White Blood Cells Nuclei Localization Using Modified K-means Clustering Algorithm and Seed Filling Technique

Mela G. Abdul-Haleem

Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq

Abstract

The presence of White Blood Cells (WBCs) in the body of human has a great role in the protection of the body against many pathogens. The recognition of the WBC is the first important step to diagnose some particular diseases. The pathologists usually use an optical microscope to recognize WBCs, but, this process is a quite tedious, time-consuming, error prone, very slow, and expensive. In addition, it needs experts with long practice in this field. For these reasons, a computer assisted diagnostic system that helps pathologists in the process of diagnosis can be effective, easy and safe. This research is devoted to develop a system based on digital image processing methods to localize WBCs nuclei. The proposed system involved a collection of pre-processing and segmentation algorithms that are capable of allocating the nuclei in different shapes of WBCs from a microscope images. To accomplish this task, a combination of local enhancement using histogram statistics, modified k-means clustering, normalization, convert to binary image using a suitable global threshold, islands removing and holes filling based on seed filling technique, and nucleus localization algorithms were performed. The features of WBCs images in the tested dataset make the WBC nuclei extraction process representing a great challenge. The test results indicate promising ability to completely isolate the nucleus from other parts of the cell. The analysis presents a high similarity between the ground truth samples and the results obtained by the proposed method. The precision percentage of the proposed method applied on the tested dataset images is 97.21% and F-score percentage is 96.23%.

Keywords: White blood cells, modified k-means clustering, nucleus segmentation, binary image, seed filling.

<div dir="rtl">

تحديد نوى خلايا الدم البيضاء باستخدام خوارزمية التجميع k-means المطورة وتقنية تعبئة البذور

ميلا غازي عبد الحليم

قسم علوم الحاسوب، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

ان وجود خلايا الدم البيضاء في جسم الإنسان له دور كبير في حماية الجسم ضد العديد من مسببات الأمراض. يعتبر التعرف على خلية الدم البيضاء الخطوة الأولى والمهمة لتشخيص بعض الأمراض المحددة. يستخدم علماء الأمراض عادةً مجهرًا ضوئيًا للتعرف على خلايا الدم البيضاء ، ولكن هذه العملية مضجرة للغاية ومستهلكة للوقت وعرضة للخطأ وبطيئة جدًا ومكلفة. بالإضافة إلى ذلك فهي تحتاج إلى خبراء ذوي خبرة طويلة في هذا المجال. ولهذه الأسباب ، يمكن لنظام تشخيصي بمساعدة الكمبيوتر يساعد أخصائيي علم الأمراض في عملية التشخيص أن يكون فعالاً وسهلاً وآمناً. يُخصص هذا البحث لتطوير نظام قائم على طرق معالجة الصور الرقمية لتحديد نوى خلايا الدم البيضاء. حيث تضمن النظام المقترح مجموعة من خوارزميات

</div>

---
*Email:a.mayla@yahoo.com

ما قبل المعالجة والتجزئة التي تكون قادرة على تخصيص النوى في أشكال مختلفة من خلايا الدم البيضاء
المحصلة من صور مجهرية. لإنجاز هذه المهمة ، تم الجمع بين تحسين محلي باستخدام إحصائيات الرسم
البياني ، تجميع K–means المطورة، والتطبيع ، والتحويل إلى صورة ثنائية باستخدام عتبة عالمية مناسبة ،
وإزالة الجزر وملء الثقوب على أساس تقنية تعبئة البذور ، وخوارزميات تحديد النواة. ان ميزات صور خلايا
الدم البيضاء في مجموعة البيانات المختبرة تجعل عملية استخراج نواة خلية الدم البيضاء تمثل تحديًا كبيرًا.
تشير نتائج الاختبار إلى قدرة واعدة على عزل النواة تمامًا عن أجزاء أخرى من الخلية. يقدم التحليل تشابهًا
كبيرًا بين عينات ground truth والنتائج التي تم الحصول عليها بواسطة الطريقة المقترحة. ان نسبة الدقة
للطريقة المقترحة المطبقة على صور مجموعة البيانات التي تم اختبارها هي 97.21٪ ونسبة درجة– ف هي
96.23٪.

1.   Introduction

Blood is a constantly circulating fluid through the body and half of the blood volume is composed of the Red Blood Cells (RBCs), White Blood Cells (WBCs), and platelets. WBCs contain nucleus and cytoplasm which can be classified into five main groups: basophils, Eosinophil, lymphocytes, neutrophils, and monocytes. WBCs are a momentous part of the immune system, so extracting information from them can assist pathologists in the diagnosis of different blood-related diseases [1]. Microscopic image is one of the most important resources to use for diagnosis and decision making on diseases. It requires analyzing the big amount of images and samples to assess the problem and discover better recovery procedure. The diagnosis and analysis of these images are mainly performed manually with the help of medical personnel. Since manual classification requires direct visual inspection, the process is prone to error and directly depends on the experience of a medical researcher. In addition, the medical devices used for cell analysis are expensive. This ineffective technique is highly time-consuming and tedious. And also, there is a possibility of different assumptions and interpretations about the same sample by different medical researchers. The medical personnel data input may not have uniformity. As it is known medical studies require data with high accuracy and reliability for decision making and interpretation [2]. Thus, a computer assisted diagnostic system that assists pathologists in the diagnosis process can be so effective.

Automatic identification technology of WBCs known for low-cost and homogenous accuracy has gained more and more focus in the domain of hematic related disease diagnosis. In general, the system consists of four parts: WBC segmentation, feature extraction, classification, and counting. Segmentation means decomposition of image in different part [3]. Since resulted segmentation directly affects the accuracy of classification and counting, it has considered a very hot topic in clinical diagnosis [4]. The complexity of internal characteristics of the WBC cells in addition to some external factors, such as illumination and different microscopic observations make the WBCs segmentation the greater challenge.

In this study, a simple method based on a set of image processing techniques is proposed to segment nucleus of WBCs in microscopy images efficiently. This set composes of modified k-means clustering to draw image in 12 color, thresholding, normalization, and segmenting processed image using the segmented color equation to use for matching and localization of nuclei of WBCs.

Some of related works are shown in the next section. A brief description for basic tasks involved in the proposed system is introduced in the materials and methods section. The main theoretical concepts are explained in theoretical level section. The results of the image enhancement and isolation of WBC nucleus, and the evaluation of the performance of the proposed model are explained in the experimental data and results section. Finally, the main conclusions derived in this work and a list of recommendations for future work is presented in the conclusion section.

