



# Classification of Banana Maturity Levels Based on Skin Image with HSI Color Space Transformation Features Using the K-NN Method

Adhe Irham Thoriq<sup>(1)</sup>, Muhamad Haris Zuhri<sup>(2)</sup>, Purwanto<sup>(3)</sup>,  
Pujiono<sup>(4)</sup>, Heru Agus Santoso<sup>(5)</sup>

*Universitas Dian Nuswantoro, Indonesia*

E-mail: <sup>(1)</sup>[adheirham90@gmail.com](mailto:adheirham90@gmail.com), <sup>(2)</sup>[hariezzouhry@gmail.com](mailto:hariezzouhry@gmail.com),  
<sup>(3)</sup>[purwanto@dsn.dinus.ac.id](mailto:purwanto@dsn.dinus.ac.id), <sup>(4)</sup>[pujiono@dsn.dinus.ac.id](mailto:pujiono@dsn.dinus.ac.id),  
<sup>(5)</sup>[heru.agus.santoso@dsn.dinus.ac.id](mailto:heru.agus.santoso@dsn.dinus.ac.id)

Received: 12 February 2022; Revised: 20 May 2022; Accepted: 25 May 2022

## Abstract

Banana fruit has a Latin name is *Musa Paradisiaca* is one type of fruit that is often found in Southeast Asia, even in the world. There are many types of bananas, the most popular of which is the Raja banana (*Musa paradisiaca* L.). The advantage of the plantain is that it has a fragrant aroma and is of medium size and has a very sweet taste that is appetizing when it is fully ripe. While the drawback of plantains is that they ripen quickly, if not handled properly, it can change the nutritional value and nutrients contained in plantains. In this study, the author focuses on identifying the level of ripeness of bananas by using the image of a plantain fruit that is still intact and its skin. Processing of the image of the plantain fruit using HSI (Hue Saturation Intensity) color space transformation feature extraction. The tool used to extract the HSI (Hue Saturation Intensity) color space transformation feature is Matlab. The attribute values obtained from the extraction are the Red, Green, Blue values obtained from the RGB values. Hue, saturation and intensity attributes were obtained from HSI extraction. Classification of the level of ripeness of plantain fruit is done with the help of the rapidminer tool. The method used is K-NN. The results obtained from this test are the accuracy value of 91.33% with a standard deviation value of +/- 4.52% with a value of  $k=4$ . The RMSE value obtained is 0.276.

**Keywords**— Banana, HSI (Hue Saturation Intensity), K-NN.

## Introduction

Banana fruit has a Latin name is *Musa Paradisiaca* is one type of fruit that is often found in Southeast Asia, even in the world. Banana plants can generally grow in tropical and humid areas (tropical fruits) (Sohaib Ali Shah et al., 2020). There are many types of bananas, the most popular of which is the Raja banana (*Musa paradisiaca* L). This fruit plant is often found in areas with tropical and subtropical climates. This banana is one type of banana that is widely consumed in Indonesia. Plantains have a fragrant aroma and are of medium size and taste very sweet when fully

ripe, and are appetizing.

Banana cultivation has been developed to meet market demand. In Indonesia, it has not been able to fill in the world market. Lack of knowledge of banana farmers to cultivate on a large scale is one of the main factors.

Bananas ripen quickly so that lack of accuracy in cultivation can reduce the quality so that it can change the nutritional and vitamin content contained in (Indarto & Murinto, 2017).

In this case the author tries to overcome this problem by identifying the level of fruit maturity by taking samples from ripe bananas with skin images using the HSI (Hue Saturation Intensity) color space transformation feature

using the K-NN method.

## Material and Method

### A. Image Processing

Image is a number, in terms of the beauty of language, image is a collection of colors that can look beautiful, patterned, abstract and so on. Because it is a number, the image can be managed digitally. Digital image processing or Image Processing is a method used to process images so as to produce other images that are in accordance with the wishes of the processor. The main function of Image Processing is to improve the quality of the image so that it can be seen more clearly without any harder work process on the human sense of sight. In addition, image processing is used to process data obtained in machine perception, namely the steps used to extract information from images, information in a form suitable for the digitization process (Areni et al., 2019).

Image processing involves changing the nature of the image to make the image more attractive. the image is more interesting in the perception of the human view. It also makes the image more suitable from the standpoint of an autonomous machine (Sripaurya et al., 2021). The algorithm used is very reliable with careful design. This is very contrary to the other plots and is an image manipulation so that more explicit differences are produced (Mohd Ali et al., 2020). In its development, image processing cannot be separated from the field of computer vision (Li et al., 2018).

