FEATURE EXTRACTION OF MALARIA PLASMODIUM DETECTION

[1]Tsedeke Temesgen Habe
Dept. of Computer science and Engineering
Parul Institute of Engineering and Technology, Vadodara, India

Abstract*— Malaria is a grave infectious disease, according to the World Health Organization (WHO), it is accountable for nearly half million deaths each year in the year 2016 report. In the world there are three different manual malaria diagnosis techniques such as sign and symptoms, Rapid diagnosis and Microscopic. The aim of this study is work on analysis of suitable feature extraction techniques which is discovering to classify the complete life cycle stages of malaria. The proposed new feature extraction techniques are Gray level co-occurrence matrix (GLCM) textural feature. It is gives us the 12 features like Entropy, Energy (angular second moment), Correlation, mean smoothness, skewness, kurtosis, shade measure to compute results. Finally, after the feature extraction we applied the classifier techniques using artificial neural network used to detecting malaria parasite.

***Keywords*— GLCM, Parasite, Histogram, RBC, Plasmodium, Malaria;**

I. INTRODUCTION

Malaria is a grave infectious disease, according to the World Health Organization (WHO), it is accountable for nearly half million deaths each year in the year 2016 report. Malaria happens in more than 100 countries and territories. Most of Africa, South Asia, Southeast Asia and parts of Central, South America, the Middle East, and Oceania are measured regions where malaria spread [12]. There are various techniques to diagnose malaria of which manual microscopy is considered to be the bullion standard. Conversely due to the number of steps required in manual assessment, this diagnostic method is time consuming (leading to late identification) and lying face down to human error (leading to mistaken diagnosis), even in qualified hands.

Types of Plasmodium malaria diseases are falciparum, vivax, and oval and malaria parasite. Each malaria plasmodium diseases can have any of three stages called schizonts, trophozoites or gametocyte. Malaria is headache of world health organization why because it is over control problem in many countries of the world. According to World Health Organization (WHO) report 2016 [12], in spite of the development in reducing cases as well as deaths due to malaria, it is estimated that a million cases and death of malaria occurred worldwide in 2016. The precise and appropriate diagnosis of malaria disease is necessary to manage and treat the disease. There are three common malaria diagnosis techniques in the world: sign and symptoms, rapid Diagnostic Test (RDT) and microscopic diagnosis

Sign and symptoms: The traditional method of diagnosis which can be takes place with the area of poor country because of scarcity of material. The accuracy of the disease is too poor based on estimated by assumption of the malaria by depending on traditional characteristics of the malaria. Rapid Diagnostic Test (RDT) is sometime known as finger pick blood sample test or dipstick uses immune chromatographic techniques. Microscopic diagnosis of malaria is which is processing by using examination of malaria from thin blood film stained with gemia. This examination techniques better for accurate than other two methods because of using instrument which is not automated by image processing but shows half accurate malaria information from the blood film. Before the parasites can be observed, however, a blood layer must be ready, dried up, and discolored finally, check up under the microscope. When the blood analyst examines discolored parasites, the identification of malaria is confirmed.

Based on the manual diagnostic method we proposed the automatic diagnosis method from microscopic blood film image by using different algorithm techniques which is reviewed from different papers. This diagnosis method is quality based and high accuracy of malaria examination result yielded from the image segmented. In this proposed paper there are two tasks such as detecting and extracting malaria from red blood cells by using feature extraction techniques such as color feature extraction and geometric feature extraction. The focal point of this study is to develop a feature extraction technique to modify/eliminate gray level co-occurrence matrix and one that has a pro over other techniques in that it minimizes human dependence and is, therefore, more constant in applying feature extraction techniques.

II. BACKGROUND OF THE STUDY

The work illustrated in this dissertation arose from collaboration with where there have been several journal survey researchers on various aspects of malarial detection and identification due to malaria is caused by a parasite in the RBC. Many detection techniques are referred from journal survey papers. Techniques like segmentation, feature extraction and classification sub techniques are observed in the review paper. The main background of the research is finding and improves feature extraction techniques for detecting malaria parasite to improve accuracy rate of the complete malaria stages. Dealing with this, many researchers have proposed different algorithms to solve the problem, however, accuracy of distinguishing the presence of parasites and approximation parasitemia under the happening of cells with multiparty structural arrangement is still observed as a challenging task. Hence, in this research, developed a better technique to address malaria disease discovery and will shape parasitemia on the base of malaria in red blood cells (RBCs), also called erythrocytes that are take out from microscopic thin RBC films is looked for.

