Identification Of Leukemia Diseases Based On Microscopic Human Blood Cells Using Image Processing

Riyanto Sigit
Departement of Informatics and Computer
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
riyanto@pens.ac.id

Mochamad Mobed Bachtiar
Departement of Informatics and Computer
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
mobed@pens.ac.id

Moh. Irsyadul Fikri
Departement of Informatics and Computer
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
irsyaduljeozed@gmail.com

Abstract—White blood cell cancer or what is often referred to as leukemia is a very dangerous disease. Until now the process of recognizing white blood cell cancer or leukemia is still done conventionally, which method can cause a diagnosis difference if done by different medical personnel. To answer these problems, a desktop-based application will be built. This application is made to help the process of identification and classification of types of leukemia using the technique of processing microscopic images of human blood cells. By doing several approaches, such as pre-processing, using the median filtering method, conversion color (RGB to HSV) to clarify the blood image and tresholding to get the image pattern of blood cells that have. Segmentation will be carried out to separate between objects that will be taken with unnecessary objects. This system will use extracted form features from each cell in microscopic blood images. This feature retrieval will be used as a classifier input which is divided into two classification classes, namely Acute Lympotic Leukemia (ALL) and Acute Myolegenous Leukemia (AML). With this method an accuracy of 80% can be generated for the detection of ALL cells separately (one cell), 100% for AML detection separately (one cell) and 90% for cell detection throughout (many cells).

Keywords — Leukemia, Microscopic, Image Processing, Human Blood Cells.

I. INTRODUCTION

Blood is a liquid that is very important for humans, because it serves as a means of transporting substances such as oxygen, material metabolism, the body's defense from germs, regulating acid-base balance, and has many other uses to support life. Circulating blood is an indication of the condition of the human body in a healthy condition or sickness. One of the blood diseases is white blood disease (granulopoetic system) caused by abnormal growth of white blood cells [1].

The objects captured by the microscope will be processed using image processing. From the experimental results, the original image was obtained from a microscope [1]. The process of diagnosing a disease that is fast and precise is necessary in connection with the accuracy of the data to determine program policies in the prevention and cure of diseases, determine the right and correct treatment, and treatment evaluation.

In addition, the development of technology in pattern recognition, which generally aims to recognize an object by extracting important information contained in an image, can help detect an abnormality in the human body through the image produced by the scanner that is used in this case [2]. By researchers as a supporting tool for diagnosing cancer of white blood cells (leukemia). Abnormalities contained in the blood, can diagnose diseases suffered by humans. Along with technological developments, blood image retrieval with a scanning electron microscope (SEM) can be done [3]. In the process of recognizing this disease, a system that has acquired a blood image will do the preprocessing process and make the image into median blur. By utilizing the RGB image changes to HSV model the characteristics of each image will be adapted [3].

This project is a combination of two previous studies which will identify two types of leukemia using a method that has been used to identify one type of leukemia with an accuracy rate of 93%. The resulting blood image can be used for disease recognition through an image processing process using certain method to detect a blood disease. In this study will be used method of recognition of blood disease through blood images as input for image processing, and the results of image processing will be analyzed using the if else branching method. With the results of this research, it is expected to be able to support healthcare kiosks where these tools are used to check human health conditions [4].

II. METHODOLOGY

This section represents the design and manufacture of systems used for leukemia identification systems based on human blood cells. Processing in the identification of leukemia is explained in the block diagram of Figure 1.
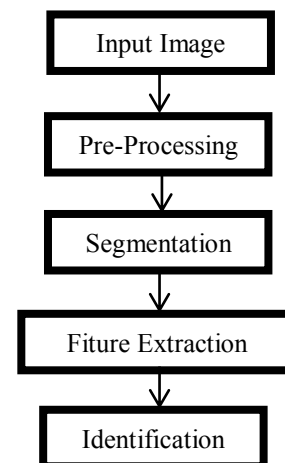

Figure 1. Leukimia identification block diagram

Input image is original image was obtained from a microscope. Preprossesing is used to improve image quality, segmentation is used to separate image objects taken with the background. Extraction feature is used to find leukemia features. After getting the features, we can detect leukemia.

