International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-4, Issue-3, March 2016

Automated Detection of Acute Myelogenous Leukemia Using Neural Classifier

Mr. Rajeev R Menon, Mr. Ranjith S

Abstract— **"Acute myeloid leukemia (AML)" is a form of cancer. In this case, abnormal myeloblasts (a type of white blood cell), red blood cells, or platelets are get formed within the bone marrow. AML is a quickly developing cancer of the blood and bone marrow. It is deadly if left untreated, because of its quick spread into the circulatory system and other fundamental organs. This is more prevalent among adults with an average age 65 years. The present strategies for AML detection include manual examination of the blood smear as the first step. Diagnosing leukemia depends on the way that white cell tally is expanded with immature blast cells (lymphoid or myeloid), and neutrophils and platelets are decreased. Thusly, hematologists routinely look at blood spread under magnifying instrument for legitimate identification and classification of blast cells. The presence of the abundance number of blast cells in blood is a significant sign of leukemia. It is difficult to detect leukemia because blood smear images are of complex nature. The imitation of similar signs of other disorders are also a main factor that make leukemia detection difficult. Moreover, the detection process need more time to diagnose and sometimes it is susceptible to errors. Hence, there is a need for automation of leukemia detection. This paper makes a survey that helps in analyzing the methodologies in detecting AML using the algorithms from neural networks. The proposed method is relied upon to deliver better results in accuracy and time consumption. Fault tolerance of neural algorithms are expected to produce more realistic results in very short time as compared with others.**

***Index Terms*— Feature Extraction, Hematology, Image Segmentation, K-means Clustering, Leukemia, Myeloblasts, Neural Networks**

I. INTRODUCTION

Leukemia or blood cancer is a condition in which abnormal blood cells formed in the bone marrow. Normally, leukemia involves the production of abnormal WBCs. But, the abnormal cells in leukemia do not function in the same way as normal WBCs. The leukemia cells keep on developing and gap, in the end swarming out the normal blood cells. It might then be exceptionally troublesome for the body to battle against diseases, control dying, and transport oxygen.

Based upon how rapidly the illness creates and the kind of anomalous cells delivered, we could group leukemia into taking after sorts: Leukemia is called an intense or acute leukemia in the event that it grows quickly. Substantial quantities of leukemia cells amass rapidly in the blood and bone marrow. Intense leukemia requires quick and forceful treatment. In any case, unending or chronic leukemia grow gradually after some time. These leukemia may not bring about particular side effects toward the start of their course. On the off chance that left untreated, the cells might in the end develop to high numbers, as in intense leukemia.

Leukemia are further classified as myeloid or lymphoid, depending upon the type of white blood cell that makes up the leukemia cells. Normally blood cells develop from stem cells that have the potential to differentiate into many cell types. Myeloid stem cells mature in the bone marrow and become immature white cells. These are called myeloid blasts which further mature to become either RBCs, platelets, or certain kinds of WBCs. Lymphoid blasts are formed by the development of lymphoid stem cells mature in the bone marrow. The lymphoid blasts later form into T or B lymphocytes. Myeloid leukemia are comprised of cells that emerge from myeloid cells, while lymphoid leukemia emerge from lymphoid cells. Knowing the kind of cell included in leukemia is critical on the grounds that it a crucial component for picking the suitable treatment.

The real reason for AML is still obscure and for the same reason AML is regularly difficult to analyze. Additionally, the side effects of the infection are fundamentally the same to flu or other regular ailments, for example, fever, shortcoming, tiredness, or pains in bones or joints [1]. It is predominant among grown-ups. Thinks about uncover that AML likewise makes up 15–20% of youth leukemia, approximately 60% of cases happen in individuals matured more youthful than 20 years. That is around 500 kids and teenagers in the U.S. every year are affected by AML [2], [3]. There are around 54,000 new instances of leukemia every year in the U.S. what is more, around 24,000 passing's because of leukemia. Around 3% of all new disease cases are made from leukemia. A noteworthy recognizing highlight of AML is that, there is no staging for AML. Fig. 1 indicates six distinct pictures, three delineating solid cells from non-AML patients and three from AML patients.

