



Automated Ripeness Detection of Oil Palm Fruit Using a Hybrid GLCM-HSV-KNN Model

Mirfan¹, Billy Eden William Asrul², Mila Jumarlis³, Juliani⁴

^{1,2,4}Department of Computer Science, Faculty of Computer Science, Handayani University Makassar, Makassar, Indonesia

³Islamic Communication and Broadcasting, STAIN Majene, Majene, Indonesia

¹mirfan@handayani.ac.id, ²billy@handayani.ac.id, ³mila.jumarlis@stainmajene.ac.id, ⁴juliani@handayani.ac.id

Abstract

Accurately determining the ripeness of oil palm fruit is crucial for ensuring the quality of palm oil. However, traditional manual methods are often time-consuming and less accurate. This study aimed to develop an automated system for detecting the ripeness of oil palm fruit by combining the Hue Saturation Value (HSV) model, Gray Level Co-occurrence Matrix (GLCM), and K-Nearest Neighbor (KNN) algorithms. This system utilizes K-Nearest Neighbors to classify the relationship between color features extracted using the HSV model and texture features derived from GLCM analysis to categorize fruit ripeness. The color features represent the fruit's chromatic characteristics associated with ripeness, while the texture features provide information regarding surface patterns related to ripeness. The color features represent the fruit's color characteristics associated with ripeness, whereas the texture features provide information about the surface patterns related to ripeness. The results indicate that the system can classify oil palm fruit into four distinct categories: Over-Ripe, Ripe, Half-Ripe, and Raw. The dataset was divided with an 80:20 ratio, where 80% was allocated for training data and the remaining 20% for test data. An accuracy rate of 85% was achieved. The results of this study demonstrate that the developed system effectively classifies oil palm fruit images based on ripeness levels. This system supports a sustainable automated palm oil production model through accurate ripeness detection, thereby reducing reliance on manual methods and enhancing consistency and productivity in palm oil processing. These findings indicate that the proposed hybrid method is feasible for integration into an automated classification system to support decision-making in oil palm harvesting.

Keywords: HSV; GLCM; KNN; image processing; oil palm fruit ripeness

How to Cite: Mirfan, Billy Eden William Asrul, Mila Jumarlis, and J. Juliani, "Automated Ripeness Detection of Oil Palm Fruit Using a Hybrid GLCM-HSV-KNN Model", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 5, pp. 1066 - 1077, Oct. 2025.

Permalink/DOI: <https://doi.org/10.29207/resti.v9i5.6683>

Received: May 28, 2025

Accepted: September 20, 2025

Available Online: October 12, 2025

This is an open-access article under the CC BY 4.0 License

Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Oil palm (*Elaeis guineensis* Jacq) is one of Indonesia's leading commodities and plays a strategic role in the national economy, particularly in supplying raw materials for the palm oil industry (CPO – Crude Palm Oil). The quality of the produced palm oil is influenced by several factors, including moisture content, impurity levels, free fatty acid content, and the ripeness level of the oil palm fruit at the time of harvest. Among these factors, the ripeness level is one of the most crucial variables that determine the quality and quantity of the oil produced [1].

Several companies in the palm oil industry apply a selection or sorting process for fresh fruit bunches (FFB) as an initial stage before processing [2]. This sorting is generally performed manually, relying on

visual inspection of color and texture by field workers to determine fruit ripeness. Although widely practiced, this manual method has notable limitations. It depends heavily on the subjective judgment of workers, is prone to fatigue-related errors due to repetitive tasks, and is time-consuming, especially across large plantation areas[3]. Misjudging the ripeness stage—whether unripe, under-ripe, ripe, or overripe—can significantly affect both the quality and quantity of the palm oil produced. Such inaccuracies reduce oil yield and impact the chemical and physical characteristics of the final product [4]. As a result, there is a growing need for more objective and efficient sorting methods. Technological solutions involving image processing and artificial intelligence offer promising alternatives, enabling faster, more consistent, and more accurate classification of fruit maturity levels [5]. As technology

advances, the application of digital image processing systems offers a more robust and streamlined alternative for detecting oil palm fruit ripeness [6]. Image processing provides objective and consistent visual analysis based on the color and surface texture characteristics of the fruit. The implementation of this technology not only improves classification accuracy but also speeds up the ripeness identification process, thereby supporting the optimization of harvesting and production processes [7]. Additionally, the use of automated systems reduces dependence on manual labor in the sorting process, which in turn contributes to overall operational effectiveness.

Several studies have explored the classification of oil palm fruit ripeness using digital image analysis, employing various approaches for both feature extraction and classification [8] - [10]. One such study was conducted, which detected the ripeness level of Fresh Fruit Bunches (FFB) of oil palm based on color composition using the K-Nearest Neighbor (K-NN) algorithm. The system achieved an accuracy rate of 92% under controlled indoor lighting conditions but saw a decrease in accuracy to 84% when images were taken outdoors under unstable natural lighting. This indicates that lighting plays a critical role in determining the accuracy of classification systems. [11], [12].

In previous studies, a color-based approach was commonly employed, such as feature extraction using RGB or HSV combined with classification methods like Naïve Bayes, Decision Tree, and Support Vector Machine (SVM). While color-based approaches are generally effective in differentiating ripeness stages based on visible surface characteristics, they are often limited in robustness when faced with variations in lighting conditions or inconsistent image backgrounds, which can significantly affect classification accuracy. [13] - [15].

To address these limitations, recent studies have begun incorporating texture features as additional parameters to enhance classification performance. Methods such as the Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and Haralick Texture Features have been utilized to extract surface information that is not captured by color features alone. Research combining both texture and color features has demonstrated improved performance compared to single-feature approaches, as this integration enables a more comprehensive representation of the complex visual changes that occur during the ripening process of oil palm fruits [16].

In terms of classification algorithms, SVM and Artificial Neural Networks (ANN) have been widely adopted. SVM is particularly effective in handling high-dimensional and non-linear data but requires intensive parameter tuning and a relatively complex training process. ANN, on the other hand, offers the ability to learn deep, non-linear representations but typically

demand large training datasets and considerable computational resources [17]. In contrast, this study adopts the K-Nearest Neighbors (K-NN) algorithm as the classification method due to its simplicity, efficiency with medium-sized datasets, and its ability to achieve competitive performance without the need for complex training procedures. K-NN operates based on the proximity of features in a multidimensional space, making it well-suited for the simultaneous integration of both color and texture information. [18], [19].

HSV (Hue, Saturation, Value) color model is a color representation that separates hue, color saturation, and brightness components, making it more aligned with human visual perception compared to the RGB model. In the context of oil palm fruit ripeness identification, the skin color changes during the ripening process from green, yellow, orange, to deep red can be quantitatively captured through the Hue (H) component [20]. Hue reflects the dominant color present on the fruit's surface and serves as a primary indicator of ripeness level. Meanwhile, Saturation (S) represents the color intensity, where ripe fruits typically exhibit more vivid and saturated colors than unripe ones. [21] The Value (V) component denotes brightness, with ripe fruits generally appearing brighter due to surface changes that influence light reflection. For feature extraction purposes, the palm fruit images are converted from the RGB color space to HSV, and the average values or histograms of each HSV component are computed to serve as numerical features in the classification process.

