EasyChair Preprint

№ 4418

Malaria Detection Using ANN

Tasabih Ibrahim Alamin, Malaz Saifeldeen Mahmooud and Mohammed Y. Esmail

October 17, 2020

Malaria Detection Using ANN

Tasabih Ibrahim Alamin[2]
Department of biomedical engineering
University of Science and Technology
Omdurman- Sudan
E-mail: tsabeehibrahim@gmail.com

Malaz saifeldeen Mahmooud [2]
Department of biomedical engineering
University of Science and Technology
Omdurman- Sudan
Email: moza0912825951@gmail.com

Mohammed.Y. Esmail[1]
Department of biomedical engineering
Sudan University of Science and Technology
Khartoum- Sudan
E-mail: mohyagou@hotmail.com

Abstract- Malaria is a life-threatening disease so rapid, accurate diagnosis is required to control the disease. The detection of Malaria parasites is done by pathologists manually using microscopes. The manual microscopic examination is still the gold standard for malaria diagnosis, it is tedious, time consuming and requires special training and considerable expertise, where the lack of access to malaria diagnosis is largely due to shortage of expertise. The identification and counting the number of RBCs in the image is very important to diagnose. In this study proposed a system to detects and identification malaria parasites and counting the number of RBCs in the image from microscopic images. The dataset was collected from CDC dataset site. The proposed system began by pre-processing, to remove unwanted objects and noise from the image by morphological reconstruction. After pre-processing the image region of interest (Erythrocyte or RBCs) was segmented from the background by using Marker controlled watershed method applied on green channel color image. Next, the cell which has a plasmodium was extracted by k mean cluster. Then the statistical features were used in all of the cells to show the infected cells, and healthy cells. The ANN classified the normal and infected cells. The framework was tested to classify the RBCs in blood cell images, and the results are; accuracy (99.3%), sensitivity (98.6%), and specificity (94.8%).

Key words – Parasite, K-means, Morphological reconstruction, Back propagation, Segmentation.

I. INTRODUCTION

Malaria is a disease caused by four species of a parasite that is carried from person to person by a mosquito and transmitted by the bite of an infected female, anopheles mosquito. When a female anopheles mosquito ingests blood containing malarial parasites, these parasites reproduce in the mosquito's gastrointestinal tract, and then move to the salivary glands. When this mosquito bites another person, the parasites injected along with the mosquito's saliva. Inside the human, the parasites move to the liver, where they multiply. The parasites infect red blood cells, multiply inside the red blood cells and eventually cause the infected cells to rupture. Malaria is characterized by extreme exhaustion associated with paroxysms of high fever, sweating, shaking chills and anemia. Malaria is endemic in parts of Asia, Africa, Central and South America. It is a protozoan disease caused in humans by four species of the genus Plasmodium namely, P. falciparum, P. vivax, P. ovale and P. malariae. Identifying the parasite in a blood sample confirms the diagnosis [1].

Image segmentation is the division of an image into regions or categories, which correspond to different objects or parts of objects. Every pixel in an image is allocated to one of a number of these categories. A good segmentation is typically one in which: - Pixels in the same category have similar greyscale of multivariate values and form a connected region, - neighboring pixels which are in different categories have dissimilar values. Segmentation is often the critical step in image analysis: the point at which we move from considering each pixel as a unit of observation to working with objects (or parts of objects) in the image, composed of many pixels. If segmentation is done well then all other stages in image analysis are made simpler [2].

Watershed applied directly to the gradient image results in over segmentation due to irreverent minima

or noise patches or other image irregularities. The concept of markers can be used to solve this over-segmentation problem whose goal is to detect the presence of the homogenous regions from the image by a set of morphological simplifications. They spatially locate object and background ensuring to keep up the interior of the object as a whole. The Markers are connected components belonging to an image [3].

