Computer Aided Leukemia Detection using Digital Image Processing Techniques

Shaikh Mohammed Bilal N, Department of Computer Engineering, University Of Mumbai, Bilals.sb@gmail.com
Sachin Deshpande, Department Of Computer Engineering, University Of Mumbai, Sachin.deshpande@vit.edu.in

***Abstract*—The machine-controlled identification of malignant cells from microscopic diagnostic assay pictures of blood samples helps in assuaging the diagnostic problems of leukemia and provides higher results if the biologically explainable and clinically important feature primarily based approaches are used for the identification of malignancy and the severity of the disease. Identification method might have error rates up to 40%; Due to this drawback it is necessary to develop a second opinion tool which helps the doctors and pathologists in a classified detection of the leukemia disease. The analysis shows a discourse method for leukemia and its subtypes detection from microscopic blood cell images which can be done by applying image processing techniques followed by a detailed classification analysis on the dataset images.**

Keywords—leukemia, detection, segmentation, feature extraction, classification, subtypes.

I. INTRODUCTION

Leukemia is a type of cancer that originates in the bone marrow of the subject. It occurs as a result of large production of immature leucocytes which override the normal blood cells (WBC, RBC, and platelets). The body gets exposed to multiple other diseases due to the decreased WBC efficiency to fight against the same. within the diagnosing of cancer, additionally to think about the signs and symptoms of the subjected patient, it's needed to identify the malignant or blast cells. To get the amount of the blood cells in every unit volume a blood count is calculated for different types of cells in blood(RBC, WBC, and platelets). For the determination of the number of malfunctioning leucocytes, a bone marrow cellular analysis is carried out after looking for abnormalities in the blood count. A diagnostician examines the microscopic cell images in correlation to a research to determine the abnormalities in the leucocytes in order to observe the occurrence of blood cancer and further its type and subtype. Classification decides the prescription for the subjected patient of blood cancer, hence it is very important. Although Flow Cytometry is used by various pathologists as a dependable mechanism for blood cancer diagnosis, still in many medical institutes an hospitals it is not feasible, mainly the public sector medical hospitals, as described in [11]. During this work, the type and subtype of Acute Leukemia can be identified by working upon the information provided by the microscopic blood cell images of the dataset of different human subjects or patients.

II. PROBLEM DEFINITION

A. Statement Of Problem

To analyze blood cells digital pictures from blood smear samples, then perform a discourse approach for the detection and classification of Acute leukemia subtype after determining the type of Acute leukemia.

Acute leukemia is of two types namely, Acute Lymphoblastic Leukemia (ALL) and Acute Myeloblastic Leukemia (AML). Further, acute leukemia is divided into the subsequent subtypes:

1. ALL subtypes are L1, L2, and L3.
2. AML subtypes are M0, M1, M2, M3, M4, M5, M6, and M7.

B. Exiting Detection Systems

Because of the irregular staining and overpopulated cell smears the segmentation of cytoplasm and nucleus is difficult. SVM along with Discrete Fourier Transform and segmentation was used in some works, which in turn provided an automated learning mechanism to differentiate the blood smear image specifications (i.e cytoplasm cells, nucleus and platelets). This method is more acceptable and efficient in contrast with the "Tresholding and the Watershed Algorithms", as in [1]. Moreover, a similar version of this method named as "Simulated visual attention via learning by on-line sampling" is projected in [2]. Due to the issue of overlapping cells in blood images some algorithms were introduced. They usually worked at splitting and merging cells at the edges, as shown in [3]. Detailed classification was deployed by Mohapatra et. al. using various classifiers to classify the healthy and cancerous cells, which is stated in [4]. Lim et. al. projected on a method thereby showcasing "thresholding, morphological operations, and watershed mechanism" as shown in [5].

