Journal of Applied Data Sciences
Vol. 5, No. 4, December 2024, pp. 1989-2001

Efficient Fruit Grading and Selection System Leveraging Computer Vision and Machine Learning

Deshinta Arrova Dewi[1,*], Tri Basuki Kurniawan[2], Rajermani Thinakaran[3], Malathy Batumalay[4], Shabana Habib[5], Muhammad Islam[6]

[1,3,4]*Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia*

[2]*Faculty of Science and Technology, Universitas Bina Darma, Palembang, Indonesia*

[5]*Faculty of Computer, Qassim University, Saudi Arabia*

[6]*Onizah College, Saudi Arabia*

(Received: July 18, 2024; Revised: September 21, 2024; Accepted: October 24, 2024; Available online: November 7, 2024)

Abstract

Automated fruit grading is crucial to overcoming the time and accuracy challenges posed by manual methods, which are often limited by subjective human judgment. This study introduces an intelligent grading system leveraging computer vision and AI to improve speed and consistency in assessing fruit quality. Using high-resolution imaging and advanced feature extraction, including grayscale processing, binarization, and enhancement, the system achieves non-destructive, efficient sorting for fruits like apples, bananas, and oranges. Grayscale processing reduces image complexity while preserving essential details, binarization isolates the fruit from its background, and enhancement highlights critical features. Notably, the Edge Pixel method proved most effective, achieving 79.20% accuracy in grading, while the Grayscale Pixel method reached 93.94% accuracy for fruit types. Edge Pixel also achieved 80.32% in differentiating grading types, showcasing its ability to capture essential shapes and edges. Fruits are classified into four grades: Grade_01 (highest quality), Grade_02 (minor imperfections), Grade_03 (notable defects but consumable), and Grade_04 (unfit for consumption). A specialized dataset supports model training, ensuring practical real-world application. The study concludes that this automated system offers significant improvements over traditional grading, providing a scalable, objective, and reliable solution for the agricultural sector, ultimately enhancing productivity and quality assurance.

Keywords: Automatic Fresh Fruit Selection, Grading, CNN, Deep Learning, Process Innovation, Product Innovation

1. Introduction

Automated fruit selection and grading systems are essential for efficient fruit commercialization, influencing packaging, storage, and sales revenue. Traditional methods, involving manual weighing and grading, are often inefficient and error-prone, leading to fruit loss. Manual and mechanical devices struggle to evaluate fruit color and surface defects adequately, introducing further limitations, such as tedious processes, timing delays, and high processing costs, which increase production and distribution costs in the fruit industry [1].

Computer vision algorithms mimic human vision by electronically interpreting images, aiming to automate fruit grading with minimal human intervention. These systems utilize RGB images to capture crucial exterior quality features, enabling automatic classification into predefined quality categories defined by users [1]. Computer vision is widely used in food and agriculture for rapid, economical, consistent, and objective quality assessment, and has shown success in objectively measuring various agricultural products. Recent advances in digital imaging technology have spurred numerous studies to develop systems that evaluate the quality of diverse food products [2], [3].

Such systems can segment fruit images to classify them by maturity level and size, distinguishing between ripe and unripe fruit based on quality parameters [4]. A prototype computer vision system automatically grades fruits into

***Corresponding author: Deshinta Arrova Dewi (deshinta.ad@newinti.edu.my)**
DOI: https://doi.org/10.47738/jads.v5i4.443

categories (grades 1, 2, and 3) based on expert-defined standards [5], [6]. Traditional quality detection relies on subjective manual inspection, impacted by lighting, grader skill, and stress levels. Consequently, a robust system to assess fruit quality during handling, processing, and packing is necessary, and computer vision systems are increasingly effective for evaluating surface and internal attributes in food products [7], [8], [9].

In the past decade, advances in digital image processing have spurred the use of computer vision for agricultural applications. Studies such as [10] on hyperspectral imaging to detect surface defects in apples and [11] on RGB imaging with machine learning for classifying tomato ripeness show high accuracy. Agarla [12] demonstrated an automated defect detection system for golden apples, achieving high accuracy, while Hameed [13] used deep learning to classify citrus fruits based on surface texture, outperforming traditional methods. Internal quality assessments have also advanced, with techniques like NIR spectroscopy and MRI for measuring firmness and sugar content, as reviewed by [14], and X-ray imaging for detecting internal browning in apples, as shown by Huang & Liang [15].

Integrating these methodologies has been a focus, as seen in Akter et al. [16], who developed a real-time inspection and grading system, successfully implemented in industrial settings by combining imaging and machine learning for comprehensive quality assessment. Although challenges remain, such as lighting variability, dataset requirements, and computational complexity, future research aims to enhance robustness, reduce costs, and improve system speed. Integrating IoT and edge computing with computer vision systems is also expected to further revolutionize the agricultural sector. Smart fruit selection and grading systems based on computer vision represent a transformative advancement in agriculture, enhancing efficiency, objectivity, and speed. Ongoing research promises even greater improvements, driving agricultural innovation [17].

2. Methodology

A Smart Fruit Selection and Grading System follows a series of steps to ensure accurate and efficient fruit quality assessment. It begins with image acquisition, where high-resolution images of the fruits are captured. These images are then pre-processed to enhance their quality by reducing noise and adjusting contrast. In the feature extraction phase, key attributes such as color, texture, size, and shape are analyzed [18]. Using these features, machine learning algorithms classify the fruits into different quality grades. Finally, the system undergoes evaluation and validation to ensure the accuracy and reliability of the grading process. Together, these steps provide a robust method for precise and consistent fruit grading that benefits farmers and consumers. Some steps should be followed in Smart fruit selection and grading systems are, such as:

2.1. Image Acquisition

High-quality images are captured using cameras under controlled lighting conditions. This step is crucial as it ensures that the images are uniform and free from shadows or reflections, which could interfere with the accuracy of subsequent analyses. Various types of cameras, such as RGB, hyperspectral, and infrared, can be employed depending on the assessment's specific requirements.

2.2. Image PreProcessing

Preprocessing techniques enhance the quality of the captured images. These may include noise reduction, contrast enhancement, and color correction. These steps are essential to ensure the images are suitable for further analysis and feature extraction.

2.3. Feature Extraction

The next step involves extracting relevant features from the preprocessed images. For surface quality assessment, features such as color, texture, shape, and size are identified using advanced techniques like histogram analysis, edge detection, and morphological operations. For internal quality assessment, non-destructive imaging techniques like X-ray or MRI are used to extract features related to the internal structure of the fruits.

