ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 4, April 2016

A comparative study of algorithms used for leukemia detection

Dr. T. Karthikeyan[1], Dr. B. Ragavan[2] , N.Poornima[3]

***Abstract*: In medical diagnosis systems, classification of blood cell is most important to evaluate and diagnosis the disease. Identification of blood disorders, it can lead to classify the diseases related to blood. Leukemia is a blood cancer that begins with bone marrow. It should be treated at earlier stage, if left untreated it cause death. This paper describes algorithms which is mostly used to detect leukemia. The algorithms described here are Support Vector Machine(SVM), Otsu's thresholding method, kNN , k-means clustering method, Watershed Transformation.**

***Index terms*: Leukemia, classification, detection, SVM, Otsu's, kNN, k-means clustering method, Watershed Transformation.**

I.INTRODUCTION

Leukemia is a cancer of blood cells. In leukemia, abnormal blood cells are produced in the bone marrow. Usually, leukemia involves the production of abnormal white blood cells which are responsible for fighting infection. The abnormal cells in leukemia do not function in the same way as normal white blood cells. The leukemia cells continue to grow and divide, eventually crowding out the normal blood cells. There are different types of leukemia, based upon how quickly the disease develops and the type of abnormal cells produced. Leukemia can also cause symptoms in organs that have been infiltrated or affected by the cancer cells. For example, if the cancer spreads to the central nervous system, it can cause headaches, nausea and vomiting, confusion, loss of muscle control, and seizures[23].

Leukemia is a cancer that starts in blood stem cells. Stem cells are basic cells that develop into different types of cells that have different jobs. [15]Blood stem cells develop into either lymphoid stem cells or myeloid stem cells.

- **Lymphoid stem cells** develop into lymphocytes, a type of white blood cell. Lymphocytes help fight infection and destroy abnormal cells. The 3 types of lymphocytes are B cells, T cells and natural killer (NK) cells.
- **Myeloid stem cells** develop into red blood cells, granulocytes, monocytes or platelets. Red blood cells carry oxygen to all tissues of the body. Granulocytes and monocytes are types of white blood cells that destroy bacteria and help fight infection. Platelets form clots in damaged blood vessels to stop bleeding.

STEPS FOR PROCESS OF AUTOMATING BLOOD RECOGNITION

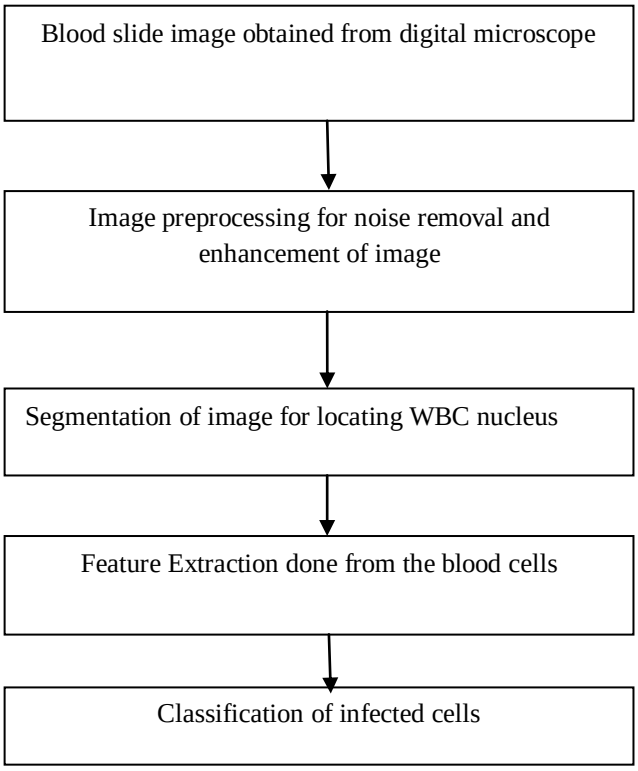

Fig 1. Process of blood recognition

The above figure describes how the blood cell can be identified automatically. First the image can be obtained from dgital microscope. Then preprocessing can be done which includes noise removal of an image and enhancement of an image. Then segmentation should be done for segmenting the blood cells individually. Segmentation can be done with the help of segmentation algorithms. After segmentation feature can be extracted. Feature extraction involves reducing the amount of resources required to describe a large set of data. After features extracted, it then applies to classification for identifying infected cells.

