International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 9, No. 1, (2025), pp. 1-7

Image Feature Extraction for Determining the Ripeness Level OF Oil Palm Fruits Using the K-Nearest Neighbor Algorithm Based on Color Features (Case PTPN IV Aceh Utara)

Atrida Sudarti[1], Munirul Ula[2], Fajriana[3]
[1.2.3]Informatics Engineering, Universitas Malikussaleh, Lhokseumawe, Indonesia
Email: atridasudarti72@gmail.com

Abstract

The availability of oil palm fruits at the appropriate ripeness level is crucial to achieving optimal oil production. Farmers often struggle to accurately determine fruit ripeness, resulting in inconsistent quality and reduced efficiency. This study aims to develop a classification system to determine the ripeness level oil palm fruits using the K-Nerest Neighbor (K-NN) algorithm based on color features extracted from fruits image. Color is a key indicator of maturity and directly influences oil yield. The data was collected through image acquisition and direct observation at the Cot Girek Palm Oil Mill (PKS) of PTPN IV, Aceh Utara. Image preprocessing was carried out to enhance and nomalize the data before feature extraction. The extracted color features were then used to classify the fruits into ripe and unripe categories using the K-NN algorithm. The results show that K-NN successfully classifies the ripeness level of oil palm fruits with an accuracy of 72.80%. This system provides a recommendation for fruit feasibility before processing, helping reduce production losses caused by immature or overripe fruits. Overall, this research contributes to improving decision-making in the palm oil industry through the application of image processing of machine learning techniques.

Keywords: *Oil Palm Fruits, Image processing, Machine Learning, Clasification*

1. Introduction

Oil palm is a monocotyledonous plant that produce oil and serves as one of Indonesia's primary commodity products. As the world's largest exporter of Crude Palm Oil (CPO), Indonesia consistently maintains and adheres to international standards for oil palm fruit quality, making its CPO products highly sought after by both food and non-food industries worldwide [1]. The palm oil industry, including its derivatives, is a major export commodity that significantly contributes to Indonesia's economy. Variations in perception among farmers or selectors often lead to suboptimal fruit selection based on ripeness. To address this, image processing offers an effective method for accurately determining the ripeness level of oil pal fruits [2]. The presence of digital image classification has led to various studies utilizing different algorithms [3]. Digital images are clear representations of objects that can be processed by computers. Larger image sizes (pixels) require more storage space. Image processing itself. Preprocessing include segmentation by converting the image to grayscale and then to black and white [4]. By using the K-Nearest Neighbor algorithm based on RGB and HSV color features, this study offers a more objective and consistent automates to improve classification accuracy compores to previous methods that used only a single color feature.

2. Research Methodology

2.1. Oil Palm Fruit

The oil-producing part of the oil palm plants is it fruit. Fruit maturity in plantations is determined by the natural detachment of the fruit, know as "membrondol". Detached fruits are called "brondolan". Oil palm fruits vary in color-black, brown, to reddish but are

ISSN : 2580-7250

generally reddish and grow in large buches. The oil content increases as the fruit ripens [5].

2.2. Feature Extraction

Feature extraction is the process of obtaining characteristics from objects within an image [6]. Feature extraction involves retrieving feature values or characteristic from an image. After acquiring oil palm fruit data in .jpg image format, the next step to perform feature extraction. The purpose of tis process is to extract features from the image so they can be used in the classification process [3]. In this study, feature extraction refers to obtaining RGB and HSV values. Color feature extraction is carried out by calculating the average of each RGB and HSV component in the image. The mean RGB values obtained from the training and testing data are then used as input for the classification process [7].

2.3. Digital Image Processing

Digital image processing is a technique for processing image using various algorithms. For example, a slightly dark photo can be processed to appear brighter this in one the processes that can be performed through digital image processing [2]. Digital image processing has been applied in many applications and fields. It is not only limited to adjusting the spatial resolution of an image or enhancing its brightness but is also capable of analyzing elements that can be captured by a camera yet are invisible to the human eye, such as color spectra displayed by electromagnetic waves [8].

