Oil Palm Fruits Ripening Prediction using Supervised Machine Learning Techniques

Chuan-Chin Teo 1, a) Han-Foon Neo 1, b) and Sheng-Qin Lim 1, c)

1Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Bukit Beruang, 75450 Melaka, Malaysia   
  
a) Corresponding author: ccteo@mmu.edu.my

b) hfneo@mmu.edu.my

c) limshengqin@gmail.com

**Abstract.** This paper focuses on using Gabor filtering, Haralick feature extraction, and a combination of nine different algorithms to construct image classification techniques for fruit ripeness assessment. To improve image quality, preprocessing techniques are used, and segmentation algorithms are used to extract the regions of interest containing the fruits. Texture features are extracted using Gabor filters of various orientations and frequency, whereas Haralick feature extraction computes statistical measures capturing textural qualities. The data is subsequently separated into training and testing sets for the training of the algorithms. Multiple algorithms, including Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Random Forest Classifier (RF), and Linear Support Vector Classifier (Linear SVC), are individually trained and combined to explore the fusion of different algorithms. The trained models are evaluated using the testing data, and the results, including accuracy scores, confusion matrices, and classification reports, are analyzed to measure the proposed algorithms' performance.

# INTRODUCTION

The world population is expecting to surge in future and the pressure are on agriculture to ramp up yields sustainably [1]. There are a number of factors that can affect the production of crops. For example, weather, irrigation practices, pest control measures and even fruit ripeness assessment. The implementation of AI-driven solutions in agriculture are possibly the best approach to target the mentioned factors and farmers now have access to cutting-edge solutions that greatly improve crop management efficacy.

Nowadays, AI technology is especially helpful in oil palm agriculture, which is largely practised in tropical countries such as Indonesia and Malaysia, which are the world's largest and second-largest producers of oil palms, respectively. Consumers choose high-quality vegetable oils generated by oil palm. As a result, there is presently a high demand for oil palm products. proposes to monitor the ripeness of oil palm fruits in order to achieve these standards and ensure effective output.

With the exact and trustworthy data on the level of fruit ripeness through data collection and the usage of filtering algorithms. Farmers will be able to optimize harvesting schedules, eliminate waste, and increase the quality of oil palm produce with this knowledge. By applying AI technology to assess the freshness of the fruit, oil palm growers can make more educated decisions, enhance their operations, and increase output. The initiative's purpose is to assist the oil palm industry's long-term growth and productivity, which will benefit both farmers and consumers.

# Related Works

Computer vision techniques such as deep learning and machine learning can be very useful in agriculture. In general, there are four stages of oil palm ripening (unripe, under-ripe, ripe, and over-ripe). An indicator of ripeness or maturity of given fresh fruit bunches (FFB) from oil palms depends on the use of an indicator which is RGB color scheme [2]. As a results, numerous methods have been proposed for fruits’ color detection. For example, a Support Vector Machine (SVM) is proposed by [3] for fruit recognition, [4] uses Artificial Neural Network (ANN) for oil palm grading, and [5] uses fuzzy classification based on color features for the prediction of apple ripeness.

On the other hand, a supervised machine learning technique named Linear Discriminant Analysis (LDA) is effective for feature extraction in classification and dimensionality reduction [6]. It maximizes the ratio of between-class variance to within-class variance and therefore effectively splitting both classes. On the other hand, [7] used Random Forest method in regression and classification tasks. [7] does not rely on a single decision tree but it generates predictions from each tree and anticipates the results using the majority of the guesses. The greater number of trees in the forest prevents higher accuracy and overfitting. Next, Quadratic Discriminant Analysis (QDA) is a classification task-oriented supervised machine learning technique. QDA starts by determining the key color characteristics pertinent to both the product and its sensory perception. For example, one significant attribute for fruit ripeness classification is color, this attribute could encompass factors like lightness, darkness, redness, or yellowness or anything that can directly influence its classification result. [8] used Quadratic discriminant analysis model for assessing the risk of cadmium pollution for paddy fields in China. The results showed that the accuracy rate was 74% with QDA in comparison using the decision tree and logistic regression models that the accuracy rate was 67% and 68% [8]. [9] proposed an idea and implementation on how a support vector network used for inter-class classification. This method is further improved in [10] for generic object categorization. Furthermore, Linear Support Vector Machines (SVMs) is ideal in addressing classification due to their computational efficiency in handling large-scale datasets [11]. The Linear Support Vector Classifier is categorized under the Support Vector Machines (SVMs) family which aims to find a hyperplane that splits the two classes in the input data. Linear SVC is widely utilized in text and picture categorization.