In addition to color transformation, ripeness level can also be characterized by surface texture variations. Therefore, texture analysis plays a significant role in digital image processing for fruit maturity classification. One of the most widely used statistical methods for texture feature extraction is the Gray Level Co-occurrence Matrix (GLCM) [22]. GLCM calculates the frequency of pixel pairs with specific gray levels at defined distances and orientations within a grayscale image, generating a matrix that captures the spatial distribution of texture patterns. Several statistical descriptors can be derived from the GLCM, including: Contrast, which measures the local intensity variation and highlights differences between adjacent pixels; Correlation, which quantifies the linear dependency between neighboring pixels and reflects textural regularity; Energy (Angular Second Moment), which indicates image uniformity and compactness of texture; and Homogeneity, which reflects how close the distribution of elements is to the GLCM diagonal, indicating local smoothness of the image. Feature extraction is performed across multiple directions (0°, 45°, 90°, and 135°) and pixel distances to comprehensively capture textural characteristics. These features are then utilized as input for the classification of oil palm fruit ripeness levels [23], [24].

Based on those findings, a combination of color feature extraction using Hue, Saturation, Value (HSV) and texture features using the Gray Level Co-occurrence

Matrix (GLCM) is considered relevant to improve the precision of classifying oil palm fruit ripeness[25]. This combination can more comprehensively represent the fruit's visual characteristics, where HSV captures the color spectrum and GLCM provides information on the surface texture pattern[26]. To enhance classification accuracy, the extracted features are then processed using the K-Nearest Neighbor (K-NN) algorithm, which classifies data based on the similarity distance to previously labeled training data [27].

This study proposes a novel approach to oil palm fruit ripeness classification through a multi-feature digital image processing system that integrates both color and texture characteristics. The main novelty lies in the system's ability to classify fruits into four ripeness categories: unripe, half-ripe, ripe, and overripe—unlike previous studies that were typically limited to only two or three categories [28]. The inclusion of half-ripe and overripe categories is a significant distinguishing aspect as it enables more accurate detection of the ripeness transition phase. This is crucial in selective harvesting practices, where misclassification during the transitional stage often leads to harvesting fruits either prematurely or too late.

The system employs the HSV color model to capture more representative and lighting-robust color gradients, and extracts texture features using the GLCM method to identify visual differences such as surface wrinkling or roughness. These combined features are classified using the K-Nearest Neighbor (K-NN) algorithm, which demonstrates a high capability in handling multidimensional features with strong performance on limited data [29]. With this approach, the study makes a tangible contribution to the development of a more precise, objective, and applicable oil palm fruit ripeness classification system, supporting optimized harvesting automation in plantation industries.

Accurate classification of oil palm fruit ripeness brings significant benefits in enhancing the effectiveness and productivity of the palm oil industry. Proper ripeness identification allows for selective and timely harvesting, which can increase oil yield and reduce losses due to harvesting fruits that are either unripe or overripe. Fruits harvested at optimal ripeness produce higher oil content with better quality and lower free fatty acid (FFA) levels, which negatively impact oil quality. Furthermore, the implementation of an image-processing-based automated classification system can reduce reliance on manual labor, increase assessment objectivity, and accelerate field inspection processes. Therefore, this technology plays a vital role in realizing sustainable precision agriculture practices in the oil palm plantation sector.

2. Methods

The research method aims to develop a classification system for the ripeness level of oil palm fruit based on digital image processing, utilizing a combination of

color and texture feature extraction techniques as the primary approach in image analysis. This system is developed through several integrated stages, including data acquisition, image preprocessing, feature extraction, classification using the K-Nearest Neighbor (K-NN) algorithm, and system performance evaluation.

2.1 Data Acquisition

The dataset used in this study comprises oil palm fruit images collected directly from plantations located in West Sulawesi, Indonesia. Image acquisition was carried out using a high-resolution digital camera, specifically the Canon EOS 700D with an 18-megapixel sensor, under semi-controlled lighting conditions to ensure consistent image quality. The shooting distance and angle were standardized to minimize inter-sample variability, and all images were captured against a neutral background to facilitate segmentation and subsequent preprocessing steps. Dataset was divided with an 80:20 ratio, where 80% was allocated for training data and the remaining 20% for test data. The labeling process was conducted manually through visual observation by experienced plantation experts, who possess practical knowledge in assessing oil palm fruit ripeness. The ripeness levels were categorized into four distinct classes: unripe, half-ripe, ripe, and overripe, based on visual features such as skin color and surface texture. Labeling was performed in accordance with commonly accepted industrial harvesting standards, and to ensure label consistency, each image was reviewed by at least two experts, with discrepancies resolved through discussion and consensus. In total, the dataset consists of 600 images, evenly distributed across the four ripeness categories, with 150 images per class. The dataset was further split into 80% for training and 20% for testing to maintain class balance and prevent bias during the classification process. To ensure the reliability and reproducibility of the labeling process, each sample was cross-validated by multiple experts. This expert-driven and standard-based approach ensures that the dataset annotations are consistent, traceable, and applicable to real-world agricultural practices. This approach ensures the reliability and reproducibility of the labeling process. Sample of an overripe oil palm fruit can be seen in Figure 1, while a ripe oil palm fruit is shown in Figure 2.

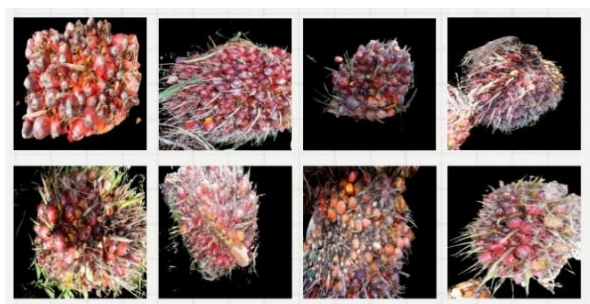


Figure 1. Overripe oil palm fruit image

The half-ripe oil palm fruit is illustrated in Figure 3, and the unripe oil palm fruit is depicted in Figure 4. All images were collected with the informed consent of the plantation owners, and no ethical issues were encountered during the data collection process.

