Acute Leukemia Classification by Using SVM and K-Means Clustering

Jakkrich Laosai and Kosin Chamnongthai

Department of Electronic and Telecommunication Engineering, Faculty of Engineering
King Mongkut's University of Technology Thonburi, Bangkok, Thailand
Email: jakkrich1@hotmail.com, kosin.cha@kmutt.ac.th

ABSTRACT

The proposed system takes as input, Color images of stained peripheral blood smears and identifies the class of each of the White Blood Cells (WBC). The process involves segmentation, feature extraction and classification. Our work focuses on classification of Foil of Bretagne (Lymphoid) and Almeida Lloyd (Myeloid). So that, physicians can analyze, detect anomalies and ensure the diagnosis. The experiment results showed that the performance of identification leukemia using our image processing techniques could classify 100 sample images to Lymphoid stem cells and Myeloid stem cells The method has been evaluated using K-Means clustering. Features extracted from the segmented cytoplasm and nucleus, are motivated by the visual cues of shape and texture. Various classifiers have been explored on different combinations of feature sets. The results presented here are based on trials conducted with normal cells. The highest performance using SVM was of 92%.

Keywords : *Classification , k-means clustering Segmentation, Support Vector Machine(SVM), White Blood Cells (WBC)*

I. INTRODUCTION

White blood cells (WBC) or leukocytes play a significant role in the diagnosis of different diseases (including Leukemia), and therefore, extracting information about that is valuable for hematologists. Leukemia refers to a progressive, malignant disease of the blood-forming organs. Complete blood count process is the first step for leukemia screening [4]. The pathology is characterized by the uncontrolled accumulation of immature white blood cells. The four main types of leukemia are Acute Myelogenous Leukemia (AML); Acute Lymphoblastic Leukemia (ALL); Chronic Myeloid Leukemia (CML); Chronic Lymphocytic Leukemia (CLL). In this paper, we build up a decision support tool to improve classification of Acute Myelogenous Leukemia (AML). AML is a fast-growing cancer of the blood and bone marrow. In AML, the bone marrow produces many unformed cells called blasts. Blasts usually develop into white blood cells that are used for defense mechanism in the body. However, the blasts are not fully formed in AML and hence cannot fight infections. Acute Myelogenous leukemia is often difficult to diagnose since the precise cause of AML is still unknown. Also the common diseases, like fever, weakness, tiredness, or aches in bones or joints. Since there is no staging for acute myelogenous leukemia, choosing the type of treatment can vary from chemotherapy, radiation therapy, bone marrow transplant, and biological therapy. [1]. AML is one of the most common types of leukemia in adults, affecting more than 10,000 adults that are diagnosed every year in the U.S. AML also makes up 15 to 20 percent of childhood leukemia. About 500 children and adolescents in the U.S. each year are affected by AML. [2]

This paper presents a method for classifying leukemia. Leukemia is divided into 2 categories which are acute leukemia and chronic leukemia and other subordinate types. Because of categorization, some types of this disease show similar blood cell characteristics are hardly distinguishable. For this reason, the idea of studying the differentiation of acute leukemia with image processing techniques is applied. Our work focuses on classification of Foil of Bretagne (Lymphoid) and Almeida Lloyd (Myeloid). The performance evaluation of the system based on the parameters extracted. The method has been evaluated using K-Means clustering. Features extracted from the segmented cytoplasm and nucleus, are motivated by the visual cues of shape and texture. Various classifiers have been explored on different combinations of feature sets.

II. AML SYSTEM OVERVIEW

Nucleus Segmentation

The aim of nucleus segmentation is separating the nucleus from the other parts of a cell and a microscopic blood smear image. A typical peripheral blood smear image consist of four components: red cells (nucleated cells), white blood cells nucleus , cytoplasm, and background which contains platelets and spot noise. WBCs appear in a different color from red cells and the other parts of a peripheral blood smear images. In

978-1-4799-3174-3/14/$31.00 ©2014 IEEE

this system using color information , after applying several preprocessing steps, nuclei are extracted by means of K-means method. The proposed approach aims to present a more robust system with an efficient segmentation of blood images for high performance.

