INTERNATIONAL JOURNAL OF SCIENTIFIC PROGRESS AND RESEARCH (IJSPR) ISSN: 2349-4689
Volume 29, Number 03, 2016

An Automated Approach Based on K-Means And Fuzzy Inference System For The Detection of Leukemia

R. Suriyagrace[1], Dr. M. Devapriya[2]

[1]Research Scholar, [2]Assistant Professor, Department of Computer Science, Government Arts College, Coimbatore, India

Abstract:* Cancer is a disease which involves abnormal cell growth and spread to other parts of the body. The base work uses the method of Wavelet Transformation for image improvement, image segmentation for segmenting the different cells of blood, edge detection for detecting the boundary, size, and shape of the cells and Fuzzy Inference System is used for the final decision of blood cancer based on the number of different cells. However, the blurred region boundaries obtained on the segmented image can create destruction results. In order to overcome the problems faced in the existing system, the proposed work includes a clustering algorithm with the image processing concepts. By using the min-max relation with R value, the infected cells produce a new fuzzy set value.

Keywords:* medical imaging, fuzzy inference system, leukemia, k-means algorithm, image segmentation.

I. INTRODUCTION

Leukemia is a type of cancer which begins in the bone marrow and results in a high numbers of abnormal white blood cells. The white blood cells which are fully not developed are called blasts or leukemia cells. Bone marrow biopsy and blood tests are done to diagnose blood cancer.[1].

Leukemia is believed to have different causes such as inherited and environmental (non-inherited) factors are believed to be involved. Smoking, ionizing radiation, some chemicals (such as benzene), prior chemotherapy, and Down syndrome are some of the risk factors which lead to blood cancer. Leukemia is classified into four main types.

- Acute lymphoblastic leukemia (ALL)
- Acute myeloid leukemia (AML)
- Chronic lymphocytic leukemia (CLL)
- Chronic myeloid leukemia (CML)[2]

II. BACKGROUND

Blood is used as a transport to supply oxygen, nutrition and removes the waste in the Body. The blood volume is composed of 3 types of blood cells:

- Red blood cells - carry oxygen to the tissues
- White blood cells - fights infections
- Platelets - helps to clot blood

Leukaemia is a malignancy (cancer) of blood cells. In general, leukaemia may be classified when abnormal blood cells are produced in the bone marrow. Infections are fought by the White Blood Cells (WBC). Leukaemia involves the production of the abnormal WBC.[3]

III. TESTS FOR DIAGNOSING CANCER

Samples collected for cancer blood tests or biopsy of a suspicious area are analyzed in a lab for signs of cancer. Blood tests used to diagnose cancer include:

- Complete blood count (CBC)
- Blood protein testing
- Tumor marker tests
- Circulating tumor cells test[5]

IV. EXISTING WORK FROM THE LITERATURE VIEW

The existing system includes image processing techniques and fuzzy system for detecting cancer cells. Image pre-processing was applied to remove the irrelevant or unnecessary parts from the image and leads to image enhancement. Threshold techniques were used for segmenting the blood cells image based on the pixel information. The number of infected cells was observed from the threshold output. Segmentation was followed by feature extraction. Feature extraction was based on shape based feature and statistical based feature.[9]

Fuzzy rule based decision system was used on extracted feature for cancer classification. Fuzzy logic includes membership function that defines fuzziness of an image and also defines information contained in the image[. Thus, fuzzy rule provides relationship among membership of number of cells and their shape with respect to threshold outputs and automatically classifies three types of blood cancer such as Leukemia, Lymphoma and Myeloma.

