Classification of *Acute Myelogenous Leukemia (AML M2 and AML M3)* using Momentum Back Propagation from Watershed Distance Transform Segmented Images

Esti Suryani[1], Wiharto[2], Sarngadi Palgunadi[3], and Nurcahya Pradana TP[4]

Ilmu Rekayasa dan Komputasi Research Group, Sebelas Maret University Jl. Ir. Sutami No 36 Kentingan Surakarta
[1]estisuryani@staff.uns.ac.id, [2]wi_harto@staff.uns.ac.id , [3] palgunadi@staff.uns.ac.id , [4]nurcahya.pradana@yahoo.com

Abstract. This study uses image processing to analyze white blood cell with leukemia indicated that includes the identification, analysis of shapes and sizes, as well as white blood cell count indicated the symptoms of leukemia. A case study in this research was blood cells, from the type of leukemia Acute Myelogenous Leukemia (AML), M2 and M3 in particular. Image processing operations used for segmentation by utilizing the color conversion from RGB (Red, Green dab Blue) to obtain white blood cell candidates. Furthermore, the white blood cells candidates are separated by other cells with active contour without edge. WBC (White Blood Cell) results still have intersected or overlap condition. Watershed distance transform method can separate overlap of WBC. Furthermore, the separation of the nucleus from the cytoplasm using the HSI (Hue Saturation Intensity). The further characteristic extraction process is done by calculating the area WBC, WBC edge, roundness, the ratio of the nucleus, the mean and standard deviation of pixel intensities. The feature extraction results are used for training and testing in determining the classification of AML: M2 and M3 by using the momentum backpropagation algorithm. The classification process is done by testing the numeric data input from the feature extraction results that have been entered in the database. K-Fold validation is used to divide the amount of training data and to test the classification of AML M2 and M3. The experiment results of eight images trials, the result, was 94.285% per cell accuracy and 75% per image accuracy

1. Introduction

Acute Myelogenous Leukemia (AML) is a heterogeneous group of hematologic malignancies that are classified according to the predominant type of malignant cells [1]. Myeloid cells are clearly derived from myeloblasts include basophils, neutrophils, eosinophils, monocytes, megakaryocytes and erythrocytes [1]. In 1976, FAB (French American British) Cooperative Group Developed a classification system, classifying AML based on Marrow (marrow), blood (blood morphology) and Cytochemical staining. According to the FAB classification of AML based on morphologic, there are eight, namely M1, M2, M3, M4, M4 E0, M5, M6 and M7 [1]. Particularly in AML M2 characteristic is myeloblastic leukemia with maturation, and there is little Auer rod, round or oval shape of the nucleus and smooth cytoplasm [2], [1]. While the AML M3 is acute promyelocytic leukemia (APL), it has more auer rod, round and lobulated nucleus and smooth cytoplasm [2], [1].

A way to identify leukemia using the traditional procedure to calculate blood cells under a microscope manually, which is time-consuming, laborious, and one of the most expensive routine tests in clinical hematology laboratory [3]. Mechanical counting blood cells that WHO recommended is the technique of using Immunophenotyping. Immunophenotyping is a technique using samples of blood that flowed through the detector (flowcytometer) and then discarded. It is still caused waste, whereas in medical diagnostics is very necessary to record the accuracy of diagnosis [4]. Problems regarding the length of

time the calculation and tracking of the diagnosis can be overcome by using image processing techniques to identify blood cells from images taken from digital microscope [4].

Many types of research have been done on the identification of blood cells by image processing techniques, among others, the segment blood particles contained noise by separating the red blood cells and leukocytes [5]. The segmentation process begins with an image denoising walvet bivariate method and cell edge obtained by the method Kuwahara, binarize images process using a combination of Otsu and Niblick methods, after the process of calculating the blood particles used Immersion Watershed algorithm [5]. The last research carried by authors are identifying leukemia types ALL, based on the morphological image of the white blood cells with Fuzzy Rule Based System, with resulting accuracy of 73.68% [6]. Followed by identifying leukemia types ALL and AML M3 based on the morphological image of white blood cells by the Fuzzy Rule based, the identification process is done by the WBC area ratio and the ratio of nucleus granules with 83.65% system accuracy results [7]. Segmentation is done by using Canny Detection and Ellipse Detection. The study was developed further in 2014, with segmentation using a bounding box for case studies of AML detection that kind of M2 and M4 [7]. The weakness of the research is on segmentation by using Ellipse Detection, and Bounding Box can not separate the good cells that overlap with [6] [7]. This study was done to resolve the problems on segmentation cells overlap, i.e., the method of Active Contour Without Edge (ACWE) and Watershed Distance Transform with case studies AML M2 and M3. ACWE method considered to be suitable for segmentation because it can be used to segment images like pictures with low contrast, varied objects, objects with different intensity, smooth contours and high noise [8]. Watershed Distance Transform is suitable for separates the blood cells that overlap by utilizing the watershed transformation distances combined with roundness cells information [9].