III. PRE-PROCESSING: COLOR-CORRELATION

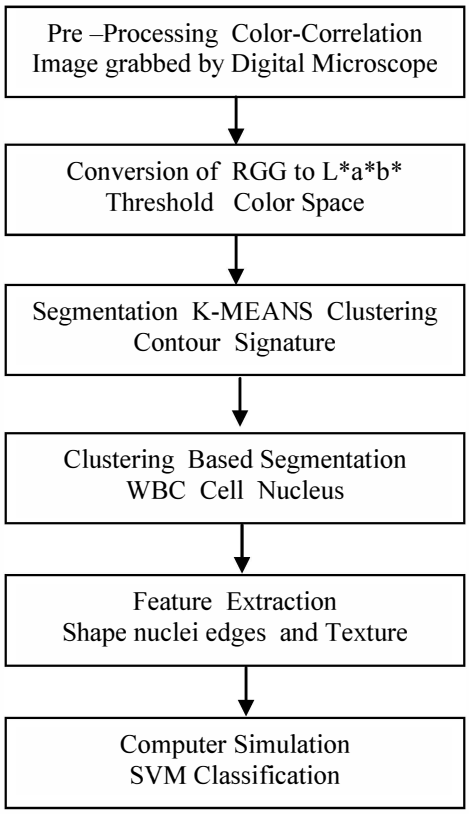

Figure 1: System Over-view

IV. SEGMENTATON K-MEANS CLUSTERING

In this paper, at first, nuclei are extracted by clustering the microscopic images into three color clusters in Luv color system using *K*-means method [3]. The K-means is a clustering method which is one of the most popular unsupervised learning algorithm due to it' s simplicity. The K-means clustering has been used for image segmentation. K-means clustering is based on minimizing the objective function

To consider an objective function of the method, it can rewrite as Equation (1).

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - C_j \right\|^2 \qquad (1)$$

Where, *J* is the distance between the pixel (x) and the cluster center(c),k is the number of cluster and n is the number of data

King Mongkut's University of Technology Thonburi, Bangkok,Thailand.

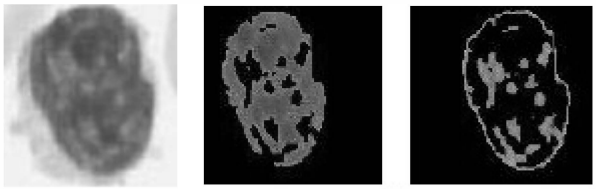

Figure2: Examples of Segmentation using K-means Clustering

While performing K-means segmentation, it was observed that in some of the segmented images, we obtained only the nuclei edges (refer Figure 2:) as opposed to the whole nuclei. Hence to make the system more robust, all segmented images were made to undergo morphological filtering [5]. Texture based features were then extracted from the resultant images.

V. FEATURE EXTRACTION

Pattern recognition is selecting the proper diagnostic features, describing the image by the numerical values, and enabling the automatic system to perform the recognition. Usually many of the features using texture, geometrical and statistical analyses of the image .The classification of acute leukemia by FAB system. FAB (French-American-British Corporative group) is the original classification scheme proposed which divides AML into 8 subtypes (M0 to M7) and ALL into3 subtypes (L1 to L3) [7]. The FAB classification of ALL and AML is based on morphology and Cytochemical staining of blasts. Those are differentiated based on morphology, including cell size, prominence of nucleoli, color of cell and the amount and appearance of cytoplasm. The FAB classification is shown in the Table1.

VI. COMPUTER SIMULATION

SVM is a powerful tool for data classification based on hyper plane classifier. This classification is achieved by a separating surface (linear or nonlinear) in the input space of the data set. [6] They are basically two class classifiers that optimize the margin between the classes . The classifier training algorithm is a procedure to find the support vectors. Shape and texture based features are extracted for the image nucleus sample and recorded. Few measurements are tabulated in Table 1. Among all the features the most relevant features are selected and used to train the SVM [8] such as number of nuclei lobes it differentiate the nucleus of each type, ratio of nuclei to cell ,ratio of perimeter to nuclei and entropy. The number and structure of nucleus lobes is one of the prominent features used to identify the class of the WBC. [10]

VII. PERFORMANCE EVALUATION

Precision, Specificity, Sensitivity and F-Measure are all defined in relation to the possible outcomes of the classifier system. *True Positive* (cancerous cell correctly identified), *False Positive* (non -cancerous cells identified as cancerous), *True Negatives* (non-cancerous correctly identified), *False Negatives* (Cancerous cells identified as non-cancerous).

