

Computerized Detection System for Acute Myelogenous Leukemia in Blood Microscopic Images

Yogesh Ambadas Gajul¹, Rupali Shelke²P.G. Student, Department of Electronics Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India¹Associate Professor, Department of Electronics Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India²

ABSTRACT: Acute myelogenous leukemia (AML) is a subtype of acute leukemia which is a type of cancer where the blood cells and bone marrow get affected. This type of cancer is mostly observed among adults with an average age of 65 years. The present method for the detection of acute myelogenous leukemia (AML) is not efficient, as this method needs manual observation of blood smear under the microscope which depends on observer's fatigueness, which is not good for patient. This paper presents an efficient method that automatically detects AML cells in the blood smear. The proposed approach mainly comprises of four stages, viz. pre-processing stage, segmentation stage, feature extraction stage and classifier stage respectively.

KEYWORDS: Blood Microscopic Images, WBC, Acute Myelogenous Leukemia (AML), HD, LBP.

I. INTRODUCTION

Many diseases can be found out by examining the blood cells. The blood cell gives lot of information by which we can identify the type of disease caused to a particular person. These blood cells are diagnosed under the microscope. The microscopic images of the blood cells are analyzed by the expert doctors and depending on that analysis and observation conclusion is made. Also the changes in the blood condition can show the development of disease in an individual person.

Blood contains many cells such as WBC (leukocytes), RBC (erythrocytes) and platelets. Leukemia is detected only by observing the WBC's. So, this paper is focused only on WBC's. Leukemia is categorized into chronic leukemia and acute leukemia. These are further classified into four types, viz. Acute Lymphocytic Leukemia (ALL), Acute Myeloid Leukemia, Chronic Lymphocytic Leukemia (CLL) and Chronic Myeloid Leukemia (CML). In this paper we develop an algorithm that will automatically classify the Acute Myelogenous Leukemia. AML is a fast growing cancer of the blood and bone marrow. In this the bone marrow produces many unformed cells called blasts. Blasts are developed mostly in the white blood cells that are used as a defense mechanism in the body. However, the blasts are not fully formed in the blast due to this it cannot fight completely with the infections. Acute Myelogenous Leukemia is difficult to diagnose because the main cause of AML is still not known. Most of the symptoms and signs of AML are caused by the replacement of normal blood cell with the leukemia cells. Susceptibility to infections is caused due to lack in the production of normal white blood cells. Decrease in red blood cell count can cause fatigue, platelets, and shortness of breath. Similarly, lack of platelets may lead to bleeding with minor trauma.

Many times there is confusion during diagnosis due to the similarities in the signs of other disorders. In recent years cells are analyzed accurately because of the improvement in the field of digital imaging. But still the required information cannot be properly extracted from WBC due to many complications. The proposed system will help to detect the AML cells in the blood smear automatically.

After studying all the systems it can be concluded that these systems are designed only for sub images. So, there is a need to develop a system that will work on whole image. In the proposed system classification is done on the entire

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

image. Linear Support Vector Machine is used for the classification and finally the results will be compared with the existing system. Section II discuss about the previous work that has been carried out by various researchers. Section III describes about the methodology of the project. Section IV discuss about performance of the system for the proposed work & Section V discuss the experimental results.

II. RELATED WORK

In last few years many researchers are putting their efforts to classify AML cells from the blood smear. S.Serbouti [5] proposed the use of classification and regression trees (CART) statistical software for the classification of malignancies using the cell markers that are extracted from the images, but did not mention the scheme used for segmentation and feature extraction. D.J. Foran [6] developed a prototype of distinguishing the malignancies. The system used two major components such as distributed telemicroscopy system and an intelligent image repository for discriminating among lymphoma and leukemia. F.Scotti[7] proposed a method for automatic classification of ALL in gray level peripheral blood smear images. From experiments it has been concluded that lymphoblast recognition is feasible from blood images using morphological features. But the recognition rate is low due to the use of Otsu thresholding in image segmentation and feed forward neural network for feature classification.

T.Markiewicz [8] used SVM as a classifier and exploits the features of the image of the blood cells related to texture, geometry and histograms. Myeloblast is recognized with this system but the work for lymphoid series is still remaining. Hazwani [9] reported a system which automatically counts blasts for acute leukemia detection in blood microscopic images. J.Angulo [10] used watershed transformation for lymphocyte image segmentation. After this morphological features are extracted for the classification of lymphocytes. Madhloom included in research some arithmetic and threshold operations for finding the white nuclei. Problem faced while selecting the thresholding method, due to which the results obtained are not satisfactory. Kovalev came up with the system that classifies five types of leukocytes from the images of blood smear. But classified the system only for sub images. Rangayyan[11] explains about the unsupervised segmentation algorithm for the separation of white blood cells. Nallaperumal [12] presented a watershed segmentation algorithm for the separation of the nucleus from the surrounding cytoplasm.

