Malaria Identification from Thin Blood Sample Images using Artificial Intelligence Methods: LeNet and Snapshot Ensemble

Tripti Singh
Dept. of Computer Science & Engineering
Graphic Era Hill University
Bhimtal, India
triptisd99@gmail.com

Ananya Mehta
Dept. of Computer Science & Engineering
Graphic Era Hill University
Bhimtal, India
mehtaananya40@gmail.com

Shashi Kumar Sharma
Dept. of Computer Science & Engineering
Graphic Era Hill University
Bhimtal, India
12sksharma@gmail.com

Navin Garg
Dept. of Computer Science & Engineering
Graphic Era University
Dehradun, India
ngarg@geu.ac.in

***Abstract*—Malaria is a serious and sometimes fatal disease disease results from parasites transmitted to humans through the bites of infected female *Anopheles* mosquitoes. These parasites belong to the *Plasmodium* family, with *Plasmodium falciparum* identified as the most deadliest species. Malaria has a substantial impact on health of public, and socio-economic development, particularly in endemic regions. It disproportionately impacts children aged five and under and pregnant women. Automating malaria detection offers numerous advantages, including early identification of infection, high accuracy in parasite detection, and faster diagnosis compared to traditional methods. These automated systems improve access to diagnosis, particularly in remote areas, and are ultimately cost-effective. Wide-ranging deep learning methodologies are employed for the detection of the disease, two of them used are LeNet and Snapshot Ensemble with the accuracy 95% and 99.3% respectively.**

***Index Terms*—Malaria, Light Microscopy, RDTs, ML, SVM, Image Processing, VGG, Deep Learning, CNN, Xception, Snapshot Ensemble, Artificial Intelligence.**

I. Introduction

Malaria is the deadliest disease spread worldwide through the bite of *Anopheles* mosquitoes, specifically females. The most extensive propagation of Malaria is in the tropical and sub-tropical countries. Malaria is more prevalent in warm-climate nations due to the favorable conditions for the breeding of *Anopheles* mosquitoes, the primary vectors of the disease [1].

P. falciparum ($\sim$75%) and *P. vivax* ($\sim$25%) are the two most noxious species of the disease known as Malaria. The other ones are *Plasmodium ovale, Plasmodium knowlesi* and *Plasmodium malariae* [1].

The transmission of the *Plasmodium* parasite from mosquitoes to humans and back. Infected mosquitoes inject *sporozoites* into humans during a bite, which travel to the hepatic gland (liver) and multiply. Then, they infect RBCs (red blood cells), causing symptoms. When a mosquito bites an infected person, it ingests the parasite, which multiplies in the mosquito's gut and eventually moves to its salivary glands, ready to infect another human during the next bite.

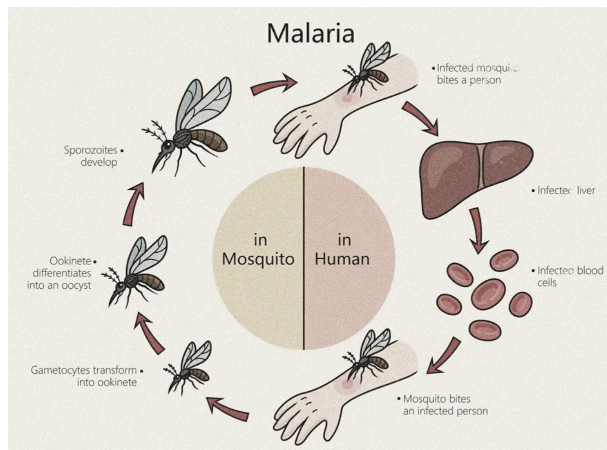

Fig. 1. Life Cycle of Malaria

The two prevalent traditional methods for malaria detection are light microscopy and RDTs. In these manual approaches, a drop of an individual's blood is spread onto a testing slide by a skilled laboratory expert for examination under a microscope. While both methods are cost-effective, they suffer from being time-consuming, prone to errors, and reliant on expert microscopists with proficiency in slide interpretation.

TABLE I
Yearly Analysis of Malaria Cases and Deaths

Year	Cases (in thousand)	Deaths (in thousand)
2018	228,000	567
2019	232,000	568
2020	245,000	628
2021	247,000	619
2022	247,000	608

If the situation demands solely to detect whether a person is infected with Malaria, the experts use 'thick blood films' also known as ’thick blood smears’. Conversely, for the precise determination of the species of the Malaria parasite, ’thin blood films’ also known as ’thin blood smears’ are utilized.