III. MOTIVATION

The general motivation of this work is to develop feature extraction of malaria from microscopic thin blood films. The motivation of the study is:-

- ✓ Understanding of Image processing techniques.
- ✓ Finding suitable feature extraction techniques and improving the complete stages of malaria parasitimia.

IV. PROBLEM STATEMENT

- ✓ Lack of effectual counting of RBCs and Plasmodium parasites with being there of multifaceted arrangement of red blood cell such as overlap and broken cells that lay around the border of image.
- ✓ Poor system performance of the accuracy for detection malaria. The outcome of this proposed study is believed to improve detection and counting of RBCs through image analysis.

V. GENERAL OBJECTIVE

The general objective of this study is investigate feature extraction to malaria detection system that will enable us to analysis of suitable features and optimal feature selection technique will be discover to classify the complete life cycle stages of malaria.

VI. SCOPE

The main focal point of this study will be analysis of suitable features technique will be discover to classify the complete life cycle stages of malaria [3]. Malaria detection will be considered only for two species namely Plasmodium vivax and Plasmodium falciparum as availability of data and degree of occurrence being high relative to other species.

VII. LITERATURE REVIEW

In this [1] paper have presented three main components, they were preprocessing, feature extraction, and classification Author proposed a precise techniques to classify plasmodium vivax from digitalization microscopic thick blood film by means of jointed of second order statistic feature extraction with K-Nearest Neighbor (K-NN) classifier technique. In this feature extraction paper, author's proposed scope were Gray Level Co-occurrence Matrix to get contrasts, correlations, energy, and homogeneity values [1]. For further research developer the feature works are using other methods of classification that can get advanced the accuracy rate of classification malaria parasites in thick blood performed. Based on the result of experiments classifier, in this paper has an advantages to calculate a high accuracy for categorize plasmodium vivax with average accuracy rate 95%. The main disadvantage is in this authors is because of the k value is small, and then this leads to the comparison is less [1].

In [2] the paper presented aims are to contrast the stage of accuracy amongst a number of methods that are applied to categorize malaria plasmodium falciparum. The scope this paper is, author used Gray Level Co-occurrence Matrix (GLCM) to get texture feature standards such as energy, contrasts, homogeneity, and correlations. Those standards are used as an input of classification techniques. That means, input for back propagation, LVQ, and KNN are same matrix [2]. The future work in this paper, other researcher can apply segmentation algorithm to identify plasmodium falciparum automatically. This is possible to get high accuracy result of plasmodium falciparum detection. The advantage of this paper is the author used small number of training data to learn the device [2]. The drawback of this paper is lowest detection rate 73.3334% but on each classifier got average accuracy rate 82%. In many of paper the rate is accuracy rate more than 98% . In this paper there is a gap on accurate detection of malaria plasmodium falciprarum [2].

In [3] paper presented are prediction error, LBP-GLCM, Chrominance channel histogram, R-G color channel difference histogram are the proposed features which are added with the some of the existing features scope like Green channel histogram ,Shannon entropy, Renyi entropy, Saturation channel histogram, Havarda and Charvat entropy, Kapur entropy. Author [3] presented the future scopes are, the optimal feature selection and analysis of suitable features technique will be look at to classify the complete life cycle phases of malaria. The scope of this paper was improving the feature extraction techniques as we describe in the beginning line. The accuracy rate of the paper is 98.5% this is sufficient output rate since his proposed work was improved 7.78% of accuracy rate. In the classifier techniques author used Naive Baye, SVM, K-NN and Hybrid classifier. In the paper the drawback is K-NN algorithm have showed the less performance than its remaining classifier techniques. And the advantage is SVM and hybrid methods have showed good accuracy rate 93.12 and 93.88 repeatedly of feature extraction [3].