2. Related Works

Some of the published works related to WBC or RBC segmentation techniques are given in the following paragraph:

In [5], the authors used active contours to follow WBC boundaries, but their method could not handle occluded cells accurately. In [6], the authors segmented nuclei using Otsu's thresholding method. They used active contours to obtain a precise boundary of cytoplasm. In [7], the authors compared Self-Organization Map (SOM) Based Fuzzy C-Means clustering algorithm to taken different types of remote sensing data to test the performance. They compared the quality measures

standard deviation, variance and also could detect the images based on various edge detection techniques for efficient segmentation. In [8], the authors proposed an adaptive method to extract, detect and calculate red blood cells in microscopic blood sample images. They used both k-means clustering technique and Hough transform to perform their task. In [9], the authors separated cells clusters using the distance transform and the watershed algorithm. The separation was less affected by cells shape, but it worked only for small or simple cells clusters. In [10], the authors discussed the use of Haar cascade classifier to determine the locations of white blood cells in an image. The results showed that their method has the ability to localize WBCs with 95% of precision values and 74% of recall values. In [11], the authors applied the Circular Hough transform (CHT) for recognizing the circular shapes inside the clumped cells after applying the segmentation phase. In [1], the authors used K-means algorithm for the segmentation of nuclei and enhanced the segmentation results by using morphological operators. In [12], the authors recognized the leukocytes using the Vector Field Convolution (VFC). They used VFC to determine cells edges and recognize them from the other blood cells. Finally, Miao and Xiao [13] presented a method to segment both of WBCs and RBCs in blood smear images. They modelled the edge gradient and shape features of these cells by identifying two transformations and efficiency using these transformations in a marker-controlled watershed.

3. Theoretical Level

3.1 Local Enhancement Using Histogram Statistics

The histogram statistics are used to enhance details over small areas in an image. In the context of image processing, the histogram means a histogram of intensity values of the pixel. This histogram shows the number of pixels in an image in each different intensity value. For colored images, the individual histogram for each of the red, green and blue channels can be calculated or three dimensions (3D) histogram with three axes representing the red, blue and green channels can be produced, while the number of pixels represents the brightness at each point [14, 15].

3.2 K-means Clustering

The K-means algorithm is the most used clustering algorithm depend on distance. This algorithm relies on distance as an evaluation parameter for similarity. This means that the shorter distance is more similar between the two objects. The clustering appears to group a multidimensional dataset by comparing similarity or difference values. K-means clustering splits n observations into K clusters in such a way so that each observation in k-means clustering associates to the cluster with the nearest mean, which acts as the prototype of the cluster. This process works efficiently, if the clusters are well separated and integrated. The steps of K-means algorithm are given below [16, 17]:

(1) Choose a set of k-cluster centroids at random.

(2) The algorithm continues alternately between two steps until the steps converge:

• The assignment step: determine the proper cluster that a sample must belong to depending on the calculated value of the squared Euclidean distance between the sample and each cluster centroid. Then, repeat this step for each sample.

• Update step: Calculate the new means for each cluster. These means become the centroids of the observations in new clusters.

The major problem of this algorithm is its reliance on the values of initial prototypes. If the initial prototypes are not selected carefully, the calculation will run on the chance of convergence to the local minimum instead of the global minimum. Thus, the proper selection of initializing prototypes has a significant impact on K-Means performance. In this research, a set of modifications related to data points is performed to overcome on the K-Means problem.

3.3 Image Normalization

Normalization is used to locate grey-level values within a desired range of values by adjusting the range of grey-level values [18, 19]. The removing of noise, illumination, or occlusion which is the result of image acquisition conditions is done by image normalization. The aim of this step is to get a standard image without any artifacts arising from the conditions of image acquisition [20]. To reduce the dynamic range of the gray-scale in an image, Hong *et al.* [21] proposed the following algorithm. Let N(i, j) denotes the normalized value to the gray-level value at the pixel (i, j). The normalized image is defined as follow:

$$N(i,j) = \begin{cases} \mu 0 \ + \ \sqrt{\dfrac{VAR_0 \ (V(i,j)-\mu e)^2}{VAR}}, \ if \ V(i,j) > \mu e \\ \mu 0 \ - \ \sqrt{\dfrac{VAR_0 \ (V(i,j)-\mu e)^2}{VAR}}, \qquad otherwise \end{cases} \quad \text{.... (1)}$$

The $\mu 0$ is the desired mean value and $VAR0$ is the desired variance value. In this research, the desired mean value is set 128 and the desired variance is taken 10000. The local gray-scale information is determined using a moving window with dimensions 11x11. At each window position instance, the mean μe and variance (VAR) is determined using the following equations:

$$\mu e(I) = \frac{1}{w \times h} \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} V(i,j) \qquad \text{... (2)}$$

$$VAR(I) = \frac{1}{w \times h} \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} (V(i,j) - \mu e(I))^2 \qquad \text{... (3)}$$

3.4 Binary Image

A binary image is an image that contains only two values per pixel. Black and white are the most used colors for a binary image. The used color for the objects in the image represents the foreground color and the rest of the image represents the background color. Binary images are usually called one-bit because each pixel in the image is stored as an individual bit. In digital image processing, binary images usually created as masks or result of specific operations such as thresholding, segmentation, and dithering operations [22].

3.5 Region-based Segmentation

In thresholding (binarization) and normalization approaches, every pixel is treated independently. However, region growing approach checks connectivity among pixels, in order to decide whether these pixels belong to the same region (object) or not. Mainly region growing algorithms use predefined similarity criteria for grouping the pixels in to large region, known as homogeneity criteria. It starts from randomly selecting a single pixel (a seed pixel) and region is grown around it, until the constraint on the resulting region satisfies homogeneity criteria [22].

Seed filling process aims to fill up all the holes inside the ridges while maintaining all the essential details that an observer wants to see in the original image. It is done by removing the white/dark regions whose area is below a predefined threshold [23]. A seed fill is an image processing operation that includes assigning the pixels in some region of the image a label. It starts by finding the seed point and then filling the object contours starting from that seed point. So, the seed filling algorithm consists of two parts. The first part of this algorithm includes finding a starting seed and discovering the connected pixels, while in the second part seed points can be searched automatically, or they can be chosen by the user [24]. In this research seed filling process is used for nucleus segmentation.

3.6 Segmentation Quality

To evaluate the quality of the segmentation results, quantitative measures in the pixel level which include precision, recall, and F-score are used. These measures define the difference between the results of the segmentation method and the ground truth of images as follows:

$$\text{Precision (P)} = \frac{Number of \ (SR \cap GT)}{Number \ of \ SR} \qquad \text{... (4)}$$

$$\text{Recall (R)} = \frac{Number of \ (SR \cap GT)}{Number \ of \ GT} \qquad \text{... (5)}$$

$$\text{F} - \text{score (F)} = \frac{2*(P*R)}{P+R} \qquad \text{... (6)}$$

Where SR is the Segmentation Result and GT is the mask of the Ground Truth of images.