Therefore, automated systems are necessary to address the limitations of traditional manual methods for malaria detection. These systems facilitate early disease detection by speeding up diagnosis and alleviating the burden of microscopists.

II. Literature Survey

The literature survey conducted for the advancement of our paper and project is summarized in Table II.

III. Our Work

A. Dataset

1) Description: This Malaria dataset from Tensorflow comprises 27,558 cell images from segmented thin blood smear slide images. It features an equal distribution of parasitized and uninfected cells (13779 each), providing balanced representations of both conditions [13]. The exemplar dataset is displayed in Fig. 2.

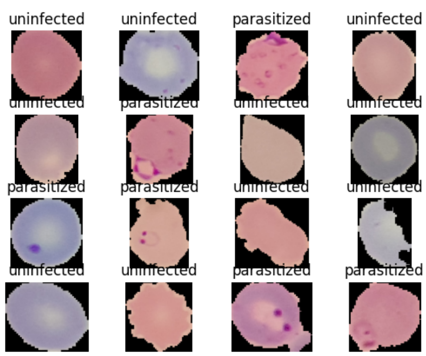

Fig. 2. Images from dataset

2) Data Analysis: Image Preprocessing standardizes image formats and removes noise, ensuring uniformity across all images. Resizing and rescaling are performed in this step, converting image dimensions to (224, 224), enhancing consistency for subsequent analysis. The dataset with this image dimension is displayed in Fig. 3.

Normalization was also performed so that all images fall in a given range.

$$X = \frac{X - Xmin}{Xmax - Xmin}$$

3) Data Segmentation: Algorithms based on deep learning necessitate massive training data to understand the patterns. Therefore, the dataset is segmented in the ratio 8:1:1. The data is divided such that 80% is allocated to the training set, and 10% and 10% are assigned to the test and validation sets, respectively.

TABLE II
Summary of Literature Analysis

Sr No	Author	Year	Methodology	Learning
1	Divyansh Shah et al. [2]	2020	CNN	Aid in identifying malaria and specific other diseases as well.
2	Vijayalakshmi A et al. [3]	2019	VGG, SVM	In reference to performance, the VGG19-SVM integration surpassed that of traditional CNN models.
3	Yuhang Dong et al. [4]	2017	LeNet, GoogLeNet, AlexNet, SVM	Deep convolutional neural networks have achieved higher accuracy than SVM methods and demand minimal human intervention for input.
4	Sumit Kumar et al. [5]	2023	CNN	Accurate malaria detection can still be achieved with just a 2-layer CNN.
5	WHO Malaria Report [1]	2023	Light microscopy and RDTs	RDTs and light microscopy are being supplemented by an expanding vector control toolbox under development.
6	Mehedi Masud et al. [6]	2020	CNN and SGD	CNNs can effectively and accurately assist in detection of malaria from input images in real-time.
7	Rajesh Mayya et al. [7]	2022	Various pre-trained learning approaches and CNN	The Inception-V3 and MobileNetV2 models exhibit higher accuracy levels compared to machine learning algorithms.
8	Aliyu Abubakar et al. [8]	2021	Deep Convolutional Networks	Comprehensive evaluations are furnished concerning precision, recall, accuracy, and computational time.
9	Saurav Mishra [9]	2021	Snapshot Ensemble	A system aimed at precise and rapid malaria detection is crucial for realizing a malaria-free environment.
10	Christonson Berin Jones et al. [10]	2022	CNN, SVM	The proposed scheme is accurate, beneficial, and dependable for parasite detection.
11	Charles Ikerionwu et al. [11]	2022	CNN, various Machine learning algorithms	Systematic analyses are utilized to understand the standard ML algorithms implemented for automating blood film interpretation with an affordable microscope.
12	Suman Kunwar et al. [12]	2018	Image Processing	Construct a new image processing system to recognize the malaria parasites.

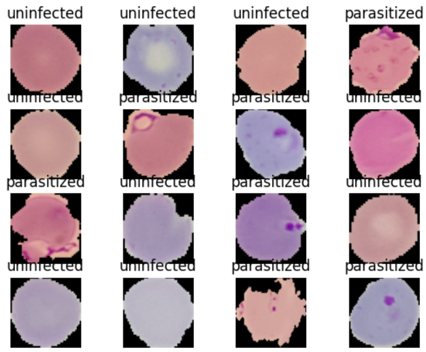

Fig. 3. Images after preprocessing

B. Methodologies

1) LeNet: LeNet, a pioneering convolutional neural network architecture, was introduced by Yann LeCun and his team in 1998 for the task of recognizing handwritten digits. [14], LeNet marked a significant advancement in the field of deep learning. It comprised four layers, including three convolutional layers along with a single fully connected layer. The convolutional layers utilized small kernel sizes and max-pooling operations to derive features from the input images, while the fully connected layer performed classification based on these extracted features [15]. We have used the architecture shown in Fig. 4 with the image size (224, 224).