The outputs of the various techniques from the original image after the techniques being followed in the existing system are

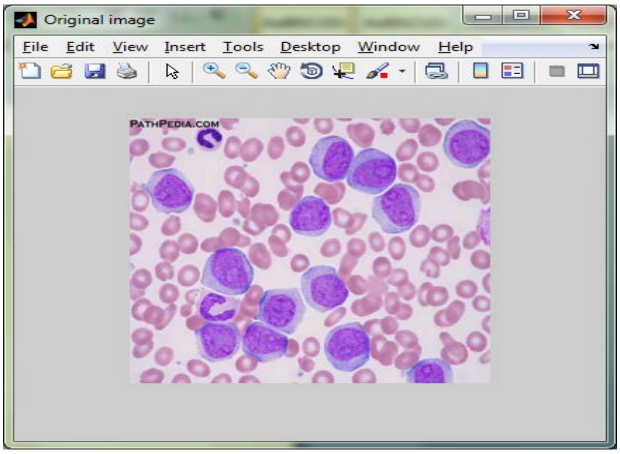


Fig a: Original image

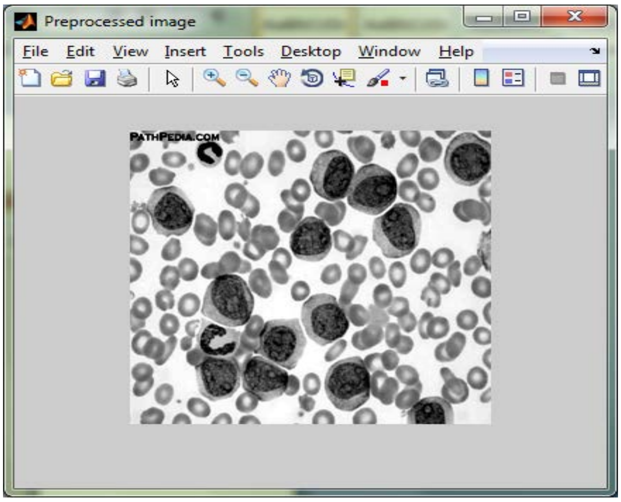


Fig b: Pre-processed image

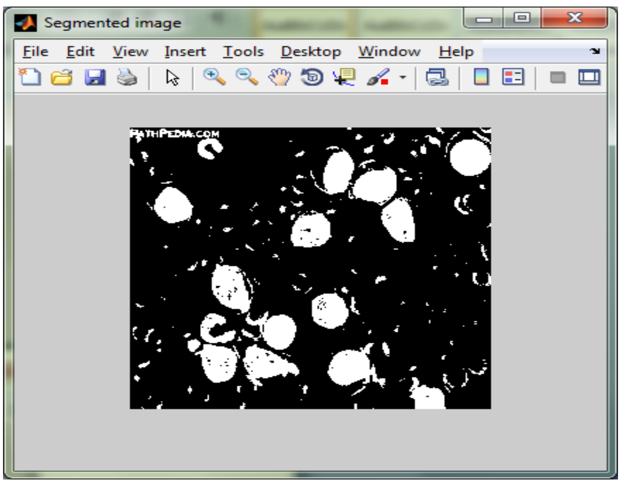


Fig c: Segmented image

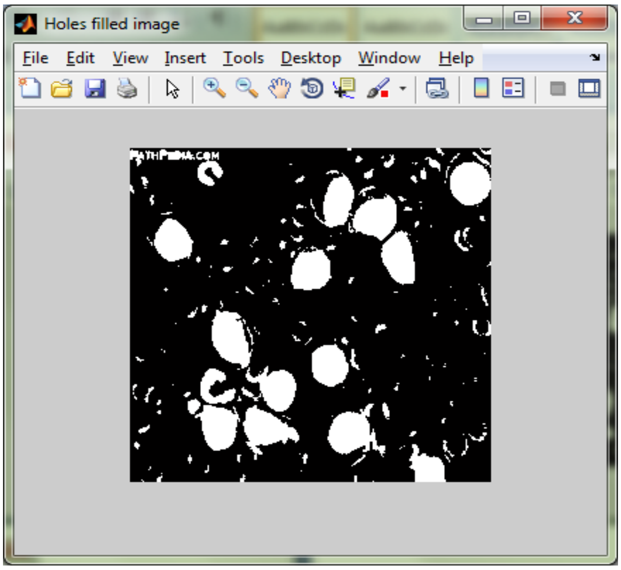


Fig d: Holes filled image

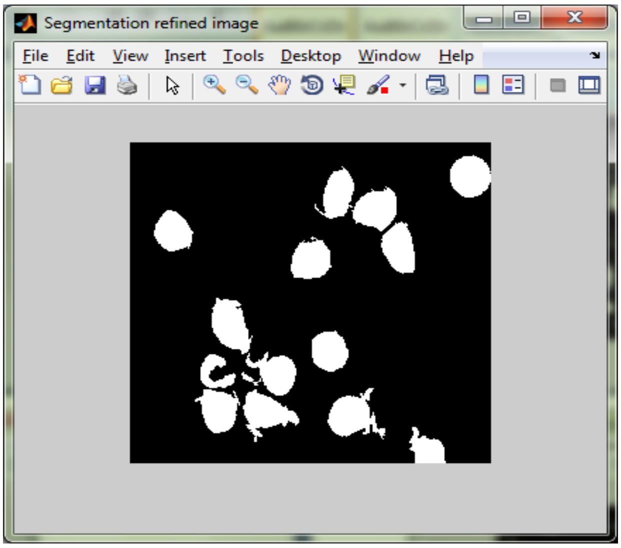


Fig e: Segmented refined image

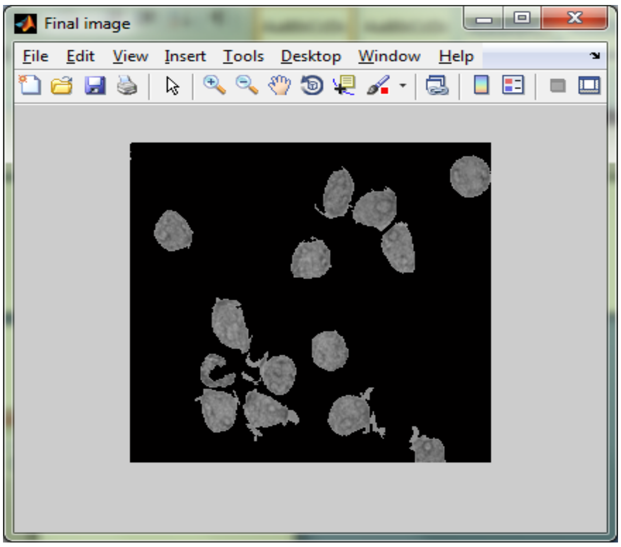


Fig f: Final image

V. PROPOSED WORK OVERVIEW

The main focus of the proposed work is to identify leukemia cells automatically from digital images.

The blood cell images from various sources are pre-processed by image preprocessing step. After preprocessing, images are enhanced by k-means clustering for highlighting the affected White Blood Cells (WBC).

During image segmentation, the nuclei of WBCs are segmented based on global and local curvature properties and eliminate the normal white blood cells from the microscopic blood image. Features such as texture, geometry, color and statistical features of nuclei are calculated to determine the various factors of leukemia cells.[6]

The extracted features are trained by Fuzzy rule based decision system of single row feature vector of each cell which is used for classifying leukemia cells from white blood cells.

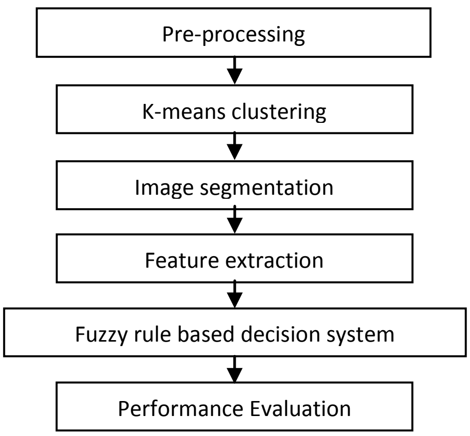
The relationship among features in microscopic image is calculated as membership function and will proceed for each microscopic image.