2. Proposed Method

Figure 1 Shows the steps of this research:

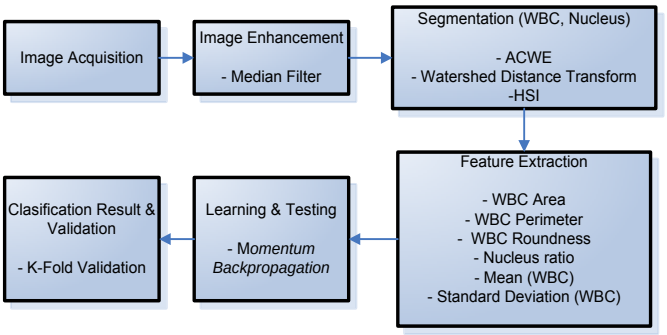


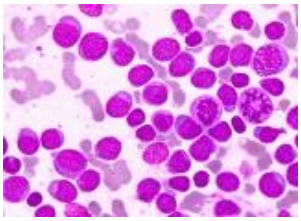
Figure 1. Research Steps

2.1. Images acquisition

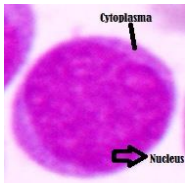
Image acquisition is made by collecting images of AML M2 and M3 blood obtained from the image microscopic blood smear of bone marrow. The image is reserved in Installation of Clinical Pathology, Dr. Moewardi Hospital. The process of collecting data was done by a specialist doctor from the Clinical Pathology Clinical in Dr. Moewardi Hospital. Digital images are obtained by using a digital microscope with a magnification of 1000 times of bone marrow blood smear preparations with Giemsa staining.

2.2. Image Enhancement and WBC Segmentation

Figure 2a is an example of the image which identified as AML M2 that will be used as the input image. WBC cells contained in the figure 2a.



(a)



(b)

Figure 2. a. AML M2 Image; b.WBC Cell with nucleus and cytoplasm

Image data which will be processed (see Figure 2a), must first be repaired with the image enhancement process using a median filter to remove noise image. After enhancement, the clear image then will be segmented to get the WBC cell. Figure 3 shows the steps of WBC segmentation.

