The Implementation of K-Means Algorithm as Image Segmenting Method in Identifying the Citrus Leaves Disease

F G Febrinanto¹*, C Dewi ¹, A Triwiratno ²

¹Brawijaya University, Malang, Indonesia
²The Research Institute for Citrus and Subtropical Fruits (Balitjestro), Batu, Indonesia

E-mail: falihgozifeb@gmail.com, dewi_candra@ub.ac.id

Abstract. The purpose of this study is to identify the disease on citrus leaves. A digital imagery makes it possible to identify disease automatically. Three diseases to be identified in this research are CVPD (Citrus Vein Phloem Degeneration), Downy Mildew, and Cendawan Jelaga. The research will study the implementation of an image segmentation to analyze the citrus leaves diseases. The method that will be used to do image segmentation is K-Means. The segmentation which will be carried out consist of two kinds, namely a leaf segmentation and a disease segmentation. After segmentation process, the results of disease segmentation are classified by using the K-Nearest Neighbor (K-NN) algorithm to know its disease class. From data analysis, the results of the optimal cluster shows that the leaf segmentation consist of 2 clusters and the disease segmentation consist of 9 clusters. While the obtained optimal parameter K gives score of 4. The accuracy percentage for disease identification in this study is 90.83%. Furthermore, the analysis states that the accuracy can be more increased by using a minimum bound parameters. Finally, overall the results show the optimal value at the minimum bound of 3%, its accuracy can be increased to 99.17%.

1. Introduction

Citrus plants are one of the horticultural commodities that are specifically considered by the Indonesian government. The Indonesian government, through the Ministry of Agriculture, said that citrus plants are annual crops that are already around 70-80% developed in Indonesia. Every year, citrus plants has a development in their cultivation, including land area, amount of production and even market demand (Rizal, Bachrian, & Retno, 2011).

The quality of citrus plants is a concern to increase the production and quality of citrus fruits. One factor that causes the low quality of citrus plants is the disease that attacks the leaves. Leaves have a function for photosynthesis, but if the leaf is attacked by disease then it is resulting the inability to provide food to the plant. Plant growth becomes obstructed and in the end the plant produces fruit with low quality which can cause the plant to wither or die. For examples of diseases that attack the citrus leaves are CVPD (Citrus Vein Phloem Degeneration), Cendawan Jelaga, and Downy Mildew (Triwiratno, 2018).

Now days, the method that can be done to identify citrus diseases is by looking directly or by laboratory tests. To be able to identify citrus diseases, more knowledge is needed so that control
can be done properly and correctly. A solution is needed to identify citrus plants automatically so that concerns about the lack of knowledge of the citrus farmers can be overcome (Triwiratno, 2018).

As the advance of technology, many innovations have been developed by humans, including in the field of digital image processing. Segmentation is one of the methods in digital image processing to differentiate objects of an input image. One algorithm that can be used for image segmentation is K-Means.

In previous studies, the implementation of the K-Means method for segmentation was used to create an applications for processing digital image of meat detection based on openCV and Eclipse (Arsy, et al., 2016). In this study, the application of digital image processing uses beef meat as an object. The results of the identification of the quality of beef in the application of digital image processing with the K-Means clustering method has produced a good output with a percentage of 80% success.

In other similar studies, researchers have made an application to detect fruit based on geometric properties of roundness and also to detect the level of fruit maturity using color characteristics. Some researchers also segmented using the K-Means method by using the L * a * b * color space. After segmentation, the researcher then classifies fruit and detects maturity using color characteristics. It produces a good output with an accuracy of 93.89% (Andri, et al., 2014). Furthermore, the others researchers uses the K-Means algorithm for segmentation of white blood cells. The focus of segmentation is on the separation of the nucleus and cytoplasm of the cell. The utilization of the subtract image aims to produce the desired part. The results of this study shows the accuracy of 95.7% of nucleus segmentation and 93.1% of cytoplasmic segmentation (Zhang, et al., 2014).

