

# Malaria Detection using Image Processing and Machine Learning

Cynara Gomes<sup>1</sup>, Abhishek Kanojiya<sup>1</sup>, Abhishek Yadav<sup>1</sup>, Kirti Motwani<sup>2</sup>

<sup>1</sup>U. G. Student, Department of Computer Engineering, Xavier Institute of Engineering, Mahim, Mumbai-400016

<sup>2</sup>Professor, Department of Computer Engineering, Xavier Institute of Engineering, Mahim, Mumbai- 400016

\*\*\*

**Abstract** - Malaria is a deadly, infectious and life-threatening mosquito-borne blood disease caused by Plasmodium parasites. The conventional and most standard way of diagnosing malaria is by visually examining blood smears via microscope for parasite-infected red blood cells under the microscope by qualified technicians. This method is inefficient and time consuming and the diagnosis depends on the experience and the knowledge of the person doing the examination. Automatic image recognition technologies based on image processing have been applied to malaria blood smears for diagnosis before. However, the practical performance has not been up to the mark so far. This gives us all the motivation to make malaria detection and diagnosis fast, easy and efficient. Our main aim is to build a model that can detect cells from images of multiple cells in thin blood smear on standard microscope slides and classify them as either infected or uninfected with early and effective testing using image processing. And also perform classification on the infected cell image using machine learning.

**Key Words:** Malaria, Falciparum, Watershed, Morphological Segmentation, Edge Detection, Segmentation.

## 1. INTRODUCTION

Malaria is a deadly, infectious disease caused by the Plasmodium parasite which is transmitted by the bites of female Anopheles mosquitoes. According to the World Malaria Report 2019 published by WHO [1], there were an estimated 405,000 malaria related deaths in the preceding year. The disease is curable but early detection holds the key. Existing methods used to detect Malaria include microscopic detection of infected cells in a laboratory. The method is both expensive and tedious. An estimated 93 percent of all Malaria cases in 2018 were reported in the WHO African region. The region also has one of the lowest per capita incomes across the world. A faster, low cost, and reliable alternative to microscopic detection of Malaria is proposed in this model.

### 1.1 Problem Statement

We propose an image processing model for detection of malaria infected cells. We use image processing techniques to detect parasite-infected red blood cells in thin smears on standard microscope slides. The most widely used present

day method is analyzing thin blood smears under a microscope, and visually searching for contaminated cells. A clinician manually counts the number of parasitic red blood cells - sometimes up to 5,000 cells (according to WHO protocol) [2]. Malaria could be forestalled, controlled, and relieved all the more adequately if an increasingly precise and effective symptomatic techniques were accessible. We have utilized image processing procedures to identify the nearness of malaria contaminated cells. And to classify the stage of malaria whether it is falciparum which is the most deadliest stage in malaria or non-falciparum, for this we use machine learning technologies.

### 1.2 Scope of the Project

The Malaria Detection from thin film blood smear images demands segmentation of single blood cells from the microscopic blood slide images which can be taken from a pathologist and the dataset would contain cell images that are not segmented. Hence, segmentation in the proposed method is done using a variety of image processing techniques.

Edge detection techniques and segmentation techniques used in this system overcomes the issue of overlapping of cells by eliminating the noise and finding the discontinuities of the cells. It differentiates each cell and detects the infection in the cell using morphological segmentation.

Also, all the images are raw and have different intensities, and since there is no uniformity in all the images, detection of cells and infection is very difficult. To overcome this problem, the proposed method uses histogram matching where all the images are standard and has the same intensity which in turn increases the accuracy level.

## 2. RELATED WORK

Malaria being one of the most fatal disease has been at the focal point of some major studies in the recent past. Some of them are described briefly here.

### 2.1 CNN-based image analysis for malaria diagnosis

The paper [3] uses a CNN model to detect parasite infected red blood cells in thin smears on standard microscopic slides prepared using routine methods. It is inspired by experiments on the underlying physiological mechanisms in the visual cortex of felines for recognizing objects.