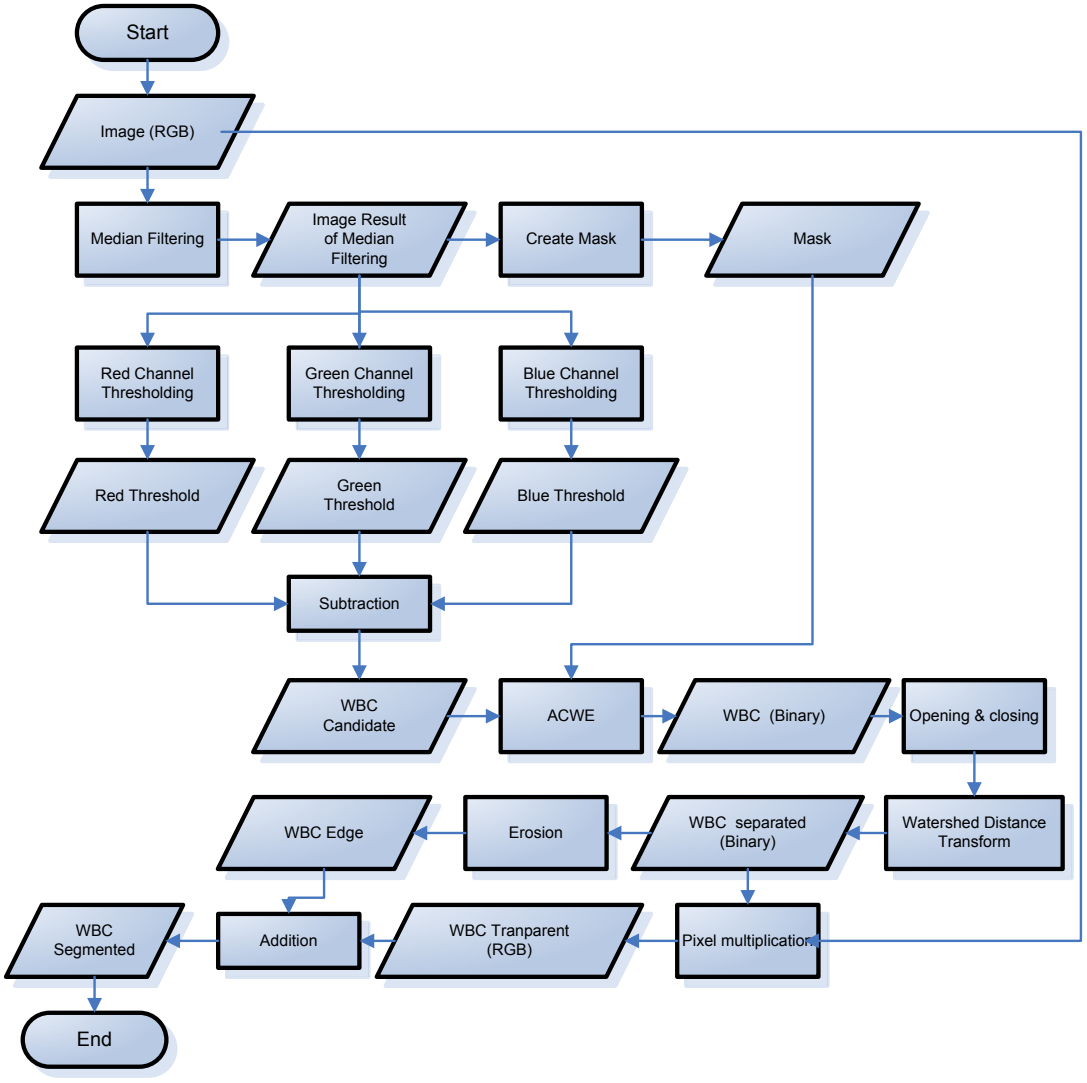
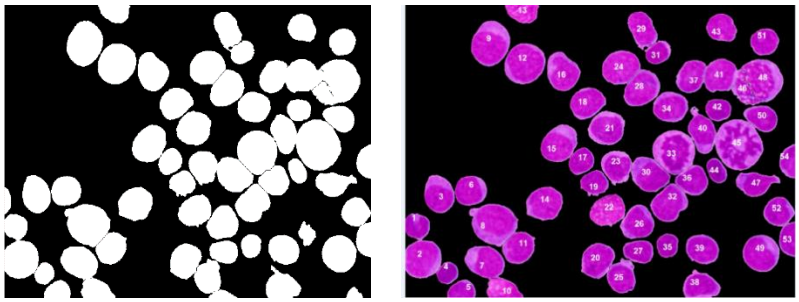


Figure 3. WBC Segmentation Steps

Input for the steps shown in Figure 3 is Figure 2a (AML M2). Median Filtering is done to eliminate noise or small spots contained in images that may interfere in the subsequent segmentation process Furthermore; the following process is creating a mask to get the WBC candidates. Candidates of white blood cells obtained from RGB threshold for each channel red, green, and blue, after a median filter process. The next step is to calculate the size of the resolution of the original image then shaped curve or a mask with a margin from the edge of the image is 100 pixels. Having obtained the candidate of white blood cells and mask, then both combined to do the segmentation process ACWE by calculating

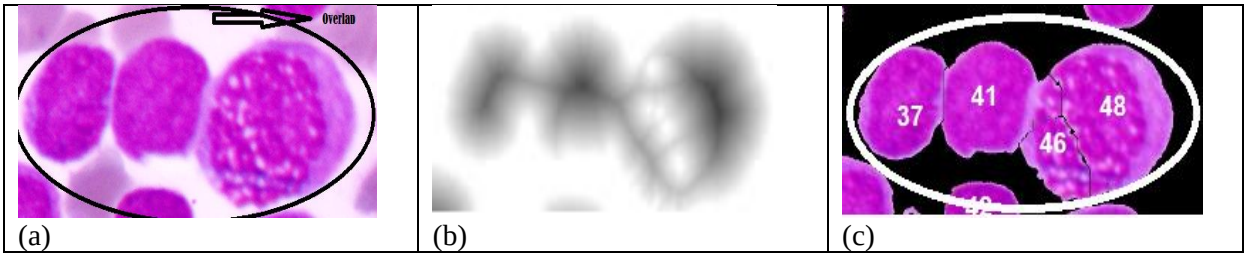
the comparable levels of energy outer and inner energy. The results of ACWE process were a binary image of white blood cells that still contained some cells overlap, and therefore needs to be done by using a cell separation distance transform watershed. Catchment basin needs to be sought by the flooding to some extent to get the position of the cell that can be separated by watershed. The flooding process is done with due regard to the existence of local minima value on each object. From the catchment basin, the centroid point can be obtained which later became the benchmark watershed in conducting separation. The results of the WBC watershed separation process is a WBC binary image (see Figure 4a). The binary image is then multiplied by the pixels of the original image and then counted the number of WBC cells (54 cells), (See Figure 4b).



