# MALARIA DETECTION USING CNN

## A PROJECT REPORT

*Submitted by*

## SAMIKSHA MAHESHWARI – 20BHI10061

*in partial fulfillment for the award of the degree*
*of*

## BACHELOR OF TECHNOLOGY
*in*
### CSE Spl. In Health Informatics



## SCHOOL OF COMPUTING SCIENCEAND ENGINEERING

## VIT BHOPAL UNIVERSITY

## KOTHRIKALAN, SEHORE
## MADHYA PRADESH - 466114

APRIL-2022

# ACKNOWLEDGEMENT

First and foremost, we would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

We wish to express our heartfelt gratitude to Dr. Pandi Murugan Head of the Department, School of Computer Science for much of his valuable support encouragement in carrying out this work.

We would like to thank our internal guide Dr Subash Chandra Bose, for continually guiding and actively participating in our project, giving valuable suggestions to complete the project work.

We would like to thank all the technical and teaching staff of the School of Computer Science, who extended directly or indirectly all support.

Last, but not least, we are deeply indebted to our parents who have been the greatest support while we worked day and night for the project to make it a success.

# LIST OF FIGURES

# ABSTRACT

The main objective of our project is to improvise the detection of Malaria diseaseby implementing machine learning. We initiated our project by overviewing some research papers to get a theoretical viewpoint of our project. Once our research will be completed, we will be collecting the data sets from various trusted organization's websites. After the collection of data as per our need. For our project we collected the blood smear image datasets and through our model we will be classifying it into either infected or uninfected . In this investigation, we gone through various research paper and selected the best possible method that can overcome the limitations of current diagnostic approach. Once our analysis is done, we will be creating a portal for those who want to check their datasets with more accuracy.

# TABLE OF CONTENTS

# 1. INTRODUCTION

## 1.1 INTRODUCTION

Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes. It is an acute febrile illness caused by *Plasmodium* parasites. It is preventable and curable. Left untreated, they may develop severe complications and die. It is a life-threatening disease caused by the infection of red blood cells with protozoan parasites of the genus Plasmodium.

Malaria is a potential medical emergency and should be diagnosed accordingly. Delays in diagnosis are leading causes of death in many countries, malaria is the most widespread cause of death, disability and economic loss in India especially among the poor who have limited access to timely and effective treatment. Moreover, fatal form of malaria that accounts for almost all malaria-related deaths - Plasmodium falciparum (Pf) – has been rising rapidly in India since the 1980s. Early and accurate diagnosis is crucial to attaining the target of a malaria-free India but Clinicians may forget to consider malaria among the potential diagnoses for some patients and not order the necessary diagnostic tests. Technicians may be unfamiliar with, or lack experience , and fail to detect parasites when examining blood smears under a microscope and therefore important to introduce efficient diagnostic techniques. This research focuses on designing an accurate malaria diagnosis model that can be implemented without any dependencies on skilled technicians and testing the model accuracy to get high-quality results .

## MOTIVATION FOR THE WORK

Malaria known as a disease of poverty because of its dominance in poverty-stricken areas. India is one the most vulnerable states for malaria morbidity and mortality. Socio-economic, environmental and demographic factors present challenges in malaria control and elimination.

After doing a lot of research, we found out that there are still many healthcare organizations or medical centers which lag in the accuracy and prediction of this disease which leads a human to a life-threatening condition also current diagnostic methods of malaria required microscopist's skill .To overcome this condition, we decided to collect

medical data from nationally recognized organizations in order to do an analysis. Once we have completed our analysis, we can create a diagnostic method that will study and predict the input dataset.

## 1.2 PROBLEM STATEMENT

The purpose of this study is to compare and contrast the effectiveness of deep learning models, specifically Convolutional Neural Networks, in categorizing human blood smear images as Malaria Parasitized vs. Healthy. Our project also focuses on uncovering the effects of preprocessing techniques on model accuracy.

## 1.3 OBJECTIVE OF WORK

The main aim of our malaria detection system is to address the challenges in the existing system by automating the process of malaria detection using Machine learning and image processing .

## 1.4 SUMMARY

Knowing the fact that on average over a millions of people die due to Malaria . Malaria is life-threatening disease caused by parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes. It is an acute febrile illness caused by Plasmodium parasites. It is preventable and curable. Left untreated, they may develop severe complications and die.  It is a life-threatening disease caused by the infection of red blood cells with             protozoan             parasites             of             the             genus             Plasmodium.

