ISSN:0975-9646

K.Dhana Shree et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4770-4773

A Survey for detecting Acute Lymphoblastic Leukaemia using Image Processing Techniques

K.Dhana Shree[1], M.Saranya[2], G.Poorani[3], Dr.A.Grace Selvarani[4]

PG Scholar, Dept of CSE (PG), Sri Ramakrishna Engineering College, Coimbatore[1, 2, 3]
Head of the Department, Dept of CSE (PG), Sri Ramakrishna Engineering College, Coimbatore[4]

Abstract: Leukaemia is a type of blood cancer that causes severe damage to humans if left untreated. Blood microscopic images usually consist of Red Blood Cells(Erythrocytes),White Blood Cells(Leucocytes),Platelets(Thrombocytes) and Plasma. White Blood Cells can be divided into different types and sub types. The major types of white blood cells are granulocytes Monocytes and lymphocytes. The granulocytes are further divided into Neutrophils, Eosinophils and Basophils. The monocytes are further divided into dendritic cells and macrophages. The lymphocytes are further divided into B-lymphocytes and T-lymphocytes. Many image processing techniques for WBC detection, segmenting, identifying grouped leucocytes and extracting their features have been proposed in this paper.

Keywords: Acute Lymphoblastic Leukaemia (ALL)

I. INTRODUCTION

Acute lymphoblastic leukaemia, sometimes called as acute lymphocytic leukaemia or acute lymphoid leukaemia (ALL) is an acute form of leukaemia, or cells. It is characterized by the overproduction and accumulation of cancerous, immature white blood cells, called as lymphoblast. The lymphoblast are produced more in the bone marrow and they multiply continuously forming more number of immature cells, causing damage and death by inhibiting the production of normal cells (such as red and white blood cells and platelets) in the bone marrow .They further spread to other organs in our body. ALL is mostly common among children of age 2-5 years than in adults.

The symptoms of ALL are indicative of a reduced production of functional blood cells, because leukaemia wastes the resources of the bone marrow that are normally used to produce new, functioning blood cells. These symptoms can include fever, increased risk of infection (especially bacterial infections like pneumonia, due to neutropenia; symptoms of such an infection include shortness of breath, chest pain, cough, vomiting, changes in bowel or bladder habits), increased tendency to bleed (due to thrombocytopenia), and signs indicative of anaemia, including pallor, tachycardia (high heart rate), fatigue, and headache[1].

Leukaemia can be cured if it is detected in the early stage and treated. The need for automation of leukaemia detection arises since current methods involve manual examination of the blood smear as the first step toward diagnosis. This consumes more time and also its accuracy is greatly dependent on operator's ability. ALL can be treated using chemotherapy, immunotherapy, and biological therapy but before treating ALL must be detected. ALL can be detected automatically and this avoids use of manual operations.

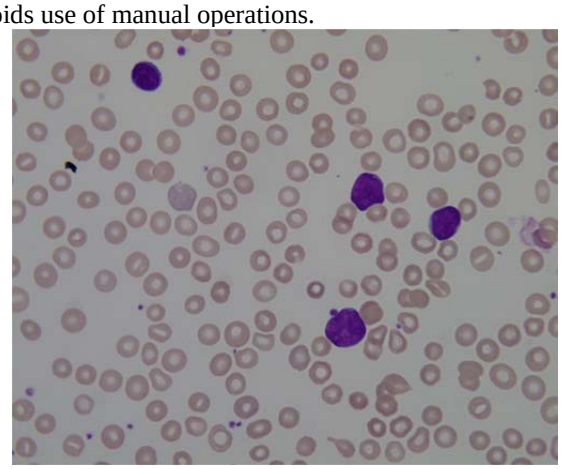


Fig.1: A blood sample suffering from Acute Lymphoblastic Leukaemia

The main advantages of using automated techniques are:
1. It reduces the computational time.
2. No need for skilled operators.
3. The overall accuracy and efficiency can be improved.

Acute Lymphocytic Leukaemia detection consists of the following steps:-
1. Image pre-processing
2. Leucocyte Identification
3. Nucleus and cytoplasm selection
4. Feature extraction
5. Classification

Usually leukaemia is detected by performing the complete blood count [2]. If the count is found to be abnormal, study of morphological bone marrow smear is carried out to confirm the leukaemia [3]. Two main analyses are done for detection: cell classification and counting. Blood sample is not required for morphological analysis and hence is suitable for low cost accurate and remote screening system, as it just needs an image [4].

