31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Identification of different types of Lymphoblasts in Acute Lymphoblastic Leukemia using Relevance Vector Machines

Lalit Gupta, Sanjay Jayavanth and Aruna Ramaiah

Abstract— **Acute Lymphoblastic Leukemia (ALL) is a blood cancer, which is characterized by an abnormal proliferation of lymphoblasts (a form of white blood cells). It leads to debilitating condition of the patient and necessitates an early therapeutic intervention. For diagnosis of ALL, a pathologist examines bone marrow smear for the count of different types of lymphoblasts to arrive at the diagnosis. The different counts of three different lymphoblasts can affect the type and intensity of the therapy to be used as well as the likely course of the disease. However, bone marrow smear examination and its interpretation is a tedious time consuming process. Therefore, any automation in terms of easing and pacing up the workflow is required. In this paper, we exactly do that by proposing an automatic method for recognition of different types of cells in ALL. The proposed algorithm mimics the pathological and ontological descriptions of bone marrow cells, in identifying abnormalities that lead to ALL. The proposed method uses discriminative shape, colour and texture features, which supposedly contains information for better discrimination of bone marrow cells. Further feature selection techniques, based on mutual information distribution and recursive feature elimination along with Relevance Vector Machines (RVM) are used for effective classification. The results are analyzed on more than 345 cell images. Sensitivity and specificity of above 93% has been achieved.**

I. INTRODUCTION

Leukemia is a blood or bone marrow cancer, which is characterized by an abnormal proliferation of blood cells, usually white blood cells (leukocytes). Leukemia is typically diagnosed by examination of bone marrow cells through microscopy. A sample of bone marrow is usually aspirated from the hip (pelvic bone) and smeared on a glass slide. The sample is then stained with the standard procedure of Giemsa or Leishman staining and analyzed under a microscope. The bone marrow cells are evaluated from their morphology, viz., size, shape, and granularity. Leukemia is classified based on severity of condition i.e. acute or chronic and its association to granulocytes or lymphocytes as:

1) Associated with granulocytes
   a) Acute Myelogenous Leukemia (AML)
   b) Chronic Myelogenous Leukemia (CML)
2) Associated with lymphocytes
   a) Acute Lymphoblastic Leukemia (ALL)

Lalit Gupta is with Philips Research Asia Bangalore, Philips Innovation Campus, Manyata Tech Park, Nagavara, Bangalore,560043 India (phone: 91 80 4189 2249; fax: 91 80 4189 2265; email: lalit.gupta@philips.com).

Sanjay Jayavath is with Philips Research Asia Bangalore, Philips Innovation Campus, Manyata Tech Park, Nagavara, Bangalore,560043 India (phone: 91 80 4189 2280; fax: 91 80 4189 2265; email: sanjay.jayavath@philips.com).

Aruna Ramaiah is with Philips Research Asia Bangalore, Philips Innovation Campus, Manyata Tech Park, Nagavara, Bangalore,560043 India (fax: 91 80 4189 2265;).

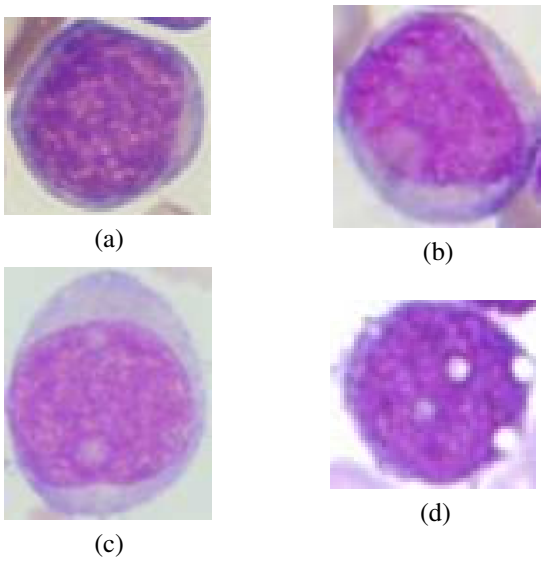

Fig. 1. Different types of lymphoblasts: (a) L1; (b) L1; (c) L2; (d) L3.

b) Chronic Lymphocytic Leukemia (CLL)

