Comparative Study of Various Techniques for Leukemia Detection

Rupinder Kaur Sandhu
PhD student , Computer Engineering, Punjabi University, Patiala, India
rupinderkrsandhu@gmail.com

Dr. Raman Maini
Professor, UCOE, Punjabi University, Patiala, India

***Abstract*— Leukemia is a type of blood cancer. The treatment depends on the type of Leukemia so proper diagnosis of the type is very essential for the recovery of the patient. Type of Leukemia can be diagnosed manually or automatically. Manual classification is sometimes prone to the errors as it depends on the experience of the hematologist. A detailed study of the techniques used in the literature for Leukemia detection is presented in this paper. Many techniques are proposed in the literature for the detection of the Leukemia. Watershed algorithm combined with Optimal thresholding produced best results for the cell segmentation, nucleus segmentation and cytoplasm segmentation i.e. 99.85%, 99.92% and 99.63% respectively.**

Keywords- Neural networks, SVM, Cytoplasm and thresholding

I. Introduction

Leukemia is cancer of blood or the bone marrow in which immature white blood cells also known as "blasts" increase abnormally. It is a treatable disease. The treatments involve medical radiation therapy, chemotherapy, bone marrow transplant or hormone treatments. Cure rate depends on the type of leukemia and also on the age of the patient. Permanent cure is more likely for the children than for the adults. Even though complete cure is unlikely, most people with leukemia can be successfully treated for years. There are three main constituents of blood namely Red blood cells which carry oxygen all over the body, Platelets which help in stopping bleeding, White blood cells help in fighting infection. Blood cells die when they become old or damaged and the new blood cells are produced by the blood stem cells [20].

A. Normal Myeloid and Lymphoid Cells

As shown in the figure 1 blood stem cells produce all the blood cells. Two pathways are shown in the Fig. 1 i.e. myeloid as well as lymphoid stem cells both are produced by the blood stem cell. Red blood cells and platelets are produced by the Myeloid stem cells. Myeloblasts are also produced by the Myeloid stem cells. Different types of white blood cells also known as Granulocytes are in turn produced by the Myeloblasts. Lymphoblasts are produced by the lymphoid stem cell, which in turn can produce different types of white blood cells which are different from the Granulocytes.

B. Broad Classification of Leukemia

Clinically and pathologically, leukemia is subcategorized into several subtypes. The main division is between its acute and chronic forms:

- Acute leukemia is a rapidly progressing disease which affects the cells which are not fully developed. Immature blood cells rapidly increase in this. These cells are not capable of carrying out their normal functions. Due to the overcrowding by these cells the bone marrow is not able to produce the healthy blood cells. Due to the rapid progress of the malignant cells immediate treatment is required in acute leukemia. Such cells then spread to the other organs of the patient. Acute forms of leukemia mostly occur in the children.
- In the case of chronic leukemia there is excessive build up of relatively mature white blood cells but they are still abnormal. Chronic leukemia usually progress slowly as compared to the Acute Leukemia. In chronic leukemia the cells are produced at a higher rate than normal so it typically takes months or years to progress. As a result there is the accumulation of the abnormal white blood cells. These cells are capable of carrying out some of their normal functions. As the treatment approach is based on the subtype so knowing the subtype of the leukemia is very important. Chronic leukemia mostly occurs in older people, but it may occur in any age group.

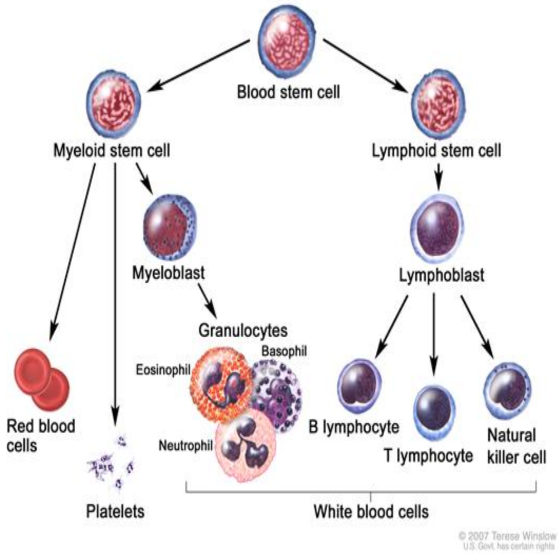

Figure 1: Cells formed in the blood [20]

C. *Four main categories of Leukemia*

Four main categories can be obtained by the Combination of the two broad classifications. There are several subcategories within each of these four main categories.

