Acute Leukemia (ALL and AML) Classification Using Learning Vector Quantization (LVQ.1) With Blood Cell Imagery Extraction

Faisal Asadi, Chiung-An Chen, Ting-Wei Liu, and Fadhilah Syafria

***Abstract*—The biggest cancer disease invading children based on the health ministry 2015 is blood cancer or leukemia. One of the type leukemias is acute leukemia which consists of acute lymphoblastic leukemia (ALL) and acute myelogenous leukemia (AML). Acute leukemia can be diagnosed according to the calculation of a complete blood in the bone marrow, but the calculation process still has several problems, such as when leukemia blood cells are manually counted by microscope, it needs more power, takes too much time, and costs very expensive. This disease can be identified and classified by combining neural network and imaging processing techniques. Learning Vector Quantization (LVQ.1) is used as the neural network approach by extracting leukemia cells of ALL and AML. The image extraction used in this study is to use the color extraction of Hue saturation value color space and the texture feature of Gray level co-occurrence matrix. The experimental results show that the highest accuracy achieved by the proposed algorithm in identifying ALL is about 93.33% trained with 80% training data and tested with 20% testing data. On average, the proposed work yields about 70.31% accuracy to identify both blood cell types. In this sense, the proposed algorithm can classify ALL and AML well.**

***Index Terms*—Acute leukemia classification, learning vector quantization, blood cell imaging extraction, digital image processing.**

I. Introduction

Leukemia is a blood cell cancer, specifically, it is a cancer of the white blood cells. This cancer is recognized as the number of white blood cells grow abnormally in the bone marrow. Therefore, it is also sometimes called as a bone marrow cancer. The bone marrow is the flexible tissue in the interior of human bones which produces three types of blood cells, such as white blood cells (leukocyte), red blood cells (erythrocyte), and platelets (thrombocyte). For the human body, leucocyte is known as the cells of the immune system involved in protecting the body against both infectious disease and foreign invaders. Erythrocyte is used for delivering oxygen to the body tissues via blood flow through the circulatory system. Thrombocyte is to stop bleeding by clumping and clotting blood vessel injuries.

Leukemia is often described as being either acute (fast-growing) and chronic (slow growing). Based on its cell morphology, leukemia can differed as Acute Leukemia Lymphoblastic (ALL) and Acute Myeloid Leukemia (AML). Different types of leukemia have different treatment options and outlooks. Generally, it will need a long time to conduct leukemia diagnoses so-called gold standard procedures. It starts by checking the bone marrow aspiration and capturing the chest radiography. In addition, cerebrospinal liquid and other medical check-ups might also necessary. This gold standard procedure can diagnose about 90% case of leukemias. The remaining cases will need additional procedures such as cytochemical, immunology, cytogenetic, and molecular biology [1], [2]. In Indonesia, the Ministry of Health reports that leukemia is the biggest number of blood cancer in Indonesia among other blood cancer types and majority this cancer is suffered by children. In 2015, there are about 77% ALL patients in age 2-6-year-old, 11% of patients suffer AML, and the rest of patients have Chronic Myeloid Leukemia [3].

There have been many studies researching to find the best way in diagnosing leukemia. Adnan, et al has been published a work to identify ALL by using blood smear images and a neural classifier. This research employed classification technique of the backpropagation method which used 900 layers and 62 hidden layers. This study claimed about 96% satisfaction in identifying the ALL cases [4]. Subrajeet, et al has published similar study by combining Functional Link Artificial Neural Network (FLANN) and imagery extraction technique. The study aimed to segment Lymphoblast images based on white blood cell imagery [5]. In our proposed work, it is proposed to classify and identify Acute Leukemia. We used both backpropagation neural network algorithm to extract the characteristics of blood cells ALL and AML and digital image processing for the leukemia identification technique. The experimental results show about 86.66% accuracy in classifying the leukemia acute types.

The rest of this paper is organized as follows. Section 2 an presents overview of learning vector quantization and digital image processing used for this proposed work. In Section 3, the proposed work is presented. Section 4 shows the performance evaluation of the proposed work. And finally, Section 5 concludes this work with further research topics.