III. METHODOLOGY

Figure 1 depicts the overview of the system. It shows the sequence of steps that are to be followed for the efficient classification of Acute Myelogenous Leukemia. The system has four main stages viz. pre-processing stage, segmentation stage, feature extraction stage and classification stage. In pre-processing stage the unwanted noise content present in the image is removed. Also the RGB image is converted into L^*a^*b color space image. Pre-processing stage is followed by the segmentation stage which uses k-means clustering. After that features are extracted in feature extraction stage which majorly uses LBP and HD. This is followed by the classification stage which uses SVM. Finally validation is performed.



Fig. 1: Overview of system.

A. Input Acquisition:

American Society of Hematology (ASH) [14] is a web based library which provides images of high quality that contains collection of AML images. These obtained images are in RGB color format.

B. Pre-Processing:

An input image has some unwanted noise present in it. Also there is lot of variations in image background with respect to color and intensity. These types of variations are caused due to many reasons such as camera settings, varying

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

enlightenment and aging blemish. To overcome such kind of problems, input RGB image is converted to CIELAB/ $L^*a^*b^*$ color space. The L represents the lightness of the color, element a^* that represents its position between red/magenta, and element b^* that represents its position between yellow and blue. To make accurate color balance in the image, elements 'a' and 'b' are used.

C. Segmentation:

The main aim of the segmentation is to extract useful information from the image. Many researchers have implemented different segmentation methods. In this proposed system k-means clustering is used. Clustering is a way to separate groups of objects. K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. Here, the cluster corresponds to the nucleus, background and other cells. Then using the property of cluster center each pixel is assigned to one of these classes. After the segmentation is performed for preserving the nuclei of whole image, morphological operations are to be performed.

D. Feature Extraction:

Transforming the input data into set of features is called feature extraction. The performance of the classifier depends totally on the features that are extracted, so a correct choice of features need to be extracted. It was noted that most of the researcher used common features as they gave effective results. Hence, following features are considered.

1) Hausdorff Dimension (HD): Edges of nucleus are considered as an essential feature, so HD is used. The fractal dimension D shows how completely the space is filled by the fractals. Box-counting technique is used to detect the edge of nucleus as it is easier to implement and most widely used. It is defined by the formula given as,

$$HD = \frac{\log(R)}{\log(R(s))} \quad \dots(1)$$

where R is the number of squares in the superimposed grid, and R(s) is the number of occupied squares or boxes (box count).

2) Local Binary Pattern (LBP): The LBP is used for the local texture feature extraction, Due to its following advantages.

- i) They are robust over illumination variation, (overcomes image Acquisition and manual Errors).
- ii) They can compute at a faster rate, (As Medical fields require correct result but within time so that is added advantage to our system).
- iii) They work with minimum parameters.
- iv) They have a local feature.
- v) They do not vary for monotonic gray scale transformations and scaling.

For our image database an (8, 1) circular neighborhood was used. After the segmentation of an image LBP operator was applied then after that all other features were employed.

3) Shape Features: According to the hematologists, the important feature for classification is the shape of nucleus. The region and boundary based shape features are extracted in order to analyze the shape of the nucleus. Following are the shape features that are considered.

Area: The area is defined by counting the total number of none zero pixels within the image region.

Perimeter: Computing distance between consecutive boundary pixels.

Compactness: Measure of a nucleus is called compactness.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

$$\text{Compactness} = \frac{\text{perimeter}^2}{\text{Area}} \quad \dots(2)$$

Solidity: The ratio of actual area and convex hull area is known as solidity.

$$\text{Solidity} = \frac{\text{Area}}{\text{Convex Area}} \quad \dots(3)$$

Eccentricity: How much the shape of a nucleus deviates from being spherical?

$$\text{Eccentricity} = \frac{\sqrt{a^2 - b^2}}{a} \quad \dots(4)$$

Elongation: Abnormal bulging of the nucleus is defined by elongation.

$$\text{Elongation} = \frac{R_{\max}}{R_{\min}} \quad \dots(5)$$

Form factor: The measure of surface irregularities.

$$\text{Form Factor} = \frac{4 \times \pi \times \text{area}}{\text{perimeter}^2} \quad \dots(6)$$

4) GLCM Features: One of the image analysis techniques is the GLCM feature. Using this method different texture features are extracted such as entropy, contrast, correlation and energy. Following are the Gcm features,

Homogeneity: Measurement of degree of variance.

