Terakreditasi SINTA Peringkat 2
Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 158/E/KPT/2021
masa berlaku mulai Volume 5 Nomor 2 Tahun 2021 sampai Volume 10 Nomor 1 Tahun 2026

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. 1 (2022) 9 - 15 ISSN Media Elektronik: 2580-0760

Detection of Essential Thrombocythemia based on Platelet Count using Channel Area Thresholding

Prawidya Destarianto[1], Ainun Nurkharima Noviana[2], Zilvanhisna Emka Fitri[3], Arizal Mujibtamala Nanda Imron[4]
[1,2,3]Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Jember
[4]Program Studi Teknik Elektro, Fakultas Teknik, Universitas Jember
[1]prawidya@polije.ac.id, [2]aynunkarima@gmail.com, [3]zilvanhisnaef@polije.ac.id, [4]arizal.tamala@unej.ac.id

Abstract

Essential Thrombocythemia is one of the Myeloproliferative Neoplasms Syndrome where the mutation of the JAK2V617F gene causes the bone marrow to produce excessive platelets. For early detection of Essential Thrombocythemia disease using a full blood count and peripheral blood smear examination. The main characteristic is that giant platelets are found as large as young lymphocytes with a number of more than 21 cells in one field of view. The purpose of this research is to detect Essential Thrombocythemia by counting the number of platelets in the peripheral blood smear image. This research utilizes computer vision technique where the research stages consist of peripheral blood smear image, color conversion, image enhancement, segmentation, labeling process, feature extraction and K-Nearest Neighbor classification. There are three features used, namely the number of platelet cells, area and perimeter. The K-Nearest Neighbor method is able to classify 215 training data with an accuracy of 98.13% and classify 40 testing data with an accuracy of 100% based on the value of K = 3.

Keywords: *abnormalities, platelet count, essential thrombocythemia, channel area thresholding, k-nearest neighbor*.

1. Introduction

The World Health Organization (WHO) recognizes Myeloproliferative Neoplasms (MPNs) as one of several malignant myeloides. MPNs consist of polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (PMF) and myeloproliferative leukemia (MPL) caused by mutations in the JAK2 gene (JAK2V617F) [1]. This gene mutation causes the bone marrow to produce excessive blood cells. The focus discussed in this research is platelet cells, because there is a lot of literature that discusses about red blood cells and white blood cells. Essential thrombocythemia is one part of MPNs where the bone marrow produces excessive platelet cells. The initial diagnosis of essential thrombocythemia is thrombocytosis where the platelet count is > 450 x 10^9/L [2] while the normal platelet count is 150-400 x 10^9/L [3] on a full blood count (FBC). Generally, if abnormal cell counts are found, then a peripheral blood smear is performed [4]. On examination of the peripheral blood smear, megathrombocytes or giant platelets were found [3] (Figure 1) but basophil cells were not found in myeloproliferative disorders[5], [6].

The problem that often occurs is that giant platelets have the same size as leukocyte cells and even have similarities with young lymphocyte cells. Meanwhile, clinical pathology specialists perform manual examination of peripheral blood smears, so the identification process is subjective depending on his experience. So the purpose of this study is to assist clinical pathologists in early detection of Essential Thrombocythemia disease by automatically counting the number of platelets in peripheral blood smear images.

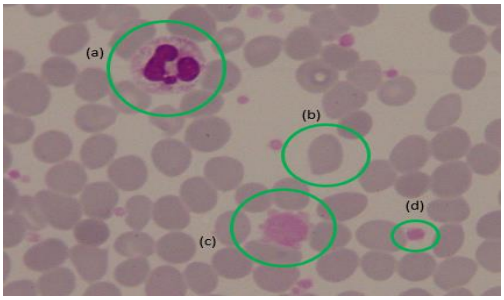


Figure 1. (a) Polymorphonuclear Neutrophil (PMN), (b) Erythrocyte, (c) Giant Platelet and (d) Normal Platelet on The Peripheral Blood Smear Image [4]

