KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
Vol.20 (2)(2023), 148-159 p-ISSN: 1693-7554, e-ISSN:2654-3990

Analysis of Tomato Ripeness by Color and Texture Using Cielab and K-Means Clustering

Asep Denih[1,*], Teguh Puja Negara[2], Ismail Marzuki[3]

[1,2,3]Department of Computer Science, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, West Java, 16143, Indonesia

Abstract

Humans have limitations, including in the identification of tomatoes. With the nature of limitations, it makes it difficult for humans to identify the ripeness of tomatoes in large quantities. So far, the selection and determination of the quality activity of tomatoes is carried out manually, resulting in a less uniform product. Manual identification of tomato ripeness has many disadvantages caused by many factors, such as fatigue, lack of motivation, experience, proficiency and so on. This study aims to create a tomato maturity level analysis system based on color and texture using CIELAB and K-Means clustering as a method to determine tomato maturity precisely and accurately. This system displays five images, namely RGB, CIELAB, K-Means clustering, binary and grayscale images, after entering the tomato image, the image will be processed using the five images and the results of extracting characteristics from the tomato will come out. The accuracy rate of tomato ripeness has an average value of 92.70%. The benefit of this research is that it can save time in classifying tomato ripeness and make it easier to determine tomato ripeness based on color.

Keywords: *K-Means; Clustering; color; texture; CIELAB*

1. Introduction

Tomato is one type of horticultural fruit. When growing and developing, tomatoes have a characteristic color. When starting to bear fruit from raw to ripe, there is a change in color in tomatoes, the maturity of tomatoes can be seen from their shape and color, some are green, yellow and red. The color of ripe tomatoes is also affected by storage temperature [1] . So far, the selection and determination of the quality activity of tomatoes is carried out manually, resulting in a less uniform product especially if the identification of tomatoes is done in large quantities. Because the results of manual selection are not satisfactory, a method is needed to select and classify tomatoes properly with maximum light intensity. Development of image processing methods [2] for classifying the ripeness of tomatoes using computers can be applied to identify the ripeness of tomatoes. Work done by computers has advantages compared to humans, because humans have limitations in terms of visual accuracy, energy, and experience. What's more, the development of artificial intelligence has been widely implanted in computers, so that they can be used to assist human work [3]. Image processing which is part of artificial intelligent is a method or techniques

*Corresponding author: ***E-mail adress:*** *asepdenih@unpak.ac.id*
Received: 3 July 2023, Accepted: 15 July 2023 and available online 31 July 2023
https://doi.org/10.33751/komputasi.v20i2.8311

that can be used to process images or images by manipulating them to be desired image data to obtain certain information [4]. Image processing application make processing easy image. Color space system transformation method is one method of image processing done in order to obtain a color space that is variations of an image in the coordinate system certain color [5].

Previous research was conducted by [6], [7]. This research made a Tomato Maturity Classification System Based on Color and Shape Using the Support Vector Machine (SVM) Method. In this study, the method of comparing color levels in red tomato varieties can be used to classify the maturity level of a tomato. This study has weaknesses when adjusting light intensity because it only uses light from sunlight, the accuracy of the level to maturity is less than maximal, only 82.83% accuracy was obtained. And only use 1 image to determine the maturity level of a tomato.Another study was conducted by [8] who built a detection system for papaya using the RGB method. However, the calculations obtained on the system made only reach 50%. The next related research was conducted by [9]. This research made a system for identification of oil Palm fruit maturity based on RGB and HSV colors using the K-Means Clustering Method. This study was able to recognize the object of the image of oil palm fruit based on the level of maturity, namely raw, moderately ripe and ripe. The level of accuracy produced in this research reached 71%.

Based on problems and research that has been done before, ananalysis of the ripeness of tomatoes will be carried out using a MATLAB with light intensity using a 5-WATT LED lamp so that the light results are not too bright or dim when photographing a tomato, the accuracy of the level of tomato ripeness can reach accuracy above 90%, and using 5 images, consisting of RGB images [10], CIELAB Images, Binary Images, Grayscale Images, and Segmented Images. The author made a tomato ripeness analysis using MATLAB 2018, with several color models added, namely CIELAB. The way it works is easy just by entering a tomato image then the results will come out, which is in the form of image processing and data results.

