AN EFFICIENT ENRICHED FEATURE EXTRACTION APPROACHES WITH SEGMENTATION AND FUZZY RULE BASED DECISION SYSTEM TO DETECT LEUKAEMIA

R.Suriyagrace

Research Scholar, Department of Computer Science, Government Arts College, Coimbatore, Tamilnadu, India

Dr.M.Devapriya

Assistant Professor, Department of Computer Science, Government Arts College, Coimbatore, Tamilnadu, India

Abstract– ***The visual examination of a patient's peripheral blood smear plays an important role in the diagnosis of blood cancer. The changes which are found in the white blood cells of the blood determine the blasts. The blasts which are found denote the cancer affected cells from the non-cancerous cells. The identification of the blasts can be carried out by using computerized system aided tools. The feature extraction techniques like identifying the cells by its colour, shape and texture makes the diagnosis process easier. In this paper various feature extraction techniques that are used to recognize the normal lymphocytes from the abnormal lymphoblast cells on the images are discussed. The methodology implies that the application of pattern recognition.***

Keywords– Leukamia detection, Feature extraction techniques, Pattern recognition.

I. INTRODUCTION

1.1 Introduction of cancer

Cancer is a disease that is caused when the abnormal cells grow uncontrollably. When the metabolic pathway is blocked by cells, it is known to be cancer. Cancer is led by certain habits like cigarette smoking, tobacco intake, alcohol intake, poor diet and exposure to UV rays. Different organs like lungs, kidney, eyes, heart, brain, etc., is being affected by cancer cells. The person who works in the chemical factories, nuclear reactors, drainage system and mining are most affected by cancer.

Doctors may find chronic leukaemia in a routine blood test, before symptoms begin. The diagnostic examinations for cancer detection includes physical examination, blood tests and bone marrow test.

1.2 Digital Image Processing

The term "Digital image processing" includes various algorithms to perform techniques based on the image processing on images to extract some useful information. When compared to analog image processing, digital image processing has many advantages. Problems such as noise and signal distortion during processing can be solved by using various algorithms as input. As we know, images are defined in two dimensions, so DIP can be modelled in multidimensional systems.

Digital Image Processing is divided into following 5 groups:

1. Visualization
2. Image sharpening and restoration
3. Image retrieval
4. Measurement of pattern
5. Image Recognition

Digital Image Processing includes the following techniques:

a) Image Acquisition
b) Image Enhancement
c) Image Restoration
d) Color Image Processing
e) Wavelets and Multi-Resolution Processing
f) Compression
g) Morphological Processing
h) Segmentation
i) Representation and Description
j) Object recognition
k) Knowledge Base

1.3 Feature Extraction

Feature extraction is a process in image processing in which the raw data is divided into small parts. The data which are divided from the image are grouped according to its features for managing purpose. The features from the dataset are selected and combined to retrieve the information from datasets. By selecting and combining variables into features, the amount of data is reduced. These features are actual data set with accuracy and originality.

The leukaemia affected cells can be diagnosed in real-time with medical investigations and system aided design and the accuracy of the disease can be recognised by combining a series of characteristic parameter values. The characteristics which determine the extraction include complexity, variability, location and the severity of the disease. In the segmentation phase, features such as colour, shape, statistical feature and texture feature were extracted. In order to improve the accuracy ratio of some features like the minor axis length, area of the nucleus, shape of the nucleus, convex area, eccentricity, perimeter, solidity, orientation, compactness, ratio of nucleus to the cell area are extracted for the blast identification [2].

II. LITERATURE REVIEW

Mohapatra & Patra (2010) in his paper has discussed about color segmentation method that is used for segregating leukocytes or white blood cells (WBC) from the collected blood sample. The features such as the nucleus shape and nucleus texture were used for the detection of leukemia. The shape features like hausdorff dimension and contour signature is implemented for classifying a lymphocytic cell nucleus. Support Vector Machine (SVM) is employed for classification. However some unwanted regions such as segmented red blood cells (RBCs) can still be seen in the image [3].