Different strategies are utilized for diagnosing leukemia. Current technique includes manual examination of the blood smear. Be that as it may, it is tedious and its precision relies on upon the administrator's capacity. There can likewise happen impersonation of comparative signs by different issue [4] prompting analytic perplexities. The present work concentrates on a procedure for the programmed identification of leukemia [1]. The principle reason for this paper is to actualize a completely automated neural classifier framework for AML detection. Another element, Hausdorff measurement (HD), is additionally utilized.

Rajeev R Menon, PG Student, Department of Computer Applications, KVM College of Engineering and Information Technology, Kerala, India, 9947088520.

Ranjith S, Assistant Professor, Department of Computer Applications, KVM College of Engineering and Information Technology, Kerala, India, 9895364228

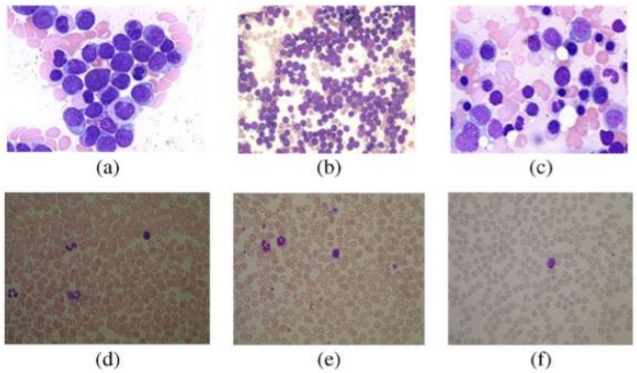

Fig.1 Images (a)–(c) Myeloblasts from AML patients. (d)– (f) Healthy cells from non-AML patients.

Rest of the paper is sorted out as takes after: Section II concentrates on the related works comparing to the proposed framework. Section III condenses the review of the framework model. Section IV portrays the procedure and outline in point of interest. Results are examined in Section V and paper is finished up in area VI.

II. Related Works

Over the years with the development of technology, digital image processing techniques have aided a lot in hematology and to analyze the cells that provide more accurate, standard, and remote disease diagnosis systems. But, there exist some difficulties in extracting the data from WBCs due to wide variation of cells in shape, size, edge, and position [5]. Also, there can be illumination imbalance and variation between the image contrast of cell boundaries and the background depending on the condition during the capturing process [6].

There are many early attempts that help in acute leukemia segmentation and classification [7]–[12]. The segmentation techniques are of mainly four classes: thresholding techniques, boundary-based, region based segmentation and hybrid techniques that combines the principle of both boundary and region criteria [13]. While examining peripheral blood or bone marrow smears, region-based or edge-based schemes are the mostly useful [14]. From the studies on color image segmentation algorithms by Ilea and Whelan [15] it was concluded that color images can produce most reliable image segmentations than gray-level images.

Many segmentation algorithms were presented in literature, including [16], [17], and [18]. Here, Otsu segmentation and automated histogram thresholding were done to segment WBCs from the blood smear image. The work in [19] used contour signature for the identification of the irregularities in the nuclear boundary. Similarly, the work in [20] is based on selective filtering to segment leukocytes from the other blood components. The work in [21] on the otherhand is based on hue and saturation value, color space, and expectation–maximization algorithm for identifying the cytoplasm and nucleus of the white blood cells.

III. System Overview

The block diagram of system model is given in Fig: 2. The AML images generated by digital microscopes are usually in RGB color space. It is subjected to preprocessing to overcome any background non uniformity due to irregular illumination. Preprocessing stage also undertakes a color correlation where RGB images are turned into L∗a∗b color space images. This step ensures perceptual uniformity.