Based on the above studies, we have decided to use the K- Means method to segment the disease in the image of a leaf. We will identify the disease in citrus leaves which are suffered of 3 diseases, namely CVPD (Citrus Vein Phloem Degeneration), Cendawan Jelaga, and Downy Mildew by using the K-Nearest Neighbor (K-NN) algorithm. The reason for segmenting the leaf disease parts is to produce optimal results in the process of classifying the test data on K-NN algorithm training data. Thus, in this study the leaf disease will be the main concern of our study.

2. Literature Review

2.1. Citrus Leaf Disease

2.1.1. CVPD (Citrus Vein Phloem Degeneration). This disease has external symptoms characterized by yellowish color on adult leaves. This disease also has characteristics such as deficiency of Zn, Mn and Fe elements, and the colored leaf bones are greener than the rest of the leaves (Wahyuningsih, 2009).

2.1.2. Cendawan Jelaga. Cendawan Jelaga disease is caused by Capnodium citri fungus. This disease attacks leaves, twigs, and fruit thus the affected part has a layer of black fungus (Syafiril, 2006).

2.1.3. Downy Mildew. Downy Mildew disease is also known as Embun Tepung. This disease is caused by white flour-shaped fungi attack. They attack the leaf surface. The effect of fungus attack is drying on the leaves (Syafiril, 2006).

2.2. Resizing

Resizing is a process to change the image size to be larger or smaller than the original size. Changing the size of the image can be resulting the different of color values. This can change the digital content of the image (Wijaya & Prayudi, 2015).
2.3. **Segmentation**

Image segmentation means dividing images into several pixel groups. In different groups, it has a high contrast. Whereas in the same group, it has a high similarities (Dhanachandra, Manglem, & Chanu, 2015).

2.4. **Rescaling**

Rescaling is the process of adding brightness to image by using a rescaling operator. The equation of the rescaling process can be seen through the following equation (Knudsen, 1999):

\[ c = \text{scaleFactor} \cdot c_0 + \text{offset} \]  

Where \( c \) is a result of the rescaling process and \( c_0 \) is the image before rescaling process. ScaleFactor and offset values are constant values that determine the brightness level of the image.

2.5. **RGB Color Space**

RGB color space is the primary color combination of red, green and blue that is commonly used for computers or televisions. The color used for each element has a 8-bit color value. (Rulaningtyas, Sukmono, Mengko, & Saptawati, 2015).

2.6. **XYZ Color Space**

The XYZ color space, also known as the CIE XYZ color space, it is obtained from the transformation of RGB color space through a 3x3 matrix transformation process. Transformation uses tristimulus values, which is a configuration of three linear light color components that correspond to CIE color matching. Calculation for transformation from RGB color space to XYZ can be done through calculation of transformation matrix with the following equation (Rulaningtyas, Sukmono, Mengko, & Saptawati, 2015):

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} = 
\begin{bmatrix}
0.412453 & 0.357580 & 0.180423 \\
0.212671 & 0.715160 & 0.072169 \\
0.009334 & 0.119193 & 0.450227
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]  

(2)

2.7. **L*a*b* Color Space**

L*a*b* color space also known as the CIELAB color space is the most complete color space set by the International Commission on color illumination (French Commission Internationale de l’Eclairage, also known as CIE). The L*a*b* color space has the ability to describe all colors that can be seen by the human eyes. (Rulaningtyas, Sukmono, Mengko, & Saptawati, 2015). The calculation of L*a*b* color space conversion can be done through the XYZ color space by the following equation (Pratt, 2007):

\[
\begin{align*}
L^* &= 116 \left( \frac{Y}{Y_n} \right)^{1/3} - 16 \\
L^* &= 903.3 \left( \frac{Y}{Y_n} \right)^{1/3} \text{for } 0.0 \leq \frac{Y}{Y_n} \leq 0.008856 \\
a^* &= 500 \left[ f \left( \frac{X}{X_n} \right) - f \left( \frac{Y}{Y_n} \right) \right] \\
b^* &= 200 \left[ f \left( \frac{Y}{Y_n} \right) - f \left( \frac{Z}{Z_n} \right) \right]
\end{align*}
\]

where,

\[
\begin{align*}
f(w) &= w^{1/3} \text{for } w > 0.008856 \\
f(w) &= 7.78(w^{1/3}) + 1.379 \text{for } 0.0 \leq w \leq 0.008856
\end{align*}
\]

