

A Novel Method for Automatic Detection of Malaria Parasite Stage in Microscopic Blood Image

Kshipra Charpe
Research Student,
Electronics Engineering
Department
AISSMS's Institute of
Information Technology
Pune, India

V.K. Bairagi
Associate Professor
E&TC Engineering
Department
AISSMS's Institute of
Information Technology
Pune, India

Shama Desarda
Assistant Professor
E&TC Department
Sinhgad College of
Engineering, Pune India

Sheetal Barshikar
Lecturer, Electronics
Department.
Sou. Venutai Chavan
Polytechnic, Pune India.

ABSTRACT

Malaria is caused due to the mosquito bite hence the parasite enter into blood through the saliva of the mosquito. The malaria parasite directly infects the red blood cells, therefore to design an automatic detection system, the red blood cells should be segmented from the artifacts and background in a microscopic image. Here in this paper watershed transform is used with the distance transform which separates even the overlapped red blood cells more efficiently, which are useful for the diagnosis of parasite and for the parasitemia too. The result shows improvement in diagnostic accuracy of detection of the parasite in Red Blood Cells and also describing the life cycle stage of the parasite. The accuracy, sensitivity and specificity achieved were as 97.7%, 97.4% and 97.7% respectively.

Keywords

Watershed, Parasitemia, Texture and Statistic Features.

1. INTRODUCTION

Malaria is a disease caused by a female *Anopheles* mosquito, which directly affects the Red Blood Cells (RBC) [2]. By the different types of malaria parasites the RBC in the blood are infected. There are mainly three life stages of malaria parasite. The medication and its dose differ accordingly by the parasite life stage in the blood of the patient. Hence, to make easy diagnosis and medication, it is a need to detect the parasitic RBC and then their life stage of the parasite. Malaria infected people in the year 2013 is 300-500 million which is given by World Health Organization (WHO) analysis report [1] [2].

The life cycle stages of malaria parasite are as shown in figure 1. They are primarily described into three stages [3]. The host female *Anopheles* mosquito injects the sporozites into the blood from their saliva [18]. These sporozites then enter into the liver and assault the liver cells. The infected liver cells then contain the full-grown sporozites which later burst and the merozoites are spread all over the blood infecting the red blood cells. These merozoites give rise to the trophozoites, schizonts, gametocytes, which are basically considered as the stages of parasite.

Trophozoite: At this stage the parasite is in a ring shape with small one or two cytoplasm dots.

Schizonts: They are matured trophozoites which can be recognized by the circular shape with many of the cytoplasm dots in it and the number of minute chromatins.

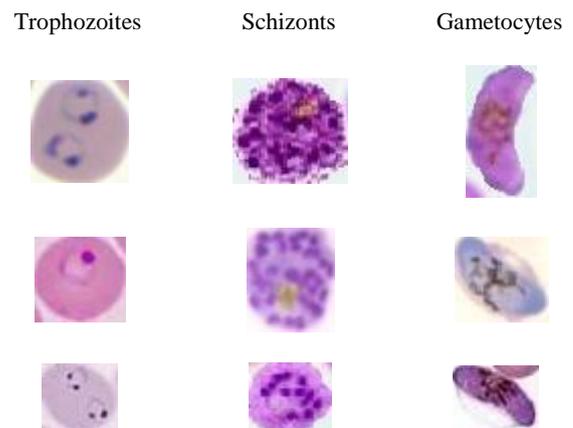


Fig. 1 Different life stages of malaria parasite

Gametocytes: This stage is known as the sexual growth stage as the male female parasite are formed in this stage. The male size is smaller than the female gametocytes, which are somewhat banana shaped.

The count of the parasite in the blood sample is called as parasitemia. Detection of parasitemia is very important in the malaria detection and medication. It helps to know the severity of the disease.

