

IMAGE SEGMENTATION USING A HYBRID CLUSTERING TECHNIQUE AND MEAN SHIFT FOR AUTOMATED DETECTION ACUTE LEUKAEMIA BLOOD CELLS IMAGES

¹FARAH H. A. JABAR, ¹WAIDAH ISMAIL, ¹ROSALINA A.SALAM, ²ROSLINE HASSAN

¹Faculty of Science and Technology, Universiti Sains Islam Malaysia, Negeri Sembilan, MALAYSIA

²Hematology Department of Universiti Hospital, Universiti Sains Malaysia, Kelantan, MALAYSIA

E-mail: fsaj79@gmail.com,

ABSTRACT

Clustering is one of the most common automated segmentation techniques used in the fields of bioinformatics applications specifically for the microscopic image processing usage. Recently many scientists have performed tremendous research in helping the haematologists in the issue of segmenting the leukocytes region from the blood cells microscopic images in the early of prognosis. During the post processing, image filtering can cause some discrepancies on the processed image which may lead to insignificant result. This research aims to segment the blood cell microscopic images of patients suffering from acute leukaemia. In this research we are using three clustering techniques which are (Fuzzy C-Means (FCM), Classic K-Means (CKM) and Enhanced K-Means (EKM) then we performed filtering techniques which are Mean-shift Filtering (MSF) and Seeded Region Growing (SRG). We tested individual clustering, from the results it show Enhanced K-Means gives the best result. We performed hybrid between EKM and MSF gave a better result from other comparison. The integrated clustering techniques have produced tremendous output images with minimal filtering process to remove the background scene.

Keywords: *Image Segmentation; Enhanced K-Means, Meanshift, Leukaemia Cells*

1. INTRODUCTION

In biomedical application, image processing becomes an interesting area that considered as important role to perform further diagnosis or other task. Segmenting images is among significant steps in image processing methods that has been, and yet still a relevant area in digital image processing due to its wide spread usage and applications. Looking at as the first step, image segmentation is a complex process which is commonly used for medical analysis. The goal of image segmentation is to partition of an image into a set of image regions, which is corresponding to certain properties or characteristics, for object identification, classification and processing [1]. Practically in haematologists' lab, the most common method for evaluating the effectiveness of a segmentation method is a human supervision and comparison with the segmented results for separate segmentation algorithms. However, this process is a tedious and inherently limits the depth of evaluation to a relatively small number of segmentation comparisons over a predetermined set of images [2]. An automated segmentation method helps to

focus on the targeting features. This method are fully automated and uses different kind of automated algorithm such as region or boundary based [3], thresholding[4] and clustering [5]. It has become a great attention for clinical researcher especially for haematologist to analyse the human blood cells and classify the area of interest such as texture, shape or colour. They can identify the clinical behaviour of the disease and predict the abnormalities of the blood cell. Many automated segmentation techniques have been proposed in the literature to overcome the issue of image segmentation specifically in blood cells analysis. All of this effort is to provide valuable information to experts in diagnosis of several diseases related to blood cells.

The rest of the paper is organized as follows. Next section illustrates previous work done in image segmentation using standard existing techniques. Section 3 described about our data set. Section 4 will discuss on our work process. Section 5 explains on our proposed method of clustering algorithm compared with other classical clustering method for image segmentation integrated with

minimal filtering process. In Section 6, the result and analysed, followed by conclusions.