A.   Pre-Processing

To improve the quality of the original image and facilitate the next processing step, preprocessing needs to be done. In this study the preprocessing consists of several steps illustrated in Figure. 2.
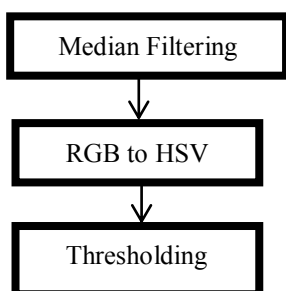

Figure 2. Preprocessing to improve image

Median Filtering

White blood cancer cells will be repaired using the median filtering method to get a better image. In OpenCV there is a separate library for image improvement using the median filtering method. By using the "cvmedianBlur" function and entering the matrix value parameter from the pixel of an image, the system will perform the repair process automatically. In this study I used a pixel matrix value of 7 which represents the value of the pixel matrix that will be searched for by its middle value. The average filter pixel value in the results image is obtained by using equation (1).

$$F(x,y) = \frac{1}{mn}\sum_{x-1}^{m}\sum_{y-1}^{n}G(x,y)$$ (1)

Where:
F (x, y) = result pixel filter value.
G (x, y) = input image pixel
m, n   = average matrix size

To see the median filter results can be seen in Figure 3. It can be seen that images of blood cells after the median filter process appear smoother.
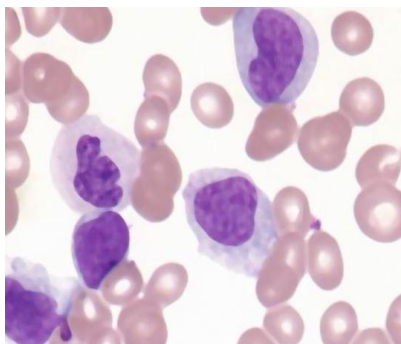

Figure 3. Median filter results

RGB to HSV

After doing the first processing, then the output of the process will be converted from RGB to HSV. HSV defines colors in terms of Hue, Saturation and Value. The advantage of HSV is that there are colors that are the same as those captured by the human senses. While the colors formed by other models such as RGB are the result of a mixture of primary colors. To do the conversion process from RGB to HSV images can be done with the following equation:

$$r = \frac{R}{(R+G+B)}, \ g = \frac{G}{(R+G+B)}, \ b = \frac{B}{(R+G+B)}$$ (2)

$$V = \max(r,g,b)$$ (3)

$$S = 0 \ if \ v = 0 \ or \quad S = 1 - \frac{\min(r,g,b)}{v} \ if \ v > 0$$

(4)

$$H = H + 360 \ jika \ H < 0$$ (5)

Where :
r,g,b   = red, green, blue
H,S,V   = Hue, Saturation, Value

After the process is done, it will get the HSV image from the RGB image that has been converted, this is done because the image in the HSV color space has the same colors as those captured by the human senses. To see the results of color convertion from RGB to HSV can be seen in Figure 4. It can be seen that the image of blood cells after the RGB to HSV color conversion process will change significantly.
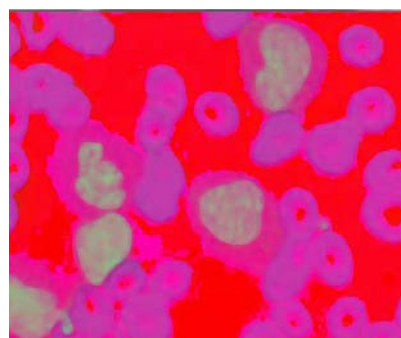

