



## A SYSTEM FOR ACUTE LEUKEMIA CELLS SEGMENTATION AND CLASSIFICATION

R. Mohammed

O. Nomir

I. Khalifa

T. Hamza

Computer Science Department,  
faculty of computer and informatics  
Mansoura University , Egypt

Computer Science Department,  
faculty of computer and  
informatics  
Helwan University , Egypt

Computer Science Department,  
faculty of computer and  
informatics  
Mansoura University , Egypt

reham\_elshamy2006@yahoo.com

o.nomir@umiami.edu

dr\_iraky@hotmail.com

Taher\_hamza@yahoo.com

**Abstract**— *This research paper presents a system for the acute leukemia blast cells segmentation and classification. The research objective is to generate the features characterizing normal and infected cells. The proposed system consists of one segmentation method and one classification method of acute leukemia. The features extracted from the cell and adopted features are used as the input signals to the Multi Layer Perception (MLP) neural network classifier. The experimental results show that our proposed system is robust and effective in identifying acute leukemia blast cells.*

**Keywords**— *Image Segmentation, ALL , RGB, C-Y color model, Features Extraction, MLP.*

### 1. Introduction

Leukemia is a blood cancer that affects the blood cells and the bone marrow. The four main types of leukemia include ((ALL),(CLL), (AML), (CML)) [1]. This type of cancer can be discriminated according to the growth speed overproduction of leukemic cells [2]. This research work focuses on the segmentation of Acute Lymphoblastic Leukemia (ALL), which is called also childhood leukemia. There are many techniques that have been adopted in the segmentation of white blood cells areas. In [3], they applied Watershed distance transform process to the binary image obtained by Otsu's thresholding on the red channel of the acquired images, which resulted in providing suitable segmentation of the white blood cells. Another technique is presented in [4] to segment white blood cells, that proposed a scheme on color space images utilizing feature space clustering techniques. Other method utilized morphological operators and investigating the scale-space characteristics of toggle operator to enhance segmentation accuracy is proposed in [5]. In [6], they proposed a segmentation technique that depends on the automatic morphological analysis of white blood cells. Another segmentation technique [7] utilized teager energy operator for segmentation, nucleus based on the edges, that were detected effectively by teager energy operator. Many approaches provided different techniques that are utilizing transformation of RGB color space into another one like HSV [8]. In [9] they utilized an image processing method for color segmentation of ALL using HSI (Hue, Saturation, Intensity), and RGB color spaces. The evaluation of color spaces is carried out using a segmentation algorithm, such as HSL, YIQ, l\*a\*b [10], etc.

In [11], they utilized the ANN to discriminate and arrange morphological granulometric features of nucleus in automatic bone marrow white blood cell. In this research, according to their introduced technique, four features that are extracted from each nucleus were tested by using Bayes classifier and artificial neural network. The results demonstrated that the features using nucleus alone can be used to achieve a classification rate of 77% on their test sets. This research paper proposes a system for acute leukemia cells segmentation and classification. The goal of the segmentation step is to segment the color image of acute leukemia by transforming the RGB color space to C-Y color space. In the C-Y color space, the luminance component is used to segment (ALL) by extracting the Y component of the converted image and choose a threshold value from the histogram of the image of Y component in the C-Y color space. The proposed algorithm is evaluated using 100 microscopic ALL images. The experimental results prove that the proposed technique can achieve a good segmentation of ALL from its complicated background and it also proves that the segmentation accuracy of the proposed technique is 98.38% compared to the result of the manual segmentation technique provided by expert. After the segmentation of the cells is done, a Matlab tool is applied to extract 24 unique features from each segmented cell. In order to classify the cells, the extracted features are input to a Multi Layer Perceptron (MLP) neural network classifier utilizing WEKA (Waikato Environment for Knowledge Analysis). The accuracy rate of our classification technique is 93.585%.

