

Article

Not peer-reviewed version

---

# Assessment of Physicochemical Properties of Cashew Apple through Computer Vision

---

[Mathala Juliet Gupta](#), [C. Igathinathane](#)<sup>\*</sup>, [Jyoti Nishad](#), [Humeera Tazeen](#), [Astina Joice](#), [Sunoj Shahjahan](#), [Anand Mohan](#), [Parveen Kumar](#), Jamboor Dinakara Adiga

Posted Date: 21 October 2025

doi: 10.20944/preprints202510.1697.v1

Keywords: automation; horticultural crops; imagej; image processing; postharvest processing; wet chemistry



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Assessment of Physicochemical Properties of Cashew Apple through Computer Vision

Mathala Juliet Gupta<sup>1</sup>, C. Igathinathane<sup>2,\*</sup>, Jyoti Nishad<sup>3</sup>, Humeera Tazeen<sup>2</sup>, Astina Joice<sup>2</sup>, Sunoj Shahjahan<sup>4</sup>, Anand Mohan<sup>5</sup>, Parveen Kumar<sup>1</sup>, and Jamboor Dinakara Adiga<sup>3</sup>

<sup>1</sup> ICAR-CCARI, Ela, Old Goa 403402, India

<sup>2</sup> Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo, ND 58102, USA

<sup>3</sup> ICAR-Directorate of Cashew Research, Puttur, Uttara Kannada, Karnataka 574202, India

<sup>4</sup> Department of Agricultural and Biosystems Engineering, University of Illinois, Urbana-Champaign, IL 61801, USA

<sup>5</sup> Food Science & Technology, University of Georgia, Athens, GA 30602-2610, USA

\* Correspondence: Igathinathane.Cannayen@ndsu.edu; Tel.: +1-701-667-3011

**Abstract:** Cashew apples, a byproduct of the cashew nut industry with an estimated global production of 38 million tonnes, are rich in several essential nutrients and are widely processed into juice, syrup, wine, pickles, and other value-added products. However, their morphological and physicochemical properties vary significantly across varieties, complicating in-field characterization, maturity assessment, and biochemical analysis. These challenges originate from the reliance on costly chemicals, skilled manpower, limited time, and sophisticated equipment. This study employed a user-developed computer vision-based ImageJ batch processing plugin to assess 15 physicochemical properties across six diverse cashew apple varieties from the images of slices and whole samples. Five methodologies—color-grid, surface morphology, gray level co-occurrence matrix, local binary pattern, and color-indices—generated image-based metrics rapidly ( $2.87 \pm 0.79$  s/image). Correlation of wet chemistry with image-based parameters, linear modeling, and wet chemistry parameters prediction with an independent dataset were successfully performed, and successfully modeled properties include acidity, antioxidant, carbohydrates, carotenoids, crude fat, flavonoids, pH, phenolics, proteins, tannins, vitamin C, and total soluble solids. The results demonstrated the feasibility of predicting 80% physicochemical properties of cashew apples ( $R^2 > 0.5$ ), and that can be successfully modeled include acidity, antioxidant, carbohydrates, carotenoids, crude fat, flavonoids, pH, phenolics, proteins, tannins, vitamin C, and total soluble solids. This methodology offers a faster, safer, and cost-effective alternative to wet chemistry and can be extended to other horticultural crops.

**Keywords:** automation; horticultural crops; imagej; image processing; postharvest processing; wet chemistry

## 1. Introduction

Cashew (*Anacardium occidentale* L.) apple is a pseudo fruit formed from the pedicel of the flower. The cashew nut, a highly sought-after commercial product component, is attached to the cashew apple (Figure 1). Cashew apples are elongated, round, or pear-shaped fibrous fruits that naturally occur in yellow, orange, and bright red colors [1]. They weigh between 40 and 80 g and have lengths between 60 and 100 mm [2]. The cashew apple is sometimes used directly for table purposes [3]; however, it is primarily utilized in various processed forms.



**Figure 1.** Orange colored cashew apple on the tree with nut of Bhaskara variety developed in ICAR-DCR, Puttur, Karnataka, India.

Although the cashew tree originated in Brazil, it was spread to many of its colonies through Portuguese traders and is now globally distributed in approximately thirty countries between 25° N and 30° S latitude of the equator. Currently, Ivory Coast has emerged as the largest producer of cashew, followed by India and Vietnam [4,5]. However, just a few years ago, India was the top producer, followed by Vietnam, Brazil, and East Africa [6]. The production of cashew apples has been indirectly estimated to be 9 to 10 times that of cashew nuts [7,8]. The global production of cashew nuts in 2023 was  $4.2 \times 10^6$  t [5], which amounts between  $37.8 \times 10^6$  t and  $42 \times 10^6$  t of cashew apples.

Cashew apple is a storehouse of vitamins, minerals (potassium, magnesium, sodium, and iron), amino acids, carotenoids, phenolics, flavonoids (myricetin and quercetin), organic acids, and antioxidants [9–15]. Cashew apple has several proven health benefits, such as an instant energy booster [16], immunobooster enhancing bioactive compounds functionality [13,16,17], anti-inflammatory, antimicrobial [18,19]. Beyond its nutritional value, cashew apple has been traditionally used in various medicinal practices [19,20]. Its bagasse, the byproduct obtained after juice extraction—also known as pomace, is rich in cellulose, lignin, and hemicellulose [14,21,22].

Numerous processes and products have been developed from the cashew apple, including its juice and bagasse/pomace. The fruit is commonly processed into candies, pickles, jams, jellies, preserves, and syrup, and is also used as animal feed [23–25]. Cashew apple juice contains five times more vitamin C than orange juice and has been utilized in the production of ready-to-serve beverages, squash, wine, feni, and bioethanol [26–28]. The pomace serves various applications, including as a flour substitute in bakery products, alternatives to meat and dairy, and as a raw material for nanocomposites, polylactic acid, and other bioproducts [29,30]. Despite these diverse and well-documented uses, the cashew apple remains underutilized commercially due to its highly perishable nature [2] and the presence of undesirable flavors and compounds [8,31,32].

Rapid and timely quality assessment immediately after harvest is essential for optimal utilization and processing. However, the short harvesting season limits the feasibility of traditional wet chemistry methods and expensive equipment like HPLC, GCMS, Raman, NIR, MIR, UV, Vis/SWIR spectroscopy, texture analyzers, NMR, X-ray imaging, etc., which are time-consuming, expensive, and require skilled personnel and sophisticated equipment. There is currently a lack of rapid, affordable, and labor-efficient alternatives for estimating the physicochemical properties of cashew apples.

Developing such alternatives is crucial not to replace wet lab analyses, but to support them by identifying optimal harvest stages and enabling in situ rapid screening. This approach would

minimize the number of samples requiring full laboratory analysis. Such innovations would be globally beneficial: they could alleviate skilled labor shortages in developed countries and compensate for limited research infrastructure in developing regions. Moreover, the methodology could be adapted for use with minimal training, enabling even unskilled workers to assess cashew apple quality and potentially extendable to other fruits and vegetables.

Open-source tools, being free and accessible to all, could play an impactful role in developing these alternative methodologies. In the domain of image analysis, "ImageJ," a free and open-source image analysis system, has been widely used to estimate various physical and chemical properties of fruits, vegetables, and other food products [33,34]. Numerous studies have demonstrated a correlation between surface texture and color and functional compounds, biochemical, and physicochemical properties [35–37]. The integration of ImageJ with computer vision technologies enables rapid, non-destructive, and repeatable assessments of cashew apple quality in both field and laboratory settings. These approaches not only reduce reliance on traditional wet lab analyses but also make quality evaluation more accessible, thus supporting the scalability and commercial viability of cashew apple processing.

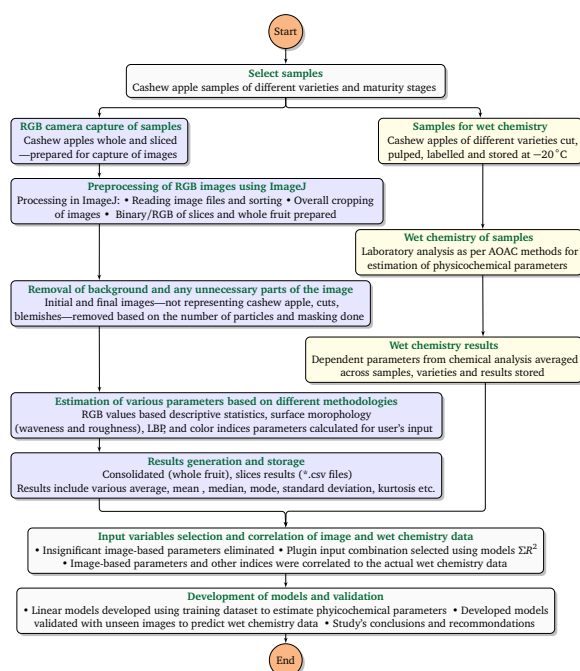
Therefore, it is hypothesized that the cashew apples' image-based parameters, derived from various formulas, are correlated to the physicochemical properties, which are estimated through standard wet chemistry methods, and consequently, simplified models for physicochemical properties can be developed using image-based parameters as independent variables. Therefore, this study aims to correlate image-based parameters derived from the images (whole and sliced) of different varieties of cashew apples, varying in color, shape, and size, to the physicochemical properties estimated through the standard laboratory protocols; and develop rapid tools (plugins and models) for in-field estimation of cashew apple physicochemical properties.

The specific objectives of this study are to: (i) derive image-based parameters for whole fruits and cut slices of cashew apples; (ii) correlate the derived parameters to the actual wet chemistry physicochemical property results; and (iii) develop models for estimation of physicochemical properties of cashew apples based on image analysis parameters and their validation.

The results of this study present a rapid, cost-effective, and user-friendly methodology for estimating physicochemical properties of cashew apple. This methodology will serve as a valuable tool for researchers, breeders, food processors, technicians, farmers, and other stakeholders; thereby reducing project budgeting costs, requiring minimal training, supporting data security, enhancing product quality, enabling the analysis of more samples, and minimizing the time required for evaluation.

## 2. Material and Methods

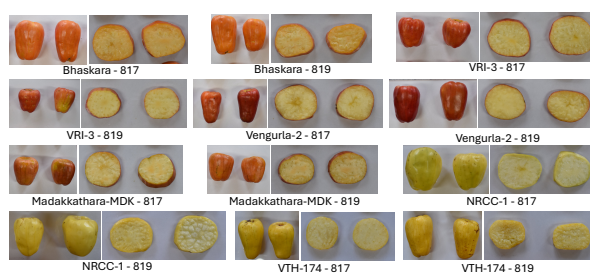
The methodology followed in the study consisted of four stages (Figure 2). First was the sample collection from the field, their image capture as a whole and sliced, processed, and stored for further analysis. The second stage was the wet chemistry in the laboratory for estimation of the physicochemical and biochemical properties (dependent variables) and simultaneously pre-processing and processing of the images acquired using a user-developed ImageJ plugin to derive various parameters (independent variables) for estimation of the physicochemical properties (dependent variables). The third stage was the determination of the correlation of the dependent and independent variables. In the ultimate stage, based on the best-ranked parameters, models were developed and validated.



**Figure 2.** Flowchart elucidating the research methodology of cashew apple physicochemical properties assessment through computer vision.

### 2.1. Samples Description

Six varieties of cashew apples of varying flesh and skin colors *viz.* orange (Bhaskara, VRI-3), red (Vengurla-2, Madakkathara-MDK), and yellow (NRCC-1, VTH-174) were harvested from orchards of the ICAR-Directorate of Cashew Research (12° 44' 44" N, 75° 13' 54" E) and (12° 44' 24" N, 75° 13' 50" E), Puttur, Karnataka, India (Figure 3). Cashew apples were harvested based on a phenological scale developed by Adiga et al. [38], namely, modified BBCH (Biologische Bundesanstalt, Bundessortenamt, and Chemische Industrie). This scale has five nut and fruit maturity stages: 811, 813, 815, 817, and 819. The final two stages, '819' representing fully ripened and '817' representing approximately 70% ripened cashew apples, were chosen for the study.



**Figure 3.** Six varieties of cashew apples with cut slices, harvested at two stages of maturity. The numbers indicated by 817 and 819 are the two stages of maturity. Complete set of original images is available from the uploaded data [39].

In this study, images of whole and sliced cashew apples were utilized as input images to assess the external and internal characteristics and their effect on the determination potential of physicochemical properties. After harvesting, whole cashew apples and their slices (images captured shortly after slicing to prevent discoloration) were photographed using a Nikon D5600 camera fitted with a Nikon F lens (maximum resolution: 6000 × 4000, 24 MP). The apples were then pulped and stored at −20 °C until used for physicochemical properties analysis. All analyses were performed in triplicate. Original images of the cashew apple samples used in the study, along with samples wet chemistry physicochemical properties measurements, can be found in the "Mendeley Data" repository (<https://doi.org/10.17632/67kn43rj5c.1>; accessed on 16 October 2025) [39].

## 2.2. Wet Chemistry

The traditional wet chemistry methods were used for estimation of the physicochemical, namely, moisture content, fat, protein, total carbohydrates, total sugar, reducing sugar, vitamin C, and biochemical properties like total phenolic content, total flavonoids content, antioxidant activity, tannin content, and carotene content of the cashew apple varieties at the two maturity stages using AOAC standard methods. For the physicochemical properties analysis, the cashew apples were pulped using a mortar and pestle, and three replications were employed for the wet chemistry analysis. All chemicals and solvents used were of analytical grade and obtained from Sigma (Sigma-Aldrich Co., Bangalore, India) and Merck (Darmstadt, Germany). A total of 15 wet-chemistry-based physicochemical properties was measured, and they were considered as dependent variables in the data analysis. Details of the wet chemistry protocols for physicochemical properties considered are presented in the Appendix A.

