Development of a Machine Learning Model for Automated Palm Fruit Ripeness Classification

Tiara Laurensia Trisely[1*], Dwi Wahyu Prabowo[2]

[1,2] Information Sistem, Darwan Ali University, Jl. Batu Berlian No.10, Mentawa Baru Hulu Kec. Mentawa Baru Ketapang Kabupaten Kotawaringin Timur, Kalimantan Tengah, Indonesia

Keywords

Machine Learning; Oil Palm; Ripeness Classification; RGB Features; Support Vector Machine; Multi-Layer Perceptron; Precision Agriculture

****Correspondence Email:***
tiaranyyy@email.com

Abstract

Oil palm is one of Indonesia's most important plantation commodities, yet determining the optimal harvesting time remains a challenge because ripeness assessment is often conducted manually, leading to subjective and inconsistent results. This study aims to conduct a comparative analysis of machine learning models for oil palm fruit ripeness classification using digital images. The dataset, obtained from Mendeley Data, consists of 254 fruit images categorized into three ripeness levels: ripe, half-ripe, and unripe. The research process involved data preprocessing, RGB (Red, Green, Blue) feature extraction, model training using several machine learning algorithms including Support Vector Machine (SVM), Naive Bayes, k-Nearest Neighbors (KNN), Random Forest, Extreme Gradient Boosting (XGBoost), Multi-Layer Perceptron (MLP), and Ensemble Learning, followed by model evaluation using accuracy, precision, recall, and F1-score. The findings show that the SVM model achieved the best performance with an accuracy of 79.41%, followed by MLP with 78.43%. These results indicate that SVM and MLP are capable of effectively distinguishing ripeness levels based on color characteristics. The developed model can support the digitalization of agriculture by enabling more objective, efficient, and consistent fruit classification, thus improving harvest accuracy and productivity. This research demonstrates the potential of machine learning as a reliable tool for automated oil palm ripeness classification in precision agriculture.

1. Introduction

Oil palm (Elaeis guineensis) plays a vital role in Indonesia's economy, producing crude palm oil (CPO) and palm kernel oil (PKO) as essential raw materials for both food and non-food industries. As the world's largest palm oil producer, Indonesia contributes more than 50% of global supply, with major production centers located in Sumatra and Kalimantan. However, determining the optimal harvesting time remains a challenge, as the ripeness level significantly affects oil yield and quality. Unripe fruits produce low oil content, while overripe fruits risk oxidation that reduces oil quality(Khan et al., 2021).

Traditional ripeness assessment relies heavily on visual inspection, which is subjective and inconsistent across varying field conditions. Therefore, the application of artificial intelligence, particularly machine learning, offers a promising approach for automating ripeness detection based on objective data derived from digital images. Machine learning allows computers to learn from visual patterns and perform classification tasks with minimal human intervention(Knott et al., 2023).

This research focuses on conducting a comparative analysis of multiple machine learning models for palm fruit ripeness classification using RGB color features extracted from digital images. The dataset, obtained from Mendeley Data, includes 254 fruit images categorized into three classesripe, half-ripe, and unripe. The study involves data preprocessing, feature extraction, model training, and evaluation using accuracy, precision, recall, and F1-score. By comparing the performance of several algorithms, including SVM, KNN, Naïve Bayes, Random Forest, XGBoost, MLP, and Ensemble Learning, this study aims to identify which model performs best in classifying ripeness levels efficiently and accurately.

1.1 Literature Review

The study conducted by Sathe, (Sathe, 2021) evaluated a broad range of supervised learning algorithms including C5.0, J48, CART, Naïve Bayes, KNN, Random Forest, and SVM across three distinct datasets to predict students' academic performance. Their findings revealed that Random Forest and C5.0 consistently yielded the highest classification accuracy, particularly when hyperparameter tuning was applied to optimize model performance. The study identified demographic attributes, historical academic records, attendance levels, and learning behaviors as critical predictors, thereby emphasizing the necessity of robust feature engineering and model optimization. This study provides strong evidence for the superiority of tree-based ensemble models in handling complex tabular educational data with heterogeneous features. However, despite its methodological relevance to algorithmic benchmarking, the domain and data modalities fundamentally differ from the current study, which focuses on RGB image–based agricultural ripeness classification rather than structured academic datasets.