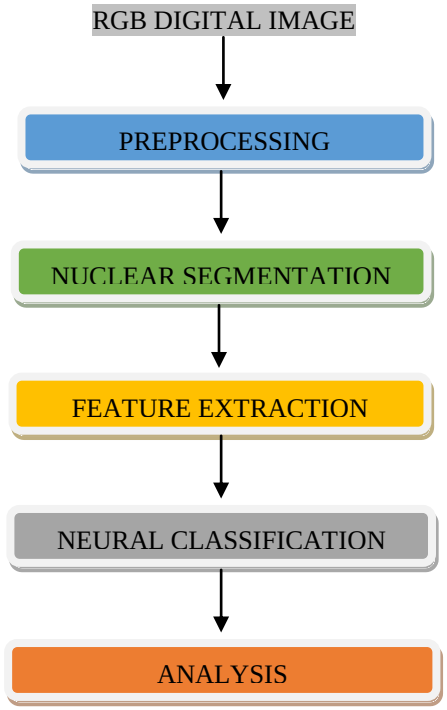

Fig.2 AML detection system overview

The preprocessed image is given as input for segmentation, where k-means clustering is used to bring out the nucleus of each cell. Segmentation is followed by feature extraction. That is, from the segmented image various features such as shape feature, GLCM features, color feature, Hausdorff dimension with and without LBP is extracted. This feature plays an important role in classification. Neural Algorithm is then used for classification. Finally, analysis is carried out for proper validation.

IV. Methodology

A. Preprocessing

The microscopic images are predominantly in RGB format. Along these lines, it is hard to segment. It is obvious that, blood cells and background might differ as for shading and power. A few reasons, for example, camera settings, shifting brightening and so forth are the contributing components for this issue. To get a definite yield by conquering these issues, the RGB picture is changed over to CIELAB. CIELAB is otherwise called L*a*b shading space. Here L speaks to the softness of the shading. a* speaks to its position in the middle of red and green. b* speaks to its position in the middle of yellow and blue. [1] There is yet another point of interest of CIELAB shading space. Utilizing this shading space, the perceptual contrast between hues is relative to the Cartesian separation in the CIELAB shading space. So the shading distinction between two examples can be ascertained by utilizing Euclidean separation.

B. Nuclei Segmentation

The process of segmentation involve separating a digital image to multiple parts. Here, in this technique a label is

assigned to every pixel. The pixels having the same label have certain characteristics. Within this system, segmentation is done to extract the nuclei from the AML image. Segmentation is performed here by employing K-means segmentation.

Segmentation using K-Means Clustering

Cluster analysis is the science of clustering objects according to measured intrinsic characteristics or a formal algorithms. K-mean is one of the common clustering techniques. A given data set is classified into certain number of clusters fixed a priori using this technique. K mean is one of the easiest unsupervised learning algorithm. The image in CIEL*a*b* is the input to the K-mean clustering and the output is three clusters.

The algorithm also require three user specified parameters: the number of clusters k, cluster initialization, and distance metric. Thus, the image which is converted into the CIEL*a*b* color space is given as input. Using the corresponding *a and *b values in the L*a*b color space each pixel is classified into the matching clusters. The three clusters we used here corresponds to nucleus, background and other cells. Here the cluster that contains the blue nucleus, which is then required for the feature extraction process.

C. Feature Extraction

Feature extraction is a technique used to transform the input data into set of features and is a form of dimensionality reduction. The important information from the input data forms the features set. The features used here are: Hausdorff dimension with and without LBP, Shape features, Texture features, Color feature.

Hausdorff dimension

The fractal estimation D is a value that gives an indication of how absolutely a fractal appears to fill space. Hypothetical fractal measurements are the packing dimension, the HD, and the Renyi measurement. All these methods are very easy to implement. In real time cases box-counting method is used.

In box counting, the number of boxes covering the point set is a power-law function of the box size. Here, the exponent of such power law is estimated as D. All fractal dimensions are real numbers and that will characterize the roughness of the objects. The fractal dimension D. The perimeter roughness of the nucleus can be used to differentiate myeloblasts.

The procedure for HD measurement using the box counting method is described below:

1) Binary image is obtained from the gray-level image of the blood sample.
2) To trace out the nucleus boundaries, edge detection technique is employed
3) Edges are superimposed by a grid of squares.
4) The HD can then be defined as follows:

$$HD = \frac{\log(R)}{\log(R(s))} \qquad (1)$$

Where R is the number of squares in the superimposed grid, and R(s) is the number of occupied squares or boxes. Higher HD signifies higher degree of roughness.

Local Binary Pattern

For texture classification, the concept of Local Binary Patterns (LBP) was introduced. The method incorporates both the structural and the statistical image analysis approaches into a single high efficiency transformation. However, it is similar to monotonic gray scale transformations and scaling.

In the LBP strategy every pixel is supplanted by a paired example that is gotten from the pixel's region. Every dark scale pixel P of a picture is utilized as a focal point of a circle with sweep r. The quantity of tests M decides the measure of focuses that are taken consistently from the form of the circle. These focuses are added from adjoining pixels if necessary. The specimen focuses are looked at against the pixel P one by one with a straightforward examination operation which come about a binary zero if the inside point is bigger than the present example point and a binary one otherwise. While doing this operation for instance clockwise from a specific beginning stage the outcome will be a paired example with length M.

