

# Microscopic Digital Image Segmentation And feature Extraction of Acute Leukemia

Alaa Shakir Ahmed  
AL Mansoura University Faculty of Engineering  
Electronics and Communication Department  
Mansoura - Egypt.

Dr. M. Morsy  
AL Mansoura University Faculty of Engineering  
Electronics and Communication Department  
Mansoura - Egypt.

Mohy Eldin A. Abo-Elvoud  
AL Mansoura University Faculty of Engineering  
Electronics and Communication Department  
Mansoura - Egypt.

**Abstract**— The goal of this paper is to identify and show the differences in the properties of Acute lymphoblastic leukemia (ALL) and normal white blood cells. This paper was conducted on a set of microscopic digital images of blood samples that got it from the “Oncology Center - Faculty of Medicine - Mansoura University Hospital - Egypt” is made up of 50 microscope image samples of Cancerous blood and 50 microscope image of the blood samples is not Cancerous (normal blood). The microscope blood images are undergo to chain of pre-processing steps which include resizing image such as (512\*512, 256\*256, 128\*128) and contrast enhancement. By executing K-means clustering on the resultant images, the cell's nucleus under consideration is obtained then these segmented images enter sub-image stage. The next step is Extracted Features that included: Shape features (Area, Perimeter, Compactness, Solidity, Eccentricity, and Elongation); texture features (Homogeneity, Energy, Correlation, Entropy, and Contrast); color features and Fractal Dimension.

**Keywords** — (Digital Image Processing , Acute Leukemia , Normal White Blood Cell , Contrast Enhancement , K-Means Clustering , Features Extraction)

## I. INTRODUCTION

Acute lymphoblastic leukemia (ALL), also known as acute lymphocytic leukemia, or acute lymphoid leukemia, is an acute form of leukemia, or cancer of the white blood cells, distinguished by the overproduction and cumulation of cancerous, immature white blood cells, known as lymphoblasts [1]. Leukemia is "a cancer of the blood cells (WBC). It starts in the bone marrow when abnormal cells (infected WBC) redouble out of control To the extent that normal blood cells (uninfected WBC) are incapable to develop". Which usually affects blood, bone marrow, and lymph nodes. It is distinguished by proliferation of abnormal white blood cells (leukocytes) in the bone marrow without responding to cell outgrowth inhibitors [2]. Four Most Common Kinds of Leukemia:

1. **Acute myeloid leukemia (AML)** influence myeloid cells and grows rapidly. Leukemic blasts cells gather in the bone marrow and blood. About 15,000 Americans was diagnosed with AML in 2013. Most (about 8,000) was 65 or older, and about 870 children and teens will infect this disease.

2. **Acute lymphoblastic leukemia (ALL)** affects lymphoid cells and grows rapidly. Leukemic blast cells usually collect in

the blood and bone marrow. More than 6,000 Americans was diagnosed with ALL in 2013. Most of them (more than 3,600) were children and teens.

3. **Chronic myeloid leukemia (CML)** influences in myeloid cells and typically grows slowly at first. Blood tests show a rise in the number of white blood cells. The abnormal blood cells work okay. May be there are a small number of leukemic blast cells in the bone marrow Nearly 6,000 Americans was diagnosed with CML in 2013. Nearly half (about 2,900) were 65 or older and only Nearly 170 children and teens will infect this disease.

4. **Chronic lymphocytic leukemia (CLL)** influences in lymphoid cells and usually grows rapidly. Blood tests show a rise in the number of the white blood cells. The actin of the abnormal cells nearly as well as the normal white blood cells [3], [4], Figure (1) shows the Kinds of Leukemia.