According to the L*a*b* model, L represents the level of brightness (Luminance), where a* represents a space between green to red, b* represents a space between blue and yellow. The L*a*b* model is designed to be an independent model. Thus this model can manage colors regardless of devices such as printers, monitors, or computers (Hapsari & Hidayattullah, 2013).
2.8. **K-Means**

K-Means is a non-hierarchical data clustering method. Grouping a data using the K-Means method in general can be done with the following basic algorithm (Agusta, 2007):

1. Determine the number of initial clusters.
2. Placing centroids according to the number of clusters randomly.
3. Get the data that has the closest distance.
4. Allocate data into clusters according to the centroid using the nearest distance calculation.
5. Allocate data into clusters according to the new centroid using the nearest distance calculation.
6. Return to step 4, if there is still moved data from cluster or in the centroid value above the specified threshold value or if the objective function value changes above the threshold value.

2.9. **Segmentation Using K-Means**

In the segmentation process using K-Means, the first step is to determine the number of clusters in the pre-processing results and calculate the centroid value randomly. Next, calculating the pixel distance to the centroid and classifying the pixels based on the closest distance. After the pixel value of the image becomes one group according to its closest distance, then recalculating the centroid to produce a new centroid through the calculation of pixel averages. Next, allocating each data into the cluster according to the new centroid. If there is still moved data from cluster, a new centroid is calculated, but if there is no moved pixel value from the cluster, the clustering process is stopped (Mardhiyah & Harjoko, 2011).

2.10. **K-Nearest Neighbour (K-NN)**

K-Nearest Neighbor (K-NN) is one of the supervised learning algorithms where the output of a new data is classified according to the majority group of the closest neighbors. The following are the steps of the K-NN algorithm (Priandana, Zulfikar, & Sukarman, 2016):

1. Determine the value of k.
2. Calculate the distance between the test data and each training data.
3. Get the data that has the closest distance.
4. The amount of data that has the closest distance is the class that appears the most frequently.
5. The most appearing value is considered as the class of the test data.

2.11. **Euclidean Distance**

Euclidean Distance is used for classification or identification process by calculating the distance between a vector training and vector testing in the existing database. Euclidean Distance calculations can be calculated through the following equation (Wibowo & Usman, 2010):

\[
d = \sqrt{\sum_{i=1}^{n} (P_i - Q_i)^2}
\]

Where,

\(d\) = Distance score
\(n\) = The number of data
\(P_i\) = The i\textsuperscript{th} testing data
\(Q_i\) = The i\textsuperscript{th} training data
2.12. Silhouette Coefficient

Silhouette Coefficient is a method used to see the quality and strength of a cluster and as well as an object placed on a cluster. The following is the calculation formula of the Silhouette Coefficient value (Asana, Widyantara, Wirastuti, & Adnyana, 2017):

\[ s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \]  

Where,

- \( s(i) \) = The value of Silhouette Coefficient
- \( a(i) \) = The distance mean between point \( s(i) \) and the other points of the same cluster
- \( b(i) \) = The distance mean between point \( s(i) \) and the other points of all different clusters

2.13. Accuracy

Accuracy calculation is done to test the accuracy of the method. It is studied to compare the processing results with the actual output target. Accuracy is represented by using the numbers 0 to 100 in percent (%). Accuracy calculation is obtained through the following equation:

\[ \text{ACCURACY} = \frac{\text{The number of the true observed data}}{\text{The number of all observed data}} \times 100\% \]  

