Acute Lymphoblastic Leukemia diagnosis in microscopic blood smear images using Texture features and SVM classifier

S. Alagu [*], K. Bhoopathy Bagan [**]
[*] Teaching Fellow, Department of Electronics Engineering, Anna University, India
[**] Professor, Department of Electronics Engineering, Anna University, India

Abstract- Acute lymphoblastic leukemia (ALL) is the most common cancer in children and adults. Leukemia produces a large number of immature blood cells in the bone marrow. The detection of leukemia in the earlier stage is important before it spreads into the blood streams and other vital organs. For decades, the diagnosis of leukemia has been done by experienced operators and it is a time consuming task for pathologists. The computer aided detection of acute lymphoblastic leukemia using supervised learning is discussed in this paper. The proposed method reduces the diagnostic time and gives better accuracy. The microscopic blood smear images from the database are preprocessed and segmented as three clusters based on shape, color and texture using k-means clustering algorithm .The texture features are extracted by grey level co-occurrence matrix (GLCM) and local binary Pattern (LBP). Support vector machine (SVM) with Gaussian radial basis function (RBF) as kernel is used for classification. The proposed methodology is tested for 367 images from ALL-IDB. The accuracy of 90.5% for ALL-IDB1 and 95.3% for ALL-IDB2 are obtained using SVM classifier and the results are compared with other standard classifiers such as Linear Discriminant (LD), Ensemble (Bagged trees) and KNN.

Index Terms- Acute lymphoblastic leukemia (ALL), Blood smear images, Acute lymphoblastic leukemia image database (ALL-IDB), Support vector machine (SVM) and K-Nearest Neighbors (KNN).

I. INTRODUCTION

Bone marrow is the soft and flexible tissue available in bone cavities which can generate millions of blood cells every day. Three types of blood cells substantially produced by the bone marrow are platelets, erythrocytes and leukocytes (white blood cells-WBC). WBC is responsible for the human immune system. Acute lymphocytic leukemia known as acute lymphoblastic leukemia is the type of cancer that occurs in white blood cells [4, 6]. Mostly children are affected by acute lymphocytic leukemia rather than adults. The most common types of leukemia are acute lymphoblastic leukemia, acute myeloid leukemia (AML), chronic lymphoblastic leukemia (CLL) and chronic myeloid leukemia (CML) [2]. Several methods have been proposed for the diagnosis of leukemia over the past ten years.

In this paper, the blood smear images of healthy persons and patients with acute lymphoblastic leukemia are obtained from the database available in the internet. The images were preprocessed and segmented using k- means clustering as three clusters based on the shape, color and texture .GLCM and local binary pattern techniques are used as texture operators. SVM classifier is adopted for classification purpose.

Acute lymphoblastic leukemia is diagnosed using texture features and SVM classifier with Gaussian radial basis function. The proposed method gives better accuracy than other standard classification

algorithms such as Linear Discriminant (LD), Ensemble (Bagged trees) and KNN and results are tabulated in the section IV.

The remaining of the paper is organized as follows. Related works are presented in section II and the proposed method is introduced in section III. In section IV, Image acquisition, the segmentation and classification results are shown and the metrics used for performance evaluation are also discussed. Finally, the conclusion and perspectives on future works are given in section V.

II. RELATED WORKS

Several methods have been used for the diagnosis of leukemia over recent years and some of the works related to the detection of leukemia cells are discussed in brief in this section.

AimiSalihah et al [1] proposed colour image enhancement techniques and morphological features for leukemia cell detection. Contrast stretching, Bright stretching and Dark stretching are used to identify the blast cell as either Acute lymphoblastic leukemia cell or Acute Myeloid leukemia cell. Histogram technique is proposed for validating the results. The contrast stretching gives more accurate results than other methods. In [6], Himali Vaghela et al recommended histogram equalization and linear contrast stretching methods to detect the leukemia cells either as acute or chronic. Watershed transform method was adopted for nucleus segmentation to detect cancer cells. K- means clustering and shape based features are proposed to identify the blast cells with the accuracy of 72% and 73.5% respectively.

Luis Vogado et al [9] proposed leukemia cells segmentation based on the multi-space color channel. K-means clustering and morphological operations were used for cancer cell detection. The results are validated by different performance measures of accuracy and kappa index with the normalized values of 0.912 and 0.93 respectively. In [10], Silva and Kelson Aires proposed method for classification of leukemia cells using convolutional neural network. Alexnet, Caffenet and Vgg-f networks architectures are developed and images features are extracted according to the gain ratios for further classification. Transfer learning is adopted for the classification of images with different characteristics obtained from the different image databases.