4.   Materials and Methods

The proposed model has the essential steps to perform segmentation and all its related tasks, implementation of the segment system consists of seven major stages. The first stage includes computing histogram of RGB color spaces. The modified k-means algorithm is adopted to draw the image with twelve colors. The image is normalized to adjust gray-scale to the desired mean and standard deviation. Then, an automatic binarization is applied on the gray level image; it is based on

determination of optimum threshold value (i.e., global thresholds); it should cause efficient separation of objects from their background. The next step is segmentation using the island removal technique with certain conditions. The result of this step is a binary image contains only nucleus as a foreground and white background. The filling of holes is then applied to fill the closed pore in the nucleus object that may be created because acquiring conditions. Finally, the resulted image is masked with the original image to obtain isolated nucleus. The general structure of the proposed segmentation system is shown in the Figure-1.
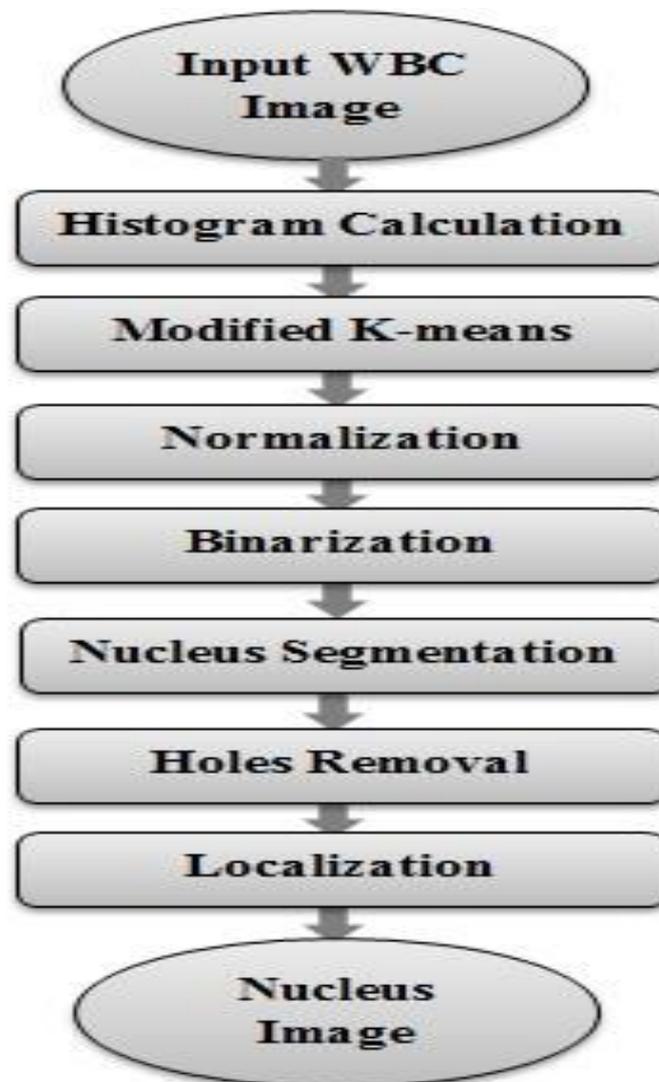
Figure 1- The proposed system architecture.

4.1   Read Input Image

The image is fed to the system as a Bitmap (BMP) image file; the color resolution of the image is taken 24 bit per pixel. The image data (i.e., RGB components) is loaded.

4.2   Local Enhancement Using Histogram Statistics

The operation is very easy. It includes scanning all image pixels in one pass and then the number of pixels found at each intensity value is retained.

4.3   Modified K-Means Clustering

To overcome on the K-Means problem, a set of modifications related to data points is performed. The intent is that to sort data point in descending order and select first twelve importance data points respectively, the new prototype positions is observed and each data point is then assigned to the nearest prototype. This data point will become a member of the cluster identified by that prototype. Algorithm (1) presents the implemented steps for histogram statistic and the modified K-Means method:

Algorithm (1): Histogram Statistic and Modified K-Means Clustering

Input : RGB () image
       h // height of image
       w // width of image
Output: org () //array of clustering areas
Goal : *convert image into 12 colors only*

Step 1: Define array cent (w, h).
Step 2: Define his () array with a 3D size, color-index (3 represent RGB, number of centroids).
Step 3: Represent each color in each pixel with 4 bits (division each color by 64).
  For all pixels in the image // x as rows and y as columns
    $i \leftarrow$ red (x, y) / 64
    $j \leftarrow$ green (x, y) / 64
    $k \leftarrow$ blue (x, y) / 64
Step 4: Calculate histogram.
    his (I, j, k) $\leftarrow$ his (I, j, k) + 1
  End For
Step 5: Select the centroids and initialize centroids counter c1.
  Loop on his
  Check if his (I, j, k) > 2 then
    cent (c1) $\leftarrow$ his (I, j, k)
    Increment c1
  End if
Step 6: Sort the centroid's points in descending order.
  Loop on centroids points
  Check if cent < cent (c + 1) then
  cent (c) $\leftrightarrow$ cent (c + 1)
  End if
Step 7: Obtain color index Look Up Table (LUT) and fill it with RGB values by multiplying each color by 64.
      For all z = 0 $\rightarrow$ c1 – 1
 check If cent(z) = his(I, j, k) Then
 a $\leftarrow$ (I + 0.5) * 64 //red
 b $\leftarrow$ (j + 0.5) * 64 //green
 c $\leftarrow$ (k + 0.5) * 64 //blue
      color-index (0, z) $\leftarrow$ a
      color-index (1, z) $\leftarrow$ b
 color-index (2, z) $\leftarrow$ c
End if
End for
Step 8: Clustering each pixel value of the image based on LUT to the nearest centroid.
For all pixels of image
index $\leftarrow$ nearest (red (I, j), green (I, j), blue (I, j))
new-red $\leftarrow$ color-index (0, index)
    new-green $\leftarrow$ color-index (1, index)
new-blue $\leftarrow$ color-index (2, index)
org(I, j) $\leftarrow$ Abs((new-red + new-green + new-blue) / 3)
End For
Step 9: Return org ().

4.4 Image Normalization

In this research, the process is performed using equation (1). The desired mean value is set 128 and the desired variance is taken 10000. The local gray-scale information is determined using a moving window with dimensions 11x11. At each window position instance, the mean (μe) and variance (VAR) is determined using equations (2) and (3). The normalization is applied to red, green, and blue separately because each of these colors is represented by 256 bits / one byte (i.e., gray-level).

4.5 Binary Image

In the developed system, this process includes calculating a global threshold automatically by taking the average of pixel multiplied with $\alpha \in [0.1\text{-}0.9]$. The value of α depends on the brightness of the image. When the threshold value is taken as the average of intensity values, the dark region is located in the range of $[0, T\text{-}1]$ while the bright region is located in the range of $[T, 255]$. Multiplying the threshold value by α will lead to decrease the bright region and increase in the dark region or increase the bright region and decrease the dark region depending on the selected value of α. The effect of using inclusion parameter α is discussed in section 5.2 and shown in Figure-4. After selection the threshold value, comparing the value of each pixel belong to the normalized gray image with the determined threshold value to decide whether each pixel belong to foreground or background. The main steps of this stage are explained in algorithm (2).