## 2.3. Computer Vision Methodology

The computer vision methodology, in essence, involves acquiring images of whole and sliced cashew apple varieties at two maturity stages, followed by preprocessing using the ImageJ software (version 1.54p), and subsequent image-based parameter extraction using a custom-developed ImageJ plugin (Figure 2). The selection of image-based parameters is guided by a logical framework that ensures its relevance to the measured physicochemical properties. Finally, the correlation between the extracted parameters and the physicochemical properties is analyzed to derive the desired insights and model development.

### 2.3.1. Image Acquisition

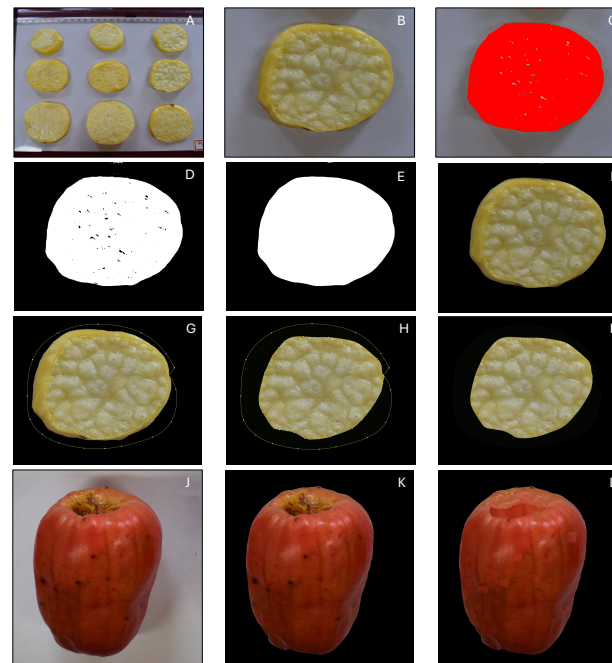
Cashew apples harvested from the field were cleaned, and their images were captured in two forms: whole and sliced, to study how these forms correlate with physicochemical properties. The samples (whole and slices) were placed on a uniform colored background that has better contrast, and photographed either using a Nikon D5600 camera equipped with an F-mount lens (maximum resolution:  $6000 \times 4000$ , 24 MP). The images were saved in \*.jpg format and organized into specific folders for subsequent analysis.

### 2.3.2. Image Preprocessing

The captured images were preprocessed to make them suitable for image-based parameter extraction using the user-developed ImageJ plugin. The sequence of preprocessing operations followed is illustrated in Figure 4. The original images are collections of many samples (slices or whole fruits) on the original background (Figure 4a) that need to be suppressed for further image processing. Hence, the collections are split into individual samples, and their background are eliminated. Preprocessed images were made to have a uniform black background, enabling the automatic extraction of objects of interest by the plugin, thereby facilitating the determination of analysis parameters.

Initially, the images were opened in ImageJ, and each slice or whole fruit was selected (rectangle selection tool) and cropped (Figure 4b). Each of these cropped images serves as a replication for the analyses. Segmentation of the samples from the background is a prerequisite for the extraction of the sample features. Usually, segmentation of color images is performed via the creation of grayscale images, from which the final segmented binary images are obtained. However, this method was not applicable as the images acquired on-field, as well as in this study, may lack a contrasting background. Therefore, the application of color-based segmentation was the most suitable option. Although several color spaces are available, the selection of the appropriate color space facilitates effective segmentation. For the selection of the color space, the presence of distinct and separated histogram peaks in the component channels serves as a discriminating criterion. Preliminary studies showed that "color thresholding" in the YUV color space (Figure 4c) produced the best segmented images for both cashew apple slices and whole fruits. The 'U' channel of the YUV color space exhibited distinct and separated histogram peaks, assisting the segmentation process.

The 'U' channel was utilized to select the region of interest (polygonal), which was converted to a binary image (Figure 4d) employing the "make binary" command. The binary images were then further refined by filling holes (Figure 4d), as necessary. A median filter was employed to smooth (Figure 4e) the edges (10 pixels) and eliminate minor particle artifacts. Finally, the preprocessed binary image (mask) was overlaid onto the corresponding cropped image sample, utilizing the image calculator command with the 'AND' operator.



**Figure 4.** Various steps followed in the preprocessing of cashew apple slice samples. (A) Collection of samples; (B) cropped individual sample; (C) sample segmentation utilizing 'U' channel histogram and selection; (D) sample binary image showing unsegmented regions as holes; (E) binary sample image mask after filling holes and median filter; (F) original color image after overlaying with mask; (G) selection of skin areas with polygon tool and smoothening with fit spline; (H) background black filling of selected area; (I) final segmented sample image; (J) cropped whole fruit original image; (K) preprocessed whole fruit showing pedicle cavity and blemishes showing dark spots; and (L) dark regions rectified by pasting similar colored patches giving the final segmented sample image.

Any issues, such as the presence of skin of the fruits appearing in slice images or blemishes or pedicel cavities in whole fruits, which had a darker color and required rectification, as they would misrepresent the natural color of the samples analyzed. This correction was performed using the polygon tool, followed by fit spline to smooth the sharp edges and corners. In slices, the skin portion was manually selected using the smoothed polygon and filled with a black background (Figure 4g, h). For whole fruits, the blemishes or pedicel cavities were removed by copying and pasting the nearby similar-colored patches. For ease of processing multiple images, shortcuts were created for the often-used operations, such as color threshold, make binary, fill holes, median filter, convert to mask, save as Gif, image calculator, fit spline, and save as Jpeg.

### 2.3.3. Image-based parameters

The preprocessed images (Figure 2) were analyzed to extract color, texture, and surface morphology descriptors using various methodologies described below, to study how they relate to wet chemistry-based physicochemical properties. A total of 15 wet chemistry parameters (dependent variables) and 144 image-based parameters (independent variables) derived from various methodologies are listed in Table 1. These parameters were assessed separately for slice and whole fruit images due to their distinct surface properties in terms of color and overall surface texture.

**Table 1.** Dependent wet chemistry-based physicochemical properties variables and independent image-based variables employed in the study.

Methodology	Applicable parameter variables	Total: 159
Wet chemistry *	Acidity (titrable), Antioxidant activity, Carbohydrates (total), Carotene (content), Fat crude, Flavonoids (total), Moisture (content), pH, Phenolic (total content), Protein, Reducing sugars, Tannin (content), Total soluble solids, Total sugars, and Vitamin C (Section 2.2).	Total: 15
Color grid **,†	Mean <i>R</i> , Mean <i>G</i> , Mean <i>B</i> , Standard deviation of <i>R</i> , Standard deviation of <i>G</i> , Standard deviation of <i>B</i> , Median <i>R</i> , Median <i>G</i> , Median <i>B</i> , Geometric mean <i>R</i> , Geometric mean <i>G</i> , Geometric mean <i>B</i> , Maximum <i>R</i> , Maximum <i>G</i> , Maximum <i>B</i> , Minimum <i>R</i> , Minimum <i>G</i> , Minimum <i>B</i> , Number of <i>R</i> values, Number of <i>G</i> values, Number of <i>B</i> values, Skewness <i>R</i> , Skewness <i>G</i> , Skewness <i>B</i> , Kurtosis <i>R</i> , Kurtosis <i>G</i> , and Kurtosis <i>B</i> (Section 2.3.3.i).	Total: 27
Waveness and roughness **	Descriptive statistics parameters of waveness plot (Number of waveness plot data points, Minimum waveness, Maximum waveness, Mean waveness, Geometric mean waveness, Standard deviation waveness, Median waveness, Geometric mean waveness, Standard deviation of waveness), Waveness parameters calculated based on formulae from references (Arithmetic mean waveness, Root mean square waveness, Maximum valley depth of waveness plot, Maximum peak height of waveness plot, Average valley depth of waveness plot, Average peak height of waveness plot, Maximum peak to valley height of waveness plot, Descriptive statistics parameters of roughness plot (Number of roughness plot data points, Minimum roughness, Maximum roughness, Mean roughness, Geometric mean roughness, Standard deviation roughness, Median roughness), Waveness parameters calculated based on formulas from references (Arithmetic mean roughness, Root mean square roughness, Maximum valley depth of roughness plot, Maximum peak height of roughness plot, Average valley depth of roughness plot, Average peak height of roughness curve, and Maximum peak to valley height of roughness plot) (Section 2.3.3.ii).	Total: 28
GLCM **	Angular second moment, Contrast, Correlation, Inverse difference moment, and Entropy (Section 2.3.3.iii).	Total: 5
LBP **,‡	Number of gray pixels analyzed, Minimum gray, Maximum gray, Mean gray, Geometric mean gray, Quartile mean gray, Standard deviation gray, Median gray, Skewness gray, Kurtosis gray, Number of <i>R</i> pixels analyzed, Minimum <i>R</i> , Maximum <i>R</i> , Mean <i>R</i> , Geometric mean <i>R</i> , Quartile mean <i>R</i> , Standard deviation <i>R</i> , Median <i>R</i> , Skewness <i>R</i> , Kurtosis <i>R</i> , Number of <i>G</i> pixels analyzed, Minimum <i>G</i> , Maximum <i>G</i> , Mean <i>G</i> , Geometric mean <i>G</i> , Quartile mean <i>G</i> , Standard deviation <i>G</i> , Median <i>G</i> , Skewness <i>G</i> , Kurtosis <i>G</i> , Number of <i>B</i> pixels analyzed, Minimum <i>B</i> , Maximum <i>B</i> , Mean <i>B</i> , Geometric mean <i>B</i> , Quartile mean <i>B</i> , Standard deviation <i>B</i> , Median <i>B</i> , Skewness <i>B</i> , and Kurtosis <i>B</i> (Section 2.3.3.iv).	Total: 40
Color indices **	Color indices derived from overall <i>R</i> , <i>G</i> , and <i>B</i> values (27 existing + 17 new) are given in Table 2.	Total: 44

\* Dependent variables (total 15); \*\* Independent variables (total 144); † Overall values derived from the pixels of the selected ROIs based on user inputs; ‡ Set of values analyzed by splitting the color image into the component *R*, *G*, and *B* channels; GLCM - gray level co-occurrence matrix; LBP - local binary pattern; *R* - red, *B* - blue, and *G* - green pixel intensity values.

### (i) Color-based analysis

The overall color profiles of each image (slices or whole fruits; e.g., Figure 4i) based on their pixel color intensities of the pixels ( $R$  - red,  $G$  - green, and  $B$  - blue) were captured for the data analysis. These values were collected after identifying several regions inside the sample outlines, and the descriptive statistics were obtained for further analysis. For this, the images were systematically partitioned into rectangular regions of interest (ROI) along the length and breadth of the sample outline (Figure 5.A2). The number of ROIs employed in the color analysis was based on the user-specified inputs (Section 2.3.4). For each ROI, the pixels'  $R$ ,  $G$ , and  $B$  intensity values were extracted, and statistical descriptors, including mean, median, geometric mean, standard deviation, maximum, minimum, skewness, and kurtosis, were computed (Table 1). The total number of independent variables in the color-based analysis was 27.

### (ii) Surface morphology—waviness and roughness

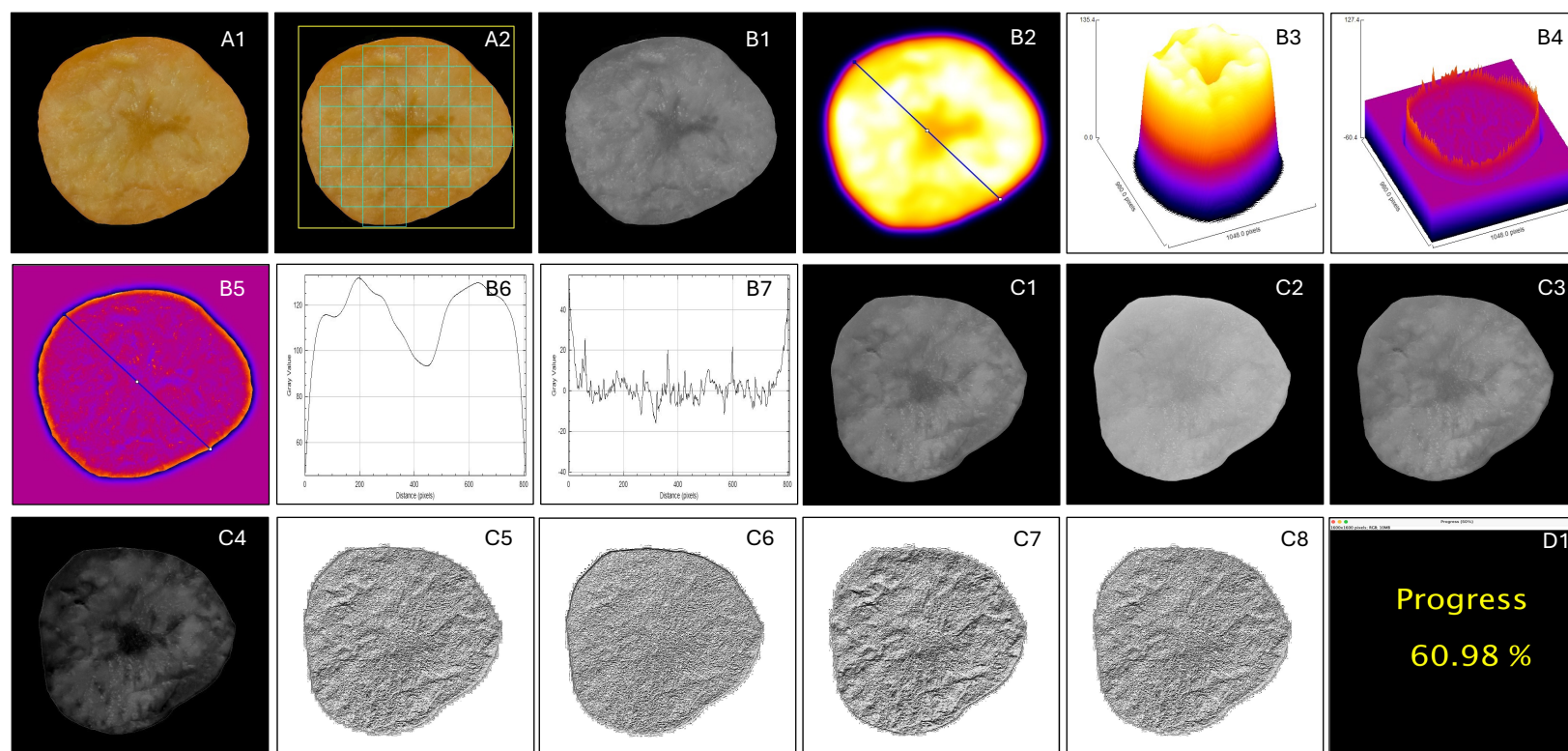
Surface morphology or texture was evaluated in terms of waviness and roughness. The waviness represents large-scale surface deviations, while roughness refers to the fine irregularities on the surface. Before surface morphology parameter extraction, the following preprocessing operations were applied to minimize artifacts: (i) Levelling—the surface was leveled to correct tilt using a polynomial least squares plane subtraction, (ii) Filtering—a gaussian or median filter was applied to reduce noise while preserving edge features, and (iii) ROI selection—a consistent ROI was selected across the samples to ensure comparability.