In this paper, we have addressed different forms of acute lymphoblastic leukemia. Acute Lymphoblastic Leukemia (ALL) is a malignant disease caused by the abnormal proliferation of early lymphocytes called as lymphoblast, in the bone marrow. This unabated proliferation of lymphoblasts in bone marrow could cause bone marrow failure leading to an acute condition. ALL could also originates in the blast cells of thymus and lymph nodes. About 80% of childhood leukemia's are due to ALL and most cases occur in children in the age group of 3 to 7 years. A confirmed diagnosis is essential for timely therapeutic intervention and hence a bone marrow study is essential. Lymphoblasts can be classified into three different subtypes L1, L2 and L3. The different morphological types of lymphoblasts are shown in Fig. 1 and their morphological features along with type of differential features are given in Table I. The prevalence of L1, L2 and L3 is 85.1%, 14.1% and 0.8%, respectively with respect to each other. For diagnosis of ALL, a pathologist looks for the count of different types of lymphoblasts to arrive at the diagnosis. The different counts of three different lymphoblasts can affect the type and intensity of the therapy to be used as well as the likely course of the disease. Children with greater than 25% of L2 lymphoblasts have a significantly higher relapse rate and significantly poorer survival [1].

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

TABLE I

MORPHOLOGICAL FEATURES OF DIFFERENT TYPES OF LYMPHOBLASTS.

Morphology	L1	L2	L3	Feature type
Cell size	Small	Large	Large	Shape
Nuclear chromatin	Fine/clumped	Fine	Fine	Color, Texture
Nuclear shape	Regular,cleft/indentation	Irregular, cleft/indentation	Regular, oval to round	Shape
Nucleoli	Indistinct	one or more/cell,large prominent	one or more/cell,large prominent	Color, Shape
Amount of cytoplasm	Scanty	Moderate	Moderate	Color, Texture
Cytoplasmic basophilia	Slight	Slight	Prominent	Color, Texture
Cytoplasmic vacuoles	Variable	Variable	Prominent	Shape, Texture, Color

In this paper the proposed algorithm mimics the pathological and the ontological descriptions for automated identification and analysis of bone marrow cells as described above. An algorithm for identification of three types of lymphoblasts is proposed in this paper. In [2][3] authors have reported bone marrow cells identification. Contrary to these methods we propose to used more discriminative features; colour and texture features such as mean in HSV color space and texture features using wavelets, which contains information that enable better discrimination. Further feature selection techniques, based on mutual information distribution and recursive feature elimination along with RVM are used for affective classification.

Rest of the paper is organized as follows: Section II describes the complete algorithm in detail. It discusses various features and algorithms used for feature selection and classification. Results are discussed in Section III and the paper is concluded in Section IV with a discussion on future work.

II. AUTOMATIC RECOGNITION OF BLASTS CELLS

The proposed classification methodology is described in Fig. 2. A cell image is first segmented to separate nucleus region and cytoplasm region. Features are then computed from these two regions, separately. The most relevant features are selected using a feature selection algorithm and finally, the relevant features are passed to Relevance Vector Machine (RVM)/ Support Vector Machine (SVM) for classification. The rest of the section describes these steps in detail.

A. SEGMENTATION

Bone marrow cells are segmented to differentiate between nucleus region and cytoplasm region. It can be observed from the images shown in Fig. 1 and Fig. 3(a) that these regions are very similar and there is variance within and across cell images therefore segmenting the image using any adaptive threshold would be difficult. To handle these variations, first histogram of the green component of the image is projected on the 'S' curve (A sigmoid curve is produced by a mathematical function having an "S" shape. Often, sigmoid function refers to the special case of the logistic function [4]) to enhance the image to increase the contrast between the nucleus region from the rest and Otsu algorithm [5] is used to segment the nucleus. The region surrounding the nucleus is considered as cytoplasm [6]. Example images showing these steps are shown in Fig. 3.

