Automatic Recognition of Acute Lymphoblastic Leukemia Cells from Microscopic Images

Pouria Mirmohammadi, Afsoon Taghavi , Ali Ameri*
Department of Biomedical Engineering and Medical Physics, Faculty of Medicine,
Shahid Beheshti University of Medical Sciences, Tehran, Iran
*Corresponding author: ameri@sbmu.ac.ir

Abstract:- **This paper provides an algorithm for automatic recognition of Acute Lymphoblastic Leukemia (ALL) cells from microscopic images. In the proposed system after segmenting nucleus of cell using fuzzy c-means clustering algorithm, numerous geometric and statistical features were extracted. These features were reduced using principle component analysis (PCA), and were used as inputs to a support vector machine (SVM) classifier. This algorithm shows promising results, and its performance for detection of ALL and its sub-types makes it an assistant diagnostic tool for hematologists.**

Keywords-component; Acute lymphoblastic leukemia, Fuzzy c-means clustering, Principle component analysis, SVM classifier

I. INTRODUCTION

Leukemia is one of the most common cancers worldwide, which affects bone marrow and blood. It develops within the body with duplication of a large number of abnormal white blood cells. Acute lymphoblastic leukemia (ALL) is the most common form of children cancer of white blood cells. The French-American-British (FAB) classification categorizes ALL into three morphological subtypes of L1, L2 and L3.The first step to identify this type of leukemia is observation of cells blast in the peripheral blood smear or increase in the smear of bone marrow by pathologists [1]. However, this approach is time-consuming, boring and grinding for pathologists, and also, diagnosis is dependent on the pathologist's skill. To overcome these problems, several scholars have worked on automatic systems for ALL diagnosis from microscopic images.

For white blood cell segmentation in microscopic images, Sinha et al. [2] ,applied two-step procedures in HSV color space using fuzzy c-means or FCM clustering. After feature extraction from the segmented nucleus and cytoplasm, several classifiers have been used on various combinations of features. Theera-Umpon et al. [3] presented an automatic system for segmentation of nucleus and cytoplasm of WBC based on the FCM clustering and morphology operations in image processing. For assessment of their proposed system, they evaluated their outcomes with the cell images manually segmented by an expert. Madhloom et al. [4] performed WBCs segmentation by a fusion of automatic contrast stretching supported using image arithmetic operation, minimum filter and global threshold methods. They reached correctness between 85-98%. Halim et al. [5] applied segmentation on HSV color space to remove the WBCs in background. To separate the overlapping cells, they applied an erosion morphological operator. Their method presented the mean accuracy of 97.8% for counting both acute lymphocytic leukemia and acute myeloid leukemia (AML) cases. Abbas et al. [6] segmented the Nuclei of cells by image processing techniques such as OTSU global thresholding and morphological operation dilation. They employed 380 microscopic images and achieved a segmentation sensitivity of 96.5%. In [7] MoradiAmin et all. used k-means process to segment cell nuclei. Then nuclei geometric and statistical features were extracted. Features that led to highest classification accuracies were chosen as the inputs of an SVM classifier.

In this paper, the procedure is similar to that in [7] except that we have altered the nuclei segmentation process, i.e. k-means, with fuzzy c-means since the prior rarely yields empty cluster while running. Another difference is pertinent to feature selection procedure. Here we applied the first 6 principle components of the feature space that were taken by PCA dimensionality reduction procedure.

II. MATERIALS AND METHODS

A. Data Collection

In this study, blood smear of 14 patients with ALL from Isfahan Al-Zahra and Omid hospital s' pathology laboratories were provided and in total 146 images were taken by high resolution digital camera, Nikon1 V1, connected to Nikon Eclipse 50i light. A total number of 643 cells were acquired (277 L1, 215 L2 and 151 L3).

B. Proposed Algorithm

Our proposed algorithm for ALL detection contains four steps as shown.in Fig. 1.

