Classification of acute myeloid leukemia subtypes M1, M2 and M3 using active contour without edge segmentation and momentum backpropagation artificial neural network

Agus Harjoko[1], *Tri* Ratnaningsih[1], *Esti* Suryani[2], Wiharto[2], *Sarngadi* Palgunadi[2], and *Nurcahya Pradana Taufik* Prakisya[1*]

[1]Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia
[2]Universitas Sebelas Maret, Surakarta, 57126, Indonesia

Abstract. Acute Myeloid Leukemia (AML) is a type of cancer which attacks white blood cells from myeloid. AML has eight subtypes, namely: M0, M1, M2, M3, M4, M5, M6, and M7. AML subtypes M1, M2 and M3 are affected by the same type of cells, *myeloblast*, making it needs more detailed analysis to distinguish. To overcome these obstacles, this research is applying digital image processing with Active Contour Without Edge (ACWE) and Momentum Backpropagation artificial neural network for AML subtypes M1, M2 and M3 classification based on the type of the cell. Six features required as training parameters from every cell obtained by using feature extraction. The features are: cell area, perimeter, circularity, nucleus ratio, mean and standard deviation. The results show that ACWE can be used for segmenting white blood cells with 83.789% success percentage of 876 total cell objects. The whole AML slides had been identified according to the cell types predicted number through training with momentum backpropagation. Five times testing calibration with the best parameter generated averages value of 84.754% precision, 75.887% sensitivity, 95.090% specificity and 93.569% accuracy.

1 Introduction

Leukemia is a cancer that attacks the white blood cells (WBC) formed in bone marrow. People with leukemia have a very large number of abnormal WBC. The excessive amount of WBC will result in decreasing number of healthy blood cells [13]. Leukemia is divided into four types, namely: Acute Myeloid Leukemia (AML), Acute Lymphocytic Leukemia (ALL), Chronic Myeloid Leukemia (CML) and Chronic Lymphocytic Leukemia (CLL). The AML type is subdivided into eight subtypes: M0, M1, M2, M3, M4, M5, M6, M7 [3][4][8]. There are three cell types whose percentages indicate AML subtypes M1, M2 and M3, namely: *myeloblast*, *promyelocyte*, and *monoblast* [6].

Leukemia can be detected by counting the total number of blood cells then performing a comparison of the number of white blood cells and red blood cells through a microscope by morphology by haematologists. However, the procedure of calculating blood cells with this microscope is quite time consuming and tiring [9]. The microscopic image inspection of bone marrow blood also has variative results and does not have clear standards due to the process relies heavily on the expertise of haematologists [20]. The other techniques used to calculate blood cell counts such as immunophenotyping, flowcytometer and molecular probing are also still used as a standard diagnostic of leukemia hence these techniques are considered relatively costly [17][22]. To help to overcome these problems, we proposed a combination of digital image processing and artificial neural network [18, 23].

This research combined active contour without edge segmentation (ACWE) process for WBC segmentation in blood microscopic images and artificial neural network with backpropagation momentum algorithm based on morphology and colour features of three types of cells mentioned above to identify AML subtypes: M1, M2 and M3.

2 Proposed Method

The research began from image acquisition of blood preparations. The image data were enhanced, segmented and then extracted to get the features value as input data in training and testing stage. The results from cell type classification were used to identify what AML class contained in each preparation. Research steps can be shown in Figure 1.

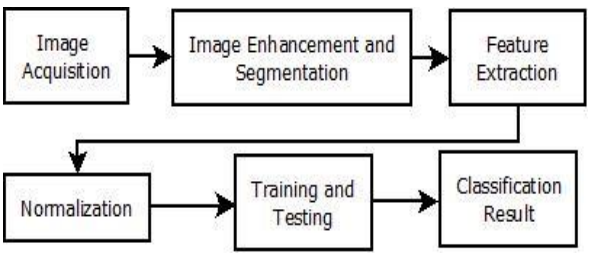


Fig 1. Research steps

* Corresponding author: nurcahya.pradana@gmail.com

2.1 Image Acquisition

AML M1, M2 and M3 preparations were obtained from Dr. Sardjito Hospital Yogyakarta through ethical clearance procedure. Image data were taken using a 21 megapixels camera attached to Olympus microscope's ocular lens with 1000 times magnification. The shooting was done 33 times for each preparation with different slide position.

