Detection of Malarial Parasites using Deep Learning

[1]Dr. Prabhanjan S
Guide & HOD Dept of CSE
Jyothy Institute of Technology
Bangalore India

[2]Satvik R Kundargi
Dept of CSE
Jyothy Institute of Technology
Bangalore India

[3]Swaroop S Kulkarni
Dept of CSE
Jyothy Institute of Technology
Bangalore India

[4]Ranjith Kumar R
Dept of CSE
Jyothy Institute of Technology
Bangalore India

Abstract:- Malaria is a fatal disease that leads to the death of lakhs of individuals every year. Malaria is caused by a microbe belonging to the Plasmodium group. Five types of these organisms cause this disease in a foreign body. *P. falciparum*, *P. vivax*, *P. ovale*, and *P. malariae* are the five varieties of parasites that cause malaria. P. Falciparum infected are the ones who are more susceptible to keywords as others are mildly infectious. Malaria is spread by a mosquito species called anopheles it is the same that spreads dengue too. However, this disease could be cured if detected at early stages. Detecting malaria is an extremely challenging aspect considering the morphological aspects of the parasite. Malaria is more prevalent in tropical and sub-tropical climatic conditions. Owing to the parts of our country the monsoon is the season where we could find increased cases. Malarial parasites enter the human body through mosquito saliva, which in turn is transmitted to the blood. In the body of an organism, it develops in the liver and matures there itself, and starts to reproduce. Generally, malaria symptoms are seen after 10-15 days the parasites intrude on the body. The real challenging part is to check the growth of malaria during the monsoon season in the rural part which faces problems like a lack of doctors, nurses, equipment, testing centers, and so on. The traditional methods of detection are time-consuming and not so accurate. Thus, our project aims to recognize and solve these issues.

Keywords: Image Processing, Convoluted Neural Network (CNN), Deep Learning, Machine Learning.

I. INTRODUCTION

➢ *How is Malaria Detected?*

Malaria detection is challenging but not impossible. There are primarily three types of malarial diagnostic techniques. Each has its own advantages and limitations. One of the diagnostic methods is using a microscope. This technique involves creating a blood smear of the suspected sample and detecting if any parasites are present using a microscope. This technique requires skilled pathologists and equipment. Even the tested samples need not be accurate considering the complex life cycle of the parasite. The second way involved in diagnosis is the Rapid Diagnostic Test (RDT). This technique involves molecular and DNA analysis of the infected blood sample. RDT was used extensively in the early 19's. However, a major drawback of this technique is that it is costly and takes a minimum of 8 hours. In poor and developing countries this technique is quite impossible. PCR i.e., Polymerase Chain Reaction test is the third type of detection technique. To be frank this procedure is an extended version of the RDT. It is more accurate than the latter but requires a large number of specialized lab technicians' extravagant materials and it also takes not less than 12 hours. Owing to these kinds of serious issues malarial detection is quite challenging. The advent of new technology and the synchronization of many technologies led to a new ray of hope in the medical field. Interpersonal fields are merged today, to pave the way that yields better results. Detection of contagious diseases, cancerous cells, and genetic disorders is the emerging scope that utilizes intrapersonal fields like computer science, electronics and communication, biomedical, biotechnology, and several applied fields. Integrating these technologies leads to a new field that solves the various hitches that evolved emerge and emerge. Looking at the lifecycle of the malarial parasite that is Plasmodium one can get to know how difficult it is to detect it. Especially the stage once these parasites reproduce, they remain in the dormant stage and remain active after 36 hours. Thus, while detecting malaria timing also matters. The technique utilized in our work envisages using some of the modern algorithms, techniques, and research.
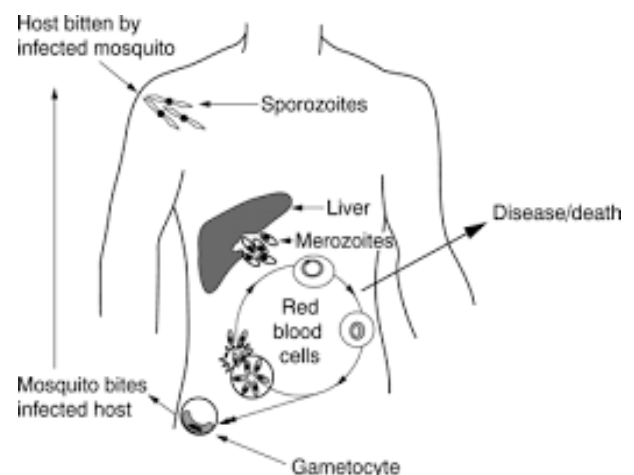

