Malaria Parasite Detection in Giemsa–Stained Blood Cell Images

Leila Malihi[1], Karim Ansari-Asl[1]
[1]Electrical Department, Engineering Faculty,
ShahidChamram University, Ahvaz, Iran.
Email: L-malihi@mscstu.scu.ac.ir
Email: karim.ansari@scu.ac.ir

Abdolamir Behbahani
School of Public Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.
Email: behbahani-a@ajums.ac.ir

***Abstract*—this research represents a method to detect malaria parasite in blood samples stained with giemsa. In order to increase the accuracy of detecting, at the first step, the red blood cell mask is extracted. It is due to the fact that most of malaria parasites exist in red blood cells. Then, stained elements of blood such as red blood cells, parasites and white blood cells are extracted. At the next step, red blood cell mask is located on the extracted stained elements to separate the possible parasites. Finally, color histogram, granulometry, gradient and flat texture features are extracted and used as classifier inputs. Here, five classifiers were used: support vector machines (SVM), nearest mean (NM), K nearest neighbors (KNN), 1-NN, and Fisher. In this research K nearest neighbors classifier had the best accuracy, which was 91%.**

Keywords—Malaria diagnosis; Blood cell image; K nearest neighbors rule; Area granulometry; Fisher's linear discriminator; Nearest Mean.

I. INTRODUCTION

Today, one of the most significant elements in diagnosing diseases is the appearance of the living tissue. Since the visual diagnosing in diseases is a difficult and time consuming way, the automatic diagnosing makes a significant help in saving the time and reducing the possible errors.

Malaria is a serious infectious disease caused by a peripheral blood parasite of the genus Plasmodium [1]. Every minute, all over the world, a child under the age of five dies because of malaria. It is estimated that every day 1,500 children die. About 216 million people got malaria in 2010 [2]. Malaria is the most important parasitic disease in Iran which has a long outbreak history. Malaria is transmitted by the infected female Anopheles mosquitoes which carry Plasmodium sporozoites in their salivary glands. The genus Plasmodium has four species that can cause human infection: falciparum, vivax, ovale, and malariae. The complete life cycle has separate development stages in the human and mosquito body. When an infected mosquito feeds on a person's blood, the sporozoites enter the blood stream and move to the liver where they multiply asexually for a period. Then they produce merozoites to enter the peripheral blood stream again to invade Red Blood Cells (RBCs) . Parasite grows in RBC till it becomes a full grown up, in order to pour the extra merozoites into blood, then, it pierces the cell—about 6-24 parasites come out from each RBC [3,4].The malaria disease happens due to one of the four-type Plasmodium parasite: falciparum, vivax, ovale, and malariae. Staining blood slides with giemsa, while working with microscope, is used to detect malaria parasite. Giemsa stain is used to differentiate nucleus and cytoplasm of parasites, morphology of platelets, red blood cells and white blood cells. Giemsa staining solution stains up nucleic acids and, therefore, parasites, white blood cells, and platelets, which contain DNA, are highlighted in a dark purple color. Red blood cells are usually colored in slight pink colors [5].

Unfortunately, visual detection is time consuming and causes great differences among microscopists. Hence in new remedies of this disease, automatic detection of malaria by using giemsa stained blood is of significance importance. Ruberto *et al.* [6] used the hue and saturation components from the HSV color space to detect the parasite regions. They assumed a paraboloid model for the non-uniform illumination of the scanned images, which is not always the case. Ross *et al.* [7] have proposed a histogram based thresholding method to detect RBCs and parasites but their technique is heavily dependent on image quality and fails when the histogram does not have distinct valleys.

Automated image analysis-based software "MalariaCount" for parasitemia determination, i.e. for quantitative evaluation of the level of parasites in the blood, has been described in Weiling Sun *et al.* [8]. The presented system is based on the detection of edges representing cell and parasite boundaries. Their proposed technique includes preprocessing, edge detection, edge linking, clump splitting, and parasite detection steps. The preprocessing of the image, which involves the enhancement of the image contrast via adaptive histogram equalization, is followed by an edge detection, where a pixel is determined to belong to the boundary edge of the red blood

978-1-4673-6184-2/13/$31.00 ©2013 IEEE

cells if a defined edge correlation coefficient exceeds an empirically determined threshold. The resultant edge contours are linked together through their terminal points to form closed boundaries. The system requires well-stained and well-separated cells in order to provide accurate result. Moreover, artifacts, 'holes' inside red blood cells and noise can lead to a false interpretation of a red blood cell. Their software is not intended for studies involving patient samples [8].