## 2. Methodology

The aim of this research work is the process of segmentation and classification of the microscopic (ALL) images. This process includes two phases, the first phase is the segmentation process by using C-Y color space and applying the median filter on the resulted images. The second phase is the classification process including extracting 24 unique features from each segmented cell. The classification step is carried out by presenting the extracted features to MLP neural network classifier to discriminate the infected cells.

The implementation of the introduced technique includes the following steps:

Step 1 : Converting RGB images into their equivalent C-Y color space.

Step 2 : Finding the Y component of the C-Y.

Step 3 : Selecting from the histogram of Y component the threshold value.

Step 4 :Preserving sharp edges by applying (5 × 5) median filter.

Step 5 : Transforming the resulted segmented image into RGB.

Step 6 : Extract 24 features to uniquely identify each segmented cell.

Step 7 : Apply MLP classifier on the extracted features by using WEKA.

The proposed system for both ALL image segmentation, and classification processes is depicted below in Figure 1.

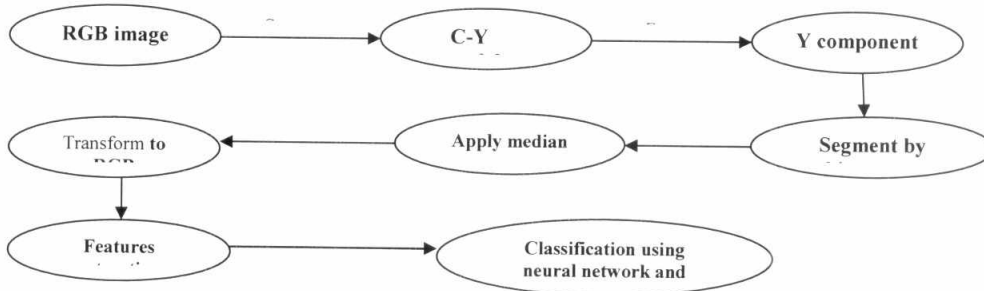


Figure 1: The introduced technique of segmentation and classification of ALL steps.

The computational results are evaluated under Processor Core Intel Dual-core (2 Core), Cache 3 MB, under Operating System Windows 7.

### 3. The Segmentation Algorithm

The goal of the segmentation of ALL introduced technique is to extract the blast cell from its complicated background by using C-Y color space .

The C-Y is chosen here in this research because it is very important and attractive color model for image processing applications. C-Y color model can exemplify color similarly as how the human eye identifies color. C-Y has three basic components, only two of them are to define a color . And another one is for the luminance component Y [12].

The following matrix is for transforming the RGB images to C-Y :

$$\begin{bmatrix} Y \\ R - Y \\ B - Y \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.701 & -0.587 & -0.114 \\ -0.299 & -0.587 & 0.886 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Equation 2 , 3 for the saturation and hue value .

$$s = \sqrt{(R - Y)^2 + (B - Y)^2} \quad (2)$$

where s is the saturation , Y is the luminance component ,  
R for red and B for blue components of C-Y color model .

$$\theta = \begin{cases} \tan^{-1}\left(\frac{R-Y}{B-Y}\right) & \text{for } s \neq 0 \\ \text{undefined} & \text{for } s = 0 \end{cases} \quad (3)$$

where  $\theta$  is the hue value

We maintain the sharp edges by using the median filter. We tried different values of N, the dimension of the filter , N is chosen equal to 5 which gives better segmentation results.

Equation 4 for transforming the resulted segmented image into RGB.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.1 & 0.1 & 0.0 \\ 0.1 & -0.509 & -0.194 \\ 0.1 & 0.0 & 0.1 \end{bmatrix} \begin{bmatrix} Y \\ R - Y \\ B - Y \end{bmatrix} \quad (4)$$

## 4. Features Extraction and Cell Classification

### A. Feature Extraction

Visual information plays an essential role in clinical diagnosis. It is widely known that there is no common distinct group of features that can be suitable for all kinds of computer vision applications. Image features extraction stage is an important step that utilizes algorithms and techniques in order to detect and isolate several desired portions or shapes (features) of the image.