2.4. K-Nearest Neighbor

The K-Nearest Neighbor (KNN) method is a technique for grouping new data based on the k-nearest neighbor distance between training and testing information [7]. The K-Nearest Neighbor algorithm is a supervised learning algorithm in which the result of a new instace is classified based on the majority class of its k-nearest neighbors. The main objective of this algorithm is to classify new objects based on their attributes and the samples from the training data. The K-NN algorithm uses classification as a predictor for the value of a new instance. K-Nearest Neighbor is a cased-based reasoning method thet works by measuring the closeness between a new case and existing cases, based on the similarity or wight matching of several features [9]. The following are the steps of the K-NN process:

1. Determine the parameter K (the number of nearest neighbors)
2. Calculate the squared Euclidean distance of each object to the given sample data.

$$d = \sqrt{\{(x_2 - x_1)^2 + (y_2 - y_1)^2\}} \quad (1)$$

   Explanation:
   x_1y_1: Coordinates of the first point.
   x_2y_2: Coordinates of the second poin.
   d: Distance.
3. Sort the calculated distances in ascending order.
4. Select the corresponding distance up to k.
5. Assign the corresponding classes to the selected neighbors.
6. Count the number of classes among the k nearest neighbors and assign the most frequent class as the class of the data to be evaluated.

2.5. Color Features

Color features are important in images because color can be perceived by the human eye. The color features that can be extracted from an image include the mean, standard deviation, and skewness [10].

International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 9, No. 1, (2025), pp. 1-7

2.5.1. Color Space RGB (Red, Green, Blue)

According[2], the human eye's sensitivity to red, green, and blue forms the basis of the RGB (true color) image theory. When these three color components are combined, they produce other colors known as additive colors. RGB color settings involve red, green, and blue, where the color of each pixel is determined by the combination of red, green, and blue intensities stored in each color channel at as specific pixel location. Graphic file formats store RGB images as 24-bit images, with each red, green, and blue component allocated 8 bits.

2.5.2. Color HSV (Hue, Saturation, and Value)

In addition to RGB, another model used in color analysis for digital image processing is HSV, which stands for Hue, Saturation, and Value. Hue represents the actual color, such as red, violet, or yellow, and is used to determine color characteristic like redness, greenness, and others. Saturation (or chroma) refers to the purity or intensity of the color. Value indicates the brightnees of the color and ranges from 0 to 100%. If the value is 0, the color appears black; the higher the value, the brighter the color becomes, producing various shades of the color [2].

2.6. Visual Studio Code

Visual Studio Code is are free code editor developed by Microsoft for Windows, Linux, and macOS. It supports multiple programming languages like C++, C#, Java, Python, and PHP. The editor cam detect the programming language and apply syntax highlighting. It also integrates with GitHub and allows developers to extend its functionality through plugins[11]

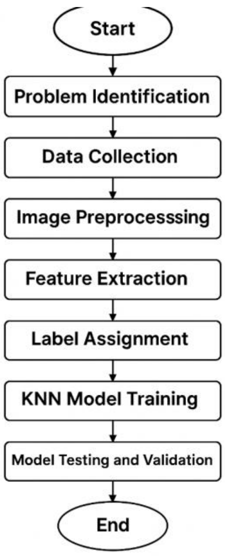

Figure 1. Research Flowchart

3. Result and Discusstion

3.1. System Testing

This testing is conducted to ensure that designed system functions properly and meets user requirements. The user interface in this study consists of three pages: The main page, the output or classification result page, and calculation result page of the K-Nearest Neighbor (KNN) algorithm. The implementation result of the main page can be seen in Figure 2. The output page or classification result page and accuracy result are used to display the analysis results of the palm oil fruit image that have been previously uploaded by the user. The interface of this page is shown in Figure 3.