Algorithm (2): Binary Image
Input : N () // array of normalized image
Output: Bin() // array of binary image
Goal :   *Convert to gray and then binary image*

Step 1: Convert RGB components of N () to gray image.
 For all pixels of image N ().
    Gray (I, j) ← (R (I, j) + G(I, j) + B(I, j)) / 3
 End For
Step 2: Calculate minimum and maximum gray values.
 min ← Gray (0, 0)
 max ← Gray (0, 0)
For all pixels of image // I as row index & j as column index
   Check If (max < Gray (I, j)) Then max ← Gray(I, j)
Check If (min > Gray (I, j)) Then min ← Gray(I, j)
End For
Step 3: Find the threshold value.
  T ← ((max + min) / 2) * α
Step 4: Find the white and black pixels.
        For all pixels of image
 Check If Gray (I, j) < T Then
   Bin (I, j) ← 0
            Else
               Bin (I, j) ←1
        End If
End For
Step 5: Return Bin ().

4.6 Nucleus Segmentation

The nucleus of WBC is seen as a special island in the image. Using the seed filling principle with appropriate constraint lead to optimal segmentation of the nucleus by removal all other islands in the image. Also, the seed filling method can be used to fill the closed pores in the nucleus object after segmentation.

In this technique, all holes found in the image will be filled in and all islands will be removed. This process uses seed filling method explained in algorithm (3) either for filling each hole (close pores) or removing the island as shown in algorithm (4). The first stage of this technique is the identification of seeds. In case of filling the hole the seed point has background value and in case of island removal the seed point has foreground value. The second stage includes constraints to prevent each region from overfilling. In this work, the selected constraint is that the selected filling region in case of segmentation must form a very small ratio of the image area (C= 0.03) otherwise it will be ignored. The effect of this parameter is explained in section 5.2 and Figure-3. On the other hand, the region of hole must form the ratio (C=0.2) of the image size. The same steps in algorithm (3) are applied for filling the holes by replacing each 0 to 1 and each 1 to 0.

Algorithm (3): Seed Filling Function

Input : p // target pixels
       I // x-position of the target pixel
       J // y-position of the target pixel
Output: *SS* // the number of pixels in seed filling region
Goal :   *The number of pixels in seed region*

Step 1: Enter the pixel positions in an array Z.
Step 2: Initialize pixel count SS to zero.
Step 3: Search in the four directions of the seed point.
      While *Z* is not empty
 b ← the last element of *Z*
 Remove the last element from *Z*
 SS ← SS +1
      Check If the color of b is equal to *target-color*
     Check If the west pixel value equal to the target value
     Add west pixel to end of Z
     End If
     Check If the east pixel value equal to the target value
      Add east pixel to end of Z
     End If
     Check If the north pixel value equal to the target value
     Add north pixel to end of Z
     End If
     Check If the south pixel value equal to the target value
      Add south pixel to end of Z
     End If
 End If
     End While
Step 4: Return SS.

Algorithm (4): Island Removal

Input : Bin () // array of binary image
     h // height of image
     w // width of image
Output: y () //array of island removal image
Goal : Nucleus Segmentation

Step 1: Enter the constant criteria parameter C.
Step 2: Count the region ratio.
     rat← C × (h × w)
Step 3: Search for each target pixel in all image and call seed function to each ones.
 For all pixels in the image// x as row number, y as column number
   Check If pixel value = 1 Then
      S ← seed (1, x_position, y_position) // algorithm (2)
Step 4: Compare the number of pixels in the seed region with the region ratio.
Check If S ≤ rat Then// remove region from the image
For all pixel in the image // i as row number, j as column number
Check If seed region (i, j) = 1Then y (i, j) = 0 End If
    End For // i& j
 End If
 End If
End For
Step5: Return y ().

Algorithm (4): Island Removal

Input : Bin () // array of binary image
        h // height of image
        w // width of image
Output: y () //array of island removal image
Goal : *Nucleus Segmentation*

Step 1: Enter the constant criteria parameter C.
Step 2: Count the region ratio.
       rat← C × (h × w)
Step 3: Search for each target pixel in all image and call seed function to each ones.
 For all pixels in the image// x as row number, y as column number
   Check If pixel value = 1 Then
        S ← seed (1, x_position, y_position) // algorithm (2)
Step 4: Compare the number of pixels in the seed region with the region ratio.
Check If S ≤ rat Then// remove region from the image
For all pixel in the image // i as row number, j as column number
Check If seed region (i, j) = 1Then y (i, j) = 0 End If
       End For // i& j
  End If
  End If
End For
Step5: Return y ().

4.7 Localization

The last stage in the system is the localization. This process contains only using of a mask from the original image to the image resulting from the previous step as shown in algorithm (5).

Algorithm (5): Nucleus Localization

Input : y () //array of island removal image
        RGB () image// original image
Output: New-pixel () // WBC nucleus image
Goal : *Nucleus Localization*

Step 1: Put origin image pixels to xx ().
For all pixels of origin image
x ←RGB (I, j)
Step 2: Mask with segmented image pixel y ().
Check If y (I, j) = 1 Then 'nucleus pixel
  New-pixel(I, j)← x
Else
  New-pixel (I, j) ← white //background value
End If
Step 3: Return New-pixel ().

5.   Results and discussion
5.1 Test Material

The system was tested on a dataset of WBCs images downloaded from the standard reference database called CellaVision blog _ dataset2. The images of this dataset contain many RBCs existed around the WBCs and have generally purple color. The number of the images in this dataset is one hundred colored images with 300×300 dimensions [25]. The cell images have purple color in general and many RBCs may exist around the WBCs. Figure-2 shows an example of each type of WBC with its final localization obtained by the proposed model.
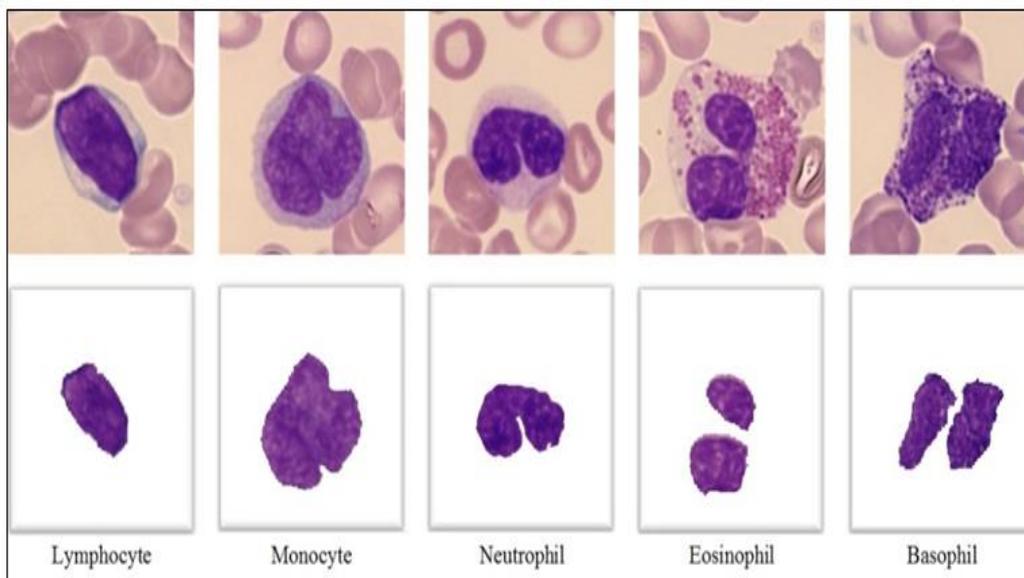
Figure 2- A sample of each type of WBC with its final localization obtained by the proposed model.

