A SURVEY ON ACUTE MYELOGENEOUS LEUKEMIA DETECTION IN BLOOD MICROSCOPIC IMAGES

Smita Bhole

Abstract - Acute myelogeneous leukemia (AML) is one of the types of acute leukemia. Acute leukemia progresses quickly and is mostly prevalent among adults. Mostly AML is observed in a person with the age in between 60-65 years. In medical field current method for detection of cancer involves manual examination because of that need of automation in detection of AML arises. The current method is time consuming and accuracy of result is totally depends on operator's ability of doing tests for detection. A simple technique that detects and segments AML in blood smear is presented in this paper. The proposed method includes simplicity of the developed approach and Classification of complete blood smear images

Index Terms –Acute myelogeneous leukemia (AML), White blood cells (WBCs)

1 INTRODUCTION

Leukocytes plays major role in diagnosis of diseases. Leukemia is a greek word that means "Leukos+ mia" = "White+ blood". Leukemia is a cancer of blood and bone marrow. Blood cells are produce in bone marrow. Leukemia is based on the fact that white blood cell count is increased with immature blast cells (Lymphoid or myeloid) and platelets are decreased [1]. The presence of blasts is a significant symptom of blasts.

Leukemia is classified into two types: the first one is acute leukemia (which progresses quickly) and second is chronic leukemia (which progresses slowly) [2]. Acute myelogeneous leukemia (AML) is a heterogeneous disorder of blasts of cells, due to this it is difficult to differentiate. This loss leads to bleeding, fatal infection and organ infiltration. When the marrow contains more than 30% blasts of cells then the presence of AML is confirmed. It is a fast growing blood and bone marrow cancer and if it is untreated then its fatal [3]. AML is difficult to diagnose because its precise cause is still unknown. Its symptoms are similar to most of diseases like flu, fever, tiredness or aches in bones or joints. Many blood tests are there for detection of leukemia like blood count, electrolytes and renal function, full blood count etc.

Treatment of AML is varying from bone marrow transplant, chemotherapy, radiation therapy etc. These techniques are depends on the operator's ability. Diagnostic confusion occurs due to symptoms are similar to other disease [4]. Detection task is difficult because of variety of features and due to the unclear images there are missing out features on vital indicators. Due to the variety in slide preparation techniques and because of complex nature of blood smear images lot of work has to be done for clinical demands.

In past, techniques of digital image processing have helped to analyze the cells that give accurate, remote, standard disease diagnosis systems. But there are lots of complications while we are extracting the data from white blood cells (WBCs) because WBCs have variations in their size, shape, edge and position [5]. So there is imbalance in illumination, image contrast between background and the cell boundaries variation depends on the condition during capturing process. The automated screening system is necessary because the treatment for leukemia is too much costly.

AML is a cancer and it starts inside bone marrow of human body. Acute myelogeneous leukemia is also called as acute myeloid leukemia. White blood cells plays important role in our body they fight with infections. In leukemia, bone marrow produces blasts i.e. immature blood cells. Due to the blasts of cells it is difficult for blood to do its work. The main objective of this project is to detect and classify Acute myelogeneous leukemia (AML) since manual examination of AML gives less accuracy in its results and automated screening system improves accuracy in results of detection.

The rest of the paper is organized as follows. In section basics to understand terms related to AML Section 3 gives techniques required for AML detection. Section 4 states the evaluation of classification. Section 5 concludes the paper.

- *Smita Bhole is currently pursuing masters degree program in Computer engineering in Babasaheb Ambedkar University, India. E-mail: smitabhole11@gmail.com*

2 BASICS FOR ACUTE MYELOGENEOUS LEUKEMIA

To understand the leukemia, it helps to have some basic knowledge about the blood and lymph systems.

2.1 Bone Marrow

Bone marrow is a inner part which is very soft and made up of skull, shoulder blades, ribs, pelvis and backbones. The bone marrow mostly contains small numbers of blood stem cells and it also have more mature blood-forming cells, fat cells and supporting tissues which helps in cell growth. Blood stem cells go through changes to make a new blood cells. During this process of development of cells, the cells develop into lymphocytes (a kind of white blood cells) and other forming cells [6].There are 3 types of blood forming cells: Red blood cells, White blood cells, Platelets

2.2 White Blood Cells

White blood cells help the body to fight with infections. One type of white blood cell is Lymphocytes and the other types are granulocytes (neutrophils, basophils, and eosinophils) and monocytes. These other types are called as myeloid cells.

2.3 Development of AML

To understand the term leukemia affects you, it is need to know how blood cells are produce and what the blood cells do. Normally, production of blood cell is done in bone marrow.

