

Mobile Application for Identification of Coffee Fruit Maturity using Digital Image Processing

Oka Sudana^{a,1}, Deden Witarsyah^b, Adhitya Putra^{a,2}, Sunia Raharja^{a,3}

^a Department of Information Technology, Udayana University, Badung, 80361, Indonesia
E-mail: ¹agungokas@unud.ac.id; ²dewa.adhitya21@gmail.com; ³sunia.raharja@unud.ac.id

^b Department of Information System, Faculty of Industrial Engineering, Telkom University, Bandung, 40257, Indonesia
E-mail: dedenw@telkomuniversity.ac.id

Abstract— Indonesia is an agricultural country that relies on the agricultural sector and is well known in producing various plantation commodities, one of which is coffee. Coffee is a leading export commodity developed in Indonesia. Community coffee plantations play an important role because most of the coffee production comes from community plantations. However, the condition of community coffee plantations can be said to be still hampered, due to the quality of coffee is still relatively low. It is caused by coffee fruit sorting, which is still done conventionally. The conventional sorting process of coffee fruits is still carried out with the help of operator knowledge, so the level of operator knowledge dramatically influences the results of sorting. The ease of sorting coffee ripeness can be done by implementing a mobile application using digital image processing. Techniques used in digital image processing are the HSV color space to get color features of coffee fruit and the K-Nearest Neighbor (KNN) classification method to classify coffee fruit ripeness. The results of the identification are in the form of ripe, half-ripe, or unripe fruits. The mobile application of this research has two main features, namely training data feature and non real-time identification feature. The results of the testing conducted resulted in an accuracy rate of 95.56% with the best membership value (k) of 3.

Keywords— coffee fruit maturity; HSV; K-Nearest Neighbor; digital image processing.

I. INTRODUCTION

Advances in technology, especially in the computer field, have penetrated all aspects. One of them is the aspect of agriculture, where computer technology plays an essential role in the world of agriculture in the quality of crops through digital image processing. Indonesia is an agricultural country that relies on the agricultural sector and is well known for producing various plantation commodities, one of which is coffee [1]. Coffee is a leading export commodity developed in Indonesia because it has a relatively high economic value in the world market [2]. The economic value of coffee is a substantial contributor to foreign exchange for the Indonesian state. Smallholder coffee plantations play an important role because 96% of coffee production comes from smallholder plantations, 2% comes from the State Large Plantation (PBN), and the remaining 2% comes from Large Private Estates (PBS) [3].

However, the condition of the smallholder coffee plantations is still relatively poor compared to the State Large Plantation (PBN). This condition can be said to be still hampered because it is caused by the quality of the coffee produced is still quite low. The influencing factor is the

process of sorting coffee fruit, which is still inadequate, whereas the sorting of good coffee is one of the important factors in strengthening the competitiveness of coffee products from Indonesia.

The process of sorting coffee fruit (henceforth will be referred to as coffee) in people's coffee plantations is still done conventionally. This method is easy to do, but the operator needs to have sufficient knowledge in sorting the maturity coffee so that the results of sorting are very much influenced by the level of knowledge of the human operator in conducting the sorting process. Digital image processing-based technology is a solution to overcome the problem of sorting the maturity coffee, which is still done conventionally with the help of operator knowledge.

Along with the development of technology, digital image processing, and classification techniques are widely used in identifying objects. Many studies have been conducted regarding identification, such as researching the classification of tomatoes by using the K-Nearest Neighbor (KNN) classification method [4]. Pattern recognition of tomato images using the GLCM and HSV methods. Accuracy results obtained from the study are equal to 100%, where the p parameter on the GLCM method is 9, and the k

parameter on the KNN method is 3. The prediction of the presence of hajj fails so that it can be useful for others who attend the pilgrimage that is on the waiting list to prepare themselves [5]. This prediction uses the K-Nearest Neighbor (KNN) algorithm. The highest accuracy results in this study were 96.2% using the optimum neighbor value (k) = 9.

This study aims to facilitate the identification of coffee maturity based on the color of the image using digital image processing by utilizing the HSV color space and K-Nearest Neighbor (KNN) classification method. The HSV color space was chosen because humans use this color space in seeing real colors. The K-Nearest Neighbor (KNN) classification method was chosen because this method is very suitable for problems related to classification or identification.

