International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Erythrocyte Classification using Multi-Layer Perceptron, Naïve Bayes Classifier, RBF Network and SVM

Dyah Aruming Tyas, Sri Hartati, Agus Harjoko, Tri Ratnaningsih

Abstract*: Several diseases can be diagnosed based on the appearance of abnormal erythrocytes, among others anaemia and thalassemia. Process of examination peripheral blood smear manually is time-consuming and subjective. Currently, the process of examination peripheral blood smear by laboratory assistants can be assisted with digital image processing technology so that it can speed up the examination time and avoid subjectivity. This research begins with the process of microscopic image acquisition, then preprocessing, segmentation, feature extraction and classification. The microscopic image acquisition is carried out using an additional special camera on a microscope. In this study, we used peripheral blood smear of thalassemia patients and healthy individuals. We convert the RGB image to grayscale image and perform the median filtering in the preprocessing stage. In the segmentation stage, we used the watershed distance transform method. As a segmentation result, we got 7108 erythrocyte images consisting of nine types of erythrocytes. In feature extraction, we used shape, color and texture characteristics to represent erythrocytes. The combination of these three features is used as classifier's input. One crucial stage in digital image processing technology is object classification. In this study, erythrocyte classification is done by comparing four types of the classifier to determine the best classifier performance in this case. Multi-Layer Perceptron (MLP), Naïve Bayes classifier, RBF Network, and SVM used as classifiers in this study. Experimental results showed that MLP got the highest performance with 89.6% accuracy, 89.3% precision and 89.6% recall. Furthermore SVM came in second place, followed by RBF Network and Naïve Bayes classifier.

***Keywords*: Classification, erythrocyte, MLP, Naïve Bayes classifier, RBF Network, SVM.**

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Dyah Aruming Tyas, Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. Email: dyah.aruming.t@mail.ugm.ac.id

Sri Hartati*, DepartmenttofCComputerrScienceeanddElectronics, Facultyyof Mathematicssand NaturalSSciences, UniversitassGadjahhMada, Yogyakarta, Indonesia. Email: shartati@ugm.ac.id

Agus Harjoko, Departmentt ofCComputer Scienceeand Electronics, FacultyyoffMathematicssanddNatural Sciences, Universitas Gadjah Mada, Yogyakarta,,Indonesia.Email:aharjoko@ugm.ac.id

Tri Ratnaningsih, Department of Clinical Pathology, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia. Email: triratnaningsih@ugm.ac.id

I. INTRODUCTION

Abnormal red blood cells (erythrocytes) have various shapes. The appearance of abnormal erythrocytes in a person's blood can indicate the presence of certain diseases. Diseases that are closely related to the appearance of abnormal erythrocytes are anaemia and thalassemia. Apart from its shape, erythrocyte abnormality can be seen from its size and color [1].

Currently, research related to medical imaging processing is growing rapidly. Some medical cases that can be helped with image processing are glaucoma [2], amniotic fluid [3], breast cancer [4], leukaemia [5]–[7], abnormality in red blood cell [8], [9], and many more. In research related to erythrocytes, there are several potential research topics to be developed. First, improving the quality of image acquisition results. Second, overlapping erythrocyte segmentation. Third, choosing the right features and increasing classification accuracy [10]. In this study, we focus on the third topic. The classification stage is one of the essential stages of medical imaging processing. Some classifiers that have been used in the classification of erythrocytes were neural networks, SVM, deep learning, Naïve Bayes classifier, logistic regression and K-nearest neighbor [9], [11]–[13].

Tyas et al. have made a performance comparison between backpropagation neural networks and deep learning. The results showed that backpropagation neural networks performance was superior to CNN [11]. Aliyu et al. also comparing the performance of deep learning and SVM to classify erythrocytes. The result showed that SVM superior to deep learning [12]. One reason for deep learning performance is lower than backpropagation neural networks, and SVM performance is that the data is scantiness for the training process in deep learning.

Lee and Chen combined shape and texture features for erythrocyte abnormality classification and erythrocyte disease classification. They proposed a hybrid neural network classifier for erythrocyte classifiers. Types of erythrocytes classified in the study are burr cell, elliptocyte cell, Horn cell, and sickle cell. The total images used in the study based on disease type were 200 images. The study achieved an accuracy of 88.25% for abnormality classification and 91% for disease classification [14].

Furthermore, Shirazi et al. classified erythrocyte images into normal and abnormal class, using extreme learning machine (ELM).