2. IMAGE SEGMENTATION

Image segmentation is considering the initial stage of understanding an image. During this stage, it is the crucial part and hardest problem of image processing. In this experiment, will detect the features of the object required. In automated segmentation technique, the existing techniques for image recognition and visualization are highly depending on the segmentation results. Segmentation is the process of partitioning a digital image into sets of pixels. In biomedical image processing, image segmentation is typically used to locate objects and boundaries in various type of medical images such as MRI (magnetic resonant imaging), x-rays and microscopic. Recently many segmentation tools have been proposed and develop to produce better segmentation on medical images such as clustering [6], active contour [7], thresholding[4] and region-based [8]. Thresholding is among the early techniques developed for image segmentation although the early work on optimal global thresholding is proposed in the classic paper in 1972 by Chow and Kaneko. The technique is significant due to its simplicity and unique properties that provides a central position to the image thresholding. The active contour model known as snake model introduced by Kass in 1988 using the deformable model to extract features, had also been used to segment white blood cells in bone marrow [7]. A proposed model based algorithm to solve the cluster-separation problem in leukocytes cluster using moving interface models and model-based combinatorial optimization scheme [9]. A combination technique of watershed clustering together with feature space clustering is used to segment white blood cells on colour space images for cytoplasm extraction and scale-space filtering for nucleus extraction [10]. Boundary-based methods and region-based methods are another type of segmentation techniques which based on two basic pixels properties that related to their local neighborhood. Some approaches have been developed recently utilized Otsu method together with combination cellular automata and heuristic search [11] while hybrid technique of heuristic search and thresholding to perform automated detection for leukaemia cells [4].

Image clustering is a method of segmenting set pixels with similar properties into a single cluster, then producing different clusters

according to coherence between pixels. This method was one of the first techniques used for segmenting natural images due to its simplicity and efficiency [12]. Generally, a good quality of a clustering method is depends on its ability to discover most of the hidden patterns. Recently there has been a great interest in developing efficient methods for image clustering. K-Means clustering algorithm is also one of the recent techniques that have been proposed in the area of blood cells analysis. K-Means algorithm is one of the clustering algorithms that classify the input data points into multiple classes based on their minimum distance. In medical imaging, many researchers have proven that K-Means clustering has produced good segmentation image due to its performance in clustering of huge datasets [13]. Enhanced K-Means which proposed by [14] has successfully segmented the fish photographic image which significantly produce better and faster segmentation process as compare to conventional K-Means. In blood cells analysis, k-means clustering method is used with thresholding method prior to distinguish nuclei from red blood cells and other particles [15]. Author [16] suggested hybrid K-Means merging with median-cut algorithms for blood cell image segmentation produce better segmented image of the blood cells. In 2000, [17] had proposed a modified technique called Moving K-Means that has proven the blast cell in acute leukaemia blood samples is successfully segmented from its background and unwanted noises [18]. Our work is using an Enhanced K-Means which imposed an efficient way of choosing the initial step for better segmentation of the leukaemia cells. This method shows that the proposed enhanced method would yield better segmentation as compared to standard K-Means [19]. Due to its significant finding in image clustering, this paper is to apply the Enhanced K-Means method integrated with Mean Shift algorithm to remove the background noise for further computer processing on pattern recognition and classification.

3. DATA SETS

The datasets used in this study consist of 100 samples taken from patients suffering from acute leukaemia type AML. The size of the image is 1280 by 960 pixels. All of these images are provided by the Department of Hematology in University Sains Malaysia (USM) located in Kota Bharu, Kelantan, Malaysia.

4. WORKPROCESS

Below show the work process in performing the experiments. In this paper, we perform three methods of clustering which are Fuzzy C-Means (FCM), Enhanced K-Means (EKM) and Classic K-Means (CKM) to find the best clustering among threes. Later in the result and analysis section show, Enhanced K-Means (EKM) is the best. Then we perform the filtering between Means Shift Filtering (MSF) and Seeded Region Growing (SRG) to find the best filtering method in removing the background.