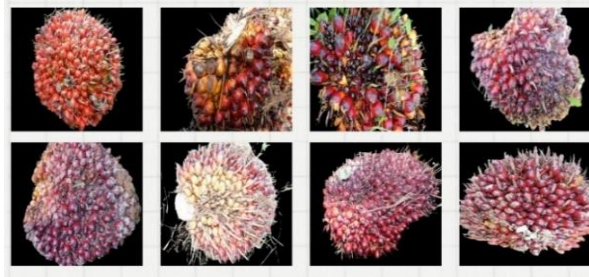


Figure 2. Ripe oil palm fruit image

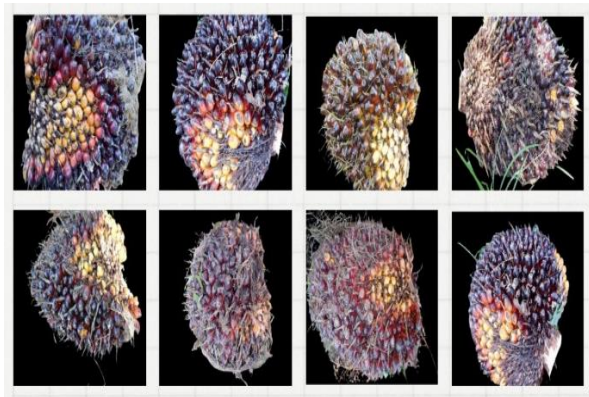


Figure 3. Half-ripe oil palm fruit image

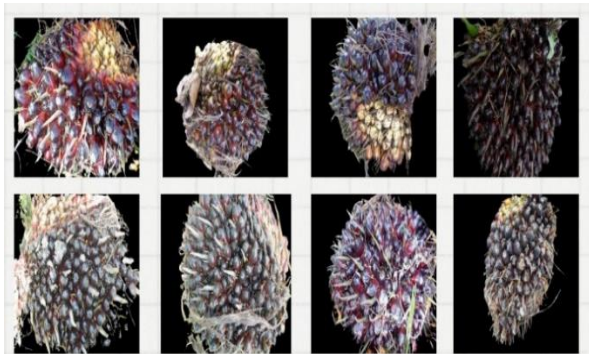


Figure 4 Unripe oil palm fruit image

2.2 Image Preprocessing

After the images are obtained, preprocessing is performed to enhance image quality and standardize the format before the feature extraction process. The preprocessing steps include cropping, resizing, and color conversion. Image cropping process is intended to isolate and extract only the region of interest (ROI) that contains the primary object—in this case, the oil palm fruit—so that it becomes the sole focus of subsequent analysis. This step helps eliminate irrelevant elements such as background, leaves, stems, or even the harvester's hand, thereby allowing the system to concentrate on the most significant visual features. Cropping can be performed either manually or automatically. In manual cropping, the operator selects

and trims the fruit region directly using software tools such as LabelImg or OpenCV-based applications. This approach is suitable for small datasets or during the initial labeling phase. In contrast, automatic cropping typically involves image segmentation techniques such as background subtraction, color thresholding (e.g., targeting reddish or orange hues indicative of ripeness), or contour detection to identify object boundaries. Once the fruit region is detected, a bounding box is generated to define the cropping area. The resulting cropped image is then resized to a fixed dimension, typically 128×128 pixels, to ensure uniformity across all dataset samples. This process not only enhances computational efficiency and processing speed but also significantly improves the accuracy of feature extraction—both in terms of color (HSV) and texture (GLCM)—as it focuses exclusively on the relevant region. Consistent and precise cropping contributes to the overall robustness and optimal performance of the classification system by minimizing visual disturbances from the surrounding environment.

The resizing process of oil palm fruit images, which involves altering the original image dimensions, must be conducted carefully, as it can result in the loss of critical visual information that directly affects the accuracy of image-based classification systems. When images are resized to smaller dimensions, fine visual details such as surface texture, color gradients, and small object contours may be lost due to pixel merging or elimination. This issue is particularly crucial in the classification of oil palm fruit ripeness, where both color and texture serve as primary indicators.

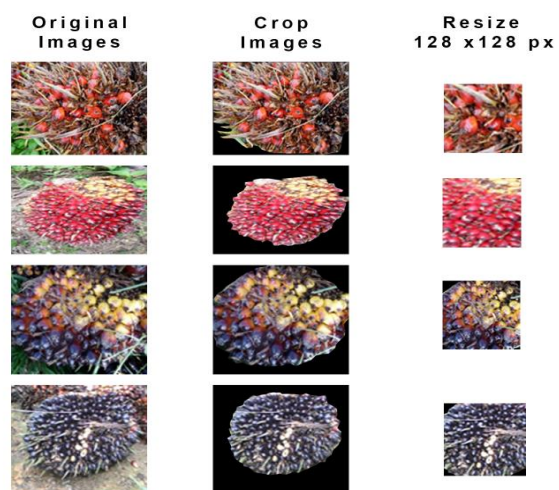


Figure 5. Preprocessing Palm Fruit Images

The loss of such details can negatively impact the accuracy of feature extraction using the HSV color model and GLCM texture analysis, as both methods rely heavily on the clarity of local visual structures within the image. Additionally, if resizing does not maintain the original aspect ratio, the shape of the object may become distorted, potentially misleading classification algorithms, particularly those sensitive to geometric features. Although resizing contributes to

data standardization and improves processing speed, selecting dimensions that are too small may result in less informative features, whereas excessively large sizes increase computational demands. Therefore, resizing must be approached with caution, balancing computational efficiency and the preservation of essential visual information—typically by adopting standard sizes such as 128×128 pixels and applying appropriate interpolation methods to maintain image quality, as shown in Figure 5.

2.3 Feature Extraction

Feature extraction is performed to obtain the color and texture characteristics of oil palm fruit images. HSV (Hue, Saturation, Value) and GLCM (Gray Level Co-occurrence Matrix) are two widely used feature extraction methods in digital image processing, each capturing distinct aspects of an object's visual representation. HSV is a color model that separates visual information into three components: Hue, which represents the dominant color (e.g., red or green); Saturation, which measures the intensity or purity of the color; and Value, which indicates brightness. In the context of fruit ripeness analysis, HSV is particularly useful, as changes in the skin color of oil palm fruits throughout the ripening process—from green to yellow, orange, and deep red—can be quantitatively represented through these components. HSV features offer a clear visual indicator of maturity stages but are sensitive to lighting variations and background inconsistencies.

On the other hand, GLCM is a statistical method for extracting texture features from grayscale images by calculating the frequency of co-occurring pixel intensity pairs at specific distances and orientations. From the resulting co-occurrence matrix, several texture descriptors can be derived, including contrast (measuring local intensity variation), correlation (assessing the linear relationship between neighboring pixels), energy or Angular Second Moment (representing texture uniformity), and homogeneity (indicating the closeness of the pixel distribution to the diagonal of the matrix).[30] Unlike HSV, which focuses on color-based visual appearance, GLCM provides insight into the surface structure and patterns of the object—such as roughness or smoothness—which also evolve as the fruit matures.

The fundamental difference between HSV and GLCM lies in the type of information they extract. HSV emphasizes color and visual cues that align with human perception, while GLCM captures spatial patterns in pixel intensity that reflect surface texture characteristics. HSV is generally more sensitive to lighting conditions, whereas GLCM tends to be more robust in such environments. The integration of these two feature types allows a classification system to comprehensively represent fruit characteristics, as both color and texture serve as crucial indicators in determining the ripeness level of oil palm fruits, which

are used to recognize the texture patterns of oil palm fruit at various ripeness levels.

2.4 K-Nearest Neighbor (K-NN) Classification Algorithm

The data obtained from feature extraction is classified using the K-Nearest Neighbor (K-NN) algorithm, which works by comparing the similarity between test data and training data based on Euclidean distance. This study uses variations of $K = 3$, $K = 5$, and $K = 7$ to determine the optimal parameter. The distance calculation between features is performed using the Euclidean Distance formula.