Mohapatra et al., (2011) in his paper discussed about fuzzy based two stage color segmentation for segregating leukocytes or white blood cells (WBC) from blood components. The features such as the nucleus shape and nucleus texture are used for detecting of leukemia. Contour signature and hausdorff dimension is used for classifying a lymphocytic cell nucleus. Support Vector Machine (SVM) is employed for classification. The major disadvantage of this paper is that the local minima point leads to improper clustering of centroids. [4].

Mohapatra et al., (2012) in his paper has considered image segmentation as a pixel classification problem. Neural architecture is implemented to classify each pixel into cytoplasm, nucleus or background regions. A neural network based lymphocyte image segmentation scheme is designed for detected of leukemia automatically. The segmentation accuracy in terms of nucleus and cytoplasm extraction is found to be high in automated disease recognition system. Disadvantage found is the feature does not contain enough spatial information for precise boundary generation [5].

Laosai & Chamnongthai (2014) in his proposed acute leukemia classification by using SVM and K-Means clustering. The images of stained peripheral blood smears are taken as input and identifies the class of each of the White Blood Cells (WBC). The process that are involved are the segmentation, feature extraction and classification. The work mainly focuses on classification of Foil of Bretagne (Lymphoid) and Almeida Lloyd (Myeloid). The experiment results shows that the performance of identification of leukemia using image processing techniques could classify 100 sample images to Lymphoid stem cells and Myeloid stem cells. The evaluation is performed by using the K-Means clustering algorithm and other classifiers are used to explore different combinations of feature sets. The results thus achieved are based on trials conducted with normal cells [6].

Goutam, D., & Sailaja (2015) in his paper has proposed a work using the supervised classifier for classifying acute myelogenous leukemia in microscopic blood images. The paper mainly focuses on known techniques such as K-mean clustering, Local Directional path (LDP), and support vector machine (SVM) respectively. The overall system performance is evaluated and calculated by using the defined parameters such as sensitivity, specificity, f-measure, and precision. 90 microscopic blood images were tested, and the proposed framework managed to obtain 98% accuracy [7].

III. PROPOSED METHODOLOGY

3.1 Overview of Methodology

1. The Blood Cell Images (BCI) is pre-processed by image pre-processing steps using the Optimal Contrast Stretching (OCS) technique.
2. The segmentation process is improved by using K-Means Cluster (KMC) along with the Hidden Markov Random Field (HMRF).
3. In order to improve the accuracy, the normal lymphocytes and blast lymphocytes are differentiated.
4. Existing features such as texture, geometry, color and statistical features of nuclei leads to poor detection of efficiency.
5. In order to improve the detection efficiency and to reduce the cost of leukemia detection, features such as the minor axis length, area of the nucleus, shape of the nucleus, convex area, eccentricity, perimeter, solidity, orientation, compactness, ration of cytoplasm to nucleus area and ration to nucleus to cell area are extracted.
6. With addition to the above features some more features such as the Fractal dimension, shape features including contour signature, Grey Level Co-currence Matrix (GLCM), texture features and angular second moment (energy) are extracted.
7. The extracted features are traced by Fuzzy Decision Support System to classify the leukemia affected cells from WBC's.

3.2 Proposed Methods

a. Image Pre-Processing

Image pre-processing is carried by using the Optimal Contrast Stretching (OCS) methods. K-means based segmentation is applied to identify and segment nuclei based on the images. By using these techniques 80 features such as shape, texture, colour and statistical based information of the nucleus and cytoplasm sub-images are extracted. A fuzzy rule based classifier is employed to recognize healthy and blast cells. The images of the blood samples are collected from the website. The images include both segmented areas of normal cells and blast cells which are obtained from the peripheral blood sampled of both normal and leukemia affected patients.

b. Refined Segmentation using K-Means and HMRF Clustering

The images are analysed by segmenting the cells using the K-means cluster and builds a cell image representation model by HMRF. Expectation Maximization probability is used to estimate the model parameters. This carried out convergence iteration until optimal value.

c. Enriched Feature Extraction Process

Feature extraction plays a vital role in the classification system of the blood cell images. The most common features such as the geometrical feature, textural feature, colour features are used for the adoption of features. The geometrical feature includes area, radius, perimeter, convex area, major axis length, compactness and orientation. The textural feature includes momentum, contrast, entropy, skewness. The colour feature includes colour distributing and histogram.