2.2 Image Enhancement and Segmentation

Image enhancement for reducing image noise was conducted by applying median filter to the original image [7, 19]. Image segmentation consists of two important parts: WBC body segmentation and nucleus segmentation.

The concept of body segmentation is utilizing ACWE as the primary method of segmenting WBC candidates. ACWE minimizes the difference in energy value from background and foreground. The stopping term does not depend on the gradient of the image but related to a particular segmentation of the image [1, 5]. Having obtained the image of white blood cells intact, overlapping cells were separated by using Watershed Distance Transform [19, 24]. The result of the WBC body segmentation was a binary mask which was then pixel-multiplied with the corresponding pixel in original image to get the WBC object in RGB channel. The segmentation flow chart of the WBC body is shown in Figure 2.

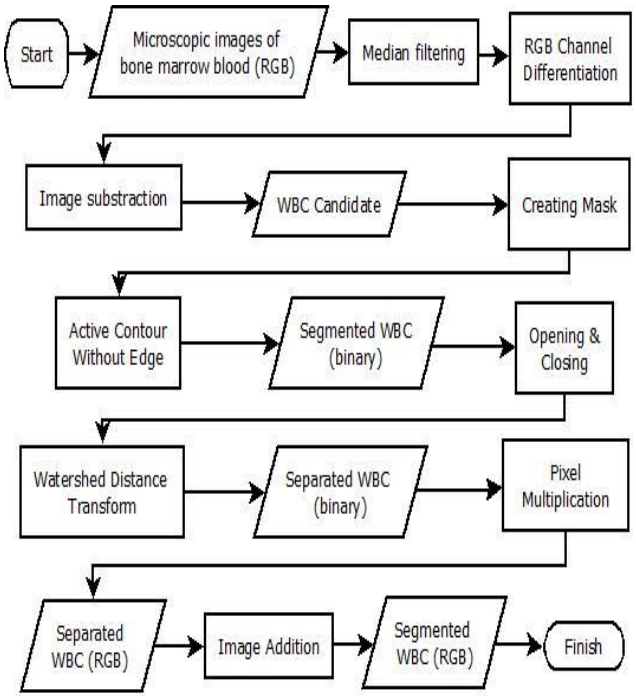


Fig 2. WBC body segmentation flowchart

Nucleus needed to be segmented to calculate nucleus ratio. Nucleus image should be converted to HIS channel because the nucleus had a high saturation making it easier to do thresholding on the saturation channel [2]. The segmentation flowchart of nucleus is shown in Figure 3.

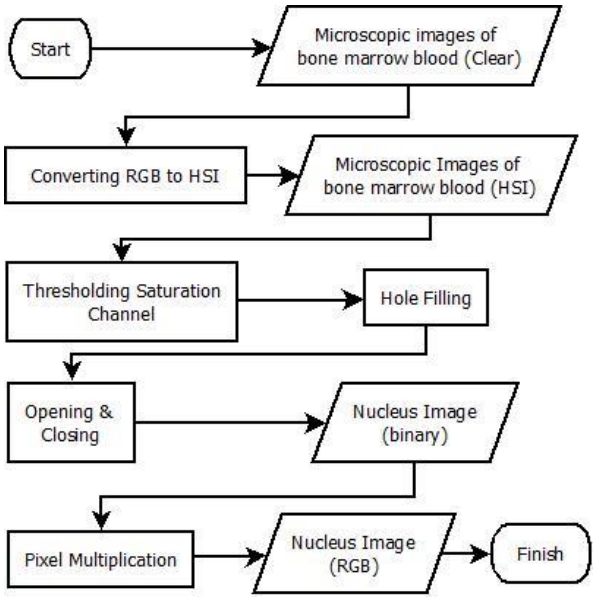


Fig 3. Nucleus segmentation flowchart

2.3 Feature Extraction

The feature extraction produced six numerical data which represent image characteristics: cell area, cell perimeter, cell circularity, nucleus ratio, mean and standard deviation. These six features were then used as artificial neural network inputs.