### B. RGB Image

The RGB color space has three basic components, namely R (red), G (green), and B (blue). Each pixel is formed from these components. This model is usually displayed in the form of a three-dimensional cube, which has red, green, and blue colors located at the corners of the axis. Each image has R, G, and B values, each of which has its own value contained in the histogram of each R, G, and B layer. Color image extraction is a step to obtain different information or characteristics. between one digital image and another digital image (Rulaningtyas et al., 2015).

#### 1. Image R (Red)

Red color element which is the result of RGB image extraction. Each color has an intensity level of 0 – 255. A value of 0 is the

darkest value in intensity. While the value of 255 is the value of the brightest intensity. With this, the image for the red color element will be visible on the screen if what is displayed is (Red, 0, 0). So the image that looks dark red. If it is displayed as (Red, 255,255) then the image will turn bright red.

#### 2. Image G (Green)

Red color element which is the result of RGB image extraction. Each color has an intensity level of 0 – 255. A value of 0 is the darkest value in intensity. While the value of 255 is the value of the brightest intensity. With this, the image for the red color element will be visible on the screen if what is displayed is (Red, 0, 0). So the image that looks dark red. If it is displayed as (Red, 255,255) then the image will turn bright red.

#### 3. Image B (Blue)

Is a type of basic color that cannot be made by mixing other colors because it is a primary or primary color. To get a blue image in an RGB image, it is necessary to compare each of the basic colors (red, green, and blue).

### C. HSI method

This method is usually used a process for feature extraction on an object. The HSI Color Space Transformation Method is a method of changing the color representation from RGB to HSI color representation, where Hue means the actual color of a color, Saturation means the level of color purity, how much white is in the color and Intensity is the amount of light. received by the eye when seeing the color of the object (Pratama et al., 2019).

#### 1. Hue

Hue is combination of reference color with vector S (Saturation). The color obtained in Hue usually tends to be red, but it is possible to have other colors.

#### 2. Saturation

Saturation is color scheme that describes the original color. This parameter depends on the number of wavelengths that contribute to the perception of the resulting color. The wider the range of wavelengths, the purer the color.

### 3.Intensity

Intensity is term used to describe colors other than Hue and Saturation. The value 1=0 represents black. As it is known that the intensity color which is the gray level is very precise in interpreting the monochromatic color level. So with the gray color, it can be measured easily.

### D. K-Nearest Neighbor

KNN is supervised learning model by looking for new patterns in the data, to find pattern recognition using the K-NN Algorithm, the Unsupervised learning K Nearest Neighbor approach is carried out based on the data regression value of the K-neighbor value to predict the data output value. The working principle of K-NN is to find the value of the adjacent distance between the training data and K neighbors (neighbors) (Yodha & Kurniawan, 2014). Furthermore, the data is combined with some data to form several patterns that are used as features of the data, and become parts based on training data. The accuracy value of the K-Nearest neighbor algorithm is strongly influenced by the relevance of the data, if the weight of the attributes used is equivalent to the relevance of the data.

The example of the calculations used to predict the price of batik cloth used the K-NN algorithm with the following model:

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Information:

$x$  = sample data

$y$  = test data

$D$  = distance

Distance calculation except by using Eq. Where  $x = x_1, x_2, \dots, x_n$  and  $y = y_1, y_2, \dots, y_n$  represents the n attribute values of the two records.

$$fknn(x') = \frac{1}{k} \sum_{i \in N_k(x')} y_i$$

$x'$  = Forecasting data

$K$  = number nearest neighbor

$N_k(x') y_i$  = closest neighbor

$y_i$  = output of the closest neighbor

### E. Research Design

In this research, the K-Nearest Neighbor method approach and by using features that can be obtained from the result of extracting with RGB and HSI image. It shwon at Figure 1.

In this study the dataset used is private data. The dataset used is an image of a plantain fruit. The image of plantains is categorized into 3, namely raw, half-ripe, and ripe. The following is the proposed method used atFigure 2.

