AUTOMATED ACUTE MYELOGENOUS LUKEMIA DETECTION IN BLOOD MICROSCOPIC IMAGE

JESLY JAMES
M.G UNIVESITY COLLEGE MUTTOM
KERALA,INDIA

Abstract

Acute Myelogenous leukemia (AML) is a subtype of acute leukemia. The current method of leukemia detection is by manual examination of the blood cells through the microscope. It very time consuming and also it depend on the operator's ability. Current method is very costly .The leukemia is very dangerous if it is left untreated. So the need of automatic detection of leukemia essential. Leukemia can be automatically detected mainly by 5 steps .The first step is preprocessing the image. In preprocessing the RGB color space is converted into CIEL*a*b color space. This is done due to three reasons. First the RGB (red ,green, blue) color space is difficult to segment, second in RGB image background varies greatly with respect to color and intensity. After the preprocessing, noises in the image will remove using a wiener filter. Then the image is segmented into three clusters using K-means clustering algorithm. The segmented white blood cell nucleus feature is then extracted. By using the SVM classifier the image is classified into cancerous or noncancerous image.

Keywords: Acute myelogenous, leukemia, k-means clustering, RGB color space, Segmentation.

1. INTRODUCTION

Blood is the fundamental component of human life. A human body has approximately 70 litters of water of which five litters are blood. Blood is essential for maintaining homeostasis. Now these days number of blood cancer patients increases day by day. White blood cell (WBC) is larger than a Red blood cell (RBC). White blood cells play a major role in the diagnosis of different diseases. Leukemia is another name of blood cancer. It is prevalent among adults. Various methods are used for diagnosing leukemia. Current method involves manual examination of the blood smear. But it is time consuming and its accuracy depends on the operator's ability. In this paper a method developed which will automatically detect the leukemia.[1]

2. PROCESS OVERVIEW

2.1 Pre processing

The image generated by digital microscopes is mainly in RGB colour space. That is difficult to segment. Practically blood cells and image background varies with respect to colour and intensity. It may be because of multiple reasons such as camera settings, varying illumination etc. So for getting correct output by overcoming these problems in this method the RGB image is converted to CIELAB. The advantage of CIELAB is the perceptual difference between colours is proportional to the Cartesian distance in the CIELAB colour space. so the colour difference between two sample can be calculated by using Euclidean

distance. CIELAB is also known as L*a*b colour space. Here L represents the lightness of the colour. a* represents its position between red and green. b* represents its position between yellow and blue[1].

2.2 Noise Removal

Removal of noise from an image is still a challenging problem in image processing system. Here a efficient method to remove the noise, wiener filter is proposed. The purpose of the wiener filter is to filter out the noise that has corrupted a signal. This filter is based on the statistical approach. Mostly all the filters are designed for a desired frequency response. Wiener filter deal with the filtering of a image from a different view. The goal of wiener filter is reduced the mean square error as much as possible. This filter is capable of reducing the noise and degrading function. One method that we assume we have knowledge of spectral property of noise and original signal. We used the linear time invariant filter which gives output similar as to the original signal as much as possible [6].

This filter uses the central pixel value and local statistical value, therefore the filtering process is adaptive, resulting in a high quality restored image. The degraded image y is assumed to be

$$y = x + n$$

where x is the original image and n is the noise whose standard deviation is σ_n. The adaptive wiener filter (AWF) is expressed by the following equations:

$$\hat{x}(i,j) = \frac{\sigma_x^2(i,j).y(i,j) + \sigma_n^2.\bar{y}(i,j)}{\sigma_x^2(i,j) + \sigma_n^2}$$

$$\sigma_x^2(i,j) = \max\{0, \sigma_y^2(i,j) - \sigma_n^2\}$$

Where $\hat{x}(i,j)$ is the filtered output at (i,j), σ_n^2 is the noise variance which is constant over the image, and $\bar{y}(i,j)$ and $\sigma_y^2(i,j)$ are local mean and local variance of the input image y at (i,j). The output is calculated using the central pixel value and local statistics, it means that the filtering process for each pixel.[9]

A. *Nuclei Segmentation*

The k-means is a clustering method which is one of the most popular unsupervised learning algorithms due to its simplicity. Here, the k-means clustering has been used for image segmentation.k-means clustering is based on minimizing the objective function,J as

$$J = \sum_{i=1}^{n} \sum_{j=1}^{k} \left\| x_i - c_j cj \right\|^2$$

Where n is the number of data is the number of cluster,xi is the i-th sample and cj is the j-th Centre of cluster. In the developed system each data is clustered into 3 groups for all analyses. The algorithm as follows

1. Select randomly k points into the space represented by pixels that are being clustered. These points represent the initial cluster Centre cj

