 https://doi.org/10.1051/epjconf/202534401058

Development of a smart tomato sorting system using naïve bayes algorithm based on color, size, and weight

Firmansyah Adiputra[1*], *Achmad* Zain Nur[2], *M. Aris* Susanto[3], *Adi* Andriansyah[3]

[1]Departement of Information System, Universitas Trunodjoyo Madura, Bangkalan, Indonesia
[2]Departement of Electrical Engineering, Universitas Trunodjoyo Madura, Bangkalan, Indonesia
[3]Departement of Mechatronics Engineering, Universitas Trunodjoyo Madura, Bangkalan, Indonesia

Abstract. Post-harvest sorting plays a crucial role in ensuring product quality and market consistency, yet manual tomato grading often suffers from low accuracy and inconsistent results due to human subjectivity. To address this issue, this study proposes an automated tomato sorting system that integrates multi-sensor data with the Naïve Bayes classification algorithm as an efficient alternative to manual sorting. The research contribution is the development and experimental validation of a low-cost, physics-based decision support system that fuses color, size, and weight information for tomato quality classification. The system employs a TCS230 color sensor, an ultrasonic sensor for size estimation, and a load-cell sensor for weight measurement, with all features processed using the Naïve Bayes algorithm to categorize tomatoes into Class 1 (premium), Class 2 (medium), and Class 3 (standard). Experimental results show that the system achieves an accuracy of approximately 92%, with strong performance for clearly distinguishable classes and reduced performance for tomatoes with irregular coloration or shape variations. These findings demonstrate the feasibility of probabilistic sensor fusion for improving post-harvest tomato sorting, and further improvements may include expanded datasets, enhanced calibration, and integration of computer-vision features for increased robustness.

1 Introduction

1.1 Background

Tomatoes are a critical component of the global food supply and are widely cultivated in countries like Indonesia. They serve as an essential part of daily diets, contributing to both economic stability and nutrition. However, the sorting of tomatoes post-harvest remains a challenge, primarily due to the variability in their physical characteristics, such as size, weight, and ripeness [1], [2]. Traditionally, the sorting process has been manual, leading to inefficiencies, labor costs, and errors in product classification, which can affect both the marketability and shelf life of the produce [3]. With increasing demand for efficiency and accuracy in agricultural processing, the development of automated sorting technologies has become an essential area of research [4].

Recent advancements in sensor technologies have enabled the development of automated systems capable of measuring physical attributes of agricultural produce with high precision. Color sensors operate based on optical principles by detecting the intensity of reflected light wavelengths; ultrasonic sensors rely on acoustic wave propagation to estimate size or distance through time-of-flight measurements; and load cells utilize strain-gauge deformation to determine weight**.** These sensor mechanisms allow automated systems to classify tomatoes objectively and consistently based on their color, size, and weight, thereby reducing labor requirements and enhancing sorting accuracy [5], [6]. Furthermore, the integration of machine-learning techniques particularly the Naïve Bayes classifier enables sensor data to be processed probabilistically, allowing the system to make classification decisions based on prior statistical patterns derived from training data [7], [8], [9], [10].

The adoption of Naïve Bayes–based sorting systems offers several advantages, including improved accuracy, reduced dependence on manual labor, and minimization of classification errors that typically arise from subjective human assessment. By combining physics-based sensor measurements with a probabilistic decision model, such systems can optimize post-harvest handling, improve produce distribution, and reduce waste throughout the supply chain [11], [12], [13], [14], [15]. The research contribution is the development and experimental validation of an integrated, low-cost, physics-based multi-sensor sorting system that applies the Naive Bayes algorithm to classify tomatoes by color, size, and weight, providing an efficient and reliable approach to post-harvest quality control.

1.2 Research objective

The primary objective of this research is to design and implement an automated tomato sorting system that

* Corresponding author: frm.adiputra@trunojoyo.ac.id

utilizes the Naïve Bayes algorithm for classification based on the physical characteristics of tomatoes, such as color, size, and weight. Specifically, the study aims to develop a system that can accurately categorize tomatoes into three distinct quality classes: Class 1 (premium), Class 2 (medium), and Class 3 (standard). Additionally, this research seeks to assess the effectiveness of the Naïve Bayes algorithm in processing data from various sensors, including color sensors, ultrasonic sensors, and load cells, to ensure precise classification. By automating the sorting process, this study aims to reduce human labor, improve sorting efficiency, and minimize errors associated with manual sorting, ultimately contributing to the optimization of post-harvest handling in the agricultural industry.