The aim of [4] paper is improving the feature extraction techniques for detection of malaria parasites. The scope is feature extraction based on histogram-based texture is applied to extract feature infected cell from RBC. Multilayer perceptron back propagation algorithm is used to classify all features. The results show that the proposed method in this paper used to achieves accuracy rate of 87.8%, sensitivity of 81.7%, and specificity of 90.8% for detecting infected red blood cells [4]. This is advantages to improving decision-making for malaria diagnosis for the author. Feature work is further studies are necessary to increase the classification with more feature extraction methods for identification. The weakness of this paper is the performance of accuracy rate is not efficient [4].

In [5] paper the author proposed to develop automated detection of malaria parasite for assumption of parasitemia by used pre-processing, Segmentation, feature extraction and classifier. Author [5] used different sub techniques for the above main methods like median filtering for pre-processing to remove salt and pepper noise. Morphological operators joined together with watershed algorithm for blood cell segmentation in to different region of interest. Discrete Wavelet Transform and Gray level co-occurrence Matrix for RBC image feature extraction to re defining of bulky set of unneeded information into set of characteristic reduced dimension [5]. Also the author [5] used Support Vector Machine (SVM), Naïve bayes and neural network as classifier of red blood cells in to two parts such as infected RBC and non-infected RBC. The author feature work is developing Android Application to Diagnose Malaria by using different feature extraction. We concluded this paper main advantages are DWT feature extraction have high performance than GLCM feature techniques. The main drawback is GLCM has less performance of accuracy [5].

In [6] paper detection is achieved by using few methods for detecting malaria parasites like feature extraction. Aim is the system will generate the output either the person is infected by malaria or not. Feature Extraction consists of various methods but these authors' system uses two main methods: such as color extraction and edge detection and used several types of edge detectors methods like Sobel Operator, Robert's Operator, Log Operator, Canny Operator, Prewitt Operator and Zerocross Operator for generation of malaria RBC image detail [6]. Finally, the advantage part in this paper due to good feature extraction techniques helped to achieved high percentages of sensitivity, specificity, positive prediction value. Drawback of this paper is having not proper and limited accuracy range [6].

The author [7] proposed to identify Malaria Parasite Stage in Microscopic blood image by using acquisition, segmentation, feature extraction and then classification techniques. For each techniques author is use different sub method. Basic feature, texture based feature and statistical feature for RBC image feature extraction [7]. And support vector machine (SVM) for classification of RBC image and to determine the output. The advantages in [7] paper is high detection accuracy rate was performed while the effect shows upgrading in diagnostic accuracy of detection of malaria parasite in Red Blood Cells and also describing the life cycle stage of the parasite. The accuracy, sensitivity and specificity achieved were as 97.7%, 97.4%

and 97.7% respectively. Disadvantage is required more training data for more performance accuracy In future work this system can be used for diagnosis of other hematological disorders and also the RBC and WBC count which help out to diagnose lots of disorders in human body [7].

VIII. COMPARISON OF FEATURE EXTRACTION TECHNIQUES

Table 1.1: Comparison of Feature extraction technique

Methods	Advantages	Disadvantage
GLCM (Gray Level Co-occurrence Matrix)[1]-[2][8] -[10],[21],[5]	Accepting the spatial interaction between image pixels. That means it is obtained from two pixels	They require a lot of computation (many matrices to be computed) 1. Features are not invariant to rotation or scale changes in the texture
Prediction error [4],	Sufficient output rate with regard of improving accuracy rate	K-NN algorithm have showed the less performance
Histogram-based texture [4] , [21]	Estimate properties using of individual pixel values	Ignoring the spatial interaction between image pixels.
DWT: Discrete Wavelet Transform	DWT is more reliable feature extraction	less accuracy
Edge Detection and Corner Detection[6]	Achieved high sensitivity, specificity, positive prediction value	High computing time required

IX. METHODOLOGY

To implement RBC image feature extraction we must use standard techniques because the red blood cells have complex arrangements. Based on this complexity of RBC arrangement we proposed GLCM feature extraction such as Entropy, energy, correlation, variance, homogeneities, prominence and cluster shade. The scope is analysis of Gray level co-occurrence matrix features extraction technique combined with histogram based textural feature that will help us to discovering classified life cycle stages of malaria. See figure 1.1

Table 1.2: Architectural techniques

Pre-processing	Image enhancing , Median filter [4]
Segmentation	K-means algorithm, morphological operation [4]
Feature Extraction	**GLCM (Proposed)**
	Histogram based texture [4]
Classification	Artificial neural network