The body produce them in controlled way so they can work properly. The earliest phase of development of cell is known as stem cell and after this stem cell develops into one of these:

- Red blood cells (erythrocytes)
- Platelets (thrombocytes)
- White blood cells (granulocytes, monocytes or lymphocytes)

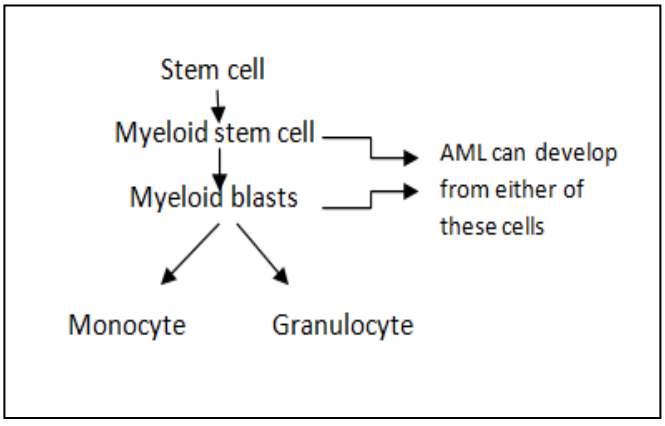

Fig 1 -Development of AML

As shown in Figure-1, the myeloblast is an immature cell of myeloid white blood cells. A mature white blood cell will developed from normal myeloblasts. In acute myelogeneous leukemia, myeloblast produces such a type of changes that can stop the cell in its state of immaturity and differentiation will be prevented.

At a number of different steps, leukemic transformation can occur along with the differentiation which causes AML stems with diversity and heterogeneity. Classification schemes for AML are to recognize the characteristics and behaviour of the leukemic cell (and the leukemia) and it depends on the stage of differentiation. In blood, there is Growth of leukemic cells which causes symptoms of acute myelogeneous leukemia. The symptoms of AML are depends on minimum of normal blood cells.

2.4 Causes of Acute Myelogeneous Leukemia

The most common type of leukemia among adults is acute myeloid leukemia (AML). This type of cancer is rare under the age of 40. Acute myelogeneous leukemia is common in men than women. Person with this type of cancer have lots of abnormal cells inside their bone marrow. The bone marrow helps the body to fight with infections and it makes other blood components may stops working correctly. Person with AML have risk of bleeding because number of healthy blood cells decrease. Problems regarding to your genes may play important role in AML development [7].

The things that can cause types of leukemia are blood disorders like thrombocythemia, and myelodysplasia, presence of chemicals in human body such as benzene, existence of chemotherapy drugs in human body, presence of drugs called as alkylating agents, Exposure to harmful chemicals and substances, exposure to radiation.

Symptoms of acute myelogenous leukemia are fever, bone pain, Lethargy and fatigue, Shortness of breath, Pale skin, frequent infections, easy bruising, Unusual bleeding, such as frequent nosebleeds and bleeding from the gums

3 ACUTE MYELOGENEOUS LEUKEMIA DETECTION TECHNIQUES

3.1 Segmentation

The main goal of image segmentation is to extract the required information from an input image. Segmentation plays very important role in AML detection because efficiency of feature extraction and correct identification of the myeloblasts is very important term for classification. For the purpose of AML detection, segmentation is performed which extract the nuclei of the leukocytes by using colour-based clustering. Cluster analysis is the study of algorithms for clustering, the objects according to measured intrinsic characteristics or similarity. For AML detection K-means algorithm is used for segmentation and K-means algorithm is one of the popular and a simple unsupervised algorithm. For this detection process very first choose clusters which corresponds to nucleus with the high saturation, background with a high percentage in

radiance and low percentage in diffusions, other blood cells such as erythrocytes and leukocyte cytoplasm. In segmentation, each pixel is assigned to one of these clusters using the properties of the cluster centre.

3.1.1 K-Means Clustering Algorithm

K-means is simplest algorithms that solve the clustering problem. The procedure of clustering follow a simple way to classify the data set given among certain number of clusters but there is assumption of K clusters [8],[9]. The main idea used in K-means is to define K centroid for each and every cluster. The centroid should be placed in such a way that they gives different results for this centroid should be placed at different locations. For this purpose centroids should be placed far from each other. The next step of K-means is to take point from given data set and associate that point to nearest centroid. When there is no point pending then the first step is completed. After this recalculate K new centroids of clusters using results from previous step. Repeat the same procedure above. A loop is generated. As a result of this loop the K centroid continuously change their location step by step and this process id done until there are no changes. Finally, Aim of this algorithm is minimizing an objective function using a squared error function. The objective function is calculated using following formula,

$$J = \sum^{k} \sum^{x} ||X^{j}_{i} - c||^{2} \qquad (1)$$

Steps of K-means algorithm:

1. Place K points which is represented by objects in given space then all these objects together make clusters. Points represents initial group of centroids.
2. Each object is assigned to closest centroid.
3. When each object is assigned then recalculate the position of K centroid.