5.2 Performance Parameters

In this work, nucleus segmentation applied by using the seed filling method by removing the islands which represent the other objects in WBC. The removed islands must have an area less than a specific ratio of the image area. This ratio has a great effect on the performance of the developed system. Figure-3 shows the resulted image using different ratio values. When the ratio equals to (0.01%) or (0.02%) of the image area, the other objects of the cell will appear in the image and when the ratio equals to (0.05%) from the image area, the nucleus of the cell will disappear completely in the image. While a good nucleus segmented is obtained when the ratio equals to (0.03%) or (0.04%).

Another effective parameter is the inclusive factor α used in calculation the threshold value. Figure-4 shows the effect of using or not using this factor in the process of transforming the image into a binary image and its effect on the performance of the system. When the threshold value is calculated as the average value of the intensity values without multiplying by α (i.e., α is set 1), the background is not completely separated from the foreground as shown in Figure-4B and when α is set 0.8, the seed filling technique will not work correctly because of the appearance of gaps in the border of the nucleus (i.e., erosion will happen) as shown in Figure-4D. Figure-4C shows that the good segmentation is obtained when α is set 0.9. For the selected dataset, when α is set to a value less than 0.8, a white image will be obtained at the final stage of the system.
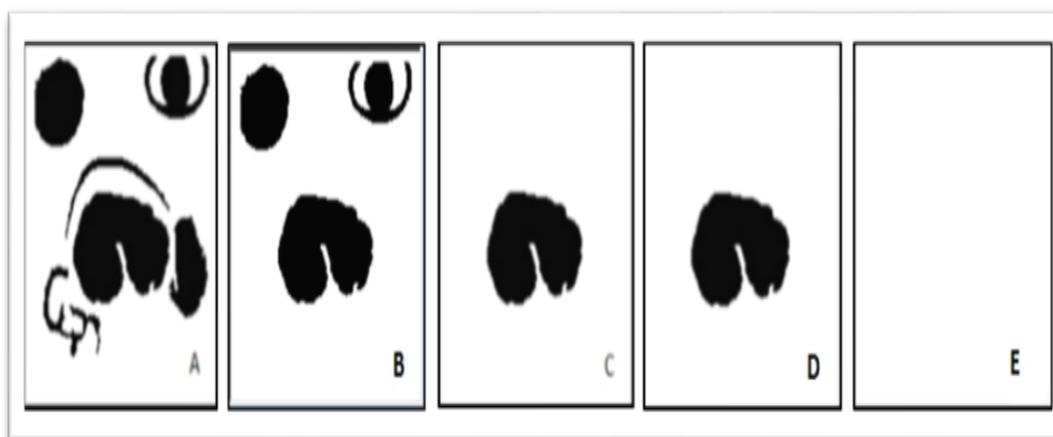
Figure 3- The result after island removal using the different ratio. (A) The ratio equals to (0.01%) from the image area; (B) the ratio equals to (0.02%) from the image area; (C) the ratio equals to (0.03%) from the image area; (D) the ratio equals to (0.04%) from the image area; (E) the ratio equals to (0.05%) from the image area.
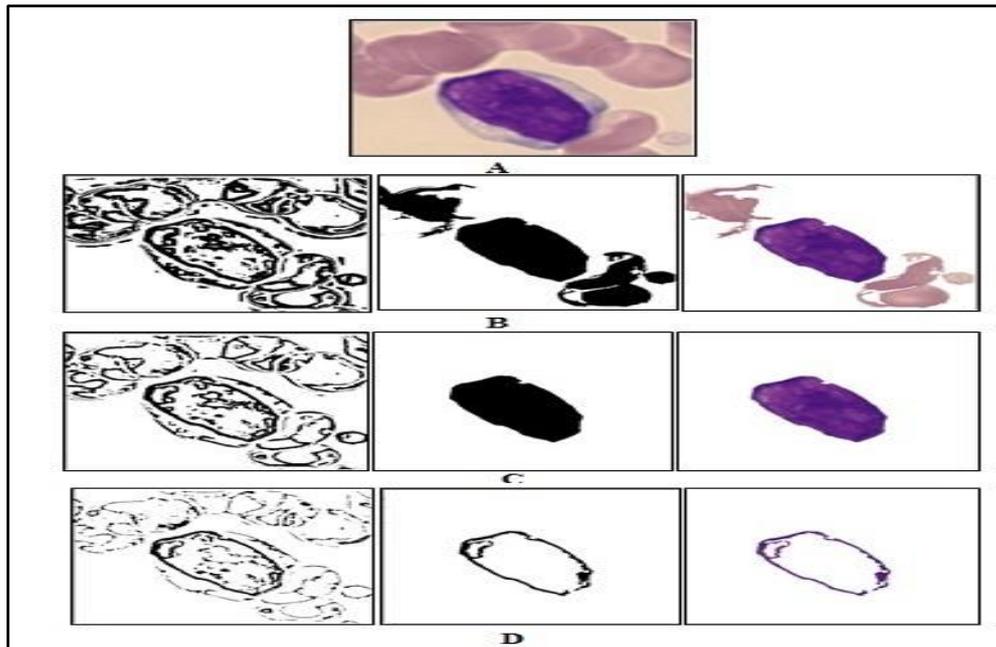
Figure 4- The effect of α on the performance of the system. (A) the input image; (B) the binary image and subsequent stages when the selected value of α=1; (C) the binary image and subsequent stages when the selected value of α=0.9; (D) the binary image and subsequent stages when the selected value of α=0.8.