2. Methods

The framework consists of 5 main steps discussed sequentially including problem identification, design formulation and concept,design and drawing analysis, creation of tomato maturity program, and functional test seen in Figure 1. In the system performance test, a ripeness test of tomatoes was carried out which was known to be unripe and ripe as much as 3 each which is called training data. Then, 6 tomatoes were tested for ripeness, which were not known to be ripe or ripe, which is called the test data. To determine the validity of the system, the accuracy calculation is performed.

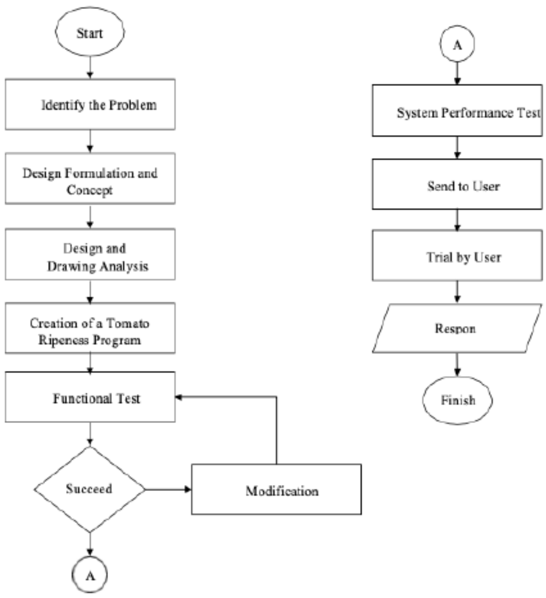


Figure 1. Research Flowchart

2.1. Sample Data Capture

The data used is in the form of images of ripe and unripe tomatoes. Tomato sampling is carried out indoors using a Smartphone camera. Budding is carried out on tomatoes at a distance of 25 cm from the object.Researchersconductedresearchbytakingtomatosamplesassamplestoidentify ripeness. The data sample in this study was 40 images divided into two maturity level classes, namely mature class and raw class.The complexity of the background should be considered with the primary goal of preprocessing to ensure improved image quality which in this process helps make the subsequent identification and recognition phases easier.

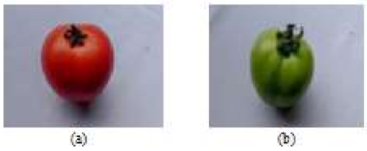


Figure 2. Tomato ripeness: (a) ripe and (b) raw

2.2. Color Feature Extraction

The extraction feature used in this study is the color feature [11]. Therefore, in this study, color became very important in determining the degree of ripeness of tomatoes [12], [13]. There are four color models that will be used in this study namely CIElab, K-means Clustering, Binary Image, Grayscale Image. The feature extraction process starts by converting the RGB image color space (image acquisition) [14] to CIElab, RGB to K-means Clustering, RGB to Binary Image and RGB to Grayscale Image seen in Figure 3 [15].

Figure 3. Color Feature Extraction

2.3. Feature Matching

The K-Means Clustering model [17] is the process of dividing data items into one cluster. K- Means is an unsupervised classification method. In the 30 unsupervised classifications pattern learning about class division is not given, so more focus is on understanding patterns in understandable clusters to find similarities and differences between patterns and to obtain useful conclusions seen in Figure 4.

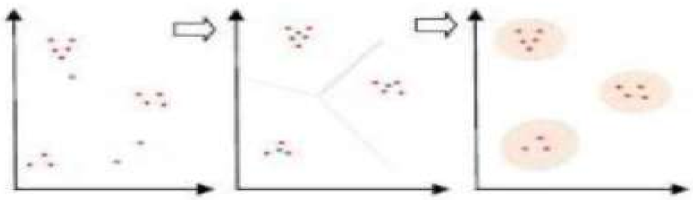


Figure 4. Model K-Means Clustering

2.4. System Design

Broadly speaking, this system has five subsystems including tomato recognition subsystem with CIELAB, tomato recognition with K-Means Clustering, tomato introduction with binary, tomato introduction with grayscale, and tomato characteristic extraction. Basically, the subsystem that is built consists of two component parts, namely software and hardware. Hardware is built using the needs of devices such as laptops, cameras, ring lights, white cloths, and tripods according to what has been described in the needs of tools and materials. The software is made with the MATLAB programming language to be able to run hardware that has been designed according to the system scenario that recognizes tomato ripeness based on color and texture using CIELAB and K-Means Clustering. The display can be seen in Figure 5.

