

## White Blood Cells and Lymphoblast Cells Segmentation using Moving K-Means

Ika Candradewi<sup>\*1</sup>, Reno Ghaffur Bagasjvara<sup>2</sup>

<sup>1</sup>Electronics and Instrumentation Research Lab, FMIPA UGM, Yogyakarta, Indonesia

<sup>2</sup>PT. EPSON INDONESIA, Bekasi, Jawa Barat, Indonesia

e-mail: <sup>\*1</sup>ika.candradewi@ugm.ac.id <sup>2</sup>reno.ghaffur.b@gmail.com

### Abstract

Salah satu prosedur diagnosis untuk leukemia limfoblastik akut adalah pemeriksaan sel darah oleh seorang ahli menggunakan mikroskop. Proses ini relatif lama dan akan memperlambat proses penyembuhan penyakit yang membutuhkan tindakan yang cepat. Cara lain untuk melakukan screening terhadap penyakit ini adalah dengan menggunakan pemrosesan gambar digital mikroskopik apusan darah untuk mendeteksi sel limfoblast dan jenis sel darah putih. Salah satu langkah penting dalam pemrosesan citra digital darah adalah segmentasi, karena proses ini mempengaruhi proses selanjutnya dalam mendeteksi dan mengklasifikasikan penyakit Acute Lymphoblastic Leukimia. Oleh karena itu, penelitian ini mengajukan metode segmentasi sel darah putih menggunakan algoritma moving k-means. Beberapa proses dilakukan untuk mengurangi gangguan pendeteksian seperti sel darah merah, sel darah putih yang tampak bertumpuk. Postprocessing dilakukan untuk meningkatkan kualitas segmentasi dan memisahkan sel darah putih yang terhubung. Dataset dalam penelitian ini telah divalidasi oleh ahli patologi klinis dari Rumah Sakit Sardjito, Yogyakarta Indonesia. Penelitian ini menghasilkan performa sistem dengan sensitivitas 85,6%, presisi 82,3%, Fscore 83,9% dan akurasi 72,3%. Berdasarkan hasil proses pengujian dengan jumlah dataset yang jauh lebih besar pada sisi tingkat kesulitan segmentasi sel yang beragam baik dalam hal pencahayaan dan sel yang tumpang tindih, metode yang diusulkan dalam penelitian mampu mensegmentasi sel yang tumpang tindih lebih baik.

**Kata kunci** — Acute Lymphoblastic Leukemia, Segmentasi Sel Darah Putih, Moving K-Means, Transformasi Watersheed

### Abstract

One of the diagnosis procedures for acute lymphoblastic leukemia is screening for blood cells by expert operator using microscope. This process is relatively long and will slow healing process of this disease which need fast treatment. Another way to screen this disease is by using digital image processing technique in microscopic image of blood smears to detect lymphoblast cells and types of white blood cells. One of essential step in digital image processing is segmentation because this process influences the subsequent process of detecting and classifying Acute Lymphoblastic Leukemia disease. This research performed segmentation of white blood cells using moving k-means algorithm. Some process are done to remove noise such as red blood cells and reduce detection errors such as white blood cells and lymphoblastic cell that's appear overlap. Postprocessing are performed to improve segmentation quality and to separate connected white blood cell. The dataset in this study validated with expert clinical pathologists from Sardjito Regional General Hospital, Yogyakarta, Indonesia. This research produces systems performance with results in sensitivity of 85.6%, precision 82.3%, Fscore of 83.9% and accuracy of 72.3%. Based on the results of the testing process with a much larger number of datasets on the side of the variations level of cell segmentation difficulties both regarding illumination and overlapping cell, the method proposed in this study was able to detect or segment overlapping white blood cells better.

**Keywords**— Acute Lymphoblastic Leukemia, White Blood Cell Segmentation. Moving K-Means, Watershed Transformation

## 1. INTRODUCTION

Acute lymphoblastic leukemia is a type of leukemia. The diagnosis of acute lymphoblastic leukemia is based on an increase in the number of immature cells (limphoblast) [1]. This disease will be very fatal if it is not treated immediately because it spreads rapidly to the bloodstream and vital organs. Observation of blood samples by expert operators is the method used to detect the presence of lymphoblast cells [2]. Analysis in this way tends to be slow and has results that are not based on a standard because the results of this analysis are very dependent on the capabilities and conditions of the operator [3]. The slow process of diagnosis of acute lymphoblastic leukemia disease would be very disruptive in the healing of the ALL disease patient who needs a diagnosis as soon as possible so that ALL diseases can be tackled immediately.

