

## A new approach for fully automated segmentation of peripheral blood smears

Abdullah Elen <sup>1,\*</sup>, Muhammed Kamil Turan <sup>2</sup>

<sup>1</sup>Department of Computer Technologies, Vocational School of T.O.B.B. Technical Sciences, Karabük University, Karabük, Turkey

<sup>2</sup>Department of Medical Biology and Genetics, Faculty of Medicine, Karabük University, Karabük, Turkey

### ARTICLE INFO

#### Article history:

Received 19 August 2017

Received in revised form

27 October 2017

Accepted 22 November 2017

#### Keywords:

Blood cell segmentation  
 Automatic blood analyses  
 Peripheral blood smear  
 Graham scan  
 Medical image processing

### ABSTRACT

Peripheral blood smear is microscopically examining technique for blood samples from patients by painting special dyes in clinic laboratories. Blood diseases can be diagnosed by examining morphology, numbers and percentages of leukocyte, erythrocyte and thrombocyte cells in blood samples. However, this method is a considerably time-consuming process and requires an evaluation performed by a hematology specialist. It is not often provided a definitive assessment due to the expert's clinical experience and judgment during review. Although there are considerable studies about the segmentation of blood smear images in the literature, there is no method to segment all blood cells. In this study, a new segmentation algorithm is proposed, which automatically extracts leukocyte, erythrocyte and thrombocyte cells from peripheral blood smear images. Purpose of this study here is to make highly accurate and complete blood count. The algorithm treats each image as a universal set and represents each object in the image as a subset as a result of the applied operations. In the developed method, leukocytes and thrombocytes achieve better success than other studies. However, it has been observed that the average success rate of stacked erythrocytes decreases. Statistical tests of the developed method were performed using 200 blood smear images in experimental studies. According to the obtained results, it is seen that high accuracy (leukocyte 99.86%, thrombocyte 98.4%, erythrocyte 93.4%) and precision (leukocyte 94.77%, thrombocyte 90.14%, erythrocyte 95.88%) were achieved in all three blood cells.

© 2017 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

Segmentation can be defined as the subdivision of an image, each of which has meaningful and similar properties (gray level, color, texture, brightness, contrast, etc.). It is one of the most important (Kaur et al., 2015) and difficult stages of digital image processing (Dhiman and Talwar, 2014). Although many methods and approaches are proposed in the literature, there is no absolute solution that can still be applied on all image types and produce excellent results (Kaur et al., 2015; Agrawal and Xaxa, 2014). For this reason, it stays as a problem that cannot be solved in image processing and computer vision (Agrawal and Xaxa, 2014). Besides general-purpose images, segmentation of

medical images is also a very important field of study. Medical image segmentation has become very important with the development of complex medical imaging modalities, especially with the ability to produce high resolution two-dimensional (2D) and three-dimensional (3D) images in large quantities (Sumengen et al., 2002). Medical image segmentation refers to the separation of two or three-dimensional images into corresponding cluster of biological constructs. As in general-purpose images, there is no universal method or algorithm for segmenting medical images (Sharma and Aggarwal, 2010). The use of manual methods (Madhloom et al., 2012) in medical image analysis brings with many drawbacks like increasing time cost (Mohamed and Far, 2012; Nazlibilek et al., 2014) and leading to calculation errors. In recent years, various studies have been carried out by a significant number of scientists to remove these problems. Some of them are automatic image processing methods. Although the concept of automating is relative, it means that the system

\* Corresponding Author.

Email Address: [aelen@karabuk.edu.tr](mailto:aelen@karabuk.edu.tr) (A. Elen)

<https://doi.org/10.21833/ijaas.2018.01.011>

2313-626X/© 2017 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

needs fewer parameters while running and the system can manage its decision-making processes with its own internal calculation methods without requiring any intervention.

Blood smear slides are usually prepared manually by clinical laboratory staff. For this reason, an ideal blood film cannot always be obtained. The main reasons for this are the inability to take blood drops as much as the patient, the pushing of the spreader slide in a jerky manner, the spreading angle that is not being constant, the physical deformations in the slides and it is not clean enough. Such situations affect the performance of the automated blood analysis.

In recent years, many studies have been performed for the segmentation of blood smear images. These studies are generally based on White Blood Cells (WBC) or Red Blood Cell (RBC) segmentation (Dey et al., 2015). Madhloom et al. (2012) have developed a method of segmenting lymphoblast cells with high accuracy using microscopic blood images. They have combined morphological reconstruction with the color properties of the cell to differentiate lymphoblast cells from other blood cells. They stated that they achieved 100% success with 180 blood images. Liu et al. (2012) used mean shift clustering, color space transformation and nucleus mark watershed operation methods for peripheral blood smearing and segmentation of white blood cells on bone marrow images in different illumination environments. Tareef et al. (2016) developed a method for automatically segmenting nucleus and cytoplasm of white blood cells into five main classes based on color and tissue healing. They used Discrete Wavelet Transform (DWT) and morphological filters to make the cytoplasm more prominent in the segmentation process and to remove the details of the cell nucleus. Accordingly, they achieved better results than other color-based methods. Jiang et al. (2006) developed a new WBC segmentation technique using scale-space filtering and watershed clustering methods. They performed scale-space filtering method to obtain nucleus of the WBC and watershed clustering on 3D HSV histogram to remove cytoplasm. They stated that the proposed method provided much better performance than the previous studies. Salem (2014) transformed microscopic blood images of RGB format into  $L * a * b$  color space for segmentation of white blood cells and then used color components  $a$  and  $b$  as parameters for K-means clustering technique. Sadr et al. (2010) successfully transformed blood images into YCbCr color space for automatic segmentation of WBC cell nucleus, dividing each Cb color component into Y color components to make the cell nucleus more prominent. In order to improve the performance of the method, they applied Max and Min filters to the Cb and Y components. Yang et al. (2014) used the S (Saturation) component and the B (Blue) component in the RGB color space for segmentation of leukocytes. According to these color

components, they managed to determine the cell nucleus and cytoplasm boundaries of the WBC by applying “AND” and “XOR” logical process on two different binary images of an RGB image. Arslan et al. (2014) used color and shape based algorithms for WBC segmentation. The study was performed in two steps as transformation and cell segmentation. In the first stage, they obtained a density map by taking the B and G color components of the blood image in RGB space. Then, binary conversion and distance transform were performed to mark the locations of the cells. They successfully segmented white blood cells by applying watershed and connected component labeling methods.

In this study, we have developed an algorithm that can be easily integrated into real-time systems that can fully segment blood cells such as leukocytes, erythrocytes and thrombocytes on microscopic blood images. This method composes of five phases based on color and shape. Each microscopic blood image can be processed within a few seconds and can be divided into sub-images whose blood-shaped elements are segmented and can calculate their numbers (WBC, RBC, Thrombocyte) at a high degree of accuracy.

## 2. Background

Blood consists of two parts, formed elements of the blood and the fluid called plasma, which these cells contain. Blood cells are three types: red blood cells (RBC), leukocytes (WBC), and thrombocytes (platelets). The physiological characteristics of these cells (Marieb, 2006) are detailed in Table 1.

### 2.1. Connected-components labeling

Each item in an image is used to determine the component to which they are attached by looking at neighborhood relations, and is often preferred in jobs where automatic checking (Gonzalez and Woods, 2007) is performed. The components on images are measured based on the proximity or colors of neighboring pixels at a certain degree and are labeled with a number in the unique structure. Based on neighborhoods there are commonly used types with 4 and 8 connections. The image is represented by an  $R$ . If we consider that the segmented  $R$  image  $S$  comes from the discrete region, the objects in the image can be calculated as in Eq. 1.

$$R_f = R_b^C = \bigcup_{x=1, x \neq b}^S R_i \quad (1)$$

where  $R_f$  is objects in the image,  $R_b$  is background of the image, and  $C$  is complement of the set.

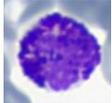
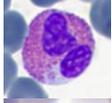
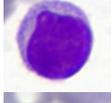
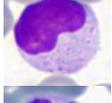
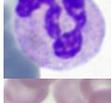
### 2.2. Convex hull

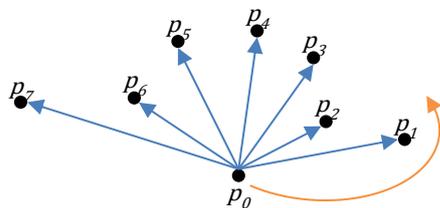
The Convex Hull is the smallest convex set, which spans a set of points  $S = \{p_0, p_1, \dots, p_N\}$  in the

Euclidean plane. In this study, [Graham \(1972\)](#) Scan algorithm is used as a Convex Hull method. In this method, firstly the point  $p(x_0, y_0)$  which has the smallest value with respect to the y axis is detected ([Fig. 1](#)). Then, based on this point, all the remaining points are sorted from small to large according to

their polar angle  $\tan^{-1}(y - y_0/x - x_0)$ . The orientation is calculated together with the previous ( $p_{h-1}$ ) and the next ( $p_{h+1}$ ) points for each point ( $p_h$ ) that will form the outer shell. If there is a change of direction at the selected point, this point is eliminated.