For the purposes of ALL classification, other techniques used different features (geometric, texture , color ,statistical) . They used shape and color features from the extracted cells like, mean and standard deviation of RGB and l\*a\*b color spaces , and shape features (area , perimeter, eccentricity , orientation, compactness, Nucleus: Cytoplasm (N:C) ratio. NC is a ratio of the size of the cell's nucleus to the size of the cell's cytoplasm.) .

In our research study, and for the purposes of ALL classification, 24 unique features in two types of features, shape and texture features are extracted. The shape features is a set of numbers that are defined to describe a given shape property (Area ,Convex Area, Major Axis Length, Minor Axis Length, Bounding Box (4 values), Eccentricity, Perimeter, Compactness for 8-connected neighborhood). Two types of texture features were used that is the First-order and second-order texture (Mean, Dispersion, Variance, Average Energy, Skewness, Kurtosis, Median, Mode, Energy, Entropy, Correlation, Inertia, Homogeneity). Figure 2 shows sample of the images which are used in our research.

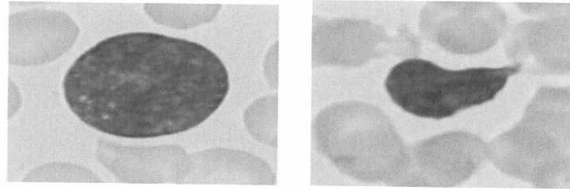


Figure 2: example of database

### B. Cell Classification

After extracted the cell features, a neural network classifier (MLP)[13] is applied on the extracted features by using WEKA environment. We use MLP with input layer which contains 24 nodes , one hidden layer with 2 nodes and two class as an output layer as shown in Figure 3

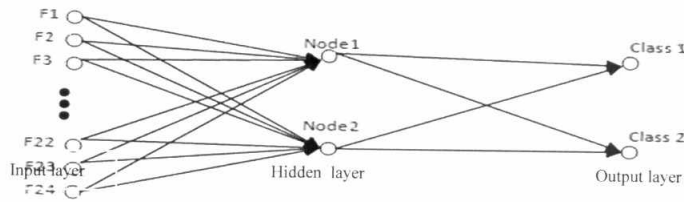


Figure 3:The MLP Network topology .

## 5. The Experimental Results

### A. The Dataset

The database of ALL-IDB dataset [14] is used to evaluate the proposed system. We use 100 images of ALL-IDB2 version. Figure 4 shows an example of the images in the ALL dataset.

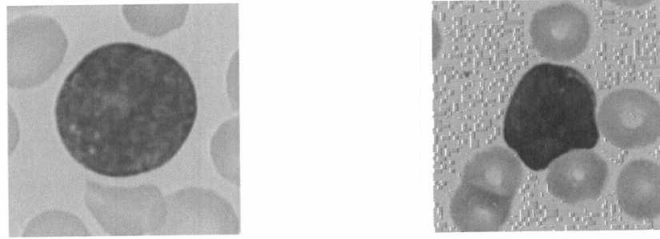


Figure 4: Example of ALL Dataset

### B. The Segmentation results

The first phase of this research work is a segmentation process of ALL by using C-Y color space , using 100 ALL microscopic images to run this technique [14] . This procedure works on the luminance of C-Y model and Compares the accuracy of the result with RGB and the expert results by area pixels on them in a quantitative manner as introduced in Figure 5.



Figure 5: Example of the segmented RGB images

The steps of obtaining the results after applying the proposed algorithm are shown below in the following Figures.

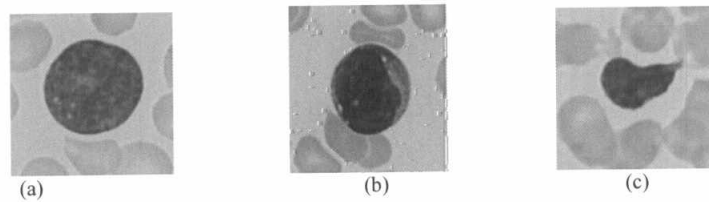


Figure 6: Original RGB images (the input )

Figure 6 represents the RGB color images that are the input to our algorithm. Then apply equation (1), to convert the RGB images to the C-Y color images as shown in Figure 7.