The paper by Diaz *et al.* [9] evaluates a color segmentation technique, based on standard supervised classification algorithms, for separation of pixels into three different classes: parasite, red blood cell and background. The article presents a simple method for red blood cell and parasite detection with no classification of parasites. Their approach is based on a classification process that finds boundaries that optimally separate a given color space. No details on the filtering process performed to separate the relevant objects of interest are given. The system assumes constant color tone in the input images, since only luminance differences are corrected. To classify the stained pixels as parasite or non-parasite, they used a distance weighted K-nearest neighbors classifier.They also used four discriminative features – color histogram, Hu moments, relative shape measurements vector, and color auto correlogram. The relative shape measurements vector is formed of simple measurements representing the object shape. According to the results of the study, the most successful feature to classify the stained objects as parasite/non-parasite was the combination of correlogram, Hu moments and relative shape measurements [10].

This research deals with a method to detect malaria parasite automatically in images which have been stained with giemsa. Using the red blood cell mask is the significant advantage of this method that causes processing done just on the part of image containing red blood cells. Since most of malaria parasites exist in red blood cells, using this mask increases the accuracy of the method. In the next stage, the stained elements of the image (including parasite, white blood cell, platelets and some exterior elements) were extracted. Then, the mask was collocated with stained elements. This method doubled the ability of detecting the parasite. After these stages, to increase the accuracy of detecting, classification and extraction of features were done. Despite its simplicity, the proposed method has a high accuracy in malaria parasite detection.

This research includes the following sections. The second section describes the proposed method; the third section indicates the obtained results of the method and forth section shows the conclusion.

II. METHODOLOGY

The steps of the algorithm were described briefly in figure 1.

A.Pre-Processing

Green component of the true original color is primarily used by the detection system, since it has the least noise [6]; and the parasites, which stain a purple color, are most visible. At first, in order to omit series of undesired elements, median filter is used. After this stage, to eliminate the effect of lack of uniform illumination in preparing slides, image illumination was corrected. In the usage of thresholding for classification, the necessity of this stage is felt. Using closing morphology operator and structural element about 1.5 times of the red blood cell size, the illumination of the image was estimated and then by subtracting it from the main image, the illumination of the image was corrected [11]. Figure 2 shows the result of pre-processing action.

B.Extracting Red Blood Cell Mask

Most of the malaria parasites exist in red blood cells, thus if red blood cells are separated from other elements of blood, the accuracy and speed of detecting malaria parasites is enhanced significantly. To this end, we made a binary image by using Otsu's thresholding. It is worth mentioning that in Otsu's method the amount of threshold is obtained from the minimum amount of inner variance of the given weight, and according to primary hypothesis it acts on grayscale images. This method, especially in blood scope samples, is a good solution to access binary images without losing data. To have a perfect binary image and suit the early tissue of the object, it is necessary to optimum accurateness of the binary image with the sum of the two edge detection images from Canny's method and the binary image from Otsu's method. Then the parts which are not filled, because of the lack of accurateness in thresholding, will be filled again. Since the finding red blood cells in the image is the aim of this method, undesired elements which were smaller than red blood cells are deleted by a morphology operation. For doing this, a disc-shape structural element is used. The result of this action on image is represented in figure 3.

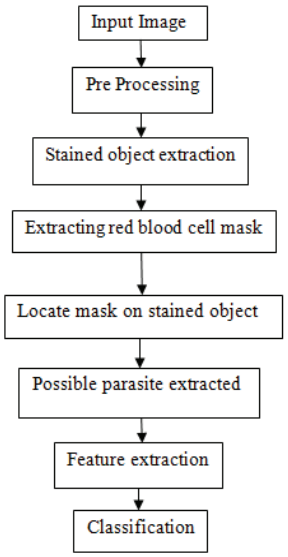
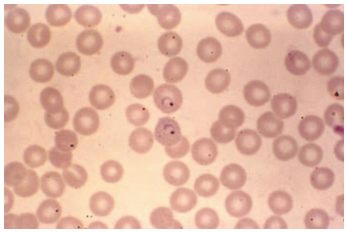


Fig. 1. Flowchart describing the steps of the proposed algorithm.

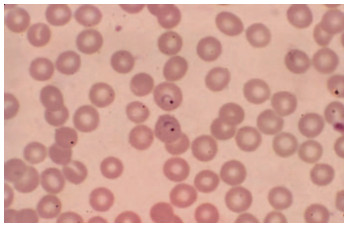
C.Stained Objects extraction

In images stained by giemsa, red blood cells are seen in slight pink colors, white blood cells, parasites and platelets, are seen in a darker color. According to this rule, the darker elements could be separated from the whole image and the other processes will be done on them. In order to separate white blood cells, parasites and platelets from the other elements of the image, the following way has been done. Otsu's method was applied on Multiplication of the edge detected and original images. It is worth noting that this method was done on about 300 images and in all of them it was able to extract successfully stained elements from the image. Figure 4 shows the result of this action.

