Analysis & Classification of Acute Lymphoblastic Leukemia using KNN Algorithm

Pratik M. Gumble
Dept of Electronics & Telecommunication
C.O.E.T.Sipna
Amravati, India
pratikmgumble@gmail.com

Dr.S.V.Rode
Prof., Dept of Electronics & Telecommunication
C.O.E.T.Sipna
Amravati, India
s_rode@rediffmail.com

***Abstract*—** The early Detection of leukemia in cancer patients can greatly increase the chances of recovery. The leukemia can be identified by specific tests such as Cytogenetics and Immunophenotyping and morphological cell classification made by hematologist observing blood & marrow microscope images. This Diagnostic methods are costly and time consuming. We propose the use of morphological analysis of microscopic images of leukemic blood cells for the identification purpose, the morphological analysis just requires an image not a blood sample and hence is suitable for low cost and remote diagnostic system . The proposed system firstly individuates in the blood image the leucocytes from the others blood cells, then it select the lymphocyte cells (the ones interested by acute leukemia), it evaluates morphological indexes from those cells and finally it classifies the presence of the leukemia. 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*-** *Morphological Analysis, Lymphocyte cells, Morphological Indexes, Segmentation*

*****

I. INTRODUCTION

Leukemia is a cancer of white blood cells, this disease develops in the bone marrow spongy tissue. Images of the infected cells blood cells have unique features which can be visually observed by a trained expert using microscopic images of the infected cells. There are four major forms of leukemia; Acute lymphoblastic leukemia (ALL), Acute myelogenous leukemia (AML), Chronic lymphocytic leukemia (CLL) and Chronic myelogenous leukemia (CML) [4]. The early diagnosis of the leukemia type provides the appropriate treatment for that particular type. Using morphological analysis methods for identifying the different leukemia types based mainly on images can greatly reduce the cost of performing type identification tests. Acute Lymphocytic Leukemia (ALL), also known as acute lymphoblastic leukemia is a cancer of the white blood cells, characterized by the overproduction and continuous multiplication of malignant and immature white blood cells (referred to as lymphoblast's or blasts) in the bone marrow. ALL produces a lack of healthy blood cells due to an abnormal number of malignant and immature white blood cells Once the blast-cells invasion starts, blast cells can be detected into the peripheral blood. Principal cells present in the blood are the red blood cells, and the white cells (leucocytes). ALL symptoms are associated only to the lymphocytes [3].

Hence, the observation of the peripheral blood film by expert operators is one of the diagnostic procedures available to evaluate the presence of the acute leukemia. Morphological analysis just requires an image not a blood sample and hence is suitable for low-cost, standard-accurate, and remote diagnostic systems. The system firstly individuates in the blood image the leucocytes from the others blood cells, then it extract the lymphocyte cells (the ones interested by acute leukemia), it extracts morphological indexes from those cells and finally it classifies the presence of the leukemia using classifier. The overall system focuses on segmentation of lymphocyte cell and how to extract a suitable set of morphological indexes from the leucocytes in order to identify the blast cells by a classifier.

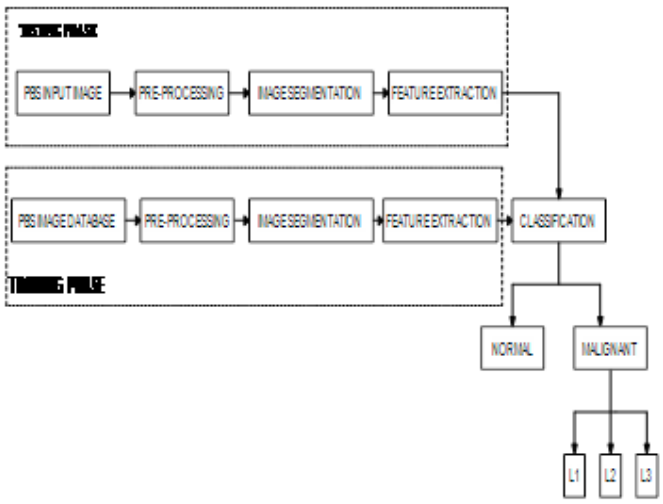

Fig1. Block diagram of proposed work.

