International Journal of Engineering Research and Science & Technology

ISSN : 2319-5991
Vol. 1, No. 2
April 2015

2nd National Conference on "Recent Advances in Science Engineering & Technologies" RASET 2015

Organized by

Department of EEE, Jay Shriram College of Technology, Tirupur, Tamil Nadu, India.

www.ijerst.com

Email: editorijerst@gmail.com or editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Special Issue, Vol. 1, No. 2, April 2015
2nd National Conference on "Recent Advances in Science, Engineering&Technologies" RASET-2015

Research Paper

DETECTION OF LEUKEMIA IN BLOOD MICROSCOPIC IMAGES USING FUZZY LOGIC

D Kavi Priya[1], S R Krithiga[1]*, P Pavithra[1] and J Rajesh Kumar[1]

*Corresponding Author: **S R Krithiga** ✉ krithu125@gmail.com

Blood cancer is a form of cancer which attacks the blood, bone marrow, or lymphatic system. To automatically recognize lymphoblasts and detect leukemia in peripheral blood samples, an efficient methodology such as pre-processing, nuclei segmentation, feature extraction, classification is proposed. The need for automation of leukemia detection arises since current methods involve manual examination of the blood smear as the first step toward diagnosis which is time-consuming, and its accuracy depends on the operator's ability. Morphological, textural and color features are extracted from the segmented nucleus and cytoplasm regions of the lymphocyte images, which helps hematologists for easier identification and early detection of leukemia from blood microscopic images which will improve the chances of survival for the patient. Various image processing techniques have helped to analyze the cells that lead to more accurate, standard, and remote disease diagnosis systems. However, there are a few complications in extracting the data from WBCs due to wide variation of cells in shape, size, edge, and position The most Common drawback observed from various survey from different literature and many more which as a problem of over-segmentation of cluster of nucleus and accuracy of detection of affected nuclei problems. Computer simulation involved the following tests: comparing the impact of Hausdorff dimension on the system before and after the influence of local binary pattern, comparing the performance of the proposed algorithms on subimages and whole images, and comparing the results of some of the existing systems with the proposed system. Eighty microscopic blood images were tested, and the proposed framework managed to obtain 98% accuracy for the localization of the lymphoblast cells and to separate it from the subimages and complete images.

INTRODUCTION

Leukemia

Cancer is a generic term to describe a group of malignant diseases with cells displaying uncontrolled and invasive growth along with metastasis. It can develop in almost any organ or tissue, such as the blood, lymph node, bone, breast, skin, colon, or nerve tissue. Leukemia is a fast-growing cancer of the blood and bone marrow, broadly classified as:1) acute leukemia

[1] Department of Information Technology, SNS College of Technology-Coimbatore, Tamilnadu, India.

(which progresses quickly); and 2) chronic leukemia (which progresses slowly) It is also known as liquid cancer which develops from cells in the blood, bone marrow, and lymphatic system. And it is different from other cancers as it does not produce solid masses or tumors. In leukemia, the abnormal growth white blood cells flood the marrow, providing no room for red blood cells and other platelets.

Overview

The automated differential blood counting system assists in the diagnosis of many ailments .From the literature on leukocyte image segmentation it is observed that most of the schemes thrust upon extracing the nuclei and very few schemes are able to extract the cytoplasm that too with lesser accuracy. One possible reason for higher cytoplasm segmentation error is direct use of gray level intensity or color (Red–Green–Blue) as features which are linearly unseparable in the image plane. It is seen through simulation that performance of many pre–existing methods fail to classify boundary pixels (nucleus–cytoplasm and cytoplasm–background) in leukocyte images due to color overlapping. Some segmentation schemes have been developed specifically for lymphocyte images.

Introduction To The Project

Diagnostic confusion occurs due to imitation of similar signs by other disorders. Moreover, the identification task is usually difficult due to the variety of features and the often unclear images cause missing out on vital indicators as to which form of leukemia is being observed.It is also predicted that the total number of deaths during the same year due to leukemia and lymphoma will be 23, 720 and 20, 200 respectively. Many attempts have been made in the past to construct systems that aid in leukemia segmentation and classification.There are four main categories in segmentation techniques: thresholding techniques and boundary-based and region-based segmentation and hybrid techniques that combine boundary and region criteria.A proper combination of both boundary and region information may present better results,color images present more reliable image segmentations than gray-level images.

