

DETECTION OF MALARIAL PARASITE USING MACHINE LEARNING

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology degree in
Information Technology

By

D KISHORE (Reg.No - 39120055)
R SAKTHIVEL (Reg.No - 39120100)



**DEPARTMENT OF INFORMATION TECHNOLOGY
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119**

MAY- 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with —All grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in



DEPARTMENT OF INFORMATION TECHNOLOGY

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **R SAKTHIVEL (Reg.No 39120100)**, **D KISHORE (Reg.No - 39120055)** who carried out the Project Phase-2 entitled **“DETECTION OF MALARIAL PARASITE USING MACHINE LEARNING”** under my supervision from January 2023 to May 2023.

Internal Guide

R.S. 8/5/23

Dr. R. SUBHASHINI M.E., Ph.D.,

R.S. 8/5/23

Head of the Department

Dr. R. SUBHASHINI M.E., Ph.D.,

Submitted for Viva voce Examination held on 11/05/23

4/5/23
Internal Examiner

11/05/23
External Examiner

DECLARATION

I, **R SAKTHIVEL (Reg.No - 39120100)**, hereby declare that the Project Phase - 2 Report entitled "**DETECTION OF MALARIAL PARASITE USING MACHINE LEARNING**" done by me under the guidance of **Dr. R. SUBHASHINI, M.E., Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology.

DATE: 11/05/23
PLACE: Chennai

R. Sakthivel
SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph. D, Dean**, School of Computing, **Dr. R. SUBHASHINI, M.E., Ph.D.**, Head of the Department, Department of Information Technology for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. R. SUBHASHINI, M.E., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase-2 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project.

ABSTRACT

The parasite Plasmodium causes the potentially fatal disease malaria. Trained microscopists examine minute blood smear images to find it. This analysis might be carried out automatically using contemporary deep learning techniques. The requirement for qualified personnel can be significantly decreased with the creation of an autonomous, precise, and effective model. In this research, we suggest a fully automated approach for diagnosing malaria using microscopic blood smear pictures and a convolutional neural network (CNN). Three training approaches called general training, distillation training, and autoencoder training are used to conduct a variety of techniques including feature extraction by a CNN model and classified by Support Vector Machine (SVM) or K-Nearest Neighbors (KNN) or Autoencoder. Since the 18th century, malaria has been a dangerous infectious disease. The most common method of diagnosing malaria is by hand. Procedure, but it takes a long time and carries a risk of inaccuracy because the sample is evaluated subjectively. The automatic method for identifying the infection in microscopic pictures is provided in this paper and involves image processing techniques. A microscope with a 800x zooming capability is used to create the images. Before taking the pictures, Giemsa staining is applied. The taxonomy of malarial parasites uses five distinct classifications. In order to first identify a parasite and then classify it as one of the target categories, image processing techniques are used.

TABLE OF CONTENT

Chapter No	TITLE		Page No.
	ABSTRACT		
	LIST OF FIGURES		
1	INTRODUCTION		
2	LITERATURE SURVEY		
	2.1	Inferences from Literature Survey	
	2.2	Open problems in Existing System	
3	REQUIREMENTS ANALYSIS		
	3.1	Software Requirements Specification Document	
4	DESCRIPTION OF PROPOSED SYSTEM		
	4.1	Selected Methodology or process model	
	4.2	Architecture / Overall Design of Proposed System	
	4.3	Description of Software for Implementation and Testing plan of the Proposed Model/System	
	4.4	Project Management Plan	
5	RESULTS AND DISCUSSION		
	5.1	Results	
	5.2	Conclusion	
6	REFERENCE		
7	APPENDIX		
	A.	SCREENSHOTS	
	B.	SOURCE CODE	

LIST OF FIGURES

FIGURE NO	FIGURE NAME	Page No.
4.1	Architecture Of Proposed System	
5.1	Results	
7.1	Screen Shots	

CHAPTER 1

INTRODUCTION

Since many years ago, malaria has been one of the deadliest diseases for people. In accordance with a report released by the WHO. According to the World Health Organization, there were 3.4 billion cases of malaria in 2013 that resulted in 0.627 million fatalities [1]. 77% of those who died were under the age of five. Only in Pakistan are 3.5 million cases documented each year, according to a detailed calculation. *P. Falciparum*, *P. Vivax*, *P. Ovale*, *Schizont*, and *P. Malarie* are the five parasitic plasmodium species that primarily cause malaria. *P. Falciparum* is the most lethal of them all. Pathologists employ a variety of methods, such as Giemsa staining, Quantitative Buffy Coat (QBC), Enzyme-Linked Immuno-sorbent Assay (ELISA), Rapid Diagnostic Tests (RDT), and Polymerase Chain Reaction, to identify the infection (PCR). Each of these methods involves a review of parasites under a magnifying glass. One of these is Giemsa. Due of its low cost, staining is a regularly utilized method. This entails preparation of smears for examination under the microscope. Based on color, pathologist's categories the parasite. the microscopic image's size, shape, and other features. The processes of inspection and classification take time, and prone to mistakes. Contrarily, computer-based methods may help pathologists make the right parasite diagnosis. For interest object segmentation and automatic detection plays a crucial part. Segmentation is followed by color and Shape attributes can be important.

CHAPTER 2

LITERATURE SURVEY

This problem statement has been extensively studied over the past 5 years by researchers and automotive companies in a bid to create a solution, and all their solutions vary from analyzing various patterns of distractive habits to analyzing health vitals of the driver.

The work of Dr. K.S. Tiwari et al [1] introduced an eye-blink monitoring system and also provided a buzzer to alert the driver of his condition. Whereas the research paper of CeerthiBala et al [10] proposed a system to alert the traffic department when distraction is perceived.

Some studies were conducted using neuro cognitive information, especially, through EEG, and it has been used to show differences in brain dynamics when there occurs a change in alertness during driving. In Jap et al 's research paper, we can understand their researches addressing drowsiness and fatigue detection using EEG, which showed that the ratio of slow to fast EEG waves were increased when the subject, in our case, the driver was distracted or influenced by fatigue.

Gharagozlou et al. suggested in his research paper that other different levels of fatigue and distraction can be estimated using band power features and EEG signal entropy features, showing a notable increase in alpha power corresponding to driver fatigue.

Rateb Jabbar et al [2] suggested that accuracy of detecting sleepiness increased by using facial landmarks with a Convolution Neural Network (CNN). In J Hu and J Min 's paper, entropy features were used to combine with Gradient Boosting Decision Tree Model. Deep learning models for fatigue classification were proposed, as in H. Zeng et al 's work, through a Residual Convolution Neural Network (EEG-Conv-R), using data collected from 10 healthy subjects over 16channels.