Being a student of the healthcare sector, we decided to build a project which will help people know if they are having this disease or not using with a higher accuracy level than the existing projects.

# 2   LITERATURE SURVEY

## 2.1 INTRODUCTION

As the standard method to identify and diagnosing the common fatal acute febrile illness 'malaria' is by visual examination of blood smears for parasite infected blood cells is time consuming , inefficient and dependent on microscopist's skill and experience. Therefore , there is a  trend that has uplifted the performance and working of various medical domains is the involvement of deep learning tools and methods. Currently , there are many certified works and research suggested and proposed the use of deep-learning methods for the diagnosis of malaria. Many of them approached their work/research by applying a transfer learning model or by implementing various classification algorithms which classify if a given blood-smeared cell image is infected or uninfected.
Some approached their work by using an only machine learning algorithms in their model like  DT , CNN , SVM , NB .

## 2.2 EXISTING ALGORITHMS / METHODS

The existing models use different machine learning algorithms for classification such as Naive Bayes, Support Vector Machine (SVM), K-NN, Random Forest,  Decision tree and current diagnostic methods are microscopy , molecular , immunology , RDT Now let us see the uses of all these techniques one by one in detail to better understand.

### 2.2.1 SVM:

 Implement architecture process of Support Vector Machine(SVM) classification involves image acquisition , pre-processing , segmentation , feature extraction and eventually the classification comparison with database  but  using only SVM can be hectic as hand engineered feature extraction is required. Another method purposed is to use SVM for classification  and pre-trained CNN model (i.e., AlexNet) for feature extraction. This transfer learning approach achieved  91.99% accuracy . In another better-proposed model of CNN-SVM , the  accuracy of 98.93 was accomplished by the

usage of 'RBF' kernel and setting gamma and regularization parameter values to 0.001 and 10 respectively . They are preety high in precision but sometimes computational work takes more than 24 hours [3]. Generally , it is observed that CNN is more advantageous over SVM when it comes in terms of the ability of learning features of image automatically .

### 2.2.2 : LOGISTIC REGRESSION

Another detection model proposed by Mandal et al. by logistic regression based classification for identifying P.vivax malarial parasite but reached about 88.77% of accuracy.

### 2.2.3 : KNN

K-Nearest Neighbors (KNN) model was also enacted under three process training named general, distillation and autoencoder training for optimizing and improving the accuracy of model and inference performance[6] like In CNN-KNN , a CNN model was used to extract the features. Later these features were classified through SVM or KNN .

### 2.2.4 Naive Bayes:

Another machine learning approach using Naïve Bayes classifier is the poorest among above discussed algorithms . It classifies five stages of malaria-infected erythrocytes (ring, schizon, gametocytes for P. Vivax and ring, gametocyte for P. falciparum) and non-infected erythrocyte.Bayesian only achieved 84% of screening accuracy, 98.10% sensitivity and 68.91% specificity for both P. vivax parasite  and P. falciparum . Das et al. used Naive Bayes algorithm to classify only ring and gametocyte stages of P.vivax and P.falciparum and achieved 96.73% accuracy , which is better than the previous one. As mentioned above , out of other classifiers Naive Bayes algorithm training time is the poorest to detect the malaria parasite in blood-smear samples .

### 2.2.5 OTHER DIAGNOSTIC METHODS
 Current diagnostic methods which are microscopy , molecular , immunology ,  RDT   -
1) Required of highly trained personnel , are Low sensitive
2)required expensive equipment , Time-consuming
3) Cannot differentiate between past and present infections
4) Unable to quantify parasite density and cannot able   to differentiate between past and  present infections

## 2.3 OBSERVATIONS FROM EXISTING RESEARCHES

 No doubt , a few notable exceptional works are done to increase the efficiency rate of machine learning models . After various implementations of the transfer learning models, it is generally observed that the use of deep learning techniques, specifically those who leverage transfer learning methods yields superior outcomes when considering the parameter of  classification metrics but unfortunately, the models proposed in those studies are quite expensive in terms of the required computational

resources . The study also repels that the CNN model shows better results than transfer learning model with a sensitivity of (96.99% vs. 89.00%) and F-score (97.36% vs. 90.24%.