International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 9, No. 1, (2025), pp. 1-7

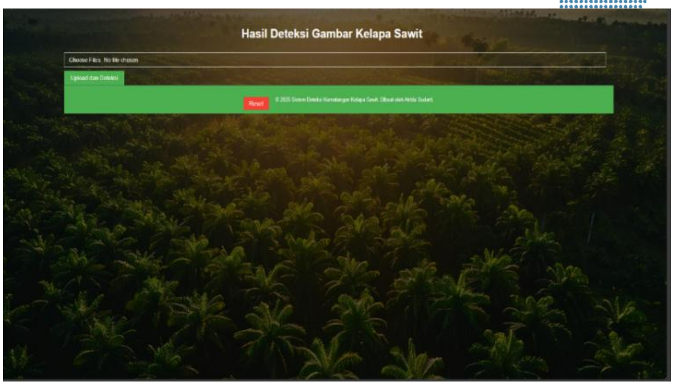

Figure 2. Main Interface Layout

The image above is part the main page (homepage) or the initial user interface of the Palm Oil Ripeness Detection System application. This page typically functions as the home screen where users begin the main process, which is uploading an image of palm fruit to be analyzed.

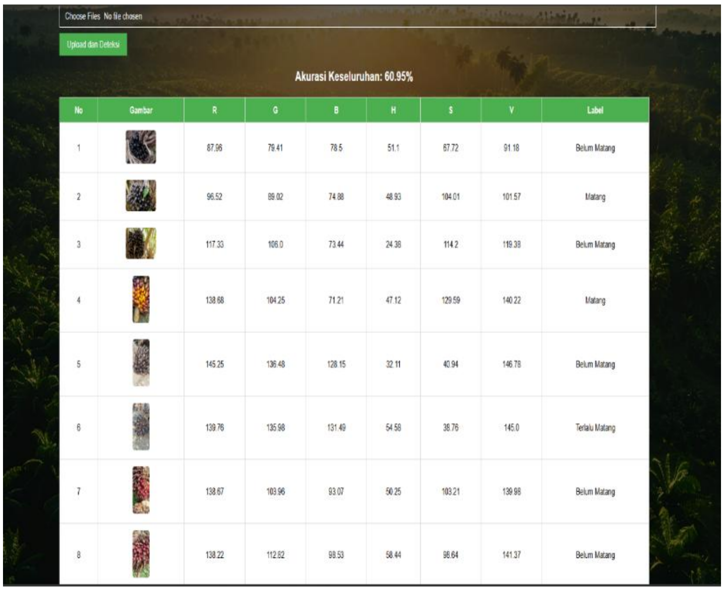

Figure 3. Clasification Result Page

This page is used to display the results of the analysis from oil palm image previously uploaded by the user.

3.2. Image Color Extraction RGB to HSV

The results of the color image extraction from RGB to HSV features can be seen in the image below.

Figure 4. RGB to HSV color extraction

International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 9, No. 1, (2025), pp. 1-7

The image above shows the stages of oil palm fruit image extraction. The first image is the original, converted from BGR to RGB. The second is grayscale transformation, retaining light intensity and removing color. The third is the HSV conversion, separating Hue, Saturation, and Value, which are used for fruit ripeness.

3.3. K-Nearest Neighbor Classification results

The provided images show tabular data for classification oil palm fruit ripeness. Each row represents a sample with colorimetric features (R, G, B, H, S, V) and a "Label" indicating ripeness levels, such as "Ripe" or "Unripe". "The accuracy of "72.80%" suggests the data is from a classification model evaluation.