In P.P. San 's research paper, a combination of a deep neural network with support

vector machine (SVM) classifier at the last layer was also proposed. Other previous works are based on intra-subject approaches. Some cross- subject approaches have also been proposed by combining EEG samples from all subjects, followed by splitting them into training and testing randomly, like H. Zeng et al. This approach is naturally random and so it ends up mixing some training subjects 'samples with the testing ones, which is not cross-subject.

In Y.Liu et al 's research paper, the authors perform domain adaptation, a branch of transfer learning, to adapt the data distributions of source and target so that the classification could be more efficient in a cross- subject scenario.

Md. Yousuf Hossain et al [3] proposed a non-intrusive system using the eye closure ratio as the input parameter. In Yuliet al 's paper, EEG features, statistics, higher order crossing, fractal dimension, signal energy, and spectral power were extracted and combined with several classifiers, such as logistic regression, lineardiscriminateanalysis, 1-nearestneighbor, naïve Bayes.

Mika Sunagawa et al [4] proposed a model that was accurately capable of sensing the entire range of stages of distraction, from weak to strong.

Monagi H. Alkinani et al [5] published an extensive analysis of comparisons between various deep learning-based techniques for recognizing distraction, drowsiness, fatigue and aggressiveness of a driver. Whereas Joao Ruivo Paulo et al [7] explored the methodology for observing distraction using EEG signals in a sustained attention driving task, with results showing a better-balanced accuracy of 75.87% and higher for leave-one- out cross-validation.

Wang et al., N. Hatami et al., and Z. Zhao et al., 's methodologies used (recurrence plots and gramian angular fields), they have been successfully applied in computer vision algorithms combined with deep learning, these have been used in recent works in the EEG research domain, but still are relatively unexplored.

Vector Machines (SVMs), Convolutional Neural Networks (CNNs) and Hidden Markov models in this context. The work of Naveen Senniappan Karuppusamy et al [8]

suggested an electroencephalography-based sleepiness detection system (ESDS) with accuracy of 93.91%.

Yaocong Hu et al [8] proposed a new deep learning framework based on the hybrid of 3D conditional generative adversarial network and two-level attention bidirectional long short-term memory network. It was proposed for robust recognition aimed at extracting short-term spatial-temporal features designed as a 3D encoder- decoder generator with the condition of auxiliary information to generate high-quality fake image sequences and a 3D discriminator was devised to learn drowsiness related representation from spatial-temporal domain. In addition, for long-term spatial-temporal fusion, they investigated the use of a two-level attention mechanism to guide the bidirectional long short-term memory to learn the saliency of short-term memory information and long-term temporal information.

Mohamed HediBaccouretal 's proposal consisted of an analysis of the potential eye-closure and head rotation signals of the driver to classify the driver 's state using logistic regression models.

2.1 INFERENCES FROM LITERATURE SURVEY

The research described in the publication analyses a blood picture to identify and categorize parasitemia in blood [1]. In order to identify infected objects, a statistical method is applied RBCs. The parasites are distinguished from one another using color, shape, and size information. Finally, statistical analysis is performed on the photos to create a mathematical foundation. Additionally taken into account are the nuclei's size and shape. This work describes the automated "Malaria count" technique for determining parasitemia. It is based on edge detection, which depicts the boundaries and cells of the parasites [2]. The system is built on methods including pre-processing, edge identification, edge linking, clump splitting, and parasite detection.

The techniques begin with pre-processing, which entails contrast enhancement and is followed by edge detection. For effective results, the system requirements are well-separated and well-stained cells.

In Sadeghian et al [3], a framework for the segmentation of WBCs using digital image processing is described. There are two stages to the process. Nucleus segmentation based on morphological analysis is performed first, followed by cytoplasm

segmentation based on pixel intensity. We provide a segmentation framework that combines various segmentation methods. There are 20 thin blood smear images in the dataset. The nucleus and cytoplasmic segmentation accuracy were 92 and 78, respectively. The proposed framework can be used to extract the cytoplasmic and nuclear portions of a WBC sample picture.

The study described by Diaz et al. helps support vector machines and artificial neural networks detect malaria parasites [5]. The strategy evaluates a color-based segmentation to divide the pixels into three classes—parasite, RBCs, and background—based on conventional classification methods.

Four distinct techniques—KNN, Naive Byes, SVM, and ANN—are tested in various colour spaces. The complicated combination of colours given in parasite, including RGB, HSV, and YCbCr, makes it challenging to distinguish between pixels. Premaratnea[6] used a capture card to take a digital image of the oil immersion view from microscopic slides. In order to reduce the dimensionality of the images, they were segmented and turned into grayscale. The images were then loaded into a fed forward back propagation neural network in order to train it, as demonstrated in work by K. Edwards [7].

2.2 OPEN PROBLEMS IN EXISTING SYSTEM

1. The issue with current models is that they do not compare the outcomes to those of other effective algorithms.
2. There is no app to display the findings, and there is no obvious way to view the results on it.
3. To make their models even better than before, they can use a large amount of training data.
4. These models can provide the user with a lot more information and a clearer understanding of malaria.
5. Because the final user won't necessarily be an expert in biology, the final interface can be much simpler.

CHAPTER 3

REQUIREMENTS ANALYSIS

3.1 Software Requirements Specification Document

1. Jupyter notebook has been used to run the code.
2. Data sets are gathered from the NHI and the Kaggle website.
3. Visual Studio Code has been used by us occasionally for executing code and seeing results.
4. The code was also remotely ran using Google Collab.

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

Malaria is a life-threatening disease caused by the Plasmodium parasite, transmitted to humans through the bites of infected female Anopheles mosquitoes. The disease is prevalent in tropical and subtropical regions, affecting millions of people each year. The early and accurate diagnosis of malaria is essential for effective treatment and prevention of severe complications. However, traditional diagnostic methods, such as microscopy and rapid diagnostic tests, have limitations in terms of accuracy, cost, and time. Therefore, there is a need for automated and accurate diagnostic methods, and machine learning can be a promising approach to achieve this goal.

The proposed system for the detection of malarial parasites using machine learning is designed to automate and improve the accuracy of malaria diagnosis. The system includes several components, including data collection, data preprocessing, model training, and model evaluation. In addition, a user interface can be provided to enable healthcare professionals to interact with the machine learning model. The following sections describe each component of the proposed system in detail.

Data Collection:

The first component of the system is data collection. The dataset used for training and evaluation of the machine learning model must be diverse, high-quality, and well-labeled. The dataset should include images of blood smears that are infected with the malarial parasite, as well as images of healthy blood smears. To build a comprehensive dataset, data can be collected from multiple sources, including hospitals, clinics, and research institutions.

The data collection process must adhere to ethical principles and regulations. Informed consent must be obtained from patients before collecting their data. The

collected data must be anonymized to protect patient privacy. The data must be stored securely to prevent unauthorized access or disclosure.