# Research Method

The implementation process of this research includes several fundamental steps for image classification using Gabor filtering, Haralick feature extraction, and a blend of nine separate algorithms. A total amount of 1988 palm oil fruits’ images were used, with the training and testing data partitioned into four distinct ripeness stages, namely unripe, ripe, over-ripe, and under-ripe.

Image preprocessing techniques such as noise suppression and enhancement are used to enhance the photo quality. The fruit regions of interest are determined by segmentation techniques such as edge detection and color-based segmentation. The construction of Gabor filters with various orientations and frequencies follows. These filters are then applied to the segmented fruit sections to obtain tangible data and information. The Haralick feature extraction procedure, which dictates statistical measurements collecting textural information, will be applied to the Gabor-filtered images.

Succeeding the feature extraction stage, the data is classified into two sets. The training data attributes are utilized to train all nine various types of algorithms, including Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Random Forest Classifier (RF), and Linear Support Vector Classifier (Linear SVC). Each algorithm learns the discriminative patterns and links between the characteristics and the appropriate fruit ripeness categories. Besides individual algorithms, several combinations of the four primary algorithms (RF, Linear SVC, LDA, and QDA) are examined. These combinations include RF + Linear SVC + LDA + QDA, RF + LDA + QDA, LDA + QDA, RF + LDA, and RF + QDA. These combinations put to the test the theory that combining different strategies can improve classification results.

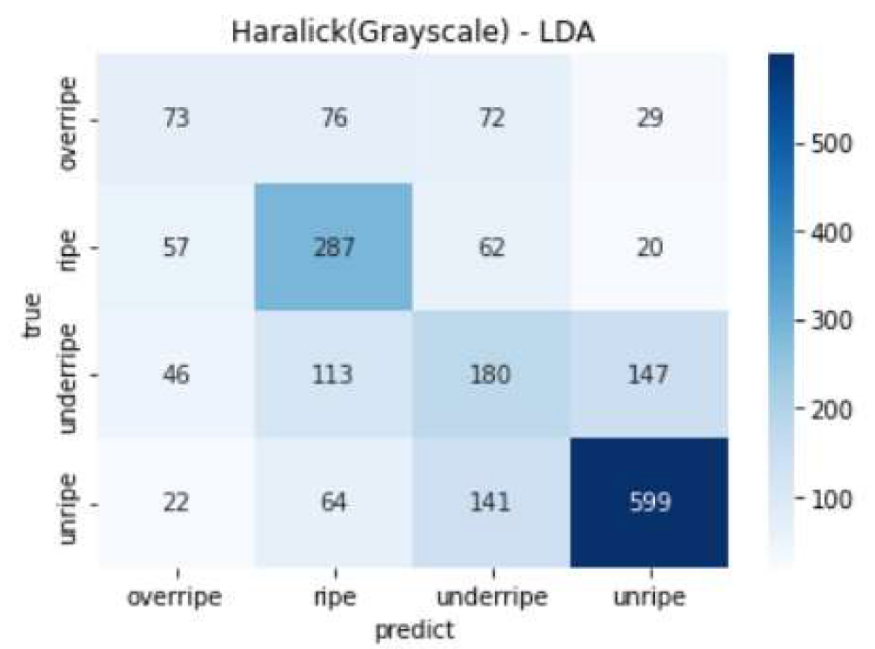
During the testing step, the trained models are applied to testing data, which undergoes the same preprocessing, segmentation, and feature extraction operations as the training data. Each algorithm, including the integrated models, uses the retrieved features to estimate the fruit ripeness categories in the test images. The implementation process yielded accuracy scores, confusion matrices, and classification reports. The accuracy score represents the status each algorithm fared overall in correctly distributing different stages of fruit ripeness. The confusion matrix provides a detailed breakdown of the correct and incorrect predictions for each category, demonstrating the advantages and weaknesses of the algorithms. The categorization report includes measurements such as precision, recall, F1-score, and support to provide a comprehensive assessment of the algorithms' presentation for each category.

Furthermore, the deployment approach involves recording relevant metrics and outcomes in an Excel file. This makes it straightforward to analyze the data that has been collected, making it effortless to measure and compare the effectiveness and speed of the algorithms. This method employs Gabor filtering and Haralick feature extraction to achieve accurate fruit ripeness rating. While the variation in algorithms and their combinations includes different ranges of approaches to the classification issue, the preprocessing, segmentation, and feature extraction procedures are critical in assembling data for categorization.