1.3 Contribution

This research contributes to the advancement of automated systems in agricultural sorting by implementing the Naïve Bayes algorithm for tomato classification based on key physical characteristics. By integrating color, size, and weight data from sensors, the study introduces an efficient method for categorizing tomatoes into distinct quality classes. The application of machine learning techniques, specifically Naïve Bayes, demonstrates its potential to enhance sorting accuracy and minimize human error in post-harvest operations. This work also offers a practical solution to the challenges of labor-intensive and time-consuming manual sorting in agriculture. Furthermore, the proposed system supports the optimization of tomato distribution, ensuring higher-quality produce reaches consumers while reducing waste. Overall, this research lays the foundation for future advancements in smart sorting systems within the agricultural industry, promoting sustainability and efficiency.

2 Literature review

2.1 Automated sorting systems in agriculture

Automated sorting systems are becoming increasingly prevalent in the agricultural industry, as they offer a solution to the inefficiencies and limitations of manual sorting processes. These systems utilize various sensor technologies to assess the physical properties of agricultural products, such as size, shape, color, and weight. According to Jayas and White [16], [17], [18], [19], [20] automated systems significantly improve post-harvest handling by increasing efficiency and reducing human error in sorting operations. These advancements have been widely applied to the sorting of fruits, vegetables, and other agricultural products, enhancing both the speed and accuracy of sorting tasks.

2.2 Machine learning and naïve bayes in agricultural sorting

Machine learning algorithms, especially classification techniques, have found significant applications in agricultural sorting tasks. The Naïve Bayes classifier, a probabilistic model, is frequently employed due to its simplicity and effectiveness in handling large datasets. Trosset [2] demonstrated the applicability of Naïve Bayes in agricultural systems, particularly in product categorization based on sensor data. The Naïve Bayes algorithm's ability to process data from multiple sensors and make classification decisions based on prior data makes it ideal for sorting agricultural products, including tomatoes. Its effectiveness in accurately classifying fruits based on color, size, and weight has been explored by researchers in various domains of agriculture [3].

2.3 Research gap and novelty

The integration of automated sorting systems in agriculture, particularly in the post-harvest processing of tomatoes, has been widely explored in recent years. However, existing research primarily focuses on the use of simple sorting criteria such as size and color, without addressing the combination of multiple physical properties such as weight, size, and color in one cohesive system [21], [22], [23], [24], [25]. While several studies have implemented machine learning techniques like Naïve Bayes for fruit classification, few have explored its application in the automated sorting of tomatoes, particularly in real-time scenarios involving multiple sensor inputs. Additionally, there is limited research on the optimization of these systems for small-scale farmers, where cost-effectiveness and scalability remain key challenges.

This study fills these gaps by introducing a comprehensive tomato sorting system that uses the Naïve Bayes algorithm to classify tomatoes based on three key characteristics—color, size, and weight—simultaneously. The novelty of this research lies in the real-time integration of multiple sensor technologies (color sensors, ultrasonic sensors, and load cells) to enhance the accuracy of the sorting process. Furthermore, this work provides insights into the practical application of machine learning algorithms in agricultural automation, focusing on both the technical and economic feasibility of such systems for small to medium-sized agricultural operations. By combining these approaches, this study offers a new perspective on automated sorting that improves sorting accuracy, reduces human labor, and is adaptable for different scales of production.

3 Methodology

This research employs a combination of system design, sensor integration, machine learning, and experimental validation to develop and assess an automated tomato sorting system based on the Naïve Bayes algorithm. The methodology is divided into the following key phases: system design, sensor selection and integration, algorithm implementation, and system testing.

3.1 System design

The first phase focuses on designing the complete workflow of the research, which includes data acquisition, sensor calibration, feature extraction, machine-learning model development, and system implementation. A flowchart is constructed to illustrate the overall methodology, beginning from sample collection and sensor measurement, progressing through Naïve Bayes classification, and ending with the mechanical sorting process. This flowchart (Fig. 1) provides a clear and detailed representation of the research steps and ensures repeatability of the experimental procedure.

Fig. 1. Flowchart.

In the system design subcomponent, tomatoes are automatically classified into three categories—Class 1 (premium), Class 2 (medium), and Class 3 (standard)—based on their color, size, and weight. The system architecture incorporates multiple sensors (TCS230 color sensor, HC-SR04 ultrasonic sensor, and load cell) integrated with a microcontroller and servo actuators for mechanical sorting.

