IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.16, No.3, July 2022, pp. 227~238
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.74531

Comparison of K-Means Clustering and Otsu Thresholding Methods in the Detection of Tuberculosis Extra Pulmonary Bacilli in the HSV Color Space

Bob Subhan Riza[*1], Jufriadif Na'am[2], Sumijan[3]
[1]Faculty of Engineering and Computer Science, Universitas Potensi Utama, Medan, Indonesia
[2,3]Computer Science Faculty, Putra Indonesia University, Padang, Indonesia
E-mail : [*1]**bob.potensi@gmail.com**, [2]jufriadifnaam@gmail.com, [3]sumijan@upiyptk.ac.id

Abstrak

Tuberkulosis Ekstra Paru (TBEP) adalah salah satu penyakit menular yang disebabkan oleh bakteri Mycobacterium Tuberculosis dan dapat menyebabkan kematian. Pasien yang menderita penyakit ini harus cepat diobati tanpa menunggu waktu yang lama. Saat ini, setiap orang yang akan dideteksi penyakit akibat bakteri ini membutuhkan waktu yang lama dan biaya besar. Biopsi merupakan salah satu teknik yang digunakan dalam mengambil cairan pada paru-paru dan diluar paru pasien. Cairan ini diberi pewarnaan kimiawi Ziehl Neelsen selanjutnya dilakukan pemeriksaan menggunakan mikroskop untuk mengidentifikasi penyakit Tuberkulosis. Penelitian ini bertujuan membantu mendeteksi bakteri dengan cepat dan tepat dengan melakukan pengolahan citra berbantu komputer dengan membuat sebuah sistem aplikasi. Teknik yang digunakan adalah melakukan pengembangan terhadap metode segmentasi. Penelitian ini diawali dengan melakukan transformasi ruang warna HSV (Hue Saturation Value) dimana Kelebihan menggunakan ruang warna HSV adalah terdapat warna-warna yang sama dengan yang dicapture oleh indera manusia. Proses segmentasi yang dilakukan adalah mengembangkan dengan teknik K-Means dan Otsu Thresholding. Dari hasil kedua metode yang digunakan ternyata metode Otsu Thresholding yang dapat mendeteksi basil TBEP dengan tingkat akurasi yang lebih baik dari metode K-Means. Sehingga metode yang dikembangkan ini sangat membantu dalam mempercepat dan meminimalisasi biaya untuk mendeteksi TBEP.

***Kata kunci*—** *Tuberkulosis Ekstra Paru, K-Means, Otsu Thresholding, Hue Saturation Value (HSV), Segmentasi.*

Abstract

Tuberculosis Extra Pulmonary (TBEP) is an infectious disease caused by the bacterium Mycobacterium tuberculosis and can cause death. Patients suffering from this disease must be treated quickly without waiting long. Currently, anyone who will be detected caused by this bacterium takes a long time and costs a lot. The biopsy is one of the techniques used to take the patient's lung fluid and give Ziehl Neelsen chemical dye and then observe using a microscope to determine this TBEP disease. This research aims to help detect bacteria quickly and precisely by performing computer-aided image processing by creating an application system. The technique used is to develop the segmentation method. The segmentation process is to develop a Hue Saturation Value (HSV) color space transformation technique with the K-Means and Otsu Thresholding techniques. From the results of the two methods used, it turns out that the Otsu Thresholding method can detect TBEP results with more accuracy than the K-Means method. So the method developed is beneficial in accelerating and minimizing costs for detecting TBEP.

***Keywords*—** *Tuberculosis Extra Pulmonary, K-Means, Otsu Thresholding, Hue Saturation Value (HSV), Segmentation.*

Received May 9th,2022; Revised May 13th, 2022; Accepted May 31th, 2022

1. INTRODUCTION

Tuberculosis is a life-threatening infectious disease worldwide caused by the bacterium Mycobacterium tuberculosis. These bacteria are in the form of acid-fast bacilli or often called acid-fast bacilli (AFB). This bacillus measuring 1-4 m long and 0.3-0.56 m wide, as shown in Figure 1, is not spore-forming, non-motile, and facultative. Bacterial cell walls contain long-chain glycolipids that are mycolic and rich in acids and phosphopoglycans [1][2][3].

Tuberculosis (TB) is a chronic and infectious disease that affects the world's human population and requires complex treatment. It is a public health problem with more than 9 million estimated new cases and 1.5 million deaths annually worldwide [4]. Of the estimated 9 million people who contracted TB in 2013, more than 80% were in Southeast Asia, the Western Pacific, and Africa. The majority of the infected population comes from poor and marginalized communities with weak health services infrastructure[5].

