An automated nuclei segmentation of leukocytes from microscopic digital images

Naveed Abbas¹, Tanzila Saba², Zahid Mehmood³, Amjad Rehman⁴*, Naveed Islam⁵ and Khawaja Tehseen Ahmed⁶
¹Department of Computer Science, Islamia College University, Peshawar, KPK, Pakistan
²College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia
³Department of Software Engineering, University of Engineering and Technology, Taxila, Pakistan
⁴College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia
⁵Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan

Abstract: Leukemia is a life-threatening disease. So far diagnosing of leukemia is manually carried out by the Hematologists that is time-consuming and error-prone. The crucial problem is leukocytes’ nuclei segmentation precisely. This paper presents a novel technique to solve the problem by applying statistical methods of Gaussian mixture model through expectation maximization for the basic and challenging step of leukocytes’ nuclei segmentation. The proposed technique is being tested on a set of 365 images and the segmentation results are validated both qualitatively and quantitatively with current state-of-the-art methods on the basis of ground truth data (manually marked images by medical experts). The proposed technique is qualitatively compared with current state-of-the-art methods on the basis of ground truth data through visual inspection on four different grounds. Finally, the proposed technique quantitatively achieved an overall segmentation accuracy, sensitivity and precision of 92.8%, 93.5% and 98.16% respectively while an overall F-measure of 95.75%.

Keywords: Leukocytes, cytoplasmic granules, granulometry measure, K-means clustering

INTRODUCTION

Leukocytes are the important components of blood for the immune system in the human body. The types of leukocyte include basophil, eosinophil, lymphocyte, monocyte and neutrophil which are differentiated from each other on the basis of cytoplasmic granules, the size of the cell and shape of the nucleus (Fahad et al., 2018; Iqbal et al., 2018). The various types of leukocytes nuclei are the areas of interests in various diseases like Leukemia, folate deficiency, liver diseases and cancer. The biomedical technologists through visual microscopic inspection could note the disorder. The manual diagnosis involves three burden key factors, labor, time and jeopardy of error on the part of biomedical technologists (Rehman et al., 2018a,b,c; Mughal et al., 2018; Saba et al., 2018). An alternate solution in this modern era is the automatic diagnosis through digital image processing techniques and procedures. The automatic diagnosis will not only assure the precision but also trim the time, reduce the labor and also make up the deficiency in the knowledge of biomedical technologists (Saba et al., 2018; Sadad et al., 2018). Automated medical imaging diagnosis grasps the attention of the scientists in a recent decade because of its efficient and accurate assistance in almost all areas of biomedicine. In the same connection, the automatic study of microscopic images not only trims the time of the biomedical technologists, but also reduces their labors and efforts, but makes up their unawareness and paucity with reference to the new perception. The segmentation of leukocytes nuclei in microscopic images is important and laborious (Iqbal et al., 2017). The main object of this study is the automatic segmentation of the nuclei of the leukocytes from microscopic thin blood smear digital images in an accurate and efficient way as compared to the current state-of-the-art. This accurate segmentation of nuclei will help the biomedical technologists in better visualization and understanding for the classification of leukocytes and in the diagnosis of various diseases like blood cancer or acute leukemia, infection, liver diseases, folate deficiency, lymphoblastic leukemia as mentioned in the latest study of (Genschaft et al., 2013) and many other diseases (Saba et al., 2012). Leukocytes are the most important components of blood and play a key role in case of foreign bodies attack and are considered protector against various diseases. The types of Leukocytes are broadly divided into granulocytes and agranulocytes on the basis of presence or absence of granules in the cytoplasm. The sub-types of granulocytes having granules in their cytoplasm are basophil; eosinophil and neutrophil, while in agranulocytes there is lacking granules in the cytoplasm including lymphocyte and monocyte. The characteristics of all these five types of leukocytes are important and are described next and Fig. 1 (a) presents the diagrammatic representation while (b) shows the micrographic images of the five types of leukocytes.

*Corresponding author: e-mail: rkamjad@gmail.com
In order to diagnose the diseases caused by the leukocytes disorder, the microscopic study is deemed very instrumental. The biomedical technologists could distinguish the five different types of leukocytes via their cytoplasmic granules, size and shape of the cell and the nucleus (Chan et al., 2010; Norouzi et al., 2014; Rad et al., 2013) However, the manual methods involve the burden of labor, time, high expertise for such activities and the chances of errors are obvious because of visual inspection under a microscope. In this connection, the alternate solution is an automatic classification of leukocytes for the diagnosis of various fatal diseases in which, the basic and challenging step is the nucleus segmentation (Jamal et al., 2017; Husham et al., 2016).

The experiments are performed on a set of 365 images obtained from Calgary laboratory services, school of medicine Cairo University and experimented with the microscopic images marked by experts in the correct identification of leukocyte in (Abbas et al., 2015; Mohamed et al., 2012; Mohamed, 2012) on the MATLAB file exchange with written permission of use for research purposes (Mohamed, 2012). The details of each type of leukocyte in the dataset are presented in Table 1.

