ISSN No. 0976-5697

Volume 8, No. 3, March – April 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

# A Proposed Novel Method for Detection and Classification of Leukemia using Blood Microscopic Images

A. R. Jasmine Begum
Assistant Professor
Dept of Information Technology
Cauvery College for Women
Trichy-620 018, India

Dr.T. Abdul Razak
Associate Professor
Dept of Computer Science
Jamal Mohammed College
Trichy-620 020, India

***Abstract:*** Leukemia is a form of cancer in the blood cells. Most forms of Leukemia occur in the White Blood Cells. The aim of the study is to recognize leukemia White Blood Cells from blood microscopic images through segmentation. The proposed work, an Enhanced Hybrid Fuzzy C-Means with Cluster Center Estimation is to separate the nucleus from the White Blood Cell image which is received as an output of the Hybrid Fuzzy C-Means with Cluster Center Estimation algorithm. The structural image modelling technique called morphological operation is utilized to extract the nucleus followed by feature extraction to extract the geometric features of the nuclei. The Support Vector Machine which is an empirically good performance pattern recognition tool to perform the classification and prediction of cancer affected blood cell images. An analysis is carried by considering the extracted feature of White Blood Cell and nucleus. Finally the performance of this method is evaluated on the basis of image quality measures such as Peak Signal to Noise Ratio and Mean Square Error. Thus, this method results in a high Peak Signal to Noise Ratio and low Mean Square Error, which indicates the better reconstruction of the images in comparison with other existing methods.

*Keywords:* Leukemia, Segmentation, Morphology, Support Vector Machine, Peak Signal to Noise Ratio, Mean Square Error

## I. INTRODUCTION

Blood smear is a prime mean to track the conditions of affected blood cells as well to monitor patients undergoing treatment for such conditions. There are many adverse conditions that may negatively impact the quantity and quality of blood cells production, its function and lifespan. Examples include anemia, myeloproliferative neoplasms, bone marrow disorders, and leukemia. In general normal, mature or almost mature cells are only released into the bloodstream, but some particular situations can prompt the release of sub-standard cells into the circulation by bone narrow. The detection of substantial increase in abnormal cells is called for further analysis and testing by doctors [1].

Leukemia is an illness caused by unspecified aetiology, it is characterized by an unrestrained and abnormal multiplying of one or more of the WBCs. Leukemia is typified either acute or chronic. Acute leukemia is described as sudden onset of the condition and it progresses rapidly. Whereas chronic leukemia takes a long time to onset. Acute Myeloid Leukemia (AML) is formed in the WBC due to early development and abnormal growth of non-granular WBC.

Acute Lymphocytic Leukemia **(**ALL) is the most common form of leukemia affects the children aged below 19. ALL is caused by an injury to the DNA of a cell in the bone marrow. A leukemic cell replaces normal bone marrow, resulting in the unrestrained and amplified growth and accumulation of "lymphoblasts" or "leukemic blasts". These cells don't function as normal blood cells [2].

## II. LITERATURE REVIEW

From the review of related work and published literature, it is observed that numerous researchers utilized diverse segmentation strategies such as threshold strategy, region based methodologies, edge detection approach, clustering approaches, artificial neural network, fuzzy procedure, watershed algorithm and so on for the segmentation of leukemia images and attempted to discover better result. In [3] used the Ensemble Particle Swarm Model Selection (EPSMS), which is an automated tool for the selection of classification models, in the context of acute leukemia classification. This algorithm does not mandates prior domain knowledge and its enable the selection of highly accurate classification models minus user intervention. It assists the diagnosis of some important blood diseases been developed, it is also tested and the results are presented. The image captured by the camera is processed by a microscope, which produces the results needed to diagnose the diseases such as the number of white blood cells, its size and types. In [4] proposed methodology which allows the analysis of blood cells automatically subjected to image processing techniques, and it evolves as a medical tool to overcome the limitations exist with manual observation. This process could also be utilized for counting, as it exhibits excellent performance and paves way for early diagnostic suspicion, which can be later substantiated by a haematologist through specialised techniques. The proposed work in [5] has unsubstantiated pixel segmentation using clustering algorithm for segmenting the blast in ALL and AML images. Three different clustering algorithms KM, FCM and MKM have been utilized and their performances were compared. The performances of these clustering algorithms have been subjected to analysis both qualitatively and quantitatively. Qualitatively, it is observed that segmentation using MKM algorithm enables the obtaining of the fully segmented blast in the most opt banner compared to the usage of KM and FCM clustering algorithms. In [6] a method is proposed for the extraction of brain tumour for MRI medical images using Particle Swarm Optimization (PSO) and a heuristic global optimization method based on swarm intelligence. Two different planes, axial and coronal planes of the MRI image are considered and conclude that PSO algorithm is best suitable and efficient for the coronal plane. In [7], a hybrid

