Detection of Malaria Parasites Using Digital Image Processing

Ahmedelmubarak Bashir[1], Zeinab A.Mustafa[2], Islah Abdelhameid[1], Rimaz Ibrahem[1]
Sudan University of Sciences and Technology, College of Engineering
School of Electronics Engineering[1], Biomedical Engineering Department[2]
Ahmedelmubarak512@gmail.com, zenab@k-space.org, islahabdelhameid@gmail.com, rimazabbo2011@gmail.com

***Abstract*—Malaria is a very serious infectious disease caused by a peripheral blood parasite of the genus Plasmodium. Conventional microscopy, which is currently "the gold standard" for malaria diagnosis has occasionally proved inefficient since it is time consuming and results are difficult to reproduce. As it poses a serious global health problem, automation of the evaluation process is of high importance. In this work, an accurate, rapid and affordable model of malaria diagnosis using stained thin blood smear images was developed. The method made use of the intensity features of Plasmodium parasites and erythrocytes. Images of infected and non-infected erythrocytes were acquired, pre-processed, relevant features extracted from them and eventually diagnosis was made based on the features extracted from the images. A set of features based on intensity have been proposed, and the performance of these features on the red blood cell samples from the created database have been evaluated using an artificial neural network (ANN) classifier. The results have shown that these features could be successfully used for malaria detection.**

Keywords— Malaria, erythrocyte, Parasite, Digital Image Processing, grayscale image.

I. INTRODUCTION

Malaria is a very serious infectious disease caused by a peripheral blood parasite of the genus Plasmodium. According to the World Health Organization statistics, in 2000, it was estimated that there were 262 million cases of malaria globally, leading to 839.000 deaths. By the year 2015, it was estimated that the number of malaria cases had decreased to 214 million, and the number of deaths decreased to 438.000. Majority of these deaths are children from Sub-Saharan Africa. This is due to the fact that, the environmental conditions are suitable for mosquitoes, in addition to the poor socio-economic conditions which make access to health care and disease prevention resources difficult [1]. There are various techniques to diagnose malaria, of which manual microscopy is considered to be "the gold standard". However due to the number of steps required in manual assessment, this diagnostic method is time consuming (leading to late diagnosis) and prone to human error (leading to erroneous diagnosis), even in experienced hands. As mentioned, this manual approach of diagnosis is time consuming and may lead to inconsistency. Thus, this demand trained and experienced technicians or pathologists. This approach once digitized will reduce the time taken for screening the disease. This will improve the consistency in diagnosis [2]. This study investigates the use and application of digital image processing for detecting malaria parasites using microscopic color images. An efficient method is proposed for parasite detection based on intensity and texture features. Parasite detection is the fundamental function of this semi-automated diagnosis.

The rest of this paper is organized as follows. Section II presents a discussion of related studies, Section III describes the employed system architecture, Section IV presents results and related discussions, and finally Section V concludes the paper.

II. RELATED WORK

In [3], the authors proposed a system for automatic detection of malaria parasite from blood images. This system employs image segmentation techniques to detect malaria parasites from images acquired from Giemsa stained peripheral blood samples. Another comparative study, [4], has presented an enhanced technique for Malaria Parasite Detection. This technique employed cell segmentation, where the segmentation process consists of various steps, such as image binarization using Poisson's distribution based minimum error thresholding, followed by morphological opening for the purpose of refinement. Furthermore, in [5], detection of Plasmodium infected erythrocytes, was investigated using a technique called holography.

Fig.1 illustrates five species that can cause human infection. Four of these are P. falciparum, P. vivax, P. ovale, and P. malariae , which are human malaria species that are spread from one person to another via the bite of female mosquitoes of the genus Anopheles. Current information suggests that P. knowlesi malaria is not spread from person to person, but rather occurs in people when an Anopheles mosquito infected by a monkey, which then bites and infects humans[1].