Problem Statement

Diagnosing leukemia is based on the fact that white cell count is increased with immature blast cells (lymphoid or myeloid), and neutrophils and platelets are decreased. Therefore, hematologists routinely examine blood smear under microscope for proper identification and classification of blast cells . The presence of the excess number of blast cells in peripheral blood is a significant symptoms of leukemia. In this method intensity and color variations of the image is work out without accord of computational speed, and the result shows that cancer cells in the background with heterogeneous intensities and colors are properly segmented.

Objective of the Project

The main objective of this work is to identify and early detection of leukemia from blood microscopic images which will improve the chances of survival for the patient. In order to enhance the contrast, without affect to the color information algorithm simply call for converting the image and the approach endure with histogram equalization, processing only intensity component. To give improved input for the segmentation, algorithm encompasses color normalization of the image that deals with the ambiguous color occurred because of satin

artifacts and poor illumination. *C*-means clustering algorithm is applied to segment RGB information.

Organization of the Report

Section I and II focuses in detail on the process overview of the proposed model. Sections V and VI show the image processing methods being used to perform enhancement and segmentation. Section VII builds in extracting the feature. Sections IX present the experimental results of the classifier system based on the features extracted. Section X and XI contains conclusions and future work.

Project Overview

The project overview gives a detailed depiction of the sequence of steps that are to be followed for efficient classification of acute and chronic leukemia. The first step involves preprocessing the complete images to overcome any background nonuniformity due to irregular illumination. Preprocessing also includes color correlation where RGB images are converted to L*a*b color space images. This step ensures perceptual uniformity. This step is followed by *fuzzy c-means clustering* to bring out the nucleus of each cell. Segmentation is followed by feature extraction based on which classification we use *feed forward classifier* and validation are performed.

EXISTING SYSTEM

Leukemia is often difficult to diagnose since the precise cause of AML is still unknown. Thus the symptoms of the disease are very similar to flu or other common diseases, such as fever, weakness, tiredness, or aches in bones or joints. If the described symptoms are present, blood tests, such as a full blood count, renal function, liver enzyme and blood count, have to be done. Thus the identification task is usually difficult due to the variety of features and the often unclear images cause missing out on vital indicators as to which form of leukemia is being observed. To the complex nature of blood smear images and variation in slide preparation techniques, much work has to be done to meet real clinical demands. Thus, these factors can lead to wrong diagnosis. However, there are a few complications in extracting the data from WBCs due to wide variation of cells in shape, size, edge, and position.

Drawbacks of Existing System

- Time Consuming.
- Results in Noise and Blur Images.
- Lack of Accuracy.

PROPOSED SYSTEM

The system overview gives a detailed depiction of the sequence of steps that are to be followed for efficient classification of acute leukemia. The first step involves preprocessing the complete images to overcome any background nonuniformity due to irregular illumination. Preprocessing also includes color correlation where RGB imagesare converted to L*"a"*b color space images. This step ensures perceptual uniformity.

The next step is Nuclei segmentation which uses modified Fuzzy c-means means clustering to bring out the nucleus of each cell. Segmentation is followed by feature extraction in which we can extract feature based on shape,size and texture. classification are performed using neural network classifier(Feed forward classifier)and validation are performed.The other employed selective filtering to segment leukocytes

from the other blood components. The work in employed hue, saturation, and value (where hue represents color, saturation indicates the range of gray in the color space, and value is the brightness of the color and varies with color saturation), color space.