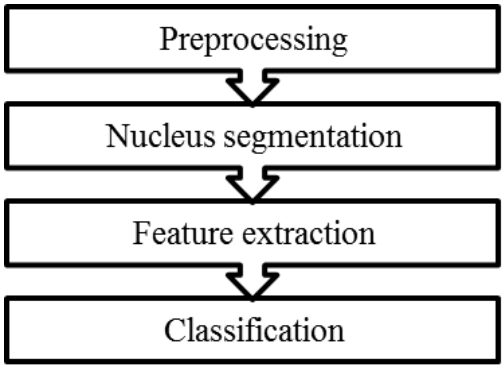
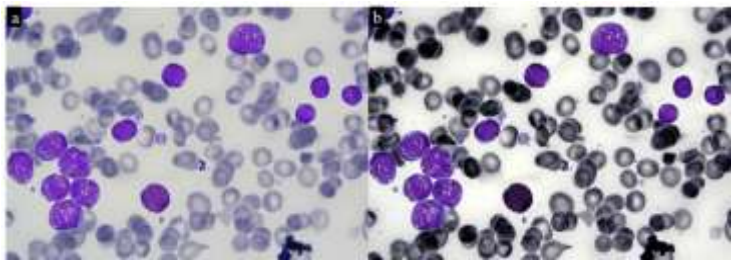


Fig. 1: A schematic diagram of our proposed algorithm.

1. Preprocessing

Image preprocessing which was used to enhance the image quality was performed by histogram equalization on V-band in HSV color space. HSV is the most common cylindrical-coordinate representation of points in an RGB color model. It stands for Hue, Saturation, Value and describes colors (hue or tint) in terms of their shade (saturation or amount of gray) and their brightness (value or luminance). An example of the preprocessing results is shown in Fig. 2.



(a) Original image (b) Enhanced image

Fig. 2: An example of preprocessing results.

2. Nucleus segmentation:

Nucleus segmentation plays a vital role in ALL recognition procedure, because it directly affects the subsequent steps (i.e. feature extraction and classification). Fuzzy c-means clustering is a method of vector quantization, and is popular for cluster analysis in data mining. In this work, the segmentation of ALL cell nucleus was performed using fuzzy c-means algorithm with four clusters on H and S bands of HSV color space.

3. Feature extraction

Two sets of geometric and texture features were extracted from nucleus. Area, perimeter, solidity, eccentricity and extent as geometric features and entropy, mean, standard deviation, energy and skewness as texture features were extracted. Geometric features were extracted from binary images which were obtained from the original images by applying thresholding based methods. In the contrary, texture features that are usually obtained from gray scale images, provides us with information about the spatial arrangement of color or intensities in an image or selected region of an image. These features were extracted from the gray scale image histograms of the red, green and blue, in addition to the hue, saturation and value bands from the original and improved image. In total, 72 texture features were generated using this method.

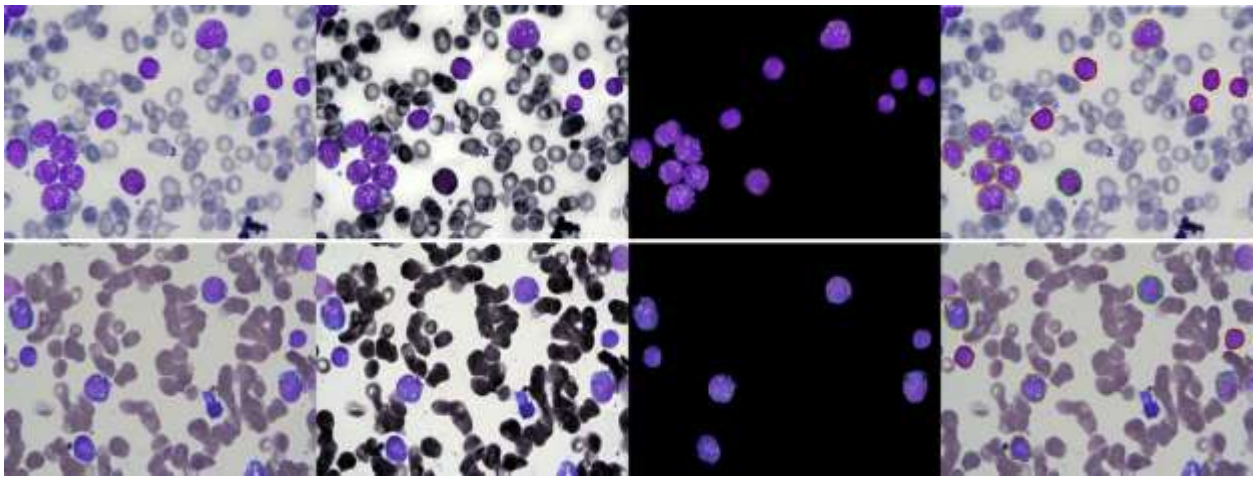
The above approaches of feature extraction yielded 77 features from each nucleus. Some of them were powerfully correlated with the others and some acted as noise. Therefore, there was a motivation to decrease the size of the feature set. For feature reduction we applied PCA and the first 6 components were selected as inputs to the classifier.

