Weight and volume estimation of single and occluded tomatoes using machine vision

Innocent Nyalala, Cedric Okinda, Qi Chao, Peter Mecha, Tchalla Korohou, Zuo Yi, Samuel Nyalala, Zhang Jiayu, Liu Chao, and Chen Kunjie

Department of Agricultural Machinery, College of Engineering, Nanjing Agricultural University, Jiangsu, P.R. China; Faculty of Agriculture, Department of Crops, Horticulture and Soil Sciences, Egerton University, Njoro, Kenya

ABSTRACT
The fundamental characteristics of agricultural products are appearance, size, and weight, which affect their market value, consumer preference, and choice. Thus, food and agricultural industries seek rapid, simple, and nondestructive approaches to assess real-time measurements at the post-harvest stage before packaging for the consumer market. While sorting and grading may be performed by humans, it is unreliable, time-consuming, complicated, subjective, onerous, expensive, and easily influenced by surroundings. Therefore, an astute sorting and grading method for tomato fruit is required. We evaluated two tomato configurations on a conveyor belt: single tomatoes (no occlusion) and multi-tomatoes (partially occluded). We used polygon approximation for concave and convex point extraction algorithms to segment the occluded tomatoes. We developed seven models for regression using single-tomato image features. The Bayesian regularization artificial neural network outranked all the trained models in weight estimation with a root-mean-square error (RMSE) of 1.468 g and R² of 0.971. For volume estimation, the RBF SVM had the best performance with R² of 0.982 and RMSE of 1.2683 cm³. It is feasible to implement a proposed system as a noninvasive in-line sorting technique for tomatoes.

INTRODUCTION
Tomatoes are downright one of the world’s most highly produced and consumed fruits. It is an essential horticultural plant and the most exported fleshy fruit globally.[1–3] China produced 35% of the world’s total production of 180 million tons of tomatoes in 2019, which was approximately 62 million tons making it the top producer, followed by India, Turkey, and the United States of America, respectively.[4] Tomatoes are native to South America (Mexico) and Central America and can now be grown worldwide, especially in temperate climates and greenhouses. Thus, there is a need to automate the sorting and grading techniques for tomato processing and quality inspection before consumption.[5]

In fruits, morphometric measurements, such as weight, size, density, surface area, and volume, are correlated and essential.[6,7] In grading and sorting of vegetables and fruits, these measurements are often used as crucial descriptors.[6] The fruit size and weight directly influence storage and transport costs, eventually influencing sales and marketing prices.[9,10] Fruit volume and density relate to fruit consistency, carbon consumption, and water content used to forecast harvesting periods.[7,11] The average density, critical in identifying hidden defects such as frost and internal damage, can be determined when the weight is combined with the volume.[9]

CONTACT Chen Kunjie kunjiechen@njau.edu.cn College of Engineering, Nanjing Agricultural University, Jiangsu, 210031, P. R. China

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
An alternative way of describing an object’s weight is through evaluating the volume. Volume is primarily defined by weight per unit mass. Weight also shows food products’ geometric properties since using a digital weighing scale is reasonably straightforward.\textsuperscript{[12]} Likewise, it is complex and challenging to measure irregular objects’ volume using geometry, especially when the surface area is uncertain. Water displacement is the most effortless, most direct, and traditional process in which a product gets entirely immersed in water, and the rise in the amount of water signifies the product.\textsuperscript{[12,13]} However, in large-scale processing, these manual processes are labor-intensive and costly, besides being invasive, unsanitary, individualized, and susceptible to error.\textsuperscript{[9,10,14,15]}

Classifications of size, volume, and weight are essential steps after harvesting in grading and sorting most agricultural products. The sorting operation helps customers effectively recognize product quality and contributes to better-coordinated distribution and delivery of agricultural products. Professionals performed quality control within the food industry until a few decades ago. The efficiency is limited, costly, inconsistent, and unreliable in the conventional approach to adjust to the rise in customer needs.\textsuperscript{[9,15,16]} Image processing is a revolutionary innovation that has seen tremendous improvement in recent years, both practically and theoretically. With the advances in machine vision technologies, numerous researchers have indicated computer vision systems in the weight and size measurement of several agricultural products as a precise, nondestructive, efficient, and non-intrusive technique.\textsuperscript{[9,10]}

Classifications of size, volume, and weight are essential steps in grading and sorting most food and agricultural products. Due to its nondestructive capability and high efficiency, machine vision has been applied to the weight (weight), volume, and size assessment in the agriculture and food industries. Computer vision systems were used to determine tomato geometric parameters, including weight and volume,\textsuperscript{[5,9,13,17]} mushrooms,\textsuperscript{[13]} strawberries,\textsuperscript{[13,17,18]} pomegranates,\textsuperscript{[19,20]} citrus fruits (lemon, orange, lime, and tangerine), poultry and products.\textsuperscript{[21]}

Besides, food products mostly go via a conveyor belt as a cluster instead of as a single object in a vast commercial production-line environment. Consequently, this poses an occlusion dilemma as a constraint on computer vision systems deployed in production facilities. Single-object assessment requires separating occluded items in a multi-object environment like harvesting and yield prediction.\textsuperscript{[22]} Separating identical connecting products, including grains\textsuperscript{[23]} and cells,\textsuperscript{[24]} have been published. No literature exists on tomato fruit splitting in a production environment. Hence, this research presents an innovative method that separates connecting tomatoes depending on the shape information analysis.

