
An Intelligent Multi Feature Learning Approach for Palm Fruit Ripeness Classification: A Comparative Study of K-nearest Neighbor and Naïve bayes

Basiroh ^{a,1*}, Widya Novita Al Afifah Irwanto ^{a,2}, Siti Nurlaili Karim ^{b,3}

^a Informatic, Engineering, Universitas Islam Batik, Surakarta, Central Java, Indonesia

^b Departemen of Computer Science, Universiti Teknologi Petronas, Bandar Seri Iskandar, Perak, Malaysia

e-mail: ¹shira@uibs.ac.id, ²widyairwanto31@sma.belajar.id, ³nurlaili.karim812@gmail.com

*Corresponding author

Abstract

Digital technology has become increasingly integrated into the agricultural sector, with Computer Vision emerging as one of the most prominent artificial intelligence technologies that utilizes visual data to learn, analyze, and recognize objects automatically. The identification of oil palm fruit ripeness represents a form of collaboration between technology and agriculture. Accurate classification of oil palm fruit ripeness plays a crucial role in improving harvesting efficiency, maximizing oil yield, and maintaining the quality of crude palm oil. Conventional visual inspection performed by human operators is often subjective and inconsistent, leading to classification errors that may reduce production quality and economic value. Although previous studies have applied various machine learning algorithms for oil palm fruit classification, most have focused on a single feature type or individual classification algorithm, making it difficult to determine the most effective combination of features and classifier. To address this gap, this study presents a comparative analysis of two widely used classification algorithms, K-Nearest Neighbor (KNN) and Naïve Bayes, using color, texture, shape, and combined feature sets to identify the optimal approach for oil palm fruit ripeness classification. The dataset consists of images of unripe, ripe, and rotten oil palm fruits, comprising 600 training images and 150 testing images. Color features include mean RGB, RGB standard deviation, RGB skewness, and RGB entropy. Texture features consist of grayscale mean, grayscale standard deviation, contrast, correlation, energy, and homogeneity, while shape features include area, perimeter, metric, major axis, minor axis, and eccentricity. The novelty of this research lies in the comprehensive evaluation of multiple feature groups across different classification algorithms to determine the most discriminative feature–algorithm combination for oil palm ripeness assessment. Experimental results demonstrate that the KNN algorithm with $k = 3$ consistently outperforms Naïve Bayes, achieving accuracies of 91.3% using color features, 89.3% using texture features, 83.3% using shape features, and 98.0% when all features are combined. In comparison, Naïve Bayes achieves accuracies of 65.0%, 68.0%, 42.0%, and 77.0% for the respective feature sets. These findings indicate that combining color, texture, and shape features with the KNN classifier provides the most effective solution for automated oil palm fruit ripeness classification and can support the development of intelligent harvesting and quality control systems.

Keywords: Feature extraction, Intelligent Multi-feature learning, KNN, Naïve Bayes, Oil palm fruit

1. Introduction

Recent technological advancements in agriculture have accelerated the adoption of computer vision techniques[1], enabling the development of intelligent systems for automated classification and decision-making[2]. Although numerous studies have investigated oil palm fruit ripeness classification using machine learning and computer vision techniques, several research gaps remain[3]. First, most previous studies have focused on evaluating a single classification algorithm or a single type of image feature, such as color or texture, making it difficult to determine which algorithm–feature combination provides the most reliable performance[4]. Second, comparative studies between K-Nearest Neighbor (KNN) and Naïve Bayes under identical experimental conditions are still limited, particularly when multiple feature groups, including color, texture, and shape, are evaluated simultaneously[5]. Consequently, the relative effectiveness of these classifiers for multiclass oil palm fruit ripeness classification has not been comprehensively established[6]. Furthermore, many existing studies primarily report overall classification

accuracy while providing limited analysis of the contribution of individual and combined feature sets to classification performance[7]. This limitation leaves an unanswered question regarding whether integrating multiple visual features can consistently improve classification accuracy compared with using a single feature type[8]. Therefore, a comprehensive comparative evaluation is necessary to identify the most effective combination of image features and classification algorithms for oil palm fruit ripeness assessment.

The application of algorithms in the classification process also requires support from relevant features for recognition[9]. These features include color, texture, and shape. Images are processed to extract feature characteristics, and the resulting feature values are then used as elements to support the implementation of oil palm fruit ripeness recognition using the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms[10]. Each algorithm has different characteristics in the classification process. The KNN algorithm performs classification by calculating the distance to the nearest neighbors of each data point to be classified [11]. On the other hand, the Naïve Bayes algorithm performs classification based on probability theory, assuming that the features and attributes used are independent of one another. Oil palm has two species within its genus, namely *Elaeis guineensis* and *Elaeis oleifera* [12]. The most dominant type of oil palm in Indonesia is *Elaeis oleifera*. The process of harvesting oil palm fruit is carried out directly by harvesters. Oil palm fruit that yields high-quality Crude Palm Oil (CPO) is fruit that has reached an optimal level of ripeness. Crude Palm Oil is one of the products derived from oil palm fruit with high economic value [13]. Harvesters often make mistakes due to various factors, resulting in the harvesting of unripe or immature fruit. Some of the factors influencing these errors include weather conditions, lighting levels, fatigue, visibility, and other related factors [14]. Errors in harvesting occur frequently, leading to the collection of under-ripe or nearly ripe fruit, and in some cases, overripe fruit is missed and eventually left to rot.

The Recent advances in computer vision and machine learning have significantly improved the automation of agricultural product quality assessment, including oil palm fruit ripeness classification. Numerous studies have investigated image-based classification methods using color, texture, and shape features combined with conventional machine learning algorithms or deep learning models [15]. Several previous studies have demonstrated that color features are effective for distinguishing oil palm fruit ripeness because the fruit surface undergoes noticeable color changes during maturation. However, methods relying solely on color information are highly sensitive to illumination variations, shadows, and camera settings, which may reduce classification accuracy under real-world conditions. Other researchers have incorporated texture features, such as Gray-Level Co-occurrence Matrix (GLCM) descriptors, to capture differences in fruit surface characteristics. Texture-based approaches generally improve robustness against color variations but may not adequately distinguish fruits with similar surface patterns at different ripeness stages.

