International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.10, October 2021 E-ISSN: 2347-2693

Effective Image Pre-Processing Techniques with Deep Learning for Leukemia Detection

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

[1,2]Department of Computer Science, Government Arts College (Autonomous), Coimbatore, India

**Corresponding Author: gracesuriya@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v9i10.2836 | Available online at: www.ijcseonline.org

Received: 16/Oct/2021, Accepted: 20/Oct/2021, Published: 31/Oct/2021

Abstract- Leukemia is a cancerous disease characterised by an uncontrollable development of abnormal White Blood Cells (WBC). The identification of acute leukaemia is based on the percentage of WBC in the peripheral blood. In practice, the manual microscopic examination methods are used for acute leukemia detection. Despite the use of hardware autofocus mechanisms, large image collections acquired by automated microscopes often contain some fraction of low quality, out-of-focus images. More complicated cell morphology, with a wide range of size, border, position, and colour contrast were also obtained. Moreover, when the images are captured, the contrast between the cell border and the background in peripheral blood smears is influenced by the lighting position, and the effects of unwanted noise on blood leukemia images can results .in inaccurate diagnosis. So, an efficient pre-processing method is required to highlights the edges of nuclei. This paper describes in detail about the proposed Image Pre-Processing Techniques with Deep Learning Method for Detecting Leukemia in Microscopic Blood Images. This automated system will detect leukemia cells from the blood cancer affected patient's collected blood sample. The image processing techniques used for the diagnosis include optimized contrast stretching (OCS) to enhance the image and detect the nuclei, also the k-means clustering algorithm for nuclei segmentation. A features extraction based on geometry, colour, texture, and statistics information are extracted, as well as fuzzy rule based decision system are performed to get better results of leukemia detection.

***Keywords*:** Leukemia, microscopic examination, Deep Learning method, contrasts stretching, k-means clustering.

1. INTRODUCTION

Leukemia is a type of blood cancer that affects the bone marrow, which is responsible for the production of blood cells. Leukemia is characterised by an abnormal development of blood cells, specifically leukocytes (WBC). Immature blood cells, primarily white cells, have their DNA broken in some way. Because of this defect, blood cells develop and divide in a chaotic manner [1]. Normal blood cells die after a while and are replaced by new bone marrow cells. Because abnormal blood cells do not die rapidly, they accumulate and take up more and more space. The normal blood cells die and replaced by new cells in the bone marrow. The abnormal blood cells do not die so easily, and accumulate, occupying more and more space. As these disrupted blood cells take up more and more space, there is less and less space for normal cells so where the person suffers from severe illness or diseases. In simple, the bad cells outnumber the good cells in the blood. After accidents, leukemia is the second biggest cause of death that affects the children aged from 1 to 15 years [2].

Early detection of the disease is critical for the patient's recovery, especially in children. Several research groups have concentrated their efforts on upto date for the developement of computer networks that can analyse various types of medical images and extract important information for medical professionals [3]. The majority of available approaches rely on images taken during a diagnostic process [4]. However, in certain circumstances, the resulting images will be in low contrast, blurred, and filled with unpleasant noises. These issues can make it harder to evaluate crucial leukaemia morphologies, resulting in a higher rate of incorrect diagnoses. To overcome this condition, image pre-processing techniques such as image enhancement are required.

Hence, an optimized contrast enhancement technique with deep learning for parameter optimization of contrast stretching is proposed in this paper to enhance the acute leukaemia images. Initially, a blood cell images from various sources are pre-processed by image pre-processing step using OCS. After pre-processing, images are segmented by k-means clustering for highlighting the affected WBC and generated initial *K* lables. The nuclei of WBCs are segmented based on global and local curvature qualities during image segmentation, and the normal WBC is removed from the microscopic blood image. It supports in the convergence of image representation models. By using this image representation model, the space constraints were estimated using the Expectation Maximization (EM) probability.

Image segmentation becomes more efficient and accurate when two-step segmentations are used together. For improving the segmentation accuracy, Mahalanobis distance was utilised as the similarity measure of spatial

clustering in Lab colour space by combining both colour and spatial information of cell images. This method can segment cell domains like adhesion, coincidence, and complicated morphology with pinpoint accuracy. It makes easier to extract all cell features and correctly detect the nucleus cytoplasm by segmenting the cell image into proper background and Noise-free procedure. The numerous components of leukaemia cells are then estimated using parameters such as texture, geometry, colour, and statistical properties of nuclei. The extracted features are utilised to train a Fuzzy rule-based decision system that uses a single row feature victor of each cell to categorise leukaemia cells from WBC. The relationship among features in microscopic image is calculated as membership function and will proceed for each microscopic image. The fuzzy rule detects the leukemia automatically from microscopic blood images.

The remaining sections are structured as follows: Section II discusses the previous work associated with this research. Section III explains the methodology of the proposed framework and Section IV exhibits its effectiveness. Section V summarizes the entire work and provides the future scope.

2. LITERATURE SURVEY

An automatic technique has been introduced [5] for identifying and detecting Acute Myelogenous Leukemia (AML) and its common subtypes like M2–M5, Initially, microscopic blood smears images were obtained from patients with AML and healthy cases. Then, the WBC was segmented from the other component structure using an image pre-processing and colour segmentation approach. After that, discriminative properties such as irregularity, nucleus-cytoplasm ratio, Hausdorff dimension, form, colour, and texture were retrieved from the complete nucleus images with many nuclei segments. By using a binary Support Vector Machine (SVM) classifier with a 10-fold cross validation technique, images are identified as malignant or noncancerous. However, this strategy needs to improve data quantity to provide large sample numbers to the classification models during the training phase.
An automated detection algorithm was developed [6] for image based acute leukemia detection using K-means clustering algorithm. This method uses a basic enhancement, filtering and segmenting techniques to extract regions of interest by utilizing the K-means clustering systems. The morphological, color, geometric, textual and statistical features were integrated to categorize a matured lymphocyte and a leukemic lymphocyte using Nearest Neighbor (kNN) and Naïve Bayes Classifier. On the other hand, the computational time of this model was high.