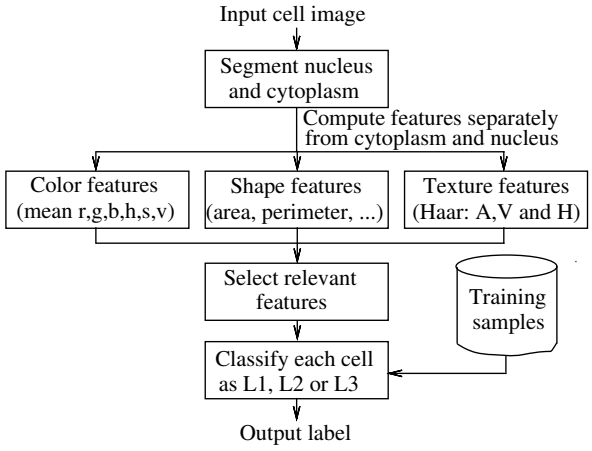

Fig. 2. Block diagram of the complete system.

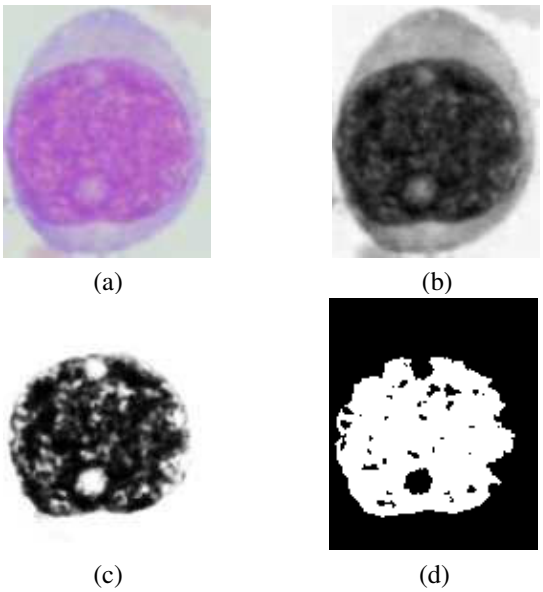

Fig. 3. Segmentation of nucleus region from a cell image: (a) input image; (b) green component of the image; (c) enhanced image after projecting image histogram on 'S' curve; (d) segmented output (nucleus region).

B. FEATURE EXTRACTION

Features extracted from each of the images are; colour, texture and shape characteristics. They are extracted separately from the nucleus and the cytoplasm. Each type of the feature is normalized and concatenated to form the augmented feature vector:

$$\mathbf{X} = [\mathbf{x}^{\text{cyto}}_{\text{colour}} \mathbf{x}^{\text{cyto}}_{\text{texture}} \mathbf{x}^{\text{cyto}}_{\text{shape}} \mathbf{x}^{\text{nuc}}_{\text{colour}} \mathbf{x}^{\text{nuc}}_{\text{texture}} \mathbf{x}^{\text{nuc}}_{\text{shape}}]^T. \tag{1}$$

where, 'cyto' refers to cytoplasm and 'nuc' refers to nucleus based features. In the following, each type of features used for classification is discussed.

1) Colour: For each segment of the image, the mean colour values in RGB and HSV color spaces are taken as the feature

$$\mathbf{X}_{\text{colour}} = [\mu_R \quad \mu_G \quad \mu_B \quad \mu_H \quad \mu_S \quad \mu_V]^T \tag{2}$$

where, μ represents the mean for the red (R), green (G), blue (B), hue(H), saturation(S) and value(V) components.

2) Texture: Haar wavelet is used for computing texture features [7] from the cell images. A feature vector for each segment is computed by taking mean of all the features associated with the pixels in a segment as

$$\tilde{\mu}_{A_R} = \frac{1}{P} \sum_{(x,y)\in\xi} \mu_{A_R}(x,y) \tag{3}$$

$$\tilde{\sigma}_{V_R} = \frac{1}{P} \sum_{(x,y)\in\xi} \sigma_{V_R}(x,y) \tag{4}$$

$$\tilde{\sigma}_{H_R} = \frac{1}{P} \sum_{(x,y)\in\xi} \sigma_{H_R}(x,y) \tag{5}$$

where, P is the cardinality of the set ξ of pixels in a segment s of the image; A_R is approximation subband from red component of the image. Similarly, mean features are computed for each of the three components of the RGB image. There features are computed using three subbands of Haar wavelet (one level of decomposition), approximation, horizontal and vertical. Diagonal co-officients have not been used because they do not contain much information for classification. The texture feature vector thus obtained is:

$$\mathbf{X}_{\text{texture}} = [\tilde{\mu}_{A_R} \quad \tilde{\sigma}_{V_R} \quad \tilde{\sigma}_{H_R} \quad \tilde{\mu}_{A_G} \quad \tilde{\sigma}_{V_G} \quad \tilde{\sigma}_{H_G} \quad ...]^T \tag{6}$$

A total of 18 texture features are obtained 9 each from cytoplasm and nucleus regions.