Fig 1 Life Cycle of Plasmodium Parasite

II. LITERATURE SURVEY

Accessing, analyzing and exploring from papers to papers led to an understanding of the loopholes in the current or existing ecosystem. About the following above domain also several researchers have spent meaningful time producing their resplendent work. Many of the emerging students and researchers have utilized these papers and reproduced the existing system to produce more significant work. Our work also tries to envision subsisting spadework. After exploring many papers, we realized the methodology procedure undertaken by the cognizant research outline. As a part of the literature survey, we deeply analyzed the manifesto of research which and quoted below as it is. In [1] includes the following "Detection of malaria parasites using digital image processing", This method makes judicious usage of the potential features of the intensity of the microbial activities and the erythrocytes being acquired, pre-processed, relevant features extrapolated from the visual objectives. A set of characteristic qualities based on the vigor assessment has been put forward and the performance of these features on the blood cell samplings from the created database has been computed using artificial neural networks or ANN classifier. In [2] "Detection of malaria parasite in Giemsa blood sample using image processing, the proposed system model is implemented by using two segmentation aisHSV segmentation and watershed segmentation. The result of both segmentations is mapped so that all the malarial parasites can be counted. This is done to improve accuracy". In [3] "Detection of malarial parasite in blood using Image Processing, the design takes the form of a standard pattern recognition and classification system. The system architecture of malaria parasite detection includes grey-scale image conversion, thresholding, and thinning labeling algorithms. To detect the infected cells, color range and image segmentation techniques are there from Giemsa-stained peripheral blood samples". In [4] "Color image segmentation approach for detection of malaria parasite using various color models and k-means clustering, This paper shows that the detection of the malaria parasite can be also done through color image segmentation which can be further applied to malaria images of P. vivax species. The use of partial contrast stretching is introduced here to obtain segmented RBC infected within the malaria parasite. After that, an unsupervised segmentation technique is used which is known as k-means clustering. It segments the infected cell from the background different color components of RCB, HIS, and C-Y color modes need to be checked to identify color components". In [5] "Automatic diagnosis of malaria based on complete circle-ellipse fitting search algorithm, thin blood smears are used in this study and are mainly based on curve fitting to detect the infected cells in the blood smear. The method is composed of six phases- the stain object extraction step, preprocessing phase which uses nonlinear diffusion filtering followed by the detection of parasite nucleus from the result of the previous step according to the image intensity, the fourth step is to introduces a complete search process in which circle search step in defines the direction and initial points for direct least square ellipse fitting algorithm, furthermore ellipse searching process is applied finally a series of decision rules on the presence of infected cells in the blood smear". In [6] "Computer automation for malaria parasite detection using Linear Programming, Mathematical modeling technique is used by means of LP as an efficient tool to solve problems related to malaria parasites problems related to malaria diagnosis through microscopy imaging problems. Two applications are approached: 1. Formulation of LP model based on given data. 2. Solving and displaying the result using a graphical method approach for detecting infectious cells. The application mainly consists of developing a linear mathematical model from the collected information and in addition to it, the problem is solved by a graphical approach. However, the system is complex and adds a drawback of delay".

➢ *Objectives*

The main objective of our research work is the perception of malarial parasites using the concept of machine learning and image processing. The objective of this task is to develop a fully integrated and automated image relegate system that defines pragmatic identification of the microorganisms present in the thin blood smears and differentiates its species. The algorithm initiated will be helpful in the field where experts in microscopic analysis may find it hard to look over. The sole purpose of this algorithm is to detect the presence of malarial foreign bodies in the bloodstream at any growing stage. The effort of this algorithm is to detect the stage in which the microbe resides. The microbe takes around 7 to 8 days to mature and reproduce in the human body. So if this algorithm is incorporated into routine tests, then the presence of parasites can be detected based on the concept of color histograms. If we look over the condition of the research being done, we propose to induce a solution for the detection-related issues with two successive categories.

The objectives of our study can be summarized below,

- To track down the region we desire to extract vital hallmarks from the standard microscopic visuals of the red blood cells using the image processing domain.
- To train various machine learning archetypes or a model using extracted features for differentiating the infectious and the ones which are normal.
- To discover the most significant advance towards based on various computation grid for the diagnosis of malaria.