They used a combination of geometrical features and texture features as input data for the classifier. The average accuracy of the ELM classifier result was 96% [8]. In 2018, Ahmad et al. conducted a comparison of nine types of erythrocyte classification. The dataset consists of 725 abnormal erythrocytes images and 99 images of normal erythrocytes.

They used Logistic Regression, Naïve Bayes RBF Network, MLP, and CART as classifiers. From the five classifiers, Logistic Regression provides the best performance, which is 83.5% accuracy [9].

The process of examination peripheral blood smear manually is time consuming, laborious and subjective [13]. The assistance of digital image processing technology in this process is expected to accelerate the process of reading the preparations and have accurate results. In previous studies, the used images were only around 100-800 erythrocyte images. In this study, we compared the performance of several classifiers to classifying nine types of erythrocytes. We used 7108 erythrocyte images obtained from peripheral blood smear of thalassemia patients and healthy individuals. The input of classifier is a combination of texture, color, and shape features. By knowing the performance of several classifiers, we can find out the accurate classifier to classify the nine types of erythrocyte in thalassemia case.

II. RESEARCH METHOD

Fig. 1 show the research stages in this study. In the collecting data process, the first step taken is image acquisition. Then, we did the preprocessing stage to improve image quality. After that, the segmentation stage is implemented to obtain sub-images containing a single erythrocyte image. Furthermore, feature extraction is performed to obtain the characteristic values, which are the characteristics of each cell. The final stage is the erythrocyte classification.

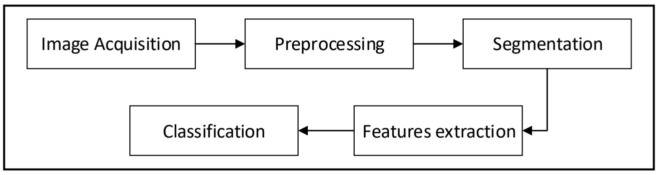

Fig. 1. Research stages in this study

A. Image Acquisition and Preprocessing

In this study, we used thalassemia patients and healthy individuals peripheral blood smear. The acquisition stage of peripheral blood smear's images is carried out using an additional special camera (Optilab) on a microscope. Furthermore, the size of the image obtained is reduced to ease the computational process. After that, the RGB image is changed to a grayscale image and then the median filtering is performed to reduce noise. Fig.2 shown the hardware used in the image acquisition stage, whereas Fig.3 shown the acquisition result image.

Fig.2. The Hardware in the image acquisition process

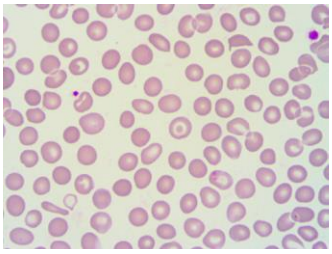

Fig.3. Erythrocyte image of thalassemia minor patients

B. Segmentation

We apply the segmentation phase to obtain a sub-image in the form of a single erythrocyte. The grayscale image is converted to a binary image. Furthermore, object detection and removal of small objects are done based on the threshold area. Then, a watershed distance transform method is used to separate attached erythrocytes. Fig. 4 shown the sub-images of the segmentation results.

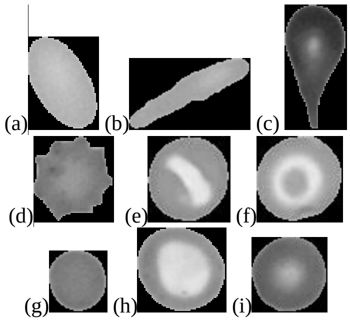

Fig. 4. Erythrocytes types: (a) Eliptocyte cell, (b) Pencil cell, (c) Teardrop cell, (d) Acantocyte cell, (e) Stomatocyte cell, (f) Target cell, (g) Sperocyte cell, (h) Hipocromic cell, (i) Normal cell.

C. Feature Extraction

Feature extraction of an object aims to obtain the feature value that describes the characteristics of the object. In this study, we use the moment invariant feature extraction method, geometry parameters, color, and GLCM. From the moment invariant, we will get seven-moment invariant values [15], while major axis, minor axis, area, perimeter, solidity, compactness, and eccentricity were the geometry parameters that we used. We use the values of mean color, standard deviation, skewness, and kurtosis as the color feature. The feature values used from GLCM [16] are energy, contrast, homogeneity, and correlation in the direction of 0 °, 45 °, 90 °, and 135 ° and the mean of each characteristic. The number of features we obtained from feature extraction are 38, as shown in Table I.

