A Classification of Platelets in Peripheral Blood Smear Image as an Early Detection of Myeloproliferative Syndrome Using Gray Level Co-Occurrence Matrix

Arizal Mujibtamala Nanda Imron1, Zilvanhisna Emka Fitri2

1 Department of Electrical Engineering, Universitas Jember, 37 Kalimantan Street, Jember 68121, Indonesia
2 Department of Information Technology, Politeknik Negeri Jember, Mastrip Street, Jember 68121, Indonesia

Abstract. Platelet disease is usually caused by abnormalities of the number or form of platelets, for example in Essential thrombocythemia (one of the groups of myeloproliferative syndrome). The characteristic of ET disease is that if a lot of giant platelets are found as large as leukocytes and cannot be detected using FBC, microscopic examination must be done manually by a clinical pathologist. The classification process begins with image processing techniques on the peripheral blood smear image, then texture features are taken using the Gray Level Co-Occurrence Matrix (GLCM) which consists of ASM, IDM and Entropy features. This feature is input into the classification system using Backpropagation. The test results, Backpropagation was able to accurately identify cells in BG images, namely leukocytes 91.84%, normal platelet cells 92.86% and giant platelet cells 84.69%. Whereas in the AL image, the accuracy of leukocyte cells is 90.82%, normal platelet cells are 96.94% and giant platelet cells are 87.76%. The average accuracy of the Backpropagation method at 84.69% BG images and AL images was 87.76%. So this classification system is able to be used as a tool for doctors or medical analysts to speed up the process of early detection, especially in myeloproliferative syndrome patients.

Keywords: Gray Level Co-Occurrence Matrix; Backpropagation; Myeloproliferative syndrome; Giant Platelets; Peripheral Blood Smear

1. Introduction

Many research related to blood cells, especially red blood cells and white blood cells, but a little research that discusses platelet cells. Whereas platelets have an important role in the blood clotting process. Platelet disease is usually caused by abnormalities in the number or form of platelets. One of the platelet diseases is Essential thrombocythemia. Essential thrombocythemia (ET) is a group of Myeloproliferative Neoplasms (MPNs) in which the bone marrow produces excessive platelets[1]. Usually in ET patients there will be an increase in platelet anisocytosis and the discovery of large (giant) platelets[2]. In myeloproliferative or myelodysplastic syndromes, we must pay attention if
some giant platelets are found as large as white blood cells. These platelets are not identified and are sometimes counted as erythrocyte cells or leukocytes in a full blood count (FBC) examination[3]. The disadvantage of FBC examination is FBC cannot detect the abnormalities of platelet morphology so it requires microscopic examination of peripheral blood smear. The process of microscopic examination in a hospital or health clinic is still done manually by a medical analyst or pathologist so that microscopic examination is subjective and depends on the experience of the medical analyst itself. for that, an automatic system is needed that can classify platelets that are objective and accurate in helping the process of early detection of myeloproliferative syndrome.

The research using the image of peripheral blood smear preparation has several factors that influence the lighting conditions, coloration and thickness of the smear so as to produce different lighting and color distribution on the image[4]. In Figure 1, there is a similarity of purplish pink between platelets and monocytes only differing in color intensity, so the classification process becomes difficult. This is caused by the concentration of blood samples that affect the density of giemsa staining so that the color differences produced between platelets and leukocytes are more similar.

![Figure 1. Digitalization microscopic peripheral blood smear](image)

Previous research on platelet classification in digitalization microscopic peripheral blood smear by image processing techniques and classification techniques using Learning Vector Quantization (LVQ) and K-Nearest Neighbor (KNN)[4]. The researcher converts the image from the RGB color space to the CIELab color space because the CIELab color space can be used to reduce the color dimensions from three to two when compared to RGB color spaces that are difficult to segment. Then the researcher used texture feature extraction with the Gray Level Co-occurrence Matrix (GLCM) technique to obtain the classification feature.

2. MATERIALS AND METHODS

Some of the steps carried out in this research are First, data in the form of peripheral blood smear image, Second, the image is carried out by color space conversion process, Third, color conversion result is segmentation image, Fourth, texture feature extraction process uses GLCM technique, Fifth, texture feature The classification process is carried out using the backpropagation method, and Sixth, the classification results are divided into 3 classes are leukocytes, normal platelets and giant platelets as in Figure 2. and are explained in the next sub-section.

In this study there were 348 data consisting of 250 training data and 98 testing data. Each data consists of 7 cells which are classified into eosinophils, lymphocytes, monocytes, neutrophil bands, PMN,
normal platelets and giant platelets as shown in Figure 3. The data has been validated based on literature and consultation with a clinical pathology specialist[4].