Phoenix et al. (Phoenix et al., 2024) introduced DateNET, a specialized CNN architecture designed for classifying date fruit varieties based on a combination of color and geometric attributes. Their results demonstrated a substantial improvement in validation accuracy from 85–87% to 93.41% when both color-based and geometric descriptors were integrated, indicating that multimodal feature representation strengthens discriminative capability. The study highlights the significant advantage of deep learning in capturing complex visual patterns that are difficult to extract manually. While the work is relevant to fruit image analysis, its focus lies in varietal identification rather than maturity classification. Moreover, its reliance on custom CNN architectures contrasts with the current research, which centers on evaluating classical machine learning models using handcrafted RGB features to classify palm fruit ripeness.

Villegas-camacho et al. (Villegas-camacho et al., 2025) applied multiple machine learning and deep learning algorithms such as k-NN, SVM, Naïve Bayes, Random Forest, MLP, and CNN to classify six industrial plastic types using FTIR spectral data. Their study revealed that Z-score normalization produced the most consistent and reliable performance across models. CNN, MLP, and Random Forest achieved near-perfect accuracy, while Naïve Bayes lagged considerably, illustrating the importance of both feature scaling and model complexity when dealing with high-dimensional spectral information. Although the comparative approach mirrors the analytical structure of the current research, the fundamental difference lies in its use of spectral FTIR data rather than color-based image features, placing it outside the agricultural domain of fruit ripeness detection. Li & Chen, (Li & Chen, 2020) examined five ensemble learning approaches Random Forest, AdaBoost, XGBoost, LightGBM, and Stacking to improve financial credit scoring accuracy. Their findings showed that ensemble models consistently outperformed baseline classifiers such as neural networks, SVM, logistic regression, decision trees, and Naïve Bayes, with Random Forest emerging as the most stable and reliable method across performance metrics including AUC, KS, ACC, and Brier Score. XGBoost and LightGBM also delivered strong performance while offering computational efficiency. Although this study reinforces the established superiority of ensemble methods, its focus on financial risk prediction differs substantially from the present agricultural context, which involves image-based ripeness classification of palm fruits using RGB features.

A more recent study by Taner et al. (Taner et al., 2023) proposed a classification model for ten apple varieties using a combination of texture features (Histogram of Oriented Gradients HOG) and color moments extracted from RGB images. Four machine learning models SVM, Random Forest (RF), Multilayer Perceptron (MLP), and K-Nearest Neighbor (KNN) were trained with K-fold cross-validation and optimized via GridSearch. The findings indicated that MLP and SVM achieved the highest accuracies, demonstrating the effectiveness of combining color and texture features for fruit classification tasks. This study provides valuable insights into feature fusion and classifier selection; however, it focuses on distinguishing fruit varieties rather than predicting ripeness levels and does not address tropical fruits with heterogeneous surface textures, such as oil palm fruit. Wang & Liang, (Wang & Liang, 2025) proposed a novel diversity metric, Double Fault Disagreement (DFD), to improve classifier selection in ensemble learning for software defect prediction. Their findings highlighted the importance of classifier diversity in preventing ensemble degradation and enhancing predictive robustness, especially in imbalanced datasets. This research aligns with global advancements in ensemble methodology but operates in the software engineering domain rather than agricultural image classification. Consequently, its contribution is methodological rather than directly applicable to image-based ripeness detection using handcrafted features.

Gulzar, (Gulzar, 2023) implemented transfer learning using MobileNetV2 for fruit classification, obtaining a remarkable accuracy of 99%, surpassing classical deep models such as AlexNet, VGG16, InceptionV3, and ResNet. Transfer learning reduced data requirements and accelerated training convergence, demonstrating the efficiency of lightweight architectures for real-world agricultural applications. Nevertheless, the study targets fruit type identification rather than ripeness classification and relies on deep architectures rather than classical ML models using raw RGB features, distinguishing it from the present research. (Surapunt, 2024) explored Bayesian Maximal Information Coefficient (BMIC) integrated with machine learning models such as SVM, Random Forest, neural networks, and XGBoost for high-uncertainty data prediction, achieving accuracy up to 96.3%. Their results demonstrate that Bayesian ensemble approaches outperform standard models, especially with noisy or incomplete datasets. Although methodologically relevant in terms of ensemble and uncertainty modeling, this study does not involve image data or agricultural classification, highlighting a clear divergence from the present research focused on RGB-based palm fruit maturity prediction.