ADVANTAGE OF PROPOSED SYSTEM

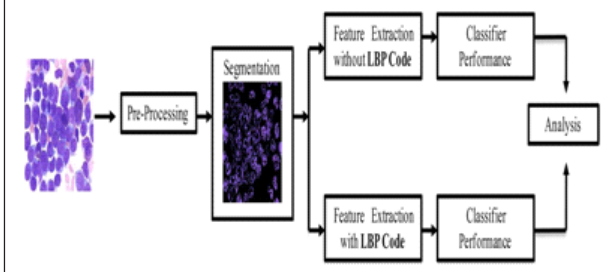
- This method tries to avoid the duplicate data in images as well as improve the performance of the system.
- More Efficient.

LIST OF MODULES

- Pre-processing
- Segmentation
- Feature extraction
- Classification

Screening System of Leukemia

Module Description



PRE-PROCESSING

Image Acquisition: For AML, the American Society of Hematology (ASH) for their online image bank of leukemia cells were accessed. The ASH image bank is a web-based image library that offers comprehensive and growing collections of images relating to a wide range of hematology categories Our database for AML comprised 80 images—40 from AML patients and 40 from non-AML patients. The resolution used for our classication was 184×138 pixels.

CIELAB Color Features and Color Correlation: The images generated by digital microscopes are usually in RGB color space, which is difcult to segment. The blood cells and image background varies greatly with respect to color and intensity. It is caused by multiple reasons such as camera settings, varying illumination, and aging stain. In order to make the cell segmentation robust an adaptive procedure is used: the RGB input image is converted into the CIELAB or, more correctly, the CIEL”a”b” color space. The key reasons for these are, rst, to reduce memory requirement and to improve the computational time. Second, the perceptual difference between colors is proportional to the Cartesian distance in the CIELAB color space. Therefore, the color differences between two samples can be calculated by using a Euclidean distance. Third, it has two color components (a and b), and it is designed to approximate human vision closely matches human perception of lightness or it can be used to adjust the lightness contrast using the L component. Finally, the a and b components can be used to make accurate color balance corrections.

SEGMENTATION

Image segmentation is the process of partitioning digital image into multiple segments(set of pixels).It is used to extract important information from an input image.The efciency of subsequent feature extraction and classication relies greatly on the correct identication of the myeloblasts. Segmentation in this system is performed for extracting the nuclei of the leukocytes using color-based clustering. Cluster analysis study of methods and algorithms for grouping, or

clustering, objects according to measured intrinsic characteristics or similarity. Cluster analysis does not use category labels that tag objects with prior identiers. Fuzzy c-means is the most popular unsupervised learning algorithm and is also a simple clustering algorithm. Clusters corresponding to nucleus (high saturation), background (high luminance and low saturation), and other cells (e.g. erythrocytes and leukocyte cytoplasm). Here, every pixel is assigned to one of these classes using the properties of the cluster center.Once these actions are performed, the following texture and shape-based features are then extracted from these whole images:

1) Edge Enhancement (used by the Sobel operator), to enhance the borders of the membranes and the cells (this helps in segmenting grouped cells and subsequent edge detection;
2) Canny Edge Detection, to obtain outputs with continuous edges, in general;
3) Dilation, to connect the separated points of the membrane in a better way (it gives a good outline of the perimeter of the nuclei);
4) Hole-Filling, to fill internal holes of the connected element having the largest area.

PROPOSED ALGORITHM

Fuzzy c-means algorithm:

The FCM clustering algorithm was first proposed by Dunn et. al. and promoted as the general fuzzy c-means clustering algorithm by Bezdek et. al. The main purpose of FCM algorithm is to make the vector space of a sample point be divided into a number of sub-spaces in accordance with distance measure. The algorithm divides the data set I = {I1, I2. In} into c clusters and n is the number of all the pixels in the image. Let the membership function uik, uik [0, 1] show the degree of the pixel Ik, k=1, 2. . . n belonging to cluster i(1d"id"c). Then the result can be denoted by a matrix of fuzzy membership function matrix U = [uik]c×n. We represent typicality by tik, tik [0, 1] and the typicality matrix by T = [tik]c×n. According to the definition of the theory, we have c i=1uik = 1 for every pixel in the image.

