BIO Web of Conferences **199**, 01003 (2025) https://doi.org/10.1051/bioconf/202519901003
SAGE-Grace 2025

Mango Ripeness Classification Based on Skin Image Shape Features Using Decision Tree

Anna Nur Nazilah Chamim[1*], *Muhammad Luthfi* Padantya[1], *Erika* Loniza[2], *Itmi Hidayat* Kurniawan[3], *Norazian* Subari[4], and *Siti Nurul Aqmariah Mohd* Kanafiah[5]

[1]Department of Electrical Engineering, Universitas Muhammadiyah Yogyakarta, Bantul, Yogyakarta 55183, Indonesia
[2]Department of Medical Electrical Engineering, Universitas Muhammadiyah Yogyakarta, Bantul, Yogyakarta 55183, Indonesia
[3]Department of Electrical Engineering, Universitas Muhammadiyah Purwokerto, Banyumas, Jawa Tengah 53182, Indonesia
[4]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, Pekan Pahang 26600, Malaysia
[5]Department of Mechatronic Engineering, Universiti Malaysia Perlis, Padang Besar, 02600 Arau, Perlis, Malaysia

Abstract. Mango ripeness classification plays a vital role in the post-harvest process to ensure optimal quality during storage and distribution. This study proposes a lightweight and interpretable classification method based on Hu Moment shape descriptors extracted from mango skin images. These features were evaluated using three variants of the Decision Tree (DT) algorithm: Fine, Medium, and Coarse. The dataset was split into 90% training and 10% testing, and performance was assessed using standard metrics including accuracy, precision, recall, specificity, F-score, and training time. The results show that the Fine DT model achieved the best performance with a testing accuracy of 86.941% and the shortest training time of 4.0016 seconds, outperforming the Medium and Coarse DT variants in terms of efficiency. The findings demonstrate that Hu Moment features, despite their simplicity, are effective in representing mango shape for ripeness classification. This approach offers a promising solution for non-destructive fruit evaluation, particularly in embedded or real-time applications.

1 Introduction

Mango (Mangifera indica) is the fruit of the mango tree, which belongs to the Anacardiaceae family [1]. Native to South Asia, particularly India, mango trees have spread widely across tropical regions worldwide [2]. Mangoes exhibit a wide range of varieties, differing in skin color, shape, size, texture, and taste. Automatic fruit ripeness detection plays a vital role in modern agricultural practices, especially in ensuring product quality, determining optimal harvest time, minimizing post-harvest losses, and delivering fruit to

* Corresponding author: anna_nnc@umy.ac.id

consumers in optimal condition [3]. Manual methods that rely on human visual inspection are subjective, inconsistent, and time-consuming [4].

Therefore, technology-based approaches such as digital image processing and artificial intelligence are increasingly being developed to classify fruit ripeness levels in a faster, more accurate, and objective manner [5]. These automated classification systems can assist farmers, distributors, and industry stakeholders in determining optimal harvest timing and market segmentation based on fruit maturity, thereby increasing market value and reducing waste [6].

In recent years, there has been growing interest in applying data modeling techniques to support more effective decision-making in agriculture and food processing [7]. One promising technique is the decision tree algorithm, which has proven effective for classification and prediction tasks across various domains [8]. The use of decision trees in fruit ripeness detection offers the potential for a more objective, automated, and efficient solution [9]. By utilizing observable physical attributes such as skin color, texture, and size, a decision tree model can be developed to classify fruit ripeness levels with high accuracy [10]. This can enhance the efficiency and accuracy of mango processing and distribution, as well as support better decision-making in orchard management.

Several studies have addressed mango ripeness classification using image-based features, particularly color and texture. Techniques such as RGB and HSV color histograms, as well as texture descriptors like GLCM and LBP [11], have been widely adopted in combination with machine learning models such as Support Vector Machines (SVM) and Neural Networks, often achieving high classification accuracy [12–13]. Meanwhile, shape descriptors such as Hu Moments have shown promise in fruit classification and grading, but their use in ripeness detection particularly for mangoes, remains limited. Additionally, while Decision Tree (DT) classifiers are known for their simplicity and interpretability, prior studies have not explored the integration of shape-based descriptors like Hu Moments with specific DT variants for mango ripeness classification.

To the best of our knowledge, there has been no prior study that combines Hu Moment shape descriptors with Decision Tree (DT) classifiers, specifically Fine, Medium, and Coarse variants for the classification of mango ripeness levels [14]. This study introduces a lightweight and interpretable approach by utilizing Hu Moment features extracted from mango skin images and evaluating their classification performance using different DT configurations. The novelty lies in the integration of shape-based moment features with interpretable machine learning models, providing an efficient solution for non-destructive mango ripeness classification.

