Identification the Maturity Level of Carica Papaya Using the K-Nearest Neighbor

I B Suban¹, A Paramartha², M Fortwonatus³, and A J Santoso⁴

¹,²,³,⁴ Universitas Atma Jaya Yogyakarta, Jl. Babarsari No.43, Janti, Catertunggal, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281

Email: albjoko@staff.uajy.ac.id

Abstract. The agricultural sector plays an essential role in economic growth in Indonesia. It can see from the grouping of economic activities. In the grouping of economic activities, the agricultural sector is classified in the primary sector [1]. One result of the agriculture sector is Carica Papaya. Carica Papaya is a fruit that belongs to the type of berries. In 2013 world Carica Papaya production reached 1.25 × 10⁷ mt [2][3]. Determination of proper maturity at harvest is essential to get good quality fruit. Overripe fruits tend to be softer and more chewy and produce a bland taste [4]. In determining the maturity of Carica Papaya in large numbers is still done manually by human operators which can cause misclassification in determining the maturity of Carica Papaya. In determining the maturity of Carica Papaya by human operators, there are still many errors found in determining the maturity of Carica Papaya. In measuring the maturity of papaya, several methods can be used; one of them is the K-nearest neighbor method.

K-Nearest Neighbor is a relatively simple method of data classification with reasonable accuracy, which is based on the closest distance from the training data to the testing data by checking the Cityblock Distance and Euclidean Distance. K-Nearest Neighbors are an algorithm that functions to classify data based on learning data taken from the nearest K neighbors (nearest neighbors). With K is the number of closest neighbors. From the closest K selected, then choose a class from the nearest neighbor. The class with the highest number of neighboring votes is given as a prediction class label on the test data [5].

Based on the above problems, the writer wants to predict the maturity of the papaya fruit using the K-Nearest Neighbor method. The authors have categorized two levels of papaya fruit maturity that
is ripe, and unripe. The aim of this research is to assist papaya farmers in recognizing the maturity level of papayas so that they can effectively determine the maturity level of the papaya fruit.

2. Literature Review

According to M.P. Vaishnave1, K. Suganya Devi, P.Srinivasan, G. Arum General Jothi in a study entitled "Detection and Classification of Groundnut Leaf Diseases using KNN classifier" Peanuts are one of the most high-income agriculture in India. However, now the economic income, especially on peanut farming in India is declining due to a disease that attacks peanuts. This paper aims to identify diseases of peanut leaves. The method used is the KNN because it is beneficial for classifying and identifying diseases in plants. In this paper, there are steps to detect peanut disease, namely first shooting, second pre-processing, third image segmentation, fourth extracting features, and finally, classifying. After testing and getting the results, the authors grouped four different types of diseases. After getting the results, it can be concluded that detecting and classifying diseases in peanut leaves using KNN is very efficient [6].

According to Suresha M, Shreekanth KN, and Thirumalesh BV, in a study entitled "Recognition of Diseases in Paddy Leaves Using KNN Classifier" Rice plants are very important plants for countries in the world, especially in Asian countries, because rice is a staple food for people Asia. Because of diseases in rice plants farmers experience a lot of losses and decreased yields during harvest. In this paper, the K-NN method is used to identify and classify rice plant diseases. The diseases identified only include Brown Spot and Blast. The results obtained with an accuracy level of 76.59% using the K-NN method. Seeing the results that have been obtained, it can be concluded that using the K-NN method to identify and classify rice plant diseases is quite accurate [7].

Febri Liantoni et al. conducted a study on classifying the maturity of watermelon (Citrullus Lanatus Tunb / Citrullus Vulgaris Schrad). The reason researchers chose watermelon is that people often have difficulty identifying watermelon maturity levels. Based on that, the researchers used the k-nearest neighbor method to identify watermelon maturity. Researchers measured the level of watermelon maturity based on first-order statistical extraction. In conducting first-order statistical extraction, the authors use the parameters mean, variance, skewness, and kurtosis. The value of this parameter will be used by researchers for the testing process and the training process. According to researchers using the K-Nearest Neighbor method to classify the watermelon maturity is quite good [5].

3. K-Nearest Neighbor

The K-Nearest Neighbor algorithm is a simple algorithm used for classification and regression [8][9]. It is one of the simplest methods for solving classification problems. The K-Nearest Neighbor method often produces better results than other similar methods [10]. Following is the formula used in the K-Nearest Neighbor algorithm:

\[ d_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \]  

Information:

- \( X_{jk} = \) sample data.
- \( x_{ik} = \) testing data.
- \( d_{ij} = \) distance between two vector i and j.

4. Design

In the design process for the application of identification of papaya fruit maturity based on color using K-Nearest Neighbor. In carrying out the process of identifying the maturity of papaya fruit, the writer divides it into the testing process and the identification process.
A. Training Process

The following is the architecture of the training process.

![Training Process Architecture](image)

**Figure 1. Training Process Architecture**

The authors do the training process on the dataset image. The training process has four stages which are explained as follows:

1. **Image**

   The authors use six papaya images, which consist of 3 ripe papaya images and three unripe papaya images. Then the authors make a folder that serves to accommodate the six images of the exercise. The program will read the six images through a folder that was created by the previous authors.

2. **Feature Extraction**

   Feature extraction is the process of getting weights from 6 images based on RGB values. Before the feature extraction process, the authors make the stage of reducing the size of the image. This aims to speed up the transfer process.

3. **Labeling**

   Labeling is the process of giving a name based on the weights that have been previously known. In the labeling process, the authors categorize it into two types, namely ripe papaya, and unripe papaya.