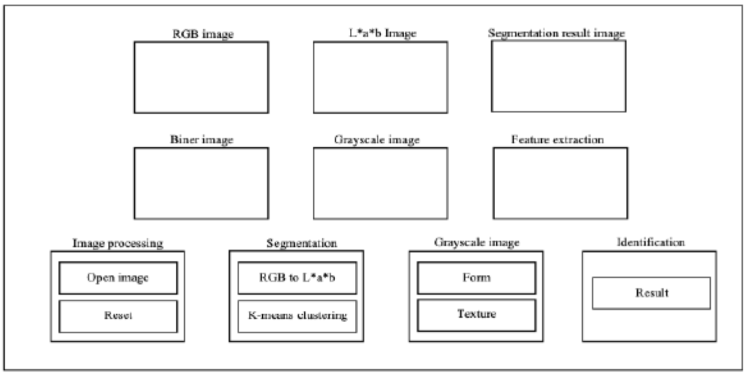


Figure 5. Display Design

2.5. System Optimization

The optimization stage is carried out to improve the performance of the system created. In carrying out the optimization process, researchers will conduct an analysis related to the initial process of the system so that the final results issued by the system, whether it has run well and there are no errors in the system.

3. Result and Discussion

Shooting from each Smartphone is done perpendicular to the tomato, aiming to reduce the shadow effect caused by the tomato. The shooting was done outdoors using a white cloth background so that the image focused on the tomatoes. In this study, the feature extraction process used l*a*b, K-Means Clustering, biner, and grayscale. The test was carried out by performing tomato fruits using uniform lighting seen in Figure 6.

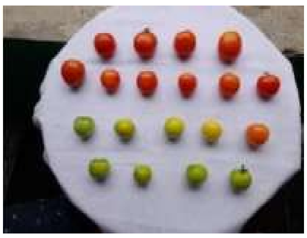


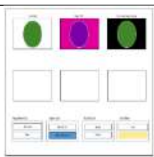
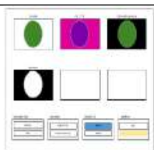
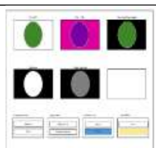
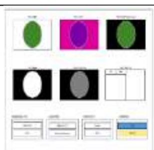
Figure 6. The Process of Shooting Tomatoes

3.1. Structural Testing

Structural testing is carried out to ensure whether the program is running well and as expected. Structural testing tests each view that has been designed by running the view in the application.

This aims to produce the desired results. If an error occurs or the results obtained are not appropriate ,it will go through the process again. The results of structural testing can be seen in Table 1.

Table 1. Structural Testing

No	Appearance	Design	Implementation	Result
1	The view opens on the image			In accordance
2	RGB image display			In accordance
3	RGB segmentation display to L*a*b			In accordance
4	Segmentation result display K-means			In accordance
5	Display of shape feature extraction on Binary Image			In accordance
6	Texture feature extraction display on Grayscale Imagery			In accordance
7	Result identification display			In accordance
8	Reset display			In accordance

3.2. Functional Testing

Functional testing is done by looking at the functions in the system. The results of this functional test can be seen in Table 2.