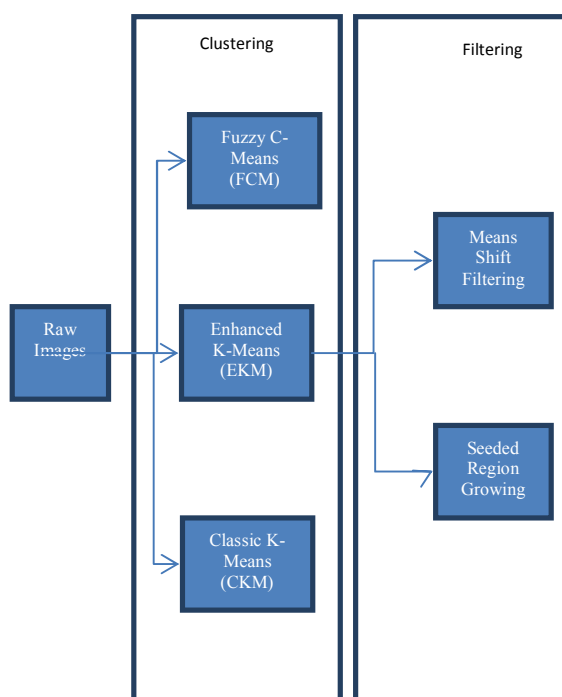


Fig. 1 Work Process For The Experiments

5. PROPOSED METHOD

In this journal we perform a comparison using three methods which are clustering algorithms, Fuzzy C-Means (FCM), Standard K-means (SKM) and Enhanced K-Means (EKM). After we performed the clustering process, then we perform the filtering techniques using Mean Shift Filtering (SMF) and Seeded Region Growing (SRG) in order to remove the background scene.

5.1. Clustering Algorithms

Three clustering techniques have been used to perform the automated segmentation processes which are FCM, SKM and EKM. The result image of these techniques were compared and analyzed. The result image using the best performance technique among the three is applied to the MSF and SRG algorithm to remove the background noise for better segmentation without applying any filtering process. All processes have been applied with Sobel edge detector to highlight the cell boundaries for better viewing.

5.1.1. Fuzzy C-Means (FCM)

The FCM clustering algorithm is a soft segmentation method that has been widely used for microscopic image segmentation [21, 22]. FCM clustering approach could retain information from the original image than hard segmentation methods known as K-Means clustering. The clustering methods are considerable benefits especially for images which contain huge sets of pixel data. In particular, the FCM algorithm, assign pixels to fuzzy clusters without labels. Unlike the K-Means clustering which force pixels to belong exclusively to one class, Fuzzy C-means allows pixels to belong to more than one cluster based on degrees of membership. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the centre of cluster. However, its main drawbacks include its computational complexity and the fact that the performance degrades significantly with increased noise[23].

FCM is a clustering method which allows single data belong to more than one clusters. This method [20] is a pattern recognition based on minimization of the following objective function given as Equation (1) below: where m is any real number greater than 1, is the degree of membership of in the cluster ij , is the i th of d -dimensional measured data, is the d -dimension center of the cluster, and $|$ expressing the similarity between any measured data and the center. Fuzzy clustering approach is carried out through an iterative optimization of the objective function shown in (1) with the update of membership and cluster center .

$$I_f = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|X_i - c_j\|^2 \quad (1)$$

5.1.2. Standard K-Means (SKM)

K-means [24] is one of the simplest unsupervised learning algorithms that solve the clustering problem. The main idea of the learning process is based on the centre based clustering method. The final clustering result of the SKM clustering algorithm is highly dependable on the correctness of the initial centroids, which are normally selected by random. Recent paper suggested hybrid K-Means merging with median-cut algorithms for blood cell image segmentation to produce better image segmentation of the blood cells [16].

The method is divided into 2 phases: first phase is defining the k-centroids, one for each cluster which contributes to the initial steps for the whole process. Then each point from data set will be mapped to the nearest centroids until all points are assigned based on minimum Euclidean Distance. Second phase is to updates each point. The k-centroids need to be recalculate as new k-centroids and new mapping is produced between the points and the new k-centroids. This process will give changes in k-centroids location step by step until the location of k-centroids fixed.

In this paper, the general steps involved in the centre based clustering process. First step is to initialize with initial (c)k-centroids value by using random value. Then for each data point , compute its minimum distance with each centre . Third steps are to find the new centre from all data points for each centre , that belong to this cluster. The final step is to repeat second and third step until the iteration stopped or converged.

