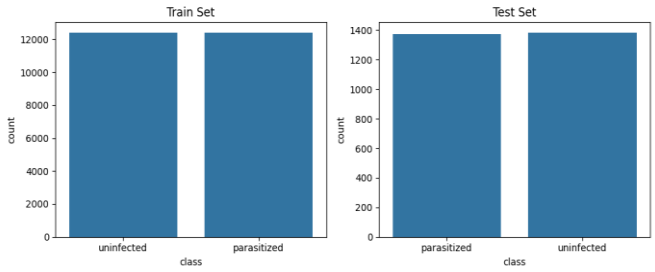

Fig 2. Training Set and Testing Set graph

C. System Architecture

Convolutional Neural Networks (CNNs) serve as the foundation for all models and experiments conducted. Raw image data is represented as arrays of pixel values, with neighboring pixels exhibiting high correlation and forming the basis for feature extraction. CNNs leverage this correlation through techniques such as local receptive fields, weight sharing, pooling, and multiple layers, resulting in a deep architecture.

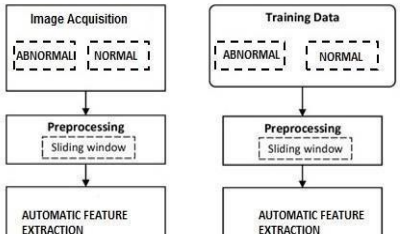

Fig-3 System Architecture

The process begins with the critical decision of determining whether a cell is infected or healthy. Training the machine involves providing all the attributes of the images. Gathering a substantial amount of images from the internet, we collected a total of 27,558 images. The next step involved dividing these images into training, validation, and testing sets, with 17,361 images allocated to training, 1,929 to validation, and 8,268 to testing. Using OpenCV, contour detection was applied to the cells. Contours are curves joining continuous points with the same color or intensity, useful for shape analysis and object detection. The presence of dark spots inside the cell led to the drawing of contours near these spots, forming circles around them.

After contour detection, the threading process was initiated. Threading allows for separate flows of execution, with multiple tasks happening concurrently. In this paper, threading was facilitated by the Thread Pool Executor, which creates a context manager to manage worker threads. The library produced essential statistics such as minimum, average, median, and maximum dimensions in an array format. Following this, image loading and resizing occurred through the Thread Pool Executor for each training, validation, and testing set. The XYZ points were then acquired as the images were processed through the setup configuration settings, scaling, and label encoding. Epochs, representing a point where time starts, were introduced during the encoding and scaling stage. Moving on to the implementation of the Basic Convolutional Neural Network (CNN), TensorFlow's Keras library was utilized. Keras provides the Conv2D attribute for working with two-dimensional images. Max-pooling was employed to down sample the neural network, while the Flatten layer prepared the data for the fully connected layers. Finally, the model was built using the sigmoid activation function to calculate accuracy.

Following the completion of the model, the next step involves printing the model summary to obtain a comprehensive overview of the parameters, including both trainable and non-trainable parameters. The model comprises a total of 15,102,529 trainable parameters and zero non-trainable parameters. After printing the model summary, training commences with the specified number of epochs, batch size, call-backs, verbosity, and validation data. The model is trained over 25 epochs, during which the loss gradually decreases until it stabilizes. Upon reaching the 25th epoch, the following readings are obtained: loss: 0.0054, accuracy: 0.9985, validation accuracy: 0.9430, and validation loss: 0.4244. However, the observed loss value is higher than expected, indicating the need for further refinement.

Graphs depicting Accuracy vs. Epoch and Loss vs. Epoch reveal a noticeable gap between the dark line representing training accuracy/loss and the grey line representing validation accuracy/loss. Notably, the loss graph exhibits significant deviation between the two lines, providing a graphical representation of the CNN's performance. Once the model is trained and achieves satisfactory accuracy, it is saved with the extension ".h5" denoting a hierarchical data format used for storing scientific data.

