International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 4 (2012) pp. 477-485

Unsupervised Blood Microscopic Image Segmentation and Leukemia Detection using Color based Clustering

Subrajeet Mohapatra[1], Dipti Patra[1] and Sanghamitra Satpathy[2]

[1]Department of Electrical Engineering,
National Institute of Technology Rourkela,
Rourkela- 769008, Odisha, India.
subrajeets@gmail.com, dpatra@nitrkl.ac.in

[2]Department of Pathology,
Ispat General Hospital Rourkela,
Rourkela- 769005, Odisha, India.
sanghamitra.satpathi@gmail.com

Abstract: **In real clinical trial there is no substitute for final assessment of any disease independent of diagnostic test. Diagnostic hematology is a speciality that deals with the understanding of the essential pathological processes of the blood through appropriate morphological or biochemical analysis. Microscopic analysis of peripheral blood and bone marrow by a hematologist are subjected to various shortcomings like inter observer variations, slowness, operator experience and tiredness. Whereas biochemical tests, immunophenotyping, molecular probing etc are expensive for routine examination. Thus microscopic image analysis serves as an impressive automated diagnostic tool for hematological disorders i.e. leukemia, malaria, psoriasis, AIDS etc. Acute Lymphoblastic Leukemia (ALL) is a serious hematological disorder of blood which needs to be diagnosed early for faster cure. This paper introduces a comparative approach to Acute Lymphoblastic Leukemia (ALL) detection based on WBC nucleus image segmentation and morphological analysis. Color based clustering is employed for segregating various blood components and obtaining the nucleus of the white blood cell. Further fractal geometry, contour signature and texture based techniques are employed for nucleus feature extraction which leads to automatic leukemia detection using a Support Vector Machine (SVM) classifier. The proposed approach is validated with the collected blood microscopic images and satisfactory results have been obtained.**

Keywords: Data Clustering, Acute Leukemia Detection, Hausdorff Distance, Quantitative Microscopy, Blood Image Analysis

I. Introduction

Ailments which are marked by abnormal functioning of blood cells or blood forming tissues are termed as hematological disorders [1]. Various disease conditions of the human body are reflected in blood and its components. Among all, abnormalities of erythrocytes, leukocytes and platelets are considered crucial and needs immediate medication. The present work basically deals with diseases of white blood cells (WBC) or leukocytes. Alterations in WBC can be neoplastic or non-neoplastic, one of the neoplastic disorder i.e. leukemia is considered as our subject of study. Leukemias are the neoplastic proliferations of hemopoietic cells. Specific genetical changes are responsible for malignant transformation of cells and their progeny forming a clone of leukemia cells. Leukemia can be understood as a hematological malignancy with increased numbers of myeloid or lymphoid blasts. Leukemia can be acute or chronic depending on the severity of the disease. Practical classification of leukemia is quite complicated and can be categorized on the basis of morphologic findings, genetic abnormalities, putative etiology, cell of origin, immunophenotypic qualities, and clinical characteristics. French, American, British (FAB) classification and World Health Organization (WHO) classification are two widely used protocols for leukemia categorization [2]. But both fundamentally divide leukemia's into myeloid and lymphoid types, depending on the origin of the blast cell. In the present paper we have only considered acute lymphoblastic leukemia (ALL) for our research.

ALL is the most common malignancy diagnosed in children representing nearly one third of all pediatric cancers [3]. This disease also affects adults, especially those above the age of 65 and older. The annual incidence of acute lymphoblastic leukemia is approximately 9-10 cases per one lakh population in childhood [4]. Leukemia diagnosis serves as a pillar to all therapies, thus diagnostic tests are of utmost importance and need to be executed with precision. Single tests i.e. morphological, cytochemical, immunophenotyping, cytogenetic and molecular genetic analysis or a combination of two tests is performed on the lymphocytes for confirmation and classification of leukemia. In spite of advanced techniques i.e. fluorescence in situ hybridization (FISH) and others morphology still remains the first choice of the hematologists for initial screening of ALL. Microscopic examination

of blood smear is a low cost, preferred and reliable evaluation technique which is essential for all suspected patients. Other methods are employed for selective patients depending on the phenotype results. Since aspiration biopsy is sometimes traumatic so it is avoided for initial screening purposes. Thus in the present work automation of microscopic examination of blood smear is only considered for initial screening. The orthodox way of evaluation of leukocytes using a microscope is still prevalent across the globe. Human evaluation is based on visual examination of the blood film based on their clinico-pathological understanding and expertise [5]. Such techniques are prone to perverted results because of inter and intra observer variations and are also subjected to factors like slowness, operator tiredness etc resulting erroneous interpretation. An automated leukemia detection technique based on image analysis is presented to assist the hematologist with accurate and objective results for disease mapping.

