ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Image Processing-Based Machine Learning for Multi-Class Apple Bruise Classification

Daranat Tansui[1]

ABSTRACT

This research develops an automated apple bruise detection system that combines digital image processing and machine learning to enhance quality control. The study employs preprocessing techniques (resizing, grayscale conversion, thresholding) and extracts key features including pixel count, mean intensity, maximum and minimum pixel values, and estimated bruise area. A dataset of 513 apple samples was created and divided into training (70%), validation (15%), and test (15%) sets. Among the five evaluated classifiers, Decision Tree and Gradient Boosting demonstrate identical peak performance (99.35% accuracy, 0.9941 F1-score), with Decision Tree offering superior computational efficiency (200 times faster training). Random Forest achieves 98.70% accuracy, outperforming conventional methods (SVM, KNN, Logistic Regression). Notably, the Decision Tree accurately classifies severe bruises (Class 2), which is crucial for quality assurance. The system's effectiveness is validated through comprehensive metrics, including confusion matrices and ROC analysis. These results highlight the practical viability of implementing Decision Tree-based solutions in commercial fruit grading systems, offering an optimal balance between accuracy (99.35%) and operational efficiency (0.0012 seconds of training time). The findings enhance automated post-harvest inspection capabilities while addressing critical industry needs for rapid and reliable bruise detection.

Article information:

Keywords: Bruise Detection, Apple Quality, Image Processing, Feature Extraction, Machine Learning

Article history:
Received: August 12, 2025
Revised: November 20, 2025
Accepted: December 25, 2025
Published: January 17, 2026 (Online)

DOI: 10.37936/ecti-cit.2026201.263517

1. INTRODUCTION

In modern agriculture, quality control of agricultural products, especially fresh fruits, is a critical step both for commercial purposes and to ensure compliance with consumer safety standards. One of the key factors affecting the quality and economic value of fruits is "bruising", which can occur during harvesting, transportation, or storage [1]. Bruising is a common occurrence in apples that can lead to gradual fruit decay and substantial economic losses, with 10-25% loss reported in the Belgian apple industry during transportation alone. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery [1].

Bruises on fruits such as apples not only affect their external appearance but also influence spoilage rates, shelf life, and consumer satisfaction. Currently, bruise detection is commonly performed by human visual inspection, which has limitations in accuracy, consistency, and speed, particularly in high-volume or high-speed production lines [2]. Manual detection or utilizing machine vision on RGB images to detect bruises on apples with various skin colors in the early stages is quite challenging, especially when bruises are not easily distinguishable by the naked eye (PDF). Comparison of Logistic Regression, Random Forest, SVM, and KNN Algorithm for Water Quality Classification Based on Contaminant Parameters [2].

To address these limitations, researchers have explored integrating digital technologies with automation systems, such as digital image processing and machine learning, which offer potential to improve accuracy and reduce inspection time [3]. Supervised classification techniques including support vector machines, linear logistic regression, neural networks, decision trees, k-means clustering, random forest, and Extreme Learning Machine have been successfully applied in bruise detection, producing better results

[1]The author is with the Department of Computer and Informatics for Management, Faculty of Communication Sciences, Prince of Songkla University, Thailand, Email: daranat.t@psu.ac.th

than pure statistical and image processing methods (PDF) Comparison of Logistic Regression, Random Forest, SVM, KNN Algorithm for Water Quality Classification Based on Contaminant Parameters [2]. Previous studies have demonstrated the effectiveness of various imaging techniques for bruise detection, including Near-Infrared Imaging [4], Hyperspectral Imaging [5], and visible/near-infrared hyperspectral reflectance imaging for multispectral detection of apple bruises with optimized wavelength-specific spectral resolutions [6].

Recent advances in computer vision have shown promising results in apple quality assessment. Convolutional neural networks have been successfully applied to apple quality identification and classification, achieving training accuracy of 99%, validation accuracy of 98.98%, and an overall accuracy of 95.33% on independent test datasets [7]. This study used Decision Tree, Extra Trees, and XGBoost to predict fruit waste at a South African wholesale market (2021–2023). Decision Tree and Extra Trees had the lowest MAE (112.19), showing potential to guide market decisions and reduce postharvest waste [8].

Therefore, this research aims to develop a multi-class bruise classification system for apples using image analysis and machine learning techniques. The goal is to accurately classify the severity and characteristics of bruises, enabling efficient sorting of produce and supporting the implementation of automated quality assessment systems in commercial applications. Comparative studies have shown that classical machine learning algorithms such as Random Forest and Support Vector Machine can achieve high accuracy, with RF outperforming SVM by 1-2% in overall accuracy and 3% in Kappa coefficient [9].

In this work, we propose a Multi-Class Apple Bruise Classification system using digital image processing and machine learning, extracting features such as area, intensity, and pixel count to classify bruise severity into three levels. SVM offers high prediction accuracy, KNN performs well on small datasets, and Decision Trees provide strong interpretability. Consistent with [10], which achieved over 98% accuracy in apple variety classification using SVM, Random Forest, MLP, and KNN, our workflow combines feature extraction with multi-class performance evaluation. This paper is organized as follows:

Section 1 introduces the research background, motivation, and objectives. Section 2 reviews related works in image-based fruit quality assessment and machine learning classification. Section 3 presents the research methodology, including image preprocessing, feature extraction, and dataset preparation. Section 4 presents experimental results, a performance comparison, and an evaluation using confusion matrices to statistically validate the significance of the performance. Section 5 concludes the study and discusses possible directions for future work.

2. RELATED WORK

2.1 Digital Image Processing for Fruit Quality Assessment

Digital image processing has become a key enabler for non-destructive fruit quality assessment, offering greater speed and objectivity than traditional inspection. Hyperspectral imaging has achieved recognition rates above 92% for early bruise detection [5], with wavelength-specific analysis further supporting real-time sorting applications [6]. Near-infrared (NIR) methods have reported classification accuracy exceeding 99% using deep learning models, such as Faster R-CNN and YOLO variants, demonstrating their suitability for real-time deployment [4]. Similarly, short-wave infrared (SWIR) has demonstrated advantages in detecting subsurface bruises, achieving a 97.4% mean average precision (mAP) [11]. Although hyperspectral and infrared systems deliver high accuracy, RGB-based approaches remain attractive for cost-sensitive settings, achieving over 95% accuracy with CNN models trained on images under complex backgrounds [12]. These findings highlight the feasibility of integrating both advanced and conventional imaging modalities with machine learning to enhance automated quality assessment.

2.2 Machine Learning Algorithms in Agricultural Applications

Classical machine learning algorithms remain central to agricultural image analysis, balancing accuracy, interpretability, and computational efficiency. Random Forest (RF) often outperforms Support Vector Machines (SVM), as shown in land cover mapping, where RF achieved higher accuracy (0.86) and Kappa scores (0.83) in mixed-class scenarios [9]. Nevertheless, SVM maintains strong predictive power in high-dimensional tasks, with accuracies above 97% in multiclass apple quality classification [13]. K-Nearest Neighbor (KNN) also demonstrated 96.6% accuracy, making it suitable for small datasets and resource-limited applications. In contrast, Decision Trees (DT) achieved 93% accuracy, offering transparent decision rules that support practical deployment [14]. Comparisons of optical property mapping further demonstrated the superiority of SVM at 98.33% and CNN at 99.16% (two-class), although CNN required substantially longer processing time, confirming the trade-offs between performance and efficiency.