Table 2. Functional Testing

No	Page	Function	Testing Techniques	Result
1	Open pictures of tomatoes	Open image button	Click the open image button then it will go to the tomato folder and select the tomatoes that will be analyzed for ripeness	In accordance
2	Open image	Open image RGB button	Once selected, a tomato image with an RGB image will appear	In accordance
3	RGB to $L*a*b$	RGB to L*a*b segmentation result button	Click the RGB to $L*a*b$ button to find out the segmentation of the L*a*b image	In accordance
4	K-means clustering	K-means clustering segmentation result button	Click the K-Means Clustering button to view the segmented image	In accordance
5	Shape	Binary image extraction button	Click the shape traits extraction button to view the binary image as well as the traits and values of the metric and eccentricity	In accordance
6	Texture	Grayscale image texture traits extraction button	Click the texture characteristic extraction button and grayscale images will come out and the values of contrast, correlation, energy, and homogeneity	In accordance
7	Result	Identify results button	Click the results button to see the ripeness of the tomato	In accordance
8	Reset	Reset button	Click the reset button to repeat the tomato ripeness analysis again	In accordance

The program can determine the maturity of tomatoes using CIELAB images and also the K-Means Clustering method, as well as other images such as biner, RGB, and grayscale to maximize the ripeness of the tomatoes. The program view is seen in Figure 7

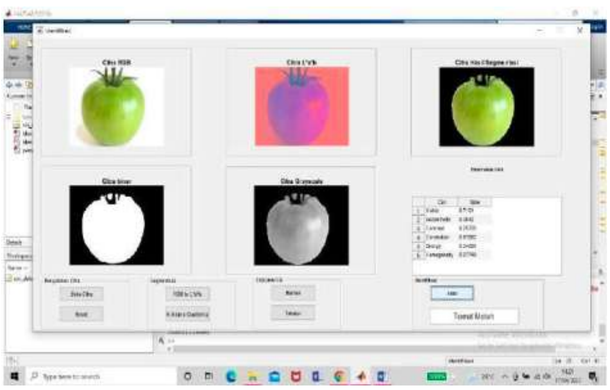


Figure 7. Tomato Ripeness Level Program

The results of classification using the K-Means Clustering Method for the training data are shown in Table 3. From the results of testing tomatoes that are known to be unripe, tested using the K-means cluster method produces the appropriate output.

Table 3. K-Means Clustering Classification Results for Training Data

Image Data	Image Data	System Output Results	Description	Output
Figure 1		Raw	True	Matrix = 0.875 Eccentricity = 0.463 Contrast = 0.011 Correlation = 0.995 Energy = 0.775 Homogeneity = 0.997
Figure 2		Raw	True	Matrix = 0.738 Eccentricity = 0.374 Contrast = 0.020 Correlation = 0.985 Energy = 0.838 Homogeneity = 0.992
Figure 3		Raw	True	Matrix = 0.896 Eccentricity = 0.314 Contrast = 0.008 Correlation = 0,994 Energy = 0.825 Homogeneity = 0.998

The results of classification using the K-Means Clustering Method for the training data are shown in Table 4. From the results of testing tomatoes that are known to be ripe are tested by producing the appropriate output.

Table 4. K-Means Clustering Classification Results for Test Data

Image Data	Image Data	System Output Results	Description	Output
Figure 1		Ripe	True	Matrix = 0,509 Eccentricity = 0,446 Contrast = 0,005 Correlation = 0,995 Energy = 0,859 Homogeneity = 0,998
Figure 2		Ripe	True	Matrix = 0,337 Eccentricity = 0,417 Contrast = 0,009 Correlation = 0,992 Energy = 0,848 Homogeneity = 0,996
Figure 3		Ripe	True	Matrix = 0,336 Eccentricity = 0,393 Contrast = 0,011 Correlation = 0,981 Energy = 0,835 Homogeneity = 0,994

Classifying the maturity level of tomatoes using K-Means Clustering is said to be quite good, it is necessary to test the accuracy and error of the classification of training data and test data for improvement. Here is the calculation process with K-Means Clustering on a feature extraction. There are 4 pieces of image data that will be grouped into 2 clusters, named C1,C2.