2.1. Color Space Conversion and Segmentation Image
Basically the original image in an RGB color space. But the RGB color space is difficult to segment so that the image must be converted to another color space as in the CIELab color space[4]. Because the CIELab color space can be used to reduce the color dimensions from three to two when compared to RGB color spaces[5]. Following the previous research, the image was converted to the CIELab color space. Then the conversion image is separated and displayed to each component, namely the L component (representing the lightness value), component a (representing the red and green colors) and component b (representing the yellow and blue colors). The next step is to reduce each color component in CIELab. this is done to reduce other cells (erythrocytes) in the segmentation image. as a comparison also, in the RGB image the same process is also done as in the CIELab image to get the segmentation image. From the previous research, it was found that there are two segmentation images that describe objects namely BG images and AL images[4]. The BG image is obtained from the reduction of blue and green components in the RGB color space. Then the AL image is obtained from the reduction in component a and Lightness in the CIELab color space. However, some platelet cells are still not able to be separated from the background (erythrocyte cells) because both cells have similar gray level values, so we do a cropping process based on the texture of platelet cells, giant
platelet cells and leukocyte cells. This has an impact on changing the image size from 251x251 pixels to 91x91 pixels as shown in Figure 4.

Figure 4. (a) Original image of RGB (b) Image of AL texture

2.2. Texture Feature Extraction

The extraction process feature is the most important thing in the classification process. The feature extraction that we use is texture-based feature extraction using the Gray Level Co-Occurrence Matrix (GLCM). The features used are Angular Second Moment (ASM), Invers Different Moment (IDM) and Entropy with angles used are 0º, 45º, 90º and 135º[4]. Retrieving the value of texture based features using the equation functions as follows:

2.2.1. Angular Second Moment (ASM)

ASM (energy) functions to measure uniformity (homogeneity) where ASM will be of high value when pixel values are similar to each other and small ASM values indicate heterogeneous GLCM Normalization. Calculation of ASM values is expressed using the function of equation 1.

\[
ASM = \sum_{i=1}^{L} \sum_{j=1}^{L} p(i, j)^2
\]  

L states the number of levels used for computing. \( p (i, j) \) is the co-occurrence matrix of the extracted image.

2.2.2. Invers Different Moment (IDM)

IDM also functions to measure homogeneity function to measure uniformity. IDM value will be high if all pixels have the same value. Calculation of IDM values is expressed using the function of equation 2.

\[
IDM = \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{p(i, j)}{1 + |i - j|}
\]  

2.2.3. Entropy

Entropy is a measure of gray value irregularities in images. The value is high if the elements in GLCM have Entropy will be of high value when the image is not uniform. While the value is low if the
elements in the GLCM approach the value of 0 or 1. Calculation of entropy values is expressed using
the function of equation 3.

\[
\text{Entropy} = \sum_{i=1}^{L} \sum_{j=1}^{L} p(i, j) \log(p(i, j))
\] (3)

3. Classification

The classification process is divided into 3 (three) classes including Class I representing leukocyte
cells, Class II representing normal platelet cells and Class III representing giant platelet cells. The
classification method used in this research is the backpropagation classification method.
Backpropagation is one of the neural network models with supervised learning algorithms.
Backpropagation is also known as the Multi Layer Perceptron (MLP) where there are many hidden
layers that are used to update the weighting value. The steps in the classification of the
Backpropagation method are:
1. Initiate all weights with small random numbers.
2. Each input neuron receives a signal and passes it to a hidden neuron layer (hidden layer)
3. Calculate all outputs on hidden neurons \( z_j \) (\( j = 1, 2, ..., p \))
4. Calculate all the outputs on hidden neurons to the neurons \( y_k \) (\( k = 1, 2, ..., m \))
5. Calculate the factor \( \delta \) output neurons based on errors in each output neuron \( y \) (\( k = 1, 2, ..., m \)).
   Where \( \delta_k \) is a unit of error that will be used in changing the weight at the output layer to the
   hidden layer.
6. Hitung perubahan bobot \( w_{kj} \) (\( \Delta w_{kj} \)) dengan laju pembelajaran \( \alpha \).
7. Calculate the factor \( \delta \) output neurons based on errors in each output neuron \( z_j \) (\( j = 1, 2, ..., p \)).
8. Calculate changes in \( v_{ji} \) weight (\( \Delta v_{ji} \)) with the rate of learning \( \alpha \).
9. Calculate the change in line weight leading to the output neuron.
10. Calculate changes in line weights leading to hidden neurons.