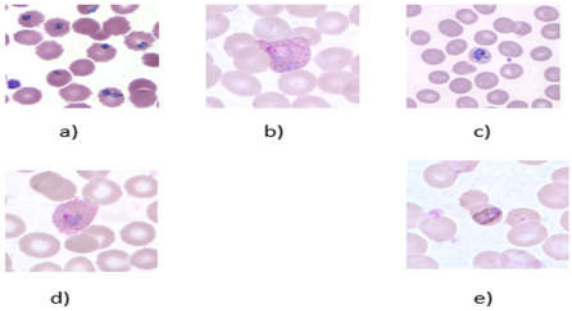

Fig. 1. Species of Malaria parasite, (a) Plasmodium Falciparum (b) P. Vivax (c) P. knowlesi (d) P.ovale (e) P.malariae

978-1-5090-1809-3/17/$31.00 ©2017 IEEE

There are three phases of development in the life cycle of most species of plasmodia[6]: exo-erythrocytic stages in the tissues, usually the liver; erythrocytic schizogony (i.e. protozoan asexual reproduction) in the erythrocytes; and the sexual process, beginning with the development of gametocytes in the host and continuing with the development in the mosquito. When an infected mosquito bites humans, several hundred sporozoites (the protozoan cells that develop in the mosquito's salivary gland and infect new hosts) may be injected directly into the blood stream, where they remain for about 30 min and then disappear. Many are destroyed by the immune system cells, but some enter the cells in the liver. Here they multiply rapidly by a process referred to as exo-erythrocytic schizogony. When schizogony is completed, the cells produced by asexual reproduction in the liver termed merozoites are released and invade the erythrocytes. In Plasmodium vivax and P. ovale, some injected sporozoites may differentiate into stages termed hypnozoites which may remain dormant in the liver cells for some time before undergoing schizogony causing relapse of the disease.

When the released merozoites enter erythrocytes, the erythrocytic cycle begins. This process is referred to as erythrocytic schizogony. Within an erythrocyte, the parasite is first seen microscopically as a minute speck of chromatin surrounded by scanty protoplasm. The plasmodium gradually becomes ring-shaped and is known as ring or immature trophozoite (Fig. 2. a-b). It grows at the expense of the erythrocyte and assumes a form differing widely with the species but usually exhibiting active pseudopodia (i.e. projections of the nuclei). Pigment granules appear early in the growth phase and the parasite is known as a mature trophozoite (Fig. 2. c). As the nucleus begins to divide, the parasite is known as a schizont (Fig. 2. d-f).

Dividing nucleus tends to take up peripheral positions and a small portion of cytoplasm gathers around each. The infected erythrocyte ruptures and releases a number of merozoites, which attack new corpuscles, and the cycle of erythrocytic schizogony is repeated. The infection about this time enters the phase in which parasites can be detected in blood smears.

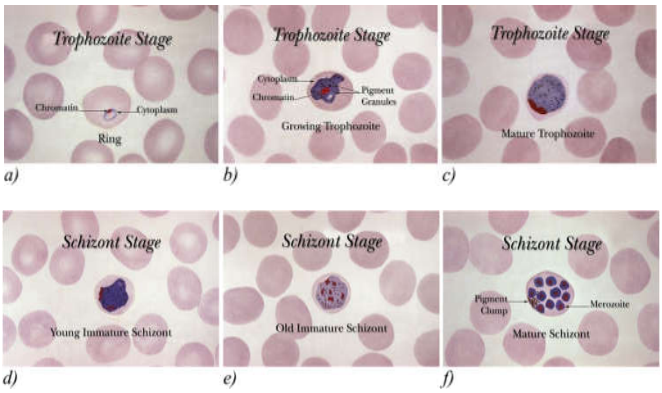

Fig. 2. Development stages of the Plasmodium parasite. Image courtesy of CDC.

Some merozoites on entering red blood cells become sexual gametocytes, instead of asexual schizonts. When a mosquito ingests gametes, the cells rapidly undergo gamete production. This is the third phase of development in the life of plasmodium, the sexual process of reproduction in a mosquito.

III. System Architecture

The system model is implemented using six main processes, namely; image acquisition, image preprocessing, image segmentation, feature extraction, comparison and classification as shown in Fig. 3.