2. Assign each data to the nearest Centre.

3. When all data have been assigned, recalculate the new Centre position

4. Repeat step 2 and 3 until the centers are no longer move. This produce a separation of the object into group from which the matric to be minimizes can be calculated.[2]

D. *Local Binary Pattern*

The concept of Local Binary Patterns (LBP) was introduced for texture classification . The LBP combines the structural and statistical image analysis approaches into a single high efficiency transformation which is invariant with respect to monotonic gray scale transformations and scaling. In the LBP method each pixel is replaced by a binary pattern which is derived from the pixel's neighborhood. Each gray scale pixel P of an image is used as a center of a circle with radius r. The number of samples M determines the amount of points that are taken uniformly from the contour of the circle. These points are interpolated from adjacent pixels if needed. The sample points are compared against the pixel P one by one with a simple comparison operation which result a binary zero if the center point is larger than the current sample point and one otherwise. When doing this operation for example clockwise from a certain starting point the result will be a binary pattern with length M.[4]

E. *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.in this developed method feature extraction can be mainly classified into 4.Those are color feature, hausdroff dimension, texture feature, and shape feature

Hausdorff Dimension

Fractals have been used in medicine and science in the past for different quantitative measurement. It indicate how completely a fractal appears to fill space. Here box-counting method is used for the HD determination. In a box counting algorithm, the number of boxes covering the point set is a power-law function of the box size. Fractal dimension is estimated as the exponent of such power law. All fractal dimension are real numbers that characterize the fractalness(roughness)of the object. Myeloblast can be differentiated using perimeter roughness of the nucleus. The procedure for HD measurement using the box counting method is elaborated below

1.Binary image is 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 super imposed by a grid of squares

4.The HD may then be defined as follows

$$HD=\frac{log(N)}{log(N(S))}$$

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[1].

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 nonzero pixels representing the nucleus region.

•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 successive boundary pixels

•Compactness: Compactness is the measure of nucleus

$$Compactness=\frac{Perimeter^2}{Area}$$

•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 as

$$Solidity=\frac{Area}{Concex\ Area}$$

•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 blast.

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

•Elongation: Abnormal bulging of the nucleus is also an feature which signifies towards leukemia. Hence 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 center of gravity to the nucleus boundary and is given by

$$Elongation=\frac{R_{max}}{R_{min}}$$

•Formfactor: This is an dimensionless parameter which changes with surface irregularities and defined as[3]

$$\text{Formfactor} = \frac{4 \times pi \times Area}{Perimeter^2}$$

Texture Feature

•Energy: Also known as uniformity.it is measure of homogeneity of image

•Contrast: The contrast feature is a difference moment of the regional co-occurrence matrix and is a measure of the contrast or the amount of local variations present in an image

•Entropy: This parameter measure the disorder of an image. When the image is not texturally uniform, entropy is large

•Correlation: The correlation feature is a measure of regional-pattern linear dependence in the image[1]

Color Feature

•Cell energy: also known as the measure of uniformity, it is defined as[1]

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

F. Classification

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one target value" (i.e. the class labels) and several attributes" (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very high dimension feature space. In this feature space a linear decision surface is constructed Z through some non-linear mapping chosen a priori. An optimal hyper plane is here defined as the linear decision function with maximal margin between the vectors of the two classes. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. The SVM is only directly applicable for two-class tasks. The effectiveness of SVM depends on the selection of kernel. In general, the RBF kernel is a reasonable first choice. This kernel nonlinearly maps samples into a higher dimensional space so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear. So in this thesis a machine learning approach to automatically classify the feature vectors is

employed. Here SVM is trained with the cancerous nucleus value and noncancerous values.so that SVM can classify the image into cancerous or noncancerous.

3. SIMULATION AND RESULTS

B. *Data Set*

The algorithm was implemented in 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 60 images used for training. In that 30 were cancerous and the other 30 were noncancerous. Then the input image is tested with the help of SVM classifier.

C. *Noise Removal*

By using wiener filter the noise in the blood microscopic image is removed. Figure 4.1 shows the input image with the noise. Figure 4.2 shows the noise removed input image. If the noise is not removed then there is a chance of wrong output.

D. *Segmentation*

After filtering the image the image is segmented using k-means clustering algorithm. Output of k-means clustering gives 3 outputs. Those are cluster correspond to nucleus, background and other cells. Figure 4.1 shows the cluster corresponds to nucleus after the segmentation. The cluster corresponds to nucleus was used for further steps like feature extraction and classification
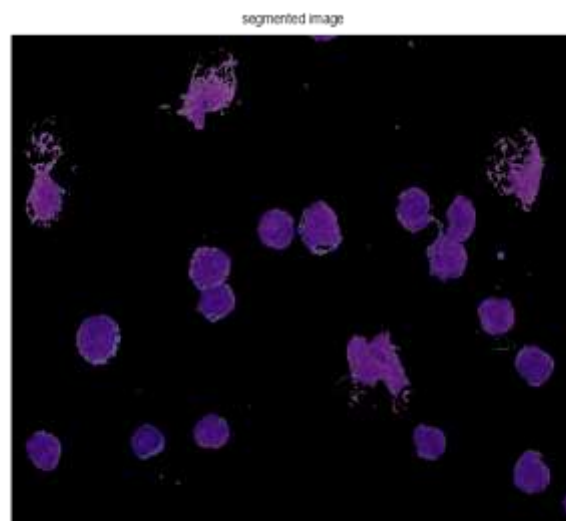