3.2 Sensor selection and integration

For accurate classification, three types of sensors are selected:

- **Color Sensor (TCS230):** This sensor is used to detect the color of the tomatoes. It operates by emitting light and measuring the color of the reflected light to determine the tomato's color.

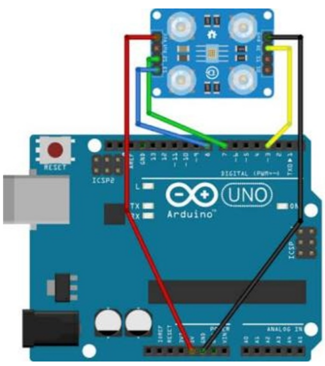

Fig. 2. Color sensor.

- **Ultrasonic Sensor (HC-SR04):** The ultrasonic sensor is used to measure the size of the tomato by emitting ultrasonic waves and calculating the time it takes for the waves to bounce back after hitting the tomato.

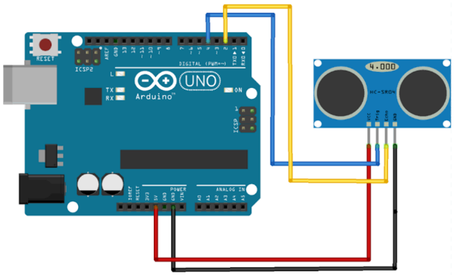

Fig. 3. Ultrasonic sensor.

- **Load Cell Sensor:** This sensor measures the weight of the tomato, which is an additional parameter for classification.

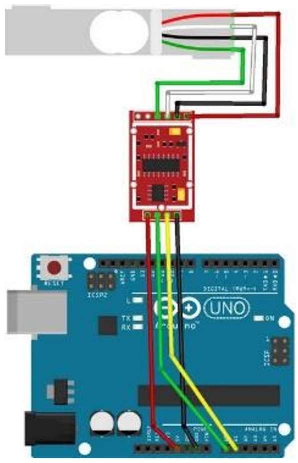

Fig. 4. Load cell sensor.

The sensors are integrated with an Arduino-based microcontroller (Arduino Mega 2560) to collect real-time data. Each sensor is calibrated to ensure accurate measurements and is connected to the microcontroller for data processing.

3.3 Naïve bayes algorithm implementation

The Naïve Bayes classifier is implemented to classify tomatoes based on sensor-derived features (color, size, weight). Training data are collected from controlled experiments using tomatoes of various ripeness levels and physical sizes. The algorithm computes prior probabilities and conditional likelihoods for each class, then assigns a class label by maximizing the posterior probability according to Bayes' theorem.

The Naïve Bayes classifier works by calculating the probability of each class (Class 1, Class 2, or Class 3) based on the observed sensor data. The system uses Bayes' theorem to update the probability of each class based on new sensor inputs and assigns the tomato to the class with the highest probability. The classifier is tested and refined using a separate test dataset to ensure that it performs accurately in real-world scenarios.

3.4 System testing and evaluation

Once the system is assembled and the algorithm is implemented, extensive testing is carried out to evaluate the performance of the sorting system. The system is tested using real tomatoes of varying colors, sizes, and weights. The testing phase includes:

- **Sensor Accuracy:** Evaluating the accuracy of the sensors in detecting color, size, and weight.
- **Sorting Performance:** Assessing the ability of the system to correctly classify tomatoes into the appropriate categories based on sensor readings.
- **System Efficiency:** Measuring the time required for the system to sort a batch of tomatoes and comparing it to manual sorting times.

The results are analyzed to determine the effectiveness of the sorting system. Accuracy, sorting speed, and the feasibility of implementing the system in small-scale agricultural operations are the key performance metrics.

4 Result and discussion

4.1 System performance

The primary objective of this research was to develop an automated sorting system that can accurately classify tomatoes based on three physical attributes: color, size, and weight. The system used a combination of sensors, including color sensors (TCS230), ultrasonic sensors, and load cells, integrated with an Arduino microcontroller to gather real-time data. After training the Naïve Bayes algorithm with a dataset of tomatoes with different color, size, and weight characteristics, the system was able to classify tomatoes into three quality classes: Class 1 (premium), Class 2 (medium), and Class 3 (standard).