The equation show below is the objective function:

$$I(k) = \sum_{i=1}^N \sum_{X_j \in S_i} \|X_j - \mu_i\|^2 \quad (2)$$

The algorithm show below for SKM

```

Input: Images (I)
Find centroid k
For i = 1 to ITER
    Let new centroid
    applied to I(K) equation (2)
End for
Output : New centroid with
segmentaion.
```

Algorithm 1

5.1.3. Enhanced K-Means (EKM)

The EKM method were used by [18] based on originally based on the SKM but the difference the initial value based on the minimum and maximum values in the RGB colour space. In author [18] proposed an enhanced K-Means to segment the leukaemia cell in acute leukaemia blood samples. The clustering was performed after applying threshold method using saturation component formula successfully produced better segmentation on the leukaemia cells.

Despite of using the classic randomly choose initial k-centroids, this Enhanced method manipulates the local minimum and maximum values based on the RGB colour space during the initialization step. The enhanced initialization method returns a two-element array of minimum and maximum RGB values from the whole pixel area. The operator computes the maximum and minimum pixel values for each band of a rendered image within the region of interest. The Enhanced method is an iteration-based clustering that produces an optimal value of initial k-centroids by minimizing the objective function in (3). This approach requires that initial first stage provides reliable input parameters, and that the feature extraction process is controlled only by very few tuning parameters corresponding to intuitive measures in the input domain. where is the Euclidean distance between minimum RGB value and maximum RGB value , n represent the number of image pixels and c is the total number of cluster.

The initial k-centroids can be obtained by using the following objective function:

$$\alpha = \sum_{i=1}^{c_j} \sum_{k=1}^n \| \dot{X} - \ddot{X} \|^2 \quad (3)$$

where $\| \dot{X} - \ddot{X} \|^2$ is the Euclidean distance between minimum and maximum RGB value, n represent the number of image pixels and c_j is the cluster center.

5.2. Filtering

After clustering process, the resultant images will be further segmented using the filtering techniques which are MSF and SRG in order to remove the background scene. The resultant images from both techniques is compared and analysed.

5.2.1. Mean Shift Filtering (MSF)

MSF is a non-parametric clustering algorithm for damping certain properties such as shading or differences in tonality. [25]. It does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. For each pixel of an image have a spatial radius (r) and a particular colour distance (d). The clustering method will be based on the difference scale for r and d .

$r \leq d$ - The cluster image are complex and compact

$r \geq d$ - Produce a cluster

The MSF clustering algorithm is a practical application of the mode finding procedure: (1) starting on the data points, run MSF procedure to find the stationary points of the density function, (2) prune these points by retaining only the local maxima.

5.2.2. Seeded Region Growing (SRG)

SRG is a simple region-based algorithm that groups pixel into larger region which representing distinct image regions and grow them until the entire image is covered [26]. In this experiment the seeds are based on the total number of leukocytes exist in each image. SRG methods can easily separate the regions that have the same properties defined and provide the original images which have clear edges with good segmentation results. The concept is simple which require only a small number of seed points to represent the property to grow the region.

6. RESULTS AND ANALYSIS

This section is explaining the result between Fuzzy C-Means (FCM), Classic K-Means (CKM) and Enhanced K-Means (EKM). As for CKM and EKM, we tested with $k=4$ because in the blood images consists of red blood cells, white blood cells, plasma and background. Then 10 iterations were choosing to make sure the output was

convergence enough. For FCM parameter, we selected the degree of $m = 2$ and $\epsilon = 0.01$. We choose the degree of m and based on the paper [27] that has proven the optimal results.

