Published online on: **https://journal.ilmudata.co.id/index.php/RIGGS**

Journal of Artificial Intelligence and Digital Business

(RIGGS)

Vol. 1 No. 1 (2022) 25 - 30 e-ISSN: 2963-914X

Classification Of Tomato Maturity Levels Based on RGB And HSV Colors Using KNN Algorithm

Lidya Ningsih[1], Putri Cholidhazia[2]
[1]Ilmu Komputer, Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor
[2]Department Informastika, Universitas Al Azhar Indonesia
[1]lidyaningsih8@gmail.com*, [2]putri_cho@apps.ipb.ac.id

Abstract

Tomatoes (Lycopersiconeculentum Mill) are vegetables that are widely produced in tropical and subtropic areas. According to (Harllee) tomatoes are grouped into 6 levels of maturity, namely green, breakers, turning, pink, light red, and red. One way that can be used to classify the level of maturity of tomatoes in the field of informatics is to utilize digital image processing techniques. This study classifies the maturity of tomatoes using K-Nearest Neighbor (KNN) based on the Red Green Blue and Hue Saturation Value color features. The KNN algorithm was chosen as a classification algorithm because KNN is quite simple with good accuracy based on the minimum distance using Euclidean Distance. The research conducted received the highest accuracy result of 91.25% at the value of K = 7 with the test data 80. This shows that the KNN algorithm successfully classified the maturity of tomatoes by utilizing the color image of RGB and HSV.

Keywords: *Tomatoes, K-Nearest Neighbor, Euclidean Distance, Red Green Blue (RGB), Hue Saturation Value (HSV*

1. Introduction

Tomatoes (Lycopersiconeculentum Mill) are vegetables that are widely produced in tropical and subtropical areas [1]. Tomato plants are horticultural commodities that are needed by the community and become a basic need in Indonesia. However, in the industry and tomato farmers themselves when detecting tomato maturity is still done manually, namely by visual observation directly on the fruit [2].

Manual observations produce a level of maturity that is less uniform and unsatisfactory [3]. This process is very dependent on the subjectivity of officers when sorting tomatoes. This manual observation requires a long time and the products produced are also very diverse. This is due to human visual limitations, fatigue when working, and differences of opinion about the quality of the fruit [4].

Because the manual method has many weaknesses, so it takes a method that can choose and classify the level of maturity of tomatoes well. This process is carried out to reduce the risk of rotten to the tomatoes [5]. There are several factors that can be used as guidelines in seeing the level of maturity of tomatoes, including from the size, shape, texture, and color. Color is the most easily used characteristic in seeing the level of tomato maturity [3]. According to (Harllee) tomatoes are grouped into 6 levels of maturity, namely Green, Breakers, Turning, Pink, Light Red, and Red [5]. One way that can be used to classify the level of maturity of tomatoes in the field of informatics is to utilize digital image processing techniques. Digital image processing techniques are used because digital images are able to choose agricultural products automatically [4]. So that it can reduce the risk of rotten in tomatoes.

Research on the classification of tomato maturity levels has been carried out by [6]. The study used the HSV algorithm as a color feature and LVQ algorithm as classification. The study used tomato data set from one side and got an accuracy of 83.75%. Based on research related to the classification of the maturity level of tomatoes, research is needed to be carried out on the four sides of tomatoes. This is because not all parts of the tomatoes have the same color.

Research [7] Comparing the Hue Saturation Intensity (HSI) and Hue Saturation Value (HSV) (HSV) color features in detecting rose flowers. HSV color features get better accuracy compared to HSI. Research [8] Classifying the image of beef and pork using KNN get the percentage of accuracy of 93.33%.

Based on the problems and related research that has been explained, this study classifies the level of tomato maturity using K-Nearest Neighbor (KNN) based on

Accepted: 29-07-2022 | Received in revised: 30-07-2022 | Published: 31-07-2022

RGB and HSV color features. The KNN algorithm was chosen as a classification algorithm because KNN is quite simple with good accuracy based on the minimum distance using Euclidean Distance. The algorithm used is expected to be able to classify the maturity of tomatoes so that it can reduce the problem of spoilage tomatoes and get better results from previous studies.

