KNN FOR CLASSIFICATION OF FRUIT TYPES BASED ON FRUIT FEATURES

Moh. Afis Azman[1)], Nur Nafi'iyah[2)]
[1)]mahasiswa Teknik Informatika, Universitas Islam Lamongan
[2)]dosen Teknik Informatika, Universitas Islam Lamongan
Jalan Veteran 53A Lamongan
e-mail: afiszmn@gmail.com[1)], mynaff@unisla.ac.id[2)]

ABSTRACT

Research related to the recognition of fruit types has been done previously. Research related to the recognition of many types of fruit applies computer vision and artificial intelligence. The purpose of this research is to apply artificial intelligence science with the KNN method to identify the type of fruit. The KNN method has a good performance in previous studies. We tried to use KNN by determining the most optimal K value. There are five types of fruit images used in this study, namely Apples, Grapes, Oranges, Mangoes, and Strawberries. The fruit image is extracted with color, texture, and shape features with a total of 15 features, namely the average value of R, the average value of G, the average value of B, the value of skewness R, the value of skewness G, the value of skewness B, the value of grayscale entropy. , grayscale contrast value, grayscale energy value, grayscale correlation value, grayscale homogeneity value, binary area value, binary circumference value, binary major axis value, and binary minor axis value. The dataset used in this study was taken from Kaggle, with a dataset of 2750 images, each type of fruit contained 550 images, 2500 training images were used and 250 images were used for testing. The experimental results show that the KNN method with K=1 has the highest accuracy, which is 99.6%. The KNN method can be used optimally in classifying fruit types based on color, texture, and shape features.

***Keywords**: Fruit classification, Image features, KNN.*

I. INTRODUCTION

Classification or automatic introduction of fruit in sales stores or supermarkets needs to be researched. The development of vision technology based on intelligence can make it easier to recognize the type of fruit-based on imagery. Ordinary people not yet aware of many types of fruit can utilize computer vision technology to recognize fruit. We can simply photograph the fruit in a sales store or supermarket, and the application will recognize the fruit. The field of research is constantly being developed to produce the most appropriate and correct introduction of the fruit. So that several methods continue to be tried to be applied to the vision computer to recognize the fruit with the correct accuracy. One of the studies introduced the type of mango fruit by applying the Perceptron Multilayer algorithm. Research results [1] used the Neural Network method to recognize the type of mango fruit with an accuracy of approximately 95%. Another method used for the introduction of fruit ripeness is KNN. KNN also has good accuracy results to recognize the maturity of mangoes, which is approximately 80% [2]. However, the KNN method is also not good in determining the type of ripeness of Manalagi mango fruit, which is only approximately 62% [3]. This is because the image data used is too little, so it does not represent the diversity of fruit maturity, and too low the relationship between image features and fruit maturity. Several previous studies have used the KNN method for classification or recognition, including the introduction of leaf patterns or leaf types based on image features, namely GLCM features, *canny edge detection* features are very good results 98% [4]; The KNN method is also used to recognize egg quality based on GLCM egg image features and the result is good, which is 82% [5]; Based on previous research KNN is also used to identify the quality or quality of guava based on GLCM image features, but the results of the KNN method are very low only have an accuracy of approximately 45% [6].

Research related to artificial intelligence computer vision used for agriculture, especially fruit among others; Classification of fruit maturity starts from mango fruit, identifies the type of quality of guava fruit, and is also used to identify the ripeness of strawberries [7]. Based on previous research that identifying the type of fruit or fruit maturity has been done a lot, either using the KNN method or others. However, we would like to propose a KNN method to identify the type of fruit with five types apples, grapes, oranges, mangoes, and strawberries. This study will identify types of fruit with five kinds of fruits to be recognized by the KNN method, because previous studies of KNN have fairly high accuracy, and determine the optimal K value.