Repeat Steps 2 and 3 until the all centroids have no changes. Due to this process all objects binds into groups from which minimized metric should be calculated.

It can be proved that if the algorithmic procedure will terminate then K-means algorithm doesn't find the optimal configuration corresponding to objective function. The algorithm is sensitive to the initially selected cluster centres. This effect can be reduced by running K-means algorithm multiple times. K-means is a simple algorithm that has adapted for problem domains.

3.1.2 Edge Based Segmentation

This is one of technique for segmentation of images. In this technique edges of objects in images are identified and those edges are considered to be the boundaries of objects. The boundaries of an object are used for segmentation. But there is drawback of this segmentation technique such as edges do not guarantee to form closed boundaries. To avoid this drawback edges should be processed so that only closed boundaries remaining. From these object boundaries segmented image is produced. A segmentation technique is based on the concept of extraction of region of interest from large image around cell nuclei which is threshold proposed by Katz [10].

3.1.3 Region Based Segmentation

The objective of this type of segmentation is to group pixels into regions and share similar characteristics. A disadvantage of this technique is that it causes failure if the definition for region uniformity tends too strict. . N. H. Abd Halim, M. Y. Mashor, A. S. Abdul Nasir, N. R. Mokhtar, H. Rosline [11] used this technique of segmentation to segment nucleus using S component of HSI model for acute leukemia detection. In this region based segmentation for segmenting a nucleus an optimal threshold value is obtained using Otsu's methods of segmentation. A region growing segmentation methods is proposed in [12]. A good results give assurance of good segmentation results so multi-scale local energy feature is constructed and using seed point in region growing based segmentation this segmentation is performed. This technique gives two-step process for segmentation of abnormal white blood cells and nucleus.

3.2 Feature Extraction

In image processing feature extraction is a technique which redefines a large dataset into a set of feature of dimensions that are reduced. The input data is transform into the set of features known as feature extraction. Classifier performance depends upon feature selection so the correct choice of features is very important step. The features were considered to improve classifier performance.

3.2.1 Shape Features

The good measure for the classification of AML from various shape features is compactness.According to the haematologists, the shape of nucleus is important for discrimination of myeloblasts. For the analysis of shape of nucleus, shape features those are based on region and boundary are extracted. From the binary equivalent image of nucleus all features are extracted and the nonzero pixels represent nucleus regions.

3.2.2 GLCM Feature

A texture is known as a function of spatial variation in pixel intensities. The associated texture feature calculations and the GLCM both belongs from image analysis techniques. The second order statistics like the probability of two pixels having particular special relationships of gray levels which gives the detailed structure about the term gray level pixel distribution. The 2-D gray level co-occurrence matrices depicted information which can be computed for various distances. To extract the textual characteristics Haralick defined some statistical measures like energy, contrast, entropy, correlation [13].

- Energy: is called as uniformity or angular second moments, energy is a measure for homogeneity of image.
- Contrast: It is the measure of contrast of image or the amount of local variations present in an image and contrast feature is difference term of the regional co-occurrence matrix.

3.2.3 Color Features

For the detection of AML, colour features are used such as cell energy.

Cell Energy: cell energy is called as the measure of uniformity. Feature is defined as "δ" and it is calculated by,

$$\delta = \sum i \sum j \, p^2(i,j) + \left(\sqrt{-1}\right)\left(\frac{\sqrt{\sum(x_i - x')^2}}{n-1}\right) \quad (2)$$

3.3 Classification

The classification technique selection for classification is a challenging task because an appropriate choice should be given for available data that can be significantly help for improving the accuracy. The blood sample of AML contain very highly overlapped cell clusters, making linear differentiation between normal and cancerous samples is not possible. With the use of kernels, the data can be transform into high dimensional spaces due to this linear separation is easily achieved. There are two candidates that can deal perfectly with highly overlapped data are neural networks (NNs) and support vector machine (SVM). Both have very simple training procedures and are easy to model. SVMs are popular method for classification of two categories and it is simple in use [14]. It's based on principles of structural risk minimization and mostly focuses on finding the support vectors [15]. A set of critical points that are used to represent the decision function of the classifier. This classifier performs classification by transforming the data into a feature space by means of kernel functions.