The development of an effective, precise, and accurate sorting and grading system used in an in-line production system is critical. This study’s primary objective was to develop an RGB-based sorting system for tomato volume and weight estimation. To achieve this aim, this article proposes a method of automated tomato volume and weight determination based on image processing and machine-learning techniques, where RGB images of the tomatoes were acquired from three orientations, and an efficient splitting algorithm for touching tomatoes (partially occluded) was implemented. Volume and weight estimation models were developed from the extracted feature variables. This proposed technique can be employed for tomato grading and sorting challenges as an alternative.

**MATERIALS AND METHODS**

**Tomato samples and the image acquisition system**

Three hundred tomato samples of the CM20 variety of different shapes and sizes were obtained from a farm in Nakuru County, Kenya, and included in the experiment between September 2020 and December 2020. Having inspected the ripe tomatoes manually and ensuring they had no visible damage and external defects. Each tomato was labeled, and the weight measured three times using an electronic precision scale (model TP-10000, Xiangyi Balance Instrument Equipment Co Ltd, Changsha, China) with ±1 g accuracy and the average value calculated. Volumetric measurement of
each tomato was determined by the drainage method based on the Archimedes’ principle (Lang and Thorpe, 1989) using a graduated cylinder (±1 cm³).

RGB images of tomato fruits were obtained using a digital camera (Panasonic Lumix DMC-FZ40, Panasonic Corporation, Osaka, Japan) installed 1.5 m perpendicularly above the imaging stage. The images were taken with a dark background and fluorescent lamps for lighting. The data were transferred and stored for subsequent analysis onto external storage. The marked tomato samples were put on a conveyor belt as the scanning stage (0.30 m by 0.30 m) as shown in Figure 1. The conveyor line was a lab fabrication of an online production system as shown in Figure 1 (b). The conveyor speed was set at 0.017 ms⁻¹ to synchronize with the image acquisition rate of 10 frames per minute.

Two modes for image capture defined by the tomatoes’ configuration (single and partially occluded) were performed in the imaging stage as displayed in Figure 2; (a) single-tomato images (for model creation) (M1) were acquired in the scanning stage at three different orientations. (b) Images of partially occluded tomatoes (M2). Nine images were captured for each tomato in M1, three for every orientation and position in the imaging stage. The calyx facing upward was the first orientation, the calyx facing downward second and, finally, the calyx’s longitudinal axes parallel to the imaging stage’s surface. A total of 1500 images was acquired in M1. The cluster sizes (the number of tomatoes partially occluded) in M2 were raised from 3 to 15 tomatoes per image (tomatoes’ number

**Figure 1.** Tomato image capture system.

**Figure 2.** Acquired color image with their corresponding processed image.
in each cluster was added by one tomato from 3 to 14 tomatoes per cluster). Two hundred images (50 tomatoes) were obtained for the partially occluded tomatoes (creating different patterns as in Figure 2).

**Image pre-processing and volume/weight prediction model**

This study’s objective was to implement an automated algorithm for the segmentation and prediction of tomatoes’ weight and volume under partial occlusions. First, tomatoes in M1 were used to establish weight and volume estimation models based on extracted features using pre-processed RGB images. Second, a polygon approximation-based tomato-splitting technique was employed to M2. Third, the split tomatoes’ weight and volume (after splitting) were predicted using the established model (the developed model in the first step). Using the total number of tomatoes identified in a cluster versus the count of image objects after separation, the tomato-splitting algorithm was assessed and comparative accuracy analysis of the volume after separation versus the actual volume from the water displacement method and estimated weight against the actual weight. Figure 3 displays the introduced tomato weight and volume prediction system’s algorithmic flow.

**Image preprocessing algorithm**

The raw RGB images were analyzed using the steps below:

Step 1: Due to the background and the camera being static, the image subtraction procedure was used to remove the background:

![Algorithm flow diagram](image-url)
\[ S_{x,y} = \begin{cases} 0, & \text{if} \left| B_{x,y} - T_{x-y} \right| \\ B_{x,y} & \text{otherwise} \end{cases} \] (1)

where \( S_{x,y} \) is the resultant image from background separation, the original image is \( B_{x,y} \) and the background image is \( T_{x-y} \).

Step 2: The images were smoothened by a Gaussian kernel filter:

\[ F_{x,y} = [G_{x,y}] \otimes h_{s,t} \] (2)

where the resultant smoothened image is \( F_{x,y} \), and \( h_{s,t} = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \) is the Gaussian filter kernel.

Step 3: Binarization was subsequently applied using Otsu’s method\(^{25}\) to transform the image into a binary image.

Step 4: A morphological opening procedure was done using a size 9-pixel disk structural element to eliminate holes in the image object:

\[ A_{x,y} = F_{x,y} \bullet Es = (F \oplus Es)Es = \bigcup \{ (Es)_z \mid (Es)_z \} \] (3)

After morphological opening, \( A_{x,y} \) was the resultant image, \( Es \) the structural element, and \( z \) the translation.

Step 6: Region filter was then performed based on the area to remain with only tomatoes’ images.

**Weight and volume prediction model**

For tomato weight and volume estimation, we explored seven regression models: Support Vector Regression (SVR) (linear, cubic, quadratic, and RBF (Gaussian)) and artificial neural networks (ANNs) (Levenberg–Marquardt, Bayesian regularization, and scaled conjugate gradient training algorithms). These models were established based on the features extracted from the training dataset using a 10-fold cross-validation-based parameter search and subsequently evaluated on the testing dataset. For model evaluation, the whole dataset (1500 images) was split into training (70%) and testing dataset (30%). Image identification criteria for feature extraction were dependent on manually selecting the images, provided that the tomatoes did not touch the edge of the image because of their movement while images were captured.