clustering algorithm is proposed and named as Rough Intuitionistic Fuzzy C-Means (RIFCM) and it proved to be more efficient than the existing standard and other algorithms used in this approach utilizing different datasets. In order to obtain noise free images, a Refined Bit Plane (RBP) algorithm was introduced by the author. A seamless integration of the RBP and RIFCM proposed and applied on leukemia images. The prime purpose of this paper is to institute a superior approach in medical diagnosis better than the existing standard and other ambiguous approaches. The other objective is to propose a computer-aided diagnosis system to assist the doctors in assessing medical images in general, and also a simple and improved diagnosing of the disease in leukemia patients in particular.

## III. PROPOSED METHOD

### *A. Methodology*

The Methodology is used to further improvise the image by evacuating the commotion by involving middle channel. The Figure 1 shows the Flow diagram of the proposed novel method. The Proposed Algorithm an Enhanced Hybrid Fuzzy C-Means Cluster Center Estimation (EHFCMCCE) is defined as below:

A Proposed EHFCMCCE Algorithm

Step 1. Input the HFCMCCE [8] Segmented Image.

Step 2. Applying Histogram Equalization to enhance the dark White Blood Cell (WBC) image.

Step 3. Performing Morphological operations on the histogram equalized image such as Erosion followed by Dilation and Close followed by Open resulting in a nuclei part.

Step 4. Feature extraction is carried out on the resultant nuclei part of the blood cell image.

Step 5. Image classification using SVM classifier is applied to classify the image.

**Input:** HFCMCCE Segmented WBC Image **Output**: A Leukemia affected Nuclei part

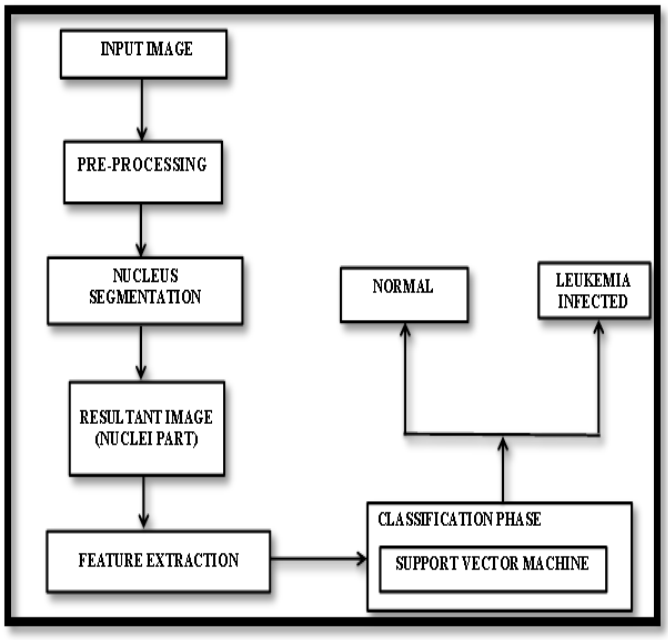



**Figure 1. Block Diagram of Proposed Method**

### *B.Morphological Operations*

Morphological operations are defined as a technique used to study and process the geometrical structure, with regard to lattice hypothesis, set hypothesis, topology, and arbitrary functions. It turns to be very productive after segmenting an image. It is often applied on digital images, further it can be applied on surface meshes, solids, graphs, and any other spatial structures. It subjects the image to a small form or template termed as structuring element. The structuring element is positioned at all probable locations in the image and it is matched with the ensuing region of pixels.