| Author | Accuracy | Author | Accuracy |
|---|---|---|---|
| Abu Seman et al. [2] | 89.80% | Chayadevi & Raju [8] | 94.45% |
| Khot & Prasad [19] | 73.57% | Le et al. [22] | 92.69% |
| Das et al. [11] | 96.73% | Malini et al. [27] | 91% |
| Anggraini et al. [4] | 93.3% | Memeu [30] | 79.7% |
| Mandal et al. [28] | 88.77% | Prasad et al. [35] | 96% |
| Yunda [47] | 77.19% | Kumarasamy et al. [21] | 86% |

Table 1 – Accuracy of various Machine Learning Models proposed by different authors



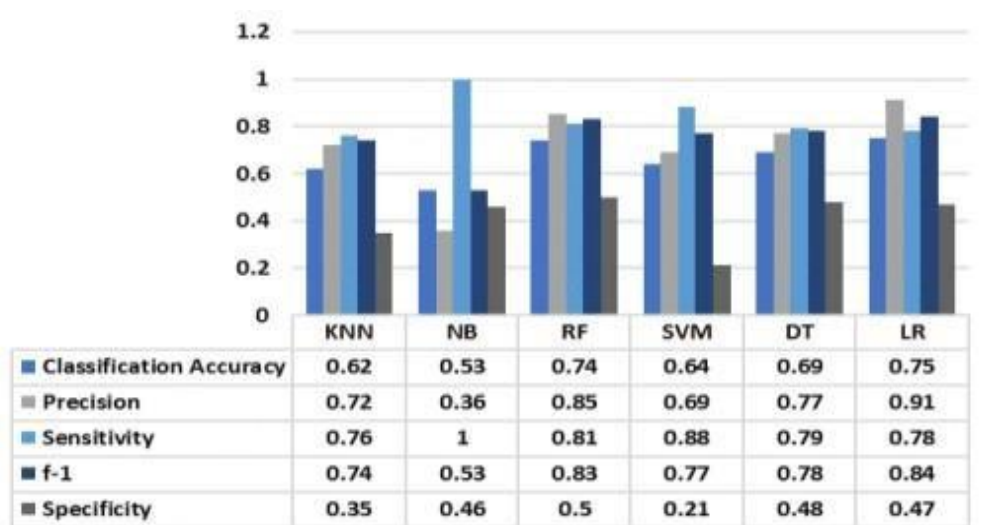| | KNN | NB | RF | SVM | DT | LR |
|---|---|---|---|---|---|---|
| Classification Accuracy | 0.62 | 0.53 | 0.74 | 0.64 | 0.69 | 0.75 |
| Precision | 0.72 | 0.36 | 0.85 | 0.69 | 0.77 | 0.91 |
| Sensitivity | 0.76 | 1 | 0.81 | 0.88 | 0.79 | 0.78 |
| f-1 | 0.74 | 0.53 | 0.83 | 0.77 | 0.78 | 0.84 |
| Specificity | 0.35 | 0.46 | 0.5 | 0.21 | 0.48 | 0.47 |

Fig 1: performance comparison of all the algorithms used

# 3. SYSTEM ANALYSIS & DESIGN

## 3.1 INTRODUCTION

## 3.2 PROPOSED SYSTEM

We are planning to create a CNN model which will identify the cells in blood smear image as infected or not. Over here one has to upload their blood smear image datasets of equal instances of infected and uninfected cells.  To start with our project, we initiated by loading our dataset using python libraries such as pandas, NumPy, matplotlib, seaborn, etc. All the required libraries are also imported for further use. Once the dataset is loaded, we done some pre-processing task  and feature extraction .