II. Overview of Learning Vector Quantization and Blood Cells Imagery Extraction for Acute Leukemia Classification

In this section, learning vector quantization neural network, blood cells imagery extraction technique, and acute leukemia

Manuscript received October 18, 2018; revised April 1, 2019. This work was supported in part by the Formosa Plastic Groups and Ming Chi University of Technology – Taiwan. Acute Lymphoblastic Leukemia (ALL) And Acute Myelogenous Leukemia (AML) Classification Using Learning Vector Quantization (LVQ.1) With Blood Cell Imagery Extraction.

The authors are with the Department of Electrical Engineering Ming Chi University of Technology, Taiwan (e-mail: M07128021@o365.mcut.edu.tw, joannechen@mail.mcut.edu.tw, pooh61215@gmail.com, fadhilah.syafria@uin-suska.ac.id).

DOI: 10.7763/IJMO.2019.V9.705

are briefly discussed. To easily read, we separately discussed those parts in this section.

A. Learning Vector Quantization (LVQ)

Learning Vector Quantization (LVQ) is a classification method for pattern recognition with a competitive layer. where each output unit represents a particular class or category. During training the data, the output unit is modified (by changing the weight value through supervised training), to estimate the decision surface of the Bayes classifier theory. It is assumed that a set of patterns for training is provided together with an initial distribution of reference vectors (each vector represents a class). Learning vector quantization (LVQ) neural network architecture is essentially the same as the self-organizing mapping (SOM) architecture, but without a topological structure and each output unit knows the class it represents. Fig. 1 depicts an illustration this learning vector quantization neural network architecture. In this research, we using learning vector quantization with first model (LVQ.1)

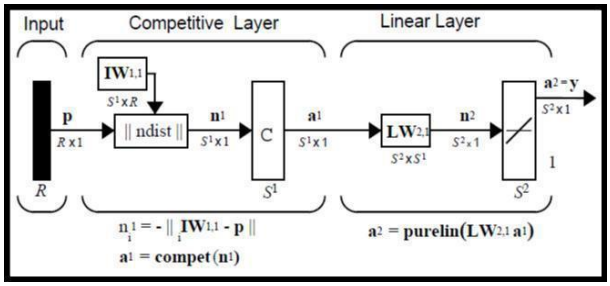

Fig. 1. Learning vector quantization neural network architecture.

B. Hue Saturation Value (HSV)

HSV color space model is known as a color representation which defines color based on the terminology of Hue, Saturation, and Value. Hue is used to distinguishing colors representing redness and greenness of lights and so forth. Saturation denotes a pureness level of color which indicates how much white color is given to a color. Value is an attribute of lights which can be received by human eyes regardless what the color is. Hue Saturation Value color space has been widely used in digital image processing in order to covert Red, Green, and Blue color space (RGB) to HSV. HSV before the conversion to RGB can be expressed as follows:

$$H = \tan\left[\frac{3(G-B)}{(G-B)+(R-B)}\right] \quad (1)$$

$$S = 1 - \frac{\min(R,G,B)}{V} \quad (2)$$

$$V = \frac{R+G+B}{3} \quad (3)$$

To convert RGB color space to HSV color space, the value of red, green, and blue need to be priory normalized which can be defined as:

$$r = \frac{R}{R+G+B} \quad (4)$$

$$g = \frac{G}{R+G+\mathrm{B}} \quad (5)$$

$$g = \frac{G}{R+G+\mathrm{B}} \quad (6)$$

where R, G, B denote red, green, and blue color of a pixel before normalization and *r*, *g*, and b express the normalized value of RGB, respectively. Therefore, the conversion formula RGB color space to HSV color space is formed as:

$$V = \max(r,g,b) \quad (7)$$

$$S = \begin{Bmatrix} 0 - \dfrac{\min(R,G,B)}{V}, \; if \; v = 0 \\ V - \dfrac{\min(R,G,B)}{V}, \; if \; v > 0 \end{Bmatrix} \quad (8)$$

$$\begin{Bmatrix} 0, & if \; S = 0 \\ 60 \times \left[0 + \dfrac{g-b}{S \times V}\right] & if \; V = r \\ & if \; V = g \\ 60 \times \left[4 + \dfrac{r-g}{S \times V}\right] & if \; V = b \\ H + 360 & if \; H < b \end{Bmatrix} \quad (9)$$