Accepted by Editor: 24-10-2021 | Final Revision: 06-12-2021 | Online Publication: 01-02-2022

This research is a development of previous research, namely the classification of platelets based on the Gray Level Co-Occurrence Matrix on peripheral blood smear images in 2017[7], then the researcher compared two classification methods to classify platelets based on texture features and found that the accuracy of KNN was 83.67% better than the accuracy of LVQ which was 74.75%[4]. The research was continued in 2019, researchers used the Backpropagation method to classify platelet cells and found that the AL image had an accuracy of 87.76% while the BG image had an accuracy of 84.69%[8]. The K-Nearest Neighbor (KNN) method is also used by researchers in classifying white blood cell abnormalities with an accuracy of 94.3% at the value of K = 23[9]. Then compared with the backpropagation method, obtained an accuracy of 91.82% at a learning rate of 0.05 and 0.3[6].

Based on the results of previous studies, the KNN method has the advantage of accuracy when compared to the backpropagation method, so that in the research of early detection of Essential Thrombocythemia based on abnormalities in the number of platelet cells using the K-Nearest Neighbor method. It is hoped that the KNN method is able to detect and classify the platelet count into three classes, namely thrombocytopenia, normal and thrombocytosis with a high level of accuracy.

2. Research Method

The research stages consist of peripheral blood smear image, color conversion, segmentation, labeling process, feature extraction and K-Nearest Neighbor classification as shown in Figure 2.

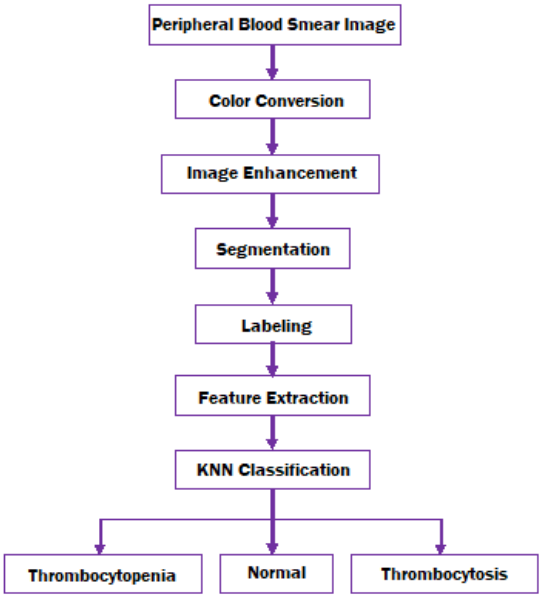


Figure 2. The Stages of Research on The System

2.1. The peripheral blood smear image

The data used in this study are peripheral blood smear images in previous studies [4], [8], where there are erythrocytes, leukocytes, normal platelets and giant platelets. In addition, there are four groups of blood smear images based on Giemsa painting colors, namely gray images, purple images, red images and orange images as shown in Figure 3. The grouping is based on the color of the red blood cells or erythrocytes.

(a) (b) (c) (d)