Over years many researchers have focused on the development of automated techniques that can model the human vision and analysis task into machine vision and recognition task. The main thrust on all such techniques is to provide automatic image segmentation followed by feature extraction and classification. These schemes mainly differ in their basic methodologies applied to WBC nucleus separation from background. The segmentation techniques employed in existing literature can be broadly categorized as region based or edge based schemes [6]. A two step segmentation process using HSV color model is used in [7]. Cell segmentation using active contour models is presented in [8]. Color segmentation procedure applied to leukocyte images using mean-shift is described in [9]. Clustering techniques has also been incorporated delicately in certain image segmentation techniques [10, 11]. Active contour models (snakes) and its variations has been explored successfully for WBC nucleus segmentation [12, 13]. Morphological operators are also used along with few other methods and also independently for WBC background separation [14]. A performance comparison between nine image segmentation techniques is presented in [15], which concludes no single method can be considered as bench mark for blood image segmentation. As a general purpose segmentation method, feature space clustering has the advantage that is straight forward for classification [16]. Drawbacks associated with clustering algorithms is to predetermine the number of clusters [6]. As per [17] selection of color space is quite vital in color based clustering. Segmentation of blood microscopic images serves as an essential platform for disease diagnosis hence is an crucial step in the entire automation process [18, 19]. Thereupon it was concluded from existing literature study that the automated leukemia detection process depends upon proper WBC nucleus segmentation, feature extraction and classification.

The aim of the paper is to present WBC nucleus segmentation using color based clustering followed by feature characterization and assessment of lymphoblasts. We characterize each lymphocyte as a mature lymphocyte or a lymphoblast based on WBC nucleus features i.e. hausdorff dimension, contour signature and other existing standard features like geometrical, textural and statistical. This approach will stimulate better, faster and accurate diagnostic results. The main focus of the paper is to study, simulate and analyze the results of various clustering algorithms for color based image segmentation. The segmented nucleus was used for feature extraction and the features were used as input for SVM based classification of lymphoblasts. The rest of the paper is organized as follows. Clustering techniques for image data are presented in section II. Section III describes the complete framework of the proposed approach. The experimental results are presented in Section IV. Finally concluding remarks are considered in Section V.

II. Data Clusterng

Clustering is an unsupervised classification of data patterns into homogeneous groups or clusters. Clustering is a difficult problem which has been addressed by various researchers in diversified areas such as pattern recognition, data mining, image processing, biology, psychology, and marketing etc. This section provides an overview of widely used clustering techniques from an image segmentation perspective. Popular clustering algorithms such as K-means, K-medoid, Fuzzy C Means, Gustavson Kessel and Fuzzy Possibilstic C means are discussed under two heads such as hard clustering and fuzzy clustering.

K-means is a center-based clustering algorithm which is efficiently employed for clustering large databases and high-dimensional databases. The objective of a center-based algorithm is to minimize its objective function and is well suited for convex shape clusters and fails drastically for clusters of arbitrary shapes [20]. The conventional K-means algorithm was first proposed by MacQueen (1967). This technique clusters the data into fixed number of clusters and the mean of one cluster is placed as far away as possible from another. Every data point is associated to the nearest mean and belongs to one of the clusters [21]. Numerous variations of similar theme are available in the literature which is usually based on changing the dissimilarity or centering. K-medoid is a similar clustering technique like K-means which tries to minimize a squared error criterion but the cluster center is chosen from the set of data points rather than mean. The element whose average dissimilarity to all the objects in the cluster is minimal is selected as medoid of that cluster [22]. It is immune to noise and outliers hence more suitable than K-means. Under fuzzy clustering we have Fuzzy C Means (FCM) which was developed in 1973 by Dunn and improved by Bezdek in 1981. In FCM each data point is associated with every cluster using a membership function, which gives degree of belongingness to the clusters. Gustafson Kessel (GK) is a variation of FCM algorithm which associates each cluster with a cluster centre and with a covariance matrix. Original FCM implicitly considers each clustering data as spherical, while GK technique is not subjected to such assumptions and can also deal with non-spherical geometry of data. Another important clustering technique is Possibilistic C means (PCM). Problems are also associated with PCM thus Fuzzy Possibilistic C means (FPCM) [23] is considered here for improved image data clustering. In the present work clustering algorithms like K-means, K-medoid, FCM, GK and FPCM are employed for blood microscopic image segmentation and comparative analysis is presented.

III. Materials and Methods

Automatic leukocyte classification from blood microscopic images consists of preprocessing, segmentation, feature extraction and classification. The overall working principle is depicted in Figure 1. The blood smear image consists of red blood cells (RBC), white blood cells (WBC) and platelets. The proposed method is based on color image segmentation and our objective is to separate WBC from the background and finally separate nucleus and cytoplasm. As per acute leukemia is concerned the cytoplasm is scanty so we have considered only the nucleus as the region of interest and its essential features are extracted.