2.3 Apple Defect and Bruise Detection Systems

Advanced imaging has proven effective in identifying apple defects and bruises at early stages. Short-wave infrared hyperspectral imaging detected latent bruises invisible to the naked eye, with SVM providing the best detection and ESD achieving over 85% accuracy in severity classification [15]. Optical property mapping has also enabled detection within 1–2 hours after the occurrence of a bruise, enhancing visibility and measurement precision [16]. High-speed systems utilizing pruned YOLOv4 on NIR images achieved 93.9% mAP at 5 fruits per second [3], [18], while grading frameworks tailored to specific varieties, such as Golden Delicious, reached 92.5% accuracy. Recent model comparisons have further indicated that SSD outperforms YOLOv2 in RGB-based defect detection [17], [19], confirming that algorithm choice remains critical, even with conventional imaging.

2.4 Machine Learning Approaches for Fruit Classification

Machine learning has been widely applied to fruit classification, offering efficiency, accuracy [20], and consistency in automated quality control. Random Forest (RF) has shown strong generalization, often outperforming SVMs in land cover mapping and disease detection [21], while K-Nearest Neighbors (KNN) and Decision Trees (DT) remain valuable for small datasets and for interpretable decision rules [10]. Logistic Regression also remains relevant for probabilistic classification [19]. These traditional machine learning models continue to play an important role in fruit quality assessment. Comparative studies indicate that no single model consistently dominates across all contexts, emphasizing the need to balance accuracy, interpretability, dataset size, and computational constraints. In agricultural applications, DT offers real-time sorting transparency due to its interpretability, whereas ensemble methods such as RF provide greater robustness at a higher computational cost.

3. RESEARCH METHODOLOGY

3.1 Research Framework

This study employs a systematic methodology for bruise classification in apples by integrating digital image processing with classical machine learning techniques. The framework (*Fig.*1) consists of data acquisition, preprocessing, feature extraction, label assignment, dataset preparation, and model training [18],[20]. The dataset, collected from Roboflow, comprises 513 images after augmentation. Each image was converted to grayscale, thresholded, and processed using contour detection to identify bruise regions. Extracted features were used to label bruise severity, followed by dataset partitioning and predictive modeling with machine learning [22].

3.1.1 Research Objectives

First Objective: To develop an image processing–based feature extraction process from apple images for classifying bruise severity levels.

Second Objective: To develop and optimize a Decision Tree model for classifying bruise severity in apples with high accuracy and fast processing speed.

Third Objective: To compare the performance of seven machine learning algorithms and statistically validate the significance of performance differences to identify the most suitable model for real-world deployment in automated fruit sorting systems.

Fig.1: *The research framework for multi-class bruise classification.*

3.1.2 Four-Phase Framework

The research framework is structured into four sequential phases:

Phase 1: Data Collection — A representative dataset was obtained from Roboflow and expanded to 513 images through augmentation from the original 244 images. This addressed the limitation of small, labelled agricultural datasets while preserving class distribution.

Phase 2: Data Processing — Standardized preprocessing (grayscale conversion, thresholding, and contour detection) ensured consistent feature extraction and reduced variability introduced by acquisition conditions.

Phase 3: Model Training — Seven machine learning algorithms were systematically compared using hyperparameter tuning and cross-validation. This phase ensured that the Decision Tree was selected as the optimal model based on evidence rather than arbitrary preference.

Phase 4: Model Evaluation — A comprehensive performance assessment employed multiple metrics (accuracy, F1-score, confusion matrices, and ROC curves) to confirm both statistical significance and real-world deployment readiness. Together, these four phases establish a rigorous pipeline from data acquisition to model validation, ensuring that the proposed system is not only statistically sound but also practically feasible for automated apple grading applications [23].

3.1.3 Technology Stack and Implementation Platform

The research leverages open-source tools to ensure reproducibility and transparency. Image processing is performed using OpenCV 4.5 or later and PIL, while machine learning algorithms and evaluations are implemented with scikit-learn 1.0 or later and XGBoost 1.6 or later. NumPy and Pandas handle numerical and structured data operations, with Matplotlib and Seaborn used for visualization. All experiments are conducted in Google Colab Pro, ensuring consistent cloud-based execution across different hardware environments.

3.2 Data Collection and Preprocessing

3.2.1 Dataset Characteristics and Sources

This research utilizes an apple bruise detection dataset from Roboflow, a platform that provides high-quality, annotated images. The dataset initially contained 244 RGB images with varying bruise severity [24], each professionally annotated with bounding boxes delineating bruise regions.

Dataset specifications:

- Total images: 244 high-resolution samples
- Resolution: 640 × 640 pixels (standardized)
- Annotation: Bounding box coordinates with "bruise" labels
- Source: Robo flow [25]

The dataset's consistent quality, standardized resolution, and precise annotations ensure suitability for machine learning, with background uniformity minimizing noise and bounding boxes providing accurate spatial coordinates for feature extraction.

3.2.2 Data Augmentation Techniques

To overcome the limited sample size, systematic data augmentation was applied to the 171 training images, expanding the dataset to 513 samples. Four transformations were used:

- Horizontal flipping – increases positional variance while preserving bruise characteristics.
- Rotation (±15°) – simulates natural handling variations.
- Brightness adjustment (±15%) – accounts for diverse lighting conditions.

Controlled blur (≤2.5 pixels) – mimics motion or focus variations.

This augmentation strategy expanded the dataset without additional collection, improved robustness through realistic variation, and reduced overfitting. All transformations preserved biological realism while maximizing diversity (*Fig*.2).

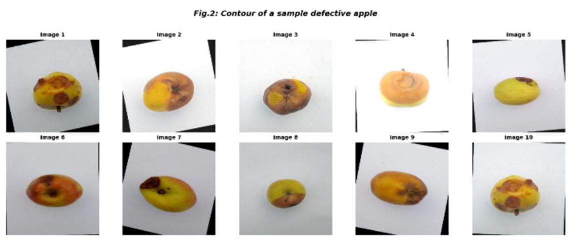

Fig.2: *Contour of a sample defective apple.*

3.2.3 Image Preprocessing with OpenCV [26]

Image preprocessing is a crucial step for transforming raw visual data into formats suitable for quantitative analysis. This study utilizes OpenCV 4.5+ as the primary framework, implementing a three-stage pipeline specifically designed for apple bruise detection. First, RGB images are converted into 8-bit grayscale using the standard luminance-weighted transformation (Eq. 1).

$$G(x,y) = 0.299 \cdot R(x,y) + 0.587 \cdot G(x,y) + 0.114 \cdot B(x,y) \quad (1)$$

where $G(x,y)$ represents the grayscale intensity at pixel coordinates (x,y), and R, G, and B denote the red, green, and blue channel values, respectively. The Grayscale Conversion and Threshold Mask is shown in Fig. 3.