Shape Features

The compactness of the image can serve as one of the shape features that help to classify the AML and NON-AML images. Region-based and boundary-based shape features are used for the shape analysis of the nucleus, which are extracted from the binary-equivalent image of the nucleus where the nucleus region is represented by the nonzero pixels.

Some of the shape features are:

1) ***Area:*** This feature is determined by counting the total number of non-zero pixels within the image region.
2) ***Perimeter:*** It is measured by calculating distance between successive boundaries pixels.
3) ***Compactness:*** It is defined as the measure of nucleus.
4) ***Solidity:*** This is the ratio of actual area and convex hull area. This is an essential feature for blast cell classification.

$$\text{Solidity} = \frac{\text{Area}}{\text{Convex Area}} \qquad (2)$$

5) ***Eccentricity:*** This feature is used to measure how much a shape of a nucleus deviates from being circular. As lymphocytes are more circular than the blast calculating this feature is of great importance.

$$Eccentricity = \frac{\sqrt{a^2 - b^2}}{a} \qquad (3)$$

6) ***Elongation:*** There will be abnormal bulging of the nuclei for leukemia affected cells. Hence this feature is used to signify this. Thus, the nucleus bulging is measured by a ratio called elongation. Elongation is defined as the ratio between maximum distance max R and minimum distance min R from the center of gravity to the nucleus boundary.

$$Elongation = \frac{Rmax}{Rmin} \qquad (4)$$

GLCM features

GLCM stands for gray-level co-occurrence matrix. Texture is defined as a function of the spatial variation in pixel intensities. The GLCM and associated texture feature calculations are important image analysis techniques. A second order statistics can be used to describe gray-level pixel distribution. Further, this information can be depicted in 2-D gray-level co-occurrence matrices, which can be computed for various distances and orientations. In order to use information contained in the GLCM there are some statistical measures to extract textual characteristics [22].

Some of these features are the following.

1) ***Energy:*** This is also known as uniformity (or angular second moment). It measures homogeneity of the image.
2) ***Contrast:*** This feature is a difference moment of the regional co-occurrence matrix. It measures the contrast or the amount of local variations present in an image.
3) ***Entropy***: This is used for measuring the disorder of an image. Non uniformity in the image is represented by very large entropy.
4) ***Correlation:*** The correlation feature is a measure of regional-pattern linear dependence in the image.

Color features

A color based feature call cell energy is evaluated. It is also known as measure of uniformity. We define feature δ as follows:

$$\delta = \sum_i \sum_j P^2(i,j) + (\sqrt{-1}) \left(\frac{\sqrt{\sum_{i=1}^{n} (x_i - x')^2}}{n-1} \right) \quad (5)$$

Where:

1) $x' = \sum_{i=1}^{n} \left(\frac{x_i}{n}\right)$
2) P(i, j) represents the normalized GLCM element for the i^{th} row and j^{th} column
3) $\sum_i \sum_j P^2(i, j)$ represents the ASM.

D. Classification

The challenging problem is in the selection of a classifier for classification. Here a neural classifier is used for making a decision surface for bisecting the two categories, i.e. AML and NON AML, and also for maximizing the margin of separation between two classes.

Artificial neural networks are relatively crude electronic networks of "neurons" based on the neural structure of the brain. They process records one at a time, and "learn" by comparing their classification of the record (which, at the outset, is largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm the second time around, and so on for many iterations.

Artificial Neural Network

An artificial neural network (ANN) is an information processing paradigm that is inspired by the way biological nervous system works. The key element of this paradigm is a novel structure of information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working together to solve specific problems.

Neural networks are realized by first trying to deduce the essential features of neurons and their interconnections.