Certain operations examine whether the element "fits" within the region, while some others examine it "hits" or intersect the region. Morphological Operations includes:

- Dilation: It is the comprehensive compilation of points in the image, wherever the structuring element “contacts” the forefront. Taking into account each pixel in the input image, if the structuring element contacts the forefront image; mark it with “1” at the starting point of the structuring element.
- Erosion: Erosion is the complete collection of points in the image, wherever structuring element “fits into”. Consider each forefront pixel in the key in image, if the structuring element fits in; mark it with “1” at the starting point of the structuring element. Pattern matching is the simple application of erosion.
- Opening: Opening contains of an erosion followed by a dilation and is utilized to eliminate all pixels in regions that are too small to rope in the structuring element. Here the structuring element is termed as query, since it examine the image for small objects to strain out of the image.
- Closing: Closing consists of a dilation sequenced by erosion and can be useful to close the small gaps and fill in holes that are tiny in size compared to structuring element [9].

### *Classification Phase*

SVM’s, proposed by Vapnik, are well known pattern recognition tools and were associated with various fields such as text mining, bioinformatics, image classification, cancer diagnosis, and feature selection. SVM’s reputation is attributed to its robust mathematical foundation which is centred on structural risk minimization and statistical learning theory, its scalability to high dimensional datasets, its efficient managing of nonlinear classification by means of kernel functions and its precise performing [10]. To address the issues in classification, an SVM forms a hyper plane or set of hyper planes in a high dimensional space and the basic concept of the SVM is to increase the margin by dividing the input space into two parts while bringing down the total classification errors.

For a given training data $T = \{(y_j, x_j)\}_{j=1}^{q} \in S^m \; X\{-1, +1\}$ the purpose of the classification is to find a function (fun(y) = x) that properly classifies the patterns of the training data, where $y_j$ is a m-dimensional vector and $x_j$ is its label. Figure 2 (a) and (b) shows the hyper planes can be defined $\langle z.y \rangle + c = 1 : z \in S^m, C \in S$ the data is then linearly separable, if such a hyper plane exists. Hyper plane margins

$(\|z\|^{-1})$ must be maximized to find the optimal hyper plane and Lagrange multipliers $(a_i)$ are used to solve this problem. The decision function can be formulated as $fun(y) = sign\left(\sum_{j=1}^{q} x_j b_j (y.y_j) + c\right)$ SVM can also sort out nonlinear classification issues by mapping the input vectors to an upper dimensional space utilizing the kernel functions $i(y_j, y_k) = \langle \emptyset(y_j).\emptyset(y_k) \rangle$ Then, the decision function can be described as $fun(y) = sign(\sum_{j=1}^{q} x_j b_j\, i(y, y_j) + c)$ There are four regular kernel functions: polynomial, linear, sigmoid and radial basis function (RBF).:

1. Linear : $i(y_j, y_k) = y_j.y_k$
2. Polynomial: $i(y_j, y_k) = (y_j.y_k + 1)^d$
3. Radial Base Function: $i(y_j, y_k) = \exp(-\|y_j - y_k\|^2 / 2\sigma^2)$
4. Sigmoid: $i(y_j, y_k) = \tanh(i(y_j, y_k) - d)$