K-Nearest Neighbor (K-NN) was selected in this study for its simplicity and efficiency in handling medium-sized datasets. It can effectively combine multidimensional features from HSV and GLCM without requiring complex training. However, its performance depends heavily on the choice of K , where too small a value may cause sensitivity to noise, while too large a value may reduce class distinction.

2.5 System Performance Evaluation

The classification system evaluation is performed using a Confusion Matrix, which measures the accuracy, precision, and recall of the system's predictions [29]. Accuracy is calculated by comparing the number of correct predictions to the total test data, while precision and recall are used to assess the system's accuracy and sensitivity in classifying the ripeness levels of oil palm fruit. Figure 6. illustrates the system architecture of this study, which consists of two main phases: the training phase and the testing phase, designed to extract and classify visual information from palm fruit images using a combination of color and texture features. In the training phase, the system receives a collection of palm fruit images as input. These images undergo a preprocessing stage that includes resizing, color normalization, noise removal, and conversion to HSV or grayscale color spaces to ensure consistent quality and format for feature extraction. Following preprocessing, the system performs two parallel feature extraction processes. The first utilizes the HSV (Hue, Saturation, Value) method to extract dominant color attributes, including hue, saturation, and brightness. Hue captures the transition in skin color during ripening (e.g., from green to deep red), saturation measures the intensity of the color—which typically increases with maturity—and value represents brightness, which may vary depending on surface texture and reflectance. The second extraction applies the Gray Level Co-occurrence Matrix (GLCM) method to obtain texture features from grayscale images. GLCM analyzes the spatial relationships of pixel intensities by computing the frequency of gray-level co-occurrence at specified distances and orientations, producing statistical descriptors such as contrast, correlation, energy, and homogeneity, which represent surface roughness, texture regularity, and uniformity.

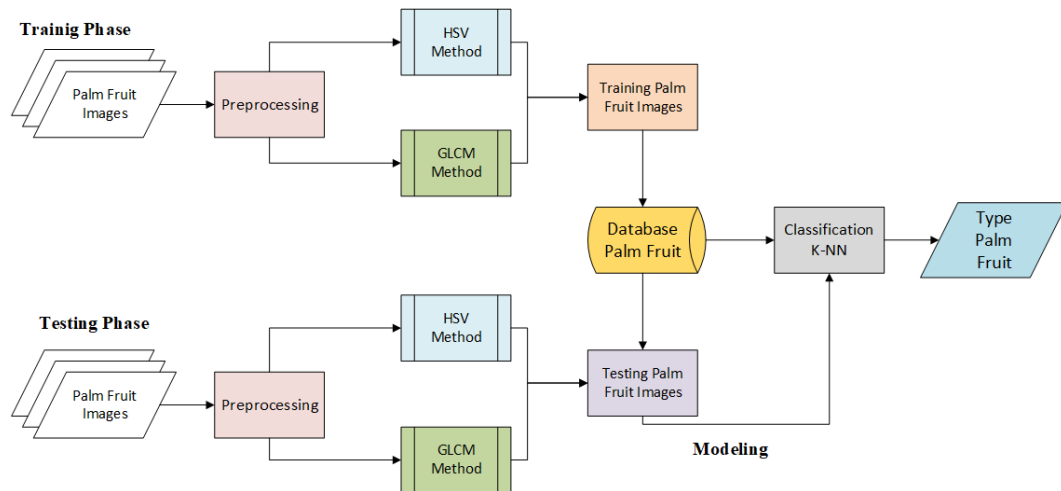


Figure 6 Simple System Workflow

The extracted color and texture features are then fused into a composite feature vector for each image and stored in the training database along with their respective ripeness labels (unripe, half-ripe, ripe, overripe). This phase results in the development of a classification model to be used in the testing phase.

In the testing phase, the system receives new images as test data. Identical preprocessing steps are applied to maintain feature consistency. The HSV and GLCM methods are again employed to extract color and texture features from the test images, forming a feature vector to be input into the classification model. The classification process uses the K-Nearest Neighbors (K-NN) algorithm, which determines the ripeness class of a fruit by identifying the majority class among the k closest training samples in the multidimensional feature space.

The strength of this approach lies in the integration of complementary color and texture features and the use of K-NN, a lightweight, non-parametric classifier that

does not require intensive training. Since K-NN relies solely on distance-based similarity, it is well-suited for numerical data derived from combined feature sets. This system is designed to deliver accurate classification results, with robustness to variations in lighting and background conditions, and practical implementation potential in field applications, particularly for selective harvesting of oil palm fruit.

3. Results and Discussions

3.1 Feature Extraction Using GLCM and HSV

The HSV method is used to extract color features based on the average (mean) values of the Hue (H), Saturation (S), and Value (V) components. The extraction process begins with preprocessing the oil palm fruit images, which includes resizing and cropping the sample area to 128x128 pixels. As shown in Table 1, After that, the color values of the image in RGB format are extracted, and then converted into the HSV color space. This

conversion is carried out to obtain color information that is more representative of human visual perception, which is then used as features in the classification process.

From the extraction of image 4.29, RGB values were obtained with a resolution of 128x128, where each pixel has R/G/B values. Then, the R/G/B values of each pixel are converted to H/S/V. For example, at pixel (0, 1) with RGB values of 244, 101, 38, the HSV values are calculated as Formulas 1 – 6.

$$R' = \frac{244}{255}, G' = \frac{101}{255}, B' = \frac{38}{255} \quad (1)$$

$$R' = 0,9569, G' = 0,3961, B' = 0,149 \quad (2)$$

$$Cmax = \max(0,9569; 0,3961; 0,149) = 0,9569 \quad (3)$$

$$Cmin = \min(0,9569; 0,3961; 0,149) = 0,149 \quad (4)$$

$$\Delta = 0,9569 - 0,149 = 0,8079 \quad (5)$$

$$H = 60x \left(\frac{0,3961 - 0,149}{0,8079} \mod 6 \right) = 1,8354 \quad (6)$$

As shown in Table 2 Result of converting the RGB values 244, 101, 38 to HSV is $H = 18$, $S = 0.8443$, and $V = 0.9569$ As shown in Table 2 with H ranging from 0–360, S from 0–1, and V from 0–1.

The steps in implementing the K-Nearest Neighbor (K-NN) algorithm begin with determining the value of K , which is the number of nearest neighbors to be used as the basis for classification. Next, the distance between the test data and all available training data is calculated. Once these distances are obtained, they are sorted from the smallest to the largest. Then from Table 3, the K exercise data process with the closest distance is selected as the nearest neighbor. From these neighbors, the class of each data point is identified. The class that appears most frequently among the neighbors is assigned as the predicted class for the test data.