The objective of this study is to compare the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms in classifying the ripeness level of oil palm fruit [16]. This research also evaluates the performance of each algorithm using color, shape, and texture features, as well as a combination of all features. To address this research gap, this study performs a systematic comparative evaluation of KNN and Naïve Bayes using four feature configurations: color, texture, shape, and a combination of all three feature groups [17]. The contribution of this research is the identification of the most effective feature–algorithm combination for multiclass oil palm fruit ripeness classification. The accuracy of the test results is used to determine the best-performing algorithm [18] for classifying the ripeness level of oil palm fruit.

2. Research Methodology

a. Data Collection

The dataset used in this study was collected specifically for evaluating the performance of the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms in classifying the ripeness level of oil palm fruit. Images were acquired directly from oil palm plantations using a smartphone camera equipped with a sensor resolution (4:3)8.0 Megapixel camera. Each image was captured at a resolution of 1920×1080 pixels, ensuring sufficient visual detail for feature extraction. The dataset consists of images representing three ripeness categories: unripe, ripe, and rotten oil palm fruits. A total of 750 images were collected, including 600 images for model training and 150 images for testing. The distribution of samples across the three classes was balanced, with 250 images per class, comprising 200 training images and 50 testing images for each category [19]. The color features consist of RGB mean, RGB standard deviation, RGB skewness, and RGB entropy. The texture features include grayscale mean, grayscale standard deviation, contrast, correlation, energy, and homogeneity[20]. Following data acquisition, image feature extraction was

performed to obtain the characteristics used for classification. Three groups of features were extracted: color, texture, and shape. Color features included RGB mean, RGB standard deviation, RGB skewness, and RGB entropy. Texture features consisted of grayscale mean, grayscale standard deviation, contrast, correlation, energy, and homogeneity, while shape features included area, perimeter, metric, major axis, minor axis, and eccentricity. These extracted features were subsequently used as inputs for the KNN and Naïve Bayes classifiers to evaluate the influence of different feature combinations on classification performance.

Table 1. Distribution of samples across classes.

Class	Training Image	Testing Image	Total image
Unripe	200	50	250
Ripe	200	50	250
Rotten	200	50	250
Total	600	150	750

The KNN algorithm performs classification by considering the similarity or proximity between data points [21]. In KNN, the closest data points are referred to as neighbors [22]. The following are the steps of the K-Nearest Neighbor (KNN) algorithm in performing classification:

1. Determine the number of neighbors (k) to be used as a reference for class determination.
 2. Calculate the distance between the new data point and each data point in the dataset.
 4. Select the K data points with the smallest distances, then determine the class of the new data.
 5. Among these K neighbors, count the number of data points in each category.
 6. Assign the new data point to the category with the highest number of neighbors.
- The model is ready for use.

The K-Nearest Neighbor (KNN) algorithm is a simple Machine Learning algorithm [23] that can be used to solve both classification and regression problems.

$$d(x-x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2} \quad (2.1)$$

Description:

x_1 : sample data

x_2 : test data

d : distance

p : data dimension

i : data variable index

In general, the Euclidean distance formula applies to one-dimensional data, and the following formula is used when there is more than one variable.

$$d_i = \sqrt{\sum_{i=0}^n i(a^2 + b^2) + (y_{2i} - y_{1i})^2} \quad (2.2)$$

To complete this study, several stages were carried out, including data collection, data preprocessing, data processing implementation, and evaluation (Figure 1).

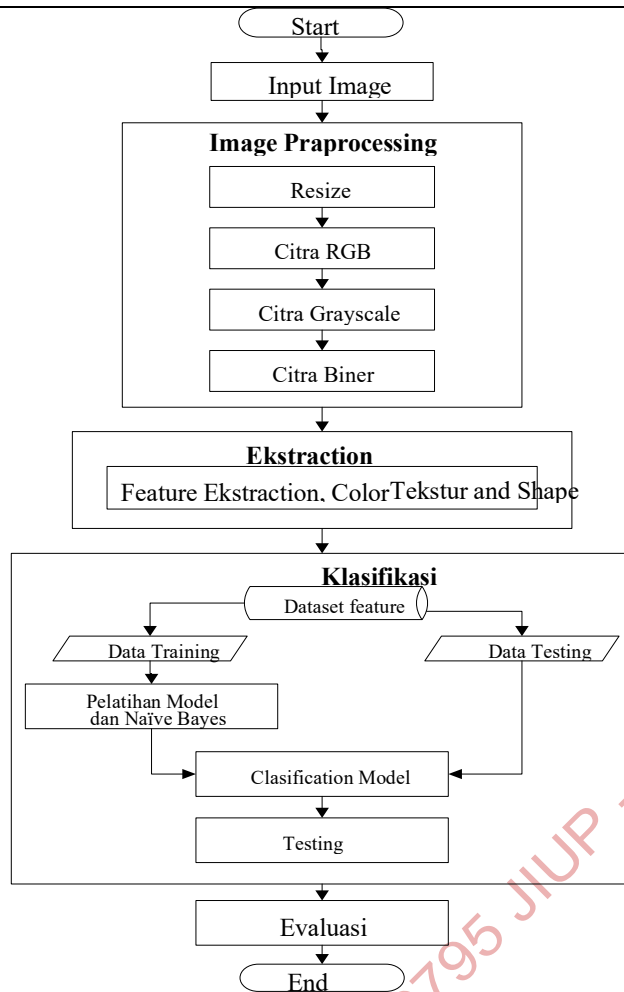


Figure 1. Research Workflow

b. Research Stages

The preprocessing stage is carried out to prepare the dataset for use. In this stage, the original images undergo background removal and are resized to a uniform dimension. The original RGB images are then converted into grayscale images, which are subsequently transformed into binary images [24], [25]. The feature comparison in this study is implemented using Matlab tools. The next step in this method is image feature extraction, which aims to obtain the relevant features. At this stage, color, texture, and shape features are collected. The color features include RGB mean, RGB standard deviation, RGB skewness, and RGB entropy. The texture features consist of grayscale mean, grayscale standard deviation, contrast, correlation, energy, and homogeneity. The shape features include area, perimeter, metric, major axis, minor axis, and eccentricity.