Figure 5. Accuracy results of the test data calculation system

3.4. Manual Calculation Using the KNN Algorithm

Tabel 1. KNN Calculation of Test Data

No	R	G	B	H	S	V	Label Asli	Jarak	Prediksi Sistem
1	133.62	118.25	109.69	49.11	51.79	136.05	Matang	0.00	Matang
2	138.02	118.00	100.31	33.64	72.26	136.67	Matang	27.77	Matang
3	136.00	114.98	102.30	49.67	55.82	133.82	Matang	10.15	Matang
4	130.54	109.66	104.77	68.61	65.77	136.60	Matang	26.52	Matang
5	139.51	130.39	123.89	55.35	46.36	146.27	Matang	22.97	Matang
6	133.78	110.69	98.89	40.88	68.90	154.90	Matang	30.16	Matang
7	127.32	107.51	102.10	72.45	67.41	132.13	Terlalu Matang	32.28	Matang
8	135.34	119.94	112.30	86.31	61.47	141.79	Matang	38.98	Matang
9	145.41	138.08	132.95	71.68	45.78	153.61	Matang	43.34	Matang
10	122.72	111.40	101.01	54.99	64.54	128.19	Matang	38.98	Matang
11	118.88	105.05	98.91	64.70	56.72	124.16	Matang	43.34	Matang
12	137.87	121.91	125.09	92.75	47.56	144.54	Matang	22.76	Matang
13	102.57	126.88	126.75	72.35	40.67	140.84	Matang	30.79	Matang
14	129.25	122.54	122.50	80.01	34.12	154.53	Matang	47.44	Matang
15	132.17	121.85	118.37	65.41	44.16	138.35	Matang	20.24	Matang
16	133.56	109.96	96.55	59.48	82.81	137.09	Matang	36.53	Matang
17	119.18	100.58	83.30	52.03	71.72	122.85	Matang	76.71	Matang
18	132.72	119.69	115.83	83.59	53.56	137.38	Matang	35.07	Matang
19	133.09	107.50	93.47	48.70	85.03	134.60	Matang	38.92	Matang
20	119.66	81.87	63.89	21.54	117.83	118.70	Matang	95.55	Matang
21	143.88	95.48	78.82	28.28	133.13	143.75	Belum Matang	93.51	Matang
22	141.15	103.36	131.89	91.26	45.61	143.79	Belum Matang	51.65	Matang
23	126.35	108.83	100.45	79.40	75.75	134.47	Belum Matang	41.76	Matang
24	107.58	160.78	148.79	24.97	86.11	167.59	Belum Matang	81.67	Matang
25	146.47	117.91	101.03	20.37	92.35	149.48	Belum Matang	53.82	Matang

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

$$d_{17} = \sqrt{\{(x_1 - x_{17})^2 + (y_1 + y_{17})^2}$$

$= \sqrt{\left((133.62 - 199.18\,)^2 + (118.25 - 100.58)^2\right) + \left(109.69 - 83.3\right)^2 + \;(49.11 - 52.03)^2 +)\right)}$
$\sqrt{(51.79 - 71.72\,)^2 \;+ (136.05 - 122.85)^2))}$
$= \sqrt{(-65.56\,)^2 \;+ (17.67)^2 + \;(26.39)^2 + \;(-2.92)^2 + \;(-19.93)^2 + \;(13.20)^2))}$
$= \sqrt{4298.11 + 312.22 + +696.42 + 8.52 + 397.21 + 174.24}$
$= \sqrt{5886.72}$
= 76.71

Repeat the calculation as follows for all other training data:

1. Select the K nearest neighbor (K=3)
2. Check the labels of these 3 neighbors.
3. Use majority voting to determine the most frequent label → this is KNN prediction result.

For K=3
If we take the 3 nearest neighbors:
No.6 → 30.16
No.23 →41.79
No.17 →76.71
Majority label among the neighbors:

1. Ripe: 2 (No.6 and No.17)
2. Unripe: 1 (No.23)

Why choose those 3 specific neighbors (No.6,23,17)?

1. Because their distance to the test data are the smallest compared to others.
2. KNN select neighbors based on the closest distances.
3. Neighbors that are closer are considered more "similar" to the test data.

So, the smaller the distance, the more relevant they are for class voting.
For example, distance to No.6 = 30.16, to No =41.79, to No = 1776.71 and these are the three closest compared to other training data. Using the Formula:

$$Akurasi = \frac{Jumlah\ Prediksi\ benar}{Jumlah\ data\ uji} \times 100\ \%$$

$Akurasi = \frac{18}{25} \times 100\% \;= 72\ \%$ (Systematic rounding)

And on the system display, the author shows an accuracy value of 72.80%.