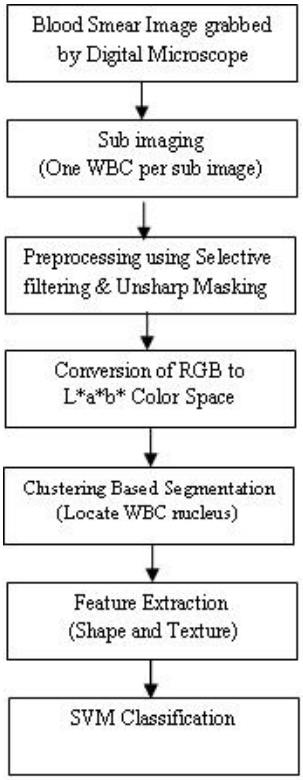

Figure. 1: System Overview

A. Blood Smear Preparation

Blood samples were collected at Ispat General Hospital, Rourkela, India through randomization. Subsequently blood smear is prepared and stained using Leishman for visualization of cell components. The images were captured with a digital microscope (Carl Zeiss India) under 100X oil immersed setting and with an effective magnification of 1000. Few images with permission from University of Virginia were also considered for experimental purposes. The data set is a mixture of lymphocytes and lymphoblasts. There are 100 images collected from Ispat General Hospital, Rourkela, India and 8 images are collected from University of Virginia. Manual segmentation was performed by Dr. Sanghamitra Satpathy, Hematologist, Department of Pathology, Ispat General Hospital, Rourkela, India. Each hand segmented image consists of nucleus, cytoplasm and back ground.

B. Sub Imaging

As peripheral blood smear images are relatively larger and our region of interest (ROI) are single cells only so the sub images are required to be cropped each containing a single cell. For accurate leukemia detection each WBC nucleus feature has to be extracted individually for classifying it as a blast cell. Sub images containing single nucleus per image are obtained using bounding box [24] technique. Conflict arises in distinguishing a WBC and a RBC, which can be solved using clustering and image morphology [25]. We use simple K-means color based clustering to obtain all the blue WBC nucleus of the entire image. Using image morphology we obtain the centroid of each nucleus and a square image is cropped around each nucleus such that entire cell will be within the cropped sub image as shown in Figure 2. Again remapping with the original image we can restore the color components and color sub images are obtained and is shown in Figure 3. Sub images containing lymphocytes only were obtained and can be used for further processing. The images of neutrophils, eosinophils, basophils are not considered for feature extraction as they are not associated with lymphocytic leukemia.

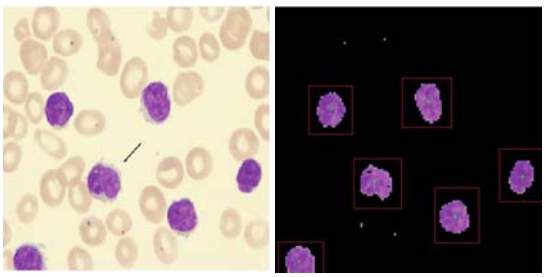

Figure. 2: Initial K-Means Segmentation

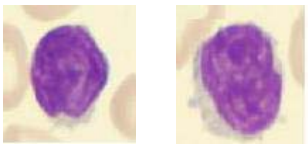

Figure. 3: Cropped Sub Images

C. Preprocessing

Noise may be accumulated during image acquisition and due to excessive staining. All the test images are subjected to selective median filtering followed by unsharp masking [26]. Incorporation of adaptive threshold into the noise detection process led to more reliable and more efficient detection of noise. Minute edge details of the microscopic images are perfectly preserved due to selective median filtering.

D. Color Conversion

Typically an image can be represented with the help of three color components. Images generated by the digital microscopes are usually in RGB color space which is visually difficult to segment. For better color based segmentation we map the RGB image into L*a* b* color space. This color space

consists of a luminosity layer L*, chromaticity layers a* and b*. Since all the color information is in the a* and b* layers we use these two components for nucleus segmentation.

E. Color based Clustering

Clustering technique is used to create K clusters from n observations. It attempts to achieve partition such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible [27]. Here each pixel of an object has two values (a* and b*). Depending on these two values we classify each pixel into four clusters. Here K is considered as 4 because we want to segment the entire image into four regions i.e. nucleus, cytoplasm, RBC and background stain. Since we want to extract nucleus features we have considered the cluster which contains blue nucleus. This cluster is used to obtain the sub images containing one nucleus per image. All the clustering algorithms as described in Section II were used separetly for segmenting the blood images.

