VOL. 10, NO. 16, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

NAÏVE BAYESIAN CLASSIFIER FOR ACUTE LYMPHOCYTIC LEUKEMIA DETECTION

Sriram Selvaraj[1] and Bommannaraja Kanakaraj[2]
[1]Department of Biomedical Engineering, P.S.N.A College of Engineering and Technology, Dindigul, Tamilnadu, India
[2]K. P.R. Institute of Engineering and Technology, Coimbatore, India
E-Mail: sriram@psnacet.edu.in

ABSTRACT

Leukemia is a malignant neoplasm of the blood and is one of the major causes of death throughout the world. Acute Lymphocytic Leukemia is the most common type of Leukemia and it generally affects children and adults above fifty years of age. Examination of Peripheral blood smear Image is one of the most widely used technique for Leukemia detection though it suffers from problems such as subjective interpretations, operator tiredness and efficiency. The objective of this work is to develop a method that can classify lymphocyte and lymphoblast nuclei from the ALL-IDB2 dataset. In the present paper, from a set of forty Images of Lymphocytes and Lymphoblasts, nucleus is segmented using K-means clustering method after which a set of features are extracted. Naive Bayesian classifier was used in this work for Lymphocyte classification which yielded 75% accuracy.

Keywords: Leukemia, acute lymphocytic Leukemia, segmentation, K-means clustering, Naïve Bayesian classifier.

1. INTRODUCTION

Human blood mainly consists of three main cells namely, Red Blood cells (RBC'S), White Blood cells (WBC's) and Platelets. Examination of the number and shape of these cells will give us vital information about the well being of an individual [1]. Excess number or Lack of either Red Blood cells or White blood cells will indicate certain characteristic diseases such as Anemia or Leukemia. Leukemia is a state in the human body where the human body produces excess number of immature white blood cells. These excess White blood cells prevent the production of normal white blood cells and hamper their normal functioning. Leukemia can be broadly classified into Acute or chronic based on the quickness with which the disease develops and spreads [2]. It can be further classified into the following types based on the affected cells

a) Lymphocytic Leukemia

b) Myelogenous Leukemia.

Leukemia results in suppression of hematopoiesis and thus resulting in anemia, thrombocytopenia and neutropenia. Diagnosis of Leukemia poses serious challenges because of the nonspecific nature of the signs and symptoms. Though there are advanced techniques available for diagnosis of Leukemia such as in situ hybridization (FISH), immunophenotyping, cytogenetic analysis and cytochemistry, they are found to be time consuming and costly [3].So the most simple and efficient method used to diagnose Leukemia is examination of Blood smear Image. But the Problem with this method is that it is subjective, tedious and susceptible to human errors. Over the years some methods have been devised to automate this process.

2. RELATED WORK

Nucleus segmentation

SubrajeetMohapatra *etal*.,[3] proposed a method in which segmentation is performed in two stages for segmenting the White Blood cell Nucleus using color based clustering. Initial segmentation is achieved by K-means clustering followed by nearest neighbour classification in L*a*b space.

SubrajeetMohapatra *etal*. then used other methods such as fuzzy clustering technique [4] and kernel induced Rough C Means clustering [5] for nuclei segmentation in White Blood cells. The same authors employed a functional Link Neural architecture [6] to segment Cytoplasm, Nucleus and Background region in a lymphocyte cell. Abd.Halim *etal.,* [7], devised a segmentation based on HIS color space to improve the Image quality

Feature extraction

Apart from features such a Area, Perimeter, Convex Area, Solidity, Major Axis Length, Orientation, Filled area and Eccentricity, Fabio Scotti[8]also extracted rectangularity and circularity of the cell for identification of Leukemia.

In almost all their works related to Leukemia detection SubrajeetMohapatra *et al*.,[2,3,4,6]extracted the following features.

a) Fractal Features Which included Hausdorff Dimension and Contour signature

b) Shape Features such as area Perimeter, Compactness, Solidity, Eccentricity& Form factor

c) Texture Features-Homogeneity, Energy, Correlation and Entropy.

Classification

Fabio scotti [8, 9] used the K-Nearest Neighbor classifier and considered various norms such as Euclidean, Cubic and Manhattan for classifying a normal lymphocyte and a lymphoblast to aid in Acute Leukemia detection. The same method was used by him to classify Leukocytes [9]. He also used Feed forward neural network and Linear Bayes normal classifier and Radial Basis Function to accomplish the same task and compared their performances. With the set of features extracted SubrajeetMohapatra*et al*., [2, 3, 4, 6] preferred to use Support Vector Machine (SVM) as their classifier for Acute Leukemia detection.

