**Journal of Asian Scientific Research**
*ISSN(e): 2223-1331*
*ISSN(p): 2226-5724*
*DOI: 10.55493/5003.v15i4.5758*
*Vol. 15, No. 4, 849-863.*

*URL: www.aessweb.com*

# Machine learning-based classification of macadamia nut quality using physical features

**Sirikunya Nilpanich**[1]
**Wilawan Rukpakavong**[2+]
**Kannikar Subsomboon**[3]

[1,2,3]*Faculty of Science and Technology, Thammasat University, Thailand.*
[1]*Email: nsirikun@tu.ac.th*
[2]*Email: rwilawan@tu.ac.th*
[3]*Email: kns@tu.ac.th*

*(+ Corresponding author)*

## Article History
Received: 29 August 2025
Revised: 14 October 2025
Accepted: 7 November 2025
Published: 26 November 2025

## Keywords
Agricultural processing
Macadamia nuts
Machine learning
Postharvest technology
Quality classification
Specific gravity
Wet floating.

## *ABSTRACT*

Macadamia nuts in shell must be classified into excellent and faulty categories to maintain market value, optimize yield efficiency, and ensure consistent product quality. Although traditional methods such as wet floating and dry specific gravity (SG) remain widely used due to their simplicity and low cost, their accuracy and consistency are often limited in large-scale or automated operations. Machine learning offers a more advanced and efficient alternative, enabling higher levels of automation, precision, and reliability. This study evaluates and compares the performance of wet floating, dry SG at different threshold levels, and machine learning models using a dataset of 1,260 macadamia nuts collected during the peak harvest period from multiple orchards in Loei Province, Thailand. The wet floating method achieved an accuracy of 78.33% with high precision (90.00%) but relatively low recall (72.97%), indicating its tendency to misclassify a considerable portion of high-quality nuts. The dry SG method demonstrated the most balanced performance at the 0.9 threshold, with 89.50% accuracy, 90.00% precision, 89.11% recall, and an F1-score of 89.55%, while threshold variation revealed clear trade-offs between precision and recall. Machine learning outperformed the traditional methods, with the Random Forest model yielding the highest performance (accuracy 92.06%, precision 94.44%, recall 91.07%, and F1-score 91.07%). These findings highlight the potential of integrating machine learning–based classification to enhance accuracy, increase operational efficiency, strengthen product quality assurance, and support more sustainable and competitive agricultural value chains.

**Contribution/ Originality:** This study is one of the few that has investigated the application of machine learning, particularly the Random Forest algorithm, to macadamia nut classification, demonstrating significant improvements in accuracy and balance over traditional wet floating and dry specific gravity methods. By enhancing precision, recall, and F1-score, the approach increases processing efficiency, product quality, and market value, supporting smarter and more consistent agricultural operations.

## 1. INTRODUCTION

Macadamia, a valuable perennial crop native to Australia and New Zealand, is recognized for its excellent taste, rich nutritional content, and high market demand [1, 2]. The significant levels of monounsaturated fatty acids, particularly oleic and palmitoleic acids, in the kernels render them particularly important due to their association with anti-inflammatory, cholesterol-regulating, and heart health advantages. Due to these traits and their appealing taste and texture, macadamia is currently a highly valued component in both its natural and processed states, such as roasted nuts, sweets, baked items, and gourmet cooking uses. Figure 1 illustrates the major stages involved in the macadamia nuts processing [3].