Table 1. RGB Values

X, Y	0	1	2	...	127
0	237/94/28	244/101/38	247/106/37	...	196/128/79
1	229/86/20	236/93/27	239/98/29	...	246/177/128
2	225/85/24	229/89/28	232/92/30	...	248/191/138
...	...	...	...	...	...
127	205/84/13	213/92/21	204/87/20	...	187/63/29

Table 2. HSV Values

x, y	0	1	2	...	127
0	19/0,882/0,929	18/0,844/0,957	20/0,85/0,969	...	25/0,597/0,769
1	19/0,913/0,898	19/0,886/0,925	20/0,879/0,937	...	25/0,48/0,965
2	18/0,893/0,882	18/0,878/0,898	20/0,875/0,91	...	29/0,444/0,973
...	...	...	...	...	...
127	22/0,937/0,804	22/0,901/0,835	22/0,902/0,8	...	13/0,845/0,733

Table 3. Values of the training data

No	HSV features			GLCM features				Class
	H	S	V	Contrast	Correlation	Energy	Homogeneity	
1	0,068	0,624	0,872	81,0135	0,9672	0,0004	0,2748	Ripe
2	0,081	0,615	0,701	134,6177	0,9660	0,0002	0,2119	Ripe
3	0,171	0,778	0,342	117,4453	0,9467	0,0008	0,2994	Ripe
4	0,216	0,601	0,556	1270,791	0,7643	0,0002	0,1410	Ripe
5	0,561	0,279	0,316	173,4997	0,9823	0,0011	0,3144	Unripe
6	0,594	0,331	0,307	461,9428	0,9289	0,0004	0,2168	Unripe
7	0,230	0,365	0,255	845,4458	0,8975	0,0005	0,1970	Unripe
8	0,594	0,331	0,307	461,9428	0,9289	0,0004	0,2168	Unripe
9	0,270	0,543	0,588	372,1295	0,9576	0,0002	0,1870	Half Ripe
10	0,397	0,475	0,646	454,6180	0,9442	0,0002	0,2090	Half Ripe
11	0,152	0,457	0,717	670,2418	0,9161	0,0002	0,1715	Half Ripe
12	0,315	0,505	0,314	95,2730	0,9727	0,0004	0,2614	Half Ripe
13	0,253	0,431	0,771	1051,697	0,7493	0,0001	0,1276	Overripe
14	0,449	0,342	0,748	1008,035	0,8222	0,0001	0,1307	Overripe
15	0,272	0,644	0,396	284,4817	0,9129	0,0003	0,2105	Overripe
16	0,519	0,554	0,537	246,9567	0,9456	0,0002	0,2262	Overripe
17	0,139	0,745	0,660	223,0101	0,9600	0,0002	0,1850	?

First, determine the parameter K. For example, we set the number of nearest neighbors to K = 3, 5, and 7, presented in Table 4. Second, calculate the distance between the new data and all training data using the Euclidean Distance, as shown in Table 5. As a calculation example, training data number 1 and the data to be classified are used as Formulas 7 – 9.

$$d_1 = \sqrt{(0,234 - 0,114)^2 + (0,958 - 0,802)^2 + (0,757 - 1)^2 + (0,175 - 0,064)^2 + (0,977 - 0,985)^2 + (0,021 - 0,042)^2 + (0,410 - 0,609)^2} \quad (7)$$

$$d_1 = \sqrt{0,0144 + 0,0243 + 0,059 + 0,0123 + 0,00006 + 0,00004 + 0,0396} \quad (8)$$

$$d_1 = 0,3869 \quad (9)$$

As shown in Table 6 to evaluate the classification performance of the proposed method, a confusion matrix was generated based on a dataset comprising 600 images of oil palm fruit, equally distributed across four ripeness categories: overripe, ripe, half-ripe, and unripe (150 images per class). The classification results demonstrate that the model was able to correctly classify 120 overripe fruits, with 30 instances misclassified as ripe. All 150 ripe fruits were correctly identified, indicating a perfect recall for the ripe class. In the half-ripe category, 120 images were correctly classified, while 30 were misclassified as ripe. Similarly, for the unripe class, 120 images were

correctly predicted, and 30 were erroneously labeled as half-ripe. The overall classification accuracy achieved was 85.00%, reflecting the model's robustness in distinguishing among the four ripeness stages. Precision values for overripe and unripe classes reached 100.00%, indicating that all instances predicted as such were indeed correct, with no false positives. Conversely, the ripe class showed a lower precision of 71.43%, suggesting that a number of samples from other classes (particularly half-ripe and overripe) were mistakenly predicted as ripe. The recall metric was highest for the ripe class (100.00%) and remained consistent at 80.00% for the other three classes.

Table 4. Result of the calculation Euclidean Distance

No	Euclidean Distance	Class
1	2,517	Ripe
2	2,493	Ripe
3	2,512	Ripe
4	2,044	Ripe
5	2,850	Unripe
6	2,589	Unripe
7	2,436	Unripe
8	2,589	Unripe
9	2,485	Hal Ripe
10	2,485	Hal Ripe
11	2,382	Hal Ripe
12	2,581	Hal Ripe
13	1,979	Operripe
14	2,251	Operripe
15	2,352	Operripe
16	2,518	Operripe

Table 5. Classification Result K=3, K=5, dan K=7

No	Euclidean Distance	Class	K=3	K=5	K=7
13	1,979	Overripe	Overripe	Overripe	Overripe
4	2,044	Ripe	Ripe	Ripe	Ripe
14	2,251	Overripe	Overripe	Overripe	Overripe
15	2,352	Overripe	Overripe	Overripe	Overripe
11	2,382	Half-ripe	Half-ripe	Half-ripe	Half-ripe
7	2,436	Unripe	Unripe	Unripe	Unripe
9	2,485	Half-ripe	Half-ripe	Half-ripe	Half-ripe
10	2,485	Half-ripe	Half-ripe	Half-ripe	Half-ripe
2	2,493	Ripe	Ripe	Ripe	Ripe
3	2,512	Ripe	Ripe	Ripe	Ripe
1	2,517	Ripe	Ripe	Ripe	Ripe
16	2,518	Overripe	Overripe	Overripe	Overripe
12	2,581	Half-ripe	Half-ripe	Half-ripe	Half-ripe
6	2,589	Unripe	Unripe	Unripe	Unripe
8	2,589	Unripe	Unripe	Unripe	Unripe
5	2,850	Unripe	Unripe	Unripe	Unripe

Table 6. Confusion Matrix for K=3

Actual Data \ Model Prediction	Overripe	Ripe	Half-ripe	Unripe	Total
Overripe	120	30	0	0	150
Ripe	0	150	0	0	150
Half-ripe	0	30	120	0	150
Unripe	0	0	30	120	150
Total	120	210	150	120	600

Table 7. Values of TP, FP, and FN for K=3

Class	TP (True Positive)	FP (False Positive)	FN (False Negative)
Overripe	120	30	0
Ripe	150	0	60
Half-ripe	120	30	30
Unripe	120	30	0

Table 7 demonstrates that the classification performance varies across the different ripeness categories of oil palm fruit. For the Overripe class, the model achieved a True Positive (TP) of 120, with 30 False Positives (FP) and no False Negatives (FN), indicating that while some other classes were mistakenly predicted as overripe, all actual overripe samples were correctly classified. The Ripe class obtained a perfect TP of 150 with no FP, but had 60 FN, suggesting that a significant number of ripe samples were misclassified as other categories, likely due to visual similarity during ripeness transitions.

In the Half-ripe category, the model achieved 120 TP, accompanied by 30 FP and 30 FN, indicating moderate classification accuracy with both over-prediction and under-prediction. Similarly, the Unripe class showed 120 TP, 30 FP, and 0 FN, meaning that while all actual unripe fruits were correctly identified, some samples from other classes were incorrectly labeled as unripe. These results suggest that the model performs well in detecting both extreme classes (overripe and unripe), while transitional stages such as ripe and half-ripe present higher classification challenges due to overlapping visual characteristics.