The automatic methods used for segmentation of leukocytes segmentation could be broadly divided into two types on the basis of their approaches i.e. global or local.

In global approaches, the image is directly converted to binary or grayscale and then the nuclei segmentation took place (Mughal et al., 2017a). The local approaches directly started after image pre-processing for illumination correction and noise reduction and then the segmentation of nuclei of leukocytes took place. The latter approach in our own proposed method is adopted. The majority of the previous approaches achieved high segmentation accuracy, but lack of universality. The study made by (Dorini et al., 2007; Mughal et al., 2017b) employs watershed transform for the segmentation of nuclei of leukocytes in grayscale images and then also extract the cytoplasm on the Granulometry measure of the size of leukocytes. They also used some morphological operators, but the results are validated by visual inspection on image set of only two images as the problem we faced in our study that there are a lot of variations in the size and shape of nuclei. The work by the same authors with the only change in the number of images in the results is carried out in (Dorini et al., 2012; Mughal et al., 2017c). However, both of the studies, (Dorini, et al., 2012, 2007) are highly dependent on the size and shape of nuclei of the leukocytes. In the study of (Mohamed and Far, 2012), the authors used edge-based

Table 1: Details of images dataset (Mohamed, et al., 2012; Mohamed, 2012)

<table>
<thead>
<tr>
<th>Types of Leukocyte</th>
<th>Number of Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monocyte</td>
<td>19</td>
</tr>
<tr>
<td>Neutrophil</td>
<td>271</td>
</tr>
<tr>
<td>Eosinophil</td>
<td>40</td>
</tr>
<tr>
<td>Basophil</td>
<td>02</td>
</tr>
<tr>
<td>Lymphocyte</td>
<td>33</td>
</tr>
<tr>
<td>Total images used</td>
<td>365</td>
</tr>
</tbody>
</table>

Segmentation is the process of grouping the pixels on the basis of some common properties to form objects in a digital image (Rahim et al., 2017a). There are many approaches used to get efficient segmentation and still the research is passing through the rebuilding process. Segmentation may be threshold based, edge-based, region growing based, clustering or binning based, graph cut based and color based. The segmentation approaches are not limited to the list and many other approaches and techniques have also existed, but the main theme of all these approaches is same and that of grouping pixels on the basis of some common properties to form the object of interest in the digital images (Rahim et al., 2017b).
segmentation through a weighted Sobel operator in the grayscale images. The results are presented with two images while also some calculation had been made for finding the number oflobsin the nucleus. The results on the experimented image are good, but there is a lack of universality. TheGlobal threshold Otsu approach and morphological operator opening is used in the work of(Mohamed and Far, 2012) and create a nucleus mask but this approach will highly dependent on the size ofthenuclei of the leukocytes, having highly variations but still the author compared their work with three other methods and achieved an overall segmentation accuracy of 90.1%. The work carried out by (Mohamed, 2012), the authors used additional step of Gram Schmidt Orthogonalization for improving the processing time but we experience the same time as in previous study (Mohamed, et al., 2012) of the same authors. The same authors carried out the solutions in (Mohamed and Far, 2012) on the same basis with the addition of minimum filter of kernel size 3x3 and achieved an overall segmentation accuracy of 80%. The authors mentioned the problem of overlapped nuclei in (Mohamed and Far, 2012), (Mohamed, et al., 2012) but the actual problem is the variation in the size and shape of nuclei of leukocytes. Ramesh et al., (2012), used two-step classifications of leukocytes. First, input RGB image is converted to JCV colour-space by cropping manually and finally segmented the nuclei of the leukocytes. In the second step, they classified the leukocytes. Rezatofighi et al., (2009) employed the Gram Schmidt Orthogonalization with difference that they further classified the leukocytes on the basis of segmented nuclei with the mentioned method and achieved overall segmentation accuracy of 93.02%. However, the method adopted is affected by the variation in shape and size of nuclei of leukocytes. The study made by Theera-Umpoon and Dhompomgsa, (2007) for the classification of leukocytes on the basis of nuclei of the leukocytes. The features of nuclei are extracted to segment the nuclei. The segmented nuclei are classification by two methods and features of nuclei of leukocytes are extracted in four ways i.e. Pattern spectrum and two Granulometry moments of the pattern spectrum to extract two features while the other two features of the nuclei are extracted from the areas of nuclei and the pattern spectrum peaks of each nucleus. Thus the four features extraction technique for each leukocyte’s nucleus segmentation is employed, but the processing time is on the stack. For the segmentation of leukocytes’ nuclei, Ramoser et al., (2006) recommended the nucleus separation and background suppression through k-means clustering with fix i.e. three number of color classes to bin the background, red blood cells and nucleus pixels. However, problems are present in the sense that the number of color classes in the microscopic images is more as compared to three. For the background suppression, a multiplicative approach of the RGB color-space is used. The contour-based approach with some other techniques is adopted in the work of(Clocksin, 2003) for the segmentation of nuclei of leukocytes in fluorescence microscopic images. The overlapped nuclei problem is resolved first and latter contour-based approach is used for the segmentation of nuclei of leukocytes but the results only show the input fluorescence microscopic image. In the study of (Nilsson and Heyden, 2001), the authors claimed that their techniques worked well on solitary nucleus of leukocyte as well as on cluster nuclei of leukocytes but the method they exercised for splitting the joint nuclei is watershed which worked well in case of slight overlap nuclei and is inappropriate in case of dense cluster nuclei. Following splitting the nuclei of leukocytes in the background, having erythrocytes is removed. The results are validated through ground truth data with visual inspection only. Madhloom et al., (2010) performed segmentation on the basis of automatic global thresholding Otsu approach by first stretching the contrast linearly in one copy of the original input image in grayscale and equalized the histogram in another grayscale copy of the original input image and applied minimum filter as in the work of Mohamed and Far, (2012). According to Madhloom, et al., (2010) and Mohamed and Far, (2012), the minimum filtered is the addition of the (addition and subtraction) two copies in grayscale. Further, they claimed that it will remove all the other components of the blood and only the nuclei of the leukocytes will be left. The results are validated by comparison with the experts marked nuclei images of leukocytes and achieved segmentation accuracy max up to 98.1% on the single leukocyte nucleus. The initial work performed by the authors of proposed a unique and robust approach of color nuclei segmentation of leukocytes through a simple method of suppressing the high color values in the three channels because of the nucleus of the leukocytes always in the darkish blue or purple colors due to the staining process of Gamesa. However, the method shows high results and achieved an overall segmentation accuracy of 96.04% at which the current state-of-the-art is standing, but badly affected when the leukocytes are overlapped. The study of color processing of leukocytes microscopic images in the YCbCr color-space gives a unique identity to the nuclei of the leukocytes in terms of colors which is mentioned in the work (Abbas and Mohamad, 2013) and is simple for segmentation especially in the case of binning or clustering algorithms.