A. Image acquisition

Thin blood smear images were acquired from the Centre for Disease Control (CDC) website [7] and captured from the Reference Laboratory of Malaria, in Sudan Ministry of Health.

B. Image pre-processing

The goal of this step is to make the acquired images more suitable for subsequent processes, mainly image segmentation and feature extraction. Basically, there are three main objectives for image pre-processing. One is to resize the image for the purposes of either magnifying the image through digital zooming, or reducing the image size in order to speed up processing. The second objective of image pre-processing is to reduce or eliminate noise from the acquired image. The third objective is to enhance the image contrast for visual evaluation. All images are rescaled to have the same size using the built in MATLAB™ function imresize. Since all CDC images have a resolution of 300×300 pixels, therefore they do not need to be rescaled. The size of CDC images is adopted to be the standard image size of the pre-processed images.

Both captured images and CDC images are converted from RGB to gray scale to reduce the processing time. RGB to gray conversion is done by using the built in MATLAB™ function rgb2gray, which converts RGB, images to grayscale by eliminating the hue and saturation information while retaining the luminance [8].

Filtering operation using a square median filter is performed to images. This operation served to remove spurious noise present in the images, The length of the median filter used is 7 by 7, a value obtained from [9].

Preprocessing effect is shown in Fig. 4.

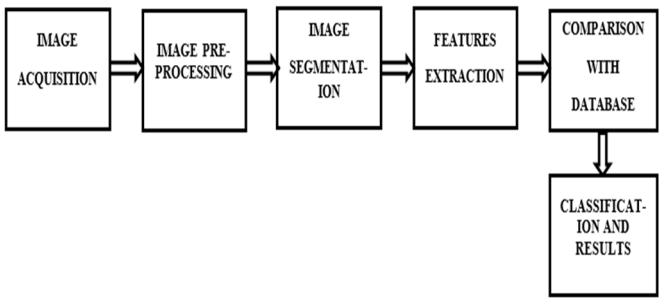

Fig. 3. Block diagram of the malaria diagnosis system.

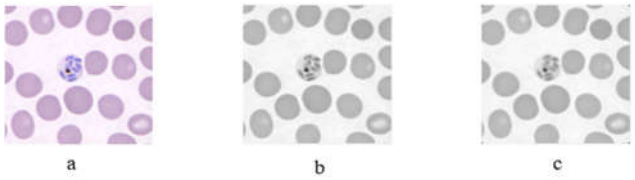

Fig. 4. Effect of image preprocessing, (a) rescaled image, (b) grayscale image, (c) filtered images.

C. Image segmentation

Image segmentation is the fundamental step to analyze images and extract data. Image segmentation is a mid-level processing technique used to analyze images and can be defined as a processing technique used to classify or cluster an image into several disjoint parts by grouping the pixels to form a region of homogeneity based on the pixel characteristics like gray level, color, texture, intensity and other features [10], [11]. The purpose of the segmentation process is to get more information about the regions of interest in an image, which helps in annotation of the object scene. The main goal of segmentation is to clearly differentiate between the object and the background in an image.

There are two objectives for image segmentation shown in Fig. 4. One is to isolate the red blood cells (RBCs) from the background and the second is to extract all the RBCs and process them individually in order to facilitate the process of feature extraction.

D. Feature extraction

In order to distinguish between infected and non-infected red blood cells, we need to extract features from the image array and compute new variables that concentrate information to separate classes. The set of features should discriminate between infected and non-infected RBCs as well as possible. An additional requirement is robustness, so that the results can be reproduced for new independently collected material. Raw images cannot be used directly as features due to high variations in morphology, which are coupled with arbitrary rotations and scales and because the raw images contain large amount of data, but relatively little information. The aim of feature extraction is to transform the input data into a reduced set of features that extract the relevant information from the input data.

Following the concept introduced in [12], the feature extraction process can be expressed in terms of the definition of the zone of measurement, and then measure the information required from that zone. This is the process generally followed in this work.

1) Intensity features

Intensity features are based only on the absolute value of the intensity measurements of the image. For most of the feature extraction methods, only one intensity value per pixel is assumed, i.e. the methods assume a gray-scale image. For some methods, the original RGB image is converted to gray-scale by eliminating the hue and saturation information while retaining the luminance. For certain features, the performance have been evaluated for several channels to choose the channel with best discriminating power.