Along with the development of technology, inspection of blood samples can be done in other ways, namely by using digital image processing techniques. One of the advantages of this method compared to manual inspection is that the results are objective and the process is faster than manual inspection. Some studies previously were using datasets same is research [4] has made the detection of white blood cells using methods Hough Circle produces an accuracy of 78 125% and research [5] has made the detection of white blood cells using methods HOG-SVM produces accuracy value of 72.3% . Research [6] used a watershed transformation method based on a marker to separate attached white blood cells. Markers are generated from thresholding against distance transform images. One disadvantage of this method is that the threshold value must be determined first and sometimes the selected threshold value does not always succeed in separating the attached white blood cells. This study uses the area values of white blood cells to determine the threshold value to be used. Based on several studies, segmentation of white blood cells using moving k-means method will be carried out by utilizing the watershed transformation algorithm to separate the attached white blood cells.

## 2. METHODS

### 2.1 *Research Method*

The system design in this study was made using OpenCV based image processing software. The data used in this study has been used in several previous studies, namely [4] and [5] consisting of four datasets, each of which is taken by a different microscope operator so that there is a slight difference in the color produced due to differences in the making of blood smears. The data used were 466 images of blood cells with a total of 763 white blood cells. The segmentation method used is Moving K-Means to segment white blood cells from microscopic images of blood cells. In the postprocessing process the Watershed Transformation [9] method is used based on a marker to separate white blood cells. Markers are generated from thresholding against the results of the distance transform. The process of finding thresholding values is done using the area values of white blood cells in the dataset so that a marker is produced which corresponds to the number of white blood cells attached.

This research is a design of white blood cell segmentation system with moving k-means algorithm and separation of white blood cells using a watershed transformation algorithm. The segmentation system diagram block can be seen in Figure 1.

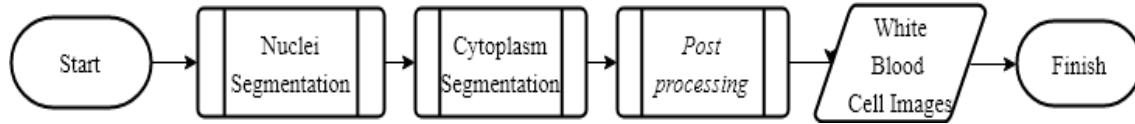


Figure 1. Systems Block Diagram in General form

The contribution of our research is we proposed design stages to segment lymphoblastic and white blood cells from blood smears image. There are three stages in the segmentation system, namely nucleus segmentation, cytoplasmic segmentation, and postprocessing. The process of nucleus and cytoplasmic segmentation is separated because each uses a different color domain. Postprocessing is done to combine the results of the nucleus and cytoplasm, eliminate the noise of red blood cells, and separate the attached white blood cells. The end result of the system is white blood cells that have been segmented from microscopic blood images.

### 2.1.1. Moving K-Means

K-means method is a clustering method which is one of the most popular unsupervised learning algorithms because it is simple. Moving k-means itself is a modification of k-Means with a modified algorithm of k-means [8] and is named moving k-means [7]. This algorithm can avoid dead centers and center redundancy issues. In addition, this algorithm has been proven to reduce the effect of the middle value trapped in the local minima. The algorithm from Moving K-Means is as follows:

