Computer Assisted Detection and Counting for Diagnosis of Blood Cancer

M.Shakunthala[1], P. Nandhini[2], B. Pooja Harini [3]
[1]*Assistant Professor, Department of ECE, R.M.D Engineering College, Tamilnadu 601206*
[2,3]*UG Students, Department of ECE, R.M.D Engineering College, Tamilnadu 601206*

Abstract- **Leukemia is a type of cancer which damages blood and bone marrow .It can be fatal illness if not diagnosed at earlier stage .Typically complete blood count (CBC) or morphological image analysis is employed to manually diagnose the malignant neoplastic disease cells. These ways are time consuming and fewer corrective measures has to be taken. In this paper it is planned for the detection of acute lymphocytic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML) by microscopic blood image analysis. Initially various kinds of cells are separated from the image i.e white blood cells, red blood cells and platelets and then lymphocytes are separated from white blood cells. Watershed segmentation is performed to separate grouped lymphocytes for counting of cells. After that form and colour options are extracted from these lymphocytes and given to SVM, KNN and CNN classifiers to classify into traditional and blast cells. Counting of the WBCs is also done for accurate diagnosis. This type of malignant neoplastic disease detection system is found to be more practical, fast and accurate as compared to manual method.**

Index terms- **Leukemia classification, White blood cell count, Watershed segmentation, SVM, KNN and CNN classifiers.)**

I.INTRODUCTION

Globally, oncology dominants the ongoing trials on human-subject research among all the therapeutic areas as its diagnosis and treatment are still emerging. Cancer has a major impact on society in the United States and across the world as it is among the leading causes of death. In the present research, we studied one of the forms of cancer; Leukemia. Leukemia is a malignancy of the blood cells generated in the bone marrow. It is characterized by the uncontrolled monoclonal multiplication of abnormal white blood cells. Leukemia is classified into different types including acute lymphoblastic leukemia (ALL), acute myelogenous leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myelogenous leukemia (CML). Acute leukemia causes the fast deterioration of the patient, whereas chronic leukemia is characterized by slow progression and may be lymphocytic or myelogenous. Hematologist and clinicians often face challenges in examining colored microscopic blood images. The process is tedious, time-consuming, less efficient, and not suitable for analyzing a large number of cells. Nevertheless, some mathematical approaches and technologies have been developed to discriminate the blood cells. As a result, machine learning (ML) methods have become a popular tool for medical researchers. These techniques can discover and identify patterns and relationships between them, from complex datasets, while they are able to effectively predict future outcomes of a cancer type. Machine learning is a branch of artificial intelligence (AI) research that employs a variety of statistical, probabilistic and optimization tools to learn from past examples and to then use that prior training to classify new data, identify new patterns or predict novel trends.

The three general types of machine learning algorithms are supervised learning, unsupervised learning, and reinforcement learning. Interestingly, almost all machine learning algorithms used in cancer prediction and prognosis employ supervised learning. Furthermore, most of these supervised learning algorithms belong to a specific category of classifiers that classify on the basis of conditional probabilities or conditional decisions. In the present research, we used the neural network (FNN- feed forward neural network and CNN- Convolutional neural network), support vector machine (SVM), and KNN (K Nearest Neighbor). Neural networks are a form of artificial intelligence that has the ability to recognize complex and highly non-linear relationships, such as are likely to characterize medical circumstances. Given the significance of personalized medicine and the growing trend on the ML techniques in developing countries, like India, where outpatient department experience heavy patient load, we proposed this study of automatic detection of acute lymphoblastic leukemia using ML techniques. We used the blood microscopic images with heterogeneous staining and pixel features, such as color or texture, with the aim to develop a system to identify the whether the sample is cancerous or not according to their morphological features. The objectives of this study were to employ ML techniques (FNN, CNN, SVM and KNN) to auto-detect the acute leukemia blast cells and develop a user-friendly Graphic-user interface (GUI) which provides flexibility regarding deployment.

Rest of the paper is organized as follows: Section II presents the literature survey of various related papers, section III describes the methodology. Sections IV presents in depth implementation details of system and the analysis of results obtained. Section V provides the concluding remarks.

