Diagnosis of Leukemia Using Digital Image Segmentation of Blood Cells

Pandipriya.J[1], Mathan.N[2]

M.Tech, Nanotechnology, Sathyabama University, Chennai, India[1]

Assistant Professor, Department of ECE, Sathyabama University, Chennai, India[2]

Abstract: The early identification of leukemia form in cancer patients can greatly increase the likelihood of recovery. Diagnostic methods that distinguish among the disease's many forms are either costly or do not exist. Amongst the existing diagnostic methods are immune- phenotype and cytogenetic abnormality, which require time to obtain results and are costly to perform due to their requirement of well equipped laboratories. Thus, there is a need for fast and cost-effective method that results in the identification of the different leukemia forms or types. Therefore, we propose the use of digital image morphological analysis of microscopic images of leukemic blood cells for the identification purpose. We present in this paper the first phase of an automated leukemia form identification system, which is the segmentation of infected cell images. The segmentation process provides two enhanced images for each blood cell; containing the cytoplasm and the nuclei regions. Unique features for each form of leukemia can then be extracted from the two images and used for identification.

Keywords: Cancer Diagnosis, Leukemia Type, Blood Cell, Image Segmentation, Morphological Analysis.

I.   INTRODUCTION

Leukemia is a cancer of white blood cells, where the disease basically develops in the bone marrow which is the spongy tissue that fills the inside region of the bones. There are four major different forms or types of leukemia, which develop in cancer patients according to the growth speed and the improper overproduction of leukemic cells [1]. These forms have unique features which can be visually observed by a trained expert using microscopic images of the infected cells. However, the identification task is usually difficult due to the variety of features and the often unclear images which cause missing out on vital indicators to which form of leukemia is being observed.

There are four major forms of leukemia; *Acute lymphoblastic leukemia* (ALL), *Acute myelogenous leukemia* (AML), *Chronic lymphocytic leukemia* (CLL), and *Chronic myelogenous leukemia* (CML) [2],[3],[4]. ALL is the most common type of leukemia in young children. This disease also affects adults, especially those aged 65 and older. The survival rates vary by age: 85% in children and 50% in adults [5]. AML occurs more commonly in adults than in children, and more commonly in men than women.

The five-year survival rate is 40% [6]. CLL most often affects adults over the age of 55. It sometimes occurs in younger adults, but it almost never affects children. Two-thirds of affected people are men. The five-year survival rate is 75% [7]. CML occurs mainly in adults. A very small number of children also develop this disease. The five-year survival rate is 90% [8],[9].

The early and fast identification of the leukemia type, greatly aids in providing the appropriate treatment for that particular type. The currently used diagnostic methods rely on analyzing immune-phenotype, cytogenetic abnormality [10],[11],[12]. These diagnostic methods require sophisticated expensive laboratories, and have been reported to provide a high ratio of misidentification; as reported in [13]: "The best labs rely on as many as a dozen different, labor-intensive technologies, all of which require highly trained specialists. Even so, perhaps 50% of patients are misdiagnosed in regard to subtype".

Morphological analysis methods, on the other hand, rely mainly on analyzing images of leukemic cells, and can be also used in correlation with cytogenetics [14]. Thus, it can be suggested that using morphological analysis methods for identifying the different leukemia types; based mainly on images, can greatly reduce the cost of performing type identification tests. Figure 1 shows examples of leukemic cell images that are used in morphological analysis.

The first task performing morphological analysis is the segmentation of the infected blood cell. Previous works on cell image segmentation include the work in [15] where WBCs in bone marrow microscopic images were segmented using a method based on pixel classification and spectral features to separate the different regions in the infected cell. In another recent work [16] that used morphological features of blood cells, focus was made on identifying the different blood cell types whether infected or leukemic. This work used a neural network to arbitrate the features of a blood cell, in order to classify the blood cell type.

Lately, the work in [17] suggested the use of image segmentation, based on K-means algorithm, to obtain colon cell morphological features which are used then to recognize the different colon cell types.