2.4. Classification

Extracted features classify fruits into different quality categories. Machine learning algorithms such as support vector machines (SVM), neural networks, and deep learning models like convolutional neural networks (CNN) are commonly employed. These models are trained on labeled datasets to learn patterns associated with various quality levels.

2.5. Evaluation and Validation

The computer vision system's performance is evaluated using accuracy, precision, recall, and F1-score metrics. Validation is performed on a separate dataset not used during training to ensure robustness and generalizability. The process of grading fruits involves several crucial steps that leverage advanced technologies such as computer vision and machine learning algorithms. Firstly, the selection phase is initiated, where fruits are scanned using high-resolution cameras to capture their images. These images are then processed to extract various features such as color, texture, shape, and size. Computer vision algorithms allow for precise and efficient analysis of these attributes, ensuring that only the fruits meeting the predefined quality standards proceed to the next stage [19].

Following the selection, the grading process commences. In this step, machine learning algorithms are employed to classify the fruits into different grades based on the extracted features. These algorithms are trained on large datasets of fruit images, enabling them to predict each fruit's quality accurately. The grading criteria can include factors like ripeness, presence of defects, and overall appearance. Machine learning not only enhances the accuracy of the grading process but also significantly reduces the time required compared to manual inspection. After the fruits have been graded, they are sorted accordingly. This sorting mechanism ensures that fruits of similar grades are grouped, facilitating efficient packaging and distribution. Integrating computer vision and machine learning in the selection and grading process ensures high consistency and quality control, ultimately leading to better consumer products and optimized production operations.

2.6. Selection and Grading System

Grading system design defines the architecture, components, modules, interface, and data for system development. Machine learning algorithms are used to train the system with an image dataset, and the system will come up with model deployment to be used in the testing phase [20]. As shown in figure 1, accurate, fresh fruit data are used in the training and testing phases. Figure 1 shows the process of the intelligent fruit selection and grading system. The training phase is conducted using a set of image datasets we will build ourselves. After reading the training image, the pre-processing stage will be carried out, and accordingly, the model deployment will be launched for testing using machine learning algorithms.

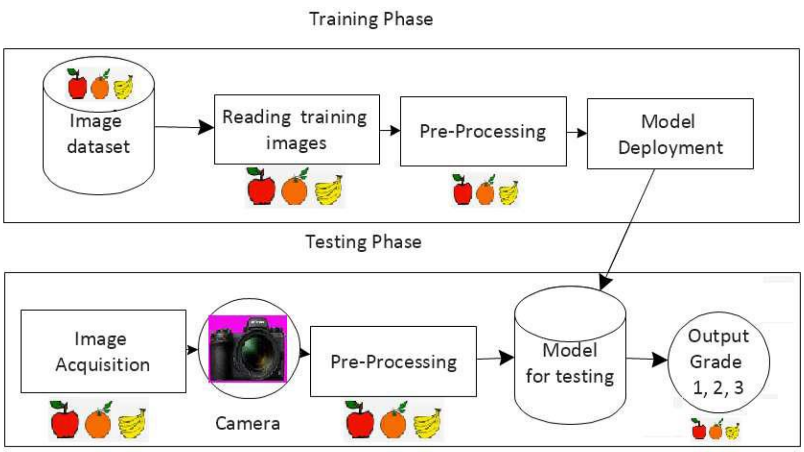


Figure 1. A proposed overview of smart fruit selection and grading system using a machine learning approach

During the testing phase, the camera detects and captures images different from the training phase and then classifies them using a computer vision algorithm. The fruits received are graded via a computer vision using an estimated size and shape analysis algorithm into Grades 1, Grade 2, Grade 3, and Grade 4, respectively. The captured image is processed for classification, as shown in figure 2.

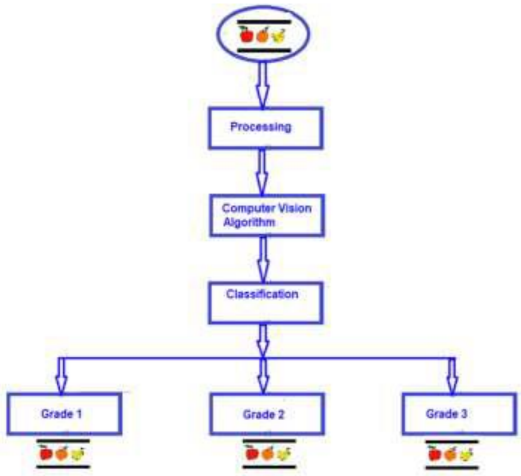


Figure 2. Proposed prototype system flowchart

The image is converted into a gray image where the filter is applied to extract the parameter most closely related to the fruit's information. Then, the number of pixels ranging from zero to 255 in the converted gray image is counted to calculate the image threshold. Then, the proposed system sends the results of fruit grading to the sorter, where the fruits are sorted based on the type and grade of the fruit. A dedicated camera with a lighting system to avoid shadow is set up for this purpose, as illustrated in figure 3.

Figure 3. The camera is statically mounted for fruit grading and sorting

2.7. Sorting System

Each fruit type is graded based on the grade-setting standard in the sorting system, and only it goes for sorting. Figure 4 shows the proposed fruit sorting process after grading.

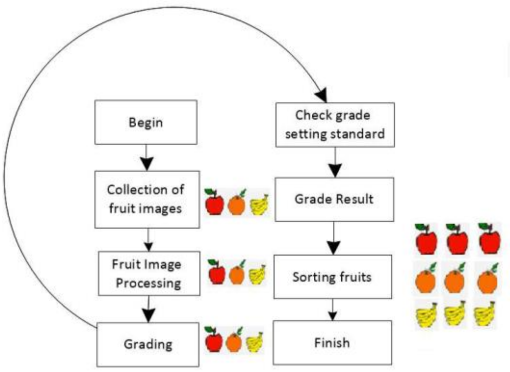


Figure 4. Proposed fruit sorting process after grading

3. Results and Discussion

Automated grading and sorting systems are transformative for the agricultural industry, particularly for fruit farmers seeking to enhance produce quality and consistency. Designed to handle a range of fruit types, these systems offer accessibility and benefits to farmers regardless of the specific crops they grow. By integrating this technology, farmers

ISSN 2723-6471

can ensure precise grading based on critical quality parameters such as size, color, texture, and ripeness. The central server manages the entire process, including a graphical user interface (GUI) parser for seamless interaction and a robust database for storing vast data on fruit quality and grading outcomes. This data repository supports ongoing analysis and optimization, allowing continuous monitoring and adjustments that improve grading efficiency and reliability. Beyond quality control, automated grading systems provide valuable managerial insights that help farmers and agricultural managers make informed decisions regarding harvesting schedules, storage conditions, and market readiness. These systems allow even small-scale farmers to access advanced technology, democratizing tools that enhance competitiveness in the market. By implementing automated grading and sorting systems, farmers can achieve better quality control, higher productivity, and improved profitability. The widespread adoption of this technology in agriculture represents a significant advancement, fostering greater efficiency and market value in fruit production. The flow of the classifier process is shown in figure 5.

