

Cancerous Cell Detection in Bone Marrow Smear

Tanuja K. Sarode,
PhD
Asst. Professor,
Thadomal Shahani Engg.
College, Bandra (W),
Affiliated to Mumbai
University

Bhavik K. Thakkar
B.E. Student,
Thadomal Shahani Engg.
College, Bandra (W),
Affiliated to Mumbai
University

Sanket J. Purandare
B.E. Student,
Thadomal Shahani Engg.
College, Bandra (W),
Affiliated to Mumbai
University.

Vikas M. Gupta
B.E. Student,
Thadomal Shahani Engg.
College, Bandra (W),
Affiliated to Mumbai
University

ABSTRACT

In recent years, Cancer as a disease has grown widely and it has been the major cause of mortality in humans. In this paper the main focus is on a specific type of blood cancer known as leukemia. The detection of leukemia is done via observation of the physical properties of the cells found in the bone marrow smear. This paper proposes novel leukemia detection using discrete cosine transform (DCT), and the results obtained are compared with the well-known image segmentation and feature extraction method given in the literature. From the results it is observed that the proposed method outperforms the image segmentation in terms of accuracy by 8.5% and precision by 5.4% as compared to the existing image segmentation method.

Keywords

Leukemia, Lymphocytes, DCT, Nucleus, Cell, Cytoplasm

1. INTRODUCTION

Application of Computer Engineering in Medical Sciences has gained significance importance over the past few years. Computer Engineering has increased the accuracy of prediction, diagnoses, detection and treatment providing the medical field with the state of the art software and embedded systems. Life expectancy of the human species has increased exponentially due to the advent of technology in recent years. Hence a solution has been proposed in this report for the detection of cancer in the early stages.

Acute Lymphocytic Leukemia (ALL), also known as acute lymphoblastic leukemia is an important hematic disease. It is fatal if it is left untreated due to its rapid spread and increase into the bloodstream and other vital organs and it mainly affects young children and adults over 50. Early diagnosis and checking of the disease is very important for the recovery and betterment of patients especially in the case of small children. The symptoms of ALL are common also in much such other disease and for this reason; the diagnosis of this disease is very difficult. One of the steps in the diagnostic procedures comprises the microscope inspection of blood peripherals. The inspection consists on the research of white cells with malformation due to the presence of a cancer. From decades, this operation is performed by experienced operators, which basically perform two main analyses: the cell classification and counting (now performed by cytometers) [2,3]. Interestingly, the morphological analysis just requires an image, not a blood sample and hence is suitable for low-cost, standard-accurate, and remote screening systems.

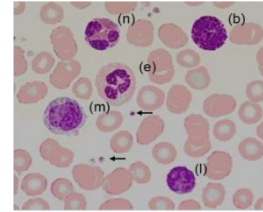


Fig 1: Blood's white cells marked with colorant: basophil (b), eosinophil (e), lymphocyte (l), monocyte (m), and neutrophil (n). Arrows indicate platelets. Others elements are red cells.

A general blood microscope image is plotted in Fig 1. The major cells present in the peripheral blood are red blood cells (RBCs), and the white cells (leucocytes or WBCs). Leucocyte cells possessing granules are called granulocytes (which include neutrophil, basophil, and eosinophil). Cells without granules are called agranulocytes (lymphocyte and monocyte). The composition of leucocytes in human blood typically lies between the following values: neutrophils 50-70%, Eosinophils 1-5%, Basophils 0-1%, Monocytes 2-10%, lymphocytes 20.45% [1].

The ALL [4] disease is mainly associated to the lymphocytes in the bone marrow and into the peripheral blood. The colorant used in the formation of the blood tends to concentrate only in white cells, particularly in their nuclei that are center-positioned (the darker elements in Fig 2). In majority of the cases, the white cells are also larger than the red cells.