II. RELATED WORK

The detection and classification of leucocytes to its different classes: Basophil, Eosinophil, Lymphocyte, Monocyte and Neutrophil is presented in [1]. The results of this classification basically point that further analysis of bloods white cells is achievable. Fabio Scotti proposes a system which classifies the presence of Acute Lymphocytic Leukemia (ALL) using a methodology to achieve a fully automated classification of ALL from microscopic blood-film images. MATLAB is used for image processing and the total numbers of features extracted are 23. Algorithms like Feed-Forward Neural Network, Linear Bayes and K-Nearest Neighbor are used for classification. This automated system offers remarkable accuracy [2]. Whereas in [3], the features extracted using MATLAB are only 5 with classifiers kNN, SVM, ANN and Kmeans clustering having accuracies 95.23%, 90.47%, 95.23% and 85.71% respectively. In [4] a greater accuracy of 95% is achieved with the use of SVM classifier. K-means clustering algorithm, Zack’s algorithm and Histogram equalization are applied for identifying and grouping WBCs. SVM is used for classification of images with an accuracy of 93.5% but further classification of type of blasts is not shown [12]. Furthermore, comparison between image segmentation method by using HIS and RGB color space for detection of blast in blood slide images is presented in [5]. The shape of blast obtained using RGB color space is not quite similar to the original blasts but the method based on HSI color space using S component provides almost similar pixel values and shape to original blasts.

Cell image segmentation techniques like watershed and fuzzy c means are discussed of which, marker-controlled watershed algorithm yield better result in detection of ALL [10]. Along with blood slide image segmentation followed by feature extraction to detect leukemia is proposed in [11]. Then classification was carried out using KNN classifier with an overall accuracy of 93%. Easy comparison can be carried out if the dataset for ALL detection used is common for the classification. The description of such a public dataset and metrics for evaluation of performance of different classification algorithms are discussed [6]. Furthermore, this public dataset was normalized and 7 most distinct features were obtained after processing of blood image which resulted in best classification performance [7]. In [8] automatic detection and counting of lymphoblasts using MATLAB with the aim to provide additional support to medical practitioners is implemented with an average accuracy greater than 90%. Additionally, this mechanism provides high speed, accuracy and scope for early detection of disease. A novel approach of selection of features based on correlation was introduced for diagnosis of ALL. The number of features extracted using this approach were 16 which gave accurate result of 92.3% [9].

III. METHODOLOGY

We have implemented the system in two modules: Image processing and classifiers. All the components are written using python programming language. In figure 1 the client interacts with the web interface or standalone application and selects the image to be tested. Now, if user chooses FNN, SVM, kNN classifier then first Image Processing module is called and then results are passed to respective classifier module which gives the final result or if user selects CNN then image is directly passed to CNN classifier and result is obtained.

Modules
- Image acquisition.
- Image preprocessing.
- Image segmentation.
- Feature extraction.
- Classification

A. Image Acquisition
Microscopic images of blood cells are acquired with the help of digital microscope. Digital microscope which has inbuilt camera inside it is in trend to acquire digital images of cell.

B. Image Preprocessing
Due to excessive stains and manual intervention microscopic images which are acquired possesses noise. Here noises present are mainly shadows of nuclei. Our region of interest is blood cell nucleus, so we process images to remove unwanted noises and recover important one. Some previous studies proved that the image enhancement technique like contrast enhancement can improve medical image quality. In this enhancement process, images are improved to make it suitable for further stages of processing. Blood cell images are enhancement with the help of linear contrast enhancement technique. Popular contrast enhancement technique is histogram equalization which adjusts the contrast and image intensity as per required.
From the literature review, the following are the steps to be followed for automatic detection of blood cancer shown in figure 1.