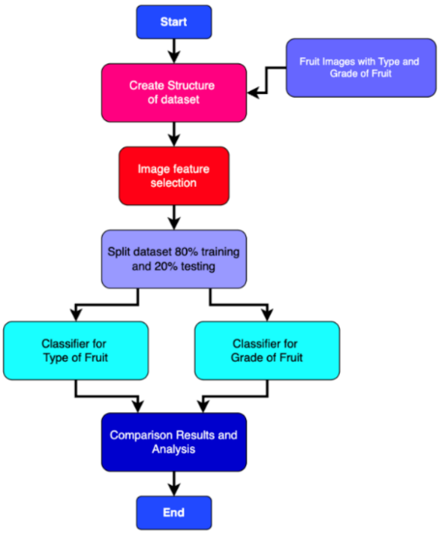


Figure 5. The flow of the classifier processing

The dataset used in this research was collected from Kaggle at https://www.kaggle.com/datasets/ryandpark/fruit-quality-classification/data. There are six classes of fruits such as apple, banana, guava, lime, orange, and pomegranate. Then, we modified the quality of fruit classes from Bad Fruit and Good Fruit into four quality gradings of fruit, such as grade_01, grade_02, grade_03, and grade_04, as shown in figure 6.

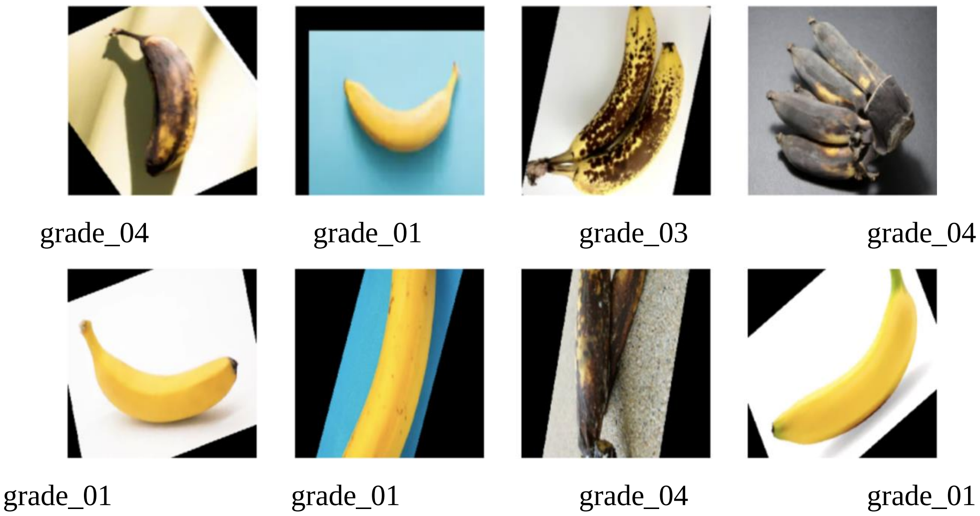


Figure 6. Sample of Banana Fruit with different quality grade fruits

We start to construct our data structure. First, we read the information about images and correspondent information for the images, like the label for the type of fruit and their quality (grade), type of fruit only, and grade of fruit only, as shown in figure 7.

	path	filename	label	fruit	grade
0	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728190031.jpg	Orange_04	Orange	Grade_04
1	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728185954.jpg	Orange_04	Orange	Grade_04
2	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728190012.jpg	Orange_04	Orange	Grade_04
3	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728190034.jpg	Orange_04	Orange	Grade_04
4	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728185953.jpg	Orange_04	Orange	Grade_04
...	...	...	...	...	...
8022	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728182316_27621.jpg	Pomegranate_02	Pomegranate	Grade_02
8023	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728182316_01_27627.jpg	Pomegranate_02	Pomegranate	Grade_02
8024	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728182316_27623.jpg	Pomegranate_02	Pomegranate	Grade_02
8025	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728182316_27620.jpg	Pomegranate_02	Pomegranate	Grade_02
8026	/content/gdrive/My Drive/FruitAndGradeClassifi...	IMG20200728182316_27622.jpg	Pomegranate_02	Pomegranate	Grade_02

8027 rows × 5 columns

Figure 7. Dataset with path, filename, label, fruit, and grade classes

Figure 7 shows 8027 images, which consist of 24 types of fruit and their quality (grade), six (6) types of fruits, and four (4) types of grades. Furthermore, each class has a composition of item values, as shown in figure 8.

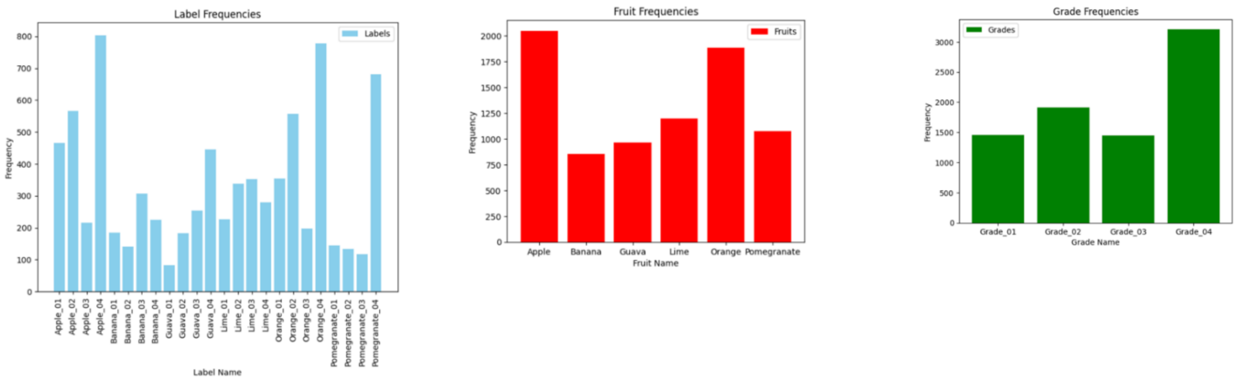


Figure 8. The balanced /unbalanced dataset for each class

Figure 8 depicts that figure discusses the impact of data imbalance across specific fruit and grading types, fruit types, and grading types on CNN model performance. The first chart shows that some labels, like Apple_02 and Orange_04, are overrepresented, leading to biased learning where the model performs well in frequent classes but struggles with less common ones like Banana_03 or Pomegranate_01. The second chart highlights that Apples and Oranges dominate, while Bananas and Pomegranates are underrepresented, further skewing model accuracy towards common classes. The third chart reveals an imbalance in grade distribution, with Grade_04 being significantly more frequent than other grades, which can hinder the model's ability to distinguish lower grades.