Convolutional Neural Network (CNN) models were used [7] to build a new approach for detecting leukaemia in blood pictures. Initially, the pre-trained CNN structure was utilised to extract features straight from the images without any pre-processing. Then, the collected features were used in the classification stage with the help of SVM. The hybrid datasets were created with one blood smears per image and many leukocytes per image to assess the model's performance. However, CNN training would necessitate a significant number of images.

A decision support system was developed [8] for automated image-based acute leukaemia detection methods. The panel selections, segmentation using K-means clustering, feature extraction, and image enhancement were used to identify the leukaemia cells. The decision-making mechanism was improved to recognise cells and their intrinsic structure. Then, the cells were categorised based on the analysis of morphological properties. A public dataset was used to test the decision support system for segmentation approaches to detect specific targeted cells in an image. However, reliable feature extraction and counting the number of cells remain difficult with this method.

A method was presented [9] for segmenting WBC cells from microscopic pictures was Pre-processing, segmentation, and post-processing are the three stages of this approach. Color correction was performed to enhance colour test photos during the pre-processing procedure. To extract picture features in the segmentation step, two approaches were used: Otsu threshold and watershed marker-controlled. An exoskeleton method was utilised to separate neighbouring cells in the post-processing step, followed by image cleaning to eliminate artefacts and tiny objects. This experiment demonstrated that the WBC segmentation, watershed marker-controlled have out-performed Otsu threshold method.

A dual-threshold approach was presented [10] to segment WBC from Acute Lymphoblastic Leukemia pictures. For WBC segmentation, this method efficiently integrates RGB and HSV colour space based single-threshold methods. Pre-processing, threshold segmentation, and post-processing are the three basic components of this approach. One contrast-stretched grey image and one H component image from modified HSV colour space were handled in the pre-processing section. This threshold method was enhanced to improve on traditional single-threshold approaches and a golden section search method which would find the best thresholds results in the threshold segmentation. The mathematical morphology and median filtering were used in the post-processing section to denoise and remove the incomplete WBCs.

An automated method was designed [11] for the detection of leukaemia in human blood samples. The difficulty of k-means clustering and thresholding was overcome in this method by applying image enhancing techniques and arithmetic operations for the nucleus segmentation from WBC. LAB colour space segmentation was used to remove the WBC from the background image. The shape-based nucleus characteristic of the WBCs was calculated using the segmented image. The K-NN classifier was used to distinguish blast cells from normal lymphocyte cells.

However, this approach does not separate overlapping cells.

A new system [12] was created to detect an acute lymphoblastic leukaemia approach from microscopic images to support the decisions of haematologists. This segmentation approach was used lymphoblasts detection by isolating it from the microscopic image. Marker-based segmentation (MBS), grey level co-occurrence matrix (GLCM)-based feature extraction, and probabilistic principal component analysis (PPCA)-based feature reduction were all used in this technique. Based on these features, the textural feature from the nucleus and cytoplasm region of the cropped image was extracted using grey level co-occurrence matrix. But, this model has a significant level of time complexity.

3. PROPOSED METHODOLOGY

In this phase, OCS with deep learning based image enhancement is used on the basis of microscopic images, followed by k-means based segmentation used to identify locate and segment nuclei. Following that, a total of 80 features representing the nucleus and cytoplasm sub-images, with their shape, texture, colour, and statistical-based information are extracted. A fuzzy rule-based classifier is used to distinguish between healthy and blast cells. The figure 1 depicts the overall architecture of the proposed method.

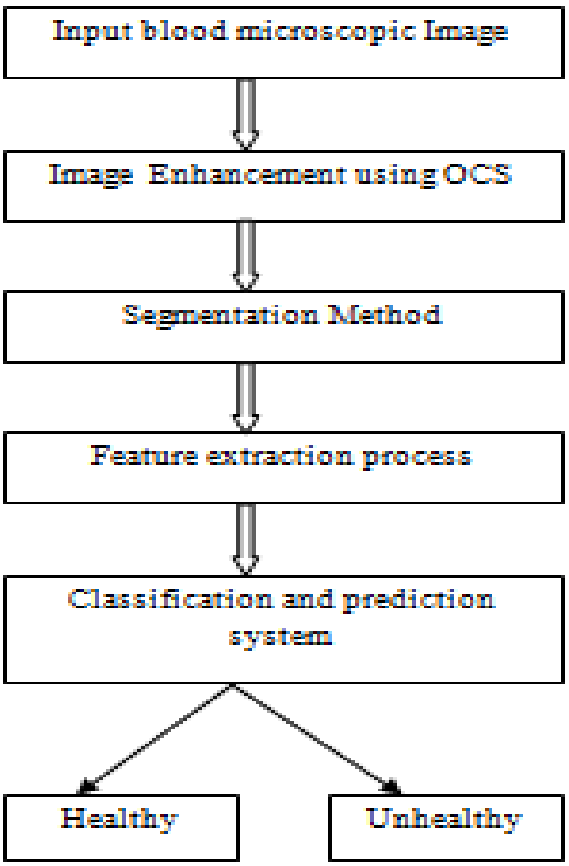


Figure 1 Overall System Architecture of proposed Leukemia Detection

A public Blood Cell Images dataset from kaggle (https://www.kaggle.com/paultimothymooney/blood-cells) was used in this research work for the system evaluation. It consists of segmented sections of normal and blast cells acquired from both leukaemia patients and healthy people's peripheral blood samples. The suggested approach outperforms similar clustering algorithms and Acute lymphocytic (or lymphoblastic) leukaemia (ALL) detection systems published in the literature in terms of accurate nucleus and cytoplasm segmentation as well as robust ALL identification.

