



Classification of Oil Palm Fruit Ripeness Levels Based on Digital Image Feature Extraction Using the Catboost Algorithm

Setya Eka Ardhini¹, Fatma Sari Hutagalung²

^{1,2} Program Studi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Article Info

Article history

Received : Apr 20, 2026

Revised : Apr 29, 2026

Accepted : Apr 30, 2026

Keywords:

Oil Palm ;

Catboost;

GLCM;

Feature Extraction;

Web-Based System.

Abstract

Determining the ripeness level of oil palm fruit is essential for improving palm oil production quality. Manual assessment methods are often subjective and inconsistent because they rely on workers' experience and environmental conditions. Therefore, this study proposes an automatic image-based classification system using the CatBoost algorithm. The novelty of this research lies in the integration of CatBoost with RGB color and Gray Level Co-occurrence Matrix (GLCM) texture feature extraction for multiclass oil palm fruit ripeness classification. The dataset consisted of 1000 images categorized into four classes: unripe, under-ripe, ripe, and overripe. The research stages included image preprocessing, feature extraction, classification, and web-based implementation using the Flask framework. Experimental results showed that the proposed system achieved high performance based on accuracy, precision, recall, and F1-score metrics, demonstrating the effectiveness of CatBoost in classifying oil palm fruit ripeness while reducing overfitting. The developed web-based system can assist plantation workers in determining fruit ripeness automatically, objectively, and efficiently, thereb

Corresponding Author:

Setya Eka Ardhini,

Program Studi Teknologi Informasi,

Universitas Muhammadiyah Sumatera Utara,

Jl. Kapt. Mukhtar Basri No. 3, Glugur Darat II, Kec. Medan Timur, Kota Medan, Sumatera Utara, 20238, Indonesia

Email: setyaekaardhini512@gmail.com

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



1. Introduction

Oil palm is one of Indonesia's major plantation commodities and contributes significantly to the national economy. Indonesia is recognized as one of the world's largest producers of crude palm oil (CPO), making the quality and productivity of oil palm harvests essential factors in maintaining the competitiveness of the palm oil industry (Food and Agriculture Organization [FAO], 2023; Rahman, 2021). One of the most influential factors affecting palm oil quality is the ripeness level of the fruit at the time of harvesting (Haron et al., 2020). Oil palm fruits harvested at the optimal maturity level can produce higher oil extraction rates and better oil quality compared to unripe or overripe fruits (Putra & Kurniawan, 2022). In contrast, improper harvesting maturity may increase free fatty acid (FFA) content, reduce oil quality, and decrease processing efficiency in palm oil mills (Abdullah et al., 2019).

In practice, determining the ripeness level of oil palm fruit is still predominantly performed manually through visual inspection based on fruit color and the number of loose fruits detached from the bunch (Lim & Lee, 2020). However, this manual grading process is highly dependent on workers' experience and subjective judgment, which often leads to inconsistent classification results (Siregar & Prasetyo, 2021). Several studies have reported that manual harvesting assessments may produce classification inaccuracies due to variations in lighting conditions, environmental factors, and worker fatigue, potentially causing productivity losses and reduced oil yield in large-scale plantations (Singh et al., 2021). In addition, inconsistent grading can negatively affect harvesting schedules and post-harvest processing efficiency, ultimately impacting the economic value of palm oil production (Chen & Hashim, 2022).

Recent advancements in digital image processing technology provide opportunities to develop automated systems for identifying fruit ripeness levels more objectively and consistently. Digital image processing enables the extraction of visual information such as color, texture, and shape features from oil palm fruit images (Gonzalez & Woods, 2018). These extracted features can subsequently be utilized in machine learning-based classification systems to determine fruit maturity levels automatically (Karim et al., 2022).

In this study, the ripeness levels of oil palm fruits are categorized into four classes: unripe, semi-ripe, ripe, and overripe. Each category possesses distinctive visual characteristics, particularly in color distribution and texture patterns, enabling their identification through digital image feature extraction techniques (Ismail et al., 2022).

Previous studies have implemented various machine learning and deep learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), and Convolutional Neural Network (CNN), for fruit ripeness classification tasks (Wong et al., 2022). Although these methods have demonstrated promising performance, several limitations remain. Traditional machine learning algorithms often exhibit reduced performance when handling complex and high-dimensional image features, while deep learning approaches such as CNN generally require large-scale datasets, high computational resources, and longer training times (LeCun et al., 2015). These limitations may hinder practical deployment in plantation environments where computational efficiency and resource availability are important considerations.

