Red blood cell classification on thin blood smear images for malaria diagnosis

Red blood cell classification on thin blood smear images for malaria diagnosis

Budi Sunarko[1*], Djuniadi[1], Murk Bottema[2], Nur Iksan[1], Khakim A N Hudaya[1] and Muhammad S Hanif[1]

[1]Department of Electrical Engineering, Universitas Negeri Semarang, Indonesia
[2]Flinders University, Australia

[*]Corresponding author's e-mail: budi.sunarko@mail.unnes.ac.id

Abstract. Parasite detection is important for the diagnosis of many blood-borne diseases including malaria. As part of a program to develop a fast, accurate, and affordable automatic device for diagnosing malaria, a critical step is to automatically classify individual red blood cells in thin blood smear images. To automatically recognize malaria parasites in an image, this paper presents a red blood cell classification study for malaria diagnosis. To diagnose malaria, the threshold-based segmentation is implemented using the Otsu's method succeeded by the distance transform and statistical classifier. The methods are applied to red blood cell images obtained from Kaggle. These experimental results show that the classification recognizes malaria parasite with 94.60% accuracy, 96.20% specificity, and 93% sensitivity.

1. Introduction

Red blood cell classification is important for the diagnosis of blood-borne diseases such as malaria. In most cases, highly trained individual inspects samples.

Malaria is an endemic disease and major cause of mortality, especially in tropical regions. Globally, 3.2 billion people in 97 countries and territories are at risk of being infected with malaria and 1.2 billion are at high risk [1].

Clinically, many diseases generate similar symptoms. Typically, malaria produces flu-like symptoms around nine to 14 days after an infective Anopheles mosquito bite; however this can vary with different malaria species. If appropriate drugs are unavailable or the parasite has gained resistance to the drugs, the infection can progress rapidly and become life threatening. If left untreated, malaria can kill by infecting and destroying red blood cells, causing anaemia and by blocking capillaries that carry blood to the brain [2].

Malaria cannot be treated until it is diagnosed and currently, microscopy is the most commonly used technique to diagnose malaria. In malaria microscopy, two kinds of Giemsa-stained blood films, thin films and thick films, are recommended [3]. A well-prepared thin blood film consists of a single layer of red blood cells and leukocytes. In thin blood films, the morphology of red blood cells and parasites is fairly easy to see and counting the number of cells in a single field of view is feasible. However, in order to distinguish between low parasitaemia and the absence of malaria, a prohibitively large number of fields must be examined.

In general, Giemsa stain enhances differences between key components of infected red blood cells [3]. Parasites appear dark red and blue, the vacuole appears clear, the host red blood cells appear light red, and the pigment appears golden brown to black. Malaria parasites infect and develop in red blood

cells and as a result, the vacuole containing clear fluid lies inside the cytoplasm of the parasites [5]. Because the various components stain differently, malaria-infected erythrocyte parasites are distinguishable from normal red blood cells.

Detection of malaria requires trained microscopists. An accurate and efficient diagnosis is possible in ideal conditions by well-trained staff as is common in well-funded medical facilities in large Western cities where malaria seldom appears. In developing countries, especially in rural areas where malaria is most prevalent, the method is often inaccurate and inefficient. Potential weaknesses of the technique arise from (a) inaccuracy due to limited training, experience or fatigue resulting from excessive workloads, (b) long processing times for inexperienced microscopists who must frequently refer to the references, (c) considerable cost of training microscopists [6][7]. Mis-observation due to the lack of technique may potentially result in mis-diagnosis or delayed diagnosis possible leading to a more severe disease state or death.

Meanwhile, computational image analysis has the potential to provide fast and consistent estimates of parasitaemia. Since red blood cells and parasites are relatively easy to segment automatically in images of thin films, and since an automatic system may, in principle, be applied to large numbers of fields of view, estimates of depth of infection based on thin film is a reasonable option for computational image analysis approaches.