TABLE I. PERFORMANCE EVALUATION PARAMETERS

No.	Evaluation parameters		
	Parameters	*Formulae*	*Value*
1	*Sensitivity* -also called recall denotes the test's ability to identify positive results	TP/(TP+FN)	92%
2	*Specificity* denotes the test's ability to identify negative results	TN/(TN+FP)	82%
3	*Precision* -gives the proportion of subjects with positive results who are correctly identified.	TP/(TP+FP)	90%
4	*F-Measure*- is a metric that gives the harmonic mean of Precision and Sensitivity. It is the overall classification performance	2 x Precision x Sensitivy / Precision + Sensitivity	88%

TABLE II Performance Evaluation Statistics

Measure	Value
Sensitivity	92%
Specificity	82%
Precision	90%
F-Measure	88%

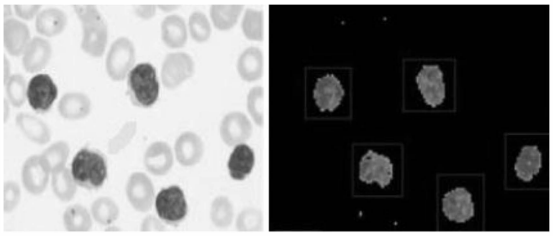

Figure 3: Initial K-Means Segmentation [11]

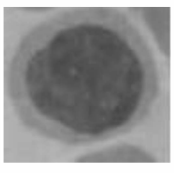

Figure 4: Cropped Sub Images

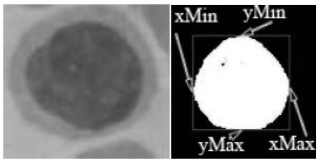

Figure 5: Pictures of white blood cells from the top and bottom Copy the image size. boundaries. $(y\max - y\min) \times (x\max - x\min)$

TABLE III FAB (French-American-British Corporative group)

FAB Classification	
Acute lymphocytic leukemia (ALL)	Descriptive Term
L1	Small monotonous lymphocytes
L2	Mixed L1- and L3-type lymphocytes
L3	Large homogeneous blast cells
Acute myeloid leukemia (AML)	Descriptive Term
M0	Acute myeloblastic leukemia, undifferentiated
M1	Acute myeloblastic leukemia, without maturation
M2	Acute Myeloblasts with maturation (best AML prognosis)
M3	Acute promyelocytic leukemia
M4	Acute myelomonocytic leukemia
M5	Acute monocytic leukemia
M6	Erythroleukemia/DiGuglielmo syndrome
M7	Acute megakaryoblastic leukemia

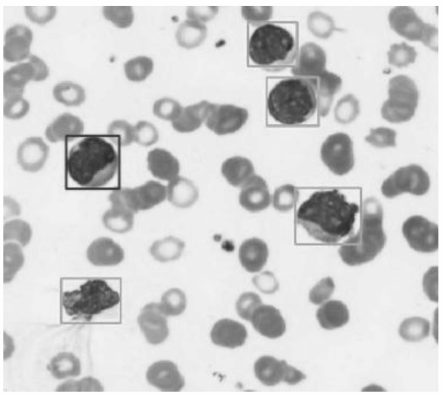

Figure 6: Classification of ALL and AML [4]