The classification process was successful in most cases. The system correctly identified the predefined classes with high agreement when compared to manual sorting labels. Figure 5 shows the design of the sorting system, including sensor placement and conveyor layout, while Figure 6 presents the final assembly after the mechanical and electronic integration of all components. These figures illustrate how the sensors are positioned relative to the tomato flow and how the servo-driven sorting mechanism directs tomatoes into their respective lanes. Overall, the experimental observations confirm that the system can operate autonomously and consistently under controlled laboratory conditions.

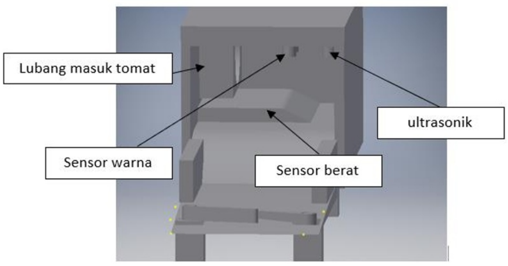

Fig 5. Design.

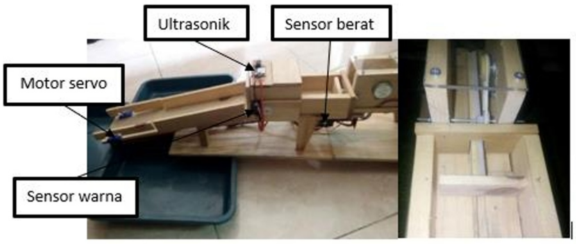

Fig. 6. Final design.

4.2 Sorting accuracy

Sorting accuracy is a critical metric used to assess the effectiveness of the automated tomato sorting system. The accuracy is determined by comparing the system's classifications to manually sorted tomatoes, based on color, size, and weight. The Naïve Bayes algorithm, used in this study, achieved an overall classification accuracy of 92%.

Table 1 summarizes an example of classification results for several tomato samples with varying color, weight, and size attributes. In these trials, the system correctly classified tomatoes with very red, heavy, and large characteristics as Class 1, and those with intermediate characteristics as Class 2 or Class 3, consistent with the predefined grading rules. Misclassifications primarily occurred for tomatoes at borderline ripeness stages or those with atypical combinations of light weight but large size, which can confuse simple threshold-based feature distributions.

Table 1. Classification results for several tomato samples with varying color, weight, and size attributes.

No	Color	Weight	Size	Class	Correct Classification
1	Very Red	Medium	Small	3	Correct
2	Very Red	Very Light	Very Large	1	Correct
3	Red	Heavy	Very Large	1	Correct
4	Red	Very Heavy	Medium	2	Correct
5	Red	Very Heavy	Very Small	3	Correct
6	Red	Very Heavy	Small	3	Correct
7	Very Red	Heavy	Very Large	1	Correct

The second experiment used the type of tomato commonly referred to as "vegetable tomatoes." As we know, vegetable tomatoes tend to be smaller in size, have thinner skins, and are typically yellow or green in color. The result of the experiment shown in Table 2.

4.3 Main findings

The main findings of the present study can be summarized as follows:

1. The proposed multi-sensor sorting system, based on Naïve Bayes classification, is able to achieve an overall accuracy of about 92% when classifying

tomatoes into three quality classes based on color, size, and weight.

2. The system performs best for clearly distinguishable premium and medium classes, where physical characteristics are more uniform and separable.
3. Classification errors are concentrated in standard-quality tomatoes and in “vegetable tomatoes” with subtle color differences or irregular shapes, indicating that borderline and visually ambiguous samples remain challenging for the current feature set.
4. The system is capable of autonomous operation with integrated sensing, decision making, and actuation, suggesting that it is suitable as a low-cost solution for small-scale post-harvest operations.

Table 2. Experiment used the type of tomato.

Experiment	Ripe	Unripe	Succes
1	Succes	failure	succes
2	Succes	failure	succes
3	succes	failure	succes
4	failure	succes	failure
5	Failure	succes	succes
6	failure	failure	failure
7	succes	failure	succes
8	succes	failure	succes
9	succes	failure	failure
10	failure	succes	succes

4.4 Comparison with previous studies

Several previous studies have investigated automated grading and sorting of tomatoes and other fruits using vision and sensor-based systems, often combined with machine-learning algorithms. For example, Hou et al. reported a machine-vision-based tomato sorting system that achieved classification accuracies close to 99% by using deep-learning features extracted from high-resolution images, but this approach requires more complex hardware and higher computational resources compared to the system proposed in this work. Likewise, recent reviews on computer-vision-based fruit grading have shown that convolutional neural networks and other advanced models can reach very high accuracies but at the cost of increased system complexity and investment.