- Cell Area: The sum of white pixel from binary segmented cell [18].

$$A = \frac{1}{n.m}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}[\, i(x,y) | \, i(x,y) = 1, f(x,y)] \quad (1)$$

The variable *i(x, y)* is the pixel value in the image *f(x, y)*. Variable *A* defines the area. The variable *n* is the number of the length of image row whereas *m* is the length of image column.

- Cell Perimeter: The sum of white pixel from the eroded binary segmented cell by 1-pixel strel [18].

$$p = A - (A \ominus B) \quad (2)$$

Variable *B* is a strel of image morphology. *A* is the area of the object and the *p* is the edge of the object. All variables are in units of number of pixels.

- Circularity: The roundness of cell within range 0 - 1 [18].

$$C = \frac{4\pi A}{p^2} \quad (3)$$

A is the area and *p* is the perimeter of the area. The more rounded an object the closer the value of 1.

- Nucleus Ration: Nucleus area and cytoplasm area ratio within range 0 - 1 [18].

$$R = \frac{A(nucleus)}{A(cell)} \quad (4)$$

Variable *A(nucleus)* defines the area of the nucleus object while *A(cell)* defines the area of the body of a WBC. *R* is the ratio of the nucleus area to the WBC.

- Mean: Pixel colour intensity of red, green and blue channel [16].

$$\bar{x} = \frac{1}{n.m}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} i(x,y) \quad (5)$$

The variable *i(x, y)* is the pixel value of the cell. The variable *n* is the number of the length of image row whereas *m* is the length of the image column. The variable $\bar{x}$ defines the average pixel intensity value of the cell.

- Standard Deviation: The root of the squared difference in pixel colour intensity and pixel mean [16].

$$\sigma = \sqrt{\frac{1}{n.m}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(i(x,y)-\bar{x})^2} \quad (6)$$

The variable *i(x, y)* is the pixel value of the cell. The variable *n* is the number of the length of image row whereas *m* is the length of the image column. The variable $\bar{x}$ defines the average pixel intensity value of the cell and σis the standard deviation of the cell.

2.4 Normalization

Due to varying values of the features, normalization needed to be done first before entering the training process with the following formula:

$$normalization = \frac{data[x]-\min(data)}{\max(data)-\min(data)} \quad (7)$$

The value of x is the value of the x-index, min (data) is the smallest value of the data set on each feature, while max (data) is the largest value of the data set on each feature.

2.5 Training and Testing

AML M1, M2 and M3 classification training stage used backpropagation momentum with one hidden layer as learning algorithm. The architectural design of this training can be explained in Figure 4.

Six normalized features were used as input neurons. The output layer consisted of two neurons called y1 and y2 in binary value. Cell type results were determined by adjusting the class to the output neuron pattern as shown in Table 1.

Learning rate (α) used were 0.1, 0.5, and 0.9. For the number of hidden layer (z) neurons used were 6, 8 and 10. Maximum epoch (ε) were 1000, 5000, and 10000. Momentum (μ) used were 0.1, 0.5, and 0.9. The tolerance limit for error (ξ) was 0.000001 thus the testing was done 81 times according to the combination of parameters.

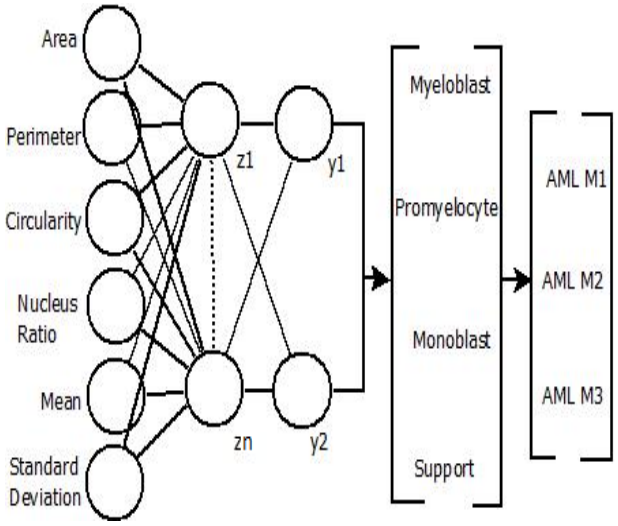


Fig 4. Momentum backpropagation architecture

Table 1. Output class based on output neuron pattern

No.	Cell Type	y1	y2
1	*Myeloblast*	0	0
2	*Promyelocyte*	0	1
3	*Monoblast*	1	0
4	Support	1	1

