

# Automated Detection of Acute Lymphocytic Leukemia-A survey

Sulaja Sanal

Dept. of Computer Science & Engg.  
Sree Buddha College of Engg. for Women.  
Pathanamthitta, Kerala, INDIA  
[sulajasanal.07@gmail.com](mailto:sulajasanal.07@gmail.com)  
+919400923731

**Abstract**— ALL is the most common type of leukemia in children. It is fatal if left untreated. The early detection of ALL is an important factor for the proper treatment. The manual checking of blood smear is time consuming and depends on the operator's ability. Hence automated techniques are introduced to enhance the performance. In this paper different automated method used for the detection of ALL is described.

**Keywords**— ALL, K means clustering, GVF snake, Otsu's thresholding, Zack algorithm, image preprocessing, CIELAB.

## INTRODUCTION

Acute Lymphocytic Leukemia (ALL) is a cancer of the white blood cells. It is characterized by the overproduction and continuous multiplication of immature white blood cells in the bone marrow. It is also known as Acute Lymphoblastic Leukemia. ALL is a fast-growing cancer and it is fatal if left untreated, due to its rapid spread into the bloodstream and other vital organs [1]. But the diagnosis of this disease is very difficult because the symptoms are very similar to flu and other common diseases such as pain in joints and bones, tiredness, weakness etc. Blood test such as full blood count, and liver function test should do if the above mentioned symptoms are present. But this diagnosis only depends on the operator's skill. ALL is mostly seen in children.

The automatic detection of Acute Lymphocytic Leukemia from blood microscopic images generally avoids the problems of manual testing of blood smear and also increases the accuracy. It reduces the computational time and thus increases the efficiency. Automated Acute Lymphocytic Leukemia detection consists of the following steps:-

1. Image preprocessing
2. WBC identification
3. Nuclei extraction
4. Feature selection
5. Classification

The images that are produced by the digital microscope are normally in RGB color space. But these images are difficult to segment because of the change in quality due to the variation in illumination, camera settings etc. Hence the images are converted to CIE L\*a\*b color space images in [2].

The main advantages of this step are

- It reduces memory requirements.
- Increases the computational time.

Segmentation is used in image processing to extract the desired portion of the image for further processing. In this WBCs are extracted to check whether it is cancerous or not. For this, K-means clustering algorithm is used. It is the most popular unsupervised learning algorithm and was published in 1955. The selection of total number of clusters is an important step while using K-means clustering algorithm. In this 3 clusters are selected and that corresponds to nucleus, background and other blood cells such as erythrocytes and leukocyte's cytoplasm. But while using this algorithm, sometimes the edges of some nuclei were obtained instead of the whole nuclei. This problem can be avoided by using some morphological filtering methods such as

- Edge enhancement by Sobel operator.
- Canny edge detector to obtain continuous edge.
- Dilation to connect separated points of the membrane.
- Hole filling to fill the internal holes of the connected element having largest area.

The next step after getting the nuclei is feature extraction. Transforming the input data into a set of features is called feature extraction. Feature selection is an important step because it influences the performance of the classifier. In [2] four different types of features are used.

1. Texture features such as homogeneity, energy, contrast and correlation. Local Binary Pattern (LBP) is used for texture classification.
2. Color features such as mean, standard deviation and nucleus energy.
3. Shape features such as area, perimeter, compactness, major axis, minor axis, eccentricity, form factor, elongation and solidity.
4. Hausdorff dimension (HD) is an additional feature used in [2].

The main advantage of [2] is that the system is applied to complete blood smear images containing multiple nuclei. Many other systems process only sub images and it requires more computational time and memory. Two new features, such as cell energy and Hausdorff dimension (HD), have been used. The result is then compared with the results of the existing models.

In [3], the same procedure such as identification of WBC, extraction of nuclei from that, feature extraction and classification are done, but the methods for doing these steps are different from that used in [3]. The main difference is that, in [3] the leukocytes are separated as sub images from the whole image at first and then identifies the nucleus from the sub images and it classifies the presence of leukemia using neural network. In [3], there are five main modules.