The converted RGB values to HSV can be usually used for a color feature extraction in digital image processing for further analysis depend on what application is being developed. For this purpose, mean formula is employed to perform the color feature extraction by using HSV color space. The mean equation can be found as follows:

$$\mu = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} pij \quad (10)$$

C. Gray Level Co-occurrence Matrix (GLCM)

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of the pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. GLCM can be computed to show the relationship of two adjacent pixels with a certain intensity, distance, and angle. Distance in pixel-based is usually expressed as 1,2,3, and so forth. Angle orientation defines a degree value which can be 0º, 45º, 90º, or 135º. This value represents so-called the gray-level value transformed into a co-occurrence matrix. This transformation is obtained with a sliding window which can be in size 3×3, 5×5, 7×7, 9×9, and so on. This window can be then to determine a spatial relationship of distance and angle. Fig. 2 illustrates a 3×3 window size where showing the corresponding of distance and angle *(bv)* of this matrix.

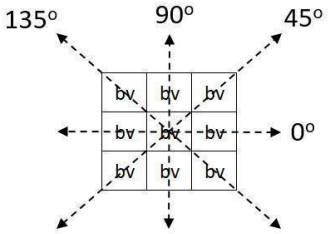

Fig. 2. Corresponding of distance and angle of a 3×3 matrix.

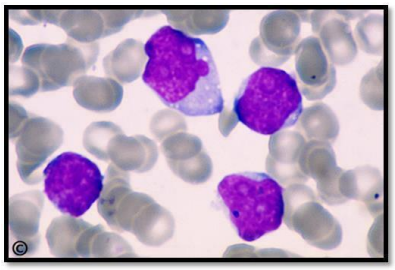

Fig. 3. Blood cells of acute lymphoblastic leukemia.

To get the gray-level co-occurrence matrix, an image with RGB color space needs to be converted first to a grayscale image. This conversion can use the following formula:

$$Y_{srgb} = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (11)$$

The following statistic equations are used in this study to get the texture feature from GLCM:

1) *Angular Second Moment (ASM)*

$$ASM = \sum_i \sum_j \{P(i,j)\}^2 \quad (12)$$

2) *Contrast*

$$CON = \sum_K K^2 \left[\sum_i \sum_j P(i,j) \right]_{[i-j]=k} \quad (13)$$

3) *Correlation*

$$COR = \frac{\sum_i \sum_j (i,j) \times p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (14)$$

4) *Variance*

$$VAR = \sum_i \sum_j (i - \mu_x) \times (j - \mu_y) \times p(i,j) \quad (15)$$

5) *Inverse Different Moment*

$$IDM = \sum_i \sum_J \frac{1}{1 + (i - j^2)} \times p(i,j) \quad (16)$$

6) *Entropy*

$$ENT = -\sum_{i,j} p(i,j) \times \log p(i,j) \quad (17)$$

D. Acute Lymphoblastic Leukemia (ALL)

ALL is one of leukemia diseases which forms normal blood cells into a deadly lymphoblastic. This lymphoblastic will then quickly grow its number of the blood cells in the bone marrow. The following figure shows how the acute lymphoblastic leukemia grows in the human's bone marrow [1].

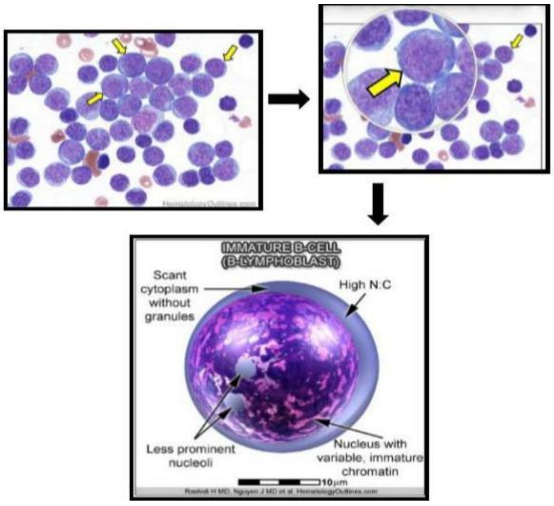

Fig. 4. Details of acute lymphoblastic leukemia (ALL).