Generally, an automatic method for the detection of lymphoblasts in microscopically color images can be divided into the following steps as described below:-

1.1 Segmentation

The cells are differently classified from the background by using algorithms based on the various characteristics of the cells (e.g. shape, color, inner intensity).

1.2 Detection of white cells

The cells are divided into white blood cells and red blood cells. These classifiers can determine the presence of the nucleus by using color information.

1.3 Detection of lymphocytes

These lymphocytes can be identified from the other white blood cells by studying the shape of the nucleus.

1.4 Identification of candidate lymphoblasts

Candidate lymphoblasts can be distinguished in a set of lymphocytes by studying the morphological deformations of the cell.

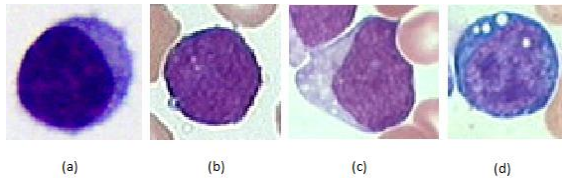


Fig 2: Morphological variability associated to the blast cells according to the FAB classification: (a) healthy lymphocytes cell from non-ALL patients, (b-d) lymphoblasts from ALL patients where (b), (c) and (d) are L1, L2 and L3 respectively.

ALL, the candidate lymphoblasts are classified as follows.

- **L1** - Lymphoblasts are small in size and homogeneous. The nuclei are round and regular with little clefting and unfavorable nucleoli. Cytoplasm is large in number and usually devoid of vacuoles.
- **L2** - Lymphoblasts are large in size and heterogeneous. The nuclei are not in well-formed shape and often clefted. Usually large nucleoli are present. The volume of cytoplasm generally varies, but often abundant and may contain vacuoles.
- **L3** - Lymphoblasts are moderate to large in size and homogeneous. The nuclei are in well-formed and round-oval in shape. One or more prominent nucleoli are present. The volume of cytoplasm is neither too less nor too much and it also contain prominent vacuoles.

The first approach as proposed in the literature [5, 6] is based on image segmentation and feature extraction technique. The input to the system is a microscopic digital image of the bone marrow smear. As this image contains different types of cells described above but the main focus is on white blood cells. So image segmentation is used to segment the WBCs. This is important to separate out the unwanted RBCs, blood platelets and other WBCs not under consideration currently. The image which is obtained is in a colored format and it is then converted to the grey scale image, as the processing of grey scale image is easier, this thus reduces the operational complexity. As the exact boundary of the nucleus is required for its geometrical properties, the sobel edge enhancing method is used for the further precise extraction of the nucleus. The nucleus is generally surrounded by the cytoplasm which overlaps the boundary of the nucleus, so in order to remove it, the opening morphological operation is used which includes erosion followed by dilation. The next step is removal of excess background portion to obtain the image such that its boundaries are tangential to the edges of the nucleus, which is done by cropping operation. The nucleus is further subjected to the geometrical properties of the nucleus such as Area, Centroid, Major Axis Length, Minor Axis Length, Eccentricity, Convex Area, Solidity, Perimeter Circularity, and Rectangularity. These properties form the feature vector which is used for constructing the training sets of cancerous and healthy cells. The training set is then used in classification of the input smear using a Euclidian Distance algorithm [1].

The second technique proposed is based on image transforms; in this paper Discrete Cosine Transform is used. The first technique focuses purely on the geometric properties of the nucleus. From analysis, it can be inferred that not only the geometric properties are essential but the texture and the consistency of the image is equally important. The transform based technique represents the changes in the frequencies in

the image and its energy. The process from obtaining the input image up to obtaining the cropped nucleus is identical to the first method. The image is scaled to the required resolution. Discrete Cosine Transform has been applied and the transformed image is stored as a feature vector, thereby constructing the training set. This training set is used for classification of cancerous cells.

The Section 2 provides a detailed description of the cancerous cell detection using image segmentation and extraction. Section 3 describes Discrete Cosine Transform based method.