The only drawback of SVM is that it depends on a limited amount of kernels. PNNs is exactly opposite to SVMs because PNNs can be extended to multi-category classification, And they can re-construct their feature space by changing neuron interconnections and activation functions so each configuration of PNN represents different kernel.

4 Evaluation of Classification

As we know selection of a technique for classification is a challenging problem. But there are different methodologies which are used for classification of AML. Each of them gives results up to certain accuracy.

4.1 Support Vector Machine (Svm):

In this, by using bone marrow images taken from microscope they can built a system in which leukemia cell is recognized. They build the system by making use of support vector machine classifier i.e. SVM classifier where features are extracted within a blood images which are interconnected with textures, geometry and statistical inference. Mostly it depends on generation and selection of features which results in best recognition [16].

For textural feature extraction mean value, angular momentum, contrast and entropy these are used. Radius, perimeter, area, compactness, concavity, symmetry are the geometrical features. The parameters for statistical analysis are mean value and standard deviation for gradient matrix, skewness is measured for image. There is error in training data up to 11.87%, error of testing data is 21.13%. The only drawback of SVM is that it depends on a limited amount of kernels. In case of classification using SVM, a statistical method named as cross validation is used for evaluating and comparing learning algorithms. Cross-validation gives independent result sets by generalizing the results of statistical analysis. There are three types of validation techniques: K-fold, holdout, leave-one-out.

4.2 K-Nearest Neighbour (Knn):

The modified k nearest neighbour (KNN) is used to classify leukemia from microscopic images of blood smear and produce results with 0.01 percentage of error. This method is an instant-based learning algorithm that designates objects based on adjacent feature space in the training set. Multi-dimensional feature space is used to map the training sets and they are partitioned into regions established on the category of the training set. One of the points in the feature space is allocated to a particular category if it is the most continual category among the k nearest training data. Euclidean Distance is consistently used in computing the distance between the vectors. In KNN method, key element is the possibility to have similarity measure to identify neighbours of a particular cell pixel. The training phase subsists only of storing categories of the training set and feature vectors. In the classification phase consists of distances from the new vector which represents a cell pixel to all stored vectors are evaluated and there is a selection of k closest samples. The blast cell has been assigned to a particular category which is anticipated on the nearest point.

4.3 EM Algorithm:

EM-Algorithm is use to identify all types of leukocyte. The image pattern is changes into a lower dimensional space by using principal component analysis. The EM algorithm is used to obtain the Gaussian functions parameters which are used to model the each class probability distribution function of cells. Bayes' theorem is used for classifications of the images and at the end class with the highest probability is chosen [17]. To find out the maximum likelihood solution for models with latent variables, the Expectation maximization algorithm or EM-algorithm is used. The model is a Gaussian mixture of distribution. The

latent variables zk are labels that use to indicate from which component k of the mixture the measurements came from. The probability density functions for a mixture model with K Gaussian distributions is given by,

Enhancement, segmentation and classification	Methods	Advantage	Drawbacks
[16]	1) Support vector machine 2) Naïve Bays classification	Exploits the features of the image of the blood cells related to the texture, geometry and histograms	Focuses on the features generation and selection only
[17]	1) Gaussian mixture using the EM algorithm. 2) Posteriori decision rule.	1) Accuracy over manual counting(near about 80 times more cells are contened) 2) Precise method for normal blood samples.	Not for pathological images.
[18]	1) Image grabbing 2) Fuzzy C clustering 3) Sub imaging 4) Colour Conversion RGB to L*a*b	Accuracy and Measuring nucleus boundaries along with shape, colour and texture.	Classification of lymphoblast using into subtypes.

$$p(x) = \sum_{k=1}^{K} \pi_k \, N(x|\mu_k, \sum_k) \qquad (3)$$

Comparative study between different classification approaches:

In general, the performance of AML classification is evaluated by using four terms. They are Precision, Specificity, Sensitivity, F-measure.

Precision, specificity, sensitivity, and f-measure
These are all possible outcomes of the classifier system. When classification is to attempt then there are four possibilities: true positives (TP) when cancer is correctly classified ; False positive (FP) when non- cancer is identified correctly as cancerous ; True negatives (TN) when non-cancerous cells correctly indentified and last one False negatives (FN) when cancer cells are classified as non-cancerous.

