IMAGE PROCESSING FOR DETECTION OF DENGUE VIRUS

Christin Erniati Panjaitan[1],Yoel Panjaitan[2] , Delima Sitanggang[3], Sri Wahyuni Tarigan[4]
[1,2,3,4] Fakultas Sains dan Teknologi- Universitas Prima Idonesia
Jl. Sampul No.4, Medan – Sumatera Utara - Indonesia
Email: christinpanjaitan@unprimdn.ac.id

ABSTRACT

Dengue is a significant health problem in tropical and Asia-Pacific regions, which typically spreads rapidly in a number of infected patients. Knowing that most of the world's population lives in risky areas, highly skilled experts and human resources are needed to diagnose and treat the disease. However, in some cases, human error may occur. Therefore, we developed a model in this research that can diagnose dengue fever disease. This study used blood smear images taken under a digital microscope with 400 x magnification specifications using image processing techniques such as color transformation, image segmentation, edge detection feature extraction, and white blood cell classification. In this study, we used white blood cell counting of the role of cell differentiation as a new feature that can classify patients' dengue viral infections via decision tree methods. The results showed that the white blood cell classification modeling technique of 167 cell images resulted in 92.2% accuracy, while the dengue classification modeling technique of 264 blood cell images resulted in 72.3% accuracy.

Keywords: Image Processing, Decision Trees, White Blood Cell, Dengue Virus, Image Segmentation.

INTRODUCTION

Dengue Hemorrhagic Fever (DHF) is a public health problem that tends to spread widely, in line with population density. This disease is one of the health problems in Indonesia, where the number of sufferers tends to increase, and its spread is increasingly widespread. Dengue Hemorrhagic Fever (DHF) is an infectious disease caused by the genus flavivirus's dengue virus transmitted through the bite of the Aedes aegypti mosquito. One DHF is characterized by continuous high fever for 2 to 7 days accompanied by bleeding and shock, which, if not treated immediately, can cause death. Dengue Hemorrhagic Fever (DHF) is a disease caused by the Dengue virus belonging to the Arthropod-Borne Virus, genus Flavivirus, and family Flavivirus. Dengue fever is transmitted through the bite of mosquitoes from the genus Aedes, mainly Aedes aegypti or Aedes albopictus. DHF can appear

throughout the year and can attack all age groups. This disease is related to environmental conditions and people's behavior.[1]

The development of dengue fever has been increasing worldwide in recent years. Before 1970, only 9 countries were endemic to dengue. The disease is present in 100 countries in the WHO, including Africa, the Americas, Southeast Asia, and the West Pacific region. The Americas, Southeast Asia, and the Western Pacific region are the most affected regions (WHO 2017). From 1968 to 2009, the World Health Organization (WHO) also noted that Indonesia had the highest dengue cases in Southeast Asia. [1,2]

In Indonesia, the prevalence of dengue hemorrhagic fever fluctuates every year and tends to increase in morbidity and spread to a broader area. In 2016, dengue hemorrhagic fever was infected in 463 regions/cities with a morbidity rate of 78.13 per 100,000 people, but the mortality rate could be reduced below 1 percent, namely 0.79 percent. 5 Dengue hemorrhagic fever is also one of the significant cases in North Sumatra. In 2020, the number of cases of dengue hemorrhagic fever in North Sumatra reached 7,584 cases, of which the number of cases of dengue hemorrhagic fever in Medan was 681 cases. [3,4]

In general, the DHF morbidity rate in Indonesia is fluctuating and relatively increasing, with the spreading area also getting wider every year. In 2016, DHF occurred in 463 regencies/cities with a morbidity rate of 78.13 per 100,000 population, but the mortality rate that occurred could be reduced below 1%, namely 0.79%. Almost every year, DHF outbreaks occur in different places with unpredictable events (Kemenkes RI, 2017). Research in 2019 found a significant increase in the incidence rate (IR), where the incidence rate (IR) of DHF in Indonesia over the last 50 years increased from 0.05 per 100,000 people in 1968 to 77.96 per 100,000 people in 2019. 2016. The IR for DHF shows a cyclical pattern, with cases peaking every six to eight years. The incidence (IR) of DHF decreased dramatically in 2017, with 68,407 points, an IR of 26.12 per 100,000 persons, and the crude death rate (CFR)) by 0.72%. A further decline in cases and IR also accompanied this decrease in 2017. In 2018, there were 65,602 cases with IR of 24.75 per 100,000 people and 467 deaths, bringing the CFR down to 0.71%. [3-5]