The geometrical feature such as the area feature (Area), length feature ($Length_{var}$) and the compactness feature (Comp) is used for the feature extraction method. In the area feature, area is the amount of pixels which belong to the segmented cell region.

d. Feature extraction based on shape of the nucleus

The detected boundaries for the nuclei are expected to present an ellipse like shape and several features to describe the characteristics. Six features such as the circularity, eccentricity, major and minor axis length, equivalent diameter of the circle with the same area as the region and the perimeter of the detected region are specified.

e. Feature extraction using Grey Level Co-Occurrence Matrix (GLCM)

Initially GLCM of the image was calculated and the features such as the contrast, energy, correlation, homogeneity, mean, standard deviation, entropy, skewness, kurtosis, inverse difference moment were extracted.

f. Feature Rule Based Decision Support System (FRDSS)

This is the process of formulating the mapping from a given input to an output using Fuzzy logic with decision tree.

IV. PERFORMANCE EVALUATION

The results of applying proposed method show satisfactory classification of cells and high values of statistical evaluation parameters. The performance of the classifiers of exiting Gray Level Run Length Matrix (GLRLM) based feature extraction and Support Vector Machine (SVM) for classification (GLRLM+SVM)proposed by Mishra et al., (2018)and proposed Enriched Feature Extraction with OCS+ RS +FRDSS (OCS+ RS + EFE+ FRDSS) is evaluated by these parameters: precision, recall, f-measure, segmentation accuracy and classification accuracy.

4.1.1. Precision comparison results

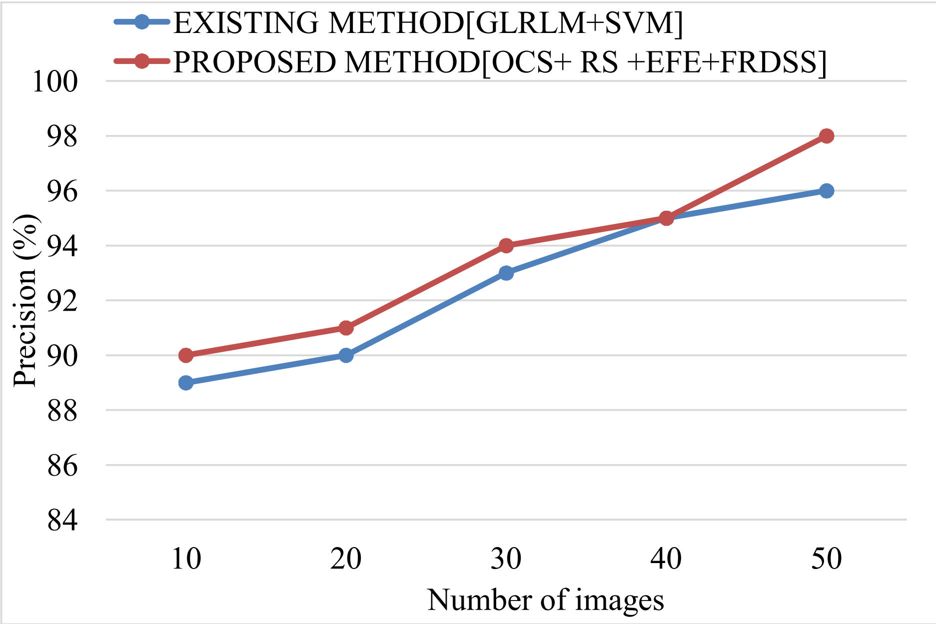

Figure 4.1. Precision performance comparison

Figure 4.1 shows that the precision comparison results of microscopic image dataset between existing and proposed based on number of images. From the figure, the proposed method can obtain high precision rate when compared to existing method. It is an effective way of getting the leukemia nucleus exactly with the high precision rate of 98%. When comparing the precision with the existing method providing low precision rates of 96% which is lower than the proposed method.