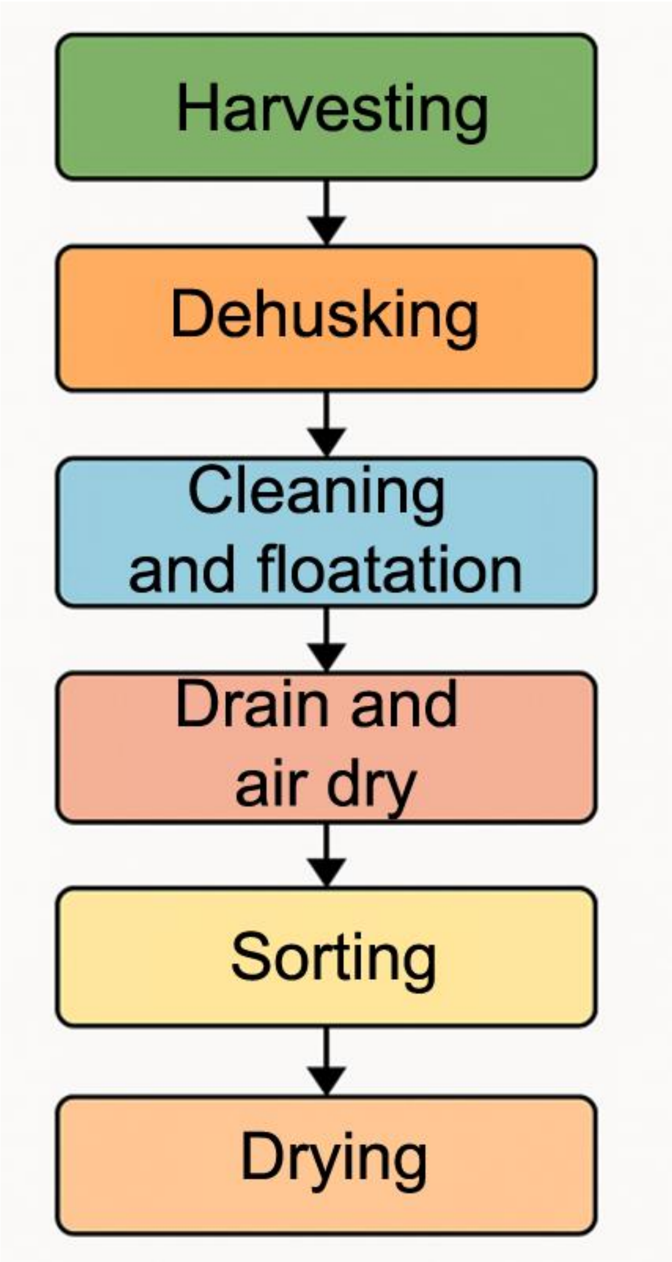


**Figure 1.** The processing stages of Macadamia.

The macadamia processing begins with harvesting, during which mature nuts are collected promptly from the orchard floor to minimize quality deterioration caused by pest infestation, excess moisture, or fungal growth. Once collected, the nuts undergo dehusking to remove the thick green outer husk, which retains a high level of moisture and can promote mold development if left intact. The macadamia nuts prior to dehusking are shown in Figure 2, and those after dehusking are presented in Figure 3. Following this step, the nuts are subjected to cleaning and floatation, where they are thoroughly washed to remove soil, leaves, and other debris, and then placed in a water floatation system. During this stage, immature or damaged nuts float to the surface, while mature, good-quality nuts sink, enabling an effective preliminary quality separation. The cleaned nuts are then drained and air-dried to remove surface moisture, thereby reducing the risk of mold growth during subsequent handling and processing. Next, the nuts are sorted to eliminate cracked, damaged, or discolored shells, ensuring that only high-quality nuts advance to the drying phase. Finally, the drying process lowers the internal moisture content from approximately 20–25% to 10–12% using either natural air drying or controlled heated-air systems, thereby stabilizing the nuts for safe storage, transportation, and further processing.


**Figure 2.** Macadamia nuts before Dehusking.


**Figure 3.** Macadamia nuts after Dehusking.

Over the past few decades, macadamia cultivation has gradually increased in Thailand, particularly in the northern and northeastern areas such as Loei, Phetchabun, Chiang Mai, Chiang Rai, and Chaiyaphum. The cultivation area exceeds 50,000 rai, or roughly 8,000 hectares, according to agricultural statistics. Annual production is approximately 12,000 tons, contributing more than 1.2 billion THB to the national economy. Macadamia is now a major cash crop in Thailand's agricultural diversification strategy, driven by growth in both domestic consumption and export potential [4]. To ensure that only premium kernels reach the market, postharvest quality sorting is a critical step in the macadamia value chain. The most widely used technique in Thailand is the water-floating method, in which fully developed nuts with intact kernels sink while immature or defective nuts float. Although this method is simple, cost-effective, and widely adopted, it has a major drawback: moisture absorption. Excess moisture reduces storage life and market value by promoting mold growth, microbial contamination, and rancidity. It also increases drying time and energy consumption, leading to higher operating costs and slower processing. Another common method is manual sorting by visual inspection; however, this method is subjective, time-consuming, and prone to errors. Human performance varies depending on experience, lighting conditions, and weariness, which can result in misclassification and reduced product consistency. As market demand continues to rise, these conventional methods are increasingly insufficient to meet industry needs for efficiency, scalability, and reliability [4-6].

Modern advancements in image processing and machine learning (ML) provide a robust alternative to traditional sorting methods. ML algorithms enable automated and objective product classification by learning from examples to detect patterns and relationships within data. Image-based features such as color, texture, and shape have proven to be reliable indicators of produce quality in agricultural quality assessment [7-14]. Particularly for macadamia, physical characteristics such as size, weight, and shell color can be used to predict kernel quality with high accuracy. Defective kernels are frequently associated with lighter nuts, whereas differences in shell color can reveal internal flaws, storage history, or maturity differences.

The primary challenge lies in replacing the water-floating method with a non-invasive, moisture-free, and scalable alternative. Such a solution must not only maintain classification accuracy but also improve processing speed, reduce post-harvest losses, and be implementable within existing processing infrastructure. Furthermore, to gain industry adoption, the method should be cost-effective and require minimal specialized training for operators. To address this gap, this study uses easily measurable physical features weight, size, and shell color extracted from high-resolution images to create an ML-based classification system.

The objectives of this study are to:

1. Evaluate the effectiveness of physical features in predicting macadamia nut quality.
2. Compare traditional methods (wet floating and dry SG) with machine learning approaches for classification accuracy.
3. Demonstrate the potential of ML-based sorting to enhance productivity, quality consistency, and efficiency in Thailand's macadamia industry.

By achieving these objectives, the research aims to provide a practical, data-driven tool for Thailand’s macadamia sector, supporting the transition toward “smart agriculture” and improving competitiveness in domestic and international markets.

The paper is organized as follows. Section 2 provides a comprehensive review of the existing literature on nut and seed quality classification. The methodology and methods are described in Section 3. The findings and discussion are presented in Section 4. Finally, Section 5 summarizes the key conclusions and suggests directions for future research.