3) Shape: Shape features have been used for discriminating different types of blast cells. In this work, we have used eight shape features; area, length of major and minor axis, convex area, filled area, solidity, perimeter and eccentricity to represent cells. These shape features are known to be invariant to translation, rotation, and scaling. We consider such invariance important for obtaining a robust classification. These shape features are concatenated to form the shape feature vector.

$$\mathbf{X}_{\text{shape}} = [x_{\text{area}} \quad x_{\text{lengthofmajor}} \quad x_{\text{lengthofminor}} \quad ...]^T \tag{7}$$

Combining all three types of features, we have obtained 46 total features of which 12, 18 and 16 are color, texture and shape features, respectively.

C. FEATURE SELECTION

A desired algorithm should extract the most relevant features and eliminate the irrelevant and redundant ones. It is important because throwing away irrelevant features reduces the risk of overfitting and decreases computational complexity. The selection of an optimal subset of features can be carried out by using an appropriately designed performance measure to evaluate their ability to classify the samples. It could not be done using a brute force method, if the number of features are huge. In this paper, we have explored a method for Feature Selection based on Mutual Information Distribution (FS-MID) [8] and a popular method Support Vector Machine - Recursive Feature Elimination (SVM-RFE) [9]. Both the methods (FS-MID and SVM-RFE) are similar in performance however earlier is faster compared to RFE. The results section compares the performance of these methods.

D. CLASSIFICATION

In this paper we have used Relevance Vector Machine (RVM) for classification. The RVM has been introduced by Tipping [10] as a Bayesian treatment alternative to SVM. The RVM introduces a prior over the model weights governed by the set of hyperparameter associated with each weight, and the most probable values are iteratively estimated from the training data. The most compelling feature of the RVM is that it typically utilizes significantly lesser number of functions compared to SVM, while providing a similar performance [11]. In this paper we propose to use RVM for classification, results have also been presented using SVM for comparison.

III. RESULTS

In the present work, we have performed experiments on identification of three types of lymphoblasts. The number of different images used are 137, 110 and 100 of L1, L2 and L3 types of lymphoblasts, respectively.

The classification of cells have been performed for all available samples. In all the cases, 50% of cells are used for training and rest 50% for testing. One against rest strategy has been used for classification. In Table II the results obtained by using different types of features and RVM for classification are shown. This comparison shows the importance of different types of features for identification of different types of blasts. It can also be observed that alone shape feature are not affective in classification. Colour and texture features and their combination is as good as using all the three types of features for classification.

Table III shows the classification accuracy using two different features selection methods along with SVM and RVM for classification. It can observed that RVM with any

TABLE II

SENSITIVITY (SN) AND SPECIFICITY (SP) OF RECOGNITION OF DIFFERENT BONE MARROW CELLS USING DIFFERENT TYPES OF FEATURES WITH RVM AS CLASSIFIER

	L1		L2		L3	
Feature set	Sn	Sp	Sn	Sp	Sn	Sp
Shape	52.6	66.7	69.2	75.0	66.7	78.2
Colour	96.9	94.5	93.0	94.3	94.0	95.3
Texture	86.2	88.1	87.0	93.3	91.8	96.0
Shape + Colour	88.1	90.7	87.7	95.7	93.0	92.4
Shape + Texture	78.8	84.3	84.6	91.0	89.1	93.0
Colour + Texture	**95.5**	**95.3**	**92.9**	**94.3**	**94.0**	**95.3**
Shape + Colour + Texture	94.0	94.4	94.1	94.3	93.8	95.3

TABLE III

CLASSIFICATION ACCURACY USING FS-MID (FEATURE SELECTION BASED ON MUTUAL INFORMATION DISTRIBUTION) AND SVM-RFE (SUPPORT VECTOR MACHINE - RECURSIVE FEATURE ELIMINATION) FOR FEATURE SELECTION ALONG WITH RVM AND SVM (NOF: NUMBER OF SELECTED FEATURES).

