ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Automated AML Detection from Complete Blood Smear Image Using KNN Classifier

Nayana B. Sen[1], Mercy Mathew[2]

PG Student [Communication], Dept. of ECE, BC Caarmel Engineering College, Kerala, India[1]

Assistant Professor, Dept. of ECE, BC Caarmel Engineering College, Kerala, India[2]

ABSTRACT:The detection of cancers related to blood and bone marrow is done by analysing microscopic images of the blood cells. A basic idea about the disease is given by the difference in count of white blood cells. Manual counting of white blood cells is a, time-consuming and tiresome process which is also susceptible to error. So, an automatic system for detection has to be introduced. This work presents such a simple technique which automatically detects Acute MyelogenousLeukemia (AML) which is a subtype of acute leukemia. The proposed method is simple and it differ from other papers by the use of complete blood smear image for classification. The system will analyse the features in microscopic images and examine changes in texture, geometry, colour, Hausdorff Dimension and statistical characteristics.The segmentation of nucleus from blood smear image is performed using colour based algorithm which is efficient than existing algorithm andthe classification process is done using KNN classifier.KNN is a straight forward classifier, where samples are classified based on the class of their nearest neighbour.

KEYWORDS:AML, Hausdorff Dimension, KNN classifier

I.INTRODUCTION

Leukemia is a cancer affecting the blood cells. The broad categories of bloodcells are red blood cells (RBCs), white blood cells (WBCs) and platelets. Diagnosis of leukemia is based on the fact that the blood consists of excess of white cells with immature blast cells (lymphoid or myeloid). Acute myelogenousleukemia (AML) is a cancer affecting the blood and bone marrow.The word "acute" in acute myelogenousleukemia denotes the disease's rapid progression. It is known as myelogenousleukemia because it affects a group of white blood cells called the myeloid cells, which develop into the various types of mature blood cells, such as red blood cells, white blood cells and platelets.Other names of acute myelogenousleukemiaare acutemyeloblasticleukemia, acute myeloid leukemia, acute granulocytic leukemia and acute non-lymphocytic leukemia.

As for every disease, blood cancer has several symptoms like anemia, weakness, sweating, extreme fatigue, breath shortness etc. The similar symptoms may be occurred due to other diseases, so that diagnosing blood cancer from symptoms is not an advisable method. In the proposed system, microscopic blood images can be used to find out the cancer cells. Pre-processing steps can include the image acquisition and colour correlation, which can be done on complete images. The converted images can be segmented using a colour based segmentation algorithm. Feature extraction can be done by extracting shape, colour and GLCM features along with Hausdorff dimension. For classification, KNN classifier is used.

Different methods for automatic segmentation and detection of leukemia in blood smear images have been proposed [1-10]. Diagnosis of leukemia in a traditional techniques require a keen observation of peripheral blood and bone marrow samples under a microscope and it relies on the morphological characters of white blood cells. This process is time consuming and greatly depends on the skills and experience of an expert. One of the main difficulties in the diagnosis is that it is impossible to exactly discriminate between cancerous cells and normal cells which are almost identical in size and shape. The texture of nuclei at high magnification is one of the most important parameters for the classification of blast cells and, hence, for the diagnosis of leukemia. In all the existing systems, a common drawback found, that is, they classify only sub-images. But in the proposed system, entire image classification is done. And here the classification is done with a linear support vector machine. The result is then compared with the existing models.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

This paper focus on the implementation of fully automated system for AML classification. The system is applied to complete blood smear images containing multiple nuclei. Two new features have been used for classification: cell energy and Hausdorff dimension (HD).The segmentation and classification is done with better algorithm. The system works on a step-by-step procedure. The proposed system consist of following stages: (a) Image Acquisition, (b) Segmentation, (c) Feature Extraction and (d) Classification. The blocks of the system are explained in subsequent sections.