1. Single cell selector module :- It enhances the image first and then identifies single cell
2. White cell identifier module :- It selects WBCs present in the image by separating them from other components
3. Lymphocyte identifier module :- It is used to recognize a lymphocyte with respect to other selected WBCs
4. Feature extraction module: - It takes the image coming from lymphocyte identifier module as input and produces a set of morphological indexes as output. It mainly consists of 3 steps.
  - a. Lymphocyte membrane selection: - This can be achieved by using the techniques sobel enhancement, adaptive canny edge detection, structured image dilation, hole filling, structured image erosion.
  - b. Nucleus and cytoplasm selection: - Otsu's method is the threshold used in this to segment the nucleus from the cytoplasm in the cell image.
  - c. Feature extraction:- The feature set used in this are area, perimeter, Convex Area, Solidity, Major Axis Length, Orientation, Filled Area and Eccentricity
5. Classification module: - It processes the morphological indexes and will classify whether it is cancerous or not.

In [4] ALL-IDB is the image database used for getting the blood microscopic images for processing [5]. WBC identification in [4] consists of several phases.

1. Conversion from RGB to CMYK color model  
This conversion is made because leukocytes are more contrasted in Y component of CMYK color model, because the yellow color is present in all elements except leukocytes.
2. Histogram equalization or contrast stretching operations  
In order to make the segmentation easier, redistribution of image grey level is necessary. For that histogram equalization is used.
3. Segmentation by threshold using Zack algorithm  
Many threshold techniques are present. In this, threshold value based on triangle method or Zack algorithm is used.
4. Background removal operation.  
Background removal processes do not produce a clean result for the whole image. To clean up the image, area opening is used to delete all the objects with size smaller than the structuring element, which are circular in shape. Then the size is calculated based on the average size of the objects in the image.
5. Identification and separation of grouped leukocytes

This phase mainly consist of 2 steps.

- a. Agglomerate identification through roundness analysis- Roundness is defined as the measure of circularity that avoids local irregularity. If roundness equals 1, then it is a circular object and less than 1 indicates deviation from circularity.

b. Watershed segmentation operation- It is used to separate adjacent leukocytes.

6. Image cleaning

This method is used to remove all non-leukocytes and the leukocytes located on the edge of the image, which prevent errors in the later stages of the analysis process. Solidity is the feature used for image cleaning. It calculates the density of an object. If solidity value is 1, then it indicates a solid object, and if value less than 1, it indicates an object with irregular boundary.

7. Feature extraction

The shape descriptors such as area, perimeter, major axis, minor axis and orientation are used as feature set in [4]. These are used to calculate elongation, rectangularity, compactness, convexity, roundness and the solidity.

The main disadvantage of shape feature is that they are more susceptible to errors in segmentation. For reducing these errors, shape descriptors are used together with regional descriptors.

1. Classification

For classification, SVM [6] is used. To evaluate the performance of SVM model the results is compared with k-Nearest Neighbor (k-NN) using the Euclidean distance measure with different values of k.

WBCs with giant nuclei are the main symptom of leukemia. But it is not sufficient to prove this disease and also other symptoms must be investigated. Another symptom of leukemia is the existence of nucleolus in nucleus. In [7] to diagnose this symptom and to discriminate between nucleoli and chromatins, curvelet transform [8] is used. It is a multi-resolution transform for detecting 2D singularities in images.

At first the image is separated into R G and B components. The median filter is then applied to R and G components. Then enhance the image using histogram equalization and Luv color transform. Then nuclei are extracted using K-means clustering algorithm. Then curvelet transform is applied on extracted nuclei and the coefficients are modified, and finally reconstruct a new image is used to extract the candidate locations of chromatins and nucleoli. For extracting the candidate zone of nucleolus feature based on the gradient of saturation channel is used. The method is applied on 100 microscopic images. The main advantage of [7] is that it also considers the nucleolus in addition to the nuclei for ALL detection.

Based on a new segmentation framework, WBCs are segmented into nucleus and cytoplasm in [9]. Twenty microscopic blood images were tested in this. At first the RGB images are converted to grey scale images. All further operations are performed in this grey scale image. Then nuclei of leukocytes are extracted by using Gradient Vector Flow (GVF) Snake model (active contours). Snakes are deformable curves that can move and change the shapes to deform to boundaries of objects in an image. Then nucleus is extracted and hole filling is done to enhance the image. In order to get the cytoplasm, the nucleus is subtracted from the grey scale image. But the image should not be clean. In order to enhance the image, Zack thresholding techniques are used. The nucleus segmentation part is based on morphological analysis, and the cytoplasm segmentation is based on pixel-intensity thresholding. The results show that the proposed method is able to yield 92% accuracy for nucleus segmentation and 78% for cytoplasm segmentation.