2. Research Methods

This study proposes a strategy to see the level of tomato maturity.

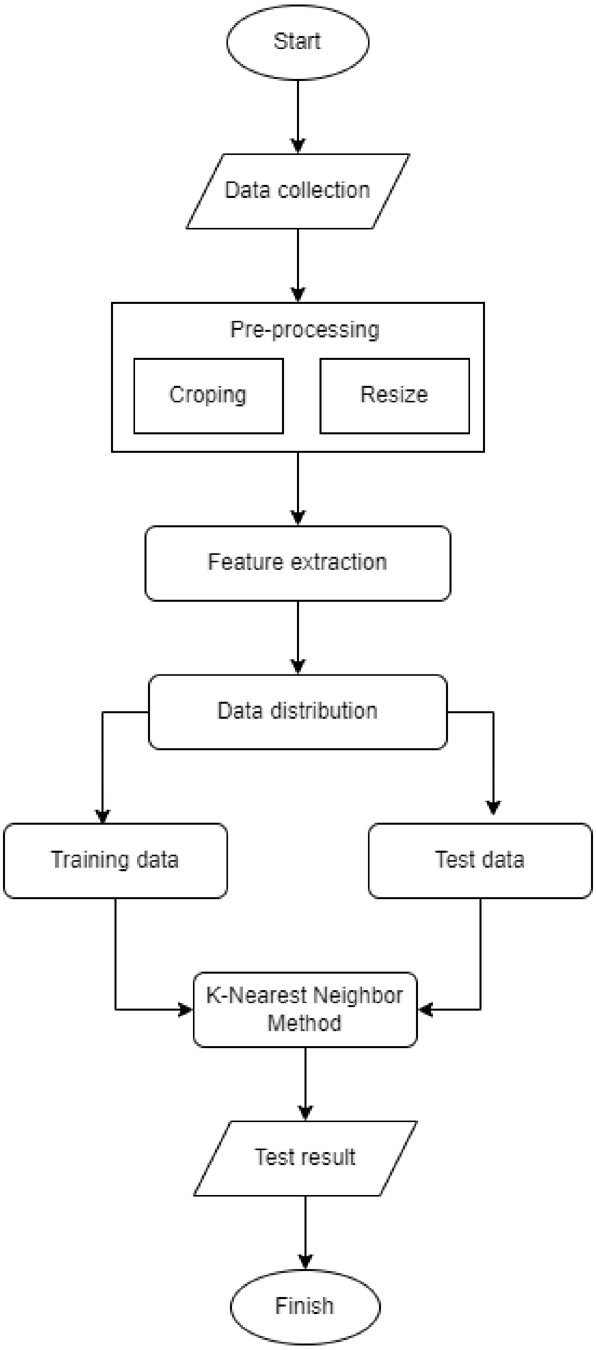

Figure 1 Research Stages

The level of tomato maturity based on the red, green, and blue (RGB) color features and from the Hue Saturation Value (HSV) from the tomatoes is classified using the K-Nearest Neighbor (KNN) algorithm.

The classification process of tomato maturity levels consists of training and testing. The training process is used to build and train a model of the image data used, the testing process is used to see the success rate of the model built.

Research conducted consists of preprocessing stages, feature extraction, modeling and evaluation. Preprocessing stage is done to prepare image data by removing background and uniforming image pixel size. The pre-processing stage conducted in this study is cropping and resize.

The feature extraction stage in this study uses color features consisting of RGB color features and HSV color features. The feature extraction process is done to get the features needed from a image. The feature value obtained from the process of extraction of color features is used as input in the classification process. The classification process used in research utilizes the Machine Learning technique using the K-Nearest Neighbor (KNN) algorithm.

2.1. Image data

The data used in this study is the image of tomato fruit classified in 5 class levels representing 5 levels of maturity, namely Green, Turning, Pink, Light Red, and Red according to Figure 2.