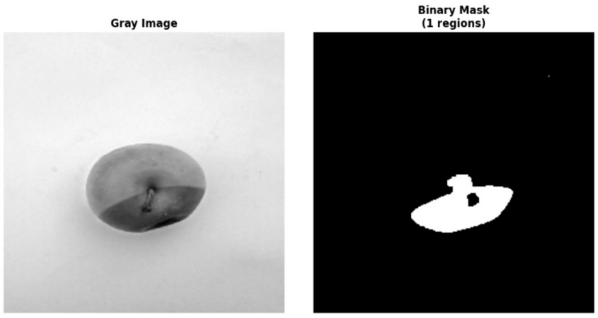

Fig.3: *Grayscale Conversion and Threshold Mask.*

$$B(x,y) = \left\{ \begin{array}{ll} 255 \; if \; G(x,y) \; \leq \; 120 \\ 0 \quad if \; G(x,y) \; > \; 120 \end{array} \right\} \tag{2}$$

where $B\ (x, y)$ represents the binary mask and 120 is the empirically determined threshold.

Following thresholding, the Suzuki–Abe algorithm detected bruise contours with sub-pixel accuracy, generating coordinate-based perimeters for feature extraction. Morphological operations were employed to remove noise, and only contours exceeding a minimum area were retained, thereby ensuring an accurate geometric representation of bruise regions.

3.2.4 Feature Extraction Methods

Mean Intensity (μ):

The mean intensity, a primary indicator of bruise severity, is calculated as:

$$\mu = \frac{1}{N} + \sum\nolimits_{i=1}^{N} I_i \tag{3}$$

where μ represents mean intensity, N denotes total pixels in the bruise region, and I_i represents individual pixel intensity values.

Post-thresholding contour detection precisely maps bruise boundaries, with the total affected serving as a key indicator of severity - larger areas correlate with greater damage ($Fig.$4). The method's sub-pixel boundary tracing ensures accurate geometric feature extraction for subsequent classification.

Contour Area (A):

Post-thresholding contour detection maps bruise boundaries, with the total affected area serving as a key indicator of severity. The area is computed as in equation (4)

$$A = \sum\nolimits_{j=1}^{n} Area(C_i) \tag{4}$$

Where A represents the total bruise area in pixels^2, C_j

C_j denotes the $j-th$ individual contour, and n is the total number of contours detected in the binary image.

Alternative approach using pixel counting:

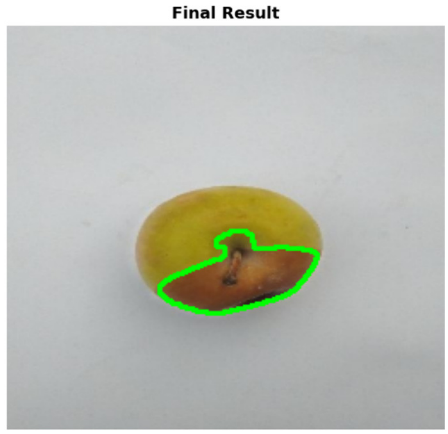

Fig.4: *Contour detection.*

$$P = \sum\nolimits_{j=1}^{n} |C_j| \tag{5}$$

where P represents total pixel count and $|C_j|$ denotes the number of pixels within contour j.

Intensity Range Analysis:

The intensity range within bruise regions provides critical information on severity and uniformity. The minimum intensity represents the most severely damaged tissue, while the maximum intensity corresponds to the transition zone between bruised and healthy areas. Together, these values strengthen the discriminative power of the feature set for Decision Tree classification.

Minimum Intensity Extraction:

The minimum intensity identifies the darkest pixels within detected bruise regions, calculated according to equation (6):

$$I_{min} = min_{i \in \text{bruise regionmin}} \tag{6}$$

The maximum intensity captures the lightest pixels within bruise boundaries, providing information about damage periphery, expressed as equation (7):

$$I_{max} = max_{i \in \text{bruise regionmin}} I_i \tag{7}$$

where I_{min} and I_{max} represent the minimum and maximum intensity values, respectively, and I_i denotes the intensity value of individual pixels within the identified bruise regions.

The complete intensity range characterizes the full spectrum of damage severity within each bruise, calculated as:

$$\Delta I = I_{max} - I_{min} \tag{8}$$

where ΔI represents the intensity range, providing a measure of bruise heterogeneity and internal variation.

In the process of extracting five key characteristics from each bruise region as described in **Algorithm 1**, the number of pixels in the bruise was 4,088, the average brightness was 89.52739726027397, the minimum brightness was 23, and the maximum brightness was 120.

Algorithm 1: Comprehensive Feature Extraction

Input: Grayscale image G (x, y), Contour set C
Output: Feature vector F = [μ, A, P, I_{min}, I_{max}]

1. Initialize feature vector F ← [0, 0, 0 ∞, -∞]
2. Create binary mask M from contour set C
3. Extract bruise pixel intensities:
   For each pixel (x, y) where M (x, y) = 1:
   a. Collect intensity value G(x, y)
   b. Update I_{min} and I_{max}
4. Calculate statistical measures:
   a. μ ← mean of collected intensities
   b. A ← sum of contour areas using Green's theorem
   c. P ← count of bruise pixels
5. Construct feature vector F ← [μ, A, P, I_{min}, I_{max}]
6. Return normalized feature vector F

The extraction process systematically analyses pixels within the detected contour, computing statistical measures that quantify the bruise's properties. The resulting feature vector provides a compact numerical representation of complex visual information, enabling quantitative comparison and automated classification. Algorithm 1 formalizes this comprehensive feature extraction workflow.

3.2.5 Dataset Splitting and Preparation

To prevent data leakage and ensure unbiased model evaluation, the 513-image dataset was partitioned prior to any augmentation. Stratified sampling was used to maintain class proportions, resulting in a 70% training (359 images), 15% validation (77 images), and 15% testing (77 images) split. Data augmentation was performed exclusively on the training set, while the validation and test sets were left untouched. This split-then-augment strategy follows established best practices in deep learning workflow design [27] and aligns with the methodological approaches used in agricultural bruise detection studies [28]. Augmentation techniques included rotation, flips, brightness adjustments, and minor geometric transformations to improve generalization without altering the inherent bruise characteristics.

Dataset integrity was verified through systematic quality checks assessing image consistency, completeness, and label accuracy. Table 1 summarizes the results of these quality assurance procedures, confirming the dataset's reliability for machine learning experiments.

Understanding feature characteristics enables appropriate preprocessing decisions and parameter optimization strategies. Table 2 presents comprehensive statistical summaries for all five extracted features across the complete 513-sample dataset.