Preeti Jagadev and Virani [14] proposed the method for the detection of leukemia and its types by using image processing techniques. Marker controlled water shed algorithm, k means clustering algorithm and HSV color based algorithms are used segmentation. SVM classifier is adopted for classification of different types of leukemia cells. Ruggero Donida Labati et al [15] discussed the details about acute lymphoblastic leukemia image data base where samples have been collected by experts of Tettamanti research Centre for childhood leukemia and hematological diseases, Monza, Italy and the morphological operations were used for identifying leukemia cells.

Subrajeet Mohapatra and Dipti Patra [16] proposed a method for leukemia detection using Hausdorff dimension and contour signature in blood microscopic images. K-means clustering is used to separate white blood cells from platelets and erythrocytes. Shape and texture features are used to detect the leukemia cells. Fractal features i.e. Hausdorff dimension is implemented to estimate the perimeter roughness and to classify lymphocytic cell nucleus. The extracted features are given as input for SVM classifier. The classification accuracy of 90% is obtained.

In [18], Shailesh Mishra and Deshmukh recommended the method for detecting leukemia using morphological features and image processing techniques. Dilation and border extraction are carried out to identify the blast cells by using morphological features. Thanh et al [19] proposed a system for identifying leukemia cells using convolutional neural network for clinical decision support system. Convolutional neural network with two convolutional layers of size 50x50x30, one max pooling layer of 25x25x30, one fully connected layer and one soft max layer is recommended and obtained the accuracy of 96.03%.

In [20], Worawut Srisukkham et al suggested an intelligent leukemia diagnosis with Bare-Bones PSO based Feature optimization technique. Marker –controlled watershed algorithm is adopted for lymphocytic membrane identification. Separation of nucleus and cytoplasm is achieved by stimulating discriminant measure (SDM) - based clustering algorithm. Initially 80 raw features are obtained. Feature optimization techniques are adapted to exhibit the optimized features. SVM classifier is used to classify lymphocytes and lymphoblasts using the identified optimized feature subsets.

III. PROPOSED METHODOLOGY

The proposed method for the automatic detection of acute lymphoblastic leukemia in microscopic blood smear images consists of four steps such as image pre-processing, image segmentation, feature extraction and classification. Fig.1 shows the system architecture for efficient detection and classification of acute lymphoblastic leukemia.