Table 1.3: Proposed Feature extraction techniques

Feature Extraction	F1	GLCM	**Proposed**
	F2	Histogram based Textural [4]	Existing

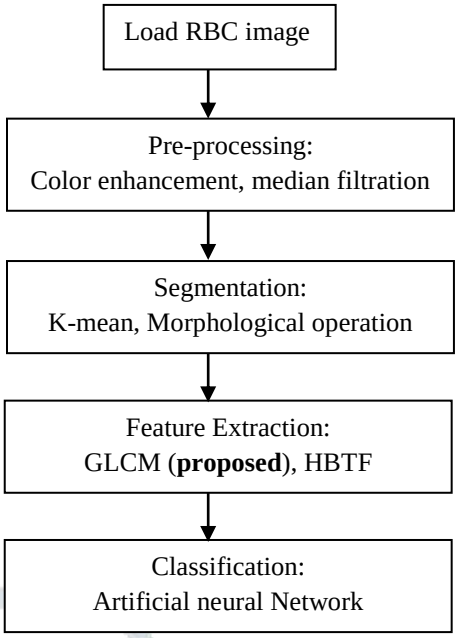


Figure 1.1: Architectural operational Techniques

a) Preprocessing

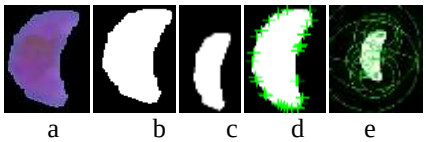
In this step we applied color 5x5 matrix size median filters. Finally, the median image filter technique is computed by first sorting all the pixel values from the framework into mathematical ordering, and then substituting the pixel being considered with the middle pixel value [4].

a b c

Figure 1.2: a) original image, b) after conversion RGB color to Lab color space and c) after median filtration

b) Segmentation

Segmentation procedure pass through is K-means algorithm to segment colour image and after segmented nuclei or more nearest ROI into a separate image morphological operation applied on result of k-means algorithm to masking hole, removing small object, extracting region of interest [4].



a b c d e

Figure 1.3: a) After k-means color image segmentation and morphology result b) after conversion of Lab color to binary image and filling holes ,c) after removing small objects , d) after found connected components and e) after sketched centre of connected components.

c) Feature Extraction

In this paper we proposed seven (7) features of Gray level co-occurrence matrix by combining five (5) features of histogram textural to detect the three (3) malaria life cycle stages such as Schizonts Trophozoites Gametocyte.

Table 1.4: System generated feature extraction

Feature	Values	selected
1. Energy	6120	✓
2. Entropy	-1.3589e-12	✓
3. Contrast	11069249	✓
4. Correlation	1.0000	✓
5. Homogeneity	445.5336	✓
6. Prominence	7.7009e+28	✓
7. Shade	-4.0888e+22	✓
8. Mean	0.0680	✓
9. Standard deviation	0.2517	✓
10. Variance	0.0634	✓
11. Skewnes	3.4320	✓
12. Kutosis	12.7788	✓

In [4] the author use histogram based feature extraction to driving 7 features of gray RBC image. In our study, we will apply both gray level co-occurrence matrix textural and histogram based textural features by combining their 12 features. In the table below the first seven are proposed from GLCM and the last five features are from existed paper [4] were applied.

As we discussed on the proposed work, we applied GLCM and HBT feature by combining their features. The total feature extractions are 12 and their questions are described below in detail. GLCM is working in the following based on the following questions

Normalization: $P_{i,j}=\frac{v_{i,j}}{\sum_{i,j=0}^{N-1} v_{i,j}}$ (2)

Based on this equation we generated the equation of 7 second order static features.