Table 1: *Stratified Class Distribution Across Partitions.*

Dataset	Class 0	Class 1	Class 2	Total
Training	122	140	97	359
Validation	27	31	19	77
Test	26	30	21	77

Table 2: *Presents comprehensive statistical summaries.*

Parameter	Options Tested	Selected	Reason
Max depth	5,8,10,12	8	Best CV score
Min samples split	10,15,20	15	Prevents overfitting
Min samples leaf	5,8,10	8	Stability
Criterion	Gini, entropy	Gini	Multi-class optimal

3.3 Model Training Methodology

3.3.1 Training Protocol and Implementation

The training phase centered on developing the Decision Tree model while systematically comparing alternative algorithms to optimize bruise severity classification. All experiments were performed on Google Colab Pro using Python 3.8+, scikit-learn 1.0+, NumPy, and Pandas, with a fixed random seed (42) for reproducibility. The 513-sample dataset was preprocessed before training, with feature scaling applied only to algorithms that require normalized inputs (SVM, KNN), while tree-based models used raw features to preserve interpretability.

3.3.2 Decision Tree Hyperparameter Optimization

Decision Tree optimization employed grid search with cross-validation to balance accuracy, interpretability, and efficiency while avoiding overfitting. The optimization objective function is defined as: Optimization Objective Function:

The hyperparameter selection prioritizes balanced performance across multiple criteria expressed as EQ:

$$\text{Obj} = \alpha{\cdot}\text{Accuracy}+\beta{\cdot}\text{Speed}+\gamma{\cdot}\text{Interpretability} \quad (9)$$

where $\alpha = 0.6$, $\beta = 0.3$, and $\gamma = 0.1$ reflect the relative importance of accuracy, computational efficiency, and model interpretability, respectively.

3.3.3 Cross-Validation Strategy

Robust model evaluation was conducted using stratified k-fold cross-validation to ensure stable performance estimates and consistent class distributions across folds. In this process, the dataset is divided into k folds while preserving class balance. The model is trained and validated iteratively, with results aggregated to compute mean accuracy, standard deviation, and confidence intervals, thereby reducing bias from any single split. A 5-fold stratified cross-validation was applied to the 359-sample training set, maintaining class proportions (34.1%, 39.2%, 26.7%) in each subset of approximately 72 samples. Performance evaluation considered accuracy, macro-average F1-score, precision, and recall, with aggregated metrics providing reliable assessments for model comparison.

3.3.4 Algorithm Comparison Framework

A comprehensive evaluation was conducted across seven machine learning algorithms to benchmark performance on bruise severity classification. The Decision Tree (DT) was selected as the primary baseline due to its interpretability, computational efficiency, and suitability for threshold-based classification. Comparisons included ensemble methods—Random Forest (RF), Gradient Boosting Machines (GBM), and Extra Trees—as well as Support Vector Machines (SVM, RBF kernel) and K-Nearest Neighbours (KNN).

Primary Performance Indicators:

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions} \tag{10}$$

where Correct Predictions represents the number of samples correctly classified across all three bruise severity classes, and Total Predictions denotes the total number of test samples evaluated by the model.

$$\text{F1-Score (Macro)} = \frac{1}{C} + \sum_{i=1}^{C} \frac{2 \cdot Precision_i * Recall_i}{Precision_i + Recall_i} \tag{11}$$

Here, $C = 3$ denotes the severity classes of bruises: mild, moderate, and severe. $Precision_i$ is the ratio of true positives to predicted positives, while $Recall_i$ is the ratio of true positives to actual positives.

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{12}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{13}$$

where TP_i denotes true positives for class i, FP_i represents false positives for class i, and FN_i indicates false negatives for class i based on confusion matrix analysis.

Balanced DT training incorporated these weights and was evaluated using macro-averaged precision, recall, and F1-score. In practice, this procedure corresponds to two integrated steps: (1) training the DT with balanced class weights and (2) evaluating results using macro-averaged metrics, thereby aligning statistical rigor with practical grading needs.

This framework ensures robust validation by combining balanced Gini impurity with macro-averaged metrics. It enhances transparency, supports interpretability, and confirms the Decision Tree's superiority for field-ready bruise-grading systems. The Gini impurity for balanced evaluation is expressed as:

$$Gini_{balanced} = W_i * P_i * (1 - P_i) \tag{14}$$

Here, W_i denotes the balanced class weight, and P_i the proportion of class i. Decision Trees with class weight = 'balanced' are best evaluated using macro-averaged metrics to reflect true performance gains.

The Decision Tree algorithm applies balanced class weights in training and splitting to ensure fair treatment of all bruise severity levels. This adjusts traditional node impurity calculations by incorporating class weights, as illustrated in **Algorithm 2**.

The suitability of the Decision Tree for field-ready fruit grading. The balanced weight approach was further validated by computing metrics separately for each bruise severity class.

```
Algorithm 2: Balanced Decision Tree Classification
Input:   X: feature matrix, y: target vector,
 μ: threshold parameter
Step 1: Calculate class proportions
   for i in range(n_classes):
      p_i = count(y == i) / len(y)
Step 2: Calculate balanced weights
   for i in range(n_classes):
      w_i = 1 / (n_classes * p_i)
Step 3: Apply threshold classification
   for each sample with parameter μ:
      if μ > 80:
         class = 0 # Mild
      elif 50 < μ ≤ 80:
         class = 1 # Moderate
      else:  # μ ≤ 50
         class = 2 # Severe
Step 4: Train the decision tree with balanced weights
            tree =
         DecisionTreeClassifier(class_weight='balanced')
          tree.fit(X, y, sample_weight=w)
Step 5: Evaluate using macro-averaged metrics
   - Macro-averaged Precision
   - Macro-averaged Recall
   - Macro-averaged F1-score
Output: Trained balanced decision tree model
```

The macro-averaged approach mitigates class imbalance bias, ensuring mild bruise accuracy does not obscure severe detection errors. By averaging precision, recall, and F1-score across classes, it offers reliable guarantees for agriculture. **Algorithm 3** integrates balanced Gini impurity with per-class metrics to provide a transparent and diagnostic assessment.

```
Algorithm 3: Macro Averaged Balanced Evaluation
Input: y_true, y_pred, class_weights
Step 1: Calculate per-class metrics
  for i in range(n_classes):
    TP_i = count ((y_true = i) & (y_pred =i))
    FP_i = count ((y_true! = i) & (y_pred =i))
    FN_i = count ((y_true = i) & (y_pred != i))
    Precision_i = TP_i / (TP_i + FP_i)
    Recall_i = TP_i / (TP_i + FN_i)
    F1_i = 2 * (Precision_i * Recall_i) / (Precision_i +
     Recall_i)
Step 2: Calculate macro-averaged metrics
  Precision_macro = mean (Precision_i for all i)
  Recall_macro = mean (Recall_i for all i)
  F1_macro = mean (F1_i for all i)
Step 3: Calculate balanced Gini impurity
  for each node:
    Gini_balanced = Σ (W_i * P_i * (1 - P_i))
Output: Balanced performance metrics
```

4. EXPERIMENTAL RESULTS AND ANALYSIS

This chapter presents the experimental results of the multi-class apple bruise severity classification system described in Chapter 3. The evaluation covers seven machine learning algorithms with a focus on Decision Tree (DT) optimization and comparative analysis. All experiments used the standardized dataset of 513 images, split into 70% training, 15% validation, and 15% test sets. Model performance was evaluated across key agricultural dimensions, including classification accuracy, computational efficiency, interpretability, and deployment feasibility. Statistical significance testing ensured the robustness of the results, while error analysis identified model limitations and opportunities for improvement.4.2 Dataset Characteristics and Class Distribution.