Using max-min composition with the relation R will produce new value for Fuzzy set which indicates the infected cells. The fuzzy rule detects the leukemia automatically from microscopic blood images.

VI. PROPOSED MODULES

The proposed work is basically classified into 6 modules which includes Pre-processing in step 1, k-means clustering in step 2, image segmentation in step 3, feature extraction in step 4, fuzzy rule based decision system in step 5 and performance evaluation in step 6.

The following diagram illustrates the step by step process which is followed in the proposed work.



Step 1: PRE-PROCESSING

Image pre-processing makes a sample image suitable for a particular application. This generally involves enhancement of image, which includes cropping or resizing, sharpening, de blurring, brightening, change in image contrast, noise removal, edge highlighting. The pre-processing step removes the unwanted parts, enhances the image[7].

In the pre-processing step the image is converted into the format which suits for performing the analysis. Unwanted parts or the excess pixels are removed from the image.

Step 2: K-Means Clustering

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem which follows an easy way to classify certain number of clusters (assume k clusters) fixed apriori from the given data set.

k centers, one for each cluster is defined. A new binding is done between the same data set points and the nearest new center after the k new centroids are created. A loop has been generated. As a result of this loop, the k centers may be noticed to have a change in their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function knows as squared error function given by:

$$J(V) = \sum_{i=1}^{c}\sum_{j=1}^{c_i}\left(\| x_i - v_j \|\right)^2$$

Where, '$\| x_i - v_j \|$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centers.

The clustered pixels are segmented into small portions.

Step 3: Image Segmentation

Image segmentation is a process of image partitioning into multiple segments or regions or structures of interest, so that the contents of each region have similar characteristics. Image segmentation results in a set of regions that collectively cover the entire image or set of contours extracted from the images.

In this work, the pixels of the blood cells are selected, since they are the areas of interest. The images are segmented using a range of values. By using this step, the regions having the same characteristics are grouped together.

Step 4: Feature extraction

Feature extraction is a sub-division of improved image for identifying or interpreting meaningful object forms. The features are of two types: shape based feature and statistical based feature.

In this work, the segmented pixels are calculated based either by the shape based features or by the statistical based feature. The differentiation among the various pixels groups are differentiated either by shape, by color or by texture.

Step 5: Fuzzy rule based decision system

Fuzzy set is a generalized of classical set theory. Fuzzy sets try to capture the way humans represent and reason with real world knowledge in fuzzy set, many degrees of membership(between 0 and 1) are allowed. Fuzzy logic is a form of many valued logic; it deals with reasoning that is approximate rather than fixed and exact.

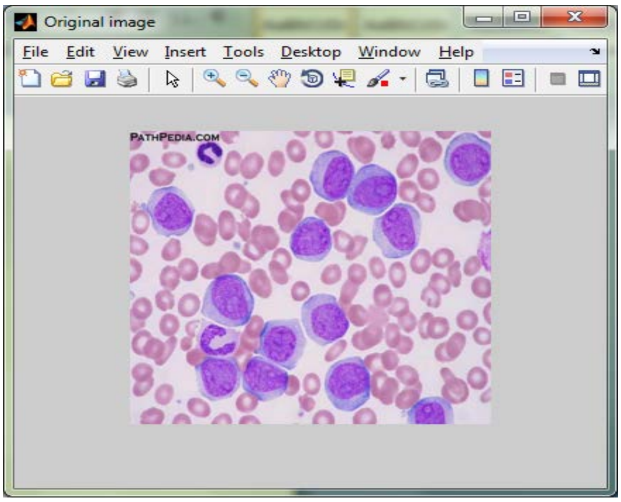


Fig i. Original image

Fuzzy rules are linguistic IF-THEN- constructions that have the general form "IF A THEN B" where A and B are (collections of) propositions containing linguistic variables. A is called the premise and B is the consequence of the rule.