The further paper is organized into five main sections, proposed methodology, material & methods, experimental results, analysis & discussion and final section concludes the research.

Proposed methodology
The basic purpose of this study is to make the segmentation process automatic with reference to nuclei regarding leukocytes from the microscopic thin blood smear, digital images in an accurate way. We propose a
An automated nuclei segmentation of leukocytes from microscopic digital images

novel method in the area for the segmentation of nuclei leukocytes in digital microscopic images of blood by considering universal factors in the microscopic images of leukocytes. The proposed method is the Gaussian Mixture Model for the segmentation of leukocytes nuclei on the basis of its unique color in the whole image when it is stained with Giemsa. The color of the leukocytes becomes purple or darkish blue which we consider our region of interest and another interesting factor about the nuclei of leukocyte is that of its large size which distinguishes it from any noise in the image. We applied the Gaussian Mixture Model because it is one of the most widely used and versatile probabilistic models. Moreover, the unique numerical and statistical property of Gaussians and the linearity of the combination make it an ideal method. Another advantage of the GMM is flexibility in terms of memory because it can store only three values for each color component. An overview of the proposed methodology is presented in fig. 2.

Pre-processing

In the pre-processing step, we first convert the input RGB microscopic image of blood to YCbCr color-space through the equation 1. The reason behind selecting this specific color-space is that the perception of human about images is in three parts i.e. brightness, hue and saturation thus the possible candidates color-spaces which can divide the color into brightness, hue and saturation are YCbCr, YIQ, HIS and HSV. Among the listed color-spaces we select the YCbCr because it preserves the luminance as detailed information than any other color-space in the category as also mentioned by (Kang et al., 2011). Moreover, the use of YCbCr color-space reduces the processing time of 4.98 seconds per slide processed by the Saliency-based approach as because of distinguishing colors of the parasites and other components. Second, we reduce the noise through Median Filter as mentioned in equation 2, of kernel size of [7x7] only in the Cb component in the YCbCr color-space because it trims the processing time and enhances the results because the nuclei of leukocytes are more visible than the other two components.

$$\begin{bmatrix}V \end{bmatrix} = \begin{bmatrix} 16 & 128 & 255 \\ 128 & 255 & 255 \\ 255 & 255 & 255 \end{bmatrix} \begin{bmatrix} V \end{bmatrix}$$

$$\begin{bmatrix} C\mathrm{b} \end{bmatrix} = \begin{bmatrix} 128 & 255 & \vdots \\ 128 & 255 & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \{V \} \end{bmatrix}$$

where V = [r, g, b], \{V \} = [r, g, b],

$$\begin{bmatrix} C\mathrm{r} \end{bmatrix} = \begin{bmatrix} 255 & 255 & \vdots \\ 255 & 255 & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \{V \} \end{bmatrix}$$

where V = [r, g, b], \{V \} = [r, g, b],

Leukocytes’ Nuclei Segmentation. Segmentation is the process of grouping pixels to form objects on the basis of some similar characteristics. The concept has a lot of variations in terms of the approach to carry it out. We consider the groups of pixels on the basis of their intensities and weighted the color on the basis of Gaussian Mixture Model through expectation maximization in the single Cb component of the YCbCr color-space. Then through factor analysis, we found the normalized weights of the groups of pixels sharing the same intensity values horizontally and vertically of the Cb component of the provided YCbCr Color-space. And on the basis of these weights, we select the group of pixels of those intensities which appears less distributed or lower in weights in the whole image. We consider the selected groups of pixels as our object of interest or the salient objects. In the next section, we explained all these in detail.