978-1-5090-3774-9/17/$31.00 © 2017 IEEE

III. LITERATURE SURVEY

Table 1

Sr. No.	Paper	Author	Advantages	Drawbacks
1.	**"Computer Aided Detection of Skin Cancer" [7][16]**	Aswin R. B. and Abdul J.	Hybrid Classification Approach, Effective feature extraction [7].	Cell overlapping not addressed, Can be used for detecting melanoma only, Accuracy is comparatively less
2.	**"Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features" [6]16]**	Rajesh K.	Texture, shape and morphology, HOG, wavelet color, Tamura's feature, and LTE used innovatively with an accuracy of 92 % [6].	It was observed that the proposed method is performing better for connective tissues type sample and was not tested for leukemia blood smear samples.
3.	**"Detection Of Skin Cancer Using Hybrid of SVM-ID3 Algorithm" [13]**	Greeshma Rajan1, G Shivaraj	Cell segmentation, Third Harmonic Generated Microscopy, SVM, ID3, Hybrid of ID3-SVM Algorithm, Nucleus-to-Cytoplasm (NC) ratio implemented in a simple effective manner [13].	Cell overlapping not addressed, Texture features not addressed.
4.	**"Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms" [14] [17]**	Ammara M.	Detailed comparison and classification of various models used in diagnosis of cancer such as ultrasound and optic coherence tomography [14].	Limited morphologic Information. Blood cancer not involved.

As shown in Table 1, Aswin. R.B and J. Abdul et al. in "**Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features"** 2014(IEEE) used a hybrid GA ANN classifier after applying image pre-processing (dull razor & filtering), segmentation (color threshold seg) and feature extraction (grey level co occurrence matrix-GLCM) techniques on images of skin for melanoma detection with an accuracy of 88% [7].

Rajesh Kumar in his paper "**Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features**" 2015 showed an approach to detect and classify cancer from blood cell images specific pathological features, which is shown in [6].

For segmentation of images color *k*-means based method is used. The various hybrid features which are extracted from the segmented images include shape and morphological features, GLCM texture features, Tamura features, Law's Texture Energy based features, histogram of oriented gradients, wavelet features, and color features. For classification purposes, *k*-nearest neighbor based method was proposed to be used with an average accuracy of 92.19%.

IV. PROPOSED SYSTEM

A. Issues Focused

We propose cell classification model using various cellular morphological features. For this purpose, it is necessary to carry out pixel level pre processing and the separation of cytoplasm and nucleus. This approach considers segmentation as an important task for the differentiation of blood cells.

978-1-5090-3774-9/17/$31.00 © 2017 IEEE

Segmentation proves to be accurate and has effective results on such medical diagnostic machine based approach. Hence as an option we use a cell segmentation procedure that considers colors, textures, and other features even in overpopulated cell smears of blood at the pixel level image evaluation thereby extracting the cytoplasm and nucleus in the same. The cytoplasm background cells and nucleus differentiation makes this work enhance its results in contrast with other works.

B. *Proposed Methodology*

The proposed mechanism is as follows:

1) Differentiating the Cytoplasm (background) and Nucleus(foreground) :

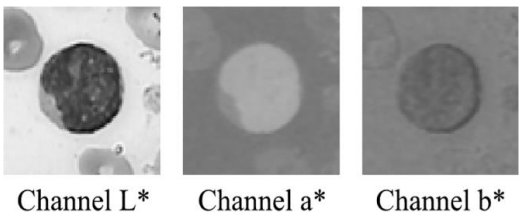


Fig. 1. L A B color space by CIE for the dataset image.

Fig. 1 shows the three channels for a same blood cell dataset image We set to use the 'CIE l_a_b_color format' as in [18]. we have a tendency to took this call as it marks the changes in color groups. Moreover it provides a color differntiation feature based on senses and has high results of evalualation as shown in [12]. As the dataset's blood cell images are in RGB color space, it has to converted to CIE L a b space using mathematical rules.

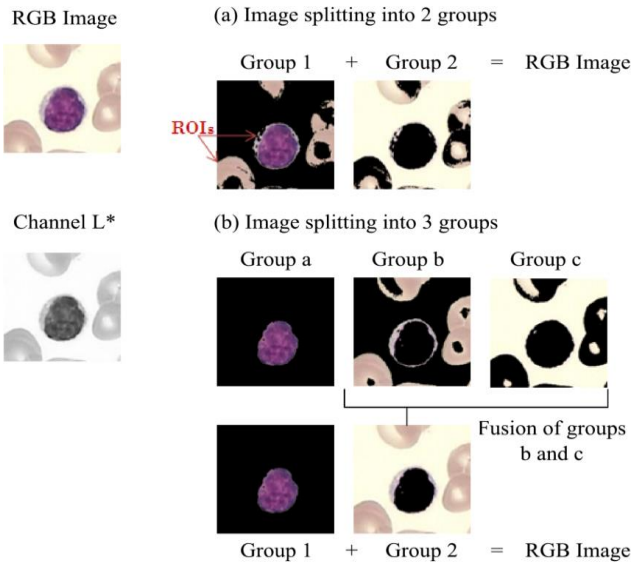


Fig. 2 Forming groups by the similarity features from channel intensity

Channel L is called the luminosity channel which focuses on the brightness or luminosity constrain of the image with specific distinctive reflections as shown in Fig 2. Color description of cells is done by channel B, which marks the presence of blue and purple shades. A clustering approach can be used keeping in picture the channel L and B values for different cells for producing two groups having values of k as 2 and 3 respectively as shown in fig 2.