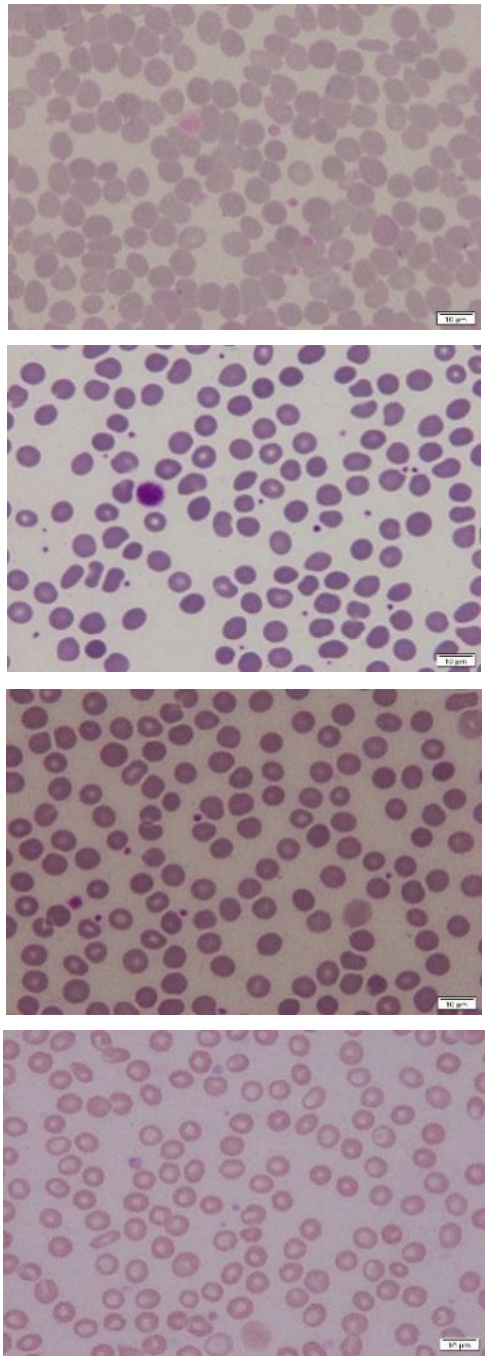


Figure 3. The Color Image (a) Gray, (b) Purple, (c) Red and (d) Orange

DOI: https://doi.org/10.29207/resti.v6i1.3571

The normal number of platelets in the peripheral blood smear is found 7-21 cells per one field of view with a magnification of 100X[10]. So for a platelet count of less than 7 cells it is called thrombocytopenia and if it is more than 21 platelets it is called thrombocytosis. The size of the peripheral blood smear image used is 1920 x 1440 pixels.

2.2. Color Conversion

Based on previous research, the four groups of images were carried out by the RGB color component splitting process. Because RGB images are difficult to segment, so it is necessary to do a component splitting process or convert them to another color space[8]. In the past research, RGB images were separated from red, green and blue color components. Then the subtraction process is carried out on each component. In this research, the red and green components were reduced, as well as the blue-green components were reduced in the entire image, but for example in the purple image as shown in Figure 4. That figure shows that from the process of reducing these components, new color spaces are obtained, for example RG and BG. Based on the color space, BG is the color space that best represents platelet cells more clearly. because it looks quite clear differences between platelet cells with erythrocyte cells that become the background.

2.3. Image Enhancement

Figure 4 shows that platelet cells have a higher gray level value when compared to the surrounding erythrocyte cells. In order to clarify the platelet cells, brightness is added which aims to improve image quality. Adding brightness using the formula equation :

$$Io(x,y) = Ii(x,y) + b \quad (1)$$

Where $Io(x,y)$ is the output image, $Ii(x,y)$ is the input image and b is the brightness constant.

2.4. Segmentation

In this research, two segmentation processes were carried out, namely grayscale thresholding and channel area thresholding. Grayscale thresholding is segmentation based on the threshold value of the gray level of the image, the process uses the formula equation[6] :

$$A(x,y) = \begin{cases} 1, I(x,y) \geq T \\ 0, I(x,y) < T \end{cases} \quad (2)$$

Channel Area Thresholding (CAT) is a segmentation based on the object area using the formula equation [11]:

$$Area_{new} = Area_{old} \geq A \quad (3)$$

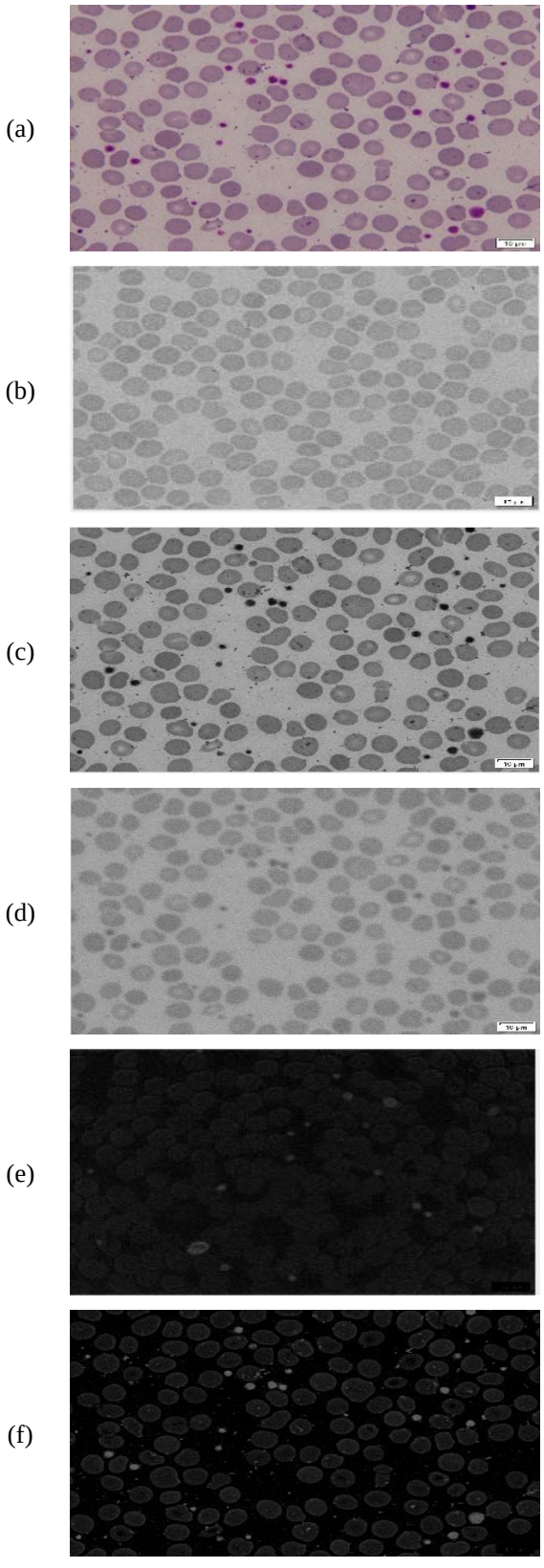


Figure 4. (a) Original Image, (b) Red Image, (c) Green Image, (d) Blue Image, (e) RG Image dan (f) BG Image.

DOI: https://doi.org/10.29207/resti.v6i1.3571

2.5. Labeling

The labeling process aims to label each white object as a result of the segmentation image. This process is needed before doing the segmentation process using Channel Area Thresholding. Each white object is grouped according to the proximity of neighbors in either four directions or eight directions as shown in Figure 5.

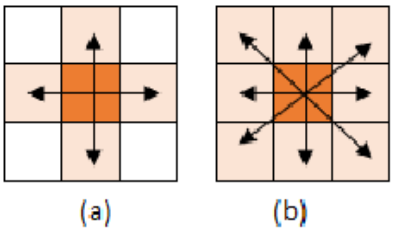


Figure 5. (a) 4 Neighboring Directions and (b) 8 Neighboring Directions[11].

2.6. Feature Extraction

Feature extraction is the final process of a digital image processing technique in which the features (parameters) of the object we study are taken. In this research, three features that represent abnormal platelet counts were taken, namely the number of platelets, area and perimeter.

To get the number of platelets, the image is segmented using the CAT, the labeling process is carried out again, so that the number of labels is the same as the number of platelets. Area is a parameter that serves to measure the size of the object while the perimeter is a parameter that functions to measure the circumference of the object. both are calculated using the formula equation [11]:

$$Area = \sum number\ of\ pixels\ by\ chain\ code \quad (4)$$

$$Perimeter = \sum even\ code + \left(\sum odd\ code\ x\sqrt{2}\right) \quad (5)$$