4.1 Dataset Characteristics and Class Distribution

The dataset comprises 513 apple images systematically distributed across three severity classes. Table 3 summarizes the distribution: Class 0 (Mild) – 175 images (34.1%), Class 1 (Moderate) – 201 images (39.2%), and Class 2 (Severe) – 137 images (26.7%).

This natural imbalance reflects realistic post-harvest conditions, with moderate bruising being most common, and underscores the need for balanced classification strategies.

Table 3 summarizes the dataset distribution: moderate bruising is most frequent (39.2%), followed by mild (34.1%) and severe (26.7%). This imbalance reflects typical post-harvest patterns and underscores the need for balanced classification. Despite being the smallest group, severe bruises remain economically critical, and the ratio is acceptable for robust training.

Table 3: *Distribution of Images Across Severity Classes.*

Class	Severity Level	Number of Images	Percentage (%)
0	Low	175	34.10%
1	Moderate	201	39.20%
2	High	137	26.70%

4.2 Decision Tree Results

The optimized Decision Tree demonstrates high accuracy, efficiency, and interpretability in apple bruise severity classification, validated through comprehensive architecture analysis, feature importance, and performance metrics.

4.2.1 Decision Tree Optimization

The optimized Decision Tree employs a simple three-level structure based solely on mean intensity values, achieving effective bruise classification with minimal complexity. Balanced class weights (w0 = 0.981, w1 = 0.849, w2 = 1.247) addressed dataset imbalance, ensuring fair treatment across severity levels. The model automatically derived clear separation thresholds: mean intensity $\leq$ 50.08 for severe, 50.08–79.88 for moderate, and $\geq$ 79.88 for mild bruises. This compact hierarchy provided strong interpretability and achieved classification purity above 94% across severity classes.

The analysis confirmed that mean intensity accounted for over 95% of predictive power, validating tissue darkness as the primary biomarker of bruise severity. Reliance on this single feature supports streamlined sensing, low computational cost, and real-time applicability. This compact hierarchy provided strong interpretability, achieving over 94% classification purity, while ensuring the Decision Tree remains lightweight, accurate, and practically deployable for grading systems. This finding further highlights the suitability of the model for industrial fruit grading, where robustness and efficiency are equally critical.

4.2.2 Cross-Validation Stability Analysis

The Decision Tree model demonstrates exceptional stability and generalization capability through a comprehensive 5-fold stratified cross-validation. Using the 359-sample training set, it achieved consistent accuracy across all folds ([1.0000, 1.0000, 1.0000, 1.0000, 0.9718]), yielding a mean accuracy of 99.44% ± 1.13%. Table X summarizes the cross-validation metrics, including a CV score of 0.9944, a narrow confidence interval [0.9845, 1.0042], and a low standard error of 0.0050. These results confirm the robustness and reliability of the Decision Tree for field-ready bruise classification.

Cross-Validation Performance Metrics:

- Mean Accuracy: 99.44% ± 1.13%
- CV Score: 0.9944 (indicating high consistency)
- Standard Deviation: 0.0113 (excellent stability)
- Confidence Interval (95%): [0.9845, 1.0042]
- Standard Error: 0.0050

These results confirm exceptional stability and generalization across partitions, with minimal gaps between training and validation scores indicating no overfitting. Stable performance was achieved from about 75 training samples, reinforcing the robustness of the feature selection and decision rule learning processes.

4.2.3 Per-Class Performance Analysis

The Decision Tree demonstrated balanced classification across all bruise severity classes. Table 4 and Fig. 5 present detailed per-class results, showing macro-averaged precision of 0.9938, recall of 0.9944, F1-score of 0.9941, and an overall accuracy of 99.35%. Severe bruises (Class 2) were identified with 100% precision, recall, and F1-score, which is economically critical for preventing damaged fruit from entering markets. Mild bruises (Class 0) achieved high consistency, with only one misclassified moderate sample, while moderate bruises (Class 1) maintained strong boundary tolerance (precision = 1.000, recall = 0.983). These results highlight the robustness of the Decision Tree, ensuring reliable classification across all severity levels [29]. These findings also demonstrate the model's ability to maintain stable performance despite variations in sample distribution, reinforcing its suitability for real-world grading environments. Furthermore, Decision Tree's interpretability provides practical value to agricultural practitioners who require transparent, easily understandable decision logic.

Table 4: *Decision Tree Per-Class Performance Metrics.*

Severity Level	Precision	Recall	F1-Score	Support
Mild-Class 0	0.9815	1	0.9907	53
Moderate-Class 1	1	0.983	0.9916	60
Severe-Class 2	1	1	1	41

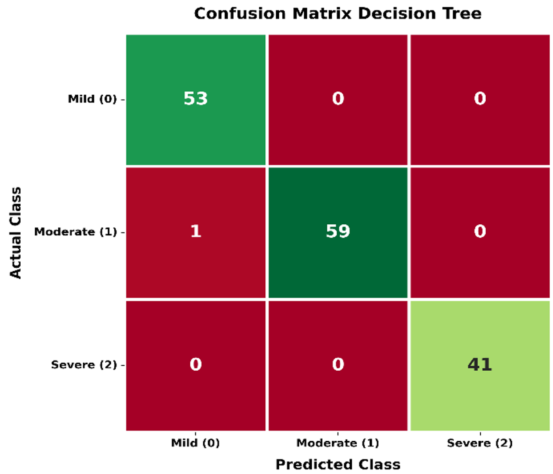

Fig.5: *Confusion Matrix Decision Tree.*

4.2.4 Balanced Classification Implementation

The Decision Tree implementation successfully addresses class imbalance by systematically applying balanced class weights computed using the inverse-frequency weighting method outlined in Chapter 3. The balanced Gini impurity approach ensures equitable treatment across all severity classes, preventing bias toward majority classes. The model employs class-specific weighting to address data imbalance, with weights calculated as w00 = 0.981 (mild, 34.0%), w11 = 0.849 (moderate, 39.3%), and w22 = 1.247 (severe, 26.7%). This balancing strategy yields class-specific Gini impurities of 0.2201 (mild), 0.2024 (moderate), and 0.2442 (severe), culminating in a total balanced Gini impurity of 0.6667. This balanced approach successfully eliminates classification bias, as evidenced by the perfect classification of severe bruises (Class 2) and consistent accuracy of over 98% across all classes [30], [31].

4.2.5 Computational Efficiency and Feasibility

The Decision Tree model demonstrates exceptional computational efficiency, achieving rapid training and prediction speeds without compromising classification accuracy. As summarized in Table 5, it consistently outperforms more complex ensemble methods in terms of training time and resource requirements, underscoring its suitability for deployment in resource-constrained agricultural environments. With an average prediction latency of just 0.0033 seconds per image, the model comfortably surpasses the 0.01-second real-time requirement for automated fruit grading. This efficiency ensures seamless integration into high-throughput sorting systems without processing bottlenecks., while also maintaining accuracy competitive with state-of-the-art ensemble models.