2.6 Classification Result and Validation

Validation for classification results used 3-fold cross-validation. Out of 734 total data, 250 data were used as fold-1 test data. 248 data were used in fold-2 and 236 data were used as fold-3 test data. Each result was then summed and inserted into a confusion matrix based on the prediction of each cell type. Those values were later used to calculate four predictive analysis values: precision, sensitivity, specificity and accuracy. The combination of parameters having the highest total predictive analysis was then used to recalibrate five times.

The identification of AML subtypes on each preparation was done by accumulating the number of predicted results of each cell and grouping them according to the origin of preparations. AML subtype later could be determined based on the percentage criteria of each cell type as outlined in Table 2. Preparation which was not meeting cell count criteria would be classified as "support" type cell.

Table 2. AML subtype class based on cell type

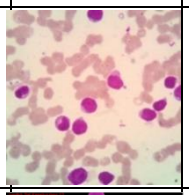
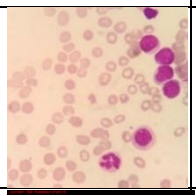
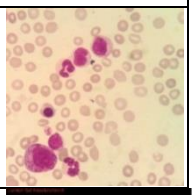
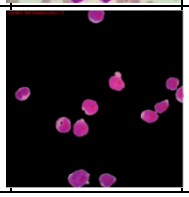
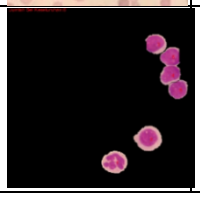
Subtype	*Myeloblast*	*Promyelocyte*	*Monoblast*	Support
AML M1	**>89%**	<10%	<10%	<10%
AML M2	**30-89%**	**>10%**	**<20%**	<10%
AML M3	**<30%**	**>20%**	<10%	<10%
NON-M1M2M3	Does not meet cell count criteria			

3 Research Result

3.1 Segmentation Result

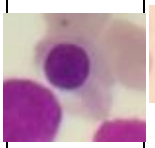
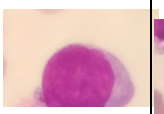
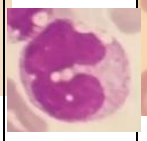
Segmentation was done on each slide preparation containing dozens of blood cells. The number of each cell of each image on each preparation needed to be calculated to ensure the quantity of each cell type of a preparation. The samples of the segmentation results on AML M1, M2 and M3 preparations are shown in Table 3.

Table 3. Segmented image of AML M1, M2 and M3

Subtype	AML M1	AML M2	AML M3
Original			
Segmented			

There were some cells which experience errors when segmented. The incorrectly segmented cells were not included as an artificial neural network training input data. Examples of incorrectly segmented cells and its reasons can be shown in the Table 4.

Table 4. Segmented image of AML M1, M2 and M3

Reason	Non-WBC *(erythroid cell)*	Cell body is cut off by the image's edge	Cell body is cut off by WDT	Cytoplasm area beyond the cell body
Original				
Segmented				

The result of segmentation shows that from the total of 99 white blood cell images, 876 data of white blood cell object consisting of 734 segmented data is correct and the remaining 142 segmented data is wrong thus the percentage of cells segmented correctly amounted to 83.789%. Detailed comparison of the number of cell objects resulting from correctly and incorrectly segmented cell for each subtype of AML can be shown in Table 5.

Table 5. The number of objects resulted in the segmentation of each preparation

Subtype	Correct	Incorrect	Total
AML M1	224	30	**254**
AML M2	225	72	**297**
AML M3	285	40	**325**
Total	**734**	**142**	**876**

3.2 Feature Extraction Result

The feature set of the cells was entered as artificial neural network input data. The average results of WBC feature extraction can be seen in Table 6. We can see that the values obtained from each feature are in accordance with the characteristics of the real cell type. For example, *myeloblast* has a higher average circularity while *monoblast* has a larger size of area than the other cell types.