II. Related work

In 2016, they proposed method involves acquisition of the thin blood smear microscopic image at 100x magnification, pre-processing by partial contrast stretching, separation of infected cell from the image by applying k-means clustering on the a*b component of L*a*b color space, feature extraction (shape and textural) of the infected cell, finally training the K-nearest neighbor classifier to test the images. Instead of extracting features for the entire group of erythrocytes present in the image. The KNN classifier is trained with 386 images to detect three lifecycle stages (trophozite, schizont and gametocyte) for each of the four species of malarial parasites (P.falciparum, P.vivax, P.malariae, and P.ovale) with an accuracy of 90.17% and sensitivity of 90.23% [4].

In 2014, they proposed a computer vision based approach to identify the malaria parasite from light microscopy images. This research deals with the challenges involved in the automatic detection of malaria parasite tissues. The method is based on the pixel based approach, This method has the particularity of being based on the usage of the b*-color channel of the CIE (Commission International de l'Élairage) L*a*b* (lightness, green/red coordinate, blue/yellow coordinate) color space. A total of 330 Leishman-stained microscopic images were used, they used K-means clustering (unsupervised approach) for the segmentation to identify malaria parasite tissues. Results being reported of 76% and 60% for SE and SP , respectively. However, the proposed approach simply consists of classifying the images as infected/not infected [5].

In 2014, they proposed a computer-aided system is proposed to automate the process of detection and identification of RBC from blood smear image. Initially RBCs region are extracted from the background by using global threshold method applied on green channel color image. 100 sample images, Next, noise and holes in the RBCs are abolished by utilizing morphological filter and connected component labeling. Following that, information from the RBCs' are extracted based on its geometrical properties. Eventually, the RBCs were classified as normal/abnormal by using Artificial Neural Network (ANN) classifier. The framework was tested to classify the RBC in blood cell images, and the results are encouraging with an average of 83% accuracy [6].

In 2012, they proposed automatic system for classification of infected erythrocytes and life stage identification of malaria parasite uses O from Giemsa stained blood Images. After segmentation, watershed transform is used for separation of overlapping cells. Geometrical and statistical features are extracted from each cell and given to SVM binary classifier to classify erythrocytes as infected or normal. SVM binary classifier gives 96.26% sensitivity and 99.09% specificity. Color and geometrical features are extracted from infected erythrocytes and given to SVM multiclassifier which identifies life stage of Malaria parasite. SVM kernels such as linear, polynomial and RBF are used on 71 images. SVM multiclass classifier (i.e. RBF kernel) gives 96.42% accuracy for correct identification rate of life stage of parasite [7].

III. METHODOLOGY

Propose algorithm to detect the malaria parasite includes image acquisition, pre-processing, image segmentation by k-Means clustering, separation of infected cell by counting the number of pixels (image with the least number of pixels is selected), feature extraction of the segmented parasite, feature reduction and classification.

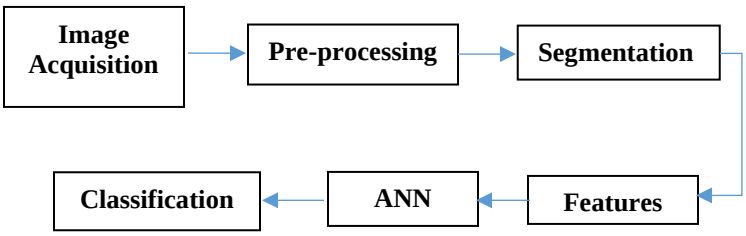


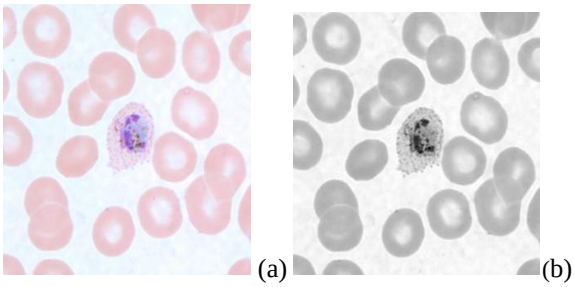
Fig 1: Proposed system for malaria detection

a. Image Acquisition

The dataset was collected from the CDC dataset, The CDC is The Centers for Disease Control and Prevention [8]. 70% of the images were used for training and 30% of the images were used for the test., normal images contain healthy RBCs, while abnormal images contain parasite inside the RBCs. The total of the normal and abnormal images is 81, 65 images used for training and 16 images used for test, MATLAB was used for implement the system.