PRAPSOED WORK

The proposed systems has in input a color image and it produces in output a list of sub-images containing one-by-one the white cells present in the input image and a estimation of the mean cell diameter. Subsequent modules of the final system will exploits this outputs to detect the presence of acute leukemia. The system is designed to be capable to process images grabbed by a commercial digital camera during normal microscope observation of a blood film. In the image are present the principal components of blood (basophil,

eosinophil), lymphocyte, monocyte, neutrophil. If acute Leukemia is present it shows heterogeneous morphological variations in lymphocytes (area, circularity, compactness, nucleus non-uniformities, etc.). Those features can suitably be measured in the following modules of the final system only if the selection of white cells is accurate and successful.

The main modules which compose the overall system are plotted in Figure 1. The cell Selector module firstly selects the peripheral blood smear image. After selection of image, pre-processing is performed which filter out the noise and adjust the contrast of the image. It has been composed by pre-filtering. Secondly, the morphological operations performed on the input image. White-cells Identifier module selects the white cells present into the image by separating them from others blood's components (red cells and platelets). Typically, the main source of error is related to the strong morphological similarities between the components of the leucocytes family (lymphocytes, monocytes, etc.), conversely is much less probable to classify the other blood components (such as red cells) as lymphocyte than vice versa [17]. In general, the first three modules of the system in Figure 1 can select sub-images of lymphocytes from the blood film image with high accuracy. The sub-system which has to recognize if a lymphocyte is blast or normal, the features of input image are extracted. From the features extracted the input image is classified as normal Lymphocyte or Lymphoblast.

Further Classification of Acute Lymphoblastic Leukemia uses KNN classifier for the classification of ALL subtypes.

The propose system has follows main processing steps:

- To preprocess the image in order to reduce acquisition noise and background non uniformities.
- To perform segmentation for achieving a robust identification of white cells.
- Extraction of heterogeneous morphological variations in lymphocytes features from segmented images
- Classification of Acute Lymphoblastic Leukemia

I. MTERIALS & METHODS

A. Peripheral Blood Smear Data Base

The database is provided by the M. Tettamanti Research Center for Childhood Leukemias and Hematological Diseases, Monza, Italy. The dataset consists of 80 images and it globally contains about 8400 blood cells, 150 of them are lymphocytes labeled by expert oncologists as normal or blast. The Image data base is provided by Fabio Scotti, Department of Information Technologies, University of Milan, Crema, Italy. For each image in the dataset, the classification/position of ALL lymphoblasts is provided by expert oncologists. The Images of the dataset has been captured with an optical laboratory microscope coupled with a Canon Power Shot G5 camera. All images are in JPG format with 24 bit color depth, resolution 2592 x 1944.

B. Preprocessing.

Noise may be accumulated during image acquisition. All the test images are subjected to selective median filtering. Minute edge details of the microscopic images are perfectly preserved even after median filtering. Unsharp masking is performed to sharpen the image details making the segmentation process easier.

C. Image Segmentation.

Recognition of leukemia in blood samples is based on morphological variation of WBC. Such alterations can only be measured with segmented nuclei and cytoplasm. Accuracy of leukemia detection solely depends on leukocyte segmentation; thus, a suitable method has to be employed for morphological region extraction.

Morphological operations are performed on the input image. Whereas erosion and dilation are considered the primary morphological operations and the operations of opening and closing are secondary operations and are implemented using erosion and dilation operations. Mathematical Morphology refers to a branch of nonlinear image processing and analysis that concentrates on the geometric structure within an image, it is mathematical in the sense that the analysis is based on set theory, topology, lattice, random functions, etc.

For the selection of membrane of lymphocyte sobel edge enhancing technique is used. This step enhances the borders of the membranes [26] in order to better perform the subsequent edge detection steps. An edge-enhanced gray level image is hence produced. This processing step is recommended since it helps to better segment grouped cells.

D. Feature Extraction module

Processes image containing a lymphocyte cells and it produces in output a set of morphological indexes. The classification module processes those indexes in order to classify the cell as lymphoblastvblast or normal. The two modules perform the automatic morphological analysis of lymphocyte images. The following morphological, textural features are measured from the binary gray image version of the nucleus and cytoplasm image regions respectively of each lymphocyte image. Area, Total White Cells, Total Black Pixels, Perimeter, Eccentricity, Solidity, Form Factor.