## 3.3 MODULE DESIGN & IMPLEMENTATION

Our proposed model is training CNN model from scratch contained 2 convolution layers with a max-pooling layer after each Conv2D RELU layer followed by Flatten , Dense layer which achieved a classification accuracy around 96% .
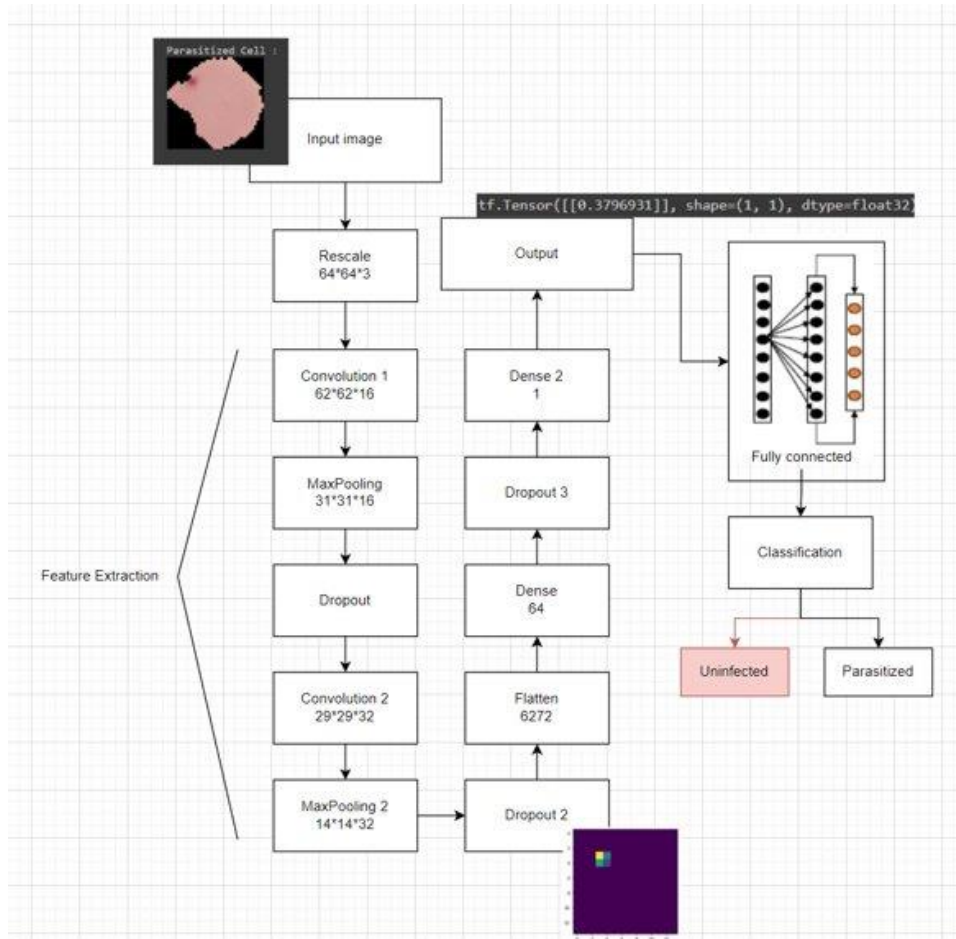


Fig 1: Module Architecture

1. Conv2D RELU layer : This layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs.
    a. Input Shape : 64*64*3
    b. Output Shape : 62*62*16

2. MaxPooling2D : Downsamples the input along its spatial dimensions (height and width) by taking the maximum value over an input window for each channel of the input. The window is shifted by strides along each dimension.
    a. Input Shape : 62*62*16
    b. Output Shape : 31*31*16

3. Dropout : The Dropout layer randomly sets input units to 0 with a frequency of rate at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by 1/(1 - rate) such that the sum over all inputs is unchanged.
    a. Input Shape : 31*31*16
    b. Output Shape : 31*31*16

4. Conv2D RELU layer :
    a. Input Shape : 31*31*16
    b. Output Shape : 29*29*32

5. MaxPooling2D :
    a. Input Shape : 29*29*32
    b. Output Shape : 14*14*32

6. Dropout :
    a. Input Shape : 14*14*32
    b. Output Shape : 14*14*32

7. Flatten : Flattens the input. Does not affect the batch size. It is used to make the multidimensional input one-dimensional, commonly used in the transition from the convolution layer to the full connected layer.
    a. Input Shape : 14*14*32
    b. Output Shape : 6272

8. Dense : A Dense layer feeds all outputs from the previous layer to all its neurons, each neuron providing one output to the next layer. It performs a matrix-vector multiplication.

      a. Input Shape : 6272

      b. Output Shape : 64

9. Dropout :

      a. Input Shape : 64

      b. Output Shape : 64

10. Dense :

      a. Input Shape : 64

      b. Output Shape : 1

Now, this 1 output shape is used to classify a particular image as parasitized or uninfected.