Features’ extraction: Any machine learning model’s precision and performance depend significantly on the model training’s feature variables.\(^{10} \) Only 2D features, i.e., area \( Y_A \), perimeter \( Y_P \) and eccentricity \( Y_E \), major-axis length \( Y_1 \), minor-axis length \( Y_2 \), and radial distance \( Y_D \) were extracted from the RGB images. These extracted features are summarized in Table 1.

The \( Y_A \) (Eq. (4)) is the number of pixels in an image object or the projected area. This feature has been extensively applied in several agricultural product weight and volume estimations.\(^{9,10} \) \( Y_P \) (Eq. (5)) is the number of pixels around the closed contour of a 2D object. If the closed contour has \( N \) vertices, the contour can hence be described as \( C = \{(x_n,y_n)\}, \) where \( n = 1, 2, \ldots, N \), and \((x_1, y_1)(x_N, y_N)\), then \( Y_P \) is defined by (Eq. (6)). \( Y_E = \frac{Y_1}{Y_P} \) equals the ratio of the Eigen values \((Y_1 and Y_2)\) of the covariance matrix of a fitted ellipse.\(^{9,10,26} \) \( Y_1 \) (Eq. (7)) is the pixel distance between the major-axis endpoints. Here, the major-axis endpoint coordinates are defined by \((X_1, Y_1)\) and \((X_2, Y_2)\). The result is the Euclidean distance between the two points. \( Y_2 \) (Eq. (8)) is the pixel distance between the minor-axis endpoints. Here, the minor-axis endpoint coordinates are defined by \((M_2, N_2)\) and \((M_1, N_1)\). The result is the Euclidean distance between the two points. The \( Y_D \) (Eq. (9)) is the average distance from the center of gravity of an image to the boundary points. The radial distance is measured from the central point (centroid) of the object of each pixel \( x(n), y(n) \) on the boundary.

Regression models: ANNs are non-linear statistical models whose learning technique is a problem of function optimization that determines the optimal network conditions to reduce network failure.\(^{10,27} \) ANNs act in response to unpredicted inputs, and they require less statistical training. The learning difficulty of an ANN is a loss function reduction via an optimization algorithm. The
Consider that the contour of a 2D image \( I(x, y) \) is defined by \( S = \{ (x_n, y_n) \} \), where \( n = 1, 2, 3, \ldots, N \) and \( (x_1, y_1) = (x_N, y_N) \), when the 2D image is fitted with an ellipse \( \{ (x_1, y_1) \) and \( (x_2, y_2) \) and \( \{ (M_2, N_2) \) and \( (M_1, N_1) \) are the endpoint coordinates of the major-axis and minor-axis, respectively.

The training of ANNs ends as soon as the optimization algorithm satisfies a particular requirement or stop criterion.\(^{[9,10,27]}\) ANNs essentially classify inputs into a set of target categories. These optimizations (training) algorithms vary in speed, precision, and memory complexity. In this study, three ANNs were explored; each had a 1-hidden layer feed-forward network trained using backpropagation algorithms activated by the tangent sigmoid function (Bayesian Regularization, Scaled Conjugate Gradient, and Levenberg–Marquardt training algorithms).

SVR is an extension of the SVM employed in high dimensional spaces to solve a regression problem. SVR is considered a non-parametric statistical learning algorithm based on kernel functions.\(^{[28]}\) SVR models’ strengths are that their training is easy, are explicitly controlled, offer better generalization performance (accuracy), provide direct geometrical interpretation, lack an optimal local solution, elegant mathematical tractability, and prevent overfitting by not necessitating a massive set of training samples.\(^{[9,29,30]}\) Similarly, the SVR models rely only on a subset of the training dataset since the cost function does not consider the training points outside the margin during model construction, thereby conveniently preventing overfitting.\(^{[31]}\) SVR is typically a quadratic programming problem (QP) to distinguish support vectors from the other training data vectors.\(^{[31]}\) The kernel functions that map input data into the necessary forms solve this quadratic problem. Thus, an SVR model’s overall performance is greatly affected by the kernel function’s choice.\(^{[9,10,26,29,32,33]}\) For this study, we applied the linear (Eq. (10)), polynomial (Eq. (11)), and Gaussian (Eq. (12)) kernel functions. Table 2 displays a description of SVR kernel functions.

### Tomato splitting algorithm

Concave and convex points’ extraction: The concave point detection method was applied to fit an object boundary with a sequence of line segments (Polygon approximation). Concave points are

---

**Table 1. The extracted features’ variables.**

| Extracted features | Defining equations |
|--------------------|--------------------|
| Area | \( Y_A = \frac{1}{2} \sum_{n=1}^{N-1} (y_{n+1}x_n - x_{n+1}y_n) \) |
| Perimeter | \( Y_P = \sum_{n=1}^{N-1} (x_{n+1}, y_{n+1}) - (x_n, y_n) \) |
| Eccentricity | \( Y_e = \frac{Y_1}{Y_2} \) |
| Major-axis length | \( Y_1 = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \) |
| Minor-axis length | \( Y_2 = \sqrt{(M_2 - M_1)^2 + (N_2 - N_1)^2} \) |
| Radial distance | \( Y_d(n) = \sqrt{|X(n) - \bar{X}|^2 + |Y(n) - \bar{Y}|^2} \) |

\( n = 0, 1, \ldots, N - 1 \)