Ahsan et al. (Ahsan et al., 2021) introduced a multimodal deep learning framework combining RGB imagery with hyperspectral data to classify papaya ripeness across six categories. By fusing morphological and biochemical information, the model achieved an F1-score of 0.90, demonstrating the superiority of multimodal sensing in fine-grained agricultural assessments. While technically advanced, this study differs from the present research by integrating hyperspectral imaging rather than relying solely on RGB images, and by adopting deep learning instead of classical ML approaches. Another related work by (Eriksson & Tabachnikova, 2022) study highlights that a wide variety of machine learning algorithms such as k-Nearest Neighbors (KNN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN) have been effectively used across agricultural applications. These models demonstrate strong capability in recognizing complex visual patterns when supported by appropriate feature descriptors, including color statistics, texture features, and shape-based information. Furthermore, the integration of multiple feature types (feature fusion) has been shown to enhance classification accuracy, particularly in products with diverse surface characteristics.

Alfatni et al. (Alfatni et al., 2022) proposed a real-time FFB ripeness classification system incorporating multivariate feature extraction techniques such as statistical descriptors, histograms, Gabor filters, GLCM, and BGLAM, which were later classified using SVM, ANN, and KNN. ANN with BGLAM features produced the highest accuracy and AUC (>93%) with rapid processing times (0.44 seconds), demonstrating feasibility for real-time field deployment. Despite its relevance, this study relies heavily on texture-based and multivariate features rather than simple RGB attributes, differentiating it from the current research's focus on lightweight RGB-based modeling.

Goh et al. (Goh et al., 2025) introduced a comprehensive multimodal outdoor FFB dataset comprising high-resolution RGB images, depth maps, and point clouds captured under diverse environmental conditions. This dataset addresses the shortcomings of previous datasets collected predominantly under controlled lighting (87%). While the work offers substantial contributions for detection and localization tasks, it does not directly address RGB-based ripeness classification, distinguishing it from the intent of the current study.

Worasawate & Sakunasinha, (Worasawate & Sakunasinha, 2022) classified mango ripeness using K-Means, Naïve Bayes, SVM, and ANN, incorporating physical, chemical, and electrical measurements to compensate for subtle visual differences among ripeness stages. ANN attained the highest accuracy of 89.6%, illustrating the importance of multisensor features. This study, although relevant to maturity assessment, differs significantly from the current research, which relies solely on RGB visual characteristics without additional physical or chemical inputs. Mamat et al. (Mamat et al., 2023) implemented YOLO-based automatic annotation for fruit detection, including palm oil FFB, achieving high mAP scores (98.7% for palm oil; 99.5% for other fruits). Their work demonstrated the potential of deep learning to automate dataset creation and reduce annotation workload. However, the focus is on object detection and annotation rather than supervised ripeness classification using handcrafted RGB features, distinguishing it from the current study's scope.

Across the reviewed literature, prior studies have demonstrated substantial progress in fruit classification, ripeness estimation, multimodal sensing, ensemble learning, and deep learning–based detection. However, there remains a lack of research dedicated specifically to evaluating multiple classical machine learning algorithms on purely RGB-based features for automated palm fruit ripeness classification involving three maturity levels (unripe, half-ripe, ripe). Many studies rely on deep learning architectures, multimodal datasets (RGB + hyperspectral), multisensor inputs (physical, chemical, electrical), or specialized feature extraction pipelines (texture, PCA, Gabor, HSI). Conversely, some palm-oil–specific research focuses on detection, dataset development, or ANN-PCA models, rather than systematic algorithm comparison using simple RGB information. Therefore, the current study fills this gap by providing a comprehensive evaluation of classical ML models using straightforward, easily extractable RGB features, offering a lightweight and practical solution for ripeness classification that does not depend on complex sensors, deep architectures, or high-cost multimodal imaging systems.

2. Research Methods

The research process was carried out through several systematic stages, as shown in the research flow diagram (Figure 1). Figure 1 illustrates the research steps implemented in this study.