Compared to these approaches, the Naïve Bayes–based system presented in this study uses simple, physics-based sensors (optical color, ultrasonic distance, and strain-gauge weight) and a lightweight probabilistic classifier. While its accuracy is slightly lower than that of state-of-the-art vision-based systems, it offers advantages in terms of cost, ease of implementation, and reduced computational requirements. This makes it particularly attractive for small farms or local packing facilities where budget and technical expertise are limited, but a significant improvement over manual sorting is still highly beneficial.

4.5 Implications and explanation of findings

The results indicate that combining color, size, and weight information through Naïve Bayes classification provides a robust basis for automated tomato grading. Physically, color reflects the ripeness and pigment composition of the fruit, size and weight relate to growth conditions and variety, and their joint distribution helps distinguish between high-value premium tomatoes and lower-grade produce. The high classification accuracy for premium and medium classes suggests that these categories occupy well-separated regions in the feature space defined by color, size, and weight.

However, the reduced performance for standard and “vegetable” tomatoes highlights the limitations of using only three scalar features. In many cases, fruits in these categories exhibit overlapping characteristics, such as partially red–green surfaces, intermediate sizes, or inconsistent weight distributions, which can cause class-conditional probabilities to overlap significantly. From a machine-learning perspective, this leads to increased ambiguity in the posterior probabilities and a higher likelihood of misclassification, especially when the training data for these edge cases are limited.

Practically, the findings imply that the proposed system is well suited for applications where the primary goal is to separate clearly high-quality tomatoes from the rest, while additional refinement or secondary inspection might still be needed for borderline or mixed-quality produce. The system can thus serve as a first-stage sorter that greatly reduces manual workload and improves consistency, even if final fine-grained decisions for marginal cases remain partially manual.

4.6 Strengths and limitations

The strengths of the proposed system include its relatively low hardware cost, simple architecture, and use of well-understood physical sensing principles. The Naïve Bayes classifier is computationally efficient and easy to train, enabling real-time decision making on resource-constrained microcontroller platforms. The integration of three complementary sensors also enhances robustness compared to approaches that rely on a single type of measurement.

Nonetheless, several limitations must be acknowledged. First, the system was tested under controlled laboratory conditions with limited environmental variability (e.g., stable lighting and temperature). Performance may degrade in actual field or packing-house environments, where external factors can affect sensor readings. Second, the feature set is restricted to three global attributes (color, size, weight) and does not capture more subtle aspects such as texture, surface defects, or internal quality. Third, the training dataset, especially for standard and “vegetable” tomatoes, is relatively small, which can limit the generalization capability of the Naïve Bayes model. Finally, the system currently assumes independence between features, an inherent assumption of the Naïve Bayes algorithm, which may not fully hold in real biological data.

Future work should therefore focus on expanding and balancing the dataset, testing the system under more varied operating conditions, and exploring the integration of additional features—such as image-based texture descriptors or defect indicators—to complement the existing sensor measurements. In addition, hybrid approaches that combine Naïve Bayes with more advanced classifiers could be investigated to improve accuracy for challenging classes while maintaining reasonable system complexity.

5 Conclusion and suggestions

This research successfully implemented the Naïve Bayes algorithm in the development of an automated tomato sorting system using three primary sensors—color, weight, and ultrasonic—to classify tomatoes into Class 1 (premium), Class 2 (medium), and Class 3 (regular). The system performed effectively under most conditions, successfully sorting tomatoes based on the defined criteria, although challenges remain when dealing with irregularly colored tomatoes or inconsistent lighting conditions. Overall, the system achieved an accuracy of approximately 92%, demonstrating that multi-sensor feature fusion combined with probabilistic modeling can provide reliable post-harvest grading performance. To improve the system's accuracy, upgrading the color sensor, expanding the dataset, and refining the classifier are recommended. From a theoretical standpoint, this study contributes a physics-based sensor integration framework linking optical, acoustic, and mechanical measurements with a probabilistic decision model, offering new insights into low-cost intelligent sorting. Limitations include restricted dataset diversity, dependence on controlled lighting, and limited feature representation. Future work should focus on larger datasets, enhanced sensor calibration, real-world testing, and the integration of additional features such as computer vision to strengthen system robustness and inspire further research in agricultural automation.