b. Pre processing

The purpose of image pre-processing is to remove unwanted objects and noise from the image so that it becomes ready for the subsequent image segmentation process. The input image starts by two color images, first RGB image fig. 2(a) **(**in which the green channel from RGB Image give high contrast for RBCs fig. 2(b)**)**, the second image is LAB image fig. 2(c) **(**in which the b channel from LAB image give good representation for parasite fig 2(d)**)**. Then the filtration process applied by morphological reconstruction. Morphological operation removing the noise, and backlight without changing the RBSs size or shape.


(a) (b)

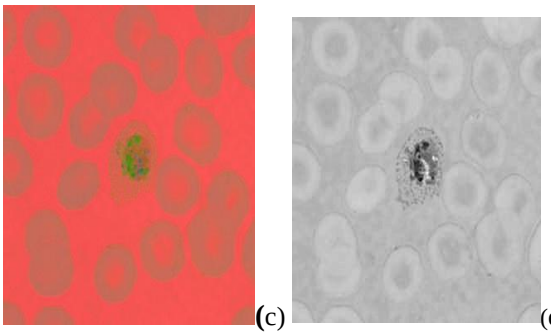
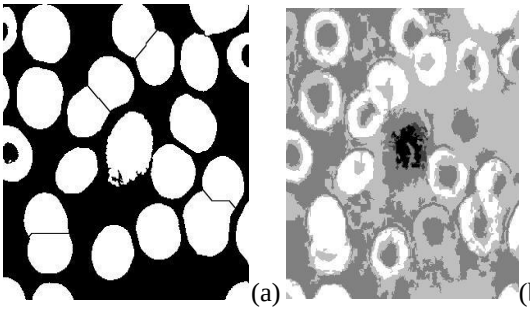

(c) (d)

Fig 2: (a)RGB image. (b)green channel. (c)LAB image. (d) B channel.

c. Image segmentation

The first segmentation for region of interest (Erythrocyte or RBCs) was segmented from the background by using the Marker controlled watershed, watershed is applied in green image to segment the RBCs and to separate overlapping cell fig 3(a). K - Mean clustering is then applied in green image, and it forms into 5 clusters fig. 3(b)**.** The cell which has a plasmodium was extract from lowest level fig. 3(c). In some cases, the segmentation process leaves behind some holes. Area **filter** takes the output of the k-mean process as it's input, to ensure that image have only infected cell. The result of area filter leaves holes in parasite, so **filling function** is used to fill the holes. And the **morphological closed** is used to fill all the tiny holes and noise that were missed, so the output is the infected cells. Finally, **XoR function** is used to determine the normal cell. Fig 3(d).


(a) (b)

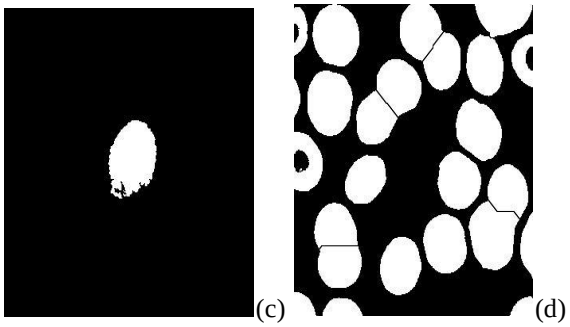


Fig 3: (a) Marker controlled watershed. (b) K-mean cluster. (c) the infected cells. (d) Normal cells.

d. Feature Extraction

Statistical features were calculated for the normal and infected cells and it was made using 10 functions. (mean, standard deviation, variance, energy, entropy, standard deviation errors, Skewness, kurtosis, mean square error, and correlation The features were calculated of each cell (normal and abnormal) in the image.

e. Artificial neural network

Feed forward network training was used; a network structure was selected by used 50 hidden layer. When training a feed forward, the information is passed into the net, and were calculated the weighted sum of its inputs and then a sigmoid function was applied to normalize the sum. If the net classification is incorrect; the weight is adjusted backward through the net in

the direction that would give it the correct classification. Each training sample on the testing data sets was analyzed in the manner as the training data set; a feature was extracted from the image sample. The feature set was identified by the trained classifier. The performance evaluation was calculated include sensitivity, specificity, and accuracy.

f. Classification performance

In medical image analysis the performance of a classifier was evaluated typically by the measured sensitivity (the true positive rate) and specificity (The true negative rate) instead of measuring direct accuracy of the classification.