4. **Training**

   The authors carry out the Training process on the dataset, which aims to improve the identification process.

B. Identification process

The following is the architecture of the identification process.

![Identification Process Architecture](image)

**Figure 2. Identification Process Architecture**

1. **Image**

   The authors use six papaya images, which consist of 3 ripe papaya images and three unripe papaya images. Then the authors create a folder to hold the six images of the exercise. The program will carry out an introduction to each papaya image. The authors enter the papaya image one by one to carry out the process of recognition of the image.

2. **Feature Extraction**

   Feature extraction is the process of obtaining weights from input images based on Red, Green, Blue values. Before the feature extraction process, the authors make the stage of reducing the size of the image. This aims to speed up the identification process.
3. Labeling

Labeling is the process of giving a name based on the weights that have been previously known. In the labeling process, the authors categorize them into two types, namely papaya ripe and papaya unripe.

4. Identification

Identification is the process of determining the input image. This identification process aims to determine whether the input image has a raw or cooked category.

5. Results and Implementation

In the results and implementation section, the authors build a simple software to predict the maturity of papaya fruit. Also, the authors' purpose in building this software is to prove the algorithm used by the authors and to realize the results of the authors' design.

Figure 3. and Figure 4. is the implementation of the training section. In Figure 3., the authors display the weight of feature extraction based on RGB color. The author uses six images for feature extraction in the training section. Then the authors also display unfilled database. The database will be used for the labeling process.

In Figure 4. the authors display the results of labeling based on data sets. In this labeling process, the authors also conducted the training process. The purpose of the writer to do the labeling and training process so that the application can predict the maturity of papaya fruit well.
Figure 5. is an implementation of the identification section. Figure 5. shows six insert images that have been identified by the software built by the authors. Where before the software identifies the input image, the input image has passed the stage of image reduction, feature labeling extraction, and identification process.

In the process of identifying the maturity of papaya fruit, the authors used 12 images consisting of 6 training images and six testing images. Six training images consist of 3 images of ripe papaya fruit and three images of unripe papaya fruit. In the process of identification, the authors also use six images — the author's test to determine the accuracy of the K-Nearest Neighbor method. The authors use the equation below to measure the accuracy of the K-Nearest Neighbor method.

\[
\frac{\text{Amount of identification data}}{\text{Total retrieval of all data}} \times 100\% \quad (2)
\]

From the equation above, the authors can find out the level of accuracy as follows:

\[
\frac{12}{12} \times 100\% = 100\%
\]

In general, the success rate of papaya fruit identification, in general, is 100%. The level of accuracy in ripe and unripe images can be seen in table 1.

| Papaya Category | Number of Samples | Matching | Error | Level of Accuracy |
|-----------------|-------------------|----------|-------|------------------|
| Ripe            | 6                 | 6        | 0     | 100\%            |
| Unripe          | 6                 | 6        | 0     | 100\%            |

6. Conclusion

Based on the results of research conducted, several conclusions can be drawn as follows:
1. In identifying the maturity of papaya using the K-Nearest Neighbor method during the identification process, one process can only accept one color parameter input (red, green, blue).

2. After testing with the K-Nearest Neighbor method, the accuracy rate is 100% in ripe and unripe images.

For further research, the authors suggest adding the amount of training data with testing data and comparing it with other methods.

7. References

[1] “Indonesian economic growth,” Badan Pusat Statistik, pp. 1–9, 2014.
[2] F. Faostat, “Agriculture Organization of the United Nations Statistics Division Production,” 2016. [Online]. Available: http://faostat3.fao.org/browse/Q/QC/S.
[3] L. F. Santos Pereira, S. Barbon, N. A. Valous, and D. F. Barbin, “Predicting the ripening of papaya fruit with digital imaging and random forests,” Comput. Electron. Agric., vol. 145, pp. 76–82, 2018.
[4] Y. A. Purwanto, P. M. Pandjahitan, Sutrisno, E. Darmawati, Y. Makino, and S. Oshita, “Maturity prediction of papaya using NIR spectroscopy,” Acta Hortic., vol. 1213, pp. 447–452, 2018.
[5] F. Liantoni et al., “Watermelon classification using k-nearest neighbours based on first order statistics extraction,” J. Phys. Conf. Ser., 2019.
[6] M. P. Vaishnnave, K. S. Devi, P. Srinivasan, and G. A. P. Jothis, “Detection and Classification of Groundnut Leaf Diseases using KNN classifier,” in Proceeding of International Conference on Systems Computation Automation and Networking 2019, 2019, pp. 1–5.
[7] M. Suresha, K. N. Shreekanth, and B. V. Thirumalesh, “Recognition of diseases in paddy leaves using knn classifier,” in 2017 2nd International Conference for Convergence in Technology, I2CT 2017, 2017, pp. 663–666.
[8] H. Y. Riskiawan, T. D. Puspitasari, F. I. Hasanah, N. D. Wahyono, and M. F. Kurnianto, “Identifying Cocoa ripeness using K-Nearest Neighbor (KNN) Method,” pp. 354–357, 2019.
[9] N. S. Altman, “An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression,” Am. Stat., vol. 46, no. 3, pp. 175–185, 1992.
[10] D. A. Adeniyi, Z. Wei, and Y. Yongquan, “Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method,” Appl. Comput. Informatics, vol. 12, no. 1, pp. 90–108, 2014.

Acknowledgments

Authors expressed their appreciation for financial support from the Informatics Engineering Magister Study Program, Postgraduate Faculty, Universitas Atma Jaya Yogyakarta. The authors also thank to Cipta Thegar for helping the authors by providing code samples. Thank you to all those who have supported this research.