2) Finding overlapping cells in highly populated smears :

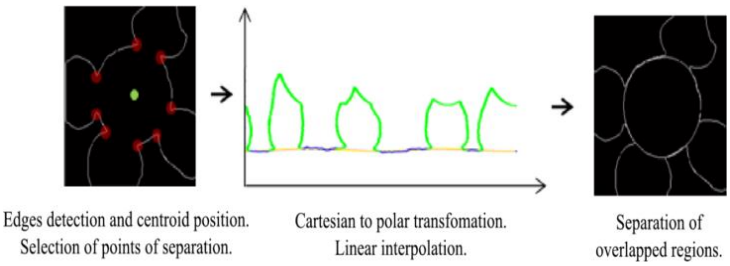


Fig. 3. Cell Separation process

Fig 3 shows a cell separation procedure. In some works the overlapped region was identified initially and the separation was done by getting the edges from the region of interest and then finding the midpoint surrounded by the curved portions as the portion of operation as shown in [3]. Then, we carry out linear interpolation or Discrete Fourier Transform to create a graph of the discontinuities to determine overlapping cells in the smear image followed by morphological operations in MATLAB.

3) *Wold's decomposition:*

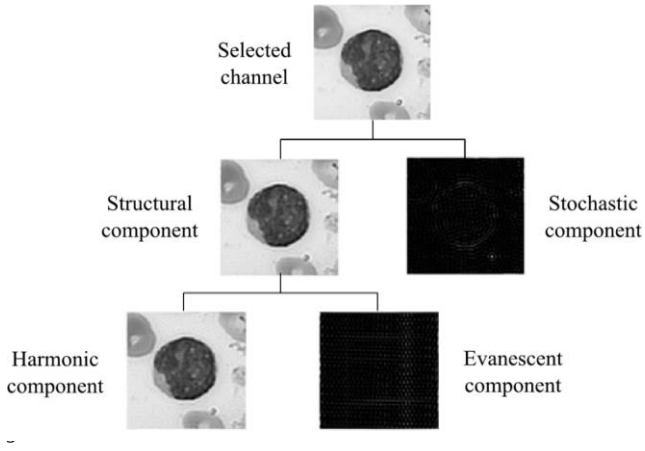


Fig 4. Wold's Decomposition Procedure.

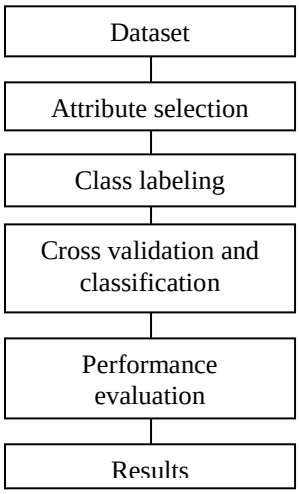
We need to determine the texture specifications of the dataset image of the blood cells. For this, 'Wold's Decomposition' can be used. It consists of fields for texture evaluation of the image, namely a) Harmonic Field, b) Stochastic Field and c)

978-1-5090-3774-9/17/$31.00 © 2017 IEEE

Evanescent Field, as stated in [9]. The texture function is the total of evanescent and harmonic elements, generalized. As in Fig 4, It can be seen that, in order to initialize the harmonic field, it has to be followed by a sine transformation by DFT or FFT for the specific channel of all orthogonal components. Hough transform along with DFT can be used to activate the evanescent field, generating four evanescent lines. The three fields correspond to three texture specifications as regularity, direction constrain, and random variation constrain.

4) *The Classification Process*

The following flowchart represents the classification paradigm of the system.