Table- I: Features obtained from feature extraction

Type of feature	Features	No of features
Shape	Invariant moment (hue's moment)	7
	Geometry parameter of cell	7
Color	Color moment	4
Texture	GLCM set of features	20
Total of features		38

D. Classification

In this study, we used 7108 images of a single erythrocyte consisting of nine types of the erythrocyte. We used four classifiers that are multi-layer perceptron, Naïve Bayes classifier, RBF Network, and SVM in this study. The four classifiers were compared in their performance in classifying nine types of erythrocytes. The k-fold cross validation method was used for the validation training and testing data. We used k set to 10 in the process. The input classifier data are all the attribute values obtained from the feature extraction, which are 38 attributes.

- Multi-Layer Perceptron (MLP)

MLP is one type of artificial neural network. In MLP, neurons are arranged in a layer configuration consisting of one input layer, at least one hidden layers and one output layer. Every single neuron in this network is connected to all next layer's neurons. The learning algorithm used in this network is backpropagation [17]. Fausett states that MLP can solve more complex problems than a single layer network, but the training stage will be more difficult [18].

- Naïve Bayes Classifier

Naïve Bayes Classifier is a machine learning method that can predict the probability of membership of a class based on the Bayes theorem with the assumption that features are all independent [17].

- Radial Basis Function (RBF) Network

Broomhead and Lowe introduced RBF Network in 1988. RBF Network consists of input, hidden, and output layer. The number of hidden layers on the RBF Network is limited to only one layer which usually uses radially symmetric functions, for example, Gaussian kernels [19]. Besides the network architecture, one of the differences between MLP and RBF Network is the activation function. The activation function that RBF Network uses is the radial basis function.

- Support Vector Machine (SVM)

SVM is a machine learning method that aims to find computationally efficient ways to learn the best hyperplane, which can separate feature spaces in high dimensions. The basic principle of SVM is a linear classifier which is further developed so that it can work on non-linear problems. The kernel in a high-dimensional workspace concept can be used to solve the non-linear problem [20][21].

III. REULT AND DISCUSSION

This study aims to determine the comparison of classifier performance in producing high accuracy. The accuracy comparison is shown in Table II. Based on Table II, MLP got the highest average accuracy, the second was SVM, third was RBF Network, and the last was Naïve Bayes Classifier. The MLP accuracy for each class was 34% -98.9%. MLP can classify teardrop cells, spherocytes, elliptocytosis, and acantocytes very well. It is shown with classification accuracy above 90%. Furthermore, normal cells, target cells, and hypochromic can be classified with an accuracy of 66% -88% whereas pencil cells and stomatocytes cannot be classified properly, seen from accuracy below 55%.

The highest classification result by Naïve Bayes Classifier is teardrop class which reaches an accuracy of 93.4%. Naïve Bayes classifier result, unlike MLP, can classify pencil cells with high accuracy of 90.6%. It followed by hypochromic, elliptocytosis, acantocyte, spherocytes, and normal cell. Naïve Bayes classifier is not good at classifying stomatocytes and target cells; this is shown with an accuracy below 42%.

Similar to MLP and Naïve Bayes classifier, the best classification result by RBF Network is in teardrop cell class with 95.9% of accuracy. Next is Elliptocytosis, acantocyte, target cell, spherocytes, normal cell and pencil cell classes. The worst classification results are in the stomatocyte and hypochromic classes with an accuracy below 30%. The teardrop cell class is also the class with the highest accuracy in the SVM classification results. It followed by the elliptocytosis, spherocytes, normal cell, acantocyte, target cell, and pencil cell. The result of poor accuracy by SVM is in the hypochromic and stomatocyte classes, which only achieve accuracy below 32%.

Table- II: Accuracy for each class

Type of cells	Accuracy (%)			
	MLP	Naïve Bayes	RBF Network	SVM
Elliptocytosis	**96.6**	87.3	87.9	**95.7**
Pencil cell	34	**90.6**	71.7	62.3
Teardrop cell	**98.9**	**93.4**	**95.9**	**98.6**
Normal cell	88.3	70.8	72.8	87.6
Stomatocyte	54.5	40.1	29.6	19.1
Target cell	77.9	34	82	83.2
Hypochromic	66.7	87.4	0	31.1
Sperocyte	**96.8**	79	79.4	89.5
Acantocyte	**93.5**	85	86.4	83.3
Average	**89.6**	**76.1**	**79.8**	**85.9**

The classification results of the stomatocyte class by four classifiers cannot achieve accuracy above 55%. The shape of the stomatocyte cell that similar to the shapes of normal cell, target cell, hypochromic, and spherocytes causes this. The same reason also applies to the target cell and hypochromic classes. The Naïve Bayes classification result for target cell was 34% of accuracy, while the RBF Network classification result for hypochromic was 0% of accuracy. Therefore, it is necessary to consider using the central pallor characteristics of the circular cells as additional features for the classifier input data. Table III shows a performance comparison of the four classifiers. MLP is a classifier that has the highest performance followed by SVM in the second-best.