Fig. 2 - 7 (a) shows the real blood images from microscopic that we choose from 100 images we executed. The result on Fig.2 - 7 (b), (c) and (d) show the clustering techniques FCM, SKM and EKM respectively. Fig. 2 until Fig.7 for FCM show more unwanted particles and noise regions are left over which can be clearly seen using the FCM algorithm. As discussed previously in method section that FCM allows pixels to belong to more than one cluster based on degrees of membership. This makes that the points on the edge of a cluster may be in the cluster to a lesser degree than points in the centre of cluster. Meanwhile, in Fig 2 until Fig 7 for SKM clustering algorithm did not produce good segmentation of the leukaemia cells due to dead centre and centre redundancy problems. At the same times show the inconsistencies of segmented region especially at the interest area of leukaemia cells. But as for EKM procedures is more reliable with respect to noise. Fig. 1 until Fig. 6 show that some of cytoplasm regions for leukaemia cell images can be detected.

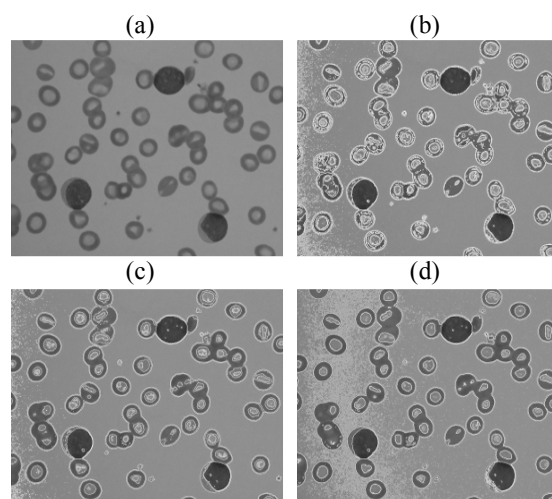


Fig. 2 Sample 1 (A) Real Image (B) FCM (C) SKM (D) EKM

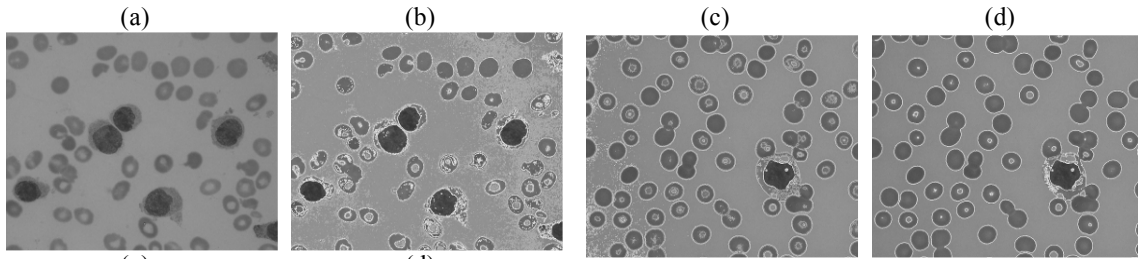


Fig. 3 Sample 2 (A) Real Image (B) FCM (C) SKM (D) EKM

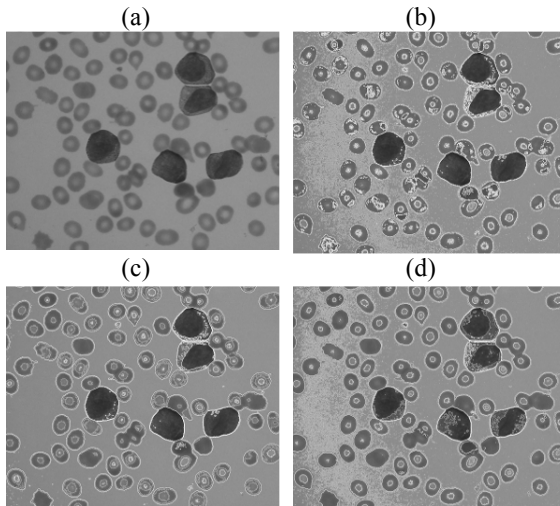


Fig. 4 Sample 3 (A) Real Image (B) FCM (C) SKM (D) EKM

