Abbas, N, Saba, T, Rehman, A, Mehmood, Z, Kolivand, H, Uddin, M and Anjum, A

Plasmodium life cycle stage classification based quantification of malaria parasitaemia in thin blood smears.

Plasmodium Life Cycle Stage Classification Based Quantification of Malaria Parasitaemiain Thin Blood Smears

Naveed Abbas[1], Tanzila Saba[2], Amjad Rehman[3], Zahid Mehmood[4], Hoshang kolivand[5]

Mueen Uddin[6], Adeel Anjum[7]

[1]Department of Computer Science, Islamia College Peshawar, KPK, Pakistan
[2]College of Computer and Information Sciences, Prince Sultan University, Riyadh, 11586, Saudi Arabia
[3]College of Computer and Information Systems, Al Yamamah University, Riyadh, 11512, Saudi Arabia

[4]Department of Software Engineering, University of Engineering and Technology Taxila, 47050, Pakistan
[5]Department of Computer Science, Liverpool John Moores University, Liverpool, UK, L3 3AF

[6] Information System Department, College of Engineering, Effat University of Jeddah, Saudi Arabia
[7] Department of Computer Science COMSATS University Islamabad Pakistan

Abstract

Visual inspection for the quantification of Malaria Parasitaemiain (MP) and classification of life cycle stage are hard and time taking. Even though, automated techniques for the quantification of MP and their classification are reported in the literature. However, either reported techniques are imperfect or cannot deal with special issues such as Anemia and Hemoglobinopathies due to clumps of RBCs. The focus of the current work is to examine the thin blood smear microscopic images stained with Giemsa by digital image processing techniques, grading MP on independent factors (RBCs morphology) and classification of its life cycle stage. For the classification of the life cycle of malaria parasite the *k*-nearest neighbor (*k*-NN), Naïve Bayes and Multi-Class SVM are employed for classification based on HOG and LBP features. The proposed methodology is based on inductive technique, segment malaria parasites through the adaptive machine learning techniques. The quantification accuracy of RBCs is enhanced; RBCs clumps are split by analysis of concavity regions for focal points. Further, classification of infected and non-infected RBCs has been made to grade MP precisely. The training and testing of the proposed approach on benchmark dataset with respect to ground truth data, yield 96.75% MP sensitivity and 94.59% specificity. Additionally, the proposed approach addresses the process with independent factors (RBCs morphology). Finally, it is an economical solution for MP grading in immense testing**.**

Keywords: Malaria Parasitaemia; Hybrid classifiers; Malaria Parasitaemia quantification and grading

1. Introduction

According to the world health organization (WHO) annual malaria report 2013 (WHO, 2013), malaria caused for 1300 young's life each day. Malaria is an infectious disease caused by genus Plasmodium, a parasite injected by female Anopheles into the human blood. The plasmodium attacks the red blood cells (RBCs) (Abbas et al., 2015; Mughal et al., 2018). The quantification of red blood cells (RBCs) infected by Plasmodium parasites shows the degree of malaria Parasitaemia (MP). It is of worth knowing to classify its life cycle stage specifically to treat it in a befitting manner. Light microscopy is an economical method to examine the blood for MP. Malaria Parasitaemia (MP), quantitatively measures the RBCs that are infected by the parasites is used to grade the severity of malaria (Tek et al., 2009). For this purpose visual inspection for the quantification of MP through light microscopy is still the most prevalent and economical practised method (WHO, 2009). However, the microscopic visual enumeration is laborious, time-consuming and highly subjective to the expert's knowledge (WHO, 2009). The microscopy test involves two types of smears that are thick and thin blood smears. The thick blood smear tests are mainly used for the detection of Plasmodium; presence and absence indicate positive and negative malaria respectively (Rehman et al., 2018a,b)

The skinny blood smear tests are used for distinct tests of malaria such as grading of MP, species and life-cycle classification. In fact, the thin blood smear ought to be tested under 100 to 200 windows even as diagnosing thru the microscopy of malaria (WHO, 2004). The wide variety of inflamed RBCs may be counted among 100 RBCs in each window. Physicians regularly ask for the thin blood smear test. However, in excessive levels of malaria underneath the microscope through visible inspection and quantification, which has proven to be too laborious, time-taking and the effects are regularly inaccurate because of the wide variety of on-going tests (Kettelhut et al., 2003; Abbas, 2015; Iqbal et al., 2017, 2018).

Nonetheless, tools and strategies so far reported in the literature are effective to the referred issues but cannot deal with dependent elements of RBCs morphology (Abbas et al., 2016; Husham et al., 2016) as shown in Figure 1. As an example, the circularity of RBCs isn't the regular case and previous studies conducted by Walliander et al (2013), Khan et al., (2011) and Guan et al., (2011) considered RBCs as round or elliptical shape and red color. Size, area, and red color RBCs are also risk factors (Tek et al., 2010; Sio et al., 2007; Savkare et al., 2011). However, a minor divergence from the proposed models of these studies will unexpectedly decrease precision and effectiveness. "Moreover, clumps and overlaps of RBCs are also serious issues that have not properly addressed in the past (Tek et al., 2011). Clump means to glue, and RBCs glued to each other in the form of long chains, an indication of iron deficiency (common in malaria) in the blood. Overlapped RBCs are formed due to inappropriate slide preparation. The mentioned problems affect the accuracy in terms of MP (Abbas, 2015; Iftikhar et al., 2017; Rahim et al., 2017a,b). MP is the percentage ratio of infected RBCs to all RBCs present on the slide (DPDS, 2002)" (Abbas et al., 2016; Norouzi et al., 2014; Saba et al., 2012).

$$\%\mathrm{MP} = \frac{\mathrm{iRs}}{\mathrm{aRs}} \times 100 \quad (1)$$

The iRs and aRs stand for infected RBCs number and total size RBCs in a window.