## 2. LITERATURE REVIEW

### *2.1. Related Studies on Nut and Seed Classification*

Classification of Cashew Kernels: Baitu et al. [7] examined the use of color characteristics to classify cashew kernels into good and defective categories. After testing several machine learning algorithms and extracting color moments from RGB and HSV images, they found that the Random Forest algorithm, with 100 trees, outperformed Decision Tree (98.4 percent accuracy) and other models, achieving an overall accuracy of 99.8 percent. Evaluation of Hazelnut Quality: Ünal and Aktaş [8] used deep learning to categorize hazelnut kernels into three groups: intact, damaged, and empty shells. In comparison to models trained from scratch, they achieved accuracies above 99 percent and significantly reduced misclassification rates by using architectures like EfficientNetB2 and EfficientNetB3 with transfer learning from ImageNet weights. Identification of Hazelnut Flaws: Erbaş et al. [9] divided hazelnuts into five quality classes using the AlexNet deep neural network. The model demonstrated high accuracy, achieving 99 percent, with precision and recall values above 98 percent. Huang et al. [10] identified the species of pine nuts using image-based machine learning in conjunction with near-infrared spectroscopy to classify seven species of pine nuts. For NIR data, they found that a Multilayer Perceptron achieved an accuracy of 99 percent, while the InceptionV3

Convolutional Neural Network (CNN) achieved an accuracy of 96.4 percent for image data. Additionally, the study identified the most suitable wavelength bands for classification. Estimating Canarium Indicum quality: Han et al. [11] used CNNs and hyperspectral imaging, categorizing Canarium Indicum kernels into three quality levels based on peroxide value. With high class-specific precision and recall values, the top-performing model achieved an overall accuracy of 93.48%. Ünal et al. [12] proposed a method to detect and classify bruised apples after harvest. Their findings indicate that using Super Chief RGB datasets for training produced the highest accuracy, at 86 percent. Gulzar et al. [13] introduced an innovative deep learning model for seed classification. A dataset of 15 seed types, comprising approximately 3,018 RGB images, was created to develop a precise and effective model capable of high accuracy in seed classification.

A recurring theme in these studies is that when given informative feature sets, machine learning models especially ensemble approaches and deep learning architectures perform exceptionally well at classifying nut quality. However, there is a dearth of research specifically addressing macadamia nuts, with the majority of existing work concentrating on cashews, hazelnuts, pine nuts, and Canarium indicum. Furthermore, even though they are very accurate, some sophisticated imaging methods, such as NIR spectroscopy or hyperspectral imaging, might be too expensive for small- to medium-sized processors. Because of this, there is a chance to create a classification system tailored to macadamias that is precise, reasonably priced, and flexible enough to accommodate different production sites.

### *2.2. Post-Harvest Quality Challenges in Macadamia Nuts*

There are multiple steps involved in the process of turning nuts in their shells from harvest to packaged product. The following are some possible risks of quality degradation at each stage:

- Mold and fungal contamination can occur if nuts are stored with high moisture content.
- Physical damage resulting from improper handling or mechanical equipment.
- Kernel discoloration caused by enzymatic browning.
- Pest damage, particularly from macadamia nut borer and stink bugs.

Sorting and grading are essential components of quality control because faulty nuts can jeopardize food safety and shelf life, in addition to lowering batch value [14].

### *2.3. Conventional Quality Sorting Methods*

#### *2.3.1. Manual Sorting*

The most conventional method for quality sorting is manual inspection, in which trained workers visually identify defects such as discoloration, mold spots, or shell cracks. Although manual sorting can achieve relatively high accuracy, it relies heavily on skilled labor and is associated with several notable disadvantages. It is labor-intensive and costly, particularly for large-scale operations; it requires considerable time, making it unsuitable for high-throughput processing; and it is inconsistent, as performance may vary depending on individual worker fatigue or subjective judgment. Previous studies conducted on other crops, such as cashew nuts and coffee beans, have shown that the accuracy of manual sorting typically declines after two to three hours of continuous work, underscoring its limitations as a sustainable and scalable method [15]. These challenges highlight the need for more efficient, objective, and scalable approaches, paving the way for the adoption of machine learning and image processing technologies as promising alternatives for macadamia quality classification.

#### *2.3.2. Wet Flotation Sorting*

The macadamia industry extensively uses the water flotation method, which is based on the principle that high-quality nuts with fully developed kernels sink, whereas low-density nuts typically caused by empty shells or immature kernels float [15]. This method offers several advantages. It provides relatively high accuracy in detecting voids or hollow kernels, and it is simple to implement without the need for advanced technology or specialized equipment.

However, despite these benefits, the method also has notable limitations. Because it introduces additional moisture, it increases the risk of mold development, thereby reducing shelf life and quality. Furthermore, the nuts require post-sorting drying, which adds both time and operational costs to the processing chain. Another critical drawback is that the method cannot detect defects unrelated to density, such as surface mold or other external blemishes that do not result in void formation. These limitations reduce its effectiveness and highlight the need for more advanced, non-invasive sorting technologies for macadamia quality classification.

### *2.4. Dry Specific Gravity Quality Sorting*

Dry specific gravity (SG) sorting uses measured mass (weight) and volume to determine nut density as an alternative to wet methods. For post-harvest handling that is sensitive to moisture, this eliminates water contact and the need for drying. The specific gravity is calculated as [16-18].

$$SG = \frac{\rho}{\rho_w} \qquad (1)$$

Where $\rho$ is the density of a substance and $\rho_w$ is the density of water. The density can be calculated as the amount of mass (m) per unit volume (v), or the amount of mass packed into a given volume [18].

$$\rho = \frac{m}{v} \qquad (2)$$

A weighing scale can be used to determine an object's mass (m), where the volume (v) can be calculated using the following equation. Since macadamia nuts are roughly spherical objects, the volume can be determined by measuring the sphere's radius (r), which is half its diameter, as illustrated in Figure 4.