STEPS IN DIGITAL IMAGE PROCESSING:

A) Image Acquisition: It deals with capturing of images. It involves preprocessing such as scaling.

B) Image enhancement: Process of manipulating an image so that the result is more suitable than the original for a specific application.

C) Image restoration: Deals with improving tha appearance of an image.

D) Color image processing: It is an area that has been gaining in importance because of the significant increase in the use of digital images.

E) Wavelets: These are the foundation for representing images in various degrees of resolution.

F) Compression: Deals with techniques for reducing the storage required to save an image.

G) Morphological processing: Deals with tools for extracting image components that are useful in the representation and description of shape.

H) Segmentation: Deals with partitioning an image into different segments.

II.LITERATURE REVIEW

S.Jagadeesh , Dr.E.Nagabhooshanam , Dr.S.Venkatachalam proposed segmentation of bone marrow image aspirate using watershed algorithm[16]. T. Markiewicz, S. Osowski, B., Marianska, L., Moszczynski use Support Vector Machine (SVM) classifier[17] and exploit features in blood cell images to classify leukemic cells. Ms. Sneha Dhakne, Ms. Kumudini K. Borkute, Ms. Priyanka Ikhar presented two techniques such as Watershed algorithm and Combine clustering along with filtering for identification of leukemic cell[18]. Monika Mogra , Vivek Srivastava., proposed watershed transform and morphological image processing technique[19] to identify affected White Blood Cell. Subhan, Ms. Parminder Kaur., discussed KNN and Hough Transform algorithms to detect abnormal cells that cause leukemia[20]. Khot S.T, Sneha Bhalekar, Divya Jaggi and Dolly Rani used Support Vector Machine classifier to detect leukemic cells.[21]. Ms. Minal D. Joshi, Prof. Atul H. Karode, Prof. S.R.Suralkar proposed a method to detect acute leukemia. They discussed global thresholding otsu method[22].

III.SEGMENTATION ALGORITHMS

Computer analysis of image objects starts with finding them – deciding which pixels belong to each object. Process of separating objects from the background, as well as from each other is called image segmentation. In Digital Image Processing Textbook written by Rafael C. Gonzalez and Richard E. Woods said that ," segmentation of non trivial images is one of the most difficult task in image processing. Segmentation accuracy determines the success or failure of computerized analysis procedures". To detect Leukemia from human blood sample images, segmentation of cells is done through some algorithms. Widely used segmentation algorithms are discussed below:

A. K-Means Clustering Algorithm:

It is a patitioning algorithm and it use an optimization function to partition the cells into successive clusters. K-means and most of its variants originated in statistics[12]. It utilizes one of the distance metrics to demarcate the clusters. We should specify a positive integer k, that represents the best guess on the number of clusters present in the data. Possible number of clusters is hidden or ambiguous in image, audio, video, and multimedia clustering applications, in such cases we may try successive values of k starting with 2. This process is stopped when two consecutive k values produce more or less identical results. This algorithm is used for segmenting the cells. Normal cells are grouped as a cluster and abnormal cells can be identified among the blood cells. Based on the result we can easily identify a patient is affected from leukemia or not.

B. Watershed Transformation:

Emad A. Mohammed, Mostafa M.A.Mohammed, Christopher Naugler, Behrouz H. Far, proposed watershed transformation for detecting leukemic cells[8].Watershed transformation can be classified as a region based segmentation approach. It is often applied on its gradient image. It is a powerful tool for image segmentation and searches for pixel and region similarities. Image gradient is a directional change in the intensity or color in an image. Image gradients may be used to extract information from images.