**Table 1:** Summary of formed elements of the blood

Cell Types	Image	Diameter	Range (per $\mu\text{L}$ )	Nucleus	Cytoplasm	Granules
<i>Basophil</i>		10-14 $\mu\text{m}$	0.5-1% of WBC 20-50	Bi-lobed or tri-lobed	Pale blue	Large purplish-black cytoplasmic
<i>Eosinophil</i>		10-14 $\mu\text{m}$	2-4% of WBC 100-400	Bi-lobed	Full of granules	Orange-red
<i>Lymphocyte</i>		5-17 $\mu\text{m}$	25-40% of WBC 1500-3000	Spherical or indented	Clear, Pale blue	-
<i>Monocyte</i>		14-24 $\mu\text{m}$	3-8% of WBC 100-700	U or kidney shaped	Gray-blue	Fine reddish (azurophil)
<i>Neutrophil</i>		10-12 $\mu\text{m}$	50-70% of WBC 3000-7000	2 to 5 segments or lobes	Pale blue-pink	Inconspicuous cytoplasmic
<i>Erythrocytes</i>		7-8 $\mu\text{m}$	4-6 million	Biconcave, anucleate disc; salmon-colored		
<i>Thrombocytes</i>		2-4 $\mu\text{m}$	150-500 thousand	Discoid cytoplasmic fragments containing granules; stain deep purple		



**Fig. 1:** Sort a series of point according to  $P_0$

If there is no change of direction, this point is determined as a coordinate of the outer shell. As a result of this, all the points in the plane are scanned. Characteristic features of the three points (a, b and c) in the same plane are classified as counter clockwise (CCW), clockwise (CW) and on the same line (COLL, Collinear). The orientation of these points is characterized by the sign of  $\Delta(a, b, c)$  of the determinant. Orientation calculation is performed by finding the determinants of the vertical and horizontal axis values of these points in the Euclidean plane (Eq. 2).

$$\Delta(a, b, c) = \begin{vmatrix} x_a & y_a & 1 \\ x_b & y_b & 1 \\ x_c & y_c & 1 \end{vmatrix} = x_a \begin{vmatrix} y_b & 1 \\ y_c & 1 \end{vmatrix} - y_a \begin{vmatrix} x_b & 1 \\ x_c & 1 \end{vmatrix} + \begin{vmatrix} x_b & y_b \\ x_c & y_c \end{vmatrix} \quad (2)$$

According to the result of the orientation process; if  $\Delta(a, b, c) < 0$  the characteristic feature is CCW, if  $\Delta(a, b, c) > 0$  it is CW and if  $\Delta(a, b, c) = 0$  it is COLL.

In this case, it is decided whether the selected point is convex by looking at the obtained results and the points that are not convex are removed from the list. In order for a point to be considered as convex, the characteristic feature must be CCW.

### 2.3. Watershed transform

Watershed transform is a method used for image segmentation in the field of mathematical morphology. It was first proposed in 1979 by [\(Beucher and Lantuejoul, 1979\)](#) for segmentation of grayscale images. As a result of the topographical structure, the watershed is the whole of the areas where rain water has been collected at a certain stage through the superficial flow and under the influence of them. In the same way, the slope of an image resembles a topographic surface. Here, the elevation is determined by the gray values of the pixels in the image. Low gray values mean valleys and high gray values mean peaks.

### 3. Materials and methods

In this study, two different peripheral blood films stained with Giemsa for educational purposes were used. Prepared blood films were taken under a light microscope with a light microscope at the magnification power of 100X given in [Table 2](#).

**Table 2:** Specifications of the biological microscope

Features	Values
Objectives	4X, 10X, 40X, 100X
Eyepiece	Wide Field Eyepiece WF 10x/18
Condenser	Sliding-in condenser NA1.25
Focusing	Coaxial Coarse, Moving Range 20mm
Illumination	6V/ 20W Halogen Lamp

These images were taken with the digital camera in the specifications given in Table 3.

**Table 3:** Specifications of the colored digital camera

Features	Values
Effective Pixels	16.0 MP
Sensor	Panasonic MN34120
Sensor Size	1/2.33 inch
Resolution (Max)	4608H x 3456V
Color Temp.	1900K-8000K
Frame Rates	5fps @ 16MP (4608x3456) 25fps @ 4MP (2304x1728)

First of all, proposed method produces a grayscale output image based on green channel after normalizing input image for leukocyte and thrombocyte cells. Next, Nucleus Template Function (NTF) is applied to generate a template of the leukocyte nucleus and thrombocytes as a polygon. The normalized input image is given to the input of the Cytoplasm Template Function (CTF) and the cytoplasm polygon is obtained at the function output. Leukocyte and thrombocyte extraction is performed by applying set and subset combination operations for these polygons. In the final step, the Erythrocyte Template Function (ETF) generates the erythrocyte polygon. ETF polygons are compared with NTF and CTF polygons and segmentation of

erythrocytes is performed. Thus, the segmentation process of all blood cells is completed. Let  $F$  be image function is defined as Eq. 3.

$$F: S \rightarrow [0,1,2, \dots, 255]^c \quad (3)$$

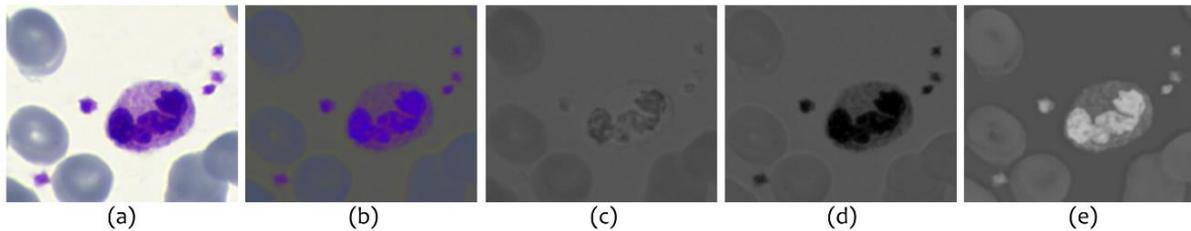
where  $S = \llbracket 0; m - 1 \rrbracket \times \llbracket 0; n - 1 \rrbracket$  is image size,  $m$  and  $n$  are column and row numbers of the  $F$  image, and  $c$  is the number of color channels in the image  $c \in \{1, 3, 4\}$ . The definition for the pixel values in a color image is as in Eq. 4.

$$v = F[x, y] = (R, G, B) \quad (4)$$

Irregular color distributions that occur due to illumination (Liu et al., 2012) during the acquisition of peripheral blood smear images are a serious problem especially in color based image processing. The normalization process ensures that the pixel values in these images are brought to a certain standard. This helps the algorithm to produce accurate results. For image normalization, each color component is calculated as shown in Eq. 5.

$$v_R^* = \frac{v_R}{v_R+v_G+v_B} \times 255, \quad v_G^* = \frac{v_G}{v_R+v_G+v_B} \times 255, \quad v_B^* = \frac{v_B}{v_R+v_G+v_B} \times 255 \quad (5)$$

where  $v_R^*$  is normalized red channel,  $v_G^*$  is normalized green channel and  $v_B^*$  is normalized blue channel. A normalized pixel is expressed as  $v^* = (v_R^*, v_G^*, v_B^*)$ . Fig. 2 shows the results of normalized image and RGB channels. Fig. 3 shows the general flow diagram of the algorithm



**Fig. 2:** The RGB color components of a sample blood image; a) Input RGB image, b) Normalized image, c) Normalized red channel, d) Normalized green channel, e) Normalized blue channel

