

CLASSIFICATION OF MATURITY LEVELS IN ARECA FRUIT BASED ON HSV IMAGE USING THE KNN METHOD

Frencis Matheos Sarimole^{1*}, Anita Rosiana²

Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya

Informatika, East Jakarta, Indonesia^{1,2}

matheosfrancis.s@gmail.com

Received : 10 August 2022, Revised: 01 September 2022, Accepted : 02 September 2022

*Corresponding Author

ABSTRACT

Areca nut (Areca catechu) is a kind of palm plant that grows in Asia and Africa, the eastern part of the Pacific and in Indonesia itself, areca nut can also be found on the islands of Java, Sumatra and Kalimantan. At the stage of classifying the maturity of the betel nut so far, it is still using the manual method which at that stage has subjective weaknesses. Based on these problems, researchers will create a system that is able to classify the level of maturity of areca nut using HSV feature extraction with assistance at the classification stage using the KNN method. In this study, 842 datasets were used which were divided into 3 types of classes, namely ripe, unripe and old fruit. The dataset was divided into 683 training data and 159 test data. In the next stage, the data is tested using the K-Nearest Neighbor method by calculating the closest distance using $k = 1$. From the results of the calculation of the closest distance $k1$ produces an accuracy rate of 87.42%.

Keywords : Matlab, Areca Ripeness, KNN, HSV.

1. Introduction

Areca nut is a kind of palm plant that grows in Asia and Africa, the eastern part of the Pacific. Which has a height that reaches 12 to 30 m, rooted in white hairs, upright stems with a diameter of about 15 to 20 cm, unbranched with loose leaf marks that are clearly visible. Stem formation occurs at 2 years and bears fruit at the age of 5 years to 8 years depending on soil conditions, soil with good moisture and having a pH range of 5-8 strongly supports the growth of areca nut. And Indonesia itself, areca nut are also widely available on the islands of Sumatra, Java, Kalimantan, Sulawesi, and Papua. Indonesia is included in a country that can be said to be the largest betel nut exporting country with export volumes reaching one hundred and ten thousand tons in 2007(Warnakulasuriya & Chen, 2022).

Increased in the following years, 80% of the demand for betel nut comes from Indonesia. In developed countries such as Belgium, the Netherlands, Japan, Germany, South Korea, and China, areca nut is used as a pharmaceutical raw material, betel nut is also used as a mixture of betel nut. And areca nut is also known as a medicinal ingredient such as being able to treat toothache in children and adults, dysentery, bloody diarrhea, and scabies. Areca nut seeds contain alkaloids, such as arekoline (C₈ H₁₃ NO₂), arekolidine, arekain, guvakoline, guvasine, and isoguvanine, condensed tannins, hydrolyzed tannins, flavans, phenolic compounds, gallic acid, latex, lignin, volatile and non-volatile oils, and salt. This research is related to the current areca nut plantation sector, which is often raised as a topic of discussion, such as determining the betel nut based on the age of the betel nut, which is still not practical, due to the age difference between one betel nut and another. And the determination of harvest can also be determined from changes in the color of the skin on the areca nut, the hardness of the skin of the fruit, and the occurrence of loss of the fruit and also the broken skin of the fruit also determines the maturity of the betel nut, and the distribution of areca nut in various regions makes it important in classifying the fruit. areca nut based on the level of maturity. The ripeness of the betel nut starts from raw, ripe and old so that the color of the areca nut can be an important indicator to be able to determine the level of maturity of the fruit and the quality of the fruit. The classification of areca nut maturity aims to reduce the risk of unripe betel nut. The classification of betel nut maturity manually still has several weaknesses and requires a long process, which has low accuracy and is inconsistent, this is because the determination is made subjectively by areca nut farmers. As for the