4. LEUKAEMIA IMAGE ENHANCEMENT, SEGMENTATION AND CLASSIFCATION

4.1 Image enhancement

The application of CS to pre-process the microscopic blood images improves the global homogeneity, local sensitivity, and geometry of the blood cells. It is implemented through morphological addition and subtraction operations to reduce computing complexity. The suggested technique improves blood cell contrast, which is reflected in WBC nuclei projection. The contrast is increased by 50%, allowing for a more precise target on the nuclei of WBCs in the blood microscopic image. However, the automatic selection of these values will improve the efficiency of the CS. The deep learning is utilized for the automatic selection of the a and b value using the ranked probability score.

To be more specific, the a and b values were determined using a deep neural network trained on the microscopic image. Given training examples of $16-bit$ 84×84 pixel input image patches and the corresponding degree of defocus (one of 11 discrete classes or defocus levels ordered from least to most defocused), the model predicts the probability distribution pd_i for $i \in \{1, \dots, CT\}$ for $CT = 11$ defocus levels, yields a measure of certainty Ce in the range [0.0, 1.0], calculated by normalising the distribution's information entropy.

$$Ce = 1 - \frac{(\sum_{i=1}^{CT} pd_i \log pd_i)}{\log CT} \quad (1)$$

Most probably, Class and the prediction certainty can be represented as coloured borders for each picture patch, with the hue indicating the anticipated class (defocus level) and the lightness indicating the certainty. The model works at the image-patch level and outputs a measure of prediction certainty, allowing for interpretable predictions. The model comprises of convolutional layer with 32 filters of size 5×5, a 2×2 max pool, a convolutional layer with 64 filters of size 5×5, a 2×2 max pool, a fully connected layer with 1024 units, a dropout layer with probability 0.5, and a fully connected layer with 11 units, one for each of the defocus levels. This model was trained with a ranking probability score loss function and the Adam optimizer to accurately penalise model errors on the ordered class categories. From this values a and b values are chosen as from the first two ranked values.

4.2 Segmentation

In this section, K-means clustering is improved to identify the nucleus subspaces that are classed using K-means clustering [14] to cluster the objects into various clusters based on the Euclidean distance metric. In k-means clustering algorithmic rule, the input medical image is pre-processed and converted into grey scale image and enhanced by OCS. Then, the nucleus identification drawback is reduced by exploiting nucleus segmentation and conjointly uses the region growing to extract the nucleus borders. K-means agglomeration issued to count the conventional and abnormal nuclei present within the medical image. Supported cell nuclei distributions in keeping with space, the conventional cells have distribution ranges from thirteen to twenty six pixels and abnormal cells have twenty five to forty five pixels. The projected algorithmic rule identifies nuclei higher than previous nucleus segmentation algorithms by providing high accuracy.

4.3 Feature Extraction Process

In this phase, four types of characteristics are extracted for categorization process. The first feature type is geometrical, which includes the Area, Area feature, Feature Length, $Length_{var}$, and the compactness featureComp. The area feature, Area, is the number of pixels presented in the segmented cell region. The last two properties like lengthvar and compactness are calculated as follows:

$$Length_{var} = \frac{1}{N_b}\sum_{i=1}^{N_b}(L_i - \bar{L})^2 \quad (2)$$

From the above equation, L_i denotes the distance between the i^{th} pixel on the cell boundary and the cell centre. The parameter $\bar{L}$ denotes the average of those lengths.

$$Compactness = \frac{Pc^2}{Ac} \quad (3)$$

Where, Pc is the perimeter of the cell and Ac is the Area of the cell.

Color features such as Hue_{var}, $Saturation_{var}$, and $Intensity_{var}$, are three variations of the hue, saturation, and intensity components of white cell pixels. The texture of nuclei varies depending on the type of WBC. In comparison to other WBCs, the lymphocyte nucleus seems homogeneous. The nucleus of a basophil contains granules, which give the nucleus a black and rough appearance. As a result, in this research work, first order statistic measures such as mean, variance, skewness, and kurtosis, as well as Spatial Gray Level Dependence Matrix (SGLDM) features [17] is used in this in the proposed work. The most recent features as second order statistics are related to co-occurrence matrices such as contrast, angular second moment, correlation, entropy, variance, homogeneity, sum average, sum variance, sum entropy, difference variance, difference entropy, and two information measures of correlation [18].

4.4 Classification and Prediction of Leukaemia

In this phase, a Fuzzy Rule based Decision Support System (FR-DSS) is developed to assist an experts in making decisions for leukaemia diseases which are particular of relevance due to their clear medical diagnostic importance. The fuzzy system's inference is carried out by using the min and max functions, which determine the rule activation strength and rule aggregation, respectively. Once the input-output configuration has been complied successfully, it is determined that the knowledge base would be incorporated in the FISDeT tool (FIS) [19] software designed to assist in building and maintaining the fuzzy rule systems.

The system output is often defuzzified, and the resulting fuzzy sets are merged using an aggregation operator from the consequences of each rule in the input. A single if-then rule is written as follows: IF "X" is A, THEN "Y" is B, or in mathematical notation as

$$\{IF\ (premise)_i\ THEN\ (consequent)_i\}\ where(i = 1,2,\dots,N)$$
$$If\ x_1 = A_1\ and\ x_2 = B_1\ then\ z_1 = class\ 1$$
$$If\ x_1 = A_2\ or\ x_2 = B_2\ then\ z_2 = class\ 2 \quad (4)$$

Where A and B are linguistic variables defined by fuzzy sets on the regions X and Y. The if-part of the rule "$x\ is\ A$" is known as the antecedent or premise, whereas the then-part of the rule, "$Y\ is\ B$" is known as the consequent or conclusion.