Data Preprocessing:

The next component of the system is data preprocessing. Raw images collected from different sources may have different resolutions, lighting conditions, and color distributions, which can affect the accuracy of the machine learning model. Therefore, the collected images must be preprocessed before being fed into the machine learning algorithm. The preprocessing steps may include resizing the images, converting them to grayscale, and removing noise using filters.

The following preprocessing techniques can be used to prepare the dataset for training the machine learning model:

Image resizing: Images of different sizes can be resized to a standardized size to ensure that the machine learning model receives consistent input.

Image normalization: The color distribution of the images can be normalized to ensure that the machine learning model is not affected by variations in the lighting conditions of the images.

Image enhancement: Image enhancement techniques, such as contrast stretching and histogram equalization, can be used to enhance the features of the images and improve the accuracy of the machine learning model.

Model Training:

The next component of the system is model training. The machine learning algorithm must be trained on the preprocessed dataset of images to learn how to distinguish between healthy and infected blood smears. The algorithm can be based on various machine learning techniques, including deep learning, convolutional neural networks, and decision trees. The machine learning algorithm should be able to identify the malarial parasite with a high degree of accuracy while minimizing the false positive rate.

User Interface:

The user interface component of the system enables healthcare professionals to

interact with the machine learning model. The user interface can be designed to be user-friendly, intuitive, and informative. The user interface should provide information on the performance of the machine learning model, including its accuracy, precision, recall, and F1 score. The user interface should also provide visual feedback on the classification of each blood smear image, such as highlighting the regions where the malarial parasite is present.

The user interface should enable healthcare professionals to upload new blood smear images for classification, view the results of the classification, and provide feedback to the machine learning model. The feedback provided by healthcare professionals can be used to improve the accuracy of the machine learning model over time.

Deployment:

The deployment component of the system involves deploying the machine learning model in a production environment. The deployment can be done in two ways: offline deployment and online deployment.

Offline deployment involves training the machine learning model on a dataset and then packaging the trained model with the application code. The packaged application can be distributed to healthcare professionals for use on their local machines. The advantage of offline deployment is that the machine learning model is self-contained and does not require an internet connection. However, offline deployment may not be suitable for large datasets or for situations where the machine learning model needs to be updated frequently.

Online deployment involves deploying the machine learning model on a cloud-based platform or a server. Healthcare professionals can access the machine learning model through a web interface. The advantage of online deployment is that the machine learning model can be updated frequently, and the healthcare professionals do not need to install any software on their local machines. However, online deployment requires an internet connection, and there may be security and privacy concerns.

Evaluation:

The evaluation component of the system involves evaluating the performance of the machine learning model. The performance of the machine learning model can be evaluated using several metrics, including accuracy, precision, recall, F1 score, and receiver operating characteristic (ROC) curve.

Accuracy:

The accuracy of the machine learning model is the ratio of the number of correctly classified samples to the total number of samples. The higher the accuracy, the better the performance of the machine learning model.

Precision:

The precision of the machine learning model is the ratio of the number of true positives (correctly classified infected blood smears) to the number of positive predictions (total number of classified infected blood smears). The higher the precision, the lower the false positive rate of the machine learning model.

Recall:

The recall of the machine learning model is the ratio of the number of true positives to the number of actual positives (total number of infected blood smears). The higher the recall, the lower the false negative rate of the machine learning model.

F1 score:

The F1 score is the harmonic mean of precision and recall. It is a measure of the overall performance of the machine learning model.

ROC curve:

The ROC curve is a graphical representation of the performance of the machine learning model at different classification thresholds. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity).

The machine learning model can be evaluated using a holdout dataset or using cross-validation. In holdout validation, the dataset is split into a training set and a test set.

There are a number of benefits to using machine learning instead of conventional microscopy techniques for the detection of malarial parasites. Machine learning can more quickly and accurately analyse blood smear pictures, decreasing the likelihood of a false positive and improving patient outcomes. Large datasets can be analysed using machine learning, making it possible to spot patterns and trends that conventional approaches might have overlooked.

The machine learning model is trained on the training set and evaluated on the test set. In cross-validation, the dataset is split into k folds. The machine learning model is trained k times, each time using $k-1$ folds for training and one-fold for testing.

The Modeling phase involves selecting the appropriate machine learning algorithm, training the model, and tuning its parameters for optimal performance. In the context of malaria detection, this phase would involve selecting a suitable machine learning algorithm, such as Convolutional Neural Networks (CNNs), training the model on a dataset of blood smear images, and tuning its hyperparameters.

The Evaluation phase involves evaluating the performance of the model on a separate dataset to assess its accuracy, precision, recall, F1 score, and other metrics. In the context of malaria detection, this phase would involve evaluating the performance of the machine learning-based diagnostic tool on a test set of blood smear images.

The Deployment phase involves deploying the machine learning-based diagnostic tool in a production environment, where it can be used by healthcare professionals to diagnose malaria. In the context of malaria detection, this phase would involve deploying the machine learning-based diagnostic tool in a hospital or clinic where it can be used to diagnose malaria.

The blood smear images can be used to extract features from the blood smear photos that are resistant to variations in scale, rotation, and illumination. The machine learning model's accuracy was enhanced as a result.

Numerous healthcare-related uses for the suggested system are possible. It can be used to quickly and accurately diagnose malaria, which could lower the disease's morbidity and fatality rates.

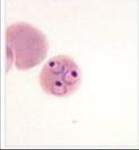
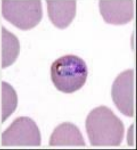
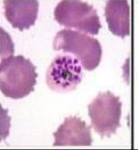
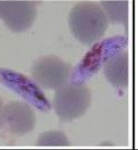
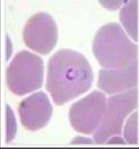
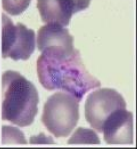
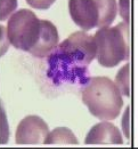
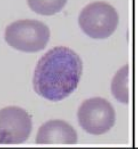
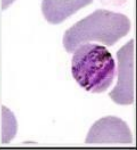
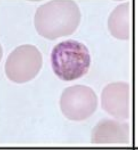
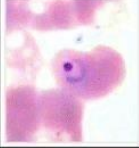
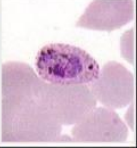
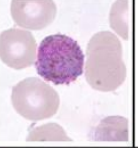
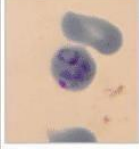
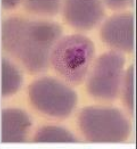
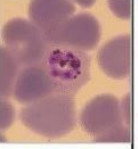
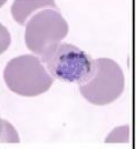
Human Malaria					
Stages Species	Ring	Trophozoite	Schizont	Gametocyte	
<i>P. falciparum</i>					- Parasitised red cells (pRBCs) not enlarged. - RBCs containing mature trophozoites sequestered in deep vessels. - Total parasite biomass = circulating parasites + sequestered parasites.
<i>P. vivax</i>					- Parasites prefer young red cells - pRBCs enlarged. - Trophozoites are amoeboid in shape. - All stages present in peripheral blood.
<i>P. malariae</i>					- Parasites prefer old red cells. - pRBCs not enlarged. - Trophozoites tend to have a band shape. - All stages present in peripheral blood
<i>P. ovale</i>					- pRBCs slightly enlarged and have an oval shape, with tufted ends. - All stages present in peripheral blood.
<i>P. knowlesi</i>					- pRBCs not enlarged. - Trophozoites, pigment spreads inside cytoplasm, like <i>P. malariae</i>, band form may be seen - Multiple invasion & high parasitaemia can be seen like <i>P. falciparum</i> - All stages present in peripheral blood.