Several studies have reported on automatic systems for diagnosis of malaria. In an early study of automatic malaria diagnosis, morphological approaches were considered for identifying red blood cells and malaria parasites [8]. Red blood cells and parasites were detected by granulometric functions based on size features and regional maxima based on colour features. Later, Tek and colleagues reviewed work on computer-aided systems for estimating malaria parasitaemia based on thin blood film smears [9]. Some studies reported by Tek produced useful systems to estimate malaria parasitaemia. However, these studies were not compared to parasitaemia estimated by microscopists on thick film. At the same time, Tek et al. used probability density to determine stained cells and local area granulometry to estimate cell size. A modified K nearest neighbour (KNN) classifier was used to detect malaria-infected red blood cells on thin film [10]. However, the parasitaemia level was not presented.

In a more recent study [11], the Otsu method [12] was used to segment red blood cells from background. The Otsu method resulted in a binary image in which red blood cells were represented by a larger area than in the grayscale image. As a result, adjacent grayscale red blood cells were identified as an occluded binary red blood cells. To separate the occluded binary red blood cells, the distance transform [13]–[16] followed by the watershed transform [17] were applied. Subsequently, statistical features skewness, kurtosis, energy, and standard deviation of saturation histogram value were used as input to support vector machine to classify red blood cells as normal or infected. Finally, the percentage of parasitaemia based on thin film was reported. At the same time, another study showed that intensity histogram values corresponding to the unique colour of malaria parasites is one of the strongest features for the identification of parasites [18]. However, in these studies, the combination of features on colour and grayscale images was not studied.

The objective of this study is to extract features on colour and grayscale images for classifying red blood cells as healthy or infected.

2. Material and Methods

2.1. Data

Red blood cell images were provided by Kaggle hosting a repository of the images from the Malaria Screener. Slides of Giemsa-stained thin blood smear were collected from 150 parasitic and 50 healthy patients, and scanned at Chittagong Medical College Hospital, Bangladesh, by using a conventional light microscope coupled with a standard Android smartphone built-in camera running a mobile application developed by the Lister Hill National Center for Biomedical Communications, the National Library of Medicine. PNG cell images were captured for each microscopic field of view with a various resolution. The images were manually annotated by an expert slide reader at the Mahidol-Oxford

Tropical Medicine Research Unit in Bangkok, Thailand. A repository of the segmented parasitic images from the Malaria Screener were hosted in Kaggle.com. A total of 2000 cell images with equal instances of parasitic and healthy cells were involved for training system. Another dataset, a total of 2000 cell images consisting of 1000 parasitic and 1000 healthy cells were included in this study for testing. Examples of original color cell images are shown in figure 1.

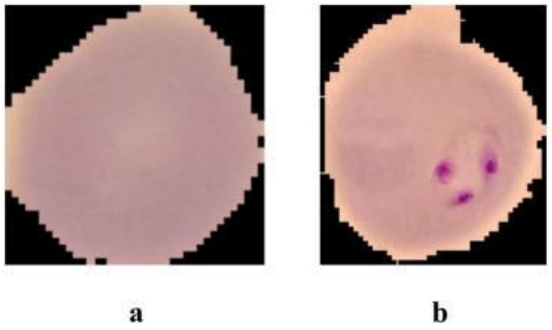

Figure 1. Original cell images. a) a healthy cell, b) an infected cell.

2.2. Features for identifying red blood cells as healthy or infected.

Classifying red blood cells as healthy or infected requires a process for automatically segmenting the parasites from background. The red blood cells were segmented by Otsu method [12]. As a preliminary step, 2000 of segmented colour red blood cell images were involved to extract a statistical feature, skewness. Skewness was a potential feature for classifying red blood cells as infected or healthy cells [4]. Of these, 1000 were healthy and 1000 were infected with malaria parasites. The intensity values for red blood cell were extracted from the grayscale images by using the footprint as a template. For the footprint, the skewness of the distribution of grayscale intensity values were extracted.

Also, the colour dominant of the colour intensity values (RGB) was also extracted. Visually, an infected cell image consists of three colour dominants representing three objects: background, cell, and parasites (if present). However, due to imperfect process of image segmentation, a cell body may have two different dominant colours: border area and cells. Thus, based on the grayscale values, a K-means clustering method [19] was employed to cluster the image pixels into four groups nominally representing the parasite, the cell body, border area, and the black (background beyond the cell body) (figure 2). The cluster C_3 and C_0 were suspected as a background and a border line, respectively. Meanwhile, the cluster C_2 with the second lowest mean value was assumed to be the one corresponding to the parasites as this cluster is reliably the darkest within the cell body (C_2) if parasites are present. The average of RGB values of every pixels in each cluster was calculated. The difference in the average values between the cluster 2 and cluster 1 was calculated, equation (1). For ease of exposition, the term “distance colour” will be used to refer to the difference. For red blood cells with parasites present, the distance colour is expected to be noticeably farther than the distance colour for red blood cells with no parasites. Accordingly, the distance colour was used along with the other features to classify red blood cells.