Tuberculosis can affect every human being regardless of region. Usually, the ones who develop this disease are adults, with 30 countries affected by Tuberculosis where almost 90% of those affected fall ill. Those of productive age are susceptible to TB between 15 to 50 years and children. TB usually comes out through phlegm and coughing. If the saliva is scattered at low temperatures, the possibility for germs to survive will be long enough to allow the transmission process to occur. There are two types of Tuberculosis, namely Tuberculosis Pulmonary (TBP) and Tuberculosis Extra Pulmonary (TBEP). TBP affects the lungs, whereas TBEP can affect any organ of the body except the spine, heart, pancreas, skeletal muscle, and thyroid.

So far, to detect Tuberculosis Extra Pulmonary (TBEP) through a biopsy, namely by taking fluid from a person's lymph nodes which a doctor or health analyst will detect, then placed on the preparation and viewed through a microscope for readings on the preparation to see the presence of germs or tuberculosis bacilli. Noticing what is happening will take a long time because the liquid preparations are viewed under the microscope carefully, and the liquid preparations contain 150 fields of vision[6]. Tuberculosis Extra Pulmonary can cause complications. Based on this, we need a system for good reporting and recording for TB control [7].

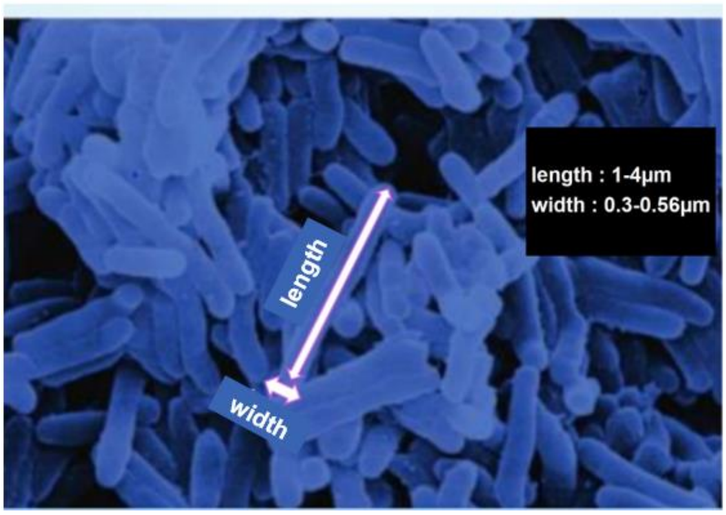


Figure 1. *Mycobacterium Tuberculosis* [7]

WHO explained that the impact of the Covid 19 Pandemic increases the number of deaths due to Tuberculosis globally could increase by around 0.2 -0.4 million in 2020. The estimated effect of the Covid 19 Pandemic on the number of deaths due to Tuberculosis is shown in Figure 2[8].

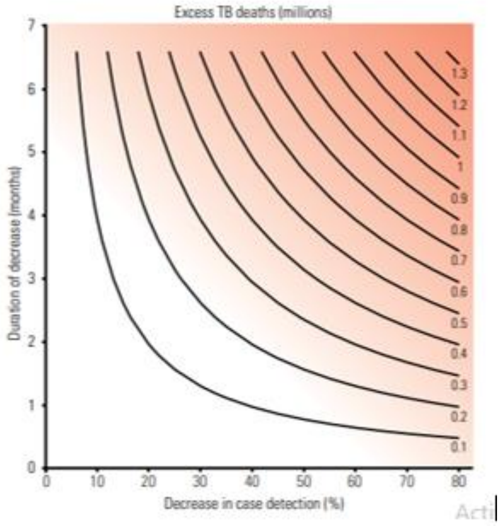


Figure 2. Estimated impact of the COVID-19 pandemic on the number of TB deaths globally in 2020, for various combinations of decreased case detection and duration of decline [8].

Tuberculosis can damage the human lungs or other parts of the body and cause severe illness. Tuberculosis is spread through the air when a person who has TB in the lungs coughs, sneezes, or talks and transmits it through the air. Further spread occurs if a person inhales the germ. Tuberculosis mainly spreads from frequent and prolonged associations, for example, with family members or friends [9][10][11].

Several image processing methods have been reported for the detection of Tuberculosis Extra Pulmonary (TBEP). Riza et al., in their research, made a system or process automatically in diagnosing Tuberculosis Extra Pulmonary (TBEP) with segmentation stages using the thresholding method and feature extraction using the invariant moment method.[12].

Chetan detects tuberculosis images from the screening results using the thresholding technique[13], whereas Jorge researched Tuberculosis Extra Pulmonary (TBEP) from the results of screening with a segmentation process[14].

Nahid et al. detected tuberculosis bacteria using the Ziehl Neelsen staining. This research uses automatic steps; for the segmentation stage, using the color-based Bayesian to identify possible Tuberculosis objects by eliminating artifacts where the resulting outcome is "definite," "probably," or "non-Tuberculosis." After testing for all data collected from the TB Hospital laboratory, the accuracy rate is very good, with an accuracy value of 90% [15].