In [10], automated detection and sub classification of Acute Lymphocytic Leukemia using image processing and machine learning is used. In this, both supervised and unsupervised frameworks are used for segmentation of nucleus and cytoplasm. These methods are based on neural networks, feature space clustering and Markov random field modeling.

The morphological components of normal and malignant lymphocyte may differ. To automatically recognize ALL from blood samples, morphological, textual and color features are extracted from the segmented nucleus and cytoplasm.

In this, the sub classification of ALL is done based on French-American-British criteria. According to this, two methodologies based on morphology and phenotype is used. This includes segmentation, nucleus and cytoplasm extraction and classification. An improved scheme is also proposed to determine the origin of blast cells. Then an ensemble of decision tree is used to map the extracted features into either lymphoid or myeloid.

In [11], the ALL detection is performed based on fuzzy logic and neural network. 17 main features based on size, shape and color are extracted from the segmented nucleus and then these features are used for the distinction between normal and cancerous cell. The Multi-Layer Perceptron network trained by Scaled Conjugate Gradient (SCG) back propagation algorithm [12] and Fuzzy

network is used to classify the segmented regions into two classes either normal or abnormal. The result shows that the MLP network trained using SCG algorithm is able to achieve an acceptable classification performance compared to the Fuzzy network.

## CONCLUSION

Automated detection of ALL minimizes the overhead of manual blood smear processing. The efficiency also increases by using the automated techniques. Many new techniques are introduced in this field for enhancement and among that image processing techniques are most popular nowadays. Some of the image processing techniques for the automated detection of ALL are studied in this paper.

## REFERENCES:

- [1] K. Breden, T. Schorr, J B Schorr, "Blood", Colliers Encyclopaedia, Vol. 4, 1978.
- [2] Sos Aгаian, Senior Member, IEEE, Monica Madhukar, and Anthony T. Chronopoulos, Senior Member, IEEE, "Automated Screening System for Acute Myelogenous Leukemia Detection in Blood Microscopic Images"IEEE SYSTEMS JOURNAL, 2014.
- [3] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in Proc. CIMSA, 2005, pp.96–101.
- [4] Lorenzo Putzu, Giovanni Caocci, Cecilia Di Ruberto "Leukocyte classification for leukemia detection using imageprocessing techniques",Artificial Intelligence in Medicine.2014.
- [5] Donida Labati R, Piuri V, Scotti F., "ALL-IDB: the acute lymphoblastic leukemia image Database for image processing in: Macq Benot, Schelkens Peter, editors".Proceedings of the 18th IEEE ICIP international conference on image processing,September 11–14. Brussels, Belgium: IEEE Publisher; 2011. p. 2045–8.
- [6] Hsu, C-W, Chang C-C, Lin C-J, "A practical guide to support vector classification. Available from: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>".
- [7] Ramin Soltanzadeh, Hossein Rabbani, and Ardeshir Talebi " Extraction of Nucleolus Candidate Zone in White Blood Cells of Peripheral Blood Smear Images Using Curvelet Transform," Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, 2012.
- [8] O. Tuzel, L. Yang, P. Meer, and D. J. Foran, "Classification of hematologic malignancies using texton signatures," Pattern Analysis and Applications, vol. 10, no. 4, pp. 277–290, 2007.
- [9] Farnoosh Sadeghian, Zainina Seman, Abdul Rahman Ramli, Badrul Hisham Abdul Kahar, and M-Iqbal Saripan, "A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing", Biological Procedures Online, Volume 11, Number 1, 2009.
- [10] Subrajeet Mohapatra, "Hematological Image Analysis for Acute Lymphoblastic Leukemia Detection and Classification", Department of Electrical Engineering, National Institute of Technology Rourkela.
- [11] Nurul Hazwani Abd Halim1, Mohd Yusoff Mashor, Aimi Salihah Abdul Nasir and Rosline Hassan, " Performance Comparison between Multilayer Perceptron and Fuzzy ARTMAP Networks for Acute Leukemia Detection", International Journal of Research and Reviews in Computer Science (IJRRCS) Vol. 2, No. 5, October 2011, ISSN: 2079-2557.
- [12] M. F. Moller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," Neural Networks, vol. 6(4), pp. 525–533, 1993