The research work was pretty successful in classifying the individual cells but the research was based on individual RBCs rather than an entire blood sample.

## 2.2 Detection of Malarial Parasite in Blood Images by two classification Methods: Support Vector Machine (SVM) and Artificial Neural Network (ANN)

The study [4] presents that there are many systems which describe the computerized methods of image analysis that commonly involves three main phases. In the first phase of pre-processing, luminance of the image is corrected and transformed to a constant color space. At the second step, a histogram-based image segmentation process is used which helps in avoiding maximum artifacts and over stained objects. Later, a back propagation neural network was used for classification.

This paper set us a go to reference for the steps at granular level yet the examination depended on singular RBCs as opposed to a whole blood sample.

## 2.3 Blood Cells Counting using Python OpenCV

A more accurate method of counting blood cells using Python OpenCV is explored in [5]. It uses images of blood obtained by keeping blood samples under a microscope to compute number of cells. Image processing is a method which involves signal processing and mathematical procedure.

In this study, the images were processed and a blob detection algorithm was used to detect and differentiate RBCs from WBCs. A cell counting method was also used to provide an actual count of the RBCs and WBCs detected. The automation comes with a GUI backed up with a database.

## 2.4 Detection of malaria parasites using digital image processing

The literature [7] proposes a parasite detection technique which is based on digital image processing. Images of thin blood smear are used and with the help of image processing approach the parasite in the cells are identified.

This approach was very much suited for us to provide enough motivation and guidance for the image processing techniques.

## 3. METHODOLOGY

At present, the recognition of Malaria parasite in single cell slide is totally manual. This procedure could be rearranged by capturing an image of the blood smear and afterwards utilizing the proposed model to arrange whether the cells are contaminated or not. The proposed model uses the utilization of image processing systems to improve existing techniques and abbreviate the time taken for recognition of malaria parasite in blood tests. The dataset is manually collected from the CDC's Division of parasitic infection and Malaria [6]. Our concentration here is to make a

mechanized capacity to distinguish the nearness of Malaria parasite in slight blood spread and measure the segment of RBC in the example that are tainted, primary assignment is to fragment the contaminations, for which segmentation of the cells is the earlier undertaking. Segmentation techniques involve methodologies based on Edge Detection[8], Watershed Segmentation[9,11] and Morphological segmentation[10]. Once the cells are segmented, the infections are segmented. This is done via using a threshold intensity pixel value for the infection, i.e. if the pixel value is in the range of the threshold the infections are identified.

### 3.1 Segmentation of cells

When we "segment" an image, we distinguish the regions of interest (ROIs) from the non-ROI portion, generally creating a binary mask of what we want to qualify, quantify, track, etc. Segmentation is a critical part of many image processing problems, and is worth considering in some depth. Here also we will segment our cell images. One of the major issues in segmenting human blood cells is the contiguity of the cells, one may find a lot of overlapping of cells. Segmenting touching object is one of the most difficult task in image processing. We use different approaches like Edge Detection, Watershed Segmentation, Morphological Segmentation.

- 1) Edge Detection(Gradient Based Techniques): The gradient based edge detections look for the first derivative of an image where the maxima and minima are occur. These techniques used sobel, prewitt and robert's cross operator for finding the edges.
- 2) Edge Detection(Gaussian Based Techniques): The main purpose of this technique is to detect the zero crossings in the second order derivative of an image to find edges. The Gaussian based techniques are Laplacian of Gaussian (LOG) and Canny edge detection.

The LOG operator works best in all of the edge detection techniques since it can eliminate the noise levels and discontinuities in the cells better as compared to the other edge detection techniques which would otherwise not prove to be good for further detection and it would be hard to segment the cells.