# 4. PERFORMANCE ANALYSIS

## 4.1 INTRODUCTION:

Performance analysis is the process of analyzing the performance of the best model obtained. The metrics that we choose to evaluate our machine learning algorithms are very important. Choice of metrics influences how the performance of machine learning algorithms is measured and compared. They influence how we weigh the importance of different characteristics in the results and our ultimate choice of which algorithm to choose.

## 4.2 PERFORMANCE MEASURES:

The performance measures of our project are dependent on the classification accuracy.

## 4.2.1 Classification accuracy:

Classification accuracy is the number of correct predictions for every prediction made. This is the most common scoring metric for classification problems. The accuracy level of the model can be predicted using the k- fold cross-validation algorithm and the validation dataset.

## 4.3 PERFORMANCE ANALYSIS
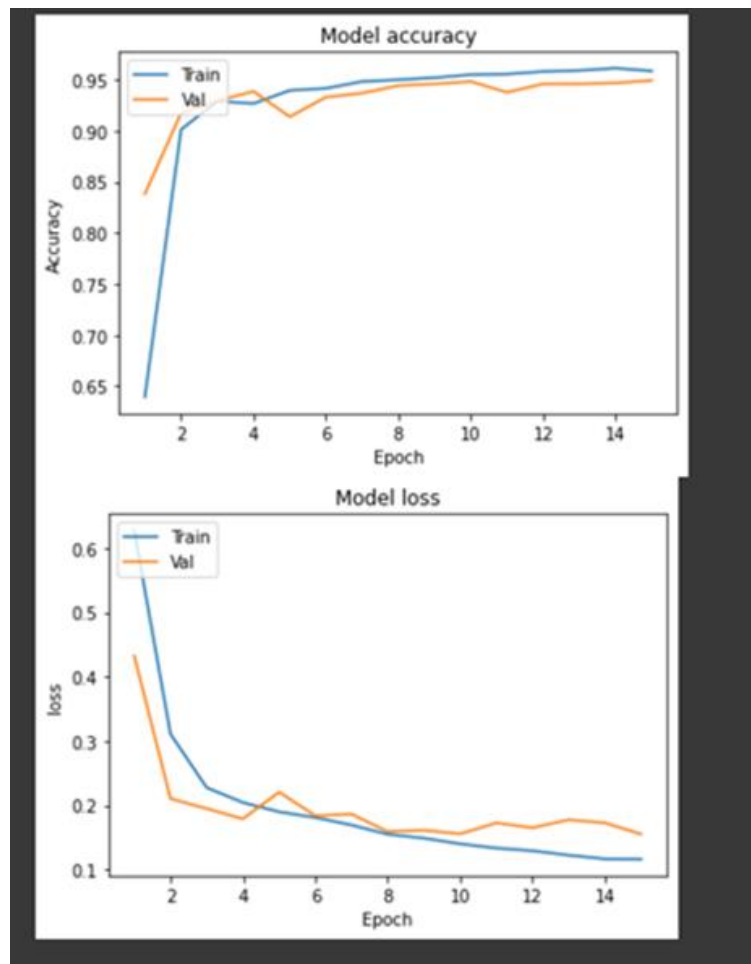
Our current model uses 15 epochs with a validation split of 0.4.
Accuracy: 96.30%
https://colab.research.google.com/drive/1gwa8Ohl80HxeSlcW7ZXW0PV5aBxLfo8m

If we change the activation functions to :
sigmoid, sigmoid, sigmoid, softmax respectively, and increase the number of epochs to 20, our accuracy peaks at 75.49%.
https://colab.research.google.com/drive/1P8NY6WP5Ccp1S8FUURKuvtrbsE5u5l31

Changing the activation functions to : relu, relu, relu, softmax respectively, keeping number of epochs at 20. Our accuracy stabilizes at 50%.
https://colab.research.google.com/drive/1CLMnX1v1PWHFsxREODcN5qsuBsBcO54Q

# 5. CONCLUSION

## 5.1 ADVANTAGES OF THE MODEL:

•The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. Since deep learning can model very complex features, a CNN provides a general-purpose learning framework not requiring beforehand feature extraction and fine-tuning, which is an advantage over traditional classifiers.

•CNN is also computationally efficient. It uses special convolution and pooling operations and performs parameter sharing. This enables CNN models to run on any device, making them universally attractive.