classification of the level of maturity of areca nut which automatically, can be faster with objective determination. And besides that it can improve accuracy and be more efficient. In a process of identification of fruit maturity can use image processing. The results of Arif Patriot's research stated that for the identification of mango fruit ripeness using images taken with the GLCM and LAB values, the accuracy of the results was 62.5% (Pamungkas, et al., 2019). GLCM is the Gray Level Co-Occurrence Matrix of the image, where the features are taken from the contrast, correlation, energy, and homogeneity of the grayscale image. While the LAB is the value of the color image taken on average, standard deviation. The process of classifying mango fruit maturity using KNN (Pamungkas, et al., 2019). Another related research in the process of image recognition pattern recognition is the identification of mango fruit maturity (Barkah, 2020). Where the image used is mango fruit, which is classified into 3 classes, namely: 38 ripe betel nut, 12 half ripe mango, and 50 unripe mango. In the process of identifying the preprocessed input image, the standard deviation feature is also taken. And the trahir stage is classification or identification using Euclidean distance with an accuracy value of 84% (Barkah, 2020).

In the year 2019 Research has been carried out by Cinantya Paramita, Eko Hari Rachmawanto, Christy Atika Sari, and De Rosal Ignatius Moses Setiadi. With the title "Classification of limes on the level of ripeness of fruit based on color features using K-Nearest Neighbor", From this study the researcher used the KNearest Neighbor method with the results of the classification of limes from the level of maturity using the color feature K, namely k=3 using k=7 on the search for Euclidean distance which produces an accuracy of 92% (Paramita, et al., 2019).

In a study conducted by Abdullah and Pahrianto in (2017) with the title "Tomato maturity classification system based on color and shape using the Support Vector Machine (SVM) method". In this study, the author uses the Support Vector Machine (SVM) method, with the test results obtained that this system is able to obtain an accuracy with an average of 82.83% and a standard deviation of 1.52 (Mubarak, 2021). Research in (2019) conducted by Husnul Khotimah, Nur Nafi'ah, and Masruroh using the title "classification of mango fruit maturity based on HSV image features using the KNN method". which is converted to HSV (Hue , Saturation , Value) where the features used are the average value of skewness, and kurtosis. The dataset used is a mango image with a value of k=1, the accuracy value is 56%. Results with raw, sufficient, ripe, and highly classified on mango images with a value of k=2, a ripe value. With training data of 129 mangoes, and the accuracy is 80% (Nafiah, 2019).

From several previous studies and existing problems, such as how to identify the maturity of areca nut is still very simple, namely by doing it traditionally only relying on the sense of sight, causing a reduction in the quality of sorting, so we need a system that is able to identify the maturity of areca nut as early as possible so that there are no mistakes. at post-harvest. So here it encourages the author to make an idea to design a system that becomes the background to be presented in this study with the title "Classification of betel nut maturity based on HSV images with KNN". With this research, it is hoped that in addition to helping betel nut farmers, there is also a classification of areca nut sold to the public so that the quality of the betel nut received by the community is better based on the level of maturity (Yang, et al., 2021; Thirani, et al., 2022; Pusadan & Abdullah, 2022).

2. Literature Review

Based on research conducted by (Paramita et al., 2019) "Classification of Lime Based on Level of Fruit Maturity Based on Color Features Using the K method. -Nearest Neighbor" According to the researcher, if the selection of limes still uses visible human judgment, it will result in weaknesses, namely being subjective and inconsistent so that the level of accuracy is low. So we need an automatic method that can improve the accuracy of the assessment with a consistent assessment in the classification of lime maturity levels based on color features. In his research, it is stated that the KNearest Neighbor method is a fairly simple classification method with fairly good accuracy, that is, it works based on the closest distance from training data to testing data by checking the distances of Cityblock Distance and Euclidean Distance. Decision makers using the K-Nearest Neighbor Method is a new data grouping technique based on the k nearest neighbor distance between training information and test information. The k values used

in this paper are 1 , 3 , 5 , 7 and 9 with the search for the distance between the training and test data is Euclidean Distance and Cityblock Distance.