---

**Table 2. Descriptions of the kernel properties.**

| Model parameters | Kernel equations |
|------------------|------------------|
| Linear | \( k_L(x_i, x_j) = x_i^j \cdot x_j \) |
| Polynomial | \( k_P(x_i, x_j) = (1 + x_i^j \cdot x_j)^d, \ d > 0 \) |
| Gaussian | \( k_G(x_i, x_j) = \exp \left( -\frac{x_i - x_j^2}{2\sigma^2} \right), \ \sigma > 0 \) |
divided into two types, the obvious concave points and the unobvious concave points. Obvious concave points connect two tomatoes and can easily be detected by angle features since their gradients change sharply, while the gradient change in the unobvious concave point is insignificant.\textsuperscript{[35]}

Polygon approximation: Polygon approximation is used to represent the object’s contours by a sequence of dominant points. The method detects dominant points on a contour by suppression of redundant points. The advantages of using this algorithm are that it smoothens the object contours, reduces complexity and calculation time, and avoids false concave point detection.\textsuperscript{[35,36]} Suppose a sequence of extracted contour points \( = \{m_1, m_2, \ldots\} \), the collinear suppression determines its dominant points. This simply means that every contour point \( m_i \) is examined for co-linearity while being compared to the previous and next successive contour points. The point \( m_i \) is considered the dominant point if it is not located on the line connecting \( m_{i-1}(a_{i-1}, b_{i-1}) \) and \( m_{i+1}(a_{i+1}, b_{i+1}) \), and the distance \( n_i \) from \( m_i \) to the line connecting \( m_{i-1}(a_{i-1}, b_{i-1}) \) to \( m_{i+1}(a_{i+1}, b_{i+1}) \) is more significant than the pre-set threshold \( n_i > n_{th} \). This distance \( n_i \) is obtained by

\[
n_i = \sqrt{\frac{((a_i - a_{i-1})(b_{i+1}-b_{i-1}) - (b_i - b_{i-1})(a_{i+1}-a_{i-1}))^2}{(a_i - a_{i-1})^2 + (b_i - b_{i-1})^2}} \] (13)

After polygon approximation and detection of the dominant points, the concave points can be identified by the techniques suggested by Bai et al.\textsuperscript{[24]} and Zhang et al.\textsuperscript{[37]} In the method adopted by Bai et al.,\textsuperscript{[24]} the dominant point \( m_{n,i} \in M_{dom} \) is defined as concave if the concavity value of \( m_{n,i} \) is within the range of \( x_1 \) to \( x_2 \) and the line connecting \( m_{n,i} \) does not pass through the inside objects.

The concavity value of \( m_{n,i} \) is defined as the angle between lines \( (m_{n,i-1}, m_{n,i}) \) and \( (m_{n,i+1}, m_{n,i}) \) as follows:

\[
\left\{ \begin{array}{ll}
\pi - |x(m_{n,i-1}, m_{n,i}) - x(m_{n,i+1}, m_{n,i})| & \text{if } |x(m_{n,i-1}, m_{n,i}) - x(m_{n,i+1}, m_{n,i})| < \pi \\
\pi & \text{otherwise}
\end{array} \right.
\] (14)

where

\[
x(m_{n,i-1}, m_{n,i})\tan^{-1}\left(\frac{b_{n,i-1} - b_{n,i}}{a_{n,i-1} - a_{n,i}}\right)
\] (15)

and

\[
x(m_{n,i+1}, m_{n,i})\tan^{-1}\left(\frac{b_{n,i+1} - b_{n,i}}{a_{n,i+1} - a_{n,i}}\right)
\] (16)

In the technique proposed by Zhang et al.,\textsuperscript{[37]} the dominant point \( m_{n,i} \in M_{dom} \) is considered to be a concave point if \( m_{n,i} \rightarrow m_{n,i} \times m_{n,i+1} \rightarrow \) is positive:

\[
M_{con} = \left\{ m_{n,i} \in M_{dom} : m_{n,i-1}m_{n,i} \times m_{n,i}m_{n,i+1} \rightarrow > 0 \right\}
\] (17)

Both Bai et al.\textsuperscript{[24]} and Zhang et al.\textsuperscript{[37]} define the threshold value \( n_{th} \) manually. For this study, we propose to employ the approach suggested by Zafari et al.\textsuperscript{[36]} to extract both concave \( (C_x) \) and convex \( (C_v) \) points. This technique combined the approaches by Zhang et al.\textsuperscript{[37]} and Prasad et al.\textsuperscript{[38]} for parameter-independent polygon approximation and dominant point detection. In this approach, the line connecting \( m_{i-1}(a_{i-1}, b_{i-1}) \) and \( m_{i+1}(a_{i+1}, b_{i+1}) \) is first digitized, and, afterward, the threshold value \( n_{th} \) is automatically selected based on the angular distance between the digitized line slope and that of the actual line. This leads to the concave point detection being fully parameter-free.\textsuperscript{[36]} Both the \( C_x \) and \( C_v \) points were extracted for all the images under partial occlusion. For both the inner and outer contour points, the points were extracted, respectively, as seen in Figure 4. Therefore, each image was tested to have inner and outer contours based on the small-hole filling technique.