SubrajeetMohapatra*etal*., [10] deployed a ensemble based classifier system which predicts by combining several diverse classifiers to classify Lymphocytes and Lymphoblasts to facilitate Acute Leukemia detection. In their ensemble they have used Naïve Bayesian, K-Nearest Neighbor, Multilayer Perceptron, Radial Basis function networks and Support Vector Machine (SVM) to get the desired output.

3. METHODS

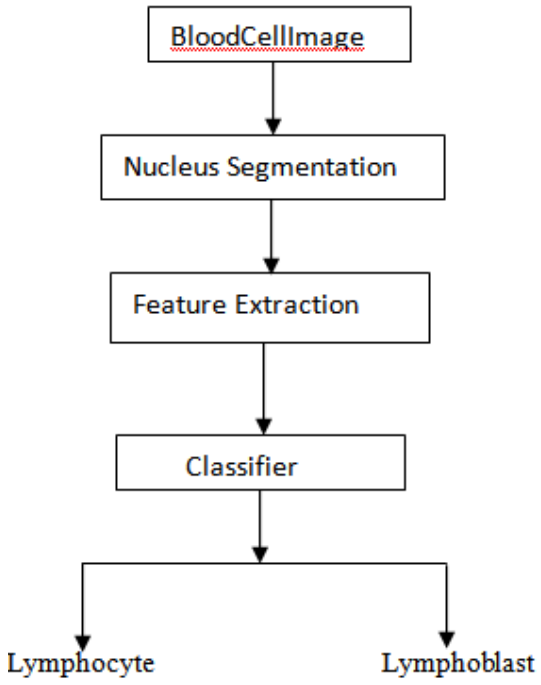

Figure-1. Flow chart of the proposed method.

Figure-1 gives the flowchart of the various steps involved in Acute Leukemia detection used in this paper.

A) Image acquisition

Images of lymphocyte cells, both from normal subjects and people suffering from Acute Lymphocyte Leukemia are taken with permission from ALL-IDB, a public Image dataset of peripheral blood samples of normal individuals and Leukemia patients, maintained by Fabio scotti, Department of Information Technology, University of Milan, Italy. The Images from ALL-IDB2 have been used for thiswork.ALL-IDB2 is a collection of normal and blast lymphocyte cells. All the Images in the dataset are in JPG format with 24 bit color depth, and a native resolution equal to 2592*1944 captured with a Power Shot G56 camera. The magnifications of the Images range from 300 to 500

B) Nucleus segmentation

The K-means clustering method is used to segment the nucleus from the normal and abnormal lymphocyte cells got from ALL-IDB2 dataset. K-Means clustering is a semi supervised clustering method that is used to create K clusters from n observations. Segmentation is achieved in such a way that objects within a cluster are as close as possible and at the same time they are far away from objects in other clusters.

C) Feature extraction

If the input data or Image is too large it may have irrelevant or redundant data. The process of selecting only the relevant information from the input data to reduce its size and complexity is called as feature extraction. Features are extracted from the segmented nuclei of lymphocyte cells from the ALL-IDB2 dataset .In this work two types of features namely the shape features and the densitometric features are extracted from the Lymphocyte nuclei.

a) Shape features

*Area: Area is determined by calculating the total number of nonzero pixels within the Image region.

*Perimeter: It is evaluated by measuring the distance between Successive boundary pixels.

*Compactness: Compactness is calculated by the following formula:

Compactness= $(\text{Perimeter})^2/\text{Area}$

*Form Factor: This feature is used to measure the surface irregularities.

Form factor is measured by the formula

Form Factor = $(4*\text{Pi}*\text{Area})/(\text{Perimeter})^2$

b) Densitometric features

Apart from the above mentioned shape features the following densitoimetric features have also been extracted from the segmented lymphocyte nuclei.

***Energy:** This feature is used to measure the degree of uniformity of a given Image.

***Entropy:** It is used to measure the randomness of a given Image

***Correlation:** This feature gives an idea about the correlation between Pixel values and its neighbourhood.

***Variance:** Variance is used to find how each pixel varies from the neighbouring pixel and is used in classification of different regions.