This research presents an automated approach for malaria-infected erythrocyte detection based on histogram features set. The image database is gathered from distinctive pathological clinics. The proposed research consists of four important levels, pre-processing, segmentation of the photo, feature is extraction primarily based on the histogram of various shade channels, choice of the functions and malaria-infected erythrocyte class is done the usage of specific classifiers including artificial neural networks (ANN), Support vector machine (SVM), *k*-nearest neighbor (*k*-NN) and naïve Bayes. In the pre-processing, the image undergoes through a median filter to enhance the quality of the microscopic images for further processing (Devi et al., 2016; Jamal et al., 2017). The proposed research utilizes a two-stage tree classifier using backpropagation feedforward neural networks (Ross et al., 2006). The SVM classifier is used for classification of the three life cycle stages Trophozoite, Schizont, Gametocytes (Sri et al., 2014; Fahad et al., 2018).

The rest of the paper is arranged into four main sections. Section 2 critically reviews the related state of the art, Section 3 presents the proposed research framework in detail, Section 4 exhibits experimental results, discussion and finally, Section 5 presents the conclusion.

2. Related work

A significant number of researchers addressed the automated MP quantification in the state of art. Most of the reported techniques focused to resolve issues of image luminance, low contrast, negative illumination and out of focus images in the pre-processing step. The most of the reported techniques employed pre-processing steps histogram equalization (HE) (Purwar et al., 2011, Mughal et al., 2017a,b; Sheeba et al., 2012), brightness preserving dynamic HE (BPD) (Mandal et al., 2010), smallest uni-fee phase assimilating nucleus (Smith and Brady , 1997) smoothing the image via Median clear out an area renovation via Laplacian (Savkare and Narote, 2011) and within the equal way, however for aspect renovation, some

researchers employed un-sharp masking (Mohapatra et al., 2011; 2010, 2011a,b). The underlying study considered image smoothing through the median filter of kernel size [3×3], high kernel size will remove the parasites, particularly at initial stages. However, an un-sharp masking is used for edges preservation of RBCs and parasites. The selection of these two methods has been made on the basis of their positive experimental performed on the standard dataset which contains 74 images (DPDx, 2002). Moreover, the maximum of those issues have been resolved due to the arrival of excellent imaging tools.

In step with the literature, we can widely group the tailored methodologies offered for automated malaria analysis (MP estimation) into deductive and inductive methods. The deductive technique is a pinnacle-down method beginning with the foreground and background separation, observed by means of RBC segmentation and subsequently, the parasites are studied. In contrast, inductive method is a bottom-up strategy. It's worth mentioning here that each of these techniques suffers from RBCs morphology dependent factors. However, the inductive technique is better for the motive that it has more freedom to pick out morphology independent elements. The color or intensities, shape, size, region, radius and all other morphology related elements of RBCs are noticeably variable factors among patients. Inside the identical way, we cannot rely upon the morphology of parasites for coloration. Furthermore, among the reported techniques, extreme dilemmas of clumps and overlapped RBCs are observed.

The deductive method is followed in few reported research conducted by Ahirwar et al.,(2012) and Khan et al., (2011), is seriously affected within the presence of dense clumped and overlapped RBCs. Ahirwar et al.,(2012) reported research is also dependent on the region of Granulometry for RBC length (only regular whilst RBC is healthful) estimation and there is no specification for clumped, overlapped RBCs. Linder et al., (2014) technique are dependent on the bimodal histogram and both mentioned studies have no clear approach on how to address the clumped and overlapped RBCs.

The research reports of Walliander et al., (2013); Guan, and Yan (2008); Berge et al., (2011) are dependent on the circularity of RBCs (detection through circular Hough transform). The circularity of RBCs is very sensitive and could be disrupted by the exertion of even slight pressure on the slide during preparation. For features extraction, the consideration of fixed area, radius, edges are the cases suited to normal RBCs but due to malaria and other diseases, these features alter frequently (DPDx,2013). However, these are adopted by a majority of the studies such as (Savkare and Narote, 2011; Khan et al., 2011; Sio et al., 2007).

Moreover, Khan et al., (2011) also counted the number of infected RBCs based on the number of parasites which is unacceptable in medical cases,(DPDx, 2013). Sio et al., (2007) stated that the infected RBC will be counted one, regardless of the number of parasites in it. Accordingly, Sio et al (2007) addressed the problems of clumped and overlapped RBCs by using the method developed by Kumar et al., (2006); however, in dense clumps of RBCs, the method affects the accuracy. The segmentation of RBCs based on nucleic approach exposes the problem, as RBCs have no nuclei and the studies consider the parasites as nuclei. The studies based on nucleic approach are seriously affected when the RBCs become really nucleated, such as when the RBCs lifespan is near the end or the RBCs are highly matured. The nucleic approach is followed by Kumarasamy et al., (2011); Khawaldeh (2013); Zoh et al., (2010) in the segmentation of RBCs. The segmentation based on chromatin dots offers no surety that on the basis of maximum and minimum intensity levels that they will be the same in all images and in addition, these studies are highly susceptible to noise (Saba et al., 2014; Rehman et al., 2011). On the same grounds, Somasekar, (2011); Makkapati and Rao, (2009) addressed the segmentation of the parasites. However, single RBCs can also have noisy chromatin dots, single dots aren't taken into consideration with the aid of experts as parasites, fake consequences might be pronounced and accuracy might be under threat (Saba, 2017; Saba et al., 2018a,b, Sadad et al., 2018).