E. Acute Myelogenous Leukemia (AML)

Acute Myelogenous Leukemia is a blood cancer which is formed similarly with myeloid of white blood cells. It can be identified based on its breeding as an abnormal blood cells in the bone marrow. However, it is generally formed and grown together with the white blood cells. AML usually strikes adults and is more often found in men bodies than women. Fig. 5 shows the AML in the bone marrow of human [1].

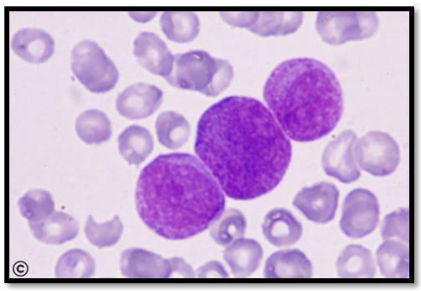

Fig. 5. Blood cells of acute myelogenous leukemia (AML).

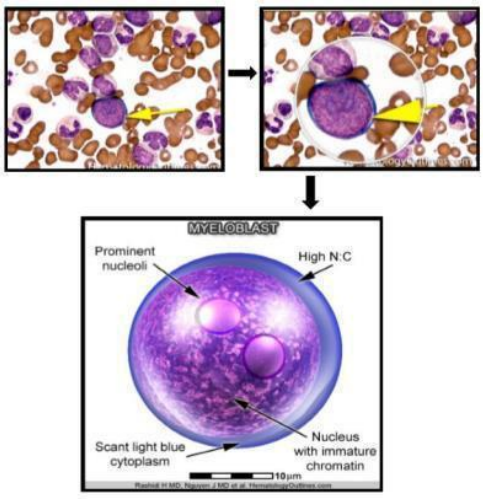

Fig. 6. Details of acute myelogenous leukemia (AML)

III. The Proposed Work of Learning Vector Quantization and Blood Cells Imagery Extraction for Acute Leukemia Classification

In this section, our implementation of learning vector quantization neural network algorithm and blood cells imagery extraction technique for classification of acute leukemia are discussed as follows.

1) *Data collecting*

We first collected the blood cells image as sample data used for both training and testing data. There are 75 samples of blood cells images captured from human's bone marrow. These images have been scaled by using Zeiss microscopic and Nikon with scale size: ×200, ×400, ×630, ×1000.

2) *Preprocessing*

Preprocessing techniques used in this work are cropping and resizing to all the sample data. This preprocessing output is only focused on one single cell with size 300×300 pixel. However, it should be noted that we do not do any background removal technique into our sample data and the cytoplasm is also kept from the entire images of blood cells.

3) *Digital image analysis*

In this step, we converted the RGB images to HSV color space by using (4) to (9).

4) *Feature extractions*

As mentioned before, there are two kinds of feature extraction used in this study: color feature extraction and texture feature extraction. For color extraction, HSV color feature was used and for texture feature extraction, we employed GLCM. Both techniques are performed onto all data samples. In the color extraction, we used (10) as further analysis after computing (4) to (9) in the third stage. In the texture extraction, we computed all formulas from (11) to (17) in this stage. Fig. 11 and Fig. 12 are given to show how both feature extractions were performed in this proposed work.

5) *Training and testing of classification*

In this stage, we do the training and testing stages to classify the acute lymphoblastic leukemia and the acute myelogenous leukemia by using learning vector quantization neural network. We used our data with three variations of data. In compositions as 90% for data training and 10% as data testing, 80% for data training and 20% as data testing, 70% for data training. and 30% as data testing.

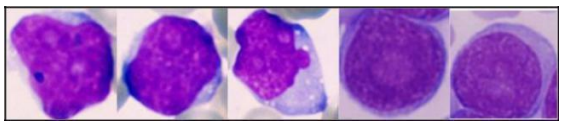

Fig. 7. Data training of all cells.