E. Classification

In pattern recognition, classifiers are used to divide the feature space into different classes based on feature similarity. Depending on the number of classes each feature vector is assigned a class label which is a predefined integer value and is based on the classifier output. Each classifier has to be configured such that the application of a set of inputs produces a desired set of outputs. The entire measured data is divided into training and testing data sets. The training data is used for updating the weights and the process of training the network is called learning paradigms. The remaining test data are used for validating the classifier performance. In this study, we propose the use of KNN classifiers for labeling each lymphocyte sub image as normal or malignant sample based upon a set of measured features.

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique. Classification is a decision-theoretic approach to identify the image or parts of the image. Image classification is one important branch of Artificial Intelligence (AI) and most commonly used method to classify the images among the set of predefined categories by using samples of a class. Image classification was categorized into two types; they are

unsupervised and supervised image classification. Since the present work is based on the techniques of supervised classification algorithms rather than unsupervised classification algorithms. Supervised classification is the most fundamental classification in machine vision classification. It requires prior knowledge of image classes. Training samples and test samples are used for classification purpose. The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point [2].

- In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.
- In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.

Both for classification and regression, it can be useful to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/*d*, where *d* is the distance to the neighbor. The neighbors are taken from a set of objects for which the class (for kNN classification) or the object property value (for kNN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. A shortcoming of the kNN algorithm is that it is sensitive to the local structure of the data [2].

Suppose that each training class is represented by a prototype (or mean) vector:

$$m_j = {}^{1}/_{N_j} \sum_{x \epsilon \omega_j} x \ \ for\ j = 1,2, \ldots., M$$

Where N_j is the number of training pattern vectors from class ω_j.

III. ALGORITHM

With the minimum distance classifier we designed the classifier as follows. In the first, the large sequence of the image is used to find the vector for the classifier. Here in the proposed system we use 20 training sequence which are divided in to the group of four i.e. normal, L1, L2 & L3. For the classification of the images K nearest neighbors is used as it is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure [2] (e.g., distance functions). The proposed KNN classification based algorithm can be summarized in the following detail steps.

A. Starting (Training Phase)

The training phase of the algorithm has storing the feature vectors and class labels of the training samples. Four classes of 20 images i.e. normal and abnormal which form the vector 8×20, proceed to step B. If the training sequence is abort, proceed to step E.

B. Process (Transformation Phase)

Noise may be accumulated during image acquisition. All the test images are subjected to selective median filtering. Minute edge details of the microscopic images are perfectly preserved even after median filtering. Unsharp masking is performed to sharpen the image details making the segmentation process easier.

Segment (Segmentation Phase)

Morphological segmentation, segments the processed image to bifurcate the light reflected area and abnormal region [2]. The features are identified from segmented area using histogram based thresholding, proceed to step D.

D. *Classify (Classification Phase)*

In the classification phase, k=2 and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point, proceed to step E.

E. Ending

Classification in two classes normal and abnormal. Depending upon the classification done in Step D, throw the message as "Abnormal" and sub-classified as L1, L2, L3 or "Normal". If the sequence of training is break, stop the process and restart again.

IV. RESULT

As per previous discussion it is well understood that ALL is detected on the basis of the presence or absence of immature lymphocytes or lymphoblasts in PBS samples. Therefore, lymphocytes in PBS samples must be characterized as malignant or normal based on certain fixed pathological criteria defined for the screening of ALL. In this regard, an automated system has been developed, and experiments are conducted using the above configuration and the results are presented in this section.

Following cases are tested for the analysis of Acute Lymphoblastic Leukemia (ALL).