F. Feature Extraction

Feature extraction in image processing is a technique of redefining a large set of redundant data, into a set of features of reduced dimension [28]. In the present paper broadly three types of features are extracted i.e. fractal dimension, shape features including contour signature and texture. In addition color features are also extracted from the nucleus image.

1) Fractal Dimension

Fractals have been used in medicine and science earlier for various quantitative measurement [29, 30]. Perimeter roughness of nucleus is an important measure that decides whether a particular nucleus represents a lymphoblast or a mature lymphocyte. Fractal geometry is a more convenient way to parameterize the cell boundary surface in comparison to euclidean geometry. Hausdorff dimension is an essential feature for fractal geometry and will be an essential quantitative measure for cell boundary roughness measurement. The procedure for Hausdorff Dimension measurement using box counting method [31] is introduced below as an algorithm:

1. Each nucleus color (RGB) image is converted to gray and successively to binary image.
2. Nucleus edge boundary is extracted using Canny [25] edge detection technique.
3. A grid of N squares is superimposed over the edges, while counting the edge occupied squares.
4. Step 3 is continued for an increasing number of squares.
5. The Hausdorff Dimension HD may then be defined as in(1).

$$HD = \frac{\log(N)}{\log(N(s))} \tag{1}$$

where, N is the number of squares in the superimposed grid and $N(s)$ is the number of occupied squares or boxes (box count). Higher HD signifies higher degree of roughness.

2) Contour Signature

Ill-defined or rough boundary is a significant feature for labeling a WBC nucleus as a blast cell. Along with the fractals contour signature method is also followed to measure the irregularity quantitatively. The nucleus boundary can be represented by a contour of dimension two. A better way of irregularity measurement of the contour is converting from coordinate based representation to distances from each contour point or edge pixels to a reference point. Since most of the nuclei have irregular shapes, centroid or centre of mass can be taken as a a convenient reference. Euclidean distance measurement from the centroid to the contour points is described as follows:

1. Nucleus boundary pixel indices are obtained from the edge image obtained during HD measurement.
2. Centroid coordinates of the nucleus region is calculated using the relation as defined in Equation(2) and (3).
3. Euclidean distance is calculated from each boundary pixel to the centroid.
4. To measure the irregularity of the nucleus boundary, variance (σ^2) of all the distances from the centroid obtained in step 3 is calculated.

$$\bar{x} = \frac{1}{M}\sum_{n=0}^{M-1} x(\mathrm{n}) \tag{2}$$

$$\bar{y} = \frac{1}{M}\sum_{n=0}^{M-1} y(\mathrm{n}) \tag{3}$$

where (x,y) are the coordinates of the pixels along the contour and M is the total no of pixels on the contour.

3) Shape Features

According to hematologist the shape of the nucleus is an essential feature for discrimination of blasts. Region and boundary based shape features are extracted for shape analysis of the nucleus. All the features are extracted from the binary equivalent image of the nucleus with none zero pixels representing the nucleus region. The quantitative evaluation of each nucleus is done using the extracted features under two classes i.e. region based and boundary based. The features are as follows:

- *Area*: The area was determined by counting the total number of none zero pixels within the image region.
- *Perimeter*: It was measured by calculating distance between succesive boundary pixels.
- *Compactness* : Compactness or roundedness is the measure of a nucleus as defined in (4).

$$Compactness = \frac{\mathrm{Perimeter}^2}{Area} \tag{4}$$

- *Solidity* : The ratio of actual area and convex hull area is known as solidity and is also an essential feature for blast cell classification. This measure is defined in (5).

$$Solidity = \frac{\mathrm{Area}}{ConvexArea} \tag{5}$$

- *Eccentricity* : This parameter is used to measure how much a shape of a nucleus deviates from being circular. It's an important feature since lymphocytes are more circular than the blast. To measure this a relation is defined in (6).

$$Eccentricity = \frac{\sqrt{\mathrm{a}^2 - b^2}}{a} \tag{6}$$

where a is the major axis and b is the minor axis of the equivalent ellipse representing the nucleus region.

- *Elongation* : Abnormal bulging of the nucleus is also an symbol which signifies leukemia. Hence the nucleus bulging is measured in terms of a ratio called elongation. This is defined as the ratio between maximum distance (R_{max}) and minimum distance (R_{min}) from the center of gravity to the nucleus boundary and is given by (7).

$$Elongation = \frac{\mathrm{R_{max}}}{R_{min}} \tag{7}$$

where R_{max} and R_{min} are maximum and minum radii respectively.

- *Formfactor* : It is a dimensionless parameter which changes with surface irregularities and can be defined as (8).

$$Formfactor = \frac{4 \times pi \times Area}{\mathrm{Perimeter}^2} \tag{8}$$

G. Color Feature Extraction

Color is an important feature that hematologist perceive while examining blood slides under a microscope. Thus mean color values in RGB and HSV color spaces were obtained for nucleus region of each lymphocyte. Healthiness of a lymphocyte can be measured upto a certain extent through this color information.