5.3 Quantitative Measures

To evaluate the quality of the proposed segmentation results, quantitative measures explained in Equations (4-6) were used. These measures define the difference between the results of the proposed nucleus localization method and the mask of the white region of the ground truth of images. Figure-5B shows the ground truth of the selected sample in Figure-5A. The white region of the ground truth can be easily obtained from the image using a binary mask as shown in Figure-5C. The segmentation result for selected image data (in Figure-5D) is compared and evaluated using the ground truth image in Figure-5C.
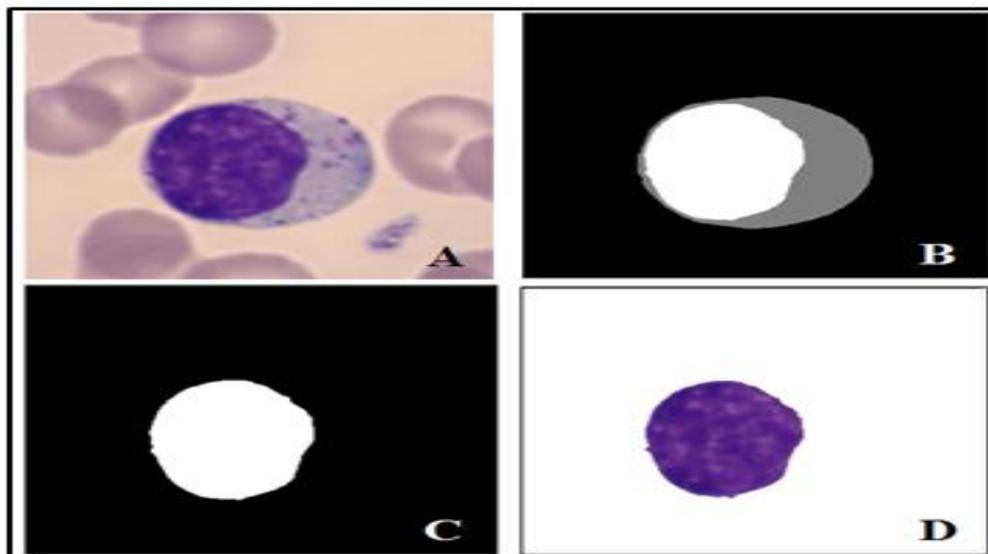
Figure 5- The segmentation of nucleus images. (A) An original sample; (B) the ground truth; (C) the mask of the white region of the ground truth; (D) the result of the nucleus segmentation method.

Each sample of the used dataset was tested using each of the quantitative measures. Table-1 shows the average values of the quality measure values for the results of the segmentation process approximated to fourth decimal points. The quality measure values for the segmentation results for all tested images are shown in Figures-(6-8). These figures show that the proposed segmentation method is robust. It has high precision values in a range (0.9541-0.9953) as shown in Figure-6 and high F-measures in a range (0.8862-0.9970), while keeping good recall values ranging between (0.8272-1).

Table 1- The average values and the percentage of the quality measure values for the results of the segmentation process.

Precision	Precision %	Recall	Recall %	F-score	F-score %
0.9721	97.21%	0.9546	95.46%	0.9623	96.23%

The performance of the nucleus segmentation method is qualitatively evaluated with the consideration that no manual outlining has been done on the WBC images. The smallest error between the segmentation result and ground truth demonstrates that the modified k-means method followed by the seed filling technique obtained robust and precise WBC nuclei segmentation results.
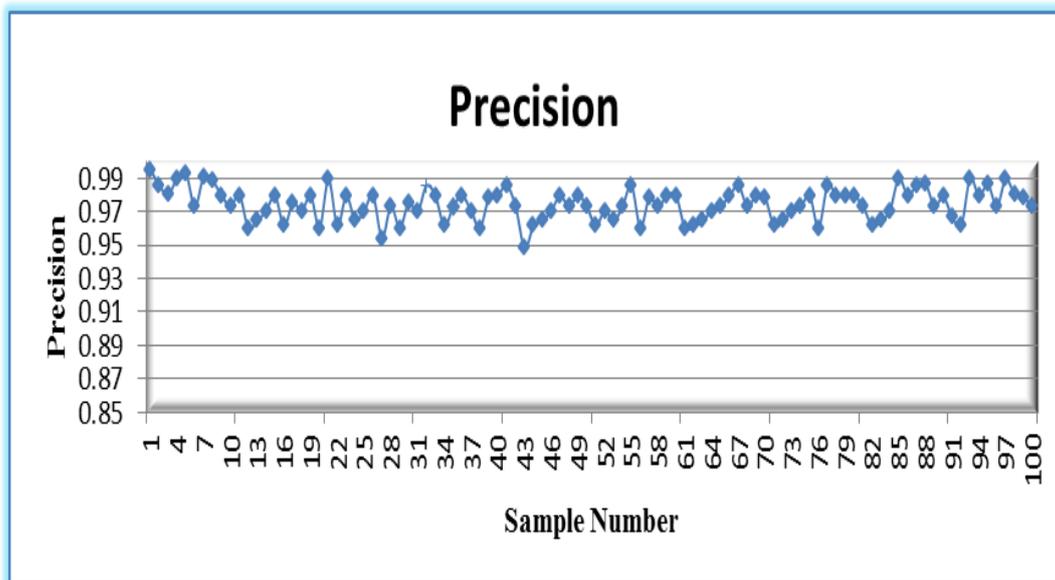
Figure 6- Precision of the proposed nucleus localization method applied on the tested dataset images.
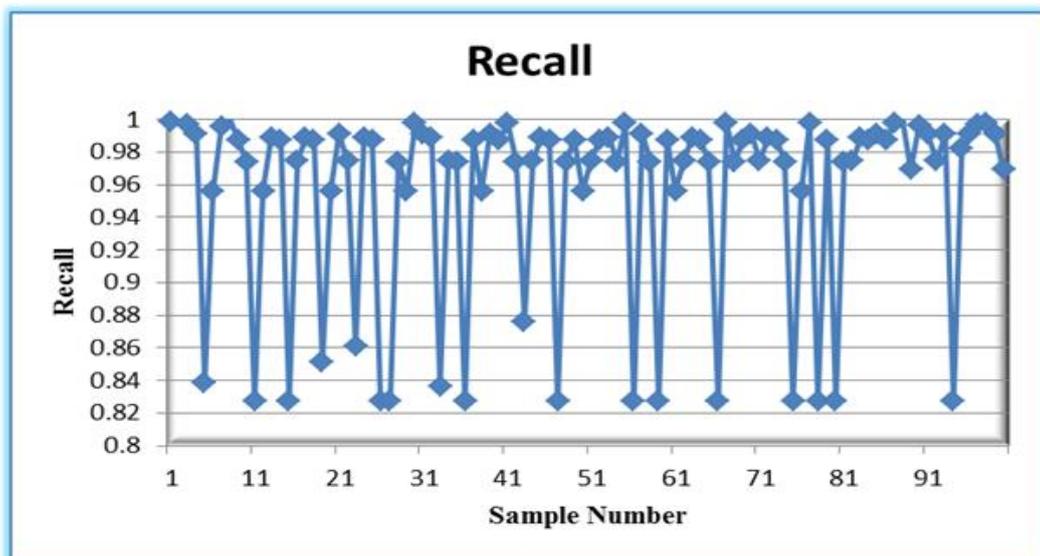
Figure 7- Recall of the proposed nucleus localization method applied on the tested dataset images.
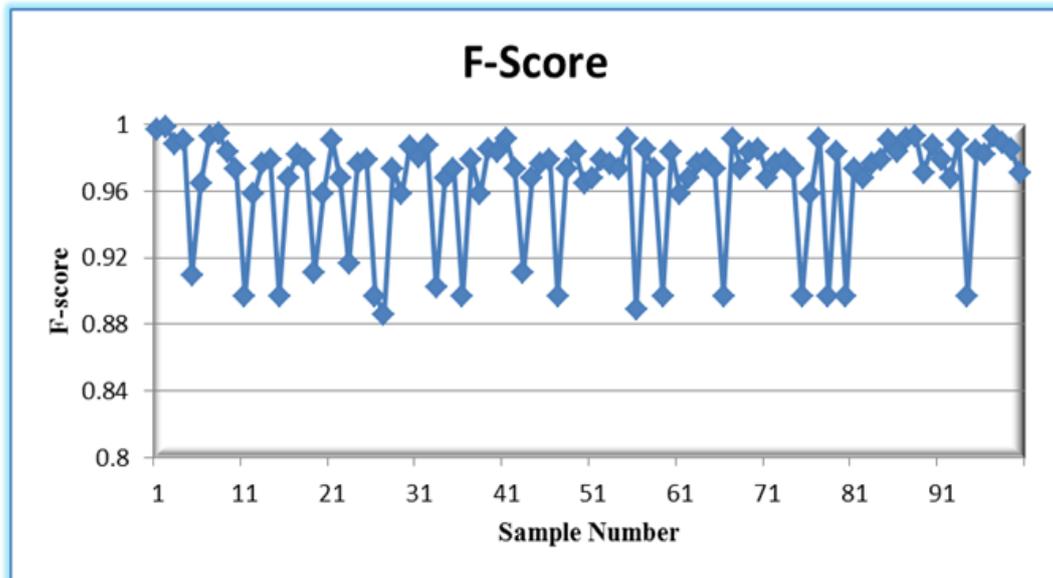
Figure 8- F-score of the proposed nucleus localization method applied on the tested dataset images.

