International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 2 (2015) pp. 4043-4055

http://www.ripublication.com

Identification of *Plasmodium Falciparum* Development Phase in Malaria Infected Red Blood Cells using Adaptive Color Segmentation and Decision Tree based Classification

Adi Pamungkas [1)], Kusworo Adi [1)], Rahmat Gernowo [1)]

[1)] Department of Physics, Faculty of Science and Mathematics,Diponegoro University, Semarang, Indonesia

Abstract

Malaria is a medical emergency that must be treated immediately at this time because it has been infecting millions of people in 90 countries every year. Malaria is caused by a parasite that infects red blood cells transmitted through the bite of Anopheles. One species of plasmodium causing malaria is plasmodium falciparum. This type of plasmodium can lead to malaria tertian which is a malignant malaria type and threatening the life of its sufferer. This study aims to design a system capable of identifying the development phase of plasmodium falciparum in red blood cells with adaptive color segmentation and decision tree based classification. Color segmentation process in this study is done by converting the original image color space based on the color components of RGB (Red, Green, Blue) to HSV color space (Hue, Saturation, Value) then thresholding against Saturation component. The used morphological parameters to distinguish the type of plasmodium falciparum development phase are the area ratio and eccentricity. While the classification process is performed using decision tree algorithm results in 87.67% of accuracy.

Keywords: *malaria, plasmodium falciparum, adaptive color segmentation, decision tree algorithms*

1. INTRODUCTION

Malaria is a global problem, which has been infecting millions of people in 90 countries every year. Malaria is caused by a parasite that infects red blood cells transmitted through the bite of Anopheles [1]. Malaria can cause death, especially in high risk groups i.e infant, toddlers, and pregnant women. Malaria directly also cause anemia and can reduce work productivity [2].

Parasite that cause malaria in humans consist of four types of plasmodium species: P. *falciparum, P. vivax, P. ovale,* and *P.malariae.P. falciparum* and *P. vivax* are the most common types, but the most malignant type is *P. falciparum*, but the most deadly is P. falciparum type. P. falciparum can lead to organ failure and blood abnormality of patient. It also causes cerebral malaria which, if not addressed promptly can lead to death. Therefore, malaria is a medical emergency that must be treated immediately [3].

One field of science that is able to perform that sample analysis is digital image processing. The process of malaria infected blood samples analysis through digital image processing has advantages over the direct observation, which is quicker and easier. Study on the application of digital image processing for analysis of malaria has been conducted by some researchers in the world [3,4,5,6,7].

Das et al (2013) characterized and classify the malaria parasite *(Plasmodium vivax and Plasmodium falciparum)*. Image segmentation use watershed method and two algorithms of classification: Bayesian learning and Support Vector Machine (SVM). The results obtained by Das et al showed that the Bayesian learning algorithm has a higher accuracy in the classification of the malaria parasite that is 84% compared with the SVM algorithm with 83.5% accuracy [3].

While, Adi et al (2014) conducted a study to identify the development phase of Plasmodium falciparum in red blood cells using Neural Network algorithm. Image segmentation is done by converting the RGB image into grayscale image and then thresholding process using otsu method. Extracted feature to identify the development phase of plasmodium falciparum is a binary pattern. Plasmodium falciparum binary pattern is used as the input in the training process using back propagation neural network algorithm. The accuracy of the identification system designed by Adi et al was 87.5% [4].

Based on the background and some supporting previous studies, this study is conducted to develop algorithm for identification system of plasmodium falciparum development phase in red blood cells. Identified development phase of plasmodium falciparum are *trophozoite,* which is the phase when the parasite in the process of growth; *schizont*, which is the phase when the parasite in the process of breeding; and *gametocyte*, which is the phase when the parasite in the process of sex formation [8]. The process of image segmentation in this study using adaptive color segmentation to convert the original image color space based on RGB component (Red, Green, Blue) to HSV color space (Hue, Saturation, Value) and then thresholding against Saturation component, while classification process is performed using the decision tree

algorithm based on morphological feature of plasmodium which are the area ratio and eccentricity.