### 3.1. Nucleus template function (NTF)

The nucleus has a strong physical absorption and chemical affinity, because it contains a dense nucleoprotein and nucleic acid. Therefore, the staining intensity of leukocyte nucleus is much higher than cytoplasm and erythrocyte cells (Gu and Cui, 2012). In addition, nucleus and thrombocytes in digital images with RGB color space have the lowest value of the Green color component compared to other objects in the image (Prinyakupt and Pluempitiwiriyawej, 2015). As shown in Fig. 2d, when a RGB image converted to gray scale using the green channel,  $G[x, y] = (v_R^*, v_G^*, v_B^*)$ , leukocyte nucleus and thrombocyte can be separated much more easily than other color components. Following the pre-processing, a simple color reduction method (Eq. 6) is performed to eliminate cytoplasm and erythrocytes in the image; where  $interval$  is the

color reduction coefficient. The  $interval$  value is set to 5 in this study. When it gets higher value, it causes significant loss of information in the image.

$$F[x, y] = round\left(\frac{G[x, y]}{interval}\right) \times interval \quad (6)$$

In addition, it results in some undesired noises in the result image. Median smoothing filter is applied to remove this problem (Eq. 7).

$$F[x, y] = median\{g[p, q]\} \quad (7)$$

where  $g[p, q]$  is convolution kernel,  $p$  and  $q$  are kernel index. In this study, a 3x3 convolution matrix is chosen for performance. In the next step, image is converted to binary image (Eq. 8). Since non-parametric functions are important in for automated segmentation process, Otsu is preferred as a thresholding method.

$$T[x,y] = \begin{cases} 255, & \text{if } F[x,y] \geq t \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

Image background is represented by 0, while objects in images are represented by 1. In this case, the nucleus appears to be background. In order to reverse it, image inversion is performed as in Eq. 9 to

allow the nucleus to be transformed into objects (foreground).

$$I[x,y] = \begin{cases} 255, & \text{if } T[x,y] = 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

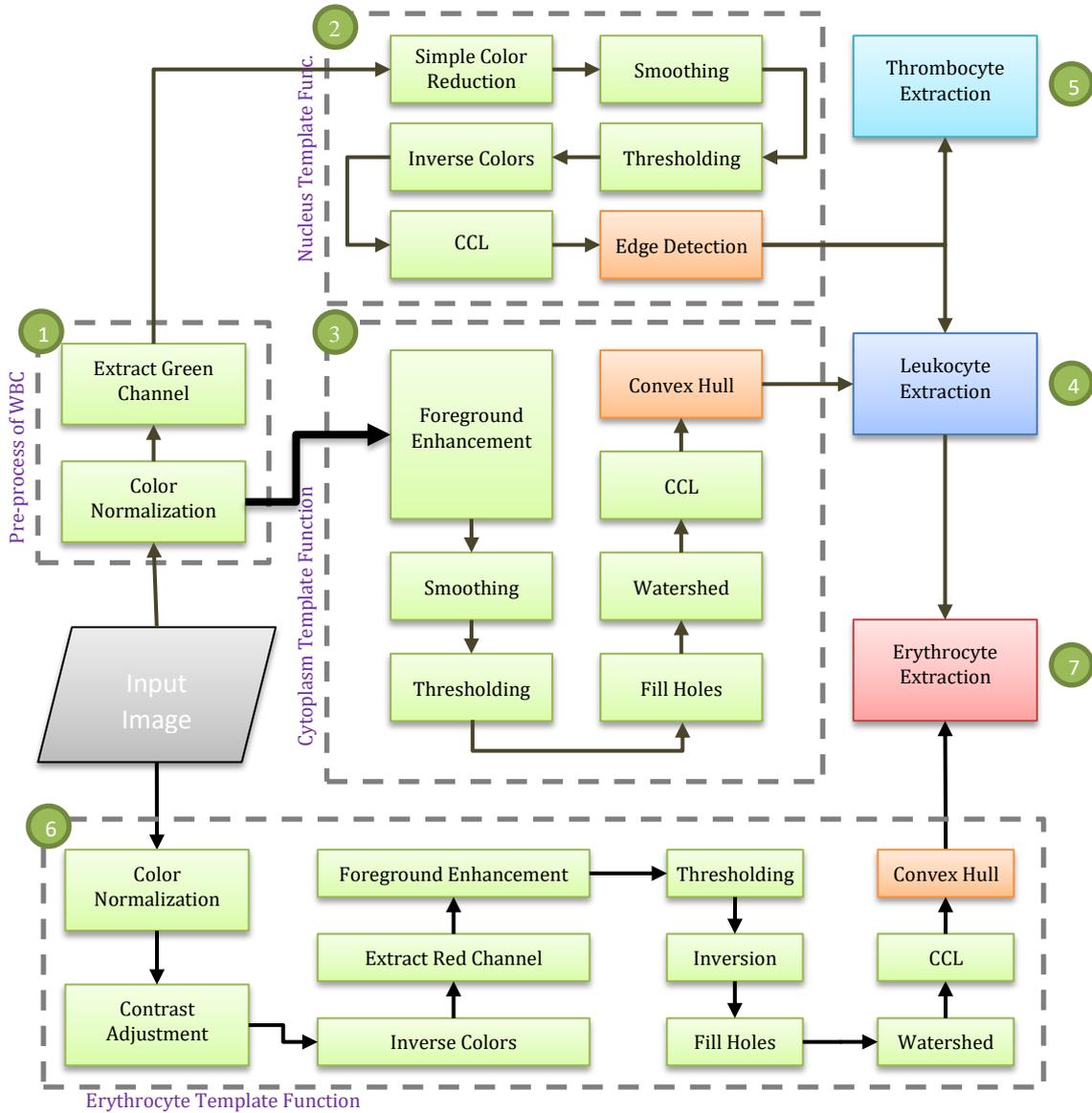


Fig. 3: General flow chart of proposed method

The Connected-Components Labeling method is used to identify these objects with a unique ID. In the next step, edge points are determined by applying the Canny edge detection algorithm on each labeled component. Thus, the points forming the outer shell of nucleus and thrombocytes according to Fig. 4e are stored as polygons of  $P_n = \{[x_0, y_0], [x_1, y_1], \dots, [x_m, y_m]\}$ . Where  $n$  is the number of objects and  $m$  is the number of points in the object of  $n$ . A list of all polygons for Leukocyte Extraction is stored as  $P_{NT} = \{P_0, P_1, P_2, \dots, P_n\}$ . Nucleus and thrombocyte in the  $P_{NT}$  list is separated from each other according to criteria determined during Leukocyte and Thrombocyte Extraction. All the processing steps of the NTF are as shown in Fig. 4.

### 3.2. Cytoplasm template function (CTF)

The cytoplasm is located between the cell membrane and the nucleus and has light colored, semi-fluid and transparent properties due to its high-water content. For this reason, cytoplasm segmentation is much more difficult than nucleus segmentation. In addition, the segmentation of the outer borders of the cytoplasm is complicated because of being adjacent to red blood cells (Tareef et al., 2016). When studies performed in this area are examined, it is seen that cytoplasm segmentation has less success than cell nucleus segmentation. In order to achieve a higher success rate, a Foreground Enhancement method is developed to optimize the

segmentation of cytoplasm with low density. In this method, the processing sequence is as follows;

- a) The average color (*avg*) between 0 and 255 is found in the image histogram.