There various different ways to detect the malaria parasite in the blood sample using image processing concepts, proposed by researchers, some of them are like- Yashasvi Purwar et.al [6] has used k-mean clustering algorithm for detection of the parasite while Subhamoy Mandal, J Chatterjee [8] used different color space like RGB, HSV, YCbCr and NTSC, and finally concluded that HSV is the best channel for detection which gives good results. R. C. S. Morling, et.al [13][14] detected parasites on the basis of size and morphology and has proposed an annular ring ratio method for segmentation of blood cells. Iis Hamsir Ayub Wahab, et.al [15] used different second order statistical measurement combined with the principal component analysis method for detection of malaria parasite.

The organization of our paper is the section second gives the detail about the work done and all the different image processing techniques used. While the third section provides all the results obtained in our study and finally the section fourth gives the conclusion part.

2. METHDOLOGY

The steps followed by the proposed system are like image acquisition, segmentation, feature extraction and then to classify the images using Support Vector Machine (SVM). Figure 6 shows the flow followed by the proposed method using flowchart.

2.1 Data Acquisition

The images of blood samples are taken using the microscope with an attached inbuilt camera of high definition. A database was created first using images from the online database like Center for Disease Control (CDC) and Prevention and from Public Health Image Library [3] [4] consisting of parasitic and non-parasitic blood samples. Some database of images used in the study was also obtained from Smt. Kashibai Navale General Hospital & Research Centre, Pune consisting of both malaria positive blood samples and the normal blood samples images. All the images are of different size, color tone, intensity; which were normalized later for use. The normalization process includes the basic processing of image enhancement and making all images of the same size. The database is very useful for training the phase of classification.

The complete database used consists images of normal RBC stained samples as shown in figure 2 and the infected RBC as shown in figure 3 (a) ring shape trophozoites, (b) and (c) are solid circle known as schizonts and (d) and (e) are the cucumber shape male/ female gametocytes.

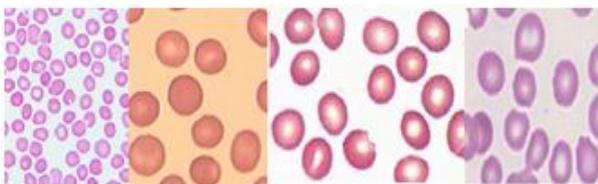


Fig.2 Microscopic blood images of normal RBC.

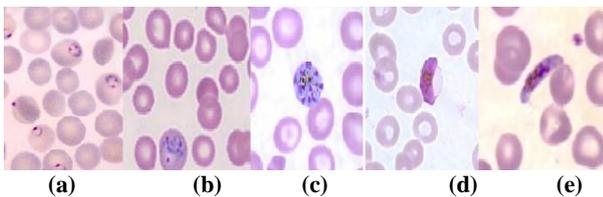


Fig.3 Microscopic blood images of the malaria parasite infected RBC.

2.2 Pre-processing

Preprocessing on an image is done before applying any algorithm. In this step image is processed and converted into the format as required for the further implementation of the algorithm. The original RGB image is firstly converted into the gray scale image. Then the RGB and Hue, Saturation, Value (HSV) bands are separated. After the successful completion it is observed that the green true color band and the value band give more proper image consisting of all the elements as in the original image. Therefore an object mask was made by combining of the green and value band. The complete preprocessing and the difference between the bands are shown in figure 4.