2. THEORITICAL BACKROUND

2.1 HSV Color Space

HSV is the color space that represents color as seen by the human eye. H is derived from the word "hue", S is derived from the word "saturation", and V is derived from the word "value". Hue refers to the color known by man, such as red and green. This property reflects the color that is catched by the human eye in respond to the light wavelength. Saturation states the level of color purity or how much white light mixed with the hue. Each pure color saturated 100% and contains no white light at all. In other words, a pure color mixed with white light has saturation between 0 and 100%. Value or sometimes called brightness states the intensity of the object reflection received by the eye. Intensity can be expressed as the change in the color from white to gray and finally reach the black, or which is known as gray level [9]. Illustration of color space transformation from RGB to HSV is shown in Figure 2.1.

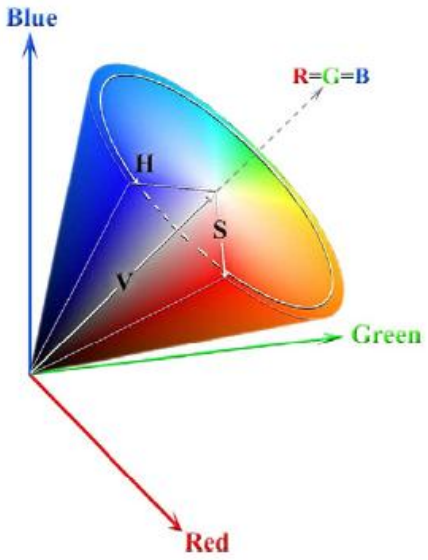

Figure 2.1 Illustration of color space transform from *RGB* to *HSV* [9]

The formulas below are using to get the value of *H, S, V* based on *R, G, and B* [10]:

$$r = \frac{R}{(R+G+B)}, g = \frac{G}{(R+G+B)}, b = \frac{B}{(R+G+B)} \quad (2.6)$$

$$V = \max(r, g, b) \quad (2.7)$$

$$H = \begin{cases} 0, if\ \ S = 0 \\ \dfrac{60*(g-b)}{S*V}, if\ \ V = r \\ 60*\left[2 + \dfrac{b-r}{S*V}\right], if\ \ V = g \\ 60*\left[4 + \dfrac{r-g}{S*V}\right], if\ \ V = b \end{cases} \quad (2.8)$$

$$H = H + 360, if\ \ H < 0 \quad (2.9)$$

2.2 Thresholding

Image thresholding become the focal point in the application of image segmentation for its intuitive properties and simplicity in implementation. Figure 2.2 shows the histogram of image intensity *f(x,y)* which consists of bright objects on the dark background, thus the object and background pixels are having intensity levels which are grouped into the dominant mode. Extracting the objects from the background is done by selecting the threshold T that divides these modes. Then an arbitrary point (x,y) for which f (x, y) ≥T called object point. The other points is called the background points. In other words, the thresholded image g (x, y) is defined as [11]:

$$g\ \ x, y\ = \begin{cases} 1, if\ \ f\ \ x, y\ \ \geq T \\ 0, if\ \ f\ \ x, y\ \ < T \end{cases} \quad (2.10)$$

Pixels are rated 1 related to the object while the pixels are given the value 0 related with the background. When T is constant, this approach is called global thresholding.

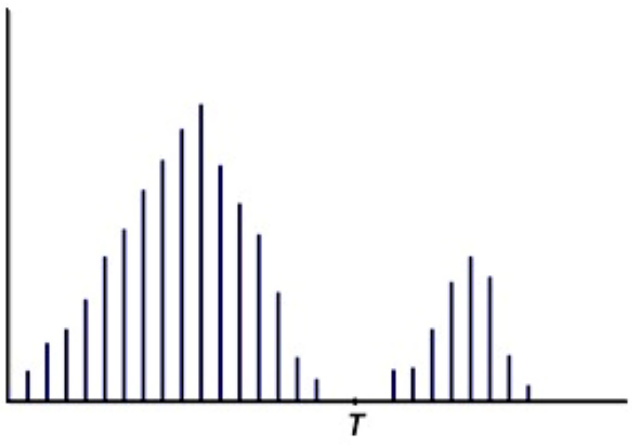

Figure 2.2. Threshold selecting using Visual histogram bimodal analysis

2.3 Morphological Feature Extraction

Morphological features which can be extracted from an object in the image are the area ratio and eccentricity. Area ratio is the ratio between number of pixels that make up an object and the image size. Eccentricity is the ratio between the distances of *foci ellipse* with major axis length [12]. Illustration of eccentricity calculation is shown in Figure 2.3.