Gaussian Mixture Model. Assume that we are using K-Gaussians in the Mixture Model presented as:

$$P(I_l | f_l) = \sum_{k=1}^{K} w_k P(I_l | f_{lk})$$

(3)

where I represent intensities in a group vector. \(\{W_k \}\) are the weights given as \((\Sigma_{l=1}^{N} w_l = 1)\) and \(f_{lk} \) are the Gaussian density parameterized by \(f_{l}(I_l \in \mu_k, \Sigma_k)\) of the \(k\)-th component given in equation 4.

$$P(I_l | f_l) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (I_l - \mu_k)^T \Sigma_k^{-1} (I_l - \mu_k)\right]$$

(4)

We initially started with selecting groups of pixel intensities on the basis of random centroids, next provided these groups or clusters to the Expectation Maximization for refinement. The EM has two steps expectation and maximization and converges through the maximum log likelihood. The maximization of the weights means and co-variance took place according to the equations (5), (6) and (7) respectively.

$$w_i^{new} = \frac{1}{N} \sum_{l=1}^{N} P(I_l | f_{lk})$$

(5)

$$\mu_k^{new} = \frac{\sum_{l=1}^{N} P(I_l | f_{lk}) I_l}{\sum_{l=1}^{N} P(I_l | f_{lk})}$$

(6)

$$\Sigma_k^{new} = \frac{\sum_{l=1}^{N} P(I_l | f_{lk}) (I_l - \mu_k)(I_l - \mu_k)^T}{\sum_{l=1}^{N} P(I_l | f_{lk})}$$

(7)

Where, \(N\) is the total number of intensities vector, number of pixels in a group. The expectation step \(P(I_l | f_{lk})\) is the probability that Gaussian \(i\) fits the pixels \(I\) in the \(f_{lk}\) component, given data \(f_{lk}\) presented in the equation 8.

$$P(I_l | f_{lk}) = \frac{w_k P(I_l | f_{lk})}{\sum_{l=1}^{N} w_l P(I_l | f_{lk})}$$

(8)

The maximization converges at the maximum log likelihood as presented by equation (9) or (10).

$$\log G(f_{lk}) = \log \prod_{l=1}^{N} P(I_l | f_{lk})$$

(9)

$$\log P(f_{lk} | I_l, \mu_k, \Sigma_k) = \sum_{l=1}^{N} \log P(I_l | f_{lk})$$

(10)
Further, we calculate the horizontal and vertical variances of the intensities in each group to map them in the Cb component of the YCbCr color-space as given in equations (11) and (12) in the same way done by (Liu et al., 2011) for the three channels of the RGB color-space.

\[ v_{HCb}(int) = \frac{1}{|H_{Cb}|} \sum_{i \in H_{Cb}} P(\{I_{Cb}\}_i) \cdot I_{Cb} \cdot \left(1 - \frac{1}{|H_{Cb}|} \sum_{i \in H_{Cb}} P(\{I_{Cb}\}_i) \cdot I_{Cb} \right) \]  

(11)

\[ v_{VCb}(int) = \frac{1}{|V_{Cb}|} \sum_{j \in V_{Cb}} P(\{I_{Cb}\}_j) \cdot I_{Cb} \cdot \left(1 - \frac{1}{|V_{Cb}|} \sum_{j \in V_{Cb}} P(\{I_{Cb}\}_j) \cdot I_{Cb} \right) \]  

(12)

where \(x\) and \(y\) are horizontal and vertical coordinates of pixels along x-axis and y-axis in the image respectively while their horizontal and vertical intensities are given in equation (13) and (15).

\[ |H_{Cb}| = \sum_{i \in H_{Cb}} P(\{I_{Cb}\}_i) \]  

(13)

\[ |V_{Cb}| = \sum_{j \in V_{Cb}} P(\{I_{Cb}\}_j) \]  

(14)

The total variance is the sum of the horizontal and vertical variances of intensities in the Cb component of the YCbCr color-space and is given in equation (15).

\[ v_{TCb} = v_{HCb} + v_{VCb} \]  

(15)

Further, we normalize the total variance of the Cb component in the YCbCr color-space i.e. \(v_{TCb}\) to [0,1] through equation (16).

\[ v_{TCb} = \frac{v_{TCb} - M_{v_{TCb}}}{M_{v_{TCb}}} \]  

(16)

Finally, we calculate the minimum weight of the intensities among the different groups as stated in equation (17).

\[ g(x,y) \propto \frac{1}{P(\{I\}_y)} \cdot (1 - P_{TCb}) \]  

(17)

The use of the Cb component of YCbCr not only reduced the computation time but also increased the appearance of the nuclei in the microscopic image as compared to RGB because it can calculate the variance in the three channels and increase the processing time as compared to proposed approach. The next section of this paper presents algorithms on how to use the above equations and segment the nuclei of the leukocytes.