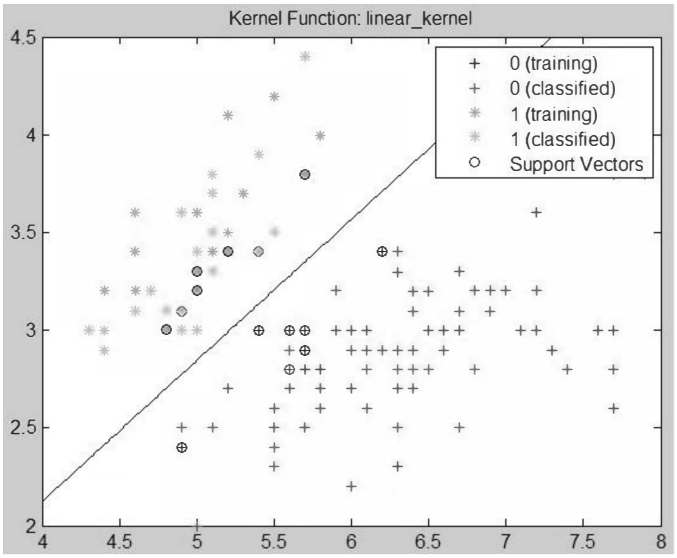

Figure 7: The image depicting use of SVM classifier for binary classification [9]

VIII. CONCLUSION

In this work, our techniques could classify leukocytes and categorize them into 2 major types which are Lymphoid stem cells and Myeloid stem. In this work a robust classifier system which segments and classifies microscopic blood images. The experiment results showed that the performance of identification leukemia using our image processing techniques could classify 100 sample images to Lymphoid stem cells and Myeloid stem cells. The method has been evaluated using K-Means clustering. Features extracted from the segmented cytoplasm and nucleus, are motivated by the visual cues of shape and texture. Our work focuses on classification of Foil of Bretagne (Lymphoid) and Almeida Lloyd (Myeloid). So that, physicians can analyze, detect anomalies and ensure the diagnosis. Various classifiers have been explored on different combinations of feature sets. The results presented here are based on trials conducted with normal cells. The highest performance using SVM was of 92%.

IX. ACKNOWLEDGEMENT

The authors would like to thank advisers, This research has been partially supported by King Mongkut's University of Technology Thonburi, Bangkok,Thailand. We are also thankful to Ubonratchathani Cancer Hospital for providing us with high quality image database.

REFERENCES

[1] Shi S., Kan G. and Liang Ch., "An Improved Segmentation Method for Color Image", Second International Symposium on Intelligent Information Technology Application (2009).

[2] Scotti,F., " Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images",in 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Application. CISMA. Pages 96-101,(2005).

[3] Sinha, N.; Ramakrishnan, A.G.; Dept. of Electr. Eng., Indian Inst. of Sci., Bangalore, India, "Automation of differential blood count". In Proceedings Conference on Convergent Technologies for Asia-Pacific Region, volume 2, pages 547–551, (2003).

[4] Ubonratchathani Cancer Hospital Pictures of white blood cells WBC blood diseases worldwide. [Online]. Available: http://www.uboncancer.go.th/

[5] A. P. Pentland, "Fractal based description of natural scenes ." IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:661 – 674, (1984).

[6] Sztojanov, I.; Voinea, V.; Stanica, J.-L.; Popescu-Mina, C.; Univ. Politeh. of Bucharest, Bucharest, Romania ," Fractal technologies for image processing in biology", in 3rd International Workshop on Soft Computing Applications, 2009. SOFA '09. Pages 139-144,(2009).

[7] J. Laosai, T. Panupong, and K. Chamnongthai "Acute Leukemia Classification using Digital Image Processing" Proceedings of The 2013 International Electrical Engineering Congress (iEECON2013) pp.0015

[8] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, vol.2, No.2, pp.121-167,(1998).

[9] B. Scholkopf, A. J. Smola, "Learning with Kernels: Support Vector Machines, Regularization Optimization and Beyond", The MIT Press, (2002).

[10] N. Cristianini, J. Shawe-Taylor, "An Introduction to Support Vector Machines and other kernel-based learning Methods", Cambridge University Press 2000.

[11] Payam aeilzadeh, Lei Tang, Huan Liu, Arizona State University , "Cross-Validation", in Encyclopedia of Database Systems EDBS), Editors: Ling Liu and M. Tamer Özsu. Springer, pp6. (2009).