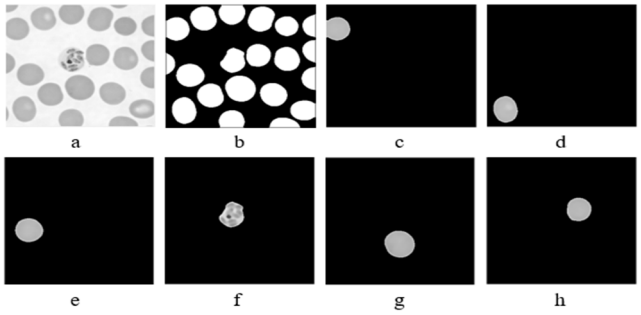

Fig. 5. RBCs extraction, (a) the grayscale image, (b) the labeled image, (c- h).extracted objects (RBCs).

Intensity values of the original image represent the transmitted light. However, for calculating some features, the pixels values from the extinction image can be more directly useful [12].

The two intensity features used in this work are variance and skewness, which are calculated as follow [13], [14]. First, let $\mathbf{h}(v) \in \mathbb{R}^{L\times 1}$, $\mathbf{h}(v) = \{h(v,i)\}_{i=1}^{L}$, denotes the frequency of pixel intensity value (histogram), where $v = 0, \dots, N$, and $p(v,i)$ is the probability distribution function (PDF), which is computed from the histogram by dividing it by the object's area A, which is given by

$$A = \sum_v \mathbf{h}(v), \tag{1}$$

And therefore the corresponding probability is calculated as

$$p(v,i) = \frac{h(v,i)}{A}. \tag{2}$$

The mean, μ_i, and the variance σ_i^2 are computed as

$$\mu_i = \sum_{v=1}^{N} v\, p(v,i), \tag{3}$$

And

$$\sigma_i^2 = \sum_{v=1}^{N} (v - \mu_i)^2\, p(v,i), \tag{4}$$

Respectively. Finally, the skewness is given by

$$\mu_3 = \frac{1}{\sigma_i^3} \sum_{v=1}^{N} (v - \mu_i)^3\, p(v,i). \tag{5}$$

Note that the mean is 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. Histogram skewness is a measure of histogram symmetry and it shows the percentage of the region's pixels that favor intensities on either side of the mean.

2) Threshold features

A suitable threshold is defined as clutch that can distinguish between the infected and non-infected RBCs. The experiments show that most of the infected RBCs have some pixels with values (i.e. gray level) less than 130, these pixels are mostly within the parasite region. In this study, the threshold is 130, where this value is obatined by the experiments.

The features described above are chosen among many sets of features, because they can lead to higher discriminative capabilities and then can improve the classification performance.

E. Database creation

The database is created by using a total of 77 images. From these images, a total of 1120 erythrocytes sub-images were cropped. 120 sub-images comprised of infected erythrocytes

while 1000 images were non-infected. The features extracted from all available erythrocytes.

F. Artificial neural network (ANN) classification

The NPRTOOL command was used to generate a MATLAB™ script which solves a Pattern Recognition problem with a Neural Network and it uses back propagation algorithm. In the script, firstly the input and target data was defined. The input dataset is a 740×3 matrix, which represents the selected features. The target data is a 2×740 matrix, which contains zeros and ones only with respect to normal and abnormal features in the input matrix. The hidden layers was set to 20, and then network training was performed using 740 erythrocytes. The network was tested using the rest of erythrocytes in addition to 244 erythrocytes randomly chosen from different images and its performance was evaluated.

G. Detection of plasmodium parasites

Detection of Plasmodium parasites was done by using a trained multilayer neural network. The network was trained with the features extracted form RBCs. The network searched through the images and identified regions infected by Plasmodium parasites. The system flowchart is depicted in Fig. 5.
Fig. 6. Shows the results after the final classification, the parasite is highlighted with red color.