Table-2 Performance Table

Model	Custom Model
Confusion Matrix	[[0 , 2016, 45 , 1992]]
F1 Accuracy	95.93
Accuracy	94.99
Precision	95.79
Recall	94.09
MCC	92.92
AUC-ROC	96.95
AUC-PR	97.92

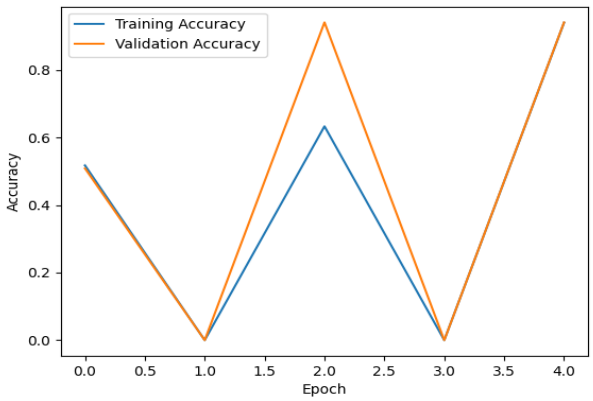

Fig. 4. Training and Testing Accuracy Graph

IV.CONCLUSION

In this study, we addressed the challenge of malaria detection, which traditionally involves time-consuming sample analysis. To overcome this limitation, we proposed the use of deep learning models for predicting malaria with high accuracy and reduced time requirements. We constructed three CNN models and identified the Fine-Tuned CNN as the most accurate among them. Our results demonstrate that the Fine-Tuned CNN achieved a significantly higher accuracy rate compared to the other models. This suggests the potential of fine-tuning pre-trained models for improving performance in malaria detection tasks

V. REFERENCES

[1]. Ahmed elmubarak Bashir ,et.al., Detection Of Malarial Parasites Using Digital Image Processing, International Conference on Communication, Control, Computing and Electronics Engineering(ICCCCEE),2017.

[2]. Ashwini Awchite et.al., A Survey on Detection of Malarial Parasites in Blood Using Image Processing, International Journal Of Innovative Research in Computer and Communication Engineering, vol.1 No.1 pp. 1096-1100, October 2011.

[3]. Akshay Y et.al. ,Detection of malaria parasites in blood using image processing, International Journal of Advanced Engineering and Research Development Department of Computer Science & Engineering, vol.4,pp.124-126, April 2011.

[4]. S. S. Savkare Moze, et.al, Automatic detection of malaria parasites for estimating parasitemia, ICASSP, vol 5, April 2013.

[5]. Pallavi T. Suradkar, Detection of malarial parasites in blood smear using image processing,International Journal Of Engineering and Innovative technology ,vol 2,April-2013.

[6]. F. Boray Tek, et.al, Malaria Parasite Detection in Peripheral Blood Images, University of South Wales,IEEE Trans.on Systems,Man and Cybernetics,vol 6, pp.325-327,2014.

[7]. Hanung Adi Nugroho ,Son Ali Akbar ,E. Elsa Herdiana Murrhandarwati, Feature extraction and classification for detection Malaria Parasites in thin blood smear., Int.Conference on Information Engineering Technology, Computer and Electrical(ICITACEE), pp.197-201, October,2015.

[8]. Zheng Zhang,et.al, Image classification of unlabeled malaria parasites in red blood cells ,Computer Vision and Image Understanding, vol 114, pp.21-32, 2015.

[9]. Pranati Rakshit, et.al., Detection of presence of parasites in human RBC in case of diagnosing malaria using image processing. , IEEE

Transactions on Pattern Analysis and Machine Intelligence,vol no.6, pp.610-621, 2015.

[10]. Isha Suwalka, et.al., Identify malaria parasite using pattern recognition technique, International Conference On Biomedical Engineering, vol.3, pp.179-187,2017.

[11]. Najeed Ahmed Khan, et.al., Unsupervised identification of malaria parasites using computer vision, IEEE Photonics J., vol.4, No.5, pp.1456-1464, 2017.