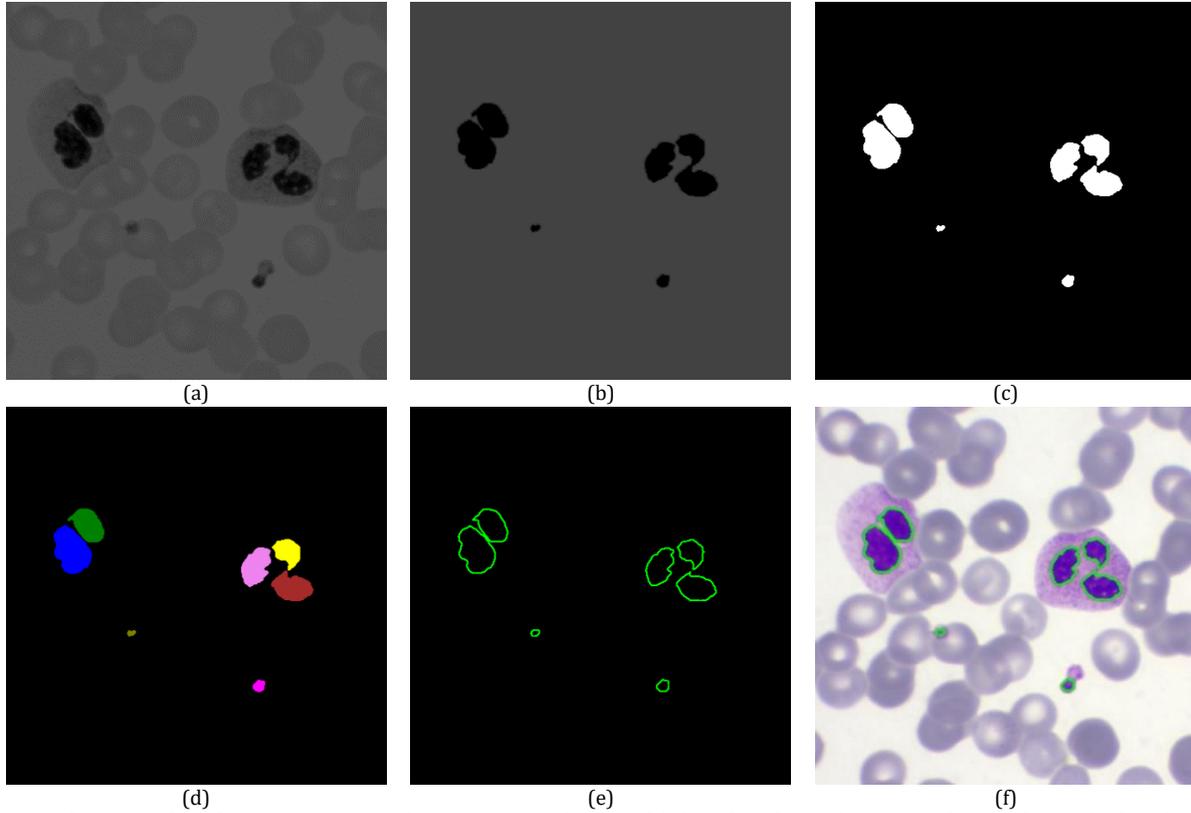
$$avg = \frac{\sum_{i=0}^{255} H_i i}{\sum_{i=0}^{255} H_i} \quad (10)$$

- b) The average color (*T*) between *avg* and 255 is found according to the image histogram.

$$T = \frac{\sum_{i=avg}^{255} H_i i}{\sum_{i=avg}^{255} H_i} \quad (11)$$

- c) For each pixel in the image, a comparison is made according to the *T* threshold value. Thus, the cytoplasm becomes more apparent.

$$F[x,y] = \begin{cases} 0, & \text{if } F[x,y] > T \\ F[x,y], & \text{otherwise} \end{cases} \quad (12)$$



**Fig. 4:** Nucleus template function process; a) Pre-processed image, b) Simple color reduction and smoothing, c) Thresholding and inverse colors, d) Connected-components labeling, e) Edge detection, f) Result image

Note that in Fig. 5a, the background pixel values are higher than the foreground. For this reason, all pixels above the brightness average are made zero and the objects are emphasized more strongly. This is followed by removal of the undesired noise from the image with Median smoothing filter. When thresholding is applied, all the objects in the cytoplasm are converted into a binary image. Before Watershed algorithm, fill-hole function is applied for all blood cells to prevent over-segmentation of the cytoplasm.

As shown in Fig. 5d, the template of the cytoplasm of both leukocytes is segmented optimally. Component labeling is then performed to ensure that each object has a unique identity. In the next step, according to Eq. 3, the actual size of each object is calculated. The points forming the outer shell of all the objects in the result image in Fig. 5f are stored as  $P_n = \{[x_0, y_0], [x_1, y_1], \dots, [x_m, y_m]\}$  polygons. Subsequently, the maximum radius of each polygon is calculated and compared with the minimum leukocyte diameter (5  $\mu\text{m}$ ) given in Table 1. Objects equal to and larger than this radius are added to  $P_{CT} = \{P_0, P_1, P_2, \dots, P_n\}$ . Note that some cells in the

$P_{CT}$  polygons except the cytoplasm (usually RBC) are added to the list. However, these cells are eliminated because they do not meet the criteria determined during Leukocyte Extraction process.

### 3.3. Leukocyte and thrombocyte extraction

In this section, full leukocyte extraction is performed by combining the binary result images obtained from the NTF and CTF procedures. Let NTF be  $f_{NT}$  and CTF be  $f_{CT}$  symbolic functions. The obtained leukocyte image is  $f_L = f_{NT} \cup f_{CT}$ . However, due to the conditions mentioned in the NTF and the CTF, other blood cells other than leukocytes are represented as  $f_L$ . To resolve this issue, a constraint rule is required for the polygons ( $P_{NT}$ ) obtained with NTF as shown in Fig. 6b and polygons ( $P_{CT}$ ) obtained with CTF as shown in Fig. 6c.

This rule is as follows; Leukocytes are detected on the merged image by searching for situations that provide the condition of  $p_n \subseteq P_{CT}, \forall p_n \{n | n \in \mathbb{N}, p_n \in P_{NT}\}$ . In other words, every cytoplasm that has at least one nucleus becomes a leukocyte. Each polygon that provides this condition in the  $p_n$  polygon and

$P_{CT}$  polygon list is added to the leukocyte polygon list in the form of  $P_L = \{P_0, P_1, P_2, \dots, P_n\}$ . In Fig. 6f,

leukocyte extraction results are shown.

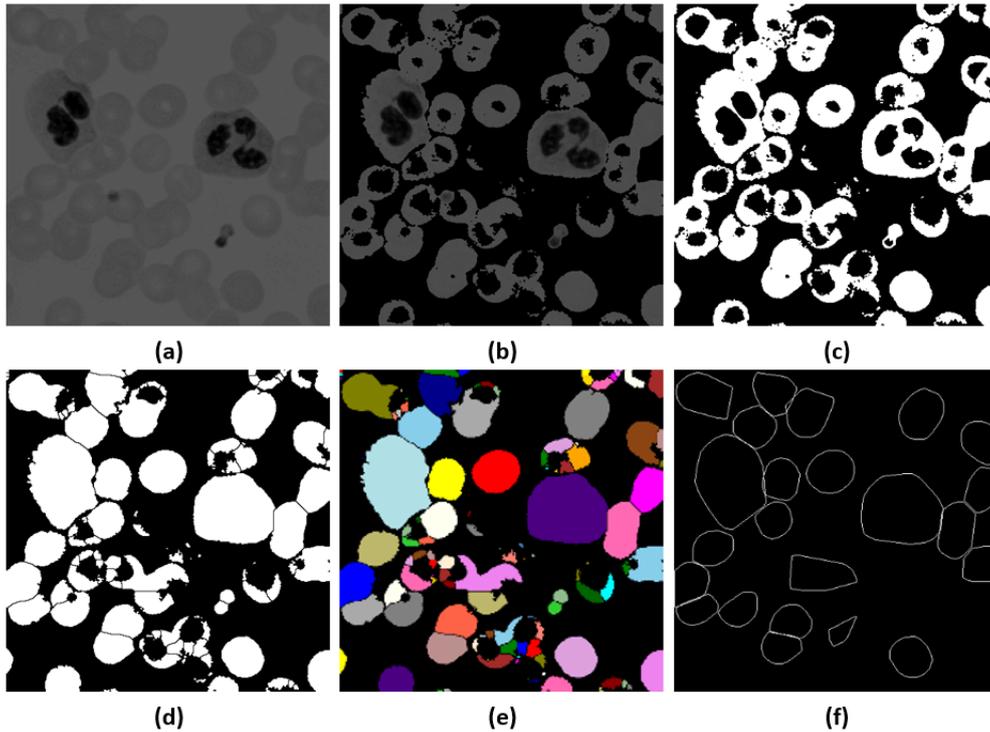


Fig. 5: Cytoplasm template steps; a) Pre-processed image, b) Foreground enhancement and smoothing, c) Thresholding, d) Fill holes and watershed, e) Connected-components labeling, f) Graham scan

However, each  $p_n$  polygon may not be a subset of  $P_{CT}$ . So, in case of  $p_n \notin P_{CT}$ , the polygons represent thrombocytes, not leukocyte nucleus. Thus, thrombocytes in the  $P_{NT}$  polygon list are added to

the thrombocyte polygon list as  $P_T = \{P_0, P_1, P_2, \dots, P_n\}$ . Fig. 6h shows the result of thrombocyte extraction.