The observed imbalances across fruit and grading types, fruits, and grades suggest that the CNN model will likely be biased and underperform in certain classes unless these distributional issues are addressed. Implementing strategies such as data augmentation, rebalancing techniques, and careful selection of evaluation metrics that focus on individual class performance (like per-class precision, recall, and F1-score) is crucial. Additionally, continuously monitoring model predictions and retraining with balanced data will enhance the model's robustness and generalizability across diverse scenarios.

The feature extraction process is the most essential part of the classification process. Especially for data sources in the form of images, it is necessary to carry out a careful feature extraction process using computer vision techniques. This is because if the feature extraction process does not obtain the correct information for the classification training process, the performance or accuracy of the classification algorithm will remain the same. To achieve that goal, four different image feature extraction methods are applied to extract information from the images, namely color, grayscale, mean pixel, and edge pixel, as shown in figure 9, figure 10, figure 11, figure 12.

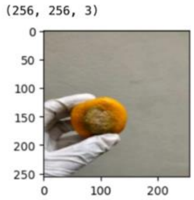


Figure 9. The result of images when extracted with color (3 channels)

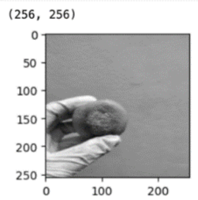


Figure 10. The result of images when extracted with grayscale (1 channel)

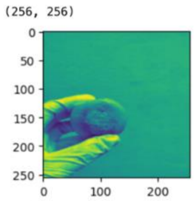


Figure 11. The result of images when extracted with mean pixel (1 channel)

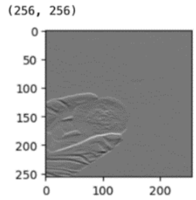


Figure 12. The result of images when extracted with edge pixel (1 channel)

In this research, we use the Convolution Neural Network algorithm to classify our data training and testing, which has 30% composition for testing and 70% for training. Figure 13 shows our code in Python programming.

```
def create_custom_cnn(input_shape, num_classes):
    model = Sequential([
        Conv2D(32, (3, 3), activation='relu', input_shape=input_shape),
        MaxPooling2D(2, 2),
        Conv2D(64, (3, 3), activation='relu'),
        MaxPooling2D(2, 2),
        Conv2D(128, (3, 3), activation='relu'),
        MaxPooling2D(2, 2),
        Flatten(),
        Dropout(0.5),
        Dense(512, activation='relu'),
        Dropout(0.2),
        Dense(num_classes, activation='softmax')
    ])
    return model

# Define the model parameters
input_shape = (256, 256, 1)  # Adjust based on your dataset preprocessing
num_classes = len_class  # Adjust based on the number of fruit classes in your dataset

# Create the CNN model
model = create_custom_cnn(input_shape, num_classes)
```

Figure 13. CNN Algorithm code in Python programming

Figure 13 illustrates the initialization of a custom Convolutional Neural Network (CNN) model designed for image classification tasks, such as fruit classification. This sequential model comprises three convolutional layers, each followed by a max-pooling layer to reduce spatial dimensions while increasing the number of filters (32, 64, 128) at each level, enabling the network to learn progressively complex features. After the convolutional layers, a Flatten layer reshapes the output into a 1D array for fully connected layers. Dropout layers are included to reduce overfitting by randomly deactivating neurons during training. The final dense layer employs a SoftMax activation function to output class probabilities, making it suitable for multi-class classification tasks. Key parameters include `input_shape = (256, 256, 1)` for grayscale images (or adjust to `(256, 256, 3)` for RGB images) and `num_classes = len_class`, which should match the number of classes in the dataset, such as six for fruit types (e.g., Apple, Banana, Guava, Lime, Orange, Pomegranate).

To optimize the model, ensure images are preprocessed by resizing to 256x256 pixels and normalizing pixel values between 0 and 1, which aids in training stability. Regularly monitor metrics such as loss and accuracy on both training and validation sets to confirm effective learning and avoid overfitting. Adjustments like tuning learning rates, dropout rates, and possibly adding data augmentation can further enhance performance. The simplicity of this CNN model provides a solid baseline for fruit classification, with potential refinements based on specific task outcomes.

According to the research framework in figure 5, four different feature extraction techniques will be applied, with each technique forming a model for specific classification tasks: fruit type and grading level (labels), fruit type, and grading type. Consequently, 12 accuracy metrics will be collected across these models, as displayed in table 1, table 2, and

 ISSN 2723-6471

table 3, which report accuracy, training and validation accuracy, and confusion matrix results, including precision, recall, and F1-score, respectively.

Table 1 highlights the accuracy of different classifiers using various image feature extraction techniques (RGB Color, Grayscale, Mean Pixel, and Edge Pixel). The results show that the Edge Pixel method consistently delivers the highest accuracy, particularly in distinguishing fruit types and grades. This method's ability to capture essential shape and edge details is crucial for differentiating subtle variations, which is less effectively achieved by simpler methods like Mean Pixel, which underperformed across all classification tasks. The Grayscale method also showed strong results, especially for fruit classification, indicating that texture and shape information are more critical than color for this task.

Table 1. Comparison accuracy results for each model

Classifier/Image feature extraction	**Accuracy (%)**			
	RGB Color	**Grayscale**	**Mean Pixel**	**Edge Pixel**
Type of Fruits and Grading (Labels)	75.92	77.92	71.44	79.20
Type of Fruits	93.86	93.94	88.54	92.36
Type of Grades	74.39	75.76	52.43	80.32

Table 2 illustrates the convergence of training and validation accuracy over epochs for each method. Grayscale and Edge Pixel methods display stable and consistent improvement, indicating that these techniques effectively learn distinguishing features without significant overfitting. In contrast, the Mean Pixel approach shows fluctuations and lower overall accuracy, underscoring its inability to capture sufficient detail for complex classifications. This performance disparity emphasizes the need for advanced feature extraction techniques to achieve higher model accuracy.