Algorithm for segmentation of the nuclei of the Leukocytes
Input: Microscopic RGB Color Image of Blood
Output: Nuclei Segmented RGB Image

Image Pre-processing
1. Read Image I(x,y,3) \(\leftarrow\) Microscopic RGB color Image of blood
2. Convert I(x,y,3) from RGB color-space to YCbCr color Space as \(\Gamma(x,y,3)\) using equation (1).
3. //Separate the channels
4. \(A_{Cb}(x,y) \leftarrow \Gamma(:, :, 2)\)

5. // For noise reduction apply a median filter
6. Apply Median Filter of kernel size of 3*3 to the \(A_{Cb}(x,y)\)

Nuclei Segmentation
1. G(x,f) \(\leftarrow\) Call function Mask(\(A_{Cb}(x,y)\))
2. Convert to pure Binary G(x,f) to G'(x,f) using OTSU Function
3. Convert to RGB the binary G(x,f) which is the nuclei segmented image of the nuclei of the leukocytes
4. Outputs:
5. Binary G(x,f)
6. RGB G'(x,f)
7. The original RGB image I(x,y,3)
8. Original YCbCr image \(\Gamma(x,y,3)\)
9. Original \(A_{Cb}(x,y)\) image of the Cb component of YCbCr color-space

Algorithm for Function Mask (Image)
Input: Cb\(^{th}\) component of the YCbCr Color-space
Output: Masked in which the Nuclei of the leukocytes are mapped
Calculate the initial centroids using K-means algorithm.
1. Refine the calculated initial cluster’s weights, means and covariance’s using equations (5), (6) and (7) and GMM equation (4) through Expectation Maximization machine learning algorithm(Sylvain, 2009) and stop at Maximum log like hood stated in equations (9) and (10) and specifically in our case as we tested experimentally to stop at In-2.
2. Further, we calculate the horizontal, vertical and total variances using equations (11) to (15).
3. The total variance is normalized using equation (16).
4. Finally, the mask map is calculated using equation (17).
5. The Mask Map is returned to the calling program.

The special k-means clustering used for the single component of Cb of the YCbCr is designed by Jose Vicente Manjon Herrera (Manjon-Herrera, 2005) for the grayscale images. We find the initials centroids of clusters of various intensities groups. Then refine these initial weights, means and co-variances with Gaussian Mixture Model through expectation maximization algorithm provided by (Sylvain, 2009). Further, the total variance is calculated to map the co-variances in the image horizontally and vertically and found the total variance of each cluster to find the class of the intensities which is less spread in the image is considered to be the area of interest. The mask map returned is further refined to make it pure binary through global Otsu function to make the brightest regions as bright while the light-dark into the pure dark.

MATERIALS AND METHODS

Digital microscopic imaging data
The digital image data set of 365 color images are obtained from the study of (Dorini, et al., 2012, 2007),
An automated nuclei segmentation of leukocytes from microscopic digital images

shared on MATLAB file exchange with the written permission of reproduction for research and study purposes. The digital image dataset as stated by (Mohamed, et al., 2012; Mohamed, 2012) is acquired from normal samples of bloodstained with standard Gimesa-right technique for regular light microscope using a 100x objective lens through a CCD color camera connected to the microscope and the video output form is fed to video grabber card to capture color images with (640x480 pixels) VGA resolution. As mentioned by (Mohamed, et al., 2012) and (Mohamed, 2012) the images are taken under the supervision of Amid Abdullah expert pathologist from Calgary laboratory services, Calgary Canada. Moreover, the images are marked manually by two medical experts Dr Yaser Hasan and Dr Mohamed Albasher from the school of medicine, Cairo University. We also duly checked the image dataset through Dr. Bakht Bilan, Associate Professor, Peshawar Medical College, KPK and Dr. Ikram UL Mabood, Associate Professor, Saidu Medical College, Swat, KPK, Pakistan both declare the marking as correct. Further, Dr. Ikramur Rehman (D-Pham) Coordinator, NSRSP, Swat Chapter, KPK, Pakistan is also consulted for the image dataset and relative marking of nuclei and he also declares that the dataset of 365 images is taken accurately and the ground truth data (manually marked images by two medical experts) are marked correctly.

Fig. 2: Proposed research framework
Fig. 3: Leukocyte’s type “Neutrophil” nucleus segmentation (a) Original Image (b and c) Segmented automatically (d) manually marked and verified by three medical experts.

Fig. 4: Leukocyte’s type “Lymphocyte” nucleus segmentation (a) Original Image (b and c) Segmented automatically (d) Manually marked and verified by three medical experts.

Fig. 5: Leukocyte’s type “Monocyte” nucleus segmentation (a) Original Image (b and c) Segmented automatically (d) Manually marked and verified by three medical experts.

Fig. 6: Leukocyte’s type “Eosinophil” nucleus segmentation (a) Original Image (b and c) Segmented automatically (d) Manually marked and verified by three medical experts.
An automated nuclei segmentation of leukocytes from microscopic digital images

**Fig. 7:** Leukocyte's type “Basophil” nucleus segmentation (a) Original Image (b and c) Segmented automatically (d) Manually marked and verified by three medical experts.