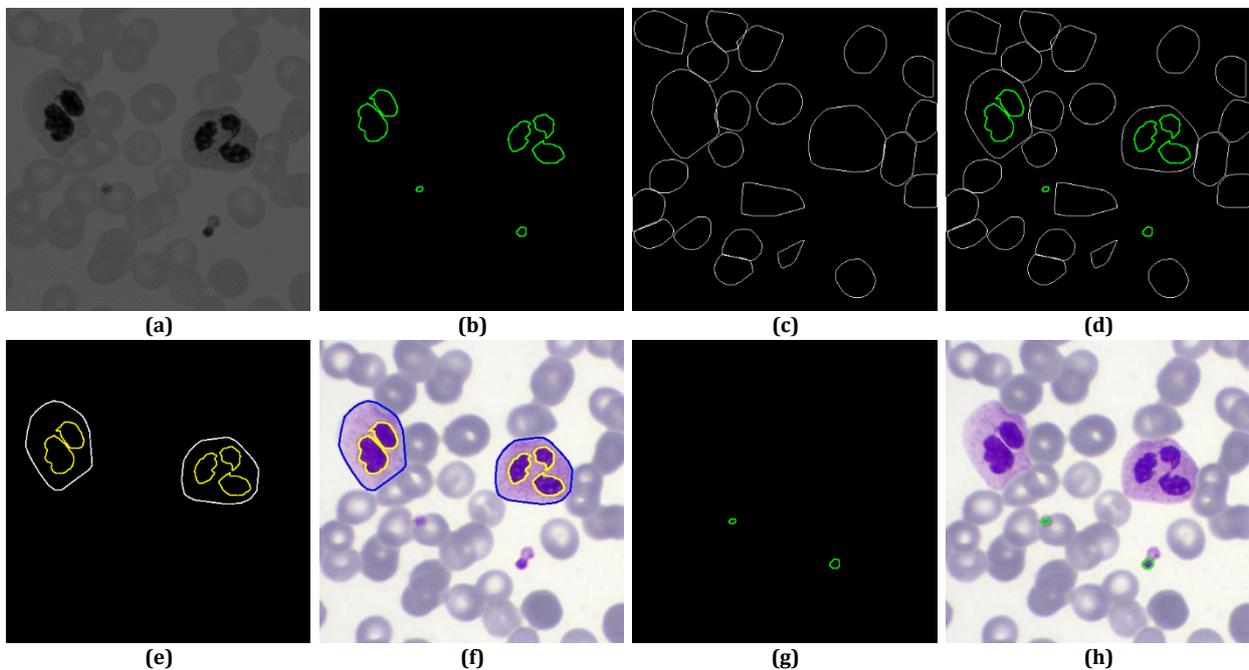


Fig. 6: Leukocyte and thrombocyte extraction process; a) Pre-processed image, b) NTF result, c) CTF result, d)  $f_L = f_{NT} \cup f_{CT}$  result, e)  $P_L$  result, f) Leukocyte extraction result, g)  $P_T$  result, h) Thrombocyte extraction result

### 3.4. Erythrocyte template function (ETF) and extraction

Erythrocytes constitute more than 90% of blood cells and almost all are of the same size. In general,

erythrocytes are like stacked discs in peripheral blood smear images, so the extraction is often difficult. In some cases, they are clustered so that the numbers of them cannot be distinguished even with the naked eye. In this study, the extraction of

erythrocytes is performed in two stages. In the first stage, it passes through the process steps as shown in Fig. 7 to obtain the ETF of the image. In the second step, the convex hull of each labeled component is

determined by the Graham Scan method. The points forming the convex hull are then stored as polygons  $P_n = \{[x_0, y_0], [x_1, y_1], \dots, [x_m, y_m]\}$ .

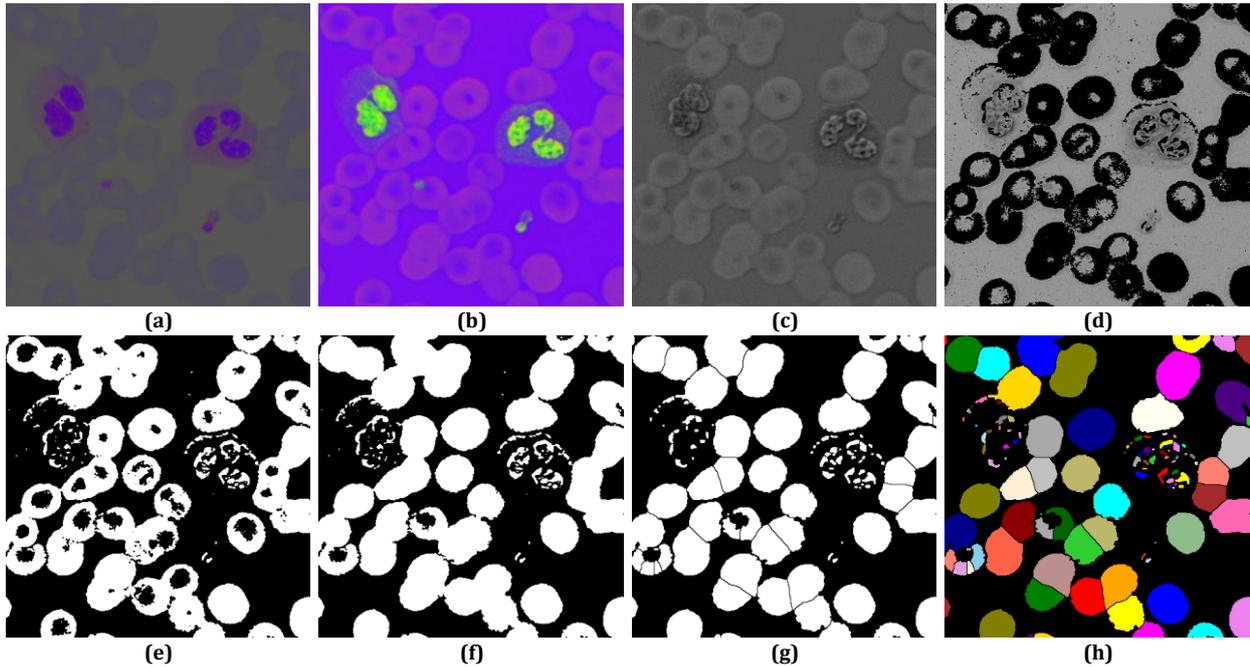


Fig. 7: Erythrocyte segmentation process; a) Pre-processed image, b) Contrast adjustment and inversion, c) Extract red channel, d) Foreground enhancement, e) Thresholding and inversion, f) Fill holes, g) Watershed, h) CCL

Finally, the non-subset of  $P_L$  or  $P_T$  from the obtained polygons is determined as  $p_n \notin (P_L \cup P_T)$ . Thus, as shown in Fig. 8, all the points forming the convex hull of erythrocytes are found and the segmentation process is completed.

shown in Table 5. Where TP, TN, FP, and FN refer to true positive (correctly approved cells), true negative (correctly rejected cells), false positive (incorrectly approved cells) and false negative (incorrectly rejected cells), respectively.

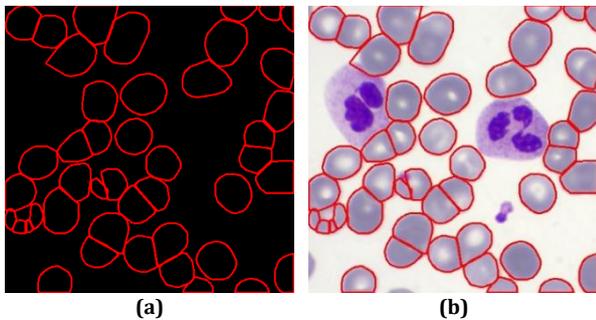


Fig. 8: Erythrocyte extraction; a) Graham Scan, b) Result image

#### 4. Experimental results

The peripheral blood smear images used in this study are obtained using a 16MP color digital camera integrated into a light microscope. The data set prepared in the two different light conditions used in the tests of the developed method is shown in Table 4. Each image contains at least one leukocyte, several thrombocytes and several erythrocytes. These blood samples used in the experiments were examined by a specialized hematologist and confirmed the validity of the results.

In this paper, we applied eleven statistical measurements to analyze segmentation results in experimental studies. These measurements are

Table 4: Data sets used in the experiments

Datasets	Image count	Magnification	Resolution
Dataset-1	100	100X	4608x3456
Dataset-2	100	100X	2304x1728

Sample peripheral blood smear images and results of Dataset-1 are shown in Fig. 9. Statistical measurement results according to Dataset-1 are given in Table 6. Sample peripheral blood smear images and results of Dataset-2 are shown in Fig. 10. Statistical measurement results according to Dataset-2 are given in Table 7.