Hamed et al.'s research developed a feature extraction through shape and color statistical model to identify Tuberculosis bacilli. Tuberculosis images are obtained based on an examination of the sputum smear under a light/fluorescent microscope. The statistical model is used as a general library to reconstruct the results with the background color and geometric feature extraction method. This study uses various methods to classify individual and overlapping bacilli from the rest of the image based on the eigenvalues of the shape and color models. Using the statistical shape model and statistical color model, the KNN classifier produces an average accuracy value of 82.7% for detecting single bacilli and overlapping bacilli. For identifying only individual bacilli from overlapping bacilli and other objects, the accuracy value is 99.1%[16].

Aeri et al. used color feature extraction to identify Mycobacterium tuberculosis using an expert system. This study uses 1266 image data from sputum, which consists of two classes. The method used in image improvement is a median filter with a color histogram to extract color features. Color histograms are used to cope with changes such as image shift, rotation, and image angle. The extracted colors are HSV colors, where a collection of hue values forms a histogram cluster. Classification method with AdaBoost and an expert system (random forest), the results of the Tuberculosis classification get an accuracy value of 85% [17].

Aeri et al. conducted a study on the classification of Mycobacterium tuberculosis based on color feature extraction using HSV (Hue Saturation Value) using adaptive boosting. Prior to identification, a filter was performed using the median filter method, where this study was to determine the effect of color in identifying Tuberculosis bacteria. Extraction of color histogram features is carried out using several quantization measures. The color histogram is formed from the combined results of the hue values of each image pixel. The hue value itself ranges from 0 to 360, which represents the color of each image pixel. The steps to be taken are 8, 16, 32, and 64 using a combination of Adaboost from decision-making to learning methods. The results of this study obtained the best Tuberculosis bacteria classification accuracy value in the testing process using the Adaboost method, namely 81.7%[18].

2. METHODS

1. Block Diagram

In making a system to diagnose Tuberculosis Extra Pulmonary (TBEP) for the segmentation stage, there are several processes carried out, which can be seen in Figure 3.

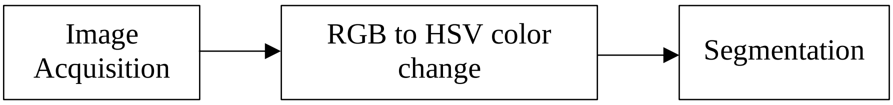


Figure 3. Block Diagram Segmentation

2. Image Acquisition

At the stage of taking images of Tuberculosis (TB) disease, namely by taking fluid from patients with Tuberculosis Extra Pulmonary through a biopsy, then the fluid is placed into preparation and given Ziehl Neelsen chemical dye so that the object to be viewed through a microscope can be seen clearly. The microscope used to take pictures or pictures of the disease is connected to a computer through an application to make it easier to capture/take pictures of Tuberculosis. The equipment used to analyze the preparations was an Olympus BX53 digital microscope with a 10x eyepiece and a 100x objective lens so that it was magnified 1000x to be able to see the structure of AFB (mycobacterium) using a 1000x magnification from the microscope shown in Figure 4. The structure of AFB cannot be seen when using a lower magnification. The captured images are 78 images consisting of 51 EPTB images and 27 Non-EPTB images with dimensions of 1920x1440 pixels. The image or picture was taken from H. Adam Malik Central General Hospital, Medan. Tuberculosis Extra Pulmonary (TBEP) images seen under a microscope can be seen in Figure 5.

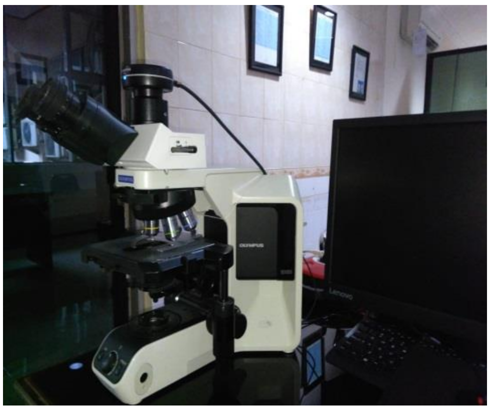


Figure 4. Digital Microscope

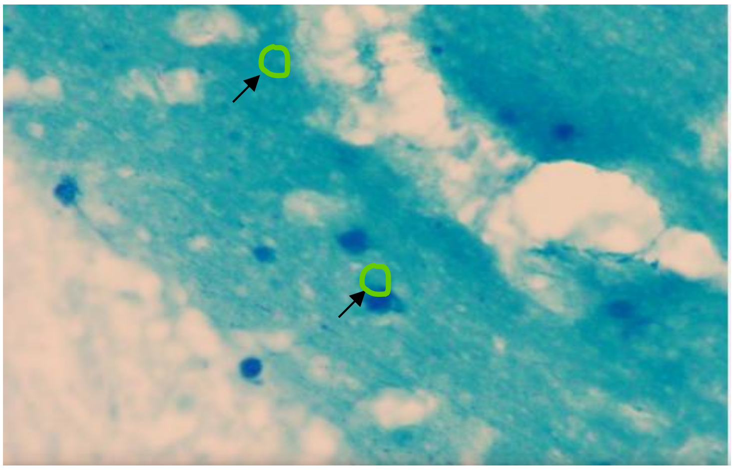


Figure 5. Tuberculosis Extra Pulmonary (TBEP) Images
(Source: H. Adam Malik Central General Hospital, Medan, 2020)

Figure 4 shows Tuberculosis Extra Pulmonary. The presence of Acid Fast Bacilli (AFB) in the figure. This BTA shows that the sample seen in the Mycobacterium Tuberculosis image is in the form of a red rod (bacillus) inside a green circle. Mycobacterium Tuberculosis is red due to the previous staining with Ziehl Neelsen. This staining makes it easier to differentiate between bacilli and non-bacilli.