3. Research Method

3.1. Data Collection

Data for this study were taken from the Research Institute for Citrus and Subtropical Fruits (Balitjestro) located in the Tiekung area of Batu city. A total of 136 leaf image data were collected and it were divided into 120 test data and 16 training data. 120 test data consists of 30 types of leaf categories, namely leaves with healthy categories, leaves with CVPD disease, leaves with Cendawan Jelaga disease, and leaves with Downy Mildew disease. Then, 16 training data were also divided into 4 leaf categories, namely Healthy Leaves, CVPD, Cendawan Jelaga, Downy Mildew. The way of how to take pictures is by picking the leaves and placed in a red cover as a leaf clamp, after that we take a picture on the leaves by using an Asus Zenfone Selfie ZD551KL 13 MP smartphone camera with a distance of 20 cm through well distributed lighting. After taking pictures of the leaves, then the cropping process is done respecting to the background of devices.

3.2. Algorithm Design

In this research, several processes were carried out, including the image input of orange leaves, pre-processing of images, segmentation by using K-Means, training process, testing process, and calculation of accuracy. The flowchart of algorithm design can be seen in Figure 1.

![Figure 1. The Flowchart of Algorithm Design](image-url)
3.2.1. Pre-processing.
In pre-processing, there are several sub-processes that must be carried out. The first process is a resizing process which function is to equalize the size of the input image such that the size is uniform. Then, we do the rescaling process to adjust the brightness of the input image in accordance with Equation 1. The last process of pre-processing, we make changes to the RGB color space into a color space $L^* a^* b^*$ where in the RGB color space processing is changed first become XYZ color space by using matrix transformation according to Equation 2, then it is processed into $L^* a^* b^*$ color space by using equation 3.

3.2.2. Segmentation Using K-Means.
In this research, the input image whose the pre-processing has been done and which has produced the $L^* a^* b^*$ color space list, is used for data process on segmentation using K-Means. The calculation of image segmentation with K-Means algorithm is only using two variables from the $L^* a^* b^*$ color space, namely variables $a^*$ and $b^*$. It aims to focus on the level of color purity by ignoring the level of brightness (luminance). In this case, there are two stages of segmentation, the first is leaf segmentation which functions to separate the leaves with cover, second is the disease segmentation which functions to separate the diseased area from the leaf part.

3.2.3. Training Process.
The training process is aimed to produce a training data that has the color space feature $L^* a^* b^*$. The training data was then inputted into the training data table as a reference of the classification process for citrus leaf disease.

3.2.4. Testing Process.
At the end of this research, it can be assessed the level of accuracy of the program and other factors that affect the accuracy of the program. Likely the training process, it is needed to do the stages, namely image pre-processing and segmentation using K-Means. In the testing process, the classification of leaf disease is carried out each cluster using the K-NN algorithm.

4. Data Analysis

4.1. Scale Factor Score Analysis
The testing of the scale factor value aims to get the best scale factor value. This scale factor value is used in the image pre-processing stage. The pre-processing stage mentioned above is a rescaling process, which is used to adjust the brightness of the image. The higher the Scale Factor value, the higher the brightness of the image.

The use of testing methods is by comparing the best accuracy output results for each parameter of the value of tested scale factor. Scale Factor parameters are tested by using values 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0. Other parameters are needed to support the testing of Scale Factor values that have been determined. Some of them are:
6. The size of image resolution: 375 x 500,
7. Number of clusters in leaf segmentation: 2
8. Number of clusters in disease segmentation: 9
9. Number of K values in the disease classification process: 3

The graphic of the testing scale factor value can be seen in Figure 2.