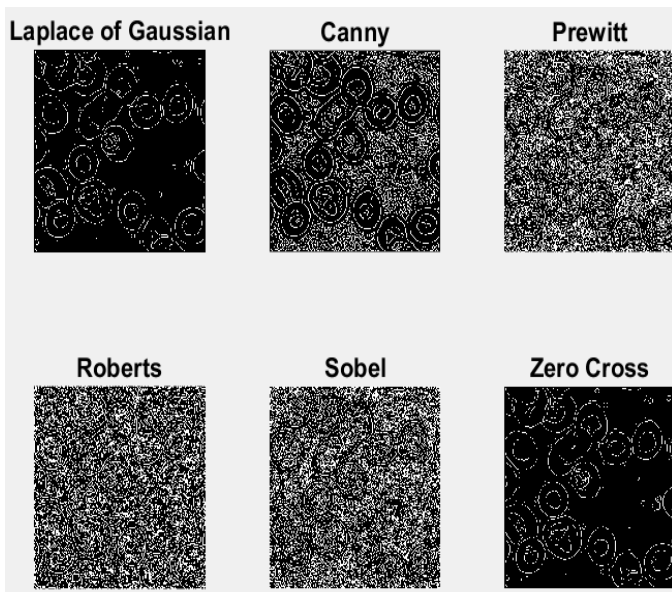


Fig -1: Comparison between edge detection techniques

3) Watershed Segmentation: Watershed segmentation algorithm is applied on the images in the dataset, it segments or divides the different cells present in the image. These divisions help in simplifying the detection and classification of malaria parasite in the blood samples.

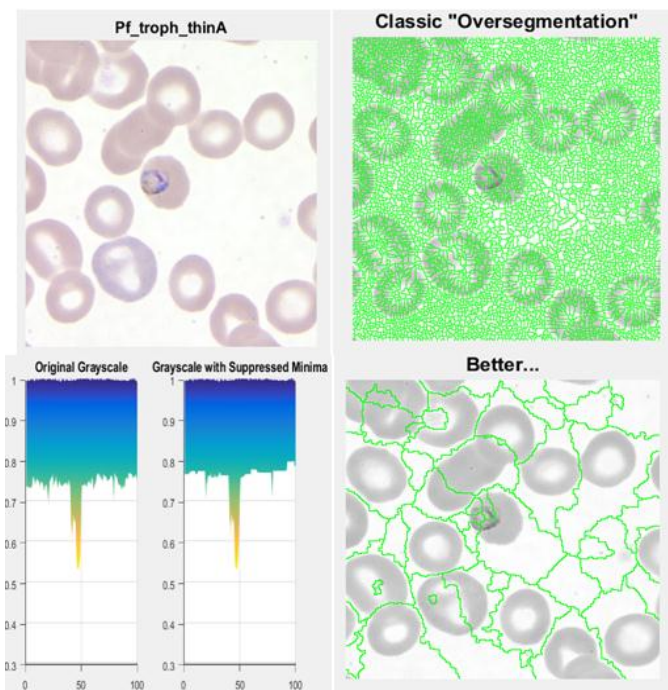


Fig -2: Watershed Segmentation

4) Morphological Segmentation: The term morphology refers to the shape and size of the objects within the image. In our study we have used the shapes of the cells, as most of the cells were circular in shape and of a

particular radii range. Thus, a circle detection algorithm was designed which could detect the cells based on the shape of the cells.

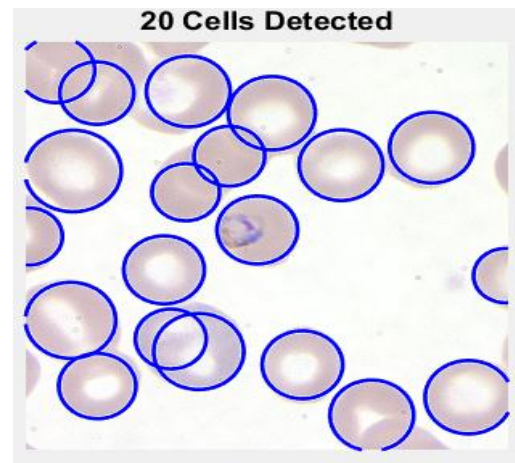


Fig -3: Morphological Segmentation

### 3.2 Segmentation of infection

After successful detection of the cells the next task is to spot the infections inside the cells. So, for this we would be using the pixel intensity values, and observe the range of intensities wherever the infections. But the problem with it is that the illumination and color scale of one image may differ to other. For this we would use histogram matching[12] to have a uniform color scale.