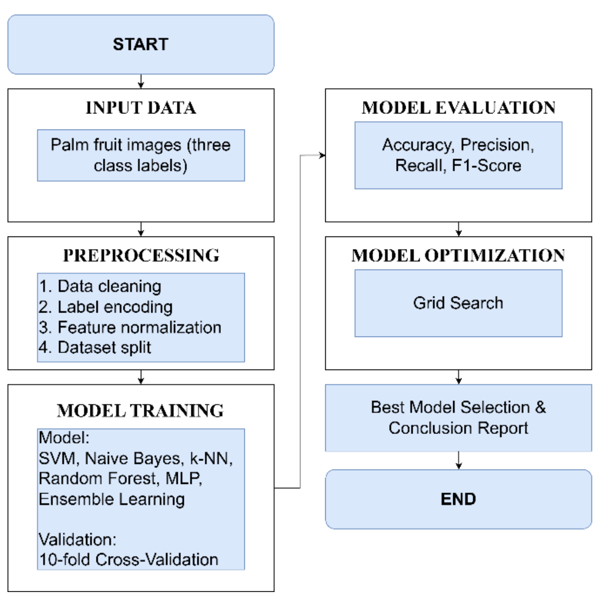

Figure 1. Research Flowchart

2.1 Sampling and Data Source

The dataset used in this research was sourced from an open-access repository on Mendeley Data and consists of 254 images of oil palm fruits categorized into three ripeness levels: Ripe with 91 images, Half-ripe with 72 images, and Unripe with 91 images. The images were captured under diverse lighting and background conditions, offering sufficient variability to support robust model generalization. Because the dataset was already labeled and publicly accessible, no manual data collection from field samples was required. Overall, this dataset represents a multiclass classification task aimed at categorizing each input image into one of the three predefined ripeness categories.

2.2 Data Preprocessing

Before model training, data preprocessing was conducted to ensure the quality and consistency of the dataset. The process began with data checking and missing value handling, where all 254 images were verified to confirm completeness and the absence of corrupted files. Next, label encoding was applied by converting the categorical labels ripe, half-ripe, and unripe into numerical values (0, 1, and 2) to ensure compatibility with machine learning algorithms. Feature normalization was then performed using Min–Max scaling to standardize the RGB features within a uniform value range. Finally, the dataset was split into training, testing, and validation subsets to effectively evaluate model performance and reduce the risk of overfitting.

2.3 Feature Extraction

The RGB color features were extracted from each image by converting the fruit image into the RGB color space and resizing it into a uniform resolution to simplify the feature extraction process (Liu et al., 2020). Each pixel contains three color intensity values corresponding to the Red (R), Green (G), and Blue (B) channels. The mean intensity value for each channel was calculated using the following formula:

$$\overline{C} = \frac{1}{N}\sum_{i=0}^{n} C_i$$

where:

- C_i represents the intensity value of the color channel at pixel i, and
- N is the total number of pixels in the image.

In addition to the mean value, the standard deviation of each color channel was also computed to capture the variation of pixel intensities across the fruit surface.

As a result, six numerical attributes were extracted from each sample:

- Mean values: mean_R, mean_G, mean_B
- Standard deviation values: std_R, std_G, std_B

These color-based features provide a compact numerical representation of the fruit's surface characteristics, which are strongly correlated with the ripeness level. The extracted feature vectors were subsequently used as inputs for various machine learning algorithms.

2.4 Machine Learning Model Training

The machine learning model training process involved implementing and evaluating multiple supervised learning algorithms to identify the most effective classifier for palm fruit ripeness detection. Machine learning is a branch of artificial intelligence that enables computers to learn from data, recognize patterns, and generate predictions without being explicitly programmed(Nadiah et al., 2022). In the context of this study, machine learning is employed to capture color patterns and visual characteristics of palm oil fruits, allowing the system to automatically and accurately determine their ripeness categories. The models used in this study included Support Vector Machine (SVM), Naïve Bayes, k-Nearest Neighbors (k-NN), Random Forest, Extreme Gradient Boosting (XGBoost), Multi-Layer Perceptron (MLP), and an Ensemble Learning approach. Each algorithm was trained using identical preprocessing outputs and parameter settings tailored to its characteristics to ensure a consistent evaluation environment. To enhance the reliability of the results, all models were assessed using 10-fold Cross-Validation, which helps reduce bias due to random data splitting and provides a more stable estimation of generalization performance(Rimal et al., 2025). The training workflow was standardized across all algorithms, allowing for a fair and systematic comparison of their predictive capabilities in classifying palm fruit ripeness levels.