Table 6. Average result of WBC feature extraction

Feature	*Myeloblast*	*Promyelocyte*	*Monoblast*	Support
Area (pixels)	7377,0530	13581,201	14923,724	9714,427
Perimeter (pixels)	322,482	454,659	470,482	378,412
Circularity (0-1)	0,8629	0,8153	0,8359	0,8245
Nucleus ratio (0-1)	0,8314	0,6016	0,6093	0,5238
Mean (intensity)	135,360	144,8222	149,6375	147,882
Standard deviation (intensity)	20,0017	24,9116	24,2194	31,0159

3.3 Training and Testing Result

Based on the results of training and testing conducted on each combination of momentum backpropagation parameters, it resulted the best parameters at α=0.1, z=8 and μ=0.1 in 10000 epochs. This pattern was then recalibrated five times. Details of the test result information of five times calibration can be presented in Table 7.

Table 7. Recalibration average result at α=0.1, z=8 and μ=0.1

No.	Presition (%)	Sensitivity (%)	Specificity (%)	Accuration (%)
1.	83,306	73,052	94,777	93,120
2.	84,005	77,671	95,295	93,665
3.	82,852	73,621	95,092	93,597
4.	87,484	73,434	94,990	93,460
5.	86,125	81,658	95,247	94,005
Avg.	**84,754**	**75,887**	**95,080**	**93,569**

The identification result of AML subtype in each preparation is shown in Table 8. Preparation 1 was identified as a subtype of AML M1 because the counted cell of *myeloblast* exceeded 89%. Preparation 2 was identified as a subtype of AML M2 with 72% *myeloblast,* 13.333% *promyelocyte* and 2.667% *monoblast*. Preparation 3 was identified as AML M3 in which the percentage of *myeloblast* was 8.070% and 34.737% for *promyelocyte*.

Table 8. Percentage of cell count in each preparation

Prep	*Myeloblast*	*Promyleocyte*	*Monoblast*	Support
#1	89,735%	1,786%	0%	8,482%
#2	72%	13,333%	2,667%	12%
#3	8,07%	34,737%	0%	57,193%

The three results of the inference were correct according to the type of preparation used. It means the classification of AML subtypes M1, M2 and M3 with backpropagation momentum can be said to be successful.

4 Conclusion

The ACWE algorithm could be used to segment WBC objects from image background with success rate 83.789%. The percentage of correctly segmented object data versus overall data from each preparations of AML M1, M2 and M3 were 88.189%, 75.757% and 87,692%, respectively. The feature extraction was performed on 734 cell objects from the total number of 876 segmented cell objects.

The best pattern of momentum backpropagation for AML subtype classification was 0,1 learning rate, eight neuron hidden layer and 0,1 momentum. Predictive analysis values obtained from five times calibration test were as follows: 84.754% average precision, 75.887% average sensitivity 95.090% average specificity and 93.569% average accuracy.

This research was supported by Grant PNBP 2016 Universitas Sebelas Maret, Faculty of Medicine Universitas Gadjah Mada and Dr. Sardjito Hospital, Yogyakarta, Indonesia.