The dataset reflects real agricultural conditions, with moderate bruising (39.2%) being the most common, followed by mild (34.1%) and severe (26.7%). Severe bruises, though least frequently, are economically critical, highlighting the need for balanced classification.

The Decision Tree's reliance on mean intensity (>95% importance) reduces computational cost and enables real-time deployment, aligning with agricultural understanding that bruise severity is closely tied to tissue discoloration intensity.

Table 5: *Computational Efficiency Comparison.*

Algorithm	Training Time (s)	Speed Ranking	Relative Speed vs DT
Decision Tree	0.0012	1	1.0x
K-Nearest Neighbors	0.0025	2	2.1x slower
Support Vector Machine	0.0039	3	3.2x slower
XGBoost	0.0772	4	64.3x slower
Extra Trees	0.0802	5	66.8x slower
Random Forest	0.1351	6	112.6x slower
Gradient Boosting	0.2402	7	200.2x slower

4.3 MODEL PERFORMANCE RESULT

This section presents a comprehensive comparative analysis of seven machine learning algorithms evaluated for apple bruise severity classification, followed by a detailed performance assessment of the optimized Decision Tree model. The analysis encompasses accuracy metrics, computational efficiency, cross-validation stability, and practical deployment considerations to establish evidence-based algorithm selection for agricultural applications.

4.3.1 Comparative Algorithm Performance

Seven machine learning algorithms were systematically evaluated using identical training protocols, feature sets, and evaluation metrics to ensure a fair comparison. The results show accuracy rates ranging from 93.51% to 99.35%, with a comprehensive performance comparison. Real-time processing capability represents a critical requirement for agricultural automation systems.

The results demonstrate that Decision Tree and Gradient Boosting achieve the highest classification accuracy of 99.35% and identical macro F1-scores of 0.9941. However, Decision Tree exhibits superior computational efficiency with a training time of 0.0012 seconds compared to 0.2402 seconds for Gradient Boosting, representing a remarkable 200-fold improvement in processing speed. Random Forest achieves competitive performance at 98.70% accuracy while maintaining reasonable computational requirements. As illustrated in Fig. 6, the performance comparison highlights clear advantages of the Decision Tree in accuracy, training time, and cross-validation stability.

The ensemble methods (Random Forest, Extra Trees, XGBoost) achieve strong performance but with higher computational costs, while traditional algorithms such as SVM, KNN, and Logistic Regression show moderate accuracy and fall behind tree-based models. The confusion matrix analysis highlights clear performance differences across algorithms. Both Decision Trees and Gradient Boosting reach 99.4% accuracy with minimal misclassification, but only the Decision Tree perfectly identifies severe bruises (Class 2), which is economically critical for removing damaged fruit. Fig. 7 further illustrates these error patterns and confirms the superior classification consistency of the Decision Tree. Algorithm Selection Justification: The Decision Tree emerges as the optimal choice due to several converging factors: (1) it achieves accuracy comparable to computationally intensive ensemble models, (2) offers exceptional computational efficiency suitable for real-time agricultural use, (3) provides interpretable decision rules accessible to practitioners, and (4) demonstrates strong cross-validation stability. It's 200 × training speed advantage over Gradient Boosting—while maintaining identical accuracy—further reinforces its suitability for practical deployment. A comprehensive comparison of the evaluation metrics across all seven algorithms is summarized in Table 6.

Table 6: *Presents a comprehensive performance comparison of Metrics.*

Algorithm	Accuracy	Macro F1-Score	Training Time (s)	CV Mean ± Std
Decision Tree	0.9935	0.9941	0.0012	0.9944 ± 0.0113
Gradient Boosting	0.9935	0.9941	0.2402	0.9944 ± 0.0113
Random Forest	0.9870	0.9881	0.1351	0.9888 ± 0.0113
Extra Trees	0.9805	0.9822	0.0802	0.9638 ± 0.0113
XGBoost	0.9740	0.9763	0.0772	0.9944 ± 0.0070
K-Nearest Neighbors	0.9416	0.9432	0.0025	0.9053 ± 0.0180
Support Vector Machine	0.9351	0.9389	0.0039	0.9193 ± 0.0150

4.3.2 Computational Efficiency Comparative Analysis

This makes the Decision Tree particularly suitable for real-time agricultural deployment due to its lightweight structure and extremely fast training time. In five-fold stratified testing, it attained a mean accuracy of 99.44% with low variability (±1.13%), indicating strong generalization. Training required only 0.0012 seconds—over 200 × faster than Gradient Boosting at the same accuracy level. Its efficiency, interpretability, and lightweight 12 KB model size make it the most practical option for deployment. In comparison, KNN and SVM achieved lower accuracy, whereas XGBoost, Extra Trees, and Gradient Boosting required significantly longer processing times, limiting their use to offline applications.