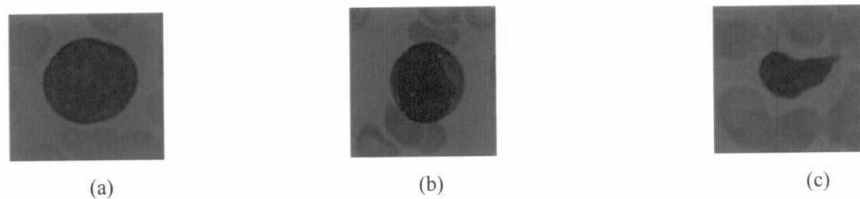


Figure 7: The transformed C-Y Images for the three images in Figure 5.

Then we extract the B-Y components of the C-Y color model as represented in Figure 8.

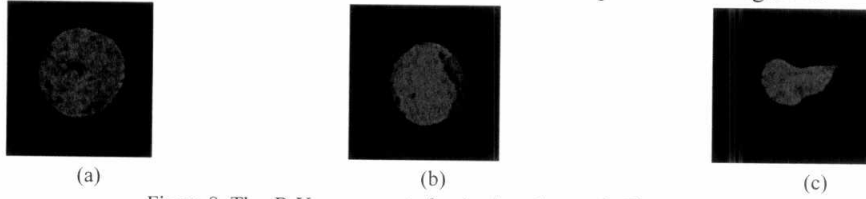


Figure 8: The B-Y components for the three images in Figure 6.

After that the C-Y images are extracted from the R-Y component as shown in Figure 9.

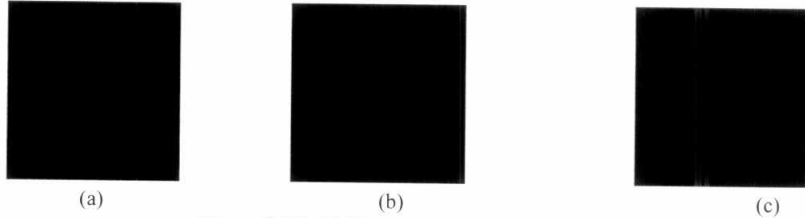


Figure 9: The R-Y components for the three images in Figure 6.

The Y components are extracted from the C-Y color images as shown in Figure 10.

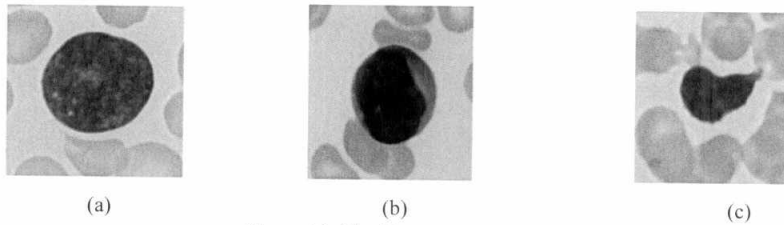


Figure 10: The Y components

the equation (2) was applied to extract S component from C-Y color images as shown in Figure 11.

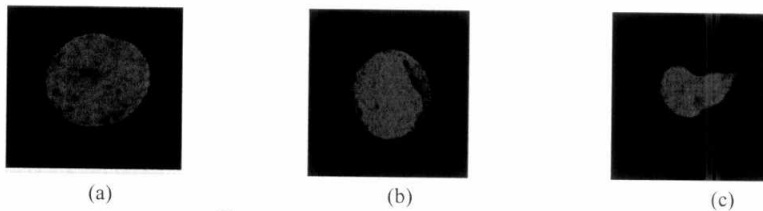


Figure 11: The S components of C-Y

From the (B-Y , R-Y ,Y) components of C- Y model, we choose the Y component in our approach to obtain the segmentation process on it . We choose the threshold value using the histogram of the Y original image.