The implementation procedure and its result enhances the comprehension of various techniques and algorithms in identifying fruit ripeness. The implemented approach focuses on using Haralick texture features extracted from grayscale images and exploring different combinations of RF, Linear SVC, LDA, and QDA algorithms for fruit ripeness classification.

# RESULTS

Figure 1 shows the LDA confusion Matrix Heatmap-Grayscale. From Figure 1, the top-left cell (row 1, column 1) shows that the model correctly predicted 73 instances as "over-ripe" when they were actually "over-ripe." The cell in row 2, column 2 indicates that 287 instances were correctly classified as "ripe" and predicted as "ripe." The cell in row 3, column 3 shows that 180 instances were correctly classified as "under-ripe" and predicted as "under-ripe." The cell in row 4, column 4 indicates that 599 instances were correctly classified as "unripe" and predicted as "unripe”.



**Figure 1.** LDA confusion matrix heatmap-grayscale

Figure 2 shows that the top-left cell (row 1, column 1) shows that the model correctly predicted 250 instances as "over-ripe". The cell in row 2, column 2 indicates that 426 instances were correctly classified as "ripe" and predicted as "ripe". The cell in row 3, column 3 shows that 482 instances were correctly classified as "under-ripe" and predicted as "under-ripe". The cell in row 4, column 4 indicates that 750 instances were correctly classified as "unripe" and predicted as "unripe".

A screenshot of a graph

AI-generated content may be incorrect.

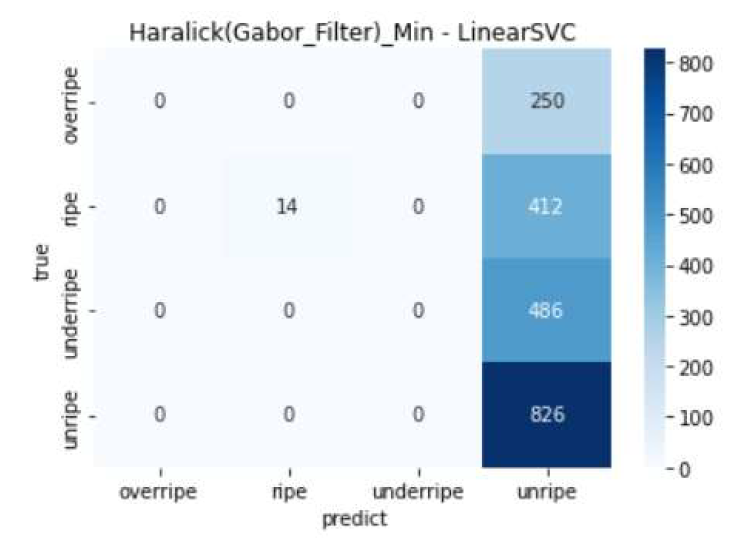
**Figure 2.** RF + QDA confusion matrix heatmap-sum

Figure 3 shows that the model correctly predicted 250 instances as "over-ripe". The cell in row 2, column 2 indicates that 426 instances were correctly classified as "ripe" and predicted as "ripe". The cell in row 3, column 3 shows that 486 instances were correctly classified as "under-ripe" and predicted as "under-ripe". The cell in row 4, column 4 indicates that 826 instances were correctly classified as "unripe" and predicted as "unripe".

In our experiment, the proposed Random Forest Classifier shows the overall accuracy of is 100% and it is very suitable to predict the ripeness or maturity of given fresh fruit bunches (FFB) from oil palms. On the other hand, Figure 4 shows the Linear SVC having the problem to determine the ripe of fresh fruit bunches (FFB) from oil palms. Therefore, it demonstrates the lowest accuracy in average. Lastly, Table 1 shows the overall accuracy of all proposed algorithms.