II. LITERATURE REVIEW

Computer-related research on vision-related to artificial intelligence in agriculture has been extensive, including

classifying fruit types [1][8][9][10][11], identifying fruit maturity [2-3][7][12][13], recognizing leaf types [4], identifying fruit quality [6][14][15]. Classification of fruit types that have been studied related to the types of mangoes, oranges, apples, and Rambutan. In some previous studies, the process of classifying fruit types is only one type, for example, the type citrus fruit, the type of apple, the type of mango, and other types of fruit. While this study will classify the types of fruit with a variety of fruits as many as five, namely: apples, grapes, oranges, mangoes and strawberries.

The methods used for the identification of fruit types in previous studies have used KNN. The KNN method also has a high accuracy value, averaging 80%. How the KNN method works will classify test data based on the calculation of the distance to training data. Equation 1 calculates the value of the closest distance based on K or neighbour.

$$d_{(x,y)} = \sqrt{\sum_{i=1}(x - y_i)^2} \quad (1)$$

Equation 1, x is test data, while y is training data, this calculation process is based on the number of variables, so there is a sigma symbol to sum each distance of the test data variable with the training data.

The variables used are fifteen, including average values R, G, B, *skewness* values R, G, B, *grayscale* entropy values of imagery, GLCM values (*Contrast, Energy, Correlation, Homogeneity*), fruit area values, fruit circumference, *major axis*, *minor axis*. How to get the value of each variable by calculating each image intensity on pixels. Equation 2 how to calculate the average value.

$$\bar{x} = \frac{\sum_{i=1}\sum_{j=1} x_{ij}}{i * j} \quad (2)$$

Explanation of Equation 2; Each red, *green, blue* component is searched for its average value, in the way x is the intensity value of each pixel component is summed, then divided by the number of rows x columns, i is the size of the image row, and j is the size of the image column.

Each variable of the KNN classification process is taking the feature value of each pixel intensity of the image. The size of the image used is not the same, so each area and circumference of the fruit image is diverse. Equation 3 how to get *skewness* variables R, G, B from the image.

$$sk = \frac{(\bar{x} - median)^3}{\sigma^3} \quad (3)$$

Equation 3, is the average value of the pixel intensity of the image of each image component ($\bar{x}$ *Red, Green, Blue*), the standard deviation value σ .

III. RESEARCH METHODOLOGY

A. Dataset

The datasets used in the study were taken from the Kaggle dataset. The total data used are described in Table 1. Each type of fruit (apples, grapes, oranges, mangoes, and strawberries) has 550 imagery. The data of 550 images of each type of fruit is divided into training data of 500 images, and 50 imagery for testing. The total data is 2750 images.

TABLE I
DATASET SHARING

Fruit Type	Training	Testing	Total
Apple	500	50	550
Wine	500	50	550
Orange	500	50	550
Mango	500	50	550
Strawberry	500	50	550
Total	2500	250	2750

B. Proposed Methods

The methods used in this study consist of feature extraction methods, and methods of classification or identification of fruit types. The feature extraction method will take features or parameters from the image ranging from the average value feature R, G, B (*Red, Green, Blue*), *skewness* value R, G, B, grayscale image entropy value, GLCM value (*Contrast, Energy, Correlation, Homogeneity*), fruit area or area value, fruit circumference or perimeter value, *major axis*, and *minor axis*. There are 15 variables or columns defined in the dataset in Table 2. Each type of fruit is taken to produce a dataset such as Table 2. Table 2 Explanation, x1 = average value R, x2=average value G, x3= average value B, x4 = *skewness* value R, x5 = *skewness* value G, x6 = *skewness* value B, x7 = grayish image entropy value, x8 = grayish image contrast value, x9 = grayish image energy value, x10 = grayish image correlation value, x11 = *grayish image homogeneity* value , x12 = grayish image broad value, x12= broad value of binary image, x13 = the circumference value of the binary image, x14 = the *major value of the*

image axis, and x15 = the minor axis value of the image, y is the label or fruit class. Each type of fruit is converted into numbers, apples = 1, grapes = 2, oranges = 3, mango = 4, and strawberries = 5.