- Acute Lymphoblastic Leukemia (ALL) mostly occurs in the young children. This disease affects adults also those of age 65 and older. The survival rate is 85% for children and 50% for adults [18]. Its various subtypes are Precursor T Acute Lymphoblastic Leukemia, precursor B Acute Lymphoblastic Leukemia, Acute Biphenotypic Leukemia and Burkitt's Leukemia.
- Mostly adults above the age of 55 get affected by Chronic Lymphocytic Leukemia. It never affects children but sometimes occur in young adults. Mostly men gets affected by this. It has 75% five-year survival rate [21]. Though there are effective treatments available but it is incurable. Its subtype is B-cell Prolymphocytic Leukemia.
- Very few children develop Acute Myelogenous Leukemia (AML) but it mostly occurs more commonly in adults. Also it occurs more commonly in men and is rare in women. The five-year survival rate for all subtypes is 40%, except that for Acute Promyelocytic Leukemia, which is above 90% [19]. Its various subtypes include Acute Myeloblastic Leukemia, Acute Promyelocytic Leukemia and Acute Megakaryoblastic Leukemia.
- A very small number of children get affected from Chronic Myelogenous Leukemia (CML) but it occurs mostly in adults. Its subtype is Chronic Myelomonocytic Leukemia.

II. LITERATURE SURVEY

Proper segmentation of cells is a very important step for the proper detection of leukemia. Further lymphocytes need to be segmented into nucleus and cytoplasm. Basically in segmentation the regions with similar characteristics are separated. The main goal of segmentation is the division of the white blood cells into its basic components like cytoplasm, nucleus, blobs etc. Segmentation is the most important step in the automatic system, therefore all other subsequent steps like extraction of features and classification totally depend on the accuracy and reliability of the previous step i.e. segmentation. If there is some mistake in the step of segmentation then all other subsequent steps will also get wrong.

Image preprocessing is also considered as an important step in the classification of leukemia. The main purpose of image preprocessing is to enhancement of the image thus making it suitable for further processing.

Various methods for the preprocessing of the images are:

- Enhancement of the contrast of the image.
- Removal of noise.
- Isolation of the objects of interest in the image.

A. The various preprocessing techniques used for image preprocessing in the literature are discussed below:

1) LPG-PCA Algorithm

Bhattacharjee, R., Chakraborty, M., [6] developed a de-noising module. Two staged noise reduction is performed using Principal Component Analysis and Local Pixel Grouping. An orthogonal transformation matrix is used in PCA to completely de-correlate the centralized matrix. The image's energy is concentrated on the small sub-sets of the PCA transformed matrix although the noise is spreading over the whole image. Hence the signal to noise ratio is reduced by this process. Refinement done in second stage improves the PSNR.

2) *Selective Filtering & Unsharp Masking*

Mohapatra, S., Patra, D., [7] uses the selective median filtering and then unsharp masking in the second step. Even after the application of the median filtering minute edge details of the microscopic images are preserved completely. The segmentation process is made easier by performing unsharp masking which sharpens the details of the images.

3) Contrast Enhancement Techniques

Salihah, A, et. al., [8] proposed three contrast enhancement techniques for the color images using the RGB components. Partial contrast, bright stretching and dark stretching are the three contrast enhancement techniques used in this paper. The morphological features of the images are enhanced using the contrast enhancement techniques hence making the classification between Acute Lymphoblastic Leukemia and Acute Myelogenous Leukemia easier. Partial contrast is proved as the best technique by the results. It improves the visibility of the images and also preserves the significant features of the Acute Leukemia images.