Table 8 presents the classification performance of the K-Nearest Neighbors (K-NN) algorithm with $K=3$ across four categories of oil palm fruit ripeness: Overripe, Ripe, Half-ripe, and Unripe based on three

evaluation metrics: accuracy, precision, and recall. The classification accuracy is uniformly 85% for all classes, indicating balanced performance. Precision scores vary, with the Ripe class achieving the highest precision at 100%, suggesting zero false positives, while Overripe, Half-ripe, and Unripe each scored 80%. For recall, Overripe and Unripe achieved perfect recall (100%), indicating that all actual instances were correctly identified. The Half-ripe class followed with a recall of 80%, whereas the Ripe class showed the lowest recall at 71.42%, indicating some difficulty in retrieving all relevant ripe instances due to feature overlaps with neighboring classes. Overall, the model achieved an average precision of 85% and recall of 88%, demonstrating that the integration of HSV and GLCM features with K-NN classification yields reliable and consistent performance in identifying oil palm fruit maturity levels.

Table 8. Accuracy, Precision, and Recall for K=3

Evaluation Metrics	Overripe	Ripe	Half-ripe	Unripe	Average
Accuracy (%)	85	85	85	85	85
Precision (%)	80	100	80	80	85
Recall (%)	100	71.42	80	100	88

Table 9. Confusion Matrix for K=5

Actual Data \ Model Prediction	Overripe	Ripe	Half-ripe	Unripe	Total
Overripe	90	60	0	0	150
Ripe	0	150	0	0	150
Half-ripe	0	60	90	0	150
Unripe	0	0	30	120	150
Total	90	270	120	120	600

Table 9. illustrates the confusion matrix results for the classification of oil palm fruit ripeness using the K-Nearest Neighbors (K-NN) algorithm with $K=5$, based on a dataset of 600 images evenly distributed across four ripeness classes. The model successfully classified all Ripe (150) and Unripe (120) fruits correctly, while misclassifying 30 Unripe samples as Half-ripe. The Overripe class showed significant misclassification, with only 90 out of 150 samples correctly identified and

60 samples misclassified as Ripe. Similarly, the Half-ripe class exhibited 90 correct classifications, with 60 samples misclassified as Ripe. These results suggest that the model performs well in classifying Ripe and Unripe classes with high reliability but faces notable challenges in distinguishing Overripe and Half-ripe fruits from the Ripe class, likely due to overlapping visual features. The concentration of errors between adjacent ripeness stages (e.g., Overripe and Ripe, Half-ripe and Ripe) indicates a need for more discriminative features or an enhanced classification strategy to improve robustness in transitional categories.

Table 10. Values of TP, FP, and FN for K=5

Class	TP (True Positive)	FP (False Positive)	FN (False Negative)
Overripe	90	60	0
Ripe	150	0	120
Half-ripe	90	60	30
Unripe	120	30	0

Table 10 shows the performance metrics of the K-Nearest Neighbors (K-NN) algorithm with K=5 for each ripeness class based on true positive (TP), false positive (FP), and false negative (FN) values. The Overripe class achieved 90 TP, with 60 FP and no FN, indicating that while all actual Overripe fruits were correctly identified, a substantial number of other classes were incorrectly classified as Overripe. The Ripe class reached perfect TP at 150 with zero FP, but suffered from a high FN of 120, revealing that many actual Ripe samples were incorrectly classified as other categories, especially Overripe and Half-ripe. For the Half-ripe class, the model produced 90 TP, 60 FP, and 30 FN, reflecting moderate precision and recall. In contrast, the Unripe class showed strong performance with 120 TP and no FN, although 30 FP occurred due to misclassifications from other classes. These findings suggest that while the classifier is effective in identifying Unripe and Overripe fruits, it struggles to maintain precision and recall in the transitional ripeness stages, particularly for the Ripe and Half-ripe classes, likely due to similarities in visual features such as color and texture.

Table 11. Accuracy, Precision, and Recall for K=5

Evaluation Metrics	Overripe	Ripe	Half-ripe	Unripe	Average
Accuracy (%)	75	75	75	75	75
Precision (%)	60	100	60	80	75
Recall (%)	100	55.56	75	100	83

Table 11 presents the evaluation metrics—accuracy, precision, and recall—for each ripeness category of oil palm fruit using the K-Nearest Neighbors (K-NN) algorithm with K=5. The classification accuracy is consistent across all four classes (Overripe, Ripe, Half-ripe, and Unripe), with each achieving 75%, indicating stable performance in predicting fruit maturity levels. However, the precision values vary, with the Ripe class achieving a perfect 100%, suggesting all instances

predicted as ripe were correct. In contrast, the Overripe and Half-ripe classes attained 60% precision, indicating the presence of false positives from neighboring categories. Regarding recall, both the Overripe and Unripe classes reached 100%, demonstrating the model's ability to correctly identify all actual instances of those classes. Meanwhile, the Ripe class yielded the lowest recall at 55.56%, indicating a considerable number of ripe fruits were misclassified. On average, the model achieved 75% precision and 83% recall, reflecting its reliable performance in detecting extreme ripeness levels (Overripe and Unripe), though classification accuracy for transitional stages such as Ripe remains a challenge due to feature overlap and class ambiguity.

Table 12. Confusion Matrix for K=7

Actual Data \ Model Prediction	Overripe	Ripe	Half-ripe	Unripe	Total
Overripe	60	90	0	0	150
Ripe	0	150	0	0	150
Half-ripe	0	90	60	0	150
Unripe	0	0	30	120	150
Total	60	330	90	120	600

Table 12 presents the confusion matrix for the classification of oil palm fruit ripeness using the K-Nearest Neighbors (K-NN) algorithm with K=7 on a dataset comprising 600 samples evenly distributed across four maturity classes. The classifier correctly identified 60 Overripe, 150 Ripe, 60 Half-ripe, and 120 Unripe samples. However, substantial misclassifications occurred in the transitional categories. Specifically, 90 Overripe samples were misclassified as Ripe, and 90 Half-ripe samples were also incorrectly labeled as Ripe, leading to an over-prediction in the Ripe class with a total of 330 predictions, compared to only 150 actual instances. The Unripe class maintained a relatively strong performance with 120 correct predictions and only 30 misclassified as Half-ripe.

These results indicate that while the model with K=7 effectively classifies the Ripe and Unripe categories, it struggles to distinguish between adjacent ripeness stages, particularly between Overripe and Ripe, as well as Half-ripe and Ripe. The dominance of Ripe predictions across multiple classes suggests that the classifier tends to favor this class, possibly due to feature similarity or class imbalance in the feature space, highlighting the need for better discriminative features or alternative classification strategies for closely related ripeness levels.

Figure 7. is a visualization of the classification distribution at K=3, showing the best predictions across almost all classes.

Figure 8. shows a decrease in classification accuracy for the Overripe and Half-ripe classes compared to K=3.