Split-line extraction: The \( C_x \) and \( C_v \) points were used for drawing the separating line in the images (splitting two partially occluded tomatoes). Splitting was carried out in three phases. First, the KNN
nearby search for each \( C_v \) point was carried out by finding its equivalent \( C_v \) points based on an Euclidean distance threshold (\( \xi \)) among the two points. If the Euclidean distance between a point \( C_{vk} \) and its neighbor \( C_{vl} \) was greater than \( \xi \), then the points were regarded as not neighbors. Second, to find the corresponding \( C_x \) points on the remaining \( C_v \) points, another KNN nearby search was conducted. Lastly, to find the corresponding \( C_x \) neighbors based on \( \xi \) of the remaining \( C_x \) points another KNN search was performed. KNN nearby search is carried out as follows: given that a point \( T \) is distinctly connected to another point \( T_j \in W_{x,y} \) (nearest) regarding the Euclidean distance among them, the nearest neighbor of \( T \) is calculated as

\[
[T, T^*] = \arg\min_{T, T_j \in W_{x,y}} \text{KNN} \{ T_j : (T_i, T_j) \in W_{x,y} \} = 1, 2, \ldots j
\]  

(18)

Figure 4. Concave and convex points extraction procedure to illustrate images with inner and outer contours (a) and (c) original image (b) and (d) polygon approximation on the outer boundary, (e) and (f) polygon approximation on the inner boundary.

Figure 5. Separation of partially occluded tomatoes using split lines for several tomato cluster sizes.
where $T^*$ is the calculated nearest neighbor of $T$. For every $C_x$ and $C_v$ points and their equivalent neighbors, a split-line was drawn linking the two points as displayed in Figure 5. Thus, detaching two touching tomatoes.

**Multi-tomato volume and weight estimation**

This paper aimed to efficiently and accurately determine each tomato’s weight and volume in a partially occluded tomato cluster based on developed volume and weight prediction models. The explored touching-tomatoes scenario results were compared in accuracy terms for predicted volume to the water displacement volume and weight estimation compared to the measured weight. Accuracies for estimated volume and weight were compared to the drainage method and measured weight, respectively, in terms of RMSE, R-squared ($R^2$), and relative error. Moreover, a dependent T-test was conducted to evaluate the variations between actual weight $W_0$, $M1$ estimated weight $W_1$ and $M2$ estimated weight $W_2$ and the difference between the drainage method volume $V_0$, $M1$ estimated volume $V_1$ and $M2$ estimated volume $V_2$. For each model, two T-test analyses were performed, i.e., for weight estimation in $M1$, $W_0$ against $W_1$ and $W_0$ against $W_2$ in $M2$, and for volume estimation $V_0$ against $V_1$ and $V_0$ against $V_2$ in $M2$. A dependent T-test contrasts the means of two similar classes using the same constant dependent variable. The data normality was calculated by the skewness value and skewness standard error values before the T-test. No data transformation was done for this research. All statistical analyses were computed using the Statistical Package for Social Sciences (IBM SPSS).

**RESULTS**

**Dataset statistical description**

In Table 3, descriptive statistics of the whole dataset are presented in aspects of central tendency metrics and variability (spread) metrics $W_0$ and $V_0$. For the 300 tomatoes, the mean weight was 163.18 g with a median of 121.50 g ranging from 36.00 g (minimum) to 507.20 g (maximum), while the mean volume was 154.69 cm$^3$ with a median of 114.00 cm$^3$ ranging from 30.00 cm$^3$ (minimum) to 505.00 cm$^3$ (maximum). The weight dataset had a skewness of 0.94 g at a standard error of 6.33 g, while the volume dataset had a skewness of 0.94 cm$^3$ and a standard error of 6.36 cm$^3$.

**Evaluation of the weight and volume estimation models**

The model parameters for the SVR and the ANN are provided in Table 4. These parameters help enhance the overall accuracy and efficiency of the model. The SVR kernel function parameters were iteratively calculated through the 10-fold-cross-validation-based model parameter search during the model training phase. Cross-validation aims to find the most accurate kernel parameters while avoiding overfitting and underfitting the model and providing an acceptable generalization capability.

| Weight | Volume |
|--------|--------|
| **Descriptive statistic** | **Value** (g) | **Descriptive statistic** | **Value** (cm$^3$) |
| Mean | 163.18 | Mean | 154.69 |
| Standard error | 6.33 | Standard error | 6.36 |
| Median | 121.50 | Median | 114.00 |
| Standard deviation | 109.68 | Standard deviation | 110.24 |
| Sample variance | 12030.70 | Sample variance | 12152.57 |
| Skewness | 0.94 | Skewness | 0.94 |
| Range | 471.20 | Range | 475.00 |
| Minimum | 36.00 | Minimum | 30.00 |
| Maximum | 507.20 | Maximum | 505.00 |
Likewise, the number of neurons in the hidden layer was adjusted to 10 (6–10-1 topology) for all the ANNs to match the computation complexity and error term by validating the model. The models were applied to the testing dataset to determine their precision in model testing. Table 5 displays the models’ accuracy in terms of $R^2$ and RMSE on the testing dataset.

The Bayesian regularization ANN had the best regression results, 0.971 $R^2$ at an RMSE of 1.468 g for weight estimation, while the linear SVM had the lowest $R^2$ 0.813 at an RMSE of 3.841 g. For volume estimation, the RBF SVM had the highest regression results of 0.982 $R^2$ at an RMSE of 1.283 cm$^3$ while the linear SVM had the lowest $R^2$ 0.835 at an RMSE of 2.917 cm$^3$. Figure 6 displays linear regression plots for the $W_0$ against estimated weight $W_1$ and $V_0$ against estimated volume $V_1$ of the best performing models.