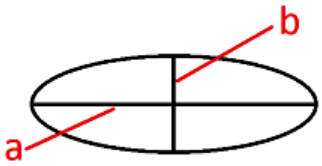

Figure 2.3 Illustration of eccentricity calculation [12]

The eccentricity is definied as [12]:

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2} \qquad (2.11)$$

e is the eccentricity value, *a* is the length of the major axis, and *b* is the length of the minor axis.

2.4 Decision Tree Based Classification

Decision tree based classification algorithm in this study using C4.5 method [14,15]. This algorithm starts with the formation of the root node, and then entropy value is calculated for all training datas on the node. Parameter with the maximum information gain is used as the node splitting into branches. Furthermore, if each node has not given the class label then entropy value calculation is repeated. However, if each node only gives one class label, then the node is used as a leaf node containing the decision [16]. Flowchart diagram of the decision tree based classification algorithm on this study are shown in Figure 2.4 below:

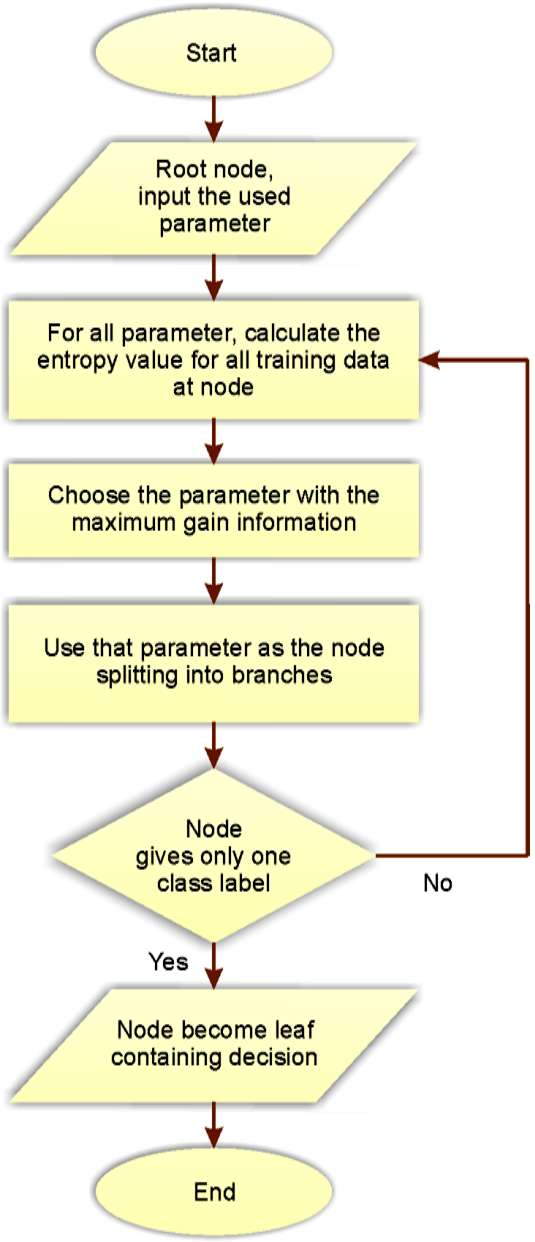

Figure 2.4. Flowchart of decision tree based classification algorithm [16]

III. METHOD

Procedure in this study included image acquisition, image quality enhancement, morphology feature extraction, and decision tree based classification. The flowchart of image processing is shown in figure 3.1.