Figure 4.1. The cluster corresponds to nucleus after the segmentation

E. *Training*

After the segmentation of the input image, the next step is the feature extraction. Almost 13 features were extracted. Then using the SVM classifier 30 cancerous image and 30 noncancerous images were trained.

Figure 4.4 shows the training of 30 cancerous image and figure 4.5 shows the training of 30 noncancerous image

E.   Output

After the training of 60 images the next step was the testing of input image. When the test image is given as the input first step is the preprocessing and noise removal.it follows the segmentation, feature extraction and at last classification so that we get the output as either cancerous or noncancerous image. This method is simple and less time consuming. It gives a perfect decision about the disease.

4. CONCLUSION AND FUTURE WORK

In this thesis, a new method for automatically detecting the acute myelogenous leukemia by extracting the features and using SVM classifier is presented. For detecting the disease the images was collected from American Society of Hematology(ASH) for their online image bank of acute myelogenous leukemia cells. The ASH image bank is a web based image library that offers comprehensive and growing collections of image relating to a wide range of hematology categories. They provide high-quality images captured using different microscopes in different resolutions. Our database for AML comprised 80 images-40 from AML patients and 40 from non AML patients .

In this thesis microscopic image of the blood cell was given as the input. The noise from the input image is then removed by using wiener filter. The goal of the wiener filter is reduced the mean square error as much as possible. The input image is then segmented into 3 categories. These are the image only consist of white blood nucleus, other two categories are first image only consist of background and image consist of other cells. For segmenting the image k-means clustering algorithm was used.

The segmented white blood cell is then used as the input. The features of the nucleus were the calculated. In this thesis the features calculated were mean, standard deviation, area, perimeter, eccentricity, elongation, form factor, roundness, energy, entropy, homogeneity , contrast, fractal dimension. Then 30 noncancerous image and 30 cancerous image were trained using a SVM classifier. In the SVM classifier feature of 30 images(cancerous) were stored in a vector and feature of another 30 images(noncancerous)is stored on another vector.

The test image was then given as the input image. The noise from test image is then filtered out and it is classified into three nucleus. The features of the image corresponds to white blood cell was found. It was then gave to SVM classifier. The SVM classifier classified it into cancerous image or into noncancerous image. So that detection of leukemia was found very easily.

The main advantages of this method is that it is very accurate and less time consuming. This method is automatically detect the acute myelogenous leukemia.

In this thesis the myelogenous leukemia detected easily. For further modification a microscope can be attached with this. If a microscope is attached with this then the leukemia can be detected directly by taking the blood sample of the patient. If this modification is done with this then the testing method of the disease become perfect.

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

[3] S. Mohapatra, D. Patra, and S. Satpathi, "Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images," in Proc. ICSMB, 2010, pp. 49–54.

[4] [4] O. Lahdenoja, "Local binary pattern feature vector extraction with CNN," in Proc. 9th Int. Workshop Cellular Neural Netw. Appl., 2005, pp. 202–205.

[5] [5] K. Nallaperumal 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.

[6] [6] Naman Chopra et al, "Despeckling of Images Using Wiener Filter in Dual Wavelet Transform Domain",International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4069-4071.

[7] [7] Fabio Scotti, "Robust segmentation and measurement techniques of white cells in blood microscope images," in Proc. IEEE Conf. Instrum. Meas. Technol., 2006, pp. 43–48.

[8] [8] Fauziah kasmin, anton satria prabuwono, azizi abdullah Faculty of Information and Communication Technology, "Detection of leukemia in human blood sample based on microscopic images": A study Universiti Teknikal Malaysia Melaka Hang Tuah Jaya, 76100, Durian Tunggal, Melaka, Malaysia.

[9] [9] Chikako Abe and Tetsuya Shimamura, "Iterative Edge-Preserving Adaptive Wiener Filter for Image Denoising" International Journal of Computer and Electrical Engineering, Vol. 4, No. 4, August 2012.

[10] [10]T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and   rotation invariant texture classification with local binary patterns,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987,Jul. 2002.

[11]G. Srinivasan and G. Shobha, "Statistical texture analysis," Proc.World Acad.   Sci., Eng. Technol., vol. 36, pp. 1264–1269, 2008.