II. MATERIALS AND METHOD

A. Coffee Plants

Coffee is a type of cash crop that has long been cultivated and has high economic value. Bali Province is one of the coffee producing regions in Indonesia. The Province of Bali produces two types of coffee, Robusta coffee and Arabica coffee, but the type of coffee most cultivated in Bali Province is Arabica coffee [6], [7].

Harvesting good coffee fruit will affect the quality of good coffee. Harvesting of Arabica coffee takes 6-8 months. Good quality coffee plants come from healthy, pithy, and red coffee beans. The maturity of the coffee is characterized by changes in the skin color of the fruit has been red. Harvesting of unripe (green fruit) and unhealthy coffee will cause the physical quality of the coffee beans to decrease and the taste deteriorate.

Sorting coffee is based on the level of maturity color of the fruit. Full red indicates that the coffee is ripe, reddish-yellow indicates that the coffee is classified as half ripe, and green or yellowish-green indicates that the coffee is still relatively unripe [8].

B. Digital Image Processing

Digital image processing has developed in everyday life, especially in digital libraries, image databases, remote sensing, and others [9]. Digital image processing is a discipline that studies techniques for image processing [10]. The image referred to here is a still image (photo) and a moving image (coming from a webcam). Digital here has the intention that image processing/image is done digitally using a computer. Mathematically, the image is a continuous function with the intensity of light in the two-dimensional plane. An image on a digital computer must be presented numerically with discrete values. A digital image can be represented by a two-dimensional matrix $f(x, y)$ consisting of M columns and N rows. The intersection between columns and rows is called a pixel (picture element) or the smallest element of an image.

C. HSV Color Model

HSV color model defines colors in three components namely hue (H), saturation (S) and value (V) [4] [11]. This HSV is closer than the RGB color model in describing the colors received by the human eye.

The hue (H) component represents the main color, such as red and green. Hue has a range of 0-360 degrees in representing the main color. The saturation component (S) represents the color clarity of the main color. Saturation has a range between 0-100%. The lower percentage of the saturation component will produce a grayish color while the higher the percentage will produce a dense color. The component value (V) represents the brightness of a color. Value has a range between 0-100%. The lower the percentage value, the main color will look dark, while the higher the percentage value, the main color will look white [11]–[13].

The HSV color model (Hue, Saturation, Value) if visualized then it looks like in Fig. 1.

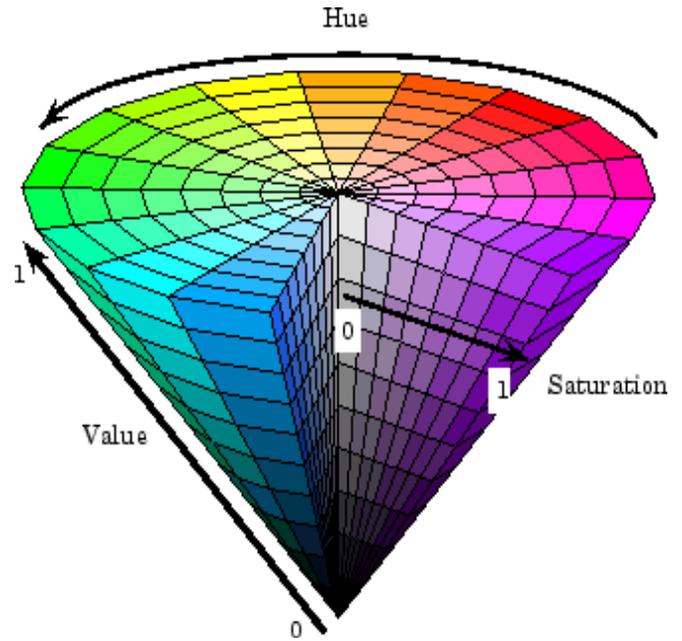


Fig. 1 HSV Color Model

The HSV color model is a derivative of the RGB color model. So the formula calculation in finding HSV values from RGB can use the following formula [13].