Table 2. Training and Validation Accuracy of the convergence curve

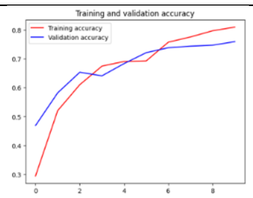
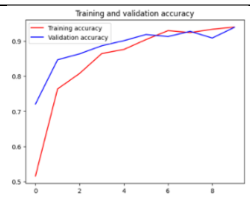
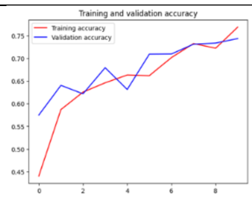
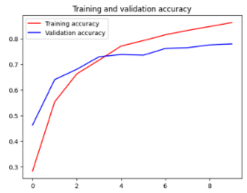
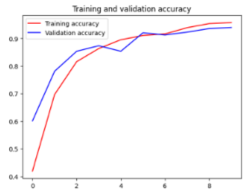
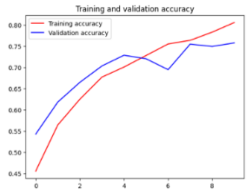
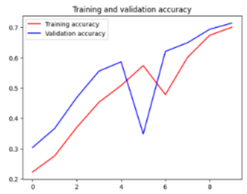
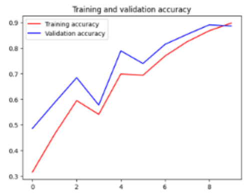
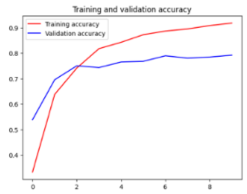
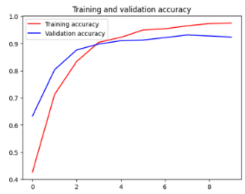
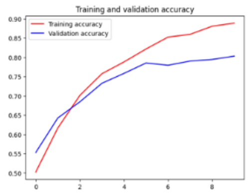
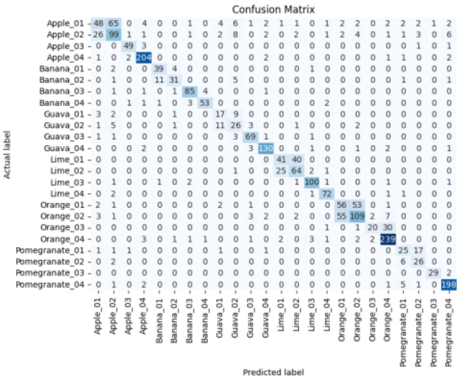
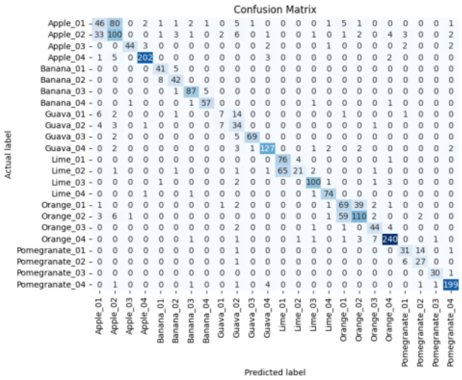
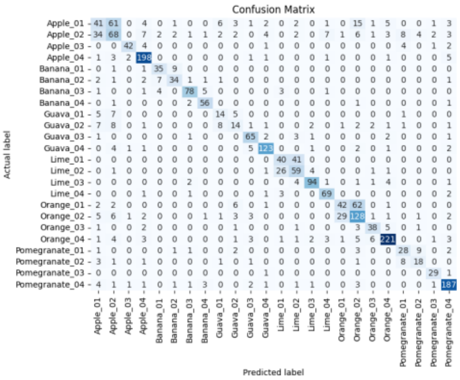
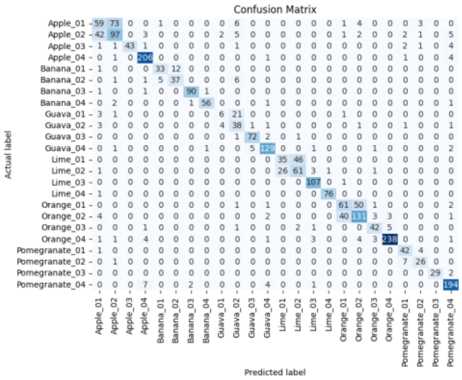
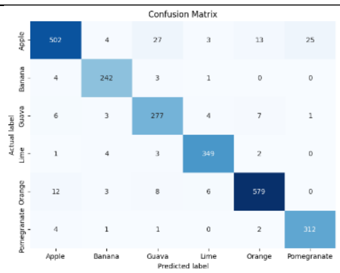
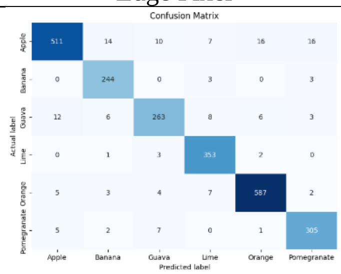
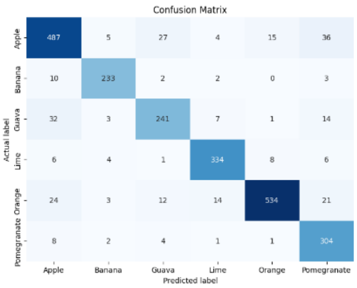
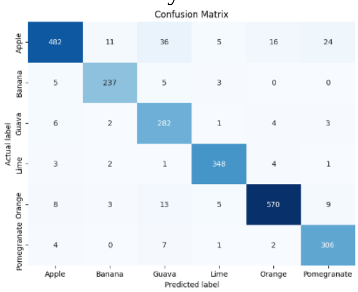
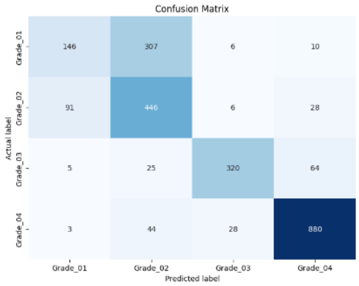
	Type of Fruits and Grading (Labels)	**Type of Fruits**	**Type of Grades**
RBG Color			
Grayscale			
Mean Pixel			
Edge Pixel			

Table 3 provides detailed performance metrics, including confusion matrices (table 3A), precision, recall, and F1-scores (table 3B), which further reveal the impact of dataset balance and feature extraction choices on classification outcomes. Misclassifications are prevalent among similar fruit types and grades, especially when class differences are subtle, such as between different grades of the same fruit. Precision and recall scores vary significantly, with higher performance in well-represented classes and lower scores in underrepresented or closely related categories. This imbalance, as depicted in figure 8, shows that the dataset is not evenly distributed across classes, with some fruit types and grades having far more instances than others. This imbalance contributes to lower performance in less frequent classes and highlights the need for data augmentation or balancing techniques.