3. Proposed Methodology

The proposed methodology consists of four main phases a) classification of parasites, b)segmentation of parasites, c) splitting clumps, and d) overlap of RBCs and MP grading, the research framework is exhibited in Figure 2.

3.1 Classification of Parasites

The main contribution of the current research is to classify the parasites according to the lifecycle phases that are: Ring stage, Trophozoite, Schizont, Gametocytes. The classification of plasmodium according to the life cycle phases in thin blood smear images is performed through SVM, *k*-NN, and Naïve Bayes classifiers. In the proposed framework, the input images are composed of infected RBCs, non-infected RBCs, and occluded RBCs. Prior to segmentation, separation of clumps or overlaps RBCs and finding infected, non-infected RBCs is performed then follows the classification of the parasites that belong to ring stage using multi-class support vector machine. There are multiple life cycle phases of parasite and the proposed research utilizes the multi-class SVM to classify the malaria parasites based on HOG and LBP features. The HOG descriptors are used for object detection. It counts the occurrences of the gradient orientation in the local portion of an image. The LBPs are used to encode the texture information and the LBP operator is computed in the local circular area that takes the difference of the central pixel with respect to the neighbour pixel.

3.2 Segmentation of Parasites

Narrow blood images are affected by various problems, to select features the data of those patients vary in every slide. The focus of the study is to come up with efficient technique to select features with more information and originates easily in every slide of MP and then keeping those features as valid justification which verifies the MP. According to literature, the researchers select features based on color information play a key role in the segmentation of parasite with Giemsa stain, which resulted in efficient accuracy. The criteria to select suitable features based on color are that the color with fewer amounts in the image will be the suitable color for the features extraction. For this, Gaussian mixture model (GMM)is used with machine learning technique (expectation maximization)to calculate the mean, co-variance and weights of the distributed colors in the images as mentioned in the following equation:

$$\left\{W_{k,}\mu_k, CV_k\right\}, \forall k_{copts} \in Color \qquad (2)$$

Where $\left\{W_{k,}\mu_k, CV_k\right\}$ represents weight, mean and covariance matrix of K *color* components respectively. Assign the pixel with normal probability distribution to Gaussian mixture model to determine the required color components as shown in equation (3).

$$P\left(k|f_x\right) = \frac{W_k N(f_x \mid \mu_k, CV_k)}{\sum_k W_k N(f_x \mid \mu_k, CV_k)} \qquad (3)$$

Where the color values are represented by k, weight withw_k, shown as $(\sum_{k=1}^{K} Wk = 1)$ and $f_x\left(W_1,\ldots,W_K; f_1,\ldots, f_K\right)$

The spatial variance is derived from vertical and horizontal divergence of the *k* color elements shown in equations 4, 5 respectively.

$$V_v(k) = \frac{1}{|Y|k}\sum_y \mathrm{P}(k|f_x)|y_v - M_v(k)|^2 \qquad (4)$$

where$M_v(k) = \frac{1}{|Y|_k}\sum_y \mathrm{P}(k|f_x)y_v$

$$V_h(k) = \frac{1}{|X|k}\sum_x \mathrm{P}(k|f_x)|x_h - M_h(k)|^2 \qquad (5)$$

where$M_h(k) = \frac{1}{|X|_k}\sum_x \mathrm{P}(k|f_x)x_h$

where x_h and y_v are x and y-coordinates of pixel value x, while $|y|\ k$ and $|x|\ k$ is set as$|y|_k = \sum_y P(k|f_x)$ and $|x|_k = \sum_x P(k|f_x)$respectively.

Total divergence of k color element is set as:

$$V(k) = V_v(k) + V_h(k) \tag{6}$$

Moreover, *V(k)*is normalized to values[0,1]

$$V(k) = \frac{\left(V(k) - \min_k V(k)\right)}{\left(\max_k V(k) - minkV(k)\right)} \tag{7}$$

Therefore sum of color special distribution feature with weights, *Fs(x,f)* are set as:

$$F_s(x, f) \propto \sum P(k|f_x).\left(1 - V(k)\right) \tag{8}$$

Color features with weights are also normalizing to values [0, 1].

The results are compared with other images marked by hematologist for visual inspection as ground truth and shown in Figure 3 and then verified from board of medical experts at Saidu Medical College Swat, KPK, Pakistan. Morphological feature dilation is applied to segmented images to enhance visibility.

At initial stage, the parasite is like threads and could spread over area of minimal 50 pixels (experiment are performed for more than 60 images in the total of 98), and small areas in the images are removed as noise. For further processes the images after segmenting parasites are converted to binary.

3.3 Clumps and Overlaps Splitting of Red Blood Cells (RBCs)

Accurate grading of malarial parasites relies on the precise quantification of cells (RBCs affected and unaffected). The clumps and overlapped cells affect the accuracy to correct quantification of RBCs (affected and unaffected).To overcome the splitting problem, pre-processing steps are performed to check the existence of clumps, overlapped cells and to separate them from RBCs.