These examples on successfully using image segmentation in the morphological analysis of microscopic cell images

are very encouraging and certainly have shown reduction in cost and efficiency in identification. However, there is still a need for an automated method that identifies the four major types of leukemia in infected blood cells. The sought after method should address the main problems with existing systems for leukemia type identification; these problems can be summarized as [13]: high financial cost ($2,400-$4,800), lengthy analysis processing time (48 hours -one week), and high ratio of misdiagnose (50% of patients may be misdiagnosed in regards to leukemia sub-type).

Therefore, we propose a novel method for automatic identification of the major leukemia types. The developed system based on our method, would classify an input image of a leukemic cell into: ALL, AML, CLL, or CML. The novel method is based on morphological analysis of microscopic images of single leukemic cells. There are three phases in our proposed method, namely, segmentation, feature extraction, and classification.

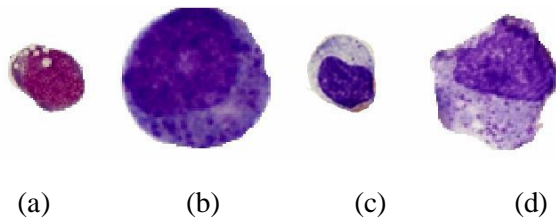(a)   (b)   (c)   (d)

Fig. 1 Examples of leukemic blood cells

Above figure shows the four different types of leukemia: (a) ALL, (b) AML, (c) CLL, and (d) CML [18]. In this paper, we present the segmentation phase, which provides two enhanced images for each input cell image. The enhanced result images contain valuable information on the cell features and shall then be used as the input images in the next two stages of the identification system.

The paper is organized as follows: Section 2 provides an overview of the proposed segmentation process. Section 3 describes in detail the segmentation process. Section 4 presents the segmentation process application results. Finally, Section 5 concludes this work.

II.   SEGMENTATION PROCESS: OVERVIEW

The unique morphological features of a leukemic blood cell are located either in the cytoplasm or the nuclei membranes or regions of the cell. Therefore, efficient extraction of the morphological features requires proper separation of both regions. The cytoplasm and nuclei regions are approximately uniform in terms of gray level intensity except for a few subcategories of the ALL leukemia infected cells; where the chromatin aggregate into masses [19]. Thus, efficient separation of nuclei from cytoplasm could be achieved using bimodal threshold segmentation which is the first stage of our proposed segmentation process.

The result of the thresholding stage is two images containing the cytoplasm and the nuclei regions. The boundaries of both regions are then traced out using the principle of chain code [20]. Chain code may result in detection of several membranes other than the primary

membranes of cytoplasm and nuclei, therefore, the enhancement of the thresholded image should eliminate the unwanted membranes; i.e. other than cytoplasm and nuclei. Furthermore, region filling is applied to the thresholded and enhanced images of the two regions in order to evaluate some region properties such as the area. Finally, image restoration is applied to the regions of interest (cytoplasm and nuclei), where the regions are assigned zero-valued pixels; i.e. shown in black. This output which contains two images can then be used for further processing, such as feature extraction, as part of the automatic leukemia form identification system. Figure 2 shows a flowchart of the segmentation process in general.

III.   THE SEGMENTATION METHOD

There are different approaches to segment an image including pixel-based segmentation, which is trivial because the spatial contents of the image are not considered, and decision is made solely using pixel gray levels.
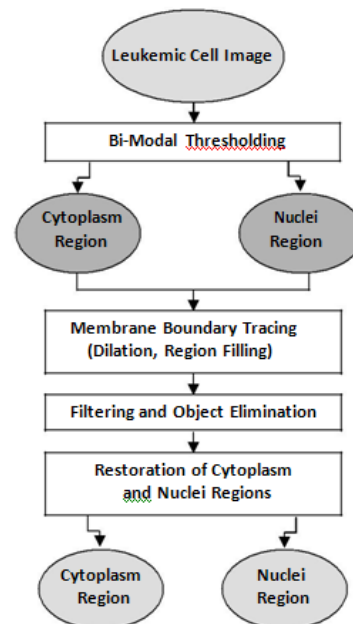

Fig 2. An overview of the segmentation process.