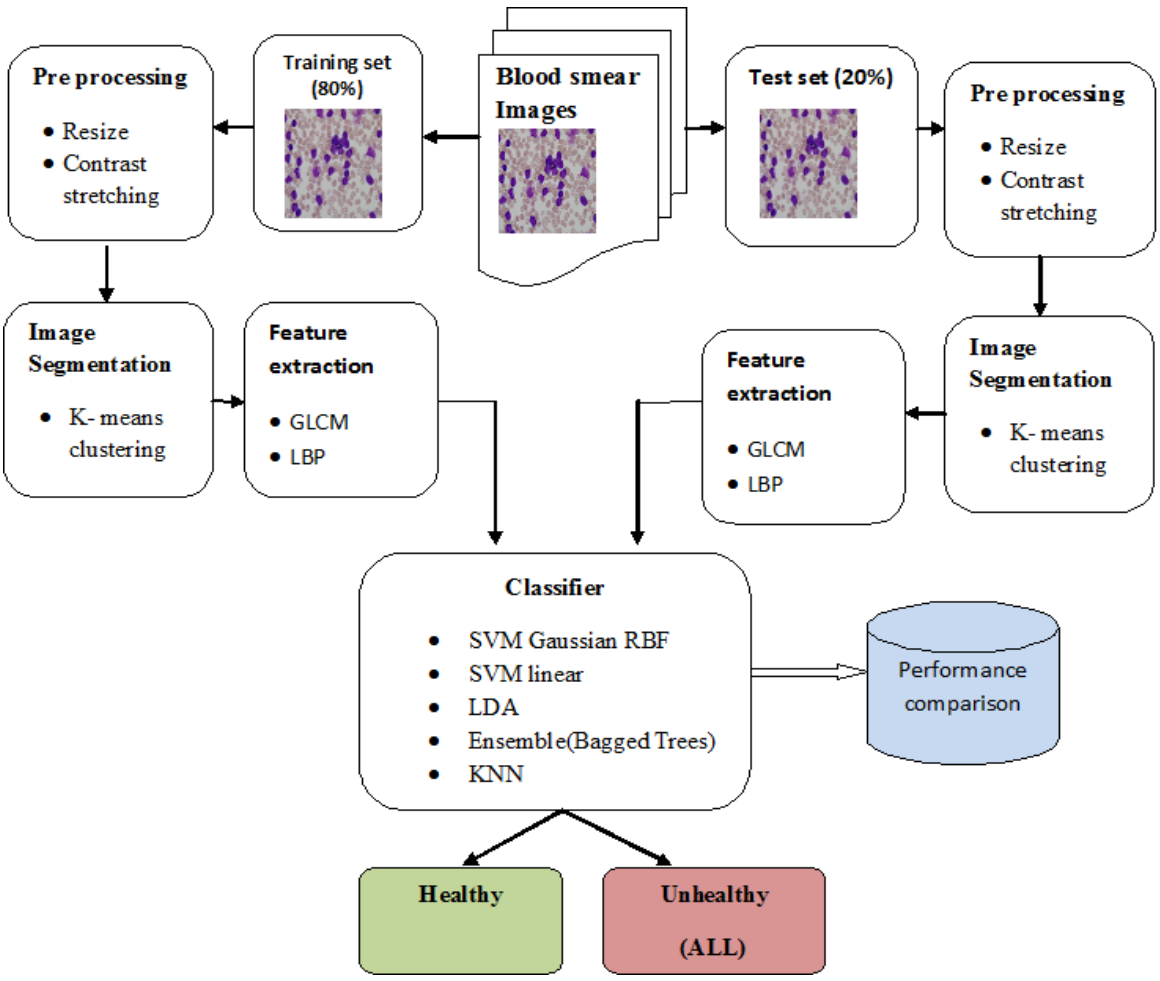

Fig.1: Proposed system architecture

A. Image Preprocessing

Image preprocessing is applied to the raw image in order to enhance the image features [1].Image resizing and contrast stretching are the two important preprocessing techniques adopted in the proposed work. The input images are resized to 256x256 sizes and the quality of the image is enhanced by contrast stretching technique.

B. Image segmentation

Image segmentation algorithms are applied to partition an image as regions which have more similarity based on some predefined criterion [3]. Region splitting and merging, region growing, thresholding and clustering are some of the techniques used for image segmentation [8]. K-means algorithm is adopted in the proposed work which is an iterative technique used to partition an image into K clusters. The selection of cluster centers are done by either randomly or based on some heuristic approach.

Each pixel in the image is assigned to the cluster that minimizes the distance between the pixel and the cluster centre. The cluster centers are re-computed by averaging all the pixels in the cluster. The steps are repeated until convergence is attained, that is until no pixels change clusters.

K-means clustering algorithm is proposed for image segmentation. The steps involved in the algorithm are given as steps. First, the microscopic blood images are read from database. The color image is

converted from RGB to L*a*b space. Colors are classified using k-Means clustering in a*b* space. Each pixel in the image is labeled from the results of k- means. The value of k is taken as three to obtain three clusters based on the shape, colour and texture.

C.Feature Extraction

The performance of a classifier depends on the features extracted. The features must be less sensitive to any actions in the image such as zooming the image, scaling and changing the orientation. The purpose of feature extraction is to represent the most relevant and important information from the image and present them in the lower dimensionality space (2). The prominent features extracted include shape, texture and color. In the proposed work, texture features are extracted from segmented image and used for classification.

D.Gray level co-occurrence matrix

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix. The GLCM is used to characterize the texture of an image by calculating a spatial relationship between pairs of pixel with specific values as matrix.

E. Local Binary Pattern

Local Binary Pattern is a texture operator used to label the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. The assigned label values are used as input for classification process.

F. Classification

Classification maps the pixels of an image into a particular class based upon the extracted features. The classification approach is implemented in two phases, i.e. training phase and testing phase [14]. The training data composed of features extracted from the images using gray level co-occurrence matrix and local binary pattern techniques. SVM is used for classification. The classification is performed by finding a hyper-plane that differentiates input data into two classes. Nonlinear transformation is also possible in SVM by adopting kernel trick. Linear, quadratic, polynomial, Gaussian radial basis function kernels are available in SVM classifier. The default kernel function is linear. SVM classifier with Gaussian radial basis function is adopted in the proposed work to obtain better classification accuracy.

G. Evaluation

The performance of the classifier is measured using various parameters such as accuracy, sensitivity, specificity and precision. A confusion matrix is generally used to describe the performance of the classifier [10]. It clearly shows the number of instances correctly classified, the number of instances which have been wrongly classified. If Class1 is a positive class and Class2 is a negative class, then the classifier's confusion matrix could be represented as in table 1. In table 1, TP stands for true positives i.e., the number of positive instances correctly classified; FP stands for false positives i.e., the number of negative instances classified as positive by the classifier. FN stands for false negatives i.e., the number of positive instances classified as negative; TN stands for true negatives i.e., the number of negative instances classified as negative.