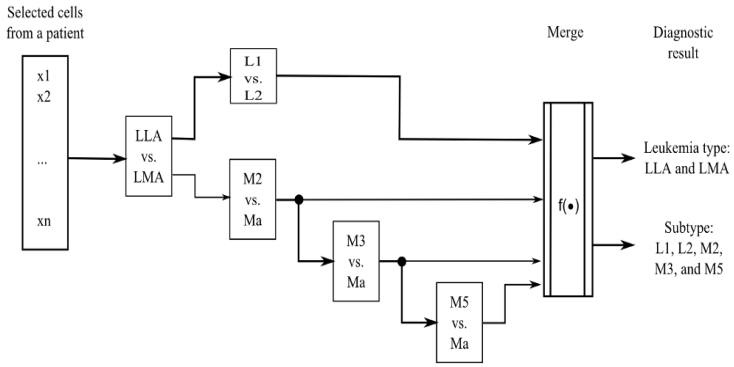


Fig 5 Classification model

Feature extraction of the cells contained in the cytoplasm as well as the nucleus leads to the classification of blood cells in terms of malignant cells and healthy cells, furthermore determining its subtypes, as shown in [6]. In this section, we evaluate the different features such as geometry, color, texture, size and statistics from the portions of interest popped up after image segmentation process, and we use them to spot the leukemia type followed by its subtype. The following table 2 shows some of the distinguishing features the analysis of whose can be done using different training sets, selection of attributes and classification programs supported in WEKA [8].

Table 2

Sr. no	Feature	Type	Information
1	Area	Geometry	Total no. of pixels in the Region Of Interest.
2	Circularity	Geometry	$Perimeter^2/4\pi area$
3	Eccentricity	Geometry	Distance(centre, focus)
4	Elongation	Geometry	Width/length of ROI.
5	Mean	Statistics	Avg. no. of pixels value in ROI.
6	Mode	Statistics	Highest Frequency value of pixels in ROI.
7	Standard Deviation	Statistics	Standard deviation of pixels value in ROI.
8	Sum	Statistics	Total of pixels value in ROI
9	Contrast	Color	Intensity variation between a pixel and its neighborhood pixel over the image
10	Correlation	Color	Similarity constrain between two neighbor pixels.
11	Entropy	Texture	Smoothness determined by gray levels.
12	Homogeneity	Texture	Correlation between the elements in co occurrence matrix to its diagonal elements.

Fig. 5 shows an easy way for developing an acute leukemia diagnosing algorithm supported by a mixture of coupled classifiers using SVM (Support Vector Machine) along with ANN (Artificial Neural Network). SVM will be functioning with the majority voting approach to determine the type and subtype of Acute Leukemia along with ANN since the dataset needs to be trained effectively to reduce the chances of error. In each step we perform classification to distinguish between class of acute leukemia (ALL and AML), subclasses of ALL (L1 and L2), subclasses of AML (M2, M3, M5). For ALL subclass, we can perform binary as well as multiclass classification to analyze the features of subtype. By Fusing different Classifiers (ANN and SVM) and the use of demonstrative features of nucleus and cytoplasm cells it is possible to reduce the number of false positives and false negatives in both classifiers.

V. OBJECTIVE

Increase efficiency & accuracy (more than 90%) thereby decreasing time complexity of each task in the algorithm i.e. using a simple fast algorithm such as DFT or FFT along with above discussed procedures in MATLAB for image processing, SVM (Support Vector Machine) and ANN (Artificial Neural Network) for classification using WEKA instead of using algorithms that increase complexity as shown in [8] and [10] respectively. Classified diagnosis leading to sublevel detection of the acute leukemia disease by applying image processing methods using MATLAB followed by the

978-1-5090-3774-9/17/$31.00 © 2017 IEEE

classification and subtype classification using SVM along with ANN.

VI. FIELD WORK

The **TATA Memorial Hospitals (TMH**), Mumbai was visited by me as part of the project detailing and research and analysis of the topic.

It was reported by the senior pathologist **Dr. Mani Subramanian** of the **Hematology** department that for the blood cells count and determination of abnormalities they used **flow cytometry** method implemented by Beckman Coulter imported by- Beckman Coulter, Inc.

The **COULTER VCS** uses unit volume, conductivity and scatter light parameters for the determination of WBC and its diagnosis, which is shown in the manual at [11], and elaborated as follows:

Volume Analysis: The unit volume of the individual blood cells is evaluated by the Beckman Coulter using very low frequency current. This method has been used since ages to evaluate the WBC count in a blood sample, which is further processed for the diagnosis of leukemia.

Conductivity Analysis: High frequency current passes easily through cell walls of the blood. As they are good conductors, the evaluation of insulation at the walls of the cell as well as the specifications of the nucleus components along with the chemical uptake and traces inside the cell is evaluated.