**Fig. 8:** Binary images comparisons of different methods with Manual marked images by experts Column(a) Method developed by (Madhloom, et al., 2010) (b) Methods developed by (Mohamed, et al., 2012) and (Mohamed and Far, 2012) (c) Proposed Method (d) Ground truth images manually marked by medical experts

**Fig. 9:** presents the qualitative comparisons of the proposed method with current state-of-the-art method (Mohamed and Far, 2012) on the basis of inaccurate segmentation, rows (a) Original RGB microscopic digital images of blood, (b) Images marked automatically and stated as inaccurate (c) Ground truth Images manually marked by medical experts while the in-set images are the segmentation results of the proposed method of this study and (d) Images marked automatically by the proposed method in the original images.
Fig. 10: presents the qualitative comparisons of the proposed method with current state-of-the-art methods (Mohamed, et al., 2012), (Mohamed and Far, 2012), (Mohamed and Far, 2012), on the basis of inaccurate segmentation as experienced by this study by implementing the methods of (Mohamed, et al., 2012), (Mohamed and Far, 2012), rows (a) Original RGB microscopic digital images of blood, (b) Images marked automatically by the proposed method in the original images (c) Images marked automatically (e) Ground truth images manually marked by medical experts.

**Comparison of the Segmentation Accuracy of the Proposed Method with current State-of-the-Art Methods**

Fig. 11: Evaluation of different types of leukocytes’ nuclei segmentation accuracy on the basis of comparing the segmented nuclei with the ground truth images marked by medical experts: different colors of line shows different types of leukocytes while the horizontal axis represents current state-of-the-art methods and vertical axis represents the segmentation accuracy of these methods for each type of leukocytes.
Fig. 12: Evaluation of total segmentation accuracy on the basis of comparing the segmented nucleus with the ground truth images marked by medical experts: The single line with multiple spots shows different types of methods’ total segmentation accuracy while the horizontal axis represents current state-of-the-art methods and the vertical axis represents the total segmentation accuracy of these methods.

Fig. 13: Sensitivity characteristics of segmented nuclei compared with the ground truth data (manually marked images) by medical experts: The single line with multiple diamond symbols represents the average sensitivity of each type of leukocyte nuclei, first compared pixel by pixel with the ground truth data marked by medical experts to find the sensitivity of each automatically segmented nuclei image (proposed method) and then the average of each type is calculated to present the average sensitivity of that particular type of leukocyte. The high recall or sensitivity rate is that of Basophile type of leukocyte while the lower precision rates are recorded for Eosinophil and Neutrophil types of leukocytes but still, it is the promising result and in the range of 90’s.

Fig. 14: Precision characteristics of segmented nuclei compared with the ground truth data (manually marked images) by medical experts: The single line with multiple diamond symbols represents the average precision characteristics of each type of leukocyte nuclei, first compared pixel by pixel with the ground truth data marked by medical experts to find the precision of each automatically segmented nuclei image (proposed method) and then the average of each type is calculated to present the average precision of that particular type of leukocyte. The high precision rate is that of Basophile type of leukocyte while the lower precision rates are recorded for Eosinophil and Neutrophil types of leukocytes but still, it is the promising result and in the range of 90’s.
Creation of ground truth dataset
The digital image dataset provided by and (Mohamed, 2012) has also a dataset of the ground truth images marked by two medical experts Dr Yaser Hasan and Dr Mohamed Albasher from the school of medicine, Cairo University is taken through an Hp Laser jet color C P6015 dn (Q3932A) printer. The image dataset is further verified by specialists for proof marking of nuclei again and they found the previously marked image dataset as accurate. The proof observer, in the same way, marked 365 microscopic digital images as marked by the previous Medical experts. Moreover, in the evaluation, the marked image dataset is converted to binary with the automatic thresholding mentioned by (Ridler and Calvard, 1978) and matched with the automatically segmented nuclei of leukocytes by the proposed method for calculation of the segmentation accuracy, sensitivity or recall, precision and F-measure of the recall and precision.

Experimental results
In this section, the results of experiments performed on 365 images dataset are presented and the performance of the proposed method is evaluated both quantitatively and qualitatively to validate its potentiality.

Qualitative evaluation of the detection results
As the proposed method performs nuclei segmentation of the leukocytes thus in the qualitative evaluation we performed the experiments on all types of leukocytes (i.e. five types Basophil, Eosinophil, Monocyte, Lymphocyte, Neutrophil) then compare the results through ground truth with visual inspection with two other state-of-the-art...
methods in the area and finally presents the results of the images with the proposed methodology which are stated by other studies as drawback in their proposed methodologies (Mohamed and Far, 2012), (Mohamed, et al., 2012). In fig. 3,4,5,6,7 the different types of leukocytes are segmented and compared with the manually marked images verified by the above mentioned three Medical experts.