[12]. Deepali A. Ghate, Prof. Chaya Jadhav, Automatic Detection of Malaria Parasite from Blood Images International Journal of Computer Science and Security, vol.5, issue.3, pp.310-315, May 2017.

[13]. J.Somasekar, B. Eswara Reddy, E.Keshava Reddy, Ching-Hao Lai, An Image Processing Approach for Accurate Determination of Parasitemia in Peripheral Blod Smear Images , IJCA,vol.10, pp.11- 22, 2017.

[14]. Ms. Snehal Suryawanshi, Prof. V. V. Dixit, Comparative Study of Malaria Parasite Detection using Euclidean Distance Classifier & SVM, vol 2 , Issue 11, November 2017.

[15]. J.SOMASEKAR, Computer Vision for Malaria Parasite Classification in Erythrocytes, International Journal on Computer Science and Engineering (IJCSE) ,vol.5, pp.653-658, June 2017.

[16]. Neetu Ahirwar, Sapnojit. Pattnaik, and Bibhudendra Acharya, Advanced Image Analysis based System for Automatic Detection and Classification of Malarial Parasite In Blood Images, International Journal of Information Technology and Knowledge Management, vol 5, No. 1, pp. 59-64, January-June 2017.

[17]. Kishor Roy, et.al., Detection of malaria parasite in giesma blood sample using image processing , International Journal of Computer Science and Information technology (IJCSJIT), vol.10 No 1,Febrauary 2017.

[18]. Sayali N Dharpal ,Dr .A.V.Malviya, Diagnosis of Malarial Parasite based on image processing , International Research Journal of Engineering and Technology(IRJET), vol.5 Issue:05, May 2017.

[19]. Ashwini Awchite,Komal pasalkar,Mukta killedar, Detection of Malaria and Anemial parasites in blood using image processing , International journal of innovative research in science engineering and technology, vol.6, Issue 5, May 2017.

[20]. Ms.Deepali Ghate, Mrs Chaya Jadhav, Dr N Usha Rani Automatic detection of parasite from blood images, International Journal of Advanced Computer Technology(IJACT), vol 4, Number 1,2011.

[21]. M. Praneesh, Dr. D. Napoleon, 2020, Extraction of Text from Compound Images using Ridler Calvard Based Median Filtering Algorithm, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NSDARM – 2020 (Volume 8 – Issue 04).

[22]. Napoleon, D., et al. "An efficient modified fuzzy possibilistic c-means algorithm for segmenting color based hyperspectral images." IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012). IEEE, 2012.

[23]. Kumar, A., Umurzoqovich, R. S., Duong, N. D., Kanani, P., Kuppusamy, A., Praneesh, M., & Hieu, M. N. (2022). An intrusion identification and prevention for cloud computing: From the perspective of deep learning. Optik, 270, 170044.

[24]. Napoleon, D., et al. "Self-organizing map-based color image segmentation with fuzzy C-Means clustering and saliency map." International Journal of Computer Application 3.2 (2012): 109-117.

[25]. Praneesh, M., and R. Jaya Kumar. "Novel approach for color based comic image segmentation for extraction of text using modify fuzzy possibilistic c-means clustering algorithm." Int J Comput Appl IPRC 1 (2012): 16-18.

[26]. Boonsatit, N., Rajendran, S., Lim, C. P., Jirawattanapanit, A., & Mohandas, P. (2022). New adaptive finite-time cluster synchronization of neutral-type complex-valued coupled neural networks with mixed time delays. Fractal and Fractional, 6(9), 515.

[27]. Napoleon, D., Praneesh, M., Sathya, S., & SivaSubramani, M. (2012). An efficient numerical method for the prediction of clusters using k-means clustering algorithm with bisection method. In Global Trends in Information Systems and Software Applications: 4th International Conference, ObCom 2011, Vellore, TN, India, December 9-11, 2011. Proceedings, Part II (pp. 256-266). Springer Berlin Heidelberg.