It was also observed that the average relative error for $W_1$ in M1 was 4.78%, with a maximum and minimum of 5.04% and 0.47%, respectively. The average relative error for $V_1$ in M2 was 3.18%, with a maximum and minimum of 4.27% and 0.96%, respectively, as shown in Table 6.

### Evaluation of the tomato splitting algorithm

The efficiency of the splitting algorithm was analyzed through the splitting accuracy (object count) and the accuracy of $W_2$ and $V_2$ in comparison to $W_0$ and $V_0$ for the same tomatoes. It can be seen that the splitting technique correctly segmented all the tomato images in clusters, i.e., the image objects’ (after splitting) number was equivalent to the number of tomatoes in all the images for M2. Moreover, to evaluate the accuracy of the splitting process with cluster size, the average relative error (in the estimation of $W_2$ and $V_2$) was assessed for each cluster size as displayed in Figure 7. The results demonstrated that the introduced system’s performance was invariant to the number of tomatoes in a cluster. In the estimation of $W_2$ the maximum and the minimum mean relative errors of 2.632% and 1.324% were

**Table 4. Regression model parameters.**

| Classifier parameters | $Mdl1$ | $Mdl2$ |
|-----------------------|--------|--------|
| SVM kernel parameters  | Scale ($\sigma$) | Degree ($d$) | Scale ($\sigma$) | Degree ($d$) |
| Linear SVM            |        |        |        |        |
| $W$ estimation        | 1.048  | –      | 1.076  | –      |
| $V$ estimation        | 0.871  | –      | 0.580  | –      |
| Quadratic SVM         |        | 2      | 0.693  | 2      |
| Cubic SVM             | 0.415  | 2      | 0.623  | 2      |
| RBF SVM               | 0.402  | 3      | 0.530  | 3      |
| Bayesian regularization ANN | 6-10-1 |        | 6-10-1 |        |
| Levenberg–Marquardt ANN | 6-10-1 |        | 6-10-1 |        |
| Scaled conjugate gradient | 6-10-1 |        | 6-10-1 |        |

**Table 5. Accuracy comparison of various regression models on the testing dataset ($R$-squared ($R^2$) and RMSE).**

| Regression models            | Weight | Volume |
|------------------------------|--------|--------|
|                              | $R$-squared ($R^2$) | RMSE | $R$-squared ($R^2$) | RMSE |
| Linear SVM                   | 0.813  | 3.841  | 0.835  | 2.917 |
| Quadratic SVM                | 0.856  | 3.967  | 0.894  | 3.395 |
| Cubic SVM                    | 0.898  | 3.182  | 0.926  | 2.979 |
| RBF SVM                      | 0.969  | 1.963  | 0.982  | 1.283 |
| Bayesian regularization ANN   | 0.971  | 1.468  | 0.954  | 1.601 |
| Levenberg–Marquardt ANN      | 0.928  | 2.346  | 0.973  | 1.701 |
| Scaled conjugate gradient ANN | 0.947  | 2.479  | 0.963  | 2.074 |
reported when the cluster sizes were 8 and 6 tomatoes. In $V_2$ the maximum and the minimum mean relative errors of 3.088% and 1.435% were reported when the cluster size was 10 and 3 tomatoes.

It was established by a dependent T-test that there is no statistically significant difference ($p < 0.05$) between $V_0$ and $V_1$. There was, however, a statistically significant difference ($p < 0.05$) between $V_0$ and $V_2$. It can be seen in Table 7 that there was a mean decrease in $V_2$ of 2.822 cm$^3$ compared to $V_0$. For weight, there was no statistically significant difference ($p < 0.05$) between $W_0$ and $W_2$ and between $W_1$ and $W_2$ as seen in Table 8.

---

**Table 6.** Results of tomato estimation under the single and touching-tomato models.

| Models | $R^2$ | RMSE | Relative error (%) | $n$ |
|--------|-------|------|-------------------|-----|
|        |       |      | Minimum | Maximum | Average |
| $M1$   |       |      |         |         |         |
| $W_1$  | 0.971 | 1.468| 0.47    | 5.04    | 4.78    | 450    |
| $V_1$  | 0.980 | 1.306| 0.73    | 3.98    | 3.52    |        |
| $M2$   |       |      |         |         |         |
| $W_2$  | 0.982 | 1.283| 0.96    | 4.27    | 3.18    | 50     |
| $V_2$  | 0.989 | 1.154| 0.38    | 2.13    | 2.09    |        |

---

**Figure 6.** Linear regression plots for actual weight and volume against estimated weight and volume, (a) actual weight ($W_0$) against estimated weight for the testing dataset ($W_1$ in M1), (b) displacement method ($V_0$) against estimated volume for the testing dataset ($V_1$ in M1), (c) actual weight ($W_0$) against estimated weight for the touching tomatoes ($W_2$ in M2), and (d) displacement method ($V_0$) against estimated volume for the touching tomatoes ($V_2$ in M2).
DISCUSSION

This paper proposes a computer vision-based automated tomato volume and weight prediction system implemented in a tomato production-line system. An RGB image preprocessing algorithm was introduced to segment tomatoes from their background. Afterward, the image feature extraction algorithm was employed for the tomato images. From prior research\textsuperscript{[9,13,14]} on volume and weight determination of tomatoes, geometric shape features are generally favored because they are accurate and descriptive. Thus, in our research, tomato image features were represented by six standard shape geometric parameters $Y_A$, $Y_P$, $Y_E$, $Y_1$, $Y_2$, and $Y_D$.