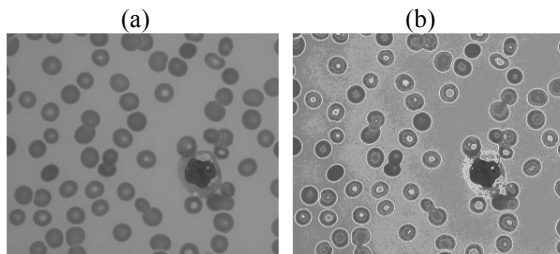


Fig. 5 Sample 4 (A) Real Image (B) FCM (C) SKM (D) EKM

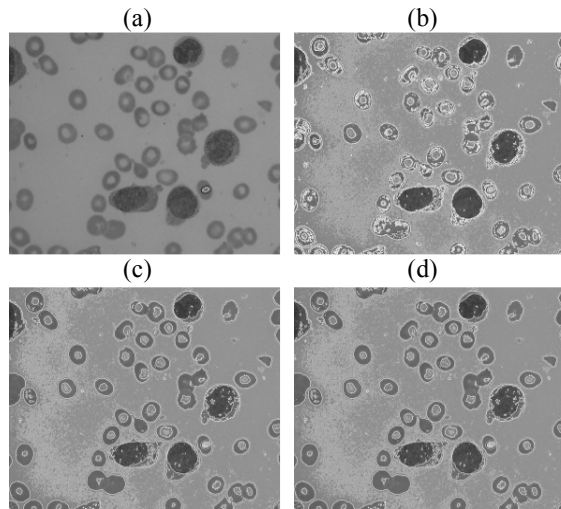


Fig. 6 Sample 5 (A) Real Image (B) FCM (C) SKM (D) EKM

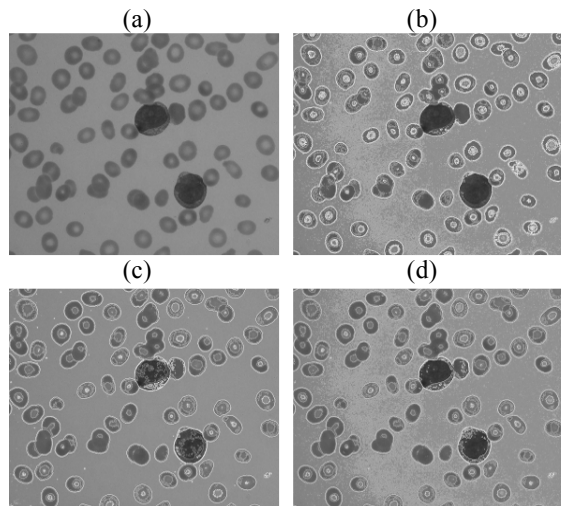
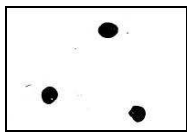

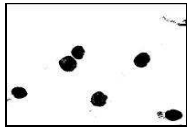

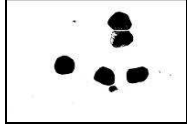
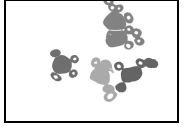
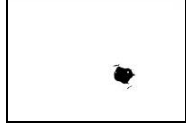
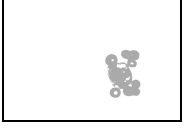


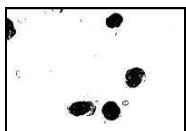
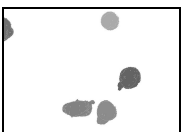

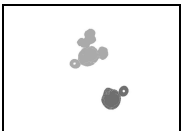
Fig. 7 Sample 6 (A) Real Image (B) FCM (C) SKM (D) EKM

From the Fig. 2 until Fig. 7, we concluded that EKM gave the “best” result with outer comparison. We performed the comparison with the eye view comparison. Then we performed all the EKM clustering images with two methods filtering which are MSF and SRG. The reason we performed the MSF and SRG is to filter out the unwanted background. In Table 1 shows the resultant images of six samples using EKM and MSF compared with EKM and SRG results. In the MSF the spatial radius is set to optimum value of 50 with colour distance of 30 as these values give the best results. For SRG, the seed value is input according to the leukaemia cells exist in each image.