To distinguish between waviness and roughness, the surface profile was decomposed using the waviness and roughness functions. A cutoff wavelength-based filter was selected to separate waviness from roughness. The 2D waviness and roughness profiles were extracted using the plugin following the reported methodology [40] and the reported plugin generating various parameters [41]. The profile data was also used to generate descriptors such as mean, median, mode, maximum, minimum, kurtosis, and skewness, employing descriptive statistics. With 14 each for waviness and roughness parameters, the total number of independent variables considered in the surface morphology analysis was 28 (Table 1). These surface morphological parameters of waviness and roughness are influenced by various inputs (Section 2.3.4). Where, the “Cutoff value (pixel)” is the cutoff radius for the Gaussian filter to be applied, it has to have a minimum value of 4, but maximum value depends on the size of the image, and the offset variables of the line ROI (ROI x1, ROI y1, ROI x2, and ROI y2) decide the position (%) of the line transect, along which the waviness and roughness were determined (Figure 5B2, B5).

### (iii) Gray level co-occurrence matrix (GLCM) method

The GLCM is a second-order statistical texture analysis method, originally developed by Haralick [42,43], that uses the spatial relationship between the gray values of pairs of pixels. Therefore, the preprocessed images were initially converted to 8-bit grayscale. The GLCM is a matrix with gray levels in the rows and columns. Each entry ( $i, j$ ) in the matrix represents the probability or frequency with which two pixels separated by a distance defined by the user have intensity values of  $i$  and  $j$ . The parameters that can influence these values are considered as user-defined inputs (Section 2.3.4), which will be the distance between the pixel pairs referred to as the “Step size” in pixels (1, 2, 3, etc.) and the “Step direction” which can be  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ , and the number of gray levels employed (256 to 8 used in the study; other values such as 16 are also possible).

The GLCM is normalized by dividing each element by the total number of pairs considered before obtaining the statistical features. Each normalized GLCM represents the joint probability distribution of gray-level intensity pairs. From these matrices, the angular second moment, contrast, correlation, inverse difference moment, and entropy were calculated. Directional values were averaged to obtain a single representative measurement



**Figure 5.** Processing of images based on different methodologies (A–C). (A1) Preprocessed slice of cashew apple processed in color-based analysis; (A2) ROIs generated for color extraction; (B1) grayscale image for surface morphology analysis; (B2) waveness image with line ROI for sampling; (B3) 3D-waveness surface plot of image; (B4) 3D-roughness surface plot; (B5) roughness image with line ROI for sampling; (B6) waveness line ROI profile plot; (B7) roughness line ROI profile plot; (C1) grayscale image for LBP; (C2–C4) red, green, and blue component channels, respectively; (C5–C6) texture plots of LBP analysis of gray, red, green, and blue channels; and (D1) analysis progress indicated during batch processing of images.

per parameter for each sample. The commonly used GCLM extracted features are (i) contrast—measures the local intensity variation, with higher contrast implies higher local variations; (ii) correlation—measures the correlation between a pixel and its neighboring pixel; (iii) energy (angular second moment)—measures textural uniformity, with higher energy means less textural complexity; (iv) homogeneity (inverse difference moment)—measures the closeness of distribution of the GLCM to its diagonal, again higher the homogeneity implies more uniform is the texture; and (v) entropy—measures randomness, with higher value signifies complexity of texture (Table 1). The total number of GLCM-based parameters derived was 5.

#### (iii) Local binary pattern (LBP) method

The LBP method operates on grayscale images by comparing the intensity of each central pixel with that of its neighboring pixels within a defined radius [44]. For each neighbor, a binary value of 0 is assigned if its intensity is lower than the central pixel, and 1 if it is equal to or higher. The resulting binary sequence is concatenated in a predefined order to form an LBP code, which is then converted into its decimal equivalent. A histogram of these LBP codes is generated to represent the frequency distribution of distinct local texture patterns within the image. Since the colors in the RGB spectra may be correlated with physicochemical properties, a modified approach of calculating the intensities separately in the three spectra was developed in this study. For this, the RGB channels of the preprocessed color image were extracted, and corresponding images were obtained utilizing the “RGB Stack” conversion and “Stack to Images” ImageJ commands. Descriptive statistics were then used to generate parameters such as minimum, maximum, mean, geometrical mean, quadratic mean, standard deviation, etc. (Table 1). No specific user-input variables are necessary. The total number of LBP-based parameters derived was 40.

#### (iv) Color indices

Numerous researchers have successfully utilized vegetative indices (VIs) derived from RGB color data of target images to assess crop yield, chlorophyll content, and various other qualitative parameters of crops. Building upon the success of VI applications, this study dealing with fruits (both slice and whole) employed 27 existing VIs, hereafter referred to as color indices. These color indices were computed based on the overall mean  $R$ ,  $G$ , and  $B$  values using relevant formulas (Table 2). Given that cashew apple samples showing the prominence of red and yellow hues in their sample images (both slice and whole), 17 additional color indices focusing mostly on ‘ $R$ ’ component and its interaction with ‘ $G$ ’ and ‘ $B$ ’ components, extending the VIs formulas, were newly developed in this study (Table 1). Therefore, the total number of color indices-based (27-existing, 17-new) was 44.



Table 2. List of RGB-based indices.

Vegetation Index	Equation	Reference
Blue chromatic coordinate (BCC)	$\frac{B}{R+G+B}$	[45]
Green-blue difference index (GBDI)	$G - B$	[46]
Blue red vegetation index (BRVI)	$\frac{B-R}{B+R}$	[47]
Color index of vegetation (CIVE)	$0.441R - 0.881G + 0.385B + 18.787$	[48,49]
Coloration index (CI)	$\frac{R-B}{R}$	[50]
Combination1 (COMB1)	$0.25ExG + 0.3EXGR + 0.33CIVE + 0.12VEG$	[51]
Combination2 (COMB2)	$0.36ExG + 0.47CIVE + 0.17VEG$	[51]
Excess green (ExG )	$2G - B - R$	[45,48]
Excess green minus excess red (ExGR)	$ExG2 - ExR$	[48,52]
Excess red (ExR)	$\frac{1.4R-G}{R+G+B}$	[48]
Green-blue vegetation index (GBVI)	$\frac{G-B}{G+B}$	[53,54]
Green chromatic coordinate (GCC)	$\frac{G}{R+G+B}$	[45]
Green-red difference (GRD)	$G - R$	[46,48]
Green-red ratio index (GRRI)	$\frac{G}{R}$	[48,55]
Kawashima index (IKAW)	$\frac{R-B}{R+B}$	[47,48]
Modified green-red vegetation index (MGRVI)	$\frac{G^2-R^2}{G^2+R^2}$	[46,48]
Normalized difference index (NDI)	$\left[128 \times \frac{G-R}{G+R}\right] + 1$	[56]
Normalized excess green (ExG2)	$\frac{2G-B-R}{R+G+B}$	[45]
Normalized green-red difference index (NGRDI)	$\frac{G-R}{G+R}$	[57,58]
Red chromatic coordinate (RCC)	$\frac{R}{R+G+B}$	[45]
Red blue ratio index (RBRI)	$\frac{R}{B}$	[59]
Red green ratio index (RGRI)	$\frac{R}{G}$	[60]
Red green blue vegetation index (RGBVI)	$\frac{G^2-BR}{G^2+BR}$	[46]
Vegetiven (VEG)	$\frac{G}{R^{0.667}+B^{0.333}}$	[46]
Visible atmospherically resistant index (VARI)	$\frac{G-R}{G+R-B}$	[61]
Visible-band difference vegetation index (VDVI)	$\frac{2G-R-B}{2G+R+B}$	[62]
Woebbecke index (WI)	$\frac{G-B}{R-C}$	[45]

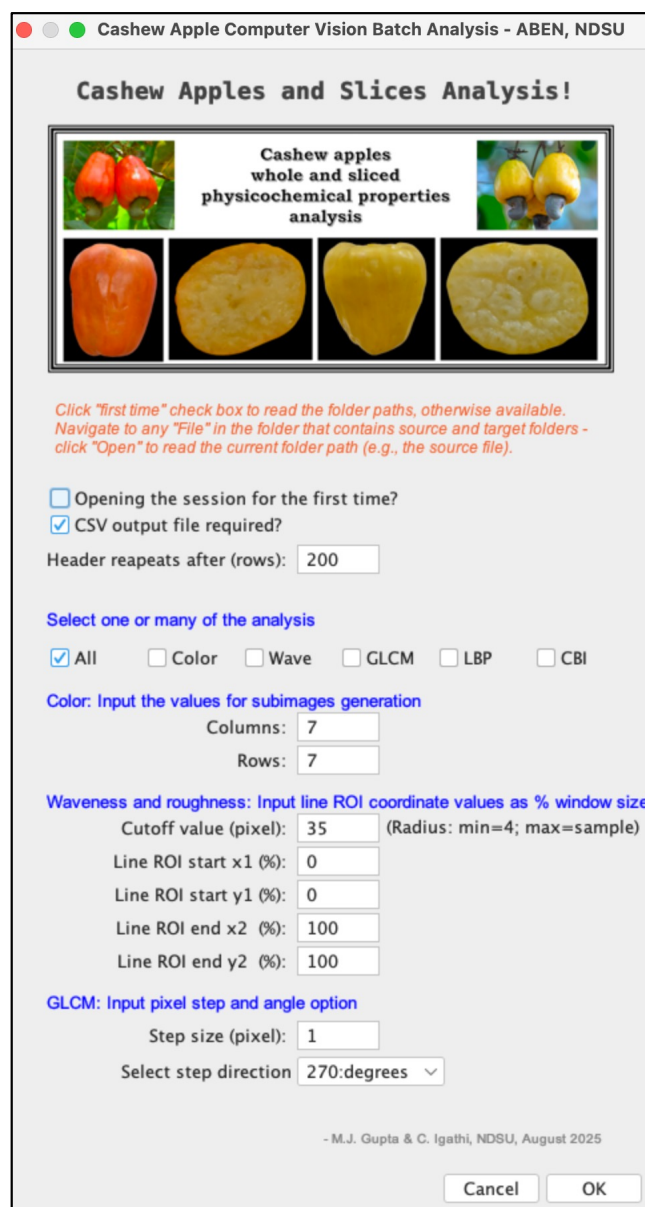
Table 2. Cont . . .

Vegetation Index	Equation	Reference
Excess modified red (ExmR)	$2R - G - B$	†
Modified coloration index (MCI)	$\frac{R-G}{R}$	†
Modified red-blue vegetation index (MRBVI)	$\frac{R^2-B^2}{R^2+B^2}$	†
Modified red-green-blue vegetation index (MRGBVI)	$\frac{R^2-GB}{R^2+GB}$	†
Modified red-green vegetation index (MRGVI)	$\frac{R^2-G^2}{R^2+G^2}$	†
Normalized excess red (ExR2)	$\frac{2R-G-B}{R+G+B}$	†
Normalized red-blue difference index (NRBDI)	$\left[ 128 \times \frac{R-B}{R+B} \right] + 1$	†
Normalized red-green difference index (NRGDI)	$\left[ 128 \times \frac{R-G}{R+G} \right] + 1$	†
Red-blue difference ratio (RBDR)	$\frac{R-B}{G-R}$	†
Red color index (RCI)	$\frac{2R-G-B}{2R+G+B}$	†
Red-green difference ratio (RGDR)	$\frac{R-G}{B-R}$	†
Red-blue difference (RBD)	$R - B$	†
Red-green difference (RGD)	$R - G$	†
Red-green vegetation index (RGVI)	$\frac{R-G}{R+G}$	†
Triangular redness index (TRI)	$R - 0.39G - 0.61B$	†
Visible atmospherically resistant red-blue index (VARRBI)	$\frac{R-B}{R+B-G}$	†
Visible atmospherically resistant red-green index (VARRGI)	$\frac{R-G}{R+G-B}$	†

n (G) bands recorded from the satellite platforms. \* Adopted from GLI. \*\* Adopted from Triangular greenness index (TGI). † Developed in the present study to capture red-tint-dominated samples.