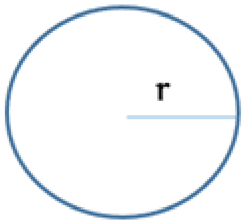


**Figure 4.** Sphere's radius.

Therefore, the value of volume (v) can be calculated from.

$$v = \frac{4}{3} \pi r^3 \qquad (3)$$

The value of pi ($\pi$) is approximately equal to 3.14159, and r is the sphere's radius. Agricultural products have been sorted using dry SG such as:

- Cocoa beans [19] for separating under-fermented beans.
- Coffee and cherries [20] for maturity grading.
- Hazelnuts [21] for defect removal.

The efficiency of SG sorting for macadamias relies on selecting the ideal density threshold to balance false positives and false negatives. While high thresholds run the risk of rejecting good nuts, low thresholds might allow defective nuts to pass.

Taken together, the literature reveals that while machine learning and advanced imaging methods have demonstrated high accuracy in classifying nuts and seeds, most studies have focused on crops other than macadamias and rely on costly or complex technologies. Traditional methods like manual sorting, wet flotation, and dry SG each have strengths such as simplicity, accessibility, or quantitative rigor but suffer from limitations in scalability, consistency, or defect coverage. This gap underscores the need for a cost-effective, scalable, and accurate ML-based classification system tailored specifically to macadamia nuts, using easily measurable physical features that can be implemented in diverse production environments.

## 3. MATERIALS AND METHODS

The methodological framework employed in this study is presented in detail in this section, including the materials used, measurement procedures, image acquisition and preprocessing, feature extraction, dataset preparation, machine learning model development, and performance evaluation. The primary objective was to establish a reproducible and scientifically rigorous workflow that enables accurate macadamia nut quality classification using both physical attributes and image-based features. This structured approach ensures methodological transparency, facilitates replication, and supports future optimization of automated sorting technologies for postharvest processing.

### *3.1. Study Area and Sampling Design*

Loei Province, Thailand, is well known for producing high-quality macadamia nuts due to its favorable elevation, temperature range, and soil conditions. The samples used in this study were collected from orchards located in this region. The province's cool, dry winters and moderate rainfall during the growing season provide optimal conditions for flowering, nut set, and kernel filling. To ensure the representativeness of the dataset, nuts were collected from multiple farms employing different cultivation methods. Variations in orchard management practices, irrigation schedules, and harvesting times helped capture the natural variability in nut quality. Sampling was conducted during the peak harvest period to ensure that the nuts were mature and representative of those typically produced for commercial purposes. A total of 1,260 macadamia nuts (10 kg) were selected for analysis. Each nut was at the stage where the outer green husk had already been removed, leaving the intact shell visible. To avoid confounding factors in the evaluation of shell-based quality indicators, nuts with visible cracks, mechanical damage, or mold growth were excluded from the study.

### *3.2. Physical Measurement*

To prevent changes in weight or shell characteristics caused by handling or moisture loss, the physical attributes of the nuts were measured immediately after sample collection. All measurements were performed in a controlled laboratory environment at the Faculty of Science and Technology, Thammasat University.

#### *3.2.1. Dimensional Measurement*

Using a precision measuring device, shown in Figure 5, the diameter of the seed was measured, with the radius being equal to the diameter divided by 2.


**Figure 5.** Size-measuring device with a precision of 0.01 mm.

#### *3.2.2. Weight Measurement*

A digital balance with an accuracy of ±0.001 g was used, as shown in Figure 6, to record the weight of the nuts. To ensure stability, each nut was positioned in the center of the weighing pan. Once the reading stabilized, the measurement was recorded. Before each measurement session, the balance was calibrated.


**Figure 6.** Digital weighting with high accuracy.

### *3.2.3. Color Feature Extraction*

To facilitate the extraction of color features, high-resolution photos were taken. To maintain a constant imaging distance and angle, the camera of a OnePlus 9 Pro smartphone was fixedly positioned 30 cm above the sample platform, as shown in Figure 7. A plain white matte background was used to simplify segmentation during subsequent processing, creating high contrast with the brown tones of the macadamia shells. Two LED light panels, angled at 45 degrees to the platform, provided illumination, reducing glare and shadows and ensuring consistent lighting.

**Figure 7.** Obtaining high-quality photos.

Every nut was positioned separately in the imaging platform's center. To maintain detail, photos were taken using the device's highest resolution setting. To keep dust and smudges from degrading the clarity of the images, the camera lens was cleaned regularly. Every image file was given a sequential name that corresponded to the dataset's sample ID, ranging from 0001.jpg to 1260.jpg.

To standardize color representation and prepare them for feature extraction, raw photos were pre-processed. The HSV (Hue-Saturation-Value) color system was used to transform RGB photos. Compared to RGB, HSV is less susceptible to changes in illumination because it distinguishes between the chromatic components (Hue and Saturation) and the intensity component (Value). Additionally, grayscale photos were produced for the possible examination of the homogeneity of shell texture and intensity. Although it was not the main focus of this study, exploratory feature analysis gained a new dimension from grayscale data. The backdrop was eliminated using a semiautomated segmentation technique. This process removed background interference in color statistics, including only the nut pixels in the analysis. The statistical distribution of pixel values in the HSV channels was compiled using the color moments method. A nine-dimensional color feature vector was produced for each nut by computing these three statistics for each of the three HSV channels (H, S, and V), which are:

- Mean: Measures average color intensity in each channel, capturing the general tone.
- Standard deviation: quantifies the variation of pixel values, indicating shell uniformity.
- Skewness: Describes the asymmetry in color distribution, detecting subtle deviations due to stains or discolorations.