➢ *Proposed Methodology*

The proposed methodology which we want to implement is an influx of conventional machine learning techniques and the advanced modern guard. We try to analyze which technique provides the most accurate results. The system itself identifies the more meticulous results. Even a 0.1% error in some respects is not entertained. We must deal with sincerity and must be firm in certain aspects like design decision-making and implementation. Dealing with the datasets appropriately is a challenge. Obtaining the dataset itself is a herculean task. But with the help of the

CIIRC lab Dept of Food Technology Jyothy Institute of Technology and a local laboratory we have effectively defined the dataset. The microscopic version of the sample is obtained with the help of a compatible device called a foldscope. Foldscope can be utilized to get a magnification of minute objects 140x times on the other hand that of the microscope is only about 100x times. The cognition of utilizing foldscope is its advantages like cost effectiveness, ease to handle, doesn't require high scale light batteries, and also space constraints. Thus, the methodology we aspire to innovate is visualized in the form of a diagram shown in the succeeding section. The three vital categories of our work can be conceptualized as shown below.
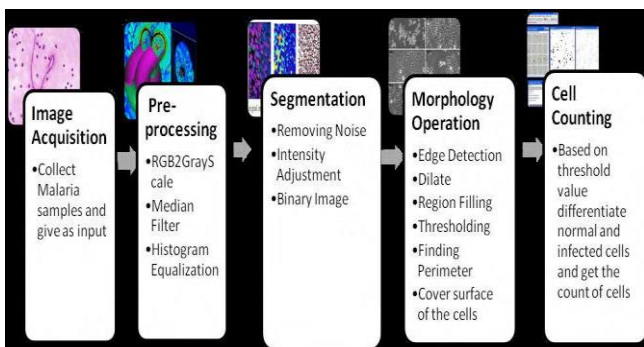

Fig 2 Pictorial Representation of Proposed Methodology

➢ *Dataset Collection*

Considering real-time images of both infected and non-infected microscopic images. However, getting microscopic real-time images of the malarial slide is quite challenging. With the help of a local laboratory and the Food Technology Department CIIRC of Jyothy Institute of Technology, we are trying to collect the respective datasets. Using a cost-effective device called Foldscope we are obtaining the dataset of our choice. Foldscope magnification is 140x times whereas microscopic magnification is 100x times. Around 27,578 images of both infected and uninfected images collected from various sources are being analyzed.

➢ *Data Preprocessing*

Data pre-processing aim is to improve the quality of the image so that we can analyze it in a better way. Involves certain image processing techniques to ensure the image is clearly filtered before being analyzed. Removing noise in the collected images requires an AMF algorithm. Converting an image to a grayscale image is carried out for quality assessment. Some of the process carried out includes image resizing removing unwanted portion and so on.

➢ *Model Implementation*

CNN is an algorithm in Machine Learning that is used to deeply analyze an image. Apart from CNN, we propose to use a random forest algorithm that works similarly to a decision tree algorithm. Also, random forest algorithm is also used to classify and regression problems. To analyze the classical machine learning and modern deep learning technique. Random forest algorithm requires feature extraction however, CNN doesn't require it.

The following requisites include as shown

- Programming Language: Python. Programming language of version 3. O or later if updated.
- Hardware: 4GB RAM (At least) required or more
- 250 GB of Hard Disk Space (At least), 1 GB of VRAM (At least) for the purpose of storage.
- Software: Windows 7 or later, Ubuntu, Mac OS X or later.
- Proposed technologies to use: Convolutional Neural Network, Tensor Flow, Google Colab, Digital Image Processing OpenCV (python library for computer vision applications), Scikit-image for specified execution.
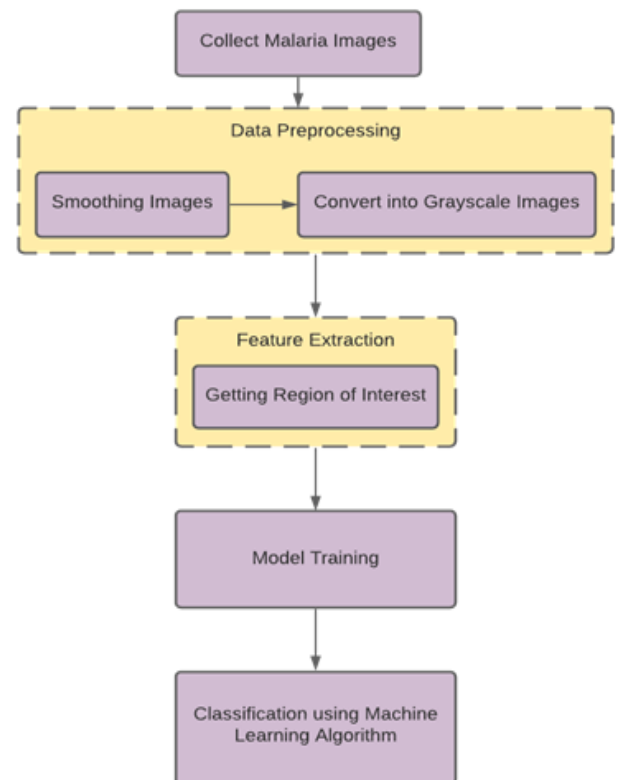