(a) (b)

Figure 4. a. WBC (binary) ; b. WBC (segmentation result)

In some cases where white blood cell objects are overlapping each other, there are some parts that can not be separated well. Figure 5a shows that there are three cells that overlap each other and form a series of fused cells. The catchment basin can be obtained by flooding the object (see Figure 5b). After the watershed transform based on the distance the catchment basin, it acquired four cells (see Figure 5c). A catchment basin is calculated by finding the value of local minima on the object. Therefore, error separation on the object instance of white blood cells (Figure 5a) is because there are four minima (Figure 5b) so that the cell is separated into four parts, namely object 37, object 41, object 46 and object 48 (see Figure 5c).



(a) (b) (c)

Figure 5. a. Overlapping Cells (3 cells, original image); b.catchment Basin; c. Result of watershed segmentation (4 cells identified)

2.3 Nucleus Segmentation

The next step is to separate the WBC nucleus. In the process of segmentation of the nucleus, it should be conducted in stages, shown in Figure 6.

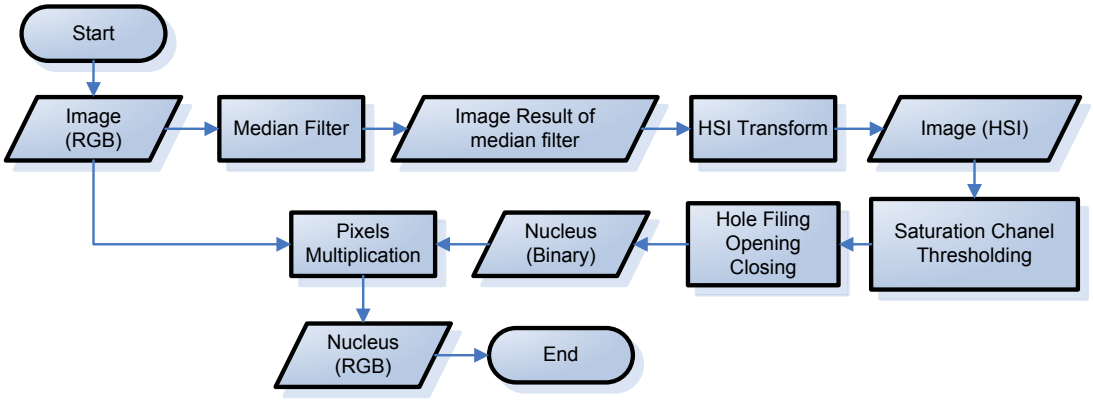
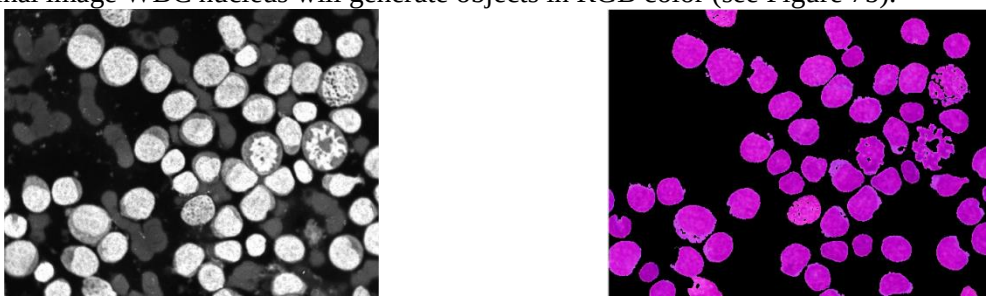


Figure 6. Nucleus Segmentation Steps

Nucleus candidates are obtained by transforming the RGB-based image into HSI. (see Figure 7a). The nucleus has higher saturation than other objects in the image, so it is just enough to perform thresholding on the saturation channel can produce binary nucleus. Then do the filling hole morphological operations, opening and closing on the binary nucleus, to obtain binary object cleaner, which is free from small spots of objects that are not the nucleus. Multiplication of binary object with the original image WBC nucleus will generate objects in RGB color (see Figure 7b).