Table 4.1. The numerical results of Precision rate

Number of images	Existing Method [GLRLM+SVM]	Proposed Method [OCS+ RS +EFE+FRDSS]
10	89	90
20	90	91
30	93	94
40	95	95
50	96	98

The numerical results of Precision rate are shown in Table 4.1. Initially 10 images were tested with the existing and proposed testing methods. The results were 89% and 90% respectively. When the images were increased to 50, the results were 96% and 98% respectively. This shows that the proposed method has higher precision value than the existing method.

4.1.2. F-measure Result Comparison

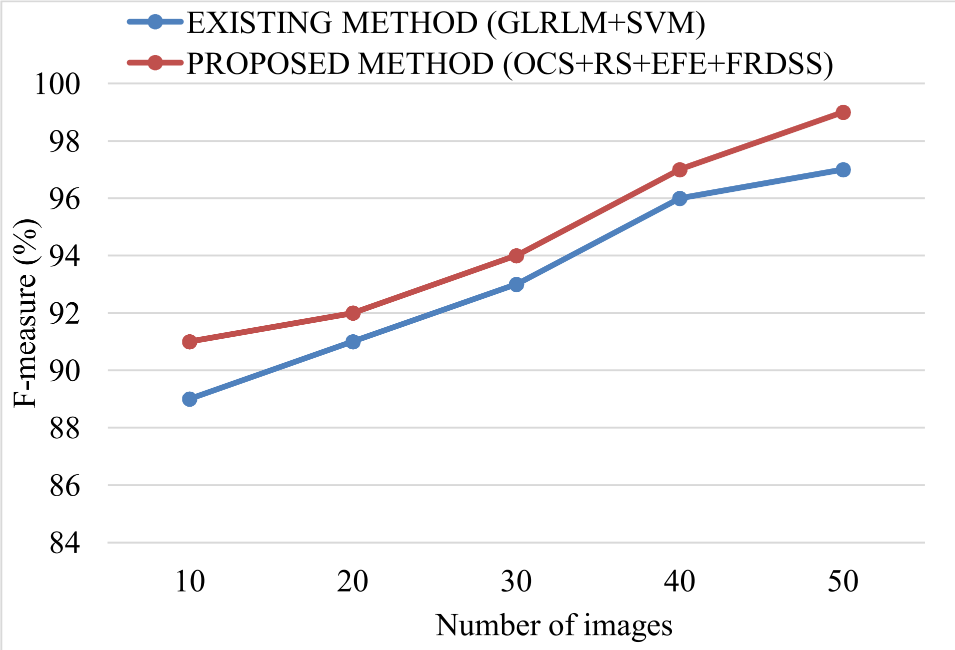

Figure 4.2 F-measure performance comparison

Figure 4.2 shows that the F-measure comparison results between proposed and existing based on number of images. The proposed method has high value of F-measure 99%, which has the advantage over the existing algorithm when the image noise is not negligible. When comparing the F-measure rate with the existing method are providing fewer rates of 97%, which indicates the proposed work can give better segmentation results of leukemia nucleus with high quality.

Table 4.2. The numerical results of F-measure rate

Number of images	EXISTING METHOD (GLRLM+SVM)	PROPOSED METHOD (OCS+RS+EFE+FRDSS)
10	89	91
20	91	92
30	93	94
40	96	97
50	97	99

The numerical results of F-measure rate are shown in Table 4.2. The numerical results of Precision rate are shown in Table 4.1. Initially 10 images were tested with the existing and proposed testing methods. The results were 89% and 91% respectively. When the images were increased to 50, the results were 97% and 99% respectively. This shows that the proposed method has higher precision value than the existing method.