Halim, N.H.A. et. al. [9] used Global contrast stretching (GCS) for improving the quality of the images. For determining the maximum and minimum values for all the RGB color images the Global contrast stretching is used which considers all range of color palates at once. Only one maximum and one minimum value for RGB color is obtained by the combination of the RGB colors. Contrast stretching process will use the maximum and minimum values.

4) Color Space Conversion

Madhukar, M., et. al., [10] converted the RGB input image to CIELAB color space. It is used basically for reducing the color dimension from three to two. The three main constituents of L*a*b* space are luminosity layer L*, chromaticity layer a* and chromaticity layer b*.

Agaian, S., et. al., [17] proposed a simple technique that automatically detects and segments the AML in the blood smears. Color correlation is used in this technique. Also RGB color space is converted to CIELAB. The main reason for doing this is reduction in the memory requirement and improvement in the computational time

B. Study of various segmentation techniques for image segmentation

In the literature various techniques are proposed for the segmentation of blood images into nucleus and cytoplasm. A brief review of the techniques used is presented below.

1) *Watershed algorithm and Optimal Thresholding*

Mohammed, E.A. et al. [1] presented in this paper a method to segment normal and CLL Lymphocytes into two parts: nucleus, and cytoplasm using a Watershed algorithm and Optimal Thresholding. The over and under segmentation error which is generally produced by the Watershed algorithm is reduced by the used method by suppressing 1% of the local minima. Maximum accuracy of 99.92% was obtained for nucleus segmentation when the algorithm was used for testing of 140 microscopic lymphocyte images (normal and CLL), and for cell segmentation the maximum accuracy obtained was 99.85%. 99.63% maximum accuracy was obtained when the cytoplasm was extracted using simple mask subtraction.

2) Automatic thresholding using contrast stretching

Madhloom, H.T., et al. [2] used arithmetic and automatic threshold for the automatic segmentation of the nucleus of the white blood cells. Accuracy obtained for nucleus segmentation using this method was 98.81% and the cytoplasm was segmented with 85.98% accuracy.

3) Pixel intensity thresholding

Sadeghian, F., et al. [3] presented in this paper a technique for the segmentation of white blood cells using a combination of morphological analysis using snake algorithm and pixel intensity thresholding. 92 % accuracy was obtained for nucleus segmentation and an accuracy of 78% for the cytoplasm segmentation.

4) Scale space filtering and watershed clustering

Jiang, K., et al. [4] proposed a technique for the segmentation of white blood cells using a combination of scale space filtering and watershed clustering. The scheme used produced 98.9% accuracy for the cell segmentation and is based on feature space clustering.

5) Morphological operations and Watershed transform

Lim Huey Nee, et al., [5] proposed a method consisting of morphological operations, gradient magnitude, thresholding, and watershed transform for performing cell segmentation. The proposed method produced 95.09% maximum accuracy for the cell segmentation.

6) Edge sensitive variational thresholding and Otsu's thresholding

Bhattacharjee, R., Chakraborty, M., [6] uses two methods for thresholding namely Edge sensitive variational thresholding and Otsu's thresholding. The threshold value obtained from the above two methods is used for the segmentation of the blast cells. 46.4975% average accuracy was produced by the application of Method 1 and an average accuracy of 80.665% was produced by the method 2.

7) Color Space Segmentation

Halim, N.H.A. et. al. [9] for the purpose of improving the image quality used segmentation based on HSI (Hue, Saturation, Intensity) color space. As there is a high correlation among the three components of RGB color model hence the chromatic information produced is not suitable for the direct processing. Due to the direct relation with the human perception it is more convenient to convert RGB to HSI color space.