Figure 9. shows that predictions for the Overripe class become increasingly inaccurate, indicating that a larger

K value negatively impacts classification performance. Based on the evaluation results using the Confusion Matrix, the parameter K=3 showed the best performance with an accuracy of 85%, precision of 85%, and recall of 88%. In second place, K=5 achieved an accuracy of 75%, precision of 75%, and recall of 83%. Meanwhile, K=7 demonstrated the lowest performance. Therefore, it can be concluded that K=3 is the optimal parameter for this classification system, as it provides more accurate results in classifying the ripeness levels of oil palm fruit.

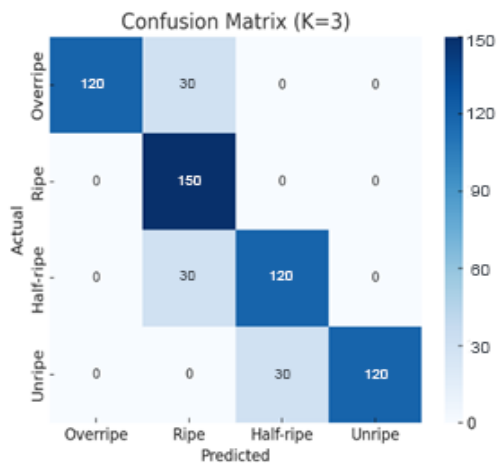


Figure 7. Confusion Matrix K=3

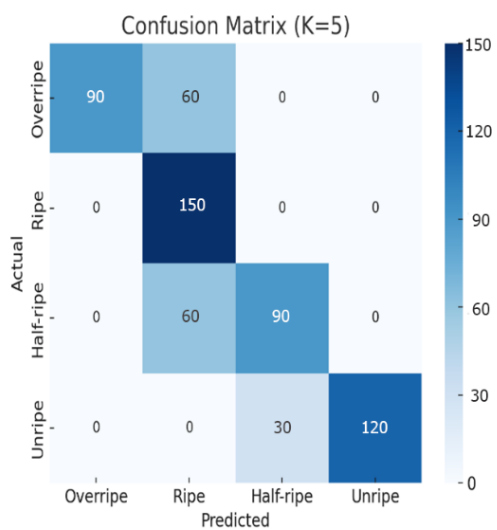


Figure 8. Confusion Matrix K=5

Figure 10 is a bar chart that summarizes the classification performance of the K-NN model with various K values. The red bars represent accuracy, indicating the percentage of correct predictions; the green bars represent precision, showing how many of the predicted labels are actually correct; and the yellow bars represent recall, measuring how well the model identifies actual positive cases. Among the tested K values, K=3 provides the best overall performance across all evaluation metrics.

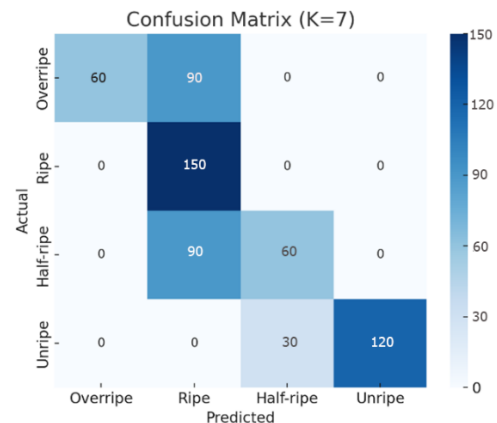


Figure 9. Confusion Matrix K=7

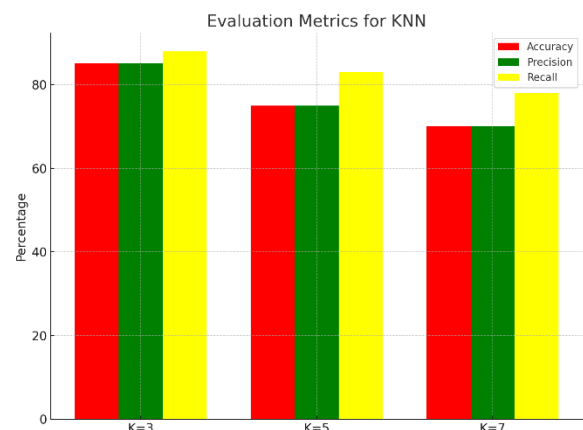


Figure 10. Evaluation Metrics Chart for KNN (K=3, K=5, K=7)

4. Conclusions

Based on the results of the research conducted, it can be concluded that the hybrid model method combining Gray Level Co-occurrence Matrix (GLCM), Hue Saturation Value (HSV), and K-Nearest Neighbor (K-NN) algorithms is effective in classifying the ripeness levels of oil palm fruit. The classification results indicate that the K=3 parameter in the K-NN algorithm yields the best performance, achieving an accuracy of 85%, precision of 85%, and recall of 88%, outperforming other values of K (K=5 and K=7). The HSV method successfully extracts informative color features (Hue, Saturation, and Value), which are critical in determining the fruit's ripeness. In parallel, the GLCM method provides texture features (Contrast, Correlation, Energy, Homogeneity) that effectively differentiate each ripeness category. The use of the K-NN algorithm as the classification method has proven to be reliable in categorizing oil palm fruit into four ripeness levels: Overripe, Ripe, Half-Ripe, and Unripe. The results of the conducted tests demonstrate that this system enhances harvesting precision and supports sustainable palm oil production by automating ripeness detection with high accuracy, thereby reducing dependence on manual methods and improving consistency and productivity in palm oil processing. These findings suggest that the proposed hybrid method

can be feasibly integrated into an automated classification system to support decision-making in oil palm harvesting operations.