By locating red blood cells mask on extracted stained elements from the image, a series of stained elements omitted and further processes are done on the remained elements. Figure 5 shows the result of this action. After extracting some elements from the image, which are suspected of being parasites, the features of the elements are extracted in order to detect parasites.

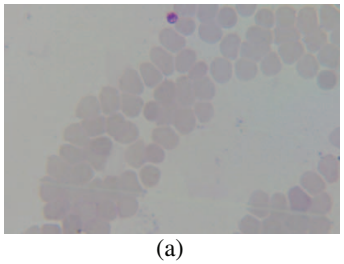


(a)

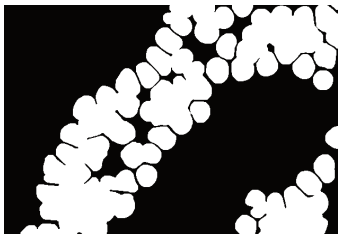


(b)

Fig. 2.The result of pre-processing action: (a) original image, (b) The result of pre-processing action on original image.

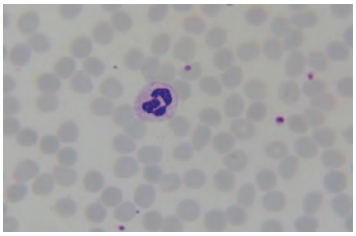


(a)



(b)

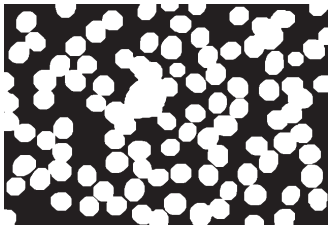
Fig. 3.Extracting red blood cell mask: (a) original image, (b) the produced mask.



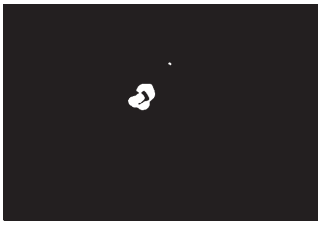
(a)

(b)

Fig. 4.Extraction stained objects: (a) original image, (b) extracting the stained objects from the image (a).



(a)



(b)

Fig. 5.Locating red blood cells on Fig. 4: (a) red blood cells mask Fig. 4(a), (b) the result of locating red blood cells on fig. 4(b).

D.Feature Extraction

Four features are used:

a) Gradient: Gradient is the selection tool for finding power and location of the edge at the place(x, y) in the image (f) which is defined*:*

$$Grad\,(f(x,y)) = (\partial f(x,y)/\partial x\ ,\ \partial f(x,y)/\partial y\ ) \quad (1)$$

$$Grad\,\big|\,(f(x,y))\,\big| = (\ (\partial f(x,y)/\partial x)^2 + (\partial f(x,y)/\partial y\ )^2\ )^{0.5} \quad (2)$$

Since the gradient is not dependent in general on values of pixels in a special area, it is not necessary to correct illumination in this case. The best detection power is obtained by computing the variance of histogram gradient of the image [12].

b) Flat texture: Flat texture is determined via computing the difference between the original image and the filtered image using median filter [13]. Flat texture image I_{FT} is computed using the following formula, where r is the size of the median operator window and $I_E(x,y)$ is the original image after pre-processing.

$$I_{FT}(x,\ y) = I_E(x,\ y) - Median\ (\{I_E(x+\mu,\ y+\epsilon);\ \mu,\ \epsilon = -r \ldots r\}) \quad (3)$$

The best performance for variance histogram of the above feature and r is between 15 and 25 [11].

c) Color Histogram: The histogram is a simple descriptor. If image I is quantised to have N distinct colors, H is the occurrences number of the color c_i:

$$H\,(c_i) = \left\|I_{ci}\right\| \quad c_i \epsilon\ C = \{c_1,\ldots,c_N\} \quad (4)$$

It is better to normalize the color before using histogram, to remove any differences if there is any because of the type of staining with giemsa. Here, the gray world normalization method is used [14].

d) Area Granulometry: In mathematical morphology, granulometry is an approach to compute a size distribution of grains in binary images, using a series of morphological opening operations. It has been used as a feature in many pattern discrimination applications [15,16]. However, area granulometry is better than granulometry (by structuring elements) for the parasite detection task since it is not possible to generalize the morphological structure of the stained objects by fixed geometric shapes. The area granulometry for a grey level image is computed as follows:

$$G_A(X) = \sum_{P\in X} \gamma_{\mu_i}^{a}(X) - \sum_{P\in X} \gamma_{\mu_{i-1}}^{a}(X) \quad (5)$$

Where $\gamma_{\mu_i}^{a}(X)$ is morphological area opening of image X with area threshold $\mu_i \epsilon A = \{\mu_1\ \ldots.\mu_n\}$ [17].