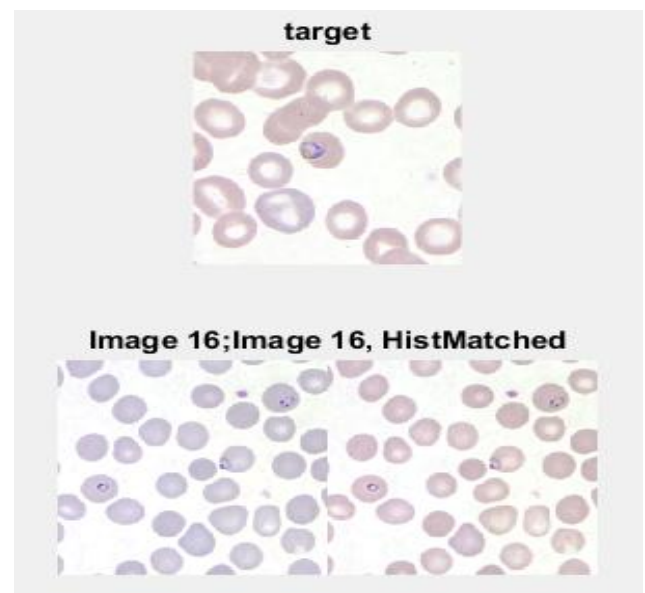


Fig -4: Histogram Matching

### 3.3 Classification of infected cells

We use machine learning to classify the cells if infected. Support Vector Machine(SVM) is a supervised machine learning algorithm that can be used for both classification and

regression tasks. An SVM classifies data by finding the best hyperplane that separates data points of one class from those of the other class. The best hyperplane for an SVM implies the one with the largest margin between the two classes. Margin means the maximal width of the slab corresponding to the hyperplane that has no interior data points.

Here cubic SVM type classifier is employed where the kernel function of the classifier is cubic given as  $k(x_i, x_j)$ .