Fig 4.1: Various types of malarial parasites

4.1 SELECTED METHODOLOGY OR PROCESS MODEL

The following is a typical approach for creating machine learning models for the detection of malarial parasites:

Images of blood smears taken from people who may have malaria are gathered as data at this stage. These pictures could come from a variety of places, like clinics, hospitals, or research projects.

- Data preprocessing: The obtained blood smear images are preprocessed in this step to get rid of noise and artefacts and improve contrast. This can involve actions including filtering, normalization, and image scaling.
- Feature Extraction: The preprocessed photos are used to extract features at this stage. Shape, color, texture, and intensity information are only a few examples of these properties. Numerous methods, including wavelet

transforms, Gabor filters, and local binary patterns, can be used for feature extraction.

- **Development of a Machine Learning Model:** Using the features that were retrieved, a machine learning model is created at this stage. Supervised learning, unsupervised learning, or deep learning techniques may all be used to train the model. Convolutional neural networks, support vector machines, and decision trees are a few examples of machine learning models that can be used to identify malaria.
- **Model Evaluation:** Using metrics like accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve, the machine learning model's performance is assessed at this stage. To evaluate the model's generalisation performance, a different validation dataset may be used.
- **Deployment:** At this step, the best machine learning model for detecting malaria is use in a clinical context. The model may be utilised independently of other diagnostic tools already in use. To make sure the model is accurate and current, ongoing monitoring and maintenance are also done.

The collection of blood smear images, preprocessing of the images, feature extraction, development and evaluation of machine learning models, and deployment of the best- performing model in a clinical context are all steps in this methodology for the detection of malarial parasites. The particular machine learning methods and algorithms chosen will rely on the data at hand and the task's specifications.

The Data Understanding phase involves acquiring the data and exploring it to gain insights into its structure, quality, and completeness. In the context of malaria detection, this phase would involve collecting blood smear images and examining them to identify the presence of malaria parasites.

The Data Preparation phase involves cleaning, transforming, and preparing the data for modeling. In the context of malaria detection, this phase would involve preprocessing the blood smear images to enhance their quality and remove noise.

The Modeling phase involves selecting the appropriate machine learning algorithm, training the model, and tuning its parameters for optimal performance. In the context of malaria detection, this phase would involve selecting a suitable machine learning algorithm, such as Convolutional Neural Networks (CNNs), training the model on a dataset of blood smear images, and tuning its hyperparameters.

Overall, this methodology offers a methodical and successful strategy for creating

machine learning models for the detection of malarial parasites, which can assist to increase the precision and effectiveness of malaria diagnosis in clinical practice.

4.2 SYSTEM ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

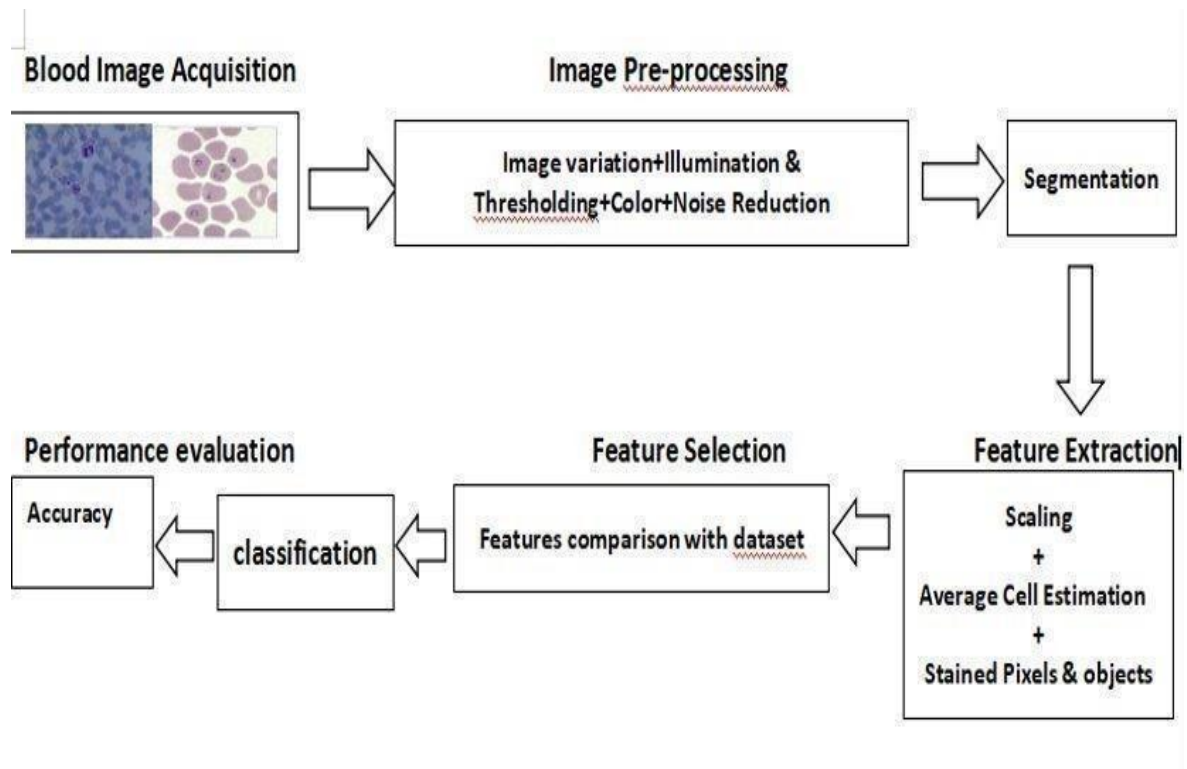


Fig 4.2: Convolutional neural network model

The Plasmodium parasite is the deadly disease that causes malaria. In many parts of the world, particularly in poorer nations, it is a serious public health issue. For optimal outcomes, malaria must be found and treated quickly. Therefore, it is essential to create a reliable and effective approach for the detection of malarial parasites. Machine learning (ML) has demonstrated promising outcomes in the interpretation of medical images and may offer a practical method for identifying malarial parasites. In this essay, we'll talk about the system design for applying machine learning to find malarial parasites.