2.7. K-Nearest Neighbor Classification

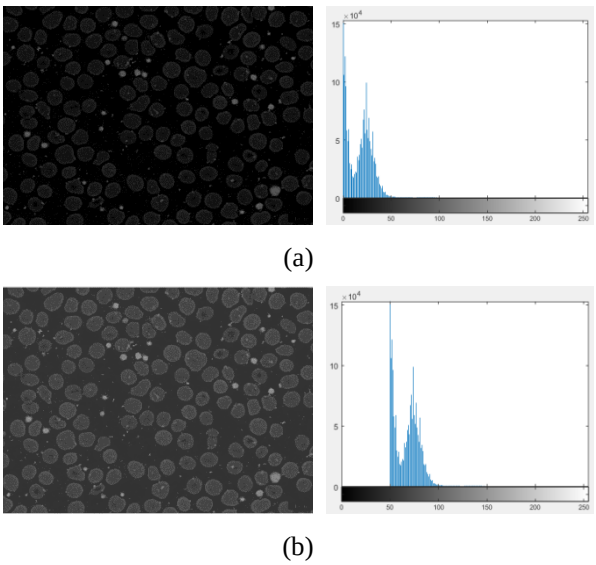
The K-Nearest Neighbor method is a simple classification method that classifies data based on the K value by calculating the closest distance using the Euclidean distance calculation in the formula equation[11] :

$$d(x,y) = \sqrt{\sum_{r=1}^{n}(xir - xij)^2} \quad (6)$$

This method includes supervised learning so that there are training data and testing data, then the number of training data is 215 and testing data is 40.

3. Result and Discussion

The process of adding brightness using equation (1) aims to improve image quality so that platelet cells are clearly visible. The brightness constant (b) used is equal to 50, resulting in changes to the histogram of the BG image as shown in Figure 6.



(a)

(b)

Figure 6. (a) BG Image and (b) BG Image after Adding Brightness

Figure 6 shows a shift in the value of the gray level in the BG image after adding a brightness of 50. In the BG image histogram after adding brightness, the graylevel value shifts to the right. This shows that the value of the graylevel is moving closer to the value of 255.

Then the segmentation process is carried out based on grayscale thresholding using the equation (2). Determination of the grayscale threshold value is different for each color image, the following Table 1 describes the optimal threshold value (T) for each color image. The grayscale threshold value is determined from the histogram of the BG image added to the brightness. The shape of platelets is small white spots, because the number of pixels is small, they are not visible on the histogram (Figure 6b). while the background tends to have a grayscale value of less than 100.

Table 1. Threshold Value (T) in Each Color Image

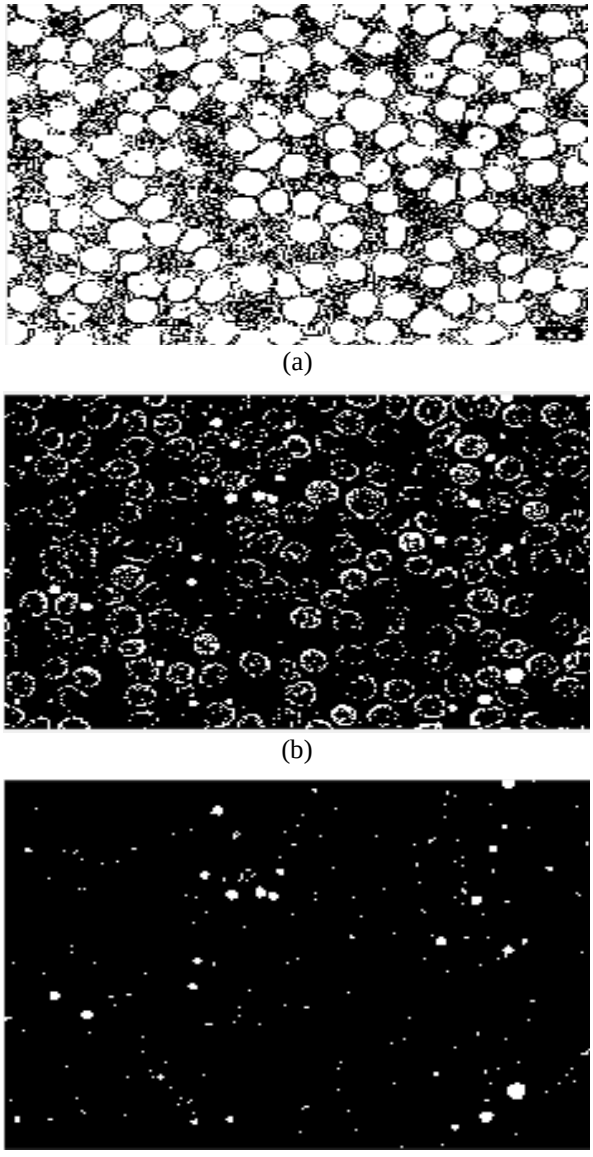
Color Image	Threshold Value (T)
Gray	80
Purple	100
Red	87
Orange	83