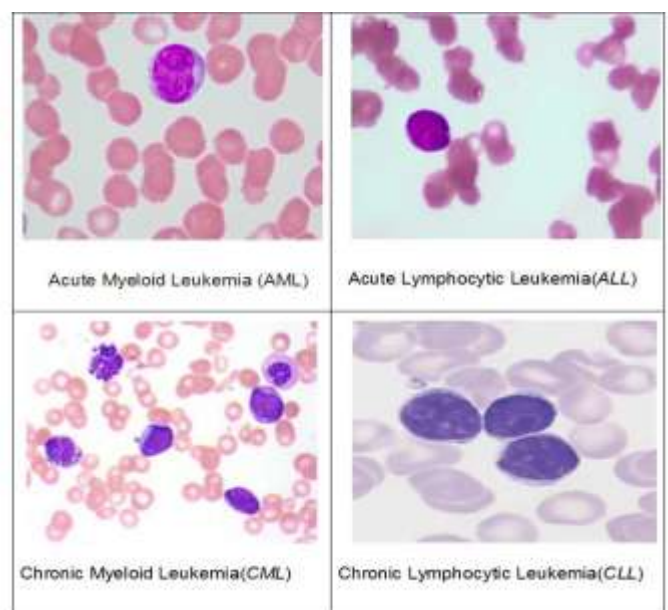


Figure (1) Kinds of Leukemia

Figure (2) and Figure (3) show the difference between normal (Non-Cancerous) blood smear and abnormal (Cancerous) blood smear the one affected with Acute Leukemia.

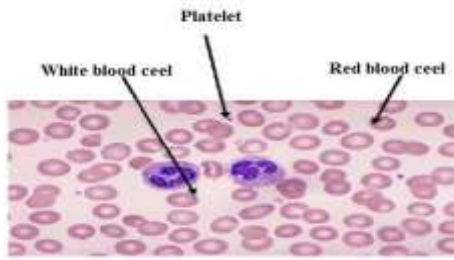


Figure (2) Non-Cancerous Blood Smear (Normal)

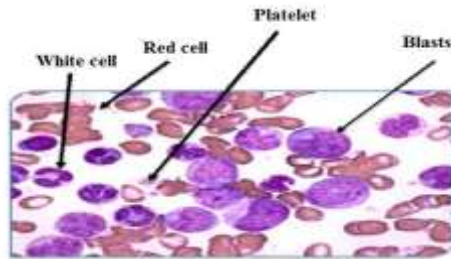


Figure (3) Cancerous Blood Smear (Abnormal)

## II. DIGITAL IMAGE PROCESSING (METHODS)

The procedure of Segmentation and Features extraction for Acute lymphoblastic leukemia (ALL) in microscopic blood images consists of pre-processing (resize and contrast image), segmentation using (k-means clustering), feature extraction (shape – texture – color – HD). The proposed system is shown in Fig. (4).

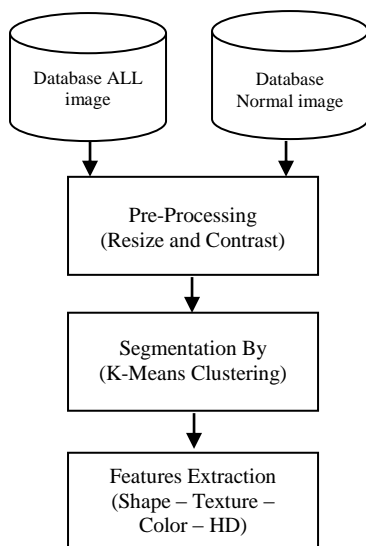


Fig.(4). System Overview

### A. Pre-processing

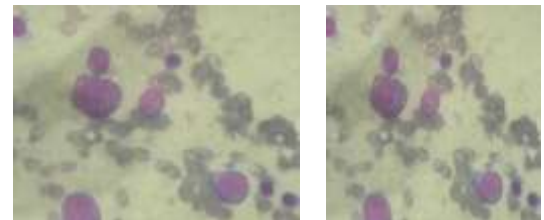
Pre-processing methods can be splitted into the two groups according to the goal of the processing:

First: - Image resizing such as 512\*512, 256\*256 and 128 \* 128.

Second: - Contrast enhancement by employ the “*fspecial*” function with the “*unsharp masking*” filter has the effectiveness of making edges and fine detail in the image more crisp [5], then apply this mask filter on image by using “*imfilter*” function with Boundary Option “*replicate*” that Input array values outside the bounds of the array are assumed to equal the nearest array border value [6].