**Qualitative results comparison of the proposed method with ground truth manually marked images of nuclei verified by three medical experts**

In this category we performed the experiments on different types of leukocytes and in each fig., the (a) represents the original input RGB image which is a microscopic digital image taken with a CCD camera as mentioned in the introduction section in detail. (b) presents the segmented nucleus of the leukocyte (c) represents the marking of the segmented nucleus in the original RGB image to view its accuracy through visual inspection while the (d) is the manually marked image by two experts and further verified by three experts mentioned in the introduction section in detail.

**Qualitative results comparison of the proposed method with current state-of-the-art methods on the basis of visual Inspection with the ground truth**

In this category of comparison, we compare the proposed method with three other current state-of-the-art methods on the basis of visual inspection with ground truth images manually marked by two medical experts and verified by three medical experts as the details are mentioned in the introduction section. Moreover, the columns from left to right (a) represents the method of (Madhloom, et al., 2010), (b) the results of the methods developed by (Mohamed, et al., 2012) and (Mohamed and Far, 2012), (c) represents the results of the proposed method of this study and (d) presents the ground truth images marked by two medical experts and verified further by three other medical experts in the area.

**Qualitative Results comparison of the proposed method with the current state-of-the-art method on the basis of the inaccurate segmentation stated by the study**

In this category, we compare the results of the proposed method with the current state-of-the-art method (Mohamed and Far, 2012) on the basis of the inaccurate segmentation mentioned in their study. The rows in fig. 9 presents the results as the images in row (a) are the Original RGB Microscopic digital images of blood, (b) are the images marked with the segmentation method of (Mohamed and Far, 2012) , (c) are the ground truth images marked by the Medical experts while in-set images in row (c) images presents the segmentation of nuclei of leukocytes with the proposed method and finally row (d) are the images in which the nuclei are marked in the original images automatically as segmentation results of the proposed method.

**Qualitative evaluation and performance comparison of the proposed method with current state-of-the-art methods on the basis of results experienced by the current study**

In this category we compare the proposed method with the current state-of-the-art methods. The coding for implementation of all these studies is available on the MATLAB file exchange for the purpose of research with written permission from the authors of the mentioned studies

**Quantitative evaluation and performance comparison with current state-of-the-art methods on the basis of segmentation accuracy**

In this category, we compare the results of the proposed method with six current state-of-the-art methods on the basis of segmentation accuracy through the metric mentioned in equation (18). The calculation of equation (18) is made on the basis taking complements of binary automatically marked image and in the same way the manually marked images and compared according to the stated equation (18).

\[
S_{\text{automatic}} = \frac{S_{\text{automatic}} \cap S_{\text{manual}}}{\text{MAX}(S_{\text{automatic}}, S_{\text{manual}})}
\]

where, \( S_{\text{automatic}} \) = Segmentation accuracy multi-methods
\( S_{\text{automatic}} \) = Automatically segmented nuclei images
\( S_{\text{manual}} \) = Manually marked images by an expert.

**Quantitative evaluation and performance measure of the proposed method on the basis of ground truth data**

In this category of results, we compared the experimental results of the proposed method with the ground truth data to validate the proposed method using different metrics.

**Metrics evaluation:** The evaluation metrics we used for the quantitative analysis of the proposed method and are presented in equations (19), (20) and (21).

\[
\text{Sensitivity or Recall} = \frac{\text{BIS}_{\text{auto}} \cap \text{BIM}_{\text{manual}}}{\text{BIS}_{\text{auto}}}
\]

(19)

\[
\text{Precision} = \frac{\text{BIS}_{\text{auto}} \cap \text{BIM}_{\text{manual}}}{\text{BIM}_{\text{manual}}}
\]

(20)

\[
F - \text{Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

(21)

Where, \( \text{BIS}_{\text{auto}} \) = Binary Image Segmented Automatically, \( \text{BIM}_{\text{manual}} \) = Binary Image Marked Manually by medical experts or Ground Truth Image. The comparison is made on pixel by pixel basis in the automatically marked image to the manual marked image. The binary image segmented automatically pixels are considered as the retrieved pixels while the binary image marked manual pixels are considered as relevant pixels. In this way the Recall and, precision are calculated.
The mentioned metrics are first applied to the individual image in each type of leukocyte and then collectively presented through its average for that type of leukocyte.

Computation of Metrics Evaluation
First the automatically segmented image is converted to binary having the leukocytes’ nuclei as white means 1’s then, in the same way, the manually marked images by the expert are converted to binary having the leukocytes’ nuclei as white means 1’s than both the images i.e. the automatically segmented images and the manually marked images are compared pixel by pixel using the metrics motioned in equation (19) and (20) and then its F-measure is calculated on the basis of the outcomes of equation (19) and (20). The results are presented in the next section in tabular form as well as in the graphical form for simplicity and understanding the validation of the proposed method.