II. IMAGE PROCESSING TECHNIQUES FOR DETECTING ACUTE LYMPHOBLASTIC LEUKAEMIA

A.Image Pre-Processing

Image pre-processing is a technique that is usually carried out for all the images before processing. The image to be

processed is first converted into CMYK image from RGB image. From CMYK the Y component is alone extracted because yellow colour is present in all elements of the image, except leucocytes. Once image has been converted into CMYK model image has to be segmented and cleaned. The Histogram or contrast stretching is applied to CMYK model followed by segmentation.

The desired portion of the image for further processing can be achieved using segmentation. In this WBCs are extracted to check whether it is affected or not. For this, K-means clustering algorithm is used. It is the one of the unsupervised learning algorithm.

- K-means algorithm
  - Initialize the center of the clusters $\mu_i = some\ value, i = 1,..,k$
  - Attribute the closest cluster to each data point

$$c_i = \{j: d\ (x_j, \mu_i) \leq d(x_{j,}\ \mu_l), l \neq i, j = 1, \dots, n\}$$

  - Set the position of each cluster to the mean of all data points belonging to the cluster

$$\mu_i = \frac{1}{|c_i|}\sum_{j \in c_i} x_j\ , \forall_i$$

  - Repeat the steps 2-3 until convergence

Various segmentation algorithms like threshold otsu method [13], automated histogram thresholding [14], snake balloon algorithm [15], Fluorescence in-situ hybridization (FISH) method [16], immune phenotyping, cytogenetic analysis and cytochemistry can be used. Further background of the image must be removed [5].

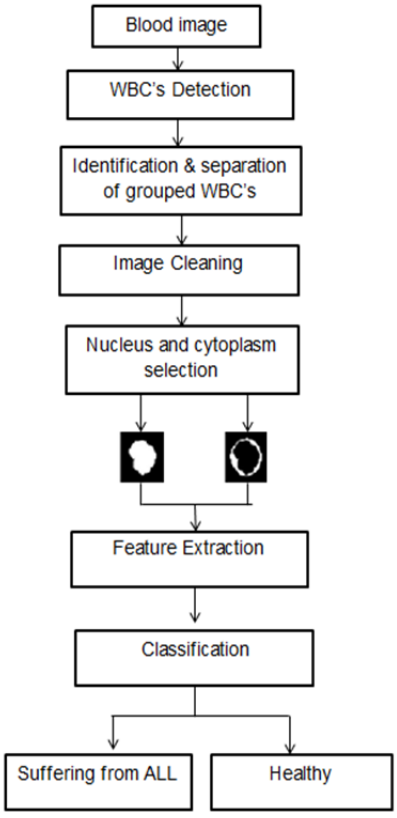


Fig.2: Steps involved in detecting Acute Lymphoblastic Leukaemia

B.Leucocyte Detection

Several methods can be used to verify the presence of adjacent leucocytes [6]**.**the presence of adjacent cells can be identified using shaped. Leucocytes are usually round in shape and using roundness value the adjacent cells are identified. Circularity can be measured using roundness (1) that excludes local irregularities and can be obtained as the ratio of the area of an object to the area of a circle with the same perimeter of the convex hull of the object**.**

$$\text{Roundness} = \frac{4*\pi*Area}{convex\ perimeter^2} \quad (1)$$

The roundness value equals to 1 indicate that it is a circular object and value greater or lesser than 1 deviates from circular object.so a threshold value of 0.80 can be used to discriminate single leucocyte from group of leucocytes. Connected components with a roundness value greater than the threshold are classified as individual leucocytes while connected components having a roundness value smaller than the threshold are classified as grouped leucocytes and these groped leucocytes must be separated. Using [7] and [8] the grouped leucocytes are extracted from the original image by cutting a square around the previously segmented nucleus. Assuming each sub-image has a single WBC, clustering around the nucleus is performed using shape and colour information. Using distance transform, the distance of adjacent cells is found and it is segmented using any one of the segmentation methods. Watershed segmentation [9] is then applied to the distance transform to yield a rough separation between adjacent leucocytes. In this way, the separation tends to be inaccurate because it uses the distance transform as a form delimiter. Thus, it only performs well in the presence of adjacent leucocytes with an early rounded shape, but it does not perform well in the presence of multiple complex forms.

Different approaches may be employed to use the watershed principle for image segmentation.

- Local minima of the gradient of the image may be chosen as markers, in this case an over-segmentation is produced and a second step involves region merging.
- Marker based watershed transformation make use of specific marker positions which have been either explicitly defined by the user or determined automatically with morphological operators or other way.