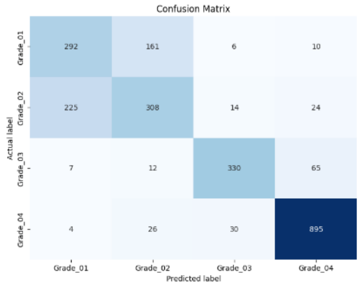
Table 3A. Confusion Matrix

	Confusion Matrix	
Type of Fruits and Grading (Labels)	 RBG Color	 Grayscale
	 Mean Pixel	 Edge Pixel
Type of Fruits	 RBG Color	 Grayscale
	 Mean Pixel	 Edge Pixel

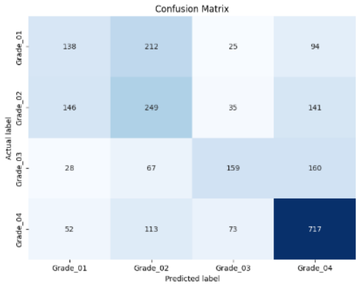
Type of Grades



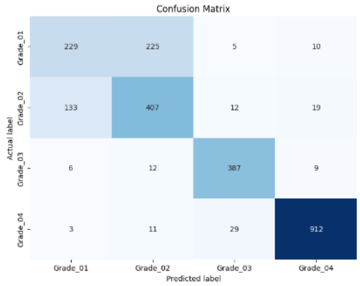
RBG Color



Grayscale



Mean Pixel



Edge Pixel

Table 3B. Confusion Matrix

Precision, Recall and F1-score

Type of Fruits and Grading (Labels)

Classification Report:

	precision	recall	f1-score	support
Apple_01	0.56	0.33	0.41	147
Apple_02	0.54	0.62	0.57	160
Apple_03	0.91	0.91	0.91	54
Apple_04	0.92	0.96	0.94	213
Banana_01	0.75	0.85	0.80	46
Banana_02	0.79	0.62	0.70	50
Banana_03	0.91	0.91	0.91	93
Banana_04	0.91	0.87	0.89	61
Guava_01	0.46	0.53	0.49	32
Guava_02	0.45	0.52	0.48	50
Guava_03	0.85	0.91	0.88	76
Guava_04	0.92	0.93	0.93	140
Lime_01	0.59	0.51	0.55	81
Lime_02	0.58	0.69	0.63	93
Lime_03	0.92	0.93	0.92	108
Lime_04	0.90	0.94	0.92	77
Orange_01	0.49	0.48	0.48	116
Orange_02	0.63	0.59	0.61	184
Orange_03	0.83	0.38	0.53	52
Orange_04	0.84	0.93	0.88	256
Pomegranate_01	0.60	0.53	0.56	47
Pomegranate_02	0.52	0.76	0.62	34
Pomegranate_03	0.97	0.94	0.95	31
Pomegranate_04	0.93	0.95	0.94	208
accuracy			0.76	2409
macro avg	0.74	0.73	0.73	2409
weighted avg	0.76	0.76	0.75	2409

RBG Color

Classification Report:

	precision	recall	f1-score	support
Apple_01	0.49	0.31	0.38	147
Apple_02	0.50	0.62	0.55	160
Apple_03	0.96	0.81	0.88	54
Apple_04	0.97	0.95	0.96	213
Banana_01	0.79	0.89	0.84	46
Banana_02	0.78	0.84	0.81	50
Banana_03	0.93	0.94	0.93	93
Banana_04	0.90	0.93	0.92	61
Guava_01	0.41	0.22	0.29	32
Guava_02	0.44	0.68	0.53	50
Guava_03	0.97	0.91	0.94	76
Guava_04	0.92	0.91	0.91	140
Lime_01	0.54	0.94	0.68	81
Lime_02	0.81	0.23	0.35	93
Lime_03	0.93	0.93	0.93	108
Lime_04	0.91	0.96	0.94	77
Orange_01	0.50	0.59	0.55	116
Orange_02	0.70	0.60	0.65	184
Orange_03	0.76	0.85	0.80	52
Orange_04	0.94	0.94	0.94	256
Pomegranate_01	0.72	0.66	0.69	47
Pomegranate_02	0.61	0.79	0.69	34
Pomegranate_03	0.94	0.97	0.95	31
Pomegranate_04	0.96	0.96	0.96	208
accuracy			0.78	2409
macro avg	0.76	0.77	0.75	2409
weighted avg	0.79	0.78	0.77	2409

Grayscale

Classification Report:

	precision	recall	f1-score	support
Apple_01	0.38	0.28	0.32	147
Apple_02	0.40	0.42	0.41	160
Apple_03	0.89	0.78	0.83	54
Apple_04	0.86	0.93	0.90	213
Banana_01	0.73	0.76	0.74	46
Banana_02	0.71	0.68	0.69	50
Banana_03	0.89	0.84	0.86	93
Banana_04	0.84	0.92	0.88	61
Guava_01	0.42	0.44	0.43	32
Guava_02	0.38	0.28	0.32	50
Guava_03	0.79	0.86	0.82	76
Guava_04	0.88	0.88	0.88	140
Lime_01	0.55	0.49	0.52	81
Lime_02	0.52	0.63	0.57	93
Lime_03	0.90	0.87	0.88	108
Lime_04	0.83	0.90	0.86	77
Orange_01	0.57	0.36	0.44	116
Orange_02	0.55	0.70	0.62	184
Orange_03	0.75	0.73	0.74	52
Orange_04	0.90	0.86	0.88	256
Pomegranate_01	0.56	0.60	0.58	47
Pomegranate_02	0.55	0.53	0.54	34
Pomegranate_03	0.83	0.94	0.88	31
Pomegranate_04	0.86	0.90	0.88	208
accuracy			0.71	2409
macro avg	0.69	0.69	0.69	2409
weighted avg	0.71	0.71	0.71	2409

Mean Pixel

Classification Report:

	precision	recall	f1-score	support
Apple_01	0.50	0.40	0.45	147
Apple_02	0.54	0.61	0.57	160
Apple_03	1.00	0.80	0.89	54
Apple_04	0.92	0.97	0.94	213
Banana_01	0.85	0.72	0.78	46
Banana_02	0.76	0.74	0.75	50
Banana_03	0.97	0.97	0.97	93
Banana_04	0.97	0.92	0.94	61
Guava_01	0.50	0.19	0.27	32
Guava_02	0.47	0.76	0.58	50
Guava_03	0.92	0.95	0.94	76
Guava_04	0.91	0.92	0.91	140
Lime_01	0.57	0.43	0.49	81
Lime_02	0.55	0.66	0.60	93
Lime_03	0.92	0.99	0.96	108
Lime_04	0.99	0.99	0.99	77
Orange_01	0.58	0.53	0.55	116
Orange_02	0.68	0.71	0.70	184
Orange_03	0.82	0.81	0.82	52
Orange_04	0.97	0.93	0.95	256
Pomegranate_01	0.76	0.89	0.82	47
Pomegranate_02	0.74	0.76	0.75	34
Pomegranate_03	1.00	0.94	0.97	31
Pomegranate_04	0.89	0.93	0.91	208
accuracy			0.79	2409
macro avg	0.78	0.77	0.77	2409
weighted avg	0.79	0.79	0.79	2409