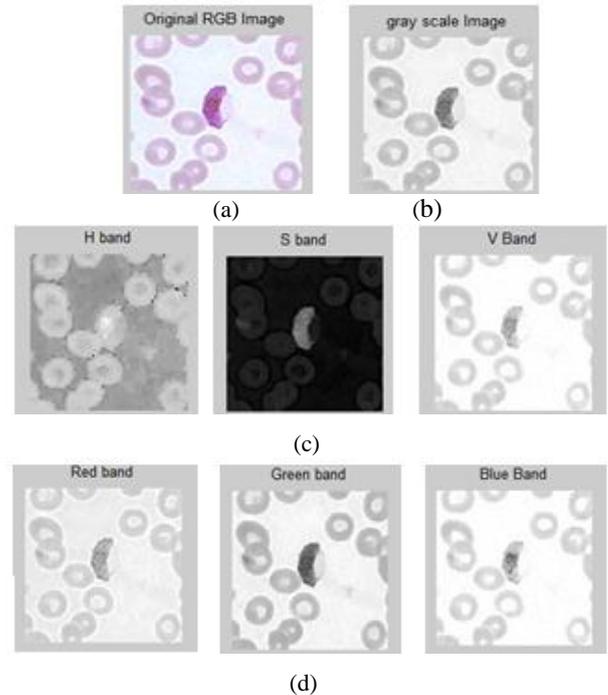


Fig.4 (a) original image, (b) grasacle image, (c) HSV band separated, (d) RGB band separated

2.3 Separation and counting of RBC

The malaria parasite affects the Red Blood Corpuscles (RBC) directly when injected into the blood [1]. There are many changes occurred in the RBC because of the presence of the malaria parasite in it. The count of RBC is also affected if the parasite multiplication starts. Hence, while working on the detection of malaria parasite, there is a basic requirement of separation of RBC from the background and from the artifacts (WBC, platelets, dirt, dye crystals etc) [13].

2.3.1 Segmentation

The purpose of segmentation is to subdivide an image into meaningful non-overlapping regions, which would be used for further analysis (i.e treatment and medication) and the classification of life cycle stages of parasite. As the malaria parasites are injected in human blood they infect the RBC and multiply in RBC only. So the RBC's has to be segmented from the background other artefacts (artefacts represents the WBC, platelets, dirt, dye crystals etc). The RBC's are located very near to each other, or sometimes overlapping each other, which leads to difficulty in counting of RBC. Hence in this step the separation of the RBC and artefacts is performed which is known as segmentation. The techniques used in our work are Distance transform and watershed algorithm.

2.3.1.1 Watershed threshold

As the RBC in blood samples are very near to each other or sometimes overlapping, it creates a problem while considering each RBC and hence there may be miscout of the RBC. Therefore Watershed transform [17] [20] is used in our work, which uses the connectivity in the given image pixel. The distance transform calculates the difference between the pixel and the nearest non-zero pixel. Euclidean distance formula is used by default in this method for calculation. This method can be used for counting the overlapped RBC in the blood smear. The output after applying the watershed algorithm is shown in Figure 5 (a), (b) which shows the segmented RBC's.

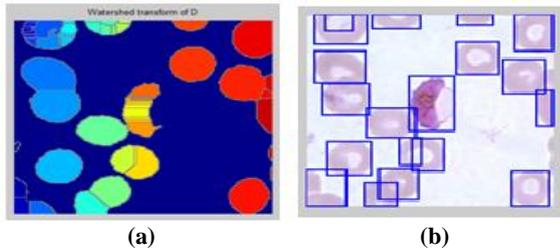


Fig. 5 (a) Output of the watershed algorithm (b) Output of the RBC segmentation after watershed

2.3.2 Counting of the total RBC and infected RBC in the sample

The process of counting number of RBC infected by malaria parasite is called as parasitemia. Parasitemia is very useful to be counted as according to the literature survey the count of RBC decides the dose of the medicines given to patients.

2.3.3 Feature extraction

Feature extraction is the process of considering some of the particular points on an image so as to identify and classify. The features used in our work include:

2.3.3.1 Basic Feature

The basic features on which the infected and non-infected RBC can be differentiated are size, shape, color, intensity values [6] [7] [8]. The RBC's which are infected have deviation in their size and shape as compared to the normal once, according to the parasite type and stage in which it is. The intensity values are also calculated to differentiate the infected once.