$$\kappa(x_i, x_j) = (x_i^T x_j)^3$$

#### 4. MODEL EVALUATION

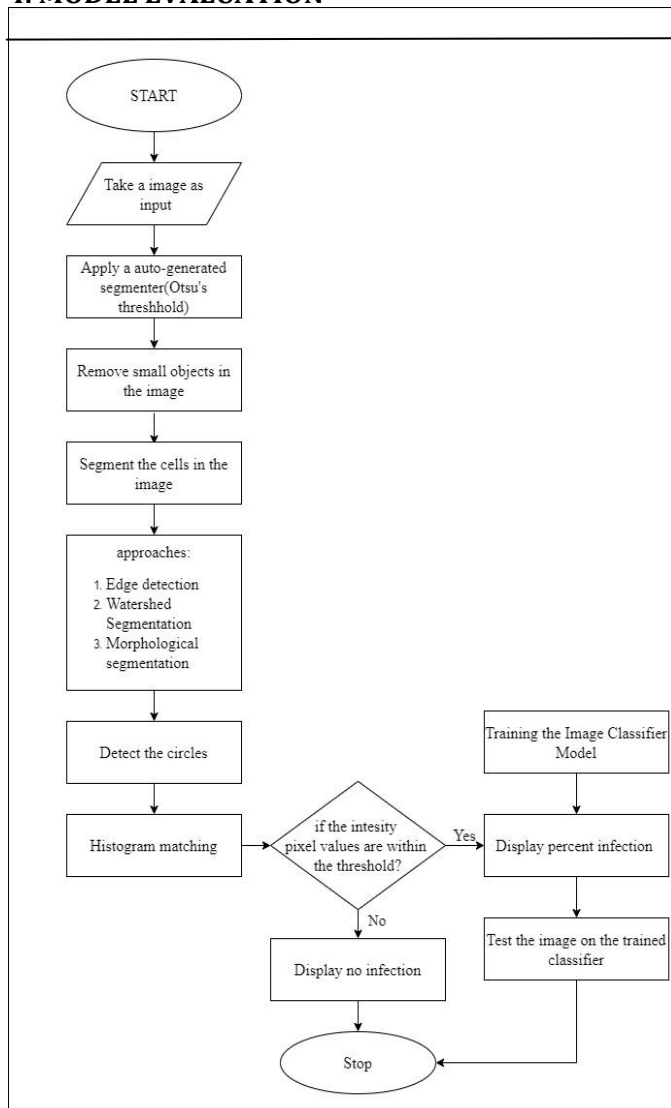


Fig -5: Flowchart

#### 4.1 Data Source

We used archived blood smear images acquired from the CDC's Division of parasitic infection and Malaria [6] as input. Our data set contains images of different malaria parasites

like Plasmodium Falciparum, Plasmodium Knowlesi, Plasmodium Malariae, Plasmodium Ovale, Plasmodium Vivax.

#### 4.2 Steps used in Image Processing

1. We apply an auto-generated segmenter that performs five steps:(1)Convert to grayscale.(2)Initialize segmentation with Otsu's threshold.(3)Filter components by area.(4)Form masked from input image and segmented image.
2. Detecting edges-We have used various edge detection techniques like Canny, Laplacian of Gaussian ,Prewitt, Roberts, Sobel and Zerocross operator out of which LOG gives us the best result.
3. Combining the edges(logically) with the segmented regions.
4. Improve edge mask by performing morphological operations like imclose, skeleton, etc.
5. Cleaning of the mask and refine the mask.
6. Use different segmentation techniques like watershed, thresholding, k-means clustering based segmentation. Of which we select watershed segmentation since the other techniques have tiny pores and do not give good results for further segmentation.
7. To improve the result and flatten the pools from the results of watershed we use imhmin function.
8. We use another segmentation technique i.e. Morphological segmentation to overcome the shortcomings in watershed segmentation. This technique detects the circles by setting a specific range of radius.
9. Histogram matching is done since all the images have different intensities and for classification we require all the images to be standard with the same intensity.
10. Finally, if the intensity values are within the threshold we display the percentage of cells infected.

#### 4.3 Classification Stage

We used Classification Learner to automatically train a selection of different classification models on our data. We use automated training to quickly try a selection of model types, then explore promising models interactively. Here, the two classes are falciparum and non-falciparum. If the cell is infected in the image processing stage, then we undergo a test for classifying whether the malaria parasite is falciparum or some other malaria parasite(non-falciparum).

Machine learning algorithms like Cubic SVM, Linear SVM and cosine KNN are used and out of which the best algorithm with the highest accuracy score is selected for classification purpose.

#### 4.4 Outputs

All the images are read in MATLAB and an objective/target image is chosen on which all the tasks are

performed and afterward on all the images. After performing all the steps of image processing, the images of blood cells are displayed whether infected or not infected, and if infected the number of cells infected and their percentage is shown and then we train the model for classification of the infected cell.