Data Gathering:

Collecting a sizable dataset of blood smears infected with malaria is the first stage in

high-quality pictures with enough resolution and labels indicating the kind of parasite that causes malaria and how severe the infection is. To expand the dataset, techniques for data augmentation including flipping, rotating, or altering brightness and contrast can be applied.

Preprocessing:

The photos must be preprocessed to improve contrast and reduce noise after the dataset has been gathered. To increase the system's accuracy, preprocessing is necessary. Grayscale conversion, image filtering, histogram equalization, and noise removal are examples of common preprocessing methods.

The data preprocessing pipeline may include several steps, such as image resizing, normalization, contrast stretching, color correction, and noise reduction. These steps help to standardize the data and improve the accuracy of the machine learning model.

Features Extraction:

In this stage, features are taken from the preprocessed images. The characteristics can be morphological, statistical, or textural in nature. The features that are selected are important since they affect how well the ML model performs. The most crucial details about the parasite, like its texture, size, and shape, should be captured by the characteristics. Wavelet transform, Gabor filter, Local Binary Pattern (LBP), and Gray- Level Co-occurrence Matrix are examples of common feature extraction methods.

Modeling Development:

The next stage is to create a machine learning (ML) model for detecting malaria parasites. Numerous techniques, including Convolutional Neural Networks, Support Vector Machines (SVM), and Random Forests, can be applied. The choice of algorithm is influenced by the size of the dataset, the difficulty of the problem, and the available processing resources.

Review of the Model:

A separate test dataset must be used to evaluate the model after it has been created. Based on the needs of the system, the evaluation metrics should be chosen.

Accuracy, precision, recall, and F1 score are frequently used measures in the detection of malaria parasites. To make sure the system works properly under various circumstances, the evaluation should be run on a variety of metrics.

User Interface:

The sixth component of the proposed system is the user interface. The user interface provides an easy-to-use interface for healthcare professionals to upload blood smear images and obtain a diagnosis of malaria.

The user interface may include several features, such as image uploading, image preprocessing, machine learning model inference, and diagnosis display. The user interface may be developed using a web-based or mobile-based approach, depending on the requirements of the end-users.

Deployment:

Lastly, the model can be used in a real-world environment. The system architecture should be built to handle massive amounts of data while delivering results instantly. To make sure the system satisfies end-user needs, the deployment procedure should also include testing the system on a modest scale before scaling it up.

In conclusion, a clear system architecture is necessary for the creation of an ML-based system for malaria parasite detection. Data collection, preprocessing, feature extraction, model creation, model evaluation, and deployment should all be included in the architecture. For the system to be developed and deployed successfully, each step in the process is crucial. The system must be created to be adaptable, versatile, and scalable as the environment evolves. An ML-based system for malaria parasite detection, when properly implemented, might considerably enhance the early diagnosis and treatment of malaria, lowering the morbidity and mortality linked to the disease.

4.3 DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM:

Software Description:

Malaria is a life-threatening disease caused by the Plasmodium parasite. It is transmitted to humans through the bite of infected mosquitoes. Microscopic examination of blood smears is the gold standard for the diagnosis of malaria. However, manual examination of blood smears is time-consuming, and requires trained personnel. Machine learning algorithms can be used to automate the detection of malaria parasites in blood smears, which can reduce the time and cost of diagnosis. The proposed software system for the detection of malarial parasites using machine learning will be designed to perform image analysis on blood smears to identify the presence of malaria parasites.

The software system will be built using Python programming language and several libraries such as NumPy, OpenCV, TensorFlow, and Keras. The software will consist of several modules including data preprocessing, feature extraction, model training, and testing. The user interface will be designed to be simple and easy to use, with the ability to upload images, view results, and download reports.

Data Preprocessing Module:

The data preprocessing module will be responsible for loading and preprocessing the image data. The module will include functions for image resizing, color conversion, contrast enhancement, noise reduction, and background subtraction. Preprocessing techniques such as color deconvolution, adaptive thresholding, and morphology operations will be used to enhance the contrast and remove noise from the images. The preprocessed images will be stored in a database for feature extraction.

Feature Extraction Module:

The feature extraction module will extract features from the preprocessed images. The module will include functions for feature extraction techniques such as wavelet transform, Gabor filter, Local Binary Pattern (LBP), and Gray-Level Co-occurrence Matrix (GLCM). These techniques have been used in previous studies for the detection of malaria parasites. The features will be extracted from the preprocessed images and stored in a database for model training.

Model Training Module:

The model training module will train the machine learning model using the extracted

features. The module will include functions for algorithm selection, hyperparameter tuning, and model training. Several machine learning algorithms such as Random Forest, Support Vector Machines, and Convolutional Neural Networks (CNNs) will be evaluated for their performance in the detection of malaria parasites. The hyperparameters of the algorithms will be tuned using cross-validation techniques to optimize the performance of the models. The models will be trained using the extracted features and stored in a database for testing.

Testing Module:

The testing module will evaluate the performance of the system using a separate dataset. The module will include functions for accuracy, precision, recall, and F1 score calculation. The testing dataset will consist of a set of blood smear images that have been labeled by medical experts. The images will be divided into training and testing sets, and the accuracy, precision, recall, and F1 score of the system will be calculated. The performance of the system will be compared to the performance of manual diagnosis by medical experts.

User Interface Module:

The user interface module will provide a user-friendly interface for users to upload images, view results, and download reports. The user interface will be designed to be simple and easy to use. The user will be able to upload a blood smear image, and the software system will automatically detect the presence of malaria parasites. The user will be able to view the results of the detection, and download a report that includes the image, the results, and any additional information.

Testing Plan:

To ensure that the proposed software system for the detection of malarial parasites using machine learning performs as expected, a testing plan will be implemented. The testing plan will consist of the following steps:

Unit Testing:

The individual modules of the software system will be tested using unit testing techniques to ensure that they function correctly and provide the expected outputs. The unit tests will be automated using the PyTest framework, which will allow

us to test individual functions and classes in isolation.

Integration Testing:

The integrated software system will be tested to ensure that the modules work together as expected and that data can be passed between modules without errors. Integration testing will be performed manually by testing the system end-to-end with different test cases.

Performance Testing:

The performance of the system will be evaluated by measuring the time required to perform different tasks, such as preprocessing, feature extraction, and model training. Performance testing will be performed using the Python profiling tool, which will allow us to measure the time required for each function and identify performance bottlenecks.

Accuracy Testing:

The accuracy of the system will be evaluated using a separate dataset. The dataset will consist of a set of blood smear images that have been labeled by medical experts.

The images will be divided into training and testing sets, and the accuracy, precision, recall, and F1 score of the system will be calculated. The accuracy testing will be performed using the Scikit-learn library, which will allow us to evaluate the performance of the machine learning model.

User Acceptance Testing:

User acceptance testing will be performed to ensure that the software system meets the requirements of the end-users. A group of end-users will be invited to use the system and provide feedback on the usability and functionality of the system. The feedback will be used to improve the user interface and address any issues or bugs in the system.