2.3.3.2 Texture based Features

Texture-based features [15] [16] are helpful for identifying the different textures of the infected and non- infected RBC. When the circular shaped schizonts are to be identified these features gives good results for this stage. While calculating the texture features the co-occurrence matrix based components are calculated using following formulas

$$\text{Contrast} = \sum_{i,j} |i - j|^2 p(i, j) \dots \dots \dots (1)$$

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \dots \dots \dots (2)$$

$$\text{Energy} = \sum_{i,j} p(i, j)^2 \dots \dots \dots (3)$$

Where Gray- Level Co-occurrence Matrix (GLCM) is calculated first and then the properties, like contrast, correlation and energy. The GLCM gives the number of times, the pixel valued i is occurred in horizontally with respect to j pixel vertically.

2.3.3.1 Statistical Features

Statistical values of the image on the basis of the histogram are also called as the histogram second order statistical features [17] [18] are calculated.

If image function is $I(x,y)$ of two space variables x and y $x=0,1,2,\dots,N-1$ and $y = 0,1,2,\dots,M-1$. Then the discrete values of function $I(x,y)$ can be taken as $i= 0,1,2,\dots,G-1$ where G represents the total number of intensity levels in the image. These all features are calculated using the histogram of the image $h(i)$ and probability density function $p(i)$. Here the $p(i)$ is the ratio of all the intensity values $h(i)$ and the total number of pixels in the image.

The mean gives an estimate of the average intensity level in the region of the cell and the variance is a measure of the dispersion of region intensity. The value of mean is 0 for black image and 1 for complete white color image.

$$\text{Mean} : \mu = \sum_{i=0}^{G-1} ip(i) \dots \dots \dots (4)$$

$$\text{Variance} : \sigma^2 = \sum_{i=0}^{G-1} (i - 1)^2 p(i) \dots \dots \dots (5)$$

Skewness is the measure of the asymmetry of the data around the input sample's mean value. Skewness is negative when the data values are spread out more to the left of the mean while it is positive when the values are spread more to the right.

$$\text{Skewness} : \mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i - \mu)^3 p(i) \dots \dots \dots (6)$$

Kurtosis is the measure of how outlier-prone a data distribution is. The kurtosis of the normal distribution is 3.

$$\text{Kurtosis} : \mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i - \mu)^4 p(i) - 3 \dots \dots \dots (7)$$

For the normal distribution of the image the skewness value is zero.

2.3.4 Classification

This is the final and crucial stage where the features of input image compared with those of the infected RBC image features and classified. In this work multiple stages Support Vector Machine (SVM) is used as a classifier. The train database matrix consisting of the feature values as discussed. Then these train feature matrix is compared with the feature matrix of each RBC from the selected image and finally the classification of the infected, non-infected RBC and stages of RBC is done.

Based on the classification results obtained in the work where calculated using following method-

Diagnostic Accuracy in following way:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \dots \dots \dots (8)$$

$$\text{Sensitivity} = TP / (TP + FN) \dots \dots \dots (9)$$

$$\text{Specificity} = TN / (FP + TN) \dots \dots \dots (10)$$

Where,

- TP - True Positive (Infected sample correctly classified)
- TN - True Negative (Normal sample correctly classified)
- FP - False Positives (Normal sample misclassified)
- FN - False Negative (Infected sample misclassified)