2 Methodology

The research method employed in this study is illustrated in the following flowchart to present the steps undertaken throughout the research process.

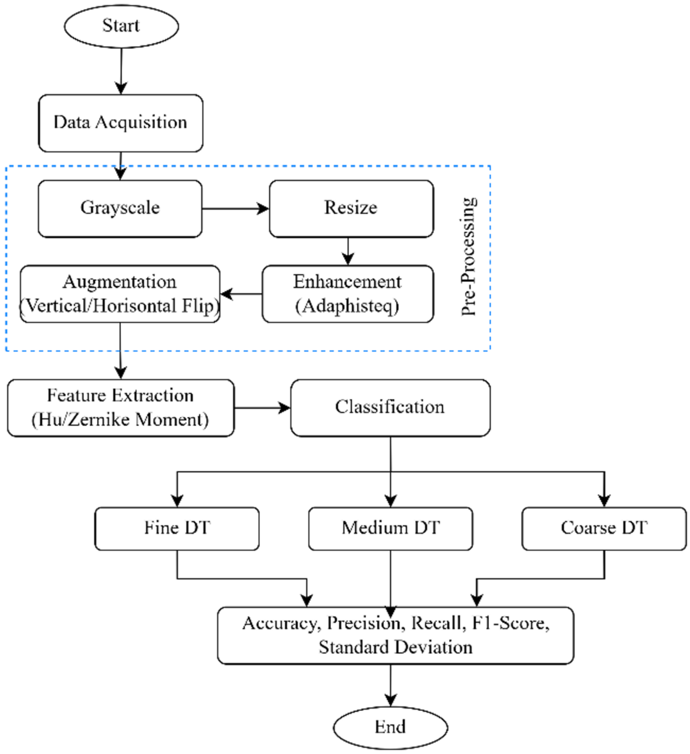


Fig. 1. Flowchart Research.

2.1 Data Aqcuisition

This study utilizes mango fruit images acquired by capturing photographs using a smartphone camera. The dataset comprises 54 images of ripe mangoes, 31 images of unripe mangoes, and 13 images of half-ripe mangoes.

2.2 Pre-Processing

The preprocessing stage involved several steps to standardize and improve the quality of the input images prior to feature extraction. Initially, the original RGB images were converted to grayscale to reduce computational complexity. Each image was then resized to a fixed resolution of 600 × 600 pixels to ensure consistency across the dataset. To enhance image contrast, adaptive histogram equalization (adapthisteq) was applied, which improves local contrast and emphasizes important structural details in the image. Furthermore, data augmentation was performed by applying vertical and horizontal flipping to each image. As a result, the dataset was expanded to a total of 294 images across all categories.

2.3 Feature Extraction

In this study, two different shape-based feature extraction techniques were applied separately: Hu Moment Invariants and Zernike Moment Invariants. Both methods are widely used in image processing for their robustness to geometric transformations, but they capture shape information in different ways. Hu Moments extract seven invariant features that are resistant to image translation, scaling, and rotation. They are computationally efficient and represent global shape characteristics. also,e features offer finer detail and are also invariant to rotation and less sensitive to noise. Each method was used independently to extract features from the preprocessed mango images. The extracted features were then classified using the Decision Tree (DT) algorithm with three model variations: FineDT, MediumDT, and CoarseDT. The goal was to evaluate and compare the classification performance of each feature extraction technique to determine which method yields the most accurate results in detecting mango ripeness levels.

2.4 Classification and Evaluation Metrics

The classification task in this study aimed to categorize mango images into three ripeness levels: ripe, half-ripe, and unripe. To achieve this, three variations of the Decision Tree (DT) algorithm were implemented: Fine Decision Tree (FineDT), Medium Decision Tree (MediumDT), and Coarse Decision Tree (CoarseDT). These variants differ in terms of tree depth and splitting granularity, which affect the model's complexity and generalization capability. Each DT model was applied independently to the features extracted using Hu Moments shape-based methods. The goal was to evaluate and compare the effectiveness of each feature extraction method combined with different DT configurations in classifying mango ripeness levels. To thoroughly assess the classification performance, the following evaluation metrics were used:

- Accuracy: The ratio of correctly classified instances to the total number of instances.
- Precision: The proportion of correctly predicted positive observations to all predicted positives, calculated per class.
- Recall (Sensitivity): The proportion of correctly predicted positive observations to all actual positives for each class.
- Specificity: The proportion of correctly predicted negative observations to all actual negatives, indicating the model's ability to detect true negatives.
- F1-Score: The harmonic meaning of precision and recall, useful in evaluating class-wise performance, especially with imbalanced datasets.
- Confusion Matrix: A table summarizing correct and incorrect predictions for each of the three classes.
- ROC Curve and AUC (Area Under the Curve): Used to analyze the model's ability to distinguish between classes in a multi-class setting.