1) ***Inputs, X_i:*** Typically, these values are external stimuli from the environment or come from the outputs of the artificial neurons. They can be discrete values for a set such as {0, 1} or real valued numbers.
2) ***Weights, W_i.*** These are real valued numbers that determine the contribution of each input to the neuron's weighted sum and eventually its output. The goal of neural algorithm is to determine the best possible set of weight values for the problem under consideration.
3) ***Threshold. U:*** The threshold is alluded to as a bias value. For this situation, a real number is added to the weighted sum. For the sake of simplicity, the threshold can be viewed as another data or weighted pair where W0 = U and Xi = - 1.
4) ***Activation Function, F*:** The Activation function for the original McCulloch Pitts neuron was the unit step-function. But now the ANN models have been expanded to include other functions such as sigmoid, piecewise, linear, and Gaussian etc.

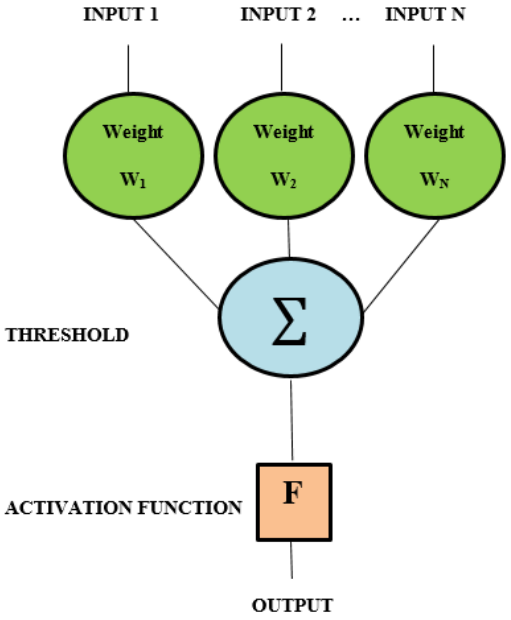

Fig.3 Artificial Neuron

Back-Propagation Algorithm

Back-Propagation is common algorithm used in neural networks. With this algorithm, the input data is repeatedly presented to the neural network. With each presentation the output of the neural network is compared with the desired output and an error is computed. This error is fed back (back-propagated) to the neural network and used to adjust the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing desired output. A schematic representation of the same is shown in Fig.4.

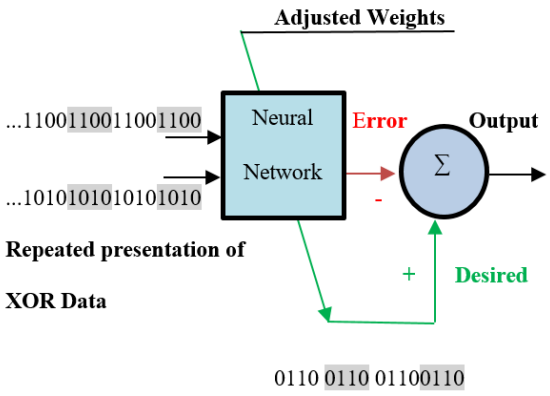

Fig.4 Training using Back-Propagation

V. Simulation And Results

Data Set: The algorithm was implemented in python and MATLAB environment. To validate the method, experiment were performed using blood microscopic image. First the image is preprocessed. Then it is segmented and features are extracted. After that 35 images were used for training. In that 15 were cancerous and the other 20 were noncancerous. Then the input image is tested with the help of neural classifier.

Segmentation: The image is subjected to segmentation, using k-means clustering algorithm. The outcome of k-means clustering is 3 clusters. Those clusters correspond to nucleus, background and other cells.

Training: Segmentation of the input image is followed by feature extraction. About 13 features were extracted for the effective training. Thirty cancerous images and thirty noncancerous images were then trained using a neural classifier.

Output: The testing of input image was done after training. When the test image is given as the input each of the sequential operations were performed, i.e.; preprocessing, segmentation, feature extraction and at last classification so as to obtain an output as either cancerous or non-cancerous. This method is simple and less time consuming. It gives a perfect decision about the disease.

VI. Conclusion

This paper has reported the design, development, and evaluation of an automated screening system for AML in blood microscopic images. The presented system performs better segmentation of the nucleated cells, feature extraction, classification and analysis than the k-mean clustering techniques. Features such as shape, texture, color is constructed to obtain all the information required to perform efficient classification. HD with LBP and color feature presents a good demarcation between AML and NON-AML cells. Finally the use of neural classifier made it more vulnerable to possible errors.