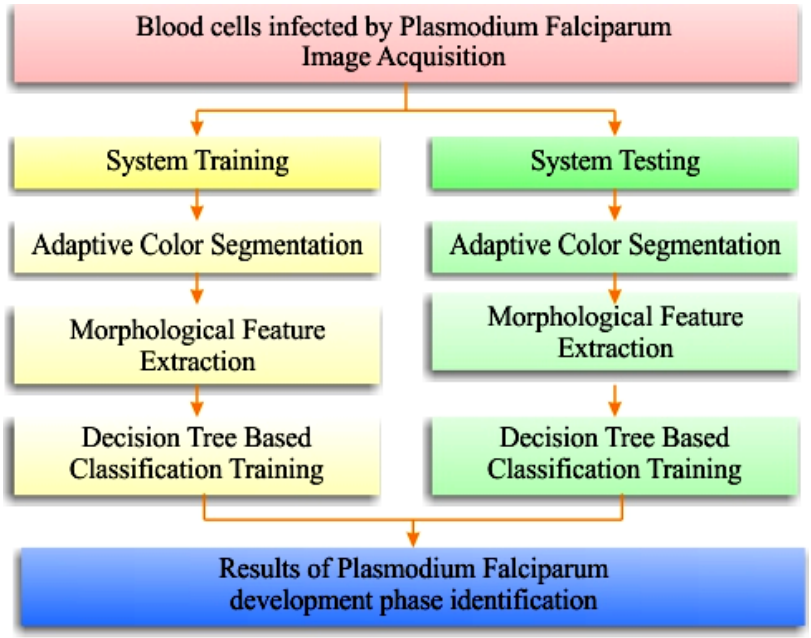

Figure 3.1. Flowchart of image processing

3.3.1 Image Acquisition

Image acquisition is done by capturing the image of a red blood cell preparation which is infected by plasmodium falciparum using a microscope with a magnification of 1000x and USB digital camera with 400 x 320 pixels resolution.

3.3.2 Adaptive Color Segmentation

Color segmentation in this study begins with the converting of the original image color space RGB (Red, Green, Blue) to HSV (Hue, Saturation, Value). After that, the thresholding is done against Saturation component.

3.3.3 Morphological Feature Extraction

Morphological parameters that are used to distinguish the phases of development of plasmodium falciparum in this study are the area ratio and eccentricity.

3.3.4 Decision Trees based Clasification

The development of P. falciparum consists of three phases which are trophozoite phase, shizont phase, and gametocyte phase. Trophozoite phase has morphological features smaller than the phase of schizont and gametocyte phase. Therefore, in this study, area ratio parameters are used to distinguish the size of trophozoite phase with other phases. While the eccentricity parameters used to distinguish between schizonts phases with gametocyte phase, in which the schizont phase has a spherical shape while gametocyte phase has an oval shape. The identification process of development phase of plasmodium falciparum in this study using decision tree based classification.

IV. RESULTS AND DISCUSSION

4.1 Image Acquisition

This study begins with the image acquisition process from red blood cells preparation infected with Plasmodium falciparum using a microscope and USB digital camera. The number of image used in this study were 155 images data, consist of 82 images data for training process and 73 images data for testing process.

4.2 Training System

The process of training systems in this study consist of three stages, which are adaptive color segmentation, morphology feature extraction, and training on decision tree based classification.

4.2.1 Adaptive Color Segmentation

Color segmentation on the training process is done to separate between objects (Plasmodium falciparum) and the background (red blood cells). Segmentation process begins with the converting of the original image color space based on RGB component (Red, Green, Blue) to HSV color space (Hue, Saturation, Value). Then *thresholding* process is done for saturation component resulting binary image of plasmodium falciparum. In the training process, this system used 82 images data of plasmodium falciparum, consisting of 5 gametocyte phases, 5 schizont phase, and 72 trophozoite phase. One of adaptive color segmentation process in the training system process is shown in Figure 4.1.