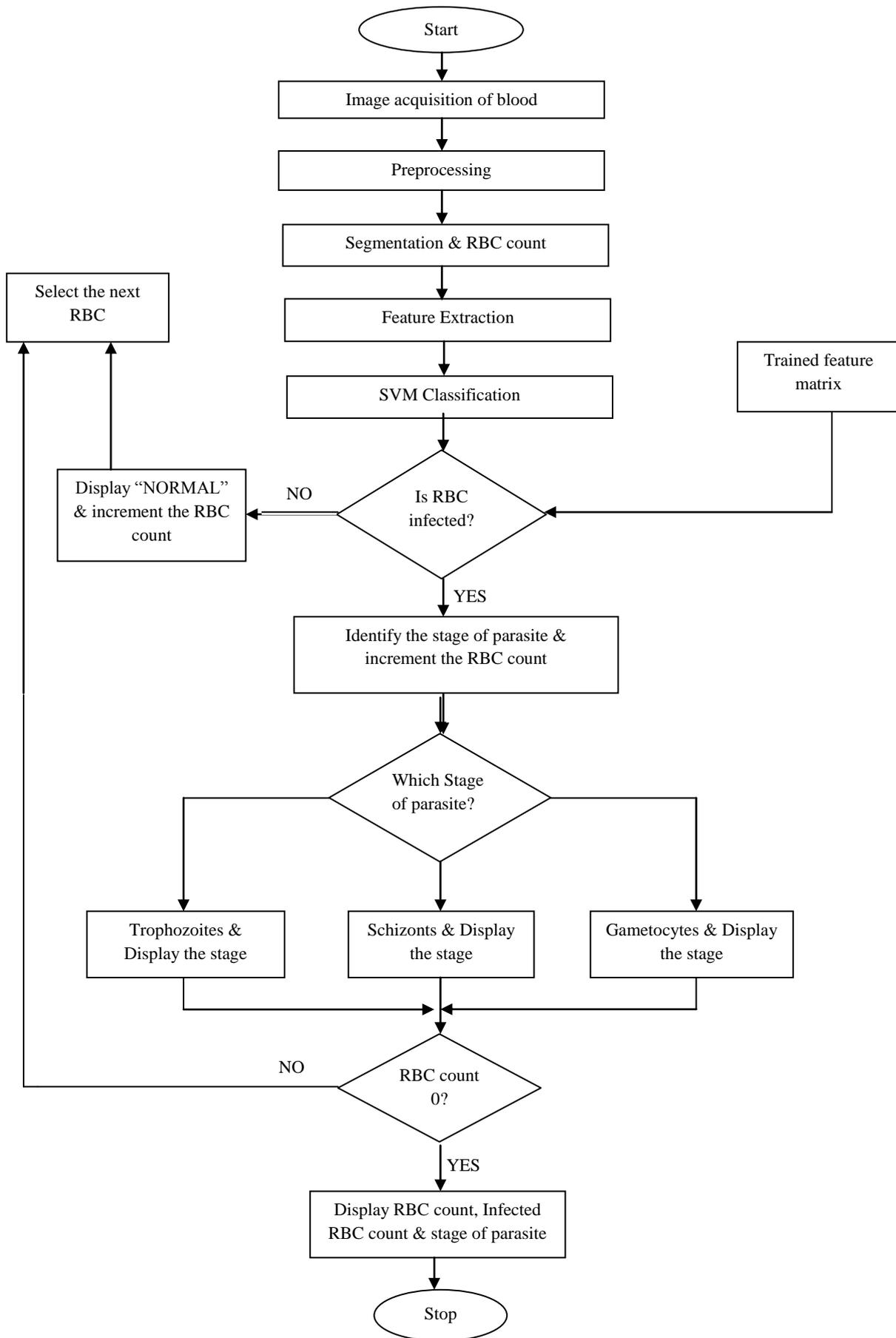


Fig.6 Flow chart of the proposed system

3. EXPERIMENTAL RESULTS

The malaria parasite varies according to the growth and change in life cycle stages. Hence to detect its presence in blood images, the above mentioned features are important to be calculated. These feature values calculated are helpful for training the classifier at its initial stage so as to gain more accuracy. Considering the above 1-7 equations the values of the features are calculated of some images from the database formed, consisting of infected and non-infected blood samples and accordingly the accuracy, sensitivity and specificity using equation 8-10 is calculated as below-

15 images are considered with total RBC 250 among which Infected RBC are 39, Normal RBC are 211.

Table.1 Accuracy, Sensitivity and Specificity calculated in %

Accuracy	Sensitivity	Specificity
97.7 %	97.4%	97.7%

The results of accuracy, sensitivity and specificity calculated in percentage are as shown in the table1.