$$H = \begin{cases} 60^\circ \times \left(\frac{G' - B'}{\Delta} \text{ mod } 6 \right) & C_{max} = R' \\ 60^\circ \times \left(\frac{B' - R'}{\Delta} + 2 \right) & C_{max} = G' \\ 60^\circ \times \left(\frac{R' - G'}{\Delta} + 4 \right) & C_{max} = B' \end{cases} \quad (1)$$

$$S = \begin{cases} 0 & \Delta = 0 \\ \frac{\Delta}{C_{max}} & \Delta <> 0 \end{cases} \quad (2)$$

$$V = C_{max} \quad (3)$$

D. K-Nearest Neighbor

The K-Nearest Neighbor method, also known as the KNN method, is one of the classification methods for supervised learning [14]. K-Nearest Neighbor can also be said as one of the non-metric methods used in classification [15] and is believed to be first introduced by Fix and Hodges in 1951 [16]. The working principle of K-Nearest Neighbor (KNN)

is to find the closest distance between the data to be evaluated with the number of neighbors (k) closest in the training data [17]. The K-Nearest Neighbor method has two phases or phases, namely the training phase and the classification phase.

The training phase process of this algorithm is to store feature vectors and sample training data classifications. The classification phase of this algorithm is that the same features are calculated for the test data (the classification is unknown).

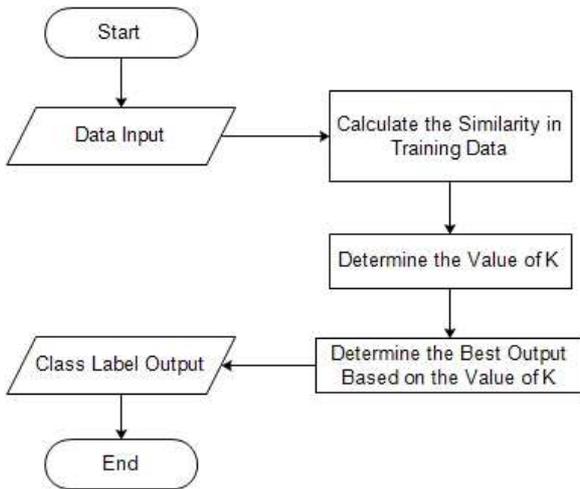


Fig. 2 Flowchart K-Nearest Neighbor (KNN)

The K-Nearest Neighbor classification process is the value of input data calculated on training data by determining the value of k . For example, determined the number of neighbors (k) which is 3, then each test data is calculated the distance to the training data and selected 3 training data the distance closest to the test data. Check the best output value based on the k value that has been determined and used as output because of classification.

E. OpenCV

OpenCV stands for Intel Open Source Computer Vision Library, which consists of at least 300 C functions, or even more. This software is very popular to be used for image processing applications such as object detection and so on. This software is free, can be used in commercial and non-commercial terms, without having to pay a license to Intel. OpenCV can operate on Windows or Linux-based computers. This software provides some image processing functions, as is the case with image and pattern analysis functions [18].

F. Android

Android is an operating system for smartphones that is very popular in the world, and the majority of smartphone users use this system [19] [20]. Android was developed by the Open Handset Alliance, but was acquired by Google in 2005 and officially released by Google in 2007 [21]. The android OS is also known to be open source and provides the Android Native Development Kit (NDK), which is a toolset for developing application parts using the C/C++ programming language [22], [23].

This system is built from 5 main components namely Applications, Application Framework, Libraries, Android

Runtime, and Linux Kernel. The processes that are served on Android are presented in the form of activity. Activity is a component of an Android application that provides a screen for users to interact such as calling, taking pictures, or sending emails [24]. Each activity is given a frame to describe the activity interface.

G. Research Methods

The method applied in this research is HSV color space for digital image processing and K-Nearest Neighbor (KNN) classification method. HSV color space was applied in this study to obtain the color value from the coffee image which is used as a feature of the maturity level of the coffee. K-Nearest Neighbor (KNN) classification method is applied in this study to classify the color features obtained from the coffee image tested, then the results of the classification show the image of coffee in the ripe, half ripe, or unripe categories.

This study used 450 images of coffee which consisted of 300 images used as training data and 150 images used as test data. The stages in the application designed can be seen in Fig. 3.