Madhukar, M., et. al., [10] converted the RGB input image to CIELAB color space. The main purpose of this conversion is the reduction of the color dimension from three to two in comparison to RGB. The L*a*b* space consists of a luminosity layer L*, chromaticity layer a* and chromaticity layer b*.

Mu-Chun Su, et al. [13] with the purpose of the recognition of five types of white blood cells namely Lymphocyte, Monocyte, Basophil, Eosinophil and Neutrophil addressed a new classification system for the classification of white blood cells. A new algorithm for the segmentation of WBCs is proposed and the main purpose of the proposed segmentation algorithm is finding a discriminating region of WBCs on the HSI color space. Ellipsoidal region describes a discriminating region and the pixels having their color lying in the defined discriminating region are regarded as the nucleus and granule of cytoplasm of a white blood cell. 99.11% highest correct recognition rate is reached.

8) K-Means Clustering

Madhukar, M., et. al., [10] used a segmentation algorithm for the extraction of the nuclei of the leukocytes using color based clustering. Every pixel is assigned to one of the clusters using a k-means clustering. On the basis of the corresponding *a and *b values in L*a*b color space each object's pixels are classified into k clusters.

Agaian, S., et. al., [17] developed a technique for the detection of AML using k-means clustering. The Euclidean distance between the pixel and each color indicator is calculated and based upon that classification of the pixels in the L*a*b color space is done into any of the k clusters.

9) Gram-Schmidt orthogonalization and Snake Algorithm

Rezatofighi, S. H., Soltanian-Zadeh, H., [14] proposed an image processing algorithm for the automatic recognition of the five types of white blood cells in the peripheral blood. For the segmentation of the nucleus and cytoplasm of the cells a method based on the Gram–Schmidt orthogonalization along with a snake algorithm is proposed. As a result of the application of the combination of the algorithms a composite image is obtained having maximum intensity of the regions of interest with the desired color while the other regions have the minimum intensity.

C. Study of various classification methods

Classification is the task of assigning a label from one of the known classes to the unknown test vector. A brief description of the methods used for the classification in the literature is given below:

1) Support Vector Machines

Mohapatra, S., Patra, D., [7] uses Support Vector Machines for the classification. The various extracted features in the proposed scheme are contour signature, fractal dimension, the various shape features like

area, compactness, perimeter, solidity, elongation, eccentricity, color and texture features. The SVM classifier is trained using the features extracted in the proposed scheme and 95% accuracy is obtained.

Rezatofighi, S. H., Soltanian-Zadeh, H., [14] proposed an image processing algorithms for the automatic recognition of the five types of white blood cells in the blood. A combination of Gram–Schmidt orthogonalization with the snake algorithm is used for the segmentation of nucleus and cytoplasm of the cells. Then, from the segmented regions a variety of features are extracted. Then using a Sequential Forward Selection (SFS) algorithm most discriminative features are selected and the performance of the two classifiers namely Artificial Neural Network and Support Vector Machine is compared. From the results it is confirmed that SVM is superior.

Markiewicz, T., et. al. [16] proposed the system for automatic recognition of the Leukemia blast cells. The proposed system uses Support Vector Machine as the classifier for the classification and SVM exploits the features of the blood cell images related to the geometry, texture and the histograms.

Agaian, S., et. al., [17] presented a technique for the detection of the AML. The SVM was used for the construction of a decision surface in the feature space that divides the cells into two categories i.e. noncancerous and cancerous and the margin of separation between two classes of points is maximized. Accuracy of 98% is obtained for the localization of the lymphoblast cells by the proposed framework. Separation from the subimages and complete images is also done.

2) Hold-out cross validation method

Madhukar, M., et. al., [10] used the hold-out cross validation method for the validation of the segmented images. In order to avoid over-fitting the hold-out validation uses an independent test set. The splitting of the available data into two non-overlapped parts is done, one for testing and other for training. The test data is kept separate and not considered during training. 93.5% of the cases were classified correctly proving that the method obtains good results for the Leukemia classification.