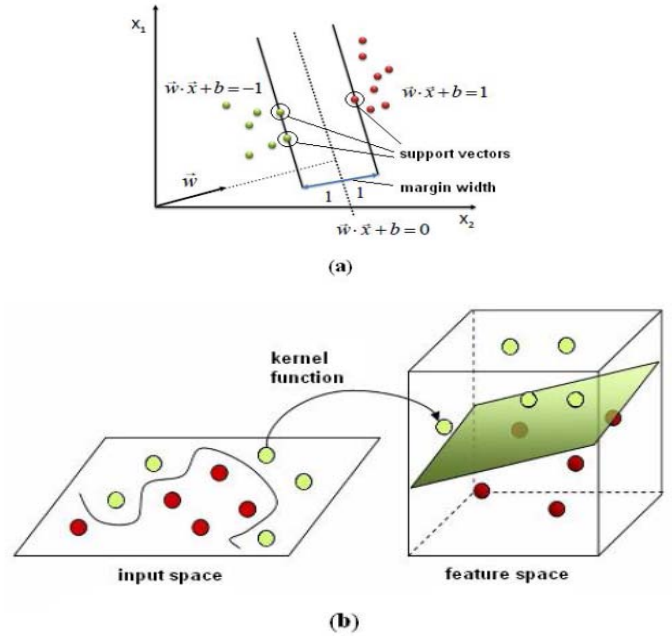



Figure 2 . (a) Classification by SVM (b) Solving nonlinear classification problems using kernel functions

## IV. RESULT ANALYSIS

This Proposed EHFCMCCE algorithm with morphological transformation is implemented using MATLAB R2013a. Figure 3 was obtained from the proposed structure calculation.

The blood smear images were collected from dataset source [11]. These images were digitalized by using a digital camera connected to a Carl Zeiss photo microscope with a magnification of 200x.

In the pre-processing stage the image enhancement is performed using Histogram Equalization which assists to equalize the gravy levels of image intensities. Using morphological operations the nuclei part is segmented and the size of structuring element is smaller than the minimum size of the nucleus. A segmented nucleus is converted into binary image by using im2bw function for feature extraction and it can be subjected to further classification.

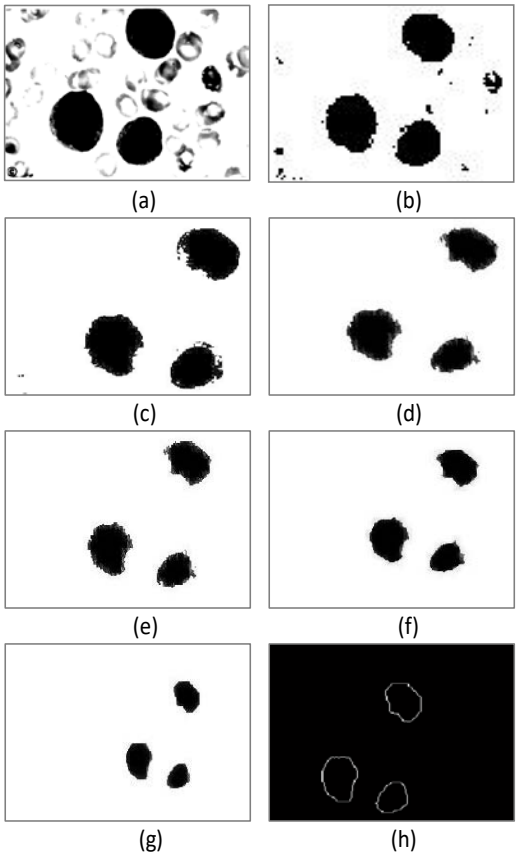



Figure 3. (a) FCM applied image (b) Input image ( HFCMCCE ) (c) Histogram Equalization (d) Dilation (e) Erosion (f) Open (g) Close (h) Binary Image

The size of the blast is one of the key morphological aspects that can be used to validate whether the segmented blast can be treated as either normal or leukemia infected. From the HFCMCCE clustered leukemia image, the types of leukemia can be analysed by the WBC. The normal monocytes generally forms the largest of the WBCs and the diameter of the normal monocytes is 12-20 µm.