if we take the example of a purple image, the optimal grayscale (T) threshold value is 100, but before finding this value, the threshold value is determined several times, for example at the threshold values of 50, 80, and 100 as shown in Figure 7. In the image, the results of segmentation with a threshold value of T = 50, there is no clear difference between platelet cells and

DOI: https://doi.org/10.29207/resti.v6i1.3571

erythrocyte cells as the background. When segmented based on the threshold value T = 80, erythrocyte cells are still partially visible but platelet cells are visible.

Then at the threshold value T = 100, the segmentation image is clearly visible platelet cells but there are still smaller white spots that become noise in the segmentation image. To overcome this, it is necessary to do a second segmentation process, namely segmentation based on Channel Area Thresholding (CAT) with equation (3). The determination of the area threshold value (A) also varies for each color image and is described in Table 2. Determination of the threshold value for CAT segmentation with the assumption that the pixel value of the platelet object is greater than the pixel value of the noise. Therefore, the threshold value used has a range of 150 to 250 pixels.



(a) (b) (c)

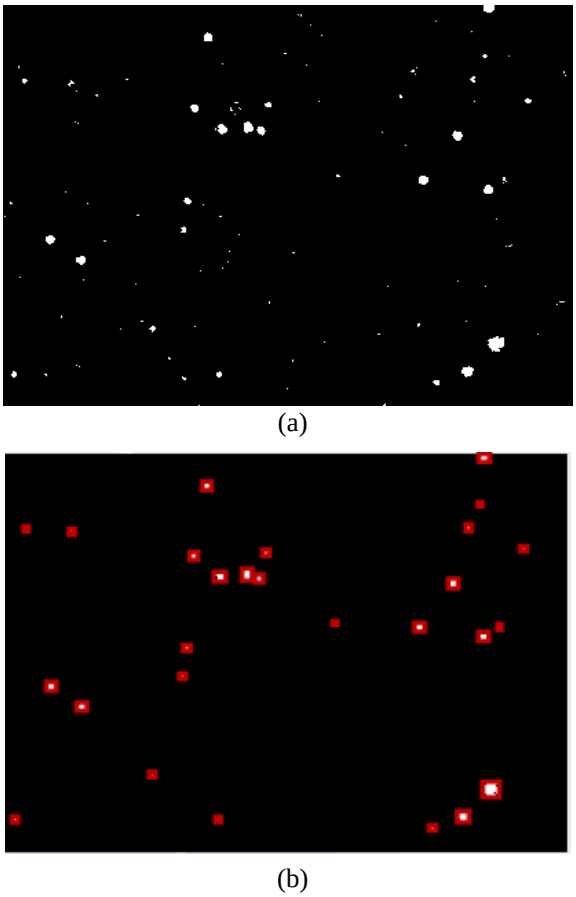
Figure 7. Image Segmentation Uses a Threshold Value of (a) T = 50, (b) T = 80 and (c) T = 100

Table 2. Threshold Value (A) in Each Color Image

Color Image	Threshold Value (A)
Gray	230
Purple	160
Red	150
Orange	250

Table 2 shows that each image has a different area threshold value (A) and this value is used to eliminate noise values in segmented images based on grayscale thresholding. However, before the CAT segmentation process, the image from the segmentation is carried out by a labeling process. each white object in the segmented image is grouped and labeled with the proximity of eight neighboring directions. This label process is also part of the chain code that functions to describe/encode the shape (contour) of an object.