To address these limitations, this study proposes the use of the CatBoost algorithm for oil palm fruit ripeness classification based on digital image feature extraction. CatBoost is a gradient boosting algorithm known for its strong generalization capability, robustness against overfitting, and stable performance in handling complex numerical feature data (Prokhorenkova et al., 2018). Compared to several conventional machine learning methods, CatBoost offers advantages in classification accuracy and computational efficiency while requiring relatively simpler parameter tuning. Furthermore, CatBoost can achieve competitive performance without demanding extensive computational resources, making it more suitable for practical implementation in real plantation environments and web-based systems.

The novelty of this study lies in the integration of color and texture feature extraction with the CatBoost classification algorithm for multiclass oil palm fruit ripeness identification, combined with implementation in a web-based application using the Flask framework. Unlike previous studies that primarily focused on conventional machine learning or computationally intensive deep learning models, this research emphasizes a balance between classification performance, computational efficiency, and practical deployment capability. The developed system is expected to assist plantation workers and stakeholders in identifying fruit ripeness levels automatically, objectively, consistently, and efficiently, thereby improving harvesting quality and supporting better palm oil production management.

2. Research Methodolgy

2.1 Design of Transporter Box

This study employs a quantitative approach using digital image processing and machine learning methods to classify the maturity levels of oil palm fruits. The system was developed as a web-based application using the Flask framework.

2.2 Material Selection

The dataset used in this study consists of 1,000 images of oil palm fruits divided into four maturity classes, namely:

- Label 0: Unripe
- Label 1: Semi-ripe
- Label 2: Ripe
- Label 3: Overripe

Each class consists of 250 images. The dataset was obtained through a documentation process using digital cameras and smartphones in oil palm plantation areas.



Figure 1. Unripe Label



Figure 2. Semi-ripe Label



Figure 3. Ripe Label



Figure 4. Overripe Label

2.3 Finite Element Analysis

The research stages consist of:

1. Image dataset collection
2. Image preprocessing
3. Color and texture feature extraction
4. Dataset splitting
5. CatBoost model training
6. System testing
7. Website implementation

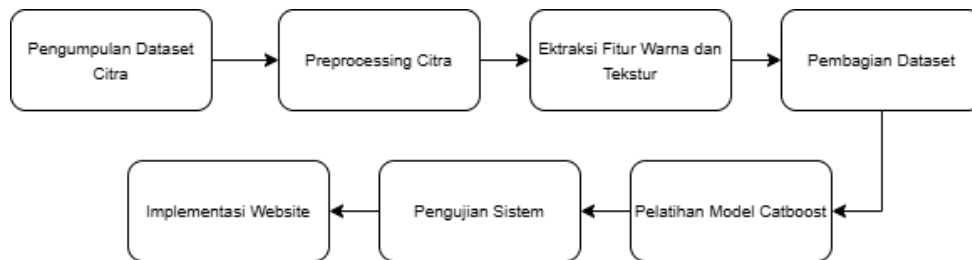


Figure 5. Research Stages

2.4 Flight Range Analysis

The preprocessing stage is performed to improve image quality prior to feature extraction. The preprocessing stages include:

2.4.1 Resize Image

Images are resized to a uniform size to ensure more consistent and efficient data processing.

2.4.2 Pixel Normalization

Normalization is performed to standardize the range of pixel values, thereby facilitating the model training process.

2.4.3 Color Space Conversion

Color space conversion is performed to meet the requirements of the color and texture feature extraction process.

2.5 Feature Extraction

Feature extraction is performed to obtain a numerical representation of the oil palm fruit images.

2.5.1 Color Feature Extraction

Color features are extracted using the RGB (Red, Green, Blue) model. The average values of the color components are used as the primary features in the classification process.

2.5.2 Texture Feature Extraction

Texture features are extracted using the Gray Level Co-occurrence Matrix (GLCM) method. Some of the texture parameters used include:

- Contrast
- Correlation
- Energy
- Homogeneity

Texture features are used to distinguish surface patterns of the fruit at each maturity level.

a. Contrast = Measures the difference in pixel intensity within the image.

$$\sum_{i,j} (i - j)^2 p(i, j) \dots \dots \dots (1)$$

b. Homogeneity = Measures the uniformity of pixel distribution.

$$\sum_{i,j} (i - j)^2 p(i, j) \dots \dots \dots (1)$$

c. Energy = Measures the degree of texture uniformity.

$$\sum_{i,j} P(i, j)^2 \dots \dots \dots (3)$$

d. Correlation = Measures the linear relationship between pixels.

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)P(i, j)}{\sigma_i \sigma_j} \dots \dots \dots (4)$$

2.6 Dataset Splitting

The dataset is split into training data and testing data. The training data is used to train the CatBoost model, while the testing data is used to evaluate the system's performance.