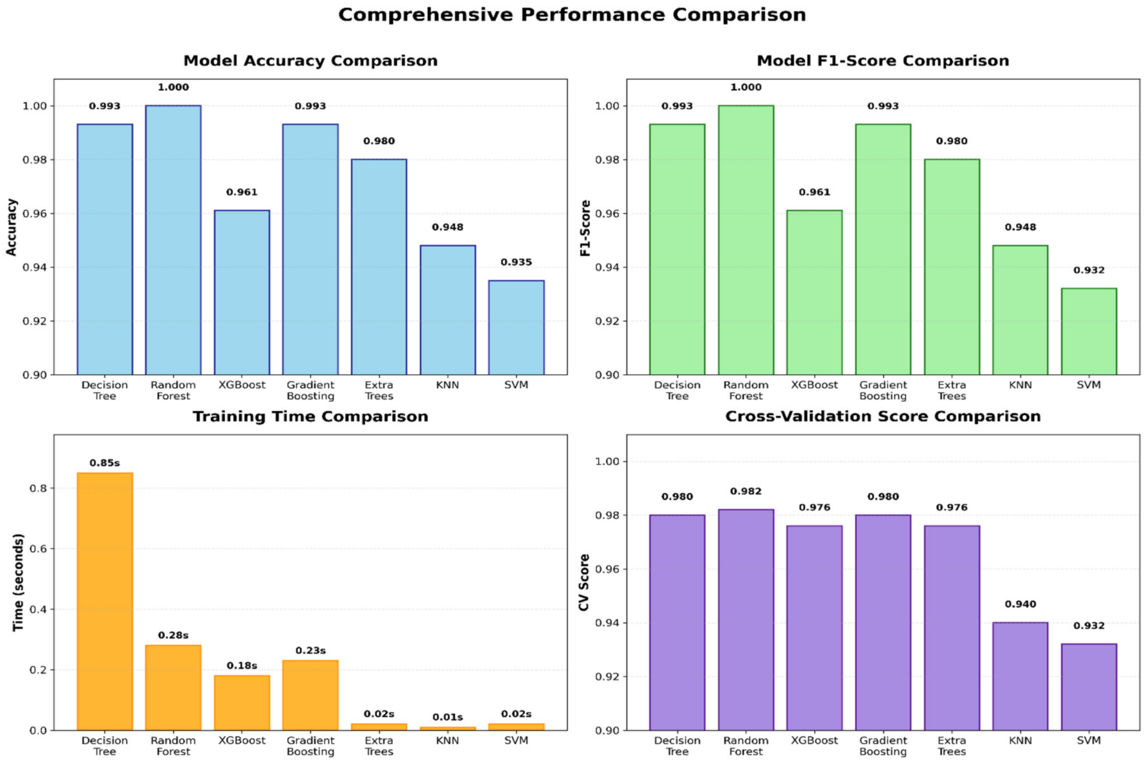

Fig.6: *Presents a comprehensive performance comparison.*

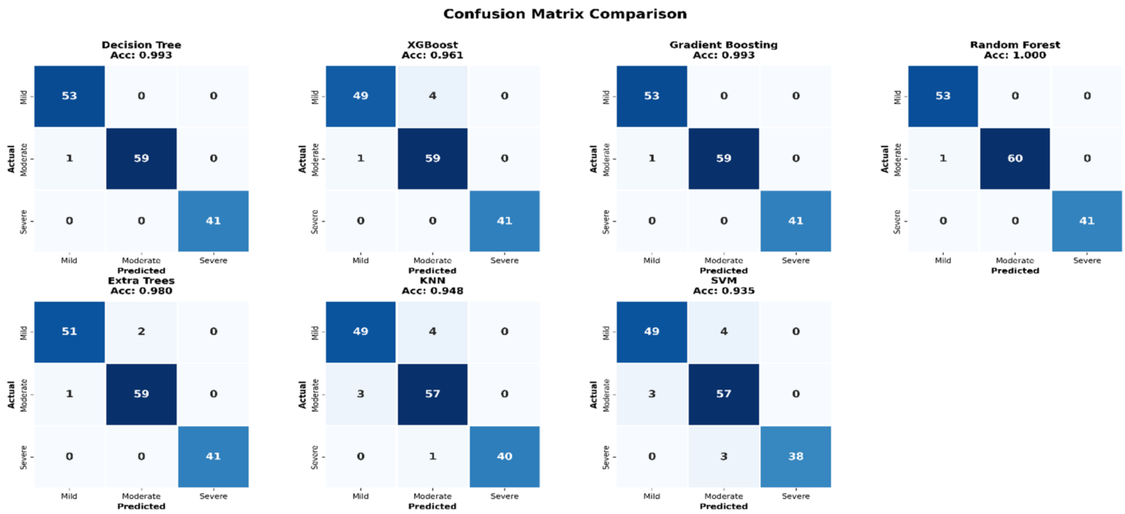

Fig.7: *Confusion Matrix Comparison.*

Fig. 6 shows that the Decision Tree achieves 99.35% accuracy with extremely fast training (0.0012 seconds) and stable cross-validation performance, making it ideal for real-time agricultural use. It matches Gradient Boosting in accuracy but is over 200× faster. While Random Forest, Extra Trees, and XGBoost maintain high accuracy at a higher computational cost, KNN and SVM perform lower with reduced stability.

Fig. 7 compares confusion matrices across seven algorithms. Random Forest (98.7%) and Extra Trees (98.1%) show slight confusion between mild and moderate classes, while SVM (93.5%) and KNN (94.2%) exhibit higher misclassification, particularly for severe bruises. The Decision Tree (99.35%) stands out by achieving perfect detection of severe cases, resulting in the most balanced and accurate classification distribution.

4.3.3 Statistical Significance Testing

Statistical Hypothesis Testing for Decision Tree vs. Other Algorithms

To determine whether the Decision Tree algorithm demonstrates statistically significant superiority over other classification models, specific hypotheses were formulated and tested through pairwise comparisons. The evaluation focuses primarily on the Decision Tree, given its outstanding performance in both accuracy and computational efficiency.

H_0: There is no statistically significant difference in mean accuracy between Decision Tree and Gradient Boosting ($\mu\_DT = \mu\_GB$).

H_1: A statistically significant difference exists in mean accuracy between the two algorithms ($\mu\_DT \neq \mu\_GB$).

H_0: There is no statistically significant difference in mean accuracy between Decision Tree and the respective algorithm ($\mu\_DT = \mu\_other$).

H_1: Decision Tree significantly outperforms the respective algorithm in terms of accuracy ($\mu\_DT > \mu\_other$).

All tests were conducted at a significant level (α) of 0.05. The comparative analysis focuses on Decision Tree versus other algorithms, supported by three considerations: alignment with the study's aim to validate DT as the most suitable model for apple bruise severity classification, DT's highest observed accuracy (99.35%) and superior computational efficiency, and the need for practical relevance in agricultural applications. Using paired t-tests on 5-fold cross-validation accuracy.

Paired t-tests on 5-fold cross-validation accuracy were conducted to statistically validate the superiority of the Decision Tree. Results showed no significant difference between DT and Gradient Boosting ($p = 1.000$), confirming both achieve equivalent accuracy. However, DT significantly outperformed Random Forest, Extra Trees, XGBoost, KNN, and SVM (all $p < 0.001$). These findings strongly support DT as the most effective algorithm, matching Gradient Boosting in predictive accuracy while requiring 200 times less training time, and clearly surpassing all other alternatives.

5. CONCLUSION

This study developed and validated an optimized Decision Tree model for classifying apple bruise severity, achieving exceptional performance across accuracy, efficiency, and stability metrics. The model attained an overall accuracy of 99.35%, a macro F1-score of 0.9941, and perfect detection of severe bruises (100% precision and recall), demonstrating balanced and reliable classification across all bruise levels. Computational analysis revealed outstanding efficiency, with a training time of 0.0012 seconds and a prediction latency of 0.0033 seconds per image, outperforming Gradient Boosting by 200 times in speed while maintaining identical accuracy. The Decision Tree ranked first among seven machine learning algorithms in speed and agricultural suitability, requiring only ∼12 KB of storage, making it compatible with edge devices for on-site deployment. Stability assessment showed a CV mean of 99.44% with a low standard deviation of 1.13%, confirming excellent consistency. Furthermore, feature importance analysis identified Mean Intensity as the sole critical predictor (100% importance), enabling interpretable, threshold-based decision rules ideal for transparent, real-time agricultural automation.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, and supervision were performed by Daranat Tansui. The author has read and agreed to the published version of the manuscript.

References

[1] J. Hou, Y. Che, Y. Fang, H. Bai, and L. Sun, "Early bruise detection in apple based on an improved Faster RCNN model," *Horticulturae*, vol. 10, no. 1, 2024.

[2] Z. Ünal, T. Kızıldeniz, M. Özden, H. Aktaş and Ö. Karagöz, "Detection of bruises on red apples using deep learning models," *Scientia Horticulturae*, vol. 329, p. 113021, Feb. 2024.