II.PROCESSES INVOLVED

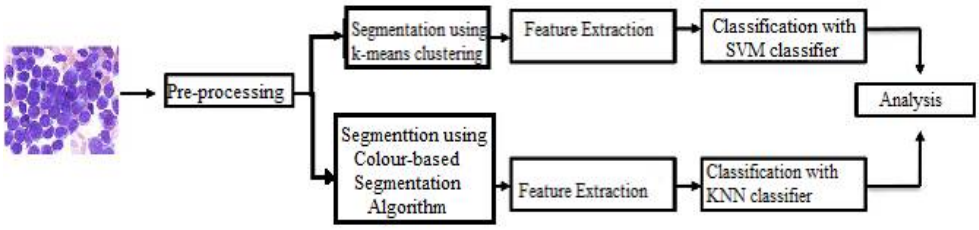


Fig.1 Proposed System

The aim of the proposed approach is to present a more robust system which provide an efficient segmentation of blood smear images for a high performance. In order to achieve the goal, the proposed system follows four main steps:

- Pre-process the image to reduce the background non-uniformities and perform colour correlation;
- Employ segmentation on whole images using a colour based algorithm thereby achieving a robust identification of the nuclei of the white cells;
- Extraction of different sets of features for a database of images;
- To classify the image under any of the two classes: Healthy or Cancerous

A. Pre-processing

It is the first step, which is followed by segmentation, feature extraction and classification. Pre-processing consists of the following steps:

Image Acquisition: For AML, we accessed the American Society of Hematology (ASH). Their online image bank of leukemia cells provided the images. The ASH image bank is a library of images based on the web and it providesa large and varyingcollection of images relating to a wide range of cancer categories. They provide high-quality images captured using different microscopes in different resolutions.

CIELAB Colour Features and Colour Correlation: The images generated by digital microscopes are usually in RGB colour space, which is difficult to segment. In practice, the blood cells and image background varies greatly with respect to colour and intensity.

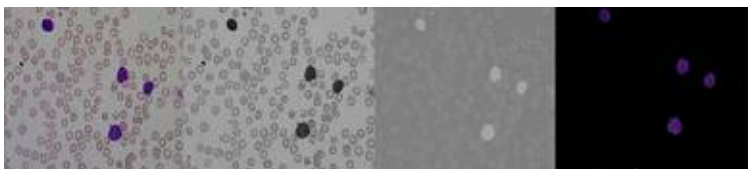


Figure.2 RBG to CIELAB conversion and segmentation example

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

B. Image segmentation

It is done to extract important portion from the input image i.e., the nucleus. Segmentation using different clusterig algorithms are employed in existing system whereas a colour based segmentation algorithm is used in proposed work. For gray level images, several algorithms for segmentation have been developed. [6]presents the segmentation process using active contours. After segmentation, the following processes will be performed: Edge enhancement, Canny edge detection, Dilation and Hole-filling.

C. Feature Extraction

In feature extraction, the input data are transformedinto a set of features. The classifier performance is greatly influenced by the feature selection and therefore, the correct selection of features is a very crucial step. The features considered for calculation are:Hausdorff dimension, Shape features,Texture features [16] and GLCM (Grey Level Co-occurrence Matrix) features. Hausdorff dimension is the ratio of log of number of squares in the superimposed grid to the number of occupied squares. The other features are explained in coming section.

D. Classification

While choosing the method for classification, great care should be taken since it will determine the accuracy of the system. For the classification tasks, many statistical approaches are available. The proposed methodology uses a KNN (K-Nearest Neighbour) classifier for classifying healthy and unhealthy cells whereas in the existing system, SVM classifier is used.

III.SYSTEM ANALYSIS

After pre-processing the acquired image, the images have to undergo segmentation, feature extraction and classification. In the existing work, the segmentation is done using k-means clustering algorithm. In this work, three clusters are formed, namely clusters of nucleus, cytoplasm and other cells. Among them, the desired cluster is to be selected by the system.In the figure below (Fig.3), the nuclei is in the first cluster. But this will not happen always. Sometimes, the second or third cluster will contain the nuclei. To avoid the confusion and selection of clusters, a new colour-based algorithm is implemented.