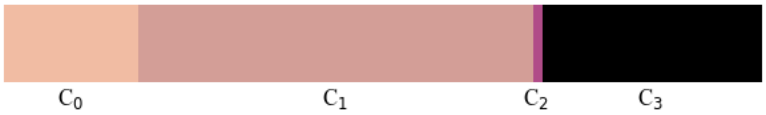

Figure 2. Colour map resulting from colour clustering

distance colour = $|mC_1 - mC_2|$ (1)

mC3 = the average of RGB values from the cluster 3 (the highest number)
mC_2 = the number of RGB values from the cluster 2.

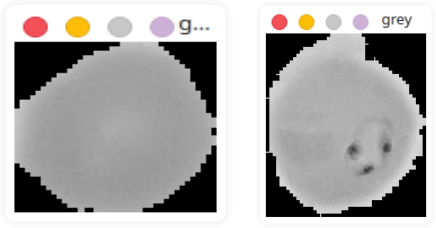

Figure 3. Grayscale images from Figure 1.

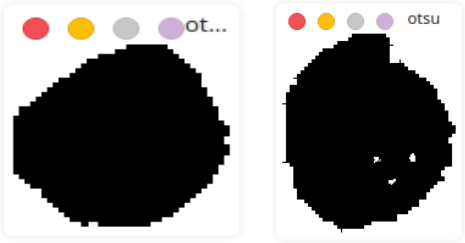

Figure 4. Parasite footprints of figure 3. White dots indicate the suspected parasites.

In addition to skewness and distance colour, the difference in angle between the highest and the lowest value of grayscale images (figure 5) was proposed. For ease of exposition, the term "distance angle" will be used to refer to the difference.

Step for picking parasites up from a cell body, the original colour red blood cell images were converted to grayscale images. Otsu method [12] was used to segment the cells from background and an operation of inversion was applied to the segmented images. Meanwhile, for obtaining cell mask, the grayscale images (*grayim*) were also segmented by binary threshold (equation 1) [20] for resulting binary images (*binim*) consisting of background and cell mask. For the next stage, the segmented images and the cell masks were operated by an operation of bitwise AND for resulting parasite footprints (figure 4).

$$binim(x,y) = \begin{cases} 0 \; if \; grayim(x,y) = 0 \\ 1 \qquad\qquad otherwise \end{cases} \quad (2)$$

After parasite footprints were obtained, Suzuki's Contour Tree algorithm [21] was used to detect extreme edge between black region and parasite footprints of the binary images. From those algorithm, two main regions: black region (i.e., zero regions) and parasite footprints were separated or segmented. The centroid of each parasite footprint was determined and the pixels around the parasite were scanned. The grayscale value of the centroid and that of the around pixels were recorded as j and k[i], i=0, 1, 2, …, n. Then, an angle colour (α) could be obtained by equation (3).

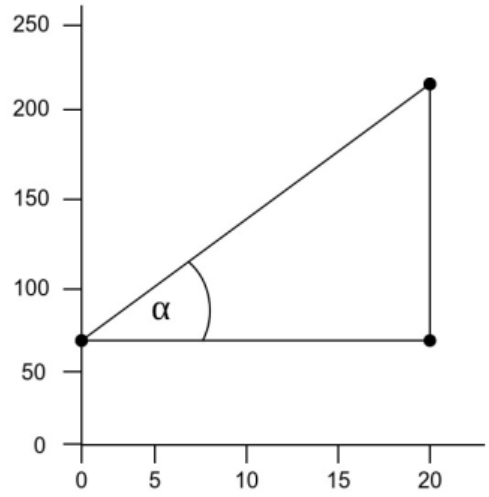

Figure 5. Angle degree.

$$\alpha = \arctan\left(\frac{|j-\max(k)|}{n}\right) \quad (3)$$

n is the number of pixels around the centroid.