After conducting testing and obtaining the accuracy results for each algorithm and feature implementation, an evaluation is carried out to determine the outcomes of the comparison between the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms. The models built from each algorithm produce accuracy values based on the testing stages.

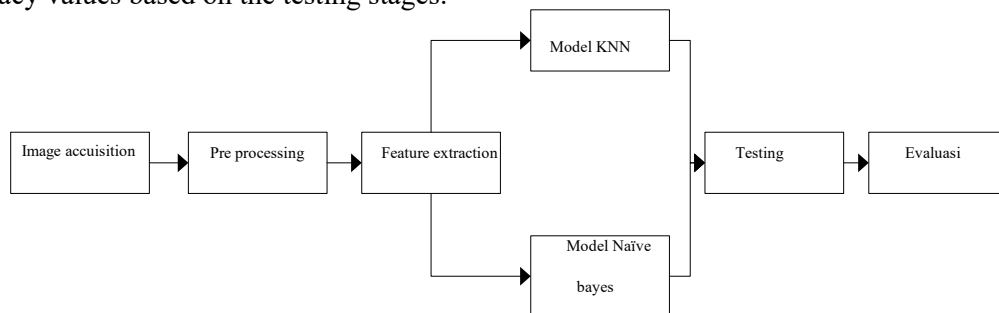


Figure 2. Image Classification Flowchart

c. Data Preprocessing

The preprocessing stage is carried out to prepare the dataset for use. The steps include problem identification, research, literature review, result analysis, and drawing conclusions [26]. The original images then undergo background removal and are resized to a uniform size. The RGB images are converted into grayscale images, which are subsequently transformed into binary images [27].

d. Feature Extraction

Image feature extraction[28] is performed to obtain the features used in this study. This stage involves collecting color, texture, and shape features [29]. The color features include RGB mean, RGB standard deviation, RGB skewness, and RGB entropy. The texture features consist of grayscale mean, grayscale standard deviation, contrast, correlation, energy, and homogeneity [20][30]. The shape features include area, perimeter, metric, major axis, minor axis, and eccentricity.

First-order feature extraction is based on the histogram characteristics of the image [31]. First-order features are generally used to distinguish macrostructure textures, which refer to locally repeated patterns with a periodic nature[32]. These features include mean, variance, skewness, kurtosis, and entropy. Meanwhile, second-order features are based on the probability of the relationship between pairs of pixels at a specific distance and angle orientation [33]. Second-order features are commonly used to distinguish microstructure textures, which refer to local patterns with less obvious repetition[34]. These features include second-order moments, contrast, correlation, variance, inverse difference moment, and entropy[35].

e. Evaluation

The experiment concludes with an evaluation stage. After implementing each of the proposed algorithms, testing results are obtained. These results show the classification accuracy achieved by each algorithm. The accuracy is calculated using a confusion matrix. The performance of the algorithms is then evaluated to determine which one achieves the highest accuracy. The performance of the K-Nearest Neighbor (KNN) and Naïve Bayes classifiers was evaluated using accuracy, precision, recall, and F1-score for each ripeness class, as well as a confusion matrix. These metrics provide a comprehensive assessment of classifier performance in the multiclass classification task. The value obtained is :

1. Precision 0,98 %

$$\text{Precision} = \frac{\text{True Positive}}{\text{True positive} + \text{false Positive}} = \frac{48}{48+1} = 0,980$$

2. Recall 0,96 %

$$\text{Recall} = \frac{\text{True Positive}}{\text{True positive} + \text{false Negative}} = \frac{48}{48+2} = 0,960$$

3. F1-Score 0,97 %

$$F1 = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}} = 2 \times \frac{0,980 \times 0,960}{0,980 + 0,960} = 0,970$$

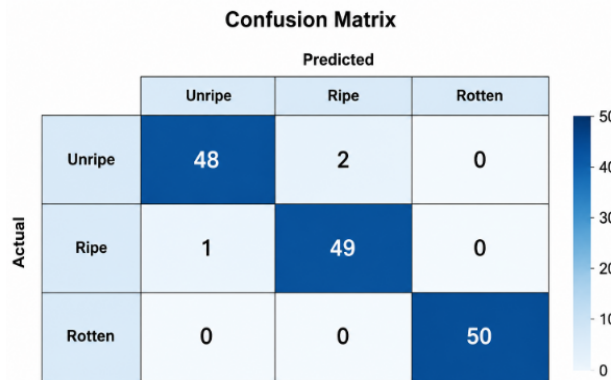


Figure 4. Confusion Matiks KNN

The figure presents a confusion matrix illustrating the performance of the K-Nearest Neighbor (KNN) classifier in classifying the ripeness levels of oil palm fruits into three classes: Unripe, Ripe, and Rotten. The results demonstrate excellent classification performance. Of the 50 actual unripe fruit samples, 48 were

correctly classified as Unripe, while 2 were incorrectly classified as Ripe, and none were misclassified as Rotten. Similarly, among the 50 ripe fruit samples, 49 were correctly classified as Ripe, whereas only 1 sample was incorrectly classified as Unripe. The classifier achieved perfect classification for the Rotten class, correctly identifying all 50 rotten fruit samples without any misclassification. Overall, the confusion matrix shows that 147 out of 150 testing samples were correctly classified, corresponding to an overall classification accuracy of 98.0%. These results confirm that the KNN classifier ($k = 3$)

Confusion Matrix			
Actual Class	Unripe	Ripe	Rotten
Unripe	36 (72.00%)	10 (20.00%)	4 (8.00%)
Ripe	8 (16.00%)	34 (68.00%)	8 (16.00%)
Rotten	6 (12.00%)	8 (16.00%)	36 (72.00%)
Total	50	50	50
Predicted Class			