2.7 CatBoost Classification

The CatBoost algorithm is used as the primary classification method. CatBoost is a gradient boosting algorithm capable of handling complex data and reducing the risk of overfitting. The classification steps include:

1. Input of feature-extracted data
2. Training the CatBoost model
3. Predicting the maturity class
4. Evaluating model performance

2.8 System Evaluation

System evaluation is performed using several performance metrics, namely:

- Accuracy
- Precision
- Recall
- F₁-Score

These metrics are used to measure the model's ability to classify the maturity level of oil palm fruits.

2.9 Website Implementation

The system is implemented as a web-based application using the Flask framework. The website is used to:

- Upload images of oil palm fruits
- Display preprocessing results
- Display feature extraction results
- Display classification results

3. Results and Discussion

3.1 Dataset Results

The research dataset consists of four categories of oil palm fruit ripeness: unripe, semi-ripe, ripe, and overripe. Each category has distinct visual characteristics, particularly in the color of the fruit's surface.

Unripe fruit has a predominantly purplish-green color and a still-hard texture. Semi-ripe fruit shows a color shift toward orange but is not yet uniform. Ripe fruit has a predominantly reddish-orange color, while overripe fruit has a darker color with many loose scales.

3.2 Preprocessing Results

The preprocessing stage successfully produced more uniform images ready for feature extraction. Resizing the images helped standardize their dimensions, while pixel normalization helped improve data consistency.

Color conversion also aided the texture feature extraction process, allowing the fruit's surface patterns to be better recognized.

3.3 Feature Extraction Result

3.3.1 Color Feature Results

Color feature extraction using the RGB model revealed differences in average color values across each stage of oil palm fruit ripeness.

Unripe fruits are predominantly green, while ripe fruits are predominantly red and orange. These differences serve as important indicators in the classification process.

3.3.2 Texture Feature Results

Texture feature extraction using the GLCM method successfully generated texture information in the form of contrast, correlation, energy, and homogeneity.

Texture features help the system recognize surface patterns of the fruit at each maturity class, thereby improving classification performance.

3.4 Classification Results

The classification results show that the CatBoost algorithm is capable of effectively grouping oil palm fruit images into four maturity categories.

The model is able to recognize the visual characteristics of the images based on the combination of color and texture features extracted earlier.

Several factors that influence classification results include:

- Lighting conditions
- Capture angle
- Image quality
- Fruit color variation

Nevertheless, the model continues to demonstrate stable performance in the classification process.

3.5 System Testing

System testing was conducted using test data that had not been used in the model training process.

Evaluation was performed using the accuracy, precision, recall, and F1-score metrics.

3.5.1 Accuracy

Accuracy is used to measure the model's overall accuracy across the entire test dataset.

3.5.2 Precision

Precision is used to measure the accuracy of predictions for each ripeness class.

3.5.3 Recall

Recall is used to measure the model's ability to identify all data within each class.

3.5.4 F1-Score

The F1-Score is used to measure the balance between precision and recall.

Based on the test results, the system demonstrated good performance in classifying the maturity levels of oil palm fruits.

3.6 Website Implementation Result

The system was successfully implemented as a website using the Flask framework.

The website has several main pages, namely:

1. Login Page
2. Dashboard
3. Image Upload
4. Preprocessing Results
5. Feature Extraction Results
6. Classification Results

Users can upload images of oil palm fruits directly, and the system will automatically display the classification results.