In North Sumatra, PS data (Central Statistics Agency) shows DHF cases in 2019 totaling 7,584 cases and 37 deaths. Of the 33 regencies/cities in North Sumatra, 3 regencies/cities have the highest coverage rates, namely Deli Serdang with 1,326 cases, Medan with 1,068 patients, and Simalungun Regency with 736 points. The leading cause of dengue hemorrhagic fever is dengue virus (DENV), which belongs to the genus Flavivirus of the Flaviviridae

family and is transmitted through the bite of the Aedes sp. The primary dengue virus vector is the Aedes aegypti mosquito. This virus has a genome stored in a span of single-strand RNA with a nucleotide base length of ± 11 kb, which is positively charged. Dengue virus infection responds differently to each individual, ranging from asymptomatic and mild symptoms to hemorrhagic fever with and without seizures. Some signs of dengue virus infection that can occur are low-grade fever, dengue fever (DF), dengue hemorrhagic fever (DHF), and dengue shock syndrome (DSS). [6]

LITERATURE REVIEW

This study uses experimental research with a simulation method using segmentation techniques to detect the formation of blood cells in an image of a blood sample and the number of white blood cells affected by the dengue virus so that it can detect the dengue virus. This research was designed to segment the blood cell image to purify one blood cell object with another blood cell (between normal white blood cells and white blood cells affected by dengue fever). [3.4]

This study begins with reading blood images as input. The process is followed by segmentation, grouping digital images into specific areas or parts to facilitate or change the image representation into something more meaningful and easy to analyze. [3-6]

This research method uses a segmentation method, namely segmentation of the Hue Saturation Lightness type al, so known as Hue Saturation and Luminosity / Luminance, namely segmentation based on RGB color components. The process separates the objects of normal white blood cells and white blood cells infected with dengue fever. Then, the color pattern is created to be detected, and the results of the color pattern are saved in JPEG and PNG format. The following process is Color Location, which is finding a color that matches the template in the input image. Calculating the number of patterns according to the template aims to count objects with colors that match the template. Then, it will calculate the number of everyday white blood cell objects and normal white blood cells infected with dengue fever that match the template. This process will generate several color locations that match the template. Marking things aims to mark the location of the color that matches the template. This process uses the coordinates obtained from the previous process. [7]

The next stage is color and component processing (component labeling). The number of each blood cell found can be calculated with the components. Components are also helpful

for noise sweeping, and this is important in analyzing objects where objects with a pixel count less than a specified number of pixels are considered noise and subsequently eliminated. Especially for closely packed cells, repeated scans are carried out so that the number of cells that corresponds to the actual number of cells is obtained. [7-8]

METHODS
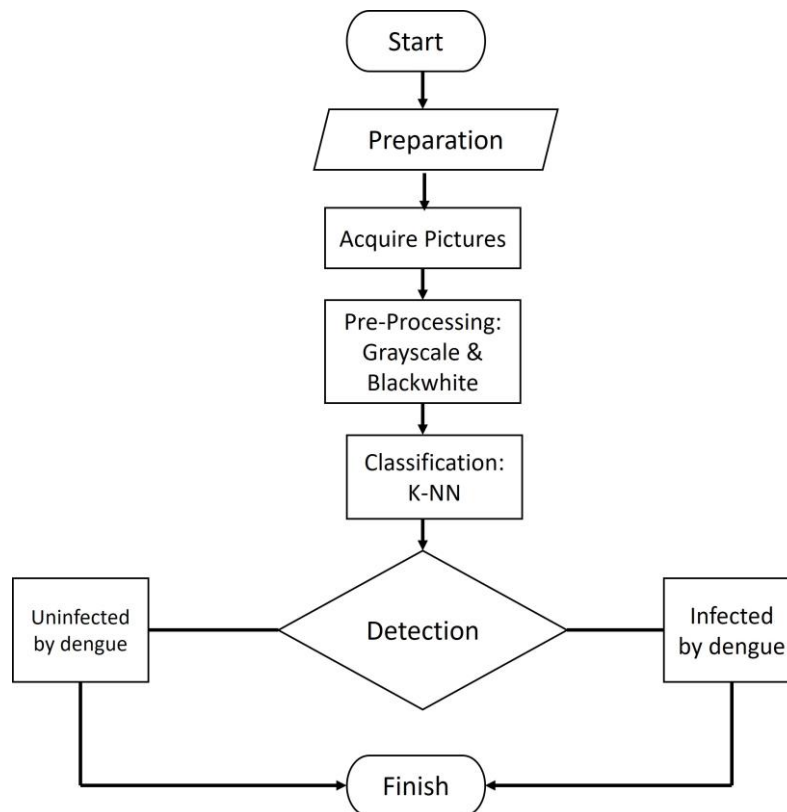
Figure 1. Research Process

The process begins with communication with the hospital to help in collecting pictures. Then, images are collected, and a pre-processing technique could be implemented. Then, the appropriate picture could pass through the classification technique.