•Using data augmentation in the Convolutional Neural Network approach decreases the chances of overfitting. Thus, Malaria detection systems using deep learning proved to be faster than most of the traditional techniques. The system is easy to use and user-friendly.

## 5.2 Possible future directions :

Stage 1: Collection of the data on malaria from different organizations (National Family Health Survey, Indian Council of Medical research (ICMR), Department of health research, Institute for Health Metrics and Evaluation). Also need to have a meeting with Madhya Pradesh (MP) government officials for collecting data and funding (1-2 months).

Stage 2: Analysis of the collected data and detection of malaria in different datasets in various districts of MP state (1 month).

Stage 3: Write and submit a proposal to various funding agency for the funding with preliminary data (3-4 months).

Stage 4: After receiving the funding, next plan will be to regularize the data with help of field studies in the MP districts (will start at Sehore district first followed by other districts).

Stage 5: Next step would be acquiring malaria data signing MoU's with various pathologies and Institutions conducting researches in the field of malaria(National Institute of malaria research - N I M R). (8-12 months).

Stage 6: Based on the collected data namely blood cells and clinical data from the field studies, we will create a prototype model for the data analysis with the help of different machine learning techniques for detection of malaria. (3-4 months).

Stage 7: Applying for the funding agency.

## 6. REFERENCES

1) Sayali N. Dharpal,2Dr.A.V. Malviya "Automated Enumeration of Malaria Parasite   Using SVM Classifier"- https://ijrti.org/papers/IJRTI1810027.pdf
2) Aliyu Abubakar , Mohammed Ajuji and Ibrahim Usman Yahya "DeepFMD: Computational Analysis for Malaria Detection in blood-Smear Images Using Deep-Learning Features"
3) Mehedi Masud, Hesham Alhumyani, Sultan S. Alshamrani, Omar Cheikhrouhou, Saleh Ibrahim,Ghulam Muhammad, M. Shamim Hossain , and Mohammad Shorfuzzaman1 "Leveraging Deep Learning Techniques for Malaria Parasite Detection Using Mobile Application" -https://www.hindawi.com/journals/wcmc/2020/8895429/

4) Sivaramakrishnan Rajaraman1, Sameer K. Antani1, Mahdieh Poostchi1, Kamolrat Silamut2, Md. A. Hossain3, Richard J. Maude2,4,5, Stefan Jaeger1 and George R. Thoma1 " Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images"

5) Aliyu Abubakar 1,2,* , Mohammed Ajuji 2 and Ibrahim Usman Yahya "DeepFMD: Computational Analysis for Malaria Detection in Blood-Smear Images Using Deep-Learning Features"

6) K. M. Faizullah Fuhad, Jannat Ferdousey Tuba, Md. Rabiul Ali Sarker, Sifat Momen, Nabeel Mohammed, and Tanzilur Rahman "Deep Learning Based Automatic Malaria Parasite Detection from Blood Smear and Its Smartphone Based Application" https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7277980/

7) Dev KumarDasaMadhumalaGhoshaMallikaPalbAsok K.MaitibChandanChakrabortya - "Machine learning approach for automated screening of malaria parasite using light microscopic images" - https://www.sciencedirect.com/science/article/pii/S0968432812002594

8) Vijayalakshmi & Rajesh Kanna ' Deep learning approach to detect malaria from microscopic images' https://link.springer.com/content/pdf/10.1007/s11042-019-7162-y.pdf

9) D. Anggraini; A.S. Nugroho; C. Pratama; I. E. Rozi; V. Pragesjvara; M. Gunawan "Automated status identification of microscopic images obtained from malaria thin blood smears using bayes decision: A study case in plasmodium falciparum' -

10) Mandal S, Kumar A, Chatterjee J, Manjunatha M, Ray AK (2010) Segmentation of blood smear images using normalized cuts for detection of malarial parasites. Annual IEEE India Conference (INDICON), Kolkata, pp 1–4 'https://www.semanticscholar.org/paper/Segmentation-of-blood-smear-images-using-normalized-Mandal-Kumar/dff12fc7dd6e7cf58fe36ea845e6962fda82fa2a'

11) P.A.PattanaikaMohitMittalbMohammad ZubairKhancS.N.Pandad " Malaria detection using deep residual networks with mobile microscopy "
https://www.sciencedirect.com/science/article/pii/S1319157820304171#b0065