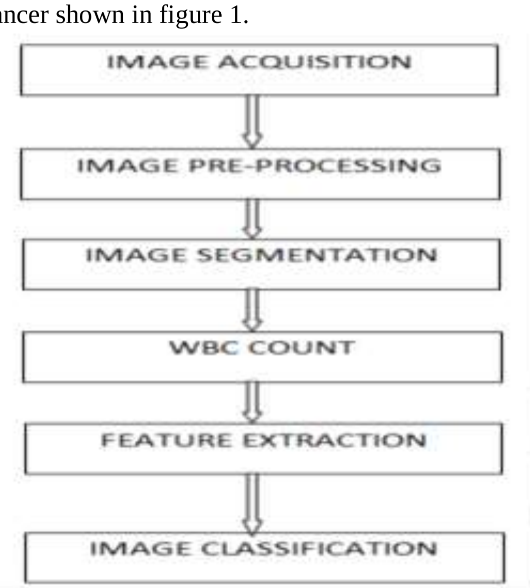

Fig 1:Block Diagram of the Proposed System

C. Image Segmentation
Image segmentation is a commonly used technique in digital image processing and it is used for analysis to partition an image into multiple parts or regions often based on the characteristics of the pixels in the image. Segmentation using computerized analysis is the process of subdividing a digital image into multiple segments and also it is the process of grouping together pixels that have similar attributes [16]. It also partitions an image into nonintersecting regions such that each region is homogeneous and the union of no two adjacent regions is homogeneous[12] .The Pixels in a region are similar based on some homogeneity criteria such as colour, intensity or texture so as to locate and identify objects and boundaries (lines, curves, etc) in an image. The accuracy of segmentation determines the eventual success or failure of computerized analysis.
Colour space conversion is the translation of the colour from one basis to another. This typically occurs in the context of converting an image that is represented in one color space to another color space, the goal is to make the translated image look as similar as possible to the original. Here we use rgb to cmyk color space conversion for blood cell segmentation.
Morphological image processing technique is a collection of non-linear operations related to the shape or morphological features in an image. Morphology is a broad set of image processing operations that process the images based on shapes. Morphological operations apply a structuring element to an input image and produces an output image of the same size.
The watershed segmentation is a classical algorithm used for the segmentation process that is for separating different objects in an image. The watershed algorithm treats pixel values as a local topography. This algorithm floods the basins from the markers until basins which are attributed to different markers meet on watershed lines. In many cases, markers are chosen as local minima of the image, from which basins are flooded.

D. Grouped Cells Identification:
Some blood images have some adjacent cells. It is really difficult to extract some features if the cells are grouped and not well separated. The features like area are very difficult to extract if the nucleus of the

cells are joined together. So our duty is to separate well these cells before studying them further. Some methods are exist to separate those grouped cells. Roundness measurement technique is used here. We are using this technique because we can identify the grouped cells easily by only analyzing the shape of them. Roundness will check whether the shape of the cell is circular or not. The theoretical value of the roundness is 1 for circular objects and less than 1 for non-circular objects. But practically after some experiments it is found that 0.9 can be used as threshold value to distinguish single and grouped cells. The objects having the value more than the threshold value are treated as single cells and objects having the value less than the threshold value are treated as grouped cells.

E. WBC Count

Algorithms exist to count WBCs in an automatic way. We are using shape based feature technique algorithm to count WBCs. We get some idea from the Viola-Jones object detection algorithm to detect and count WBCs.

F. Feature Extraction

The feature extraction starts from an initial set of measured data and builds derived values (features) which intend to be informative and non- redundant facilitating the subsequent learning and generalization steps, and in some cases feature extraction leads to better human interpretations. It is related to dimensionality reduction [15]. When a large input data is given to an algorithm and if it is suspected to be redundant, then it can be transformed into a reduced set of features (also named a feature vector). Determination of subset of the initial features is called as feature selection. The selected features contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.

Geometric Features:

Area, perimeter, centroid, eccentricity, compactness, convexity, concavity, rectangularity, elongation, solidity etc

Statistical Features:

Mean, variance, standard-deviation, skewness etc.

$$rectangularity = \frac{area}{major\ axis \times minor\ axis}$$

$$elongation = 1 - \frac{major\ axis}{minor\ axis}$$

$$compactness = \frac{4 \times \pi \times area}{perimeter^2}$$

Shape Features:

Visual features of objects are called as the shape characteristics or visual features. For example, circular object or triangular objects or other shapes, perimeter boundary of the object, the diameter of the border.

Colour Features:

Global features include colour, texture histograms and also colour layout of the whole image. Local features include colour, texture, and shape features for sub images, segmented regions, and interest points.

Texture Features:

An image texture is a set of metrics calculated in image processing which is designed to quantify the perceived texture of an image.The Image texture gives us information about the spatial arrangement of colour or intensities in an image or for a selected region of an image[19]. Here we use GLCM (Grey Level Co-occurrences Matrix) for texture feature analysis.

G. Classification

All the features are calculated and extracted are listed in a column with their values. We will get a matrix called feature extracted matrix. The image to be tested called as test image. The values of the test image features are checked with the previously calculated values based on the values of the input test image while the SVM classifier will classify that whether the test image has infected cell or not. We are also using PNN (probabilistic neural network) with the extracted feature matrix to know whether the cell is infected or not.

There are many classification algorithm available and the classification algorithm that we use are given below

1) KNN
2) CNN
3) SVM

KNN:

K nearest neighbour is an algorithm that stores all available cases and classifies new cases based on a similar measure .A case is classified based on majority vote of its neighbours, with the case being assigned to the class most common amongst its K nearest neighbours measured by a distance

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Manhattan $$\sum_{i=1}^{k}|x_i - y_i|$$

Minkowski $$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$$

In the instance of categorical variables the hamming distance must be used.