1. Initialize  $a_o$  values with values  $a_a = a_b = a_o$  (where  $a_o$  is a constant of small value with value  $0 < a_o < 1/3$  and the value must be inversely proportional to the number of centers)
2. Find the closest center for each pixel and calculate the center position with Equation 1

$$c_j = \frac{1}{n} \sum_{y \in c_j} \sum_{x \in c_j} p(x, y) \quad (1)$$

3. Calculate the suitability of each center value with Equation 2.

$$f(c_j) = \sum_{y \in c_j} \sum_{x \in c_j} (\| p(x, y) - c_j \|^2) \quad (2)$$

4. Find  $c_s$  and  $c_l$  value, which is the center that has the smallest and largest value of  $f(c_j)$ .
  - a. If  $f(c_s) < a_a f(c_l)$ , then if  $p(x, y) < c_l$ , move member  $c_l$  ke  $c_s$  where  $x, y \in c_b$ , and the remaining members are made members  $c_l$
  - b. Recalculate the center position  $c_s$  and  $c_b$ , in accordance with Equation 3 and Equation 4

$$c_s = \frac{1}{n_j} \sum_{y \in c_s} \sum_{x \in c_s} p(x, y) \quad (3)$$

$$f(c_j) = \sum_{y \in c_j} \sum_{x \in c_j} (\| p(x, y) - c_j \|^2) \quad (4)$$

$c_s$  Parameter will throw its members in step 4a, value  $n_s$  and  $n_l$  in Equation 3 and Equation 4 is the number of new members  $c_s$  and  $c_l$ , after the member reassignment process in step 4a.

5. Update  $a_a$  values, where  $a_a = a_a - a_a / n_c$  and repeat steps 3 and 4 until  $f(c_s) \geq a_a f(c_l)$
6. Re-assign all pixels to the nearest center then recalculate the center position with Equation 1.
7. Update  $a_a$  dan  $a_b$  values, where  $a_a = a_o$  and  $a_b = a_b - a_b / n_c$ , and repeat steps 2 to 7 until  $f(c_s) \geq a_b f(c_l)$ .

### 2.1.2. Nucleus Segmentation Design

Nucleus segmentation is performed with saturation or S values from the HSV color domain using a moving k-means algorithm with  $K = 4$  to separate blood images into four parts, namely red blood cells, white blood cells, blood chips and background. Based on the research [7] the saturation value can differentiate the nucleus from other objects in the blood image well because the nucleus has the highest saturation value so that the cluster resulting from the moving k-means chosen is the cluster that has the highest saturation average value. The result of the nucleus segmentation process is the mask of the nucleus in the form of a binary image with a value of 255 is the nucleus and 0 is the background. The nuclear mask is labeled nucleus with values other than 0 and 255 to facilitate the stage of cytoplasmic segmentation. The flow chart of the nucleus segmentation stage of white blood cells can be seen in Figure 2.

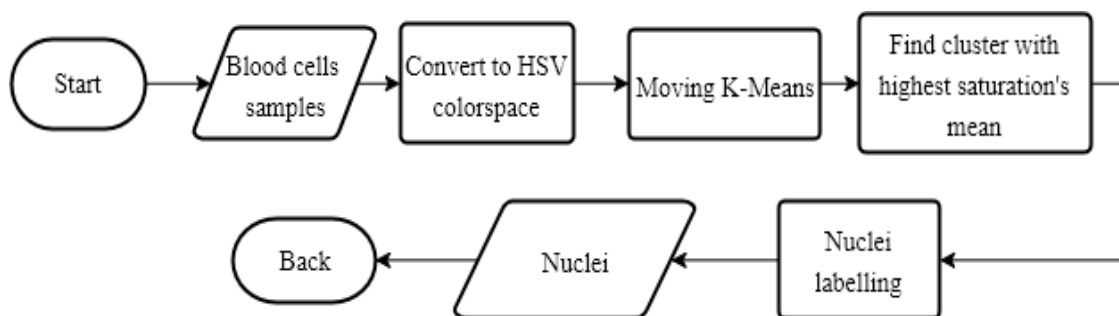


Figure 2. Flowchart of Nucleus Segmentation

### 2.1.3. Design of Cytoplasmic Segmentation

The first step of cytoplasmic segmentation is an image subtraction between microscopic images of blood cells with images resulting from nucleus segmentation resulting in images of blood cells without nucleus. This image is converted to the realm of the color  $L^*a^*b^*$  then the cytoplasm segmented with the moving k-means algorithm utilizes the color values of channels a and b with a value of  $K = 3$  to divide it into three parts namely red blood cells, cytoplasm and background. Through experiments it was found that the cytoplasm has a higher average channel b value compared to other objects on blood images so that the clusters resulting from moving k-means that are further processed are clusters with the highest average b channel values.