After the labeling process, the segmentation process is carried out based on the area threshold value (A) as described in Table 2. The application of CAT on the purple color image is shown in Figure 8. In this image only platelet cells were detected while the small dots as noise disappeared after the CAT operation.



(a) (b)

Figure 8. Image Segmentation Based on (a) Grayscale Thresholding and (b) Channel Area Thresholding

The next step is the feature extraction process which aims to take the unique characteristics contained in the platelet cell object so that the parameters used are the

DOI: https://doi.org/10.29207/resti.v6i1.3571

number of platelets, the area and perimeter of all platelet cells in one field of view. The results of feature extraction are described in Table 3.

Table 3. The Result of Data Feature Extraction

Feature		Class		
		Thrombocytopenia	Normal	Thrombocytosis
Platelet Count	Min	1	7	22
	Max	6	21	59
Area	Min	291	2671	6040
	Max	10197	20342	42221
Perimeter	Min	93.86	584.96	1557.5
	Max	1440.89	3378.85	17982.9

The K-Nearest Neighbor method is a classification method with supervised learning where the classification target is known, so there are training data and testing data. The number of training data is 215 data, while the number of testing data is 40 data. The training data is divided into 46 data for class A (Thrombocytopenia), 136 data for class B (Normal) and 33 data for class C (Thrombocytosis). The testing data is divided into 10 data for class A (Thrombocytopenia), 20 data for class B (Normal) and 10 data for class C (Thrombocytosis).

The basic principle of the KNN method is to find the Euclidean distance in each data, then classify the data based on the value of the constant K. The first step is to find the Euclidean distance in each training data, then the data is sorted based on the smallest Euclidean distance. then, the constant value of K is determined, namely 3, 5, 7, 11, 15 and 20. The data is grouped based on the value of the constant k and then classified according to its priority class. after being classified, the accuracy of both training and testing is calculated as shown in Table 4.

Table 4. The Percentage of Training and Testing Accuracy Based on Variation in K Value

K Value	Training Accuracy	Testing Accuracy
3	98.1%	100%
5	97.6%	100%
7	96.7%	100%
11	96.2%	92.5%
15	94.4%	92.5%
20	93.9%	87.5%

Based on Table 4, the highest system training accuracy is 98.1% at K = 3, while the testing accuracy is 100%. At the value of K = 20, the training accuracy is 93.9% while the testing accuracy is 87.5%. The results of the calculation of training accuracy are obtained based on the calculation of Receiver Operating Characteristics (ROC) using the help of the confusion matrix table as shown in Table 5.

Table 5. The Confusion Matrix in The System Training Process

Classification Results			Target
A	B	C	
45	1	0	A = Thrombocytopenia
1	134	1	B = Normal
0	1	32	C = Thrombocytosis

$$Accuracy = \frac{211}{215} x100\% = 98.13\%$$

Table 5 shows that from 215 training data, in class A, there is one data that is misclassified into class B (Normal). While in class B, there were two data that were misclassified, one data being class A (Thrombocytopenia) and one data being class C (Thrombocytosis). In class C, there is one data that is misclassified into class B (Normal). So that in the training process, the system still misclassifies 4 data and the accuracy rate is 98.13%. Based on the results of the training, the calculation of testing accuracy with a value of K = 3 uses a confusion matrix as shown in Table 6.

Table 6. The Confusion Matrix in The System Testing Process

Classification Results			Target
A	B	C	
10	0	0	A = Thrombocytopenia
0	20	0	B = Normal
0	0	10	C = Thrombocytosis

$$Accuracy = \frac{40}{40} x100\% = 100\%$$

Table 6 shows that from 40 test data, the system can classify the platelet count with an accuracy rate of 100%, where 10 data are classified into class A (Thrombocytopenia), 20 data are classified into class B (Normal) and 10 data are classified into class C (Thrombocytosis) according to the target.