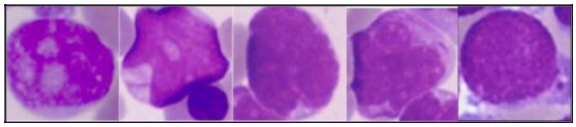

Fig. 8. Data training of AML cells.

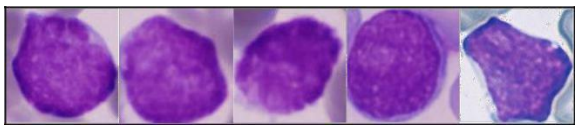

Fig. 9. Data testing of all cells.

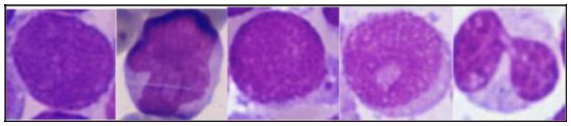

Fig. 10. Data testing of AML cells.

Below the samples of data set for data training and data testing.

TABLE I: Data Set of Acute Lymphoblastic Leukemia (AML) and Acute Myelogenous Leukemia for Data Classification

Data Training	1	2	3	4	5
Mean H	0.376265	0.372441	0.360519	0.364189	0.360726
Mean Saturation	0.245826	0.252516	0.279248	0.287212	0.252121
Mean Value	0.377909	0.375044	0.360233	0.3486	0.387153
Angular Second Moment	0.003479	0.001440	0.002913	0.004018	0.001984
Contrast	11.2179	10.4817	11.3819	8.41391	12.6142
Correlation	0.99802	0.99798	0.998293	0.998479	0.998075
Variance	2858.62	2588.86	3327.58	2761.01	3269.48
Inverse Different Moment	0.51303	0.431573	0.488703	0.529036	0.455292
Entropy	9.94011	10.4581	9.87562	9.62846	10.4082
Image	ALL55	ALL56	AML01	AML02	AML04

The data above are results of color feature extraction and texture feature extraction. The results will be classified with learning vector quantization neural network algorithm. Below are steps of feature color extraction and feature texture extraction.

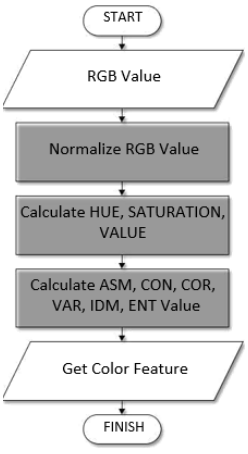

Fig. 11. Flowchart of HSV color feature extraction.

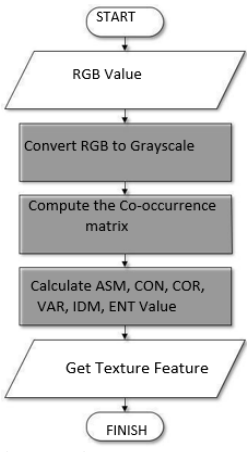

Fig. 12. Flowchart of texture feature extraction by using GLCM.

IV. Implementation and Experiment Results

In this section, we discuss the results of our implementation of learning vector quantization neural network algorithm and blood cells imagery extraction technique for classification of acute leukemia are discussed. Figure 13 and Figure 14 show the user experience of neural network training and application of this implementation, respectively. In this study, we use MATLAB to develop our proposed work. To evaluate our proposed work, we calculate the accuracy (Acc) of our model by using formula defined as:

$$Acc = \frac{\sum Truepositive\ value}{\sum Tatal\ DataTest} \quad (18)$$

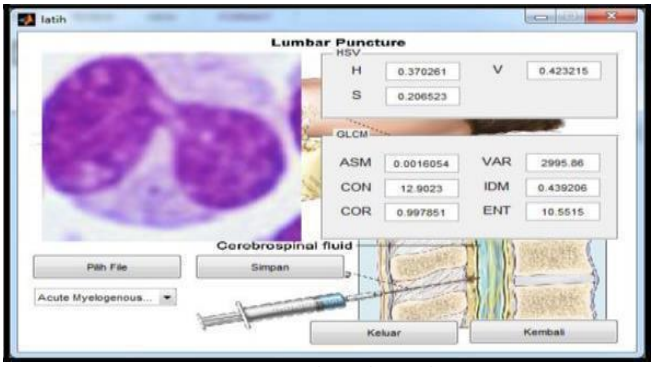

Fig. 13. User interface of classification.