Fuzzy sets and fuzzy membership functions: The system is implemented with the assumption that the input and output images acquired after defuzzification are quantized with grey levels ranging from 0 to 255. The fuzzy sets are produced to reflect the intensities of each variable in which these sets are related with the language variables "dark," "edge," and "bright". The Mamdani approach is employed for defuzzification, which implies that the fuzzy sets created by applying each inference rule to the input data are merged using the add function, and the system output is computed using the weighted average method of the resultant membership function. The obtained three values in the output membership functions are intended to distinguish the values of the image's blacks, whites, and edges.

Inference Rules: The inference rules are determined by the weights of the four neighbour grey level pixels, which are either degree of blacks or whites. This rule can directly extract all edges in the produced image. This method evaluates all of the pixels in the processed image by examining the position of each pixel's neighbour. The position of each pixel is determined by employing a floating $2x2$ mask that would scan all greys. Some of the desired rules are explained in this location. Because it takes in four pixel values, the total number of rules generated is sixteen. If at least one of the four input pixel values differs from the other three input pixel values, there is an edge. The figure 2 depicts the flowchart model of the proposed system

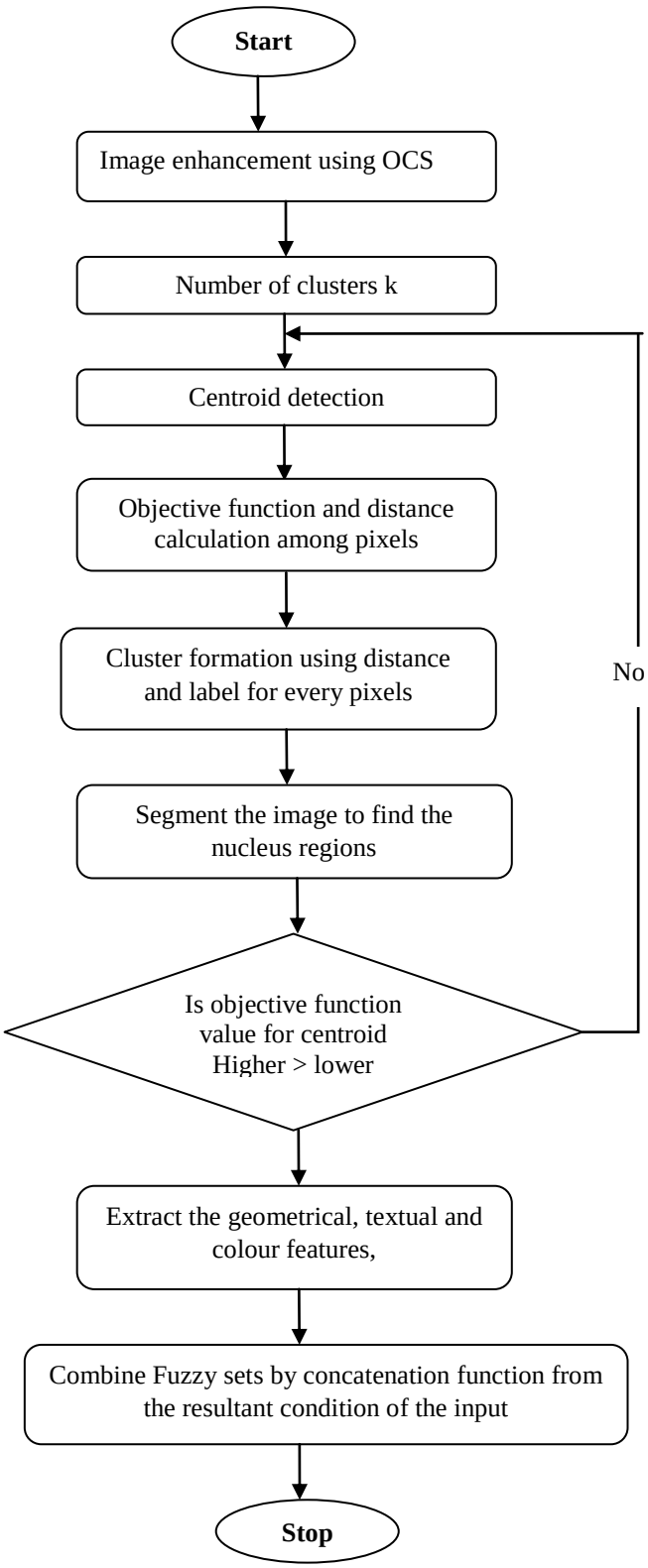


Figure 2 Flowchart for OCS based leukemia detection system

Algorithm

Inputs: Color microscopic Image $I(x, y)$

Output: Contrast Enhanced Image $C(x, y)$

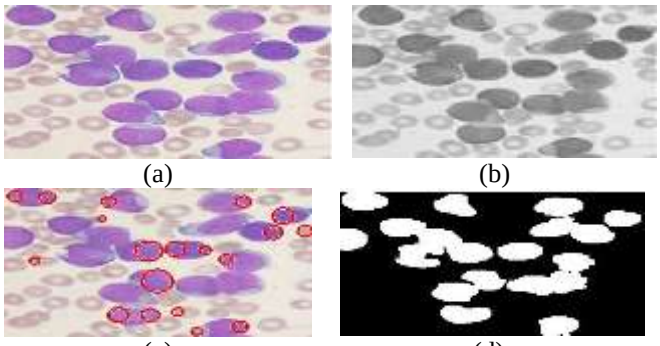
1. First, $I(x, y)$ is transformed into gray scale picture for representing the nuclei of the WBCs as dark area in $I(x, y)$ by using $gr(x, y) = 0.29R + 0.59G + 0.11B$
2. Calculate the primary cluster center K randomly.
3. Find the objective $f(x)$ of K-Means grouping for reducing the addition of squared distances among each point and the cluster centroid $f(x) = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2$, //k denotes the amount of clusters, n denotes the amount of images, x is the iamge and c is the cluster centroid and $x_i^j - c_j$ is the Euclidean distance function.
4. Determine the distance between all pixels and the center and combine them to get the fresh center depending on all cluster centers.
5. Continue processes 2 and 3 until the updated center is identical to the prior center.
6. Contrast stretching is employed to improve image contrast by extending the range of intensity values. The lower and upper boundaries are defined as a and b by the deep learning method
   i. Load the input image and create a synthetically defocused image by generating a probability distribution pd_iover the N ordered categories or defocus levels with certainty for each image patch which is given as
   $$Cw = 1 - \frac{\left(\sum_{i=1}^{CT} pd_i \log pd_i\right)}{\log CT} \quad (5)$$