2. CANCEROUS CELL DETECTION USING IMAGE SEGMENTATION AND FEATURE EXTRACTION TECHNIQUE [7]

Step 1: First the input image of the blood sample is fed into the system. This blood sample has basically four elements that are RBCs, WBCs, platelets and plasma. The amount of white blood cells in the blood gives us valuable information and plays a vital role in the identification of different diseases. So, the first module in the proposed system is white blood cell identification. The uncontrollable growth of lymphocyte is the main reason for leukemia. From these WBCs, lymphocyte cells are identified using segmentation method. As stated above the colored image is fed as an input to the system.

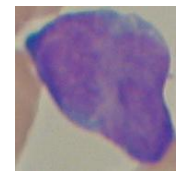


Fig 3: The basic input cell to be predicated

Step 2: This colored image is then converted into the grey scale image [2, 3]. Only the intensity information is carried by this image. It contains shades of grey varying from white which is considered to have the strongest intensity value and black, which has the least intensity value.



Fig 4: (a) Original image, (b) Grey-scaled image

Step 3: After obtaining the grey scaled image of the input object, sobel edge enhancing method is then used to intensify the line of demarcation of the cell nucleus.

Following sobel edge enhancement [2] opening operation is performed. In the opening operation, erosion operation is followed by dilation. The elementary effect of erosion on a binary image is to erode away the boundaries of regions of foreground pixels specially the white ones. These pixels thus shrink in size, and the holes which are present in those areas become larger. In order to perform erosion of a binary input image, all the foreground pixels in the image which is provided as an input are taken into account. For each of these foreground pixels, place the structuring element on top of the input image in order to see that the starting of the structuring element coincides with the coordinates of the input pixel. If for all the pixels in the structuring element, the corresponding pixel in

the image below is a foreground pixel, then the input pixel is left as it is without any changes. If any of the corresponding pixels in the image are background, then the input pixel is also set to be the same value as that of the background value. Fig. 4 shows the output after applying erosion,



Fig 5: (a) Original nucleus image (b) Eroded nucleus image

After erosion, perform the dilation operation. Generally erosion and dilation are performed in tandem. The main effect of dilation on any given set of binary image is to slowly enhance the boundaries of regions of foreground pixels which are basically the white pixels. The areas of the foreground pixels thus slowly grow in size while holes in those regions gradually become smaller. The dilation operator takes its input as two sets of data. The first set is the image which is to be dilated. The second is a small set of coordinate points which are referred as structuring elements. Fig 6 shows the output after dilation.



Fig 6: (a) Original nucleus image, (b) Dilated nucleus image

After this operation the cytoplasm has been removed and the desired nucleus is obtained.

Step 4: The next step is to crop the image in order to get rid of the excessive boundary portions. Cropping basically refers to the elimination of the outer parts of an image in order to improve the framing structure of that image. As only a specific part of the image is needed so discard the rest which is not useful, thus cropping operation has been performed. Fig 7 shows the nucleus after cropping.



Fig 7: (a) Original nucleus image, (b) Cropped nucleus image

Step 5: After cropping is the feature extraction of the nucleus. Feature extraction means extracting the different physical properties of the image on which that image can be classified. The cell is basically classified according to the physical properties. The shape and form of the nucleus are calculated with the help of many geometric functions. By using this, a complete feature vector containing the necessary information required for the physical structure of the nucleus is obtained. The different physical properties calculated are Area, Centroid,

Major Axis Length, Minor Axis Length, Eccentricity, Convex Area, Solidity, Perimeter Circularity, and Rectangularity.