3) Hausdorff Dimension and contour signature

Mohapatra, S., et. al. [11] uses two novel shape features for the classification of lymphocytic cell nucleus namely Hausdorff Dimension and contour signature. For the calculation of the perimeter roughness of each nucleus, Hausdorff Dimension is used. For measuring the irregularity quantitatively, Contour Signature method is used. An accuracy of 93% was observed by this method.

4) Artificial Neural Networks

Aeinfar, V. et. al., [12] uses Multilayer Perceptron Neural Network for the diagnosis of the five types of blood disorders. The proposed method achieved fast convergence, high accuracy and low use of memory for the diagnosis and for predicting the disorder. The multi variable nonlinear regression can analyze only one disorder at a time whereas the proposed system has a very prominent property of diagnosing all five types of disorders simultaneously.

Khashman, A., [15] developed an intelligent system which simulates a human visual inspection and classification of the three types of blood cells namely, erythrocytes (RBCs), leukocytes (WBCs), and platelets. The proposed system comprises of the two phases: the image preprocessing phase in which the features of blood cells are extracted using global pattern averaging and the neural network arbitration phase where firstly training is done and then classification. Experimental results prove that the proposed method performs very well in identifying blood cell types regardless of their irregular sizes, shapes and orientation, hence providing a fast, simple and efficient rotational and scale invariant identification system for the blood cells which can also be used in automating the laboratory reporting.

III. RESULTS & DISCUSSION

TABLE 1. PERCENTAGE OF CELL SEGMENTATION ACCURACIES FOR DIFFERENT METHODS

S. No.	Technique used	Result (%)
1	Watershed Algorithm Optimal Thresholding [1]	99.85 %
2	Scale space filtering Watershed clustering [4]	98.9 %
3	Morphological operations Watershed transform [5]	95.09 %

TABLE 2. RESULTS OF SEGMENTATION OF LYMPHOCYTES USING DIFFERENT TECHNIQUES

S. No.	Technique used	Result of segmentation	
		Nucleus (%)	Cytoplasm (%)
1	Morphological analysis using snake algorithm Pixel intensity thresholding [3]	92 %	78 %
2	Watershed algorithm Optimal Thresholding [1]	99.92 %	99.63%
3	Image arithmetic Automatic thresholding using contrast stretching [2]	98.81 %	85-98 %

IV. CONCLUSION

Many algorithms are used in the Literature for the detection of Leukemia. Detailed study of the algorithms indicates that the combination of watershed algorithm and Optimal Thresholding produced highest result for cell segmentation (99.85%), nucleus segmentation (99.92%), and Cytoplasm segmentation (99.63%).

REFERENCES

[1] Mohammed, E.A., Mohamed, M.A., Naugler, C., Far, B.H., "Chronic Lymphocytic Leukemia cell segmentation from microscopic blood images using Watershed Algorithm and Optimal Thresholding", 26th IEEE Canadian Conference of Electrical and Computer Engineering, pp. 1-5, 5-8, May 2013.

[2] H.T. Madhloom, S.A. Kareem, H. Ariffin, A.A. Zaidan, H.O. Alanazi and B.B. Zaidan, "An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold", J. App. Sci., vol. 10, pp. 959-966, 2010.

[3] Farnoosh Sadeghian, Zainina Seman, Abdul Rahman Ramli, Badrul Hisham Abdul Kahar, and M-Iqbal Saripan "A framework for white blood cell segmentation in microscopic blood images using digital image processing", Biol. Proc. Online, vol. 11, pp. 196-206, 2009.

[4] Kan Jiang, Qing-Min Liao, Yuan Xiong, "A novel white blood cell segmentation scheme based on feature space clustering", Soft Computing, vol. 10, pp. 12-19, 01/01, 2006.

[5] Lim, Huey Nee, Mohd Yusoff Mashor, and Rosline Hassan, "White blood cell segmentation for Acute Leukemia bone marrow images", International Conference on Biomedical Engineering (ICoBE), Penang, Malaysia, pp. 357-361,27-28 February 2012.