Figure 4. Confusion Matrix Naïve Bayes Model Performance oil palm fruit classification

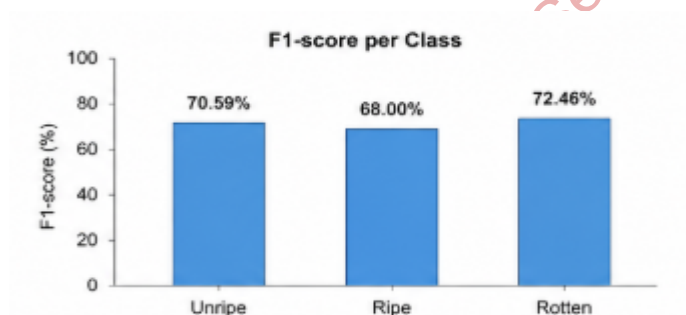


Figure 5. F1-score per Class Diagram

3. Result and Discussion

The classification of oil palm fruit ripeness is performed by selecting relevant features and comparing the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms through a series of stages for image classification of oil palm fruit. This process begins with image preprocessing, followed by feature extraction from the training data and the development of a model based on the extracted patterns. The images used are captured directly using a camera. The image objects consist of oil palm fruit seeds collected directly from a palm fruit storage area.

a. Image Preprocessing

All images used in this study consist of unripe oil palm fruit (a), ripe oil palm fruit (b), and rotten oil palm fruit (c).

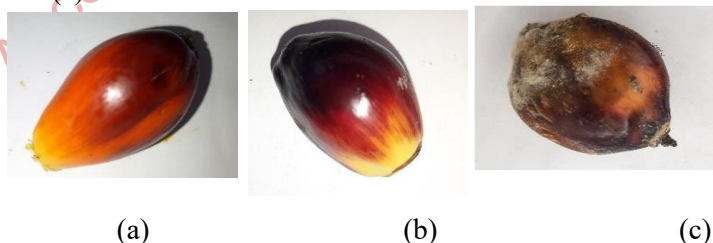


Figure 6. Original Palm Oil Image

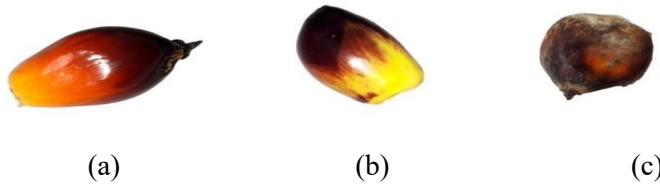


Figure 7. Original Palm Oil Image

Before performing feature extraction for classification, the images are first preprocessed. Color feature extraction is carried out using the original images in the Red, Green, and Blue (RGB) color space. The images are then converted into grayscale images for texture feature extraction. After obtaining the texture features, the images are further converted into binary (black-and-white) images to extract shape features.

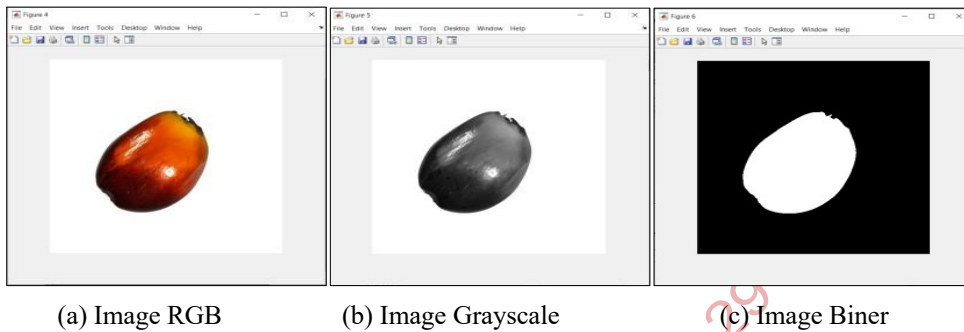


Figure 8. Images After Background Removal

The features used in this research include color, texture, and shape features. The color features consist of RGB mean, RGB standard deviation, RGB skewness, and RGB entropy. The images used for color feature extraction, as shown in Figure (a), are RGB images.

b. Feature Extraction

At this stage of the research, feature extraction is carried out to classify the ripeness level of oil palm fruit by selecting relevant features and comparing the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Naïve Bayes algorithms. The steps involved focus on implementing the algorithms for feature and algorithm comparison. The following table presents the images and color feature values used in this study.

Table 3. Image and Color Featur Values

Teks	Praproses citra		
	Citra	Color feature	Value
1		Average R	244.4234
		Average G	240.3429
		Average B	229.1418
		Standar deviasi R	27.62461
		Standar deviasi G	32.427
		Standar deviasi B	49.69119
		Skewness R	-3.98791
		Skewness G	-5.75407
		Skewness B	-2.7119
		Entropy R	0
		Entropy G	0
		Entropy B	0.00651

2		Average R	230.6969
(ripe)			
		Average G	226.1345
		Average B	224.385
		Standar deviasi R	43.92652
		Standar deviasi G	50.11445
		Standar deviasi B	52.30312
		Skewness R	-4.94015
		Skewness G	-5.61407
		Skewness B	-5.40754
		Entropy R	0.0000775
		Entropy G	0.005763
		Entropy B	0.006814
		Average R	230.6969
		Average G	226.1345
		Average B	224.385
3		Average R	239.9649
(rotten)			
		Average G	235.8278
		Average B	233.1569
		Standar deviasi R	31.52151
		Standar deviasi G	39.41311
		Standar deviasi B	44.21225
		Skewness R	-2.79784
		Skewness G	-4.40243
		Skewness B	-2.45422
		Entropy R	0
		Entropy G	0
		Entropy B	0.001366
		Average R	239.9649
		Average G	235.8278
		Average B	233.1569