2.5 Model Evaluation

Model evaluation was conducted using four primary performance metrics. Each metric is mathematically defined as follows:

- Accuracy: Measures the proportion of total correct predictions.

$$accuracy\ \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision: Indicates how many predicted positive samples are actually correct (Foody, 2023).

$$precision\ \frac{TP}{TP + FP}$$

- Recall (Sensitivity): Measures the model's ability to correctly identify all actual positive samples(Foody, 2023).

$$recall\ \frac{TP}{TP + FN}$$

- F1-Score: Represents the harmonic mean between precision and recall (Foody, 2023).

$$F - 1\, score = 2 \times \frac{precision \times recall}{precision + recall}$$

Where TP = True Positive, TN= True Negative, FP= False Positive, and FN= False Negative.

These four metrics were used as the primary indicators to assess the performance of each classification model.

3. Result and Discussion

3.1 Data Preprocessing Results

The preprocessing stage ensured that the dataset was clean, balanced, and ready for model training. A total of 254 images were used, divided into three ripeness categories: ripe (91 images), half-ripe (72 images), and unripe (91 images). Each image was verified to ensure there were no missing values or corrupted files. The category labels were numerically encoded as 0 (ripe), 1 (half-ripe), and 2 (unripe). Normalization was then applied to maintain consistent feature scales across all RGB channels.

Tabel 1. 1 Summary of Data Preprocessing Results

Component	Result Details
Number of Images per Category	Ripe: 91; Half-ripe: 72; Unripe: 91; Total: 254
Integrity Check (Corrupt Image Detection)	Ripe: 0; Half-ripe: 0; Unripe: 0; Total corrupt: 0
First Five Rows of the Dataset	All records show valid *image_path* values with label = 2 (specific category)
Label Distribution	Label 0 (Ripe): 91; Label 1 (Unripe): 91; Label 2 (Half-ripe): 72; Total: 254
Missing Value Check	image_path: 0; label: 0; Total Missing: 0

Component	Result Details
Preprocessing Summary	The dataset is clean, contains no missing values or corrupt files, and is ready for model training.

3.2 RGB Feature Extraction Results

The RGB color extraction process generated six numerical features for each image, representing both the mean and standard deviation of the red, green, and blue channels. These values characterize the color intensity patterns that differ across fruit ripeness levels. For instance, ripe fruits generally showed higher red intensity values, while unripe fruits had stronger green dominance. Table 1.6 illustrates examples of RGB feature extraction results, highlighting visible differences across classes that served as the foundation for machine learning classification.

Tabel 1. 2 Sample of Extracted RGB Features from Palm Fruit Images

No	mean_R	mean_G	mean_B	std_R	std_G	std_B	Label
1	117.39	92.88	92.31	62.89	61.98	59.59	Ripe
7	122.77	119.68	113.88	61.27	61.45	61.76	Half-ripe
11	105.50	99.53	94.63	63.21	62.72	62.24	Unripe

The differences between classes confirm that color intensity particularly, the red channel is a strong indicator of fruit maturity.

3. Model Evaluation (Train–Test Split 80:20)

The first evaluation used an 80:20 data split, where 80% of the images were used for training and 20% for testing. Seven machine learning algorithms were tested: Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Random Forest, Ensemble Learning, Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), and Naïve Bayes.

Tabel 1. 3 Evaluation Results Using 80:20 Data Split

Model	Accuracy	Precision	Recall	F1-Score
Support Vector Machine	**0.6275**	0.6028	0.5787	0.5750
Multi-Layer Perceptron	0.5686	0.5788	0.5509	0.5334
Random Forest	0.5492	0.5702	0.5231	0.5225
Ensemble Learning	0.5294	0.5167	0.5000	0.4939
XGBoost	0.4902	0.5004	0.4676	0.4669
KNN	0.4706	0.5400	0.4676	0.4634
Naïve Bayes	0.3922	0.4605	0.3704	0.3774

The SVM model achieved the highest accuracy (62.75%), demonstrating better capability in identifying the ripeness level based on RGB color features. Although the accuracy level is moderate, this indicates that SVM can learn distinct visual patterns more effectively than other algorithms at this stage.