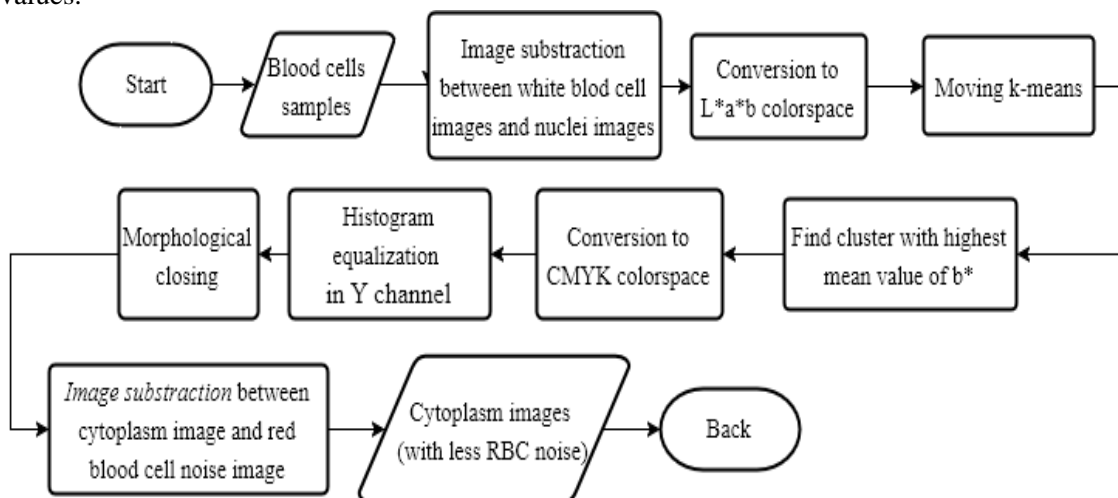


Figure 3. Flowchart of Cytoplasm Segmentation

Cytoplasmic segmentation results still have noise in the form of red blood cells. This noise is segmented because the color of the cytoplasm with red blood cells is not much different. The first step to eliminate red blood cells is the conversion of color images from cytoplasmic segmentation to CMYK color, then histogram equalization surgery is performed on channel Y. The closing morphology operation is then performed to improve image quality and obtain red blood cell image. Image subtraction operation is performed on cytoplasmic images that still have red blood cell noise with red blood cell images and obtained results in the form of cytoplasmic images of white blood cells with fewer red blood cell noise. These fewer red blood cells will be removed at the post-processing stage by utilizing the labeled nucleus at the stage of nucleus segmentation. The flow chart of the cytoplasmic segmentation stage can be seen in Figure 3.

#### 2.1.4. Post-processing design

The postprocessing process is carried out to combine the nucleus image and cytoplasmic image, eliminate the noise of red blood cells, and separate the attached white blood cells. The flow chart from the post-processing stage can be seen in Figure 4.

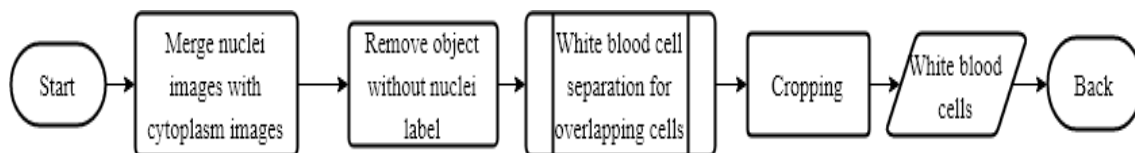


Figure 4. Flowchart of Postprocessing

The first stage in postprocessing is the incorporation of the image of the nucleus and cytoplasmic image into one image of white blood cells. After that the red blood cell noise is eliminated by removing the object that has no labeled nucleus that has been prepared at the beginning of the nucleus segmentation stage.

The stage of separation of white blood cells attached is done so that each white blood cell can be segmented without sticking with other cells. To determine whether a sticking white blood cell object is used a threshold area value based on the maximum value of the white blood cell area in the dataset. White blood cells that have an area greater than the threshold value of the area will be considered as white blood cells attached. Separation of white blood cells is only applied to the attached white blood cell object to speed up the computational process. The flow chart of the stage of separation of white blood cells can be seen in Figure 5.