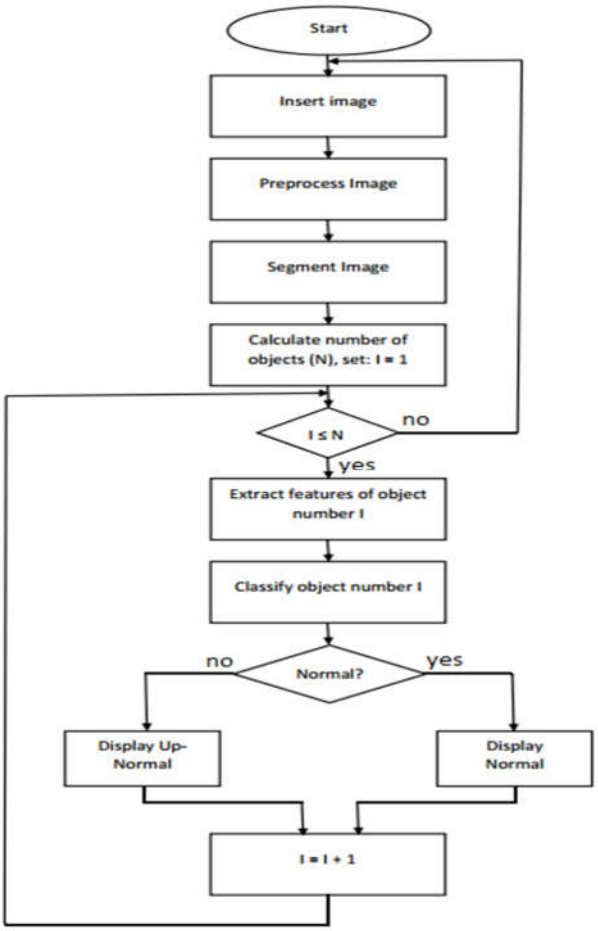

Fig. 6. System flowchart.

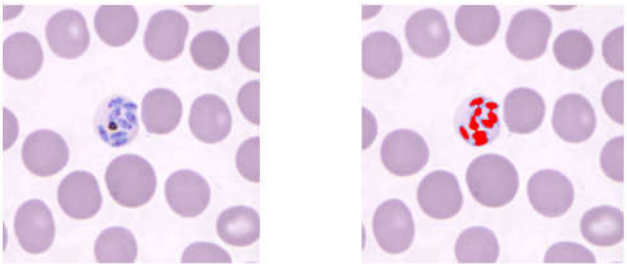

Fig. 7. Classification results, (a) original image, (b) detection of infected RBCs.

IV. Result Analysis

The performance of the employed classification method, was evaluated using the following statistical measures: (a) the sensitivity; (b) the specificity and (c) the accuracy. The *sensitivity*, $\mathcal{S}_s$, of a test (or symptom) is defined as the probability of a positive test result (or presence of the symptom) given the presence of the disease. The sensitivity is given by

$$\mathcal{S}_s = \frac{100\,\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}\,. \tag{6}$$

The *specificity* of a test, $\mathcal{S}_\mathrm{P}$, is defined as the probability of a negative test result (or absence of the symptom) given the absence of the disease. The specificity is given by

$$\mathcal{S}_\mathrm{P} = \frac{100\,\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}}. \tag{7}$$

Finally, the *accuracy* is how close a measured value is to the actual (true) value. The accuracy is given by

$$\mathcal{A} = \frac{100\,(\mathrm{TP} + \mathrm{TN})}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}, \tag{8}$$

where TP, TN, FP and FN denote the true positive, the true negative, the false positive and the false negative, respectively. The calculation process is shown in Table I.

TABLE I. PERFORMANCE EVALUATION OF CLASSIFICATION

	Abnormal	**Normal**	**Total**
Positive	**46** (TP)	**2** (FN)	**48**
Negative	**0** (FP)	**576** (TN)	**576**
Total	**46**	**578**	**624**

From Table I., TP = 46, FP = 0, FN = 2, TN = 576, Total = 624, and by applying in the above equations. The values are obtained numerically as

$$\mathcal{S}_s = \frac{46}{46+0} \times 100 = 100\ \%, \quad (9)$$

$$\mathcal{S}_P = \frac{576}{576+2} \times 100 = 99.65\ \%, \quad (10)$$

$$\mathcal{A} = \frac{46+576}{46+576+0+2} \times 100 = 99.68\ \%. \quad (11)$$

It is clear that from the performance evaluation, the accuracy is 99.68 %, which means that the artificial neural network gives more accurate result for the data used in this study.