TABLE II
DATASET FEATURE

x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	y
210	157	152	-2	-2	-2	5	0	0	1	1	543805	3863	1405	1236	1
129	116	76	1	0	1	8	0	0	1	1	1	0	1	1	1
162	119	118	0	0	-1	7	0	0	1	1	108941	2792	538	304	1
94	108	64	2	0	1	8	1	0	1	1	6	6	4	2	2
133	163	133	0	0	0	8	1	0	1	1	14883	1332	179	151	2
149	146	40	0	0	1	8	0	0	1	1	87609	4813	474	300	2

How to determine the identification of fruit types by the KNN method, by calculating the distance based on Equation 1. As for the feature capture flow as in Figure 1. Imagery is carried out in the process of converting to *grayscale* and binary to take texture features (*grayscale* imagery), and shape features (binary imagery). Figure 2 describes the process of classifying and identifying fruit types by the KNN method.

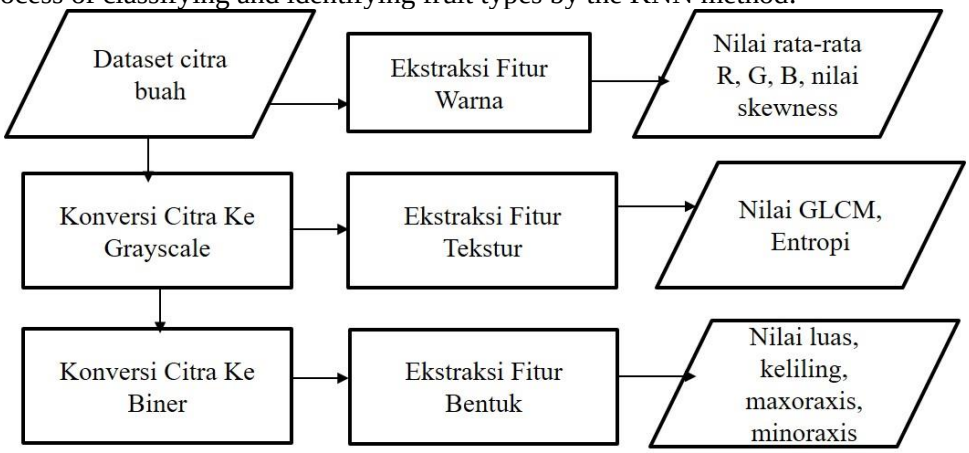


Figure. 1. Flow processing Feature Extraction

Colour image data is performed feature extraction (calculates the average values R, G, B, *skewness* values R, G, B). The feature is stored in an excel file. Other features are also used as one file so that it can be used as training data. The ready data is then trained, to produce a model from the KNN method. The results of the KNN model are used as a testing process to determine the type of fruit. The process of evaluating the model by calculating the accuracy is in Equation 4.

$$akurasi = \frac{jumlah\_data\_benar}{total\_data\_tes} x100\% \quad (4)$$

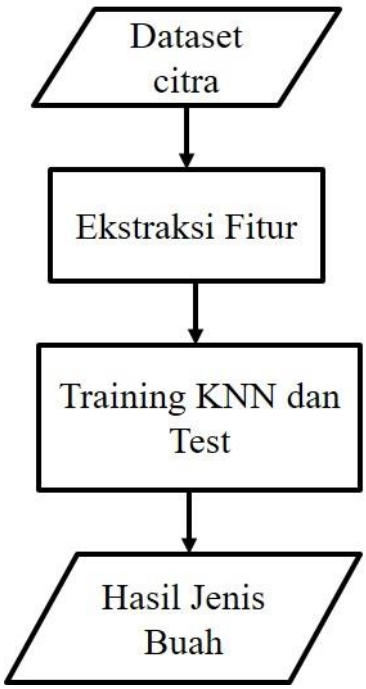


Figure. 2. Fruit Type Classification with KNN

IV. RESULT AND DISCUSSION

Figure 3 displays a view of determining the type of fruit at a time. Input the colour image and then click the *grayscale* image process to change the image to *grayscale*, after which click the binary image process to change the image to a binary shape. Next click the feature extraction process to get the values of the colour, texture and shape features, and then click the classification process to determine the image class of the fruit.