According to research conducted by (Pariyandani et al., 2019) "Classification of Formalin Fish Imagery using the K-NN Method and GLCM " Health is a very important factor in life. So to create human health, it is necessary to improve the quality of their food, namely by fulfilling 4 healthy 5 perfect. With the fulfillment of food quality, one must consume nutritious food rich in protein. One of the nutritious foods with high protein is fish. As a maritime country, Indonesia has a lot of resources, so we can find many types of fish in the seas of Indonesia. Fish is one of Indonesia's abundant and diverse marine resources, for centuries many people have made fish as their livelihood. With advances in marine fishing technology that has become increasingly important with modern equipment, approximately 60 million tons of fish are caught every year from the sea which is worth about 20 billion US dollars. There is more competition in the industrial world, not a few ways that are not well used, such as preserving fish using Formalin, which is a dangerous chemical which is used for preservation and is now also often used for the textile industry. Formalin substances that are used in the long term can cause many diseases, one of which is disturbed digestion and can also cause cancer. In this study using the K-NN and GLCM methods to determine the image of fresh fish and the image of fish that has been given formalin, in this study the classification of formalin fish images requires materials such as formalin liquid as a testing substance and also fresh fish as the main ingredient for research testing and using The research samples were 500 images of fish, 250 images of fresh fish and 250 samples of formalinized fish images. The sample trainer used was 60% of the 500 sample, 40% of the sample used as a tester. And to measure the spatial frequency of the image and GLCM moments can use contrast calculations. Contrast is the pixel gray level of the image.

According to research conducted by (Wijaya & Ridwan, 2019) "Classification of Types of Apples Using the K-Nearest Neighbors Method." enjoyed and consumed by the public. Apples have many varieties that can be distinguished based on the color and shape of the fruit. From these types of apples, besides being able to be consumed raw/directly, the fruit can be consumed by processing it first, such as candied fruit, apple chips and drinks. With so many types of apples, it is not uncommon for ordinary people to find it difficult to classify the types of apples. One of them is by utilizing computer technology. In classifying the types of apples, there are features of Hue Saturation Value (HSV) and local Binary Pattern (LBP). Used in this study as extraction of color features and also fruit shape which will then be used as characteristics of the color and shape of the apples to be studied(Shumaila, et al., 2022).

Research conducted by (Meiriyama, 2018) " HSV Color Feature Fruit Image Classification With SVM Classifier" Image classification with fruit objects is a classic problem in the area of image classification that is still attracting researchers' interest. In the fruit classification process, the feature selection process is one of the factors that affect the success rate and level of accuracy. The features obtained will be used in the training process on SVM to get the ideal hyperplane with the maximum margin. To extract the fruit, several steps are needed, such as: Converting the RGB color model into an HSV color model, Forming an HSV histogram with a Hue channel of 16 bins, Saturation 4 bin and value 4 bin, The HSV histogram is then normalized by dividing each value by the total value. from the histogram.

3. Research Methods

In this study, the research material is a public dataset downloaded by the MIHAI MINUT account through the Kaggle website with the title "fruit-262" with 256 kinds of fruit datasets. For the dataset that you want to study is the Pinang fruit label, there are 1,020 datasets which are divided into 3 classes. Each class is in jpg format with the label "Mature" containing an image of a ripe areca, the label "Raw" containing an image of an unripe betel nut, the label "old" containing an image of a ripe betel nut.

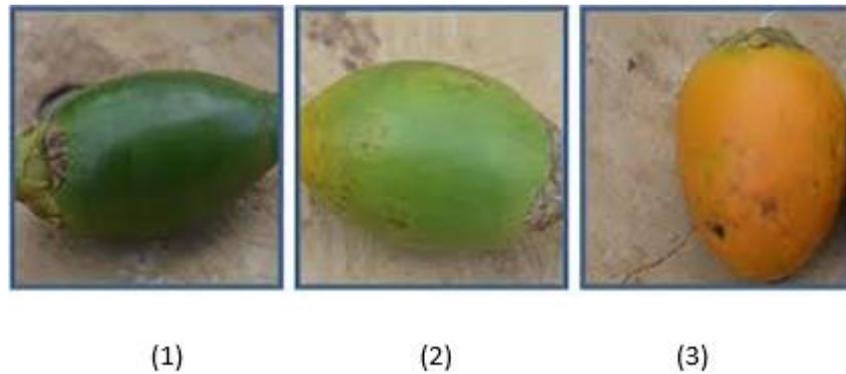


Fig 1. Description (1) Raw, (2) Ripe, (3) Old

A. System Design Flow

The following is a test design stage using Matlab to identify the level of maturity of areca nut using the HSV and KNN methods.