   ii. The CT grouped categories are aggregated using probability score from these first two values which are considered automatically as the a and b values.
7. The original histogram image (h) is measured using the equation $h = \frac{gr(x,y)}{N}$ by pixels numbers (N) to initialize the limit of $lower(c)$ and $upper(d)$ in histogram, and image $gr(x, y)$ contrast is stretched to $CS(x, y)$ using the equation $CS(x, y) = (gr(x, y) - c)\left(\frac{b-a}{d-c}\right) + a$
8. Histogram equalization (HE) of equation is used for adjusting the gray image intensities which enhances the contrast of the nuclei where li is the level of intensity $HE = floor\left(gr_{li} \sum_{i=0}^{gr(x,y)} h(li)\right)$
9. The equation of Histogram Equalisation (HE) equation is employed to alter the grey image intensities to increase the contrast of the nuclei in which li is termed as the intensity level as, $HE = floor\left(gr_{li} \sum_{i=0}^{gr(x,y)} h(li)\right)$
10. Then perform the morphological addition method $(CS + HE)$; to brighten an image except for nuclei, reducing the resultant pixels that exceeds an intensity value 255 to 255.
11. The subtraction process $SP = CS - HE$ is being used to Highlight all the objects and its borders in the image including the cell nuclei.
12. Finally combing the AP and SP process as $AP + SP$, removes the other blood components and retains the nuclei with less distortion on the nuclei part of the WBC and this result is considered as Contrast Enhanced Image $C(x, y) = AP + SP$.

5. PERFORMANCE EVALUATION

5.1 Dataset description

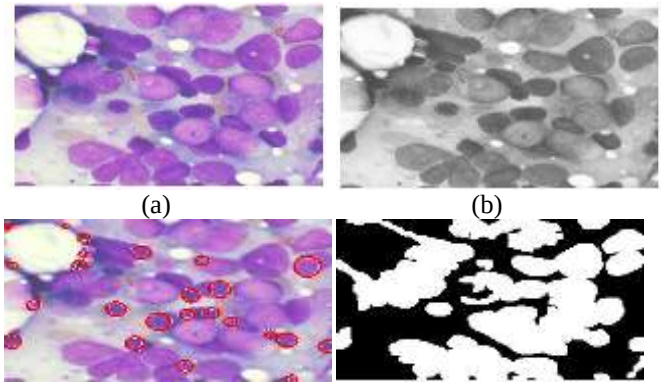
This collection includes 12,500 augmented images of blood cells (JPEG) with cell type labels (CSV). There are around 3,000 photos for each of four different cell kinds, such as Eosinophil, Lymphocyte, Monocyte, and Neutrophil, which are organised into four distinct folders (according to cell type). This dataset is complemented by another dataset that contains the original 410 photos (pre-augmentation), as well as two extra subtype labels (WBC vs WBC) and bounding boxes for each cell in each of these 410 images (JPEG + XML metadata). The folder 'dataset-master' has 410 photos of blood cells with subtype labels and bounding boxes (JPEG + XML), whereas the folder 'dataset2-master' contains 2,500 augmented images and 4 extra subtype labels (JPEG + CSV). There are nearly 3,000 augmented photographs for each of the four classes, compared to 88, 33, 21, and 207 images for each in the 'dataset-master' folder.

The outcomes of applying proposed method show satisfactory classification of cells and high values of statistical evaluation parameters. Result of classification in four types of leukemia images such as Acute myeloid (or myelogenous) leukemia (AML) Chronic myeloid (or myelogenous) leukemia (CML) Acute lymphocytic (or lymphoblastic) leukemia (ALL) Chronic lymphocytic leukemia (CLL) are shown in Figure 3 to 6.



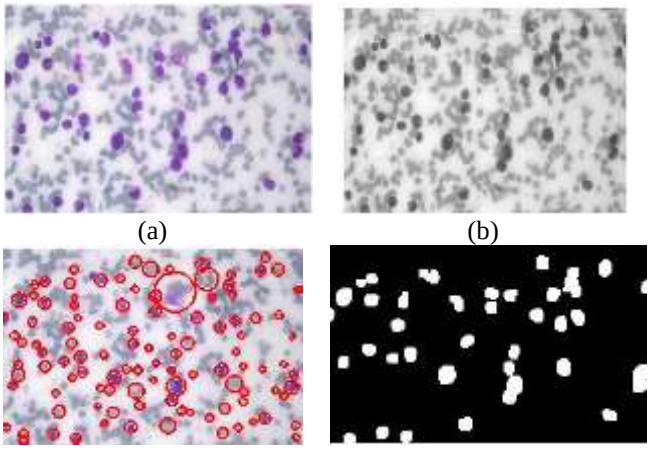
(a) (b) (c) (d)

Figure 3 Results of the proposed algorithm (a) original images (b) gray scale (c) detected images (d) segmented images of ALL using OCS