3.3.1 Searching for Clumps and Overlaps of Red Blood Cells

The twofold checking for the existence of clumps and overlaps of RBCs are median expansion and median area checking. For this, the convex hull of every RBC is traced in current window using equation (9), area and elongation of that convex hull using equations (10) and (11), respectively. By using these values we find out variance normalization of all RBCs via experimentation, the variance is 0.2 in an area with 0.5 elongation that represents overlaps of RBCs existence.

$$\sum_{i=1}^{|X|} \alpha_i x_i \mid (\forall_u : \alpha_i \geq 0) \,\&\, \sum_{i=1}^{|X|} \alpha_i = 1 \tag{9}$$

where |x| represents a finite area,x_i is position of |x| while $\alpha_i$is the given weight of x_i, sum of all weights = 1 due to normalization.

$$Area_{RBC} = No.ofPels \tag{10}$$

Total number of Pels(Pixels)shows the convex hull of the RBCs.

$$Elongation_{RBC} = \frac{L_{RBC}}{B_{RBC}} \tag{11}$$

Major axis is represented with L_{RBC} and minor axis with B_{RBC} of each(RBCs) convex hull.

$$\sigma^2 = \frac{(X-\mu)^2}{N} \qquad (12)$$

In above equation X illustrates area in one instance while elongation in other, N is the distribution terms.

3.3.2 Dissociation of Single and Clumps or Overlaps of RBCs

Following the decision of either clumps or overlaps of RBCs existence, next step is the segmentation of clumps from RBCs. To segment clumps, twofold checking method as described in equation (9), (10) and (11) is applied. When there are irregularities in data values, and finding out central tendency we calculate median as it is best for central measure. If the result obtained by dividing each convex hull area of RBC is almost equal to it will be considered as single RBCs and included in its mask. Alternatively, if obtained results are greater than 1, it will be considered as multi-RBCs and are take in as multi-RBCs mask. Then, we pass mask of single RBCs into the pixel *IDX*_list of input images and acquired the image of single RBCs. Similarly, we pass mask of multi-RBCs to get image for clumps or overlaps of RBCs. Furthermore, we use elongation in place of area to perform second check. This process of dissociation is represented in Figure 4.

3.3.3 Splitting the Clumps and Overlaps of RBCs

Following, dissociation of single and clumps or overlaps RBCs, images of clumps and overlaps RBCs is furthermore taken for splitting into single divided RBCs. For splitting the clumps or overlaps of RBCs, we first determine the concavity regions in a simple way through taking the convex hull of each clumped or overlapped object and subtracting the actual region from its convex hull. After, determining the concavity regions, the boundary of the concavity regions are analyzed for finding the concavity points, which always lie at the last point of $1/4^{th}$ part of the boundary in a clockwise direction while the same in an anti-clockwise direction. Taking these points as P1 and P2 and drawn a line in between them will split the RBCs. This process is continued until the maximum number of RBCs in each clump and overlap. We gather the segmented single RBCs as a result in isolated output image. The line between P1 and P2 is drawn through the concept of digital differential analyzer (DDA) line drawing algorithm of computer graphics as shown in Figure 5.

The proposed technique is applied to clogged RBCs images for splitting purpose that yields efficient results. This concept is derived from watershed transformation algorithm, afflicted by over and under-segmentation in the clog. If RBCs are more than four it requires significant processing time in contrast to the techniques proposed in this research. A sample of experimental results is depicted in Figure 6.

3.4 Malaria Parasitaemia Grading

For MP grading, the following steps were performed in the proposed technique.

3.4.1 Imposition of Segmented parasites

Segmented parasites in the first step are inflicted on single RBC after the splitting process if clumps and overlaps exist, else it will be applied directly after segmentation of MP. The inflictions of parasites are needed for identifying the affected RBCs and add up their number to approximate the percentage of MP. This infliction process of parasites is the simple addition of two binary images i.e. the single RBCs while other has segmented parasites as they are in opposition to drop the impact of noise and any other relics. The visual results of this process are presented in Figure 7.

3.4.2 Identifying Infected RBCs

As all the processes are done on RBCs like separation and splitting, now we need to identify the infected RBCs for counting purpose. For that we use one particular quality of infected RBCs, consideration of outside boundaries of each of the RBCs and draw red circle from those that are infected on the basis of the parent boundary having child boundary. If there is no child boundary then the RBCs are count as unaffected RBCs. From medical point of view, an RBC is count as affected of malarial parasites if they have plasmodium presence in it. However, if they have more Plasmodium in the cell then it is counted as one affected RBC. The results of this process are presented in Figure 8.

3.4.3 Segmentation of affected RBCs

The concept being employed to identify RBCs is also employed to segment affected RBCs. Accordingly, the binary image with same size is taken having all RBCs (affected and unaffected). Later, areas with (1's) or more than 1 are segmented as affected RBCs. This process is shown in Figure 9 while the results are exhibited in Figure 10,11, 12.

3.4.4 Counting affected and unaffected RBCs

Moreover, following segmentation, counting the affected and unaffected RBCs is a simple deal. We use bwlabel function of Matlab for automatic counting. Counting results are presented in Figures 13 and 14.

3.4.5 Estimation of Percentage MP

The percentage ratio of affected RBCs and all RBCs that is available in each window on a slide is MP. WHO recommends that (Tek et al., 2003; WHO, 2009), the percentage ratio should be calculated on the microscopic examination of 100-200 windows, having100 RBCs each. The total sum of affected and unaffected RBCs and the percentage ratio of MP can be calculated by the formula given in equation (1).

3.4.6 Grading of Malaria Parasitaemia

According to Iyar, (2013); Hänscheid et al., (2000) and Homel (1998), the percentage ratio of MP could be estimated by examining 100 to 200 windows. Finally, it will be graded as one of the levels reported in Table 1.