### 2.3.4. The plugins input panel or front panel

The developed plugin, titled “Cashew Apple and Slices Analysis”, features an input panel depicted in Figure 6. The panel provides the user interface for various methodologies and enables the input of various parameters for the analysis. The input panel is designed to be self-explanatory with brief instructions, distinct checkboxes, and input fields categorized by types of analysis. Users interact with the panel to generate analysis and output. This plugin operates in a “batch-mode”, where all the preprocessed images stored in the input folder are automatically analyzed, results are appended with timestamps and stored in a ‘csv’ file format for record and further analysis, and visualize the results progress of batch analysis as a pop-up window (Figure 5D1) and on the ImageJ progress bar.



**Figure 6.** Input panel of the developed “Cashew Apple and Slices Analysis” batch processing plugin for calculating various image-based parameters used for estimating physicochemical properties of cashew apples. The panel shows brief instructions, various method choices for analysis, input parameters for the methodologies, and buttons to operate the plugin. The main inputs (5 total) are color subimage column and rows (2 inputs), waviness and roughness cutoff pixel (1 input), and GLCM step size pixel and direction degrees (2 inputs).

The plugin (Figure 6) provides user-friendly prompts to ensure accurate input. The first checkbox, “Opening the session for the first time?”, establishes the default folder for storing input images during batch processing, a task that is necessary only once. When the ‘csv’ output is desired, the second box

should be checked. The type of analysis desired can be chosen from the group of checkboxes: All, Color, Wave, GLCM, LBP, and CBI. Users can choose the analysis individually or in combination, or 'All' (ignores individual analysis and generates all analysis results). The other inputs (input boxes) are related to the different analyses and are explained in Section 2.3.3. Upon clicking the "OK" button, the program executes based on the selected analysis and generates images, log files, and CSV data. The "Cancel" stops the execution, bypassing it, and makes the plugin ready for the next session.

#### 2.3.5. The Cashew Apples and Slices Analysis plugin

The "Cashew Apples and Slices Analysis" ImageJ plugin has a total of 2050 lines coded in Java using Fiji (Ver. ImageJ 1.54q). The plugin's functionality is encapsulated within various Java methods (functions), including reading inputs, performing different analyses, and generating graphical and textual outputs, which play well with the batch processing strategy used in the plugin. Some of the advanced features used in the plugin are derived from packages including Apache Commons DescriptiveStatistics; ImageJ's ImagePlus, ImageProcessor, GenericDialog, GaussianBlur, WindowManager; and Java's Stream, Date, File, BufferedWriter, FileWriter.

#### 2.4. Data analysis for predicting the wet chemistry from image-based parameters

Data analysis is aimed to (i) determine the strength of association between the wet chemistry physicochemical properties (dependent) and the image-based parameters (independent) variables and graphically visualize the correlation, (ii) evaluate the range and ranking of image-based parameters correlation to wet chemistry parameters, (iii) develop the top three linear prediction models of wet chemistry based on the image-based parameters, and (iv) perform prediction and validation with visualization, employing the top-ranked prediction model developed and applied to unseen images. To accomplish these aims, an R (Ver. 4.5.1, 2025) program was developed utilizing the RStudio IDE (Ver. 2025.09.1, Posit Software, PBC). Major R packages involved in the R program development are `tidyverse`, `dplyr`, `readr`, `corrplot`, and `ggplot2`. The R program was coded as an assembly of several "functions", each designed to perform specific tasks. These functions were appropriately invoked in the program to execute the analysis and visualization. A brief description of the various data analysis procedures is presented subsequently, and the R script containing the data is uploaded to an online data repository [39].

##### 2.4.1. Correlation analysis and visualization

The data obtained from the measured wet chemistry parameters and generated by the computer vision plugin (Section 2.3.4) as an \*.csv file, forms the input for the correlation analysis. The program reads the entire csv file and converts it into a "dataframe". The data should be preprocessed by subsetting to extract the relevant variables, discarding other informative (e.g., time, image file name) and non-varying (e.g., certain minimum and maximum values) parameters (Section 3.2). The subset data contains categorical variables, namely, cashew apple varieties (6), maturity stages (2), and sample types (2), along with their replications. The replications should be combined into mean values associated with unique combinations of the categorical variables. The analysis is performed separately for the slice and the whole samples (as their images exhibit distinct differences). These operations were carried out through `group_by`, `filter`, and `select` operations.

A matrix of correlation coefficients ( $r$ ) was generated using the `cor()` command, and this object, after being converted to a dataframe, was output as a csv file for further analysis using the `write_excel_csv` commands. The `corrplot()` command with method of "square" and color options created the  $r$  values visualization through a color-coded correlation diagram (Figure 7). Correlation diagrams were created separately for the slice, whole, and combined samples to study their effect on the association.

#### 2.4.2. Range and ranking of correlation coefficients

Although feasible with R, the MS Excel spreadsheet program was utilized to create the range data (Section 3.5) and ranking of  $r$  values. In addition, code for generating a L<sup>A</sup>T<sub>E</sub>X ranking table was created (Section 3.6). Utilizing the MIN() and MAX() functions on the slice and whole samples \*.csv files separately, the  $r$  range data was generated, excluding the  $r$  values of wet chemistry parameters (self-correlation). The generated range data was then read into the R program, reshaped using pivot\_longer, and plotted using ggplot commands. The ggplot utilized geom\_bar with the "stack" option to draw the range maximum and minimum  $r$  values, facet\_warp, and other commands to depict both slice and whole samples.

For developing the ranking table, the actual  $r$  values (both +ve and -ve) were (i) converted to absolute values using the ABS(), (ii) for each wet chemistry parameter, the image-based parameters were ranked using the RANK(), (iii) the top 10 image-based parameters for each wet chemistry parameter were collected after sorting, and (iv) those top ten image-based parameters as text were combined using the CONCATENATE() with other symbols (e.g., & and \) to generate the L<sup>A</sup>T<sub>E</sub>X table code.

#### 2.4.3. Wet chemistry prediction models development

The process of creating the top three (3) models for each wet chemistry parameter (15) with maturity stages (2) and sample types (2) from the single master subset data (total = 180 models; Section 3.7) involves a series of intricate operations. The data processing requirements to achieve this as a "repetitive" function are: select the specific wet chemistry variable while dropping the other 14; select all of the image-based parameters; filter the rows that contain the desired maturity stage and sample type; create a set of linear models using the lm() function for all the image-based parameters by generating a formula for the dependent with all the image-based parameters predictor variables using the map() function that generates a "list" of models; extract the intercept, slope,  $R^2$ , number of observation, and individual formula strings using the sapply() function; generate a dataframe with these values; sort the models based on  $R^2$  in descending order using arrange(-()); and select the top three models using slice\_head(n=3). This function was invoked 60 times (15 wet chemistry  $\times$  2 stages  $\times$  2 sample types), and each time it produced the top three models (3 rows). These three rows of output were appended to generate the 180 model result rows using bind\_rows and finally written to the file using write\_excel\_csv() command.

#### 2.4.4. Wet chemistry prediction and validation

For prediction, the top-performing models of wet chemistry in each combination of stage and sample type (60) were utilized across the test set of unseen images. The predictions are validated with the known wet chemistry measurements for the six varieties of cashew apple. The image-based parameters were generated by the developed plugin with the test images as inputs. The top-performing models in each category will specify the independent variable involved, and the value of that variable will be obtained as a mean from the image-based parameters plugin results. The mean value of the independent variable when input into the linear model with specific slope and intercept values will generate the prediction of wet chemistry physicochemical properties. The performance of prediction will be evaluated as the ratio (predicted to actual), where the ideal prediction will have a ratio of 1.0, while underestimation will have ratios <1.0 and overestimation will have ratios >1.0.

The prediction and validation can be visualized as a dodged bar graph with maturity stages for all the wet chemistry properties evaluated for each sample type. The six varieties will be depicted as facets. The dodged bar plot function takes the data (actual, predicted, and ratio, specific to slice and whole), and  $y$ -axis limits (min and max) inputs, and generates the plot. The plot function filters the ratios, the geom\_bar() command takes "dodge" as argument, passes the  $y$ -axis inputs into the ylim() command, draws a horizontal line at  $y = 1.0$  using geom\_hline with yintercept = 1.0, and makes facets using facet\_wrap() with ~Variety as the argument. The online data repository R script provides additional details [39].

### 3. Results and discussion

#### 3.1. Results: Variation in wet chemistry results within samples and varieties

Significant differences were observed in physicochemical properties of cashew apples between maturity stages 817 and 819 across all varieties (mean  $\pm$  std shown in Table 3; wet chemistry measurement protocols are given in Appendix A). With the progress of maturity stages, the acidity decreased while the pH and TSS increased consistently ( $p < 0.05$ ). Antioxidant activity, flavonoids, phenolic content, and tannins showed significant reductions ( $p < 0.01$ ), indicating a decline in bioactive compounds after maturation. Carbohydrate content, reducing sugars, and total sugars increased significantly at stage 819 ( $p < 0.01$ ), suggesting an increase in sweetness and energy value. Carotenoids and vitamin C also increased sharply, particularly in Bhaskara and VRI-3 varieties ( $p < 0.001$ ). Crude fat content and proteins declined across all varieties, with VTH-174 and MDK showing the most pronounced reductions. The variety VTH-174 exhibited the highest antioxidant activity at stage 817. Bhaskara recorded the highest vitamin C content at stage 819 (181.71 mg/100g). Variety VRI-3 showed a threefold increase in carotenoids from stage 817 to 819, and NRCC-1 had the highest sugar and total carbohydrate levels at stage 819.

**Table 3.** Cashew apple wet chemistry data for various varieties and maturity stages.

Wet chemistry	Cashew apple varieties and maturity stages											
	Bhaskara		MDK		NRCC-1		Vengurla-2		VRI-3		VTH-174	
	817	819	817	819	817	819	817	819	817	819	817	819
Acidity (%)	0.14 $\pm$ 0.0	0.09 $\pm$ 0.01	0.12 $\pm$ 0.0	0.08 $\pm$ 0.0	0.16 $\pm$ 0.0	0.12 $\pm$ 0.0	0.25 $\pm$ 0.0	0.18 $\pm$ 0.0	0.16 $\pm$ 0.0	0.12 $\pm$ 0.0	0.16 $\pm$ 0.0	0.11 $\pm$ 0.0
Antioxidants (μmol/100g)	28.59 $\pm$ 1.13	20.85 $\pm$ 0.22	25.96 $\pm$ 0.45	8.94 $\pm$ 0.11	24.39 $\pm$ 0.47	22.69 $\pm$ 0.91	26.01 $\pm$ 0.83	2.88 $\pm$ 0.28	4.21 $\pm$ 0.12	11.83 $\pm$ 0.87	30.47 $\pm$ 0.72	22.57 $\pm$ 0.87
Carbohydrates (g/100g)	14.09 $\pm$ 0.93	16.35 $\pm$ 0.71	14.67 $\pm$ 0.42	17.2 $\pm$ 0.69	16.94 $\pm$ 0.11	20.77 $\pm$ 0.35	16.81 $\pm$ 0.24	19.24 $\pm$ 0.48	14.16 $\pm$ 1.1	15.78 $\pm$ 0.41	14.67 $\pm$ 0.62	18.07 $\pm$ 0.29
Carotenes (mg/100g)	1.5 $\pm$ 0.01	1.4 $\pm$ 0.01	1.41 $\pm$ 0.02	0.82 $\pm$ 0.0	1.29 $\pm$ 0.0	0.34 $\pm$ 0.0	0.74 $\pm$ 0.0	1 $\pm$ 0.0	0.42 $\pm$ 0.0	0.86 $\pm$ 0.0	0.57 $\pm$ 0.0	0.78 $\pm$ 0.0
Crude fat (%)	2.02 $\pm$ 0.11	1.78 $\pm$ 0.04	2.73 $\pm$ 0.07	2.46 $\pm$ 0.09	5.61 $\pm$ 0.04	4.59 $\pm$ 0.19	2.46 $\pm$ 0.04	2.32 $\pm$ 0.04	2.23 $\pm$ 0.11	1.68 $\pm$ 0.07	5.44 $\pm$ 0.01	4.66 $\pm$ 0.11
Flavonoids (mg/g)	64.33 $\pm$ 13.9	106.04 $\pm$ 0.0	101.49 $\pm$ 21.60	84 $\pm$ 2.8	25.87 $\pm$ 0.4	16.78 $\pm$ 0.2	171.30 $\pm$ 56.71	52 $\pm$ 5.7	80.24 $\pm$ 8.4	12.98 $\pm$ 3.5	218.94 $\pm$ 31.82	47 $\pm$ 5.7
MC (% w.b.)	87.57 $\pm$ 0.49	85.81 $\pm$ 0.56	86.42 $\pm$ 1	85.92 $\pm$ 0.8	87 $\pm$ 0.65	85.92 $\pm$ 0.32	87.31 $\pm$ 0.25	86 $\pm$ 0.21	88.89 $\pm$ 1.47	86.93 $\pm$ 0.32	86.91 $\pm$ 0.28	85.26 $\pm$ 0.17
pH	4.56 $\pm$ 0.01	4.73 $\pm$ 0.02	4.87 $\pm$ 0.02	5.23 $\pm$ 0.15	4.49 $\pm$ 0.02	4.78 $\pm$ 0.01	4.29 $\pm$ 0.09	4.45 $\pm$ 0.02	4.53 $\pm$ 0.01	4.69 $\pm$ 0.01	4.49 $\pm$ 0.0	4.85 $\pm$ 0.04
Phenolics (mg/100g)	311.41 $\pm$ 5.1	191.77 $\pm$ 4.3	308.66 $\pm$ 6.5	134.33 $\pm$ 4.0	161 $\pm$ 7.0	124.33 $\pm$ 3.8	374.7 $\pm$ 10.7	308.33 $\pm$ 6.8	318.55 $\pm$ 3.5	170.49 $\pm$ 4.5	405.84 $\pm$ 4.3	243.5 $\pm$ 4.8
Proteins (%)	1.7 $\pm$ 0.07	1.2 $\pm$ 0.04	2.16 $\pm$ 0.06	1.12 $\pm$ 0.03	2.05 $\pm$ 0.07	1.18 $\pm$ 0.03	2.26 $\pm$ 0.09	1.26 $\pm$ 0.04	1.78 $\pm$ 0.14	1.26 $\pm$ 0.05	2.24 $\pm$ 0.09	1.45 $\pm$ 0.03
Reducing sugars (g/100g)	7.89 $\pm$ 0.17	6.75 $\pm$ 0.38	9.33 $\pm$ 0.51	10.84 $\pm$ 0.29	14.7 $\pm$ 0.59	17.68 $\pm$ 0.55	14.11 $\pm$ 0.41	16.2 $\pm$ 1.34	9.04 $\pm$ 0.07	10.4 $\pm$ 0.58	12.22 $\pm$ 0.71	9.7 $\pm$ 0.38
Tannins (mg/100g)	2.41 $\pm$ 0.11	1.65 $\pm$ 0.05	2.25 $\pm$ 0.28	0.88 $\pm$ 0.03	0.91 $\pm$ 0.02	0.76 $\pm$ 0.05	2.57 $\pm$ 0.07	2.36 $\pm$ 0.06	2.04 $\pm$ 0.06	1.11 $\pm$ 0.03	2.53 $\pm$ 0.14	1.86 $\pm$ 0.14
Total sugars (g/100g)	8.62 $\pm$ 0.15	12.05 $\pm$ 0.08	10.4 $\pm$ 1.27	13.07 $\pm$ 1.47	15.81 $\pm$ 0.5	19.88 $\pm$ 0.6	9.92 $\pm$ 0.59	11.19 $\pm$ 0.25	15.26 $\pm$ 1.26	17.81 $\pm$ 0.42	13.63 $\pm$ 0.23	17.05 $\pm$ 0.26
TSS ( $^{\circ}$ brix)	2.41 $\pm$ 0.11	1.65 $\pm$ 0.05	2.25 $\pm$ 0.28	0.88 $\pm$ 0.03	0.91 $\pm$ 0.02	0.76 $\pm$ 0.05	2.57 $\pm$ 0.07	2.36 $\pm$ 0.06	2.04 $\pm$ 0.06	1.11 $\pm$ 0.03	2.53 $\pm$ 0.14	1.86 $\pm$ 0.14
Vitamin C (mg/100g)	98.86 $\pm$ 5.1	181.71 $\pm$ 1.5	154.62 $\pm$ 3.1	252.2 $\pm$ 17.4	27.5 $\pm$ 2.3	91.12 $\pm$ 0.29	197.8 $\pm$ 12.3	305.5 $\pm$ 11.5	150.86 $\pm$ 2.6	305 $\pm$ 2.3	40.62 $\pm$ 2.3	69.24 $\pm$ 7.6