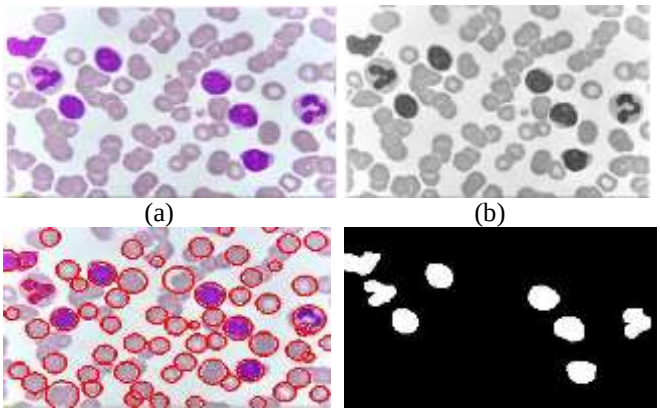
(a) (b) (c) (d)

Figure 4 Outcomes of the proposed algorithm (a) actual images (b) gray scale (c) detected images (d) segmented images of AML using OCS



(a) (b) (c) (d)

Figure 5 Results of the proposed algorithm (a) actual images (b) gray scale (c) detected images (d) segmented images of CLL using OCS



(a) (b) (c) (d)

Figure 6 Results of the proposed algorithm (a) actual images (b) gray scale (c) detected images (d) segmented images of CML using OCS

5.2 EXPERIMENTAL RESULTS

The performance of the classifiers of exiting methods Robust Classifier system (RCS + K-means + SVM) [20]; Hue, Saturation, and Value with K-means and SVM (HSV+ K-means + SVM) [21]; Contrast Improvement with Mid-Range Stretch with Fuzzy C-Means and SVM (CMS + FCM + SVM) [22] and proposed Optimal Contrast Stretching with K-means and fuzzy rule based DSS (OCS + K-means + FRDSS) is evaluated by these parameters: precision, recall, f-measure , segmentation accuracy and classification accuracy.

(1) PRECISION

It is determined as the fraction of cancerous pairs correctly put in the same class label and it is estimated as follows in equation (6)

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Table 1 shows that the precision comparison results of microscopic image dataset between proposed and existing methods based on number of images

Table 1 Numerical results of Precision rate

Number of images	RCS+K-means+SVM	HSV+K-means+SVM	CMS+FCM + SVM	OCS+K-means+FRDSS
10	74	78	82	85
20	83	85	89	91
30	81	84	87	93
40	79	83	89	95
50	83	85	92	96

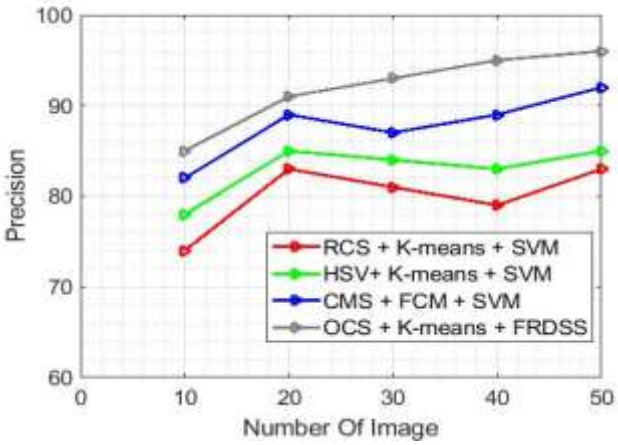


Figure 7 Comparison of Precision performance

Figure 7 demonstrates the precision of presented and classical methods based on various images. When the number of image is 50, the existing RCS+K-means+SVM; HSV+K-means+SVM and CMS+FCM+SVM methods are 15.66%, 12.94 and 4.35% respectively lower than the OCS+K-means+FRDSS. Hence, it is proved that the OCS+K-means+FRDSS achieves a superior precision than the other classical method.

(2) F-MEASURE

It is a harmonic mean of precision and recall as:

$$F - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (7)$$

Table 2 presents the F-measure of microscopic image dataset between proposed and existing methods based on number of images.

Table 2 Numerical results of F-measure rate

Number of images	RCS+K-means+SVM	HSV+K-means+SVM	CMS+FCM + SVM	OCS+K-means+FRDSS
10	80	83	88	89
20	84	87	90	91
30	86	89	92	93
40	89	92	94	96
50	87	93	95	97

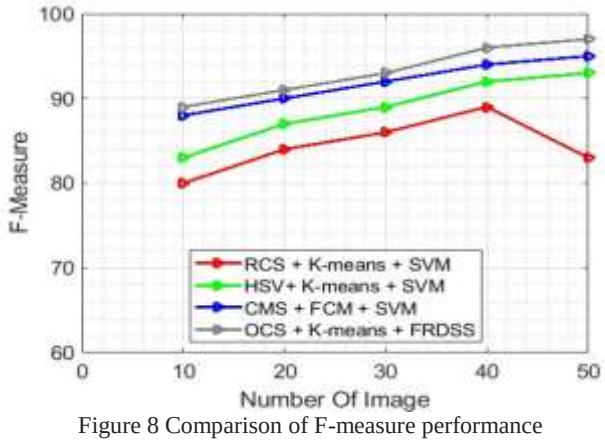


Figure 8 Comparison of F-measure performance

Figure 8 displays the F-measure of presented and classical methods under different number of images. Consider the F-measure for the number of images (50), the existing methods like RCS + K-means + SVM; HSV+ K-means + SVM; CMS+FCM+SVM provides the lower rates of 11.49%, 4.30% and 2.11% correspondingly that notices the OCS+K-means+FRDSS realizes higher segmentation efficacy.

(3) RECALL

This parameter is the probability of being cancerous among the people diagnosed as cancerous. It is defined as:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

Table 3 presents the recall comparison results of microscopic image dataset between proposed and existing methods based on number of images.