References

[1] Sos Agaian, Senior Member IEEE, Monica Madhukar, and Anthony T. Chronopoulos, Senior Member, "Automated Screening System for Acute Myelogenous Leukemia Detection in Blood Microscopic Images" in IEEE SYSTEMS JOURNAL.

[2] A. Nasir, M. Mashor, and H. Rosline, "Unsupervised colour segmentation of white blood cell for Acute leukemia images," in Proc.IEEE IST, 2011, pp. 142–145.

[3] O. Lahdenoja, "Local binary pattern feature vector extraction with CNN," in Proc. 9th Int. Workshop Cellular Neural Netw. Appl., 2005, pp. 202–205

[4] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in Proc. CIMSA, 2005, pp. 96–101

[5] D. Mandal, K. Panetta, and S. Agaian, "Face recognition based on logarithmic local binary patterns," in Proc. SPIE, Image Process., Algorithms Syst. XI, 2013, vol. 8655, pp. 865514-1–865514-12.

[6] F. Sadeghian, Z. Seman, A. Ramli, B. Kahar, and M. Saripan, "A framework for white blood cell segmentation in microscopic blood images using digital image processing," Biol. Procedures Online, vol. 11, no. 1, pp. 196–206, Jun. 2009.

[7] V. Piuri and F. Scotti, "Morphological classification of blood leucocytes by microscope images," in Proc. CIMSA, 2004, pp. 103–108.

[8] M. Subrajeet, D. Patra, and S. Satpathy, "Automated leukemia detection in blood microscopic images using statistical texture analysis," in Proc. Int. Conf. Commun., Comput. Security, 2011, pp. 184–187.

[9] H. Ramoser, V. Laurain, H. Bischof, and R. Ecker, "Leukocyte segmentation and classification in blood-smear images," in Proc. IEEE EMBS, 2006, pp. 3371–3374.

[10] C.Reta, L. Altamirano, J.A.Gonzalez, R.Diaz, and J.S.Guichard, "Segmentation of bone marrow cell images for morphological classification of acute leukemia," in Proc. 23rd FLAIRS, 2010, pp. 86–91.

[11] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S.Beksac, "Feature extraction and classification of blood cells for an automated differential blood count system," in Proc. IJCNN, 2001, vol. 4, pp. 2461–2466.

[12] S. Mohapatra and D. Patra, "Automated leukemia detection using hausdorff dimension in blood microscopic images," in Proc. Int. Conf. Emerg. Trends Robot Commun. Technol., 2010, pp. 64–68.

[13] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," J. Electron. Imaging, vol. 13, no. 1, pp. 146–165, Jan. 2004.

[14] K. Nallaperumal and K. Krishnaveni, "Watershed segmentation of cervical images using multi scale morphological gradient and HIS colorspace," Int. J. Imaging Sci. Eng., vol. 2, no. 2, pp. 212–216, Apr. 2008.

[15] R. Adollah, M. Mashor, N. Nasir, H. Rosline, H. Mahsin, and H. Adilah, "Blood cell image segmentation: A review," in Proc. IFMBE. Berlin, Germany: Springer-Verlag, 2008, ch. 39, pp. 141–144.

[16] S.Suri, S.Setarehdan, and S.Singh, "Advanced Algorithmic Approaches to Medical Image Segmentation": State-of-the-Art Application in Cardiology, Neurology, Mammography and Pathology. Berlin, Germany: SpringerVerlag, 2001, pp. 541–558

[17] F. Scotti, "Robust segmentation and measurement techniques of white cells in blood microscope images," in Proc. IEEE Conf. Instrum. Meas. Technol., 2006, pp. 43–48.

[18] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, p. 27, Apr. 2011.

[19] S. Mohapatra, D. Patra, and S. Satpathi, "Image analysis of blood microscopic images for acute leukemia detection," in Proc. IECR, 2010, pp. 215–219.

[20] S. Mohapatra, S. Samanta, D. Patra, and S. Satpathi, "Fuzzy based blood image segmentation for automated leukemia detection," in Proc. ICDeCom, 2011, pp. 1–5.

[21] N. Sinha and A. G. Ramakrishnan, "Blood cell segmentation using EM algorithm," in Proc. 3rd Indian Conf. Comput. Vis., Graph, 2002, pp. 445–450.

[22] R. M. Haralick, "Statistical and structural approaches to texture," Proc. IEEE, vol. 67, no. 5, pp. 786–804, May 1979