2) Snapshot Ensemble: The fundamental concept revolves around guiding the model towards numerous local minima points throughout the process of optimization and recording the parameters of model at these points. Throughout training, the neural network navigates via various local minima. Among them, the lowest point is termed the Global Minima. With larger models, the count of parameters and local minima increases [9]. This suggests the existence of distinct clusters of weights and biases where the model makes minimal errors. Each minima serves as a modest yet potential model for addressing the problem. Here, we are using the EfficientNetB0 Snapshot Ensemble model. And the split of dataset is in the ratio 3:1, where the train set has 75% data and the test set has 25% data. By capturing multiple snapshots of weights and biases, we can blend them to create a more robust, generalized model that minimizes errors effectively.

IV. Result Comparison

A. Performance Evaluation

The following are the techniques used for evaluating the performance of the models:

- **Accuracy:** It is the fraction of the summation of the true positive (tpr) and true negative (tnr) to the summation of the predictions.

$$Accuracy = \frac{tpr + tnr}{tpr + tnr + fpr + fnr}$$

```
Model: "sequential_5"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_10 (Conv2D)          (None, 222, 222, 6)       168

 batch_normalization_20 (Ba  (None, 222, 222, 6)       24
 tchNormalization)

 max_pooling2d_10 (MaxPooli  (None, 111, 111, 6)       0
 ng2D)

 conv2d_11 (Conv2D)          (None, 109, 109, 16)      880

 batch_normalization_21 (Ba  (None, 109, 109, 16)      64
 tchNormalization)

 max_pooling2d_11 (MaxPooli  (None, 54, 54, 16)        0
 ng2D)

 flatten_5 (Flatten)         (None, 46656)             0

 dense_15 (Dense)            (None, 100)               4665700

 batch_normalization_22 (Ba  (None, 100)               400
 tchNormalization)

 dense_16 (Dense)            (None, 10)                1010

 batch_normalization_23 (Ba  (None, 10)                40
 tchNormalization)

 dense_17 (Dense)            (None, 1)                 11

=================================================================
Total params: 4668297 (17.81 MB)
Trainable params: 4668033 (17.81 MB)
Non-trainable params: 264 (1.03 KB)
_________________________________________________________________
```

Fig. 4. LeNet model

- **Precision:** It is the fraction of the true positive to the summation of the true positive (tpr) and false positive (fpr).

$$Precision = \frac{tpr}{tpr + fpr}$$

- **Recall or Sensitivity:** It is the fraction of the true positive (tpr) to the summation of the true positive (tpr) and false negative (fnr).

$$Recall = \frac{tpr}{tpr + fnr}$$

- **F1-Score:** It combines the precision (pr) and recall (rc).

$$F1 = \frac{2 * Precision(pr) * Recall(rc)}{Precision(pr) + Recall(rc)} = \frac{2 * tpr}{2 * tpr + fpr + fnr}$$

- **Specificity:** It is the fraction of the true negative to the summation of false positive (fpr) and true negative (tnr).

$$Specificity = \frac{tnr}{fpr + tnr}$$

where,

tpr: True Positive,

tnr: True Negative,
fpr: False Positive, and
fnr: False Negative.

The variables employed for performance evaluation are defined below:

- **True Positive (tpr):** A person who is parasitized (positive) and classified as parasitized (positive).
- **True Negative (tnr):** A person who is not parasitized (negative) and classified as not parasitized (negative).
- **False Positive (fpr):** A person who is not parasitized (negative) and classified as parasitized (positive).
- **False Negative (fnr):** A person who is parasitized (positive) and classified as not parasitized (negative).

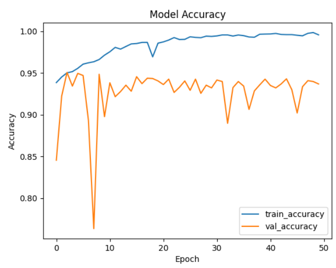

Fig. 5. LeNet model accuracy graph

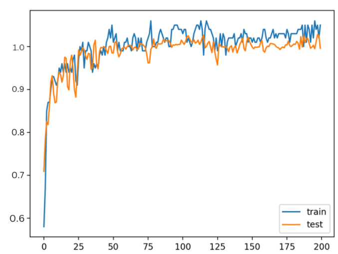

Fig. 6. EfficientNetB0 Snapshot Ensemble model accuracy graph

B. Error Sanctioning

For error sanctioning in binary classification problems, generally the binary cross-entropy loss [16]. In binary classification, where the classes, m = 2.