4.1.3. Recall Result Comparison

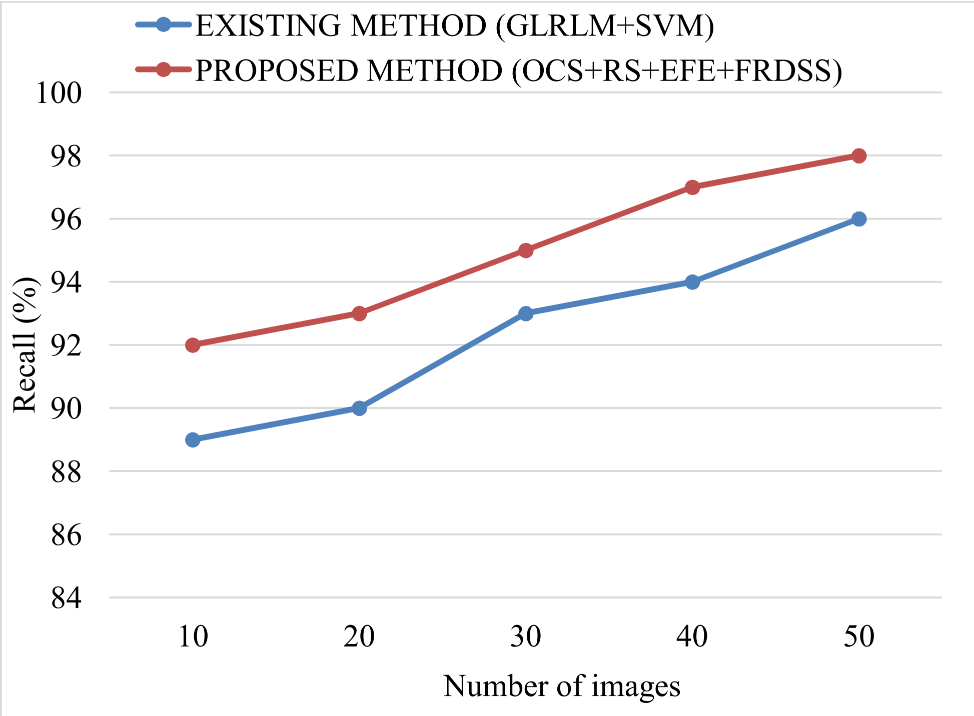

Figure 4.3. Recall performance comparison

Figure 4.3 shows that the recall comparison results between proposed and existing based on number of images. The proposed method has high value of recall rate of 97%. When comparing the recall rate among the existing methods providing recall rate of 96%. The proposed model used the HMRF model for refined segmentation and more feature extraction techniques and thus an accurate and robust segmentation of nucleus can be achieved.

Table 4.3. The numerical results of Recall

Number of images	EXISTING METHOD (GLRLM+SVM)	PROPOSED METHOD (OCS+RS+EFE+FRDSS)
10	89	92
20	90	93
30	93	95
40	94	97
50	96	98

The numerical results of Recall rate are shown in Table 4.3. Initially 10 images were tested with the existing and proposed testing methods. The results were 89% and 92% respectively. When the images were increased to 50, the results were 96% and 98% respectively. This shows that the proposed method has higher precision value than the existing method.