D.Classification

Five classifiers K-NN, 1-NN, NM, SVM and Fisher linear discriminator have been used in classification step.

a)The K nearest neighbor classifier(KNN): To find the class related to F feature vector,at first, the class related to K nearest training data vector of vector F (according to Euclidean standard distance) is considered. Then, the class which more vectors are related to, is determined as the class related to vector F [18].

b)The Nearest Mean Classifier(NM): The method in such a classifier is that first the mean of each class is computed separately; then to find the class of an unknown vector like F, its Euclidean distance with respect to the mean of each class is computed separately. Finally, the class with lesser distance is assigned to F [19].

c)Fisher's Linear Discriminator: This discriminator is based on the making the images of the classes. Suppose there are two classes whose data are two-dimensional. Fisher's discriminator in two-dimensional space seeks a line on which making image of the data of the two classes is more discriminate [20].

d)Support Vector Machine (SVM): In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models that used for classification and regression analyses. The basic SVM takes a set of input data and estimates, for each given input, which of two possible classes forms the output; thus it is a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, SVM training algorithm builds a model that assigns new examples into one category or the other. SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and determined to be belong to a category based on which side of the gap they fall on.In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using a non-linear kernel, implicitly mapping their inputs into high-dimensional feature spaces [21].

III. EXPERIMENTS AND RESULTS

All of the images prepared in size of 480×640. In order to train classifier, we used 363 images. Images consist of four types of parasite: falciparum, vivax, ovale, and malariae. To detect parasites, leave-one-out method with classifier SVM, 1-NN, NM, KNN and Fisher's linear discriminator is used. To compare the result these standard measures are used:

$$ACC = (TP+TN) / (TP+FN+FP+TN) \quad (6)$$

$$Se=TP / (TP+FN) \quad (7)$$

$$SP=TN / (TN+FP) \quad (8)$$

$$Precision=TP/ (TP+FP) \quad (9)$$

- TP: the number of parasites which has been detected parasitic correctly by classifier
- FP: the number of parasites which has been detected parasitic wrongly by classifier
- TN: the number of non-parasites which has been detected non-parasitic correctly by classifier
- FN: the number of non-parasites which has been detected non-parasitic wrongly by classifier

Table 1 shows the best performance for KNN classifier then for SVM classifier.

In table 2, by using KNN classifier with proposed algorithm, TP, TN, FP, FN, SE, SP , and precision were calculated and SE, SP paper method compared with the Tek [10].

TABLE I. THE RESULTS OF USING ALGORITHM ON 5 CLASSIFIRES

ACC	classifier				
	SVM	***K-nn***	***1-nn***	***NM***	***Fisher***
The research method	90%	91%	87%	70%	84%

TABLE II. THE RESULT OF DETECTING PARASITES AND NON-PARASITES BY THE ALGORITHM USING KNN CLASSIFIRE AND COMPARING IT WITH PREVIOUS WORK

Method	standard						
	TP	***TN***	***FP***	***FN***	***SE***	***SP***	***precision***
The research method	115	210	28	10	80%	95.5%	90%
Tek[10]	-	-	-	-	72.4%	97.6%	-

Prediction accuracy reached 88.77% in an essay published by Das[22] and Kumarasamy[23] parasite detection accuracy reached 86% which proves the usefulness of the above algorithm.

IV. CONCLUSION

In this research, a method to detect malaria parasite in blood samples stained with giemsa is presented. At the first step, the per-processing of images is done. In this stage,the illumination of images are corrected and noises are deleted. Then, by using morphology methods, red blood cell mask and stained elements of the image are extracted. In the next step, by locating red blood cell mask on the stained elements, possible parasites from the other elements in the image are separated. Extracting the features and detecting the parasite with classifier are done in the next steps. The results showed that among five studied classifiers (SVM, 1-NN, NM, KNN and Fisher), the KNN outperforms others.