A. Normal case:

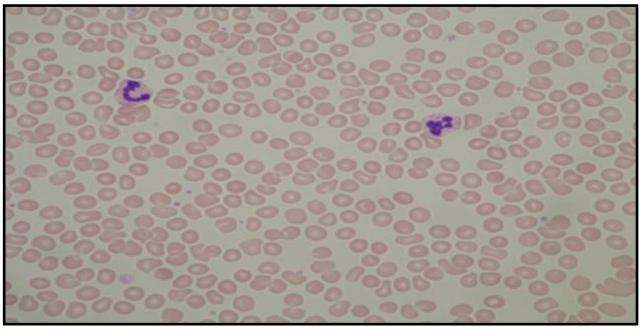

Figure2. PBS image of Normal Patient

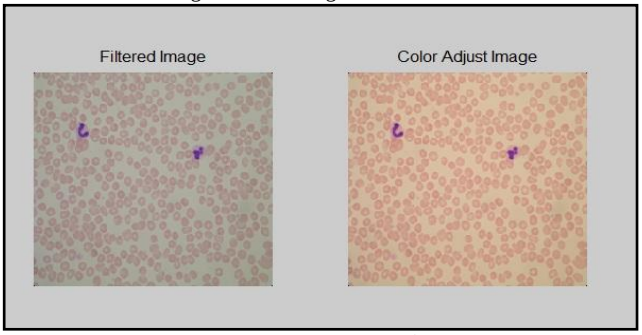

Figure3. Enhanced PBS image of Normal Patient

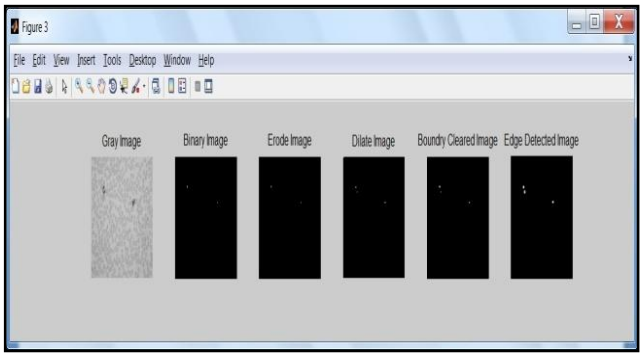

Figure4. Morphological Operation and Segmented image of Nucleus Region.

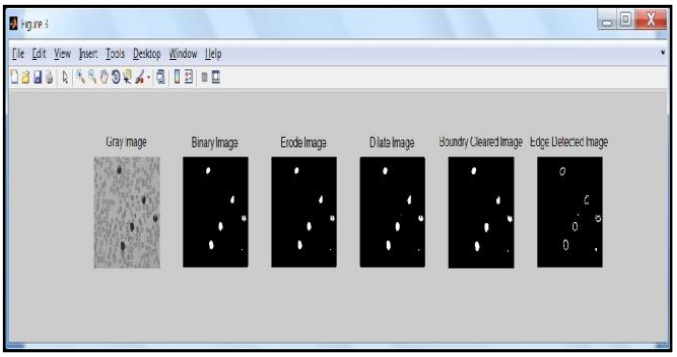

Figure8. Morphological Operation and Segmented Image of Nucleus Region.

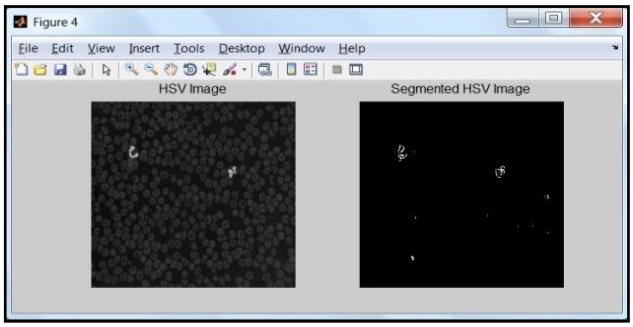

Figure5. Morphological Operation and Segmented image of Cytoplasm Region.

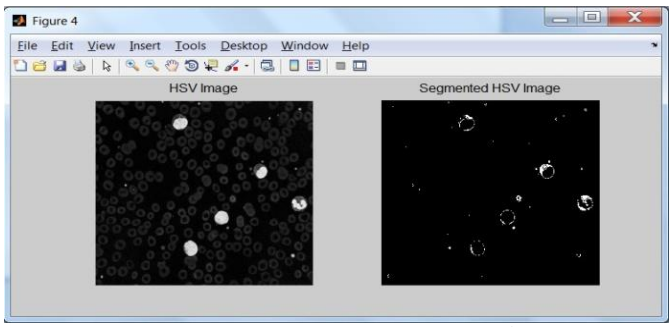

Figure 9.Morphological Operation and Segmented Image of Cytoplasm Region.

B. Abnormal Case:

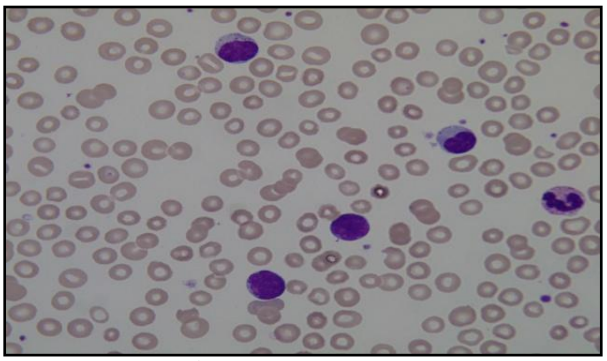

Figure6. PBS image of ALLPatient.