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \dots(7)$$

Energy: Measurement used to measure uniformity.

$$\sum_{i,j=0}^{N-1} (P_{i,j})^2 \quad \dots(8)$$

Correlation: To measure correlation between pixel values and its neighborhood.

$$\sum_{i,j=0}^{N-1} \frac{(i-\mu)(j-\mu)}{\sigma^2} \quad \dots(9)$$

Entropy: Measurement of randomness.

$$\sum_{i,j=0}^{N-1} P_{i,j} (-\ln(P_{i,j})) \dots(10)$$

5) Color Feature: An important feature that human perceive while visualizing is color. This feature is considered for extraction from nucleus region. Hence, for each nucleus image the mean color value in RGB color space is obtained. Following is the color feature that is considered,

Cell Energy: Also known as the measure of uniformity, it is the different Lab image components. We define feature “ δ ” to be

$$\delta = \sum_i \sum_j P^2(i, j) + (\sqrt{-1}) \left(\frac{\sqrt{\sum_{i=1}^n (x_i - x')^2}}{(n-1)} \right) \dots(11)$$

where, $x' = \sum_{i=1}^n x_i / n$, $P(i,j)$ represents the normalized GLCM element for the i^{th} row and j^{th} column, and represents the ASM.

E. Classification:

This stage is considered to be very important because it decides the accuracy of the system. So, the selection of method used for classification plays an important role. The proposed methodology uses linear support vector machine (SVM)

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

for classifying abnormal and normal cells. SVM is a supervised learning algorithm that analyze data for classification and regression analysis. There are different kernel tricks that can be experimented. But for the classification, a polynomial kernel is used for analyzing and comparing the learning algorithm.

IV. PERFORMANCE EVALUATION

To validate the performance of the system this paper used four statistical measures of the performance of a binary classification test. Sensitivity, Specificity, Precision and F-measure are all defined in relation to the possible outcomes of the classifier system. When we attempt to classify an image, there are four outcomes viz. True Positive, True Negative, False Positive, False Negative. True Positive is the condition when cancerous cells are identified as cancerous cells, False Positive is the condition when non-cancerous cells are identified as cancerous cells, True Negative is the condition when non-cancerous cells are identified as non-cancerous cells and False Negative is the condition when cancerous cells are identified as non-cancerous cells.

Table 1: Parameters for Performance Evaluation.

| Parameters | Formulae | Results |
|--|---|---------|
| Sensitivity - also called as recall denotes the test's ability to identify positive results. | $TP / (TP + FN)$ | 71.43 % |
| Specificity - denotes the test's ability to identify negative results. | $TN / (TN + FP)$ | 100 % |
| Precision - gives the proportion of subjects with positive results who are correctly identified. | $TP / (TP + FP)$ | 100 % |
| F-measure – is a metric that gives the harmonic mean of Precision and Sensitivity. It is the overall classification performance. | $\frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity}$ | 83.33 % |

Table 1 describes about the parameters that are required to calculate performance evaluation. The results are also shown in the table. The values that are obtained for TP, TN, FP and FN respectively are substituted in the formula in order to calculate values for each parameter. Depending on formula the results are obtained.

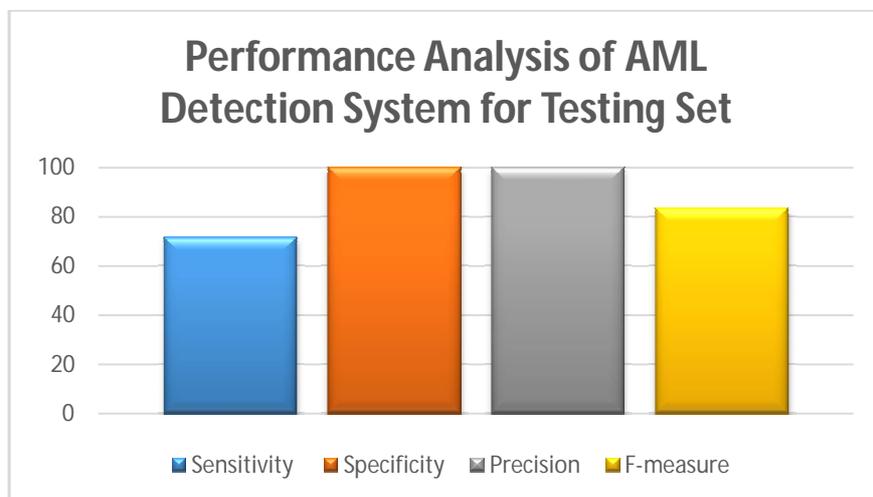


Fig. 2: Performance evaluation of the system for testing dataset.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Figure 2 shows the graphical representation of the result obtained for the above four parameters. These results were obtained for the testing images. The testing images were totally different from the training images. Accordingly the results were obtained as shown in Table1 and Fig.2.