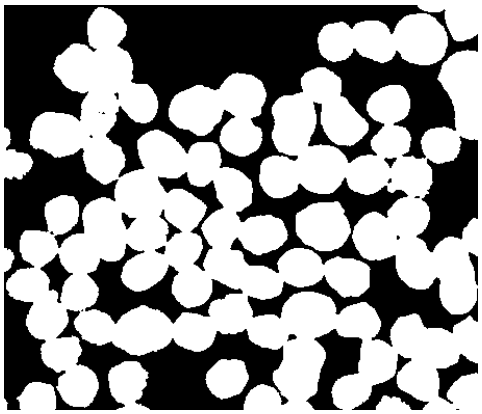

Fig 2. an example of cells segmentation using watershed transformation.

Principle of Watershed Transformation:

- Any graytone image can be considered as a topographic surface
- If we flood this surface from its minima, and if we prevent the merging of the waters coming from different sources, we partition the image into two different sets: the catchment basins and the watershed lines.
- If we apply this transformation to the image gradient, the catchment basins should theoretically correspond to the homogeneous gray level regions of the image.
- This transformation produces over segmentation due to noise or local irregularities in the gradient image.

Marker- Controlled Watershed:

A major enhancement of the watershed transformation consists in flooding the topographic surface from a previously defined set of markers. Doing this, we can prevent over segmentation.

The Segmentation Paradigm:

Segmenting an image by the Watershed transformation is there a two step process:

A) Finding the markers and the segmentation criterion(the criterion or function which will be used to split the regions, it is most often contrast or gradient, but not necessarily)

B) Performing a marker controlled Watershed .

There are three methods to implement Watershed: they are

- Distance Transform approach
- Gradient method
- Marker Controlled approach

Watershed Procedure:

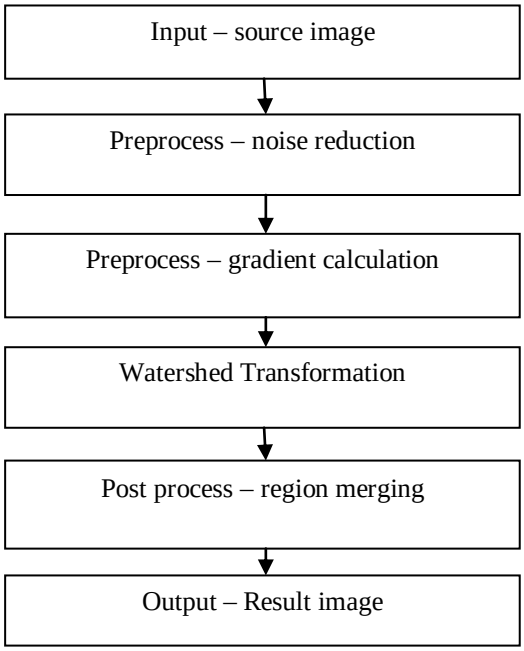

Fig 3. Flow chart of watershed procedure

C. Otsu's Thresholding Method

Miss Hetal J. Vala., et al proposed otsu thresholding for segmenting blood cells[13].Image thresholding is one of the segmentation methods to convert gray scale image to binary images. This method is an automatic thresholding method. Otsu's threshold iterate through all the possible threshold values to find the threshold value where the sum of foreground and background spreads is at its minimum.

Otsu's is used to automatically perform clustering based image thresholding. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram(foreground pixels and background pixels) it then calculates the optimum threshold separating the two classes so that their combined spread is minimal or equivalently, so that their inter- class variance is maximal. The extension of the original method to multilevel thresholding is referred to as multi otsu method. It operates directly on the gray level histogram, so it is fast.

Otsu's Assumptions:

Histogram(and the image) are bimodal. No use of spatial coherence, nor any other notion of object structure. Assumes stationary statistics but can be modified to be locally adaptive. Assumes uniform illumination (implicitly),so the bimodal brightness behavior arises from object appearance differences only.