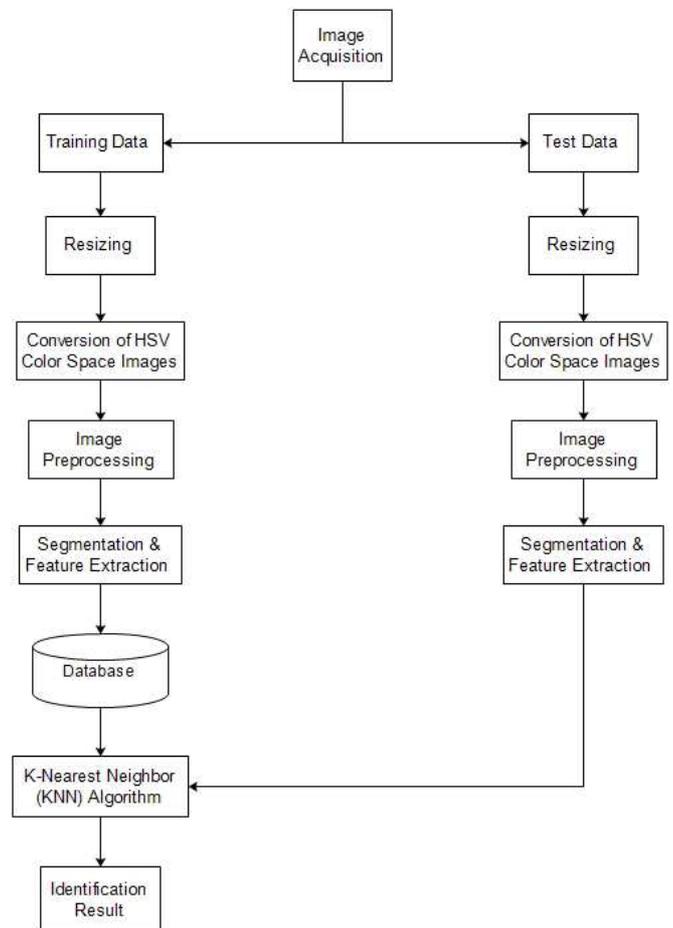


Fig. 3 General Outline of the System

General description of this system has two main stages, namely the training stage and the identification stage. The stages in the general description of the system are explained as follows.

1) *Image Acquisition*: The process of acquiring this image is to use the coffee image taken with a white background. Coffee image data used are 300 images for training data with various types of levels of maturity ranging from unripe, half ripe, until ripe. While the images of coffee used for the test data were 150 images consisting of 50 images of coffee classified as unripe, 50 images of coffee classified as half ripe, and 50 images of coffee as ripe. Coffee image that has for further processing.

2) *Resizing*: Resizing stage is the stage of changing the size of the image of coffee that has been inputted. The initial stage of this process is to check the orientation of the input image. If the orientation of the input image is in the form of portrait, then it needs to be changed to landscape orientation. If the orientation of the input image is in the form of landscape, then proceed to the process of changing the image size to a size of 667 x 500 pixels.

3) *Conversion of HSV Color Space Images*: Conversion stage of HSV color space is the stage of changing the image color space from RGB color space to HSV color space. The coffee image that has been converted into an HSV image is parsed into several channels in it, namely the hue, saturation, and value channels. One channel of the HSV color space, namely saturation channel, is used in the image preprocessing stage.

4) *Image Preprocessing*: The image preprocessing stage is the image quality improvement stage. This stage is done to get the object to be recognized. Image preprocessing stage includes the binarization stage, which is to change the image from saturation channel to binary image, the filling stage is the stage to fill empty object pixels so that they become intact objects, and finally the opening stage is to eliminate small objects that are not needed so that in the final image of the preprocessing image stage produces objects that are ready for use in the feature segmentation and extraction stages.

5) *Segmentation and Feature Extraction*: The segmentation phase is carried out to obtain objects that will be recognized in each HSV channel, namely hue, saturation, and value channels. The feature extraction step is performed to get the feature values of the coffee image object on each segmented HSV channel. The results of masking process in each channel in the form of coffee fruit image objects, then performed an average calculation of pixel values on the object image in the hue channel, saturation, until the value. The results of the average color value in each channel is used as a feature value of the coffee object. If the coffee image input is training data, the value of the object features that have been obtained will be stored in a database along with the type of classification label input by the user. The type of classification label indicates the level of maturity of the coffee that is unripe, half ripe, and ripe. If the image of the coffee input is test data, then the value of the object features of the coffee that has been obtained will be used in the classification process.

6) *Classification*: The classification stage is the stage to identify the type of coffee maturity based on the value of features of the coffee object that has been processed against the training data in the database. The color values of the hue

channel, saturation, and values that have been obtained from the previous process will be used as input values to be classified using the K-Nearest Neighbor (KNN) method with training data in the database. The results of classification show identification in the form of an estimated age and the level of maturity of the coffee fruit such as ripe, half ripe, or unripe.