	Predicted class 1	**Predicted class 2**
Action class 1	TP	TN
Action class 2	FP	FN

Table 1: Confusion Matrix Structure

Accuracy is the most common measure to check the performance of the classifier. Accuracy is the ratio of the number of correctly classified images to that of the total number of images.

$$\text{Accuracy} = (TP + TN)/ (TP + FP + FN + TN) \quad (1)$$

Sensitivity is also known as the true positive rate or recall. It states the rate at which the positive instances are correctly classified.

$$\text{Sensitivity} = TP / (TP + FN) \quad (2)$$

Specificity is known as the true negative rate. This parameter states the rate at which the negative class label are misclassified as correct label values.

$$\text{Specificity} = TN / (FP + TN) \quad (3)$$

Precision is the ratio of correctly classified positive instances to that of the total number of instances that have been classified as positive by the classifier.

$$\text{Precision} = TP / (TP + FP) \quad (4)$$

F-measure is known as F1 score or F score. F-measure is the harmonic mean of precision and sensitivity.

$$\text{F-measure} = 2*TP / (2*TP + FP + FN) \quad (5)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Image Acquisition

Leukemia is a blood cancer. The diagnosis of acute lymphoblastic leukemia is carried out in the proposed work. Totally 367 images of the blood smears of leukemia patients and non- leukemia patients obtained from Acute Lymphoblastic Leukemia-Image Database (ALL-IDB).The details of the ALL-IDB is shown in the table 2.

Images	**ALL-IDB1**			**ALL-IDB2**		
	HEALTHY	BLAST	**TOTAL**	HEALTHY	BLAST	**TOTAL**
	92	15	107	130	130	260

Table 2: ALL-IDB Data set details

ALL-IDB1 and ALL-IDB2 are the two image sets available in ALL-IDB. ALL-IDB2 images have similar grey level properties to the images of ALL-IDB1. The ALL-IBD2 is a collection of cropped area of interest of normal and blast cells that belongs to the ALL-IDB1 dataset. The description of ALL-IDB1 is as follows. The ALL-IDB1 image files are named with the notation 'Im XXX_Y.jpg" were XXX is a 3- digit Integer counter and Y is a Boolean digit equal to 0 if no blast cells are present, and equal to 1 if at least one blast cells is present in the images. The image with labelled with Y=0 are from healthy persons,

and all images with label Y=1 are from acute lymphoblastic leukemia patients. Blood smear images of healthy and leukemia patients were obtained from acute lymphoblastic leukemia-Image Database (ALL-IDB). ALL- IDB1 contains 92 healthy blood cells samples and 15 blast cells whereas in ALLIDB2, healthy and leukemia cells images are 130 each. The normal and leukemia blast cells are shown in figure 2.

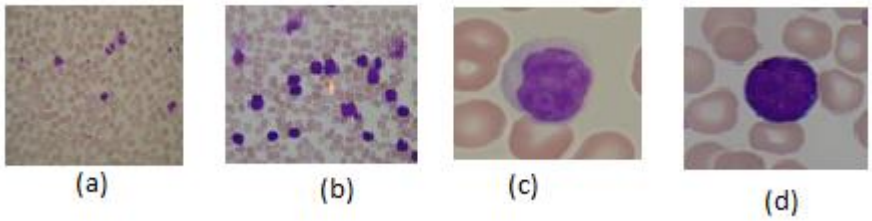

Fig. 2: Blood smear images (a) Healthy cells_ALL-IDB1 (b) Blast cells_ALL-IDB1 (c) Healthy cell_ALL-IDB2 (d) Blast cell_ALL-IDB2

ALL-IDB1 dataset images with healthy cells and blast cells are shown figures 2. (a) and (b). ALL-IDB2 dataset consists of single cell images. Figure 2. (c) Shows the healthy cell consists of nucleus and cytoplasm as separable one. In figure 2. (d), cytoplasm area are completely covered by nucleus indicates that the cell is affected by an acute lymphoblastic leukemia.

B. Image Pre-processing results

The blood smear image quality is improved by adopting image preprocessing techniques. Image resizing and contrast stretching are used in the proposed work. The images are resized to 256x256 sizes. Contrast stretching is done to preserve the edges of the blood cells as shown in figure 3.

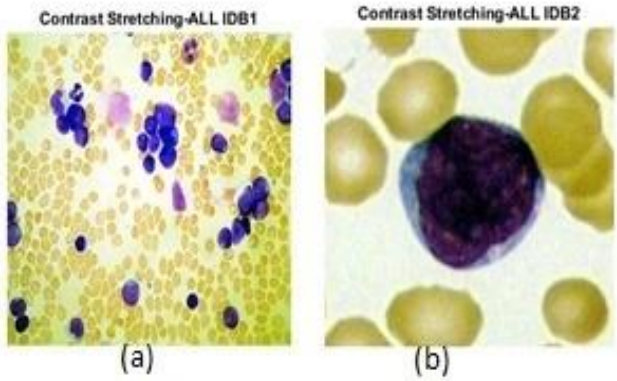

Fig. 3: Contrast stretching (a) ALL-IDB1 image (b) ALL-IDB2 image

C. Image Segmentation outputs and Feature extraction

The image is segmented after preprocessing by using k- means clustering algorithm with a cluster size of 3.Based on the colour, texture and shape the image pixels are grouped as three clusters as shown in figure 4. The texture features only separated by using the texture operators such as gray level co-occurrence matrix and local binary pattern techniques. The extracted texture features are used for further classification.