$$Cross - entropy = (-a)log(p) + (1 - a)log(1 - p) \quad (1)$$

a: actual value, p: predicted value

TABLE III
RESULTS OF MODELS USED IN THIS RESEARCH PAPER

Methodology	Epochs	Image Size	Accuracy
LeNet	50	(224, 224)	94.84%
EfficientNetB0	200	(224, 224)	99.32%

V. CONCLUSION

For this work, we have done a comparative study of how the automation of the malaria detection outperforms the traditional methods, and how the machine learning and deep learning techniques did a commendable job in the healthcare sector. In our project, we have used two CNN methodologies LeNet (with 50 epochs) and EfficientNetB0 (with 200 epochs), giving us the accuracy of 94.84% and 99.32% respectively on the portion of dataset used for testing. Therefore, further researches can be implemented considering these results.

REFERENCES

[1] W. H. Organization *et al.*, "World malaria report 2022. geneva: World health organization; 2022. licence: Cc by-nc-sa 3.0 igo," tech. rep., accessed 3–1–2023). https://www.who.int/teams/global-malaria-programme ....

[2] D. Shah, K. Kawale, M. Shah, S. Randive, and R. Mapari, "Malaria parasite detection using deep learning:(beneficial to humankind)," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 984–988, IEEE, 2020.

[3] A. Vijayalakshmi, "Deep learning approach to detect malaria from microscopic images," *Multimedia Tools and Applications*, vol. 79, pp. 15297–15317, 2020.

[4] Y. Dong, Z. Jiang, H. Shen, W. D. Pan, L. A. Williams, V. V. Reddy, W. H. Benjamin, and A. W. Bryan, "Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells," in *2017 IEEE EMBS international conference on biomedical & health informatics (BHI)*, pp. 101–104, IEEE, 2017.

[5] S. Kumar, S. Priya, and A. Kumar, "Malaria detection using deep convolution neural network," *arXiv preprint arXiv:2303.03397*, 2023.

[6] M. Masud, H. Alhumyani, S. S. Alshamrani, O. Cheikhrouhou, S. Ibrahim, G. Muhammad, M. S. Hossain, and M. Shorfuzzaman, "Leveraging deep learning techniques for malaria parasite detection using mobile application," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–15, 2020.

[7] R. Mayya, S. Chopra, and M. Vijayalakshmi, "Study on malaria detection using transfer learning,"

[8] A. Abubakar, M. Ajuji, and I. U. Yahya, "Deepfmd: computational analysis for malaria detection in blood-smear images using deep-learning features," *Applied System Innovation*, vol. 4, no. 4, p. 82, 2021.

[9] S. Mishra, "Malaria parasite detection using efficient neural ensembles," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 3, no. 3, pp. 119–133, 2021.

[10] C. B. Jones and C. Murugamani, "Malaria parasite detection on microscopic blood smear images with integrated deep learning algorithms.," *Int. Arab J. Inf. Technol.*, vol. 20, no. 2, pp. 170–179, 2023.

[11] C. Ikerionwu, C. Ugwuishiwu, I. Okpala, I. James, M. Okoronkwo, C. Nnadi, U. Orji, D. Ebem, and A. Ike, "Application of machine and deep learning algorithms in optical microscopic detection of plasmodium parasites: A malaria diagnostic tool for the future," *Photodiagnosis and Photodynamic Therapy*, p. 103198, 2022.

[12] S. Kunwar, "Malaria detection using image processing and machine learning," *arXiv preprint arXiv:1801.10031*, 2018.

[13] S. Rajaraman, S. K. Antani, M. Poostchi, K. Silamut, M. A. Hossain, R. J. Maude, S. Jaeger, and G. R. Thoma, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, p. e4568, 2018.

[14] H. Lin, Y. Wang, J. Zhou, S. Pei, Z. Geng, Z. Chen, and Z. Guo, "Research and design of open convolutional neural network based on fpga," *IOP Conference Series: Earth and Environmental Science*, vol. 714, p. 042015, mar 2021.

[15] Wikipedia contributors, "Lenet — Wikipedia, the free encyclopedia." https://en.wikipedia.org/w/index.php?title=LeNetoldid=1189193486, 2023. [Online; accessed 1-May-2024].

[16] Wikipedia contributors, "Cross-entropy — Wikipedia, the free encyclopedia." https://en.wikipedia.org/w/index.php?title=Cross-entropyoldid=1221840853, 2024. [Online; accessed 2-May-2024].