The obtained features: skewness, distance colour, and angle colour (figure 6) were processed by the support vector machine (SVM) with radial basis function to classify red blood cells as healthy or infected.

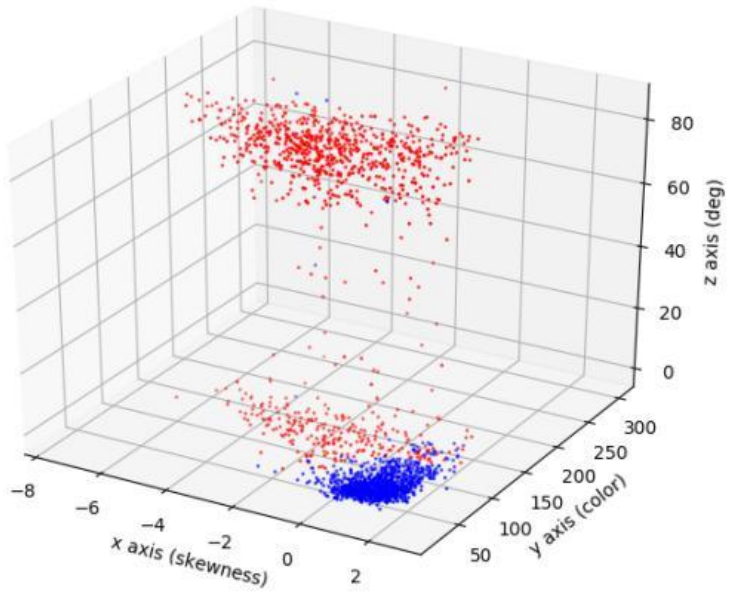

Figure 6. Feature values of red blood cells. Blue and red dots are representing healthy and infected cells, respectively.

2.3. Analysis Methods

The performance of the method for red blood cell classification was analysed in terms of accuracy, sensitivity, and specificity. Accuracy measures the total proportion of correct assignments (equation 4) while sensitivity (equation 5) and specificity (equation 6) measure the proportion of actual infected cells detected and that of actual healthy cells detected, respectively. These values are expressed in terms of true positive detection (TP), false positive detection (FP), true negative detection (TN), and false negative detection (FN).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

3. Experimental Results and Discussion

The results of comparing an expert reader to the outputs of this algorithms are summarized in table 1. Of the 2000 red blood cell images, 1892 were correctly classified, providing an accuracy of 94.60%. The classification performance of the proposed algorithms outperforms that of Tek's study [10].

Table 1. Confusion matrix for classification of red blood cells. The classification accuracy is 94.60%. The specificity and sensitivity are 96.20% and 93%.

		Predicted	
		Positive	Negative
Actual Class	Positive	930	70
	Negative	38	962

Although many red blood cells were correctly classified, there are some plausible sources of inaccuracy that could be addressed in future work. A number of infected red blood cells were detected as healthy cells (FN). It is probably caused by schizonts (parasites on early stage) which are similar colour to red blood cells (figure 3a). Accordingly, this algorithm is not applicable to detect the red blood cells as infected. On the other hand, accumulations of stain on healthy red blood cells due to faults in the staining process (figure 3b) resemble parasites in colour. Accordingly, the cells might be detected as infected (FP).

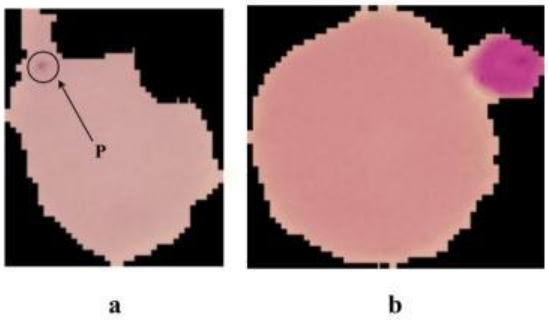

Figure 7. a is FP, b is FN.

4. Conclusion

The features and algorithm presented in the paper may be used to classify red blood cell, healthy or infected. These experimental results show that the accuracy is 94.60%. Despite the fact that plausible diagnosis of malaria parasite using these features, a final system based on this algorithm alone may not be practical for detecting parasites in schizont stage. Accordingly, a future study is needed to improve the quality of segmentation.