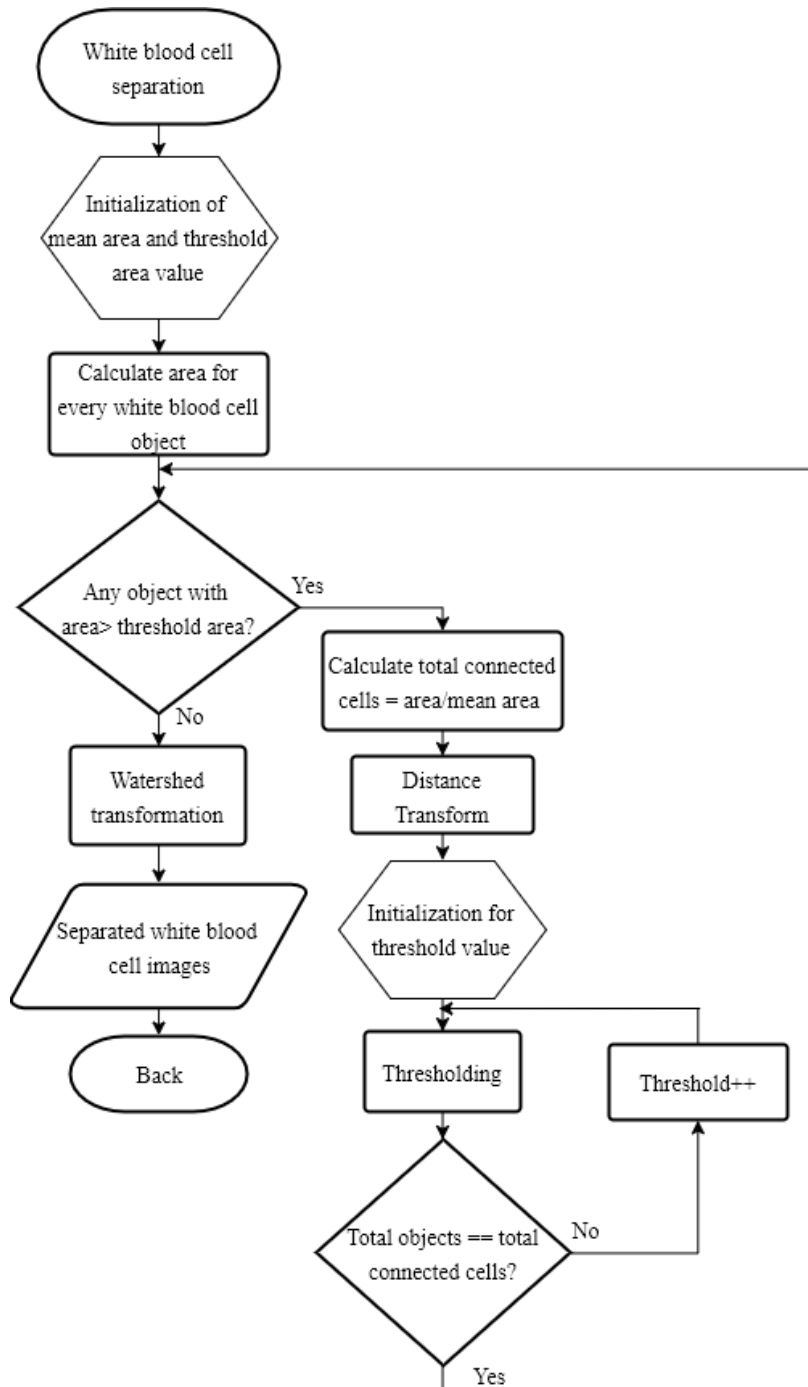


Figure 5. Flowchart of White Blood Cell Separation

White blood cell separation is performed with a marker-based watershed transformation algorithm [10]. Markers are used to limit objects that are processed by the watershed transformation to prevent excessive segmentation. The marker is generated through the thresholding process against the distance transform result of the image. The threshold value used is usually determined manually. In this study the determination of the threshold value is done by utilizing the average value of the area of the white blood cell object so that the marker of the object obtained has the same amount as the object of the attached white blood cell. The number of attached cells is calculated by dividing the value of the area of the white blood cell object with the average value of the area of the white blood cell then the result is rounded. The