Table 3 Numerical results of Recall

Number of images	RCS + K-means + SVM	HSV+ K-means + SVM	CMS+FCM + SVM	OCS+K-means+FRDSS
10	79	82	85	87
20	83	85	87	88
30	85	87	89	92
40	86	88	91	94
50	88	90	93	96

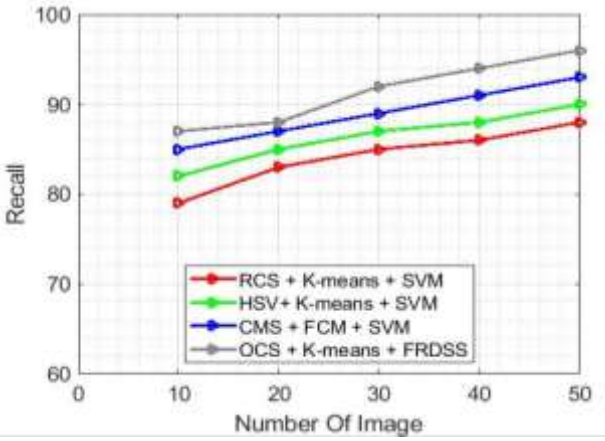


Figure 9 Comparison of Recall performance

Figure 9 demonstrates the recall for presented and classical methods based on number of images. On considering the number of images is 50, it results that the proposed OCS+K-means+FRDSS attains greater recall of 9.09%, 6.66% and 3.23% among the existing methods , RCS + K-means + SVM; HSV+K-means + SVM and CMS+FCM+SVM. This analysis shows the proposed work can give better cluster center selection and thus produce the good nucleus segmented results than the existing method.

(4) SEGMENTATION ACCURACY

To measure generalization ability, two measures are employed. The first measure is the direct pixel density of the ground truth mask and the partitioned mask produced by the segmentation algorithm. Let I_n and IM_n are the amount of non-zero pixels in the resultant nucleus and the amount of pixels in the ground truth mask. The partition accurateness is determined as:

$$segmentation\ accuracy = \left[1 - \frac{IM_n - I_n}{IM_n}\right] * 100\% \quad (9)$$

the segmentation accuracy comparison results of microscopic image dataset with respect to number of images are illustrated in Table 4.

Table 4 Numerical results of segmentation accuracy

Number of images	RCS+K-means + SVM	HSV+K-means + SVM	CMS+FCM + SVM	OCS+K-means+FRDSS
10	78	80	83	87
20	79	83	85	89
30	84	86	89	91
40	83	87	91	93
50	89	91	93	95

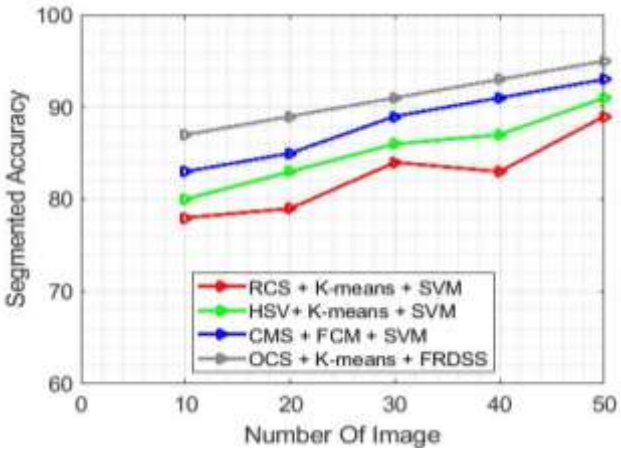


Figure 10 comparison of segmentation accuracy

Figure 10 illustrates the segmentation accuracy of presented and classical methods. The previous methods such as RCS+K-means+SVM; HSV+K-means+SVM CMS+FCM+SVM will attains accuracy rate of 6.74%, 4.39% and 2.15% which is much lower than the OCS+K-means+FRDSS when considering the number of images is 50. This proposed method enhances the image by stretching the intensity range and thus produced high quality nucleus segmentation results to detect the leukemia disease.

(5) CLASSIFICATION ACCURACY

It is the fraction of the exactly categorized images to the overall amount of images.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Table 5 presents the classification accuracy of microscopic image dataset between proposed and existing methods based on number of images.

Table 5 Numerical results of classification accuracy

Number of images	RCS+K-means + SVM	HSV+K-means + SVM	CMS+FCM+ SVM	OCS+K-means+FRDSS
10	77	80	83	88
20	78	82	85	89
30	80	83	86	90
40	81	85	88	91
50	86	88	91	93

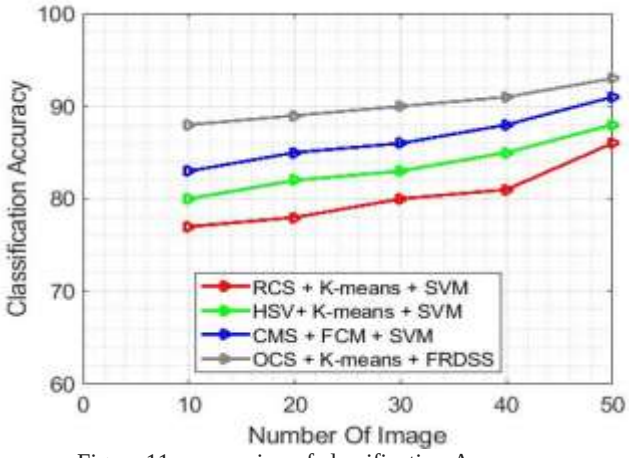


Figure 11 comparsion of classification Accuracy

Figure 11 illustrates the classification accuracy of presented and classical methods. From the given number of images 50, it is learnt that the proposed OCS+K-means+FRDSS effectively select the nucleus and detect the leukemia disease with high accuracy rate of 8.14%, 5.67% and 2.19% in comparsion to previous methods like RCS+K-means+SVM, HSV+K-means+SVM and CMS+FCM+SVM respectively. This analysis proves the proposed OCS+K-means+FRDSS provide better accuracy performance among other methods.