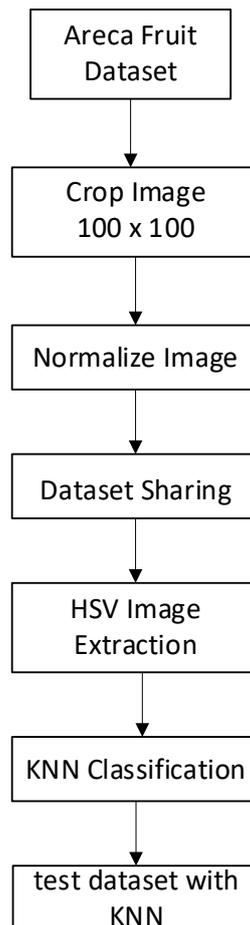


Fig 2. Test Design Flowchart

This stage is a stage. Based on the description of the test design to identify the level of maturity of the areca nut in this study :

- 1) Areca nut dataset will be formulated
- 2) The dataset will be processed through several pre-processing stages consisting of cropping, resizing image 100x100, converting to Grayscale with matlab.
- 3) Divide the dataset into training data and test data by dividing 80% of the training data and 20% of the test data manually
- 4) Furthermore, the RGB value will be converted through segmentation into the HSV color space, Hue, Saturation and Value obtained after calculating the equation.
- 5) The next test will be classified with the KNN algorithm with the calculation of K values from 1 to 9.

- 6) Test the dataset with the highest accuracy search

4. Results and Discussions

a) Implementation

This chapter describes how the system works and the results that have been created. That is successful in detecting the maturity level of Areca nut by applying HSV and K-NN color feature extraction in the classification of maturity levels. In implementing this research, the researcher uses Matlab as a research support program. The level of maturity of the areca nut in this study were raw, ripe and old.

b) Testing

- Areca Fruit Dataset

The initial stages of image acquisition, by determining the data needed for the image segmentation process. In this study, the process of searching for the Pinang fruit dataset obtained a public dataset which was downloaded by the MIHAI MINUT account through the Kaggle website.

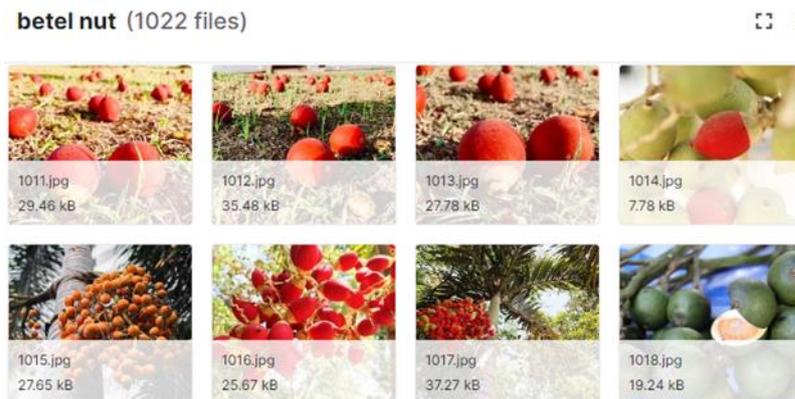


Fig 3. Image Acquisition Process

- Preprocessing Stage

The image preprocessing stage consists of image cropping, image size normalization (image resizing). The following is the process carried out at the image preprocessing stage:

• Cropping Image

The image cutting process stage is the stage of taking a certain area to be observed in the image to make it easier to analyze the image and reduce the size of the image storage. In the process of cutting this image using the Snipping Tools.