![Figure 2. The diagram of Scale Factor result](image)

In the testing of the Scale Factor value, the best value is 1.1. This value has a total of 105 classification data of true value and 15 classification data is incorrect value. Thus, it has an accuracy percentage of 87.5%.
4.2. **Optimal Cluster Score Analysis**
The optimal cluster value is needed for processing the K-Means algorithm. Testing the optimal cluster value in this research uses Silhouette Coefficient method. In accordance with Equation 5, the calculation of the Silhouette Coefficient value is performed for all data and is taken on average in the overall Silhouette Coefficient calculation results. There are two tests of optimal cluster values which include leaf segmentation and disease segmentation. The parameter value of the testing cluster has a value range of 2 to 10. In this test, it takes a leaf image by selecting one from the entire testing data.

#### 4.2.1. **Leaf Segmentation Analysis.**
Testing on leaf segmentation focuses on finding the optimal number of clusters in the K-Means algorithm when processing leaf segmentation. Leaf segmentation aims to separate the orange leaf area and cover area. Here are some supporting parameters needed in the testing process:

1. The size of image resolution: 375 x 500
2. The scale factor value: 1.1
3. The initial centroid uses the color space value L * a * b * input image on the first n data.
   The n data is adjusted by the number of parameters and the number of testing clusters. It has been done such that the results of the Silhouette Coefficient calculation does not change.

The results of testing the optimal cluster K-Means algorithm on leaf segmentation is shown in the graph in Figure 3.

![Figure 3. The diagram of Optimal Cluster Number Analysis on Leaf Segmentation](image)

In Figure 3 the graphic of the analysing results reaches the optimal point when the number of clusters is 2 with the Silhouette Coefficient value of 0.89. At the time of testing, the number of clusters is 7 and it continues with no cluster formation such that the Silhouette Coefficient value becomes # N / A (undefined). Therefore, it is concluded that the number of clusters is 2 and it will be the most optimal number of clusters in leaf segmentation.

#### 4.2.2. **Disease Segmentation Analysis.**
Testing of disease segmentation focuses on finding the number of optimal K-Means algorithm clusters when the processing disease segmentation is done. The disease segment itself functions to separate the area of the leaf according to the similarity of its features to be carried out in the identification process of the disease in the next stage. To support the testing process, several additional parameters are needed which includes:

1. The size of image resolution: 375 x 500
2. The scale factor value: 1.1
3. Number of clusters in leaf segmentation: 2
4. The initial centroid uses the color space value L*a*b* input image on the first n data. The n data is adjusted by the number of parameters and the number of testing clusters. This treatment is needed such that the results of the Silhouette Coefficient calculation does not change in each process.

The number of clusters with the largest Silhouette Coefficient value is when the number of clusters is 9. The Silhouette Coefficient value is generated when the number of clusters of 9 is 0.68. When testing the number of clusters of 10 there is no cluster formation, thus the Silhouette Coefficient value becomes # N / A (undefined). It is concluded that the number of clusters is 9 which is the most optimal number of clusters in disease segmentation.

The results of testing the optimal cluster K-Means algorithm on leaf segmentation is shown in the graph in Figure 3.
4.3. The Analysis of Optimal K Score of K-NN

K value is one of the parameters for processing using the K-NN algorithm. The function of finding the optimal K value is focused on the classification of orange leaf disease. Testing the optimal K value in this study uses an accuracy calculation method by comparing the number of correct classifications and incorrect classification number.

There are 120 testing data included in the optimal K value testing process, it consists of 30 leaf images in the healthy category, 30 leaves with Downy Mildew disease, 30 leaves with Cendawan Jelaga, and 30 leaves with CVPD disease. The parameters of the K value are tested with a value range of 1 to 20. To support the testing process, several additional parameters are needed including:

1. The Size of image resolution: 375 x 500
2. The scale factor value: 1.1
3. Number of clusters in leaf segmentation: 2
4. The number of clusters in disease segmentation: 9

According to Figure 5, the most optimal K value when K is 4. When the parameter K is 4 the results show that the number of true value classifications is 109 and the number of incorrect classifications is 11. With this program, it has an accuracy of 90.83%. The resulting accuracy for each K parameter is vary and it tends to be constant until experiment K of 18, and after that the accuracy begins to decrease, thus it is not possible to increase the experimental parameters at a greater value.