thresholding process will be carried out with an initial threshold value. If the number of objects produced is not the same as the number of cells attached, then the threshold value is added and the thresholding process is carried out again. The end result of the thresholding process is a marker used for the watershed transformation algorithm. After that the watershed transformation algorithm is executed using a marker that will separate the attached white blood cells. The last stage of post-processing is cropping of each white blood cell object.

### 3. RESULT AND DISCUSSION

This section will discuss the results of testing the white blood cell segmentation. Segmentation testing is carried out to determine the system's ability to segment white blood cells and Lymphoblastic cells through the value of sensitivity, precision and accuracy produced.

#### 3.1. Segmentation Testing

White blood cell segmentation is done to get white blood cells from microscopic blood images. Segmentation testing was carried out on four blood image datasets. Each dataset has different colors caused by differences in blood sampling techniques carried out by experts from the hospital. Examples of images from each dataset can be seen in Figure 6. White blood cell segmentation is done by moving k-means clustering algorithm based on the color values of each image. White blood cell separation is done using a watershed transformation algorithm with a distance transform.

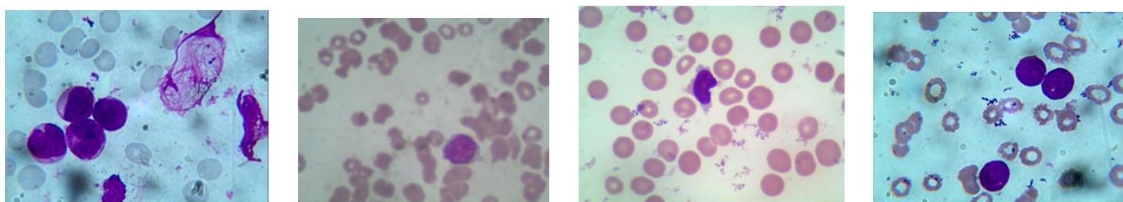


Figure 6. Blood Cells Images from Every Dataset

Thresholding is done to get the marker to be used on the watershed transformation. Area constant values and area boundary constant values are used to find the appropriate threshold value to separate attached white blood cells. Area constant values obtained from the calculation of the average value of the area of all white blood cells in the dataset and obtained a value of 18000 pixels. Area boundary constant values obtained from the maximum value of the area of all white blood cells in the dataset and obtained a value of 30000 pixels. The size of the image in the dataset used is 1280x960 pixels. There were 49 cases of white blood cells attached from all existing datasets where 22 white blood cells from 12 instances of white blood cells that were attached failed to be properly separated by the system so that the accuracy of white blood cell separation performed by the system was 75.5%.

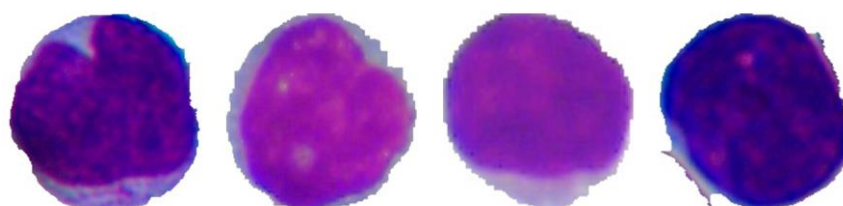


Figure 7. Some Good Segmentation Results



Figure 8. Some Bad Segmentation Results

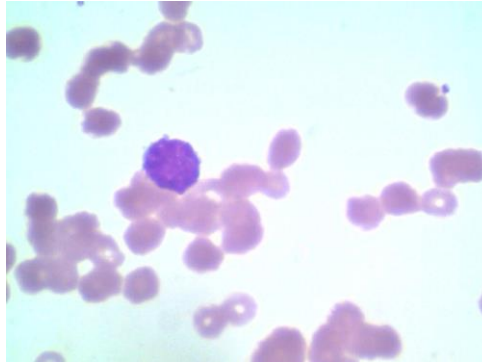


Figure 9. Sample which is failed to be segmented

The process of segmentation of white blood cells results in several failures such as the presence of some non-segmented white blood cells or poor quality segmentation results. The results of segmentation of white blood cells are said to be bad when the white blood cells attach to other objects such as red blood cells or other white blood cells and when the results of segmentation of white blood cells are not intact. Examples of good white blood cell segmentation results can be seen in Figure 7. Examples of poor segmentation results can be seen in Figure 8 and Figure 9, in the form of white blood cells that are still attached, other objects are segmented or cytoplasm fails to be segmented.