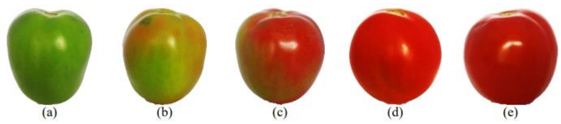

Figure 2 The level of maturity of tomatoes (a) green, (b) turning, (c) pink (d) light red, (e) red. [9]

The maturity level of tomato breakers in Figure 2 is combined with the Green class because the breakers class is more dominant in dark green, and only 10% contains a brownish yellow color on its surface [10]. The data was taken from research [9] using plum tomatoes with image acquisition using a 24.3 megapixel DSLR camera. The image data format in the study is PNG.

Collecting image data using a white background by positioning the image object in the middle. The data used in this study amounted to 400 images.

Table 1 Tomato Image Data Distribution

No	Atribut	Jumlah Citra
1	*Green*	80
2	*Turning*	80
3	*Pink*	80
4	*Light Red*	80
5	*Red*	80
	Total	**400**

DOI: https://doi.org/10.31004/riggs.v1i1.10

2.2. Preprocessing

After obtaining tomato image data, then preprocessing data is carried out to prepare data in accordance with the research needs. The initial stage of preprocessing data conducted in the study was cropping according to Figure 2.

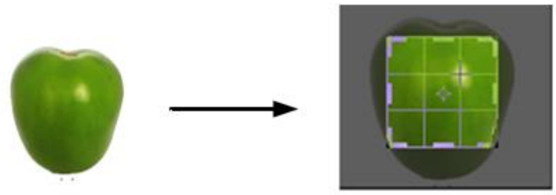

Figure 3 Cropping Process

The cropping process is carried out to facilitate the system in processing the image used by taking the object needed and removing the background in the image. The results of the cropping process in this study are square. Cropping is done manually using the Photoshop CS6 application.

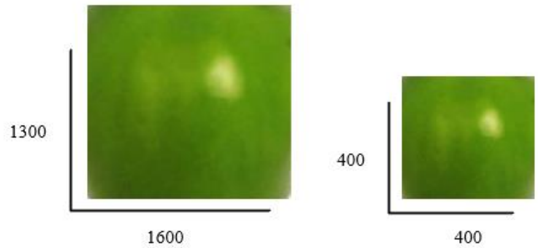

Figure 4 Resize Process

The next step is the resize process according to Figure 4. The resize process is done by changing the size of the image pixel according to the desired size in the study. This study uses a 400x400 pixel resize size.

2.3. RGB color space

RGB is a color space resulting from the acquisition of color frequency by an electronic sensor in the form of analog signals. The RGB color space consists of 3 basic colors, red, green and blue (Figure 4) [11]. Of the three basic colors, 224 or 16,777,216 colors can be formed [12].

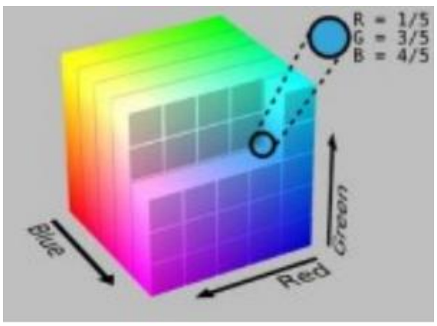

Figure 5 RGB Color Room [13]

Figure 5 can be seen a combination of RGB color space. The combination of red and green colors produces yellow, red and blue combination produces purple. The combined blue and green colors produce cyan colors. While the combination of red, green, and blue produces white when it has the same intensity, which is 255. The lower the intensity value of the three colors will produce a gray color from bright to dark (gray level) to the black color when the three colors value This is the same as zero [12].

2.4. Hue Saturation Value (HSV)

The HSV color model is a derivative of the RGB color model [14], but the HSV color model is better than the RGB color space. This is because HSV can express color shadows, color hue, color degree and color contrast [15].

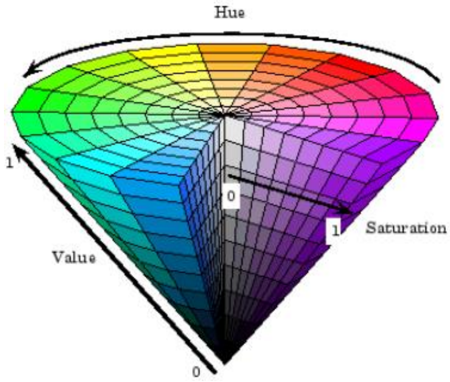

Figure 6 HSV Color Room [13]

The HSV color model has 3 main components [16], [17] which can be seen in Figure 6 based on the following information.