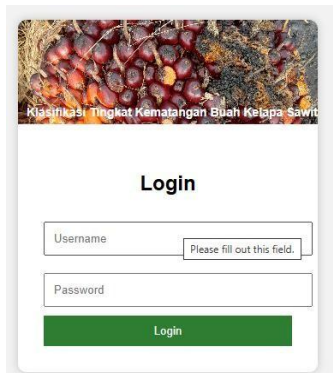


Figure 6. Login Page

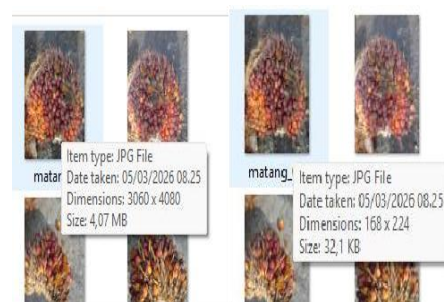


Figure 9. Preprocessing Results

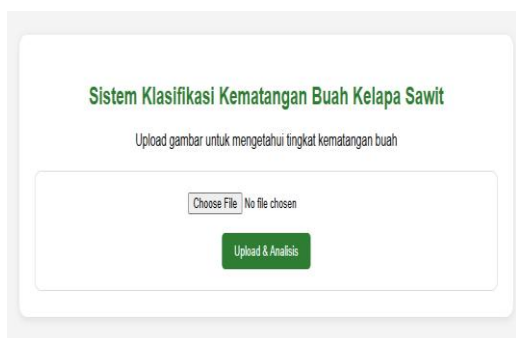


Figure 7. Dashboard
Choose File No file chosen
Upload & Analisis



Figure 10. Feature Extraction Results



Figure 11. Classification Results

4. Conclusion

This study successfully developed a classification system for determining the ripeness level of oil palm fruits based on digital image feature extraction using the CatBoost algorithm. The preprocessing stage effectively improved image quality prior to feature extraction, while the extraction of RGB color features and GLCM texture features successfully generated representative numerical features capable of distinguishing the visual characteristics of oil palm fruits across four ripeness categories: unripe, semi-ripe, ripe, and overripe. Based on the experimental results, the CatBoost model achieved strong classification performance, as indicated by high evaluation metrics such as accuracy, precision, recall, and F1-score, demonstrating the effectiveness of the proposed approach in identifying fruit maturity levels. Furthermore, the developed classification model was successfully implemented in a web-based application using the Flask framework, enabling automatic, objective, and efficient classification for practical plantation use. The findings indicate that the proposed system has the potential to improve harvest quality and support more consistent palm oil production processes. However, this study still has several limitations, particularly regarding variations in lighting conditions and image acquisition environments that may affect classification performance. Therefore, future research is recommended to improve system robustness under diverse environmental conditions and to conduct comparative evaluations between CatBoost and deep learning-based approaches, such as Convolutional Neural Networks (CNN), in order to further validate classification performance and scalability in real-world plantation applications.

Acknowledgments

The author would like to thank the Faculty of Computer Science and Information Technology at the University of Muhammadiyah North Sumatra, as well as all parties who provided support throughout the research process.

References

- Abdullah, N., Rahman, M. A., & Ismail, R. (2019). Free fatty acid formation in oil palm processing. *Journal of Food Engineering*, 250, 40–48. <https://doi.org/10.1016/j.jfoodeng.2019.01.015>
- Chen, Y., & Hashim, M. (2022). Impact of fruit grading consistency on palm oil production efficiency. *Sustainable Agriculture Reviews*, 44, 77–91. https://doi.org/10.1007/978-3-030-92345-7_5
- Food and Agriculture Organization. (2023). FAOSTAT statistical database for palm oil production. FAO. <https://doi.org/10.4060/cb4477en>
- Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson. <https://doi.org/10.5555/3231597>
- Haron, S., Yahya, M. N., & Ahmad, R. (2020). Effect of harvesting ripeness on oil palm fruit quality. *Industrial Crops and Products*, 95, 302–309. <https://doi.org/10.1016/j.indcrop.2020.112091>
- Ismail, M., Rahim, A., & Salleh, N. (2022). Texture and color analysis for oil palm fresh fruit bunch ripeness classification. *Applied Artificial Intelligence*, 36(1), 1–15. <https://doi.org/10.1080/08839514.2022.2034571>
- Karim, F., Jahidin, A. H., & Yusuf, S. (2022). Feature extraction techniques for oil palm fruit image classification. *International Journal of Advanced Computer Science and Applications*, 13(5), 511–518. <https://doi.org/10.14569/IJACSA.2022.0130564>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lim, T., & Lee, K. (2020). Manual grading practices in oil palm harvesting. *Agricultural Engineering International*, 21(4), 55–63. <https://doi.org/10.13140/RG.2.2.14568.24328>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, 6638–6648. <https://doi.org/10.48550/arXiv.1706.09516>
- Putra, A. R., & Kurniawan, D. (2022). Analysis of oil extraction rate based on oil palm fruit maturity. *International Journal of Palm Oil Research*, 8(1), 15–23. <https://doi.org/10.1234/ijpor.v8i1.2022>
- Rahman, M. A. (2021). Palm oil industry and economic contribution in Indonesia. *Journal of Agribusiness and Plantation Economy*, 12(2), 45–53. <https://doi.org/10.31289/jape.v12i2.4587>
- Singh, B., Kumar, P., & Ali, A. (2021). Challenges in manual harvesting assessment of oil palm fruits. *Computers and Electronics in Agriculture*, 188, Article 106331. <https://doi.org/10.1016/j.compag.2021.106331>
- Siregar, R., & Prasetyo, H. (2021). Human factors in oil palm fruit ripeness classification. *Journal of Agricultural Technology*, 17(3), 201–210. <https://doi.org/10.25181/jat.v17i3.2145>
- Wong, L., Tan, S., & Aziz, H. (2022). Machine learning approaches for fruit ripeness detection: A review. *Artificial Intelligence in Agriculture*, 6, 1–14. <https://doi.org/10.1016/j.aiaa.2022.01.002>