During the image preprocessing stage, color feature extraction was performed to characterize the visual properties of the oil palm fruit image. The extracted color descriptors included the mean RGB values, RGB standard deviation, RGB skewness, and RGB entropy. The extracted color features indicate that the image is dominated by relatively high intensity values, with average RGB values of 244.4234, 240.3429, and 229.1418 for the red, green, and blue channels, respectively. The standard deviation values were 27.6246 (Red), 32.4270 (Green), and 49.6912 (Blue), indicating that the blue channel exhibits the greatest variation in pixel intensity. The skewness values of -3.9879, -5.7541, and -2.7119 for the red, green, and blue channels, respectively, indicate negatively skewed intensity distributions. Entropy analysis revealed values of 0.0000 for the red channel, 0.0000 for the green channel, and 0.0065 for the blue channel. These very low entropy values indicate minimal randomness and a relatively uniform color distribution within the image. Overall, the extracted color features provide quantitative information describing the color distribution of the oil palm fruit and serve as essential inputs for the K-Nearest Neighbor (KNN) and Naïve Bayes classification models

Table 4. Results of Color Feature Data Preprocessing

Average R	Average G	Average B	Stdv R	Stdv G	Stdv B	Skew R	Skew G	Skew B	Entr R	Entr G	Entr B	Target
228.7	223.1	220.8	9.1	6.4	9.3	4.5	-4.2	-4.134	0	0.0	0	0
228.7	225.1	222.5	4.9	9.7	2.2	3.1	3.0	2.576	0	0.0008	0.0005	0
231.3	227.0	224.1	2.7	8.2	1.3	3.0	-2.8	2.922	0	0	0.0001	0

229.4	225.4	222.2	3.8	8.4	1.9	3.4	-3.5	3.364	0	0	0.0001	0
238.9	236.4	233.7	9.6	3.1	6.7	3.3	-3.2	3.570	0	0	0	0
226.7	216.1	214.0	0.9	53.0	5.7	2.1	-3.4	3.020	0	0	0.0935	1
244.1	229.9	224.7	0.3	2.9	1.7	3.6	-3.0	2.474	0	0	0.1229	1
237.7	224.6	220.7	0.2	7.1	3.0	4.1	-4.2	2.635	0	0	0.1399	1
242.0	230.1	226.2	5.2	3.1	9.8	3.9	-3.8	3.894	0	0	0.1201	1
242.6	228.1	223.0	2.3	5.6	4.1	3.8	-3.4	3.338	0	0	0.1849	1
232.3	225.2	214.6	0.5	6.9	3.8	3.0	-3.3	5.210	0	0.0070	0.0279	2
226.1	217.7	203.3	7.3	4.6	7.0	4.1	-5.8	4.178	0	0.0040	0.0673	2
209.9	203.3	201.6	1.2	6.3	7.1	3.9	-3.7	4.100	0.00021	0.0108	0.030	2
230.4	219.9	207.8	1.2	9.3	8.3	3.9	-4.1	3.470	0	0.0052	0.0079	2
231.8	228.4	221.0	3.8	7.4	1.8	2.7	-4.1	2.579	0	0.0455	0.0266	2

The color feature preprocessing stage produced a set of numerical descriptors representing the color characteristics of each oil palm fruit image. The extracted features consisted of the average values of the red (Average R), green (Average G), and blue (Average B) channels, as well as the entropy values of each channel (Entropy R, Entropy G, and Entropy B). These features were subsequently combined with the corresponding target label representing the fruit ripeness category (Unripe, Ripe, or Rotten).

Table 5. Results of Feature Tekstur Preprocessing

Average Grayscale	Std Grayscale	Contrast	Corre lation	Energy	Homoginety	Target
224.51	44.42	0.08	0.99	0.69	0.98	0
225.87	38.46	0.17	0.98	0.69	0.96	0
227.93	36.82	0.15	0.98	0.69	0.96	0
226.21	37.41	0.16	0.97	0.65	0.95	0
236.82	32.41	0.11	0.97	0.74	0.97	0
219.01	49.13	0.11	0.99	0.68	0.98	1
233.56	36.57	0.10	0.98	0.76	0.99	1
228.08	42.04	0.08	0.99	0.73	0.99	1
233.21	37.86	0.10	0.98	0.77	0.99	1
231.83	39.19	0.09	0.99	0.76	0.99	1
226.12	44.67	0.05	0.99	0.67	0.99	2
218.58	43.25	0.09	0.99	0.61	0.99	2
205.08	54.61	0.10	0.99	0.61	0.99	2
221.66	36.99	0.13	0.98	0.62	0.98	2
228.59	46.04	0.06	0.99	0.71	0.99	2

In the next stage, the data classification process is carried out using an algorithm, where testing is performed on a training dataset consisting of 600 images. After extracting the training features, a model is built based on the KNN algorithm pattern to conduct testing using the test data. The evaluation is carried out using the model developed during the training phase. There are three classification categories: unripe, ripe, and rotten. Each feature is used to compare and determine the best-performing feature for classification using the KNN algorithm.

Table 6. Accuracy Results of KNN and Color Features

Feature	K-1	K-3	K-5	K-7	K-9	K-10
---------	-----	-----	-----	-----	-----	------

color	70%	91.3%	90%	90.6%	88.6%	89%
--------------	-----	--------------	-----	-------	-------	-----

K=1 achieved an accuracy of 70%. The accuracy for K=3 was 91.3%, K=5 was 90%, K=7 was 90.6%, K=9 was 88.6%, and K=10 was 89%. As shown in Table 4.5, K=3 achieved the highest accuracy of 91.3%, while the lowest accuracy was 70%.

Next, the KNN algorithm was tested using texture features to determine the best accuracy when performing classification based on texture features.

Table 7. Accuracy Results of KNN and Tekstur Features

Feature	K-1	K-3	K-5	K-7	K-9	K-10
Tekstur	72%	89.3%	84%	81.3%	80%	82%

In the texture feature testing, K=1 achieved an accuracy of 72%. The accuracy for K=3 was 89.3%, K=5 was 84%, K=7 was 81.3%, K=9 was 80%, and K=10 was 82%. As shown in Table 4.5, K=3 achieved the highest accuracy of 89.3%, while the lowest accuracy was 72%.