Hamming Distance

$$D_H = \sum_{i=1}^{k}|x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

CNN:

CNN have an input layer, output layer and hidden layers. The hidden layers usually consists of convolution layers, ReLU layers, pooling layers and fully connected layers. CNN works are based on extracting features from images. This eliminates the need for manual extraction. The features are learned when the network trains on a set of images.

SVM:

In machine learning support vector machine (SVM) is a supervised learning model, Which is used for both classification or regression challenges. However, it is mostly used in classification problems. In this we plot each data item as a point in n dimensional space (where n is number of features you have) with the value of each feature is taken as the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well Support Vectors are just the co- ordinates of individual observation. The SVM classifier is a frontier which is best in segregating the two classes (hyper-plane/ line).

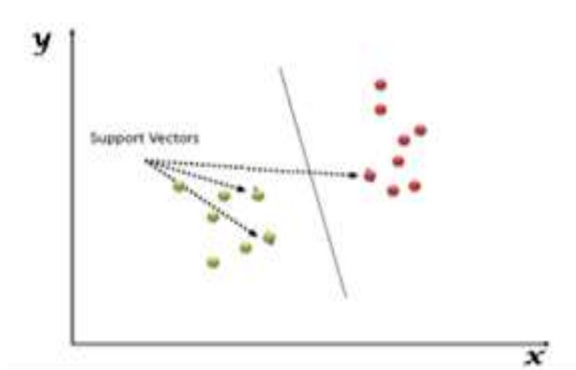

Fig.2.SVM in 2-dimensional space

IV. RESULTS AND DISCUSSION

Leukemia images obtained from online database and from Goa Medical College are used as input images. And pre-processing and segmentation steps are performed on these images. GUI is created to execute the steps of code step wise and to make it user friendly. First the RGB image is converted into HSI model. Graphical User interface has been created in order to make it user friendly. Fig 5 shows the pop up that comes after running the code.

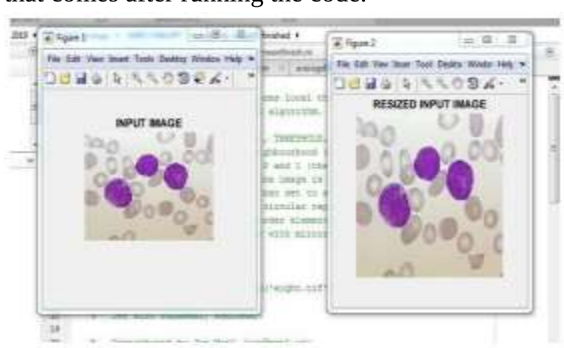

Fig 3:Input image

Fig 4: Preprocessed Image

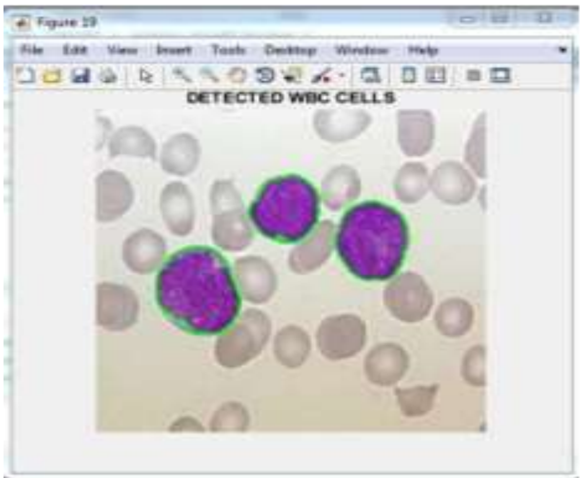

Fig 5: Segmented Image

Fig 7 shows segmented output of the selected image. The image is represented water segmented clusters. The value 4 is selected by experimenting and checking for all the images of dataset.