Step 6: Performance Evaluation

Based on all the above modules, the performance is evaluated for the existing work and the proposed work in each and every step. Thus, the overall performance is evaluated and produced as a comparative result either in terms of percentage (%) or numeric values.

The outputs of the various techniques from the original image after the techniques being followed in the existing system are

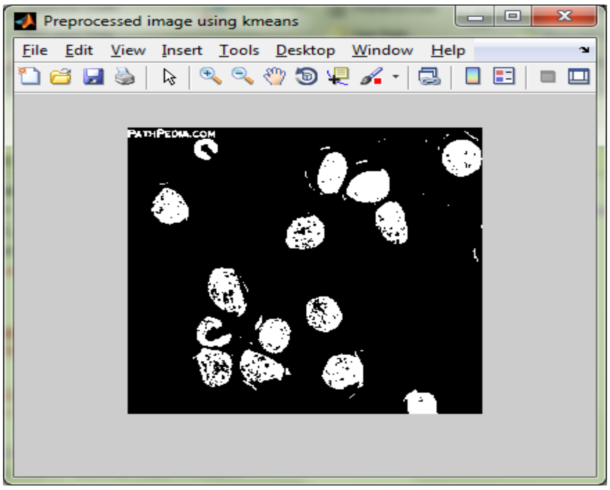


Fig ii. Pre-processed image using k-means

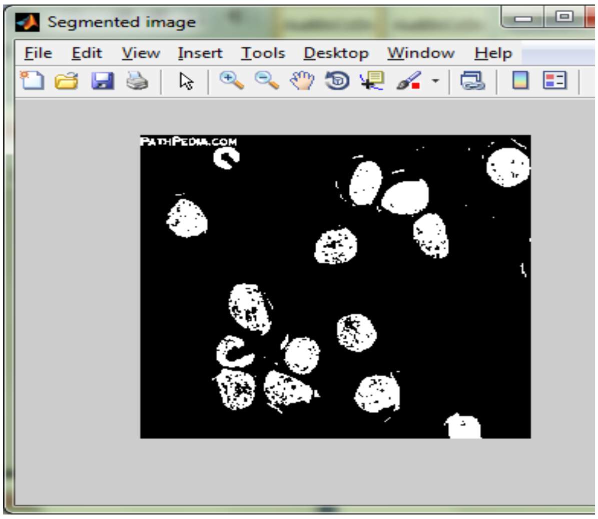


Fig iii. Segmented image

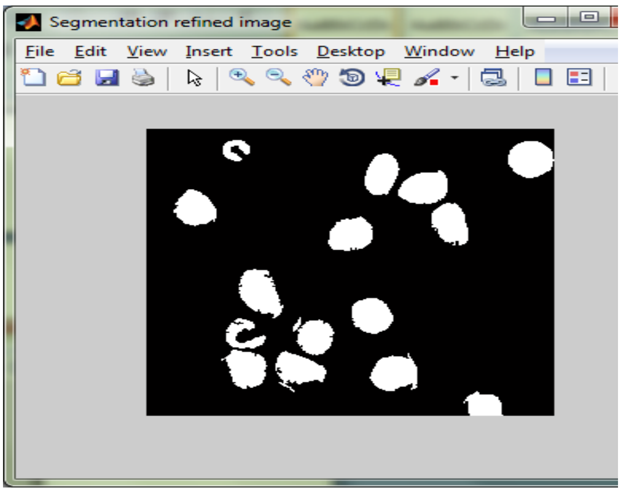


Fig iv. Segmented refined image

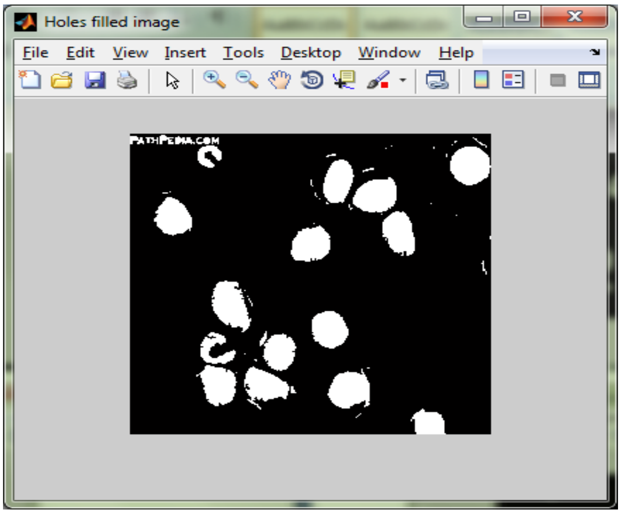


Fig v. Holes filled image

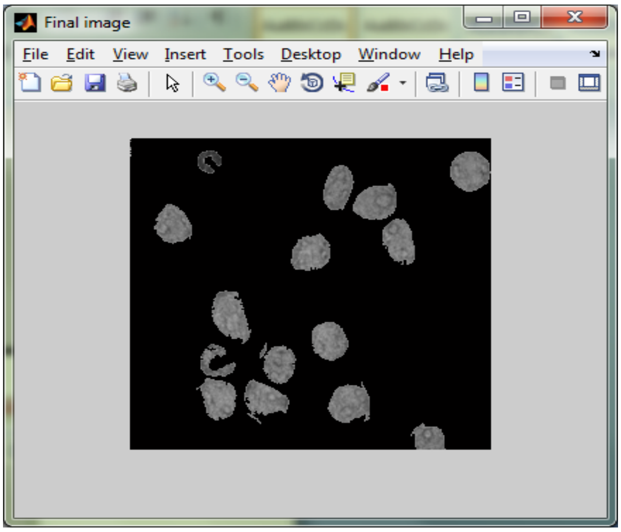


Fig vi. Final image

VII. RESULTS