From the results it show the Table 1, EKM-MSF gave a better results than EKM-SRG because MSF effectively smoothest the regions that required and preserving the boundaries. Overall the final output of the segmented images show that the leukaemia cells is segmented tremendously using the Mean Shift filter after applying EKM clustering algorithm.

Table 1 Result From EKM-MSF And EKM-SRG For 6 Samples

ITEM	EKM-MSF	EKM-SRG
SAMPLE 1		
SAMPLE 2		
SAMPLE 3		
SAMPLE 4		

SAMPLE 5		
SAMPLE 6		

6. CONCLUSIONS

Conclusively, this paper presents an Enhanced K-Means (EKM) clustering algorithm for segmenting acute leukaemia blood cells microscopic images with minimal filtering process to remove the background scene. Experimental results shows better performance of segmentation images using the proposed Enhanced K-Means (EKM) method compare to standard K-Means and Fuzzy C-Means. Then between the filtering methods, Mean Shift Filtering (MSF) performed better than Seeded Region Growing for background removal. For future works, this research will continue to perform image recognition and classification for acute leukaemia blood cells image.

Acknowledgements:

We would like to express our gratitude to Department of Hematology in University Sains Malaysia (USM) Kelantan, Malaysia for providing the images data. The research for this paper was financially supported by the Universiti Sains Islam Malaysia, grant no. PPP/UTG- 0213/FST/30/12213.

REFERENCES:

- [1] Lucchese, L. and Mitra, S. K. (2001). “ Colour image segmentation: A state-of-the-art survey.” Computer Vision ECCV 2008. 10th European Conference on Computer Vision Marseille, France, October 2008, Proceedings. pg 207 - 221.
- [2] Zhang, H., Fritts, J. and Goldman, S. (2008). “Image Segmentation Evaluation: A survey of unsupervised methods.” Journal Computer Vision and Image Understanding Volume 110 Issue 2, May 2008, Pg 260-280.
- [3] Freixenet, J., Munoz, X., Raba, D., Marti, J., & Cufi, X. (2002). “Yet another survey on