The objective function to be minimized is:

$$J_{MAFCM} = \sum_{i=1}^{c}\sum_{k=1}^{n}\left(C_F u_{ik}^{m} + C_T\, t_{ik}^{\eta}\right) * \|I_k - V_i\|^2 + \sum_{i=1}^{c}\gamma \sum_{k=1}^{n}(1 - t_{ik})^{\varsigma}$$

where V ={v1,v2,.....vi} is the characterized intensity center. The parameters CF >0, CT >0, m>1, $\gamma > 1$, the $\upsilon i > 0$ are user defined constants. The constants CF and CT define the relative importance of fuzzy membership and typicality values in the objective function. Note that uik has the same meaning of membership as that in FCM. Similarly, tik has the same interpretation of typicality as in PCM.

FEATURE EXTRACTION

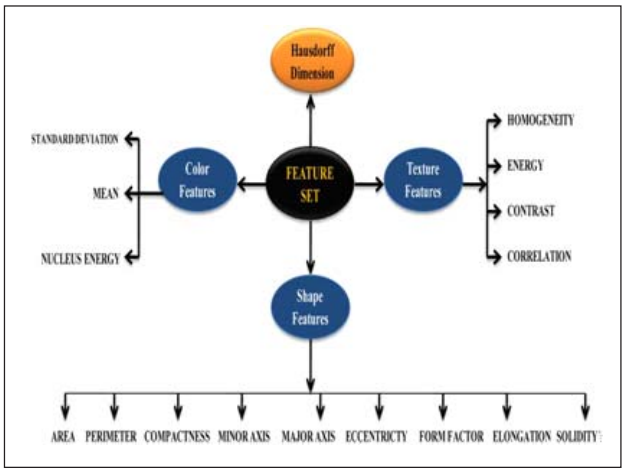
Feature extraction is a technique of redening a large set of redundant data into a set of features of reduced dimension. Transforming the input data into the set of features is called feature extraction. Feature selection greatly inuences the classier performance;.. Certain features were widely used as they gave a good classication. We implemented these features on whole images in our system. Those features were considered to boost the classier performance. In order to construct an effective feature set. We implemented following on whole images in our system:

- Hausdroff Dimension(HD).

- Local Binary Pattern(LBP).

A. Hausdroff Dimension

Fractals have been used in medicine and science in the past for various quantitative measurements . The fractal dimension D is a statistical quantity that gives an indication of how completely a fractal appears to ll space. The most important theoretical fractal dimensions are the Rényi dimension, the HD, and the packing dimension. The box-counting dimension is widely used, partly due to their ease of implementation. In a box counting algorithm, the number of boxes covering the point set is a power-law function of the box size.All fractal dimensions are real numbers that characterize the fractalness (texture/roughness) of the objects. Myeloblast can be differentiated using perimeter roughness of the nucleus. HD is an essential feature considered in our proposed system.



The procedure for HD measurement using the box counting method is elaborated below as an algorithm:

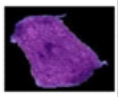
1) binary image in obtained from the gray-level image of the blood sample;
2) edge detection technique is employed to trace out the nucleus boundaries;
3) edges are superimposed by a grid of squares;
4) the HD may then be dened as follows:
5) HD =log(R)/ log(R(s))

where R is the number of squares in the superimposed grid, and R(s) is the number of occupied squares or boxes (box count). Higher HD signies higher degree of roughness. It shows how the nucleus from a noncancer cell is superimposed with a grid of squares to perform suitable box counting. It depicts the results of HD on subimages and complete images.

LOCAL BINARY PATTERN

The concept of local binary pattern (LBP) was introduced for texture classication. In order to deal with textures at different scales, the LBP operator was later extended to use neighborhoods of different sizes. Dening the local neighborhood as a set of sampling points evenly spaced on a circle centered at the pixel to be labeled allows any radius and number of sampling points. When a sampling point does not fall in the center of a pixel, bilinear interpolation was employed In the LBP method where each pixel is replaced by a binary pattern that is derived from the pixel's neighborhood. Each grayscale pixel P of an image is used as a center of a circle with radius R =1or 2 (radius R is usually kept very small).M represents the number of samples that determines the number of points that are taken uniformly from the contour of the circle. If needed, these points are interpolated from adjacent pixels.