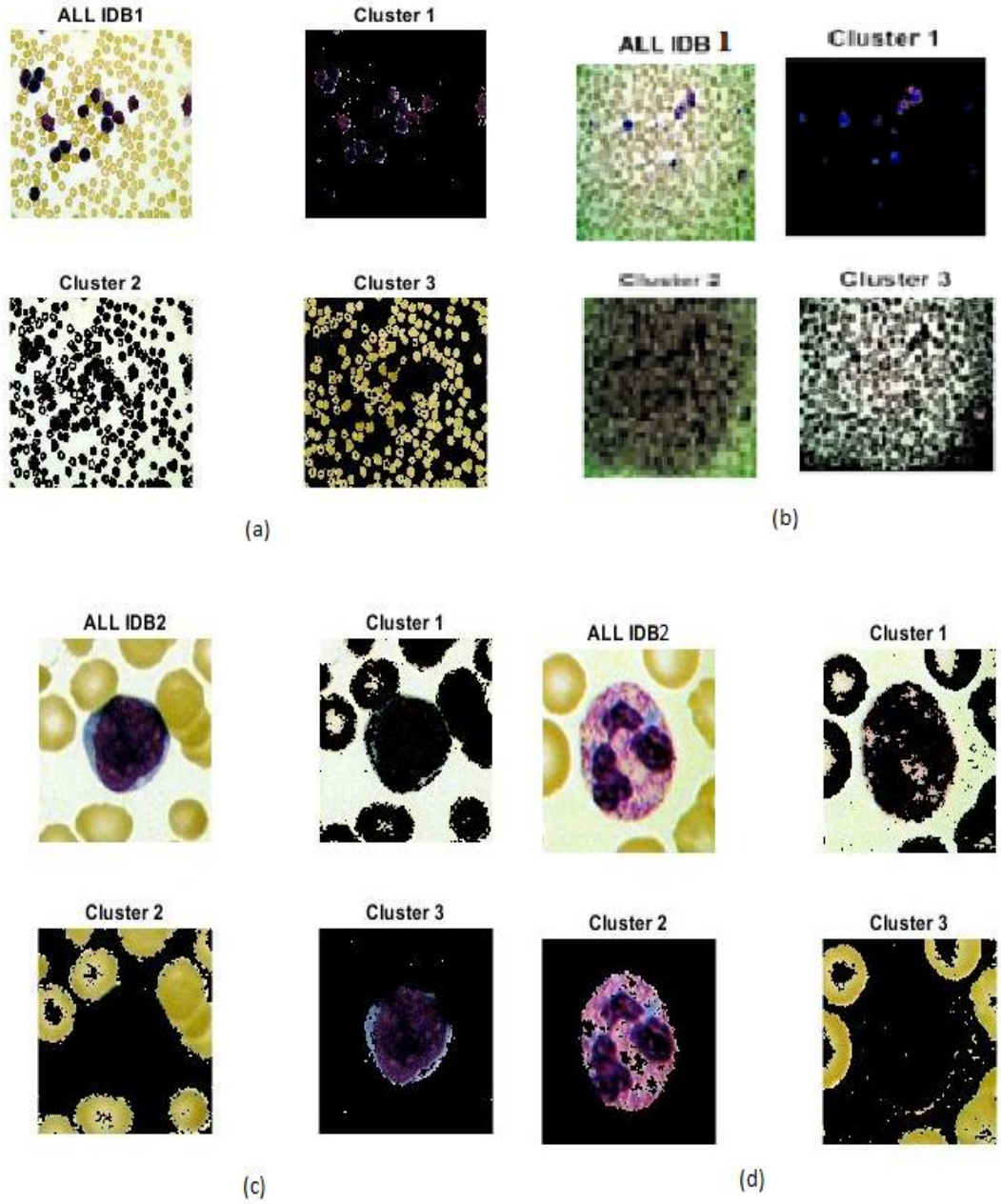

Fig. 4: Segmentation of blood smear image using K means clustering

(a) ALL-IDB1 healthy cells clusters (b) ALL-IDB1blastcells clusters

(c) ALL-IDB2 healthy cell clusters (d) ALL-IDB2 blast cell clusters

D. SVM classification results

Binary SVM is a classifier which discriminates data points into two categories. Each data point is represented by an n-dimensional vector. Maximum separation between the two classes is achieved by selecting the hyper-plane with the largest margin. The margin is the summation of the shortest distance from the separating hyper-plane to the nearest data point of both categories.