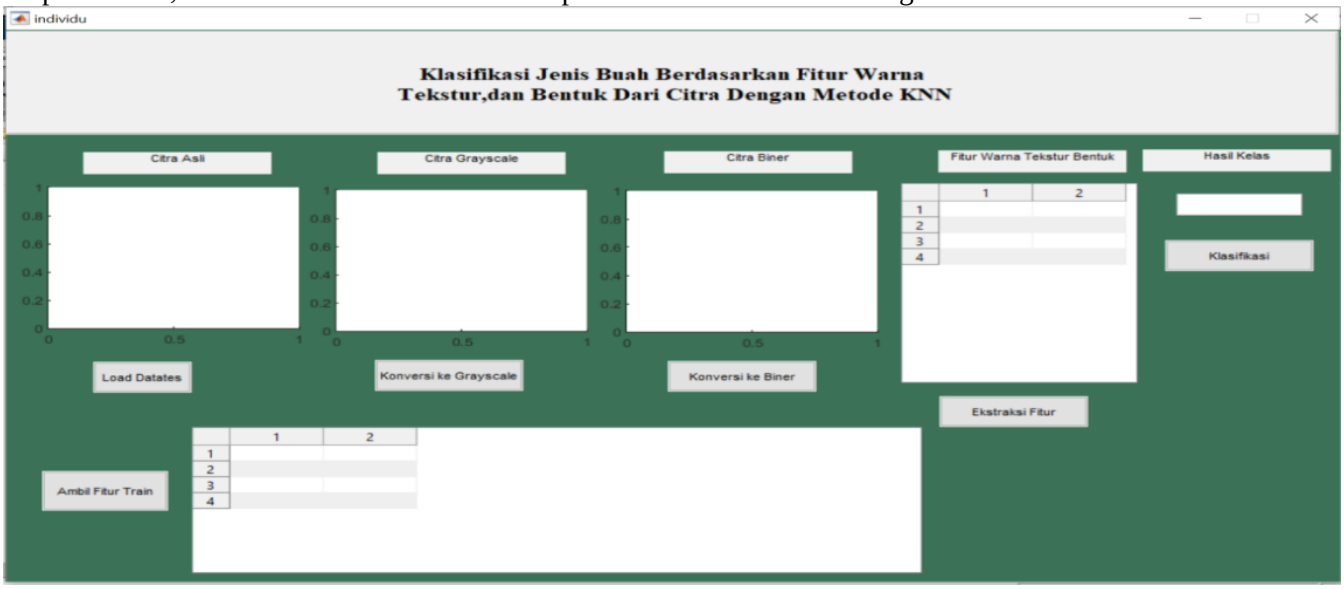


Figure. 3. Interface Process of identifying Fruit Type in Individual

Figure 4 displays a view of determining the overall type of fruit from the test data. Inside there is a menu for load data training and load data testing. How to display the value of the imagery feature extract, texture and shape click the feature extraction process. And there is a classification menu to display the results of the classification.