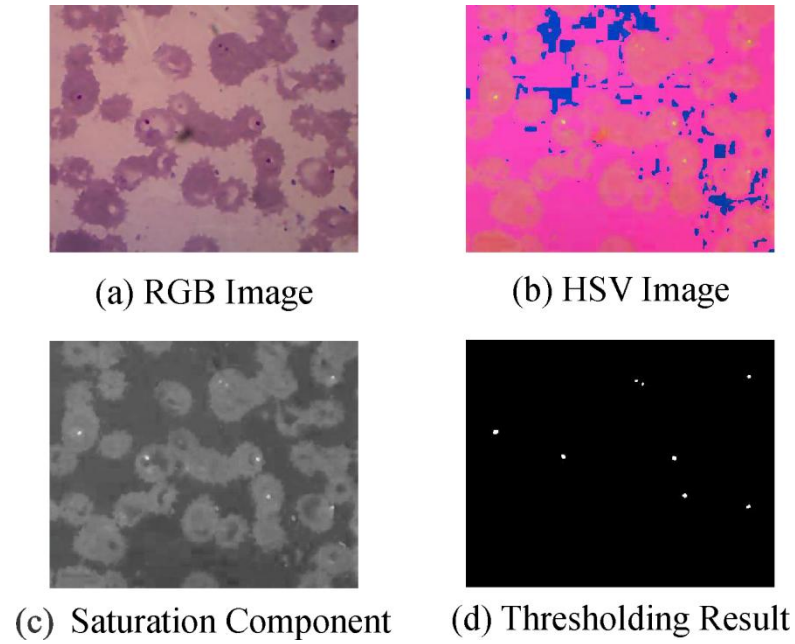

Figure 4.1. Adaptive color segmentation in system training process

4.2.2 Morphological Feature Extraction

Feature extraction in this study is based on morphological characteristics of plasmodium using area ratio and eccentricity parameters. Feature extraction results in Figure 4.2 (d) are shown in Table 4.1.

Table 4.1 Result of Morphological Feature Extraction of plasmodium falciparum

No.	*Area Ratio*	*Eccentricity*
1	*2,813E-04*	*6.135E-01*
2	*2,266E-04*	*6.744E-01*
3	*8,594E-05*	*6.847E-01*
4	*2,344E-04*	*6.658E-01*
5	*2,109E-04*	*4.539E-01*
6	*1,797E-04*	*4.758E-01*
7	*3,359E-04*	*7.524E-01*
8	*1,875E-04*	*6.577E-01*

4.2.3 Training on Decision Tree based Classification

In the training process of classification, feature extraction yield data were classified using decision tree algorithm with area ratio and eccentricity parameters to produce the output of classification rules. The results of the training process of decision tree based classification using area ratio and eccentricity parameters are shown in Figure 4.2.

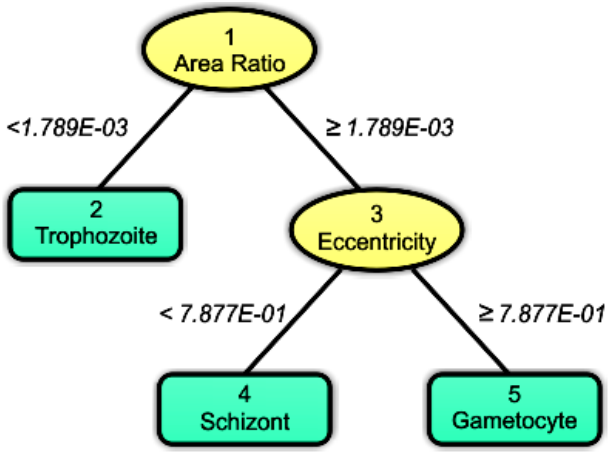

Figure 4.2. Decision tree of classification training process with the area ratio and eccentricity parameters

The generated rules from classification tree on Figure 4.2 are as follows:

1. If the area ratio < 1.789E-03 then node 2, whereas if the area ratio > = 1.789E-03 then node 3

2. Class = trophozoite

3. If the eccentricity < 7.877E-01 then the node 4, whereas if the eccentricity > = 7.877E-01 then the node 5

4. Class = Schizont

5. Class = Gametocyte

Confusion matrix generated by the decision tree is shown in Table 4.2.