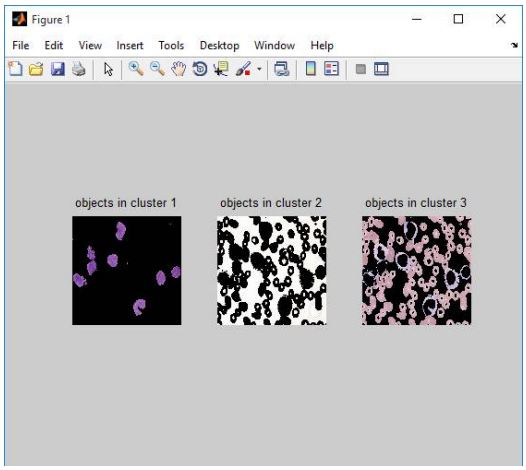


Fig.3 Segmentation result using k-means clustering algorithm

A. For segmentation, a colour-based algorithm. Steps of Segmentation process:

1) The very first step is to read an RGB image.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

2) The next step is to convert RGB image to L*a*b* colour space.
3) A matrix is formed so as to see values of L*a*b* colour space at each pixel.
4) Enter the pixel range of colours user wants to see.

B. Feature extraction

Hausdorff dimension: For various quantitative measurements, fractals have been used for a long time.Fractal dimension is a statistical quantity that indicates how completely a fractal appears to fill space. Haushorff dimension and the packing dimension are the most important theoretical fractal dimensions. In practical applications the box-counting dimension is widely used, due to the fact that it iseasyto implement. In box counting algorithms, the number of boxes covering a point set is a power law function of the box size. All fractal dimensions are estimated as the exponent of such power laws and are real numbers that characterize the fractalness (texture or roughness) of the objects. The perimeter roughness of the nucleuscan be used to differentiatemyeloblasts.

Steps to find Hausdorff dimension:

1) Obtain the binary image from the colour image
2) The nucleus boundaries are traced out using edge detection technique
3) The edges are superimposed by grid of squares.
4) Then, the hausdorff dimension is defined as

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

where R is the number of squares in the superimposed grid and R(s) is the number of occupied squares.

As the roughness increases, the HD also increase. The algorithm is implemented as below:

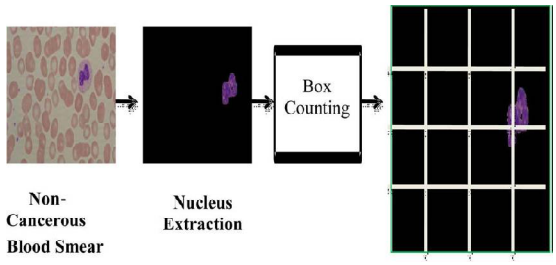


Fig.4 Implementation of box count measuring

Shape features: The shape of the nucleus is an essential feature for distinguishing the blast cells. For analysing the shape features of the nucleus, the region andboundary based shape features are extracted. The input image is converted into binary equivalent form and all the features are extractedfrom it. In this image, the nucleus region is represented by the non-zero pixels.

The features considered are as follows:

1) Area: The total number of none zero pixels within the image region.
2) Perimeter: Calculating distance between successive boundary pixels gives the perimeter.
3) Compactness: Also called roundness and it is the measure of a nucleus.
4) Solidity: An essential feature for blast classification and it is the ratio of actual area to convex hull area.
5) Eccentricity: Since lymphocytes are more circular than the blast, eccentricity is an important feature. It is a parameter that is used to measure how much a shape of a nucleus deviates from being circular.
6) Elongation: The nucleus bulging is measured in terms of a ratio called elongation. This is defined as the ratio between maximum distance (Rmax) and minimum distance (Rmin) from the centre of gravity to the nucleus boundary.
7) Form-factor: Dimensionless parameter considered which changes with surface irregularities.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

The other shape features include major axis length, minor axis length, roundness, standard deviation, mean and energy.