A nine-dimensional color feature vector was produced for each nut by computing these three statistics for each of the three HSV channels (H, S, and V).

### *3.3. Dataset Preparation*

The feature set includes weight, dimensions, and color. The dataset was organized in a tabular format, with each row corresponding to a nut sample and each column representing a feature or the target label. The features were normalized using min-max scaling to a range of [0,1] to guarantee equal importance during the model training process. The dataset was split into a training set consisting of 70% (n = 882 nuts) and a test set comprising 30% (n = 378 nuts). Stratified sampling was applied to maintain the initial ratio of high-quality and defective nuts in both subsets. Each seed was tested to determine whether it floats or sinks. Subsequently, the shell was cracked to determine whether the kernel was accepted or rejected. Mold, brown centers, insect damage, and internal discoloration are some of the causes of kernel rejection [22-27]. An accepted kernel is shown in Figure 8, whereas multiple rejected kernels are shown in Figure 9.

**Figure 8.** An accepted kernel.

**Figure 9.** Multiple rejected kernels.

### *3.4. Experimental Methods*

Using a test data set of 378 nuts, the experiment compares three approaches for quality prediction. The three techniques are wet floating, dry specific gravity, and machine learning. If the seed sinks, the wet floating prediction is considered successful (accepted kernel). However, it is a rejected kernel if the seeds float. The SG value for dry specific gravity prediction is determined using (1). A seed is expected to be an accepted kernel if its SG value is higher than or equal to a threshold value. Conversely, the kernel is rejected if the SG value is below the threshold. Five supervised learning algorithms were used for the machine learning technique: K-nearest neighbor (K-NN), Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Logistic Regression (LR). Prior to predicting a test set of 378 nuts, a training set of 882 nuts is used for training.

### *3.5. Performance Evaluation*

Accuracy, precision, recall, and F1-score are important performance indicators for techniques. Accuracy gauges how accurate forecasts are overall, while precision and recall focus on positive predictions. The F1-score combines recall and precision to provide a balanced metric. The confusion matrix value from Table 1 is used to determine each metric.

**Table 1.** A confusion matrix.

| Actual/Prediction Result | | Prediction | |
|---|---|---|---|
| | | **Accepted Kernel (Good)** | **Rejected Kernel (Bad)** |
| Actual | Accepted Kernel (Good) | True Positives | False negatives |
| | Rejected Kernel (Bad) | False Positives | True negatives |

According to Table 1, a true positive (TP) indicates that the accepted kernel was correctly predicted, and a true negative (TN) indicates that the rejected kernel was correctly predicted. A false negative (FN) indicates that the prediction is rejected but the kernel is actually accepted, and a false positive (FP) indicates that the prediction is accepted but the kernel is actually rejected. The accuracy, precision, recall, and F1-score are computed as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

$$F1-Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

## 4. RESULTS AND DISCUSSION

### *4.1. Result*

Tables 2, 3, and 4 display the performance metrics of 378 testing nuts for wet floating, dry specific gravity, and machine learning techniques.

**Table 2.** Wet floating result.

| Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|
| 78.33 | 90.00 | 72.97 | 80.60 |

Table 2 presents the performance of the wet floating method for macadamia nut classification. The method achieved an overall accuracy of 78.33%, indicating that slightly more than three-quarters of the nuts were correctly classified. The precision of 90.00% reflects the method's strong ability to correctly identify high-quality nuts among those classified as excellent, meaning it produces relatively few false positives. However, the recall of 72.97% reveals a notable limitation: a significant portion of high-quality nuts were misclassified or not detected. The F1-score of 80.60%, which represents the harmonic mean of precision and recall, confirms this imbalance between the two metrics. Overall, while the wet floating method demonstrates high precision, its lower recall indicates that it tends to overlook a considerable number of good-quality nuts, reducing its effectiveness as a reliable sorting method for large-scale postharvest operations.

**Table 3.** Dry specific gravity result.

| SG threshold | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| 0.8 | 87.50 | 77.00 | 97.47 | 86.03 |
| 0.9 | 89.50 | 90.00 | 89.11 | 89.55 |
| 1.0 | 85.50 | 97.00 | 78.86 | 87.00 |

Table 3 presents the classification performance of the dry specific gravity (SG) method at three threshold levels: 0.8, 0.9, and 1.0. At the 0.8 threshold, the method achieves the highest recall (97.47%), indicating its ability to identify nearly all good nuts, although precision is relatively lower (77.00%), reflecting a higher rate of misclassification. At the 0.9 threshold, the performance is the most balanced, with accuracy (89.50%), precision (90.00%), and recall (89.11%) closely aligned, resulting in the highest F1-score (89.55%). At the 1.0 threshold, precision reaches its maximum value (97.00%), meaning the method is highly effective at minimizing false positives, but recall drops to 78.86%, indicating some good nuts are not detected. These results illustrate a clear trade-off between precision and recall across threshold levels, with 0.9 providing the optimal balance for effective macadamia nut classification.