- image segmentation: Region and boundary information integration". Computer Vision — ECCV. Lecture Notes in Computer Science , Volume 2352, pp. 408-422.
- [4] Nasir, A. S., Mustafa, N., and Nasir, N. F. (2009). "Application of Thresholding Technique in Determining Ratio of Blood Cells for Leukaemia Detection." Proceedings of the International Conference on Man-Machine Systems (ICoMMS), Batu Ferringhi, Penang.
- [5] Mohapatra, S., Patra, D., and Satpathy, S. (2012). "Unsupervised Blood Microscopic Image Segmentation and Leukaemia Detection using Colourbased Clustering" International Journal of Computer Information Systems and Industrial Management Applications , pp 477-485.
- [6] Piuri, V., and Scotti, F. (2004). "Morphology Classification of Blood Leucocytes by Microscope Images". IEEE International Conference on Computational Intelligence International Conference on Image, Speech and Signal Analysis, Taipen, Taiwan (pp. 530–533).
- [7] Park, J., and Keller, J. (2001). "Snakes on the watershed." IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 23, No. 10. (pp. 1201–1205).
- [8] Bezdek, J., Hall, L., and Clarke, L. (1993). "Review of MR Image Segmentation Technique Using Pattern Recognition". Med. Phys. 1993 Jul-Aug;20(4) , 1033-1048.
- [9] Nilsson, B. and Heyden, A. (2002). "Model-based Segmentation of Leukocytes Clusters". Pattern Recognition Proceedings. 16th International Conference on, Vol. 1, pp. 727 – 730.
- [10] Jiang, K., Liao, Q. and Dai, S. (2003). "A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering." 2003 International Conference on Machine Learning and Cybernetics, pp. 2820–2825.
- [11] Ismail, W., Hassan, R. and Swift, S. (2011). "Detecting Leukaemia (AML) Blood Cells Using Cellular Automata and Heuristic Search". In Advances in Intelligent Data Analysis IX Lecture Notes in Computer Science, Arizona, USA. pp. 54-66.
- [12] Lloyd, S. P. (1982). "Least squares quantization in PCM". IEEE Trans. Inf. Theory, vol. IT-28, no. 2, pp. 129–136.
- [13] Malyszko, D., & Wierzchon, S. T. (2007). "Standard and Genetic K-Means Clustering Techniques in Image Segmentation." Computer Information Systems and Industrial Management Applications, 2007. CISIM '07. 6th International Conference on , vol., no., pp.299- 304, 28-30 June 2007
- [14] Samma, A. S. (2009). "Adaptation of K-Means Algorithm for Image Segmentation." International Journal of Information and Communication Engineering, Insider Science Publisher. 5:4.
- [15] Filipczuk, P., Kowal, M. and Obuchowicz, A. (2011). "Automatic Breast Cancer Diagnosis Based on K-Means Clustering and Enhanced Thresholding Hybrid Segmentation." Image Processing and Communications Challenges. Pp 295-302.
- [16] Muda, T. and Salam, R. (2011). "Blood cell image segmentation using hybrid K-Means and median-cut algorithms". IEEE International Conference on Control System, Computing and Engineering (ICCSCE). Penang, 2011.
- [17] Abdul Nasir, A.S.; Mashor, M.Y.; Rosline, H., "Unsupervised colour segmentation of white blood cell for acute leukaemia images," Imaging Systems and Techniques (IST), 2011 IEEE International Conference on , vol., no., pp.142,145, 17-18 May 2011. doi: 10.1109/IST.2011.5962188
- [18] Harun, N. M. (2011). "Automated Blasts Segmentation Techniques Based on Clustering Algorithm for Acute Leukaemia Blood Samples". Journal of Advanced Computer Science and Technology Research, pp 96-109.
- [19] Jabar, F. H., Ismail, W., Rahim, K. A. and Hassan, R. (2013). "K-Means Clustering For Acute Leukaemia Blood Cells Image". Proceedings of the International Conference on Soft Computing and Computational Mathematics.
- [20] James C. Bezdek and Robert Ehrlich (1984) "FCM: The Fuzzy C-Means Clustering Algorithms" Computers & Geoscience. Vol. 10, No. 2-3 pp. 191 - 203
- [21] Ballerini, L., Bocchi, L and Johansson, C. (2004). "Image Segmentation by a Genetic Fuzzy c-Means Algorithm Using Colour and Spatial Information". In Applications of Evolutionary Computing Lecture Notes in Computer Science (pp. 260 - 269).



- [22] Isa, N., Salamah, S., and Ngah, U. (2009). "Enhanced fuzzy moving K-means clustering algorithm for image segmentation". IEEE Transactions on Consumer Electronics, Volume:55, Issue: 4.
- [23] Ng, H., Ong, S., Foong, K. and Nowinski, W. (2005). "An improved watershed algorithm for medical image segmentation." Proceedings 12th International Conference on Biomedical Engineering. Singapore.
- [24] MacQueen, J. B. (1967). "Some Methods for classification and Analysis of Multivariate Observations". Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp.281-297
- [25] Comaniciu, D., & Meer, P. (1999). "Mean Shift: A Robust Approach toward Feature Space Analysis." The Proceedings of the Seventh IEEE International Conference on Computer Vision.
- [26] Adams, R. And Bischof, L., (1994). Seeded Region Growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(6), pp. 641-647
- [27] Okeke, F. and Karneili, A. (2005). "Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm." Ecological Informatics 1, pp 117 - 124.