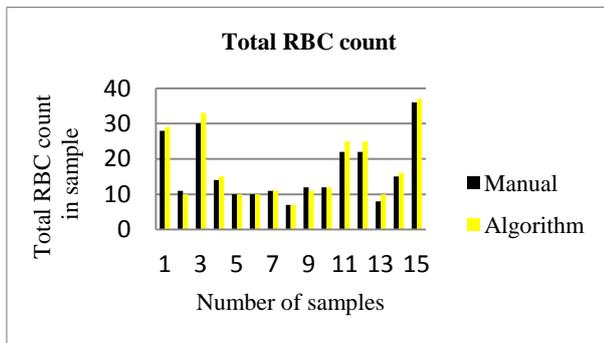


Fig. 7 Represent the graph of total RBC count by manual and algorithm.

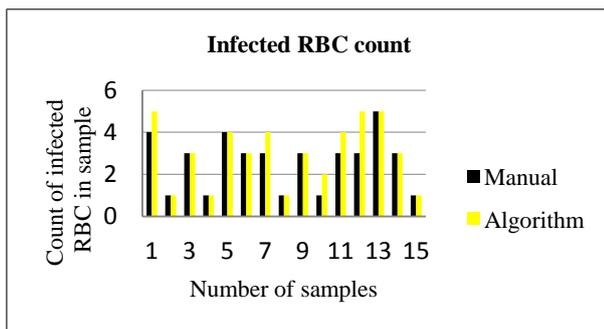


Fig.8 Represent the graph of infected RBC count by the manual and algorithm.

According to the calculation and results the output represented in graph was as shown in figure 7 and figure 8 represent the graph of total RBC count by manual and algorithm. Figure 8 Represent the graph of infected RBC count by the manual and algorithm. According to the graph it can be noticed that the manual count and algorithm count are almost equal while in some cases the parasitic RBC skipped manually are counted by the system.

4. CONCLUSION AND FUTURE SCOPE

After the classification the results obtained for the values of accuracy, sensitivity and specificity were calculated as 97.7%, 97.4% and 97.7% respectively. This system will definitely have some use in other hematological disorder detection. Further it can be extended for other organism detections with updating the database and different features to be extracted.

In future this system can be used for diagnosis of other hematological disorders like leukemia, sickle- cell anemia, any kind of infection and also the RBC and WBC count which helps to diagnose many disorders in human body.

5. ACKNOWLEDGMENTS

We would like to thank Principal of the institute Dr. P. B. Mane, HOD Prof. D. K. Shedge, ME staff technical and nontechnical support. Finally the authors gratefully acknowledge and thanks to Dr. Saurabh Joshi form Smt. Kashibai Navale General Hospital & Research Centre; he helped me in the formation of database in a proper way.

6. REFERENCES

- [1] World Health Organization (WHO) statistical analysis of Malaria diseases in the year 2013. http://www.who.int/malaria/publications/world_malaria_report_2013/en/
- [2] <http://www.nvbdc.gov.in/malaria9.html> (National Vector Borne Disease Control Program (NVDCP) Directorate General of Health Services Ministry of Health and Family Welfare, Delhi)
- [3] World Health Organization. Basic malaria microscopy, 2nd edition, WHO Press, pp-7-88, 2010.
- [4] Centres for Disease Control and Prevention: Public Health Image Library [online]. 2005 [www: http://phil.cdc.gov/phil/home.asp](http://phil.cdc.gov/phil/home.asp).
- [5] Priyamvada Jain, Babina Chakma, Sanjukta Patra, and Pranab Goswami, "Potential Biomarkers and Their Applications for Rapid and Reliable Detection of Malaria", Hindawi Publishing Corporation BioMed Research International Volume 2014, Article ID 852645, pp1-20, 2014.
- [6] Yashasvi Purwar, Sirish L Shah, Gwen Clarke, Areej Almgairi and Atis Muehlenbachs, "Automated and unsupervised detection of malaria parasites in microscopic images", Springer-Malaria Journal, Vol. 10, pp 11-22, 2011.
- [7] K. M. Khatria, V. R. Ratnaparkheb, S. S. Agrawal, A. S. Bhalchandra, "Image Processing Approach for Malaria Parasite Identification", International Journal of Computer Applications, National Conference on Growth of Technologies in Electronics, Telecom and Computers - India's Perception, 2013.
- [8] Subhamoy Mandal, J Chatterjee, "Segmentation of Blood Smear Images using Normalized Cuts for Detection of Malarial Parasites", 2010 Annual IEEE India Conference (INDICON), pp1-4, 2010.
- [9] Aimi Salihah, Abdul-Nasir, Mohd Yusoff Mashor, Zeehaida Mohamed, "Colour Image Segmentation Approach for Detection of Malaria Parasites Using