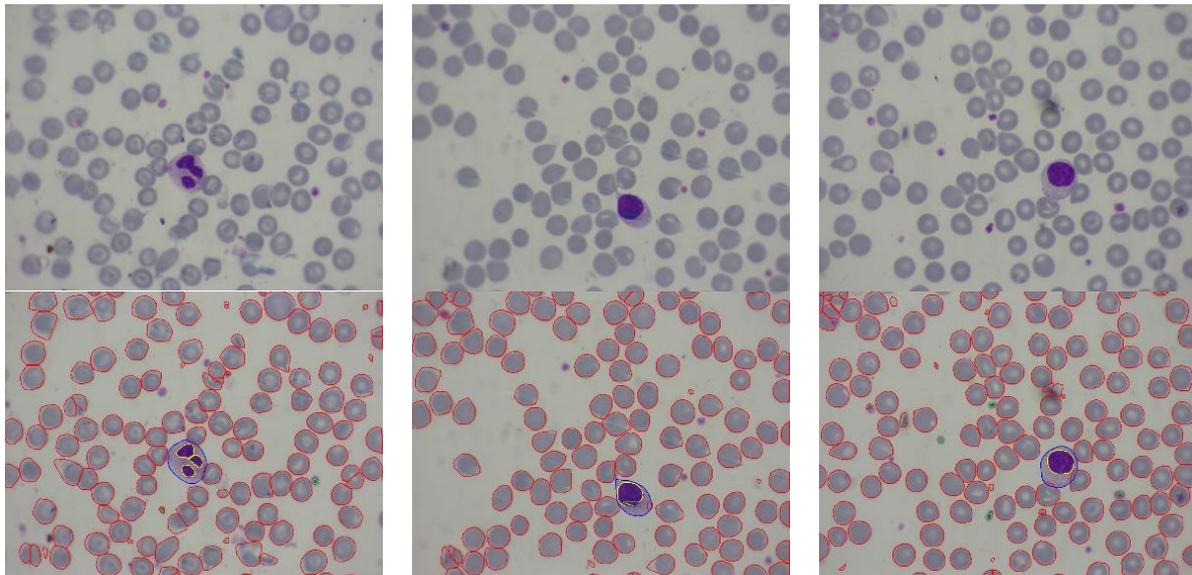
The mean values of the two datasets used to test the developed method are given in Table 8. Accordingly, the accuracy of leukocyte segmentation is 99.86% and the accuracy of leukocyte segmentation is 94.77%. The accuracy of thrombocytes is 98.4%, the precision of thrombocytes is 90.14%. Because thrombocytes have a much smaller diameter than leukocytes, paint residues on the preparation, etc. accuracy and precision decrease. According to this study, designing a separate Thrombocyte Template Function (TTF) instead of using NTF in thrombocyte segmentation gives better results. Finally, it is seen that the accuracy of erythrocytes segmentation is 93.4% and the precision of erythrocytes segmentation is 95.88%. However, when the images

in Dataset-2 are examined, it is understood that segmentation of erythrocytes according to Dataset-1 is more difficult. For this reason, the accuracy of Dataset-2 is slightly lower. Fig. 11 shows the

accuracy and precision measurement charts of the segmentation results and detailed examples showing all processes are given in Fig. 12.

**Table 5:** Statistical measurements

Sensitivity or True Positive Rate	$TPR = \frac{TP}{TP + FN}$	Bookmaker Informedness	$BM = TRP + SPC - 1$
Specificity or True Negative Rate	$TNR = \frac{TN}{TN + FP}$	Markedness	$MK = PPV + NPV - 1$
Precision or Positive Predictive Value	$PPV = \frac{TP}{TP + FP}$	Accuracy	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$
Negative Predictive Value	$NPV = \frac{TN}{TN + FN}$	F-Measurements	$FM = 2 / (1/TPR + 1/PPV)$
False Positive Ratio	$FPR = \frac{FP}{FP + FN}$	Matthews Correlation Coefficient	
False Negative Ratio	$FNR = \frac{FN}{TP + FN}$	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	



**Fig. 9:** Sample peripheral blood smears images and results of Dataset-1

**Table 6:** Dataset-1 results

Dataset-1	TPR	TNR	PPV	NPV	FPR	FNR	BM	MK	ACC	MCC	FM
Leukocyte nucleus	95,77	99,97	98,55	99,91	0,03	0,09	95,75	98,46	99,89	97,10	97,14
Leukocyte cytoplasm	96,97	99,94	94,12	99,97	0,06	0,03	96,91	94,09	99,92	95,49	95,52
Leukocyte	96,37	99,96	96,33	99,94	0,04	0,06	96,33	96,28	99,90	96,29	96,33
Thrombocyte	79,17	99,56	88,37	99,13	0,44	0,87	78,73	87,50	98,74	83,00	83,52
Erythrocyte	97,67	67,21	96,38	76,31	32,79	23,69	64,88	72,69	94,60	68,67	97,02

Segmentation results obtained in this study are compared with other studies in the literature. Leukocyte segmentation results are given in Table 9, erythrocyte segmentation results are given in Table 10, and thrombocyte segmentation results are given in Table 11.

Prinyakupt and Pluempitiwiriyawej (2015) performed nucleus and cytoplasm extraction with high success rate by performing combined thresholding, morphological operation and ellipse curve fitting process steps. However, it is known that the number of repetitions in the application of morphological operators in the nucleus segmentation section will differ from the image to the image. For example, if there is not enough erosion, the image will be noisy. For this reason, it is not a very suitable structure for automatic segmentation methods.

Alomari et al. (2014) proposed an iterative structured circle detection algorithm for segmentation of WBC and RBC cells. In the tests performed, the WBC achieved an accuracy rate of over 98% and a precision of close to 90%. However, physiological structures required by the nature of WBCs are not regular.

For this reason, circular approaches may not always be suitable for leukocytes. As predicted, this methodical approach to RBCs has yielded more successful results. However, erythrocytes in irregularly stacked or nonparametric structures did not produce good results. If we have implemented the Convex Hull method in our study, it provides flexibility, especially for irregularly structured blood cells. Vale et al. (2014) used the green component of the input images according to the RGB color space. They automatically calculated the Euclidean

distances between the peak points in the histogram and the rest of the leucocyte core, which were

obtained according to this component, by means of Fuzzy clusters.

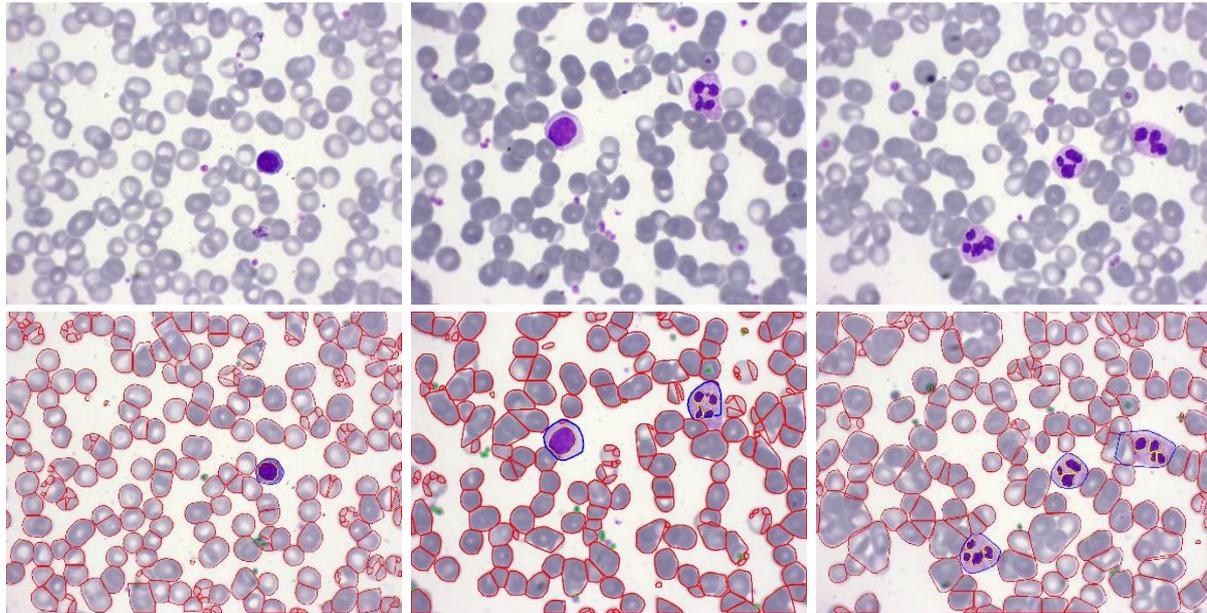


Fig. 10: Sample peripheral blood smears images and results of Dataset-2

Table 7: Dataset-2 results

Dataset-2	TPR	TNR	PPV	NPV	FPR	FNR	BM	MK	ACC	MCC	FM
Leukocyte nucleus	93,10	99,92	94,74	99,90	0,08	0,10	93,03	94,64	99,82	93,83	93,91
Leukocyte cytoplasm	89,19	99,92	91,67	99,90	0,08	0,10	89,11	91,57	99,82	90,33	90,41
Leukocyte	91,15	99,92	93,20	99,90	0,08	0,10	91,07	93,10	99,82	92,08	92,16
Thrombocyte	86,41	99,23	91,91	98,64	0,77	1,36	85,64	90,54	98,06	88,06	89,08
Erythrocyte	95,43	73,96	95,37	74,20	26,04	25,80	69,39	69,57	92,19	69,48	95,40