Table 5. Feature Extraction Data

Image Data	Metric	Eccentricity	Contrast	Correlation	Energy	Homogeneity
1	0.875	0.463	0.011	0.995	0.775	0.997
2	0.896	0.314	0.088	0.994	0.825	0.998
3	0.738	0.374	0.020	0.985	0.838	0.992
4	0.618	0.563	0.018	0.984	0.798	0.992

First determined the centroid value randomly, then to obtain the minimum distance used Euclidean Distance formula. Known initial value centroid:

$$\begin{array}{lllllll} \text{C1 } 0.875 & 0.463 & 0.011 & 0.995 & 0.775 & 0.997 \rightarrow \text{Data} - 1 \\ \text{C2 } 0.738 & 0.374 & 0.020 & 0.985 & 0.838 & 0.992 \rightarrow \text{Data} - 3 \end{array}$$

Put each document in to the most suitable cluster based on the size of proximity to the centroid.Centroids are vector terms that are considered themed point of the cluster. At this stage, the process of calculating the distance between the data and the center of the cluster is carried out using the Euclidean Distance using the following equation:

$$d(P,Q) = \sqrt{\sum_{i=1}^{n} (x_i(P) - x_i(Q))^2} \tag{1}$$

where is document/data point, is data record, is data centroid. Data distance with initial centroid based on Table 5:

$\text{C}_1 = \sqrt{(0.875 - 0.875)^2 + (0.463 - 0.463)^2 + ... + (0.997 - 0.997)^2} = 0$

$\text{C}_2 = \sqrt{(0.875 - 0.738)^2 + (0.463 - 0.374)^2 + ... + (0.997 - 0.992)^2} = 0.175513$

Data distance with second image centroid:

$\text{C}_1 = \sqrt{(0.896 - 0.875)^2 + (0.314 - 0.463)^2 + ... + (0.998 - 0.997)^2} = 0.158597$

$\text{C}_2 = \sqrt{(0.896 - 0.738)^2 + (0.314 - 0.374)^2 + ... + (0.998 - 0.992)^2} = 0.17076$

Furthermore, the distance of all data is calculated in the same way. The result of the calculation of the overall distance in the first iteration. The data is then grouped according to its cluster, which is the data that has the shortest distance.

Table 6. Iteration 1

Data	C1	C2	Cluster
1	0	0.175	1
2	0.158	0.170	1
3	0.175	0	2
4	0.277	0.227	2

Because 0 is smaller than 0.175, the first data goes to cluster 1. Similarly, the third data, which is 0, is smaller than 0.175, then the third data enters cluster 2. After the data is grouped, then the centroid value must be recalculated to determine the new minimum distance, along with the centroid calculation.

Since iteration 1 and iteration 2 are unchanged, the iteration calculation is complete. Based on training 50 times for training data, 47 correct data were obtained so that the accuracy rate produced was94%, while in testing 25 times the test data obtained 23correct data so that the accuracy level was 92%. Thus, the total accuracy of K-Means Clustering used to classify the maturity level of tomatoes on test and training data is 93%.

$$C1 = \begin{bmatrix} \frac{0.875+0.896}{2} \\ \frac{0.463+0.314}{2} \\ \frac{0.011+0.008}{2} \\ \frac{0.995+0.994}{2} \\ \frac{0.775+0.825}{2} \\ \frac{0.997+0.998}{2} \end{bmatrix} = \begin{bmatrix} 0.885 \\ 0.388 \\ 0.009 \\ 0.095 \\ 0.800 \\ 0.997 \end{bmatrix} \qquad C2 = \begin{bmatrix} \frac{0.738+0.618}{2} \\ \frac{0.374+0.563}{2} \\ \frac{0.020+0.018}{2} \\ \frac{0.985+0.984}{2} \\ \frac{0.838+0.798}{2} \\ \frac{0.992+0.992}{2} \end{bmatrix} = \begin{bmatrix} 0.678 \\ 0.468 \\ 0.019 \\ 0.985 \\ 0.818 \\ 0.992 \end{bmatrix}$$

Table 7. Iteration 2

Data	C1	C2	Iteration 1	Iteration 2
1	0.079	0.202	1	1
2	0.079	0.267	1	1
3	0.153	0.113	2	2
4	0.319	0.113	2	2

3.3. Validation Test

Validation is a test to determine the accuracy of the image results that have been entered in to the tomato matlab system.