The performance of the proposed k-means fuzzy system is evaluated in terms of precision, recall, accuracy and f-measure with existing fuzzy system and achieves high accuracy.

Precision value

Precision value is evaluated according to the feature classification at true positive prediction; false positive. It is expressed as follows in terms of point value:

$$Precision = (True\ positive)/(True\ positive + False\ positive)$$

The precision value in point value for the existing work and the proposed work is calculated by using the above formula and the values got is represented in a bar chart as shown below.

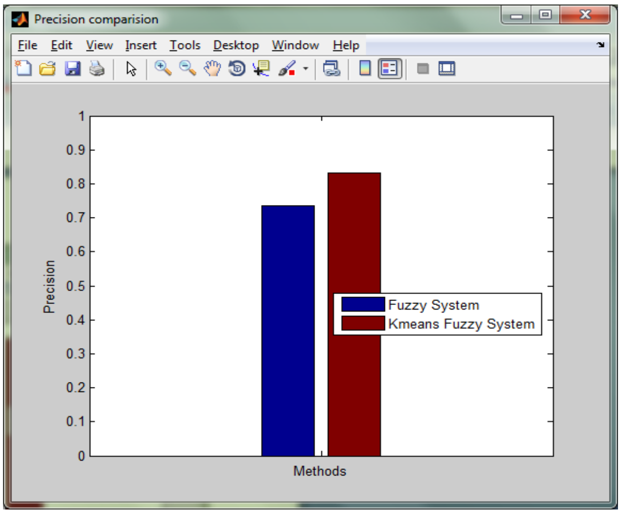


Figure 1 Comparison of precision

Figure 1 show the comparison of the existing fuzzy system with the proposed K-means fuzzy system in terms of precision values.