These properties then help in the formation of the feature vector. This feature vector is then used to construct the training set of both cancerous as well as healthy cells. Now the feature vector of the cell to be classified is obtained using the same process as explained above. Each and every property of the feature vector which is stored is then subtracted from the corresponding property of the training set of cancerous cells as well as the healthy cells respectively. The deviation obtained from all of the corresponding values is made to square up and then the mean of the resultant value is calculated. This procedure is known as mean square error, which is formulated using the below given formula. In the formula given below, 'n' denotes the number of feature vectors the training set contains, y_i is the i^{th} feature vector and \tilde{y}_i is the current feature vector. The training set which then gives the least mean square error (MSE) is the training set to which that corresponding cell belongs.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad (1)$$

3. PROPOSED METHOD

Of the many different image transformation techniques, the main focus has been put on the Discrete Cosine Transform (DCT) method in order to detect the cancerous cells.

The discrete cosine transform (DCT) helps us to separate the image into different parts or which can be called as spectral sub-bands of varying importance taking into consideration the visual quality of the image. The Discrete Cosine Transform (DCT) has been proven to be acutely appropriate for a large class of images in the energy concentration and decorrelating. The DCT disintegrates the signal into underlying spatial frequencies, which then give the permission of further processing techniques to reduce the appropriateness of the DCT coefficients consistent with the Human Visual System (HVS) model. The DCT coefficients of an image define the ability to represent the regularity, complexity and some various other texture features of an image and in the compressed domain, it can be directly applied to image data. The DCT which is two dimensional can be expressed in terms of pixel values $f(i, j)$ for $i = 0, 1, \dots, N-1$ and $j = 0, 1, \dots, M-1$ and the coefficients of frequency-domain transform $F(u, v)$:

The basic equation for a 2Dimensional DCT is described by the following equation:

$$F(u, v) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} A(i).A(j) \cdot \cos\left[\frac{\pi u}{2.N}(2i+1)\right] \cos\left[\frac{\pi v}{2.M}(2j+1)\right] \cdot f(i, j) \quad (2)$$

Where,

$$A(\xi) = \frac{1}{\sqrt{2}}, \text{ for } \xi = 0$$

$$A(\xi) = 1, \text{ otherwise}$$

The basic operation of a DCT starts by taking the input image as $n \times m$. $f(i, j)$ basically represents the intensity of the pixel

in the i^{th} row and the j^{th} column and $F(u, v)$ represents the coefficient of DCT in the k^{th} row and the k^{th} column of the DCT matrix. For most of the images, the energy of the signal lies at a comparatively lower frequencies and these frequencies appear in the upper left corner of the discrete cosine transform. As the lower right values represent the higher frequencies, compression is achieved successfully. Generally all the 8 bit pixels have levels ranging from 0 to 255 values.

The output array of the coefficients of DCT includes integers that basically range from -1024 to 1023. The DCT generally concentrates the information, thus helping in making it useful for various image compression applications plus also helping in reducing the size of the feature vector. For a complete 2-Dimensional DCT for an $N \times N$ matrix image, the total number of multiplications necessary are N^2 and the total number of additions desired are $N(N-1)$. In this transform based technique, the variations in the frequency as well as the energy of the image are also taken into consideration for better and accurate results.

Steps for training cancerous and healthy cells:

- Step 1:** Convert the input colored image into a grey scale image.
- Step 2:** Sobel edge enhancement method for intensifying the line of demarcation of the nucleus is applied.
- Step 3:** Apply the opening operation.
- Step 4:** Perform cropping of the image in order to get rid of the excessive boundaries which are not needed.
- Step 5:** Scale the image to the required necessary resolution which is of size 128x128.
- Step 6:** Apply 2- dimensional DCT.
- Step 7:** Obtain the feature vectors of various sizes by gradually reducing it from the top-left corner from the calculated DCT coefficient as shown in the Fig 8.
- Step 8:** Step 1-6 are repeated for 100 cancerous and 100 healthy training images.