References

1. Abidin, S.R., Salamah, U., Nugroho, A.S., Segmentation of Malaria Parasite Candidates from Thick Blood Smear Microphotographs Image Using Active Contour Without Edge. 1st International Conference on Biomedical Engineering (2016).
2. Agaian, S., Madhukar, M., Chronopoulos, A.T., Automated Screening System for AcuteMyelogenous Leukemia Detection in Blood Microscopic Images. IEEE System Journal, **Vol. 8, No. 3.** (2014).
3. American Cancer Society, Cancer Facts & Figures 2014. Atlanta. (2014).
4. Bell, A. dan Sallah, S., *The Morphology of Human Blood Cells. Seventh Edition*. Abbott, A Promise for Live. (2005).
5. Chan, T. F., dan Vese, L. A., Active Contour Without Edges. IEEE Transactions on Image Processing, 10(2), 266-277. (2001).
6. Dacie, Lewis, Practical Haematology: Eleventh Edition. Elsevier Churcill Livingstone. ISBN-13: 9780702034084. (2011).
7. Gonzales, R.C., Woods, R.E. *Digital Image Processing: Second Edition.* New Jersey : Pearson Prentice Hall. Upper Saddle River, 07458. (2002).
8. Hamid, A.G., Classification of Acute Leukemia, Acute Leukemia: The Scientist's Perspective and Challenge. Prof. Mariastefania Antica (Ed.), ISBN: 978-953-307-553-2. (2011).
9. Kasmin F., Prabuwono, A.S., Abdullah A., Detection of Leukemia in Human Blood Sample Based on Microscopic Images: a Study. Journal of Theoretical and Applied Information Technology. **Vol.46 Num.2.** ISSN: 1992-8645. (2012).
10. Kjeldsberg, C.R., *Practical Diagnosis of Hematologic Disorders: Second Edition.* Chicago Illinois: American Society of Clinical Pathologists (ASCP Press). ISBN : 0891894012. (1995).
11. Kusumadewi, S., *Membangun Jaringan Syaraf Tiruan Menggunakan MATLAB & EXCEL LINK*. Yogyakarta: Graha Ilmu. (2004).
12. Losen Adnyana, I. W. Dan Suega, K., Perubahan Golongan Darah Pada Penderita Leukemia Mieloblastik Akut. J Peny Dalam, **Vol. 12 No.1** (2011).
13. Leukemia & Lymphoma Society. *The AML Guide Information for Patients and Caregivers Acute Myeloid Leukemia.* 1311 Mamaroneck Avenue, Suite 310, White Plains, NY 10605. (2012).
14. Madhukar, M., Agaian, S., and Chronopoulos, A.T., New Decision Support Tool for Acute Lymphoblastic Leukemia Classification. Journal of SPIE-IS&T/ **Vol. 8295 8295181**. (2012).
15. Mishra, A. K., Fieguth, P. W., & Clausi, D. A. Decoupled Active Contour (DAC) for Boundary Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(2), 310-324. (2011).
16. Nasir, A.A.1, Mashor, M.Y.$_1$, and Hassan, R.$_2$. 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, Malaysia. (2013).
17. Patil T.G., Prof. Mr. V. B. Raskar, 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**, (2015).
18. Pradana, T.P.N., Suryani, E., Wiharto, Pemanfaatan Seed Region Growing Segmentation dan Momentum Backpropagation Neural Network Untuk Klasifikasi

Jenis Sel Darah Putih. Seminar Nasional Teknoin 2013 FTI UII. (2013).

19. Prasetyo, Eko, *Pengolahan Citra Digital dan Aplikasinya Menggunakan Matlab*. Yogyakarta: Penerbit Andi. ISBN:978-979-270-0. (2011).
20. Putzu, L., Ruberto C.D., 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).
21. Suryani, E., Wiharto, and Polvonov, N., 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. (2013).
22. Suryani, E., Salamah, U., Wiharto., Polvonov, N., Identification and Counting White Blood Study of Leukemia. International journal of Computer Science & Network Solutions. **Vol. 2 No. 6**. ISSN: 2345-3397. (2014a).
23. Suryani, E., Salamah, U., Wiharto., Wijaya., A.A., Identifikasi Penyakit Acute Myeloid Leukemia (AML) Menggunakan 'Fuzzy Rule Based System' Berdasarkan Morfologi Sel Darah Putih Studi Kasus : AML M2 dan AML M4. Seminar Nasional Semantik UDINUS Semarang. (2014b).
24. Suryani, E., Wiharto, Palgunadi, S., Pradana T.P.N., Classification of Acute Myelogenous Leukemia (AML M2 and AML M3) using Momentum Backpropagation from Watershed Distance Transform Segmented Images. International Conference on Computing and Applied Informatics. IOP Conf. Series: Journal of Physics: Conf. Series 801 (2017) 012044. doi:10.1088/1742-6596/801/1/012044. (2016).