Analysis and discussion
The main problem in the segmentation of leukocytes’ nuclei is the variation in its morphology, while also highly disturbed in cases of any disease like leukaemia, anemia and other disorder. The only universal factor of the nuclei of the leukocytes is that of its unique color of the nuclei when the blood is stained with Giemsa, thus taking the idea from the manual work of the experts we made our algorithm and is based on the principle mentioned. However, there is a problem of noise, but we reduce it while the larger size when disturbed by any diseases or disorder, will still be considered the larger one as compared to the small areas as detected as noise in the microscopic image. The results of the proposed method are compared both qualitatively and quantitatively. The qualitative analysis is made on the basis of four different aspects. First, on the basis of each type of leukocyte’s nucleus segmentation and compared with ground truth data (manually marked images by experts) through visual inspection. Second, on the basis of comparing the nuclei segmentation with current state-of-the-art methods on the basis of ground truth data through visual inspection, third on the grounds that the inaccurate segmentation results mentioned by the current state-of-the-art methods with

<table>
<thead>
<tr>
<th>Types of Leukocytes</th>
<th>Basophile</th>
<th>Eosinophil</th>
<th>Lymphocyte</th>
<th>Monocyte</th>
<th>Neutrophil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>Basophile</td>
<td>Eosinophil</td>
<td>Lymphocyte</td>
<td>Monocyte</td>
<td>Neutrophil</td>
</tr>
<tr>
<td></td>
<td>0.9091</td>
<td>0.9928</td>
<td>0.9491</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9444</td>
<td>0.9621</td>
<td>0.9531</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9447</td>
<td>0.9902</td>
<td>0.9669</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9042</td>
<td>0.9883</td>
<td>0.9444</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9729</td>
<td>0.9750</td>
<td>0.9740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.9350</td>
<td>0.9816</td>
<td>0.9575</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The code is not open source for some period of time and will be soon available as open source for research purposes, while the experimentation images with ground truth data are available free with the studies of (Mohamed and Far, 2012), (Mohamed, et al., 2012) on the site referenced as with written permission to use for research purposes while the images verified by three experts are available in the this study.

Analysis and discussion
The main problem in the segmentation of leukocytes’ nuclei is the variation in its morphology, while also highly disturbed in cases of any disease like leukaemia, anemia and other disorder. The only universal factor of the nuclei of the leukocytes is that of its unique color of the nuclei when the blood is stained with Giemsa, thus taking the idea from the manual work of the experts we made our algorithm and is based on the principle mentioned. However, there is a problem of noise, but we reduce it while the larger size when disturbed by any diseases or disorder, will still be considered the larger one as compared to the small areas as detected as noise in the microscopic image. The results of the proposed method are compared both qualitatively and quantitatively. The qualitative analysis is made on the basis of four different aspects. First, on the basis of each type of leukocyte’s nucleus segmentation and compared with ground truth data (manually marked images by experts) through visual inspection. Second, on the basis of comparing the nuclei segmentation with current state-of-the-art methods on the basis of ground truth data through visual inspection, third on the grounds that the inaccurate segmentation results mentioned by the current state-of-the-art methods with
An automated nuclei segmentation of leukocytes from microscopic digital images

ground truth data through visual inspection and finally on the basis of inaccurate results found by this study while implementing the current state-of-the-art methods and compared the results of the proposed method with ground truth data through visual inspection. The quantitative analysis is made on image set of 365 images and on the basis of four statistical measures the segmented nuclei are compared with current state-of-the-art methods and ground truth data (manually marked by medical expert).

The overall segmentation accuracy achieved is 92.8%, which is compared to the current state of the art methods and found among them as the highest. The nuclei segmented with the proposed method when compared on the basis of pixel by pixel comparison with the ground truth data manually marked by two medical experts and verified by three further medical experts achieved a sensitivity of 93.50%, while a high precision of 98.16%, which are the highest rates among current state-of-the-art methods while the F-measure of the recall and precision results in 95.75%. However, the proposed method is computationally more expensive as compared to current state-of-the-art methods, but still promising as because the approximate processing time per slide is 4 seconds on the hardware mentioned above.

CONCLUSION

The segmentation of the leukocyte’s nucleus is a challenging problem which is previously addressed by many studies and achieved the desired goals in favourable conditions means if there is a little variation in light and any disorder in the slide formation the results are badly affected accuracy wise. The proposed method is designed in such a way that it will not be affected drastically by the mentioned factors as compared to the current state-of-the-art methods. The proposed method results in accurate leukocytes nuclei segmentation and outperforms state-of-the-art methods, in precision, recall and F-measure aspects. The results show that the proposed method is highly robust and could be used for noisy images captured using low exposure light. Hence, the basic and challenging step is solved with the promising results. The future work will focus on the identification of Leukocytes type identification using neural networks or pattern matching techniques. We hope it will be an easy task since the leukocyte’s nuclei are already segmented in such a manner that the morphology of the nucleus is not disturbed and could easily be identified by any mentioned techniques.

ACKNOWLEDGEMENT

This work was supported by Artificial Intelligence and Data Analytics (AIDA) Lab Prince Sultan University Riyadh Saudi Arabia. Authors are thankful for the support. The authors are grateful for the support and equipment.

REFERENCES


Naveed Abbas et al


Pak. J. Pharm. Sci., Vol.32, No.5, September 2019, pp.2123-2138