H. Texture Features

Nucleus texture measurements are performed on gray scale version of the nucleus images. Gray level statistics i.e. mean gray level (μ_{gray}), standard deviation (σ_{gray}) and coefficient of variance (CV) of nuclear gray levels are measured. The gradient strength which signifies local rate of change in gray levels is obtained using statistical parameters i.e. mean gradient (μ_{grad}), standard deviation of the gradient (σ_{grad}) and coefficient of variance of the gradient of nuclear gray levels. Coefficient of variance (CV) is defined in (9).

$$CV = \frac{\sigma}{\mu} \tag{9}$$

I. Densitometric Features

These features were computed from the co-occurrence matrices for each nucleus image. This includes

- *Homogeneity* : It is a measure of degree of variation.
- *Energy*: Is used to measure uniformity.
- *Correlation*: This represents correlation between pixel values and its neighborhood.
- *Entropy* : Usually used to measure the randomness.

J. Classification

Classification is the task of assigning a label from one of the known classes to the unknown test vector. Since the patterns are very close in the feature space, support vector machine (SVM) is employed for classification. SVM is a powerful tool for data classification based on hyper plane classifier [32]. This classification is achieved by a separating surface (linear or non linear) in the input space of the data set. They are basically two class classifiers that optimize the margin between the classes [33]. The classifier training algorithm is a procedure to find the support vectors. Relevant extracted features as described in Section III-F are used as input to the SVM.

IV. Experimental Results

The clustering based segmentation approaches as discussed in Section III and the other subsequent techniques has been applied on 108 peripheral blood smear images to test the efficiency for leukemia detection. Visual segmentation results are presented in experiment IV-A to compare the various clustering based segmentation techniques. In experiment IV-B we use two clustering validity index i.e. Global Silhouette Index (SL) and Partition Index (SC), and computation time as quality measure to compare the proposed segmentation techniques. Section IV-C presents the feature extraction and classification results.

A. Experiment 1

Leukocyte image sample of size (128×128) as shown in Figure 4 are mapped from RGB color space to $L^*a^*b^*$ color space. The color information in the $L^*a^*b^*$ color space is represented using two components (a^* and b^*) only. This property of reduction in number of color features from three to two can be utilized in accelerating color based clustering process. Thus a^* and b^* component for every pixel is recorded and feature data set X of size 16384×2 is prepared. Each row of X represents a data pattern and redundancy among them was discarded. This concise form of X with size $N \times 2$ serves as an input towards pixel labeling problem through color based clustering. After successful clustering, background including RBC are clustered into single class whereas cytoplasm is considered in another class. However the entire nucleus is represented in two different clusters due to inconsistency in absorption of the staining material. Various standard clustering schemes as explained in Section II are simulated for obtaining the corresponding individual clusters. Segmented results obtained from different clustering schemes are presented in Figure 5 for a particular leukocyte image sample (Figure 4). Each column represents a particular cluster and each row of the image indicates a particular clustering scheme. As we have four clusters so the image indicates four cluster outputs for each clustering scheme. As we have four clusters so the image indicates four cluster outputs for each clustering scheme. Using cluster validity techniques the optimum number of clusters was found to be four and hence the four cluster outputs of each technique is presented here. The third and fourth cluster images were added to obtain the desired nucleus image as shown in Figure 6. This image containing a single nucleus is used

for various feature extraction followed by leukemia classification. In the present work nucleus is only considered for leukemia feature extraction and classification. However, the cytoplasm which is obtained in the first cluster of each clustering scheme can also be used as an indicator of leukemic condition.