IV.CLASSIFICATION ALGORITHMS

Classification of cells is more important in medical image. Based on classification results we can identify whether the human is affected with leukemia or not. The algorithms which are used to detect leukemia with highest accuracy are as follows.

A. *Support Vector Machine:*

Niranjan Chatap., et al proposed Support Vector Machine and kNN for detecting leukemic cells[2]. It is an optimization theory based classifier that finds a maximal margin surface, which seperates the data into distinct classes. There are many classifiers that originated in statistics. In addition, multiple (linear and non linear) regression, and logistic regression models can be used as classifiers[12]. Assume that for instance, in multiple regression model, error terms are normally distributed and that the independent variables are uncorrelated. SVM is a supervised classification model without any assumption on data distribution. Another name for SVM is Kernel Machines(as nonlinear SVM uses Kernel mapping). Kernel method is a popular technique to transform nonlinearly separable data to linearly separable form. They map the input data into a higher dimensional feature space using one of the kernel functions. The reason for the popularity of SVM is that it is easily scalable to very large data sets. It has a small number of user tunable parameters (example is C in SM-SVM, LS-SVM) and same model can be generalized to non linear, overlapping and fuzzy models.

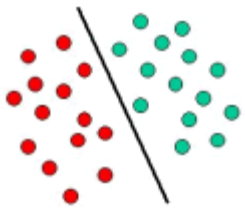

Fig 4. Example of svm classification

The above is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GREEN and RED in this case) with a line.

A classifier is well generalisable if it can predict the correct class of unseen data with minimal error. The classical primal SVM is solved as a quadratic programming problem with as many variables as the dimension of training data. Dual SVM is also a quadratic program with as many variables.

A hyperplane is an n-dimensional generalization of a straight line in 2D. it can be visualized as a plane surface in 3D, but it is not easy to visualize when dimensionality is greater than 3. Number of support vectors found depends upon the solution technique (LS-SVM gives more support vectors) and the kernel function used(in NL-SVM).

B. *K-Nearest Neighbour*

Subhan., et al discussed kNN for detecting leukemia[7].It has been widely used in the area of pattern recognition. Nearest neighbor classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n attributes[14]. Each tuple represents a point in an n-dimensional space. When given an unknown tuple, a k-nearest-neighbour classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k " nearest neighbours" of the unknown tuple.

"closeness" is defined in terms of a distance metric, such as Euclidean distance. Nearest neighbor classifiers use distance based comparisons that intrinsically assign equal weight to each attribute. So it can suffer from poor accuracy when given noisy or irrelevant attributes. The method has been modified to incorporate attribute weighting and pruning of noisy data tuples. Nearest neighbor classifiers can be extremely slow when classifying test tuples.

Table 1 shows COMPARISON OF SEGMENTATION ALGORITHMS:

SEGMENTATION ALGORITHMS	MERITS	DEMERITS
K-Means clustering segmentation	k-means computation is faster than hierarchical clustering when variables are huge	Difficult to predict k value and does not work well with clusters of different size and different density.
Watershed transformation	Marker-controlled approach overcomes the problem of over segmentation	Leads to over segmentation
Otsu"s thresholding segmentation	Segmentation is good or stable	Time complexity for multilevel selection

Table 2 shows COMPARISON OF CLASSIFICATION ALGORITHMS:

CLASSIFICATION ALGORITHMS	MERITS	DEMERITS
k- Nearest-Neighbour	Simple classifier that works well on basic recognition problems	It does not learn anything from training data and simply uses the training data itself for classification
Support Vector Machine	With the help of kernel svm gains flexibility and deliver unique solution	Choice of kernel and limitation in speed and size both in training and testing

V.CONCLUSION

Leukemia is a cancer that begins with bone marrow and leads to death if not treated earlier. To detect leukemia many algorithms are used. But this paper describes algorithms which are widely used to detect leukemia and with good accuracy. Among these specified algorithms Support Vector Machine classifier and Otsu's thresholding segmentation are popularly used with more accuracy.