The developed models are compared in Table 5, where the Bayesian regularization ANN surpassed the other ANNs and SVM models in weight estimation, and the RBF SVM performed better than the rest of the SVM, ANN models in volume estimation. This can be attributed to the need for less statistical data and better optimization of data for the Bayesian ANN and the RBF SVM’s increased ability to generalize data on the testing dataset. All of the models provided excellent estimates of weight and volume, with an accuracy of at least 0.81.

Calculation of tomato weight and volume in real time is a time-consuming and cumbersome task. The drainage approach is the most common method of measuring the volume of irregular objects or objects with different densities. For weight, tomatoes are weighed using weighing scales as the primary method to get fruits’ weight. Nevertheless, a volume-based sorting system is a cost-effective approach since it is more accurate and offers an alternative to weight sorting since fruits are usually sorted and graded with reference to size.\textsuperscript{[6,9,40]} Derived from previous researches on weight and volume and mass prediction for multiple fruits, such as tangerine, orange, lime, and lemon, the limit of agreement was set to 95% such that the agreed predicted volume ought to be below 5% error.\textsuperscript{[6,9]} This study’s results

\begin{table}[h]
\centering
\caption{The dependent $t$-test results for $V_0$, $V_1$, and $V_2$.}
\begin{tabular}{lll}
\hline
& $V_{01}$ (cm$^3$) & $V_{02}$ (cm$^3$) \\
& mean ± std & mean ± std \\
155.902 ± 110.629$^a$ & 155.129 ± 111.152$^a$ & 153.058 ± 109.493$^b$ \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{The dependent $t$-test results for $W_0$, $W_1$, and $W_2$.}
\begin{tabular}{lll}
\hline
& $W_0$ (g) & $W_1$ (g) \\
& mean ± std & mean ± std \\
163.001 ± 108.821$^a$ & 164.796 ± 110.489$^a$ & 163.365 ± 109.920$^b$ \\
\hline
\end{tabular}
\end{table}
concluded with those of Koc, Omid et al., Fellegari et al., and Nyalala et al.\textsuperscript{[6,9,40,41]} in terms of maximum and minimum errors.

In an in-line tomato production environment, tomatoes move entirely at random in clusters. Thus, in any location (partial or no occlusion) across the line, a robust weight and volume determination system is expected to predict each tomato’s volume and weight. This work introduces a touching-tomatoes separating technique for weight and volume prediction. The implemented separating method provides a remedy to the tomato occlusion problem. Algorithms used to split partially occluded objects include global minimization-based methods, watershed-based algorithms, and shape information-based algorithms. However, the watershed-based algorithms are frequently unreliable, lacking a strong gradient among objects due to occlusion.\textsuperscript{[10,24]} Shape information-based algorithms conduct the separation of occluded objects by evaluating an object contour to determine convex and concave points.\textsuperscript{[23,24]}

In shape representation, occlusion points are identified by high curvature points on an object boundary. Thus, our objective was to accurately detect occlusion points across the shape contour for split-line extraction. In contrast to Lin et al.,\textsuperscript{[25]} it was shown that extracting high curvature points centered on a curvature curve function requires the obtained curvature peaks to be transformed to be invariant to translation and rotation. However, using a set threshold on the angles between two successive contour points, concave points can be established according to Bai et al.\textsuperscript{[24]} Our research presents a robust, precise, and determinant-based approach to extract high curvature points from contour primitives for noise invariance and reduce computational complexity.

In an accurate splitting algorithm, the method’s consistency and efficiency with regard to the number of occluded objects are evaluated by the accuracy of the system developed. Figure 6 indicates that the average relative error is independent of the cluster size after the splitting technique, but only in the initial image processing procedure. Henceforth, the proposed system with an average error range of 1.324–2.632\% for weight estimation and 1.435–3.088\% can accurately estimate partially occluded tomatoes’ weight and volume of partially occluded tomatoes (tomatoes in a cluster).

However, this study had some limitations; the system is only applicable to the CM20 tomato variety. Other tomato fruit types, particularly those with large fruits, may have different weight and volume than CM20 fruits due to differences in the internal fruit structure (locule number and flesh percentage). As such, this technique requires validation on various tomato varieties and fruit types to ensure its feasibility for estimating weight and volume for other axis-symmetric fruits. Furthermore, a higher frame rate and superior computing resources are suggested to increase the proposed system’s processing capability.

**CONCLUSION**

In this article, a novel tomato weight and volume prediction method was developed for tomatoes with no and partial occlusions using image processing and machine learning techniques. The findings revealed that the weight and volume of tomato fruits could be accurately measured under two conditions: occlusion and no occlusion. Based on the image processing steps presented, errors can be easily minimized. This new approach met the study’s primary objectives of splitting tomatoes during partial occlusion and estimating weight/volume with acceptable accuracy on the test dataset. The RGB-based system can be used in a production line to calculate the volume and weight of tomatoes as a single item or with partial occlusion with reduced errors. The masses and volumes of tomatoes are accurately measured, allowing for more efficient handling and sorting processes after harvest. Since most customers and food retailers prefer homogeneous lots of a specific size, weight, and volume, producers can boost their trade agreements by having a greater understanding of classified sales volume. This technique can minimize labor costs and sorting time at any point in the production line. Hence, this computer vision system can calculate the volume and weight of tomato fruits in a nondestructive manner.


**Acknowledgments**

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

**Funding**

This work was supported by the National Science and Technology Support Program [grant number 2015BAD19806], provided by the College of Engineering, Nanjing Agricultural University, China; National Science and Technology Support Program [grant number2015BAD19806].