Figure. 4: Cell 1

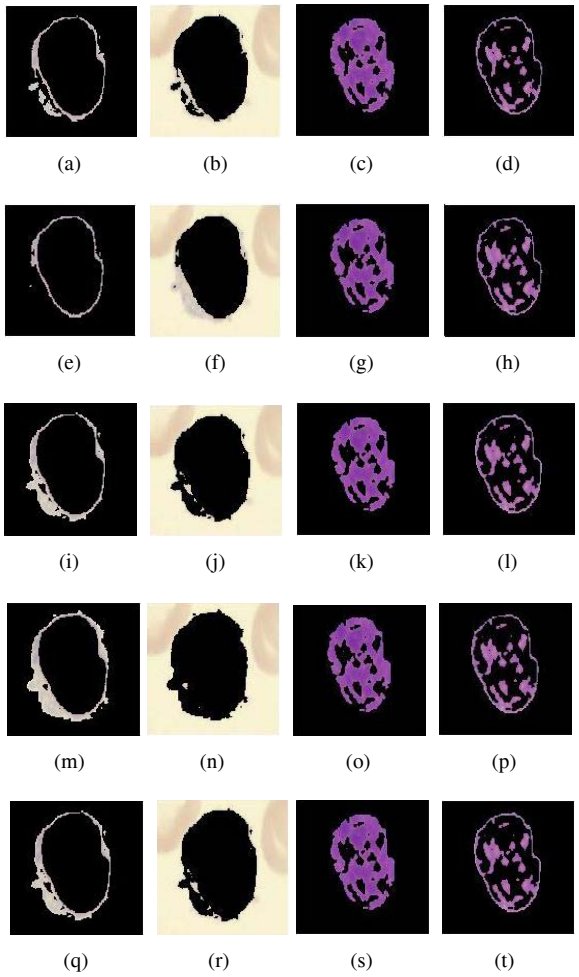

Figure. 5: Clustering results obtained by different clustering techniques. K-Means (a-d), K-Medoid (e-h), FCM (i-l), GK (m-p), FPCM (q-t)

Figure. 6: Desired Nucleus after adding cluster 3 and 4

B. Experiment 2

For quantitative evaluation of the segmentation techniques the same image used in Section IV-A was considered for performance evaluation. The input L*a*b* image is segmented using all the clustering techniques separately as discussed in section II. Performance time, Global Silhouette Index (SL) and Partition Index (SC) for each color clustering technique was measured for optimum number of clusters as four and is tabulated in Table 1. Due to limited space clustering results in terms of feature space separation for K-Medoid clustering for the experimental image data is shown in Figure 8.

Figure. 7: Separeted Nucleus Sub Images using Bounding box technique

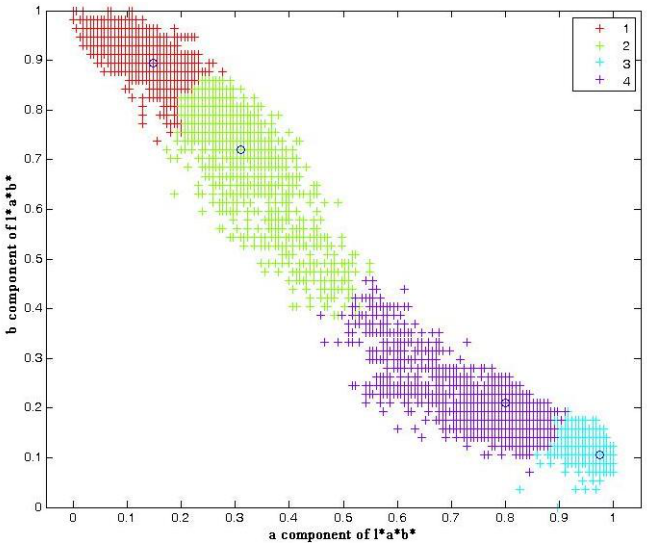

Figure. 8: K- Medoid Feature Space Clustering Results

C. Experiment 3

After the nucleus image is obtained various features are extracted. Feature extraction was done using methods as presented in Section III-F over each nucleus. Initially we applied fractal geometry i.e. Hausdorff Dimension for calculating the perimeter roughness of each nucleus using the procedure explained in Section III-F.1. Figure IV-C shows the nucleus and its boundary whose roughness is measured using Hausdorff Dimension (HD). Table 2 and table 3 depicts hausdorff measurements for few lymphocytes and lymphoblast samples. Similar analysis was conducted on 50 samples (lymphocytes and lymphoblasts) each and it was found that none of the lymphocytes had a HD larger than 1.22 and a large percentage of lymphoblasts had HD more than this value. So this can be a suitable feature for leukemia detection.

Table 1: Clustering Performance with different techniques

	Clustering Method	*Time Taken*	*SL*	*SC*
Hard Computing	K-means	3.2344	0.1656	0.0308
	K-medoid	10.8281	0.15232	0.2741
Fuzzy Computing	FCM	**2.2344**	0.15723	0.2219
	GK	3.6563	0.1848	0.2298
	FPCM	4.7502	**0.1970**	**0.0114**

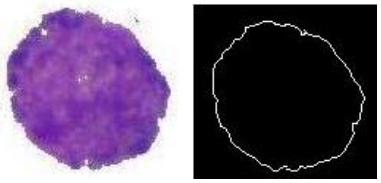

Figure. 9: Nucleus Boundary Detection

As contour signature is also an important such measure, hence this feature is extracted using the method as described in Section III-F.2. The centroid of the nucleus contour is determined using Equation(2) and (3). Euclidean distance between the centroid and boundary pixels is depicted in Figure 10. Variance of the distances of lymphocytes and lymphoblasts is measured as shown in Table 4 and Table 5 respectively. This feature acts as a hard limiter to classify a nucleus as healthy or leukemic. A threshold for variance feature is fixed as 0.02 for the given set of WBC samples. Similarly the shape features were also measured using the relations given in Section III-F.3. Few shape feature comparison between a mature lymphocyte and a lymphoblast is tabulated in Table 6. Color and texture features are also extracted for the image nucleus sample and recorded. Few densitometry measurements are tabulated in Table 7. All the extracted features are used to feed the SVM classifier for leukemia detection. Based on the above features a particular lymphocyte can be classified as healthy or lymphoblastic(leukemic). Classification accuracy of 92% was observed for the given set of images collected from IGH Rourkela, Odisha.