Figure 4. HSV image

Thresholding

Object and background pixels are grouped into two dominant modes. The way to extract objects from the background is to select the (T) threshold value that separates the two modes. Then for any point (x, y) that satisfies f (x, y)> T is called the object point, otherwise it is called the background point. The success of this method depends on how good the histogram partition technique is. If the pixel value in the image is greater than the threshold value of 0, then the pixel value will be replaced by the maxval value,

which is 255 (white). Conversely, if the gray pixel value of the image is smaller than the threshold, the pixel value will be replaced with 0 (black). This threshold value of 0 is used to segment the whole blood smear image. So it is known that the area of blood smear image is used as a segmentation object. In the research that has been done is used a slider to determine the threshold value of a threshold, using the minimum and maximum values by utilizing the "inrange" function in OpenCV. After the pattern is obtained, masking of the initial data will be carried out according to the pattern obtained during the holding process. To see the thresholding results can be seen in Figure 5. It can be seen that the image that previously still had red blood cells after the thresholding process left the remaining disease objects in white.
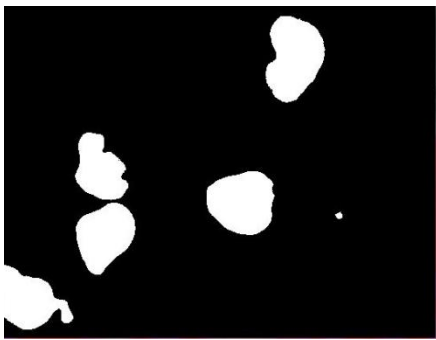

Figure. 5 Thresholding results

B. Segmentation

The segmentation process aims to separate objects in the image. In this paper blood cells will be separated from the backgroud image. For segmentation of microscopic images, the watershed method is performed. Watershed changes the image based on the intensity of the image into a topography consisting of mountains and valleys. The mountain part is a value that has high intensity while the valley part is a value that has a low intensity.

$$T[n] = \{(s,t)|f(x,y) < n\} \qquad (6)$$
$$Cn(Mi) = C(Mi) \cap T[n] \qquad (7)$$
$$C[n] = \cup_{i=1}^{R} Cn(Mi) \qquad (8)$$
$$C[max+1] = \cup C(Mi) \qquad (9)$$

Min and max notations are used to mark the minimum value and the maximum value of g (x, y). Then make T [n] a set of coordinates (s, t) where g (s, t) <n. Then create Cn (Mi) into a collection of coordinates of the point in the filling area that corresponds to the minimum Mi area that is filled in step n. Cn (Mi) can be seen as a binary image. In other words Cn (Mi) = 1 lies in location (x, y) if (x, y) ∈ C (Mi) and (x, y) ∈ T [n], besides that the value of Cn (Mi) = 0 From this it can be said that C [n] is a combination of flows in the catchment basin at stage n. Then C [max + 1] is a combination of all catchment basins. In this system is done by making markers on the cell area and also outside the cell area. In this process

to draw markers in the area, the "Mouse Callback" function is used to indicate the response that occurs when the user operates the mouse in an active working window by utilizing mouse movement. To see the segmentation results can be seen in Figure 6.
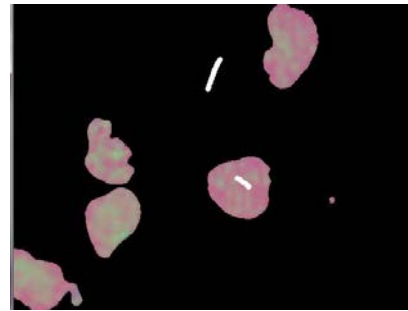

Figure. 6 Segmentation results

C. Feature Extraction

Feature extraction is carried out to obtain data from the form of cells to be identified. This stage is used to distinguish the shape of one object from another using integrals projection . Integral Projection is a method used to find the area or location of an object. This method can be used to detect the boundaries of different image areas, so that we can search the location, shape and features. This method can also be called integral rows and columns of pixels, because this integral adds pixels per line and pixels per column.

$$h(i) = \sum_{i=1}^{Nrows} x(i,j) \qquad (10)$$

$$h(i) = \sum_{j=1}^{Ncols} x(i,j) \qquad (11)$$

Where :
h (i) = integral calculation vertical.
h (j) = integral calculation horizontal.
Nrows = average of rows.
Ncols = average of column.