III. RESULT AND DISCUSSION

The results of this study are in the form of an android application to identify the maturity of coffee. This application utilizes the OpenCV library in the application of existing step processes, such as the HSV color space, image preprocessing, and the K-Nearest Neighbor (KNN) classification method.

A. System Testing Results

The results of system testing are carried out to find out which system/application has been made to function according to its function. The results of the application that have been designed can be seen in Fig. 4 in the form of the main view of the application

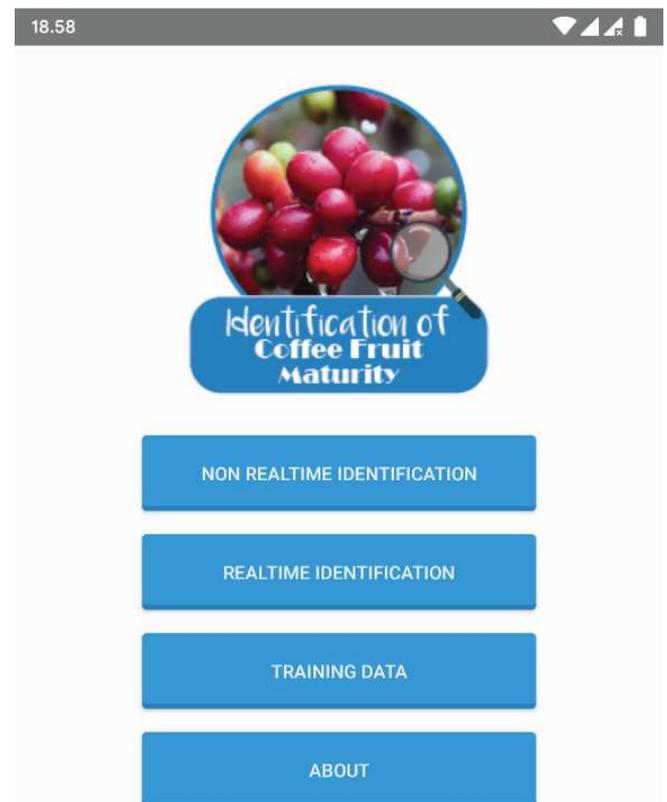


Fig. 4 Coffee Fruit Maturity Identification Application

Fig. 4 is the main display of the designed application. There are four features, namely the non realtime identification feature, the real-time identification feature, the training data feature, and the application feature.

This study discusses two features in the application, namely the training data feature and the non realtime identification feature. The training data feature functions to display and register the characteristics of the coffee which is used as training data and stored in a database. Coffee used is

has been sorted based on the type of maturity level. The coffee image process that used as training data is a image of

coffee with white background that taken using smartphone camera with a size of 4096 x 3072 pixels.

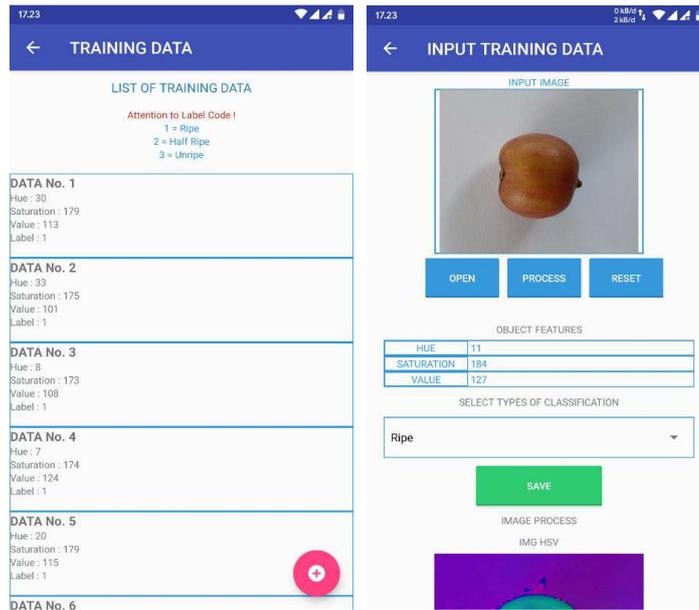


Fig. 5 Data Training View

This training data feature displays a list of training data on coffee stored in the database. The data displayed is in the form of hue, saturation, value, and label code type of maturity level. The add button is a button to add new training data. The first process is inserting the coffee image to be registered as training data and then processed to display the image processing stages such as HSV color space image conversion, image preprocessing, segmentation to feature extraction of the coffee object. The features obtained are hue, saturation, and value. Then determine the type of

maturity label type, which is ripe, half ripe, or unripe. Finally save these features in the database by selecting the Save Button.