3.4 Model Evaluation Using 10-Fold Cross-Validation

To ensure robustness and reduce random bias, the models were further tested using 10-Fold Cross-Validation. The results are shown in Table 1.8

Tabel 1. 4 10-Fold Cross-Validation Performance of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score
Multi-Layer Perceptron (MLP)	**0.7238**	0.7136	0.7257	0.7201
Ensemble Learning	0.7035	0.6686	0.6757	0.6811
Support Vector Machine	0.6928	0.6893	0.6749	0.6676
XGBoost	0.6643	0.6644	0.6518	0.6488
Random Forest	0.6603	0.6328	0.6313	0.6387
KNN	0.6137	0.5976	0.5985	0.5839
Naïve Bayes	0.4995	0.5087	0.4921	0.4789

The Multi-Layer Perceptron (MLP) showed the best overall performance with an average accuracy of 72.38%, followed by Ensemble Learning (70.35%) and SVM (69.28%). These results indicate that the MLP's ability to capture nonlinear relationships between RGB features contributes to improved classification performance.

3.5 Model Optimization Using Grid Search

The final optimization stage involved hyperparameter tuning through the Grid Search technique to achieve the most accurate configuration for each model. The performance results are summarized in Table 1.9.

Tabel 1. 5 Model Evaluation Results After Grid Search Optimization

Model	Accuracy	Precision	Recall	F1-Score
Support Vector Machine (SVM)	**0.7941**	0.7923	0.7941	0.7804
Multi-Layer Perceptron (MLP)	0.7843	0.7819	0.7843	0.7678
Ensemble Learning	0.6667	0.6438	0.6667	0.6474
XGBoost	0.6078	0.5925	0.6078	0.5949
Random Forest	0.5784	0.5537	0.5784	0.5620
KNN	0.5686	0.5378	0.5686	0.5475
Naïve Bayes	0.5294	0.5301	0.5294	0.5205

After optimization, the Support Vector Machine (SVM) achieved the highest accuracy of 79.41%, surpassing all other models. This suggests that SVM's ability to form optimal decision boundaries in high-dimensional feature spaces is particularly suitable for RGB-based classification of palm fruit ripeness.

3.6 Discussion

This study successfully developed and evaluated a machine learning–based approach for automated palm fruit ripeness classification using RGB color features extracted from digital images. Through a comparative analysis of seven machine learning algorithms SVM, MLP, Random Forest, XGBoost, KNN, Naïve Bayes, and Ensemble Learning the research validated that color information, represented through mean and standard deviation of RGB channels, provides an effective basis for distinguishing between unripe, half-ripe, and ripe oil palm fruits.

Across three evaluation stages (80:20 split testing, 10-fold cross-validation, and Grid Search optimization), the Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) consistently achieved superior performance compared to other models. After hyperparameter tuning, SVM obtained the highest accuracy of **79.41%**, followed closely by MLP with **78.43%**, demonstrating strong stability and generalization capability in learning nonlinear RGB-based patterns. These findings confirm that classical machine learning models particularly SVM and MLP remain highly effective for lightweight fruit classification systems, even without complex feature extraction techniques or deep learning architectures. The results also highlight the practicality of RGB-based features as a low-cost and easily implementable solution for agricultural automation, especially in environments where advanced imaging sensors are not feasible. By reducing subjectivity in manual ripeness assessment and enabling faster, more consistent decision-making, the proposed model supports the broader advancement of precision agriculture. Future work may incorporate additional color spaces, texture descriptors, or deep learning methods to further improve accuracy and enable real-time field deployment.

4. Conclusions

This study successfully conducted a comparative analysis of several machine learning models for classifying oil palm fruit ripeness based on RGB color features. The research demonstrates that digital image data can effectively represent color variations that correspond to ripeness levels, allowing machine learning algorithms to perform classification objectively and efficiently. Among the tested algorithms, the Support Vector Machine (SVM) achieved the best overall performance with an accuracy of 79.41%, followed by the Multi-Layer Perceptron (MLP) with 78.43%. These two models consistently delivered stable and reliable results across multiple evaluations, highlighting their effectiveness in recognizing RGB-based visual patterns. The findings indicate that machine learning, particularly SVM and MLP, can serve as reliable computational tools for automating fruit ripeness assessment, reducing human subjectivity, and enhancing decision-making in harvesting processes. The comparative results also provide valuable insights for future research aimed at optimizing and integrating machine learning approaches into agricultural digitalization and precision farming practices.