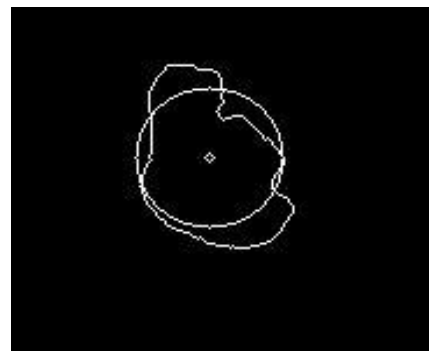To see the feature extraction results can be seen in Figure 7.

Figure 7. Feature extraction results

D. Identification

Feature extraction value is obtained from the results of the integral projection method. Identification is done using

standard deviation. The smallest difference was identified as data that had acute lhympotic leukemia.

An if statement is used to perform branching processes in the program. And in this study if-else branching is used to make decisions based on comparisons with the following formula.

$$\text{ALL Cell} > \text{AML Cell} = \text{ALL}$$
$$\text{ALL Cell} < \text{AML Cell} = \text{AML}$$

Fiture Ekstraction
3.454768->Ini adalah sel AML

Figure. 8 AML Identification

III. EXPERIMENTAL RESULT AND DISCUSSION

In the application of leukemia identification system, the system uses input in the form of images of human blood cells taken through a microscope. Then it is processed and obtained value extraction value using the Projection Integral method. From the test results of 20 test data Acute Lympotic Leukemia (ALL), 20 test data Acute Myolegenous Leukemia (AML), and 20 test data combined, there are 10 data obtained in the acute type of lhympotic disease and 10 in the acute myelogenous disease. Test results can be seen in table 1, 2 and 3.

Table 1. Result (16 data correct)

No.	Output	Medical Result	System Result	Accuracy
1.	3.231483	ALL	ALL	True
2.	2.527768	ALL	ALL	True
3.	3.455140	ALL	ALL	True
4.	3.649082	ALL	ALL	True
5.	2.582028	ALL	ALL	True
6.	2.719108	ALL	AML	False
7.	3.407445	ALL	ALL	True
8.	3.690390	ALL	ALL	True
9.	4.783291	ALL	AML	False
10.	3.420901	ALL	ALL	True
11.	4.173091	ALL	AML	False
12.	3.309829	ALL	ALL	True
13.	3.478282	ALL	ALL	True
14.	3.587272	ALL	ALL	True
15.	5.268892	ALL	AML	False
16.	3.108930	ALL	ALL	True
17.	2.724706	ALL	ALL	True
18.	3.436890	ALL	ALL	True
19.	2.765248	ALL	ALL	True
20.	3.374095	ALL	ALL	True
Avarage				80%

Based on Table 1, the table results from identification, from 20 data obtained as many as 16 data correctly. The success of the average identification using the proposed method is 80% of the test data used.

Table 2. Result (20 data correct)

No.	Output	Medical Result	System Result	Accuracy
1.	4.790278	AML	AML	True
2.	3.908270	AML	AML	True
3.	3.802810	AML	AML	True
4.	5.810801	AML	AML	True
5.	3.920927	AML	AML	True
6.	4.645810	AML	AML	True
7.	5.297010	AML	AML	True
8.	3.923169	AML	AML	True
9.	4.219028	AML	AML	True
10.	4.701020	AML	AML	True
11.	3.901727	AML	AML	True
12.	5.290710	AML	AML	True
13.	3.882561	AML	AML	True
14.	3.903287	AML	AML	True
15.	5.262728	AML	AML	True
16.	4.189891	AML	AML	True
17.	5.101028	AML	AML	True
18.	3.817681	AML	AML	True
19.	4.629810	AML	AML	True
20.	3.927320	AML	AML	True
Avarage				100%

Based on Table 2, the table results from identification, from 20 data obtained as many as 20 data correctly. The success of the average identification using the proposed method is 100% of the test data used.