Algorithm: Pre-processing using unsharp (Filtering)

- Read input image (X).
- Use  $Y = \text{imresize}(X, [512*512])$ .
- Create new variable color filtered for having same attribute that of input image using unsharp masking filter.
- Apply boundary option replicate filter algorithm on s Merge all three planes together eparated RGB planes.
- Merge all three planes together.
- Output (Y) shown in Fig. (5).



(X) Input Image (Y) Resize 512\*512 & Contrast Image

Fig. (5)

### B. Segmentation

Segmentation is executed in two stages for extracting WBC nucleus from the blood microscopic images using color based clustering. Initial segmentation are completed by K-means clustering followed by nearest neighbor classification in  $L^*a^*b^*$  space. K-means is a semi supervised clustering technique which is employ to create K clusters from n observations. It is aiming to achieve partition such that objects within each cluster are as near to each other as possible, and as far from object in the other clusters as possible [7]. Each pixel of an object is classified into four clusters based on corresponding  $a^*$  and  $b^*$  values in  $L^*a^*b^*$  color space as shown in Fig. (6). This four clusters represents four regions i.e. RBC, WBC nucleus, cytoplasm and background stain. It was observed that WBC cytoplasm and RBC are classified into same cluster. In order to overcome the undesirable overlapping of regions, a second stage segmentation is performed using nearest neighbor classification. In the second stage we choose a

sample region randomly from each of the four clusters acquired using K-means. The mean color of the each sample regions are calculated in a\*b\* space and those values act as color indicators. here each pixel in the L\*a\*b\* space is distributing into any of the four classes by computing the Euclidean distance between that pixel and each color index. Each pixel of the whole image will be labeled to a specific color depending on the minimum distance from each index. The nucleus segmented RGB image is reconstructed from the labeled image. We have only considered the cluster which contains blue nucleus as it is required for feature extraction and hence leukemia detection. Few left out holes in the nucleus creates problem during texture extraction and hence they are filled using morphological reconstruction [8].

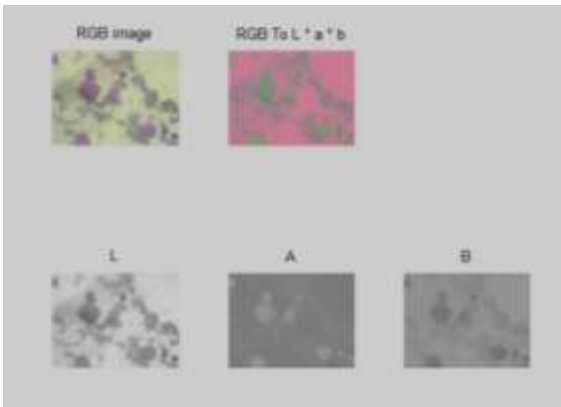


Fig. (6)

Algorithm Segmentation using K-means clustering

- Read Enhancement image (Y).
- [l a b] = Convert To L\*A\*B space
- A\_B = Merge A\*B space
- [r c] = get size (A\_B)
- A\_B\_new = reshape(A\_B, r\*c , 2)
- Apply K-means clustering and output as shown in Fig. (7).

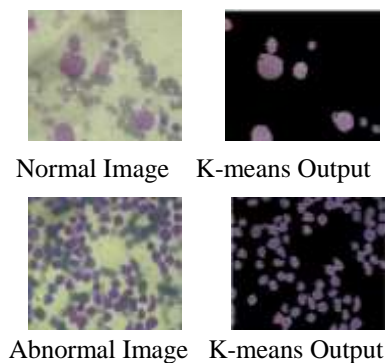


Fig. (7)

### C. Sub Imaging

Sub images including single nucleus per sub image are obtained using bounding box technique [9]. Using image morphology [10] only those sub images are selected which contains only lymphocytes. The nucleus sub images of neutrophils, eosinophil's, and basophils are not considered for feature extraction as they are not associated with lymphocytic leukemia.

Fig. (8). shows the major steps and examples of input/output images:

Step one) Input Image

Step tow) Sobel edge enhancing: - It's enhances the borders of the membranes [11].

Step three) Structured image dilation: - The morphological operator named dilation [12] has been employed to better connect to the separated points of the membrane border and make the perimeter of cell as a connected item (thicker more than one pixel).

Step four) Hole filling: -This step include of filling internal holes of the connected element with the largest area in the processed image [13, 14].