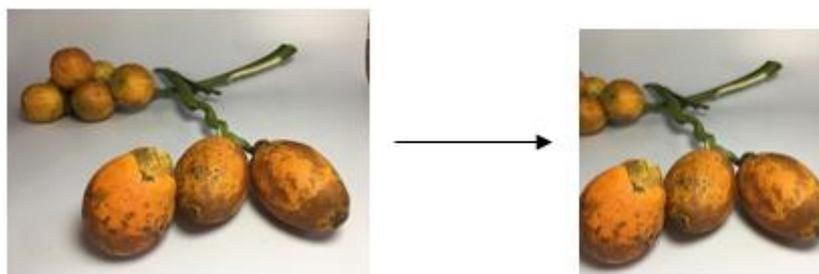


Fig 4. Cropping Image Dataset

• Image Size Normalization (Image Resize)

The stage of normalizing the image size is by changing the size of the original image into pixels. In the application of this research, the number of pixels to resize the image is 100x100, so the process steps are as follows. which is where the resize formula will later be stored into a process with the aim of resizing each image dataset that has been cropped.



Fig 5. Result of Normalization of Areca fruit image

- **Dataset Sharing**

in the distribution of datasets using split-folders as a divider for the entire dataset which will become a folder for training data and test data. By using Google Collabs. This Python library package is needed as an initial step to start sharing datasets. Next import splitfolders then `splitfolders.ratio(input_folder, output="buahpinang_2", seed=42, ratio=(.8, .2), group_prefix=None)` # default values. To divide the 80% training dataset and 20% test dataset.

- **HSV Feature Extraction**

After dividing 80% of the train dataset and 20% of the test dataset, implementing the HSV (Hue Saturation Value) feature extraction, in this stage the aim is to obtain the calculation results of the extraction value of the image object. The following is the calculation of the results of the test dataset and training dataset obtained from feature extraction using HSV (Putra, et al., 2021).

The first step to get the HSV feature extraction image is the steps that will be processed. By calling the dataset with RGB color which will go through the grayscale, binary and HSV stages.



Fig 6. RGB Color Image

The RGB dataset image segmentation will be converted to grayscale using Matlab software, with a total of 599 datasets, which have been divided into 485 trains and 114 tests. The following displays the RGB images that have been converted to Grayscale.



Fig 7. Grayscale Image

After going through the Grayscale stage. Then the dataset goes through a morphological stage, namely a binary image which is used to remove noise in the image.



Fig 8. Citra Biner

The next stage is HSV (Hue, Saturation and Value) which is the stage of image feature extraction to obtain feature data or information from the train image object and test whether the areca nut is ripe, unripe or old.

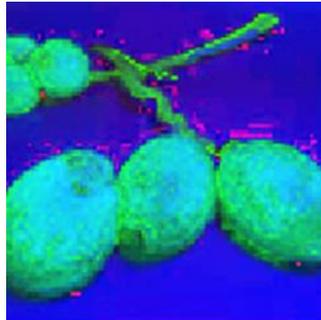


Fig 9. HSV feature extraction

From the results of the feature extraction above, the value of the image object has been obtained, namely Hue, Saturation and Value which are training data.

Table 1 - HSV Feature Extraction

H	S	V	Information
0.10449761	0.65284983	0.520627826	Ripe Areca nut

Below is a table of training data extraction results obtained through feature extraction using HSV (Hue, Saturation and Value). Table 2 shows 15 data from the total data, namely 683 data.

Table 2 - Data extraction practice table

Extraction of Train Data Features			
H	S	V	Label
0.104497610020441	0.652849833230313	0.520627825801766	ripe betel nut'
0.164378626476636	0.726115967253132	0.539204950272792	ripe betel nut'
0.124033681352992	0.813614887330634	0.616905563105330	ripe betel nut'
0.0929736385522287	0.712481321749618	0.500406530624097	ripe betel nut'
0.138702081049235	0.561940120885431	0.434187078814716	ripe betel nut'
0.322032100795644	0.230180856346037	0.498850655348603	raw betel nut'
0.214399510893484	0.547819674883151	0.274101935391720	raw betel nut'
0.229852247834494	0.527675824021506	0.224036265682573	raw betel nut'
0.316281169294299	0.464792307963316	0.343298198878942	raw betel nut'
0.222673277025372	0.544343440243290	0.435835696175945	raw betel nut'
0.107189914766225	0.547175719428006	0.515623173257457	areca nut'
0.118607173574856	0.420454373415293	0.449084510260981	areca nut'
0.199315575226905	0.661136019714119	0.465461863340880	areca nut'
0.0780859291791207	0.698450945937737	0.465266533225784	areca nut'
0.0895837544713060	0.391317513570586	0.497612984401621	areca nut'