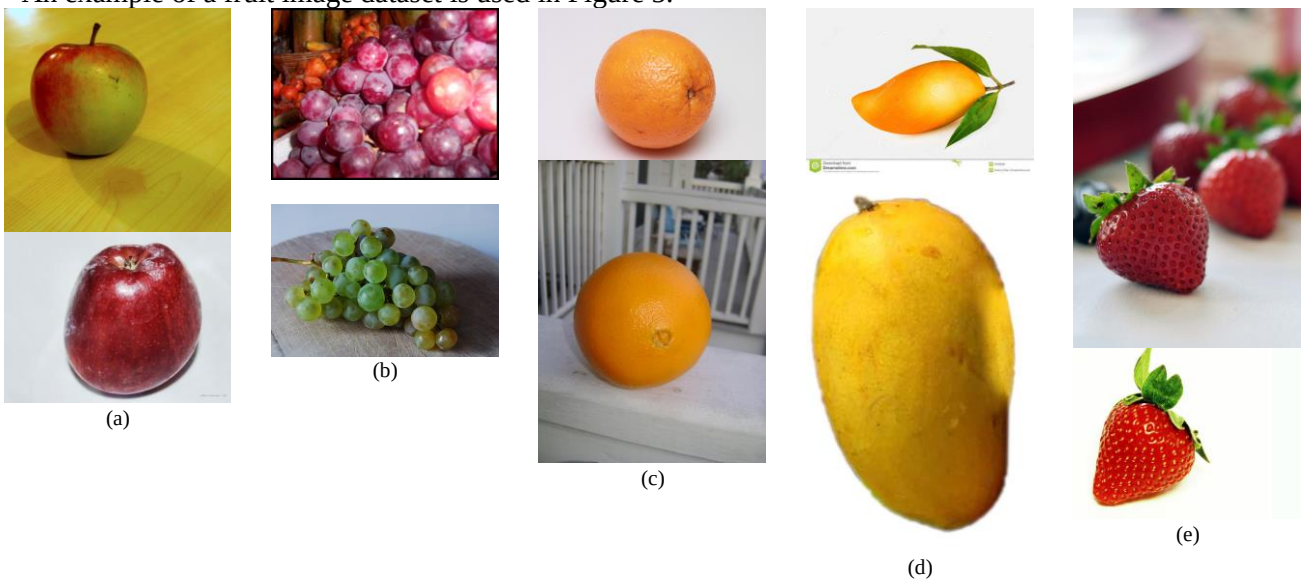
Figure. 4. Interface Process of Identifying Fruit Types in Groups

TABLE III
ACCURACY RESULTS BASED ON K

K Value	Accuaracy Result
K1	99,60%
K2	73,60%
K3	73,60%
K4	72%

K5	69,20%
K6	69,20%
K7	65,60%
K8	67,20%
K9	66,40%
K10	64,40%
Average	64.95%

An example of a fruit image dataset is used in Figure 5.



(a) (b) (c) (d) (e)

Figure 5. Fruit Data Imagery

Table 3 describes the number values of K neighbor in the KNN method for recognizing the type of fruit. The optimal K is 1 with an accuracy of 99.6%. Data tests of 250 imagery were used to test the KNN method with a value of K between 1 to 10. The results of the trial showed that the KNN method was good at recognizing the type of fruit.

V. CONCLUSION

The KNN method we use to determine the type of fruit (apples, grapes, oranges, mangoes, and strawberries) can work well. It is proven that KNN accuracy in identifying the type of fruit has the highest accuracy value of 99.6% with K = 1, while we tried K ten times from K = 1 to K = 10. The result is that K=1 has the highest accuracy and K=10 has the lowest accuracy value of 64.4%.

THANK-YOU NOTE

Thank you to the lecturers of informatics engineering at the Islamic University of Lamongan.

REFERENCE

[1] J. Jamaludin, C. Rozikin, and A. S. Y. Irawan, "Klasifikasi Jenis Buah Mangga dengan Metode Backpropagation," *Techné J. Ilm. Elektrotek.*, vol. 20, no. 1, 2021, doi: 10.31358/techne.v20i1.231.

[2] N. Nafi'iyah, "Klasifikasi Kematangan Buah Mangga Berdasarkan Citra HSV dengan KNN," 2019.

[3] A. Patriot, S. Pamungkas, N. Nafi'iyah, and N. Q. Nawafilah, "K-NN Klasifikasi Kematangan Buah Mangga Manalagi Menggunakan L * A * B dan Fitur Statistik," vol. 4, pp. 1–8, 2019.

[4] A. A. Paturrahman and I. G. P. S. Wijaya, "Analisis Pengenalan Pola Daun Berdasarkan Fitur Canny Edge Detection dan Fitur GLCM Menggunakan Metode Klasifikasi k-Nearest Neighbor (kNN)," *J. Comput. Sci. Informatics Eng.*, vol. 5, no. 1, 2021, doi: 10.29303/jcosine.v5i1.388.