# 

**Figure 3.** RF confusion matrix heatmap-sum



**Figure 4.** Linear SVC confusion matrix heatmap-min

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **TABLE 1.** Accuracy of proposed algorithms | | | | | |  | |
| **Algorithms** | | | **Mean** | **Min** | **Sum** | **Grayscale** | **RGB** | |
| LDA | | | 78.72 | 81.24 | 78.62 | 57.29 | 58.05 | |
| QDA | | | 97.23 | 97.38 | 95.98 | 64.03 | 64.34 | |
| Random Forest Classifier | | | 100 | 100 | 100 | 100 | 100 | |
| Linear SVC  RF + Linear SVC + LDA + QDA  RF + LDA + QDA  LDA + QDA  RF + LDA  RF + QDA | | | 25.3  96.28  98.14  87.27  88.23  97.28 | 42.25  98.64  98.09  88.58  89.34  97.43 | 28.57  95.57  97.83  86.77  88.63  95.98 | 23.39  67.40  77.41  60.56  77.72  73.34 | 41.55  82.14  77.52  59.66  77.67  74.65 | |

# CONCLUSION

# The project aimed to develop a detection/monitoring system for assessing the ripeness of oil palm fruits using AI technology. The execution incorporated image preprocessing, feature extraction using Gabor filtering and Haralick analysis, and the implementation of nine different algorithms for specific identification. The results demonstrated that the Random Forest Classifier (RF) consistently obtained outstanding precision invariably, demonstrating its reliability in clarifying fruit ripeness. Combining RF with Linear SVC, LDA, and QDA also prompted high mean accuracies, illustrating the possibility of accurate classification when numerous algorithms are applied. The Linear SVC algorithm, on the other hand, had lower mean accuracies, proving the urge for additional optimization or other algorithm prospects. The efficiency of the algorithms differed when grayscale and RGB photographs were utilized, emphasizing the imperativeness of embracing the appropriate image format for satisfactory categorization. Looking at the picture, the project illustrated the effectiveness of Gabor filtering and Haralick feature extraction in obtaining textural features for fruit ripeness analyzation process. Combining these strategies can significantly enhance classification accuracy and durability. The successful completion of the project has promising implications for both oil palm industry and agriculture sector. The proposed detection/monitoring system can provide essential information to farmers and stakeholders, optimizing the harvesting process while boosting the total productivity. Furthermore, the project experienced a number of challenging complications to the implementation of AI technology in agriculture, ranging from data quality and availability, cost and accessibility, interoperability and standardization, ethical and social ramifications, to technical help. Addressing these struggles are crucial for advancing AI technology implementation in agriculture and intensifying its interests especially for farmers and stakeholders. To sum up, the project has significantly aided the evolution of AI technology in agriculture, uniquely in the evaluation of fruit maturity. The research findings and final outcome can contribute to future advances in this field, encouraging farmers with creative tools and techniques to improve their agriculture productions and accomplish stunning sustainable growth.

# References

1. D. Hemathilake. and D. Gunathilake, Agricultural productivity and food supply to meet increased demands, in Future Foods. Elsevier. pp. 539-553. (2022)
2. W. I. WanIsmail, M.Z. Bardaie and A. M. A. Hamid, Optical properties for mechanical harvesting of oil palm FFB. Journal of Oil Palm Research Vol. 12 No. 2, pp. 38-45 (2000)
3. M. T. Vo, T. D. Nguyen and N. T. Dang. Embedded Machine Learning for Mango Classification Using

Image Processing and Support Vector Machine. 6th NAFOSTED Conference on Information and Computer Science (NICS). pp. 279-284. (2019)

1. M. Junos, A. Khairuddin, S. Thannirmalai, and M. Dahari. An optimized YOLO based

object detection model for crop harvesting system. IET Image Processing, 15(9), pp. 2112-2125. (2021)

1. R. Hamza, and M. Chtourou, Design of fuzzy inference system for apple ripeness estimation using gradient method. IET Image Processing, 14(3), pp. 561-569. (2020)
2. R. M Ansari, M. A. Anvari, S. K. Monfared, S. Gorgin, Dimensionality Reduction in Machine Learning, Morgan Kaufmann, pp. 109-126. (2025)
3. L. Breiman, Random Forests. Machine Learning, vol 45, no 1, pp. 5-32. (2001)
4. X. Wang, X. Li, R. Ma, Y. Li, W. Wang, H. Huang, C. Xu and Y. An, Quadratic discriminant analysis model for assessing the risk of cadmium pollution for paddy fields in a county in China, Environmental Pollution, Volume 236, pp. 36-372. (2018)
5. C. Cortes and V. Vapnik, Support-vector networks. Machine Learning, vol 20, pp. 273–297 (1995).
6. F. J. Huang, and Y. LeCun, Large-scale learning with SVM and convolutional for generic object categorization. CVPR, pp. 284–291. (2006)
7. L. H. S. Torr, Locally Linear Support Vector Machines. Proceedings of the 28th International Conference on Machine Learning, pp. 985-992. (2011)