The hyper-plane correctly classifies testing data points. SVM does the mapping from input space to feature space in order to support nonlinear classification problems. The kernel trick is helpful to make a linear classification in the feature space in to nonlinear classification in the input space.

In the proposed work, SVM with Gaussian radial basis function (RBF) is employed for classification. SVM classifier defines an optimal hyper-plane that separates the data into two different classes. The SVM classification output of the test image is either as healthy blood cell or leukemia blast cell. The classification results either diseased or normal. The results are displayed with the help of MATLAB GUI commands as shown in figure 5.

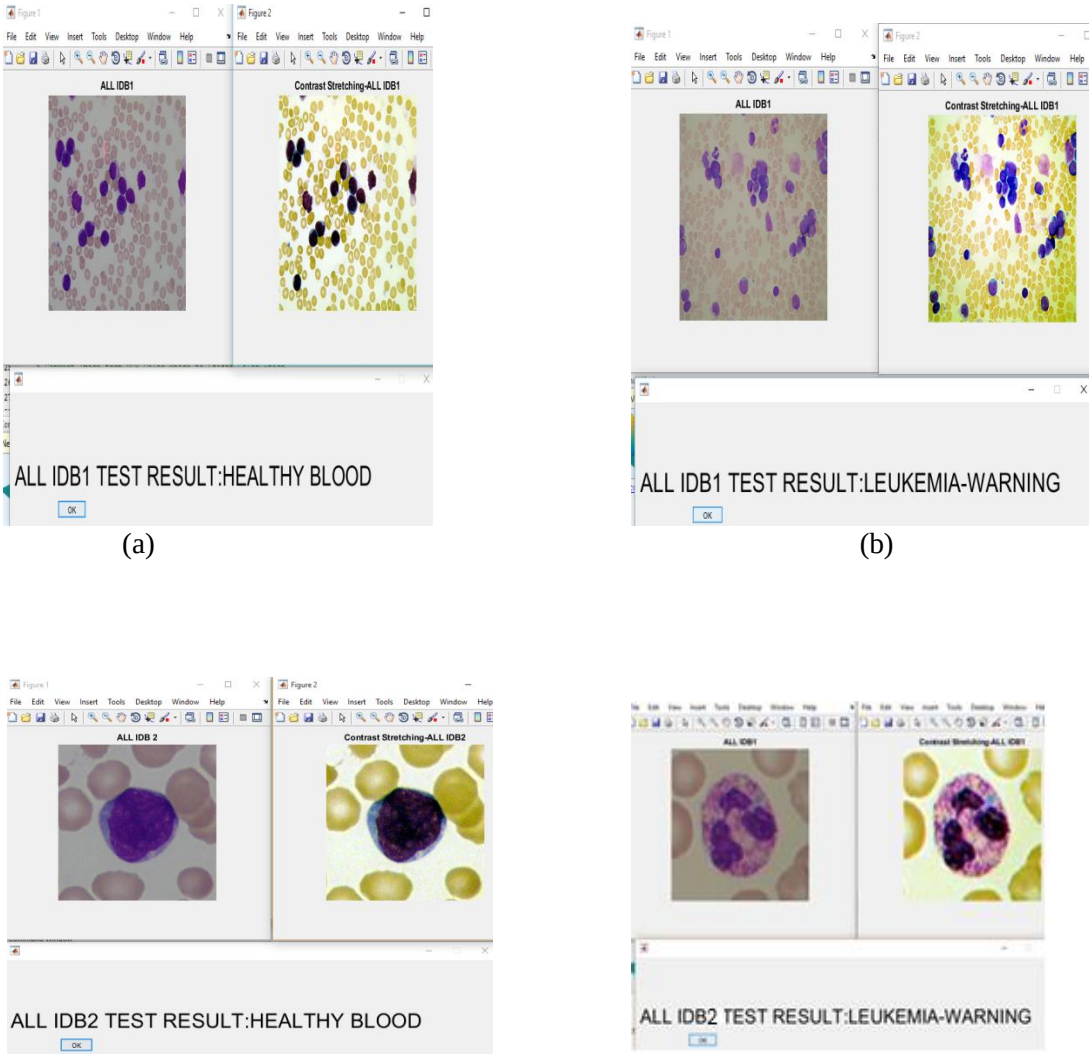

Fig. 5: SVM classifier with Gaussian radial basis function outputs

(a) Healthy person–ALLIDB1 image (b) Leukemia patient- ALLIDB1 image

(c) Healthy person–ALLIDB2 image (d) Leukemia patient- ALLIDB2 image

E. Metrics for performance evaluation

The accuracy, precision, specificity, sensitivity and F-measure of the proposed method are given in the table 3. SVM classifier with Gaussian radial function as a kernel gives the accuracy of 90.5% for ALL IDB1 and 95.30 for ALL IDB2.This is comparatively higher than SVM with linear kernel. The results are obtained by generating confusion matrix in Matlab version 2018a.The error rate is very low and the value of sensitivity is a measure that indicate the number of times the classification perfectly done. The specificity and F-measure are also calculated for the proposed SVM classifier with Gaussian radial basis function.