References

[1] W. H. Organization, "World Malaria Report 2014. Geneva: World Health Organization; 2014," *Fecha Consult.*, vol. 23, p. 238, 2017

[2] M. Aikawa, M. Iseki, J. W. Barnwell, D. Taylor, M. M. Oo, and R. J. Howard, "The pathology of human cerebral malaria," *Am. J. Trop. Med. Hyg.*, vol. 43, no. 2_Part_2, pp. 30–37, 1990

[3] W. H. Organization, "Basic malaria microscopy: Part I. Learner's guide.," *Basic Malar. Microsc. Part I. Learn. Guid.*, no. Ed. 2, 2010

[4] B. Sunarko, S. Williams, W. R. Prescott, S. M. Byker, and M. J. Bottema, "Correlation between automatic detection of malaria on thin film and experts' parasitaemia scores," in *AIP Conference Proceedings*, 2017, vol. 1818

[5] A. R. Dluzewski, G. H. Mitchell, P. R. Fryer, S. Griffiths, R. J. Wilson, and W. B. Gratzer, "Origins of the parasitophorous vacuole membrane of the malaria parasite, Plasmodium falciparum, in human red blood cells," *J Cell Sci*, vol. 102, no. 3, pp. 527–532, 1992

[6] C. C. J. Carpenter, G. W. Pearson, V. S. Mitchell, and S. C. Oaks Jr, *Malaria: obstacles and opportunities*. National Academies Press, 1991

[7] N. E. Ross, C. J. Pritchard, D. M. Rubin, and A. G. Duse, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Med. Biol. Eng. Comput.*, vol. 44, no. 5, pp. 427–436, 2006

[8] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Automatic thresholding of infected blood images using granulometry and regional extrema," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, 2000, vol. 3, pp. 441–444

[9] F. B. Tek, A. G. Dempster, and I. Kale, "Computer vision for microscopy diagnosis of malaria," *Malar. J.*, vol. 8, no. 1, p. 153, 2009

[10] F. B. Tek, A. G. Dempster, and I. Kale, "Parasite detection and identification for automated thin blood film malaria diagnosis," *Comput. Vis. image Underst.*, vol. 114, no. 1, pp. 21–32, 2010

[11] S. S. Savkare and S. P. Narote, "Automatic detection of malaria parasites for estimating parasitemia," *Int. J. Comput. Sci. Secur.*, vol. 5, no. 3, p. 310, 2011

[12] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979

[13] P.-E. Danielson, "A new shape factor," *Comput. Graph. Image Process.*, vol. 7, no. 2, pp. 292–299, 1978

[14] G. Borgefors, "Distance transformations in digital images," *Comput. vision, Graph. image*

Process., vol. 34, no. 3, pp. 344–371, 1986
[15] C. T. Huang and O. R. Mitchell, "A Euclidean distance transform using grayscale morphology decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 4, pp. 443–448, 1994
[16] L. Binghan, F. Xiuduan, W. Weizhi, and Z. Zhiyong, "Automatic separation of overlapping objects," in *Proceedings of the 4th World Congress on Intelligent Control and Automation (Cat. No. 02EX527)*, 2002, vol. 4, pp. 2901–2905
[17] H. Digabel and C. Lantuéjoul, "Iterative algorithms," in *Proc. 2nd European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, 1978, vol. 19, no. 7, p. 8
[18] Y. Purwar, S. L. Shah, G. Clarke, A. Almugairi, and A. Muehlenbachs, "Automated and unsupervised detection of malarial parasites in microscopic images," *Malar. J.*, vol. 10, no. 1, p. 364, 2011
[19] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: data mining, inference and prediction," *Math. Intell.*, vol. 27, no. 2, pp. 83–85, 2005
[20] R. Laganiere, *OpenCV 2 Computer Vision Application Programming Cookbook*, 1st ed. Birmingham: Packt Publishing, 2011
[21] S. Suzuki, "Topological structural analysis of digitized binary images by border following," *Comput. vision, Graph. image Process.*, vol. 30, no. 1, pp. 32–46, 1985