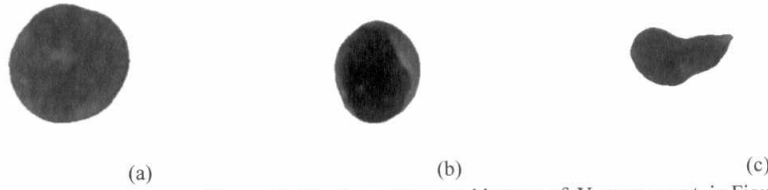


Figure 12: The three segmented images of Y component in Figure 10.

Figure 12 shows the final results after applying median filter  $5 \times 5$  in order to preserve sharp edges. Finally, equation (4) is applied on the segmented images in Figure 12 to convert the output result to RGB color space as shown in Figure 13.

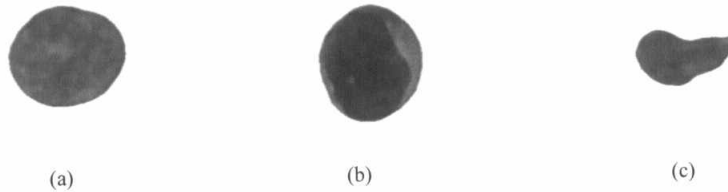


Figure 13: The final converted RGB images

As we compared our segmentation method using (C-Y) color space by other color spaces such RGB and CIELAB. CIELAB is an opponent color system depended on the earlier (1942) technique of Richard Hunter called l, a, b. In [15], they utilized  $l^*a^*b^*$  (CIELAB) color accurate blood cell segmentation of ALL cells. Table 1, presents a comparison of the accuracy of the introduced segmentation technique of ALL to the accuracy of the segmentation by RGB color model. we found that the method based on RGB has not performed well as presented in Figure 14. It does not give accurate result. Besides that, the shape of the blast after the segmentation process is not quite similar to the expert blasts. While the method depended on C-Y color space utilizing Y component can produce a good accuracy segmentation by comparing it with the expert blasts. this comparison of the accuracy presented in Table 1.

Table1: Segmentation's accuracy

color space	Accuracy
C - Y	98.38%
R G B	95.93 %

We find that there are some images were hard to segment using our proposed segmentation method, because of their damages in capture process, and sometimes because of the image poor quality. The following, Figure 14 is an example of failed segmented images.

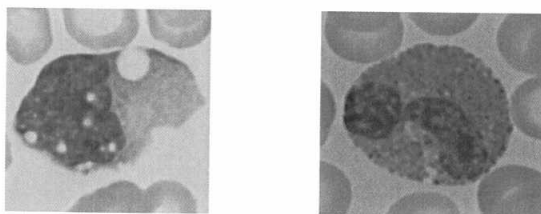


Figure 14: Some images that the system failed to segment

### C. The Features Extraction and Cell Classification

As a classifier we have applied neural network (MLP) by Weka, trains a multilayer perceptron with one hidden layer using WEKA's Optimization class.

For the purposes of ALL classification, we tried different sets of features to uniquely identify the segmented cells. We conclude that the previously mentioned 24 can uniquely classify the ALL cells. These 24 features are extracted from the previously segmented cells. To classify the segmented cells, a neural network classifier (MLP) is applied using

the extracted features. It trains a multilayer perceptron using WEKA's Optimization class as 24 input nodes (features) and 2 output nodes (for the two classes) with one hidden layer with 2 nodes. Both SVM, and MLP are easy to model and have very simple training procedures. SVMs are a common technique for two-category classification that is simple to use. It is based on the structural risk minimization principle and focuses on finding [16] the so-called support vectors, a set of critical points that can be used to clarify the decision function of the classifier. They performed classification by transforming the data into a feature space using kernel functions. The drawback of SVM is that they depend on a limited amount of kernels. We compared the accuracy when using both classifiers and we find that that MLP outperforms SVM [16] as shown in Table 2.

Table 2: The result of applying both MLP, SVM classifier.