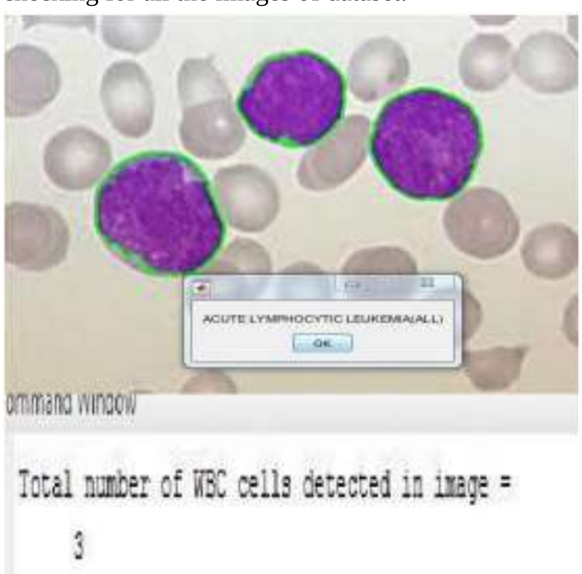

Fig.6.Cancer Classified image

V. CONCLUSION

An automated system is developed which can detect the Acute Lymphoblastic Leukemia (ALL) from the microscopic blood images to improve the accuracy and enable early detection compared to manual approach. Image processing is carried to process the microscopic blood image and extract the required attributes. Finally, the comparison between accuracy and speed of SVM, Neural Network and K-nearest Neighbour is shown. The Convolutional Neural Network (CNN) classifier model achieves the highest accuracy of 98.33% while the accuracies of the other models SVM, KNN and Feed Forward Neural Networks are 91.4%, 94.85% and 94% respectively. We have use Tensor Flow framework for implementing CNN & FNN, the network consists of 7 convolution, 7 pooling, 2 fully connected and 1 dropout layer. The developed system speeds up the process of detection of Acute Lymphoblastic Leukemia with high accuracy which makes it easy to provide early treatment to patient.

REFERENCES

[1] Vincenzo Piuri, Fabio Scotti, “Morphological Classification of Blood Leucocytes by Microscope Images”, IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, 2004

[2] Fabio Scotti, “Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images”, IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA), 2005

[3] Romel Bhattacharjee, Lalit Mohan Saini, “Robust Technique for the Detection of Acute Lymphoblastic Leukemia”, IEEE Power, Communication and Information Technology Conference (PCITC), 2015

[4] Subrajeet Mohapatra, Dipti Patra, “Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images, IEEE International Conference on Systems in Medicine and Biology, 2010

[5] Nor Hazlyna Harun, M.Y. Mashor, N.R.Mokhtar, Aimi Salihah, A.N, Rosline Hassan, R.A.A. Raof, M.K. Osman,

"Comparison of Acute Leukemia Image Segmentation using HSI and RGB Color Space", International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)

[6] Ruggero Donida Labati, Vincenzo Piuri, Fabio Scotti, "ALL-IDB: The Acute Lymphoblastic Leukemia Image Database for Image Processing", IEEE International Conference on Image Processing 2011

[7] Hayan T. Madhloom, Sameem Abdul Kareem, Hany Arin, "A Robust Feature Extraction and Selection Method for the Recognition of Lymphocytes versus Acute Lymphoblastic Leukemia", International Conference on Advanced Computer Science Applications and Technologies 2012

[8] Vasuki Shankar, Murali Mohan Deshpande, N Chaitra, S Aditi, "Automatic detection of Acute Lymphoblastic Leukemia using Image Processing, IEEE International Conference on Advances in Computer Applications (ICACA) 2016

[9] Vanika Singhal, Preety Singh, "Correlation based Feature Selection for Diagnosis of Acute Lymphoblastic Leukemia", ACM 2015

[10] Ansha Beevi, Remya R.S., "An Overview on Acute Lymphocytic Leukemia Detection using Cell Image Segmentation", IOSR Journal of Computer Engineering (IOSR-JCE) 2014

[11] Ms. Minal D. Joshi, Prof. Atul H. Karode, Prof. S.R.Suralkar, "White Blood Cells Segmentation and Classification to Detect Acute Leukemia", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 2013

[12] Nimesh Patel, Ashutosh Mishra, "Automated Leukaemia Detection using Microscopic images, Second International Symposium on Computer Vision and the Internet VisionNet 2015

[13] Madhumala Ghosh, Devkumar Das, Chandan Chakraborty, Ajoy K.Ray, "Automated leukocyte recognition using fuzzy divergence", Elsevier Micron 2010

[14] Ivan Vincent, Ki-Ryong Kwon, Suk-Hwan Lee, Kwang-Seok Moon, "Acute lymphoid leukemia classification using two-step neural network classifier", Frontiers of Computer Vision (FCV), 2015

[15] Malek Adjouadi, Melvin Ayala, Mercedes Cabrerizo, Nuannuan Zong, Gabriel Lizarraga, Mark Rossman, "Classification of Leukemia Blood Samples Using Neural Networks", Annals of Biomedical Engineering, Vol. 38, No. 4, April 2010.