K-nearest neighbor (KNN) is a method commonly used in data classification. This algorithm is used to classify objects based on learning data that have the closest neighbor or have a slight difference in value with the thing. KNN is a method that uses a supervised algorithm with the results of newly classified query instances based on most categories in KNN. This algorithm aims to classify new objects based on attributes and training samples.

The general principle of this algorithm is to find k-training data to determine the k-nearest neighbors based on the size of the distance. The formula commonly used to measure distance for this numeric data includes Manhattan Distance. Manhattan Distance is one of the most widely used measurements, including the placement of the squared difference by adding up the absolute differences of the variables.[9-10]

The images used in the study were collected using blood smears with a digital microscope at 400x magnification. This Image is the RGH color model with a resolution of 1600 × 1200 pixels with a 4:3 ratio where each pixel consists of 3 channels of 8 bits of color. This stage is the selection of important objects from the Image. The research area is the nucleus of the white blood cells using the KNN model. This study uses two classification models. The first model is a classification of white blood cell types, including phagocytes. This model uses KNN. The second classification model is the classification of dengue virus infection (positive) or standard (harmful) using KNN. The data needed is a phagocyte image. [11-12]

DISCUSSION

This system is generally designed to segment images of standard white and white blood cells infected with dengue fever. Next, count the number of each blood-forming cell (normal white blood cells and white blood cells infected with dengue fever). The system designed in this study is divided into 2 (two) stages of the process: the stage for making templates and the stage for fulfilling objects with templates. For the first to third processes, the two methods have the same action as follows:

Loading Image is the initial process of the system to select an image or images to be processed. Before being used as a source or input to the system, the Image or Image has been known for its microscope magnification size and stored in the same folder as the program. All samples must have the same size, microscope magnification, and lighting to get an ideal image. The segmentation method used is the HSL segmentation method. In this case, HSL is used to increase the color intensity of red blood cells. After this process, the R and G color components will generally produce a value of 0.

The color conversion process increases image contrast or clarity between the segmented images or images. The components used are Matlab. This color conversion output is RGB, HSL. At the same time, the second output is Luminance (LUMA). This process aims to

increase the clarity level of each color in the Image or images so that the boundary between each color can be detected.

The color manipulation process referred to here is by inverting the colors. This process uses the Matlab method. In Matlab, there are various processes. The process used in this system is the reverse process. The reverse process was chosen because it eliminates image or image customization, guaranteeing that the Image or Image used is clean and remains in its original condition.

Color equalization is an advanced process of the color conversion process. This process still uses Matlab components. This process is only carried out at the template-making stage, while the identification stage does not carry out the Color Equalize process. This aims to maintain the integrity of the converted color through color manipulation. In color location, it is finding a color that matches the template in the source image. ColorLocation does this process. Both stages, creating templates and identifying objects, carry out this stage. [13]

This process creates a color template that will be detected when creating an image template. The result of the template is saved in JPEG format. This process is done in Matlab. This process is only carried out when creating an image template. The template creation stage is complete here. The system will calculate objects with colors that match the template's color. Next, the number of things that match the template will be calculated. This process will also produce color locations that match the template. The result of this system's operation is in the form of an image that calculates and marks the position of the blood cells that indicate dengue fever.[14-15]

Table 1. Data Training

Sample	Age	Genre	NTT	NHB	NHR	NHT	Label
1	51	L	72	15	46	47	DD
2	42	L	47	13	37	38	DD
3	44	L	111	12	37	38	DD
4	50	L	76	17	34	40	DD
5	21	L	74	14	40	41	DBD
6	23	P	77	13	31	33	DBD
7	23	P	104	11	35	38	DSS
8	19	P	39	12	50	52	DBD
9	9	L	99	16	29	33	DD
10	7	L	84	11	30	31	DBD
11	10	P	72	15	46	47	DBD
12	15	P	47	13	37	38	DD
13	28	L	111	12	37	38	DD