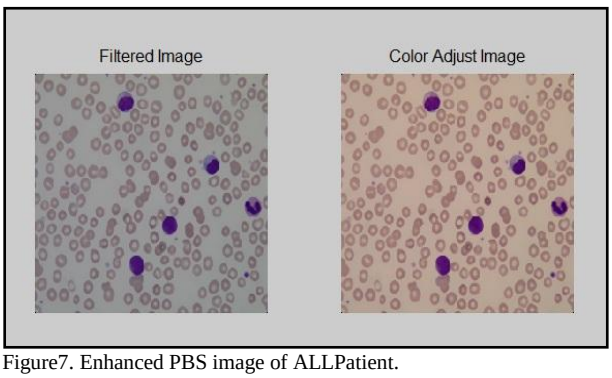

Figure7. Enhanced PBS image of ALLPatient.

Analysis of Morphological features extracted from lymphocyte of normal and malignant lymphocytes are shown in table 1

Sr.No	Features	Normal Lymphocyte	Malignant Lymphocyte
1	White Cells	00003	00007
2	Black Pixels	65501	64343
3	Area	00035	01193
4	Perimeter	012.24	056.48
5	Bounding Box min.	00036	024.50
6	Bounding Box max.	00127	00194
7	Eccentricity	0.9979	0.5910
8	Solidity	0.0338	0.0571
9	Form Factor	2.9344	4.7322

Table1. Morphological features of Normal and malignant lymphocyte

From the above results it is inferred that above system obtains promising results in recognizing lymphoblasts from peripheral blood microscopic images. However, we agree with the fact that much more research is necessary to completely fulfill the real clinical demand. Nevertheless, the results achieved demonstrate the potential of adopting a computer aided approach for assisting hematopathologists in their final decision on suspected ALL patients. Additionally, the proposed system can support initial screening of ALL patients in remote and rural parts of the country.

V. PERFORMANCE ANALYSIS & CONCLUSION

Performance evaluation is mandatory in all automated disease recognition system and is conducted in this study to evaluate the ability of the above system for the detecting the abnormality in the endoscopic images. Accuracy of the system is found out to determine the correctness of the automatic screening of system.

Accuracy factor is defines as ratio of correctly identified results to total observations.

$$Accuracy = \frac{samples\ correctly\ detected}{number\ of\ samples} \times 100$$
$$= \frac{66}{72} \times 100 = 91.66\%$$

Total 72 samples were taken out of 66 found to be correctly identified. The accuracy of the system is 91.66%.

ACKNOWLEDGMENT

We would like to acknowledge the faculties of Electronics & Telecommunication Department, Sipna College of Engineering & Technology, Amravati fir their support. I Pratik M. Gumble specially want to thank my guide Dr.Sandeeep V.Rode for their guidance and constant encouragement towards the work.

REFERENCES

[1] Pratik M. Gumble, Dr.S.V.Rode, "Study & Analysis of Acute Lymphoblastic Leukemia Using Image Procesing",IJIRCCE, Vol- 5, Issue-2 January 2017

[2] Shrikant D.Kale, Dr.S.B.Kasturiwala,"Detection of abnormality in endoscopic images using endoscopic technique", IJRITCC,Vol - 4 , Issue – 1 January 2017.

[3] "Automatic morphological Analysis for Acute leukemia Identification in Peripheral blood Microscope Images" CIMSA 2005 – IEEE International Conference on Computational Intelligence for Measurement Systems and Applications Giardini Naxos, Italy, 20-22 July 2005

[4] Ruggero Donida Labati IEEE Member, Vincenzo Piuri IEEE Fellow, Fabio Scotti IEEE Member Universit`a degli Studi di Milano, "All-IDB: the Acute Lymphoblastic Leukemia Image Database for image processing" Department of Information Technologies, via Bramante 65, 26013 Crema, Italy

[5] Dr.S.V.Rode, "Modified Approach for Feature Space Improvement of Imagery" International Journal of Innovative Research in Electrical, Electronics, Instrumentation & Control Engineering Vol. 3 Issue 10, October 2015. Page 48-50.