4.4. Program Accuracy Analysis

By using the best parameters from the testing results, then the testing is done on 120 data with 30 healthy leaf image data, 30 Downy Mildew diseased leaves, 30 Cendawan Jelaga leaves, and 30 CVPD diseased leaves, it was resulting the best accuracy of the program at 90.83%. The comparisons for each leaf type on the classification results are described in Table 1

| Leaf Class      | Number of Correct Classification | Number of Incorrect Classification | Accuracy |
|-----------------|----------------------------------|-----------------------------------|-----------|
| Healthy Leaf    | 19                               | 11                                | 63.33%    |
| Downy Mildew    | 30                               | 0                                 | 100%      |
| Cendawan Jelaga | 30                               | 0                                 | 100%      |
| CVPD            | 30                               | 0                                 | 100%      |

In Table 1, it shows the results of accuracy of each leaf class. For diseased leaves (Downy Mildew, Cendawan Jelaga, and CVPD), the program has been able to classify well and has an accuracy of 100%. However, the different conditions is shown in the healthy leaf class program, it can only reach 63.33% accuracy. The next analysis is by testing one by one of the class of leaves which has an incorrect classification. By using 3 healthy leaf samples that are not precisely classified, the processing results can be shown in Figure 6.
From the results of the identification process, the first input image was classified as Cendawan Jelaga, the second input image was classified as Cendawan Jelaga, and the third image produced was classified as Downy Mildew. According to Figure 1, the program results have an output of detecting only a few diseased areas on the edge of the leaf. Thus, 3 input images that should have a healthy output leaf tag can not be classified properly. In accordance with the existing analysis, a minimum limit of diseased areas is needed.

In the following test, to maximize the output, some steps are needed to ignore the diseased area according to a certain minimum limit such that it is considered as a healthy leaf part. The use of minimum limits is based on the percentage of leaf area with several parameters, namely 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, and 10%. The results of testing the minimum limit for all test data (120 test data) are shown in figure 7.

Figure 7 is a diagram of the testing results of the minimum limit which reach an optimum conditions when the minimum limit is 3%. Under these conditions, the program can achieve an accuracy of 99.17% with details of 119 correct classifications and 1 incorrect classification. After testing the parameters at a minimum limit of 3%, the accuracy of the program has decreased significantly. Thus, the obtained final analysis can achieve the highest accuracy if it is given a minimum limit of 3%. Meaning, for diseased parts that have a portion of 3% or less in the leaf area is considered a healthy part of the leaf.

5. Conclusion

Based on the results, discussion and data analysis of the implementation of K-Means algorithm as an image segmentation method in the process of identifying orange leaf disease, we can give some conclusions in the followings:

1. The implementation the K-Means algorithm as a method of image segmentation in the process of identifying orange leaf disease should take several steps, namely input image pre-processing which includes the process of resizing, rescaling, and changing the RGB color space to L * a * b *. We also need to do segmentation by using K-Means for the

Figure 6. Results of Healthy Leaf Testing

Figure 7. Results of The Minimum Limit
process of leaf segmentation and disease segmentation, and finally we need to do the classification using the K-NN algorithm.

2. Based on the data testing, we have used several tests, namely testing the Scale Factor value, testing the optimal cluster value, and optimal K testing of K-NN. The results show that the programming has obtained the highest accuracy of 90.83%. For scale factor values will reach the highest value accuracy when it is 1.1. Whereas, the optimal cluster value obtained in the leaf segmentation process is 2 and the optimal cluster value obtained in the process of disease segmentation is 9. Optimal K testing on K-NN has recommended that the most optimal K value when it is 4. Furthermore, the results of program accuracy can be increased by using the minimum limit parameter. Based on testing the minimum limit parameters, the results shows that the optimal value obtained when it is 3% and the accuracy is gained for the identification of the disease when it is 99.17%.