Precision gives proportion of subjects with positive results which are correctly classified.

$$precision = \frac{TP}{(TP + FP)}$$

Specificity gives ability to identify negative results

$$specificity = \frac{TN}{(TN + FP)}$$

Sensitivity denoted ability to identify positive results

$$sencitivity = \frac{TP}{(TP + FN)}$$

F-measure is metric that gives harmonic mean of precision and sensitivity

$$F - measure = \frac{2 \times Precision \times sensitivity}{Precision + sensitivity}$$

5 Conclusion

In this literature survey paper it is seen that the development of automated system gives accurate results for AML in blood microscopic images. For detection it uses high quality images i.e. microscopic images. For detection segmentation, feature extraction, classification these techniques are used. A feature set uses texture, shape and colour characteristics of cell and this information is used for efficient classification. A colour feature that is cell energy is introduced and its result shows that it gives good discrimination between cancerous and non-cancerous cells. In classification there are two methods Probabilistic neural networks (PNNs) and Static vector machine both are good but PNNs overcome the drawback of SVMs and gives more accurate results.

In future, more work is needed on further improving the performance measures which requires the further research. Also we can implement such a kind of software which is able to detect all types of cancer and also according to the results, software will tell us what precautions are required for particular type of cancer.

References

[1] C. Haworth, A. Hepplestone, P. Jones, R. Campbell, D. Evans, and M. Palmer, "Routine bone marrow examination in the management of acute lymphoblastic leukaemia of childhood," J. Clin. Pathol., vol. 34, no. 5, pp. 483–485, May 1981.

[2] MedlinePlus: Leukemia.National Institutes of Health. [Online] Available: http://www.nlm.nih.gov/medlineplus/ency/article/001299.htm

[3] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in Proc. CIMSA, 2005, pp. 96–101.

[4] S. Serbouti, A. Duhamel, H. Harms, U. Gunzer, J. Mary, and R. Beuscart,

"Image segmentation and classification methods to detect leukemias," in Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., 1991, pp. 260–261.

[5] D. Mandal, K. Panetta, and S. Agaian, "Face recognition based on logarithmic local binary patterns," in Proc. SPIE, Image Process., Algorithms Syst. XI, 2013, vol. 8655, pp. 865514-1–865514-12.

[6] American cancer society: what is acute myloid leukemia. [online] available :http://m.cancer.org/cancer/leukemiaacutemyloidaml/overviewguide/leukemia.htm

[7] MedlinePlus: Acute myloid leukemia. National Institutes of Health. [on line] Available :http://www.nlm.nih.gov/medlineplus/ency.htm

[8] Alexander Rakhlin and Andrea Caponnetto, "Stability of k-Means clustering", Advances in Neural Information Processing Systems 12, pages 216–222, MIT Press, Cambridge, MA, 2007. http://cbcl.mit.edu/projects/cbcl/publications/ps/rakhlin-stability-clustering.pdf

[9] Borah S. and M.K. Ghose," Performance analysis of AIM-K-Means and K-Means in quality cluster generation", Journal of Computing, Volume 1, Issue1,December-2009,ISSN:2151-9617, https://sites.google.com/site/journalofcomputing/

[10] A. R. J. Katz, "Image Analysis and Supervised Learning in the Automated Differentiation of White Blood Cells from Microscopic Images", Master's thesis, Department of Computer Science, RMIT, Feb. 2000.

[11] N. H. Abd Halim, M. Y. Mashor, A. S. Abdul Nasir, N. R. Mokhtar, H. Rosline, "Nucleus Segmentation Technique For Acute Leukemia", IEEE 7th International Colloquium on Signal Processing & its Application, 2011.

[12] Yang Wang, Qianqian Wang, "Image Segmentation Based On Multi Scale Local Feature", 3rd International Congress on Image and Signal Processing (CISP), 2010.

[13] R. M. Haralick, "Statistical and structural approaches to texture," Proc. IEEE, vol. 67, no. 5, pp. 786–804, May 1979.

[14] Vapnik, V. N. The Ature of Statistical Learning Theory. New York: Springer, 1995.

[15] Cristianini, N., and J. Shawe-Taylor. An Introduction to support vector machines and other kernel-based learning methods. New York: Cambridge University Press, 2000.

[16] T. Markiewicz, S. Osowski, B., Marianska,L., Moszczynski, "Automatic Recognition of the Blood Cells of Myelogenous LeukemiaUsing SVM", Proceedings of International Joint Conference on Neural Networks,Montreal ,Canada, July 31 – August 4, 2005,pp.2496-2501.

[17] M.C., Colunga, O.S., Siordia, S.J.,"Leukocyte Recognition Using EM Algorithm",MICAI 2009

[18] C.Reta, L.Alamirano, J.A.Gonzales, R.Diaz,andJ.S.Guichard, " Segmentati -on of Bone Marrow for Morphological Classification of Acute Leukemia, " in Proceding of the Twenty-Third Interational Florida Artificial Intellige - -nce Research Society Conference (FLAIRS 2010)