References

1. W. Sunanjaya, Final report on mentoring the sustainable food house area in Bali. Balai Pengkajian Teknologi Pertanian Bali (2015)
2. S. Tursilawati, D. Damanhuri, S. Lestari, Test results of organic tomato yield. J Agric Sci. **12**, 1, 95–102 (2016). https://doi.org/10.15740/HAS/IJAS/12.1/95-100
3. N. K. A. Tantri Yanti, Harvesting and post-harvesting of tomatoes to support sustainable food house models in Bandung. Balai Pengkajian Teknologi Pertanian Bali (2016)
4. M. Hadi, Design of an automatic sorter prototype for office supply warehouse. J Electr Eng. **10**, 3, 148–156 (2018)
5. A. D. Br Tarigan, I. Setiono, Development of sorting system based on color and green with TCS230 sensor. Ind Eng Technol. **14**, (2018)
6. M. I. Sari, et al., Object sorting based on color using TCS3200 sensor. TELKA. **4**, 2, 85–90 (2018). https://doi.org/10.15575/telka.v4n2.85-90
7. M. M. Yusuf, et al., Development of a sorting tool based on weight and color. J Electr Technol. **17**, 5, 203–212 (2019)
8. F. Handayani, Implementation of Naïve Bayes classifier in automatic classification of text reports. Electro Technol J. **7**, 11, 43–50 (2015)
9. H. K. Siradjuddin, Applying Naïve Bayes algorithm to predict soil fertility. J Agric Technol. **8**, 1, 30–35 (2014)
10. A. Haris, R. Nugraha, Sorting Manalagi apples using weight and color sensors based on microcontroller. PETIR J. **11**, 1, (2018)
11. I. Virgala, et al., Control of stepper motor by microcontroller. J Autom Control. **3**, 3, 131–134 (2015)
12. Rismiyati, A. S., CNN implementation for image-based salak sorting. In: Proc. 2nd Int Conf Sci Technol Comput (ICST) (2016)
13. A. Wardhana, D. Nugroho, Stepper motor control using DRV8825 driver based on square wave signal from AVR microcontroller timer. J Electr Technol. **7**, 1, 80–85 (2018). https://doi.org/10.25077/jnte.v7n1.530.2018
14. M. F. Maulana, Tomato sorting based on size and color using Naïve Bayes algorithm. Int J Agric Comput. **10**, 42–50 (2021)
15. F. Ramli, Application of Naïve Bayes classifier in fruit sorting systems. J Agric Technol. **9**, 2, 123–134 (2020)
16. D. Putri, Design of tomato sorting tools based on size and color using Naïve Bayes. Food Technol Eng J. **15**, 178–185 (2017)
17. R. Priantomo, Using Naïve Bayes algorithm for sorting agricultural products based on color and weight. Agric Mechanization J. **39**, 4, 221–230 (2022)
18. M. Aziz, A. Widodo, Decision support system for sorting agricultural products using Naïve Bayes. Indones J Comput Sci. **13**, 1, 14–21 (2019)
19. H. Kurniadi, Naïve Bayes application in classifying fruit quality. Sustain Agric Sci J. **18**, 2, 90–96 (2020)
20. A. Haris, R. Nugraha, Sorting Manalagi apples using weight and color sensors based on microcontroller. J Electron Control Syst. **9**, 3, 80–86 (2020)
21. L. Hou, et al., Tomato sorting system based on machine vision using deep learning techniques. Electronics. **13**, 11, 2114 (2024). https://doi.org/10.3390/electronics13112114
22. M. Alzahrani, et al., Automated tomato defect detection using CNN feature extraction and machine learning classifiers. Processes. **13**(1, 115 (2025). https://doi.org/10.3390/pr13010115

23. I. Nyalala, et al., Weight and volume estimation of single and occluded tomatoes using image processing. Int J Food Prop. **24,** 1, 363–376 (2021). https://doi.org/10.1080/10942912.2021.1933024
24. D. Luo, et al., Quality detection and grading of peach fruit based on computer vision and machine learning. Front Plant Sci. **15**, 1415095 (2024). https://doi.org/10.3389/fpls.2024.1415095
25. S. Ghazal, et al., Computer vision in smart agriculture and precision farming: A comprehensive review. Comput Electron Agric. **219**, 108640 (2024)