6. 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 by using OCS for developing the image, discovering the nuclei and k-means clustering for the partition of nuclei. Then feature extraction based on geometry, colour, texture and statistical are extracted and fuzzy rule based decision system are performed to get better results of leukemia detection. The comparison results of the precision value, recall value, f-measure value, segmentation accuracy and the accuracy value illustrates that the proposed OCS+K-means+FRDSS have better performance than the existing methods. However, the good segmentation accuracy of the k-means clustering results should be refined by various other clustering algorithms to perform diagnosis by reducing time and effort.

REFERENCE

[1]. Dharani, T., & Hariprasath, S, *"Diagnosis of Leukemia and its types Using Digital Image Processing Techniques",* 3rd International Conference on Communication and Electronics Systems (ICCES) . IEEE, **pp. 275-279, 2018 October.**

[2]. Van Maele-Fabry, G., Lantin, A. C., Hoet, P., & Lison, D, *"Childhood leukaemia and parental occupational exposure to pesticides: a systematic review and meta-analysis",* Cancer Causes & Control, **21(6), 787-809, 2010.**

[3]. Agaian, S., Madhukar, M., & Chronopoulos, A. T. "*Automated screening system for acute myelogenous leukemia detection in blood microscopic images",* IEEE Systems journal, **8(3), 995-1004, 2014.**

[4]. Haidekker, M. *"Advanced biomedical image analysis",* John Wiley & Sons, **2010.**

[5]. Kazemi, F., Najafabadi, T. A., & Araabi, B. N, "*Automatic recognition of acute myelogenous leukemia in blood microscopic images using k-means clustering and support vector machine*", Journal of medical signals and sensors, **6(3), 183, 2016.**

[6]. Kumar, S., Mishra, S., & Asthana, P, *"Automated detection of acute leukemia using k-mean clustering algorithm",* Advances in Computer and Computational Sciences. Springer, Singapore, **pp. 655-670, 2018.**

[7]. Vogado, L. H., Veras, R. M., Araujo, F. H., Silva, R. R., & Aires, K. R, "*Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification",* Engineering Applications of Artificial Intelligence, **72, 415-422, 2018.**

[8]. Negm, A. S., Hassan, O. A., & Kandil, A. H, *"A decision support system for Acute Leukaemia classification based on digital microscopic images",* Alexandria engineering journal, **57(4), 2319-2332, 2018.**

[9]. Salem, N., Sobhy, N. M., & El Dosoky, M, "*A comparative study of white blood cells segmentation using Otsu threshold and watershed transformation"*, Journal of Biomedical Engineering and Medical Imaging, **3(3), 15-15, 2016.**

[10]. Li, Y., Zhu, R., Mi, L., Cao, Y., & Yao, D, "*Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method*", Computational and mathematical methods in medicine, **2016.**

[11]. Choudhary, R. R., Sharma, S., & Meena, G, "*Detection of Leukemia in Human Blood Samples through Image Processing"*, International Conference on Next Generation Computing Technologies. Springer, Singapore, **pp. 824-834, 2017 October.**

[12]. Mishra, S., Majhi, B., Sa, P. K., & Sharma, L, "*Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection",* Biomedical Signal Processing and Control, **33, 272-280, 2017.**

[13]. https://www.kaggle.com/paultimothymooney/blood-cells.

[14]. Chitade, A. Z., & Katiyar, S. K., *"Colour based image segmentation using k-means clustering",* International Journal of Engineering Science and Technology, **2(10), 5319-5325, 2010.**

[15]. Sassi, O. B., Sellami, L., Slima, M. B., Chtourou, K., & *Hamida, A. B, "Improved spatial gray level dependence matrices for texture analysis"*, International Journal of Computer Science & Information Technology, **4(6), 209, 2012.**

[16]. Anuradha, K. "*Statistical feature extraction to classify oral cancers",* Journal of Global Research in Computer Science, **4(2), 8-1, 2013.**

[17]. Castellano, G., Castiello, C., Pasquadibisceglie, V., & Zaza, G, "*Fisdet: Fuzzy inference system development tool",* International Journal of Computational Intelligence Systems, **10(1), 13-22, 2017**

[18]. Laosai, J., & Chamnongthai, K, "*Acute leukemia classification by using SVM and K-Means clustering*", IEEE, International Electrical Engineering Congress (iEECON) **pp. 1-4, 2014.**

[19]. Jagadev, P., & Virani, H. G, (2017, May). *"Detection of leukemia and its types using image processing and machine learning",* IEEE, International Conference on Trends in Electronics and Informatics (ICEI), **pp. 522-526, 2017.**

[20]. Singh, A, *"Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM",* IEEE, 2nd International Conference on Signal Processing and Integrated Networks (SPIN), **pp. 98-102, 2015.**

AUTHOR'S PROFILE

R.Suriyagrace has completed M.Sc Computer Science at Bharathiar University and M.Phil in Government Arts College, Coimbatore. She is currently doing her Ph.D in computer science. Her area of interests is Image Processing, Data Mining and Computer Networks.

Dr. M.Devapriya received her M.Sc., Computer Science from PSG College of Arts and Science Coimbatore in the year 1991. She obtained her M.Phil., and Ph.D., degree from Mother Teresa Women's University, Kodaikanal in the year 2001 and 2011 respectively. She is having around 24 years of teaching experience in various colleges. At present she is working as an Assistant Professor in Computer Science, Government Arts College, Coimbatore. She has published more than 20 papers in National, International journals and Conferences. Her area of interest includes Object Oriented Programming, Image Processing, Multimedia Security and Computer Networks.