By evaluating the results across all metrics, the best-performing combination of feature extraction method and DT model was identified for mango ripeness classification. This approach supports the development of an accurate, efficient, and interpretable system for fruit ripeness detection using image-based analysis.

3 Results and Discussion

This section presents and discusses the results obtained from each stage of the proposed mango ripeness classification system, including preprocessing, feature extraction, and classification. The performance of the system was evaluated using Hu Moments feature extraction techniques and three variations of the Decision Tree classifier: FineDT, MediumDT, and CoarseDT. Each combination was assessed based on accuracy, precision, recall, specificity, F1-score, confusion matrix, and ROC analysis. The goal of this evaluation is to identify the most effective method for classifying mangoes into three ripeness categories: ripe, half-ripe, and unripe. The following subsections detail the experimental outcomes and provide interpretations of the observed patterns and performance trends.

3.1 Pre-Processing

In this stage, a series of image preprocessing techniques were applied to prepare the data for feature extraction. The original RGB images were first converted to grayscale to reduce computational complexity, followed by contrast enhancement using adaptive histogram equalization (adapthisteq). To improve model robustness, image augmentation was also performed through vertical and horizontal flipping. The results of each preprocessing step are illustrated below.

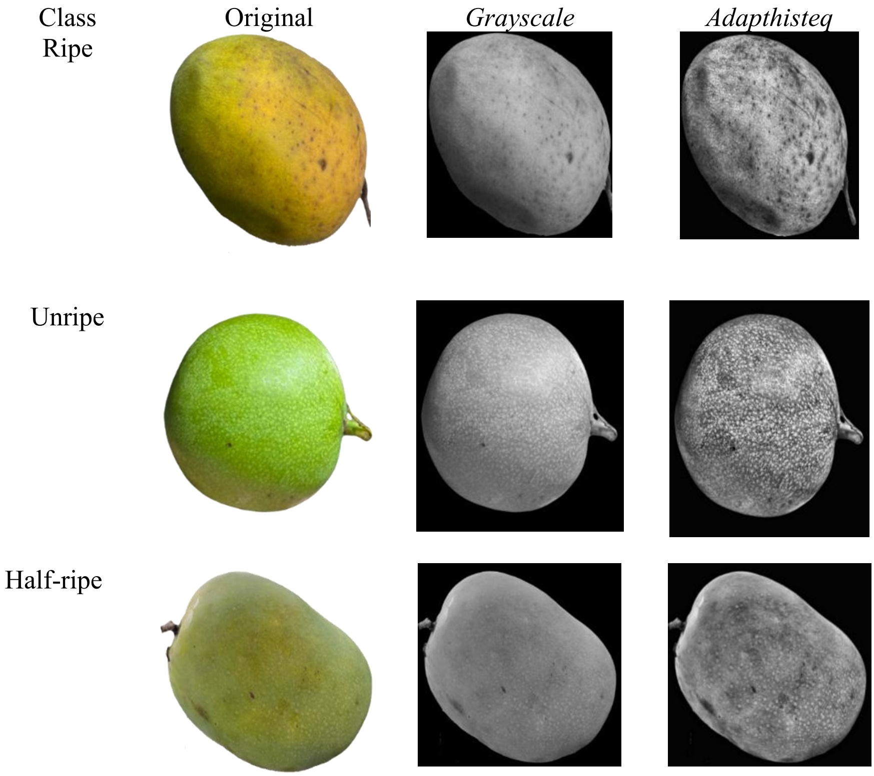
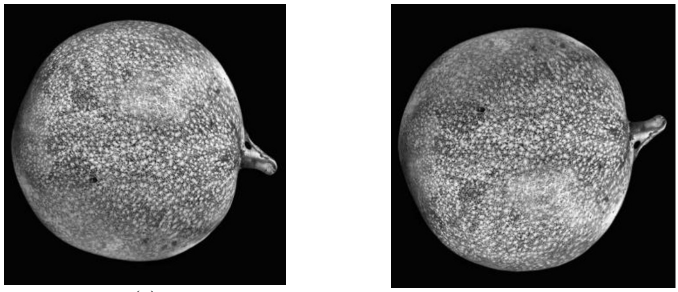


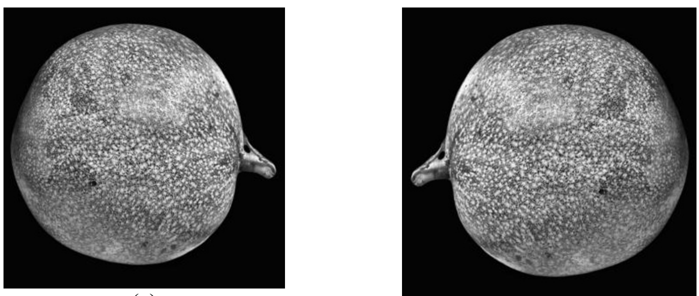
Fig. 2. Example of Original vs. Grayscale vs. Enhanced Images.