Various Colour Models and k-Means Clustering”, *Wseas Transactions on Biology and Biomedicine*, Vol. 10, Issue 1, pp 41-55, January 2013.

- [10] Yazan M. Alomari, Siti Norul Huda Sheikh Abdullah, Raja Zaharatul Azma and Khairuddin Omar, “Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm”, *Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume of 2014*, Article ID 979302, pp1-17, 2014.
- [11] V. V. Panchbhai, L. B. Damahe, A. V. Nagpure, and P. N. Chopkar, “RBCs and parasites segmentation from thin smear blood cell images”, *IJ. Image, Graphics and Signal Processing*, Vol.10, pp 54-60, 2012.
- [12] A.S.Abdul Nasir, M.Y.Mashor and Z.Mohamed “Segmentation Based Approach for Detection of Malaria Parasites Using Moving K-Means Clustering”, *IEEE EMBS International Conference on Biomedical Engineering and Science*, pp 653-658, Dec 2012.
- [13] S.Kareem, R. C. S. Morling, “Automated *P.falciparum* detection system for Post-treatment Malaria Diagnosis using Modified Annular Ring Ratio Method”, *14th International conference on modelling and simulation, IEEE computer society*, pp 432-436 2014.
- [14] S.Kareem, R. C. S. Morling, “Automated Malaria Parasite Detection in Thin Blood Films, A Hybrid Illumination and color constancy Insensitive, Morphological Approach”, *14th International conference on modelling and simulation, IEEE computer society*, pp-240-243, 2012.
- [15] Iis Hamsir Ayub Wahab, Adhi Susanto, P. Insap Santosa, Maesadji TJokronegoro, “Principal Component analysis combined with second order statistical feature method for malaria parasite classification ”, *Journal of Theoretical and Applied Information Technology*, Vol 63, Issue 1, May 2014.
- [16] S.S.Savkare, S.P.Narote, “Automatic System for Classification of Erythrocytes Infected with Malria and Identification of Parasite life Stage”, *2nd International Conference on Communication, Computing & Security [ICCCS-2012] Elsevier*, Vol 6, pp 405-410, 2012.
- [17] Cloppet F, Boucher A, “Segmentation of overlapping/aggregating nuclei cells in biological images”, *Pattern Recognition, 2008. ICPR 2008 19th International IEEE Conference*, pp 1-4, 2008.
- [18] Kshipra Charpe, Dr. V. K. Bairagi, “Automated Malaria Parasite and there stage Detection in Microscopic Blood Images”, *9th IEEE Sponsored International Conference on Intelligent Systems & Control, 9th – 10th January, 2015*.
- [19] Dr. Arora, “*Mediacal Parasitology*”, Second edition, CBS publisher and distributors, chapter 5-sporozoa , malaria parasite, pp 67-80.
- [20] Rafael C. Gonzalez and Richard E. Woods, “*Digital Image Processing*”, Third edition 2009.