Metrics for evaluation	**ALL IDB1**	**ALL IDB2**
Accuracy	90.5%	95.30%
Sensitivity	87.83%	94.59%
Specificity	92.06%	95.23%
F-measure	87.25%	93.33%

Table 3: Evaluation metrics

The measured accuracy is compared with other existing classifiers such as Linear Discriminant (LD), Ensemble (Bagged trees) and KNN classifiers as shown in figure 6.The SVM classifier with Gaussian Radial Basis Function gives the accuracy of 90.5% for ALL-IDB1 and 95.3% for ALL-IDB2 .

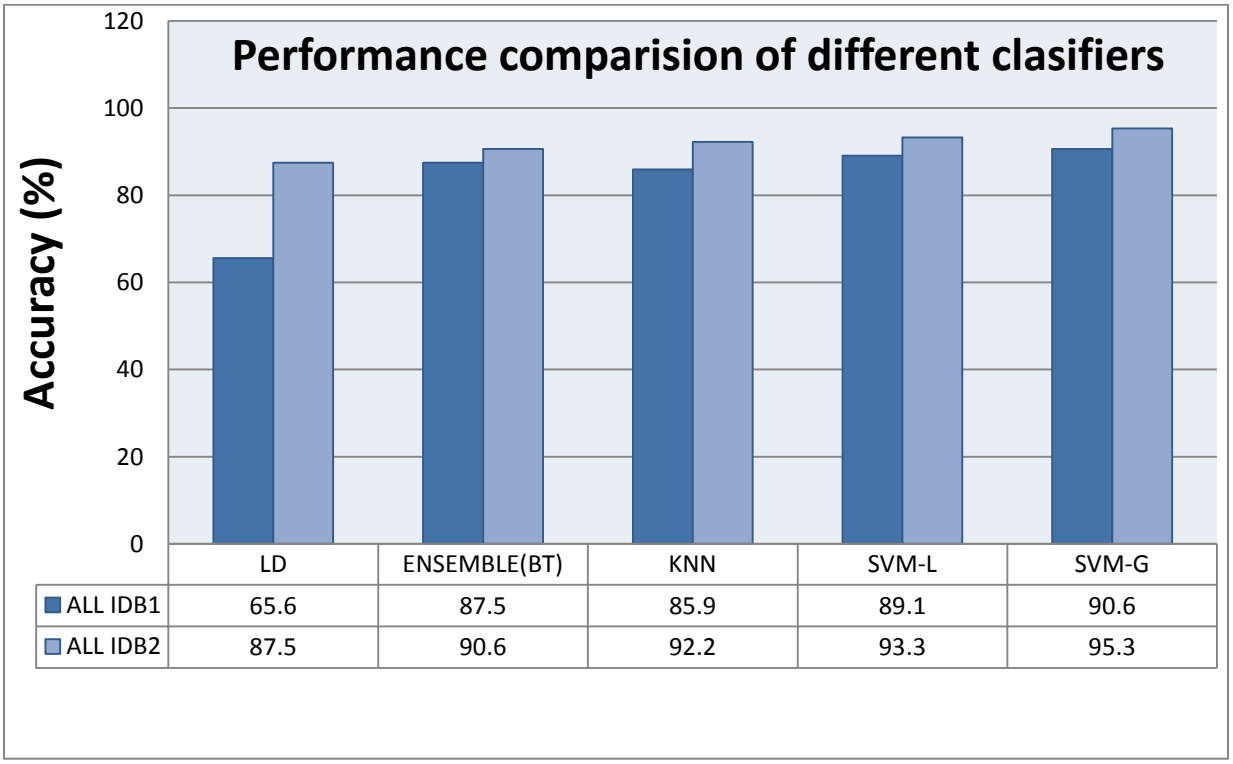

Fig. 6: Performance comparison of different classifiers

V. Conclusion

The diagnosis of acute lymphoblastic leukemia using texture features and SVM classifier is carried out and results are discussed in this paper. The proposed methodology is tested for 367 images from ALL-IDB. The accuracy of 90.5% forALL-IDB1 and 95.3% for ALL-IDB2 are obtained using SVM classifier with Gaussian radial basis function and the results are compared with other standard classifiers such as Linear Discriminant (LD), Ensemble (Bagged trees) and KNN. Deep learning algorithms, convolutional neural network (CNN) architectures will be used in future to ensure the unsupervised learning mechanisms and to obtain better accuracy. The proposed system will be validated with large amount of data and used in daily life, helping physicians and patients to diagnose the disease at the earliest.