**Table 4.** Machine learning result.

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Logistic regression | 84.13 | 87.00 | 80.56 | 83.65 |
| Decision tree | 90.48 | 91.10 | 94.83 | 92.93 |
| Random forest | 92.06 | 94.44 | 91.07 | 91.07 |
| Support vector machine | 83.07 | 85.71 | 79.84 | 82.66 |
| K-nearest neighbor | 88.36 | 90.12 | 87.50 | 88.79 |

Table 4 presents the classification performance of five machine learning algorithms applied to macadamia nut quality sorting: Logistic regression, Decision tree, Random forest, Support vector machine, and K-nearest neighbor. Logistic regression is a respectable machine learning technique; however, it is not as strong as tree-based models. Despite being marginally less accurate than Random forest, Decision tree has a very high recall (94.83%), which means it rarely misses good-quality nuts. The best overall accuracy (92.06%), precision (94.44%), and recall (91.07%) are all displayed by Random forest. Out of all the ML models examined, Support vector machine has the lowest performance. It demonstrates strong mid-level performance with decent balancing for K-nearest neighbor. In conclusion, Random forest is the model that performs the best overall, with Decision tree coming in second.

Figure 10 compares the best performance outcomes of the three evaluated classification approaches: wet floating, dry SG at a 0.9 threshold, and Random Forest. The results clearly show that the dry SG method substantially outperforms wet floating across all performance metrics, including accuracy, precision, recall, and F1-score. This finding reinforces the practical value of dry SG as a more reliable option for producers seeking an affordable and relatively simple solution without the drawbacks of moisture absorption. In contrast, wet floating demonstrates the weakest performance and may be appropriate only in environments with extremely limited resources or where technological integration is not feasible. Random Forest consistently achieves the highest performance across all evaluated metrics, demonstrating its strong capability for accurate and balanced classification. Its superior precision and recall indicate its effectiveness in correctly identifying both high- and low-quality nuts, while its high F1-score reflects the method's robustness.

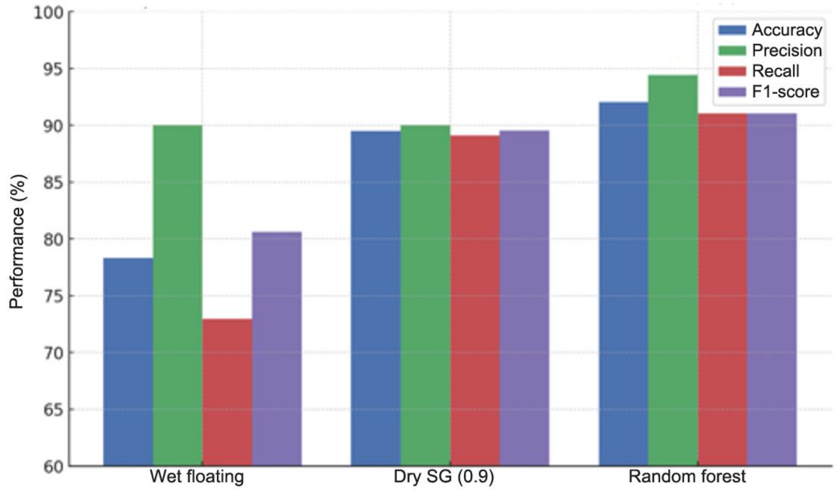


**Figure 10.** Comparing the best results of each method type.

### *4.2. Discussion*

The study's findings shed important light on the relative merits of the three main sorting strategies wet floating, dry specific gravity (SG) techniques, and machine learning algorithms for the classification of macadamia nuts. Regarding accuracy, precision, recall, and F1-score, each approach showed unique advantages and disadvantages, with varying trade-offs based on the operational goals and production environment.

Wet floating is a popular, inexpensive, and simple method. With a high precision of 90.00% and a relatively low recall of 72.97%, the study's accuracy was 78.33%. This suggests that although the approach is dependable in guaranteeing that nuts categorized as "excellent" are indeed high quality, it is unable to detect a sizable percentage of truly high-quality nuts. Since many high-quality nuts go unnoticed, the comparatively low recall leads to lost output. In practice, wet floating might still be useful in settings with limited resources or on smallholder farms when financial limitations preclude the use of more sophisticated methods. It might be more appropriate as a preliminary screening step rather than a final categorization technique. In this manner, it can assist in removing blatantly inferior nuts prior to using more exacting methods. Nevertheless, wet floating should not be used as the only technique to guarantee constant product quality, as indicated by the generally poor performance metrics. When compared to wet floating, the dry SG approach performed better and more consistently. The approach achieved 89.50% accuracy, 90.00% precision, 89.11% recall, and 89.55% F1-score at the 0.9 SG threshold. These findings demonstrate a strong trade-off between reducing false positives and increasing the identification of healthy nuts. The 0.9 threshold is the optimal choice for real-world application because of this balance. Different thresholds highlight the trade-offs associated with this strategy. Recall was maximized at 0.8 SG (97.47%), suggesting that nearly all good nuts were detected; however, accuracy decreased to 77.00%, indicating a higher number of false positives. Conversely, recall decreased to 78.86% at 1.0 SG while precision peaked at 97.00%, indicating that many genuine good nuts were missed. These findings emphasize the importance of threshold selection, with 0.9 providing the most reliable compromise.