After getting the results of the training data extraction then perform feature extraction to get the results in the test data. The training data and test data will be processed through an advanced method, namely KNN. The following is the result of the test data that has gone through the feature extraction stage, which shows 15 test extraction data from 159 test data.

Table 3 - Example Extraction test table

Test Data Feature Extraction			
H	S	V	Label
0.329845521552262	0.523663691254355	0.470404558941270	ripe betel nut'
0.172848572876017	0.518397478151850	0.391113863089095	ripe betel nut'
0.143340978057847	0.557703792923646	0.429409259597250	ripe betel nut'
0.412282866996376	0.460173534238910	0.463127806912707	ripe betel nut'
0.409336644845051	0.465532001862337	0.443082958954015	ripe betel nut'
0.322032100795644	0.230180856346037	0.498850655348603	raw betel nut'
0.265691747314418	0.560139888442877	0.359851404949444	raw betel nut'
0.240365342180436	0.645016010983586	0.328631533204282	raw betel nut'
0.149512737622453	0.543546075454673	0.452293690877339	raw betel nut'
0.187792357130913	0.659172122880860	0.299708768720852	raw betel nut'
0.122254805437172	0.408095218707927	0.388288447909284	areca nut'
0.0812957670686929	0.475068377096455	0.592531324777262	areca nut'
0.104291033437651	0.690943395992200	0.319077671175422	areca nut'
0.0913405534415003	0.405255680277492	0.474169146452654	areca nut'
0.0812957670686929	0.475068377096455	0.592531324777262	areca nut'

- KNN Classification

After the hsv feature is obtained, then the classification is carried out with knn with the closest distance $k = 3$. With the test results obtained at 83.01% as an example, the following is the result of the test data generated.

Table 4 - Example KNN Result

Test Data	Original Class	KNN Result	Information
2	ripe betel nut'	areca nut'	Not accurate
3	ripe betel nut'	ripe betel nut'	Accurate
12	ripe betel nut'	raw betel nut'	Not accurate
13	ripe betel nut'	ripe betel nut'	Accurate
14	ripe betel nut'	raw betel nut'	Not accurate
63	raw betel nut'	raw betel nut'	Accurate
64	raw betel nut'	raw betel nut'	Accurate
65	raw betel nut'	ripe betel nut'	Not accurate
66	raw betel nut'	ripe betel nut'	Not accurate
67	raw betel nut'	areca nut'	Not accurate
138	areca nut'	raw betel nut'	Not accurate
139	areca nut'	areca nut'	Accurate
140	areca nut'	areca nut'	Accurate
141	areca nut'	areca nut'	Accurate
142	areca nut'	ripe betel nut'	Not accurate

- Final Test Results

From the testing process, 159 data were tested from a total of 842 data using KNN, it was found that 132 data were classified as accurate and 27 data were inaccurate and the test results can be seen as follows :

Tabel 5 Prediction

Prediction			
Current	Ripe Betel Nut'	Raw Betel Nut	Areca Nut
Ripe Betel Nut'	37	11	2
Raw Betel Nut	3	45	2
Areca Nut	2	7	50

$$\text{Accuracy} = \frac{TP (\text{True Positive})}{\text{Data}} \times 100\%$$

$$\frac{(37+45+50)}{159} \times 100\% = 83,01 \%$$

5. Conclusion

After testing the betel nut, the researcher can draw the conclusion that in detecting the maturity level of the areca nut by applying the HSV and K-NN color feature extraction in the level classification, it can be drawn well in a ripe areca classification system. From the testing data, the accuracy obtained has an accuracy value of 87,42%.with the closest distance K = 1 by using training data as many as 485 betel nuts and test data as many as 114 betel nuts by having their respective classes, namely raw class enough class, ripe class and old class. So based on the extraction of HSV color features with the areca nut maturity classification system using the K-Nearest Neighbors (KNN) method, it is feasible to use it properly.