Based on the above conclusion, we can propose the following suggestion:

1. A further research is needed on image processing study, especially in obtaining an information of texture of the image.
2. A further research is needed to carry out the addition of a class of diseases other than those tested diseases: Downy Mildew, Cendawan Jelaga, CVPD.
3. Need to have further research to identify the disease of citrus leaves without picking leaves from the tree.
4. Need to have further research to identify the diseases with more than one leaf in one image frame.

6. Acknowledgement

The author would like to express the sincere gratitude to the department of informatics engineering department of Brawijaya University, Indonesia of year 2018.

Other paragraphs are indented (BodytextIndented style).

References

[1] Agusta Y 2007 K-Means - Application, Problems and Related Methods Journal of Systems and Information 3 47-60.
[2] Andri, Paulus, Wong N P, and Gunawan T 2014 Fruit Segmentation Using the K-Means Clustering Method and Identifying Maturity Using a Comparison of Color Levels. JSM STMIK Mikroskil 15 91-100.
[3] Asana I M., Widiantara I M, Wirastuti N, and Adnyana I B 2017 Contrast Stretching Method for Improving Image Quality in the Video Segmentation Process. Electrical Technology 16 (02) 1-6.
[4] Dhanachandra N, Manglem K, and Chanu Y J 2015 Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm Procedia Computer Science 764-771.
[5] Hapsari Y and Hidayattullah M F 2013 Detecting the Face of Various Human Race Using Skin Color Based on Color Space L * a * b. University Press
[6] Knudsen J 1999 Java 2D Graphics Sebastopol: O’Reilly & Associates.
[7] Mardhiyah A and Harjoko A 2011 Lung and Heart Segmentation Methods in XRay Thorax Image IJEIS 1 (2) 35-44.
[8] Pratt W K 2007 Digital Image Processing (4th ed.). Los Altos, California: A John Wiley & Sons, Inc.
[9] Priandana K, Zulfikar A, and Sukarman 2016 Android-Based Mobile Munsell Soil Color Chart by Using HVC Color Histogram Model with KNN Classification Journal of Computer Science Agri-Informatics 3 (2) 93-101.
[10] Rizal P, Bachrian W, and Retno 2011 Disease-Free Citrus Cultivation East Kalimantan: Agricultural Technology Assessment Center.
[11] Rulaningtyas R, Suksmono A B, Mengko T L, and Saptawati G A 2015 Color Image Segmentation Using PatchBased Clustering Method for Identification of Mycobacterium Tuberculosis Bioscience 17 (1).

[12] Syafiril 2006 Types of Important Pests and Diseases Attacking Koto Tinggi Oranges. Padang: Assessment Center for Agricultural Technology in West Sumatra. From http://sumbar.litbang.pertanian.go.id/images/pdf/hptjeruk

[13] Triwiratno A 2018 Types and Ways to Identify Orange Leaf Diseases.

[14] Wahyuningsih E 2009 CVPD on Oranges (Citrus spp) and Control Efforts. Vis Vitalis 02 (2) 65-73.

[15] Wibowo S A and Usman K 2010 Voice Activity Detection G729B Improvement Technique Using K-Nearest Neighbor Method. International Conference on Distributed Frameworks for Multimedia Applications DFmA.

[16] Wijaya E S and Prayudi Y 2015 Integration of DCS Steganography Methods on Image with Blowfish Cryptography as an Anti Forensic Model for Dual Security of Digital Content. National Seminar on Information Technology Applications (SNATi) D11-D17.

[17] Zhang C, Xiao X, Li X, Chen, Y J., Zhen W, and Chang J 2014 White Blood Cell Segmentation by Color-Space-Based K-Means Clustering. Sensors 14 16128-16147.