Energy: .The square root of Angular Second Moment (ASM) feature character is applied as energy. Its variety is [0 1] since stable image its value is 1[19].The equation of energy is

$$E = \sum_{i,j=0}^{N-1}(P_{ij}{}^2) \quad (3)$$

Contrast: In contrast measure, weight increases exponentially (0, 1, 4, and 9) as persists from the diagonal [19]. The equation of contrast is [19]

$$C = \sum_{i,j=0}^{N-1} P(i-j)^2 \quad (4)$$

Correlation: On behalf of constant image its value is NaN. Range = [-1,1] and the formula is [19]

$$Corr = \sum_{i,j=0}^{N-1} P_{ij}\frac{(i-\mu)(j-\mu)}{\sigma^2} \quad (5)$$

Homogeneity: contrary of contrast mass is homogeneity mass values, with weight reduces exponentially loose from the diagonal. The weight calculated in contrast is (i-j)^2 and in homogeneity ,it is 1/1+(i-j)^2.The equation is [19]

$$H = \sum_{i,j=0}^{N-1}\frac{P_{ij}}{1+(i-j)^2} \quad (6)$$

Entropy: Entropy point out the complexity or complication of the RBC image. By nature the human red blood cells have complicated arrangement. So identifying this arrangement is not the easy task. Due to this complexity entropy measured by the following equation [4].

$$Entr = \sum_{i,j=0}^{N-1} -\ln(P_{ij})P_{ij} \quad (7)$$

Cluster prominence: is also a measure of asymmetry gray level image. During the prominence value is increase; the image is decrease symmetric [19]. It is calculated by:

$$Prom = \sin(B)|B|^{1/4} \quad (8)$$

Cluster shade: A new "$i + j$" RGB image is created, having a range of integer values from 0 to 2 [19]. It is calculated by $Shade = \sin(A)|A|^{1/3}$

Where:

Pij = i,j Element of the normalized symmetrical GLCM

N = Number of Gray levels in the image as nominative by number of levels in under Division on the GLCM texture page of the variable Properties dialogue box[19].

μ = the GLCM mean (being an computation of the intensiveness of all pixels in the kinship that contributed to the GLCM), calculated as:[19]

σ 2 = the variant of the intensiveness of all mention pixels in the kinship that modify to the GLCM, measured as [19]:

Note: This may estimate, but is not indistinguishable to, the variant of the strength of all the pixels in the data window W (as defined by the GLCM algorithm), and it is dependent upon the selection of spatial relationship in that algorithm.

d) *Classification*

In ANN (Artificial Neural Network) multilayer perceptron is a system of training processor which has a similar characteristic to human network. ANN (Artificial Neural Network) can be used to categorize and detect a pattern of RBC image or objects. Multilayer Perceptron Back propagation is one of algorithm of ANN (Artificial Neural Network) which has hidden layer between input layer and output layer [4]. In base paper author used Multi-perceptron Back propagation method for classifying the extracted RBC gray level image feature. The performance of the classification operation is commonly enumerated in terms of its predictive accuracy, sensitivity, and specificity.

Table 1.5: Improved result on existed performance [4]

Stages	Accuracy %	Sensitivity %	Specificity %
Trophozoite	89	97	97.67
Schizont	87.5	89	91
Gametocyte	95.66	93	92

X. CONCLUSION

Malaria is one of the grave diseases in the world. This study is extract life cycle of malaria Plasmodium parasites and computing parasites by using gray level co-occurrence matrix feature extraction techniques. The methodology after reviewed and comparison of different papers the proposed feature techniques are gray level co-occurrence matrix (GLCM). The existed technique is histogram based feature. The main scope of this study is identifying suitable algorithm for life cycle growth stage of malaria parasite extraction. The GLCM (Gray level co-occurrence matrix) jointed histogram based feature techniques is our suitable algorithm to find extract malaria from RBC image. Finally, in those selected papers complexity, over-fitting, sensitive noise response and low quality of image and less accuracy problems are examined for more accurate result.

XI. ACKNOWLEDGE

I am very grateful to **Asst. Prof. Hetal Bhaidasna**, Computer Science and Engineering department, Parul Institute of Engineering and Technology who was a supporter and an eager advisor, for provide the guidance for this survey. I am also extends my thanks to **Asst.Prof. Makhduma Saiyad** to her motivations and advices.

XII. REFERENCES

[1] Farah Zakiyah Rahmanti1 , Sutojo1, Novita Kurnia Ningrum1, Niken Kartika Imania1 Mauridhi Hery Purnomo2, "Plasmodium Vivax Classification from Digitalization Microscopic Thick Blood Film Using Combination of Second Order Statistical Feature Extraction and K-Nearest Neighbor (K-NN) Classifier Method" IEEE-ICICI-BME,2015, 978-1-4673-7800-0/15,79-83

[2] Farah Zakiyah Rahmanti,"A Comparison of Plasmodium Falciparum Identification from Digitalization Microscopic Thick Blood Film" ,IEEE-ICBE ,2016 ,978-1-5090-4142-8/16.