Finally, each MP grads as presented in Table 1. Furthermore, for testing purposes, we took 40000 RBCs in each slide and calculated the results based on presumption with the result of each image in a single window. Table 2 presents the statistical analysis after examining one narrow blood smear image

4. Experimental results and analysis

To validate the proposed research framework, qualitative and quantitative analysis is performed.

4.1 Hybrid classifiers

To classify MP (according to the life cycle phases in thin blood smear images) SVM, *k*-NN and Naïve Bayes classifier based on HOG and LBP are used for classification. The equation 12 and 13 show the sensitivity and specificity of the classifiers based on HOG and LBP features. The comparison of sensitivity and specificity for HOG and LBP based features for different classifiers is given in Table 3 and Table 4.

$$Sensitivity = \frac{T_p}{T_p + F_n} \qquad (12)$$

$$Specificity = \frac{T_n}{T_n + F_p} \quad (13)$$

4.2 Ground Truth Data Preparation

To prepare the ground truth dataset, DPDx (2013) images printed as forms and distributed among three Pathologists. Each form has a single image of thin blood smear and its manually estimated statistics and marking of the parasites in the image. Another panel of three medical experts verifies these forms. The data collection has been made at the Department of Pathology, Saidu Medical College, Saidu Sharif Swat, KPK, Pakistan.

4.3 Inter-Rater Agreement

The collected data is first checked for inter-raters reliability agreement through a variation of Cohen's Kappa (Bi-Raters) called Fleiss' Kappa through equation 14.

$$\kappa = \frac{P' - P_e^{'}}{1 - P_e^{'}} \quad (14)$$

where, $P' = \frac{1}{N}\sum_{i=1}^{N} P_i$ *and* $P_e' = \sum_{j=1}^{N} p_j^2$, N=total number of subjects and *i,j=1,2,3,....N,k* represents subjects and categories respectively. The Fleiss' Kappa calculation for the collected data is K=0.96

4.4 Quantitative evaluation of the proposed clumps and overlaps RBCs splitting technique

Pearson's correlation coefficient is employed to analyze the relationship of counting RBCs automatically (after splitting clumps and overlaps of RBCs) and manually made by the experts. The relationship between these two variables is presented in Figure 14. Additionally, confusion matrix based Precision, Recall and F-measure are conducted via equation 15, 16 and 17, the results are available in Table 5.

$$Precision = \frac{T_p}{T_p + F_p} \quad (15)$$

$$\mathrm{Re}call = \frac{T_p}{T_p + F_n} \quad (16)$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (17)$$

Here T_p= Correctly counted as RBCs, T_n=Correctly counted as non-RBCs, F_p=Incorrectly counted as RBCs, and F_n=In-correctly counted as non-RBCs

The achieved precision, recall and F-measure by counting the RBCs after splitting the occluded RBCs with the proposed technique are 0.9618, 0.9803 and 0.9704 respectively.

The graph in Figure 15, demonstrates a strong positive correlation (R^2=0.98 on 74 degrees of freedom). Furthermore, it shows that the strength of the relationship between two variables i.e. manually counted RBCs by the experts and automatically counting carried out in the proposed technique.

4.5 Statistical Analysis of the proposed framework

Pearson's correlation coefficient is employed to analyze the relationship between manually and automatically estimated percentage MP presented in Figure 15. Confusion matrix based sensitivity and

specificity are calculated via equ. 18 and 19. The main strength of the proposed technique is the correct acceptance of the infected RBCs as infected through sensitivity and correct rejection of non-infected as non-infected RBCs through specificity.

$$Sensitivity = \frac{T_p}{T_p + F_n} \quad (18)$$

$$Specificity = \frac{T_n}{T_n + F_p} \quad (19)$$

Finally, the proposed research framework achieved sensitivity is 0.967 and the specificity is 0.9460.

The graph presented in Figure 16 shows a strong positive correlation between the percentages of MP estimated by the experts manually and automatically by the proposed framework.

4.6 Performance comparison in state of art

It is hard to compare results to available techniques in the literature, mainly due to a different setup and dataset employed in the experiments. However, the proposed framework is analyzed and compared to current techniques reported in the state of art using the same dataset; a few comparisons are exhibited in Table 6.

5. Conclusion

This paper has presented a complete research framework for Malaria Parasitaemiain (MP) grading in thin blood smear digital images. The proposed algorithms are implemented to classify life phases of malaria parasite using multi-class SVM, *k*-NN, and Naïve Bayes. The multi-class SVM produced best classification results using HOG and LBP features. The proposed research is tested on grounds truth from accuracy and efficiency points of view. The color of the parasites is the only feature, which is the same in all thin blood smear digital images. The only issue faced due to noise effect on images that are eliminated using noise removal technique. The enhanced reported accuracy is due to appropriate and independent segmentation of the occluded RBCs. Finally, the comparison in the state of art exhibited high accuracy 96.75% on DPDx benchmark dataset with 94.59% specificity as presented in Table 6.

Highlights

The research presents inductive machine learning approach for segmenting malaria parasites through the adaptive algorithm. The quantification accuracy of RBCs is improved, splitting clumps of RBCs via analysis of concavity regions for focal points.

Acknowledgement

"This work was supported by the Machine Learning Research Group; Prince Sultan University Riyadh; Saudi Arabia [RG-CCIS-2017-06-02]; The authors are grateful for this support and equipments."

References

Abbas, N. Saba, T. Mohamad, D. Rehman, A. Almazyad, A.S. Al-Ghamdi, J.S. (2016) Machine aided malaria parasitemia detection in Giemsa-stained thin blood smears, Neural Computing and Applications, pp. 1-16, doi:10.1007/s00521-016-2474-6.