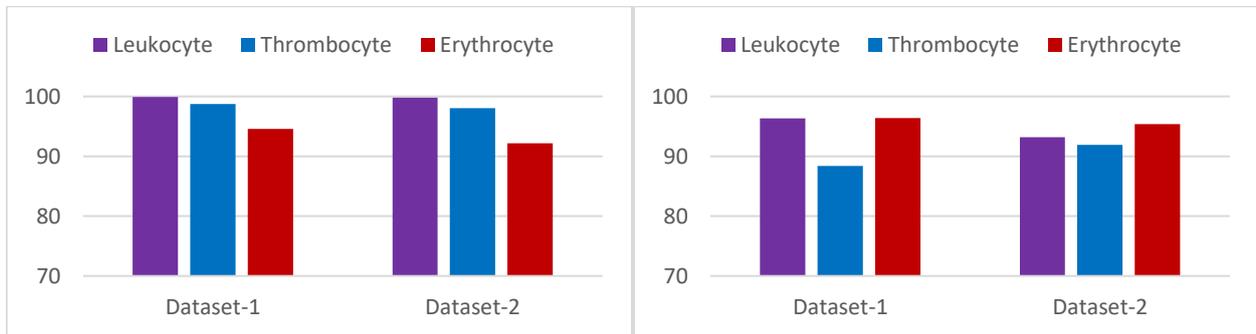


Fig. 11: Comparative results for datasets

Table 8: Mean values of Dataset-1 and Dataset-2

Average of Datasets	TPR	TNR	PPV	NPV	FPR	FNR	BM	MK	ACC	MCC	FM
Leukocyte nucleus	94,44	99,95	96,64	99,91	0,05	0,09	94,39	96,55	99,86	95,46	95,53
Leukocyte cytoplasm	93,08	99,93	92,89	99,94	0,07	0,06	93,01	92,83	99,87	92,91	92,97
Leukocyte	93,76	99,94	94,77	99,92	0,06	0,08	93,70	94,69	99,86	94,19	94,25
Thrombocyte	82,79	99,40	90,14	98,88	0,60	1,12	82,19	89,02	98,40	85,53	86,30
Erythrocyte	96,55	70,59	95,88	75,25	29,41	24,75	67,13	71,13	93,40	69,08	96,21

Table 9: Comparison of leukocyte segmentation results

Methods	TPR	TNR	PPV	NPV	ACC	MCC
Prinyakupt and Pluempitiwiriwaj (2015)	—	92,50	—	92,30	—	—
Alomari et al. (2014)	89,70	98,40	89,73	—	98,40	—
Vale et al. (2014)	94,87	99,75	99,74	95,11	97,31	94,73
Our proposed method	93,76	99,94	94,77	99,92	99,86	94,19

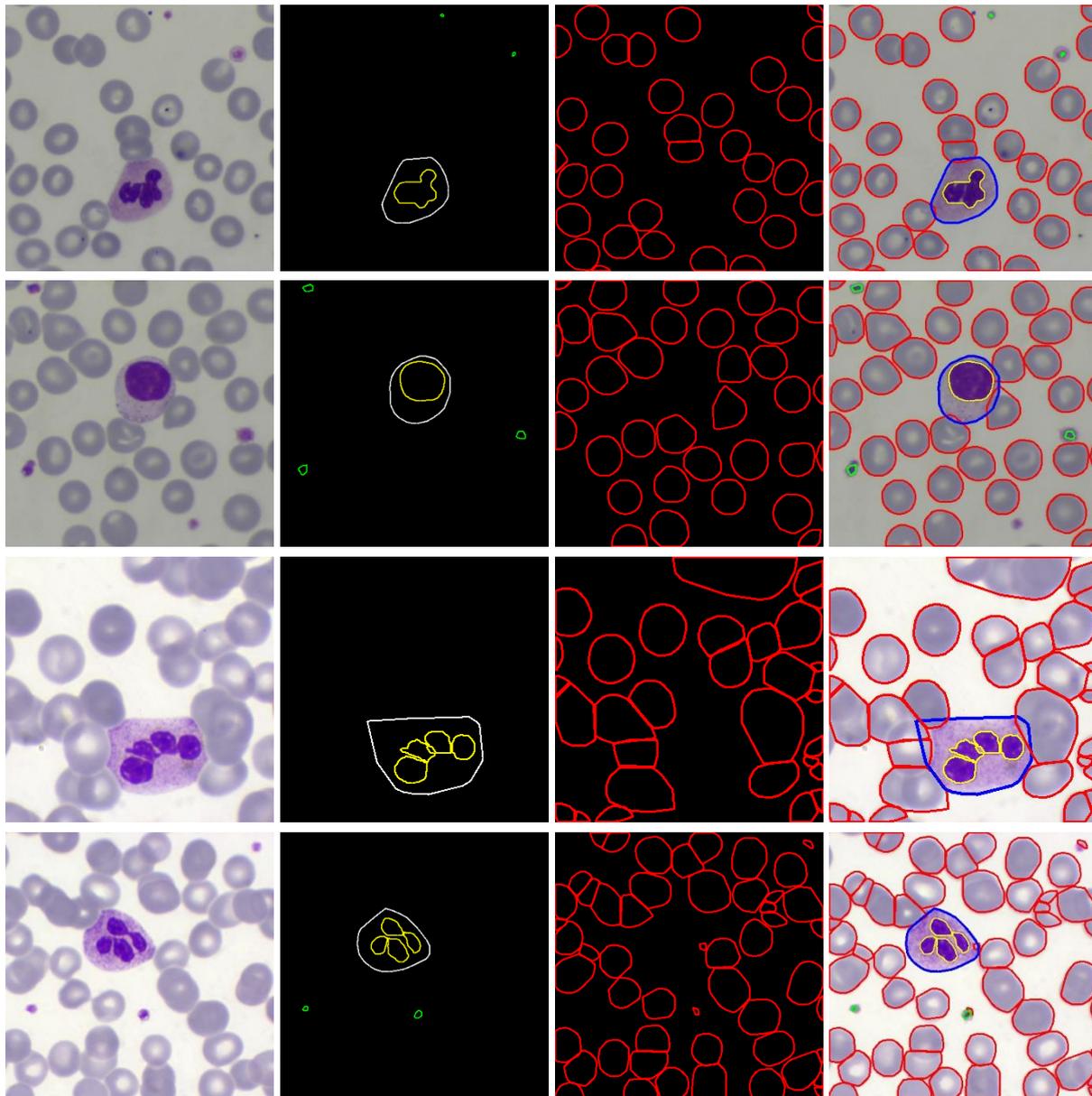
According to the results, they achieved more than 95% in both leukocyte and erythrocyte segmentation.

Khajehpour et al. (2013) applied the Euclidian distance transform to the binary transformed image.

By applying the Watershed method to the gray scale image obtained by this process, they have segmented the erythrocyte cells with high success rate. Angulo and Flandrin (2003) have proposed a two-step method for erythrocyte segmentation. In the first

step, by applying mathematical morphology methods for image analysis, erythrocyte extraction, erythrocyte centers and erythrocytes with center were obtained. In the second step, according to these

three parameters, an automatic method of segmentation processing is proposed by calculating the number of connected components, spreading and overlaps coefficients.