Number of replications employed = 3 and measurement protocols in Appendix A; MC - moisture content; TSS - total soluble solids; and values shown are mean  $\pm$  1 standard deviation.

Overall, significant fluctuations in the biochemical composition of cashew apples were observed between maturity stages 817 and 819, reflecting the typical ripening dynamics (Table 3). The decline in phenolics and flavonoids suggests reduced antioxidant activity with maturity (negative correlation between these compounds and fruit ripening), as reported in previous studies [63]. Conversely, the increase in sugar and vitamin C content indicates that with ripening, the palatability and nutritional appeal of cashew apples are improved. This is particularly evident in varieties Bhaskara and VRI-3. Therefore, it is expected that any visual or textural differences identified at these stages could be correlated with these biochemicals.

The selected quality parameters are essential for assessing the quality of the fruit; however, these methods are destructive in nature. All parameters require special instruments and an elaborate extraction procedure of the samples for analysis. Total phenols, tannins, flavonoids, antioxidant activity, sugars, carbohydrates, vitamin C, and carotenoids require solvent-based extraction of the targeted compounds followed by spectrophotometric or titrimetric estimation. Extraction and estimation process demand additional time and the use of expensive solvents and chemicals for analysis, especially for total phenols, tannins, flavonoids, antioxidant activity, sugars, carbohydrates, and carotenoids. Parameters such as carbohydrates, protein, and fat estimation involve extended extraction and thus require more time for estimation. Additionally, estimation of many of these quality parameters requires

the use of chemicals, which are corrosive and harmful to skin and eyes, which may also lead to systemic injury.

### 3.2. Independent image-based variables considered for analysis

By executing the developed ImageJ plugin with for a set of user inputs (Figure 6) with the training (75 %) batch of images of cashew apple samples (6 varieties; 2 sample types; 2 sample stages; 7–13 slice replications/variety; 4–6 whole replications/variety; 104 slices + 60 wholes; 164 total), the exhaustive 144 image-based parameters output was generated. Upon analyzing the variations of these parameters across the 15 wet chemistry physicochemical properties variables (dependent), several of the image-based variables (independent) exhibited no or insignificant variation. Therefore, the following set of 28 parameters was excluded (abbreviated names provided and expanded form presented in Table 1 for further analysis and model development: Ov-mxR, Ov-mxG, Ov-mxB, Ov-miR, Ov-miG, Ov-miB, Rn-Gmn, Ov-noR, Ov-noG, Ov-noB, Wn-no, Rn-no, L-pxGr, L-pxR, L-pxG, L-pxB, L-miGr, L-mxGr, L-medGr, L-miR, L-mxR, L-medR, L-miG, L-mxG, L-medG, L-miB, L-mxB, and L-medB. It should be noted that these parameters, when considered, would have resulted in the least correlation and have not generated practical models or those with the lowest performance. Thus, only the rest 116 image-based parameters ( $144 - 28 = 116$ ) were used for further analysis.

### 3.3. Plugin input variables combination selection

Given that user inputs and their combinations (Figure 6) influence the image-based parameters evaluated, the best inputs must be selected based on the collective performance of the prediction models. Even though five sets of image-based parameters, including color, waveness and roughness, GLCM, LBP, and color indices, were evaluated (Table 3), the input variables were only from the three sets (Figure 6), while other inputs were internally configured within the plugin. A systematic approach was employed to evaluate and identify the most suitable combination of input variables (color: columns and rows; waveness and roughness: cutoff value; and GLCM: step size and step direction) entered through the plugin's input panel (Figure 6). Every combination of the input parameters for the 4 categorical variables (types and stages), 15 dependent wet chemistry variables, with the 116 dependent parameters resulted in the generation of a total of 6960 ( $2 \times 2 \times 15 \times 116$ ) simple linear models generated internally. However, considering only the top three, totaling 180 ( $2 \times 2 \times 15 \times 3$ ), models within each category were utilized for further analysis. The sum of the coefficient of determination ( $R^2$ ) of the top three 180 models ( $\Sigma R^2$ ), with the largest value serving as an indicator for selecting the most suitable combination of input variables (Table 4).

Therefore, initially the color inputs, namely the grid size, were varied to  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $10 \times 10$ ,  $12 \times 12$ , and  $15 \times 15$ , while keeping the other inputs at their starting values. It was observed that the  $\Sigma R^2$  values attained their peak at  $7 \times 7$  and subsequently declined. Now keeping  $7 \times 7$  as a constant input for color inputs, the waveness and roughness Gaussian filter radius was gradually varied to 4, 5, 10, 15, 20, 25, 30, 35, 40, and 50 pixels. The effects of these variations were recorded, and the 35 pixels yielded the best outcome (Table 4). Keeping these two inputs constant, the GLCM step size was varied gradually to 1, 2, 3, 5, and 10 pixels, which revealed that the 1 pixel step size was the best. Proceeding similarly, a combination of 4 GLCM step direction angles ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ ) was studied, and found that  $270^\circ$  was the best (Table 4). To further validate the effectiveness of the color grid size in conjunction with the selected input combination, the models were subjected to testing for  $5 \times 5$ ,  $7 \times 7$ ,  $10 \times 10$ , a step above and below the considered  $7 \times 7$ , which reaffirmed its effectiveness with the highest value of  $\Sigma R^2$  of 76.437 (Table 4). Consequently, the best plugin input combination that is used for further analysis is: a color grid size of  $7 \times 7$ , a waveness and roughness cutoff of 35 pixels, a GLCM step size of 1 pixel, and a GLCM step direction of  $270^\circ$ . The CPU processing time for this combination was  $2.87 \pm 0.79$  s/image (HP ZBook Fury 16-inch G11 Mobile Workstation PC; Intel® Core™ i9 13950HX 5.5 GHz, 32 GB RAM, 1 TB storage).

The results underline the importance of input selection in modeling with multiple variables. The optimization strategy—based on cumulative  $\Sigma R^2$  values, proved effective in isolating input

**Table 4.** Image analysis input parameters selection based on the sum of  $R^2$  based on the top three sets of 180 models.

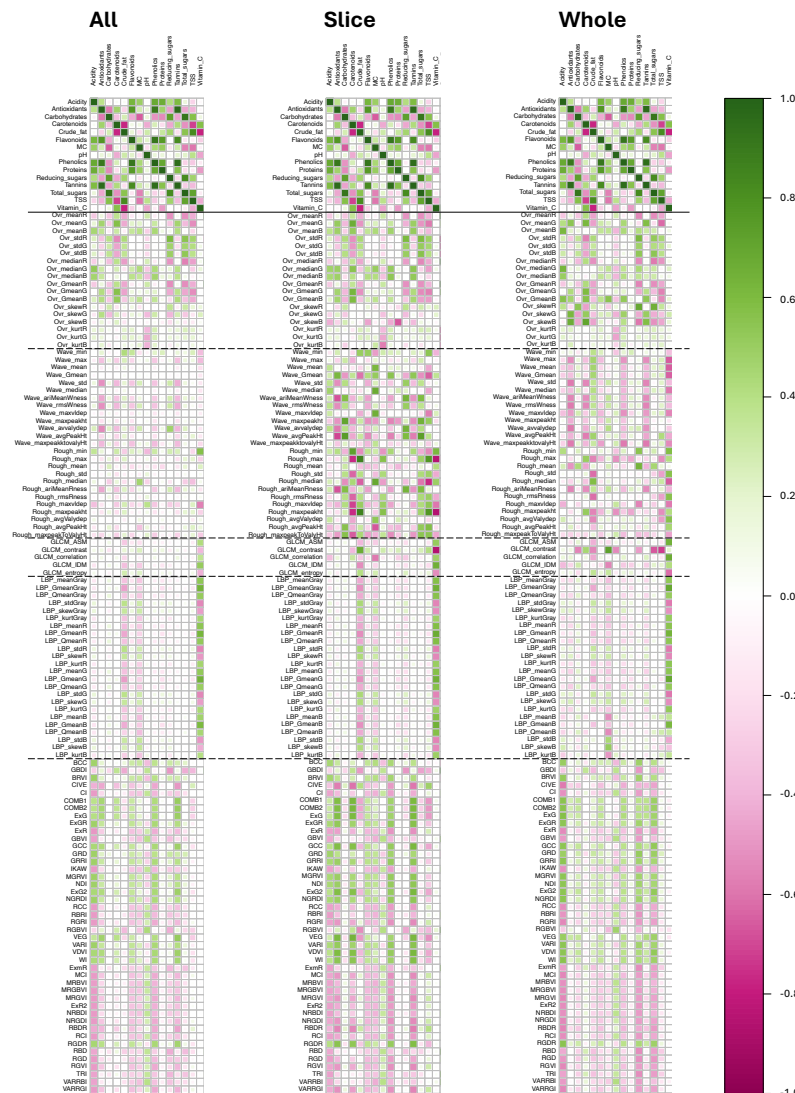
Selected input	Input combination	$\Sigma R^2$	$\Sigma R^2$ change from *reference (%)
Waveness cutoff value	$7 \times 7$ ; 4 pix; 1 pix; $0^\circ$	73.900*	0.00
	$7 \times 7$ ; 5 pix; 1 pix; $0^\circ$	73.900	0.00
	$7 \times 7$ ; 10 pix; 1 pix; $0^\circ$	75.003	1.49
	$7 \times 7$ ; 15 pix; 1 pix; $0^\circ$	75.505	2.17
	$7 \times 7$ ; 20 pix; 1 pix; $0^\circ$	75.805	2.58
	$7 \times 7$ ; 25 pix; 1 pix; $0^\circ$	75.941	2.76
	$7 \times 7$ ; 30 pix; 1 pix; $0^\circ$	76.014	2.86
	$7 \times 7$ ; 35 pix***; 1 pix; $0^\circ$	76.043**	2.90**
	$7 \times 7$ ; 40 pix; 1 pix; $0^\circ$	75.966	2.80
	$7 \times 7$ ; 50; pix 1 pix; $0^\circ$	75.966	2.80
GCLM step size	$7 \times 7$ ; 35 pix; 1 pix***; $0^\circ$	76.043**	2.90**
	$7 \times 7$ ; 35 pix; 2 pix; $0^\circ$	76.023	2.87
	$7 \times 7$ ; 35 pix; 3 pix; $0^\circ$	76.024	2.87
	$7 \times 7$ ; 35 pix; 5 pix; $0^\circ$	76.031	2.88
	$7 \times 7$ ; 35 pix; 10 pix; $0^\circ$	75.933	2.75
GCLM step direction	$7 \times 7$ ; 35 pix; 1 pix; $0^\circ$	76.043	2.90
	$7 \times 7$ ; 35 pix; 1 pix; $90^\circ$	76.434	3.43
	$7 \times 7$ ; 35 pix; 1 pix; $180^\circ$	76.056	2.92
	$7 \times 7$ ; 35 pix; 1 pix; $270^\circ$ ***	76.437**	3.43**
Color grid size	$5 \times 5$ ; 35 pix; 1 pix; $270^\circ$	74.670	1.04
	$^\dagger 7 \times 7$ ***; 35 pix; 1 pix; $270^\circ$	76.437**	3.43**
	$10 \times 10$ ; 35 pix; 1 pix; $270^\circ$	73.751	-0.20

\* Reference  $\Sigma R^2$  value from 180 models; \*\* maximum value at each input block; \*\*\* selected combination of image analysis input variables at different evaluation stages; and  $^\dagger$  final selected input variable combination used for further analysis.

combinations that maximized the predictive performance across the multiple dependent variables. This approach is particularly valuable when dealing with complex systems where interactions among variables are non-linear or context-dependent. This methodology offers a framework for model refinement. Thus, by iteratively evaluating input combinations, one could converge on combinations that yield the most reliable and generalizable models. However, caution should be exercised when using  $\Sigma R^2$  as a heuristic for input selection, as it may not fully capture nuances such as overfitting, multicollinearity, or model interpretability. Therefore, further research on additional criteria such as adjusted  $\Sigma R^2$ , or cross-validation error, etc., is necessary. Additionally, applying a larger data set (with more varieties and maturity stages), and non-linear modeling techniques may enhance the predictive capacity where linear assumptions are insufficient.