4. Experiment and Result

In this research, image processing techniques and feature retrieval followed previous research, only we
made improvements to the classification system. The previous research comparing two classification
methods are Learning Vector Quantization and K-Nearest Neighbor. Both have different principles in
classifying cells. The Learning Vector Quantization (LVQ), the class classification results depend on
the distance between input vectors so that if there are two similar vector inputs, the competitive layer
will classify the two vector inputs into the same class. While the K-Nearest Neighbor (KNN) method
depends on the value of K (nearest neighbor) used. If the K value increases, the success rate of the
system in classifying and recognizing cells is also large. In addition, the classification results from the
KNN method depend on the majority class. The average results of the accuracy of the two methods in
classifying cells are 74.75% in LVQ while 83.67% in KNN. The results of the comparison of accuracy
in the two methods are described in Table 1

| Method   | Average of Accuracy |
|----------|---------------------|
| LVQ      | 74.75 %             |
| KNN      | 83.67 %             |

Based on the research, the researchers developed the cell classification system. The classification
system used is the backpropagation classification method. The backpropagation method is a
supervised learning method that has advantages, for example when the classification results are not in accordance with the expected target, there will be an update in the weight of each hidden layer. As a result of the renewal of these weights, it is expected that the classification results will be in accordance with the target and minimize any errors. As a comparison with the previous method, the researchers used two segmentation images that have the highest accuracy in classifying cells. The two segmentation images are BG and AL images that have been classified using the KNN method. The cell classification results using the KNN method based on BG images and AL images. KNN was able to accurately identify cells in BG images, namely leukocyte cells 89.79%, normal platelet cells 92.85% and giant platelet cells 86.73%. Whereas in AL images, the accuracy of leukocyte cells is 87.88%, normal platelet cells are 96.94% and giant platelet cells are 87.76%, the results are illustrated in Table 2.

| Leukocytes | Normal Platelet | Giant Platelet |
|------------|-----------------|---------------|
| BG         | 89.79           | 92.85         | 86.73         |
| AL         | 87.88           | 96.94         | 87.76         |

In this research, the learning rate used was 0.3 and the epoch was 500. Backpropagation was able to accurately identify cells in BG images, namely leukocytes 91.84%, normal platelet cells 92.86% and giant platelet cells 84.69%. Whereas in AL images, the accuracy of leukocyte cells is 90.82%, normal platelet cells are 96.94% and giant platelet cells are 87.76%, the results are illustrated in Table 3.

| Leukocytes | Normal Platelet | Giant Platelet |
|------------|-----------------|---------------|
| BG         | 91.84           | 92.86         | 84.69         |
| AL         | 90.82           | 96.94         | 87.76         |

Overall, from the two classification systems, it was concluded that KNN was able to classify and identify cells accurately on BG images and AL images by 83.26% while in the backpropagation method it was able to accurately classify and identify cells in 84.69% BG images and AL 87.76% images. The results are illustrated in Table 4.

|                     | KNN     | Backpropagation |
|---------------------|---------|-----------------|
| BG                  | 83.26   | 84.69           |
| AL                  | 83.26   | 87.76           |

5. Conclusion

In this session, the researcher gave conclusions and inputs to be developed in future research. Based on our experimental results, the image used and able to represent platelets is the image of AL for CIELab color space and BG image for RGB color space. The Backpropagation classification method
is more capable and accurate in classifying platelet cells compared to the KNN classification method. The average accuracy of the Backpropagation method on 84.69% BG imagery and AL 87.76% image.

Although the backpropagation method is good enough in classifying platelet cells, there is still a lack of a classification system in recognizing giant platelet cells and lymphocyte cells. This is because the two cells have a gray value similarity so that the value will affect the extraction of texture features using ASM, IDM and Entropy and of course it will affect the accuracy of the backpropagation method classification system. As a solution, the researchers suggest adding new features such as features based on color and cell morphology so that the accuracy of the platelet cell classification system becomes higher.

The researcher hopes that this research will be able to provide new knowledge to the community about platelet classification methods and become a tool for medical analysts or clinical pathologists to analyze platelet abnormalities specifically as early detection of myeloproliferative syndrome accurately and objectively

6. References

[1] R. T. Silver, W. Chow, A. Orazi, S. P. Arles, dan S. J. Goldsmith, “Evaluation of WHO criteria for diagnosis of polycythemia vera: a prospective analysis,” Blood, vol. 122, no. 11, hlm. 1881–1886, Sep 2013.

[2] B. J. Bain, Blood cells: a practical guide. John Wiley & Sons, 2014.

[3] K. Kottke-Marchant dan B. H. Davis, Ed., Laboratory hematology practice. Chichester, West Sussex, UK ; Hoboken, NJ: Wiley-Blackwell/International Society for Laboratory Hematology, 2012.

[4] Z. E. Fitri, I. K. E. Purnama, E. Pramunanto, dan M. H. Pumomo, “A comparison of platelets classification from digitalization microscopic peripheral blood smear;” dalam 2017 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, 2017, hlm. 356–361.

[5] S. Mohapatra dan D. Patra, “Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images,” dalam Systems in Medicine and Biology (ICSMB), 2010 International Conference on, 2010, hlm. 49–54.

Acknowledgments

we thank the Institute of Research and Community Service (LP2M) of Jember University as the funder in this research.