1. Hue represents the basic color that has a range of 0 to 360 ° according to Figure 4. Point 0 is a color that varies from red, yellow, green, cyan, blue and magenta then return to red.
2. Saturation represents the level of purity or strength in a color that has a range of 0 to 1. The value of 0 here is a color that is nuanced gray until there is no white component.
3. Value or referred to as brightness represents how dark or how bright the color is. Value has a range of values of 0 to 100%. The value of 0 represents the black color, and the higher the value the brighter color.

The HSV color model was first introduced by A.R. Smith in 1978. According to [18] HSV values can be converted according to equation (1), (2), and (3).

$$H = arctan\left\{\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right\} \quad (1)$$

$$S = 1 - \frac{\min(R, G, B)}{V} \quad (2)$$

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

DOI: https://doi.org/10.31004/riggs.v1i1.10

However, the value of H cannot be represented if S = 0. So the RGB normalization process needs to be carried out in accordance with equation (4), (5), (6).

$$r = \frac{R}{R+G+B} \tag{4}$$

$$g = \frac{G}{R+G+B} \tag{5}$$

$$b = \frac{B}{R+G+B} \tag{6}$$

After the RGB normalization process is carried out, the RGB to HSV conversion process is then carried out using equation (7), (8), (9), (10).

$$v = \max(r, g, b) \tag{7}$$

$$s = \begin{cases} 0 \\ v \end{cases} \tag{8}$$

$$H = \begin{cases} 0 & Jika\ s = 0 \\ 60 \times (-g - b) & jika\ v = r \\ 60 \times \left[2 + \frac{b-r}{s \times v}\right] & jika\ v = g \\ 60 \times \left[4 + \frac{r-g}{s \times v}\right] & jika\ v = b \end{cases} \tag{9}$$

$$H = H + 360 \qquad Jika\ H < 0 \tag{10}$$

Where V is the maximum value (R, G, B), S is the saturation value, H is the Hue value.

2.5. K-Nearest Neighbor

K-Nearest Neighbor (KNN) is an algorithm that is often used as a classification. KNN is a supervised learning algorithm by storing training data and comparing data that has not been classified in the training data [19]. The KNN algorithm is one of the non -metric methods in the recognition of patterns. This algorithm groups objects based on the closest features by finding the closest distance between data and neighboring values (K) [20]. The following are the stages of the KNN algorithm: [21]

1. Determine the value k. This study uses 10 kinds of K values (K = 1,3,5,7,9,11,13,15,15,17 and 19).
2. Calculate the distance using Euclidean Distance for each object to new data according to equation (11) [22].

$$euc = \sqrt{\left(\sum_{i=1}^{n}(pi - qi)^2\right)} \tag{11}$$

   The PI value is a training data with Qi Data Testing, i is a data variable and n is the dimension of the data.
3. Sorting the object based on the minimum distance according to the value k.
4. Adjust the Y class label to the settings that have been set.
5. Looking for the number of classes from the closest prudence value as a basis for determining the class of new data.

3. Results and Discussions

Experimental testing conducted in the study according to table 2.

Table 2 Testing Parameters

No.	Experimental testing	Parameter configuration
1	*Distribution of training data and test data*	The percentage of training data distribution and test data used in this study consisted of 5 types, namely 50:50, 60:40, 70:30, 80:20, and 90:10
2	Value k	The process of determining the distance between the relationship between the KNN algorithm. The value of K used in the study (1, 3, 5, 7, 9, 11, 13, 15, 17, 19).

The amount of data sets used in research 400 image data taken from 100 tomato images. This study uses Plum type tomatoes. To determine the level of maturity of tomatoes can use the extraction of color features. This is because the color of tomatoes is a very important factor in determining the level of maturity of the tomatoes. Extraction of the color features used is RGB and HSV. Extraction of RGB color features can be rated in Table 3.