Table 8. Validation Testing

Figure	Class	K-Means Clustering	CIELAB	Biner	Grayscale
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1
7	1	1	1	1	1
8	1	1	1	1	1
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	2	2	1	1	1
14	2	2	1	1	1
15	2	2	2	2	2
16	2	2	2	2	2
17	2	2	2	2	2
18	2	2	2	2	2
19	2	2	2	2	2
20	2	2	2	2	2

The type of ripeness of tomatoes is written in class, if the number 1 is raw and if the number 2 is ripe. 1 That is the information that the process is wrong. From Table 8. There is a clear difference in the results of the validation accuracy of each color model. This proves the proper use of color models is necessary in determining the level of accuracy of the system. The accuracy percentage results are shown in table 12. K-Means Clustering delivers 90% accuracy rates, CIElab by 100%, Binary by 90% and Grayscale by 90%.

Table 9. Accuracy Percentage

Color Model and Segmentation	True	False	Accuracy
K-Means Clustering	18	2	90%
CIELAB	20	0	100%
Biner	18	2	90%
Grayscale	18	2	90%

Table 10. Validation Testing with Smartphone's

Figure	Class	Megapixel	Shutter	ISO	HDR
1	1	v	v	v	v
2	1	v	v	v	v
3	1	v	v	v	v
4	1	v	v	v	v
5	1	v	v	v	v
6	2	v	v	v	v
7	2	v	v	v	v
8	2	v	v	v	v
9	2	v	v	v	v
10	2	v	v	v	v

The type of ripeness of tomatoes is written in class, if the number 1 is raw and if the number 2 is ripe. v indicates that the settings of the camera are functioning properly and accurately in shooting. From Table 14.There is a clear difference in the results of the accuracy of the color model and the clarity of the image. This proves that proper use of color models and camera sensors is necessary in determining the degree of curvature of shooting an object.

4. Conclusion

The utilization and development of MATLAB 2018 has not been widely used in agriculture and plantations so research needs to be done to prove that MATLAB 2018 has many abilities, especially to determine the maturity of tomatoes. This research has several stages of the process including sample data collection, color feature extraction, feature matching and testing process. Tests carried out include structural, functional, k-means clustering, and validation trials. Based on the results of the research conducted, it can be concluded that the process of identifying the maturity of tomatoes using the K-Means Clustering method is able to recognize the object of tomato image based on the level of ripeness, namely raw and ripe. The results of the identification of tomato ripeness obtained an accuracy level for training data of 94% and an accuracy level for test data of 92%, so that the total accuracy level of both is 93%.

5. Acknowledgement

Thank you to the Department of Mathematics for providing support and input in this research. Also, thank you very much to the Faculty of Mathematics and Natural Sciences (FMIPA) which has funded this research with the PNBP 2021 funding source, University of Riau.

References

[1] Takashi, N., Maki, H., Nishina, H., Takayama, K. (2013). Evaluation of Tomato Fruit Color Change with Different Maturity Stages and Storage Temperatures using Image Analysis. *IFAC Proceedings Volumes*, 46(4), 147-149. doi: https://doi.org/10.3182/20130327-3-JP-3017.00034.

[2] [2] Mukhofifah,, M., Nurraharjo, E. (2019). Sistem Deteksi Kematangan Buah Alpukat Menggunakan Metode Pengolahan Citra. *Jurnal Dinamika Informatika*, 11(1), 12–23. doi: https://doi.org/10.35315/informatika.v11i1.8144.

[3] Mohammad, S. M. (2020). Artificial Intelligence In Information Technology. *International Journal of Innovations In Engineering Research and Technology (IJIERT)*, 7(6), 168-175. doi: http://dx.doi.org/10.2139/ssrn.3625444

[4] Jatmika, S., Purnamasari. (2014). Rancang Bangun Alat Pendeteksi Kematangan Buah Apel Dengan Menggunakan Metode Image Processing Berdasarkan Komposisi Warna. *Jurnal Ilmiah Teknologi dan Informasi ASIA*. 8(1), 1-43. Available at: https://jurnal.stmikasia.ac.id/index.php/jitika/article/view/128

[5] Astrianda, N. Klasifikasi Kematangan Buah Tomat Dengan Variasi Model Warna Menggunakan Support Vector Machine. (2020). *VOCATECH Vocational Education Technology Journal*. 1(2), 45–52. doi: https://doi.org/10.38038/vocatech.v1i2.27.