Segmentation testing produces true positive or total segmented white blood cells as much as 653 cells, with 110 cell failure and segmentation failure and segmentation errors in other objects or 140 positive cells. The resulting accuracy values are 72.3%, sensitivity 85.6%, precision 82.3%, and the fcore value is 83.9%. When looking at the system's ability to segment positive samples, namely white blood cells, the value of sensitivity, precision obtained from segmentation testing is above 80%. But the resulting accuracy is still below 75% that is equal to 72.3. The system is still unable to distinguish objects other than white blood cells so that there are still 140 objects that are not desired but also segmented. The amount of white blood cells is 763 there are 653 white blood cells that can be segmented and segmentation failure occurs in as many as 110 white blood cells. Its occur because there are different pressure method while blood smear made, so the color of blood smear images are not same each other. The differences of the images made segmentation difficult especially to distighuish white blood cells and lymphoblast from other object with same color. Table 1 perform confusion matrix tables from this research. Several studies that have been done previously using the same dataset were carried out in research [4] and [5]. In the study [4] and [5] the number of datasets was less than this study. Comparison of the value of accuracy performance with this study can be seen in Table 2. Based on the results of the testing process with same dataset but a much larger number of datasets on the side of the variations level of cell segmentation difficulties both in terms of illumination and overlapping cell, the method proposed in this study was able to detect or segment overlapping white blood cells better.

Table 1. Confusion Matrix of Testing

	Predicted
--	-----------



		Positive (WBC and Lymphoblastic cells)	Negative
Real	Positive (WBC and Lymphoblastic cells)	653	110
	Negative	140	0

The values of accuracy, sensitivity, precision and fscore are generated from the following calculations:

$$\%Accuracy = \frac{653 + 0}{653 + 0 + 140 + 110} \times 100\% = 72.3 \%$$

$$\%Sensitivity = \frac{653}{653 + 110} \times 100\% = 85.6 \%$$

$$\%Precision = \frac{653}{653 + 140} \times 100\% = 82.3 \%$$

$$\%Fscore = \frac{2 \times 653}{2 \times 653 + 1 \times 110 + 140} \times 100\% = 83.9 \%$$

Table 2. The Comparison of segmentation performance with previous research

Researcher	Method	Accuracy	Amount of image white blood cell data
[4]	<i>Hough Circle</i>	78.125 %	100
[5]	HOG-SVM	72.3 %	302
Proposed Method	<i>Moving k-means</i>	72.3 %	763

#### 4. CONCLUSION

Based on the results of research on the classification of white blood cells and lymphoblast cells it can be concluded that the system successfully segmented white blood cells with an accuracy of 72.3%, sensitivity 85.6%, precision 82.3%. The separation of white blood cells was successfully carried out with an accuracy of 75.5%. The resulting system still has a weakness in separating the color of the cytoplasm with red blood cells so that there is still some noise of red blood cells that are segmented. It because of different pressure method while blood smear made, so the color of blood smear images are not same each other. Some occlusion object can not be separated because they have same color and the edge of object not found so watershed transformation can not handled. The future works that can be applied to further research is the use of other methods or colors that can separate red blood cells from the image of white blood cells to improve segmentation performance. Also, another white blood cell separation algorithm can be applied so that it can separate white blood cells with better accuracy.