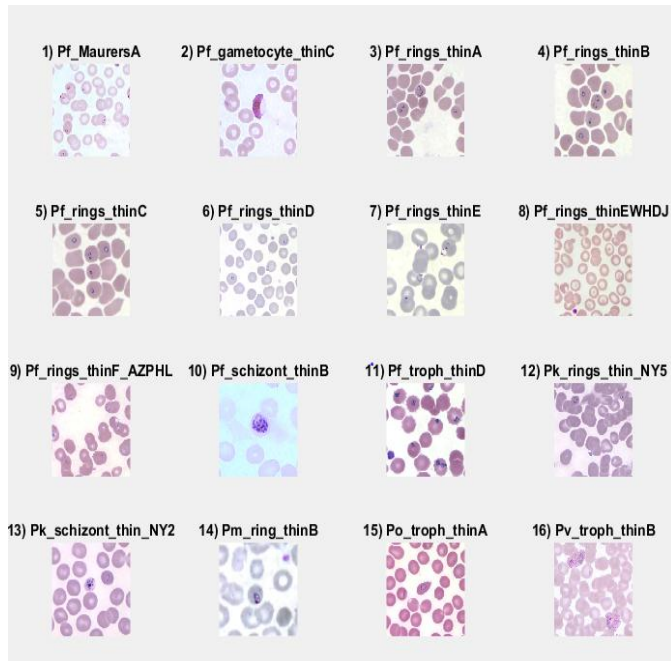


Fig -6: Input images

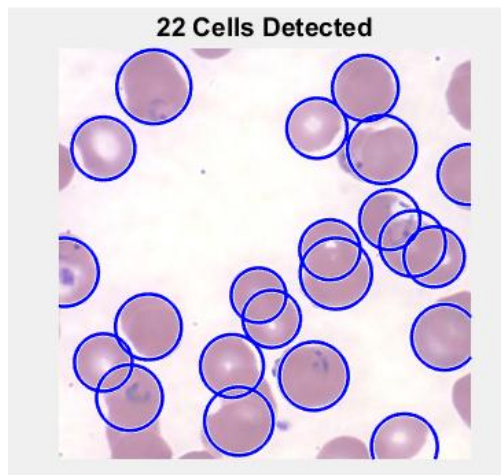


Fig -7: Detection of cells

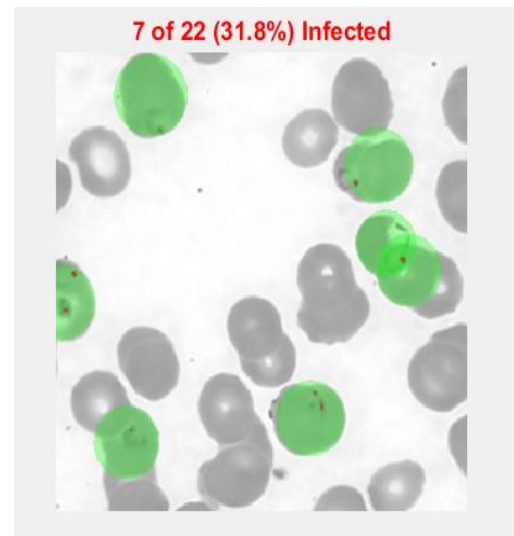


Fig -8: Output (Percentage of infected cells)

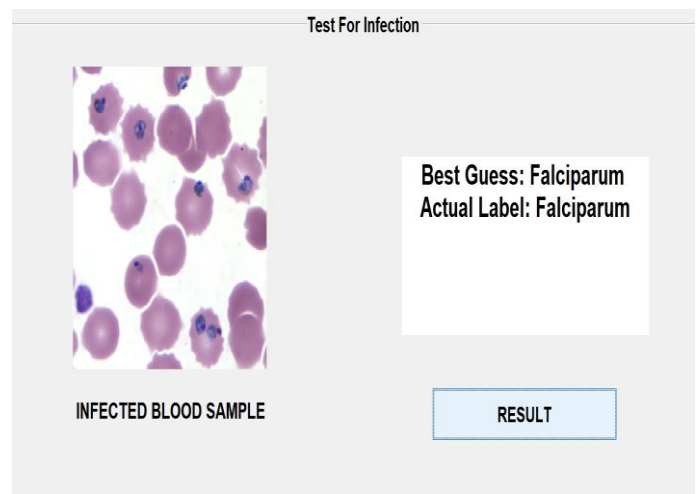


Fig -9: Output (Classification of infection)

#### 4.5 Result

Table 1. presents the accuracy of detecting the cells whether infected or not infected.