3. RESULTS AND DISCUSSION

1. HSV Color Space Transformation

There are several methods used in the transformation of the RGB color space, including CMYK, HSV, and L*a*b. HSV (Hue Saturation Value) color is a color that is derived from the RGB (Red Green Blue) color model, so to produce HSV color, you must go through the process of converting colors from RGB to HSV. In this study, we have tried various color spaces, including RGB, and the better result is HSV to get to the segmentation stage.

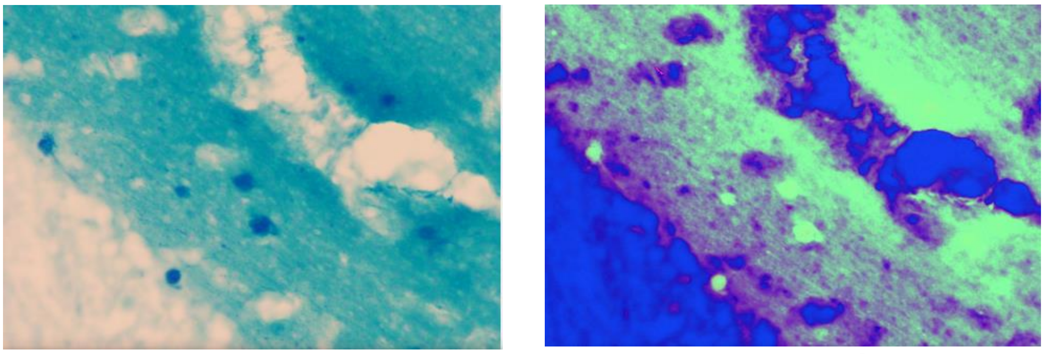
After taking Tuberculosis and Non-Tuberculosis images, the next step is to transform the HSV (Hue, Saturation, Value) color space. Hue is the actual color, saturation is the purity of the color, and value is the brightness of the color. The advantage of using the HSV color space is that there are colors that are the same as those captured by the human senses. The results of the HSV color can be seen in Figure 6.

The image resulting from the HSV transformation will be used for the segmentation stage, following equation (1) for HSV color[19]:

$$H = \begin{cases} \begin{cases} \theta & if\ B \leq G \\ 360\ \ \theta & if\ B > G \end{cases} \\ cos^{-1}\left\{ \dfrac{\frac{1}{2}\left[(\mathrm{R}-\mathrm{G})+(\mathrm{R}-\mathrm{B})\right]}{\left[(R-G)^2+\ (\mathrm{R}-\mathrm{B})\,(\mathrm{G}-\mathrm{B})\right]^{1/2}} \right\} \end{cases}$$

$$S = \begin{cases} 0 & if\ R = G + B \\ 1 - \dfrac{3}{(R+G+B)}\left[\min(R,G,B)\right] & other \end{cases} \qquad \text{..................(1)}$$

$$V = \frac{1}{3}\,(R+G+B)$$



a b

Figure 6. (a) RGB Image, (b) Image converted from RGB to HSV

Figure 6 (a) shows the original image, namely RGB, then the RGB image is converted to HSV (b). The transformation of the RGB color space into HSV can be assumed that the coordinates R, G, B (0 or 1) are a sequence of red, green, and blue in the RGB color space, min is the minimum value of value (red, green, blue). After converting the color space to HSV, the Hue component is used because the three HSV components can visually distinguish between objects and the background. The Hue component is the Hue component, where the Hue component is obtained from the HSV image. The image of the Hue component can be seen in Figure 7.

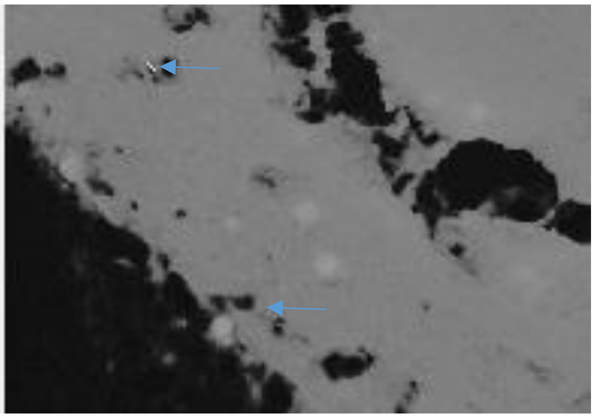


Figure 7. Hue Components of RGB to HSV Color Space Transformation

Figure 7 is the Hue component of the result of the RGB to HSV color space transformation. In Figure 7. it can be seen that the basil object can be different from the color of other objects. The color of the basil object, as shown by the arrow, is white.