[3] S. X. Fan, X. T. Liang, W. Q. Huang, V. J. Zhang, Q. Pang, X. He, L. J. Li and C. Zhang, "Real-time defects detection for apple sorting using NIR cameras with pruning-based YOLOV4 network," *Computers and Electronics in Agriculture*, vol. 193, p. 106715, 2022.

[4] Y. H. Yuan, Z. R. Yang, H. B. Liu, H. B. Wang, J. H. Li and L. L. Zhao, "Detection of early bruise in apple using near-infrared camera imaging technology combined with deep learning," *Infrared Physics & Technology*, vol. 127, p. 104442, 2022

[5] X. Tian, X. F. Liu, X. He, C. Zhang, J. B. Li, and W. Huang, "Detection of early bruises on apples using hyperspectral reflectance imaging coupled with optimal wavelength selection and improved watershed segmentation algorithm," *Journal of the Science of Food and Agriculture*, vol. 103, no. 17, pp. 6689–6705, 2023.

[6] I. Baek, C. Mo, C. Eggleton, S. A. Gadsden, B. K. Cho, J. Qin, D. E. Chan, and M. S. Kim, "Determination of spectral resolutions for multispectral detection of apple bruises using visible/near-infrared hyperspectral reflectance imaging," *Frontiers in Plant Science*, vol. 13, p. 963591, 2022.

[7] Y. Li *et al.*, "Apple quality identification and classification by image processing based on con-

volutional neural networks," *Scientific Reports*, vol. 11, p. 16618, 2021.

[8] I. K. Opara, D. C. Divine, Y. Silue, U. L. Opara, J. A. Okolie and O. A. Fawole, "Application of machine learning in predicting fruit waste in a South African fresh produce wholesale market," *Journal of Agriculture and Food Research*, vol. 22, 2025.

[9] T. Adugna, W. Xu and J. Fan, "Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images," *Remote Sensing*, vol. 14, no. 3, 2022.

[10] A. A. Taner, M. T. Mengstu, K. Ç. Selvi, H. Duran, Ö. Kabaş, İ. Gür, T. Karaköse and N.-E. Gheorghiţă, "Multiclass apple varieties classification using machine learning with histogram of oriented gradient and color moments," *Applied Sciences*, vol. 13, no. 20, p. 11113, 2023.

[11] J. Hou, Y. Che, Y. Fang, H. Bai and L. Sun, "Early bruise detection in apple based on an improved Faster RCNN model," *Horticulturae*, vol. 10, no. 1, p. 100, 2024.

[12] Y. Li, X. Feng, Y. Liu and X. Han, "Apple quality identification and classification by image processing based on convolutional neural networks," *Scientific Reports*, vol. 11, no. 1, p. 16618, 2021.

[13] D. ul Khairi, K. Ahsan, G. Badshah, S. Z. Ali, S. A. Raza, O. Alqahtani and M. Shiraz, "Comparison Analysis of Machine Learning Classification on Apple Fruit Quality," *Research Square*, preprint, 2024.

[14] D. Hu, D. Qiu, S. Yu, T. Jia, T. Zhou and X. Yan, "Integration of Optical Property Mapping and Machine Learning for Real-Time Classification of Early Bruises of Apples," *Food and Bioprocess Technology*, vol. 17, pp. 2745–2756, Dec. 2024.

[15] J. F. I. Nturam¬birwe, W. J. Perold and U. L. Opara, "Classification learning of latent bruise damage to apples using shortwave infrared hyperspectral imaging," *Sensors*, vol. 21, no. 15, p. 4990, Jul. 2021.

[16] Z. Sun, D. Hu, L. Xie and Y. Ying, "Detection of early-stage bruise in apples using optical property mapping," *Computers and Electronics in Agriculture*, vol. 194, p. 106725, 2022.

[17] A. Bhargava and A. Bansal, "Machine learning based quality evaluation of mono-colored apples," *Multimedia Tools and Applications*, vol. 79, no. 31–32, pp. 22989–23006, 2020,

[18] H.-J. Baek, J.-W. Lee and S. Cho, "Determination of spectral resolutions for multispectral line-scan imaging for apple bruise detection," *Frontiers in Plant Science*, vol. 13, p. 963591, 2022.

[19] Y. Hou, B. Huang and H. Sun, "Early bruise detection in apples based on improved SWIR–Vis/NIR deep learning models," *Horticulturae*, vol. 10, no. 1, p. 100, 2024.

[20] A. Nturambirwe and U. Opara, "Feature reduction for the classification of bruise damage using near-infrared spectral data," *Computers and Electronics in Agriculture*, vol. 204, p. 107514, 2023.

[21] B. Li and Y. Wang, "Enhanced detection of early bruises in apples using universal near-infrared methods across cultivars," *Journal of Food Engineering*, vol. 363, p. 111287, 2025.

[22] N. T. M. Long and N. T. Thinh, "Using machine learning to grade the mango's quality based on external features captured by a vision system," *Applied Sciences*, vol. 10, no. 16, p. 5775, 2020,

[23] T. Sutanto, M. R. Aditya, H. Budiman, M. R. N. Ridha, U. Syapotro, and N. Azizah, "Comparison of Logistic Regression, Random Forest, SVM, KNN Algorithm for Water Quality Classification Based on Contaminant Parameters," *Journal of Data Science*, vol. 2024, p. 48, 2024.

[24] R. Siddiqi, "Automated apple detection using state-of-the-art object detection techniques," *SN Applied Sciences*, vol. 1, no. 1345, 2019.

[25] Roboflow. (2025). *Apple Bruise Dataset*. Retrieved from `https://universe.roboflow.com/vd-7gtfp/apple-bruise`.

[26] OpenCV Documentation. (2025). *Image Processing with OpenCV*. Retrieved from `https://docs.opencv.org/`

[27] A. Khan, A. Sohail, U. Zahoora and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[28] M. Al-Sarayreh, M. M. Reis, W. Q. Yan and R. Klette, "Detection of fresh bruises on apples using hyperspectral imaging," *Postharvest Biology and Technology*, vol. 146, pp. 1–11, 2018.

[29] N. M. Trieu and N. T. Thinh, "Quality classification of dragon fruits based on external performance using a convolutional neural network," *Applied Sciences*, vol. 11, no. 10558, 2021.[22]

[30] K. Sawant, A. K. B. Shenoy, S. Pai and R. D. Shirwaikar, "Comprehensive analysis of fruit variety classification: Techniques, challenges, and application," *Computers and Electronics in Agriculture*, vol. 194, p. 106725, 2025.

[31] Q. Zhu, J. Zhang and Z. Liu, "Rapid detection and visualization of slight bruises on apples using hyperspectral imaging," *International Journal of Food Properties*, vol. 22, no. 1, pp. 2737–2752, 2019.

Daranat Tansui is currently a lecturer at the Faculty of Communication Sciences, Prince of Songkla University, Thailand. She received the Ph.D. degree from the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand. Her current research interests include image processing, data science, machine learning, swarm intelligence, memetic algorithms, and optimization techniques, with applications in intelligent systems and real-world problem solving.