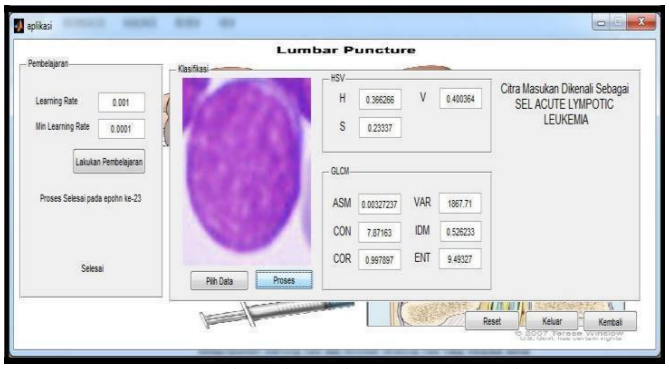

Fig. 14. User interface of classification and result of testing.

After we trained our data samples, we found the best parameter for our model, then we used the parameters for data testing that can be detailed as follows: Learning Rate (LR) is set to 0.0001, and Minimum Learning Rate (MLR) equals 0.00001 with percentage of 80% is data training and 20 % is data testing. From the results of our implementation can be tabulated as in Table II and Table III. Table II shows the classification testing of ALL and Table III presents the classification testing of AML, respectively. According to Table II, there are 14 cases of 15 identified as ALL by using our approaches. In this study, ALL can be detected with a satisfaction score of about 93.33%. However, when tested our algorithm to detect AML cases, we only achieved about 80% or 12 cases are true out of 15 cases from the percentage of data 80% is data training and 20 % is data testing. Consequently, after we testing all of the variants of the percentage of data (90%:10%, 80%:20%, 70%:30%) we can report that on average, our algorithm can argue about 70.31% accuracy in classifying and identifying the acute leukemia cases.

TABLE II: Testing of Acute Lymphoblastic Leukemia (ALL) with Percentage 80% Data Training: 20% Data Testing

No.	File	classification Result	Conclusions	Execution time (second)	Explanation
1	ALL01. JPEG	ALL	detected	0.184773	ALL
2	ALL02. JPEG	ALL	detected	0.109459	ALL
3	ALL03. JPEG	ALL	detected	0.101214	ALL
4	ALL04. JPEG	ALL	detected	0.108363	ALL
5	ALL05. JPEG	AML	not detected	0.103572	AML
6	ALL06. JPEG	ALL	detected	0.102826	ALL
7	ALL07. JPEG	ALL	detected	0.100651	ALL
8	ALL08. JPEG	ALL	detected	0.106951	ALL
9	ALL09. JPEG	ALL	detected	0.103132	ALL
10	ALL10. JPEG	ALL	detected	0.099071	ALL
11	ALL11. JPEG	ALL	detected	0.097178	ALL
12	ALL12. JPEG	ALL	detected	0.102824	ALL
13	ALL13. JPEG	ALL	detected	0.099826	ALL
14	ALL14. JPEG	ALL	detected	0.102028	ALL
15	ALL15. JPEG	ALL	detected	0.105162	ALL
Conclusions of Data Testing			Detected	14	
			Not Detected	1	
Average of Execution time (Second)				0,1036812 second	

TABLE III: Testing of Acute Lymphoblastic Leukemia (AML) with Percentage 80% Data Training: 20% Data Testing