4. Conclusion

The K-Nearest Neighbor method is able to classify platelet count abnormalities into three classes, namely thrombocytopenia, normal and thrombocytosis with training accuracy of 98.13% and 100% test accuracy based on variations in the value of K = 3. The difficulty in this research is determining the threshold value in the segmentation process, both on Grayscale Thresholding and Channel Area Thresholding. To develop further research, it is necessary to develop image conversion so that the determination of the threshold based on the grayscale threshold and the channel area threshold only gets one value for all peripheral blood smear images.

References

[1] A. Tefferi and T. Barbui, "Polycythemia vera and essential thrombocythemia: 2019 update on diagnosis, risk-stratification and management," *Am J Hematol*, vol. 94, no. 1, pp. 133–143, Jan. 2019, doi: 10.1002/ajh.25303.

[2] P. A. Beer, W. N. Erber, P. J. Campbell, and A. R. Green, "How I treat essential thrombocythemia," vol. 117, no. 5, p. 11, 2011.

[3] S. B. McKenzie, J. L. Williams, K. Landis-Piwowar, and Teton Data Systems, *Clinical laboratory hematology*. Boston: Pearson, 2015. Accessed: Oct. 22, 2021. [Online]. Available: http://VH7QX3XE2P.search.serialssolutions.com/?V=1.0&L=VH7QX3XE2P&S=JCs&C=TC0001352396&T=marc&tab=BOOKS

[4] Z. E. Fitri, I. K. E. Purnama, E. Pramunanto, and M. H. Pumomo, "A comparison of platelets classification from

DOI: https://doi.org/10.29207/resti.v6i1.3571

digitalization microscopic peripheral blood smear," in *2017 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, Aug. 2017, pp. 356–361. doi: 10.1109/ISITIA.2017.8124109.

[5] S. C. A. Young and K. B. Poulsen, *Anderson's Atlas of Hematology*, Third Edition. Burlington: Jones & Bartlett Learning LLC, 2020.

[6] Z. E. Fitri and A. M. N. Imron, "Classification of White Blood Cell Abnormalities for Early Detection of Myeloproliferative Neoplasms Syndrome Using Backpropagation," in *Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics*, vol. 746, Triwiyanto, H. A. Nugroho, A. Rizal, and W. Caesarendra, Eds. Singapore: Springer Singapore, 2021, pp. 499–508. doi: 10.1007/978-981-33-6926-9_43.

[7] Z. E. Fitri, "Klasifikasi Trombosit Pada Citra Hapusan Darah Tepi Berdasarkan Gray Level Co- Occurrence Matrix Menggunakan Backpropagation," Institut Teknologi Sepuluh Nopember, Surabaya, 2017.

[8] A. M. Nanda Imron and Z. E. Fitri, "A Classification of Platelets in Peripheral Blood Smear Image as an Early Detection of Myeloproliferative Syndrome Using Gray Level Co-Occurence Matrix," *J. Phys.: Conf. Ser.*, vol. 1201, p. 012049, May 2019, doi: 10.1088/1742-6596/1201/1/012049.

[9] Z. E. Fitri, L. N. Y. Syahputri, and M. N. Imron, "Classification of White Blood Cell Abnormalities for Early Detection of Myeloproliferative Neoplasms Syndrome Based on K-Nearest Neighbor," *Scientific Journal of Informatics*, vol. 7, no. 1, p. 7, 2020.

[10] E. M. Keohane, C. N. Otto, and J. M. Walenga, *Rodak's hematology: clinical principles and applications*, Sixth edition. St. Louis, Missouri: Elsevier, 2020.

[11] Z. E. Fitri, L. N. Sahenda, P. S. D. Puspitasari, P. Destarianto, D. L. Rukmi, and A. M. N. Imron, "The Classification of Acute Respiratory Infection (ARI) Bacteria Based on K-Nearest Neighbor," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 12, no. 2, p. 11, 2021.

DOI: https://doi.org/10.29207/resti.v6i1.3571