Table 3. Result (18 data correct)

No.	Output	Medical Result	System Result	Accuracy
1.	2.791383	ALL	ALL	True
2.	3.434466	ALL	ALL	True
3.	2.881334	ALL	ALL	True
4.	3.344678	ALL	ALL	True
5.	2.833156	ALL	ALL	True
6.	4.899104	ALL	AML	False
7.	2.855763	ALL	ALL	True
8.	3.456774	ALL	ALL	True
9.	4.678334	ALL	AML	False
10.	3.245884	ALL	ALL	True
11.	4.791383	AML	AML	True
12.	3.934466	AML	AML	True
13.	3.881334	AML	AML	True

No.	Output	Medical Result	System Result	Accuracy
14.	3.944678	AML	AML	True
15.	5.833156	AML	AML	True
16.	3.899104	AML	AML	True
17.	5.355763	AML	AML	True
18.	5.456774	AML	AML	True
19.	4.678334	AML	AML	True
20.	3.945884	AML	AML	True
Avarage				90%

Based on table 3, a table of results from identification, from 20 data obtained as many as 18 data true. The average success of identification using the proposed method is 90% of the testing data used.

IV. CONCLUSION

From this study, it can be concluded that the identification system for acute lymphotic leukemia has identified disease with an accuracy of 80% and acute myelogenous by 100% with 20 test data. Meanwhile, the success rate for identifying lymphocyte and myelogenous cells found in the same frame as a whole in this study was 90% with 20 test data. The more test data, the greater the success rate of the system.

REFERENCES

[1] Zahra Khandan Khadem Alreza and Alireza Karimian, "Design a new algorithm to count white blood cells for classification Leukemic Blood Image using machine vision system", IEEE, 2016.

[2] Fauziah Kasmin, Anton Satria Prabuwono, Azizi Abdullah, "Detection Of Leukemia In Human Blood Sample Based On Microscopic Images: A Study", JATIT, Vol. 46, No. 2, 2012.

[3] Arthania Retno Praida, "Pengenalan Penyakit Darah Menggunakan Teknik Pengolahan Citra Dan Jaringan Syaraf Tiruan", 2015.

[4] Riyanto Sigit, Zainal Arief, Mochamad Mobed Bachtiar, "Development of Healtcare Kiosk for Cheking Heart Health", EMITTER, 2015.

[5] Lorenzo Putzu, Cecilia Di Ruberto, "White Blood Cells Identication and Classication from Leukemic Blood Image", IWBBIO Proceedings, 2013.

[6] Van-Nhan Tran, Waidah Ismail, Rosline Hassan And Atsuo Yoshitaka, "An Automated Method for the Nuclei and Cytoplasm of Acute Myeloid Leukemia Detection in Blood Smear Images", WAC, 2016.

[7] D. Anggraini1, A.S. Nugroho1, C. Pratama2, I. E. Rozi2, V. Pragesjvara1, and M. Gunawan1, "Automated Status Identification of Microscopic Images Obtained from Malaria Thin Blood Smears using Bayes Decision: A study case in Plasmodium Falciparum", Agency for the Assessment and Application of Technology, 2011.

[8] Stefano Managò(1), Carmen Valente(1), Peppino Mirabelli(2), Michela Napolitano (1), Daniela Corda(1), Anna Chiara De Luca(1), "Identification and classification of acute lymphoblastic leukemia cells from peripheral blood by using Raman spectroscopy", Institute of Protein Biochemistry, National Research Council, Via P. Castellino 111, 80131, Napoli, Italy, 2013.

[9] HUANG Zhanpeng, Zhang Qi, JIANG Shizhong, Chen Guohua, "Medical Image Segmentation Based on the Watersheds and Regions Merging", International Conference on Information Science and Control Engineering, 2016

[10] Istiqamah. I, Yanuar. F, Wibawa. A. D, Sumpeno. S. "Line Hand Feature-based Palm-print Identification System Using Learning Vector Quantization", International Seminar on Application for Technology of Information and Communication, 2016.