Therefore, the accuracy result of 72.80% is the closets match to the outcome produced by the system.

4. Conclusion

This study proves that the K-Nearest Neighbor (K-NN) algorithm is effective in classifying the ripeness level of oil palm fruit based on RGB and HSV color features. The system achieved an accuracy of 68.80% on training data and 72.80 on test data, indicating that the method performs well. The developed system has also successfully automated the processes of image upload, feature extraction, and ripeness classification, showing great potential for implementation in the industry, particularly in work environments such as PTPN IV Aceh Utara.

References

[1] 2023 Pusadan et al., "The Image Extraction Using the HSV Method to Determine the Maturity Level of Palm Oil Fruit with the k-nearest Neighbor Algorithm," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 6, pp. 1448–1456, 2023, doi: 10.29207/resti.v7i6.5558.

[2] 2020 Himmah et al., “Identifikasi Kematangan Buah Kelapa Sawit Berdasarkan Warna RGB Dan HSV Menggunakan Metode K-Means Clustering,” *J. Sains dan Inform.*, vol. 6, no. 2, pp. 193–202, 2020, doi: 10.34128/jsi.v6i2.242.

[3] M. H. Hanafi, N. Fadillah, and A. Insan, “Optimasi Algoritma K-Nearest Neighbor untuk Klasifikasi Tingkat Kematangan Buah Alpukat Berdasarkan Warna,” *It J. Res. Dev.*, vol. 4, no. 1, pp. 10–18, 2019, doi: 10.25299/itjrd.2019.vol4(1).2477.

[4] Z. D. Lestari, N. Nafi’iyah, and P. H. Susilo, “Sistem Klasifikasi Jenis Pisang Berdasarkan Ciri Warna HSV Menggunakan Metode K-NN,” *Semin. Nas. Teknol. Inf. dan Komun.*, pp. 11–15, 2019.

[5] S. Salsabilla, I. Nirmala, and T. Rismawan, “Sistem Pemilah Otomatis Tingkat Kematangan Buah Kelapa Sawit Menggunakan Metode Logika Fuzzy Mamdani Dan Sensor TCS3200,” *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 144–154, 2023, doi: 10.47065/josyc.v5i1.4449.

[6] M. Widyaningsih, “Identifikasi Kematangan Buah Apel Dengan Gray Level Co-Occurrence Matrix (GLCM),” *J. SAINTEKOM*, vol. 6, no. 1, p. 71, 2017, doi: 10.33020/saintekom.v6i1.7.

[7] C. Paramita, E. Hari Rachmawanto, C. Atika Sari, and D. R. Ignatius Moses Setiadi, “Klasifikasi Jeruk Nipis Terhadap Tingkat Kematangan Buah Berdasarkan Fitur Warna Menggunakan K-Nearest Neighbor,” *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 1–6, 2019, doi: 10.30591/jpit.v4i1.1267.

[8] R. Ravikumar and V. Arulmozhi, “Digital Image Processing-A Quick Review,” *Int. J. Intell. Comput. Technol.*, vol. 2, no. 2, pp. 16–24, 2019, [Online]. Available: https://ijict.com.

[9] Ar Razi, “Klasifikasi Penerima Beasiswa Aceh Carong (Aceh Pintar) Di Universitas Malikussaleh Menggunakan Algoritma Knn (K-Nearest Neighbors),” *J. Tika*, vol. 7, no. 1, pp. 79–84, 2022.

[10] D. Hernando, A. W. Widodo, and C. Dewi, “Pemanfaatan Fitur Warna dan Fitur Tekstur untuk Klasifikasi Jenis Penggunaan Lahan pada Citra Drone,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 2, pp. 614–621, 2020, [Online]. Available: http://j-ptiik.ub.ac.id.

[11] R. Rahman and F. Andreas Sutanto, “Data Mining to Predict Gojek’s Consumer Satisfaction Level Using Naive Bayes Algorithm,” *Int. J. Inf. Syst. Technol. Akreditasi*, vol. 6, no. 158, pp. 590–602, 2023.