Texture features:Texture is defined as a function of the spatial variation in pixel intensities [17]. Texture features include homogeneity, energy, correlation, entropy.

GLCM Features: The GLCM feature calculation is an image analysis technique [12]. Different texture features like energy, contrast, entropy, correlation are extracted using this method.

C. Classification

For classification, KNN classifier is used. K nearest neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure in k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour.

IV. SIMULATION RESULTS AND SYSTEM PERFORMANCE

Experiments are done on the images obtained from ASH. The first process done on the image is CIELab colour correlation. Here, the input image in RGB colour space. The next step is to segment out the nuclei from the complete image. It is done by mentioning the pixels of corresponding region to be displayed, i.e., of the nuclei. Segmentation using the colour based algorithm is better in the sense that a single output is obtained. Using k-means clustering, a number of clusters are formed among which one of them should be selected for further process.

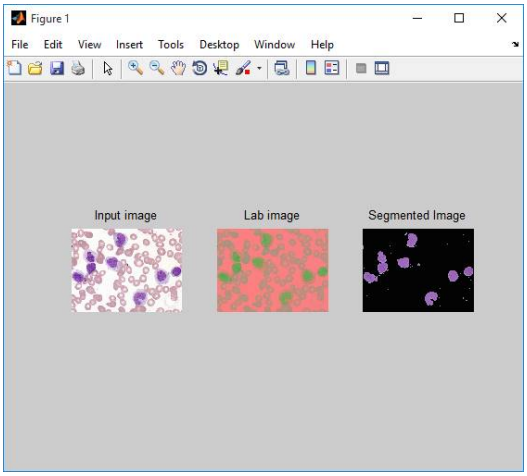


Fig.5Result of Pre-processing and Segmentation

After segmentation, a set of features of the nuclei are extracted. The first feature is the Hausdorff Dimension. The result of HD ofnormal cell and cancerous cell is given in Table-I which is equal to 1.5501 and 1.7828 respectively. The HD is found out on the basis of box counting. The number of white blood cells in the images of cancerous case is more than that of the normal case. This will result in a comparable difference in HD. That is why, the HD turns to be an important feature in the proposed system.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

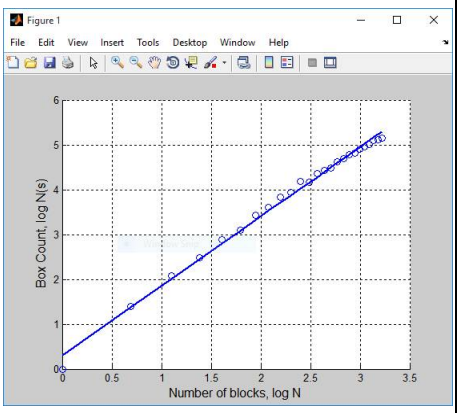
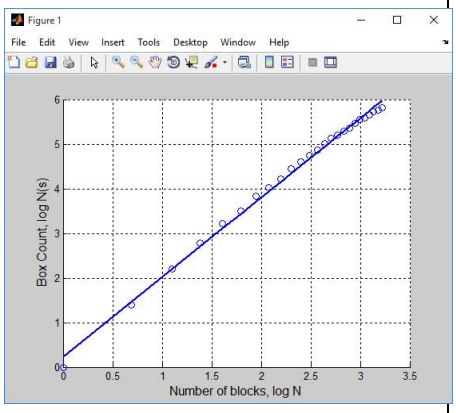
International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

TABLE-I

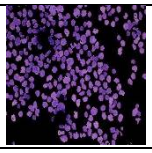
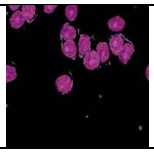
HD COMPARISON