Table 3 RGB color features

No.	**R**	**G**	**B**	**Label**
1.	0.49888	0.62067	0.28676	Green
2.	0.54057	0.64674	0.30178	Green
3.	0.48959	0.59552	0.31744	Green
4.	0.56043	0.65498	0.29296	Green
5.	0.52990	0.63966	0.28686	Green
		………	………	
81.	0.89929	0.38290	0.28013	Light Red
82.	0.81320	0.31478	0.26132	Light Red
83.	0.83032	0.30208	0.24463	Light Red
84.	0.88763	0.37773	0.257583	Light Red
85.	0.86784	0.35998	0.27433	Light Red
		………	………	
161.	0.80195	0.58022	0.33631	Pink
162.	0.79268	0.61908	0.33681	Pink
163.	0.77848	0.63888	0.30678	Pink
164.	0.81901	0.58539	0.34599	Pink
165.	0.805200	0.48772	0.32426	Pink
		……….	……….	
241.	0.64874	0.300827	0.26792	Red
242.	0.65871	0.27917	0.24715	Red
243.	0.71053	0.32585	0.28386	Red
244.	0.69099	0.32041	0.28912	Red
245.	0.65696	0.30244	0.27280	Red
		………..	……….	
396	0.71081	0.59732	0.28763	Turning
397	0.71799	0.63201	0.29956	Turning
398	0.58780	0.65584	0.33008	Turning
399	0.72743	0.59783	0.31741	Turning
400	0.69254	0.57409	0.25044	Turning

DOI: https://doi.org/10.31004/riggs.v1i1.10

After the feature extraction process is carried out, the testing process is then carried out using tomato image data based on HSV color extraction. The distribution of training data and test data is carried out with a percentage of division that is appropriate in Table 2. Testing using the KNN algorithm with several test scenarios, namely the distribution of data and the value of the Determination (K). The test results are determined by confusion matrix to calculate accuracy [23]. The results of testing the model carried out for the overall percentage of data distribution and K value can be seen in Figure 7.

Based on the results of the model testing using the KNN algorithm based on the color feature for the classification of tomato maturity levels in Figure 8. The highest accuracy results are located at the progress value (K = 7) with a percentage of accuracy of 91.25% with the amount of 80 data test data. The test scenario conducted at the highest accuracy can be seen using confusion matrix in Table 5.

Table 5 Confusion Matrix

		Prediction Class				
		Green	**Light Red**	**Pink**	**Red**	**Turning**
Actual Class	**Green**	10	0	0	2	0
	Light Red	0	14	0	0	1
	Pink	4	1	8	0	0
	Red	1	0	0	21	0
	Turning	0	0	0	0	18

Based on the test matrix shown in Table 5. The level of accuracy can be formulated using confusion matrix with equation (12).

$$Accuracy = \frac{10 + 14 + 8 + 21 + 18}{400} x100\% = 91,25 \quad (12)$$

4. Conclusion

Research conducted using 5 types of tomato maturity levels with 400 images consisting of Green, Turning, Pink, Light Red, and Red. The testing conducted in the study consisted of testing the percentage of data distribution and testing of the KNN parameter, namely the value of the progress (K). The research conducted obtained the highest accuracy result of 91.25% at the value of K = 7 with the 80 test data. The results of the accuracy obtained were quite good, but there were still classification errors because there were still tomato images that had a reflection of light with different intensity . This can cause errors in classification. Further research to be able to pay attention to the quality of the image of tomatoes by minimizing the reflection of light when taking pictures.

Reference

[1] R. Pratama *et al.*, "DETEKSI KEMATANGAN BUAH TOMAT BERDASARKAN FITUR WARNA MENGGUNAKAN METODE TRANSFORMASI RUANG WARNA HIS," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 2, pp. 81–86, 2019, doi: 10.33387/jiko p-ISSN:

[2] S. Aprilisa and Sukemi, "Klasifikasi Tingkat Kematangan Buah Tomat Berdasarkan Fitur Warna Menggunakan K-Nearest Neighhbor," in *Prosiding Annual Research Seminar*, 2019, vol. 5, no. 1, pp. 978–979.