References

- B Barkah, M. F. (2020). Klasifikasi Rasa Buah Jeruk Pontianak Berdasarkan Warna Kulit Buah Jeruk Menggunakan Metode K-Nearest Neighbor. *Coding Jurnal Komputer dan Aplikasi*, 8(1).
- Meiriyama, M. (2018). Klasifikasi Citra Buah berbasis fitur warna HSV dengan klasifikator SVM. *Jurnal Komputer Terapan*, 4(1), 50-61.
- Mubarok, H., Murni, S., & Santoni, M. M. (2021). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Tingkat Kematangan Buah Tomat Berdasarkan Fitur Warna. *Senamika*, 2(1), 773-782.
- Khotimah, H., Nafi'iyah, N., & Masruroh, M. (2020). Klasifikasi Kematangan Buah Mangga Berdasarkan Citra HSV dengan KNN. *Jurnal Elektronika Listrik dan Teknologi Informasi Terapan*, 1(2), 1-4
- Paramita, C., Rachmawanto, E. H., Sari, C. A., & Setiadi, D. R. I. M. (2019). Klasifikasi Jeruk Nipis Terhadap Tingkat Kematangan Buah Berdasarkan Fitur Warna Menggunakan K-Nearest Neighbor. *Jurnal Informatika*, 4(1).7
- Pariyandani, A., Larasati, D. A., Wanti, E. P., & Muhathir, M. (2019). Klasifikasi Citra Ikan Berformalin Menggunakan Metode K-NN dan GLCM. In *Semantika (Seminar Nasional Teknik Informatika)*. 2(1) 42-47).
- Pamungkas, A. P. S., Nafi'iyah, N., & Nawafilah, N. Q. (2019). K-NN Klasifikasi Kematangan Buah Mangga Manalagi Menggunakan L* A* B dan Fitur Statistik. *Journal of Computer Science and Visual Communication Design*, 4(1), 1-8.
- Pusadan, M. Y., & Abdullah, A. I. (2022). k-Nearest Neighbor and Feature Extraction on Detection of Pest and Diseases of Cocoa. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(3), 471-480.
- Putra, A. W., Putra, W., Mulyanto, B. S., & Gaffar, A. O. (2021). A performance of combined methods of VCG and 16BCD for feature extraction on HSV. *International Journal of Image, Graphics and Signal Processing*, 13(3), 13-32.
- Shumaila, M. N., Rehman, M., Anjum, M., Kanwal, F., & Bashir, K. (2022). Performing content-based image retrieval using rotated local binary pattern and multiple descriptors. *Mehran University Research Journal of Engineering and Technology*, 41(3), 142-148.
- Thirani, E., Jain, J., & Narawade, V. (2022). Using SVM and KNN to Evaluate Performance

- Based on Video Plagiarism Detectors and Descriptors for Global. *Journal of Soft Computing Paradigm*, 4(2), 82-100.
- Warnakulasuriya, S., & Chen, T. H. H. (2022). Areca Nut and Oral Cancer: Evidence from Studies Conducted in Humans. *Journal of Dental Research*, 00220345221092751.
- Wijaya, N., & Ridwan, A. (2019). Klasifikasi Jenis Buah Apel Dengan Metode K-Nearest Neighbors Dengan Ekstraksi Fitur HSV dan LBP. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 8(1), 74-78.
- Yang, R., Wu, Z., Fang, W., Zhang, H., Wang, W., Fu, L., ... & Cui, Y. (2021). Detection of abnormal hydroponic lettuce leaves based on image processing and machine learning. *Information Processing in Agriculture*.