Fig 3 Model Implementation

➢ *Technologies used*

- *Basic Convulational Neural Network*

Progress in modern equipment has led to deeper dependence on modern machines. Deep learning is one such young technology that can create wonders. In the vast majority of deep learning techniques there exists a domain named Convolutional Neural Network abbreviated as CNN. The primary motto of CNN is computer vision. Computer vision is one of the branches that deal with the complete acculturation of image analysis. Image interpretation, separation and visualizing the data is a prime example of vision. CNN is generally a proffer version of artificial neural networks or ANN. Principally, CNN is utilized for extracting features from the grid look alike matrix data sets. CNN consists of numerous input layers unlike ANN. CNN was animated by the multiplex biological processes that occur inside the living creatures. The connectivity motif between the neurons resembles the organization of the animal visual cortex and individual cortical neurons. The

basic CNN is enacted from genesis. So, by using the tensor flow and the Python library which consists of the keras. The keras is an open-source neural network in the programming interface of python. With this, the keras has an accredit called as the Conv2D. There exists an older version named Conv1D, but the latter is used owing to the two-dimensional aspects of the image. The max-pooling part comes here to enlarge the clump of neurons at the appropriate layers. Max-pooling is a down sampling blueprint on CNN. Max-pooling is used to coincide the various essence of the matrix to the least possible values. Consider an instance by taking a 4X4 matrix, inside the matrix, there are four values located at the four adjoining vertices. With the assistance of the max-pooling out-turn the determinants of the four values are found and the solution will be a 2X2 matrix. In the Keras layers, a layer exists called Flatten which is used to prepare a vector for the fully connected layers by using the command. The activation of sigmoid will entrust by calculating the average accuracy and once the attributes of keras is acquired, the model is built. Printing the summary of the model will occur to identify the parameters that are full-fledged with both the training and testing data sets. There are roughly around 15,102,529 trainable framework that are earned by the non-zero trainable parametric. To conclude the summary of the model, and the training of the model will occur by the details of the epochs, size of the batch, callbacks, verbose, and the data validation. Fitting the ideals of the model will begin the process of 25 epoch procedures. With this process, the total loss percentage will start to be in the downtime slowly and finally comes to an halt after the 25th epoch. At the 25th epoch, the perusal are, loss:0.054, accuracy: 0.9985, validation accuracy: 0.9430 and validation loss: 0.4244. By the visual acuity of these values, it is clear that it is not digesting part as the loss value is extremely higher of the anticipated value. To ensure lesser loss value, to the next version of the CNN may somehow produce the accurate value. The pictorial graph depicts the accuracy versus epoch and loss versus epoch graph states the huge divergence between the dark line with a grey line that shows the validation percentage of both values.

- *Frozen Convolution Neural Network*

The hierarchy of the demonstration is same for both the basic and frozen CNN however, the given attributes will change as the frozen CNN will come into the picture. In this type of CNN there exists keras with VGG. Two different types of versions of VGG exist namely, VGG16 and VGG19. The primary contrasting factor between these two types is that VGG16 ply 16 layers, on the other hand the latter uses 19 layers in the deep neural networks. In the project, VGG19 is employed to make it effective and run on the images and the trained model. VGG built under 16 layers act as the main convolutional layers and the max-pooling layer. The size of the model is ascribed to be 528MB. The first layer is the convolution layer with 64 filters that undergoes where the image goes through this underlying layer. The second layer includes both convolution and max pooling with 64 filters.
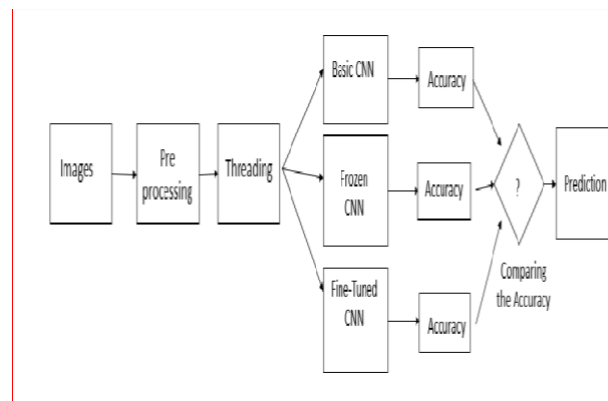