The table show classification performance the result of statistical features and Feed forward (ANN) by the training back propagation. MATLAB program was used to calculate the accuracy, sensitivity, specificity and error rate shown in table below.

Sensitivity	99.3%
Specificity	94.8%
Accuracy	98.6%

Table (1): system result.

All cells highlighted in red in the image bellow fig 4 were the unhealthy cells and the ones highlighted in green were the healthy cells, the system classified malaria based on those unhealthy cells to prove that.

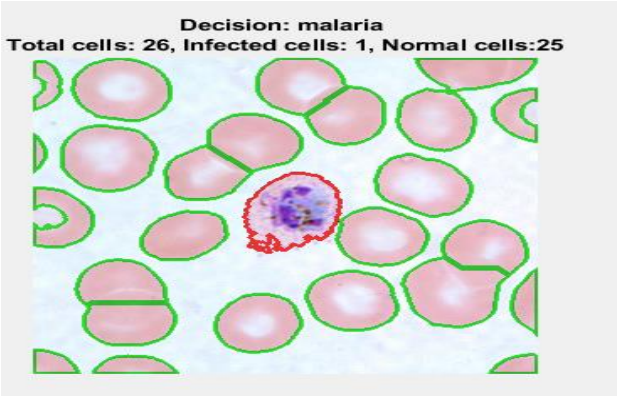


Fig 4 The final result after detection.

VI. CONCLUSION

The paper aimed to design a system to detect four types of malaria. The process began by preprocessing of the input image by two color system, first the RGB Image .and the second the LAB image the plasmodium stood out in green color, then a morphological reconstruction was done to uniform the image and get rid of the noise then the cells were segmented by watershed marker controlled which removed all of the overlapped cells and separated the cells of the image which made it easier to calculate. The following step was clustering of the image by K-Means which produced an image with an outstanding plasmodium cells then it was easier extracted, Then the XOR was used to reverse the output of the k-means which was the healthy cells, then the statistical

feature was used in all of the cells to show the infected cells, healthy cells. Next features applied to each cell, then the ANN classified the normal and up normal cells. The framework tested to classify the RBC in blood cell images, and the results are accuracy (99.3%), sensitivity (98.6%), and specificity (94.8%).

References: -

1. Bench Aiods , for Malaria Microscopy. Geneva, World health Organization, 2010
2. C. A. Glasbey, G. W. Horgan, Image analysis for the biological sciences, the University of Michigan,1995
3. Beucher, S., Meyer, F., 1993. The morphological approach to segmentation: the watershedtransformation. In: Dougherty, E. (Ed.), Mathematical Morphology in Image Processing. Marcel Dekker, New York,GOTETC, 2012.
4. Akshay Nanoti ; Sparsh Jain ; Chetan Gupta ; Garima Vyas , Detection of malaria parasite species and life cycle stages using microscopic images of thin blood smear, 2016 International Conference on Inventive Computation Technologies (ICICT)
5. Najeed Ahmed Khan ; Hassan Pervaz ; Arsalan Khalid Latif ; Ayesha Musharraf ; Saniya, Unsupervised identification of malaria parasites using computer vision, International Joint Conference on Computer Science and Software Engineering (JCSSE), 2014
6. Razali Tomaria,*, Wan Nurshazwani Wan Zakariaa, Muhammad Mahadi Abdul Jamila, Faridah Mohd Norb, Nik Farhan Nik Fuad, Computer Aided System for Red Blood Cell Classification in Blood Smear Image,2014
7. S.S.Savkarea , S.P.Naroteb, Automatic System for Classification of Erythrocytes Infected with Malaria and identification of parasite life stage,2012
8. https://www.cdc.gov, centers of disease control