Table 4.2. Confusion Matrix of training process with the area ratio and eccentricity parameters

		Predicted Class		
		Gametocyte	Schizont	Trophozoite
Actual Class	Gametocyte	5	0	0
	Schizont	0	5	0
	Trophozoite	0	0	73

Based on Table 4.2, it is clearly seen that all of Plasmodium falciparum entering each class correctly. Accuracy generated by classification training with the area ratio and eccentricity parameters is:

$$Accuracy = \frac{Number\ of\ right\ identified\ data}{Number\ of\ identified\ data} \times 100\%$$
$$= \frac{83}{83} \times 100\%$$
$$= 100\ \%$$

The rules generated by classifications are used as a rule in the testing process of development phase of plasmodium falciparum identification system.

4.3 System Testing

The testing process classification in this study using 73 images data of plasmodium falciparum. Then, all datas are classified according to the rules of the decision tree with area ratio and eccentricity parameters. Examples of test results decision tree based classification with area ratio and eccentricity parameters are shown in Table 4.3.

Table 4.3. Examples of test results decision tree based classification with area ratio and eccentricity parameters

Image data Number	Identification of *Plasmodium Falciparum* development phase	
	Developed system	Laboratory analysis
41	*Gametocyte*	*Gametocyte*
42	*not P. Falciparum*	*Trophozoite*
43	*not P. Falciparum*	*Trophozoite*
44	*Trophozoite*	*Trophozoite*
45	*Trophozoite*	*Trophozoite*
46	*Trophozoite*	*Trophozoite*
47	*Schizont*	*Schizont*
48	*Trophozoite*	*not P. Falciparum*
49	*Gametocyte*	*Gametocyte*
50	*Gametocyte*	*Gametocyte*
51	*Trophozoite*	*not P. Falciparum*
52	*Trophozoite*	*not P. Falciparum*

From the results of classification with the area ratio and eccentricity parameters, there are 9 objects which are not *plasmodium falciparum*, but identified as *plasmodium falciparum*. The accuracy of this classification testing is :

$$Accuracy = \frac{Number\ of\ right\ identified\ data}{Number\ of\ identified\ data} \times 100\%$$
$$= \frac{64}{73} \times 100\%$$
$$= 87.67\ \%$$

In general, the process of system testing in this study consisted of three stages: adaptive color segmentation, feature extraction morphology, and testing of decision tree based classification. The result of developed adaptive segmentation in this study could well segmented between objects (Plasmodium falciparum) and the background (red blood cells), although sometimes there are dust, dirt, or the rest of the painting that has same color with plasmodium falciparum, so the actual object which is not plasmodium falciparum is identified as plasmodium falciparum.

Morphological parameters for distinguish the size of the object are the area ratio parameter, also morphological parameter to distinguish the shape of the object are the eccentricity. These parameters are needed for the identification system of plasmodium falciparum in this study. Morphological parameters of the area ratio are used to distinguish trophozoite phase size which is smaller than schizont and gametocyte phase. While the morphological parameters of the eccentricity are used to distinguish between schizont phase form which hasthe spherical shape and gametocyte phase which has the oval shape.

The test results of identification system in this study using area ratio-eccentricity parameters showed the degree of accuracy, which is 87.67%.

V. CONCLUSION

It can be concluded that identification system of plasmodium falciparum development phase in preparations of red blood cells infected with malaria has been designed in this study. Image acquisition process is done microscopically using microscope and USB digital camera. The process of image segmentation using adaptive color segmentation, feature extraction process based on morphological parameters, while the process of classification using decision tree algorithm. The results accuracy of the system in identifying the plasmodium falciparum development phase is 87.67%.

ACKNOWLEDGMENT

The authors would like to thank Indonesia Endowment Fund for Education (LPDP); Parasitology laboratory of Faculty Medicine of Diponegoro University, Semarang, Indonesia; dr. Sri Hendratno, DAP&E, Sp.ParK; Prof. dr. Edi Dharmana, M.Sc., Ph.D., Sp.ParK; dr. Dian Puspita Dewi; and Rahmah Afrianti, A.Md. AK for many illuminating discussions on the red blood cell count aspects of this work.