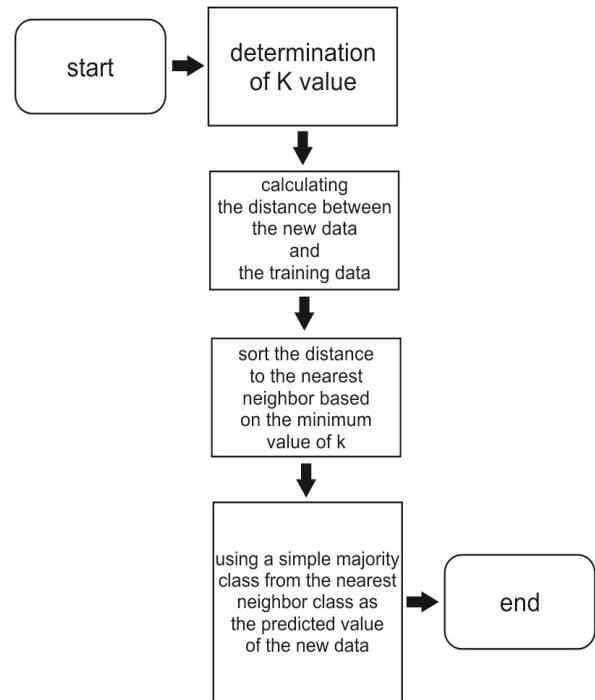


Figure 1.. Flowchart of KNN

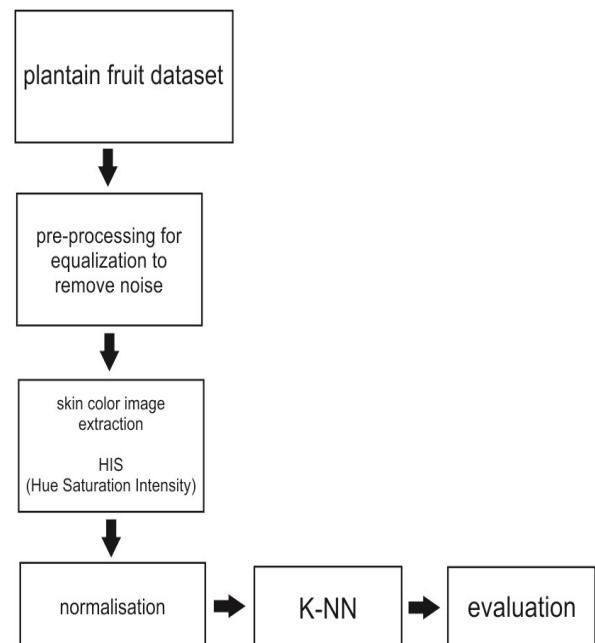


Figure 2. Proposed Method

## Result and Discussion

### A. Data Collecting

Research that uses digital images as research objects, to obtain image data in the form of numbers. digital image processing is needed in order to become numbers that can be used in the extraction and data mining process. The stages carried out in this study are based on the proposed method, the stages carried out include preprocessing, skin image transformation with color space transformation on the skin using RGB and HSI images, the classification method used is the K-nearest neighbor method. There are 300 digital images of bananas used in this study, 100 images of raw plantains 100 images of half ripe plantains, 100 images of ripe plantains.(Figure 3, Figure 4, Figure 5)



Figure 3.  
Raw Banana



Figure 4.  
half-baked



Figure 5.  
mature

### B. Pre-Processing

The preprocessing process carried out in this research uses 300 data sets with three categories of banana ripeness levels where 100 data indicates the level of raw maturity, 100 data indicates the half ripe category and 100 data indicates the ripe ripeness level. The next stage, after the dataset has been fulfilled, the next stage of the process is to flatten the size of the plantain image with a pixel size of 187x400, then replace the bright white background on the object to then look for the value of the RGB and HSI images.

### C. Test

In this study, the determination of the value of K is very influential on the value of accuracy when classifying the data, from testing the value of K in this study with a value of K = 1 to K = 10, the highest value is at K = 1.K = 1 to K = 10. The experimental results of the K-NN method on the plantain image dataset using rapidminer tools:

The results of the experiments that have been carried out using the K-NN method, the highest accuracy is found at a value with an accuracy of 91.33%, K = 1 with an accuracy of 91.33%. 89.33%, k=2 with accuracy 89.00%, k=3 with accuracy 90.00%, k=5 with accuracy 88.00%, k=6 with accuracy 89.00%, k=7 with an accuracy of 88.67%, and then k= 8 and K=9 with an accuracy value of 86.67%, K=10 with an accuracy of 88.00% and the values that have the lowest accuracy are k8 and k9 = 86, 67%. (Table 1)

### D. Evaluation

After testing the models, the classification prediction results obtained from the method used. The next step is to calculate the confusion matrix according to the prediction results of the classification of the model used.(Table 2)

The chart diagram displays the results of grouping each type. Based on x-Axis: Red and y-Axis: Green with blue color characteristics representing raw images, green representing matal or half-cooked images, and red representing mature images.