Steps for classifier:-

- Step 1:** Step 1-6 is the same as steps for training cancerous and healthy cells.
- Step 2:** Calculate the squared error between the DCT transformed image which is in question and the DCT transformed images of the known cancerous and healthy cells which had been previously stored in the feature vector. The formula for calculating the square error is as follows:-

$$e = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} [f(x, y) - f'(x, y)]^2 \quad (3)$$

Here, $f(x, y)$ is the feature vector corresponding to the test image and $f'(x, y)$ is the feature vector obtained from the training set and thus two squared errors are obtained corresponding to cancerous and healthy training sets.

- Step 3:** Calculate the mean for both the squared errors. Compare these two mean values and classify.

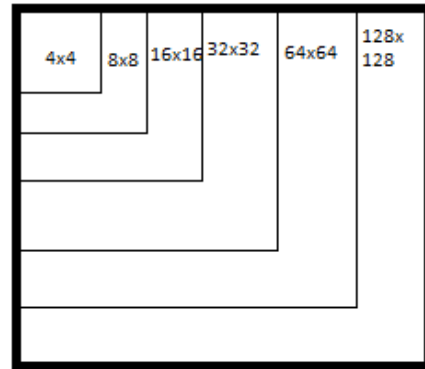


Fig 7: Image representing feature extraction from various sized DCT partial coefficients.

4. RESULT

For the experimental results, the database of cancerous and healthy cells has been taken from Acute Lymphocytic Leukemia Image Database (ALL IDB)[8,9] which consist a total of 260 smear images, 130 of which are cancerous and the remaining 130 are healthy. In the training set, the feature vectors of 100 cancerous and 100 healthy cells are stored and the training set was tested on the remaining 60 cells.

Table 1 shows the Percentage accuracy and precision of Cancerous cell detection using image segmentation and feature extraction method and DCT average of cropped grey scale image for different partial DCT transformed coefficients.

Table 2 shows the Computational Complexity required for computing e given in equation 3 for different feature vector selected using DCT Average Method.

Fig 8 shows the comparison of accuracy and precision between the Existing method and the proposed method.

Table 1 - Percentage accuracy of Cancerous cell detection using image segmentation and feature extraction technique and DCT average of cropped grey scale image

Methods		% Accuracy	%Precision
Image segmentation and feature extraction technique		78.33	72.97
Portion of the feature vector selected using Proposed DCT Average Method	4x4	80	71.42
	8x8	83.33	76.31
	16x16	85	78.37
	32x32	85	78.37
	64x64	85	78.37
	128x128	85	78.37

- Computational complexity for computing e given in equation 3
- No of subtractions required for $N \times N$ feature vector is $N * N + (N * N - 1)$
- No of Multiplications required for $N \times N$ feature vector is $N * N$.

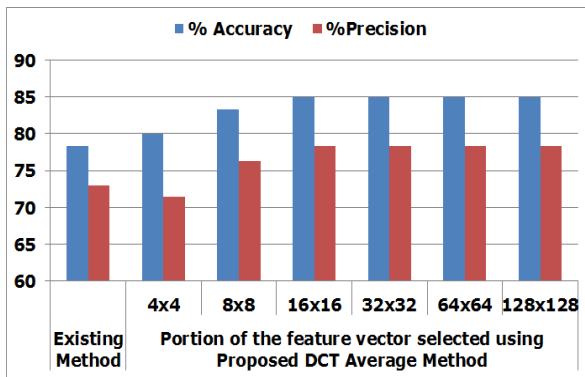


Fig 8: Comparison of Existing method and proposed method

Table 2 – Computational Complexity required for computing e given in equation 3 for different feature vector selected using DCT Average Method

Feature Vector size	No. of Subtractions	No. of Multiplications
128x128	32767	16384
64x64	8191	4096
32x32	2047	1024
16x16	511	256
8x8	127	64
4x4	31	16

From the Table 1 it is observed that proposed method gives 85 percent accuracy and 78.37 percent precision for the partial DCT coefficients of sizes 16x16 to 128x128. Further the reduction in computational complexity for 16x16 feature vectors is 64 times than that of 128x128 feature vectors. Hence 16x16 feature vector size could be considered as optimal feature vector.