Recall value

Recall value is evaluated according to the feature classification at true positive prediction, false negative. It is given in terms of point value,

$$\text{Recall} = \frac{\text{Truepositive}}{(\text{Truepositive} + \text{Falsenegative})}$$

The recall value in point values for the existing work and the proposed work is calculated by using the above formula and the values got is represented in a bar chart as shown below.

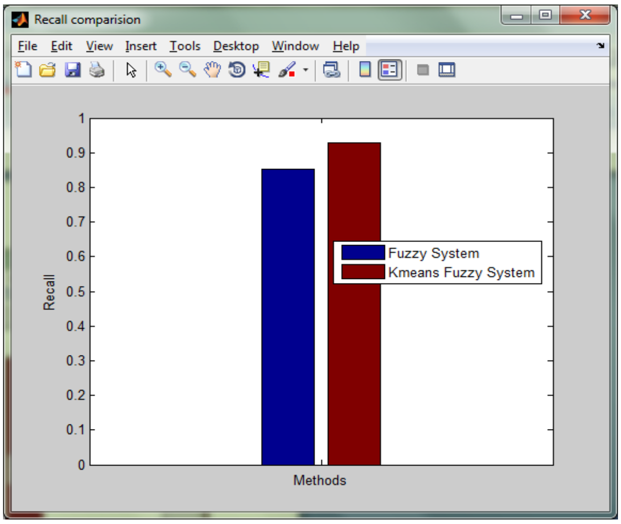


Figure 2 Comparison of recall

Figure 2 show that the comparison of the existing fuzzy system with proposed K-means fuzzy system in terms of recall values.

F-Measure value

F-measure is calculated from the precision value and recall value. It is calculated in terms of point values:

$$\text{f} - \text{measure} = 2 \times \left(\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}\right)$$

The f-measure value in point values for the existing work and the proposed work is calculated by using the above formula and the values got is represented in a bar chart as shown below.

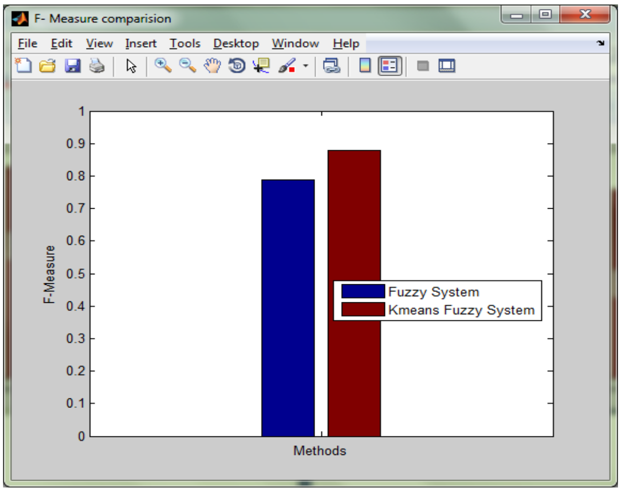


Figure 3 Comparison of f-measure

Figure 3 show that the comparison of the existing fuzzy system with proposed K-means fuzzy system in terms of f-measure.

Accuracy value

The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

Accuracy can be calculated from formula given in terms of %

$$\text{Accuracy} = \frac{\text{Tru e positiv e} + \text{Tru e negativ e}}{\text{Tru e positiv e} + \text{Tru e negativ e} + \text{Fals e positiv e} + \text{Fals e negativ e}}$$

The accuracy value in % for the existing work and the proposed work is calculated by using the above formula and the values got is represented in a bar chart as shown below.

Figure 4 show that the comparison of the existing fuzzy system with proposed K-means fuzzy system in terms of accuracy.

Performance Comparison

The performance comparison of the fuzzy classification system and the proposed k-means fuzzy system in terms of precision, recall, accuracy and f-measure is shown in below table.