In this study, machine learning specifically, the Random Forest model produced the best overall results. In terms of precision and general reliability, Random Forest outperformed both wet floating and dry SG, with 92.06% accuracy, 94.44% precision, 91.07% recall, and 91.07% F1-score. These results indicate a high potential for reducing classification errors, which would guarantee a high yield and consistent product quality. Random Forest outperformed other models, including Decision Tree (accuracy 90.48%) and K-Nearest Neighbor (accuracy 88.36%), but they were still quite effective. In this context, Support Vector Machine (83.07%) and Logistic Regression (84.13%) were less competitive. The primary drawback of machine learning techniques is their requirement for scientific expertise, labeled datasets, and computational resources. However, once implemented, these models offer automation, flexibility, and scalability. Machine learning remains the most viable approach for large-scale processors aiming to update their processes and reduce human error. It should be noted that this study was limited to macadamia nuts collected from a single region (Loei Province, Thailand) during one harvest season, which may reduce the generalizability of the findings to other regions or seasons with different growing conditions. Although the sample size was sufficient for training and testing, it may not capture the full variability present across different orchards and cultivation practices.

## 5. CONCLUSIONS

Macadamia nut processing plays a critical role in maintaining product quality, market value, and economic returns for producers. In Thailand, where cultivation has expanded significantly over recent decades, traditional post-harvest sorting methods particularly wet floating and dry specific gravity (SG) remain the most widely used due to their simplicity and low operational costs. While these methods have contributed to the development of the industry, they also present notable limitations in terms of accuracy, consistency, and scalability, especially as production volumes increase and market demands become more stringent. Wet floating, in particular, is useful as a basic initial sorting method but is constrained by moisture absorption, low recall, and the need for additional drying steps. In contrast, the dry SG method at a threshold of 0.9 provides a more balanced performance, offering a practical improvement over wet floating by combining good precision, accuracy, and recall.

However, this study clearly demonstrates that machine learning (ML) represents a transformative advancement in macadamia quality classification. By leveraging easily measurable physical and image-based features such as weight, size, and shell color, ML models can overcome the key limitations of conventional techniques. Among the models tested, Random Forest achieved the highest performance with 92.06% accuracy, 94.44% precision, and 91.07% recall, outperforming both wet floating and dry SG methods. Beyond its predictive power, ML offers additional advantages including automation, scalability, objectivity, and real-time processing capabilities making it particularly well suited for modern postharvest operations. The findings of this research highlight the potential of integrating machine learning technologies into existing processing lines to enhance classification accuracy, improve operational efficiency, and ensure product consistency. As technological solutions become more affordable and accessible, the adoption of models such as Random Forest is likely to set new standards in quality assurance and economic returns for macadamia producers worldwide. Furthermore, this work provides a methodological foundation that can be extended to other high-value crops where quality sorting plays a critical role in value chain management.

Future research should focus on validating the model with larger and more diverse datasets collected from different regions, orchards, and harvest seasons to improve its robustness and generalizability. Additional studies could also explore real-time deployment of ML-based sorting systems at the industrial scale, integration with other sensing technologies, and cost–benefit analyses to support adoption in commercial settings. Ultimately, the incorporation of intelligent, data-driven technologies into agricultural processing represents not only a technological shift but also a strategic step toward more sustainable, efficient, and competitive macadamia production.

**Funding:** This study was supported by Thammasat University Research (Grant number: TUFT 61/2566).
**Institutional Review Board Statement:** Not applicable.
**Transparency:** The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
**Competing Interests:** The authors declare that they have no competing interests.
**Authors' Contributions:** All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
**Disclosure of AI Use:** The author used OpenAI's ChatGPT (GPT-4) to edit and refine the wording of the Introduction and Literature Review. All outputs were thoroughly reviewed and verified by the author.

## REFERENCES

[1] W. Hu, M. Fitzgerald, B. Topp, M. Alam, and T. J. O'Hare, "A review of biological functions, health benefits, and possible de novo biosynthetic pathway of palmitoleic acid in Macadamia nuts," *Journal of Functional Foods*, vol. 62, p. 103520, 2019. https://doi.org/10.1016/j.jff.2019.103520

[2] C.-Y. Chen, W. Yang, S.-L. Zhao, and Y.-L. Nie, "Study on primary processing technology and equipment status of Macadamia," *Food and Fermentation Sciences & Technology*, vol. 57, no. 5, pp. 104–112, 2021.

[3] Y. Wei *et al.*, "Quality damage and control measures of Macadamia nuts during primary processing," *Journal of Food Processing and Preservation*, vol. 2025, no. 1, p. 1205390, 2025. https://doi.org/10.1155/jfpp/1205390

[4] P. Srisawang *et al.*, "Phytochemical contents and anti-colon cancer activities of Macadamia integrifolia Maiden & Betche leaves and husks extracts," *Natural and Life Sciences Communications*, vol. 23, no. 2, p. e2024017, 2024. https://doi.org/10.12982/NLSC.2024.017

[5] N. M. D. Buthelezi, L. S. Magwaza, and S. Z. Tesfay, "Postharvest pre-storage processing improves antioxidants, nutritional and sensory quality of Macadamia nuts," *Scientia Horticulturae*, vol. 251, pp. 197-208, 2019. https://doi.org/10.1016/j.scienta.2019.03.026

[6] X. Shuai *et al.*, "Comparative study on the extraction of Macadamia (Macadamia integrifolia) oil using different processing methods," *LWT*, vol. 154, p. 112614, 2022. https://doi.org/10.1016/j.lwt.2021.112614

[7] G. P. Baitu, O. A. A. Gadalla, and Y. B. Öztekin, "Traditional machine learning-based classification of cashew kernels using colour features," *Tekirdağ Ziraat Fakültesi Dergisi*, vol. 20, no. 1, pp. 115-124, 2023. https://doi.org/10.33462/jotaf.1100782