4. SVM classifier

In order to classify the cells, a Support Vector Machines (SVM) classifier was applied. SVM is a supervised learning model that analyzes data and recognize patterns and is used for both classification and regression analyses. Because of the nature of ALL , that is categorized into 3 morphological subtypes, we used multiclass SVM for classification of ALL.

III. **Results**

The proposed algorithm, automatically and robustly classifies ALL into three predefined classes, L1, L2 and L3. Fig. 3 clearly shows robustness of our algorithm in classification of ALL. In this figure, the original image, enhanced image, cluster of nuclei and detected cells are depicted for two sample images. Fig. 3 (d) distinguishes between L1, L2 and L3 cells by red, green and yellow colors respectively.



(a) original image (b) enhanced image (c) cluster of nuclei and (d) detected cells

Fig. 3: Results of the proposed algorithm for ALL detection of two sample images.

Evaluation of the proposed algorithm was performed by three statistical parameters including sensitivity, specificity and accuracy for all 3 types of L1, L2 and L3 and the values are listed in Table. 1.

Table 1. Statistical results for evaluation of the proposed method vs. similar work [7].

	L1		L2		L3	
statistical parameters (%)	Proposed method	Similar work [7]	Proposed method	Similar work [7]	Proposed method	Similar work [7]
sensitivity	92	91	86	84	97	97
specificity	97	97	96	95	99	99
Accuracy	96	95	92	92	99	99

IV. **Conclusion**

Based on the results it is obvious that our proposed algorithm has a better performance in order to [7] for the diagnosis of ALL in comparison with previous algorithms, and can be used as an assistant diagnostic tool for pathologists. Future work will include segmentation of cytoplasm and extraction of geometric and texture features from cytoplasm in addition to nucleus.

REFERENCES

[1] Bain, B.J., *Leukaemia diagnosis*2010: John Wiley & Sons.
[2] N. Sinha and A. Ramakrishnan, "Automation of differential blood count," in TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region, pp. 547-551, 2003.
[3] N. Theera-Umpon, "White blood cell segmentation and classification in microscopic bone marrow images," in Fuzzy systems and knowledge discovery, ed: Springer, pp. 787-796, 2005.
[4] H. Madhloom, S. Kareem, H. Ariffin, A. Zaidan, H. Alanazi, and B. Zaidan, "An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold," 2010.
[5] N. H. A. Halim, M. Y. Mashor, and R. Hassan, "Automatic Blasts Counting for Acute Leukemia Based on Blood Samples," International Journal of Research & Reviews in Computer Science, vol. 2, 2011.
[6] N. Abbas and D. Mohamad, "Automatic Color Nuclei Segmentation of Leukocytes for Acute Leukemia," Research Journal of Applied Sciences, Engineering and Technology, vol. 7, pp. 2987-2993, 2014.
[7] M. M. Amin, S. Kermani, A. Talebi, and M. G. Oghli, "Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using K-Means Clustering and Support Vector Machine Classifier," Journal of medical signals and sensors, vol. 5, p. 49, 2015.