RBF Network is the third best on accuracy and recall but the last for the precision, while Naïve Bayes classifier is the last one in accuracy and recall but has a precision better than RBF Network. Based on the experiments in this study, MLP obtained accuracy, precision and recall, respectively 89.6%, 89.3% and 89.6%. SVM achieved accuracy, precision and recall respectively 85.9%, 86.4% and 85.9% while the RBF Network were 79.8%, 62.9% and 79.8%, respectively.

Finally, Naïve Bayes Classifier obtained 76.1%, 78.7% and 76.1% sequential accuracy, precision and recall.

Table- III: Comparison of classification result

Classifier	Accuracy (%)	Precision (%)	Recall (%)
MLP	**89.6**	**89.3**	**89.6**
Naïve Bayes	76.1	78.7	76.1
RBF Network	79.8	62.9	79.8
SVM	85.9	86.4	85.9

IV. CONCLUSION AND FUTURE WORK

Digital image processing technology can help the process of examination peripheral blood smear to speed up the examination time and avoid subjectivity. In this study, we compared four classifiers performance in classifying nine types of erythrocytes. The features used in this study are invariant moments, geometry parameters, color moments, and GLCM. Experimental results show that MLP has the best performance in classifying erythrocytes with an average accuracy of 89.6%, followed by SVM 85.9%, RBF Network 79.8% and Naïve Bayes 76.1%. In this study, five types of circular cells causes the classification accuracy of the five cells is not optimal. Therefore, in future studies, it is necessary to consider using the central pallor characteristics of circular cells as additional features for the classifier input data.

ACKNOWLEDGMENT

The authors would like to thank the Clinical Pathology Laboratory Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada for providing data to this research. The authors also thank the Research Directorate of Universitas Gadjah Mada in the Rekognisi Tugas Akhir 2019 scheme for funding this research.

REFERENCES

1. L. Palmer *et al.*, "ICSH recommendations for the standardization of nomenclature and grading of peripheral blood cell morphological features," *International Journal of Laboratory Hematology*, vol. 37, pp. 287–303, 2015.
2. A. Septiarini, A. Harjoko, R. Pulungan, and R. Ekantini, "Automated detection of retinal nerve fiber layer by texture-based analysis for glaucoma evaluation," *Healthcare Informatics Research*, vol. 24, no. 4, pp. 335–345, Oct. 2018.
3. D. W. Ayu, S. Hartati, and A. Musdholifah, "Amniotic Fluid Segmentation by Pixel Classification in B-Mode Ultrasound Image for Computer Assisted Diagnosis," 2019, pp. 59–70.
4. S. 'Uyun, S. Hartati, A. Harjoko, and L. Choridah, "A Comparative Study of Thresholding Algorithms on Breast Area and Fibroglandular Tissue," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 1, 2015.
5. A. Setiawan, A. Harjoko, T. Ratnaningsih, E. Suryani, Wiharto, and S. Palgunadi, "Classification of cell types in Acute Myeloid Leukemia (AML) of M4, M5 and M7 subtypes with support vector machine classifier," in *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, 2018, pp. 45–49.
6. R. G. Bagasjvara, I. Candradewi, S. Hartati, and A. Harjoko, "Automated detection and classification techniques of Acute leukemia using image processing: A review," in *Proceedings - 2016 2nd International Conference on Science and Technology-Computer, ICST 2016*, 2017, pp. 35–43.
7. R. B. Jeyavathana, P. Jose, and R. Balasubramanian, "Estimation of White Blood Cells using Convolutional Neural Network," *International Journal of Engineering and Advanced Technology (IJEAT)*, no. 9, pp. 2249–8958, 2019.
8. S. H. Shirazi, I. A. Umar, N. Haq, S. Az, I. M. Razzak, and A. Zaib, "Extreme learning machine based microscopic red blood cells classification," *Cluster Computing*, 2017.
9. I. Ahmad, S. N. H. S. Abdullah, and R. Z. A. R. Sabudin, "Morphological Features Analysis for Erythrocyte Classification in IDA and Thalassemia," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 12, pp. 274–280, 2018.
10. D. A. Tyas, A. Harjoko, and S. Hartati, "A Survey of Thalassemia Identification," in *Proceedings of 25th Research World International Conference*, 2016, pp. 47–52.
11. D. A. Tyas, T. Ratnaningsih, A. Harjoko, and S. Hartati, "The Classification of Abnormal Red Blood Cell on The Minor Thalassemia Case Using Artificial Neural Network and Convolutional Neural Network," in *Proceedings of the International Conference on Video and Image Processing*, 2017, pp. 228–233.
12. H. A. Aliyu, R. Sudirman, M. A. A. Razak, and M. A. A. Wahab, "Red Blood Cells Abnormality Classification: Deep Learning Architecture versus Support Vector Machine," *INTERNATIONAL JOURNALOF INTEGRATED ENGINEERING*, vol. 10, no. 7, pp. 34–42, 2018.
13. P. T. Dalvi and N. Vernekar, "Computer aided detection of abnormal red blood cells," in *2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, 2016, pp. 1741–1746.
14. H. Lee and Y.-P. P. Chen, "Cell morphology based classification for red cells in blood smear images," *Pattern Recognition Letters*, vol. 49, pp. 155–161, 2014.
15. M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Info. Theory*, vol. IT-8, pp. 179–187, 1962.
16. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
17. G. Dougherty, *Pattern Recognition and Classification*. New York: Springer, 2013.
18. L. Fausett, Ed., *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1994.
19. D. S. Broomhead and D. Lowe, "Multivariable Functional Interpolation and Adaptive Network," *Complex System*, vol. 2, pp. 321–355, 1988.
20. N. Deng, Y. Tian, and C. Zhang, *Support Vector Machines, Optimization Based Theory, Algorithm, and Extensions*. Boca Raton: CRC Press, 2013.
21. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. United Kingdom: Cambridge University Press, 2014.