C.Image Cleaning

Image cleaning requires the removal of all of the leucocytes located on the edge of the image.The abnormal components(non-leucocytes)in blood sample are also removed.Solidity is defined as the ratio of the area of an object to the area of a convex hull of the object:

$$\text{Solidity} = \frac{area}{convex\ area} \quad (2)$$

A solidity value of 1 indicates a solid object, and a value less than 1 indicates an object with an irregular boundary or cells containing holes. The solidity value used for the

threshold is calculated directly from the image containing only the individual leucocytes, and when this image is empty, a default value of 0.90 is used.

D. Nucleus and cytoplasm selection

Automatic image selection is done using bounding box size and each individual leucocyte is isolated.Leucocyte nucleus is inside the membrane and cropping is done to the entire portion of the image outside the leucocyte. Cseke's [10] demonstrated that WBC nuclei are more in contrast on the green component of the RGB colour space that helps to select the nucleus easily. However, in this colour space, the threshold operation described by Otsu [11] does not produce clean results, especially in the presence of granulocytes, because granules are selected erroneously as part of the nucleus. The binary image obtained from the green component is combined with the binary image obtained from the a* component of the CIE Lab colour space via a threshold operation is performed to avoid this issue. The mask obtained allows us to clearly extract the leucocyte nucleus. A subtraction operation is performed between the binary image containing the whole leucocyte and the image containing only the nucleus is performed to obtain cytoplasm finally.

E.Feature Extraction

The input data is transferred into set of features. Various features of the cells are studied to identify Leukaemia. Some of these features are: color, contour signature, statistical features, geometry, texture, etc. the shape descriptors, such as area, perimeter, convex area, convex perimeter, major axis, minor axis and orientation are extracted. Elongation, eccentricity, rectangularity, compactness, convexity, roundness (1) and solidity (2) can be calculated from the above measures. Elongation (3) measures how an object is elongated. Eccentricity (4) is the ratio of the distance between the foci of the ellipse and its major axis length; this value is between 0 and1. Rectangularity (5) represents how rectangular a shape is (i.e., how well it fills its minimum bounding box). Compactness (6) is defined as the ratio between the area of an object and the area of a circle with the same perimeter; the maximum value is 1 for a circle. Convexity (7) is the relative amount that an object differs from a convex object, and this value represents the ratio of the perimeter of an object's convex hull to the perimeter of the object itself; the value is 1 for a convex object and less than 1 if the object is not convex, such as an object with an irregular boundary. These measures can be defined as:

$$\text{Elongation}=1-\frac{minor\ axis}{major\ axis} \quad (3)$$

$$\text{Eccentricity}=\frac{\sqrt{(major\ axis^2-minor\ axis^2)}}{major\ axis} \quad (4)$$

$$\text{Rectangularity}=\frac{area}{major\ axis \times minor\ axis} \quad (5)$$

$$\text{Compactness}=\frac{4\times\pi\times Area}{perimeter^2} \quad (6)$$

$$\text{Convexity}=\frac{Perimeter_{convex}}{perimeter} \quad (7)$$

F.Classification

The cells are further classified to know whether it is affected leukaemia cells or the normal cells. The classification model chosen for this phase is the Support Vector Machine SVM, since this model is particularly suitable for binary classification problems, in which the separation between the classes depends on a large number of variables. In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithm that analyse data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked for 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.

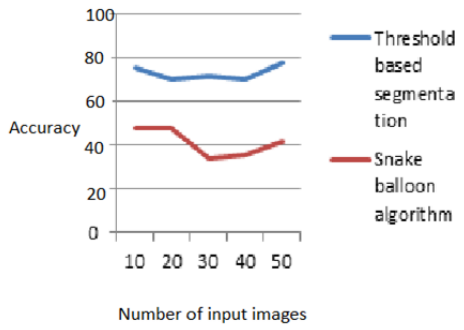
G.Dataset

ALL-IDB is a public image dataset of peripheral blood samples from normal individuals and leukaemia patients, and it contains the relative supervised classification and segmentation data. These samples were collected by the experts at the M.Tettamanti Research Centre for childhood leukaemia and haematological diseases, Monza, Italy. The ALL-IDB database has two distinct versions: in the first version (ALL-IDB1) can be used for both testing the segmentation capability of algorithms, as well as the classification systems and image pre-processing methods, and the second version (ALL-IDB2) is a collection of cropped areas of interest from normal and blast cells that belong to the ALL-IDB1 dataset, so it can be used only for testing the performance of classification systems. In both versions of the dataset, each image has an associated text file containing the coordinates of the centroid of each candidate lymphoblast, which was manually labelled by a skilled operator and can be used as a ground truth.