Figure X illustrates the visual transformation of mango images through grayscale conversion and contrast enhancement using adaptive histogram equalization (adapthisteq). Grayscale conversion preserves key structural features while reducing color complexity. The application of adaptive histogram equalization (adapthisteq) enhances local contrast, making surface textures and blemishes more visible, particularly in ripe and unripe samples. This improvement supports better feature extraction by highlighting important shape details essential for classification.



(a) (b)

Fig. 3. Examples of vertically flipped mango images.



(a) (b)

Fig. 4. Examples of horizontally flipped mango images.

3.2 Feature Extraction

This subsection presents the results of feature extraction using Hu Moment Invariants. Each preprocessed mango image was processed to generate seven numerical features that represent the global shape characteristics of the fruit. These features serve as the input for the classification stage and are expected to contribute significantly to distinguishing between the three ripeness categories. The 7 features are shown in Table 1, and the comparison visualization is shown in Figure 5.

Table 1. Feature Extraction Result.

Feature	**Class**		
	Ripe	**Unripe**	**Half-ripe**
1	0.001316116 ±6.94463E-05	0.001288951 ±3.11708E-05	0.001311038 ±2.16954E-05
2	7.64E-08 ±7.24E-08	1.85E-08 ±2.39E-08	4.51E-08 ±4.03E-08
3	2E-11 ±1.64E-11	6.21E-12 ±4.64E-12	3.2E-11 ±1.35E-11
4	4.18E-07 ±4.66E-08	3.95E-07 ±8.15E-08	4.25E-07 ±4.94E-08

Feature	Class		
	Ripe	Unripe	Half-ripe
5	-6.3E-24 ±6.28E-23	-4.9E-24 ±2.2E-23	1.21E-24 ±9.4E-24
6	-6.6E-17 ±8.92E-16	-5.7E-17 ±1.93E-16	5.86E-17 ±2.14E-16
7	1.66E-23 ±1.65E-22	-5.6E-25 ±2E-23	7.12E-24 ±2.55E-23

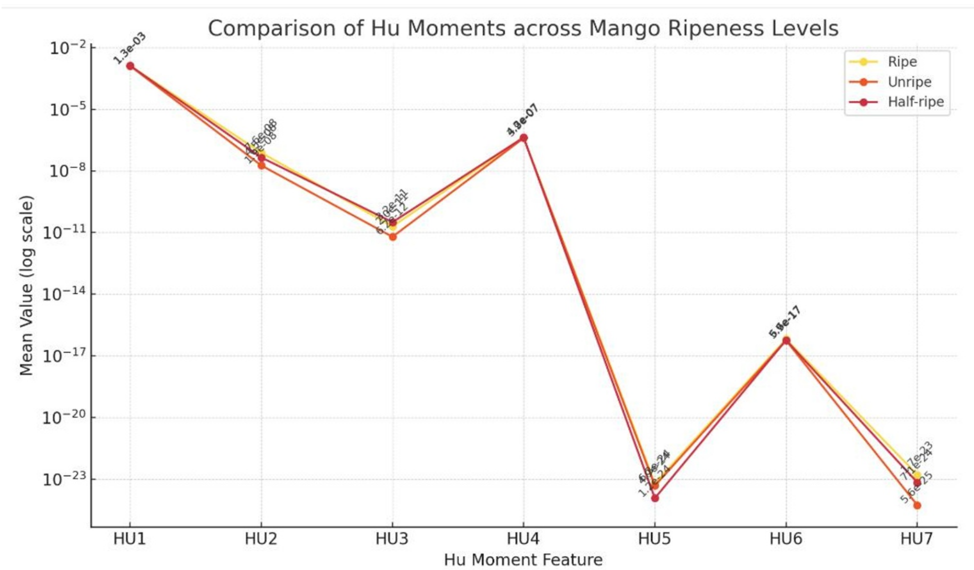


Fig. 5. Comparison of Hu Moments Results.