4.1.4. Segmentation Accuracy comparison

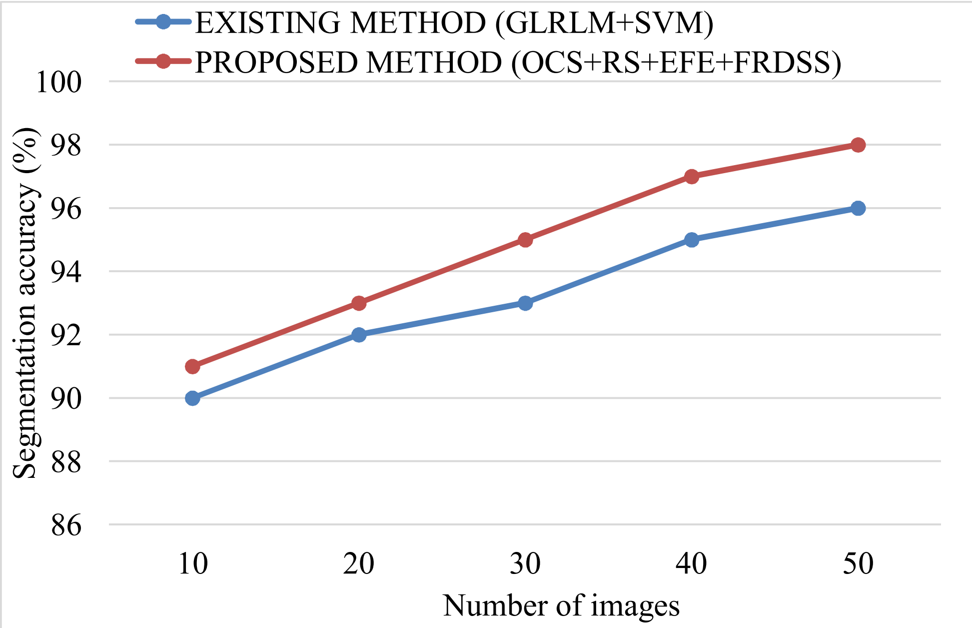

Figure 4.4. Result of Accuracy

From the above Figure 4.4, the graph explains that the segmentation accuracy comparison for prediction of nucleus between proposed and existing techniques. When number of images increased according with the segmentation accuracy value is increased linearly. From this graph, it is learnt that the proposed method effectively select the cluster centre with high segmentation accuracy rate of 98% for nucleus detection. However, the existing method attains accuracy rate of 96% which is much lower than the proposed method. More importantly, as the bias field estimation and the segmentation are mutually influenced by each other, the performance of the segmentation algorithm will have direct influence on the bias field estimation, as well as the overall performance. Also the enriched feature extraction is utilized so for the most optimal features are produced to detect the leukemia.

Table 4.4. The numerical results of segmentation accuracy

Number of images	EXISTING METHOD (GLRLM+SVM)	PROPOSED METHOD (OCS+RS+EFE+FRDSS)
10	90	91
20	92	93
30	93	95
40	95	97
50	96	98

The numerical results of Segmentation accuracy rate are shown in Table 4.4. Initially 10 images were tested with the existing and proposed testing methods. The results were 90% and 91% respectively. When the images were increased to 50, the results were 96% and 98% respectively. This shows that the proposed method has higher precision value than the existing method.