V. ACKNOWLEDGMENTS

We would like to express our gratitude to Boray Tek for his sincere assistance in providing the experimental data; moreover, we appreciate Mehdi Maboudi for his guidance and help during the research.

REFERENCES

[1] E. Korenromp, J. Miller, B. Nahlen, T. Wardlaw, and M.Young, "World malaria report 2005," World Health Organization, Geneva , Tech. Rep, 2005.

[2] UNICEF, "World Malaria Report,"Technical Report, WMR and UNICEF.http://www.unicef.org/health/drcongo_62290.html2012

[3] "Laboratory identification of parasites of public health concern," Centers for Disease Control Prevention, National Center for Infectious Diseases, Division of Parasitic Diseases, Date Accessed, June2006. [Online]. Available:http://www.dpd.cdc.gov/dpdx/HTML/Malaria.htm.

[4] G. Coatney, W. Collins, M. Warren, and P. Contacos, The Primate Malarias.U.S. Department of Health, Education and Welfare, 1971.

[5] J.K. Mui, K.S. Fu,"Automated classification of nucleated blood cells using a binary tree classifier,"IEEE Trans. Pattern Analysis and Machine Intelligence ,vol. 2(5),pp.429-443, 1980.

[6] C.D. Ruberto, A. Dempster, S. Khan, andB.Jarra, " Analysis ofinfected blood cell images using morphological operators," ImageVis. Comput,vol.20, pp.133–146 ,2002.

[7] N.E. Ross, C.J. Pritchard, D.M. Rubin, andA.G. Dusé, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," Med. Biol. Eng. Comput,vol.44,pp.427–436 2006.

[8] W.S.S Sio,"MalariaCount: An image analysis-based program for the accurate determination of parasitemia,"Journal of Microbiological Methods, vol 2(2), pp.13–16, 2006.

[9] G.Diaz , F.Gonzalez, E. Romero, "Infected Cell Identification in Thin Blood Images Based on Color Pixel Classification: Comparison and Analysis," Lecture Notes in Computer Science, vol. 4756, pp. 812-821, 2007.

[10] F.Tek , A.Dempster, I.Kale,"Parasite detection and identificationfor automated thin blood film malaria diagnosis. Computer Vision and Image Understanding," vol. 114, pp.21–32, 2010.

[11] Vit springl, Automatic Malaria Diagnosis through Microscopy Imaging,Diploma Thesis,Czech technical univercity, Prague2009.

[12] TM. Korn, GA. Korn, "Mathematical Handbook for Scientists and Engineers", [Internet], Dover Publications, New York,2000.

[13] K.Rodenacker, E. Bengtsson, "A feature set for cytometry on digitized microscopic images,"Analytical Cellular Pathology vol. 25, pp. 1-36,2003.

[14] F.B. Tek, Computerised Diagnosis of Malaria, PhD Thesis, University of Westminster, London, UK, 2007.

[15] X.Tang, K.Stewart,L.vincent,H.Huang,M.Marra,S.Gallager,and C. davis, "Automatic plankton image recognition," Artificial Intelligence Review, vol. 12,pp. 177-99, 1998.

[16] P.F. Yang, P.maragos, "Morphological systemsfor character image processing and recognition," in Proc. IEEE Int. Conf.on Acoustics, Speech, and Signal Processing, Minneapolis, USA, 1993.

[17] K.N.R.M. Rao, A. Dempster, "Area-granulometry: an improved estimator of sizedistribution of image objects", IEE ElectronLett. Vol. 37 ,pp.950-951,2001.

[18] S. Theodoridis, K. Koutroumbas, "Pattern recognition and neural networks," Machine Learning and Its Applications, vol. pp. 169-195, 2001.

[19] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," Journal of neural engineering, vol. 4, pp. R1, 2007.

[20] A. Shashua, " On the relationship between the support vector machine for classification and sparsified Fisher's linear discriminant," Neural Processing Letters,vol. 9, pp. 129-139, 1999.

[21] C.Corinna, V.Vladimir,"Support-Vector Networks", Machine Learning, vol. 20,pp.273-297, 1995. http://www.springerlink.com/ content/ k238jx0 4hm87j80g/.

[22] D. Das, M.Ghosh, and C.Chakraborty, " Probabilistic Prediction of Malaria using Morphological and TexturalInformation," International Conference on Image Information Processing (ICIIP)), India, Bengal,2011.

[23] S.K. Kumarasamy , S.H. Ong , K.S.W. Tan, "Robust contour reconstruction of red blood cells and parasites in the automated identification of the stages of malarial infection" Machine Vision and Applications springer,vol. 22(1), PP.461-469,2011.