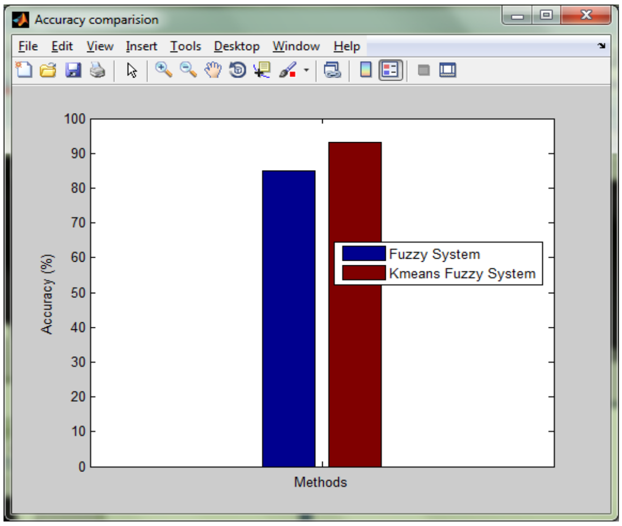


Figure 4 Comparison of accuracy

The accuracy value of the fuzzy system and k-means fuzzy system is given in terms of percentage (%) and the other values such as the precision value, recall value and the f-measure value is given in terms of numeric values.

Comparision table

	Fuzzy System	Kmeans Fuzzy System
Accuracy	85.0006	93.0008
Precision	0.7355	0.8332
Recall	0.8512	0.9287
F- Measure	0.7891	0.8784

Table 5 Performance Comparison

VIII. CONCLUSION

This paper presented an efficient automated system to detect leukemia cells from the blood cancer affected patient's collected blood sample. The microscopic image which was taken for the diagnosis undergoes image processing techniques and k-means clustering algorithm in the proposed work. The image processing techniques such as the pre-processing, image segmentation, feature extraction and fuzzy rule based decision system are performed to get better results than the previous works which are done. The features are classified based on K-means Fuzzy classification to improve the classification accuracy and sensitivity. The experimental results demonstrate that the K-means fuzzy classifier has better accuracy and sensitivity than the fuzzy classification technique. The comparison results of the precision value, recall value, f-measure value and the accuracy value illustrates that the proposed system has better performance than the existing system. Various other clustering algorithms may be used in future to perform diagnosis by reducing time and effort.

REFERENCES

[1] http://www.cancer.gov

[2] http://cancerresearchuk.org

[3] http://medicinenet.com

[4] http://www.webmd.com/heart/anatomy-picture-of-blood

[5] http://www.mayoclinic.org/diseases-conditions/cancer/in-depth/cancer-diagnosis/art-20046459

[6] Ms.Minal D. Joshi, Prof. A. H. Karode, Detection of Acute Leukaemia using White blood Cells Segmentation based on Blood Samples, International Journal of Electronics and Communication Engineering and Technology, Vol 4, Issue 3, May-June 2013,pp. 148-153.

[7] Aimi Salihah, A. N., M.Y. Mashor, Azian Azamimi Abdullah, Improving Blast Segmentation of cute Myelogenous Leukaemia (AML) Images Using Bright stretching Technique, Proceedings of International Post Graduate Conference on Engineering (IPCE 2010), 16th – 17th oct. 2010, Perlis, Malaysia.

[8] Sinha N, Ramakrishnan AG, Automation of differential blood count, In: Chock-Alignam A, editor. Proceedings on the conference on convergent technologies for the Asia-Pacific region, October 15-16. IEEE Publisher: 2003, p. 371-5.

[9] Nur Alom Talukar, Daizy Deb, Sudipta Roy, Automated Blood Cancer Detection Using Image Processing Based on Fuzzy System, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 4, Issue 8, August 2014, ISSN: 2277 128X.

[10] Pawandeep and Arshdeep Singh, A Data Mining Of Leukemia Cancer Detection Using Genetic Algorithm and Neural Network, static.ijcsce.org/wp-content/uploads/2016/01/Leukemia-Cancer.doc.