VI.REFERENCES

[1] Niranjan Chatap, Sini Shibu" Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbour" *(IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III (Sep – Oct. 2014), PP 79-87*

[2] S.Jagadeesh 1, Dr.E.Nagabhooshanam 2, Dr.S.Venkatachalam., "Image processing based approach to cancer cell prediction in blood samples" International Journal of Technology and Engineering Sciences Vol.1 (1), ISSN: 2320-8007

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

[4] http://www.svms.org/disadvantages.html

[5] C.Vidhya, P.Saravana kumar, K.Keerthika, C.Nagalakshmi, B.Medona devi.," Classification of acute lymphoblastic leukemia in blood microscopic images using svm**" (**ICETSH-2015**)** ISSN: 2348 – 8549
[6] udel.edu./kschieff/leukemiadiagnosis.html.

[7] Subhan, Ms. Parminder Kaur., " Significant Analysis of Leukemic Cells Extraction and Detection Using KNN and Hough Transform Algorithm" International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 1, Jan-Feb 2015.

[8] Emad A. Mohammed, Mostafa M.A.Mohammed, Christopher Naugler, Behrouz H. Far, "Chronic Lymphocytic Leukemia Cell Segmentation From Microscopic Blood Images Using Watershed Algorithm and Optimal Thresholding", 26th IEEE Canadian Conference Of Electrical And Computer Engineering(CCECE), 2013.

[9]http:// www.nickgillian.com/wiki//pmwiki.php/GRT/KNN

[10] http://www.statsoft.com/Textbook/Support-Vector-Machines

[11] "Digital Image Processing" written by Rafael C. Gonzalez and Richard E. Woods

[12] "Data Mining Methods" written by Rajan Chattamvelli.

[13] Miss Hetal J. Vala, Prof. Astha Baxi., "A Review on Otsu Image Segmentation Algorithm" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 2, February 2013

[14]" Data mining" typical data mining process for predictive modeling by BPB publications.

[15]http://www.cancer.ca/en/cancer-information/cancer-type/leukemia/leukemia/?region=bc

[16] S.Jagadeesh 1, Dr.E.Nagabhooshanam 2, Dr.S.Venkatachalam., "Image processing based approach to cancer cell prediction in blood samples" International Journal of Technology and Engineering Sciences Vol.1 (1), ISSN: 2320-8007

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

[18] Ms. Sneha Dhakne, Ms. Kumudini K. Borkute, Ms. Priyanka Ikhar., "Detection of cancer cell in blood samples using an Effective algorithm" (IJCESR) ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-2, ISSUE-6, 2015

[19] Monika Mogra , 2,Vivek Srivastava., " A Comprehensive Review of Analysis of Counting Blood Cells Using Different Image Processing Algorithms" *International Journal of Engineering Science Invention ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org Volume 3 Issue 6‖ June 2014 ‖ PP.29-31*

[20] Subhan, Ms. Parminder Kaur., " Significant Analysis of Leukemic Cells Extraction and Detection Using KNN and Hough Transform Algorithm" International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 1, Jan-Feb 2015.

[21] Khot S.T, Sneha Bhalekar, Divya Jaggi and Dolly Rani., " An innovative approach in myelogenous leukemia detection using attribute analysis" (IJAEEE) ISSN (Print): 2278-8948, Volume-2, Issue-5, 2013

.

[22] Ms. Minal D. Joshi, Prof. Atul H. Karode, Prof. S.R.Suralkar "White Blood Cells Segmentation and Classification to Detect Acute Leukemia" (IJETTCS) Volume 2, Issue 3, May – June 2013.

[23]http://www.healthline.com/health/leukemia#overview1.

VII. AUTHORS PROFILE

First Author: Associate Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamilnadu, India.

Second Author: Associate Professor, Department of Biochemistry, PSG College of Arts & Science, Coimbatore, Tamilnadu, India.

Third Author: Research Scholar, Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamilnadu, India.