(a) (b)

Figure 7. a. Nucleus Candidate; b. Nucleus Segmentation result

2.4 Feature Extraction

The features used are namely: WBC area, perimeter WBC, WBC roundness, nucleus ratio, mean (WBC), and the standard deviation (WBC).[10][11]. For example, the object number 15 in the image M2 (33) .jpg (see Figure 8), has a cytoplasm and nucleus. Values obtained from the characteristics of the object number 15 in a row are: area of 17 466, 481 perimeters, 0.93639 roundness, 0.74229 ratios, 160.1461 mean and 21.1403 standard deviation

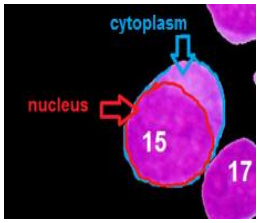


Figure 8. object example (WBC Cell)

Feature extraction performed on eight pieces of the image and obtained 315 objects of white blood cells which will then be used as input in the process of learning and testing.

3. Experimental Results

The input parameters consist of six neurons, namely WBC area, perimeter WBC, WBC roundness, nucleus ratio, mean (WBC), and the standard deviation (WBC), the result consists of two neurons that M2 and M3 which is a subtype of AML. The testing process on the classification would choose a neuron with highest accuracy value as a result. The training process uses a binary sigmoid as activation function and the momentum of its propagation as a classifier. Learning rate (α) : 0.1, 0.5, and 0.9, hidden layer neuron : 6, 8 and 10. Momentum : 0.1, 0.5, and 0.9. Error tolerance limits which

are equal to 0.000001. Determination of these parameters aimed at limiting the number of combinations of parameters used. Training using the k-fold cross validation to avoid overfitting. Fold amount used in this training are three-fold with a random distribution. The results of 27 times of training showed that the best accuracy is obtained from a combination of learning rate 0.1, hidden layers 8 and 0.9 momentum which is equal to 94.285% (see Table 1).

Table 1. Training using K-Fold Cross Validation

Learning rate	Hidden layer	Momentum	Time	Accuracy
0.1	6	0.1	39,472	89,673
0.1	6	0.5	30,89	91,184
0.1	6	0.9	25,134	90,176
0.1	8	0.1	36,855	94,285
0.1	8	0.5	28,644	91,184
0.1	8	0.9	23,894	89,421
0.1	10	0.1	34,119	91,436
0.1	10	0.5	27,526	90,932
0.1	10	0.9	36,004	89,421
0.5	6	0.1	22,831	89,421
0.5	6	0.5	20,222	88,917
0.5	6	0.9	20,551	88,917
0.5	8	0.1	22,928	89,421
0.5	8	0.5	22,435	89,169
0.5	8	0.9	21,485	89,421
0.5	10	0.1	22,481	89,421
0.5	10	0.5	23,395	89,673
0.5	10	0.9	22,592	88,161
0.9	6	0.1	22,024	89,421
0.9	6	0.5	20,699	88,413
0.9	6	0.9	20,424	88,413
0.9	8	0.1	23,315	88,917
0.9	8	0.5	22,678	88,917
0.9	8	0.9	20,403	88,413
0.9	10	0.1	21,267	88,917
0.9	10	0.5	21,162	88,917
0.9	10	0.9	19,854	88,161

Testing the whole cells carried on a combination of parameters with the best accuracy. The results obtained are based on the amount of the corresponding object classification compared to the number of objects that do not fit the classification. Total comparison both define a class type of an image. Table 2 shows the testing based on the number of cells in each image.