Fig. (9) shows separated Nucleus Sub Images using Bounding box technique.

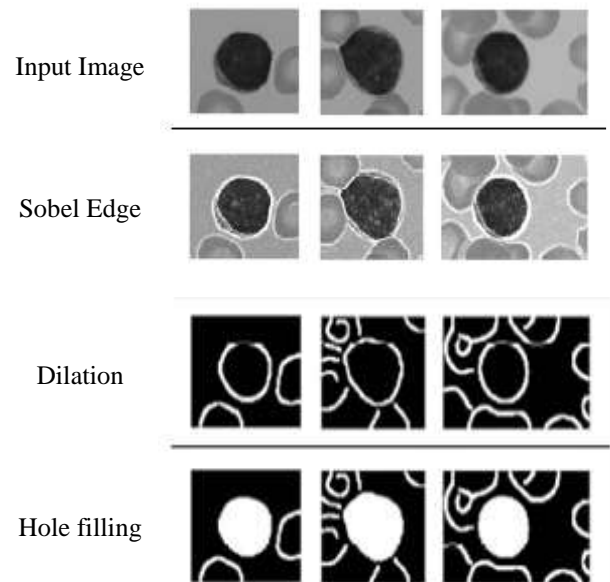


Fig. (8)



Fig. (9). Separated Nucleus Sub Images using Bounding box technique

**D. Feature Extraction:**

Feature extraction in image processing involves reducing the amount of resources required to describe a large set of data. In the present paper broadly four types of features are extracted (shape features, texture features, fractal dimension. In addition also color features are extracted from the nucleus image).

*A: Shape Feature:-* According to the hematologist the shape of the nucleus is an essential feature for distinguish of blasts. boundary based shape features and Region are extracted for shape analysis of the nucleus. All the features are extracted from the binary equivalent image of nucleus with nonzero pixels represents the nucleus region. For each nucleus we make a quantitative evaluation by using the extracted features under two classes. Region based and boundary based. These features are as follows:

- *Area:* It was determined by computation the total number of nonzero pixels within the image region.
- *Perimeter:* the perimeter was measured by computation distance between the successive boundary pixels.
- *Compactness:* Compactness or roundedness is the measure of a nucleus as defined in (1).

$$Compactness = \frac{Perimeter^2}{Area} \quad (1)$$

- *Solidity:* The solidity is the ratio of actual area and the convex hull area and is also an essential feature for classification a blast cell. This measure is defined in (2).

$$Solidity = \frac{Area}{ConvexArea} \quad (2)$$

- *Eccentricity:* This parameter is used to measure how much a shape of a nucleus deviates from being circular. It's an important feature since lymphocytes are more circular than the blast. To measure this a relation is defined in (3).

$$Eccentricity = \frac{\sqrt{a^2 - b^2}}{a} \quad (3)$$

where "a" is the major axis and "b" is the minor axis of the equivalent ellipse representing the nucleus region.

- *Elongation:* Abnormal bulging of the nucleus It's also an feature which indicates towards leukemia. Hence the nucleus bulging is measured in terms of a ratio called elongation. This is defined as the ratio between maximum distance ( $R_{max}$ ) and minimum distance ( $R_{min}$ ) from the center of gravity to the nucleus boundary and is given by (4).

$$Elongation = \frac{R_{max}}{R_{min}} \quad (4)$$

Where  $R_{max}$  and  $R_{min}$  are maximum and minimum radii respectively.

- *Formfactor:* This is a dimensionless parameter which changes with surface irregularities and is defined as (5).

$$Formfactor = \frac{4 * pi * Area}{Perimeter^2} \quad (5)$$

*B: Texture Feature:-*The nucleus texture measurements were performed on a gray scale version of the nucleus images. These features were computed from the co-occurrence matrices for each nucleus image. This includes:

- *Homogeneity:* It is a measure of degree of variation.
- *Energy:* The energy are used to measure uniformity.
- *Correlation:* This represents the correlation between the pixel values and its neighborhood.
- *Entropy:* It is usually used to measure the randomness.
- *Contrast:* The contrast is a measure of the intensity contrast between a pixel and its neighbor over the entire image.