The other available technique is region-based segmentation which looks into the probability distribution of the object and the background, which could be followed by statistical analysis to figure out the optimum threshold [20]. In this work we follow the pixel-based segmentation approach. The following subsections describe in detail the stages of the segmentation method.
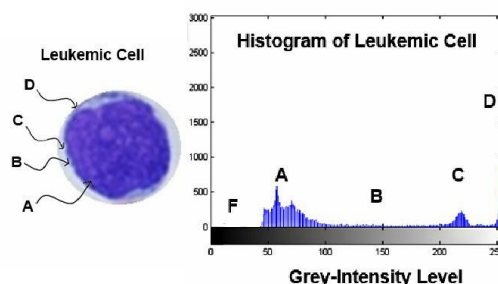

Fig 3. Histogram distribution of a leukemic cell.

A.   Bi-Modal Thresholding:

The gray level intensities of both nuclei and cytoplasm are often uniform, and thus bimodal thresholding can be used to separate the two regions. The histogram distribution of a typical leukemic cell is shown in Figure 3. The peaks at regions *A* and *C* represent the nuclei and cytoplasm areas respectively; the difference in peak values between regions *A* and *C* indicates that the nuclei area is larger in size than the cytoplasm area.

The optimum threshold to segment nuclei from cytoplasm can be defined as the threshold value which separates region *A* from *C*. The selection of the threshold value is application dependent, and there are various methods to select the value [21]. However, for selecting the optimum threshold, we suggest in this work the use of the average number of pixel distribution over the 65% to 75% range, which is defined as follows:

$$Avg_D = \frac{\sum_{I=1}^{165} P(I)}{\sum_{I=1}^{165} I}, \ \forall P(I) \neq 0 \qquad (1)$$

$$TH_N = \frac{I\big|_{P(I)=Avg_D}}{255} \qquad (2)$$

where Avg_D is average number of pixels distribution over the intensity range, discarding the zero-pixel intensity levels, *P* is the total number of pixels at a certain intensity level *I*, and TH_N is the first threshold level which is optimum to separate nuclei area from cytoplasm area.

The second threshold is optimized to separate cytoplasm from its white background. Technically, it can be acceptable selecting an intensity level just below the 255 intensity level, but since the segmented cells may have additional unwanted objects which merge with the cytoplasm membrane as a result of the manual extraction of the blast cells out of the original multi cell images, then incorrect detection of the membrane may occur.

To solve this problem, we examine the intensity levels from 75% up to 98%, after discarding the lateral 255-valued pixels outside the area of interest (cytoplasm); using a second threshold (TH_C) which is defined as:

$$TH_C = \frac{I\big|_{P(I)=Max(P(I))}}{255}, \forall \ I \in (200,250) \qquad (3)$$

B.   Boundary Tracing of Membranes:

At this stage, the separated images of cytoplasm and nuclei contain their membranes which may require smoothing. The method used to trace and outline the boundaries of the membranes is 8-neighberhood chain code or contour code [21]. To ensure successful tracing of only the outline boundaries of the desired membranes; two

additional morphological operators are also used; namely Dilation and Region Filling. Dilation is a fundamental morphological operator that adds pixels to the boundary of any image and causes it to enlarge; this operator has the ability to bridge any gaps or broken elements at the boundary which is essential before proceeding into regions filling [21]. The process of dilation is defined as:

$$A \oplus B = \{z \mid [(B)_Z \cap A] \subseteq A\} \qquad (4)$$

where *A* is the original image to be dilated, *B* is denoted by the structuring element which has to be symmetrical about its origin, then *B* will be displaced by a factor of *z* all the way over image *A*, and the output image of dilation will be the resultant of overlapping or intersecting pixels, which logically could be obtained by applying an AND operation.