D) Classification

Classification is the process of grouping an image into one or several distinct and exclusive classes. In this work Naïve Bayesian classification method is used for segregating a Lymphocyte Nucleus Image into a Normal Lymphocyte or an abnormal Lymphoblast because of its simplicity and reliability.

Naïve Bayesian classifier belongs to a family of simple probabilistic classifier based on applying Baye's theorem with strong independence between the features. Naive Bayesian classifier assumes that the value of a feature which is independent of the value of any other feature.

Naïve Bayesian classifier combines the Naïve Bayes Probability model with a decision rule. The decision rule is known as maximum aposteriori rule and it picks the most probable hypothesis. Bayes classifier is a function that assigns a class label $y=C_k$ for some K as given below:

$$y = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}} P(C_k) \prod_{i=1}^{n} P(X_i / C_k)$$

4. PERFORMANCE ANALYSIS

Performance of the Naïve Bayesian Classifier for Automated Leukemia detection is evaluated by calculating the performance metrics such as Accuracy, Sensitivity and specificity from the confusion matrix.

In Binary classification method, Positive is considered as identified and Negative as rejected. So generally TP, TN, FP and FN can be defined as follows:

TP (True Positive): Correctly Identified
TN (True Negative): Incorrectly Identified
FP (False Positive): Correctly rejected
FN (False Negative): Incorrectly Rejected

Performance measures can be calculated as follows:
Accuracy= {(TP+TN)/(TP+FP+TN+FN)}*100%
Sensitivity={TP/(TP+FN)}*100%
Specificity={TN/ (TN+FP)}*100%

5. RESULTS AND DISCUSSIONS

The method Proposed in Figure-1 is implemented on dataset ALL-IDB2 for Automated Leukemia detection and the steps given there are followed sequentially.

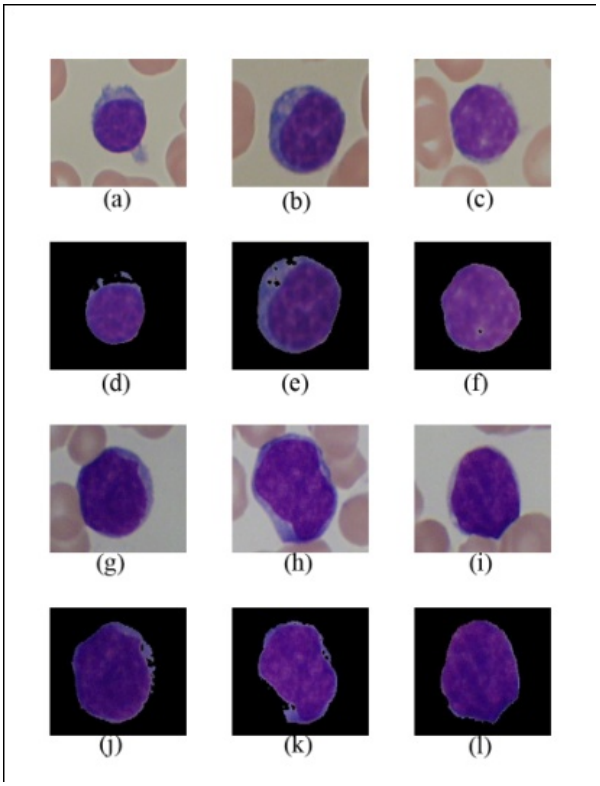

Figure-2. Results of segmentation of nuclei by K-means clustering. d,e,f represent segmented nuclei from lymphocytes a,b,c and j,k,l represent segmented nuclei from lymphoblasts g,h and I respectively.

A set of 40 Images are taken from ALL-IDB2 Dataset, out of which 20 are samples of normal lymphocytes and 20 are samples of cancer affected Lymphocytes called as Lymphoblasts. All the Image samples are subjected to Segmentation as per the method suggested in III (B).The segmented output of cell Nuclei Images of three normal Lymphocytes and three abnormal Lymphoblasts are given in Figure-2.

After the nucleus of the Lymphocyte is segmented as described in III (B), a set of features are extracted from them so that they may be fed to a classifier. The set of features extracted from two sample nuclei, one from a lymphocyte and the other from a Lymphoblast are given in Table-1.

Table-1. Feature values of a sample Lymphocyte and a Lymphoblast.