[6] "Image Segmentation of Blood Cells in Leukemia Patients" adnan khashman 1, esam al-zgoul Intelligent Systems Research Group (ISRG) Department of Electrical & Electronic Engineering Near East University Near East Boulevard, Nicosia N.CYPRUS.

[7] Hugo Jair Escalantea,b, Manuel Montes-y-Gómeza,c, Jesús A. Gonzáleza, Pilar Gómez-Gila, Leopoldo Altamiranoa, Carlos A. Reyesa, Carolina Retaa, Alejandro Rosalesa,Recent Advances in Computer Engineering and Applications. "Acute leukemia classification by ensemble particle swarm model selection.",Journal of Artificial Intelligence in medicine by Elsevier

[8] Monica Madhukar, Sos Agaian, Anthony T.Chronopoulos "Deterministic Model for Acute Myelogenous Leukemia Classification" Department of Electrical and Computer Engineering,University Of Texas at San Antonio 2012 IEEE International Conference on Systems, Man and Cybernetics. October 14-17, 2012, COEX, Seoul, Korea

[9] G. J. Meschino and E. Moler. Semiautomated image segmentation of bone marrow biopsies by texture features and mathematical morphology. Analytical and Quantitative Cytology and Histology, 26(1):31–38, 2004

[10] Q. Liao and Y. Deng. An accurate segmentation method for white blood cell images. In Proceedings of the IEEE International Symposium on Biomedical Imaging, pages 245 –248, 2002.

[11] J. Angulo and G. Flandrin. Microscopic image analysis using mathematical morphology: Application to haematological cytology. In A. Mendez-Vilas, editor, Science, Technology and Education of Microscopy: An overview, pages 304–312. FORMATEX, Badajoz, Spain, 2003.

[12] N. T. Umpon. Patch based white blood cell nucleus segmentation using fuzzy clustering. ECT Transaction Electrical Electronics Communications, 3(1):5–10, 2005.

[13] L.B. Dorini, R. Minetto, and N.J. Leite. White blood cell segmentation using morphological operators and scale-space analysis. In Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing, pages 294–304, October 2007.

[14] C. Meurie, O. Lezoray, C. Charrier, and E. Elmoataz. Combination of multiple pixel classifiers for microscopic image segmentation. IASTED International Journal of Robotics and Automation, 20(2):63 – 69, 2005. Special Issue on Colour Image Processing and Analysis for MachineVision.

[15] M. Ghosh,D.Das, C. Chakraborty, and A. K. Ray. Automated leukocyte recognition using fuzz divergence. Micron, 41(7):840–846, 2010.

[16] F. Scotti. Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images. In Proceedings of IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, pages 96 – 101, July 2005.

[17] T. Markiewicz,S. Osowski, B. Marianska and L. Moszczynski. Automatic recognition of the blood cells of myelogenous leukemia using svm. In Proceedings of IEEE International Joint Conference on Neural Networks, volume 4, pages 2496 –2501, August 2005

[18] L. Gupta, S. Jayavanth, and A. Ramaiah. Identification of different types of lymphoblasts in acute lymphoblastic leukemia using relevance vector machines. In Proceedings of the 31st Annua International Conference of the IEEE Engineering in Medicine and Biology Society, USA, 2009.

[19] P. Bamford and B. Lovell, "Method for accurate unsupervised cell nucleus segmentation," in Proc.of the Engineering in Medicine and Biology Society Conference, 2001.

[20] M. Oberholzer, M. Ostreicher, H. Christen, and M. Bruhlmann, "Methods in quantitative image analysis," Histochem Cell Biol., vol. 105, no. 5, pp. 333–355, May 1996.

[21] N.Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[22] M. Nixon and A.S. Aguado, Feature extraction and image processing, Elsevier, 2005.

[23] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," Pattern Recognition, vol. 13, no. 2, pp. 111–122, 1981.

[24] R. C. Gonzalez, R. E. Woods, and S.L. Eddins, Digital Image Processing Using MATLAB, Pearson Prentice Hall Pearson Education, Inc., New Jersey, USA, 2004.

[25] http://en.wikipedia.org/wiki/Lab_color_space

[26] http://dba.med.sc.edu/price/irf/Adobe_tg/models/cielab.htm