Next, the KNN algorithm was tested using shape features to determine the best accuracy when performing classification using all features.

Table 8. Overall KNN Results

Feature	K-1	K-3	K-5	K-7	K-9	K-10
Color	70%	91.3%	90%	90.6%	88.6%	89%
Tekstur	72%	89.3%	84%	81.3%	80%	82%
Shape	61.3%	83.3%	74.6%	69.3%	74.6%	71.3%
All	73%	98%	91.3%	93.3%	93.3%	94%

It can be seen from the table above that, at K=3, using all features achieves the highest accuracy of 98%. Other features also show the best accuracy compared to other K values. For the color features, K=3 achieves an accuracy of 91.3%, for texture features K=3 reaches 89.3%, for shape features K=3 reaches 83.3%, and using all features results in an accuracy of 98%.

In the final stage, the classification of oil palm fruit ripeness is carried out using algorithms, namely K-Nearest Neighbor (KNN) and Naïve Bayes. The classification results from each algorithm are compared to obtain the best performance. Each feature is tested in the oil palm fruit ripeness classification process. The implementation of the Naïve Bayes algorithm is carried out using color, shape, and texture features to evaluate which feature performs best in achieving the highest classification accuracy.

Tabel 9. Naive Bayes Results : Color Features, Textures, Shape Features, and All Features

Feature	Accuracy
color	65.3%
Feature	Accuracy
Tekstur	68.6%
Feature	Accuracy
Shape	42%
Feature	Accuracy
All feature	77%

Based on the test results, it can be seen that the implementation of the Naïve Bayes algorithm achieved its highest accuracy when using all features, reaching 77%. The accuracy for color features was 65.3%, texture features 68.6%, and shape features 42%. After conducting feature-based testing, the comparison between the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms applied to oil palm fruit ripeness classification shows clear results. Based on the comparison of both algorithms using color, texture, and shape features, it was found that the best-performing algorithm in this study is the K-Nearest Neighbor

algorithm with $k = 3$. The highest accuracy achieved using $K=3$ was 91.3% for color features, 89.3% for texture features, 83.3% for shape features, and 98% when using all features. Meanwhile, the Naïve Bayes algorithm achieved its highest accuracy of 77% when using all features.

The accuracy for color features was 65.3%, texture features 68.6%, and shape features 42%. After implementing and testing all algorithms, the feature comparison results between the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms in this study show that K-Nearest Neighbor consistently achieves the best accuracy across all tests. Based on this research, the K-Nearest Neighbor (KNN) algorithm using all features produces the highest accuracy of 98%. In contrast, the lowest accuracy is obtained by the Naïve Bayes algorithm using shape features, which reaches 42%. It can be observed that, using the same dataset, the K-Nearest Neighbor (KNN) algorithm with $k = 3$ is able to optimize performance and achieve the best accuracy compared to the other algorithms being evaluated. Based on previous studies on oil palm fruit ripeness classification, accuracy results vary. However, it can be seen that color features tend to dominate most studies. In this research, it is also shown that other features, such as texture features, can achieve strong performance, reaching an accuracy of 83.9% using the K-Nearest Neighbor (KNN) algorithm with $k = 3$. From the results presented, it can be concluded that the best accuracy is achieved by the K-Nearest Neighbor (KNN) algorithm at $k = 3$. In general, each algorithm has its own strengths and weaknesses depending on its working principle. The K-Nearest Neighbor (KNN) algorithm, which works by finding the nearest data points to determine the class label of a sample, has a different model compared to Naïve Bayes. The Naïve Bayes algorithm builds its model by assuming that each feature independently contributes to the class probability. As a result, its accuracy may decrease when the features used are not strongly correlated with one another.

Table 10. Mean Accuracy Value

Model	Color	Texture	shape	Combined Features	Mean Accuracy K-10
KNN (K=3)	91.3%	89.3%	83.3%	98.0%	95.5%
Naïve Bayes	65.0%	68.0%	42.0%	77.0%	63.0%

In this research, the K-Nearest Neighbor (KNN) algorithm is able to perform classification very well, achieving an accuracy of 98%. The experimental results obtained from the 10-fold cross-validation indicate that the KNN classifier consistently outperformed the Naïve Bayes classifier across all feature groups. Using the combined color, texture, and shape features, KNN achieved an average classification accuracy of 98.0%, whereas Naïve Bayes achieved 77.0%. The relatively small variation in performance across the ten folds demonstrates that the proposed classification model is stable and exhibits good generalization capability. It can be concluded that the features used show clear distinctions in their values. As a result, the K-Nearest Neighbor (KNN) algorithm is able to achieve the best accuracy in the classification process. All features (color, texture, and shape) exhibit clear differences between unripe, ripe, and rotten oil palm fruit. Therefore, the K-Nearest Neighbor (KNN) algorithm is able to perform classification by effectively utilizing the values of each feature provided during the classification process.

Comparing algorithms and features using the same dataset is essential to identify the best-performing algorithm that can maximize accuracy, as well as ensure relevance, effectiveness, and reliability in classifying oil palm fruit ripeness levels.