Type	Healthy	Cancerous
Graph		
HD Value	1.5501	1.7828

The region and boundary based shape features are extracted for shape analysis of the nucleus. The features are extracted from the binary equivalent image of the nucleus where the nucleus region is represented by nonzero pixels.The following table (Table II) displays the difference in values of the shape features for a pair of cancer and non-cancer nuclei. Beyond the HD and shape features, texture and GLCM features are the calculations which comes under the image analysis techniques. Some of the features used to extract the textual characteristics are:

1) Energy: It is the measure of homogeneity of image and it is also known as angular second moment.
2) Contrast: It is the measure of the contrast or the amount of local variations present in an image.
3) Correlation: It is the measure of regional-pattern linear dependence in the image.
4) Entropy: It measures the disorder of an image. The entropy is very large in case of texturally non-uniform images.

TABLE-II
SHAPE FEATURE COMPARISON

Features	**Cancerous**	**Normal**
Images		
Mean	198.5744	184.7889
Area	673618	599846
Standard Deviation	59.4928	46.3664
Perimeter	5.8222e+03	4.3972e+03
MajorAxis Length	1.1299e+03	1.1269e+03
Minor Axis Length	841.9264	827.0189
Eccentricity	0.6669	0.6793
Solidity	0.7282	0.8678

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

After calculating the features the image is classified under one of the categories: Cancerous or Healthy. For classification, KNN classifier is used. There is no training phase for KNN classifier and once the data is loaded into memory, it began to classify.To ensure the performance of the proposed system, the measures of accuracy are calculated. They are: precision, specificity, sensitivity and f-measure. These are all defined on the basis of possible outcomes of the classifier system.

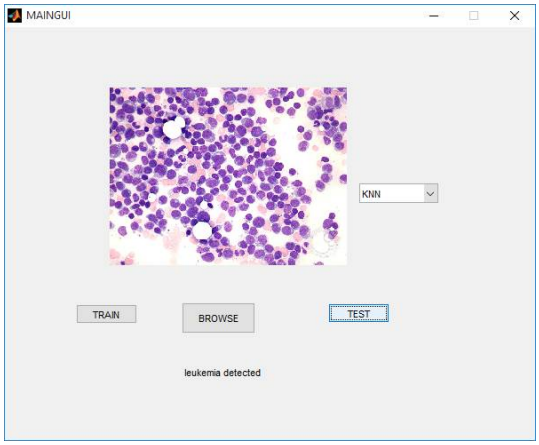


Fig.6 Final Result of Leukemia Detection

Classification of a specimen results in four possibilities: true positives (TP), when cancer cells are correctly identified; false positives(FP),when noncancerous cells are identified as cancerous; true negatives(TN), when noncancerous cells are correctly identified; and false negatives(FN), when cancer cells are identified as noncancerous. Table-III shows the evaluation of these parameters.

TABLE-III
PARAMETERS FOR PERFORMANCE EVALUATION

Parameters	Formulae
Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)
Precision	TP/(TP+FP)
F-Measure	$\frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity}$

After calculating all these performance parameters, it is found that, KNN classifier is as efficient as SVM in case of specificity and precision. But in other cases, SVM is a little better than KNN classifier.

V. CONCLUSION

Design of an AML detector is proposed which is efficient than existing systems.The presented system performs automated processing, including colour correlation, segmentation of the nucleated cells, feature extraction and classification. A feature set exploiting the shape, colour, and texture parameters of a cell is constructed to obtain all the information required to perform efficient classification. All these processes are done on a complete microscopic image of blood smear.

From the images collected from ASH, the nuclei is segmentedusing a colour based algorithm which is better than k-means clustering algorithm. The segmented images undergo the process of feature extraction, in which the features

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

including Hausdorff Dimension are calculated. For classification, KNN classifier is employed which don't have a training phase. The K-Nearest Neighbour(KNN) Classifier is a simple classifier that works well on basic recognition problems. It is as better as SVM classifier in case of specificity and precision.