**ORCID**

Innocent Nyalala [http://orcid.org/0000-0002-7969-437X](http://orcid.org/0000-0002-7969-437X)

Cedric Okinda [http://orcid.org/0000-0001-7887-4216](http://orcid.org/0000-0001-7887-4216)

Tchalla Korohou [http://orcid.org/0000-0003-0419-3674](http://orcid.org/0000-0003-0419-3674)

Samuel Nyalala [http://orcid.org/0000-0001-6496-5029](http://orcid.org/0000-0001-6496-5029)

**References**

[1] Van De Poel, B.; et al. Model-based Classification of Tomato Fruit Development and Ripening Related to Physiological Maturity. *Postharvest. Biol. Technol.* 2012, 67, 59–67. Elsevier. DOI: [10.1016/j.postharvbio.2011.12.005](http://dx.doi.org/10.1016/j.postharvbio.2011.12.005).

[2] Wan, P.; et al. A Methodology for Fresh Tomato Maturity Detection Using Computer Vision. *Comput. Electron. Agric.* 2018, 146, 43–50. Elsevier. DOI: [10.1016/j.compag.2018.01.011](http://dx.doi.org/10.1016/j.compag.2018.01.011).

[3] Wei, Y.; et al. Defense Response of Cherry Tomato at Different Maturity Stages to Combined Treatment of Hot Air and Cryptococcus Laurentii. *Postharvest. Biol. Technol.* 2016, 117, 177–186. Elsevier. DOI: [10.1016/j.postharvbio.2016.03.001](http://dx.doi.org/10.1016/j.postharvbio.2016.03.001).

[4] FAOSTAT. *Food and Agriculture Organization (FAO), Statistics, 2019. Rome, Italy: Food and Agriculture Organization of the United Nations*.

[5] Uluisik, S.; Yildiz, F.; Ozdemir, A. T., Image Processing Based Machine Vision System for Tomato Volume Estimation. In *2018 Electric Electronics, Computer Science, Biomedical Engineering’s Meeting (EBBT)*. Istanbul, Turkey, 2018, IEEE. p. 1–4.

[6] Omid, M.; Khojastehnazhand, M.; Tabatabaeeefar, A. Estimating Volume and Mass of Citrus Fruits by Image Processing Technique Elsevier Ltd. *J. Food Eng.* 2010, 100(2), 315–321. DOI: [10.1016/j.jfoodeng.2010.04.015](http://dx.doi.org/10.1016/j.jfoodeng.2010.04.015).

[7] Vanderlei, R. S.; Silva, J. V.; Moura, F. D. B. P. BIOMETRY AND NON-DESTRUCTIVE ALLOMETRIC MODEL FOR ESTIMATING BABASSU (Attalea Speciosa) FRUIT VOLUME AND DENSITIES. *Rev Oureicuri.* 2020, 9, 1–10.

[8] Moreda, G. P.; et al. Shape Determination of Horticultural Produce Using Two-dimensional Computer Vision - A Review. *J. Food Eng.* 2012, 108(2), Elsevier Ltd, 245–261. DOI: [10.1016/j.jfoodeng.2011.08.011](http://dx.doi.org/10.1016/j.jfoodeng.2011.08.011).

[9] Nyalala, I.; et al. Tomato Volume and Mass Estimation Using Computer Vision and Machine Learning Algorithms: Cherry Tomato Model. *J. Food Eng.* 2019, 263, 288–298. DOI: [10.1016/j.jfoodeng.2019.07.012](http://dx.doi.org/10.1016/j.jfoodeng.2019.07.012).

[10] Okinda, C.; et al. Egg Volume Estimation Based on Image Processing and Computer Vision. *J. Food Eng.* 2020, 283, 110041. Elsevier Ltd. DOI: [10.1016/j.jfoodeng.2020.110041](http://dx.doi.org/10.1016/j.jfoodeng.2020.110041).

[11] Hahn, F.; Sanchez, S., Carrot volume evaluation using imaging algorithms Elsevier. *J. Agri. Eng. Res.* 2000, 75(3), 243–249. DOI: [10.1006/jaar.1999.0466](http://dx.doi.org/10.1006/jaar.1999.0466).

[12] Gan, Y. S.; et al. A Statistical Approach in Enhancing the Volume Prediction of Ellipsoidal Ham. *J. Food Eng.* 290. 2020, Elsevier 110186.

[13] Concha-Meyer, A.; et al. Volume Estimation of Strawberries, Mushrooms, and Tomatoes with a Machine Vision System. *Int. J. Food Prop.* 2018, 21(1), Taylor & Francis, 1867–1874. DOI: [10.1080/10942912.2018.1508156](http://dx.doi.org/10.1080/10942912.2018.1508156).

[14] Ireri, D.; et al. A Computer Vision System for Defect Discrimination and Grading in Tomatoes Using Machine Learning and Image Processing. *Artif. Intell. Agric.* 2019, Elsevier, 2, 28–37. DOI: [10.1016/j.aiia.2019.06.001](http://dx.doi.org/10.1016/j.aiia.2019.06.001).

[15] Satpute, M. R.; Mjagdale, S. Color, Size, Volume, Shape and Texture Feature Extraction Techniques for Fruits: A Review. *Int. Res. J. Eng. Technol.* 2016, 3, 2395–56.

[16] Ehsani, R.; Toudeshki, A.; Wan, P., New Sensor Technology for Yield Estimation and Disease Detection. In *Tomato Proceedings*. Florida, USA. 2016. p. 15–16.