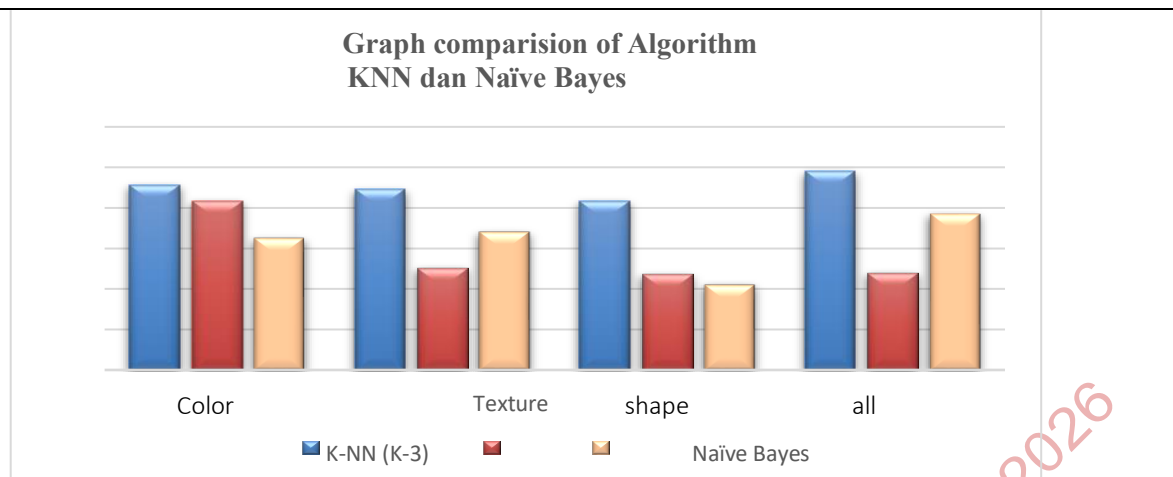


Figure 9. Graph comparison of Algorithm

The graph describe comparison of algorithm, it can be seen that the implementation using color features with the K-Nearest Neighbor (KNN) algorithm (K=3) achieved the highest accuracy of 91.3%, while Naïve Bayes achieved 65%.

The implementation using texture features with the K-Nearest Neighbor (KNN) algorithm (K=3) achieved the highest accuracy of 89.3%, while Naïve Bayes reached 68%. The implementation using shape features with the K-Nearest Neighbor (KNN) algorithm (K=3) obtained the highest accuracy of 83.3%, whereas Naïve Bayes achieved 42%. In this research, testing was also conducted using all features across both algorithms, and the highest accuracy was achieved by the K-Nearest Neighbor (KNN) algorithm (K=3) at 98%, while the Naïve Bayes algorithm achieved 77%.

4. Conclusion

Based on result of implementation and testing, it can be concluded that the comparison between the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms for oil palm fruit ripeness classification yields The findings provide empirical evidence that combining multiple visual features with a distance-based classifier such as KNN yields superior performance for multiclass oil palm fruit ripeness classification. the following findings: the best-performing algorithm is K-Nearest Neighbor (KNN) with $k = 3$, achieving an accuracy of 91.3% for color features, 89.3% for texture features, 83.3% for shape features, and 98% when using all features. This indicates that each algorithm has different performance capabilities in classification, even when using the same image data and features. From a practical perspective, the proposed approach can support automated oil palm harvesting by providing an objective and consistent method for determining fruit ripeness. Meanwhile, the best accuracy achieved by the Naïve Bayes algorithm using all features is 77%, with 65.3% accuracy for color features, 68.6% for texture features, and 42% for shape features. Future research should investigate the use of larger and more diverse datasets collected under varying environmental conditions to improve model robustness and generalization. In addition, advanced deep learning approaches, such as Convolutional Neural Networks (CNNs) and vision transformer-based models, may be explored and compared with conventional machine learning algorithms. Future studies may also integrate the proposed classification model into real-time mobile or embedded systems to facilitate practical deployment in smart agriculture applications.

References

- [1] S. Puttinaovarat and S. Chai-arayalert, "Oil Palm Bunch Ripeness Classification and Plantation Verification Platform : Leveraging Deep Learning and Geospatial Analysis and Visualization," 2024.
- [2] F. Nadhirah, M. Nain, N. Hashimah, A. Hassain, and R. Abdullah, "A Review of an Artificial Intelligence Framework for Identifying the Most Effective Palm Oil Prediction," 2022.
- [3] S. Nasyira and L. Rosnita, "Comparison of K-Nearest Neighbors Method and Naïve Bayes Method in Classifying the Quality of Oil Palm Seed Varieties," vol. 10, no. 2, pp. 413–425, 2025, doi: 10.31572/inotera.Vol10.Iss2.2025.ID539.
- [4] G. Tan, H. Tzuan, F. H. Hashim, T. Raj, A. B. Huddin, and M. S. Sajab, "Oil Palm Fruits Ripeness