5.4 Implementation

Each stage of the system plays a critical role in the localization of WBC nucleus. The results of these stages for two samples are shown in Figure-9.
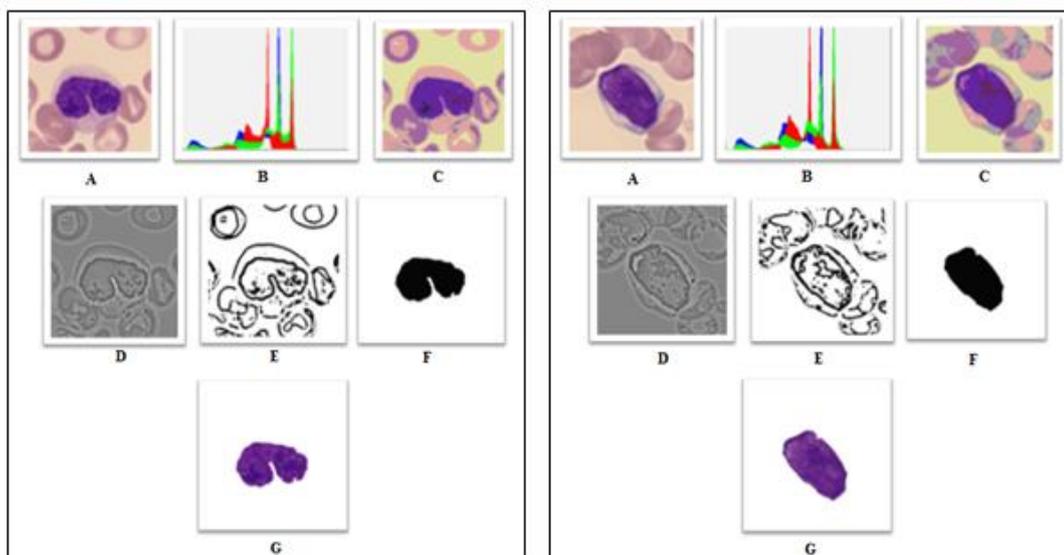
Figure 9- The implementation of the system stages for two samples. (A) the input image; (B) Histogram; (C) K-means classification; (D) the normalized image; (E) the gray image and global thresholding; (F) island/hole removal; (G) nucleus localization.

5.5 Location Quality

In this section, the proposed method is compared with other segmentation methods in [26], [27], [28], [29] and [30]. In theory, the neural networks classification in [26] could not process the background well, either the darker RBCs would be appeared, or the lighter WBCs would be disappeared. The authors in [27] performed nuclei segmentation based on color transformation and watershed techniques. Although, the recall degree is higher than the recall degree of the proposed method, but the F-score and precision values are lower than those of the proposed method. In [28] the authors segmented the nuclei using dynamic convolution filter, but the nuclei region in the ground truth images are located manually. So, the results could not be very accurate. Neither the watershed

technique in [28] nor the given learning-based algorithm to detect leukocyte depending on bottom-up saliency detection in [30] obtain the precision percentage that the proposed method gains. In [29], the authors used discrete wavelet transform, morphological filtering and Otsu's thresholding techniques. But this procedure also not works well with very dyeing light images. Table-2 shows the detailed results from the different methods. Experiments determine that the other WBCs segmentation methods may have either low recall or precision rate, but the proposed method has high degrees of all the quantitative measures.

Table 2- Comparison of the proposed method with other methods in the area

Methods	Precision	Recall	F-score
Neural networks classification in [26]	0.8534	0.7049	0.8001
Color-based segmentation	0.9409	0.9800	0.9601
Shape-based segmentation in [27]	0.8876	0.9578	0.9214
Watershed technique in [28]	0.9216	0.8955	0.9083
Otsu's thresholding in [29]	0.8981	0.7897	0.8459
Learning-based algorithm in [30]	-	-	0.9500
The proposed method	**0.9721**	**0.9546**	**0.9623**

6.   Conclusion

Several conclusions have been deduced from the obtained test results. It is clear to observe that the proposed model gave good results for totally isolation of WBC nuclei. The correct selection of the filling region in the nucleus segmentation algorithm highly influences on the desired result. The region that forms the ratio (0.03%) of the image area gave the best possible result according to the tested dataset; of course dataset with different images quality, will need another value. The successful isolation of image objects from the background in the binarization process depends heavily on the good selection of threshold value which in turns depends on the value of (α).

This work can be extended in different directions. Extend the developed system to include cytoplasm and cell segmentation, in which it could be more useful for detecting diseases. Add more enhancement methods so the system will be accepting more dataset with a low image quality. Using other segmentation methods, test their efficiencies in the WBC segmentation field and make a comparison between them like Hue-Saturation-Intensity (HSI), naïve Bayes classifiers, Edge detection and the other segmentation methods.

Acknowledgements

References

1. Ghane, N., Vard, A., Talebi, A. and Nematollahy, P. **2017**. Segmentation of white blood cells from microscopic images using a novel combination of K-means clustering and modified watershed algorithm. *Journal of medical signals and sensors*, **7**(2):92-101.
2. Karki B. J. **2016**. A Study on automated process for extracting white blood cellular data from microscopic digital injured skeletal muscle images. UNLV Theses, Dissertations, Professional Papers, and Capstones. 2689, University of Nevada, Las Vegas.
3. Joshi1, M. D., Karode, A. H. and Suralkar, S. R. **2013**. White blood cells segmentation and classification to detect acute leukemia. *International Journal of Emerging Trends & Technology in Computer Science*, **2**(3):147-151.
4. Li, Y., Zhu, R., Mi, L., Cao, Y. and Yao, D. **2016**. Segmentation of white blood cell from acute Lymphoblastic Leukemia images using dual-threshold method. In: Udupa J K (ed.). *Computational and Mathematical Methods in Medicine*, Article ID 9514707. 12 p. https://doi.org/10.1155/2016/9514707.
5. Ongun, G., Halici, U., Leblebicioglu, K., Atalay, V., Beksac, M. and Beksac, S. **2001**. Feature Extraction and Classification of Blood Cells for an Automated Differential Blood Count System. Proceedings of IJCNN '01. International Joint Conference on Neural Networks, Jul 15-19, Washington, DC, USA, USA, IEEE, p. 2461-2466.