[5] R. K. Salsabiilaa, "DETEKSI KUALITAS DAN KESEGARAN TELUR AYAM RAS BERDASARKAN DETEKSI OBJEK TRANSPARAN DENGAN METODE GREY LEVEL CO-OCCURRENCE MATRIX (GLCM) DAN KLASIFIKASI K-NEAREST NEIGHBOR (KNN)," *TEKTRIKA - J. Penelit. dan Pengemb. Telekomun. Kendali, Komputer, Elektr. dan Elektron.*, vol. 1, no. 2, 2019, doi: 10.25124/tektrika.v1i2.1740.

[6] I. G. Wirayudhana, "Klasifikasi Mutu Buah Jambu Biji Getas Merah Berdasarkan Tekstur Menggunakan Grey Level Co-Occurence Matrix (GLCM) dengan Klasifikasi KNN," *J. Indones. Sos. Teknol.*, vol. 2, no. 6, 2021, doi: 10.36418/jist.v2i6.166.

[7] I. S. Areni, I. Amirullah, and N. Arifin, "Klasifikasi Kematangan Stroberi Berbasis Segmentasi Warna dengan Metode HSV," *J. Penelit. Enj.*, vol. 23, no. 2, pp. 113–116, 2019, doi: 10.25042/jpe.112019.03.

[8] N. Wijaya and A. Ridwan, "KLASIFIKASI JENIS BUAH APEL DENGAN METODE K-NEAREST NEIGHBORS DENGAN EKSTRAKSI FITUR HSV DAN LBP," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 1, 2019, doi: 10.32736/sisfokom.v8i1.610.

[9] E. K. Ratnasari and A. Wikaningrum, "Pengenalan Jenis Buah pada Citra Menggunakan Pendekatan Klasifikasi Berdasarkan Fitur Warna Lab dan Tekstur Co- Occurrence," *J. Inf.*, vol. 1, no. 2, 2016.

[10] Ridho Ary Sumarno, "Aplikasi Klasifikasi Jenis – Jenis Buah Jeruk Menggunakan Metode K-Nearest Neighbor," *Artik. Skripsi Univ. Nusant. PGRI*

Kediri (Universitas Nusant. PGRI Kediri Fak. Tek. Prodi Tek. Inform., 2017.
[11] F. Rizal, "PENERAPAN ALGORITMA BACKPROPAGATION UNTUK KLASIFIKASI JENIS BUAH RAMBUTAN BERDASARKAN FITUR TEKSTUR DAUN," *JATIM J. Apl. Teknol. Inf. dan Manaj.*, vol. 1, no. 2, 2020.
[12] I. Siswanto, E. Utami, and S. Raharjo, "Klasifikasi Tingkat Kematangan Buah Berdasarkan Warna dan Tekstur Menggunakan Metode K-Nearest Neighbor dan Nearest Mena Classifier," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, p. 93, 2020, doi: 10.35585/inspir.v10i1.2559.
[13] S. F. Kusuma, R. E. Pawening, and R. Dijaya, "Otomatisasi klasifikasi kematangan buah mengkudu berdasarkan warna dan tekstur," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 3, no. 1, 2017, doi: 10.26594/r.v3i1.576.
[14] T. Y. Prahudaya and A. Harjoko, "METODE KLASIFIKASI MUTU JAMBU BIJI MENGGUNAKAN KNN BERDASARKAN FITUR WARNA DAN TEKSTUR," *J. Teknosains*, vol. 6, no. 2, 2017, doi: 10.22146/teknosains.26972.
[15] R. Enggar Pawening, W. Ja, and F. Shudiq, "Klasifikasi Kualitas Jeruk Lokal Berdasarkan Tekstur dan Bentuk Menggunakan Metode k-Nearest Neighbor (k-NN)," *J. Ilmu Komput. dan Desain Komun. Vis.*, vol. 1, no. 1, 2020.