### 3.4. Visualizing Correlation—Correlation Diagram And Absolute Correlation Coefficients

Pearson's correlation coefficient ( $r$ ) quantifies the strength and direction of 15 wet chemistry physicochemical properties with the 116 image-based parameters, and themselves were evaluated using the input combination ( $7 \times 7$ ; 35 pix; 1 pix;  $270^\circ$ ; Section 3.3) and visualized in the form of a correlation diagram (Figure 7). The correlations were calculated separately for (i) all sample types combined, (ii) slice images only, and (iii) whole fruit images only to determine the effect of the sample type category during model development. Table 5 presents the sum and mean absolute correlation coefficient ( $\bar{r}$ ) values, which facilitate the comparison of the performance of sample types and groups of parameters. The correlation results (Figure 7 and Table 5) indicate that combining both sample types is not an effective strategy ( $\bar{r} = 0.225$ ). Although keeping the types separate yielded a better correlation, the slice ( $\bar{r} = 0.288$ ) performed superiorly compared to the entire sample ( $\bar{r} = 0.274$ ).



**Figure 7.** Correlation diagram for cashew apple sample types demonstrating the correlation of wet chemistry and image-based parameters. Solid horizontal line separates interrelationships among wet chemistry, and dotted horizontal lines separate different parameter groups of methodologies.

**Table 5.** Effect of sample types and analysis groups in correlation analysis of wet chemistry with image-based parameters through absolute and mean Pearson's correlation coefficient ( $r$  and  $\bar{r}$ ).

Parameters group	N	All samples		Slice samples		Whole samples	
		Sum	Mean	Sum	Mean	Sum	Mean
Wet chemistry	225	99.6	0.443	99.6	0.443	99.6	0.443
Color grid	270	61.0	0.226	67.8	0.251	72.2	0.267
Waveness and roughness	375	56.7	0.151	121.6	0.324	95.1	0.254
GLCM	75	9.6	0.128	16.0	0.213	17.5	0.234
LBP	360	59.3	0.165	68.2	0.189	74.0	0.206
Color indices	660	156.5	0.237	193.0	0.292	179.7	0.272
Total image-based	1740	343.1	0.197	466.5	0.268	438.6	0.252
Total all methods	1965	442.7	0.225	566.1	0.288	538.2	0.274

Data used represent absolute  $r$  values with stages (817 and 819) combined;  $N$  is the number of parameters in each group; GLCM is gray level co-occurrence matrix; LBP is local binary pattern; Total represents all the combinations of  $r$  values of wet chemistry (15) with both wet chemistry and image-based (15 + 116) parameters.

### 3.4.1. Interrelationships among wet chemistry parameters

Overall, the correlation diagram (Figure 7; separated by a solid horizontal line) shows better intercorrelation among wet chemistry variables compared to the image-based parameters, evident from the darker-colored tiles with wet chemistry. This observation is corroborated by the absolute mean  $\bar{r}$  of 0.433, which is greater than the range of 0.128–0.324 observed for the image-based parameters (Table 5). Strong interrelationships were observed among key physicochemical properties, specifically between crude fat, vitamin C, antioxidants, and total phenol content; carbohydrates and total sugars; carotenoids and total sugars; as well as tannins and antioxidants. These robust associations also suggest, apart from serving as direct predictors, that certain physicochemical properties may serve as indirect predictors for others, which can be beneficial in gaining insights into other parameters. For example, if vitamin C shows poor correlation with texture features but is strongly linked to crude fat, then crude fat, if reliably predicted from image-derived parameters, could act as a proxy for estimating vitamin C levels. Despite these specific advantages of direct prediction of physicochemical properties from others, the wet chemistry determination is time-consuming and highly expensive.

This strategy of leveraging internal biochemical correlations to enhance predictive modeling aligns with previous studies. Hasanzadeh et al. [64] demonstrated that hyperspectral imaging combined with regression analysis could accurately estimate apple quality traits such as pH, soluble solids, and titratable acidity, with correlation coefficients exceeding 0.99. Similarly, Paulo et al. [65] demonstrated that visual descriptors extracted from fruit images captured in uncontrolled environments could be mapped to physicochemical parameters using non-linear regression models, thereby effectively estimating fruit maturity.

### 3.4.2. Correlation between wet chemistry and image-based parameters

The five groups of image-based parameters collectively had a lower correlation with wet chemistry compared to the intercorrelation of wet chemistry. While all sample images of image-based parameters were consistently low, the slice and whole samples were the best for certain groups (Table 5). For example, slices exhibited better correlation with waveness and roughness, LBP, and color indices, whereas wholes exhibited better correlation with the color grid and GLCM.

Among the image-based parameters group methodologies, slice sample images exhibited the highest correlation with surface morphology (waveness and roughness;  $\bar{r} = 0.324$ ), followed by color indices ( $\bar{r} = 0.292$ ), and whole samples color grid ( $\bar{r} = 0.267$ ), followed by GCLM ( $\bar{r} = 0.234$ ) and LBP ( $\bar{r} = 0.206$ ) were the best (Table 5). It is interesting to note that the color indices, which are easy to obtain as they are derived solely from R, G, and B values, also exhibit favorable correlation for slice and whole samples ( $0.272 \leq \bar{r} \leq 0.292$ ). This observation can be visualized in the correlation diagram (Figure 7) with methodology groups separated by dotted horizontal lines and darker color patches. Correlations derived from the whole fruit sample images indicated that, although associations were generally stronger than the combined dataset, they were consistently weaker than those observed

for slice images. Notably, several wet chemistry parameters, including vitamin C, TSS, total sugars, crude fat, and moisture content, exhibited moderately higher correlations with GLCM contrast and roughness-based texture metrics.

Overall, the  $\bar{r}$  for image-based parameters for slice, whole, and combined samples were 0.268, 0.252, and 0.197, respectively, which is lower than the inter-correlation of wet chemistry value of 0.443 as observed previously. Collectively, these findings reinforce the potential of image-based parameters employing computer vision analysis not only for direct prediction of physicochemical properties but also for inferring quality parameters through correlated biochemical proxies. Thus, it expands the scope for the development of non-destructive quality assessment tools in postharvest and processing operations.

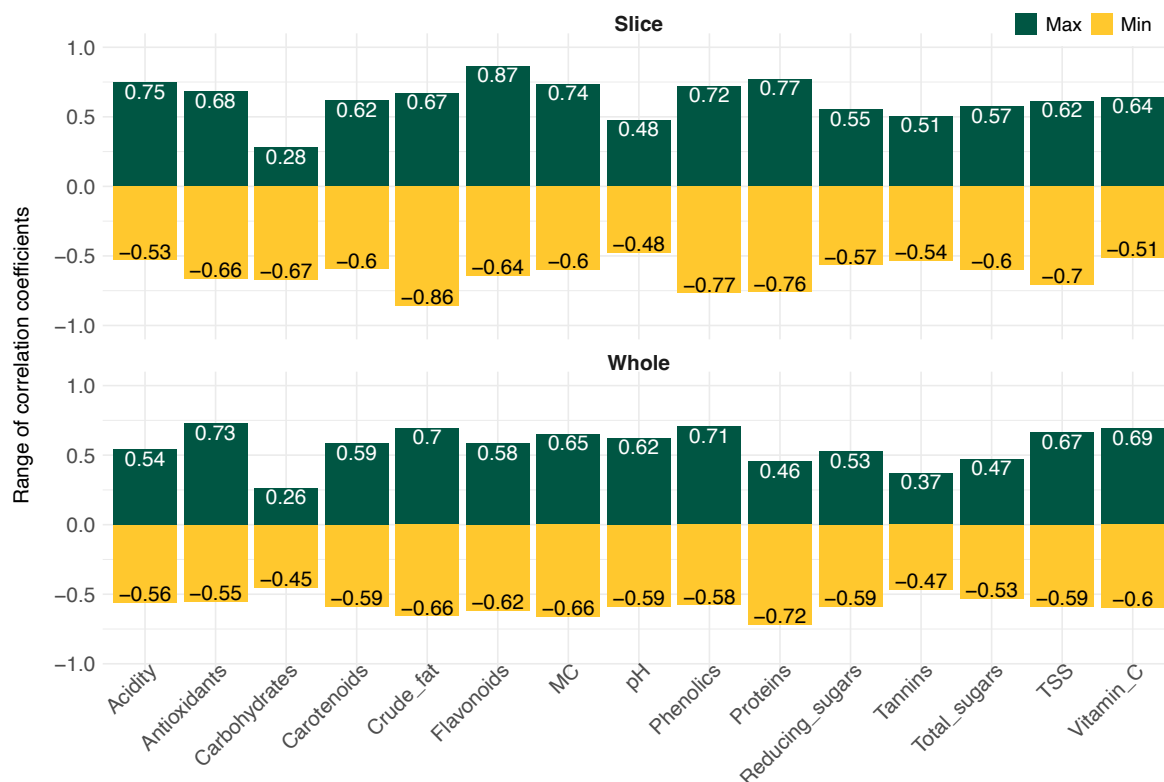
### 3.5. Correlation analysis of wet chemistry with image-based parameters—value ranges

The range of  $r$  for all the 116 image-based parameters (independent variables) against the 15 wet chemistry (dependent variables) for the finalized input combination, grouped by sample types (slice and whole) is presented in Figure 8. These categories are distinct from the image viewpoint, and their correlations are treated separately (Table 5). For slice samples,  $r$  ranged from  $-0.86$  to  $0.87$ , indicating both strong negative and positive associations. In contrast, for whole samples, the range was slightly lower and narrower, from  $r = -0.72$  to  $0.73$ , suggesting that the independent parameters computed based on slice images exhibited enhanced sensitivity to the image-based parameters.

Contrasting the association of wet chemistry parameters with image-based parameters between sample types (slice and whole), the parameters derived from slice images revealed stronger correlations for most parameters and in “internal quality” traits, particularly those related to biochemical composition (e.g., flavonoids, crude fat, proteins, total phenolics content, acidity, and moisture content). This supports our hypothesis that exposing internal tissues (through slice images) enhances the accuracy of indirect measurements, especially for parameters not uniformly distributed across the fruit. Whole fruit images favored parameters linked to “external features” (e.g., antioxidant activity and vitamin C), but underperformed for most internal chemical traits. The reduced correlation range suggests that whole fruit assessments may require more advanced imaging or hybrid sensing techniques to improve predictive power. The observed correlation patterns underscore the potential of using image-based data to predict wet chemistry traits, especially with images of sliced fruits is feasible. For improved correlation, a larger sample size for training may account for better results.

For slice sample images, the most strongly correlated dependent parameters ( $r > 0.7$ ) estimated were flavonoids (0.866), crude fat ( $-0.859$ ), proteins (0.772), total phenolic content ( $-0.768$ ), acidity (0.749), moisture content (0.736), and TSS ( $-0.705$ ). The least correlated parameters were pH, tannins, reducing sugars, and total sugars, with coefficients of  $-0.476$ ,  $-0.535$ ,  $0.553$ , and  $0.566$ , respectively. For whole fruit sample images, the most strongly correlated dependent parameters were antioxidant activity ( $r = 0.729$ ), proteins ( $-0.720$ ), and total phenolics (0.706). The moderately correlated parameters were crude fat (0.696), vitamin C (0.691), TSS (0.666), moisture content ( $-0.662$ ), and pH (0.623). The least correlated variables were total carbohydrates, tannins, total sugars, reducing sugars, and carotenoids, with correlation coefficients of  $-0.450$ ,  $-0.468$ ,  $-0.530$ ,  $-0.587$ ,  $0.587$ , respectively.

Overall, among the wet chemistry parameters, flavonoids and crude fat consistently exhibited high correlation ( $r > 0.8$ ) with the plugin image-based parameters. Proteins, total phenolic content, acidity, moisture content, antioxidant activity, carotenoids, and TSS demonstrated moderate to strong correlations ( $r = 0.6$  to  $0.77$ ) with texture and color-based parameters, especially where internal tissue exposure improves detection. Total carbohydrates, tannins, total sugars, and reducing sugars showed weak correlations ( $r < 0.6$ ) across most independent variables, suggesting limited predictability from image-based non-destructive measurements. Furthermore, the correlation coefficients between dependent and independent variables were consistently higher for slice-derived independent parameters compared to those obtained from whole-fruit images. Focusing on analyzing the dynamics of the maturity stages, varieties, and collecting more samples could contribute to improving the association between the independent and dependent parameters.