[3] N. Astrianda, "Klasifikasi Kematangan Buah Tomat Dengan Variasi Model Warna Menggunakan Support Vector Machine," *VOCATECH Vocat. Educ. Technol. J.*, vol. 1, no. 2, pp. 45–52, 2020, doi: 10.38038/vocatech.v1i2.27.

[4] S. Kusumaningtyas and R. A. Asmara, "Identifikasi Kematangan Buah Tomat Berdasarkan Warna

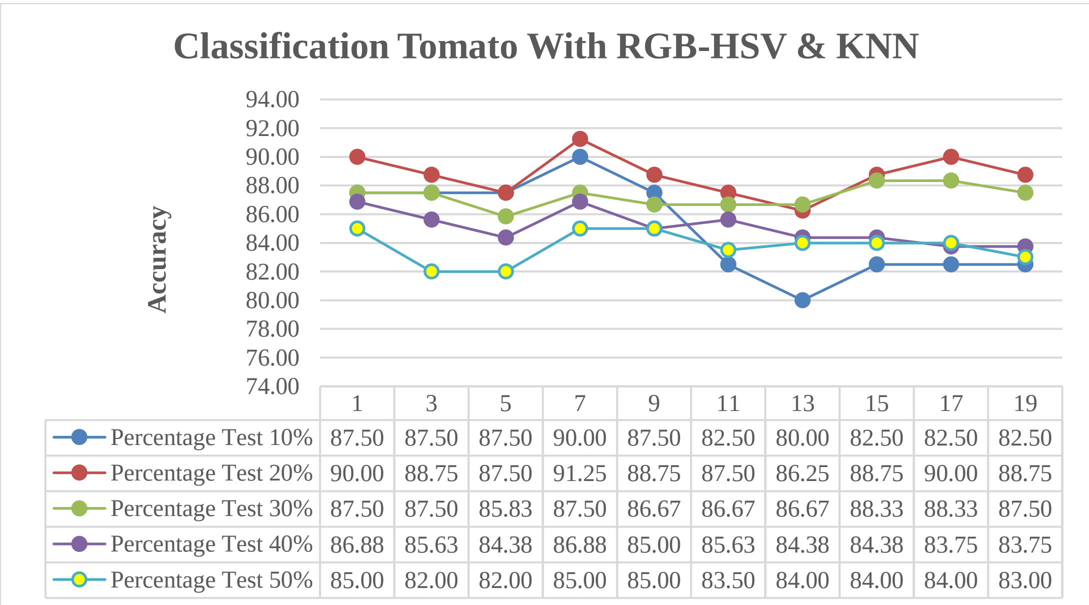

Figure 7 Model Testing Results

DOI: https://doi.org/10.31004/riggs.v1i1.10

Menggunakan Metode Jaringan Syaraf Tiruan (Jst)," *J. Inform. Polinema*, vol. 2, no. 2, p. 72, 2016, doi: 10.33795/jip.v2i2.59.

[5] S. Y. Riska and P. Subekti, "Klasifikasi Level Kematangan Buah Tomat Berdasarkan Fitur Warna Menggunakan Multi-Svm," *J. Ilm. Inform.*, vol. 1, no. 1, pp. 39–45, 2016, doi: 10.35316/jimi.v1i1.442.

[6] M. A. Anggriawan, M. Ichwan, and D. B. Utami, "Pengenalan Tingkat Kematangan Tomat Berdasarkan Citra Warna Pada Studi Kasus Pembangunan Sistem Pemilihan Otomatis," *J. Tek. Inform. dan Sist. Inf.*, vol. 3, no. 3, pp. 550–564, 2017, doi: 10.28932/jutisi.v3i3.688.

[7] D. Wandi, F. Fauziah, and N. Hayati, "Deteksi Kelayuan Pada Bunga Mawar dengan Metode Transformasi Ruang Warna HSI Dan HSV," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 5, no. 3, p. 333, 2021, doi: 10.30998/string.v5i3.8464.