14	28	L	76	17	34	40	DD
15	51	P	74	14	40	41	DBD
16	51	L	77	13	31	33	DD
17	47	L	104	11	35	38	DSS
18	39	P	39	12	50	52	DBD
19	22	P	99	16	29	33	DBD
20	9	P	84	11	30	31	DD
21	51	L	72	15	46	47	DD
22	42	L	47	13	37	38	DBD
23	44	L	111	12	37	38	DD
24	50	L	76	17	34	40	DD
25	21	L	74	14	40	41	DBD
26	23	P	77	13	31	33	DBD
27	23	P	104	11	35	38	DD
28	19	P	39	12	50	52	DD
29	9	L	99	16	29	33	DD
30	7	L	84	11	30	31	DBD
31	10	P	72	15	46	47	DD
32	15	P	47	13	37	38	DD
33	28	L	111	12	37	38	DD
34	28	L	76	17	34	40	DBD
35	51	P	74	14	40	41	DBD
36	51	L	77	13	31	33	DD
37	47	L	104	11	35	38	DD
38	39	P	39	12	50	52	DBD
39	22	P	99	16	29	33	DD
40	9	P	84	11	30	31	DD

Table 2. Data Testing

Sample	Age	Genre	NTT	NHB	NHR	NHT	Label
1	19	L	69	14	35	38	DD
2	48	L	129	14	40	43	DD
3	42	P	94	13	36	37	DBD
4	48	P	61	12	37	38	DD
5	46	L	108	13	38	39	DBD
6	46	P	95	14	36	39	DD
7	21	P	65	13	37	41	DBD
8	22	L	64	15	42	47	DD
9	10	P	184	14	38	44	DD
10	2	L	78	13	36	37	DSS

Information:

NTT   = Nilai Trombosit Terendah

NHB   = Nilai hemoglobin

NHR   = Nilai hemoglobin Terendah

NHT   = Nilai hemoglobin Tertinggi

DD   = Demam Dengue

DBD   = Demam Berdarah Dengue

DD   = Dengue Shock Syndrom

Table 3. KNN Algorithm Accuracy Results

Akurasi KNN	Normal	SDP	Terjangkit
	K2	K3	K4
Total Data Testing	8	15	31
Correct	5	11	26
Incorrect	3	4	5
Accuracy	62,5%	73,33%	83,87%
Avg-Accuracy		73,23%	

The evaluation stage is done by looking at the classification performance in the previous modeling. They are using the confusion matrix with the following results. The results of the accuracy analysis in Table 3 using the KNN algorithm show that the accuracy value of K=4 is the highest accuracy using the KNN Algorithm and 10-fold cross-validation to get an accuracy value of 83.87%. At the same time, the average accuracy value for the 3rd 〖K〗_2, K_3, and 〖K〗_4 is 73.23%.

CONCLUSION

Based on the results of data analysis using the MATLAB program according to the flowchart for the three research samples, it can be concluded that there are three samples of white blood cells infected with DHF with the following magnification image: There are 3 classifications of blood samples in this study, namely blood samples, Based on the calculation results that the closest (smallest) distance to the test data which refers to DHF (Dengue Hemorrhagic Fever) is the 3rd data calculation with the closest (smallest) distance value of 1.0467. The accuracy value of DHF detection analysis is 73.23%.

REFERENCES

World Health Organization, et al. Dengue: guidelines for diagnosis, treatment, prevention and control. World Health Organization, 2009.

NorachanAuaprasert.Chulalongkorn hematology handbook. Bangkok: chulalongkorn University, 2014.

ChitsanuPunjarean, Darin Sorsotikuland colleagues. Dengue virus diseases: New insightsandtrendsof change.Bangkok: Faculty of Medicine Chulalongkorn University.

SupinunSpeek-Saichua. Color Aids Hematology. Bangkok, 2010.

Sonka, Milan, Vaclav Hlavac, and Roger Boyle. Image processing, analysis, and machine vision. Cengage Learning, 2014.

SaicholSinsombulthong. DATA MINING.Bangkok, 2015.

Mayxay, Mayfong, et al. Predictive diagnostic value of the tourniquet test for diagnosing dengue infection in adults. Tropical Medicine &International Health 16.1, pp. 127-133, 2011.

Katz, Alfred RJ. "Image Analysis and Supervised learning in the automated Differentiation of White Blood cells from Microscopic Images." ,2000.

Joshi, Ms. Minai D., Atul H. Karode, and S. R. Suralkar. "White Blood Cells Segmentation and Classification to Detect Acute Leukemia." International Journal of Emerging Trends & Technology in Computer Science (lJETTCS) 2.3, pp.147-151, 2013.

Madhukar, Monica, Sos Agaian, and Anthony T. Chronopoulos. "New decision support tool for acute lymphoblastic leukemia classification." IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2012.

Parker, Jim R. Algorithms for image processing and computer vision. John Wiley & Sons, 2010.

Russ, John C. The image processing handbook. CRC Press, 20 II.

Bradski, Gary, and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Inc, 2008.

Nixon, Mark. Feature extraction & image processing. Academic Press, 2008.

Pang-Ning Tan. Introduction to Data Mining: Pearson New International Edition. Pearson Education inc, 2014