Fig 4 Block Diagram of Proposed Methedology

- *Fine-Tuned Convolutional Neural Network*

The general practice of training the networks from the base foundational is just to first train the network on million to billions of images. Thus, fine-tuning CNN is used to work on image datasets with high numbers. Hence, this type of CNN is used ubiquitously for large data sets. In our project we tried to implement this technique to give higher accuracy. In this type of CNN there is about 10X times less training time. It works like the latter one but have good accuracy.

✓ *Algorithm Steps*
- Step 1: Collection of all the image sources and merge them into a single file.
- Step 2: Split the images by the way user desires here train and test data are split using a module called sklearn.
- Step 3: The Use of OpenCV to understand the image and estimate the image parameter.
- Step 4: Process the dataset here image using the thread pool executor for not to face time straining.
- Step 5: Create the general eccentric CNN model from the base foundation and fit the model
- Step 6: Upload the images in the model and ensure the model is ran using Keras.
- Step 7: Use Epoch = (number of iterations * batch size) / total number of images in the training
- Step 8: Verify the accuracy of the CNN model.
- Step 9: Create the Frozen CNN model and fit the model and repeat step 6-7
- Step 10: Verify the accuracy of the CNN model and move to next if found low.
- Step 11: Create the Fine-Tuned CNN model and fit the model and repeat step 6-7
- Step 12: If the accuracy is good manageable halt here and get the accuracy rate.

III. RESULTS AND DISCUSSIONS

Detection of diseases caused by protozoans' malaria in this case is a very strenuous fidelity. This project developed by our team thus creates a healthy atmosphere which reduces the deaths caused due to the delay in the process of lab specimen technique. The revered techniques used by our team ensure there is greater accuracy. The total number of datasets specified elaborates the huge number

i.e., around 25,000 and more. The system envisages fastening the results and printing the results upon looking on the images of the malarial microscopic slides.

Infected Cells   **Uninfected Cells**
Fig 5 Sample of Blood Cell Images of both Healthy and Infected
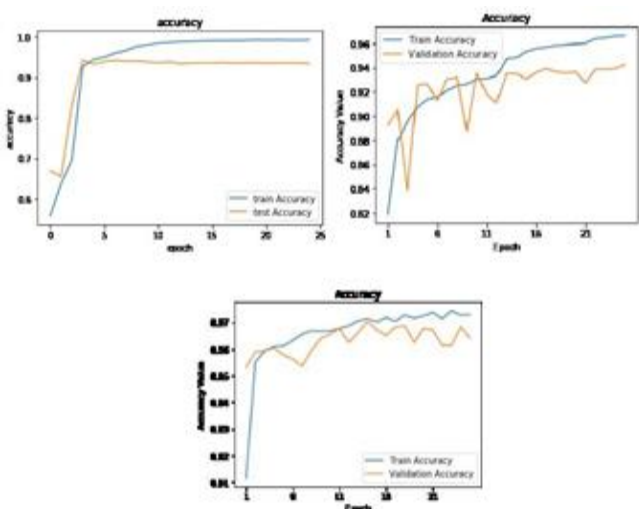
➤ *Performance Analysis*

Fig 6 The Accuracy vs Epoch Graph Starts from the Top Left is the Basic CNN Model, Frozen CNN Model, and Fine-Tuned CNN Model

The above figure tries to present the accuracy against epoch graph of the various CNN models where it tries to depict the fine-tuned CNN a higher accuracy model than the other three. In this model the train and validation accuracy line meet with the one other that gives an accurate result about the model.