After interpolating and bridging the gaps at the membranes boundary, the inner region is then filled using a combination of the morphological operators [21]. The process of region filling is defined as:

$$X_K = (X_{K-1} \oplus B) \cap A^C, \ K = 1,2,3\ldots \qquad (5)$$

Where X_K is the starting point inside the boundary, *B* is structuring element and A^C is the complement of *A*. This process is repeated until X_K is equal to X_{K-1}. Finally, the result is augmented with the original boundary.

C.   Filtering and Object Elimination:

After dilating, filling, and boundary -tracing the regions of both membranes (cytoplasm and nuclei), filtering may be required to eliminate unwanted particles or objects which appear outside the membranes.

To achieve this, the size of the connected elements at each object in the image is monitored, and then the largest connected elements will be selected. The resultant data structured matrix of the boundary tracing has a size of M_{x1}, where *M* refers to the number of objects contained inside the data structured matrix. Every index within the data structured matrix is a sub-matrix with a size of N_{x2}, where each sub-matrix stores the positions of all elements belonging to a certain object in the image.
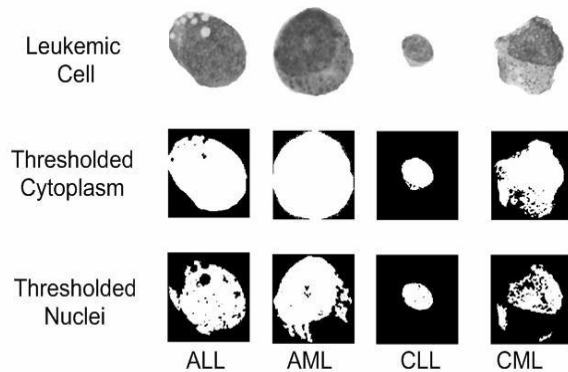

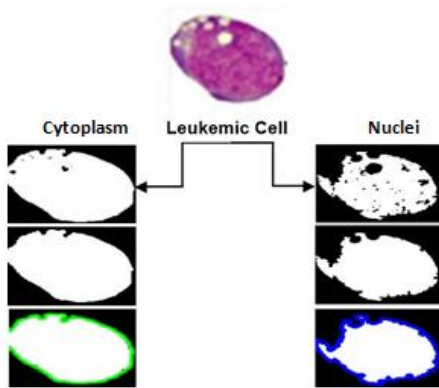Fig 4. Examples of bimodal-thresholding.

Fig 5. Cytoplasm and Nuclei membrane processing.

Hence, the desired membrane represents the biggest connected object or the one that holds for maximum *N* in the whole set.

Obtaining a single object that only represents the membrane of nuclei or cytoplasm requires a construction of a new zero- elements matrix and holding a size similar to the original image, then only elements belong to the largest object will be set to a binary value of one in the new constructed matrix without affecting the lateral zeros in the background. Figure 5 shows different stages of membranes processing, and a plot of the biggest selected object over the original image. As an advantage of boundary tracing and providing us with positions of all pixels over the boundary, then region properties such as area, perimeter, and center of the object could now be easily calculated.

D. Region Restoration:

At this stage of the segmentation process the two enhanced regions on interest (cytoplasm and nuclei) are restored in preparation for post-processing at the next phase of the automatic leukemia type identification system. The original leukemic cell image is being defined as $L(x,y)$, the nuclei filled up region as $N(x,y)$, and the cytoplasm filled up region as $C(x,y)$, then the following equations can be applied to restore back the nuclei and cytoplasm regions respectively, and setting the remaining regions to zero values.

$$L_{N(x,y)} = L(x, y).N(x, y) \qquad (6)$$

$$L_{C(x,y)} = L(x, y).N^C(x, y).C(x, y) \qquad (7)$$

L_N and L_C represent the restored regions of both nuclei and cytoplasm respectively. The nuclei region recovery could be obtained by convoluting both of the matrices L and N in their spatial domain, or simply by fining the logical intersection using an AND operator. On the other hand, restoring cytoplasm region requires firstly eliminating the nuclei region, which can be achieved by complementing the nuclei region to form N_C, and then multiply it with the whole set to extract the nuclei out. N_C is obtained by XOR

all of $N(x,y)$ elements with a matrix of ones, constructed with a size similar to *N*. Figure 6 shows both of cytoplasm and nuclei regions after the restoration process.