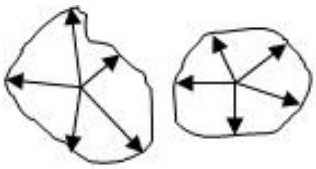

Figure. 10: Contour Signature

Table 2: Hausdorff Dimension measurement for lymphocytes

Sample	Hausdorff Dimension
1.	1.0367
2.	1.0421
3.	1.0546
4.	1.0785
5.	1.0837

Table 3: HD Measurement for Lymphoblasts

Sample	Hausdorff Dimension
1.	1.2018
2.	1.2386
3.	1.2345
4.	1.2540
5.	1.2601

Table 4: Contour Signature for lymphocytes

Sample	Variance
1.	0.0024
2.	0.0083
3.	0.0160
4.	0.0093
5.	0.0179

Table 5: Contour Signature for lymphoblasts

Sample	Variance
1.	0.0382
2.	0.0244
3.	0.0628
4.	0.0639
5.	0.0344

Table 6: Results of Various Shape Measurements

Measure	Lymphocyte	Lymphoblast
Area	2208	2715
Perimeter	186.1	208.3
Compactness	13.7	16
Solidity	1.0	1.0
Eccentricity	0.5	0.4
Elongation	1.4113	1.4181
Formfactor	0.2549	0.2504

Table 7: Results of Various Texture Measurements

Measure	Lymphocyte	Lymphoblast
Homogeneity	0.877	0.879
Energy	0.277	0.292
Correlation	0.832	0.779
Entropy	2.787	3.305

V. Conclusion and Future Work

Color based clustering for WBC nucleus segmentation of stained blood smear images followed by relevant feature extraction for leukemia detection is the main theme of the paper. Few standard clustering techniques, viz., K-Means, K-Medoid, Fuzzy C-Means (FCM), Gustafson Kessel (GK) and

Fuzzy Possibilistic C Means (FPCM) were employed for color based segmentation and their performance were compared. The paper also presents two novel procedures for measuring nucleus boundary irregularities using hausdroff dimension and contour signature. Satisfactory results were obtained for detecting leukemia with the proposed features using SVM classifier. Furthermore the system should be robust to excessive staining and touching cells. Results obtained encourage future works like stain independent blood smear image segmentation and leukemia type classification.

References

[1] *Hematological Disorders, McGraw-Hill Dictionary of Scientific and Technical Terms*. McGraw-Hill Companies, Inc, 2003.

[2] C. Douglas Tkachuk and Jan V. Hirschmann, editors. *Wintrobe's Atlas of Clinical Hematology*. Lippincott Williams and Wilkins, 1st edition, 2007.

[3] Noriko Satake and Janel M Yoon. Acute lymphoblastic leukemia,. 2010. "http://emedicine.medscape.com".

[4] JM Ribera and A. Oriol. Acute lymphoblastic leukemia in adolescents and young adults. *Hematology Oncology Clinics of North America, Elseiver*, 23(5), 2009.

[5] Madhumala Ghosh, Devkumar Das, Chandan Chakraborty, and Ajoy Kumar Ray. Automated leukocyte recognition using fuzzy divergence. *Micron*, 41(7):840–846, 2010.

[6] Kan Jiang, Qing-Min, and Sheng-Yang Dai. Red blood cell segmentation scheme utilizing various image segmentation techniques. In *2nd International Conference on Machine Learning and Cybernetics*, Xi'an, 2003.

[7] Neelam Sinha and A.G. Ramakrishnan. Automation of differential blood count. In *Conference on Convergent Technologies for Asia-Pacific Region*, pages 547–551, 2003.

[8] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksak. An automated differential blood count system. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2583–2586, 2001.

[9] D. Comaniciu and P. Meer. *Cell Image Segmentation for Diagnostic Pathology*. Springer, 2001.

[10] Nipon Theera Umpon. Patch based white blood cell nucleus segmentation using fuzzy clustering. *ECTI Transaction Electrical Electronics Communications*, 3(1):5–10, 2005.