Table 1 The Confusion Matrix of the CNN

Actual/Predicted	Healthy	Malaria
Healthy	4004	71
Malaria	260	3933

IV. CONCLUSION

The proposed project that we aspire to carry forward is a messiah for many poor people. We aspire to create an ecosystem where we can easily detect, diagnose and predict malarial parasitemia. Malarial detection tests are expensive, time-consuming, and not so accurate. With the help of our project, we hope to reduce the burden on the system and bureaucracy on contagious diseases. With the help of the advent guard, we hope to connect it to more individuals and youth who aspire to serve the poor and rural parts of the

country. With this project, we hope to reduce the mortality rates caused by this disease. We want to present a proposition and develop an algorithm for the diagnosis of malaria, automated process is being instantiated and quantification of this fatal disease. Also, we want to develop a strategy to train with machine learning, adaptable to the sensing of malaria with other types of parasites, and also consider increasing the predictive value with the yielded output. And finally, we want to juxtapose the conventional and modern ways. The future of this work includes predicting some of the complex diseases like pneumonia, breast cancer, dengue and the impact of Covid19 over the lungs.

REFERENCES

[1]. Ahmedelmubarak Bashir; Zeinab A. Mustafa; Islah Abdelhameid; Rimaz Ibrahim, "Detection of malaria parasites using digital image processing" 2017 International Conference on Communication,Control, Computing and Electronics Engineering (ICCCCEE) Khartoum, Sudan.

[2]. Kishor Roy, Shayla Sharmin, Rahma Bintey Mufiz Mukta, Anik Sen, "Detection of malaria parasite in Giemsa blood sample using image processing", International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 1, February 2018.

[3]. Pallavi T Suradkar, "Detection of malarial parasite in blood using Image Processing techniques", International Journal of Engineering and Innovative Technology (IJEIT) Vol 2 Issue 10, pp. 124-126, April 2013.

[4]. Aimi Salihah Abdul-Nasir, Mohd Yusoff Mashor, Zeehaida Mohamed, "Color image segmentation approach for detection of malaria parasite using various color models and k-means clustering", WSEAS transaction on biology and biomedicine, Issue 1, Volume 10, pp 41-55, January 2013.

[5]. Rajaraman S, Antani S, Poostchi M , Silamut K,Hossain M.A ,Maude R.J, Jaeger S, Thoma G.R: Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images.PeerJ 6 (2018)

[6]. Kumari U, Memon M, Narejo S, Afzal M.: Malaria detection using image processing and machine learning. IJERT NTASU–2020 (Volume 09–Issue 03 (2021)

[7]. Liang Z, Powell A, Ersoy I , Poostchi M, Silamut K, Palaniappan K, Guo P, Hossain M.A, Sameer, A, Maude R.J, Huang J.X, Jaeger S, Thoma G.: Cnn-based image analysis for malaria diagnosis. In: 2016 IEEE International Conference on Bioinformatics and Bio-medicine (BIBM). pp. 493–496 (2016)

[8]. Lecun Y, Bengio Y.: Convolutional networks for images, speech, and time-series (01 1995)

[9]. Mahdieh Poostchia, Kamolrat Silamutb, Richard J. Maudebcd, Stefan Jaegera, George Thoma, "Image analysis and machine learning for detecting malaria".

[10]. W. A. Mustafa , R. Santiagoo, I. Jamaluddin, N. S. Othman, W. Khairunizam and M.N. K. H. Rohani, :"Comparison of Detection Method on Malaria Cell Images" 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA), 2018

[11]. S. S. Savkare Moze College of Engineering, University of Pune; S. P. Narote Sinhgad College of Engineering, University of Pune, Pune, India

[12]. M. Sheikh Hosseini, H. Rabbani, M.Zekri, A.Talebi, "Automatic diagnosis of malaria based on complete circle-ellipse fitting search algorithm". RMS-Journal of Microscopy, Volume 252, Issue 3, pages 189–203, December 2013

[13]. Vipul Parkhi, Pooja Pawar, Archana Survi, "Computer Automation for Malarial Parasite Detection using Linear Programming", International Journal for Advanced Research in electrical, electronics and instrumentation engineering, volume 2, Issue 5, May 2013, pp.1984-1985

[14]. K M Khatri, V R Ratnaparkhe, S S Balchandra, "Image Processing Approach For Malarial Parasite Identification", International Journal of Computer Applications (IJCA), 2013.

[15]. Nicholas A. Boon, Nicki R. Colledge, Brian R. Walker, John A. A. Hunter, "Davidson's Principles and Practice of Medicine". 20th ed., published by Churchill Livingstone (Elsiever), 2006, pp. 342-348.