References

- [1] J. Segarra et al., "Multiscale assessment of ground, aerial and satellite spectral data for monitoring wheat grain nitrogen content," *Information Processing in Agriculture*, vol. 10, no. 4, pp. 504–522, Dec. 2023, doi: 10.1016/j.inpa.2022.05.004.
- [2] N. P. Sutrisna et al., "Deteksi Tingkat Kematangan Buah Pepaya menggunakan Model Convolutional Neural Network," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 3, pp. 569–578, Jul. 2024, doi: 10.25126/jtiik.938119.
- [3] L. M. Jawad, "A Novel Region of Interest for Selective Color Image Encryption Technique based on New Combination between GLCM Texture Features," in *2021 National Computing Colleges Conference (NCCC)*, IEEE, Mar. 2021, pp. 1–6. doi: 10.1109/NCCC49330.2021.9428802.
- [4] Mohammad Yazdi Pusadan, Indah Safitri, and Wirdayanti, "The Image Extraction Using the HSV Method to Determine the Maturity Level of Palm Oil Fruit with the k-nearest Neighbor Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 6, pp. 1448–1456, Dec. 2023, doi: 10.29207/resti.v7i6.5558.
- [5] A. Putra, V. Sihombing, and M. H. Munandar, "Rancang Bangun Aplikasi Deteksi Tepi Citra Digital Menggunakan Algoritma Prewitt," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 4, no. 1, pp. 83–87, 2021, doi: 10.37600/tekinkom.v4i1.214.
- [6] B. Sugianto, C. A. Sari, and M. Arief Soeleman, "KNN Algorithm Optimization in GLCM-Based Beef and Pork Image Classification," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2023, pp. 141–146. doi: 10.1109/iSemantic59612.2023.10295331.
- [7] A. Pambudi, M. A. Rofiq, M. Anwar Sadat, M. Arief Soeleman, and R. Anggi P, "Knn Algorithm For Foodstuff Classification Using Hsv Color Space And Feature Extraction," in *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2022, pp. 217–225. doi: 10.1109/iSemantic55962.2022.9920421.
- [8] Z. Y. Lamasigi, S. Serwin, and Y. Malago, "Identification of the Freshness Level of Tuna based on Discrete Cosine Transform on Feature Extraction of Gray Level Co-Occurrence Matrix using K-Nearest Neighbor," *ILKOM Jurnal Ilmiah*, vol. 15, no. 1, pp. 153–164, Apr. 2023, doi: 10.33096/ilkom.v15i1.1426.153-164.
- [9] N. H. Jusoh, H. Yazid, S. N. Basah, and S. Abdul Rahim, "Detection of Pulmonary Nodule using Shape-Based Feature Descriptor and Neural Network," *J Phys Conf Ser*, vol. 1372, no. 1, p. 012066, Nov. 2019, doi: 10.1088/1742-6596/1372/1/012066.
- [10] K. Elaiyaraja and M. Senthil Kumar, "Fusion Imaging in Pixel Level Image Processing Technique – A Literature Review," *International Journal of Engineering & Technology*, vol. 7, no. 3.12, p. 175, Jul. 2018, doi: 10.14419/ijet.v7i3.12.15913.
- [11] M. Kaur, C. Thacker, L. Goswami, T. TR, I. S. Abdulrahman, and A. S. Raj, "Alzheimer's Disease Detection using Weighted KNN Classifier in Comparison with Medium KNN Classifier with Improved Accuracy," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, May 2023, pp. 715–718. doi: 10.1109/ICACITE57410.2023.10183208.
- [12] J. Zhang, Y. Li, F. Shen, Y. He, H. Tan, and Y. He, "Hierarchical text classification with multi-label contrastive learning and KNN," *Neurocomputing*, vol. 577, 2024, doi: 10.1016/j.neucom.2024.127323.
- [13] M. Jumarlis et al., "A hybrid hue saturation lightness, gray level co-occurrence matrix, and k-nearest neighbour for palm-sugar classification," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 3, pp. 2934–2945, Sep. 2024, doi: 10.11591/ijai.v13.i3.pp2934-2945.
- [14] M. A. Aziz, P. Pujiono, M. A. Soeleman, and S. Rosidin, "Scabies Skin Classification Using SVM Method With GLCM Feature Extraction," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2023, pp. 221–226. doi: 10.1109/iSemantic59612.2023.10295283.
- [15] S. S. Sumit, J. Watada, A. Roy, and D. Rambli, "In object detection deep learning methods, YOLO shows supremum to Mask R-CNN," *J Phys Conf Ser*, vol. 1529, no. 4, p. 042086, Apr. 2020, doi: 10.1088/1742-6596/1529/4/042086.
- [16] A. G. Karegowda, V. D., and Y. Harshalatha, "Coconut Disease Detection using Traditional ML Classifiers with GLCM and SIFT Features," in *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/CSITSS64042.2024.10816784.
- [17] Z. Chen, Z. Song, T. Zhang, and Y. Wei, "IoT devices and data availability optimization by ANN and KNN," 2024. doi: 10.1186/s13635-023-00145-0.
- [18] Mardison and Yuhandri, "Detection of Kidney Cysts of Kidney Ultrasound Image using Hybrid Method: KNN, GLCM, and ANN Backpropagation," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, IEEE, Feb. 2023, pp. 222–227. doi: 10.1109/ICCoSITE57641.2023.10127703.
- [19] N. Anwar et al., "Prediction system for quality-of-service level of network infrastructure using naive bayes classifier," 2024, p. 030001. doi: 10.1063/5.0225195.
- [20] C. Y. Hu et al., "Alcohol patch test with hue-saturation-value model analysis predicts ALDH2 genetic polymorphism," *Comput Biol Med*, vol. 147, 2022, doi: 10.1016/j.combiomed.2022.105783.
- [21] F. M. Sarimole and A. Syaeful, "Classification of Durian Types Using Features Extraction Gray Level Co-Occurrence Matrix (GLCM) AND K-Nearest Neighbors (KNN)," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 4, no. 1, pp. 111–121, Sep. 2022, doi: 10.37385/jaets.v4i1.959.
- [22] G. Prasad, V. S. Gaddale, R. C. Kamath, V. J. Shekaranai, and S. P. Pai, "A Study of Dimensionality Reduction in GLCM Feature-Based Classification of Machined Surface Images," *Arab J Sci Eng*, vol. 49, no. 2, 2024, doi: 10.1007/s13369-023-07854-1.
- [23] C. Liu et al., "Improvements to a GLCM-based machine-learning approach for quantifying posterior capsule opacification," *J Appl Clin Med Phys*, vol. 25, no. 2, 2024, doi: 10.1002/acm2.14268.
- [24] F. E. Batool et al., "Offline signature verification system: a novel technique of fusion of GLCM and geometric features using SVM," *Multimed Tools Appl*, vol. 83, no. 5, 2024, doi: 10.1007/s11042-020-08851-4.
- [25] L. He, X. Cheng, A. Jiwa, D. Li, J. Fang, and Z. Du, "Zanthoxylum bungeanum Fruit Detection by Adaptive Thresholds in HSV Space for an Automatic Picking System," *IEEE Sens J*, vol. 23, no. 13, 2023, doi: 10.1109/JSEN.2023.3277042.
- [26] T. Putri Ananda, S. Viola Widayarsi, M. Ihsan Muttaqin, and A. Stefanie, "Identifikasi Tingkat Kematangan Buah Pepaya Menggunakan Metode Convolutional Neural Network (Cnn)," *Jati (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 2094–2097, Dec. 2023, doi: 10.36040/jati.v7i3.7137.
- [27] Q. Hu, M. Zhang, A. S. Mujumdar, G. Xiao, and S. Jin-cai, "Drying of edamames by hot air and vacuum microwave combination," *J Food Eng*, vol. 77, no. 4, pp. 977–982, Dec. 2006, doi: 10.1016/j.jfoodeng.2005.08.025.
- [28] M. Faisal, M. Hasan, and K. C. Pelangi, "The Implementation of GLCM and ANN Methods to Identify Dragon Fruit Maturity Level," *ILKOM Jurnal Ilmiah*, vol. 15, no. 1, pp. 64–71, Apr. 2023, doi: 10.33096/ilkom.v15i1.1504.64-71.
- [29] G. Ashour, A. Al-Dubai, and I. Romdhani, "Ontology-based Course Teacher Assignment within Universities," *International Journal of Advanced Computer Science and*

- Applications*, vol. 11, no. 7, 2020, doi: 10.14569/IJACSA.2020.0110787.
- [30] M. Jumarlis and Mirfan, "Detecting Diseases on Clove Leaves Using GLCM and Clustering K-Means," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, 2022, doi: 10.29207/resti.v6i4.4033.