III.RESULTS AND DISCUSSION

From the above surveys made, segmentation plays a vital role in segmenting the leukaemia cells and detecting whether it is infected or healthy cell. Experimental results show Threshold based segmentation works out well in segmenting the cells and watershed segmentation works out well in identifying and separating the grouped leucocytes.

The Accuracy of otsu thresholding segmentation is 75% to 80% accurate based on the quality of the image and overall percentage of detecting leukaemia using all the above steps is 78-83% accurate.



IV. CONCLUSION

In this paper we illustrated various techniques and methodologies used to segment the images, clean the images, remove the granules and finally to classify the images. A common framework for evaluation of segmentation in introduced in [12].Further the type of blasts can also be identified in blood cancer.

REFERENCES

1. Seiter, K (5 February 2014). Sarkodee Adoo, C; Talavera, F; Sacher, RA; Besa, EC, ed. "Acute Lymphoblastic Leukemia". *Medscape Reference*. WebMD. Retrieved 17 April 2014.
2. Priyankara G., Seneviratne O., Silva R., Soysa W. and Silva C., "An Extensible Computer Vision Applicaton for Blood Cell Recognition and Analysis", 2006.
3. Reta C., Altamirano L., Gonzalez J., Diaz R. and Guichard J.," Segmentation Of Bone Marrow Cell Images For Morphological Classification Of Acute Leukaemia", Proceedings of the 23rd international Florida artificial intelligence Research society Conference, Florida, U.S.A., 2010, pp. 86-91.
4. Ruggero Donida Labati, Vincenzo piuri, Fabio Scotti, All-IDB: "The acte lymphoblastic Leukaemia image Database for image Processing".
5. Scotti F. "Robust segmentation and measurements techniques of white cells in blood microscope images." In: Daponte P, Linnenbrink T, editors. Proceedings of the IEEE instrumentation and measurement technology conference, 24–27April. Sorrento, Italy: IEEE Publisher; 2006. p. 43–8.
6. Gonzalez RC, Woods RE. "Digital image processing". 3rd ed. NJ, USA: Prentice Hall Pearson Education, Inc.; 2008.
7. Sinha N, Ramakrishnan AG. "Automation of differential blood count". In: Chock-alingam A, editor. Proceedings of the conference on convergent technologies for the Asia-Pacific region, October 15–17. Taj Residency, Bangalore: IEEE Publisher; 2003. p. 547–51.
8. Kovalev VA, Grigoriev AY, Ahn H. "Robust recognition of white blood cell "images.In: Kavanaugh ME, Werner B, editors. Proceedings of the 13th international conference on pattern recognition, August 25–29. Vienna, Austria: IEEE Publisher; 1996. p. 371–5.
9. Meyer F." Topographic distance and watershed lines". Signal Process1994; 38(July):113–25.
10. Cseke I. "A fast segmentation scheme for white blood cell images. In: Proceedings of the IAPR international conference on image, speech and signal analysis", 30 August–1 September. The Hague, Netherlands: IEEE Publisher; 1992.p. 530–3
11. Otsu N. "A threshold selection method from gray level histograms". IEEE TransSyst Man Cybern 1979; 9(1):62–66.
12. .http://www.webmd.com/cancer/lymphoma/acute-lymphoblastic-leukemia
13. Qian Wang ; Tianjin Key Lab. of Intell. Comput. & Novel Software Technol., Tianjin Univ. of Technol., Tianjin, China ; Hua Zhang ; Qi Dong ; Qing Xiao, "Otsu thresholding segmentation algorithm based on Markov Random Field", Natural Computation (ICNC), 2011 Seventh International Conference on (Volume:2),IEEE Publisher, July 2011,969-972
14. Ben Chaabane, S. ; SICISI, Ecole Super. des Sci. et Tech. de Tunis (ESSTT), Tunis ; Sayadi, M. ; Fnaiech, F. ; Brassart, E., "Color image segmentation using automatic thresholding and the fuzzy C-means techniques", Electrotechnical Conference, 2008. MELECON 2008. The 14th IEEE Mediterranean, IEEE Publisher, May 2008, 857-861
15. Jinyang Sun, Fangjun Luan, and Hanhui Wu**,** "Optic Disc Segmentation by Balloon Snake with Texture from Color Fundus Image", International Journal of Biomedical Imaging, Volume 2015 (2015), Article ID 528626, 14 pages.
16. Nandy K, Gudla PR, Meaburn KJ, Misteli T, Lockett SJ, "Automatic nuclei segmentation and spatial FISH analysis for cancer detection", Conference Proceedings IEEE Engineering in Medicine and Biology Society, 2009;2009:6718-21. doi: 10.1109/IEMBS.2009.5332922.