Edge Pixel

Type of Fruits

Classification Report:

	precision	recall	f1-score	support
Apple	0.95	0.87	0.91	574
Banana	0.94	0.97	0.95	250
Guava	0.87	0.93	0.90	298
Lime	0.96	0.97	0.97	359
Orange	0.96	0.95	0.96	608
Pomegranate	0.92	0.97	0.95	320
accuracy			0.94	2409
macro avg	0.93	0.95	0.94	2409
weighted avg	0.94	0.94	0.94	2409

RBG Color

Classification Report:

	precision	recall	f1-score	support
Apple	0.96	0.89	0.92	574
Banana	0.90	0.98	0.94	250
Guava	0.92	0.88	0.90	298
Lime	0.93	0.98	0.96	359
Orange	0.96	0.97	0.96	608
Pomegranate	0.93	0.95	0.94	320
accuracy			0.94	2409
macro avg	0.93	0.94	0.94	2409
weighted avg	0.94	0.94	0.94	2409

Grayscale

 ISSN 2723-6471

Classification Report:

	precision	recall	f1-score	support
Apple	0.86	0.85	0.85	574
Banana	0.93	0.93	0.93	250
Guava	0.84	0.81	0.82	298
Lime	0.92	0.93	0.93	359
Orange	0.96	0.88	0.92	608
Pomegranate	0.79	0.95	0.86	320
accuracy			0.89	2409
macro avg	0.88	0.89	0.89	2409
weighted avg	0.89	0.89	0.89	2409

Mean Pixel

Classification Report:

	precision	recall	f1-score	support
Apple	0.95	0.84	0.89	574
Banana	0.93	0.95	0.94	250
Guava	0.82	0.95	0.88	298
Lime	0.96	0.97	0.96	359
Orange	0.96	0.94	0.95	608
Pomegranate	0.89	0.96	0.92	320
accuracy			0.92	2409
macro avg	0.92	0.93	0.92	2409
weighted avg	0.93	0.92	0.92	2409

Edge Pixel

Type of Grades

Classification Report:

	precision	recall	f1-score	support
Grade_01	0.60	0.31	0.41	469
Grade_02	0.54	0.78	0.64	571
Grade_03	0.89	0.77	0.83	414
Grade_04	0.90	0.92	0.91	955
accuracy			0.74	2409
macro avg	0.73	0.70	0.70	2409
weighted avg	0.75	0.74	0.73	2409

RBG Color

Classification Report:

	precision	recall	f1-score	support
Grade_01	0.55	0.62	0.59	469
Grade_02	0.61	0.54	0.57	571
Grade_03	0.87	0.80	0.83	414
Grade_04	0.90	0.94	0.92	955
accuracy			0.76	2409
macro avg	0.73	0.72	0.73	2409
weighted avg	0.76	0.76	0.76	2409

Grayscale

Classification Report:

	precision	recall	f1-score	support
Grade_01	0.38	0.29	0.33	469
Grade_02	0.39	0.44	0.41	571
Grade_03	0.54	0.38	0.45	414
Grade_04	0.64	0.75	0.69	955
accuracy			0.52	2409
macro avg	0.49	0.47	0.47	2409
weighted avg	0.52	0.52	0.51	2409

Mean Pixel

Classification Report:

	precision	recall	f1-score	support
Grade_01	0.62	0.49	0.55	469
Grade_02	0.62	0.71	0.66	571
Grade_03	0.89	0.93	0.91	414
Grade_04	0.96	0.95	0.96	955
accuracy			0.80	2409
macro avg	0.77	0.77	0.77	2409
weighted avg	0.80	0.80	0.80	2409

Grayscale

The analysis of table 1, table 2, and table 3, combined with the visual representation of dataset balance in figure 8, provides a comprehensive view of the classifier performance and the challenges faced due to dataset imbalance. These insights are critical for understanding the strengths and limitations of different feature extraction methods and their impact on classification accuracy.

The combined analysis suggests that while Edge Pixel and Grayscale methods are effective for classifying fruit types and grades, the model's performance could be better by class imbalance and the limitations of simpler feature extraction methods. The imbalance in the dataset, as shown in figure 8, significantly affects the model's ability to generalize across all classes, leading to poor performance in underrepresented categories. To enhance the model's accuracy, future efforts should focus on balancing the dataset, employing more sophisticated feature extraction techniques, and potentially integrating advanced classification models like deep learning. These strategies will help to address the current challenges, improve classification accuracy, and provide a more robust system for accurately identifying and grading fruits.

4. Conclusion

The implementation of an Automatic Fresh Fruit Selection and Grading system represents a crucial development in agricultural technology by leveraging advanced imaging, preprocessing, and feature extraction techniques. However, the performance of these systems is often hindered by imbalances in the dataset, affecting the accuracy and reliability of classification models, particularly in distinguishing various fruit types and grades.

The evaluation of different feature extraction methods reveals that techniques capturing edge and texture details, such as Edge Pixel and Grayscale, generally provide higher accuracy in fruit classification tasks. These methods excel in identifying the intricate features necessary for accurate grading and type classification. However, the distribution of data significantly impacts these results, as certain fruit types and grades are overrepresented, leading to biased learning. Models tend to perform exceptionally well in frequently occurring classes but struggle with underrepresented ones, resulting in uneven classification performance.

Further analysis of model learning stability shows that the convergence of training and validation accuracy can be stable and reliable for some methods. Nonetheless, the dominance of specific fruit types and grades within the dataset skews the learning process. This imbalance causes the model to favor predictions for the more common classes, thereby limiting its ability to identify less represented categories accurately. Such biases in the training data can mislead the model, affecting its capacity to generalize effectively and handle diverse classification scenarios.

Performance metrics, including precision, recall, and F1 scores, demonstrate that the classification model excels in frequently occurring categories but performs poorly in rarer ones. This disparity in performance is further exacerbated by an unequal distribution of grade levels, with higher-quality grades being far more prevalent. Consequently, the model has limited training exposure to lower grades, which are critical for applications that require detailed quality differentiation. This lack of exposure results in unreliable predictions when lower-grade classifications are essential.