The non realtime identification feature functions to identify the level of maturity of the coffee from the coffee image which is used as test data. The coffee image process that used as test data is an image of coffee with white background that taken using smartphone camera with a size of 4096 x 3072 pixels.

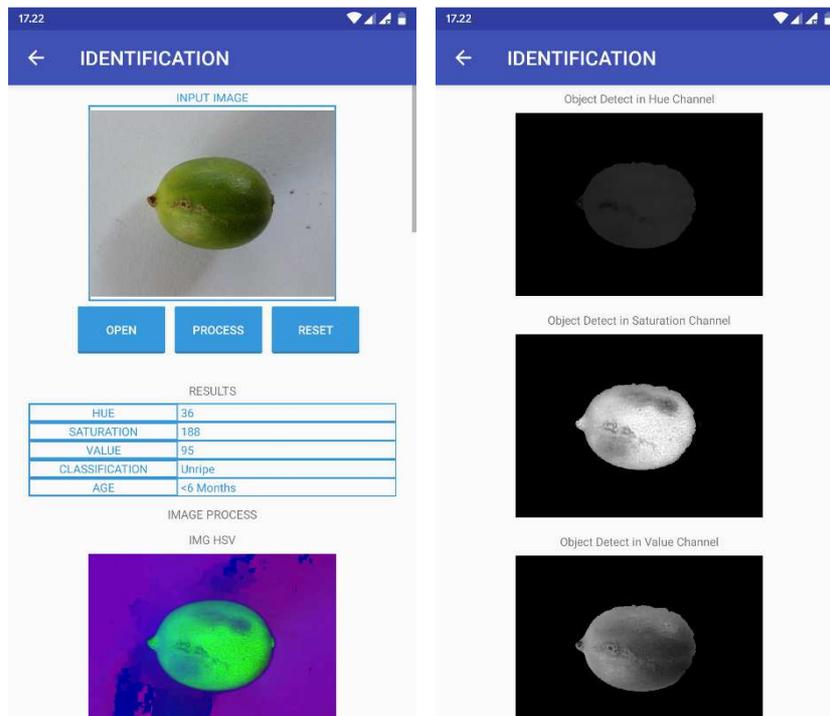


Fig. 6 Identification View

The first process in this identification feature is inserting the coffee image and then processed to display the stages of the image process such as HSV color space image conversion, image preprocessing, segmentation to feature extraction of the coffee object. The features obtained are hue, saturation, and value. Then the feature values are classified using the K-Nearest Neighbor (KNN) classification method with training data that has been stored in the database. The results obtained in the form of labels of the type of coffee maturity level and the estimated age of the coffee.

B. Analysis of Results

Testing of the identification of coffee maturity in the application made aims to test the success of the application of the method used. The coffee image that will be identified has been taken previously using a smartphone camera with a white background and conditioned lighting. Test data samples used were as many as 50 images in each type of maturity namely ripe, half ripe, and unripe. Testing is done by testing various numbers of neighbors (k) on the classification method used. The k values used in the test are $k=1$, $k=3$, $k=5$, $k=7$, and $k=9$.

The results obtained from testing with various parameters in the K-Nearest Neighbor (KNN) classification method can be seen in Table 1 as well as a comparison of the percentage of accuracy of the test results that can be seen in Table 2 and Fig. 7.

TABLE I
KNN CLASSIFICATION TEST RESULTS

Type of Maturity	Total of Test Data	Total of Correct Data				
		$k=1$	$k=3$	$k=5$	$k=7$	$k=9$
Ripe	50	46	45	43	41	40
Half Ripe	50	42	46	45	44	43
Unripe	50	48	49	49	50	50

TABLE II
COMPARISON OF KNN CLASSIFICATION TEST RESULTS

Type of Comparison	Type of KNN Testing				
	$k=1$	$k=3$	$k=5$	$k=7$	$k=9$
Accuracy	93.33%	95.56%	94.22%	93.33%	92.44%