[6] Bhattacharjee, R., Chakraborty, M., "LPG-PCA algorithm and Selective Thresholding based automated method: ALL & AML blast cells detection and counting", International Conference on Communications, Devices and Intelligent Systems (CODIS), pp. 109-112, 28-29, December 2012.

[7] Mohapatra, S., Patra, D., "Automated cell nucleus segmentation and Acute Leukemia detection in blood microscopic images", proc. of International Conference on Systems in Medicine and Biology, IIT Kharagpur, India, pp. 49-54, 16th-18th December, 2010.

[8] A.N. Aimi Salihah, M.Y. Mashor, Nor Hazlyna Harun, H. Rosline, "Colour image enhancement techniques for Acute Leukaemia blood cell morphological features ", IEEE International Conference on Systems Man and Cybernetics (SMC) , p. 3677 – 3682, 2010.

[9] Halim, NH Abd, M. Y. Mashor, A. S. Abdul Nasir, N. R. Mokhtar, and H. Rosline, "Nucleus segmentation technique for Acute Luekemia", proc of IEEE 7th International Colloquium on Signal Processing and its Applications, 2011.

[10] Madhukar, M., Agaian, S., Chronopoulos, A.T., "Deterministic model for Acute Myelogenous Leukemia classification", IEEE International Conference on Systems, Man and Cybernetics, COEX, Seoul, Korea, pp. 433-438, 14th-17th October, 2012.

[11] Mohapatra, Subrajeet, Sushanta Shekhar Samanta, Dipti Patra, and Sanghamitra Satpathi, "Fuzzy based blood image segmentation for Automated Leukemia detection ", International Conference on Devices and Communications (ICDeCom), p. 1-5, 2011.

[12] Vahid Aeinfar, Hoorieh Mazdarani, Fatemeh Deregeh, Mohsen Hayati, Mehrdad Payandeh, "Multilayer perceptron Neural Network with supervised training method for diagnosis and predicting blood disorder and cancer", IEEE International Symposium on Industrial Electronics, p. 2075 – 2080, 2009.

[13] Mu-Chun Su, Chun-Yen Cheng, Pa-Chun Wang, "A Neural-Network-Based approach to White Blood Cell classification", The Scientific World Journal, Hindawi Publishing Corporation, pp. 1-10, Jan, 2014.

[14] Rezatofighi, S. H., Soltanian-Zadeh, H., "Automatic recognition of five types of white blood cells in peripheral blood", Computerized Medical Imaging and Graphics, published by Elsevier, pp. 333-343, January, 2011.

[15] Khashman, A., "IBCIS: Intelligent blood cell identification system", Progress in Natural Science 18, published by Elsevier, pp. 1309-1314, March, 2008.

[16] Markiewicz, T., Osowski, S., Marianska, B., Moszczynski, L., "Automatic recognition of the blood Cells of Myelogenous Leukemia using SVM", proceedings of the International Jpoint Conference on Neural Networks, Montreal, Canada, July 31-August 4, 2005.

[17] Sos Agaian, Monica Madhukar, Anthony T. Chronopoulos, "Automated screening system for Acute Myelogenous Leukemia detection in blood microscopic images", IEEE Systems Journal, Vol 8, No. 3, pp. 995-1004.

[18] Jameson, J. N. St C.; Dennis L. Kasper; Harrison, Tinsley Randolph; Braunwald, Eugene; Fauci, Anthony S.; Hauser, Stephen L; Longo, Dan L. , "Harrison's principles of internal medicine", New York: McGraw-Hill Medical Publishing Division. ISBN 0-07-140235-7, 2005

[19] Colvin G. A., Elfenbein G. J. "The latest treatment advances for acute myelogenous leukemia", Medicine and Health, Rhode Island 86 (8): 243–6. PMID 14582219, 2003.

[20] http://www.cancer.gov/cancertopics/wyntk/leukemia/page2

[21] http://en.wikipedia.org/wiki/Leukemia