2. Segmentation

Image segmentation is the identification and isolation of an image into certain areas with the aim of conforming to the structural unit. Segmentation is an important operation in biomedical image processing because it is used to isolate physiological and biological structures. The general approach to segmentation can be grouped into three classes: pixels, regions, and edges. In this study, the purpose of the segmentation process is to divide the network image into two regions, TBEP, and non TBEP regions. The TBEP region refers to objects that characterize the TBEP result, while the non- TBEP region refers to the background and objects in the image

At this stage, two methods are used, namely K-Means and Otsu Thresholding. The following figure 8 is the result of several segmentation methods that have been carried out to distinguish between background and foreground:

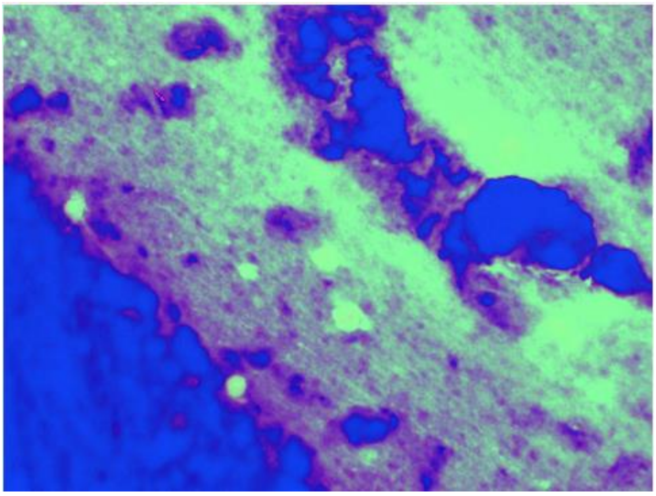


Figure 8. Image segmentation results from the K-Means. HSV staining technique

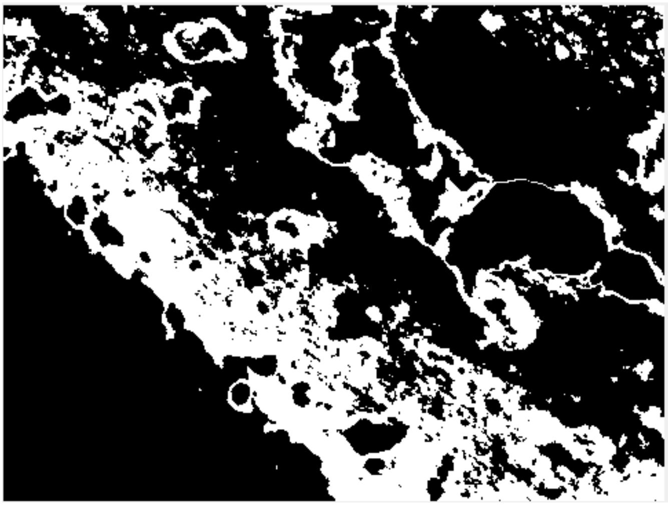


Figure 9. Image segmentation results using the K-Means method

Figure 8 is the result of segmentation using the HSV staining technique for the segmentation method. Figure 9 is a segmented image using the K-Means method with a value of k = 3. The input values are hue and saturation values, while the output is a segmented binary image. The segmentation results using the K-Means method cannot display the Tuberculosis Extra Pulmonary bacilli in the image used.

Otsu thresholding is used automatically to calculate the threshold value and maximize the variance between classes from classes separated by a threshold, which aims to separate BTA objects from the background and can detect the results in the image or image used. The Otsu thresholding equation can be seen in equation (2)[20]:

$$\sigma^2 = P_{nw}(M_{nw} - M)^2 + P_w(M_w - M)^2$$
$$M = P_{nw}.M_{nw} + P_w.M_w \qquad \text{……………….(2)}$$
$$P_{nw} + P_w = 1$$

$$t^* = \underset{a \le t \le b}{\text{Arg Max}}\{P_{nw}(M_{nw} - M)^2 + P_w.(M_w - M)^2$$

Where :
σ^2 : Variants within the Tuberculosis and Non-Tuberculosis classes
P_{nw} : pixel value probability for non-Tuberculosis class
P_w : pixel value probability for Tuberculosis class
M_{nw} : average non-Tuberculosis class pixel value
M_w : Tuberculosis class average pixel value
M : average pixel value of the image
t^* : average pixel value of the image

The segmentation results using Otsu thresholding with the HSV (Hue, Saturation, Value) coloring technique are shown in Figure 10 and 11.