**Figure 8.** Range of image-based parameters correlation coefficients with physicochemical properties wet chemistry parameters.

### 3.6. Top ranking image-based parameters correlation coefficients

To identify the most influential image-based parameters predictors for each wet chemistry physicochemical properties trait, they were ranked based on their Pearson correlation coefficients, and the top ten features of wet chemistry categorized by sample type (slice and whole), are presented in Table 6. This ranking framework enabled the identification of dominant image patterns and feature categories associated with specific chemical and physical attributes of the fruits. In general, slice and whole samples had a distinctly different set of top-ranking parameters for the same fruit physicochemical properties due to differences in color, texture, and morphology.

**Table 6.** Slice and whole cashew apple top-10 ranking for the wet chemistry properties based on correlation coefficients of image-based independent parameters.

Wet chemistry properties	Ranks of image-based independent variables based on correlation ( <i>r</i> ) values									
	1	2	3	4	5	6	7	8	9	10
<b>Slice</b>										
Acidity	Wn-avPH	Wn-mxPH	Wn-amWn	Rn-amRn	Rn-xpvh	Rn-avPH	Rn-med	Wn-Gmn	Ov-stdR	Rn-mxVD
Antioxidants	Rn-med	Wn-med	Rn-avPH	Wn-mxVD	Wn-mn	Gl-corr	Rn-avVD	Wn-Gmn	Ov-medG	Wn-mxPH
Carbohydrates	Ov-skB	Wn-xPt	Rn-mn	Rn-avPH	Rn-xpvh	Wn-med	Gl-corr	Gl-cont	Rn-avVD	Wn-avPH
Carotenes	VDVI	GCC	ExG2	VEG	COMB1	COMB2	ExG	CIVE	RGDR	Wn-avPH
Crude fat	Gl-cont	Rn-mxPHt	Rn-mx	L-GmnR	L-GmnG	L-GmnGr	Gl-IDM	L-GmnB	Rn-mi	Rn-mxVD
Flavonoids	Rn-mxPH	Rn-mx	Gl-cont	Rn-med	Wn-Gmn	L-GmnR	Gl-IDM	RGBVI	Rn-mi	L-GmnG
MC	Wn-avPH	Wn-mxPH	Rn-med	Wn-amWn	Rn-amRn	Ov-stdR	Ov-stdB	Rn-xpvh	Ov-medR	Ov-GmnG
pH	Wn-med	GRD	RGD	Ov-medG	RBRI	ExmR	TRI	RCC	ExR2	BRVI
Phenolics	Rn-mx	Rn-mxPH	Rn-mi	Rn-mxVD	Ov-mnG	Rn-xpvh	Rn-amRn	Rn-std	VEG	Ov-GmnG
Proteins	Rn-mx	Rn-mxPH	Rn-med	Wn-Gmn	Ov-mnG	Rn-avPH	Wn-mxVD	Ov-GmnG	Rn-xpvh	Wn-mi
Reducing sugars	Wn-avPH	Wn-mxPH	Wn-Gmn	Ov-medB	Rn-xpvh	ExGR	Ov-medG	VARI	VARRGI	RGRI
Tannins	Wn-avPH	Wn-mxPH	Wn-mi	Rn-xpvh	Wn-amWn	Rn-amRn	Ov-skB	Wn-rsWn	Wn-std	Wn-Gmn
Total Sugars	Ov-medG	RBRI	BRVI	IKAW	NRBDI	BCC	CI	MRBVI	MRGBVI	RCC
TSS	Wn-avPH	Wn-amWn	Rn-amRn	Wn-Gmn	VEG	VDVI	GCC	ExG2	COMB1	COMB2
Vitamin C	Wn-avVD	Wn-amWn	Rn-amRn	Ov-stdR	Wn-avPH	Ov-GmnR	Wn-rsWn	Wn-std	Ov-mnR	Ov-stdB
<b>Whole</b>										
Acidity	Gl-cont	Ov-skR	Ov-skG	Ov-skB	Rn-mn	Ov-medR	Ov-stdR	Ov-stdB	VEG	GCC
Antioxidants	Gl-cont	Ov-GmnB	Gl-IDM	L-QmnB	L-mnB	L-skB	L-kuB	L-stdB	Wn-avVD	RGBVI
Carbohydrates	Rn-mxPH	Rn-mx	Gl-IDM	Wn-mi	Wn-mxVD	Wn-avVD	Wn-amWn	Rn-amRn	L-GmnG	L-GmnGr
Carotenes	Wn-xpvh	Ov-mnG	Ov-skB	Wn-rsWn	Wn-std	Ov-GmnB	Wn-amWn	Rn-amRn	Wn-mx	Wn-avVD
Crude fat	L-GmnG	L-GmnGr	Gl-corr	Rn-med	Wn-mn	Wn-mx	Wn-Gmn	Gl-ASM	L-GmnR	L-GmnB
Flavonoids	Rn-rsRn	Rn-std	Rn-med	Rn-mxPH	Rn-mx	Rn-avVD	Gl-cont	Ov-medR	Wn-avPH	Wn-mn
MC	Gl-cont	Ov-skR	Ov-medR	Ov-stdB	Ov-stdR	Ov-GmnR	Ov-mnR	GCC	ExG2	VDVI
pH	Ov-medG	Ov-mnB	CIVE	ExG	Ov-GmnB	WI	RBDR	COMB1	COMB2	RGDR
Phenolics	Ov-skB	Ov-mnG	Ov-GmnG	Wn-avPH	Ov-GmnB	Ov-stdR	Wn-rsWn	Wn-std	Wn-amWn	Rn-amRn
Proteins	Gl-cont	Ov-mnG	Ov-GmnG	Ov-GmnB	Rn-mxPH	Rn-mx	Ov-medR	Rn-avVD	Rn-std	Rn-rsRn
Reducing sugars	Wn-xpvh	Ov-GmnB	Wn-mxVD	Wn-mx	Ov-skG	Ov-mnG	Ov-mnB	CIVE	ExG	Wn-rsWn
Tannins	Wn-xpvh	Rn-mx	Rn-mxPH	Wn-mxVD	Ov-GmnB	Wn-mx	Ov-medB	CIVE	Ov-mnB	ExG
Total Sugars	Ov-GmnB	Ov-mnB	CIVE	ExG	L-kuB	Ov-medG	COMB2	COMB1	L-skB	VDVI
TSS	Ov-skB	Ov-skG	Wn-xpvh	Wn-rsWn	Wn-std	Wn-avVD	Wn-amWn	Rn-amRn	Ov-mnG	Wn-mx
Vitamin C	Ov-skR	Rn-mn	Ov-stdR	Ov-skB	Rn-mi	Rn-mxVD	Ov-GmnR	Ov-mnR	Ov-medR	Ov-stdB

Short forms for the image-based independent parameters were used. They can be combined to make the complete image-based independent property names. Table 1 list the full form. Short forms utilized are: am - arithmetic mean, av - average, B - blue, cont - contrast, corr - correlation, G - green, Gl - gray level co-occurrence matrix, Gmn - geometric mean, Gr - gray, IDM - inverse difference moment, kurt - kurtosis, L - local binary pattern, med - median, mi - minimum, mx - maximum, Ov - overall, PH - peak height, pvh - peak to valley height, rs - root mean square, R - red, Rn - roughness, sk - skewness, std - standard deviation, VD - valley depth, and Wn - waveness; color-based indices: BCC, CIVE, COMB1, COMB2, ExG, ExG2, GCC, GRD, MRGBVI, RBDR, RGD, RGDR, RGBVI, VDVI, VEG, and WI are available from Table 2; MC - moisture content; and TSS - total soluble solids.

Among the 150 top 10 ranked image-based parameters for slice samples (15 wet chemistry  $\times$  10 ranks; Table 6), the surface methodology covered 52.0 % followed by color indices (25.3 %), color grid (14.0 %), GLCM (4.7 %), and LBP (4.0 %); and for whole samples surface methodology covered 38.0 % followed by color grid (32.6 %), color indices (14.7 %), LBP (8.7 %), and GLCM (6.0 %). Across the 15 physicochemical properties, surface morphology descriptors (waveness and roughness metrics)

consistently ranked among the top predictors, followed by color-based indices or grids, and finally GLCM and LGB methodologies with similar low performance. Among these top-ranked parameters, the color indices developed in this study (17; Table 2) performed well. The total number of color indices featured in the top-ranking 150 parameters was 22 for each slice and whole sample (Table 6). Out of these, the number of indices developed in this study was 9 for the slice and 3 for the whole samples. This, out of the total 150, was 6.0 % (slice) and 2.0 % (whole), and out of color-based indices was 40.9 % (slice); 13.6 % (whole), indicating a significant performance of the indices developed in this study.

Color-based indices methodology, which is easy to derive, also showed strong associations with biochemical properties such as phenolics and carotenoids, indicating that reflectance captured in RGB images can serve as a proxy for phenolics and carotenoids profiling. Color grid methodology parameters (e.g., mean, median, STD of blue, green, and blue channels) were particularly informative for sugar-related traits such as TSS, total sugars, and reducing sugars. Acidity and pH had dominating associations with surface morphology-based parameters and color-based indices. These results agree with earlier studies [66,67] that have demonstrated a direct correlation between color values and textural properties are directly correlated to various physicochemical properties such as acidity, pH, and phenolics.

### 3.7. Wet Chemistry physicochemical properties Prediction Model Development

The Table 7 presents the top three models, assessed based on the model performance parameter  $R^2$ , for each wet chemistry physicochemical properties parameter across all sample types and maturity stages (total = 180 models; Section 3.3), from the complete set of 116 image-based parameters independent variables. It is important to note that image-based parameters featuring in the top models (based on  $R^2$ ; Table 7) may differ from the top ten highly correlated image-based parameters (based on  $r$ ; Table 6). This is because there are differences in the linear model fitting and Pearson's correlation evaluation. While the correlation coefficient evaluates the strength of association between the independent and dependent variables, the model development provides the actual tool used for making the final prediction. Therefore, the latter (Table 7) is more suitable for the objective of the study.

Across all the wet chemistry physicochemical properties parameters, the  $R^2$  values, which represent the models' performance, ranged from as high as 0.936 to as low as 0.167 for both maturity stages, indicating varying degrees of predictive strength depending on the dependent parameter, sample type, and maturity stage categories. For slice samples, the mean  $R^2$  values for stage 817, stage 819, and both are  $0.42 \pm 0.18$ ,  $0.48 \pm 0.18$ , and  $0.45 \pm 0.17$ ; and these whole samples are  $0.37 \pm 0.12$ ,  $0.35 \pm 0.12$ , and  $0.36 \pm 0.12$ , respectively. The mean  $R^2$  for the samples and stages combined is  $0.408 \pm 0.149$ . Considering the better-performing models ( $R^2 > 0.5$ ), it was found that slices covered 24.4 % with stage 817 and 51.1 % with stage 819, and for wholes 6.7 % and 13.3 % coverage, respectively. These results indicate that the slice images and ultimate maturity stage of 819 are the most effective for evaluating the wet chemistry, hence recommended for the computer vision analysis of physicochemical properties analysis. This combination will also yield 24.4 % of high-performance models ( $R^2 > 0.6$ ). When whole samples are required, the ultimate stage of 819 should be chosen again.

The models of slice and the maturity stage 819 combination ( $R^2 > 0.5$ ) will cover 11 out of 15 (73.3 %) wet chemistry physicochemical properties. These wet chemistry properties that can be successfully modeled are: acidity, antioxidant, carbohydrates, carotenoids, crude fat, flavonoids, pH, phenolics, proteins, tannins, and vitamin C. The wet chemistry properties covered by whole samples were only five (acidity, carbohydrates, carotenoids, proteins, and vitamin C), which is a subset of the "slice+819" combination. The "slice+819" combination covered again five wet chemistry properties (carotenoids, crude fat, pH, TSS and and vitamin C), whereas the property TSS was the unique addition to the list of slice



Table 7. Cashew apple slice and whole top-three ranked linear models of wet chemistry parameters with image-based parameters as independent variables from the training image dataset.