Table 1. Accuracy of K . value

K . value	ACCURACY
1	89.33%
2	89.00%
3	90.00%
4	91.33%
5	88.00%
6	89.00%
7	88.67%
8	86.67%
9	86.67%
10	88.00%

Table 2. Confusion Matrix

	true RAW	True half-baked	true Mature	class precision
pred. RAW	96	10	0	90.57%
pred. half-baked	4	89	11	85.58%
pred. MATURE	0	1	89	98.89%
class recall	96.00%	89.00%	89.00%	

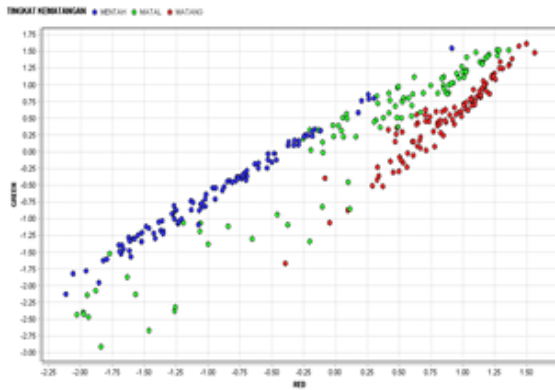


Figure 6 Plot Diagram

The RMSE obtained from testing on the rapid minner resulted in a value of 0.276, shown at Figure 6.

## Conclusion

In this study, an experiment using the image of Pisang Raja divided into three categories, namely raw, half-ripe (matal) and ripe. With extraction and HSI with the K-NN method it is proven to increase accuracy, the highest accuracy in the K-NN method is found at  $k = 4$  with a value of 91.33% and the lowest accuracy value lies at  $k = 8$  and  $k = 9$  with 86.67% while the RMSE (root mean square error) level is 0.276.

## Suggestion

In this research, it can be further developed with various scientific steps with better features and extraction methods to get better accuracy values. Therefore, it is hoped that further research can improve the shortcomings that exist in this study by adding features or using other methods.

## References

- Areni, I. S., Amirullah, I., & Arifin, N. (2019). Klasifikasi Kematangan Stroberi Berbasis Segmentasi Warna dengan Metode HSV. *Jurnal Penelitian Enjiniring*, 23(2), 113–116. <https://doi.org/10.25042/jpe.112019.03>
- Indarto, & Murinto. (2017). Deteksi Kematangan Buah Pisang Berdasarkan Fitur Warna Citra Kulit Pisang Menggunakan Metode Transformasi Ruang Warna HIS ( Banana Fruit Detection Based on Banana Skin Image Features Using HSI Color Space Transformation Method ). *Jurnal Ilmiah Informatika*, V (November), 15–21.
- Li, X., Wei, Y., Xu, J., Feng, X., Wu, F., Zhou, R., Jin, J., Xu, K., Yu, X., & He, Y. (2018). SSC and pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology. *Postharvest Biology and Technology*, 143(September 2017), 112–118. <https://doi.org/10.1016/j.postharvbio.2018.05.003>
- Mohd Ali, M., Hashim, N., & Abdul Hamid, A. S. (2020). Combination of laser-light backscattering imaging and computer vision for rapid determination of oil palm fresh fruit bunches maturity. *Computers and Electronics in Agriculture*, 169(November 2019), 105235. <https://doi.org/10.1016/j.compag.2020.105235>
- Pratama, R., Assagaf, A. F., Tempola, F., Studi, P., Informatika, T., Teknik, F., Khairun, U., Citra, P., Kematangan, K., & Classification, M. (2019). Tomato Fruit Detection Based on Color Features Using His Color Space Transformation Method. *JIKO (Jurnal Informatika Dan Komputer)*, 2(2), 81–86. <https://ejournal.unkhair.ac.id/index.php/jiko/article/view/1318/1058>
- Rulaningtyas, R., B. Suksmono, A., L. R. Mengko, T., & Putri Saptawati, G. A. (2015). Segmentasi Citra Berwarna dengan Menggunakan Metode Clustering Berbasis Patch untuk Identifikasi Mycobacterium Tuberculosis. *Jurnal Biosains Pascasarjana*, 17(1), 19. <https://doi.org/10.20473/jbp.v17i1.2015.19-25>
- Sohaib Ali Shah, S., Zeb, A., Qureshi, W. S., Arslan, M., Ullah Malik, A., Alasmay, W., & Alanazi, E. (2020). Towards fruit maturity estimation using NIR spectroscopy. *Infrared Physics and Technology*, 111, 103479. <https://doi.org/10.1016/j.infrared.2020.103479>
- Sripaurya, T., Sengchuai, K., Booranawong, A., & Chetpattananondh, K. (2021). Gros Michel banana soluble solids content evaluation and maturity classification using a developed portable 6 channel NIR device measurement. *Measurement: Journal of the International Measurement Confederation*, 173(May), 108615. <https://doi.org/10.1016/j.measurement.2020.108615>
- Yodha, J. W., & Kurniawan, A. W. (2014). Pengenalan Motif Batik Menggunakan Deteksi Tepi Canny Dan K-Nearest Neighbor. *Techno.COM*, 13(4, November), 251–262.