Light Scatter Analysis: Scattering of laser light throughout the population of cells contained in a blood sample of a patient also known as Fulwyler's method for cell analysis, is deployed in Beckman Coulter so as to determine the blast cell count, the containment proportion of the cells and the overpopulated regions.

VII. DATASET

As urged by the doctors of the Tata Memorial Hospital, Mumbai; the request to accumulate the dataset was sent to Fabio Scotti, University of Milano, Italy. The database ALL IDB was gathered that consisted of ALL_IDB1 and ALL_IDB2 datasets consisting of pictures that were captured by optical laboratory magnifier as well as a Canon PowerShot G5 camera. The resolution of the JPEG format 24 bit pictures is 2592 x 194, which is stated in [15].

ALL_IDB1 consists of 108 pictures (healthy and malignant) having 39000 blood components in which the lymphocytes are labeled by skilled oncologists.

ALL_IDB2 dataset is for testing and training the classification system. It's a set of separated regions to be considered pertaining to healthy and leukemia cells from ALL_IDB1 dataset.

Acknowledgment

I wish to offer my sincere gratitude to our principal Dr. S.A. Patekar for providing me an opportunity to be a part of the Vidyalankar Institute of Technology, Mumbai under the postgraduate program and moreover to allow me work on this final year Project on the respective domain.

I sincerely thank my Guide Prof. Sachin Deshpande for the guidance and encouragement in carrying out this work.

I also wish to thank the Expert Panel members who carried out the reviews on this seminar topic for myself and other students and rectified the mistakes in the same for the betterment of our work.

I would like to offer gratitude to Dr. Subramanian and the other pathologists of the Hematology Department, TMH, Mumbai for their support and information gathering.

References

[1] Pan C, Lu H, Cao F. "Segmentation of blood and bone marrow cell images via learning by sampling." Emerging Intelligent Computing Technology and Applications, Springer Berlin Heidelberg. 2009; 336–345.

[2] Pan C, Park DS, Yoon S, Yang JC. "Leukocyte image segmentation using simulated visual attention." Expert Systems with Applications. 2012; 39(8): 7479–7494.

[3] Wang W, Hao S. Cell cluster image segmentation on form analysis. IEEE Third International Conference on Natural Computation. 2007; 4: 833–836.

[4] Mohapatra S, Patra D. "An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images." Neural Computing and Applications. 2014; 24(7–8): 1887–1904.

[5] Lim HN, Mashor MY, Hassan R. "White blood cell segmentation for acute leukemia bone marrow images." IEEE International Conference on Biomedical Engineering (ICoBE) 2012; 357–361.

[6] Aswin.R.B and J. Abdul "Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features." 2014(IEEE).

[7] Rajesh Kumar," Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features" Hindawi Publishing Corporation Journal of Medical Engineering ,Volume 2015, Article ID 457906, 2015.

[8] Weka—Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update; ACM SIGKDD Explorations Newsletter. 2009; 11(1): 10–18.

[9] Liu F, Picard RW. A spectral 2D Wold decomposition algorithm for homogeneous random fields. IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing. 1999; 6:

[10] 3501–3504.Pankaj Agrawal, S.K.Shriwast ava and S.S.Limaye, "MATLAB Implementation of Image Segment at ion Algorithms", IEEE Pacific, pp.68–73.

[11] COULTER HmX Hematology Analyzer, PN 4237523CB (March 2011) Beckman Coulter, Inc.250 S. Kraemer Blvd.Brea, CA 9282.

[12] Paschos G. Perceptually uniform color spaces for color texture. IEEE Transactions on Image Processing. 2011; 10(6): 932–937.

[13] Greeshma Rajan et al, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.5, May- 2014, pg. 143-153.

[14] Ammara Masood, "Computer Aided Diagnostic Support System for Skin Cancer:A Review of Techniques and Algorithms" Hindawi Publishing Corporation, International Journal of Biomedical Imaging Volume 2013, Article ID 323268.

[15] F. Scotti "ALL_IDB: the acute lymphoblastic leukemia image database for image processing", 2011 IEEE, Brussels, Belgium, pp. 2045-2048, ISBN: 978-1-4577-1302-6.

[16] www.scilit.net

[17] www.semanticscholar.org / http://dx.doi.org/10.1155/2013/323268

[18] C. Reta. " Leukocytes Segmenta ion Using Markov Random Fields" , Advances in Experimental Medicine and Biology, 2011.