**Fig. 12:** Detailed segmentation samples. First column demonstrate the original images, second are leukocytes and thrombocytes, third are erythrocytes, and last column the segmentation of all blood cells results

**Table 10:** Comparison of erythrocyte segmentation results

Methods	TPR	TNR	PPV	NPV	ACC	MCC
Alomari et al. (2014)	95,00	98,00	95,39	—	97,50	—
Vale et al. (2014)	97,96	92,82	93,17	97,85	95,39	90,90
Khajehpour et al. (2013)	97,99	45,09	97,76	47,91	95,91	44,36
Angulo and Flandrin (2003)	97,29	89,53	88,88	97,46	93,12	86,59
Our proposed method	96,55	70,59	95,88	75,25	93,40	69,08

**Table 11:** Comparison of thrombocyte segmentation results

Methods	TPR	TNR	PPV	NPV	ACC	MCC
Dey et al. (2015)	—	—	—	—	92,71	—
Our proposed method	82,79	99,40	90,14	98,88	98,40	85,53

Dey at al. (2015) proposed a color-based algorithm for thrombocyte segmentation from bloodstream images (Dey at al., 2015). In the method, the input image is converted from RGB to L \* a \* b color space, and the image processing is performed using \*a and \*b components. It is known

that thrombocytes are stained in the same color as WBCs. For this reason, it is difficult to distinguish them from each other in color-based segmentation methods. However, since WBCs are usually larger in size, this problem can be solved to a considerable extent by making a size comparison. However, the

diameter of some lymphocytes may be close to the diameter of the thrombocytes. In such cases, it will be necessary to look for cell cytoplasm in order to achieve higher success.

## 5. Conclusion

In this study, a fully automated segmentation method is developed to present a new perspective for the segmentation of leukocytes, erythrocytes and thrombocytes in peripheral blood smear images obtained under different light and spread conditions. The proposed method segments all formed-elements of the blood in five phases. First, it produces a normalized grayscale image for leukocyte and thrombocyte cells. Second, it generates a polygon list for the leukocyte nucleus and thrombocyte. Third, it generates a polygon list for the cytoplasm of leukocyte. Fourth, it extracts leukocyte and thrombocyte using the polygons mentioned in the second and third steps. Fifth, it generates a polygon list for the erythrocyte and it extracts erythrocyte cells using all polygon lists. Each peripheral blood smear image and its components are considered as a set/subset, error rate of cell segmentation is decreased. In addition, storage of image components as polygons provides computational flexibility in terms of cell physiology. Thus, it is thought to help about the evaluation of abnormal cells for studies on blood diseases. The proposed method is quite successful and efficient in leukocyte cells. As previously mentioned for thrombocytes, better results can be obtained by performing an image processing separately from leukocyte segmentation. In erythrocytes, although segmentation of stacked cell clusters is an important problem, an efficient result is obtained in this study. Performance improvement can be achieved by developing a "Distance Transform" or "ROI" based local segmentation method for erythrocyte cell clusters that cannot be segmented.

## Acknowledgment

This work was supported by Research Fund of the Karabük University, Project Number: KBÜ-BAP-15/2-DR-003.

## References

- Agrawal S and Xaxa DK (2014). Survey on image segmentation techniques and color models. *International Journal of Computer Science and Information Technologies*, 5(3): 3025-3030.
- Alomari YM, Abdullah SNHS, Azma RZ, Omar K (2014). Automatic detection and quantification of wbcs and rbcs using iterative structured circle detection algorithm. *Computational and Mathematical Methods in Medicine*, 2014: Article ID 979302, 17 pages. <https://doi.org/10.1155/2014/979302>
- Angulo J and Flandrin G (2003). Automated detection of working area of peripheral blood smears using mathematical morphology. *Analytical Cellular Pathology*, 25(1): 37-49.
- Arslan S, Özyürek E, Gündüz-Demir Ç (2014). A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images. *International Society for Advancement of Cytometry (Cytometry, Part A)*, 85(6): 480-490.
- Beucher S and Lantuejoul C (1979). Use of watersheds in contour detection. In the *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, Rennes, France: 1-12.
- Dey R, Roy K, Bhattacharjee D, Nasipuri M, and Ghosh P (2015). An automated system for segmenting platelets from microscopic images of blood cells. In the *International Symposium on Advanced Computing and Communication*, IEEE, Silchar, India: 230-237. <https://doi.org/10.1109/ISACC.2015.7377347>
- Dhiman SR and Talwar R (2014). Image segmentation review a survey of image segmentation techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(9): 2584-2589.
- Gonzalez RC and Woods RE (2007). *Digital image processing*. Pearson Corporation, London, UK.
- Graham RL (1972). An efficient algorithm for determining the convex hull of a finite planar set. *Information Processing Letters*, 1(4): 132-133.
- Gu G and Cui D (2012). Polar angle detection and image combination based leukocyte segmentation for overlapping cell images. *Computing and Informatics*, 30(1): 189-199.
- Jiang K, Liao Q, and Xiong Y (2006). A novel white blood cell segmentation scheme based on feature space clustering. *Soft Computing*, 10(1): 12-19.
- Kaur N, Singh J, and Sharma V (2015). Analysis and comprehensive study-image segmentation techniques. *International Journal for Research in Applied Science and Engineering Technology*, 3(1): 241-246.
- Khajehpour H, Dehnavi AM, Taghizad H, Khajehpour E and Naeemabadi M (2013). Detection and segmentation of erythrocytes in blood smear images using a line operator and watershed algorithm. *Journal of Medical Signals and Sensors*, 3(3): 164-171.
- Liu Y, Shih H, Yang T, Yang H, Yang D, and Sun Y (2012). Quantitative measurement for pathological change of pulley tissue from microscopic images via color-based segmentation. In: Pan JS, Chen SM, and Nguyen NT (Eds.), *Intelligent Information and Database Systems, ACIIDS 2012. Lecture Notes in Computer Science*, 7198: 476-485. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-28493-9\\_50](https://doi.org/10.1007/978-3-642-28493-9_50)
- Madhloom HT, Kareem SA, and Ariffin H (2012). An image processing application for the localization and segmentation of lymphoblast cell using peripheral blood images. *Journal of Medical Systems*, 36(4): 2149-2158.
- Marieb EN (2006). *Human anatomy and physiology*. Benjamin Cummings, San Francisco, USA.
- Mohamed MMA and Far B (2012). A fast technique for white blood cells nuclei automatic segmentation based on gram-schmidt orthogonalization. In the *IEEE 24<sup>th</sup> International Conference on Tools with Artificial Intelligence, IEEE, Athens, Greece*, 1: 947-952. <https://doi.org/10.1109/ICTAI.2012.133>
- Nazlibilek S, Karacor D, Ercan T, Sazli MH, Kalender O, and Ege Y (2014). Automatic segmentation, counting, size determination and classification of white blood cells. *Measurement*, 55: 58-65.
- Prinyakupt J and Pluempitiwiriwajew C (2015). Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers. *Biomedical Engineering Online*, 14(1): 1-19.
- Sadr A, Jahed M, Salehian P, and Eslami A (2010). Leukocyte's nucleus segmentation using active contour in YCbCr colour space. In the *IEEE EMBS Conference on Biomedical Engineering and Sciences, IEEE, Kuala Lumpur, Malaysia*: 257-260. <https://doi.org/10.1109/IECBES.2010.5742239>

- Salem NM (2014). Segmentation of white blood cells from microscopic images using K-means clustering. In the 31<sup>st</sup> National Radio Science Conference, IEEE, Cairo, Egypt: 371-376. <https://doi.org/10.1109/NRSC.2014.6835098>
- Sharma N and Aggarwal LM (2010). Automated medical image segmentation techniques. *Journal of Medical Physics*, 35(1): 3-14.
- Sumengen B, Manjunath BS, and Kenney C (2002). Image segmentation using curve evolution and flow fields. In the International Conference on Image, IEEE, Rochester, NY, USA, 1: I-I. <https://doi.org/10.1109/ICIP.2002.1037970>
- Tareef A, Song Y, Cai W, Wang Y, Feng DD, and Chen M (2016). Automatic nuclei and cytoplasm segmentation of leukocytes with color and texture-based image enhancement. In the 13<sup>th</sup> International Symposium on Biomedical Imaging, IEEE, Prague, Czech Republic: 935-938. <https://doi.org/10.1109/ISBI.2016.7493418>
- Vale AMPG, Guerreiro AMG, Dória Neto AD, Cavalvanti Junior GB, Leitão VCLTD, and Martins AM (2014). Automatic segmentation and classification of blood components in microscopic images using a fuzzy approach. *Revista Brasileira de Engenharia Biomédica*, 30(4): 341-354.
- Yang Y, Cao Y, and Shi W (2014). A method of leukocyte segmentation based on s component and b component images. *Journal of Innovative Optical Health Sciences*, 7(01): 1-8.