ACKNOWLEDGMENT

The author would like to thank Labati et al for sharing the Acute lymphoblastic leukemia image database (ALL-IDB).

REFERENCES

[1] AimiSalihah, A.N, M. Y. Mashor, and Nor HazlynaHarun "Colour Image Enhancement Techniques for Acute Leukemia Blood Cell Morphological Features," IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Volume 10 – No.1, 246-251, 2010.

[2] Ben Athiwaratkun and Keegan Kang," Feature Representation in Convolutional Neural Networks", arXiv: 1507.02313v1 [cs.CV] 8 Jul 2015.

[3] EmadA.Mohammed, Mostafa M.A.Mohamed, Christopher Nauglerand, C. P. Bean, Behrouz H. Far ,'Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimum thresholding", IEEE Canadian Conference on Electrical and Computer Engineering (CCECE).,2013.

[4] Florence Yellin, Benjamin D. Haeffele, and Ren´eVidal,"Blood cell detection and counting in holographic lens free imaging by convolutional parse dictionary learning and coding", IEEE,2017

[5] Geert Litjens, ThijsKooi , Babak Ehteshami Bejnordi ,Arnaud Arindra Adiyoso Setio , Francesco Ciompi, Mohsen Ghafoorian, JeroenA.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez," A survey on deep learning in medical image analysis", Elsevier, Medical Image Analysis 42 (2017)60–88.

[6] Himali P. Vaghela, Hardik Mod, Manoj Pandya and M.B. Potdar "Leukemia Detection using Digital Image Processing Techniques," International Journal of Applied Information Systems, Volume 10 – No.1, November2015.

[7] J. M. H. Du Buf, M. Kardan and M. Spann "Texture feature performance for image segmentation," Pattern Recognition, Vol. 23, No. 3/4, pp. 291- 309, 1990.

[8] Keh-Shih Chuang, Hong-Long Tzeng, Sharon Chen, Jay Wu and Tzong- Jer Chen "Fuzzy c-means clustering with spatial information for image segmentation," Elsevier2005.

[9] Luis H. S. Vogado, Rodrigo de M. S. Veras, Alan R. Andrade and Flavio H. D. de Araujo, "Unsupervised Leukemia Cells Segmentation Based on Multi-space Color Channels", IEEE International Symposium on Multimedia, 2016.

[10]R.V. Silva and Kelson R.T. Aires." Leukemia diagnosis in blood slides using transfer learning in

CNNs and SVM for classification, Engineering Applications of Artificial Intelligence 72 (2018) 415–422.

[11]Lokman Faivdullah, FaridAzahar, ZawZawHtike, and Wei Yan NyeinNaing," Leukemia Detection from Blood Smears", Journal of Medical and Bioengineering Vol. 4, No. 6, December 2015.

[12]Nikhil R Paul and Sankar K Paul," A review on image segmentation techniques", Pattern Recognition, Vol. 26, No. 9, pp. 1277 1294. 1993

[13]Patil Tejashri G and V. B. Raskar," A Robust Segmentation Method for Acute Lymphoblastic Leukemia Detection", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January2015.

[14]Preeti Jagadev and Dr. H.G. Virani," Detection of Leukemia and its Types using Image Processing and Machine Learning", International Conference on Trends in Electronics and Informatics ICEI 2017.

[15]Ruggero Donida Labati, Vincenzo Piuri and Fabio Scotti," ALL-IDB: The acute lymphoblastic leukemia image database for image processing", 18th IEEE conference on Image processing, 2011.

[16]Subrajeet Mohapatra and Dipti Patra," Automated Leukemia Detection using Hausdorff Dimension in Blood Microscopic Images", IEEE, 2010.

[17]Sos Agaian, Monica Madhukarand Anthony T. Chronopoulos, Automated Screening System for Acute Myelogenous Leukemia Detection in Blood Microscopic Images, IEEE. Trans., 2014.

[18]Shailesh J. Mishra and A.P.Deshmukh ,"Detection of leukemia using matlab ",International Journal of Advanced Research in Electronics and Communication Engineering, Volume 4, Issue 2,February 2015.

[19]Thanh. TTP, Giao N. Pham, Jin-Hyeok Park, Kwang-Seok Moon, Suk- Hwan Lee, and Ki-Ryong Kwo, "Acute leukemia classification using convolutional neural network in clinical decision support system", pp. 49– 53, 2017. © CS & IT-CSCP 2017.

[20]Worawut Srisukkham, Li Zhang and Siew Chin Neoh, "Intelligent Leukemia Diagnosis with Bare-Bones PSO based Feature Optimization", Journal on Applied soft computing, S1568-4946(17)30148-5, 2017.