[8] Z. Ünal and H. Aktaş, "Classification of hazelnut kernels with deep learning," *Postharvest Biology and Technology*, vol. 197, p. 112225, 2023. https://doi.org/10.1016/j.postharvbio.2022.112225

[9] N. Erbaş, G. Çınarer, and K. Kılıç, "Classification of hazelnuts according to their quality using deep learning algorithms," *Czech Journal of Food Sciences*, vol. 40, no. 3, pp. 240-248, 2022. https://doi.org/10.17221/21/2022-CJFS

[10] B. Huang *et al.*, "Applications of machine learning in pine nuts classification," *Scientific Reports*, vol. 12, no. 1, p. 8799, 2022. https://doi.org/10.1038/s41598-022-12754-9

[11] Y. Han, Z. Liu, K. Khoshelham, and S. H. Bai, "Quality estimation of nuts using deep learning classification of hyperspectral imagery," *Computers and Electronics in Agriculture*, vol. 180, p. 105868, 2021. https://doi.org/10.1016/j.compag.2020.105868

[12] 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, 2024. https://doi.org/10.1016/j.scienta.2024.113021

[13] Y. Gulzar, Z. Ünal, S. Ayoub, and F. A. Reegu, "Exploring transfer learning for enhanced seed classification: Pre-trained Xception model. In Cavallo, E., Auat Cheein, F., Marinello, F., Saçılık, K., Muthukumarappan, K., Abhilash, P.C. (Eds.)," presented at the 15th International Congress on Agricultural Mechanization and Energy in Agriculture. ANKAgEng 2023. Lecture Notes in Civil Engineering, Springer, Cham. https://doi.org/10.1007/978-3-031-51579-8_14, 2024, pp 137–147.

[14] D. R. Harris, J. A. Seberry, R. B. H. Wills, and L. J. Spohr, "Effect of fruit maturity on efficiency of 1-methylcyclopropene to delay the ripening of bananas," *Postharvest Biology and Technology*, vol. 20, no. 3, pp. 303-308, 2000. https://doi.org/10.1016/S0925-5214(00)00150-2

[15] M. G. Nelson, D. Lelinski, and S. Gronstrand, *Design and operation of mechanical flotation machines. In D. Malhotra, P. R. Taylor, E. Spiller, & M. LeVier (Eds.), Recent advances in mineral processing plant design.* Englewood, CO: Society for Mining, Metallurgy, and Exploration, 2009, pp. 168–189.

[16] L. W. McKeen, *The effect of radiation on properties of polymers.* New York: William Andrew, 2020.

[17] R. L. Mason and R. B. H. Wills, "Evaluation of the use of specific gravity as an index of the quality of Australian Macadamia nuts," *Food Australia Journal*, vol. 35, pp. 245–248, 1983.

[18] J. T. Ballinger and G. J. Shugar, *Chemical technicians' ready reference handbook*, 5th ed. New York: The McGraw-Hill Companies, 2011.

[19] E. O. Afoakwa, *Chocolate science and technology*, 2nd ed. Chichester, UK: Wiley-Blackwell, 2016.

[20] R. N. Subedi, "Comparative analysis of dry and wet processing of coffee with respect to quality in Kaure District, Nepal ", Master's Thesis, Wageningen University, The Netherlands, 2010.

[21] N. Cetin, M. Yaman, K. Karaman, and B. Demir, "Determination of some physicomechanical and biochemical parameters of hazelnut (Corylus avellana L.) cultivars," *Turkish Journal of Agriculture and Forestry*, vol. 44, no. 5, pp. 439-450, 2020. https://doi.org/10.3906/tar-1905-115

[22] M. Weinert, M. Whitten, and S. Andreoli, "Growers guide to kernel quality testing," Retrieved: https://marquis.com/wp-content/uploads/Growers-Guide-to-Kernel-Quality-Testing-090622-002.pdf, 2022.

[23] R. A. Stephenson, E. C. Gallagher, and V. J. Doogan, "Macadamia responses to mild water stress at different phenological stages," *Australian Journal of Agricultural Research*, vol. 54, no. 1, pp. 67-75, 2003. https://doi.org/10.1071/AR02108

[24] T. G. Smit, N. J. Taylor, and S. J. E. Midgley, "The seasonal regulation of gas exchange and water relations of field grown Macadamia," *Scientia Horticulturae*, vol. 267, p. 109346, 2020. https://doi.org/10.1016/j.scienta.2020.109346

[25] N. N. Aruwajoye, O. O. Olarewaju, A. E. Oluwalana-Sanusi, A. Mditshwa, L. S. Magwaza, and S. Z. Tesfay, "Accelerating abscission of Macadamia nuts using ethephon: Are there implications for nut quality?," *The Journal of Horticultural Science and Biotechnology*, vol. 100, no. 2, pp. 153-163, 2024. https://doi.org/10.1080/14620316.2024.2434614

[26] S. Y. Rogiers, K. P. Quinlan, and J. D. Bright, "Challenges to optimal Macadamia (Macadamia spp.) kernel quality in a changing climate," *Functional Plant Biology*, vol. 52, no. 4, p. FP24218, 2025. https://doi.org/10.1071/FP24218

[27] D. D. Du Preez, "Studies on Macadamia nut quality," Master's Thesis, Stellenbosch University, South Africa. AGRIS, 2015.