TABLE 1
PERFORMANCE OF THE SEGMENTATION METHOD

Leukemia Type	ALL	AML	CLL	CML
Successfully Segmented Cells	(30/30) 100%	(30/30) 100%	(30/30) 100%	(28/30) 93.33%

TABLE 2
SEGMENTATION PROCESSING TIME

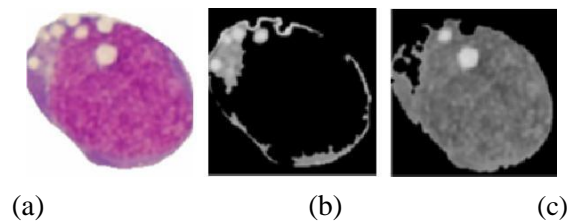| Segmentation Process Stage	Processing Time (seconds)
Image Reading	0.0310
Bi-Modal Thresholding	0.0624
Dilation & Region Filling	0.0624
Boundary Tracing	0.0780
Objects Elimination	0.2652
Regions Restoration	0.0780
Total Segmentation Process	0.5770

(a)   (b)   (c)

Fig 6. Result of region restoration. (a) Original leukemic cell image and restored: (b) cytoplasm region (c) nuclei region.

IV. EXPERIMENTAL RESULTS

The results of implementing the proposed leukemic cell image segmentation method were obtained using 2.8 GHz PC with 2 GB of RAM, Windows XP OS and Matlab 2008 software tools. The proposed segmentation system has been tested using images of leukemic single cells extracted from multi cell images provided online by the Health Center at University of Virginia [18]. The dataset used in this work consists of 120 gray-scale images of size 200x200 pixels, and comprising 30 images of each of the four types of leukemia.

Empirical tests using the proposed segmentation system on the 120 leukemic cell images have shown successful and efficient performance in the segmentation of the nuclei and cytoplasm regions in the four types of leukemia. The yielded segmented images underwent visual inspection, looking for degradation in the boundaries of the membranes, existence of unwanted particles or objects, and noisy restored cytoplasm and nuclei areas. The outcome of this evaluation is shown in Table 1 where an overall sufficient segmentation ratio of 98.33% was obtained. The computational and time costs

of our proposed segmentation method have been minimal which suggests that this method can be used in real time application. Table 2 lists the processing times for each stage in the segmentation process, in addition to the overall time which was a fast 0.577 seconds.

V.   CONCLUSION

This paper presented a novel method for segmenting single-blood-cell images of leukemia patients. The work in this paper aims at providing an automated system for identifying the four major different types of leukemia, by using microscopic images of infected blood cells. The successful segmentation and separation of an infected cell into its two sub-regions containing cytoplasm and nuclei regions is vital, since these two regions contain features that correlate to the different forms of leukemia. The output of our proposed segmentation system can then be used for feature extraction and consequently leukemia form or type identification.

The segmentation method in our work comprises the following stages: bimodal thresholding, dilation, region filling and boundary tracing, filtering and elimination of unwanted objects, and finally restoration of the two separated cell regions. Our experimental results yielded a successful overall ratio of correct segmentation by 98.33%. The yielded segmentation regions were visually inspected according to a pre-set criterion. The processing time for the segmentation process has been remarkable which suggest its applicability in real time implementations.

Future work will focus on developing an accurate feature extraction method in order to obtain unique features for each leukemia type from the segmented cell images.