No.	File	classification Result	Conclusions	Execution time (second)	Explanation
1	AML01. JPEG	AML	detected	0.095459	AML
2	AML 02. JPEG	AML	detected	0.098941	AML
3	AML 03. JPEG	ALL	not detected	0.099563	ALL
4	AML 04. JPEG	AML	detected	0.099469	AML
5	AML 05. JPEG	AML	detected	0.099782	AML
6	AML 06. JPEG	AML	detected	0.102393	AML
7	AML 07. JPEG	AML	detected	0.103215	AML
8	AML 08. JPEG	AML	detected	0.097433	AML
9	AML 09. JPEG	ALL	not detected	0.097300	ALL
10	AML 10. JPEG	ALL	not detected	0.098575	ALL
11	AML 11. JPEG	ALL	not detected	0.101379	ALL
12	AML 12. JPEG	AML	detected	0.098297	AML
13	AML 13. JPEG	AML	detected	0.110073	AML
14	AML 14. JPEG	AML	detected	0.106016	AML
15	AML 15. JPEG	ALL	not detected	0.212099	ALL
Conclusions of Data Testing			Detected	10	
			Not Detected	5	
Average of Execution time (Second)				0,108586571 second	

V. Conclusion

In this paper, an implementation of learning vector quantization neural network and blood cells extraction imagery to classify acute leukemia is proposed. We combine these two approaches and test our algorithm by using 150 data samples. In our implementation, we used learning vector

quantization neural network algorithm (LVQ.1) for classifying ALL and AML. To extract the characteristics of blood cells (ALL and AML) and digital image processing for the leukemia identification technique. The experimental results show that our algorithm can identify ALL and AML cases about 93.3% and 66.67% with the percentage of data is 80% data training and 20% data testing, or we can achieve about 80%. Finally, after we testing all of the variants of the percentage of data (90%:10%, 80%:20%, 70%:30%) we can report that on average, our algorithm can argue about 70.31% accuracy in classifying and identifying the acute leukemia cases.

REFERENCES

[1] T. Haferlach, W. Kern, S. Schnittger, and C. Schoch, "Modern diagnostics in acute leukemias," *Journal of Critical Reviews in Oncology/Hematology*, vol. 56, no. 2, pp. 223-234, October 2005.

[2] G. J. L. Kaspers and C. M. Zwaan, "Pediatric acute myeloid leukemia: Towards high-quality cure of all patients," *Hematological: Journal of the European Hematology Association*, vol. 92, pp. 15191532, November 2007.

[3] *Journal of the European Hematology Association*, vol. 92, pp. 1519153, 2007

[4] J. I. Kanker, *Pusat Data dan Informasi Kesehatan Kementerian Kesehatan Republik Indonesia*, ISSN 2088-270X Semester 1, 2015.

[5] A. Khasman and H. H. Abbas, "Acute lymphoblastic leukemia identification using blood smear images and a neural classifier," in *Proc. International Work-Conference on Artificial Neural Networks*, pp. 80-87, 2013.

[6] S. Mohapatra, D. Patra, S. Kumar, and S. Satpathy, "Lymphocyte image segmentation using functional link neural architecture for acute leukemia detection," *Biomedical Engineering Letters*, vol. 2, pp. 100-110, 2012.

[7] S. Mohapatra, P. K. Sa, and B. Majhi, "Adaptive threshold selection for impulsive noise detection in images using coefficient of variance," *Neural Computing and Applications*, vol. 21, no. 2, 2018.

Faisal Asadi was born in Padang, Indonesia on27th October, 1994. He received his bachelor of engineering (B.Eng) in Department of Informatics Engineering State Islamic University of Sultan Syarif Kasim Riau-Indonesia in September 2017. Currently He is pursuing master degree in Department of Electrical Engineering Ming Chi University of Technology, Taiwan. His research interests in biomedical and health informatics, digital image processing, Intelligence control, and Artificial intelligence, and Bioinformatics.

Previous experienced was be research assistant in Cardiovascular Engineering Centre – Institute Jantung Negara Malaysia (IJN), Faculty of Bioscience and Medical Engineering University Technology Malaysia (UTM) at Johor Bahru, Malaysia in 2018.

Chiung-An Chen received the Ph.D degree in electrical engineering in National Cheng Kung University, Taiwan. Now she is active be lecturer and teaching in Ming Chi University of Technology, New Taipei,Taiwan. She is also assistant professor in Department of Electrical Engineering Ming Chi University of Technology–Taiwan. Her research interests integrated circuit design, system on a chip, biomedical circuit system, and data compression.