Addressing these challenges requires implementing strategies that enhance the dataset's balance and improve model learning across all classes. Techniques such as class weighting, data augmentation, and oversampling minority classes can help mitigate the effects of data imbalance, allowing the model to learn from a more representative distribution of features. These adjustments would enable the model to make fair and accurate predictions, improving its generalization capabilities and performance in real-world applications where uniform accuracy is crucial.

The imbalanced nature of the dataset significantly undermines the classification model's effectiveness, leading to biased predictions and poor performance in less common classes. By employing corrective strategies in data preparation and training, the model can be refined to deliver more consistent, reliable results, enhancing its practical application in the automatic selection and grading of fresh fruits.

5. Declarations

5.1. Author Contributions

Conceptualization: D.A.D., T.B.K., R.T., M.B., S.H., M.I.; Methodology: M.B.; Software: D.A.D.; Validation: D.A.D., M.B., dan M.I.; Formal Analysis: D.A.D., M.B., dan M.I.; Investigation: D.A.D.; Resources: M.B.; Data Curation: M.B.; Writing Original Draft Preparation: D.A.D., M.B., dan M.I.; Writing Review and Editing: M.B., D.A.D., dan M.I.; Visualization: D.A.D.; All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Alturki, M. Islam, M. F. Alsharekh, M. Almanee, and A. Hassan Ibrahim, “Date Fruits Grading and Sorting Classification Algorithm Using Colors and Shape Features,” *Int. J. Eng. Res. Technol.*, vol. 2020, no. 8, pp. 1917–1920, 2020.

[2] K. M. Prabusankarlal, P. Thirumoorthy, and R. Manavalan, “Assessment of combined textural and morphological features for diagnosis of breast masses in ultrasound,” *Hum.-Centric Comput. Inf. Sci.*, vol. 2015, no. 12, pp. 1–10, 2015.

[3] A. Hassoun, F. Carullo, F. Polignano, R. L. Ayoub, and L. Xu, “Food quality 4.0: From traditional approaches to digitalized automated analysis,” *J. Food Eng.*, vol. 2023, pp. 111216, 2023.

[4] A. Bhargava and A. Bansal, “Fruits and vegetables quality evaluation using computer vision: A review,” J. King Saud Univ. - Comput. Inf. Sci., vol. 33, no. 3, pp. 243–257, 2021.

[5] Anjali, A. B., A. Banerjee, and A. Sharma, “State-of-the-art non-destructive approaches for maturity index determination in fruits and vegetables: principles, applications, and future directions,” *Food Prod. Process. Nutr*., vol. 2024, no. 1, pp. 1–12, 2024.

[6] N. Gururaj, V. Vinod, and K. Vijayakumar, “Deep grading of mangoes using Convolutional Neural Network and Computer Vision,” *Multimed. Tools Appl*., vol. 82, no. 25, pp. 39525–39550, 2023.

[7] M. Palumbo, M. Cefola, B. Pace, G. Attolico, and G. Colelli, “Computer vision system based on conventional imaging for non-destructively evaluating quality attributes in fresh and packaged fruit and vegetables,” *Postharvest Biol. Technol*., vol. 200, pp. 112332, 2023.

[8] A. Momin, N. Kondo, D. F. al Riza, Y. Ogawa, and D. Obenland, “A Methodological Review of Fluorescence Imaging for Quality Assessment of Agricultural Products,” *Agriculture*, vol. 2023, no. 7, pp. 1–12, 2023.

[9] N. D. Tai, W. C. Lin, N. M. Trieu, and N. T. Thinh, “Development of a Mango-Grading and -Sorting System Based on External Features, Using Machine Learning Algorithms,” *Agronomy*, vol. 2024, no. 4, pp. 1–15, 2024.

[10] N. M. T. Nguyen and N.-S. Liou, “Detecting Surface Defects of Achacha Fruit (Garcinia humilis) with Hyperspectral Images,” *Horticulturae*, vol. 2023, no. 8, pp. 1–12, 2023.

[11] M. Zhao, Q. Lin, J. Zhang, R. Li, and H. Wang, “Determination of quality and maturity of processing tomatoes using near-infrared hyperspectral imaging with interpretable machine learning methods,” *LWT*, vol. 183, pp. 114861, 2023.

[12] M. Agarla, P. Napoletano, and R. Schettini, “Quasi Real-Time Apple Defect Segmentation Using Deep Learning,” *Sensors*, vol. 2023, no. 18, pp. 1–18, 2023.

[13] K. Hameed, D. Chai, and A. Rassau, “Texture-based latent space disentanglement for enhancement of a training dataset for ANN-based classification of fruit and vegetables,” *Inf. Process. Agric*., vol. 10, no. 1, pp. 85–105, 2023.

[14] B. M. Nicolaï, J. Lammertyn, V. M. S. Beullens, and W. Saeys, “Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review,” *Postharvest Biol. Technol*., vol. 46, no. 2, pp. 99–118, 2007.

[15] Y. Huang and Z. Liang, “Assessment of apple bruise resistance under transient collisions through X-ray computed tomography and image processing,” *Biosyst. Eng*., vol. 2024, no. 5, pp. 16–25, 2024.

[16] T. Akter, A. Rahman, S. Hasan, M. Islam, and M. S. Uddin, “A comprehensive review of external quality measurements of fruits and vegetables using nondestructive sensing technologies,” *J. Agric. Food Res*., vol. 2024, no. 1, pp. 1–10, 2024.

[17] Y. H. Tan, Y. J. Khoo, M. K. Chai, and L. S. Wong, “Tropical fruit wastes as an organic nutrient sources for the cultivation of Chlorella vulgaris and Haematococcus pluvialis,” *Nat. Environ. Pollut. Technol*., vol. 2021, no. 2, pp. 613–618, 2021.

[18] A.R.Yadulla, G.S.Nadella, M.H.Maturi, H.Gonaygunta, “Evaluating Behavioral Intention and Financial Stability in Cryptocurrency Exchange App: Analyzing System Quality, Perceived Trust, and Digital Currency in Indonesia,” *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 2, pp. 103-124, 2024

[19] Henderi and Q. Siddique, “Anomaly Detection in Blockchain Transactions within the Metaverse Using Anomaly Detection Techniques*”, J. Curr. Res. Blockchain.,* vol. 1, no. 2, pp. 155–165, Sep. 2024.

[20] R. A. Putawa and D. Sugianto, “Exploring User Experience and Immersion Levels in Virtual Reality: A Comprehensive Analysis of Factors and Trends,” *Int. J. Res. Metav.*, vol. 1, no. 1, pp. 20-39, 2024