[11] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksak. A modified fuzzy clustering for white blood cell segmentation. In *3rd International Symposium on Biomedical Engineering*, pages 356–359, 2008.

[12] M. Kass, A. Witkins, and D. Terzopoulos. Snakes: Active contour models. In *1st International Conference on Computer Vision*, pages 259–268, 1987.

[13] G. Ongun and U. Halici. *Automated Contour Detection in Blood Cell Images by an Efficient Snake Algorithm*, volume 47. Springer, 2001.

[14] Cecilia Di Ruberto, Andrew Dempster, Shahid Khan, and Bill Jarra. Analysis of blood cell images using morphological operators. *Image and Vision Computing*, 20(2):133–146, 2002.

[15] Khoo Boon How, Alex See Kok Bin, Alex See Kok Bin, Ng Teck Siong, and Khoo Kong Soo. A novel white blood cell segmentation scheme using scale space filtering and watershed clustering. In *International Conference on Man, Machine Systems*, Langkawi Malaysia, 2006.

[16] Robiyanti Adollah, M.Y. Mashor, N.F. Mohd Nasir, H. Rosline, H.Mahsin, and H. Adilah. Blood cell image segmentation: A review. In N.A. Abu Osman, F. Ibrahim, W.A.B. Wan Abas, H.S. Abd Rahman, and H.N. Ting, editors, *4th Kuala Lumpur International Conference on Biomedical Engineering*, volume 21, pages 141–144. Springer Berlin Heidelberg, 2008.

[17] B. Ravi Kumar, Danny K Joseph, and T.V. Sreenivas. Teager energy based blood cell segmentation. In *14th International Conference on Digital Signal Processing*, Bangalore, India, 2002.

[18] Lalit Gupta, Sanjay Jayavanth, and Arun Ramaiah. Identification of different types of lymphoblasts in acute lymphoblastic leukemia using relevance vector machines. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, Minnesota, USA, 2009.

[19] S. Serbouti, A. Duhamel, H. Harms, U. Gunjer, U.M. Aus, J.Y. Mary, and R. Beuscart. Image segmentation and classification methos to detect leukemia. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 13, pages 260–261, 1991.

[20] Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data Clustering Theory, Algorithms,and Applications*. SIAM, Society for Industrial and Applied Mathematics, 2007.

[21] Nishchal K. verma, Abhisek Roy, and Shantaram Vasikarla. Medical image segmentation using improved mountain clustering technique version-2. In *IEEE Seventh International Conference on Information Technology*, pages 156–161, 2010.

[22] Bertrand Clarke, Ernest Fokue, and Hao Helen Zhang. *Principles and Theory for Data Mining and Machine Learning*. Springer, 2009.

[23] N. Pal, K.Pal, J.Keller, and J. Bezdek. A new hybrid c-means clustering model. In *Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on*, pages 179–184 vol.1, 2004.

[24] A. K. Jain. *Fundamentals of Digital Image Processing*. Pearson Education, India, 1st edition, 2003.

[25] R. C.Gonzalez and R. E. Woods. *Digital Image Processing*. Addison Wesley, 2nd edition, 1992.

[26] Subrajeet Mohapatra. Deveopmant of Impulse Noise Detection Schemes for Selective Filtering. Master's thesis, National Institute of Technolgy Rourkela, 2008.

[27] K. S. Ravichandran and B. Ananthi. Color skin segmentation using k-means cluster. *International Journal of Computational and Applied Mathematics*, 4(2), 2008.

[28] "The Wikipedia the Free Encylopedia Website". "http://en.wikipedia.org".

[29] B. T. Milne. *Measuring the fractal geometry of landscapes: Applied Mathematics and Computation*, volume 27. 1988.

[30] B. B. Mandelbrot. How long is the coast of britain? stastical self similarity and fractional dimension. *Science*, 156:636–638, 1967.

[31] A. P. Pentland. Fractal based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 1986.

[32] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[33] M. Hearst. Support vector machines. *IEEE Transactions on Intelligent Systems*, pages 18–28, 1998.

Author Biographies

Subrajeet Mohapatra is currently pursuing for Ph.D in the field of medical image processing at the National Institute of Technology Rourkela, India. He has contributed 10 research articles to different National and International Journals/Conferences. His research interest also includes Soft Computing, medical instrumentation.

Dipti Patra received her Ph.D degree in Electrical Engineering in the year 2006 from National Institute of Technology Rourkela, India. She is currently working as an Associate Professor in the Department of Electrical Engineering, National Institute of Technology Rourkela, India. Her research interest includes Stochastic Modeling, Image Processing and Signal Processing.

Sanghamitra Satpathi received her MD in pathology from VSS Medical College Burla, Odisha, India. Currently she is a consultant at Ispat General Hospital Rourkela. Her research interests includes clinical cytology, image cytometry and gynae pathology.