#### REFERENCES

- [1] M. M. Amin, S. Kermani, A. Talebi, and M. G. Oghli, "Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using K - Means Clustering and Support Vector Machine Classifier," *J. Med. Signals Sens.*, vol. 5, no. 1, pp. 49–58, Jan-Mar 2015 [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4335145/?report=printable>. [Accessed: 3-Sep-2017]
- [2] D. Goutam, "Classification of acute myelogenous leukemia in Blood Microscopic Images using Supervised Classifier," in *2015, Int. Conf. on Engineering and Technology (ICETECH)*. pp. 1-5, 2015 [online]. Available: <https://ieeexplore.ieee.org/document/727>

5021. [Accessed: 31-Agust-2017]
- [3] F. Scotti, "Automatic Morphological Analysis for Acute Leukemia Identification in Peripheral Blood Microscope Images," in *2015, IEEE Int. Conf. on Computational Intelligence for Measurement Systems and Applications*, pp. 96–101, 2005 [online]. Available: <https://ieeexplore.ieee.org/document/1522835>. [Accessed: 20-Agust-2017]
- [4] M. N. Khasanah, A. Harjoko, and I. Candradewi, "Klasifikasi Sel Darah Putih Berdasarkan Ciri Warna dan Bentuk dengan Metode K-Nearest Neighbor (K-NN)," *IJEIS (Indonesian. J. Electron. Instrum. Syst.*, vol. 6, no. 2, pp. 151–162, Oct 2016. [online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/15254>. [Accessed: 20-Sep-2017]
- [5] B. Caraka, B. A. A. Sumbodo, and I. Candradewi, "Klasifikasi Sel Darah Putih Menggunakan Metode Support Vector Machine (SVM) Berbasis Pengolahan Citra Digital," *IJEIS (Indonesian. J. Electron. Instrum. Syst.*, vol. 7, no. 1, pp. 25–36, April 2017 [online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/15420>. [Accessed: 20-Sep-2017]
- [6] F. I. Sholeh, I. Candradewi, H. Abdul, A. Jabbar, and D. A. Setiawan, "Segmentasi Sel Blast pada Otomatisasi Sistem Deteksi Leukimia Limpositik Akut ( ALL )" in *2014 Prosiding – Seminar Nasional Ilmu Komputer*, pp. 1-6, 2014.
- [7] N. H. Harun, A. S. A. Nasir, M. Y. Mashor, and R. Hassan, "Unsupervised Segmentation Technique for Acute Leukemia Cells Using Clustering Algorithms," *Int. J. Comput. Electr. Autom. Control Inf. Eng.*, vol. 9, no. 1, pp. 253–259, 2015. [online]. Available: <https://waset.org/publications/10000450/unsupervised-segmentation-technique-foracute-leukemia-cells-using-clustering-algorithms>. [Accessed: 2-Sep-2017]
- [8] M. M. Amin, S. Kermani, A. Talebi, and M. G. Oghli, "Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using K - Means Clustering and Support Vector Machine Classifier," *J. Med. Signals Sens.*, vol. 5, no. 1, pp. 49–58, 2015. [online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4335145/>. [Accessed: 10-Sep-2017]
- [9] C. Raje and J. Rangole, "Detection of Leukemia in Microscopic Images Using Image Processing," in *Int. Conf. on Communication and Signal Processing*, pp. 255-259, 2014 [online]. Available: <https://ieeexplore.ieee.org/document/6949840>. [Accessed: 20-Agust-2017]
- [10] Bagasjvara, R.G. "Klasifikasi Jenis Sel Darah Putih dan Sel Acute Lymphoblastic Leukimia Menggunakan Pengolahan Citra Digital dengan Metode Multilayer Perceptron", *Thesis*, FMIPA Universitas Gadjah Mada, 2017.
- [11] I. Vincent, K. Kwon, S.-H. Lee, and K.-S. Moon, "Acute Lymphoid Leukemia Classification using Two-Step Neural Network Classifier," in *21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, 2015, pp. 1–4. [Online]. Available: <http://ieeexplore.ieee.org/document/7103739/>. [Accessed: 13-September-2017]
- [12] M. Su, C. Cheng, and P. Wang, "A Neural-Network-Based Approach to White Blood Cell Classification," *Sci. World J.*, vol. 2014, no. 1, pp. 1–9, 2014. [Online]. Available: <https://www.hindawi.com/journals/tswj/2014/796371/>. [Accessed: 15-September-2017]