Table 1 shows the area, radius and diameter of WBC of HFCMCCE segmented WBC. The diameter of cell 1, cell 2 and cell 3 exceeds the normal range specified, for further accurate analysis the geometrical feature of the nucleus can be taken for. The Equations (1)-(3) are used to calculate the diameter and radius of the WBC [12], where 1 µm/pixel is equal to 0.129 µm [13] .

$$WBC\ Radius = \sqrt{\frac{WBC\ Area}{\pi}} \quad (1)$$

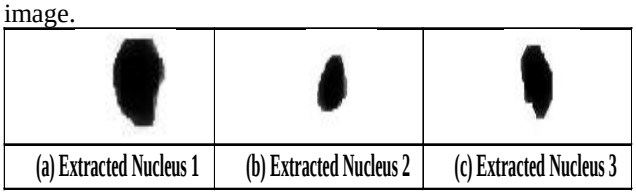
$$WBC\ Radius = \frac{WBC\ Diameter}{2} \quad (2)$$

$$WBC\ Diameter = 2X\sqrt{\frac{WBC\ Area}{\pi}} \quad (3)$$

**Table 1. WBC's Area, Radius and Diameter**

| Measure | Cell1 | Cell2 | Cell3 |
|---|---|---|---|
| WBC Area (in pixels) | 3881 | 2818 | 3186 |
| WBC Radius (in µm) | 12.62 | 10.75 | 11.44 |
| WBC Diameter (in µm) | 25.25 | 21.52 | 22.88 |

The shape of the nucleus is more important in detecting the leukmia. Figure 4 shows the extracted nucleus image.



(a) Extracted Nucleus 1 (b) Extracted Nucleus 2 (c) Extracted Nucleus 3

Figure 4. Extracted Nucleus after close operation using Region of Interest Method

The classification is based on measures of various features obtained during the feature extraction stage. The classification primarily focuses on geometrical features.

Area - The area was measured by counting the total number of none zero pixels within the image region.

Perimeter - Any pixel whose four neighbourhoods are white is certainly forms a non- boundary pixel as it exists in interior of the cell.

Circularity - This is a dimensionless parameter which varies with surface irregularities and is defined in equation (5):

$$\text{Circularity} = 4 * \text{Pi} * \text{Area} / \text{Perimeter}^2 \quad (5)$$

Eccentricity - This parameter is used to measure to what extend the shape of a nucleus deviates from being circular. It's an important feature since Monocytes are more circular than the blast. The value of eccentricity ranges between 0 and 1. An ellipse whose eccentricity is 0 is basically circular. If the value of the eccentricity is below one then it is not circular. Eccentricity is provided by the equation (6):

$$\text{Eccentricity} = \sqrt{a^2 - b^2}/a \quad (6)$$

Where a-Major axis, b-Minor axis.

Solidity - Solidity is utilized to derive out the density of a component. If the solidity value is below one then it seems to possess irregular boundaries [14]. It is given in equation (7).

$$\text{Solidity} = \frac{area}{convex\ area} \quad (7)$$

where convex area returns a scalar that determines the count of pixels in the hull filled in.

An object is convex if any two vertices in it when joined then form an edge which is contained within it [15].

Table 2 the geometrical features of the nucleus such as Area, Perimeter, Eccentricity, Solidity, and Circularity. (Referring to Table 1) shows that the given Nucleus 1, 2 & 3 is identified as a blast by the circularity value is less than 0.90, which indicates that the nucleus is not circular and also denotes the distortion in the shape of the nucleus. If eccentricity value is near to one indicates that the nucleus is a blast. Considering the solidity, if its value is below one then it seems to possess irregular boundaries which again substantiates that the input image contains a blast.

**Table 2. Result of Geometrical Features [Blast-Eccentricity≈1, Solidity<1 &Circularity<0.90]**

| Measure | Nucleus Area (in pixels) | Perimeter (in pixels) | Eccentricity (in pixels) | Solidity (in pixels) | Circularity (in pixels) |
|---|---|---|---|---|---|
| Nucleus1 | 2607 | 213.1960 | 0.8064 | 0.7204 | 0.7203 |
| Nucleus2 | 1556 | 158.8112 | 0.8314 | 0.9425 | 0.7748 |
| Nucleus3 | 2025 | 188.3675 | 0.7668 | 0.9280 | 0.7168 |

In the classification step, based on the feature vectors achieved in the feature extraction step, leukemia infected and normal images are distinguished. To ensure the effectiveness of the classifier and segmentation certain parameters are calculated. The four possible outcome of the classifier is as follows:

True Positive (TP) = No of images having leukemia/total No of images

True Negative (TN) = No of images that have not leukemia/total No of images

False Positive (FP) = No of images that have and detected positive/total No of images

False Negative (FN) =No of images have leukemia and not detected/total No of images

On the basis of the above criteria the below performance measures are defined, the performance of the binary classifier is gauged by the three parameters i.e. sensitivity, specificity, and accuracy which is defined in the equations (8)-(10).