[6] Arief, M. (2019). Klasifikasi Kematangan Buah Jeruk Berdasarkan Fitur Warna Menggunakan Metode SVM. *Journal Ilmu Komputer dan Desain Komun*. Vol. 4(1), 9–16. Available at: https://journal.unusida.ac.id/index.php/jik/article/view/53

[7] Widyasari K., Rosiani, U. D., Pramudhita, A.N. (2021). Implementasi Sistem Pendeteksi Tingkat Kematangan Buah Pepaya Menggunakan Metode RGB. SMATIKA Jurnal. 11(01), 32-36. Available at: https://jurnal.stiki.ac.id/index.php/SMATIKA/article/download/536/396/

[8] Saikin, S., Fadli, S., Ashari. (2021). Optimization of Support Vector Machine Method using Feature Selection to Improve Classification Results. *JISA (Jurnal Inform. dan Sains)*. 4(1), 22–27. doi: https://doi.org/10.31326/jisa.v4i1.881

[9] Aditya, A., Jovian, I., Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *J. MEDIA Inform.* BUDIDARMA. 4(1). doi: https://doi.org/10.30865/mib.v4i1.1784.

[10] Himmah, E. F., Widyaningsih, M., Maysaroh, M. (2020). Identifikasi Kematangan Buah Kelapa Sawit Berdasarkan Warna RGB Dan HSV Menggunakan Metode K-Means Clustering. *J. Sains dan Inform.* vol. 6, no. 2, 193–202. doi: https://doi.org/10.34128/jsi.v6i2.242

[11] Abdullah, M. N et al. (2021). Colour Features Extraction Techniques and Approaches for Content-Based Image Retrieval (CBIR) System. *J. Mater. Sci. Chem. Eng.* 09(07): 29–34. doi: https://doi.org/10.4236/msce.2021.97003.

[12] Akishin, O. V., Gudkovsky, V. A., Sutormina, A. V., Potapova, A. A., Medelyaeva, A. Y. (2021). Development of a new method for determining the degree of ripeness of tomato fruits with different colors of ripe fruits. *IOP Conf. Ser. Earth Environ. Sci.*. 845(1), 012011. doi: https://doi.org/10.1088/1755-1315/845/1/012011

[13] Ladipi, I. A. (2022). Application of Image Processing to Determine the Tomato Fruit's Ripeness. *Sriwij. J. Environ.*, 7(2), 80–90. doi: https://doi.org/10.22135/sje.2022.7.2.80-90.

[14] Sinaga, A. S. (2019). Segmentasi Ruang Warna L*a*b. J. Mantik Penusa. 33(1), 43– 46. Available at: https://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/562/336

[15] Sanusi, S. H. S., Susetianingtias, D. T. (2019). Pembuatan Aplikasi Klasifikasi Citra Daun Menggunakan Ruang Warna RGB dan HSV. *J. Ilm. Inform. Komput.* 24(3), 180–190. doi: https://doi.org/10.35760/ik.2019.v24i3.2323.

[16] S. G. Kellison, "The theory of interest," Richard D. Irwin Inc, Homewood, 1991.

[17] Putra, I. K. N. (2023). Implementasi Metode K-Means Clustering pada Ekspor Kopi Berdasarkan Negara Tujuan. Jurnal Ilmu Komputer dan Bisnis. XIV(1a), 45–55. Available at: https://ojs.stmikdharmapalariau.ac.id/index.php/jikb/article/view/332/373.

[18] Prayoga Y., Tambunan, H. S., Parlina I. (2019). Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means. BRAHMANA J. Penerapan Kecerdasan Buatan. 1(1), 24–30. doi: https://doi.org/10.30645/brahmana.v1i1.4.