[3] Salam Shuleenda Devi1 & Amarjit Roy1 & Joyeeta Singha1 & Shah Alam Sheikh2 & Rabul Hussain Laskar1, Springer "Malaria infected erythrocyte classification based on a hybrid classifier using microscopic images of thin blood smea" , Springer Science+Business Media New York 2016

[4] Hanung Adi Nugroho1, Son Ali Akbar2, E. Elsa Herdiana Murhandarwati3 ,"Feature Extraction and Classification for Detection Malaria Parasites in Thin Blood Smear" , IEEE-ICITACEE,2016, 978-1-4799-9863, 197-201

[5] TEJASHRI CHAUDHARI1, Prof.D.G.AGRAWAL, "Automatic detection of malaria parasites for estimating parasitemia" ,J.IJAREST, , ISSN(P) 2394-2444, 2017,2834

[6] Rohit. M. Kandari, Harshada. S. Bhere, Shraddha. B. Pophale, Prof. Sonali. J. Man, "Malaria Parasite Detection Using Image Processing", IJIR, Imperial Journal of Interdisciplinary Research,Vol-3, Issue-2, 2017, 467-469

[7] Kshipra Charpe Research Student, "A Novel Method for Automatic Detection of Malaria Parasite Stage in Microscopic Blood Image.", IJCA (0975 – 8887) Volume 128 – No.17, 2015, 32-37

[8] 1Ashok Sutkar, 2 Marathe N. V , "Detection and Analysis of Malarial Parasites Using Microscopic Images:", IJCAT, ISSN : 2348 - 6090 ,2014, 605-610

[9] 1Rika Rosnelly, 2Sri Hartati, 3Edi Winarko, 4Sri Mulatsih , " Identification of malaria disease and its stadium based on digital image processing",JTAIT, ISSN: 1992-8645,2017, 700-710

[10] Miss. Shruti Annaldas, "Automatic Identification of Malaria Parasites using Image Processin.", International Journal of Emerging Engineering Research and Technology , Volume 2, Issue 4,, PP 107-112 ISSN 2349-4395,2014 , 107-112

[11] VinayakK.Bairagi1 andKshipraC.Charpe2, "Comparison of Texture Features Used for Classification of Life Stages of Malaria Parasite", Article ID 7214156, 9 pages", 2016.

[12] World health organization malaria report, ISBN 978-92-4-151171-1, 2016.

[13] Rafael C. Gonzalez, digital image processing, 3rd edition , Upper Saddle River, New Jersey 07458,ISBN 0-[13]-168728-x 978-0-[13]-168728,2008, 25-27

[14] Aimi Salihah Abdul-Nasir, Mohd Yusoff Mashor, Nurul Hazwani Abd Halim, and Zeehaida Mohamed "The cascaded moving k-means and fuzzy c-means clustering algorithms for unsupervised segmentation of malaria images ",AIP Conference Proceedings 1660, 070111 (2015); Published by the American Institute of Physics

[15] DPDx - Laboratory Identification of Parasitic Diseases of Public Health Concern, May 3, 2016 //www.cdc.gov/dpdx/malaria/index.html

[16] Digital OMAX 40X-2000X Digital Lab LED Binocular Compound Microscope with Double Layer,2017 , http://www.amazon.in/com

[17] The MaMic Image Database ,http:// fimm.webmicroscope.net /Research/Momic/mamic.

[18] Daniel Zemene Mequanint Advisor: Yaregal Assabe (PhD), addis ababa university college of natural sciences," Automatic Malaria Detection from Images of Microscopic Thin Blood Films",AAU,2015

[19] GLCM Texture Feature , Last update: 28 August 2016 , http://support.echoview.com/WebHelp/Windows_and_Dialog_Boxes/Dialog_Boxes/Variable_properties_dialog_box/Operator_pages/GLCM_Texture_Features.htm

[20] K Means, Written by Chris Piech, Fall 2012, http: // stanford.edu/ ~cpiech/cs221/ handouts/ kmeans.html