Table -1: Accuracy Outcome of Infection Detection

	Actual data	System's detection(ip)	Accuracy
Infected	113	102	90.2%
Normal	11	11	100%

Table 2. presents the overall accuracy of the machine learning algorithms. The performance of the model was evaluated using the performance measures that include: Accuracy, Precision, Recall and F-Score.

**Table -2:** Overall Accuracy

Algorithm	Accuracy	Precision	Recall	F-Score
Cubic SVM	86.1	71.2	86.3	77.9
Linear SVM	79.2	51.2	84.3	63.87
Cosine KNN	74.4	70.2	64.7	67.33

### 5. CONCLUSIONS

Considering our above methodology, we come to the conclusion that the newly designed system of image processing technologies is suitable for parasite detection. The morphological segmentation technique and improved watershed segmentation technique proves to be the best in detection of the cells and segmenting them from the non-region of interest portion. To characterize the sort of plasmodium parasite for which we utilized machine learning technologies and cubic SVM proves to be the best by having the highest accuracy score. The system has been tested with around 110 thin film blood smear images and the results are satisfactory. The inconsistencies among the images was a challenge but we have tried to make our system robust.

### APPENDIX

WHO	World Health Organization
ANN	Artificial Neural Network
CNN	Convolution Neural Network
SVM	Support Vector Machine
ROI	Region of Interest
CDC	Centers for Disease Control and Prevention
LOG	Laplacian of Gaussian
KNN	K- Nearest Neighbor

### REFERENCES

- [1] World Health Organization, Malaria, <https://www.who.int/news-room/factsheets/detail/malaria-report-2019>
- [2] World Health Organization, Malaria, [https://www.who.int/news-room/factsheets/detail/malaria\(2018\)](https://www.who.int/news-room/factsheets/detail/malaria(2018))
- [3] Z. Liang et al., "CNN-based image analysis for malaria diagnosis," 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, 2016, pp. 493-496, doi: 10.1109/BIBM.2016.7822567
- [4] Afkhami, S., Rashidi Heram-Abadi, H., (2017).Detection of Malarial Parasite in Blood Images by two classification Methods:Support Vector Machine (SVM) and Artificial Neural Network (ANN), Int. J. of Comp. & Info. Tech. (IJOCIT), 5(2): 81-100.
- [5] Meimban, Raymond Joseph & Fernando, Alexa Ray & Monsura, Armil & Ranada, Jonrey & Apduhan, J.C. (2018). Blood Cells Counting using Python OpenCV. doi:50-53. 10.1109/ICSP.2018.8652384.
- [6] CDC – DPDX – MALARIA <https://www.cdc.gov/dpdx/malaria/index.html>
- [7] Bashir, A., Mustafa, Z. A., Abdelhameid, I., & Ibrahim, R. (2017). Detection of malaria parasites using digital image processing. 2017 International Conference on Computing and Electronics Engineering (ICCCCEE) doi:10.1109/iccccee.2017.7867644.
- [8] Amer, G. M. H., & Abushaala, A. M. (2015). Edge detection methods. 2015 2nd World Symposium on Web Applications and Networking (WSWAN). doi:10.1109/wswan.2015.7210349.
- [9] Chourasiya, S. (2014). Automatic Red Blood Cell Counting using Watershed Segmentation.
- [10] Bhusare, M.P., & Bhosale, S.D. (2019). Automatic Blood Cell Counting using Morphological Image Processing operations.
- [11] Mohamad Sharif, Johan & Miswan, M. & Ngadi, Md & Abdul Jamil, Muhammad Mahadi. (2012). Red blood cell segmentation using masking and watershed algorithm: A preliminary study 2012 International Conference on Biomedical Engineering, ICoBE 2012 10.1109/ICoBE.2012.6179016.
- [12] Shapira, D., Avidan, S., & Hel-Or, Y. (2013). Multiple histogram matching. 2013 IEEE International Conference on Image Processing, doi:10.1109/icip.2013.6738468.