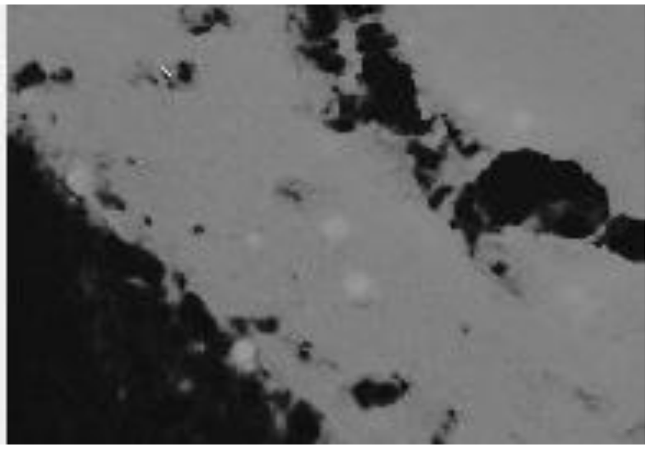


Figure 10. Image result of HSV staining technique

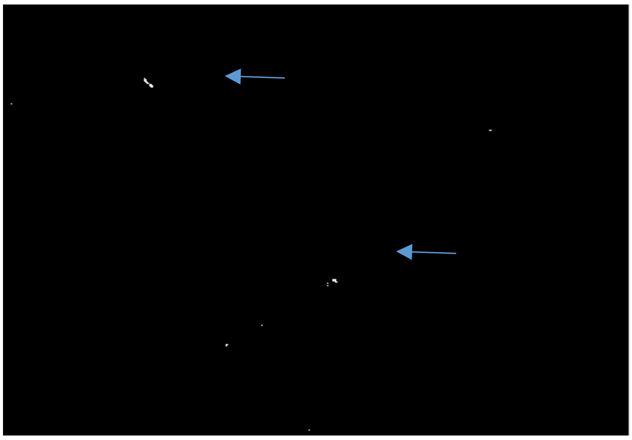
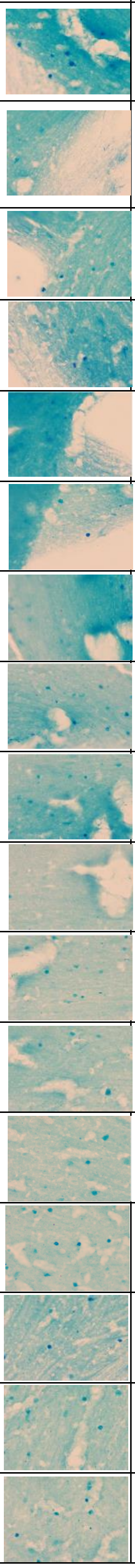
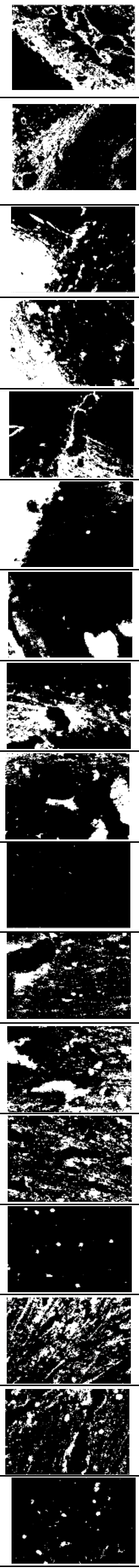
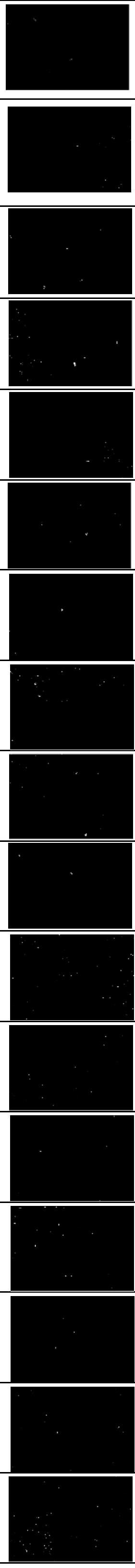
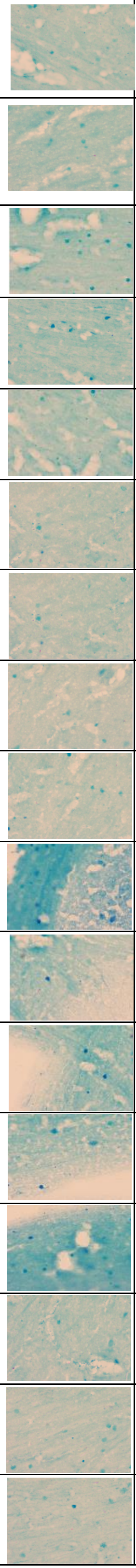
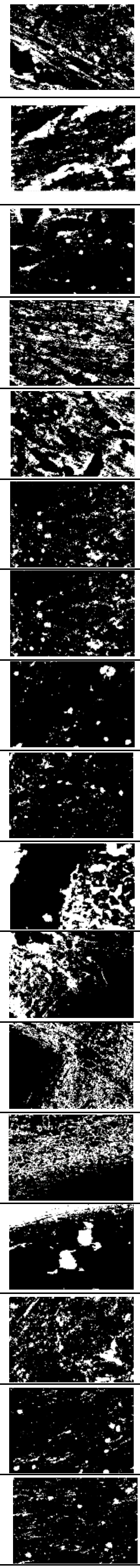