REFERENCES

[1] Hirimutugoda, Y. M., and Wijayarathna, G. 2010. *Image Analysis System for Detection of Red Cell Disorders using Artificial Neural Networks.* Sri Lanka Journal of Bio-Medical Informatics 2010;1(1):35-42.

[2] Sio, S. W. S., Sun, W., Kumar, S., Bin, W. Z., Tan, S. S., Ong, S. H., Kikuchi, H., Oshima, Y., and Tan, K. S. W. 2007. *Malaria Count: An Image Analysis-based Program for the Accurate Determination of Parasitemia.* Journal of Microbiological Methods 68 pp. 11–18.

[3] Das, D. K., Ghosh, M., Pal, M., Maiti, A. K., and Chakraborty, C. 2013. *Machine Learning Approach for Automated Screening of Malaria Parasite using Light Microscopic Images.* Micron 45 pp. 97–106.

[4] Adi, K., Pujiyanto, S., Gernowo, R., Pamungkas, A., and Putranto, A. B. 2014. *Identification of Plasmodium Falciparum Phase in Red Blood Cells using Artificial Neural Networks.* International Journal of Applied Engineering Research (IJAER) ISSN 0973-4562 Volume 9, Number 23 (2014) pp. 13917-13924.

[5] Adi, K., Pujiyanto, S., Gernowo, R., Pamungkas, A., and Putranto, A. B. 2014. *Automatic Thresholding with Otsu's Method to Identify Plasmodium falciparum Phase in Malaria-infected Red Blood Cells.* The 4th International Conference on Theoretical and Applied Physics (ICTAP-2014) 16-17 October 2014, Denpasar-Bali, Indonesia.

[6] Savkare, S. S., and Narote, S. P. 2011. *Automatic Classification of Normal and Infected Blood Cells for Parasitemia Detection.* IJCSNS International Journal of Computer Science and Network Security, Vol.11 No.2, February 2011.

[7] Savkare, S. S., and Narote, S. P. 2012. *Automatic System for Classification of Erythrocytes Infected with Malaria and Identification of parasite's life Stage*. Procedia Technology 6 (2012) 405 – 410.

[8] Tilley, L., Dixon, M. W. A., and Kirk, K. 2011. *The Plasmodium Falciparum-Infected Red Blood Cell.* The International Journal of Biochemistry & Cell Biology 43 pp. 839–842.

[9] Ledley, R. S., Buas, M., and Golab, T. J. 1990. *Fundamentals of true-color image processing. Pattern recognition*. Proceedings of the 10th International Conference, June 16–21, 1990, Vol. 1. (pp. 791–795).

[10] Acharya, T. and Ray, A.K. 2005. *Image Processing Principles and Applications*. New Jersey: John Wiley & Sons, Inc.

[11] Gonzalez, R.C. and Woods R.E. 2004. *Digital Image Processing Second Edition.* New Jersey: Pearson Prentice Hall.

[12] Adi. K., Gernowo. R., Sugiharto. A., Firdausi. K. S., Pamungkas, A., and Putrato. A. B. 2013. *Tuberculosis (TB) Identification in The Ziehl-Neelsen Sputum Sample in NTSC Channel and Support Vector Machine (SVM) Classification*. International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2/ Issue 9.

[13] Adi, K., Gernowo, R., Sugiharto, A., Pamungkas, A., Putranto, A. B., and Mirnasari, N. 2013. *Autothresholding Segmentation for Tuberculosis Bacteria Identification in the Ziehl-Neelsen Sputum Sample*. The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013, ITS, May 15-16, 2013, Bali.

[14] Quinlan, J. R. 1986. *Induction of Decision Trees*. Machine Learning 1: 81-106.

[15] Quinlan, J. R. 1996. *Improved Use of Continous Attributes in C4.5*. Journal of Artificial Intelligence Research 4: 77-90.

[16] Tan, P. and Steinbach, M., Kumar, V. 2006. *Introduction to Data Mining*. New York: Pearson Education.