Abbas, N. Mohamad, D. Abdullah A.H., Saba, T., Al-Rodhaan, M., Al-Dhelaan, A. (2015). Nuclei segmentation of leukocytes in blood smear digital images. Pakistan journal of pharmaceutical sciences 28(5), pp. 1801-1806

Abbas, N. (2015) Segmentation of parasites and splitting of red blood cells approaches for malaria parasitemia grading, PhD thesis, Universiti Teknologi Malaysia, Skudai Malaysia.

Ahirwar, N., S. Pattnaik, and B. Acharya, Advanced image analysis based system for automatic detection and classification of malaria parasite in blood images. International Journal of Information Technology, 2012. 5(1): p. 59-64.

Berge, H., et al. Improved red blood cell counting in thin blood smears, IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011.

Devi, S.S., et al., Malaria infected erythrocyte classification based on a hybrid classifier using microscopic images of thin blood smear. Multimedia Tools and Applications, vol. 77(1): p. 631-660, 2016.

Díaz, G., F.A. González, and E. Romero, A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. Journal of Biomedical Informatics, vol. 42(2): p. 296-307, 2009.

DPDx, DPDx, Laboratory Identification of Parasites, Centers of diseases Control and Prevention, D.a.m.b.C.s.D.o.P.D.a.M. (DPDM), 2002, Web site developed and maintained by CDC's Division of Parasitic Diseases and Malaria (DPDM): 1600 Clifton Rd. Atlanta, GA 30333, USA

DPDx. Determination of Malaria Parasitemia.2013 [cited 2014 7-4-2014]; Available from: http://www.cdc.gov/dpdx/resources/pdf/benchAids/malaria/Parasitemia_and_LifeCycle.pdf.

DR.B.K.Iyar. Malaria Diagnostics.2013 [cited 2014 7-04-2014]; Available from: http://www.slideshare.net/iyerbk/malaria-diagnostics.

Esteva, A., et al., Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639): p. 115, 2017.

Fahad HM, Ghani Khan MU, Saba T, Rehman A, Iqbal S. (2018) Microscopic abnormality classification of cardiac murmurs using ANFIS and HMM. Microsc Res Tech. pp. 1–9. https://doi.org/10.1002/jemt.22998.

Guan, P.P. and H. Yan. Blood cell image segmentation based on the Hough transform and fuzzy curve tracing. IEEE International Conference on Machine Learning and Cybernetics (ICMLC), 2011.

Hänscheid, T., E. Valadas, and M. Grobusch, Automated malaria diagnosis using pigment detection. Parasitology Today, vol. 16(12): p. 549-551, 2000.

Husham, A., Alkawaz, M. H., Saba, T., Rehman, A., Alghamdi, J.S. (2016) Automated nuclei segmentation of malignant using level sets, Microscopy Research and Technique, vol. 79(10), pp. 993-997, doi. 10.1002/jemt.22733.

Iftikhar, S. Fatima, K. Rehman, A. Almazyad, A.S., Saba, T. (2017) An evolution based hybrid approach for heart diseases classification and associated risk factors identification, Biomedical Research 28 (8), pp. 3451-3455

Iqbal, S. Ghani, M.U. Saba, T. Rehman, A. (2018). Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN). Microsc Res Tech. pp.1–9. https://doi.org/10.1002/jemt.22994

Iqbal, S., Khan, M. U. G., Saba, T., Rehman, A. (2017). Computer assisted brain tumor type discrimination using magnetic resonance imaging features. Biomedical Engineering Letters, vol.8(1),pp.5-28, doi. 10.1007/s13534-017-0050-3.

Jamal A, Hazim Alkawaz M, Rehman A, Saba T. (2017) Retinal imaging analysis based on vessel detection. Microsc Res Tech. 2017;00:1–13. https://doi.org/10.1002/jemt.

Homel M., H.M.G., Malaria. 9th ed. Microbiology and Microbial. Infections, ed. T.W.s.M.a.M. Infections. 1998, Arnold: Topley& Wilson's.

Kettelhut, M., et al., External quality assessment schemes raise standards: evidence from the UKNEQAS parasitology subschemes. Journal of clinical pathology, vol. 56(12): p. 927-932, 2003.

Khan, M.I., et al., Content Based Image Retrieval Approaches for Detection of Malarial Parasite in Blood Images. International Journal of Biometrics and Bioinformatics (IJBB), vol. 5(2): p. 97-108, 2011.

Khawaldeh, B.A.I., Developing a Computer-based Information System to Improve the Diagnosis of Blood Anemia, in Department of Computer Information Systems Faculty of Information Technology. Middle East University: Amman, Jordan. p. 116, 2013.

Kumar, S., et al., A rule-based approach for robust clump splitting. Pattern Recognition, vol. 39(6): p. 1088-1098, 2006.

Kumarasamy, S.K., S. Ong, and K.S. Tan, Robust contour reconstruction of red blood cells and parasites in the automated identification of the stages of malarial infection. Machine Vision and Applications, vol. 22(3): p. 461-469,2011.

Linder, N., et al., A Malaria Diagnostic Tool Based on Computer Vision Screening and Visualization of Plasmodium falciparum Candidate Areas in Digitized Blood Smears. PloS one, vol. 9(8): p. e104855, 2014.

Makkapati, V.V. and R.M. Rao. Segmentation of malaria parasites in peripheral blood smear images. . IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.

Mandal, S., et al. Segmentation of blood smear images using normalized cuts for detection of malarial parasites. IEEE Annual India Conference (INDICON), 2010.