AUTHORS PROFILE

Dyah Aruming Tyas obtained a bachelor's degree (Electronics and Instrumentation) in 2013 from Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia. She worked as a software engineer at PT Multipolar Technology Tbk, an IT Service Management Company, in Jakarta, Indonesia, for 2013-2015. In 2015, she received a PMDSU scholarship from Ministry of Research, Technology and Higher Education and she continued her master's education in Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia. She is currently working toward a doctoral degree in the Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. Her research interests are image processing and artificial intelligence.

Sri Hartati obtained her bachelor's degree from faculty of mathematics and natural sciences, Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia, in 1986. Then, she obtained her master's degree in 1990 and doctorate in 1996 from Faculty of Computer Science, University of New Brunswick, Canada. Currently, she is a lecturer and a Professor in the Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. Subsequently, she is Head of Lembaga Akreditasi Mandiri (LAM) and Chairman of the curriculum committee unit of Department of Computer Science and Electronics. Her research interests include Pattern Recognition, Decision Support System, Artificial Intelligence, Medical Computing, Soft Computing, and Machine Learning. Her latest research is Machine Learning-Assisted Medical Diagnosis.

Agus Harjoko obtained his bachelor's degree, in 1986, in Electronics and Instrumentation from Universitas Gadjah Mada (UGM), Yogyakarta, Indonesia. He received scholarship from World Bank XVII Master Scholarship, World Bank in 1987 and World Bank XXI Ph.D. Scholarship, World Bank in 1991. He obtained his master's degree in 1990 and doctorate in 1996 from Faculty of Computer Science, University of New Brunswick, Canada. He is currently an associate professor and head of the Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta. He also member of the Faculty Senate and member of Electronics and Instrumentation Research Lab. His research interest includes digital image processing, machine vision, sensor network, and multimedia IR (image, audio, video processing).

Tri Ratnaningsih graduated as a medical doctor from the Faculty of Medicine, Gadjah Mada University, Indonesia in 1995. She obtained her master's degree (Tropical Medical Science) and doctorate (Health Science) from the same institution. Subsequently, she continued her career as a clinical pathologist and became a Hematology and Transfusion Medicine consultant in 2014. She was appointed the head of the hematology laboratory at Sardjito General Hospital, Indonesia up until today. She is also an active member of the Indonesian Association of Clinical Pathology as well as the International Society for Laboratory Haematology (ISLH). Currently, she is a lecturer and holds the position of head of the Department of Clinical Pathology and Laboratory Medicine at Universitas Gadjah Mada, Indonesia. Her research interests include anaemia and thalassemia.