V. CONCLUSION

A system for detecting Plasmodium parasites was implemented. The images used in this work were collected from different sources, then the images were processed and certain features were extracted. These features were then used to detect the presents of the malaria parasite. In addition, a graphical user interface has been designed to facilitate the use of the system. A total of 1120 erythrocytes sub-images were used to train and test the performance of the system. The outputs of the system were compared to the results of expert microscopists. The results were promising and the sensitivity of the proposed method outperforms most of the other reported methods. The system recorded 99.68 % accuracy in detecting the presence of Plasmodium parasites. The neural network, which has been trained with the back propagation algorithm, improves the accuracy and performance of the system. Moreover, the automated computer based method introduced in this project is interactive; hence, it is faster and more accurate than manual process.

REFERENCES

[1] W. World Health Organization, "Who Report 2015," 2015.

[2] N. R. Shet and N. Sampathila, "An Image Processing Approach for Screening of Malaria," *Canar. Eng. Coll. Mangalore*, pp. 395–399, 2015.

[3] D. A. Ghate and P. C. Jadhav, "Automatic Detection of Malaria Parasite from Blood Images," *Int. J. Adv. Comput. Technol.*, vol. 1, no. 3, pp. 66–71, 2012.

[4] M. S. Suryawanshi and P. V. V Dixit, "Comparative Study of Malaria Parasite Detection using Euclidean Distance Classifier & SVM," vol. 2, no. 11, pp. 2994–2997, 2013.

[5] A. Anand, V. K. Chhaniwal, N. R. Patel, and B. Javidi, "Automatic identification of malaria-infected RBC with digital holographic microscopy using correlation algorithms," *IEEE Photonics J.*, vol. 4, no. 5, pp. 1456–1464, 2012.

[6] V. Špringl, "Automatic Malaria Diagnosis through Microscopy Imaging," *Czech Tech. Univ. Prague Fac. Electr. Engeneering*, 2009.

[7] CDC, "CDC - Malaria - About Malaria - Biology - Malaria Parasites," *USA Government*, 2012. [Online]. Available: http://www.cdc.gov/malaria/about/biology/parasites.html.

[8] N. Bhargava and R. Bhargava, "Threshold and binarization for document image analysis using otsu ' s Algorithm," *Int. J. Comput. Trends Technol.*, vol. 17, no. 5, pp. 272–275, 2014.

[9] Aimi Salihah Abdul-Nasir, Mohd Yusoff Mashor, and Zeehaida Mohamed, "Colour Image Segmentation Approach for Detection of Malaria Parasites Using Various Colour Models and k -Means Clustering," *WSEAS Trans. Biol. Biomed.*, vol. 10, no. 1, pp. 41–55, 2013.

[10] W. Khan, "Image Segmentation Techniques: A Survey," *J. Image Graph.*, vol. 2, no. 1, pp. 6–9, 2013.

[11] A. Verm, M. T. Scholar, C. Lal, and S. Kumar, "Image segmentation: Review paper," *Int. J. Educ. Sci. Res. Rev.*, vol. 3, no. 2, 2016.

[12] K. Rodenacker and E. Bengtsson, "A feature set for cytometry on digitized microscopic images.," *Anal. Cell. Pathol.*, vol. 25, pp. 1–36, 2003.

[13] D. P. Doane and L. E. Seward, "Measuring Skewness : A Forgotten Statistic?," *J. Stat. Educ.*, vol. 19, no. 2, pp. 1–18, 2011.

[14] G. Brys, M. Hubert, and A. Struyf, "Goodness-of-fit tests based on a robust measure of skewness," *Comput. Stat.*, vol. 23, no. 3, pp. 429–442, 2008.