*C: Color Feature:-* Since color is an important feature that human perceiv ewhile visualizing it is considered for extraction from nucleu sregions. Hence for each nucleus image the mean color values in RGB color spaces are obtained. [15].

*D: Fractal Dimension:-* Fractals have been used in medicine and science previously for several quantitative measurement [16] [17]. the important measure that decided whether a particular nucleus represents a lymphoblast or a mature lymphocyte is the Perimeter roughness of nucleus. the more convenient way to parameterize the cell boundary surface in comparison to Euclidean geometry is fractal geometry. Hausdorff dimension (HD) is a main feature for fractal geometry and will be a main quantitative measure for cell boundary roughness measurement. a procedure for Hausdorff Dimension (HD) measurement using box counting method [18]. The Hausdorff Dimension *HD* may then be obtained as in (6).

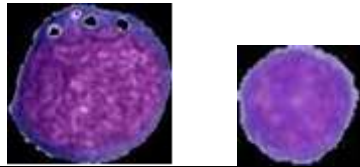
$$HD = \frac{Log(N)}{Log(N(s))} \quad (6)$$

where, N the number of squares in the superimposed grid, and N(s) the number of occupied squares or boxes (box count). Higher HD denote to higher degree of roughness.

Table 1 displays the difference in the values of the shape features for a pair of cancerous and noncancerous (normal)

cell. Values indicate steep difference among the two sub-images.

**Table 1** Shape Feature Values



| Features           | Cancerous 7 | Normal 42 |
|--------------------|-------------|-----------|
| Area               | 2089.5      | 1056      |
| Perimeter          | 407.79      | 280.98    |
| Compactness        | 80.26       | 74.76     |
| Convex Area        | 6175        | 2137      |
| Solidity           | 0.657       | 0.494     |
| Eccentricity       | 0.442       | 0.439     |
| Elongation         | 123.06      | 100       |
| Form Factor        | 64.35       | 47.20     |
| Contrast           | 0.668       | 1.362     |
| Correlation        | 0.756       | 0.634     |
| Homogeneity        | 0.839       | 0.810     |
| Energy             | 0.159       | 0.192     |
| Entropy            | 2.237       | 2.069     |
| Mean               | 221.26      | 35.21     |
| Standard Deviation | 750.99      | 128.99    |
| HausDroff          | 1.934       | 1.808     |

### III. EXPERIMENTAL RESULT

The proposed technique has been applied on 100 blood smear images obtained from “Oncology Center - Faculty of Medicine - Mansoura University Hospital - Egypt” is made up of 50 microscope image samples of blood infected , 50 microscope image of the blood samples is not infected.

#### A. Experiment

The experiment work of proposed system consist several steps, all image are preprocessed by MATLAB to contrast enhancement and resize be defined as in Table I. The system segments all dataset image (cancer and normal image) ,and we get separated nucleus sub Images using Bounding box technique to extract all features.

#### B. Result Analysis

The experimental result has been developed by taking the sub-images. The entire test images are gone through the Preprocessing – Segmentation and from the sub-images obtain Features Extraction.

**Table.2**

| Pre-processing<br>(contrast – resize) | Time               |                    |
|---------------------------------------|--------------------|--------------------|
|                                       | Cancer<br>45 image | Normal<br>45 image |
| 512*512                               | 14.7313            | 14.778             |
| 256*256                               | 13.412             | 13.7233            |
| 128*128                               | 12.8222            | 12.2991            |
| 600*400                               | 14.9824            | 145964             |
| 300*200                               | 13.2879            | 13.1211            |
| 200*180                               | 12.8286            | 12.8089            |

**Table.3**

| Segmentation<br>K-Means | Time               |                    |
|-------------------------|--------------------|--------------------|
|                         | Cancer<br>45 image | Normal<br>45 image |
| 512*512                 | 54.4611            | 52.575             |
| 256*256                 | 14.5826            | 14.9145            |
| 128*128                 | 8.1589             | 6.2723             |
| 600*400                 | 49.5682            | 48.9681            |
| 300*200                 | 13.5127            | 13.5717            |
| 200*180                 | 12.5690            | 12.3620            |