5. References

Ahsan, M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). *Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance*. 5–9.

Alfatni, M. S. M., Khairunniza-bejo, S., Marhaban, M. H. B., Osama, M., Saaed, B., Mustapha, A., Rashid, A., & Shariff, M. (2022). *Towards a Real-Time Oil Palm Fruit Maturity System Using Supervised Classifiers Based on Feature Analysis*.

Eriksson, I., & Tabachnikova, N. (2022). "Learning models." *LUMAT: International Journal on Math, Science and Technology Education*, *10*(2). https://doi.org/10.31129/lumat.10.2.1681

Foody, G. M. (2023). Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *PLoS ONE*, *18*(10 October), 1–27. https://doi.org/10.1371/journal.pone.0291908

Goh, J., Sultan, M., Ali, M., Yunos, Y., & Sheikh, U. U. (2025). *Outdoor RGB and Point Cloud Depth Dataset for Palm Oil Fresh Fruit Bunch Ripeness Classification and Localization*. 1–10. https://doi.org/10.1038/s41597-025-04953-6

Gulzar, Y. (2023). *Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique*.

Khan, N., Kamaruddin, M. A., Sheikh, U. U., Yusup, Y., & Bakht, M. P. (2021). Oil palm and machine learning: Reviewing one decade of ideas, innovations, applications, and gaps. *Agriculture (Switzerland)*, *11*(9), 1–26. https://doi.org/10.3390/agriculture11090832

Knott, M., Perez-cruz, F., & Defraeye, T. (2023). *Facilitated machine learning for image-based fruit quality assessment. 2022*.

Li, Y., & Chen, W. (2020). *A Comparative Performance Assessment of Ensemble Learning for Credit Scoring*.

Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R., & Cui, Y. (2020). Improved Kiwifruit Detection Using Pre-Trained

VGG16 with RGB and NIR Information Fusion. *IEEE Access*, *8*, 2327–2336. https://doi.org/10.1109/ACCESS.2019.2962513

Mamat, N., Othman, M. F., Abdulghafor, R., & Alwan, A. A. (2023). *Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach*.

Nadiah, N., Soim, S., & Sholihin, S. (2022). Implementation of Decision Tree Algorithm Machine Learning in Detecting Covid-19 Virus Patients Using Public Datasets. *Indonesian Journal of Artificial Intelligence and Data Mining*, *5*(1), 37. https://doi.org/10.24014/ijaidm.v5i1.17054

Phoenix, L., Rybacki, P., Niemann, J., Derouiche, S., Chetehouna, S., Boulaares, I., Seghir, N. M., Diatta, J., & Osuch, A. (2024). *Convolutional Neural Network (CNN) Model for the Classification of Varieties of Date Palm Fruits*.

Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M. P., & Gill, S. (2025). Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy. *Scientific Reports*, *15*(1), 1–14. https://doi.org/10.1038/s41598-025-93675-1

Sathe, M. T. (2021). *Comparative Study of Supervised Algorithms for Prediction of Students ' Performance*. *13*(1), 1–21. https://doi.org/10.5815/ijmecs.2021.01.01

Surapunt, T. (2024). *Ensemble Modeling with a Bayesian Maximal Information Coefficient-Based Model of Bayesian Predictions on Uncertainty Data*.

Taner, A., Mengstu, M. T., Selvi, K. Ç., Duran, H., Kabaş, Ö., Gür, İ., Karaköse, T., & Gheorghiță, N. E. (2023). Multiclass Apple Varieties Classification Using Machine Learning with Histogram of Oriented Gradient and Color Moments. *Applied Sciences (Switzerland)*, *13*(13). https://doi.org/10.3390/app13137682

Villegas-camacho, O., Francisco-valencia, I., & Alejo-eleuterio, R. (2025). *FTIR-Based Microplastic Classification : A Comprehensive Study on Normalization and ML Techniques*. 1–20.

Wang, J. I. E., & Liang, Y. A. N. (2025). A Novel Ensemble Classifier Selection Method for Software Defect Prediction. *IEEE Access*, *13*(January), 25578–25597. https://doi.org/10.1109/ACCESS.2025.3537658

Worasawate, D., & Sakunasinha, P. (2022). *Automatic Classification of the Ripeness Stage of Mango Fruit Using a Machine Learning Approach*. 32–47.