		L1			L2			L3		
Feature selection	Classification	Sn	Sp	NOF	Sn	Sp	NOF	Sn	Sp	NOF
FS-MID	SVM	92.7	77.4	10	89.8	95.4	41	85.7	95.4	28
FS-MID	RVM	83.1	90.3	10	92.5	95.0	41	91.5	94.5	28
SVM-RFE	SVM	89.4	88.5	20	92.9	95.7	45	85.6	94.6	19
SVM-RFE	RVM	**96.9**	**94.5**	**20**	**94.1**	**94.3**	**45**	**95.7**	**95.3**	**19**

feature selection based algorithm performs superior to SVM. The best results are obtained when RVM is used along with SVM-RFE. It can also be observed that these results are superior to using all the features.

IV. CONCLUSION AND FUTURE WORK

An algorithm for classification of different lymphoblasts is presented in this paper. The results show that proposed features based on color and texture information show good discrimination between different lymphoblasts. The classification is performed using RVM, which gives better discrimination than SVM. Sensitivity and specificity of greater than 94% recorded in all the cases. The algorithm enables quick and easy bone marrow differential count. This information will be useful to categorize the leukemia's into the subclassification, which is important for therapy and prognosis. As a scope of future work, we are working on testing our algorithm on a dataset of over 1000 images. We are also working on algorithm development for the classification of different cells in AML condition.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Sarif Kumar Naik, Dr. Payal Keswarpu and Dr. Puranjoy Bhattacharya for their comments, which helped us in significantly improving the paper.

REFERENCES

[1] Denis R. Miller, Sanford Leikin, Vincent Albo, Harland Sather, and Denman Hammond, "Prognostic importance of morphology (FAB classification) in childhood acute lymphoblastic leukaemia (all)," *British Journal of Haematology*, vol. 48, no. 2, pp. 199–206, 1981.

[2] E Montseny, P Sobrevilla, and S Roman, "A fuzzy approach to white blood cells segmentation in color bone marrow images," in *Proceedings of IEEE International Conference on Fuzzy Systems*, Budapest, Hungary, 2004, pp. 173–178.

[3] Tomasz Markiewicz, Stanislaw Osowski, Bozena Marianska, and Leszek Moszczynski, "Automatic recognition of the blood cells of myelogenous leukemia using svm," in *Proceedings of International Joint Conference on Neural Networks*, Montreal, Canada, July 31 - August 4 2005, pp. 2496–2501.

[4] http://en.wikipedia.org/wiki/Sigmoid function, ," .

[5] W. U. Jianhua, Zeng Pingping, Zhou Yuan, and Oliver Christian, "A novel color image segmentation method and its application to white blood cell image analysis," in *International Conference on Signal Processing*, December 2006.

[6] Kan Jiang, Qing Min Liao, and Sheng Yang Dai, "A novel while blood cell segmentation scheme using scale space filtering and watershed clustering," in *Proceedings of the second international conference on Machine Learning and Cybernetics*, Xian, 2-5 November 2003, pp. 2820–2825.

[7] E. Salari and Z. Ling, "Texture segmentation using hierarchical wavelet decomposition," *Pattern Recognition*, vol. 28, pp. 1819–1824, 1995.

[8] Marco Zaffalon and Marcus Hutter, "Robust feature selection using distributions of mutual information," in *Proceedings of the 18th International Conference on Uncertainty in Artificial Intelligence (UAI-2002)*, A. Darwiche and N. Friedman, Eds., San Francisco, CA., 2002, pp. 577–584, Morgan Kaufmann.

[9] S. K. Majumdar, N. Ghosh, and P. K. Gupta, "Support vector machine for optical diagnisis of cancer," *J. of Biomedical Optics*, vol. 10, no. 2, pp. 024034–1–14, 2005.

[10] M. E. Tipping, "The relevance vector machine," in *Advances in Neural Information Processing Systems*, Sara A Solla, Todd K Leen, and Klaus-Robert Muller, Eds. 2000, Cambridge, Mass: MIT Press.

[11] Begm Demir and Sarp Ertrk, "Hyperspectral image classification using relevance vector machines," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 586–590, October 2007.