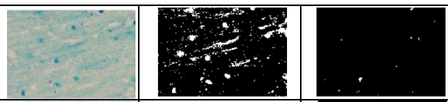
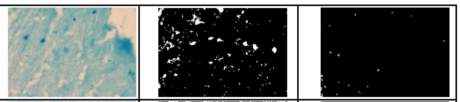
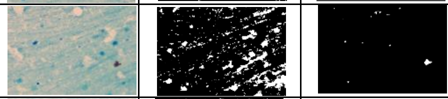
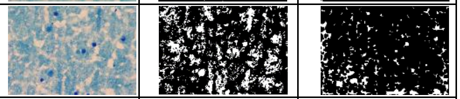
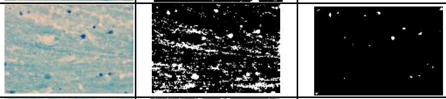
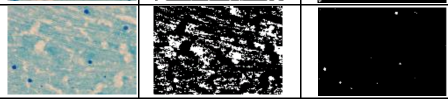
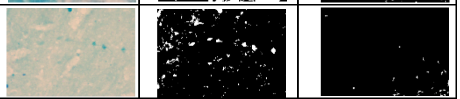
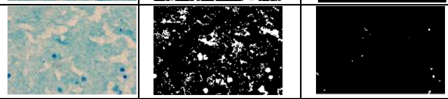
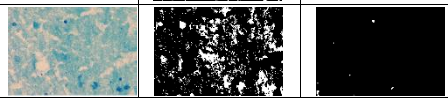
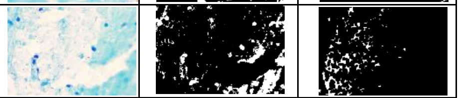
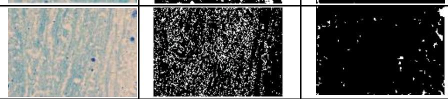
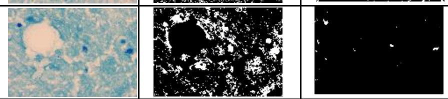
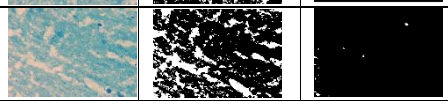
Figure 11. Image segmentation results from Otsu thresholding

Figure 10 is the result of segmentation using the HSV staining technique. Figure 11 is the result of thresholding using Otsu thresholding. The results show that the basil object can be separated from other objects. In the image, it can be seen that objects other than basil have a black background while the basil object is in the foreground with white color. The blue arrow in Figure 11 shows an image of basil that can be detected by this method. Otsu.

Table 1 describes the results of segmentation testing using the K-Means and Otsu thresholding methods, and from these tests, the results show that the Otsu thresholding method is better.

Table 1. Test results of TBEP image with K-Means and Otsu thresholding methods

No	RGB Image	K-Means Image Result	K-Means Image Result		No	RGB Image	K-Means Image Result	K-Means Image Result
1					27			
2					28			
3					29			
4					30			
5					31			
6					32			
7					33			
8					34			
9					35			
10					36			
11					37			
12					38			
13					39			
14					40			
15					41			
16					42			
17					43			

18					44			
19					45			
20					46			
21					47			
22					48			
23					49			
24					50			
25					51			
26								

4. CONCLUSION

Results of the color transformation process from RGB to HSV that the Hue component has better image results than the S component and V component, the segmentation test carried out is the K-Means method and the Otsu Thresholding method, the result is that the Otsu Thresholding method can display Tuberculosis Extra Pulmonary (TBEP) clearly on the images used in this research.

5. ACKNOWLEDGMENT

We thank the H. Adam Malik Central General Hospital, which has allowed us to take pictures or images of Tuberculosis Extra Pulmonary (TBEP). We thank the health analysts at the Health Polytechnic of North Sumatra, who have assisted us in taking the image for this research.

REFERENCES

[1] B. S. Riza, M. Y. Mashor, M. K. Osman, and H. Jaafar, "Automated segmentation procedure for Ziehl-Neelsen stained tissue slide images," *2017 5th Int. Conf. Cyber IT Serv. Manag. CITSM 2017*, Oct. 2017, doi: 10.1109/CITSM.2017.8089292.

[2] B. S. Riza, M. Y. Mashor, M. K. Osman, and H. Jaafar, "Segmentation for Tuberculosis Ziehl-Neelsen Stained Tissue Slide Image Using Thresholding," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, Mar. 2019, doi: 10.1109/CITSM.2018.8674335.