[8] E. Budianita, J. Jasril, and L. Handayani, "Implementasi Pengolahan Citra dan Klasifikasi K-Nearest Neighbour Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi Berbasis Web," *J. Sains dan Teknol. Ind.*, vol. 12, no. Vol 12, No 2 (2015): Juni 2015, pp. 242–247, 2015, [Online]. Available: http://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/1005

[9] S. Sanjaya, M. L. Pura, S. K. Gusti, F. Yanto, and F. Syafria, "K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 101, 2019, doi: 10.24014/ijaidm.v2i2.7975.

[10] USDA, "United States Standards for Grades of Fresh Tomatoes," *United States Dep. Agric.*, vol. 1991, no. January, pp. 1–13, 1991.

[11] A. M. Priyatno, "The Application of HAAR Wavelet and Backpropagation for Diabetic Retinopathy Classification Based on Eye Retina Image," *Int. J. Sci. Eng. Inf. Technol.*, vol. 03, no. 02, pp. 139–142, 2019, [Online]. Available: https://journal.trunojoyo.ac.id/ijseit/article/view/4536

[12] H. Prabowo, "Deteksi Kondisi Kematangan Buah Jeruk Berdasarkan Kemiripan Warna Pada Ruang Warna RGB Berbasis Android," *J. Elektron. Sist. Inf. dan Komput.*, vol. 3, no. 2, pp. 9–19, 2017.

[13] L. Farokhah, "Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1129, 2020, doi: 10.25126/jtiik.2020722608.

[14] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, "Human Skin Detection Using RGB, HSV and YCbCr Color Models," in *Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016)*, 2017, vol. 137, pp. 324–332. doi: 10.2991/iccasp-16.2017.51.

[15] P. LI, Y. HUANG, and K. YAO, "Multi-algorithm Fusion of RGB and HSV Color Spaces for Image Enhancement," in *2018 37th Chinese Control Conference (CCC)*, Jul. 2018, vol. 2018-July, pp. 9584–9589. doi: 10.23919/ChiCC.2018.8483674.

[16] M. H. Malik, T. Zhang, H. Li, M. Zhang, S. Shabbir, and A. Saeed, "Mature Tomato Fruit Detection Algorithm Based on improved HSV and Watershed Algorithm," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 431–436, 2018, doi: 10.1016/j.ifacol.2018.08.183.

[17] A. M. Priyatno, F. M. Putra, P. Cholidhazia, and L. Ningsih, "Combination of extraction features based on texture and colour feature for beef and pork classification," *J. Phys. Conf. Ser.*, vol. 1563, no. 1, p. 012007, Jun. 2020, doi: 10.1088/1742-6596/1563/1/012007.

[18] C. Chen, A. Dantcheva, and A. Ross, "Automatic facial makeup detection with application in face recognition," in *2013 International Conference on Biometrics (ICB)*, Jun. 2013, no. April 2018, pp. 1–8. doi: 10.1109/ICB.2013.6612994.

[19] M. Reza Noviansyah, T. Rismawan, D. Marisa Midyanti, J. Sistem Komputer, and F. H. MIPA Universitas Tanjungpura Jl Hadari Nawawi, "Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya)," *J. Coding, Sist. Komput. Untan*, vol. 06, no. 2, pp. 48–56, 2018.

[20] T. Setiyorini and R. T. Asmono, "PENERAPAN METODE K-NEAREST NEIGHBOR DAN INFORMATION GAIN PADA KLASIFIKASI KINERJA SISWA," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 7–14, Jun. 2019, doi: 10.33480/jitk.v5i1.613.

[21] F. Shidiq, "Penerapan Metode K-Nearest Neighbor (KNN) Untuk Menentukan Ikan Cupang Dengan Ekstraksi Fitur Ciri Bentuk Dan Canny," *Innov. Res. Informatics*, vol. 3, no. 2, pp. 39–46, 2021, doi: 10.37058/innovatics.v3i2.3093.

[22] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.

[23] A. M. Priyatno, "Spammer Detection Based on Account, Tweet, and Community Activity on Twitter," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp. 97–107, Jul. 2020, doi: 10.21609/jiki.v13i2.871.

DOI: https://doi.org/10.31004/riggs.v1i1.10