Each grayscale pixel P is compared with these sample points one by one. If the center point P is larger than the current neighborhood sample point I, the result is a binary zero; otherwise, the result is a binary one.

Features	Cancerous	Normal
Images		
1.Mean	32.3699	37.1222
2.Standard Deviation	47.3662	41.0178
3.Area: The total number of none zero pixels within the image region.	6453	1985
4. *Perimeter:* Calculating distance between successive boundary pixels gives the perimeter.	256	807.293
5.*Elongation:* The nucleus bulging is measured in terms of a ratio called elongation. This is defined as the ratio between the maximum distance (*Rmax*) and the minimum distance (*Rmin*) from the center of gravity of the nucleus boundary.	1.1357	1.522
6. *Eccentricity*: Since lymphocytes are more circular than the myeloblasts, eccentricity is an important feature. It is a parameter that is used to measure how much the shape of a nucleus deviates from being circular.	0.4740	0.7541
7. *Form-factor*: A Dimensionless parameter considered which changes with surface irregularities.	1.2373	0.0383
8. *Solidity*: An essential feature for blast classification equals the ratio of actual area over the convex hull area	0.5679	0.2117
9.*Compactness/***roundedne ss:** is the measure of a nucleus	10.1559	328.32

Shape Features

One of the shape features that has proven to be a good measure for classifying AML by their shape is compactness. The shape of the nucleus, according to haematologists, is an essential feature for discrimination of myeloblasts. 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 where the nucleus region is represented by the nonzero pixels. Table II displays the difference in the values of the shape features for a pair of cancer and noncancer nuclei.

GLCM Features

Texture is dened as a function of the spatial variation in pixel intensities.Gray-level pixel distribution can be described by second-order statistics such as the probability of two pixels having particular gray levels at particular spatial relationships. 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, Haralick dened some statistical measures to extract textual characteristics. Some of these features are the following.

1) Energy: Also known as uniformity(or angular second moment), it is a measure of homogeneity of image.
2) Contrast: The contrast feature is a difference moment of the regional cooccurrence matrix and is a measure of the contrast or the amount of local variations present in an image.
3) Entropy: This parameter measures the disorder of an image. When the image is not texturally uniform, entropy is very large.
4) Correlation: The correlation feature is a measure of regional-pattern linear dependence in the image.

Color Features

In addition to the features aforementioned, we have used the following color-based feature. Cell Energy: Also known as the measure of uniformity, it is the different Lab image components. We define feature "δ" to be where x =n i=1 xi/n, P(i, j) represents the normalized GLCM element for the ith row and jth column, and ij P2(i, j) represents the ASM.

CLASSIFICATION

The selection of a classication technique for

classication is a challenging problem because an appropriate choice given the available data can signicantly help improving the accuracy. There is a plenty of statistical techniques, which aim at solving binary classication tasks. Here,we use a neural network classifier for constructing a decision surface in the feature space that bisects the two categories, i.e., cancerous and noncancerous, and maximizes the margin of separation between two classes of points.

A neural network consists of units(neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand.These range from function representation to pattern recognition, which is what we will consider here.

A feedforward neural network is an artificial neural network where connections between the units do *not* form a directed cycle.The feedforward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes and to the output nodes. There are no cycles or loops in the network.

CONCLUSION

This paper has reported the design, development, and evaluation of an automated screening system for leukemia in blood microscopic images. It uses 80 high-quality 184 × 138 size images obtained from the American Society of Haematology (ASH). The system performs automated processing, including color correlation, segmentation of the nucleated cells, and effective validation and classification. A public dataset of blood samples, specifically designed for the evaluation and comparison of the performances algorithms of segmentation and image classification. A feature set exploiting the shape, color, and texture parameters of a cell is constructed to obtain all the information required to perform efficient classification. Furthermore, a color feature called cell energy was introduced, and results show that this feature presents a good demarcation between cancer and noncancer cells.