V. EXPERIMENTAL RESULTS

The proposed technique has been experimented on peripheral blood smear which are obtained from two places [14][16]. A microscopic blood images of size 184 x 138 are considered for evaluation. We have considered 80 images for our database. Out of the total 80 images, 40 images were of AML and 40 images were of normal (non-AML). These images were used for training the system. A separate database of 14 images were considered for testing the system. In HD the nucleus is extracted from the image in a distinct manner. The edges of the nucleus is extracted in a pronounced manner in LBP. Therefore, a conclusion is made that using HD and LBP together can give better performance. Their individual performance does not give as much accuracy as their combined performance gives. The combination of HD and LBP features gave an accuracy of 100% for training images and 85.71% for testing images. So, in our project we used the combination of both the features including the other features as aforementioned for analysis of the system.

To see the effectiveness of the developed algorithm, we performed a trial where we have taken some real time images that were obtained from the local hospital. These images were passed through our system and observed that the system gave an accuracy of 83.33%. This trial was done for 30 images. The obtained results further supported the impact of HD and LBP operator which were applied together. The presented system not only enables the classification of whole images but also presents a better performance for sub images when compared with some of the existing systems.

VI. CONCLUSION

The proposed work has conveyed the design, development, and evaluation of an automated screening system for AML in blood microscopic images. It uses high-quality images obtained from the American Society of Hematology [13]. The presented system performs an automated processing, including color correlation, segmentation of the nucleated cells, and effective validation and classification. A feature set exploiting the shape, color, and texture parameters of a cell is constructed to obtain all the information required to perform efficient classification. The LBP operator and HD gave a distinguishing results for this analysis. The output of the system resulted in normal or abnormal condition of the patient. The proposed system outperforms when compared with the existing system. An accuracy of 85.71% is obtained.

REFERENCES

- [1] S.Agaian, Monica Madhukar and A. T. Chronopoulos.(2014),“Automated Screening System for Acute Myelogenous Leukemia Detection in Blood Microscopic Images”, IEEE System Journal, 2014.
- [2] A.Nasir, M. Mashor, and H. Rosline. (2011), “Unsupervised color segmentation of white blood cell for Acute leukemia images,” in Proc. IEEE IST, pp. 142–145.
- [3] B. Nilsson and A. Heyden.(2002), “Model-based segmentation of leukocytes clusters,” in Proc. Int. Conf. Pattern Recognition, vol. 1, pp. 727–730.
- [4] C. Haworth, A. Hepplestone, P. Jones, R. Campbell, D. Evans, and M. Palmer.(1981), “Routine bone marrow examination in the management of acute lymphoblastic leukemia of childhood”, vol. 34, no. 5, pp. 483–485.
- [5] S. Serbouti, A. Duhamel, H. Harms, U. Gunzer, U.M. Aus, J.Y. Mary, and R. Beuscart. “Image segmentation and classification methods to detect leukemias”. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 13, pages 260–261, 1991.
- [6] D.J. Foran, D. Comaniciu, P. Meer, and L.A. Goodell. “Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy”. IEEE Transactions on Information Technology in Biomedicine, 4(4):265–273, 2000.
- [7] F.Scotti. (2005), “Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images,” in Proc. CIMS A, pp. 96–101.
- [8] T. Markiewicz, S. Osowski, B. Marianska, and L. Moszczynski.” Automatic recognition of the blood cells of myelogenous leukemia using svm”. In Proceedings of IEEE International Joint Conference on Neural Networks, volume 4, pages 2496–2501, August 2005.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

- [9] HazwaniAbd Halim, MohdYusoffMashor, and Rosline Hassan. "Automatic blasts counting for acute leukemia based on blood samples". *International Journal of Research and Reviews in Computer Science*, 2(4):971–976, 2011.
- [10] J. Angulo, J. Serra, and G. Flandrin." Quantitative descriptors of the lymphocytes". In *Proceedings of the 7th Congress of the European Society for Analytical Cellular Pathology*, pages 69–70, France, April 2001.
- [11] R.Rangayyan, *Biomedical Image Analysis*. Series title: *Biomedical Engineering*. Boca Raton, FL, USA: CRC Press, Dec 2004.
- [12] K. Nallaperumal and K. Krishnaveni. (2008), "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.
- [13] F. Scotti. (2006), "Robust segmentation and measurement techniques of white cells in blood microscope images," in *Proc. IEEE Conf. Instrum.Meas.Technol.*, pp. 43–48.
- [14] American Society of Hematology [Online]: <http://www.hematology.org>
- [15] MedlinePlus: Leukemia National Institutes of Health. [Online]. Available: <http://www.nlm.nih.gov/medlineplus/ency/article/001299.html>.
- [16] "Markandey Hospital, Solapur, Maharashtra".
- [17] F. Sadeghian, Z. Seman, A. Ramli, B. Kahar, and M. Saripan. (2009), "A frame work for white blood cell segmentation in microscopic blood images using digital image processing," *Biol. Procedures Online*, vol. 11, no. 1, pp. 196–206.