Sensitivity is the measure of how reliable a system is at making positive identifications and it is defined as:

$$sensitivity = \frac{T_P}{T_P + F_N} \quad (8)$$

Specificity is a measure of how well a system can make a negative identification and it is defined as:

$$Specificity = \frac{T_N}{T_N + F_P} \quad (9)$$

Accuracy is a criterion which measures the global performance of the algorithm about the correct decisions of the output of the classifier and it is defined as:

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (10)$$

Finally the performance of the proposed method is analysed by calculating Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) of the input image, FCM applied image, HFCMCCE image and EHFCMCCE which is shown in the Table 3 and Figure 5. The proposed

EHFCMCCE resulting in high PSNR and low MSE values indicates that more noise is reduced. The Figure 6 shows the leukemia infected blood cell image obtained by using the SVM classifier and the Table 4 shows the results of the proposed algorithm for binary SVM classifier with values 81.25%, 89.79% and 83.27%, for sensitivity, accuracy, and specificity, respectively.

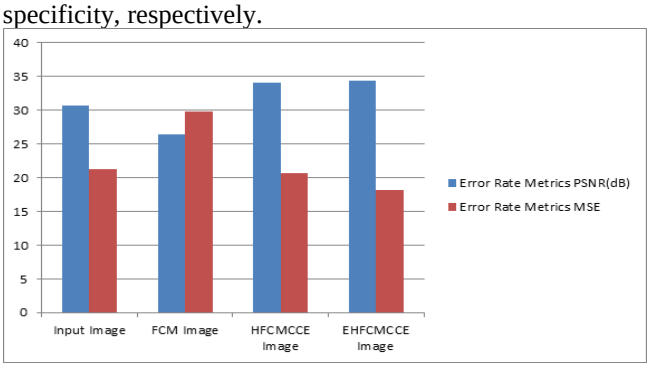



Figure 5. Comparison of the error rate of the Input Image, FCM applied Image, HFCMCCE Image and EHFCMCCE Image

Table 3: Comparison of the error rate of the Input Image, FCM applied Image, HFCMCCE Image and EHFCMCCE Image

| Types of Images | Error Rate Metrics | |
|---|---|---|
| | PSNR (in db) | MSE |
| Input Image | 30.63 | 21.23 |
| FCM Image | 26.42 | 29.78 |
| HFCMCCE Image | 34.08 | 20.73 |
| EHFCMCCE Image | 34.44 | 18.17 |

Figure 6. Classification of Abnormal blood cell images detection using SVM classifier

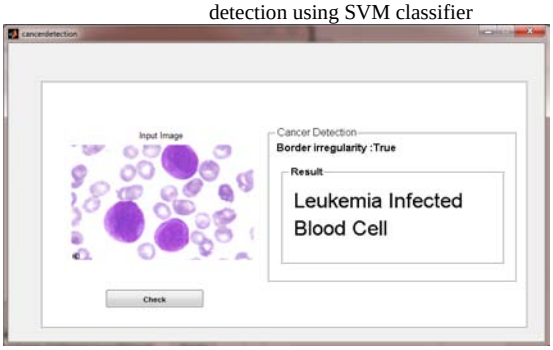
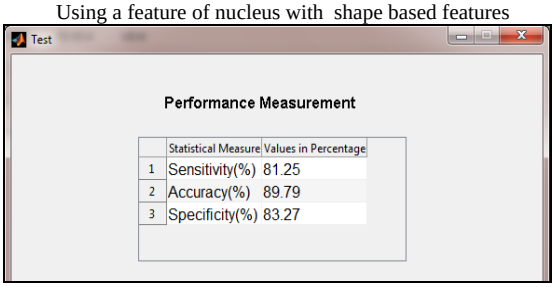