Classifier	Correctly Classified Instances	Accuracy
MLP	86	93.58 %
SVM	73	79.35%

### 6. Conclusion

In this research study, we have successfully produced a technique for the detection and classification of ALL utilizing image processing algorithms and neural network. Image quality and accuracy is the core factors of this research. One segmentation method is proposed which uses C-Y color model in separating the ALL cells from the background. Also a classification method relying on MLP classifier is proposed. The classifier uses 24 unique features to uniquely identify the segmented cells. Our introduced technique achieves a good result comparing with the others techniques.



## References

1. A. Salihah, A. Mashor, and N. Harun, "Colour Image Enhancement Techniques for Acute Leukemia Blood Cell Morphological Features", IEEE on Biomedical Engineering & Sciences, vol 72, pp. 3677-3682, 2010.
2. The Cancer Institute of New Jersey Patient Education Committee, Leukemia and Lymphoma awareness month, <http://www.cancer.gov>, New Jersey, USA, 2008.
3. L. Vincent, and P. Soille, "Watersheds in Digital Spaces; An Efficient Algorithm Based on Immersion Simulations", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp. 583-597, 1991.
4. K. Jiang, Q. Liao, and S. Dai, "A Novel White Blood Cell Segmentation Scheme Using Scale-Space Filtering and Watershed Clustering", International Conference on Machine Learning and Cybernetics Machine, vol.5, pp 2820-2825, China, 2003.
5. L. Dorini, R. Minetto, and N. Leite, "White Blood Cell Segmentation Using Morphological Operators and Scale-Space Analysis", Proceedings of Brazilian Symposium conference on Computer Graphics and Image Processing, pp. 294-304, Sibgrapi, Brazil, October 2007.
6. F. Scotti, "Automatic Morphological Analysis for Acute Lymphoblastic Leukemia Identification in Peripheral Blood Microscope Images", IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, CIMSA, pp. 96-101, Sicily Italy, July 2005.
7. B. Kumar, D. Joseph, and T. Sreenivas, "Teager Energy Based Blood Cell Segmentation," in Proceedings of the 14th International Conference on Digital Signal Processing, pp. 619-622 Bangalore, India, 2002.
8. N. Sinha, and A. Ramakrishnan, "Automation of Differential Blood Count", In Proceeding Conference on Convergent Technologies for Asia-Pacific Region, vol.2, pp: 547-551, India, 2003.
9. W. Skarbek, and A. Koschan, "Color Image Segmentation", <http://citeseer.nj.nec.com/skarbek94colour.html>, Institute for Technical Informatics, Technical University of Berlin, Berlin, 1994.
10. S. Mohapatra, and D. Patra, "Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images", International Conference On Systems In Medecine and Biology, pp, 49-54, ISBN:978-1-61284-039-0, India, 2010.
11. N. Theera-Umpon, and S. Dhompongsa, "Morphological Granulometric Features of Nucleus in Automatic Bone Marrow White Blood Cell classification", IEEE Trans on Biomedical Information Technology, vol.11, ISSN : 1089-7771, pp. 353-9, 2007.
12. E. Dougherty, "Electronic Imaging Technology", pp. 2583- 2586, 1999.
13. S. Buavirat, and C. Risa-an, "Classification For Acute Lymphocytic Leukemia Using Feature Extraction and Neural Networks in White Blood Cell Stained Images", ISBME, ISCCSP 3rd International Symposium on A new method for blood cell image segmentation and counting based on PCNN and autowave, pp. 6-9, St Julians, Malta, 2008.
14. R. Labati, V. Piuri, and F. Scotti, "ALL-IDB: the Acute Lymphoblastic Leukemia Image DataBase for image processing", <http://homes.di.unimi.it/scotti/all> Italy.
15. N. Mahmood, and P. Lim, "Blood Cells Extraction Using Color Based Segmentation Technique", [www.ijlbpr.com](http://www.ijlbpr.com), vol. 2, pp. 233-240, ISSN 2250-3137, April 2013.
16. Shutao Li, Chen Liao, "Feature Extraction for Cancer Classification Using Kernel-Based Methods", Springer-Verlag Berlin Heidelberg, pp. 162-171, LNBI 4689, 2007.