Mughal, B. Muhammad, N. Sharif, M. Rehman, A. Saba, T. (2018) Removal of pectoral muscle based on topographic map and shape-shifting silhouette, BMC Cancer, https://doi.org/10.1186/s12885-018-4638-5

Mughal, B. Muhammad, N. Sharif, M. Saba, T. Rehman, A. (2017a) Extraction of breast border and removal of pectoral muscle in wavelet domain, Biomedical Research, vol.28(11), pp. 5041-5043.

Mughal, B. Sharif, M. Muhammad, N. Saba, T. (2017b). "A novel classification scheme to decline the mortality rate among women due to breast tumor," Microscopy Research and Technique, doi. 10.1002/jemt.22961.

Mohapatra, S. and D. Patra. Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images. IEEE International Conference on Systems in Medicine and Biology (ICSMB), 2010

Mohapatra, S., D. Patra, and K. Kumar.Blood microscopic image segmentation using rough sets. IEEE International Conference on Image Information Processing (ICIIP), 2011.

Mohapatra, S., et al. Fuzzy based blood image segmentation for automated leukemia detection. IEEE International Conference on Devices and Communications (ICDeCom), 2011.

Norouzi, A. Rahim, M.S.M, Altameem, A. Saba, T. Rada, A.E. Rehman, A. and Uddin, M. (2014) Medical image segmentation methods, algorithms, and applications IETE Technical Review, vol.31(3), pp. 199-213, doi. 10.1080/02564602.2014.906861.

Opoku-Ansah, J., et al., Wavelength Markers for Malaria (Plasmodium Falciparum) Infected and Uninfected Red Blood Cells for Ring and Trophozoite Stages. Applied Physics Research, vol. 6(2): p.47-56, 2014.

Prinyakupt, J. and C. Pluempitiwiriyawej, Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers. Biomedical engineering online, vol.14(1): p. 63-79, 2015.

Purwar, Y., et al., Automated and unsupervised detection of malarial parasites in microscopic images. Malaria journal, vol. 10(1): p. 364, 2011.

Rahim, M.S.M, Norouzi, A. Rehman, A. and Saba, T. (2017a) 3D bones segmentation based on CT images visualization, Biomedical Research, vol.28(8), pp.3641-3644

Rahim, M.S.M, Rehman, A. Kurniawan, F. Saba, T. (2017b) Ear biometrics for human classification based on region features mining, Biomedical Research, vol.28 (10), pp.4660-4664

Ravikumar, S., Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine. Artificial cells, nanomedicine, and biotechnology, vol. 44(3): p. 985-989, 2016.

Rehman, A., Saba, T. (2014) Neural networks for document image preprocessing: state of the art, Artificial Intelligence Review Vol. 4(2), pp. 253-273, doi. org/10.1007/s10462-012-9337-z

Rehman, A. Abbas N. Saba, T. Mahmood, T. Kolivand, H. (2018a) Rouleaux red blood cells splitting in microscopic thin blood smear images via local maxima, circles drawing, and mapping with original RBCs, Microscopic research and technique, doi.org/10.1002/jemt.23030.

Rehman, A. Kurniawan, F. Saba, T. (2011) An automatic approach for line detection and removal without smash-up characters, The Imaging Science Journal, vol. 59(3), pp. 177-182, doi. 10.1179/136821910X12863758415649.

Rehman, A. Abbas, N. Saba, T. Mehmood, Z. Mahmood, T. Ahmed, K.T. (2018b) Microscopic malaria parasitemia diagnosis and grading on benchmark datasets, Microscopic Research and Technique, doi.org/10.1002/jemt.23071

Ross, N.E., et al., Automated image processing method for the diagnosis and classification of malaria on thin blood smears. Med BiolEng Computing, vol. 44(5): p. 427-36, 2006.

Saba T, Al-Zahrani S, Rehman A. (2012) Expert system for offline clinical guidelines and treatment Life Sci Journal, vol.9(4), pp. 2639-2658.

Saba, T. (2017) Halal food identification with neural assisted enhanced RFID antenna, Biomedical Research vol. 28(18), pp. 7760-7762.

Saba, T., Rehman,A. Mehmood, Z., Kolivand, H., Sharif, M.(2018a) Image Enhancement and Segmentation Techniques for Detection of Knee Joint Diseases: A Survey, Current Medical Imaging Reviews Vol. 13, doi. 10.2174/1573405613666170912164546.

Saba,T., Rehman, A., Al-Dhelaan, A., Al-Rodhaan, M. (2014) Evaluation of current documents image denoising techniques: a comparative study , Applied Artificial Intelligence, vol.28 (9), pp. 879-887, doi. 10.1080/08839514.2014.954344

Saba T, Bokhari STF, Sharif M, Yasmin M, Raza M. (2018b) Fundus image classification methods for the detection of glaucoma: A review, Microsc Res Tech. doi: 10.1002/jemt.23094.

Sadad, T. Munir, A. Saba, T. Hussain, A. (2018) Fuzzy C-means and region growing based classification of tumor from mammograms using hybrid texture feature, Journal of Computational Science, Vol. 29, pp. 34-45.

Savkare, S. and S. Narote, Automatic detection of malaria parasites for estimating parasitemia. International Journal of Computer Science and Security (IJCSS), vol. 5(3): p. 310, 2011.

Sheeba, F., et al. Detection of Plasmodium Falciparum in Peripheral Blood Smear Images. Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). 2013

Sio, S.W., et al., Malaria Count: an image analysis-based program for the accurate determination of parasitemia. Journal of microbiological methods, vol. 68(1): p. 11-18, 2007.