REFERENCES

[1] Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images", in Proc. CIMSA,pp.96101.,2005
[2] 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, pp. 184187. ,2011
[3] S. Mohapatra, S. Samanta, D. Patra, and S. Satpathi, "Fuzzy based blood image segmentation for automated leukemia detection", in Proc.DeCom, pp. 15., 2011
[4] A. Nasir, M. Mashor, and H.Rosline, Unsupervised colour segmentation of white blood cell for Acute leukaemia images, in Proc. IEEE IST,pp. 142145., 2011
[5] FatemehZamani Reza Safabakhsh "An unsupervised GVF Snake Approach for White Blood Cell Segmentation based on Nucleus" in ICSPProceedings, 2006.
[6] P. Bamford and B. Lovell, "Method for accurate unsupervised cell nucleus segmentation", in Proc. Eng. Med. Biol. Soc. Conf., vol. 3, pp. 27042708., 2001
[7] K. Panetta, S. Agaian, S. Nercessian, and A. Almunstashri, "Shapedependent canny edge detector",in Opt. Engg., vol.50, no. 8, pp.087008-108700812, Aug. 2011.
[8] W. Shitong, K. F. L. Chung, and F. Duan, "Applying the improved fuzzy cellular neural network IFCNN to white blood cell detection", Neurocomputing, vol. 70, no. 79, pp. 13481359, Mar. 2007.
[9] Q. Liao and Y. Deng, "An accurate segmentation method for white blood cell images", in Proc. IEEE Int. Symp. Biomed. Imaging, Atlanta, GA,USA,pp. 245248, 2002
[10] D.Foran, D.Comaniciu, P.Meer, and L.A.Goodell,"Computer-assisted discrimination among malignant lymphomas and leukemia usingimmunphenotyping, intelligent image repositories, and telemicroscopy", IEEE Trans. Inf. Technol. Biomed., vol. 4, no. 4, pp.265273, Dec. 2000.
[11] S. Serbouti, A. Duhamel, H. Harms, U. Gunzer, J. Mary, and R. Beuscart, "Image segmentation and classification methods to detect leukemias", in Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 260261. , 1991
[12] SosAgaian, Monica Madhukar, and Anthony T. Chronopoulos,"Automated Screening System for Acute MyelogenousLeukemia Detection in Blood Microscopic Images," in Proc. IEEE, Oct 22,2014.
[13] Vivek Singh Rathore,MessalaSudhirKumar,AshwiniVerma,"Colour Based Image Segmentation Using L*A*B* Colour Space Based On Genetic Algorithm, in IJETAE , June 2012.
[14] Shanhong Zhu, Weipeng Dong and Wei Liu, "Hierarchical Reinforcement Learning Based on KNN Classification Algorithms, in International Journal of Hybrid Information Technology", Vol.8, No.8 ,pp.175-184 ,2015.
[15] FarnooshSadeghian, ZaininaSeman, Abdul Rahman Ramli, BadrulHisham Abdul Kahar, and M-Iqbal Saripan "A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing",Shulin Li (ed.), Biological Procedures Online, Volume 11,2009.
[16] SubrajeetMohapatra,Dipti Patra,SanghamitraSatpathy,"Automated Leukemia Detection in Blood Microscopic Images using Statistical Texture Analysis",inICCCS, February 2011.

BIOGRAPHY

Nayana B. Sen received the B.Tech degree in Electronics and Communication Engineering from Kerala University, Trivandrum, Kerala, India, in 2014 and she is currently doing M.Tech in Electronics and Communication Engineering with specialization in Communication Engineering in BC Caarmel Engineering College, Pathanamthitta, MG University Kottayam, Kerala, India.

Mercy Mathew is an Assistant Professor in Electronics and Communication Engineering Department in BC Caarmel Engineering College, Pathanamthitta, Kerala. After B.Tech, she accomplished M.Tech in Electronics and Communication Engineering with specialization in VLSI from Sathyabhama University, Tamil Nadu, India.