Based on Figure 5, it can be interpreted that the patterns are more stable and capable of distinguishing between classes, as shown in Hu Moments 1 to 4, where Hu Moments 2 and 3 appear to be fairly discriminative and variable across classes. Hu Moments 5 to 7 show relatively small values and appear less stable.

3.3 Classification and Evaluation Metrics

The classification performance for mango ripeness detection was evaluated using three variants of the Decision Tree algorithm: Fine, Medium, and Coarse, with a data split of 90% for training and 10% for testing. The evaluation metrics included accuracy, precision, recall, specificity, and F-score. In addition, training time was recorded to assess the computational efficiency of each model. These metrics provide a comprehensive overview of the classification performance, especially in balancing true positive detection and false positive avoidance. The classification results of each Decision Tree variant are summarized in Table 2.

Table 2. Classification Result.

Model	Training Accuracy (%)	Testing Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F-Score (%)	Training Time (s)
Fine DT	90.8	86.941	84.224	87.595	94.183	85.951	4.0016
Medium DT	90.8	86.971	84.174	87.595	94.183	80.489	4.4911
Coarse DT	81.6	86.971	84.426	79.279	94.183	85.707	20.5140

Based on table x, Fine DT requires the shortest training time (4.0016 seconds), followed by Medium DT (4.4911 seconds). On the other hand, Coarse DT has the longest training time (20.514 seconds), indicating higher model complexity that is not offset by significantly better testing performance. Overall, the Fine DT model offers the best balance between training time, classification accuracy, and F-score, making it a good choice for mango ripeness classification based on shape descriptors. A slight drop in testing accuracy compared to training accuracy is observed, which is expected due to the model being evaluated on unseen data. This gap remains within a reasonable range, indicating that the model generalizes well without significant overfitting.

4 Conclusion

This study demonstrated the potential of Hu Moment shape features in classifying mango ripeness using Decision Tree classifiers. Among the three variants tested—Fine, Medium, and Coarse—the Fine Decision Tree achieved the best balance of accuracy (86.941%) and training time. Despite comparable accuracy, Coarse DT required significantly longer computation. Overall, the results show that Hu Moments can effectively represent shape information for ripeness classification using lightweight and interpretable models. However, relying solely on shape features may limit robustness, suggesting future integration with color and texture descriptors.

References

1. E. M. Yahia, J. de J. Ornelas-Paz, J. K. Brecht, P. García-Solís, and M. E. Maldonado Celis, Arabian Journal of Chemistry **16**, 104860 (2023).
2. Mehta, *History of Mango-'King of Fruits'* (2017).
3. B. O. Olorunfemi, N. I. Nwulu, O. A. Adebo, and K. A. Kavadias, J Agric Food Res **16**, 101154 (2024).
4. Z. Zhou, U. Zahid, Y. Majeed, Nisha, S. Mustafa, M. M. Sajjad, H. D. Butt, and L. Fu, Front Plant Sci **14**, (2023).
5. S. Sultana, M. A. Moon Tasir, S. M. Nuruzzaman Nobel, M. M. Kabir, and M. F. Mridha, J Agric Food Res **18**, 101474 (2024).
6. Rojas Santelices, S. Cano, F. Moreira, and Á. Peña Fritz, Sensors **25**, 1524 (2025).
7. E. Elbasi, N. Mostafa, C. Zaki, Z. AlArnaout, A. E. Topcu, and L. Saker, Applied Sciences **14**, 8018 (2024).

8. B. Charbuty and A. Abdulazeez, Journal of Applied Science and Technology Trends **2**, 20 (2021).
9. Shivendra, K. Chiranjeevi, and M. K. Tripathi, in (2023), pp. 127–139.
10. Kaur, K. S. Gill, S. Malhotra, and S. Devliyal, in *2024 4th Asian Conference on Innovation in Technology (ASIANCON)* (IEEE, 2024), pp. 1–5.
11. S. Kousar, E. R. Raj, E. Shweta Bala, and M. Tech, *A CNN-LBP Image Modeling and Classification Scheme for Mango Leaf Disease Detection* (2021).
12. H. M. Rizwan Iqbal and A. Hakim, International Journal of Fruit Science **22**, 95 (2022).
13. Z. Dong, J. Wang, P. Sun, W. Ran, and Y. Li, Journal of Food Measurement and Characterization **18**, 2237 (2024).
14. N. N. Chamim, H. Ali, Y. Jusman, S. N. A. M. Kanafiah, M. I. Yusof, and I. R. Siddik, in *2023 International Conference on Artificial Intelligence Robotics, Signal and Image Processing (AIRoSIP)* (IEEE, 2023), pp. 427–432.