**Table.4**

| Sub – Image | Time               |                    |
|-------------|--------------------|--------------------|
|             | Cancer<br>45 image | Normal<br>45 image |
| 512*512     | 9.47817            | 10.1307            |
| 256*256     | 3.8499             | 3.67331            |
| 128*128     | 3.42657            | 2.94794            |
| 600*400     | 7.92125            | 7.48196            |
| 300*200     | 3.88147            | 3.99509            |
| 200*180     | 3.7890             | 3.6025             |

Table.5

| Features Extraction | Time                   |                        |
|---------------------|------------------------|------------------------|
|                     | <i>Cancer 45 image</i> | <i>Normal 45 image</i> |
| 512*512             | 16.898                 | 17.0685                |
| 256*256             | 5.21939                | 5.31412                |
| 128*128             | 2.42911                | 2.84163                |
| 600*400             | 29.3403                | 29.1306                |
| 300*200             | 8.20683                | 8.09558                |
| 200*180             | 3.6898                 | 3.45067                |

#### IV. CONCLUSION

This paper has present segmentation (K-means Clustering) technique and features extraction (Shape – Texture – Color – HD ) and BP – SCG neural network for classify. The system was evaluated in MATLAB 2014 and using data base 50 infected images and 50 non – infected images. The system is less computational requirement this make system well suited for low cost hardware implementation.

#### V. ACKNOWLEDGMENTS

My deepest appreciation goes to my supervisor Professor Dr. Mohy Eldin A. Abo-Elvoud and DR. Mohamed El-Said Morsy for his help and support in advising me to keep improving my knowledge and to keep believing in my abilities. A similar level of gratitude is due to “Oncology Center - Faculty of Medicine - Mansoura University Hospital - Egypt”, for supplying me the medical images we have needed it. I also would like to thank the Ministry of Science and Technology on the support provided by us to achieve optimal scientific degrees.

#### VI. REFERENCES

- [1] K. Breden, T. Schorr, J. B. Schorr, “Blood” Colliers Encyclopaedia,
- [2] Children’s Hospital of Wisconsin Website. <http://www.chw.org>
- [3] Edward D. Ball, G. A. L. (2002). 100 Questions & Answers about Leukemia Jones & Bartlett Publishers
- [4] Bain, B. (2003). Leukemia diagnosis (Third ed.): John Wiley & Sons.
- [5] Website <http://www.mathworks.com/help/images/filter-images-usingpredefined-filters.html>
- [6] Website <http://www.mathworks.com/help/images/ref/imfilter.html>
- [7] K. S. Ravichandran and B. Ananthi. Color Skin Segmentation using K– Means Cluster. International Journal of Computational and Applied Mathematics, 4(2):153 – 157, 2009.
- [8] S. Mohapatra and D. Patra, “Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images ,”International Conference on Systems in Medicine and Biology. P( 51).16-18 December 2010, IIT Kharagpur, India.
- [9] R. C. Gonzalez and R. E. Woods. Digital Image Processing. Addison Wesley, 2nd edition, 1992.
- [10] A. K. Jain, Fundamentals of Digital Image Processing. Pearson Education, 1st Indian edition, 2003.
- [11] J.S. Lim “Two dimensional signal and image processing” Prentice Hall 1990.
- [12] R. C. Gonzalez, R. E. Woods, S.L. Eddins, “Digital Image Processing Using MATLAB”, Pearson Prentice Hall Pearson Education, Inc., New Jersey, USA, 2004.
- [13] H. J. A. M. Heijmans (1994): Morphological Image Operators, Academic Press, New York.
- [14] Fabio Scotti, ” Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images”, IEEE International Conference on Computational Intelligence for Measurement Systems and Applications Giardini Naxos, Italy, P98. 20-22 July 2005.
- [15] The Wikipedia the Free Encylopedia Website. <http://en.wikipedia.org>
- [16] B. B. Mandelbrot. How long is the coast of Britain? Statistical self similarity and fractional dimension. Science, 156:636 – 638, 1967.
- [17] B. T. Milne. Measuring the fractal geometry of landscapes. Applied Mathematics and Computation, 27:67 – 79, 1988.
- [18] A. P. Pentland. Fractal based description of natural scene. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:661 –674,1984.