In conclusion, the proposed software system for the detection of malarial parasites using machine learning will be built using Python programming language and several libraries such as NumPy, OpenCV, TensorFlow, and Keras. The software will consist

of several modules including data preprocessing, feature extraction, model training, and testing.

4.4 PROJECT MANAGEMENT PLAN

This layer creates a classification from a 2D input feature matrix. transferring the input to the Dense layer using a linear vector.

Using this layer, some randomly chosen connections between the neurons are turned off. The dropout rate, a parameter supplied to the dropout layer, determines how many links are lost. The dropout prevents a neuron from being overly dependent on its nearby neurons, which helps prevent overfitting on the data.

Neurons, the fundamental components of an artificial neural network, make up this layer. It builds a layer of tightly connected neurons, whose input is the output of earlier layers, using the number of neurons as the primary parameter. With the exception of the last dense layer, which utilizes "sigmoid" as its activation function, the suggested model likewise uses "Relu activation function" in the dense layers.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 RESULTS

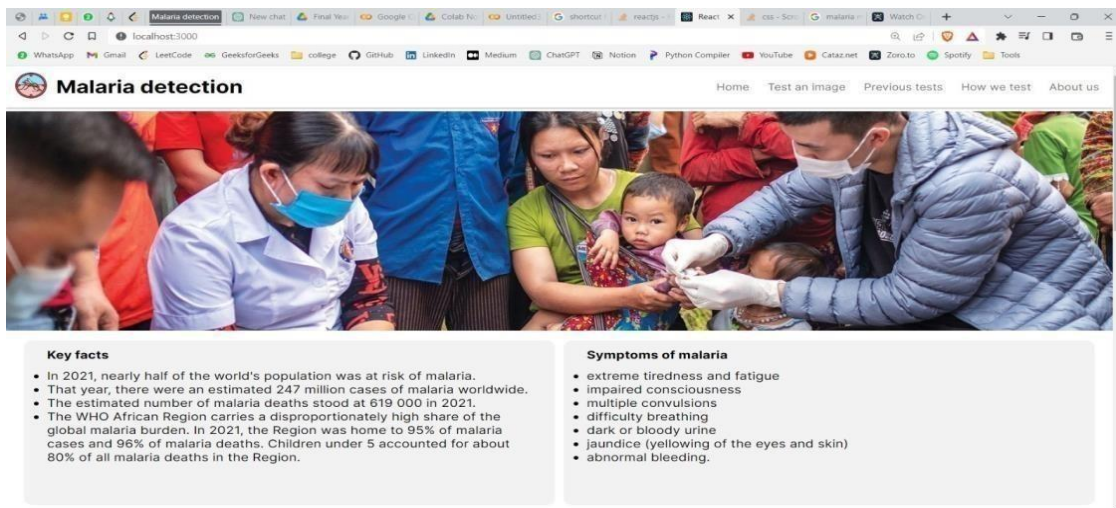


Fig:5.1: The main page of the project

5.2 CONCLUSION

In conclusion, the use of machine learning in the detection of malarial parasites offers a promising solution to the challenges faced in traditional microscopy methods. The proposed system described above offers a comprehensive approach to the development of a machine learning-based diagnostic tool for malaria detection.

The proposed system includes a data acquisition component, where blood smear images are collected and preprocessed for analysis. The preprocessed images are then used to train a machine learning algorithm to classify each image as either infected or healthy.

The machine learning algorithm is selected based on its performance on various metrics, including accuracy, precision, recall, and F1 score.

The system also includes a user interface component that enables healthcare professionals to interact with the machine learning model. The user interface provides visual feedback on the classification of each blood smear image and enables healthcare professionals to upload new blood smear images for classification, view the results of the classification, and provide feedback to the machine learning model.

The proposed system can be deployed either offline or online. Offline deployment involves training the machine learning model on a dataset and then packaging the trained model with the application code. Online deployment involves deploying the machine learning model on a cloud-based platform or a server, which healthcare professionals can access through a web interface.

The performance of the machine learning model can be evaluated using several metrics, including accuracy, precision, recall, F1 score, and ROC curve. The evaluation can be done using holdout validation or cross-validation.

The use of machine learning in the detection of malarial parasites has several smear images faster and with higher accuracy, reducing the chances of misdiagnosis and improving patient outcomes. Machine learning can also be used to analyze large datasets, enabling the detection of patterns and trends that may be missed by traditional methods.

However, there are also some challenges to the use of machine learning in malaria diagnosis. Machine learning algorithms require large amounts of data to be trained effectively. The lack of labeled data can be a significant bottleneck in the development of machine learning models for malaria diagnosis. Furthermore, the performance of machine learning models can be affected by the quality of the data and the preprocessing techniques used.

In conclusion, the proposed system described above offers a promising solution to the challenges faced in traditional microscopy methods for malaria diagnosis. The development of a machine learning-based diagnostic tool for malaria detection has the potential to improve patient outcomes and reduce the burden of malaria in endemic regions.