-
- Classification Based on the Characteristics of Protein , Lipid , Carotene , and Guanine / Cytosine from the Raman Spectra,” 2022.
- [5] Z. Hong, W. Lei, J. Ruiming, and L. Junwei, “A distance measuring method using visual image processing,” *Proc. 2009 2nd Int. Congr. Image Signal Process. CISP’09*, vol. 1, no. 2, pp. 1–5, 2009, doi: 10.1109/CISP.2009.5303258.
 - [6] J. Q. Goh, A. Rashid, and M. Shariff, “Application of Optical Spectrometer to Determine Maturity Level of Oil Palm Fresh Fruit Bunches Based on Analysis of the Front Equatorial , Front Basil , Back Equatorial , Back Basil and Apical Parts of the Oil Palm Bunches,” 2025.
 - [7] A. Noviyanto, F. Ramadhani, V. Kautsar, Y. Avianto, and S. Gunawan, “Advancing Concession-Scale Carbon Stock Prediction in Oil Palm Using Machine Learning and Multi-Sensor Satellite Indices,” pp. 1–27, 2026.
 - [8] M. S. M. Alfatni *et al.*, “Towards a Real-Time Oil Palm Fruit Maturity System Using Supervised Classifiers Based on Feature Analysis,” 2022.
 - [9] A. M. Syaifullah, “Proses Pengolahan Kelapa Sawit PT Perkebunan Nusantara XIV Unit PKS Luwu,” 2021.
 - [10] R. Hidayat, M. Fikry, Y. Yusra, F. Yanto, and E. P. Cynthia, “Penerapan Naïve Bayes Classifier dalam Klasifikasi Sentimen Publik di Twitter terhadap Puan Maharani,” *JUKI J. Komput. dan Inform.*, vol. 6, no. 1, pp. 100–108, 2024, doi: 10.53842/juki.v6i1.479.
 - [11] M. A. Zulkhoiri *et al.*, “Investigation of oil palm fruit bunch ripeness classification using machine learning classifiers,” vol. 02010, pp. 1–12, 2024.
 - [12] Y. Jusman, A. Maulana, and J. H. Lubis, “Classification of Leaf Diseases in Oil Palm Plants with Haar Wavelet Transform Features Based on Machine Learning,” vol. 01002, 2024.
 - [13] X. Jian, W. Loon, K. Shien, and W. Zhe, “Expert systems in oil palm precision agriculture : A decade systematic review,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1569–1594, 2022, doi: 10.1016/j.jksuci.2022.02.006.
 - [14] Y. Luo, A. P. P. A. Majeed, Z. Omar, S. Jagtap, G. Garcia-garcia, and Y. Chen, “Synergizing Residual and Dense Architectures for Fine-Grained Oil Palm Grading: A Deep Feature Concatenation Approach,” pp. 1–23, 2026.
 - [15] T. Dzulkarnain, D. E. Ratnawati, and B. Rahayudi, “Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Penilaian Masyarakat Terhadap Pelayanan Rumah Sakit di Malang,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 993–1000, 2024, doi: 10.25126/jtiik.2024117979.
 - [16] N. S. Suriani, N. H. Mohd, S. M. Shah, S. Z. Muji, F. Atyka, and N. Rashid, “A Feasibility Study on Synthetic RGB-NIR Image Generation for Oil Palm Fresh Fruit Bunch Grading,” vol. 17, no. 2, pp. 729–739, 2026.
 - [17] C. Li *et al.*, “Integrating Multi-Temporal Landsat and Sentinel Data for Enhanced Oil Palm Plantation Mapping and Age Estimation in Malaysia,” pp. 1–27, 2025.
 - [18] D. S. Arief, M. Shiddiq, J. Jahrizal, and M. Dalil, “Machine vision for detection of defective oil palm fresh fruit bunches using YOLO Algorithm,” vol. 3186, pp. 1–16, 2026, doi: 10.1088/1742-6596/3186/1/012026.
 - [19] S. Singh, H. Dureja, and R. Singh, “A multilingual smart farming system for real time agricultural decisions using machine learning,” 2026.
 - [20] T. Li *et al.*, “Advancements in Intelligent Monitoring Technologies for Behavioral , Physiological , and Biomarker Analysis in Cattle Health : A Review,” pp. 1–40, 2026.
 - [21] P. Chaudhary, P. Gulia, N. S. Gill, N. Alduaiji, and P. K. Shukla, “An evaluation of machine learning for soil analysis in internet of things-enabled smart farming,” pp. 1–24, 2026.
 - [22] S. Lingamgunta, J. Mudidana, D. R. Vincent, and S. A. M. Felicita, “Distilled vision transformers with CNN fusion for robust cashew apple maturity prediction,” no. April, pp. 1–20, 2026, doi: 10.3389/fpls.2026.1787609.
 - [23] L. Colaco and P. Kamat, *Artificial intelligence advances for cashew fruit maturity and quality detection : a systematic review on models , sensors , and farming applications*. 2025.
 - [24] W. Novita and A. Afifah, “ANALISIS SENTIMEN NEGATIVE PADA APLIKASI JOBSTREET MENGGUNAKAN HADOOP DISTRIBUTED FILE SYSTEM (HDFS),” vol. 4, no. 4, pp. 287–293, 2025.
 - [25] R. Wulandari and A. Abas, “Understanding the impact of climate change on oil palm plantation : a systematic literature review,” no. August, pp. 1–17, 2025, doi: 10.3389/fsufs.2025.1621217.

- [26] M. P. J. Tabe-ojong, A. G. Geffersa, and K. T. Sibhatu, "Producer organizations , productivity and sustainable intensification practices in oil palm production," pp. 1–18, 2026.
- [27] V. Lai, N. Yusma, M. Yusoff, A. Najah, and Y. F. Huang, "The benefits and perspectives of the palm oil industry," *Environ. Dev. Sustain.*, no. 0123456789, 2024, doi: 10.1007/s10668-024-04593-7.
- [28] J. Ma, M. Li, W. Fan, and J. Liu, "State-of-the-Art Techniques for Fruit Maturity Detection," 2024.
- [29] Y. Xu, H. Li, Y. Zhou, Y. Zhai, Y. Yang, and D. Fu, "GLL-YOLO : A Lightweight Network for Detecting the Maturity of Blueberry Fruits," pp. 1–22, 2025.
- [30] C. J. Huang, M. C. Liu, S. S. Chu, and C. L. Cheng, "Application of machine learning techniques to Web-based intelligent learning diagnosis system," *Proc. - HIS'04 4th Int. Conf. Hybrid Intell. Syst.*, pp. 242–247, 2005, doi: 10.1109/ichis.2004.25.
- [31] Y. Liu, Q. Yu, S. Geng, S. Guo, and L. Liu, "SSViT-YOLOv11 : fusing lightweight YOLO & ViT for coffee fruit maturity detection," no. December, pp. 1–20, 2025, doi: 10.3389/fpls.2025.1691643.
- [32] N. Wu *et al.*, "Maturity detection and counting of blueberries in real orchards using a novel STF-YOLO model integrated with ByteTrack algorithm," no. November, pp. 1–26, 2025, doi: 10.3389/fpls.2025.1682024.
- [33] S. Jiang *et al.*, "A Real-Time Detection and Maturity Classification Method for Loofah," 2023.
- [34] T. Daware, P. Ramteke, U. Shaikh, and S. Bharne, "Crop Guidance and Farmer ' s Friend – Smart Farming using Machine Learning," vol. 03021, pp. 1–8, 2022.
- [35] J. P. Mehare, "Classification and Regression Supervised Machine Learning Approach in Smart Farming for Hydroponics System with Intelligent and Precise Management based on IoT," vol. 20, no. August, pp. 9334–9347, 2022, doi: 10.14704/nq.2022.20.10.NQ55912.