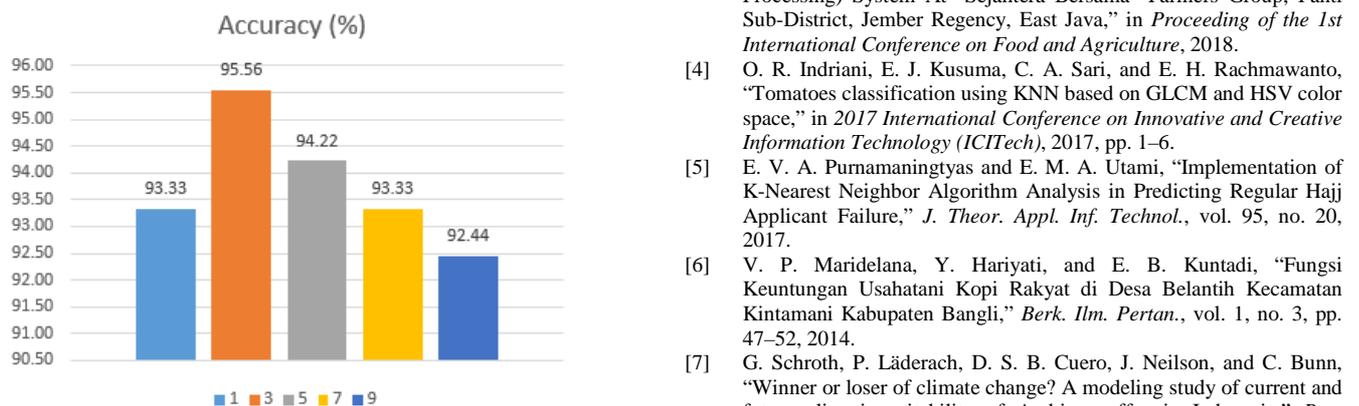


Fig. 7 Comparison of KNN Classification Test Results

C. Comparison of Research Results

The classification of tomatoes could be done using the K-Nearest Neighbor (KNN) classification method [4]. A comparison of the results of previous studies with the proposed research can be seen in Table 3.

TABLE III
RESEARCH RESULTS COMPARISON

Type of Comparison	Comparison	
	$k=1$	$k=3$
Accuracy	100%	95.56%

Research conducted by Indriani has an accuracy of 100% in the classification of tomatoes using K-Nearest Neighbor (KNN) with the best k value of 3 [4]. While this research on the identification of coffee fruits maturity using K-Nearest Neighbor (KNN) has an accuracy of 95.56% with the best k value is 3. The results of the comparison in previous studies with the proposed research have the best accuracy on the k value that is 3.

IV. CONCLUSION

Identification of coffee maturity based on the color of the image using digital image processing based on mobile applications is a solution in sorting the maturity level of coffee which is currently still mostly done conventionally. The feature used to recognize the maturity of the coffee is the color of the coffee by utilizing the HSV color space. These features are then classified using the K-Nearest Neighbor (KNN) method by displaying identification results in the form of ripe, half ripe, or unripe fruit categories and displaying fruit age estimates. The test results that have been done show the best percentage of performance with the number of neighbors (k) of 3. The percentage generated at $k=3$ is an accuracy of 95.56%.