[3] "Etiologi Tuberkulosis Paru - Alomedika."

https://www.alomedika.com/penyakit/pulmonologi/tuberkulosis-paru/etiologi (accessed May 09, 2022).

[4] T. Ferkol and D. Schraufnagel, "The global burden of respiratory disease," *Ann. Am. Thorac. Soc.*, vol. 11, no. 3, pp. 404–406, 2014, doi: 10.1513/ANNALSATS.201311-405PS.

[5] Y. Cao *et al.*, "Improving Tuberculosis Diagnostics Using Deep Learning and Mobile Health Technologies among Resource-Poor and Marginalized Communities," *Proc. - 2016 IEEE 1st Int. Conf. Connect. Heal. Appl. Syst. Eng. Technol. CHASE 2016*, pp. 274–281, Aug. 2016, doi: 10.1109/CHASE.2016.18.

[6] "(5) Detection of Tuberculosis Bacilli using Image Processing Techniques | Request PDF." https://www.researchgate.net/publication/277089955_Detection_of_Tuberculosis_Bacilli_using_Image_Processing_Techniques (accessed May 12, 2022).

[7] "Global tuberculosis report 2017." https://apps.who.int/iris/handle/10665/259366 (accessed May 09, 2022).

[8] "Global tuberculosis report 2020." https://www.who.int/publications/i/item/9789240013131 (accessed May 09, 2022).

[9] P. Bulan, "Pendahuluan DICARI PARA PEMIMPIN UNTUK DUNIA BEBAS TBC," Accessed: May 09, 2022. [Online]. Available: www.who.int/gho/mortality_burden_disease/cause_death/top10/en/.

[10] H. Hartatik, "Diagnosa Penyakit Pulmonary Tuberculosis Dan Extrapulmonary Tuberculosis Menggunakan Algoritma Certainty Factor (CF)," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 8, no. 1, pp. 11–24, Mar. 2016, doi: 10.22303/CSRID.8.1.2016.11-24.

[11] "Buku Ajar Tuberkulosis Ekstra Paru — Universitas Indonesia." https://scholar.ui.ac.id/en/publications/buku-ajar-tuberkulosis-ekstra-paru (accessed May 09, 2022).

[12] "ACADSTAFF UGM." https://acadstaff.ugm.ac.id/karya_files/selection-of-suitable-moment-invariant-features-for-mycobacterium-tuberculosis-detection-in-ziehl-neelsen-stained-tissue-images-dcfcd07e645d245babe887e5e2daa016 (accessed May 12, 2022).

[13] C. C. Pawar and S. R. Ganorkar, "TUBERCULOSIS SCREENING USING DIGITAL IMAGE PROCESSING TECHNIQUES," *Int. Res. J. Eng. Technol.*, 2016, Accessed: May 09, 2022. [Online]. Available: www.irjet.net.

[14] J. L. Díaz-Huerta, A. del Carmen Téllez-Anguiano, M. Fraga-Aguilar, J. A. Gutiérrez-Gnecchi, and S. Arellano-Calderón, "Image processing for AFB segmentation in bacilloscopies of pulmonary tuberculosis diagnosis," *PLoS One*, vol. 14, no. 7, p. e0218861, Jul. 2019, doi: 10.1371/JOURNAL.PONE.0218861.

[15] N. Kamal and A. Chakrabarty, "Tuberculosis Diagnosis through Image Processing."

[16] H. Yousefi, F. Mohammadi, N. Mirian, and N. Amini, "Tuberculosis Bacilli Identification: A Novel Feature Extraction Approach via Statistical Shape and Color Models," *Proc. - 19th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2020*, pp. 366–371, Dec. 2020, doi: 10.1109/ICMLA51294.2020.00065.

[17] "Mycobacterium Tuberculosis identification based on colour feature extraction using expert system." https://www.cabdirect.org/globalhealth/abstract/20203239846 (accessed May 12, 2022).

[18] A. Rachmacl, N. Chamidah, and R. Rulaningtyas, "Classification of mycobacterium tuberculosis based on color feature extraction using adaptive boosting method," *AIP Conf. Proc.*, vol. 2329, no. 1, p. 050005, Feb. 2021, doi: 10.1063/5.0042283.

[19] T. A. Aris, A. S. A. Nasir, L. C. Chin, H. Jaafar, and Z. Mohamed, "Fast k-means clustering algorithm for malaria detection in thick blood smear," *2020 IEEE 10th Int. Conf. Syst. Eng. Technol. ICSET 2020 - Proc.*, pp. 267–272, Nov. 2020, doi: 10.1109/ICSET51301.2020.9265380.

[20] N. Otsu, “THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS.,” *IEEE Trans Syst Man Cybern*, vol. SMC-9, no. 1, pp. 62–66, 1979, doi: 10.1109/TSMC.1979.4310076.