4.1.5. Classification Accuracy comparison

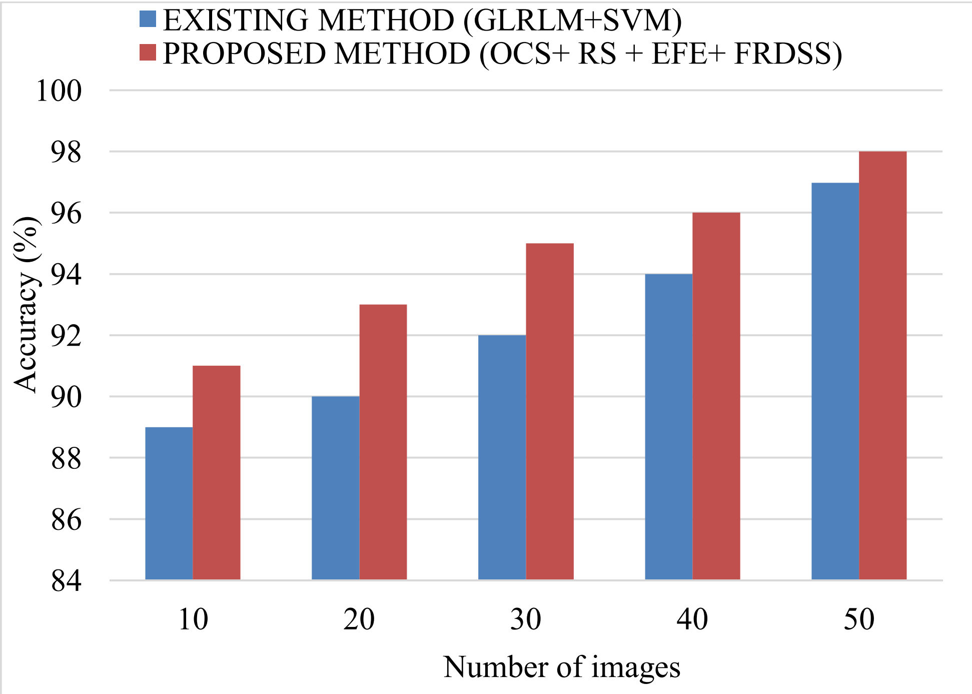

Figure 4.5. Result of Accuracy

From the above Figure 4.5, the graph explains that the accuracy comparison for prediction of leukemia with nucleus segmentation between proposed and existing methods. When number of images increased according with the accuracy value is increased linearly. From this graph, it is learnt that the proposed effectively select the nucleus and detect the leukemia disease with high accuracy rate of 98%. However, the previous method attains accuracy rate of 96.97% which is much lower than the proposed method.

Table 4.5. The numerical results of accuracy

Number of images	EXISTING METHOD (GLRLM+SVM)	PROPOSED METHOD (OCS+ RS + EFE+ FRDSS)
10	89	91
20	90	93
30	92	95
40	94	96
50	96.97	98

The numerical results of Accuracy rate are shown in Table 4.5. The numerical results of Precision rate are shown in Table 4.5. Initially 10 images were tested with the existing and proposed testing methods. The results were 89% and 91% respectively. When the images were increased to 50, the results were 96.97% and 98% respectively. This shows that the proposed method has higher precision value than the existing method.

V. CONCLUSION

In this section, with the development of clinical technologies, different nucleus features have been collected for leukemia detection. This work implements the Enriched features extracted from nucleus using microscopic images to assist in leukemia recognition. This technique resides in its ability to detect leukemia automatically with high accuracy rate of 98% performance which was difficult previously. These fuzzy rules and decision tree rules were used to build the clinical decision support system using the fuzzy inference system. Turning to the future, it is the authors' intentions to extend this work to implement a full algorithm integrated feature selection.

VI. REFERENCES

1. https://www.javatpoint.com/dip-image-transformations
2. https://www.mygreatlearning.com/blog/feature-extraction-in-image-processing/
3. Mohapatra, S., & Patra, D. (2010, December). Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images. In *2010 International Conference on Systems in Medicine and Biology* (pp. 49-54). IEEE.
4. Mohapatra, S., Samanta, S. S., Patra, D., & Satpathi, S. (2011, February). Fuzzy based blood image segmentation for automated leukemia detection. In *2011 International Conference on Devices and Communications (ICDeCom)* (pp. 1-5). IEEE.
5. Mohapatra, S., Patra, D., Kumar, S., & Satpathy, S. (2012). Lymphocyte image segmentation using functional link neural architecture for acute leukemia detection. *Biomedical Engineering Letters*, *2*(2), 100-110.
6. Laosai, J., & Chamnongthai, K. (2014, March). Acute leukemia classification by using SVM and K-Means clustering. In *2014 International Electrical Engineering Congress (iEECON)* (pp. 1-4). IEEE.
7. Goutam, D., & Sailaja, S. (2015, March). Classification of acute myelogenous leukemia in blood microscopic images using supervised classifier. In *2015 IEEE International Conference on Engineering and Technology (ICETECH)* (pp. 1-5). IEEE.
8. Gautam, A., Bhadauria, H. S., & Singh, A. (2014). White blood nucleus segmentation using an automated thresholding and mathematical morphing. In *international conference on Advances in Engineering and Technology, India. TheIRED*.