REFERENCES

- [1] N. L. C. P. S. Suyasa, I. M. R. T. Setiawan, and M. Regina, "Kopi Arabika Jempolan Catur Village, Kintamani: Product and Promotional Mix Analysis," *SINTESA*, 2018.
- [2] I. P. A. Heryana, I. M. Sudarma, and I. G. S. A. D. I. Putra, "Perbandingan Pendapatan antara Usahatani Kopi dan Usahatani Jeruk di Desa Serai Kecamatan Kintamani Kabupaten Bangli," *E-Journal Agribisnis dan Agrowisata (Journal Agribus. Agritourism)*, 2016.
- [3] T. Kustiasari, U. Setyoko, D. Arieni, and P. Prayitno, "Arabica Coffee Bean Quality Test With Wet Processing (Full Wash Processing) System At "Sejahtera Bersama" Farmers Group, Panti Sub-District, Jember Regency, East Java," in *Proceeding of the 1st International Conference on Food and Agriculture*, 2018.
- [4] O. R. Indriani, E. J. Kusuma, C. A. Sari, and E. H. Rachmawanto, "Tomatoes classification using KNN based on GLCM and HSV color space," in *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, 2017, pp. 1–6.
- [5] E. V. A. Purnamaningtyas and E. M. A. Utami, "Implementation of K-Nearest Neighbor Algorithm Analysis in Predicting Regular Hajj Applicant Failure," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 20, 2017.
- [6] V. P. Maridelana, Y. Hariyati, and E. B. Kuntadi, "Fungsi Keuntungan Usahatani Kopi Rakyat di Desa Belantih Kecamatan Kintamani Kabupaten Bangli," *Berk. Ilm. Pertan.*, vol. 1, no. 3, pp. 47–52, 2014.
- [7] G. Schroth, P. Läderach, D. S. B. Cuero, J. Neilson, and C. Bunn, "Winner or loser of climate change? A modeling study of current and future climatic suitability of Arabica coffee in Indonesia," *Reg. Environ. Chang.*, vol. 15, no. 7, pp. 1473–1482, 2015.
- [8] D. J. Perkebunan, *Pedoman teknis budidaya kopi yang baik (Good Agriculture Practices /GAP on Coffee)*. 2014.
- [9] S. Febrianto Kurniawan, I. K. G. Darma Putra, and A. A. Kompiang Sudana, "Bone Fracture Detection Using OpenCV," *J. Theor. Appl. Inf. Technol.*, vol. 64, no. 1, pp. 249–254, 2014.
- [10] W. Burger and M. J. Burge, *Digital image processing: an algorithmic introduction using Java*. Springer, 2016.

- [11] V. Chernov, J. Alander, and V. Bochko, "Integer-based accurate conversion between RGB and HSV color spaces," *Comput. Electr. Eng.*, vol. 46, pp. 328–337, 2015.
- [12] I. P. Pratama Andika, I. P. Agung Bayupati, and N. K. Ayu Wirdiani, "Rancang Bangun Aplikasi Pendeteksi Tipe Dan Nilai Resistor Berbasis Android," *Lontar Komput. J. Ilm. Teknol. Inf.*, pp. 61–5832, 2015.
- [13] G. Saravanan, G. Yamuna, and S. Nandhini, "Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models," in *2016 International Conference on Communication and Signal Processing (ICCSP)*, 2016, pp. 462–466.
- [14] M. P. A. Deole and R. Longadge, "Content Based Image Retrieval using Color Feature Extraction with KNN Classification," *IJCSMC*, vol. 3, no. 5, pp. 1274–1280, 2014.
- [15] J. Maillo, S. Ramírez, I. Triguero, and F. Herrera, "kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data," *Knowledge-Based Syst.*, vol. 117, pp. 3–15, 2017.
- [16] 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, 2016.
- [17] F. Astutik, "Sistem Pengenalan Kualitas Ikan Gurame dengan Wavelet, PCA, Histogram HSV dan KNN," *Lontar Komput. J. Ilm. Teknol. Inf.*, 2015.
- [18] R. M. Gurav and P. K. Kadbe, "Real time finger tracking and contour detection for gesture recognition using OpenCV," in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 974–977.
- [19] H. Lahiani, M. Elleuch, and M. Kherallah, "Real time hand gesture recognition system for android devices," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2015, pp. 591–596.
- [20] Z.-Q. Zhao, L.-H. Ma, Y. Cheung, X. Wu, Y. Tang, and C. L. P. Chen, "ApLeaf: An efficient android-based plant leaf identification system," *Neurocomputing*, vol. 151, pp. 1112–1119, 2015.
- [21] A. F. Waruwu, I. P. A. Bayupati, and I. K. G. D. Putra, "Augmented Reality Mobile Application of Balinese Hindu Temples: DewataAR," *Int. J. Comput. Netw. Inf. Secur.*, vol. 7, no. 2, p. 59, 2015.
- [22] A. Tedeschi and F. Benedetto, "A real-time automatic pavement crack and pothole recognition system for mobile Android-based devices," *Adv. Eng. Informatics*, vol. 32, pp. 11–25, 2017.
- [23] A. Tedeschi, A. Liguori, and F. Benedetto, "Information Security and Threats in Mobile Appliances," *Recent Patents Comput. Sci.*, vol. 7, no. 1, pp. 3–11, 2014.
- [24] H. Zhang, H. Wu, and A. Rountev, "Automated test generation for detection of leaks in Android applications," in *Proceedings of the 11th International Workshop on Automation of Software Test*, 2016, pp. 64–70.