Table 4: Classification details for healthy and infected WBC Using a feature of nucleus with shape based features




## V. CONCLUSION

Experimental results demonstrates the improved execution of segmentation images using the proposed EHFCMCCE algorithm for detection and classification of leukemia using blood microscopic images for enhanced classification. Using the proposed algorithm EHFCMCCE, the image is generated with high PSNR and low MSE in comparison with existing FCM and HFCMCCE generated image. SVM is meant to classify the images as leukemia infected images or normal images. This proposed structure performs as a programmed framework to determine the leukemia influenced cell images and normal images are provided as input info image. In future, the proposed structure can be further improvised by considering texture and statistical features of the images for classification.

## VI. REFERENCES

[1] CANCER DISEASE. HTTPS://WWW.BRITANNICA.COM/SCIENCE/CANCER-DISEASE. DATE ACCESSED: 05/03/2016.

[2] Handbook of CMAI Medical Laboratory Technology. 3rd edition. Christian Medical Association of India, pp. 246, 2012.

[3] Nazlibilek S, Karacor D, Ercan T, Sazli MH, Kalender O, Ege Y, "Automatic segmentation, counting, size determination and classification of white blood cells", Science Direct Measurement., Vol. 55, pp. 58-65, Sep 2014.

[4] Putzu L, Caoccib G, Ruberto C D, "Leucocyte classification for leukaemia detection using image processing techniques", Artificial Intelligence in Medicine., Vol. 62(3), pp.179–91, 2014.

[5] Harun N H, Nasir A S A, Mashor M Y, Hassan R, "Unsupervised Segmentation Technique for Acute Leukemia Cells Using Clustering Algorithms", World Academy of Science, Engineering and Technology International Journal of Computer, Control, Quantum and Information Engineering., Vol. 9, pp. 253-59, 2015.

[6] Mahalakshmi S, Velmurugan T, "Detection of Brain Tumor by Particle Swarm Optimization using Image Segmentation". Indian Journal of Science and Technology, Vol. 8, pp.14-19, 2015.

[7] Purushotham S, Tripathy B K, "A Comparative Analysis of Depth Computation of Leukaemia Images using a Refined Bit Plane and Uncertainty Based Clustering Techniques", Cybernetics and Information Technologies, Vol. 15, pp.126-55, 2015.

[8] Begum A R J, Razak T A, "A Proposed Hybrid Fuzzy C Means Algorithm With Cluster Center Estimation for Leukemia Image Segmentation". International Journal of Control Theory and Applications, International Science Press. Vol. 9, pp. 335-342, 2016.

[9] Atlas of hematology, http://www.hematologyatlas.com/ leukemias.htm, Date Accessed: 27/04/2016.

[10] Gonzalez R C, Woods R E, Eddins S L. Digital Image Processing using MATLAB. In: Image segmentation. Pearson Education Publication: South Asia., pp. 392-438, 2004.

[11] Pirooznia M, Deng Y. SVM Classifier – A comprehensive java interface for support vector machine classification of microarray data. BMC Bioinformatics. Vol. 7, pp. 4-25, 2006.

[12] Suryani E, Wiharto, Polvonov N, "Identification and Counting White Blood Cells and Red Blood Cells using Image Processing Case Study of Leukemia", International journal of Computer Science & Network Solutions, Vol. 2, pp. 35-49, 2014.

[13] Methods to determine the size of an object in microns http://openwetware.org/wiki/Methods_to_determine_the_ size _ of _an_object_in_microns. Date Accessed: 05/08/2016.

[14] Histology Guide blood. http://www.histology. leeds.ac.uk/blood/blood_wbc.php. Date Accessed: 23/05/2016.

[15] Region props http://in.mathworks.com/help/images/ref/ regionprops.html. Date Accessed: 02/08/2016.