CHAPTER 6

REFERENCES

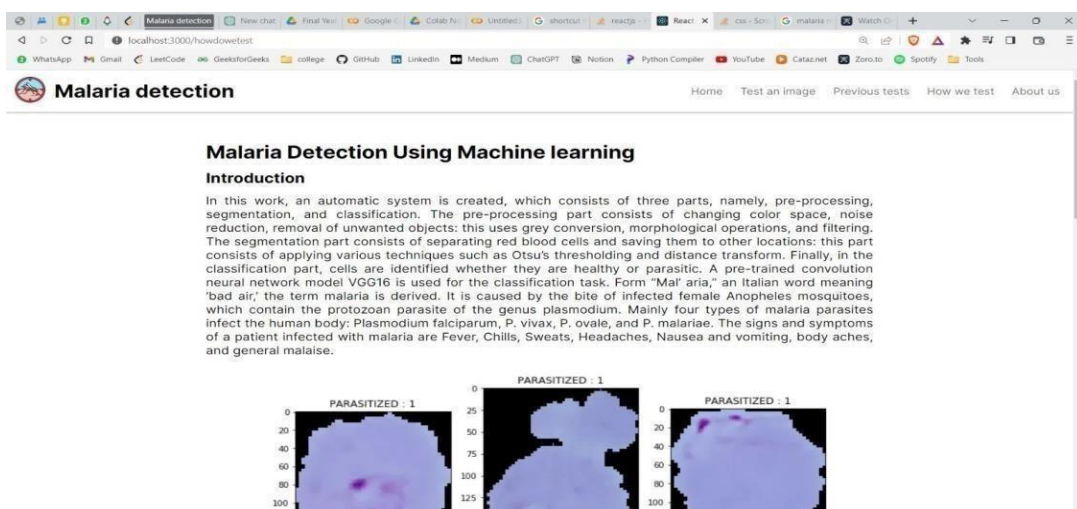
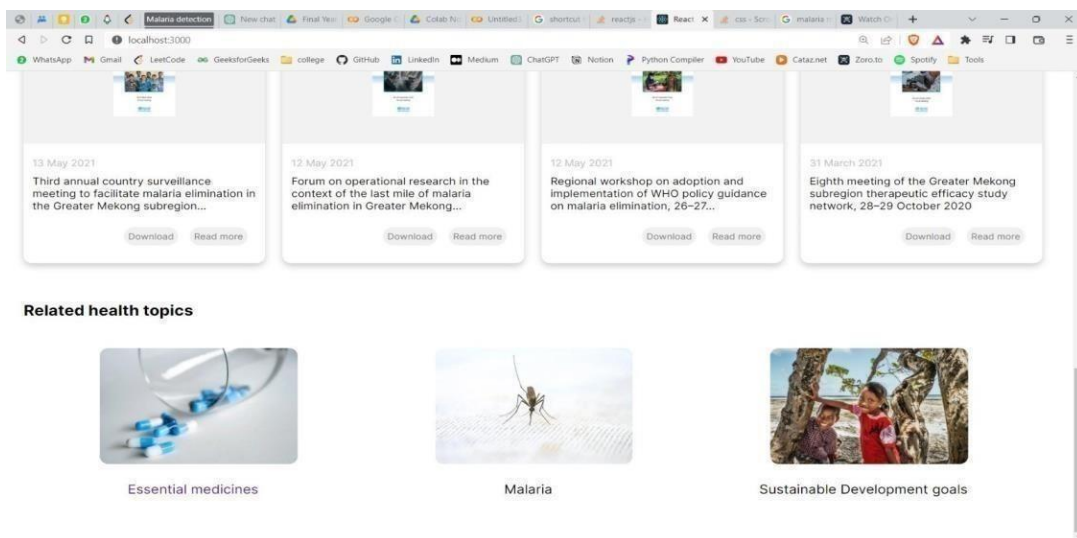
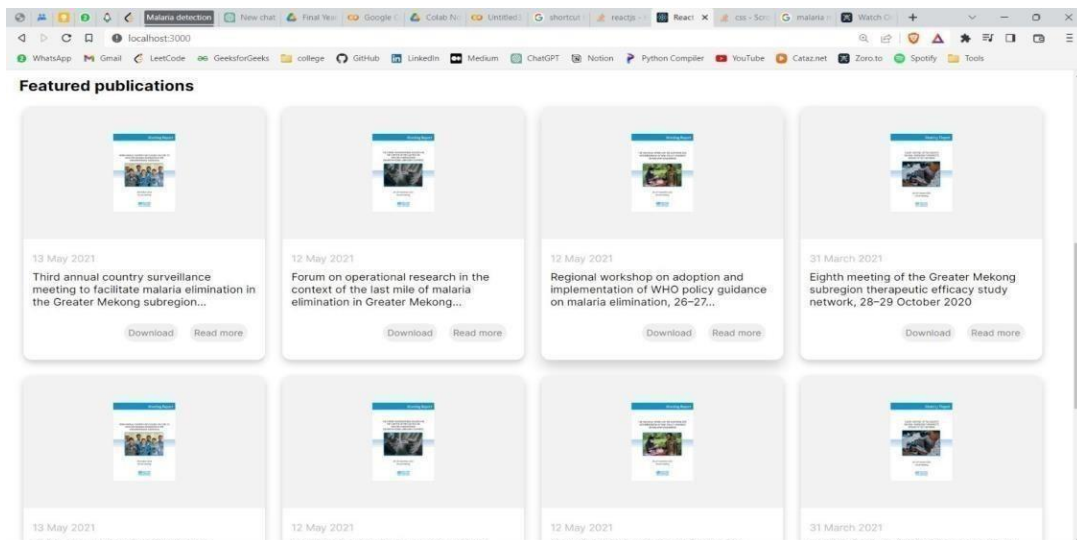
- [1] [1] Dr. K.S. Tiwari, Supriya Bhagat, Nikita Patil, Priya Nagare, IOT Based Driver Drowsiness Detection&Health MonitoringSystem.
- [2] Rateb Jabbar, Mohammed Shinoy, Mohamed Kharbeche, Khalifa Al-Khalifa, MoezKrichen, Kamel Barkaoui, Driver Drowsiness Detection Model using CNN Techniques for Android App.
- [3] Md. Yousuf Hossain, Fabian Parsia George, IOT based Real-time Drowsy Driving Detection System for the Prevention of Road Accidents.
- [4] Mika Sunagawa, Shin-ichi-Shikii, WataruNakai, Makoto Mochizuki, Koichi Kusukame, and Hiroki Kitajima, Comprehensive Drowsiness Level Detection Model Combining Multimodal Information.