6. Hamghalam, M., Motameni, M. and Kelishomi, A.E. **2009**. Leukocyte Segmentation in Giemsa-Stained Image of Peripheral Blood Smears Based on Active Contour. Proceedings of 2009 International Conference on Signal Processing Systems, May 15-17, Singapore, Singapore. USA, IEEE, p. 103-106.

7. Napoleon, D., Subramanian, M. S., Praneesh, M. and Sathya, S. **2012**. Self-organizing map-based color image segmentation with fuzzy C-Means clustering and saliency map. *International Journal of Computer Application*, **3**(2): 109-117.

8. Mogra, M., Bansel, A. and Srivastava, V. **2014**. Comparative analysis of extraction and detection of RBCs and WBCs using Hough transform and k-means clustering algorithm. *International Journal of Engineering Research and General Science*, **2**(5): 670-674.

9. Putzu, L., Caocci, G. and Di Ruberto, C. **2014**. Leucocyte classification for leukaemia detection using image processing techniques. *Artificial Intelligence in Medicine*, **62**(2014): 179-91. https://:dx.doi.org/10.1016/j.artmed.2014.09.002.

10. Budiman, R.A., Achmad, B., Faridah, Arif, A., Nopriadi and Zharif, L. **2016**. Localization of White Blood Cell Images Using Haar Cascade Classifiers. Proceedings of the 1st International Conference on Biomedical Engineering (IBIOMED)*,* Oct 5-6, Yogyakarta, Indonesia. Yogyakarta, IEEE, P. 1-5.

11. Di Ruberto, C., Loddo, A. and Putzu, L. **2016**. A leucocytes count system from blood smear images. *Machine Vision and Application*, **27**(8): 1151 - 1160. https://doi.org/10.1007/s00138-016-0812-4.

12. Porcu, S., Loddo, A., Putzu, L. and Di Ruberto, C. **2018**. White Blood Cells Counting Via Vector Field Convolution Nuclei Segmentation. In: Tremeau A, Imai F and Braz J (eds.). Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), Jun 27-29, Funchal, Madeira, Portugal, SciTePress. **4**(VISAPP), P. 227-234. DOI: 10.5220/0006723202270234.

13. Miao, H., and Xiao, C. **2018**. Simultaneous segmentation of Leukocyte and Erythrocyte in microscopic images using a marker-controlled watershed algorithm. In: Migliore M (ed.), *Computational and Mathematical Methods in Medicine*, Article ID 7235795. 9p. https://doi.org/10.1155/2018/7235795.

14. Davies, E. R. **2012**. *Computer and machine vision: theory, algorithms and practicalities*. 4th Ed. USA. Academic Press. 912 P.

15. Casado C. O. **2010**. *Image contrast enhancement methods*. Bachelor of Telecommunications Degree, Technical University, Sofia.

16. Zhang, C., Xiao, X., Li, X., Chen, Y., Zhen, W., Chang, J., *et al.* **2014**. White blood cell segmentation by color-space-based K-Means clustering. *Sensors (Basel)*, **14**(9):16128-16147. DOI: 10.3390/s140916128.

17. Gowda, P. and Kumar, P. **2017**. Segmentation of white blood cell using K-Means and Gram-Schmidt orthogonalization. *Indian Journal of Science and Technology*, **10**(6):1-6. DOI:10.17485/ijst/2017/v10i6/111137.

18. Ellerbusch S. C. **2008**. *Image stitching using partial latent fingerprints*. M.Sc. Thesis, Florida Institute of Technology, Melbourne, Florida. 136p.

19. Verma, M. and Agarwal, S. 2009. Fingerprint Based Male-Female Classification. In: Corchado E., Zunino R., Gastaldo P., Herrero Á. (eds). *Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08*. Advances in Soft Computing, Springer, Berlin, Heidelberg, 53. 251–257.

20. Bart, E. and Ullman, S. **2004**. Image Normalization by Mutual Information. In: Hoppe A, Barman S and Ellis T (eds). *Proceedings of the British Machine Vision Conference*, Sep 7-9th; Kingston University, London, BMVA Press, 35.1-35. 10 p. DOI:10.5244/C.18.35.

21. Hong, Wan, Y. and Jain, A. K. **1998**. Fingerprint image enhancement: algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(8):777-789. DOI: 10.1109/34.709565.

22. Mendoza, F. and Lu, R. **2015**. Basics of Image Analysis. In: Park B and Lu R (eds). *Hyperspectral imaging technology in food and agriculture*. New York: Springer, 9-56.

23. Zhao, F. and Tang, X. **2007**. Preprocessing and postprocessing for skeleton-based fingerprint minutiae extraction. *Pattern Recognition*, **40**(2007):1270–1281. https://doi.org/ 10.1016/ j. patcog.2006.09.008.

24. Khayal, M. S., Khan, A., Bashir, S., Khan, F. H. and Aslam, S. **2011**. Modified new algorithm for seed filling. *Journal of Theoretical and Applied Information Technology*, **26**(1): 28-32.

25. https://github.com/zxaoyou/segmentation_WBC.

26. Mohapatra, S., Patra, D., Kumar, S. and Satpathy, S. **2012**. Lymphocyte image segmentation using functional link neural architecture for acute leukemia detection. *Biomedical Engineering Letters*, **2**(2): 100-110.

27. Arslan, S., Ozyurek, E. and Gunduz-Demir, C. **2014**. A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images. *Cytometry Part A*, **85A**: 480-490.

28. Abbas, N., Mohamad, D., Abdullah, A. h., Saba T., Al-Rodhaan, M. and Al-Dhelaan, M. A. **2015**. Nuclei segmentation of leukocytes in blood smear digital images. *Pakistan Journal of Pharmaceutical Sciences*, **28**(5):1801-1806.

29. Tareef, A., Song, Y., Cai, W., Wang, Y., Feng, D. D., and Chen, M. **2016**. Automatic Nuclei and Cytoplasm Segmentation of Leukocytes with Color and Texture-based Image Enhancement. Proceeding of 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 13-16 April 2016; Prague, Czech Republic, IEEE, p. 935–938. DOI: 10.1109/ISBI.2016.7493418.

30. Pan, C., Xu, W., Shen, D. and Yang, Y. **2018**. Leukocyte image segmentation using novel saliency detection based on positive feedback of visual perception. *Journal of Healthcare Engineering,* **2018**, 5098973. DOI: 10.1155/2018/5098973.