Smith, S.M. and Brady, J.M. International Journal of Computer Vision (1997) 23: 45. https://doi.org/10.1023/A:1007963824710

Somasekar, J., Computer Vision for Malaria Parasite Classification in Erythrocytes. International Journal on Computer Science and Engineering, vol. 3(6): p. 2251-2256, 2011

Sri Widodo, Wijiyanto et al., texture analysis to detect malaria tropica in blood smears image using support vector machine. ISSN: 2349-2163.IJIRAE., vol.1(8). p. 301-306, 2014.

Tek, F.B., A.G. Dempster, and I. Kale, Computer vision for microscopy diagnosis of malaria. Malaria Journal, vol. 8(1): p. 153, 2009.

Tek, F.B., A.G. Dempster, and I. Kale, Parasite detection and identification for automated thin blood film malaria diagnosis. Computer Vision and Image Understanding, vol. 114(1): p. 21-32, 2010.

Tek, F.B., A.G. Dempster, and I. Kale.Malaria Parasite Detection in Peripheral Blood Images.in BMVC. 2006.

Walliander, M., et al., Automated segmentation of blood cells in Giemsa stained digitized thin blood films. Diagnostic Pathology, 8(Suppl 1): p. S37, 2013

Wang, H., H. Zhang, and N. Ray. Clump splitting via bottleneck detection. 18th IEEE International Conference on Image Processing (ICIP), 2011

WHO, Annual Malaria Report, W.H. Organization, Editor. 2013.

WHO, Basic Malaria Microscopy Part 1.2nd ed. 2004.1-88.

WHO, Basic malaria microscopy: Part I. Learner's guide: Part II. Tutor's guide. 1st ed. 2009: World Health Organization.

Zou, L.H., et al. Malaria cell counting diagnosis within large field of view. IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2010.

Figures legends

Figure 1: Different shapes of Red Blood Cells, a) Normal Red Blood Cells, and b) presents irregular shapes of Red Blood Cells.

Figure 2: Framework of the proposed technique to determine the degree of MP in thin blood smears

Figure 3: Parasite segmentation with the proposed technique (slight dilation is applied for clear visibility). The first two columns (left) consist of the original input image and parasites segmented images with the proposed technique while the last column (right) contains ground truth data marked by medical experts.

Figure 4: Presents the separation process, a) Presents input original image, b) Presents the binary image of the original, c) presents the single RBCs and d) presents the separated occluded RBCs

Figure 5: Overall process after separation of clumps and overlaps of RBCs from single RBCs' image A) Presents the image of Clumps and overlaps of RBCs, B) Clumps and overlaps of RBCs separated from image in A), C) presents Convex hull of objects presented in B), D) Presents the concavity regions extracted by subtracted image presented in B) from image presented in C), E) presents the concavity points in clock and anticlockwise direction and F) presents the results of DDA line drawing between the two opposite concavity points.

Figure 6: Clumps and overlaps of RBCs splitting through the proposed technique, A), C), E) and G) are original images while B), D), E) and H) are the results obtained by drawing the lines with DDA between opposite concavity points to obtain the actual cleaved number of RBCs

Figure 7: Parasite imposition process with proposed technique on images having clumps and overlaps of RBCs. A), D) and G) present original images, B), E) and H) have cleaved clumps or overlaps of RBCs and imposition of parasites on them and C), F) and I) present the imposition of parasites on single RBCs

Figure 8: Parasite imposition on images having no clumps or overlaps of RBCs. A) and C) are input images while B) and D) are the resultant after imposition of parasites

Figure 9: Identification of infected RBCs. A) and C) present original images while B) and D) present infected RBCs highlighted with the Red boundaries

Figure 10: The process of infected RBCs segmentation. A) is an original binary image, B) is the empty image with areas highlighted as the infected RBCs area, C) present infected RBCs, resulted through proposed technique and finally D) contains all non-infected RBCs.

Figure 11: The segmentation process of the infected RBCs in slide images having clumps and overlaps of RBCs. A), G) and M) represent the input images to this module while B), H) and N) present the infected RBCs in the cleaved RBCs if exist, C), I) and O) present non-infected RBCs the cleaved RBCs, D), J) and P) present both infected and non-infected RBCs (single) images while E), K) and Q) present the segmentation of the infected RBCs presented in D), J) and P). Finally, the F), L) and R) presents the non-infected RBCs in the single RBCs

Figure 12: The segmentation process of the infected RBCs in slide images having all single RBCs. A) and D) represent the original input images, B) and E) present the infected RBCs while C) and F) present the non-infected RBCs.

Figure 13: Counting Process results in images having clumps and overlaps of RBCs. A) presents cleaved infected RBCs, B) presents non-infected cleaved RBCs, C) presents single infected RBCs and D) presents non-infected single RBCs

Figure 14: Counting process results slides having RBCs without clumps and overlaps. A), D) and G) are original input images, B), E) and H) are images labelled as infected RBCs, while images C), F) and I) are labelled as non-infected RBCs

Figure 15. Graph present correlation between manually and automatically (by splitting clumps and overlaps of RBCs via Concavity Regions analysis for concavity points to draw lines between them with DDA) counted RBCs

Figure 16. Correlation between automatic and manual MP Estimation

Tables legends

Table 1: Percentage of Malaria Parasitaemia Grading

Table 2: Statistics of single thin blood smear image

Table 3: Comparison of classifiers based on HOG Features

Table 4: Comparison of classifiers based on LBP features

Table 5: Confusion Matrix

Table 6: Performance comparison in state of art