Feature	Lymphocyte	Lymphoblast
Area	4873	5536
Perimeter	603	1306
Compactness	74.6170	308.0989
Form factor	0.1683	0.0407
Energy	1.5203	1.5314
Entropy	3.2310	3.1383
Correlation	-0.3158	-0.3405
Variance	2.6846e+05	2.6717e+05

The features mentioned in thetable.1 are used as inputs to the classifier described in III (D)

The Naïve Baye's classifier that we have used in this work for classification of Lymphocytes and Lymphoblasts in the datasetAL-IDB2 yielded 75% accuracy. Other Performance metrics of the classifier are given in the Table-2.

Table-2. Performance values of Naïve Bayesian classifier used for Leukemia detection in ALL-IDB2.

Performance metric	Value
Accuracy	75%
Sensitivity	77.78%
Specificity	72.72%

6. CONCLUSION AND FUTUTRE WORK

In this work a Naïve Bayesian classifier has been proposed for classification of normal Lymphocytes and abnormal Lymphoblasts from ALL-IDB2 which yielded 75% accuracy. The results give us encouragement to use other modern classification techniques for the same dataset so that Acute Lymphocytic Leukemia can be detected more accurately. In future same methods can be applied to real time Blood samples Images which will help the clinical expert in Acute Leukemia Detection.

ACKNOWLEDGEMENT

The authors would like to thank Professor Fabio Scotti, Department of Information Technology, and University of Milan, Italy for granting them Permission to use the dataset ALL-IDB Prepared and maintained by him [11] which is an image dataset of Peripheral Blood samples of Normal individuals and Leukemic Patients.

REFERENCES

[1] Vincenzo Piuri and Fabio Scotti. 2004. Morphological Classification of Blood Leucocytes by Microscopic Images. In:Proc. IEEE International conference on computational intelligence for Measurement systems and Applications, Boston, MA.USA. pp. 103-108.

[2] Fabio Scotti. 2005. Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscopic Images.In:Proc of IEEE International conference on computational intelligence for Measurement systems and Applications, Giardini Naxos,Italy. pp. 96-101.

[3] Fabio Scotti. 2006. Robust segmentation and Measurement Techniques of White Cells in Blood Microscopic Images. In: Proc. IEEE Instrumentation and Measurement Technology conference, IMTC, Sorrento, Italy. pp. 43-48.

[4] SubrajeetMohapatra and DiptiPatra. 2010. Automated Leukemia Detection using Hausdorff Dimension in Blood Microscopic Images. In: Proc. IEEE international conference on Emerging Trends in Robotics and communication Technologies (INTERACT), Chennai, India. pp. 64-68.

[5] SubrajeetMohapatra and DiptiPatra. 2010. Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images. In:Proc IEEE International Conference on Systems in Medicine and Biology.pp. 49-54, IIT Kharagpur,India.

[6] SubrajeetMohapatra, SushantaShekarSamanta, DiptiPatra and SanghamitraSatpathi. Fuzzy Based Blood Image Segmentation for Automated Leukemia Detection. In:Proc in IEEE international conference on Devices and communication (ICDeCom),Mesra. pp. 1-5.

[7] N.H.AbdHalim, M.Y.Mashor, A.S.Abdul Nasir, N.R.Mokhtar and H.Rosline. Nucleus segmentation Technique for Acute Leukemia. Published in IEEE 7thInternational colloquiumon Signal Processing and its Applications, Penang. pp. 192-197.

[8] SubrajeetMohapatra, DiptiPatra,SunilKumarandSanghamitraSatpathi. 2012. Kernel Induced Rough c-means clustering for Lymphocyte Image Segmentation. In: Proc. 4thInternational conference on Intelligent Human Computer Interaction. pp. 1-6, Kharagpur.

[9] SubrajeetMohapatra, DiptiPatra, Sunil Kumar and SanghamitraSatpathi. 2012. Lymphocyte Image Segmentation Using Functional Link Neural

Architecture for Acute Leukemia Detection. Biomedical Engineering Letters. 2: 100-110.

[10] SubrajeetMohapatra, DiptiPatra, and SanghamitraSatpathi. 2013. An ensemble classifier system for early diagnosis of acute Lymphoblastic Leukemia in blood microscopic images. Neural Computing and Applications. l24: 1887-1904.

[11] RuggeroDonida, Labati, Vincenzo Piuriand Fabio scotti. 2011. ALL-IDB: The Acute Lymphoblastic Leukemia Image Database for Image Processing. Published in 18th IEEE International conference on Image Processing (ICIP), Brussels.