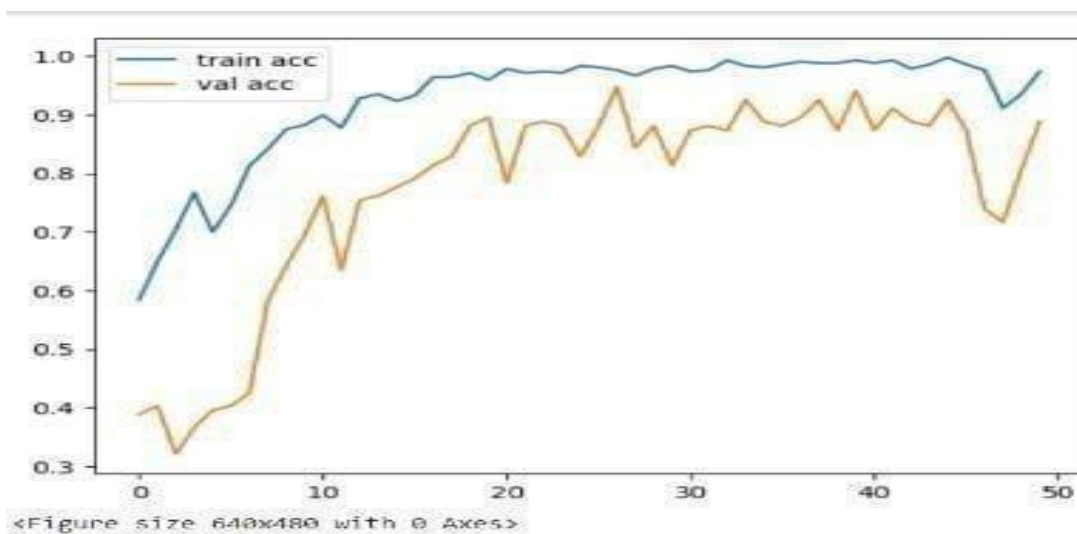
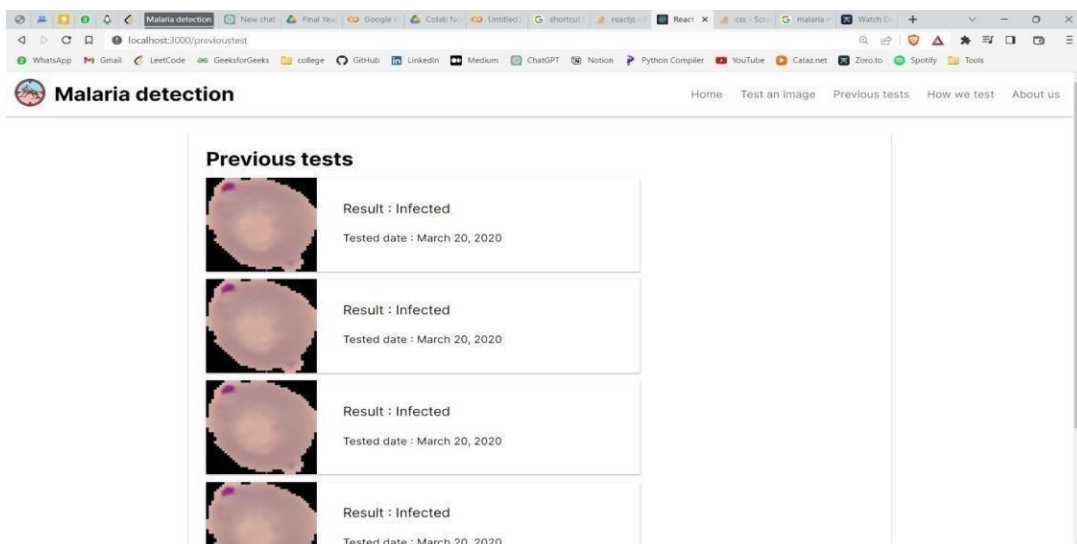
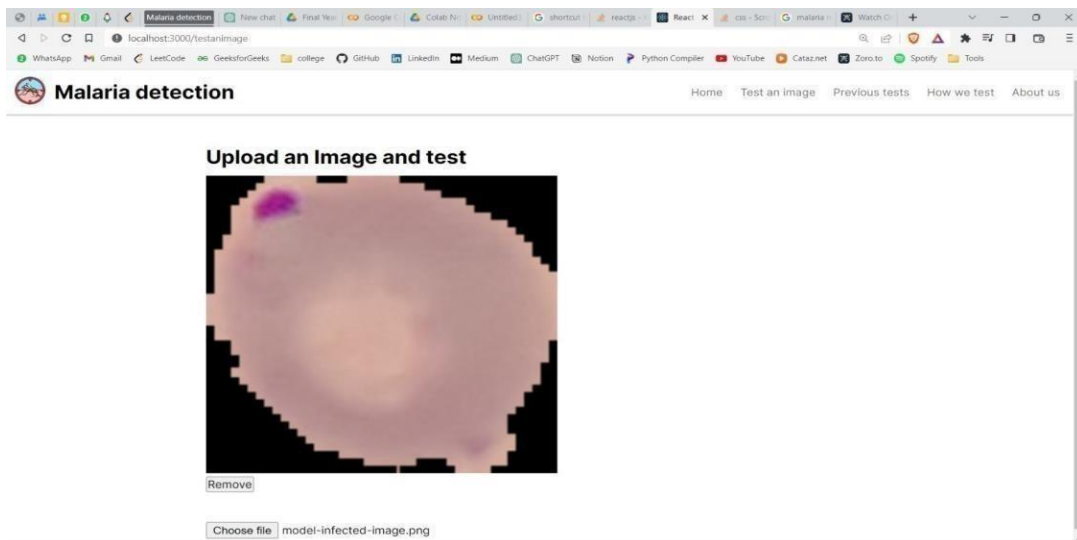
- [5] Monagi H. Alkinani, Wazir Zada Khan, and Quratulain Arshad, Detecting Human Driver Inattentive and Aggressive Driving Behavior using Deep Learning: Recent Advances, Requirements, and Open Challenges.
- [6] MkhuseleNgxande, Jules-Raymond Tapamo and Michael Burke, Driver drowsiness detection using Behavioral measures and Machine Learning Techniques: A review of state-of-art techniques.
- [7] Joao Ruivo Paulo, Gabriel Pires, and Urbano J. Nunes, Cross-Subject Zero Calibration Driver's Drowsiness Detection: Exploring Spatiotemporal Image Encoding of EEG Signals for CNN Classification.
- [8] Naveen Senniappan Karuppusamy, Bo-Yeong Kang, Multimodal System to Detect Driver Fatigue using EEG, Gyroscope, and Image Processing.
- [9] Yaocong Hu, Mingqi Lu, Chao Xie, and Xiabo Lu, Driver Drowsiness Recognition via 3D Conditional GAN and Two-level Attention BiLSTM.
- [10] CeerthiBala U.K, Sarath TV, Internet of things based Intelligent Drowsiness Alert System, Fifth International Conference on Communication and Electronics Systems (ICCES2020).

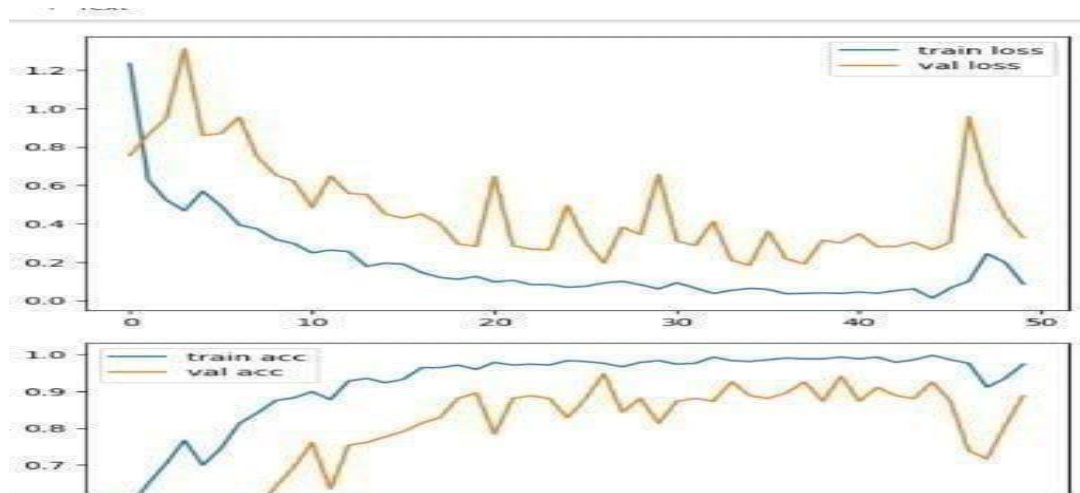
CHAPTER 7

APPENDIX

7.1 SCREENSHOTS







7.2 SOURCE CODE