REFERENCES

[1]   The Cancer Institute of New Jersey Patient Education Committee, *Leukemia and Lymphoma awareness month*, New Jersy, USA, 2008.
[2]   M.J. Horner, L.. Ries, M. Krapcho, N. Neyman, R. Aminou, N. Howlader, S. Altekruse, E.J. Feuer, L. Huang, A. Mariotto, B.A. Miller, D.R. Lewis, M.P. Eisner, D.G. Stinchcomb, B.K. Edwards (eds), *SEER Cancer Statistics Review, 1975-2006*, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2006/, based on November 2008 SEER data submission, posted to the SEER web site, 2009.
[3]   S. Faderl, H.M. Kantarjian, M. Talpaz, Z. Estrov, Clinical Significance of Cytogenetic Abnormalities in Adult Acute Lymphoblastic Leukemia, *The Journal of The American Society of Hematology*, Washington, USA , Vol. 91, No. 11, 1998.
[4]   Revolution Health. [Online] Sep 06, 2008. Available:http://www. revolutionhealth.com/cond itions/cancer/leukemia/types/acute-lymphoblastic.
[5]   J.N. Jameson, L.K. Dennis, T.R. Harrison, E. Braunwald, A.S. Fauci, S.L. Hauser, D.L. Longo, Harrison's principles of internal medicine. New York: McGraw-Hill Medical Publishing Division, 2005.
[6]   G.A. Colvin, G.J. Elfenbein, The latest treatment advances for acute myelogenous leukemia, *Med Health R I*, Vol. 86, No. 8, 2003, pp. 243-246.
[7]   National Cancer Institute, Finding Cancer Statistics, Cancer Stat Fact Sheets, *Chronic Lymphocytic*, Accessed online November 2009, http://seer.cancer.gov/statfacts/ html/clyl.html.
[8]   Medscape Medical News, Patients with Chronic Myelogenous Leukemia Continue to Do Well on Imatinib at 5-Year Follow-Up 2006. Accessed online November 2009, http://www.medscape.com/viewarticle/536049.
[9]   Conference Summaries, Updated Results of Tyrosine Kinase Inhibitors in CML, *ASCO 2006*, Accessed online November 2009, http://   professional.cancerconsultants.com/conference_a sco_2006.aspx?id=37519.
[10]   R.S. Riley, M. Idowu, Principles and Appl-ications of Flow Cytometry, *Report*, Virginia Commonwealth University, USA. pp. 2-4.
[11]   J. Graham, Automated Image Interpretation in Mammography, Cytogenetics and Retinopathy, *IEE Colloquium on Applications of Image Processing in Mass Health Screening*, London, 1992, pp. 2/1-2/3.
[12]   American Society of Cancer. Accessed online November 2009. http://www.cancer.org/ docroot/home/index.asp
[13]   K. Capell**,** Meeting Leukemia's Diagnostic Challenge, *Business Week*, The McGraw-Hill Companies, London, 5 Sep. 2005.
[14]   Online resources: Available: http://www. cancer.org/docroot/c ri/content/c ri_2_4_3x_how_ I s_leukemia_ diagnosed_ 62.asp.
[15]   N. Guo, L. Zeng, Q. Wu, A method based on multispectral imaging technique for White Blood Cell segmentation, *Computers in Biology and Medicine*, 2005, Vol. 37, pp. 70-76.
[16]   A. Khashman, IBCIS: Intelligent blood cell identification system, *Progress in Natural Science*, 2008, Vol. 18, pp. 1309-1314.
[17]   M. Krukc, S. Osowskia, R. Koktyszd, Recognition and classification of colon cells applying the ensemble of classifiers, *Comp.in Biology and Medic.,* 2009, Vol. 39, pp. 156-165.
[18]   Online   resources: Available: http://www. healthsystem.virginia.edu/internet/hematology/H essEDD/MalignantHematologicDisorders/Leuke mias/AML-M4.cfm.
[19]   B.J. Bain, *Leukemia Diagnosis*, Blackwell Publishing Ltd, London, UK, 2003.
[20]   B. Jane, *Practical handbook on image processing for scientific and technical applications*, CRC Press, 2nd Edition, 2004.
[21]   R.C. Gonzalez, R.E. Woods, Digital Image Processing Using Matlab, Pearson Prentice Hall, Jeresy, USA, Inc., 2004.