5. CONCLUSION

The paper presents the different approaches used for the detection of cancerous cells. In the first approach which is based on the geometrical and the physical properties according to the literature is similar to the already in use approach to manually examine the detection of cancerous cells. The proposed method is based on extracting the texture of the image, which is done by using the Discrete Cosine Transform method. From the results it is clearly obvious that the later one produces better results and is more accurate by 8.5% than the former one.

6. REFERENCES

- [1] Fabio Scotti, "Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images", CIMS 2005 – IEEE International Conference on Computational Intelligence for Measurement Systems and Applications Giardini Naxos, Italy, 20-22 July 2005
- [2] J.S. Lim "Two dimensional signal and image processing" Prentice Hall 1990
- [3] R. C. Gonzalez, R. E. Woods, S.L. Eddins, "Digital Image Processing Using MATLAB", Pearson Prentice Hall Pearson Education, Inc., New Jersey, USA, 2004
- [4] F. Cillesen, W. Der Meer, "Atlas of Blood Cell Differentiation", Elsevier Science B.V., Amsterdam, The Netherlands, 1998

- [5] Serbouti S., Duhamel A. Harms, H. Gunzer U. Aus, H.M. Mary J.-Y, Beuscart R., "Image segmentation and classification methods to detect leukemias", Annual international conference of IEEE engineering in Medicine and Biology society, Vol.13, No.1, 1991
- [6] Foran DJ, Comaniciu D, Meer P, Goodell LA "Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy", IEEE Trans. Inf. Technol. Biomed., pp. 265-73, December 2000
- [7] Piuri, F. Scotti: "Morphological Classification of Blood Leucocytes by Microscope Images" Proc. International Symposium on Computational Intelligence for Measurement Systems and Applications, Boston, MD, USA, July 2004.
- [8] R. Donida Labati, V. Piuri, and F. Scotti, "ALL-IDB: the Acute Lymphoblastic Leukemia Image Database for Image Processing", in IEEE International Conference on Image Processing (ICIP), September 11 - 14, 2011
- [9] Ms. Minal D. Joshi, Prof. Atul H. Karode, Prof. S.R.Suralkar "White Blood Cells Segmentation and Classification to Detect Acute Leukemia" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)" Volume 2, Issue 3, May – June 2013.

7. AUTHOR PROFILE

Dr. Tanuja K. Sarode has Received Bsc.(Mathematics) from Mumbai University in 1996, Bsc.Tech.(Computer Technology) from Mumbai University in 1999, M.E. (Computer Engineering) degree from Mumbai University in 2004, Ph.D. from Mukesh Patel School of Technology, Management and Engineering, SVKM's NMIMS University, Vile-Parle (W), Mumbai, INDIA. She has more than 17 years of experience in teaching and currently working as Associate Professor in Department of Computer Engineering at Thadomal Shahani Engineering College, Mumbai. She is life member of IETE and ISTE. Her areas of interest are Image Processing, Signal Processing and Computer Graphics. She has more than 160 papers in National /International Conferences/journal to her credit.

Bhavik K. Thakkar is currently pursuing B.E in Computer engineering from Thadomal Shahani Engineering College. He has keen interest in networking. He aims to complete his B.E degree and then wishing to pursue a Masters degree in the same field.

Sanket J. Purandare is currently pursuing B.E in Computer engineering from Thadomal Shahani Engineering College. He has keen interest in designing algorithms and programming. He aims to complete his B.E degree and then wishing to pursue a Masters degree in the same field.

Vikas M. Gupta is currently pursuing B.E in Computer engineering from Thadomal Shahani Engineering College. He has interest in coding and teaching. He aims to complete his B.E degree then wishes to pursue a Masters degree in the same field.