Table 2. Classification Result

Num	Image	Correct cells	Wrong cells	Type
1	M2 (10).jpg	89	1	M2
2	M2 (14).jpg	76	0	M2
3	M2 (15).jpg	63	0	M2
4	M2 (33).jpg	53	2	M2
5	M3 (2).jpg	0	5	M2
6	M3 (3).jpg	3	6	M2
7	M3 (6).jpg	10	1	M3
8	M3 (7).jpg	3	3	M3

Table 2 shows that from eight pieces of the images used for the experiment, there are two pieces of the image that are misclassified. Misclassified occurred because the number of cell object belongs to which class M3 relatively less than classes M2 so that the patterns formed better recognize the objects in the class M2. The value of accuracy from the classification per cell is:

$$Cell\ Accuracy = \frac{correct\_cell}{total\_count\_cell} * 100\% = \frac{297}{315} * 100\% = 94{,}285\%$$

While the value of the accuracy of the classification per image is:

$$Image\ Accuracy\ = \frac{correct\_image}{total\_count\_image} * 100\% = \frac{6}{8} * 100\% = 75\%$$

Thus, the test using momentum backpropagation classifier was able to classify subtypes of AML M2 and M3 where per cell classification results show an accuracy of 94.285% and per image classification results have an accuracy of 75%.

4. Conclusion

Watershed Segmentation Distance Transform can separate the white blood cells that overlap. However there is still an error occurred object separation of white blood cells that overlap, due to the minimal local number found during the process of calculating catchment basin on watershed segmentation. Based on experiments conducted on eight pieces of images, obtained 315 pieces of white blood cells which will be used as input in the process of learning and testing. The test using momentum backpropagation classifier was able to classify subtypes of AML M2 and AML M3 where per cell classification results show an accuracy of 94.285% and per image classification results have an accuracy of 75%.

5. Acknowledgment

This research was supported by Grant PNBP UNS 2016 and the image's data AML M2 and AML M3 obtained from Dr Moeward Hospital, Surakarta, Central Java.

6. References

[1].Bell, A &Sallah S 2005 *The Morphology of Human Blood Cells*, Seventh Edition, Abbott, A Promise for Live.

[2].Hamid AG 2011 *Classification of Acute Leukemia, Acute Leukemia - The Scientist's Perspective and Challenge*. Prof. Mariastefania Antica (Ed.), ISBN: 978-953-307-553-2, InTech.

[3] Houwen B 2001 *The Differential Cell Count* Loma Linda, California: Carden Jennings Publishing Co., Ltd.

[4] Supriyadi E dkk 2011 *Immunophenotypic Patterns of Childhood Acute Leukemias in Indonesia* AsianPacific Journal of Cancer Prevention, Vol 12.

[5] Habibzadeh M et al. 2011 *Counting of RBCs and WBCs in noisy normal blood smear* Montréal, Canada; Tehran, Iran: Dept. of Computer Science and Software Engineering, Concordia University, Dept. of Electrical Engineering, Sharif University.

[6] Suryani E, Wiharto, and Polvonov N 2013 *Identifikasi Penyakit Acute Lymphoblastic Leukemia (ALL) Menggunakan Fuzzy Rule Based System Berdasarkan Morfologi Sel Darah Putih* Prosiding Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK), Udinus, Semarang, ISBN:979-26-0266-6.

[7] Suryani E, Salamah U, Wiharto, and Polvonov N 2014 *Identification and Counting White Blood Study of Leukemia.* International Journal of Computer Science & Network Solutions. Juni 2014. Vol. 2 No. 6. ISSN: 2345-3397

[8] Chan, T. F., dan Vese, L. A., 2001. *Active Contour Without Edges*. IEEE Transactions on Image Processing*, 10*(2), 266-277.

[9] Putzu, L., Ruberto C.D., 2013. *White Blood Cells Identification and Counting from Microscopic Blood Image. World Academy of Science, Engineering and Technology*. International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering Vol:7, No:1, 2013.

[10]. Pradana T P N, Suryani E, and Wiharto 2013 *Pemanfaatan Seed Region Growing Segmentation dan Momentum Backpropagation Neural Network Untuk Klasifikasi Jenis Sel Darah Putih* Seminar Nasional Teknoin 2013 FTI UII.

[11] Nasir A A_1, Mashor MY_1, and Hassan R_2 2013 *Classification of Acute Leukaemia Cells using Multilayer Perceptron and Simplified Fuzzy ARTMAP Neural Networks* $_1$Electronic and Biomedical Intelligent Systems Research Group, Universiti Malaysia Perlis, Malaysia. $_2$Department of Haematology, Universiti Sains Malaysia, MalaysiaPatil T.G., Prof. Mr. V. B. Raskar, 2015. Automated Leukemia Detection By Using Contour Signature Method. E & TC Department JSPM's, Imperial College of Engineering & Research Wagholi, Pune. The International Arab Journal of Information Technology, Vol. 10, No. 4, July 2013.