FUTURE WORK

Further research will focus on collection of more samples to yield better performance and building an overall system for cancer classifications.

REFERENCES

1. Lopez C and S Agaian, “Iterative local color normalization using fuzzy image clustering,” in *Proc. SPIE, Mobile Multim./Image Process., Security, Appl.*, 2013, vol. 8755, pp. 875518-1–875518-12.
2. ASH Image Bank: American Society of Hematology. [Online]. Available: http://imagebank.hematology.org/Default.aspx
3. Scotti F, “Robust segmentation and measurement techniques of white cells in blood microscope images,” in *Proc. IEEE Conf. Instrum. Meas. Technol.*, 2006, pp. 43–48.
4. Pentland A, “Fractal based description of natural scenes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 661–674, Nov. 1984.
5. Agaian S, J. Astola, and K. Egiazarian, *Binary Polynomial Transformsand Nonlinear*

Digital Filters. New York, NY, USA: Marcel Dekker,1995.

6. Sinha N and A. G. Ramakrishnan, "Blood cell segmentation using EM algorithm," in *Proc. 3rd Indian Conf. Comput. Vis., Graph.*, 2002, pp. 445–450.
7. Bordes A, S. Ertekin, J. Weston, and L. Bottou, "Fast kernel classifiers with online and active learning," *J. Mach. Learn. Res.*, vol. 6, pp. 1579– 1619, Sep. 2005.
8. Wharton E, K. Panetta, and S. Agaian, "Logarithmic edge detection with applications," *J. Comput.*, vol. 3, no. 9, pp. 11–19, Sep. 2008.
9. Hu J, J. Deng, and M. Sui, "Color space conversion model from CMYK to LAB based on prism," in *Proc. IEEE GRC*, 2009, pp. 235–238.
10. Hunter Labs. Hunter lab color scale, Hunter Associates Lab., Reston, VA,USA, *Insight on Color 8 9 (August 1–15, 1996)*. [Online]. Available:http://www.hunterlab.com/appnotes/an08_96a.pdf
11. Haworth C, A. Hepplestone, P. Jones, R. Campbell, D. Evans, and M. Palmer, "Routine bone marrow examination in the management of acute lymphoblastic leukaemia of childhood," *J. Clin. Pathol.*, vol. 34,no. 5, pp. 483–485, May 1981.
12. Sezgin M and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *J. Electron. Imaging*, vol. 13, no. 1, pp. 146–165, Jan. 2004.
13. Nallaperumal K and K. Krishnaveni, "Watershed segmentation of cervical images using multiscale morphological gradient and HSI color space," *Int. J. Imaging Sci. Eng.*, vol. 2, no. 2, pp. 212–216, Apr. 2008.
14. Rangayyan R, *Biomedical Image Analysis. Series Title: Biomedical Engineering*. Boca Raton, FL, USA: CRC Press, Dec. 2004.
15. Adollah R, 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. Nasir A, M. Mashor, and H. Rosline, "Unsupervised colour segmentation of white blood cell for Acute leukaemia images," in *Proc. IEEE IST*, 2011, pp. 142–145.
17. Ilea D and P. Whelan, "Image segmentation based on the integration of colour-texture descriptors—A review," *Pattern Recognit.*, vol. 44, no. 10/11, pp. 2479–2501, Oct./Nov. 2011.
18. Inaba H, M. Greaves, and C. Mullighan, "Acute lymphoblastic leukaemia," *Lancet*, vol. 381, no. 9881, pp. 1943–1955, Jun. 2013. [49] O. Lahdenoja, "Local binary pattern feature vector extraction with CNN," in *Proc. 9th Int. Workshop Cellular Neural Netw. Appl.*, 2005, pp. 202–205.
19. Mandal D, 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. [51] *Acute Myeloid Leukemia*, National Cancer Institute, Bethesda, MD, USA, 2006.
20. Sadeghian F, 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.

International Journal of Engineering Research and Science & Technology
Hyderabad, INDIA. Ph: +91-09441351700, 09059645577
E-mail: editorijlerst@gmail.com or editor@ijerst.com
Website: www.ijerst.com