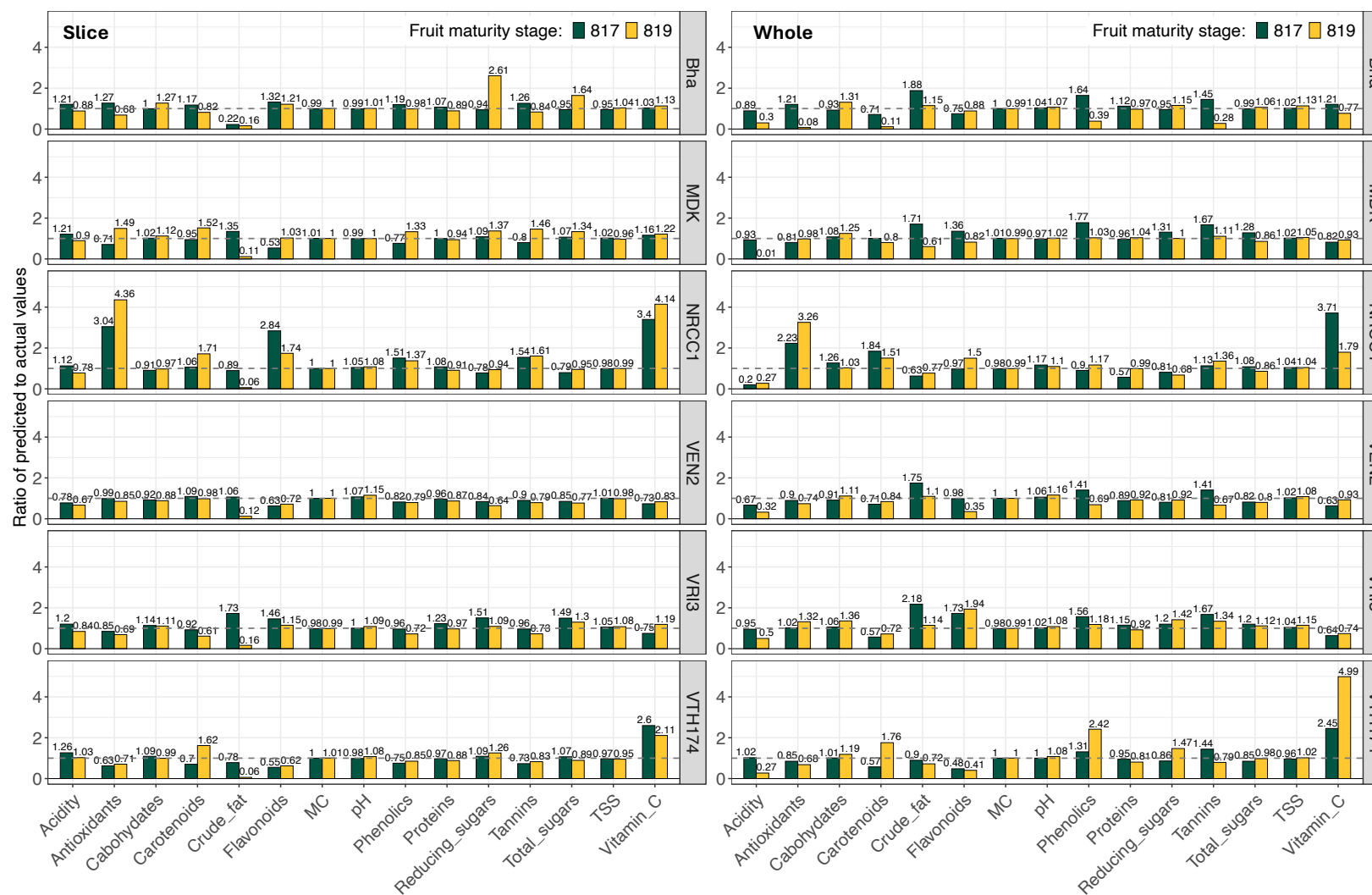
Wet chem. (y)	x	Model (R <sup>2</sup> )	x	Model (R <sup>2</sup> )	Wet chem. (y)	x	Model (R <sup>2</sup> )	x	Model (R <sup>2</sup> )
<b>Slice</b>									
<b>Stage = 817</b>					<b>Stage = 819</b>				
Acidity	Ov-medR	-0.0043x + 0.9284; (R <sup>2</sup> = 0.223)	WI	0.229x - 0.235; (R <sup>2</sup> = 0.669)	Antioxidant	VDVI	481.89x - 0.2153; (R <sup>2</sup> = 0.429)	VEG	170.16x - 17.33; (R <sup>2</sup> = 0.505)
	Ov-mnR	-0.003x + 0.6866; (R <sup>2</sup> = 0.206)	RBDR	-0.229x - 0.464; (R <sup>2</sup> = 0.669)		GCC	1044.72x - 347.99; (R <sup>2</sup> = 0.426)	Ov-mnG	1.9564x - 233.88; (R <sup>2</sup> = 0.394)
	Ov-GmnR	-0.0028x + 0.6437; (R <sup>2</sup> = 0.206)	Ov-medG	0.0154x - 2.0769; (R <sup>2</sup> = 0.651)		ExG2	348.24x + 0.2505; (R <sup>2</sup> = 0.426)	COMB2	12.68x - 120.77; (R <sup>2</sup> = 0.385)
Carbohydrates	Ov-stdR	0.1099x + 13.42; (R <sup>2</sup> = 0.240)	Ov-skG	-6.5953x + 2.4957; (R <sup>2</sup> = 0.672)	Carotenoids	VEG	3.8925x + 0.0719; (R <sup>2</sup> = 0.936)	Rn-xpH	-0.0206x + 2.1379; (R <sup>2</sup> = 0.502)
	Ov-GmnB	-0.0583x + 18.42; (R <sup>2</sup> = 0.232)	Ov-kuB	-0.0009x + 18.59; (R <sup>2</sup> = 0.216)		COMB1	0.5129x - 2.9892; (R <sup>2</sup> = 0.930)	Rn-mx	-0.0205x + 2.147; (R <sup>2</sup> = 0.501)
	Ov-stdB	0.0722x + 13.33; (R <sup>2</sup> = 0.204)	Ov-stdR	0.1938x + 14.22; (R <sup>2</sup> = 0.167)		COMB2	0.3559x - 3.0132; (R <sup>2</sup> = 0.930)	Rn-mpvh	-0.0214x + 2.5627; (R <sup>2</sup> = 0.428)
Crude fat	VEG	-12.76x + 5.994; (R <sup>2</sup> = 0.542)	Gl-cont	0.1355x + 0.255; (R <sup>2</sup> = 0.518)	Flavonoids	ExmR	-2.0486x + 424.89; (R <sup>2</sup> = 0.195)	VEG	550.93x - 67.64; (R <sup>2</sup> = 0.509)
	COMB2	-1.121x + 15.60; (R <sup>2</sup> = 0.497)	Rn-mx	0.0608x - 0.5068; (R <sup>2</sup> = 0.503)		TRI	-3.9558x + 443.18; (R <sup>2</sup> = 0.193)	COMB2	44.00x - 434.37; (R <sup>2</sup> = 0.446)
	COMB1	-1.6147x + 15.53; (R <sup>2</sup> = 0.496)	Rn-xpH	0.061x - 0.4769; (R <sup>2</sup> = 0.503)		RBD	-3.3652x + 481.89; (R <sup>2</sup> = 0.181)	COMB1	62.76x - 426.48; (R <sup>2</sup> = 0.444)
MC	Ov-kuB	0.0006x + 80.00; (R <sup>2</sup> = 0.430)	Rn-mx	-0.0218x + 87.18; (R <sup>2</sup> = 0.385)	pH	RBD	0.0166x + 2.7368; (R <sup>2</sup> = 0.509)	Ov-GmnB	-0.07x + 8.1414; (R <sup>2</sup> = 0.612)
	RBD	-0.0604x + 93.88; (R <sup>2</sup> = 0.358)	Rn-mxPH	-0.0219x + 87.17; (R <sup>2</sup> = 0.385)		TRI	0.0186x + 2.9997; (R <sup>2</sup> = 0.495)	CIVE	0.1321x + 0.40447; (R <sup>2</sup> = 0.608)
	TRI	-0.0685x + 92.98; (R <sup>2</sup> = 0.356)	Rn-mpvh	-0.0224x + 87.60; (R <sup>2</sup> = 0.318)		ExmR	0.0095x + 3.1075; (R <sup>2</sup> = 0.483)	ExG	-0.0658x + 6.2548; (R <sup>2</sup> = 0.606)
Phenolics	GRD	6.4798x + 586.02; (R <sup>2</sup> = 0.273)	VEG	1840.87x - 152.75; (R <sup>2</sup> = 0.686)	Proteins	Ov-GmnB	-0.0147x + 2.8153; (R <sup>2</sup> = 0.426)	Ov-kuR	0.0017x + 0.9997; (R <sup>2</sup> = 0.674)
	RGD	-6.4798x + 586.02; (R <sup>2</sup> = 0.273)	COMB2	151.10x - 1421.89; (R <sup>2</sup> = 0.635)		Ov-stdB	0.0191x + 1.5097; (R <sup>2</sup> = 0.414)	Ov-kuG	0.0009x + 1.1169; (R <sup>2</sup> = 0.570)
	ExmR	-2.8428x + 741.30; (R <sup>2</sup> = 0.255)	COMB1	215.56x - 1395.25; (R <sup>2</sup> = 0.633)		Ov-mnB	-0.0216x + 3.4392; (R <sup>2</sup> = 0.308)	Ov-medR	-0.015x + 3.8487; (R <sup>2</sup> = 0.470)
R-sugars	Ov-stdR	0.3289x + 5.6368; (R <sup>2</sup> = 0.463)	Ov-skG	-12.74x - 17.62; (R <sup>2</sup> = 0.477)	Tannins	VDVI	33.54x + 0.509; (R <sup>2</sup> = 0.466)	VEG	16.07x - 1.5984; (R <sup>2</sup> = 0.678)
	Ov-stdB	0.2153x + 5.3745; (R <sup>2</sup> = 0.391)	Ov-stdR	0.6897x - 0.9805; (R <sup>2</sup> = 0.401)		GCC	72.71x - 23.70; (R <sup>2</sup> = 0.463)	COMB2	1.2671x - 12.11; (R <sup>2</sup> = 0.579)
	Ov-GmnG	-0.1451x + 29.32; (R <sup>2</sup> = 0.379)	Ov-stdB	0.6667x - 8.1091; (R <sup>2</sup> = 0.332)		ExG2	24.24x + 0.5414; (R <sup>2</sup> = 0.463)	COMB1	1.8024x - 11.86; (R <sup>2</sup> = 0.574)
T-sugars	Ov-stdR	0.3531x + 6.2683; (R <sup>2</sup> = 0.478)	Ov-skG	-10.73x - 9.88; (R <sup>2</sup> = 0.491)	TSS	VEG	-3.61x + 10.12; (R <sup>2</sup> = 0.547)	Rn-mx	0.0184x + 8.63; (R <sup>2</sup> = 0.393)
	Ov-stdB	0.2345x + 5.89; (R <sup>2</sup> = 0.416)	Ov-medR	-0.3703x + 79.45; (R <sup>2</sup> = 0.246)		COMB1	-0.4595x + 12.84; (R <sup>2</sup> = 0.507)	Rn-mxPH	0.0185x + 8.6394; (R <sup>2</sup> = 0.393)
	Ov-GmnG	-0.1575x + 31.91; (R <sup>2</sup> = 0.401)	Ov-stdB	0.4693x + 0.9376; (R <sup>2</sup> = 0.239)		COMB2	-0.3189x + 12.86; (R <sup>2</sup> = 0.507)	Rn-mpvh	0.0181x + 8.3368; (R <sup>2</sup> = 0.298)
Vitamin C	L-GmnR	2.4462x - 255.55; (R <sup>2</sup> = 0.512)	Rn-mpvh	-6.1257x + 650.14; (R <sup>2</sup> = 0.625)					
	L-GmnGr	2.4646x - 257.40; (R <sup>2</sup> = 0.508)	Rn-mx	-5.2789x + 499.37; (R <sup>2</sup> = 0.592)					
	L-GmnB	2.4933x - 262.33; (R <sup>2</sup> = 0.507)	Rn-mxPH	-5.2914x + 496.49; (R <sup>2</sup> = 0.591)					
<b>Whole</b>									
<b>Stage = 817</b>					<b>Stage = 819</b>				
Acidity	WI	0.1894x - 0.1214; (R <sup>2</sup> = 0.423)	Ov-medG	0.0162x - 2.1898; (R <sup>2</sup> = 0.564)	Antioxidant	Ov-skG	21.74x + 74.51; (R <sup>2</sup> = 0.401)	Ov-skG	18.44x + 58.08; (R <sup>2</sup> = 0.355)
	RBDR	-0.1894x - 0.3108; (R <sup>2</sup> = 0.423)	Ov-GmnB	0.0092x - 0.3184; (R <sup>2</sup> = 0.538)		Wn-mxPH	-0.3435x + 39.66; (R <sup>2</sup> = 0.291)	Wn-avVD	-1.0144x + 27.94; (R <sup>2</sup> = 0.279)
	RGDR	1.1605x + 0.6284; (R <sup>2</sup> = 0.405)	Ov-mnB	0.0069x - 0.3073; (R <sup>2</sup> = 0.517)		Wn-mx	-0.2186x + 53.37; (R <sup>2</sup> = 0.289)	Wn-mxVD	-0.2122x + 27.25; (R <sup>2</sup> = 0.230)
Carbohydrates	Ov-skG	-4.4682x + 4.8171; (R <sup>2</sup> = 0.758)	Gl-cont	-0.2345x + 21.48; (R <sup>2</sup> = 0.292)	Carotenoids	Ov-GmnG	0.0498x - 5.0505; (R <sup>2</sup> = 0.638)	Ov-skG	1.1342x + 3.6336; (R <sup>2</sup> = 0.440)
	Ov-kuB	-0.0006x + 15.73; (R <sup>2</sup> = 0.214)	COMB1	2.5718x - 1.1543; (R <sup>2</sup> = 0.254)		Ov-stdR	-0.0584x + 2.0017; (R <sup>2</sup> = 0.606)	Wn-mxVD	-0.0139x + 1.7822; (R <sup>2</sup> = 0.322)
	Ov-stdR	0.1121x + 13.11; (R <sup>2</sup> = 0.167)	COMB2	1.8006x - 1.4471; (R <sup>2</sup> = 0.254)		Ov-stdG	-0.0362x + 2.1947; (R <sup>2</sup> = 0.495)	Wn-xpvh	-0.0083x + 1.8279; (R <sup>2</sup> = 0.192)
Crude fat	Ov-mnG	-0.2689x + 37.55; (R <sup>2</sup> = 0.337)	VEG	17.06x - 0.3989; (R <sup>2</sup> = 0.339)	Flavonoids	Wn-xpvh	-1.3864x + 258.44; (R <sup>2</sup> = 0.258)	Ov-kuB	0.0114x + 24.51; (R <sup>2</sup> = 0.282)
	Ov-GmnG	-0.1594x + 22.36; (R <sup>2</sup> = 0.322)	L-GmnG	-0.0806x + 15.78; (R <sup>2</sup> = 0.32)		Ov-kuR	0.602x + 36.06; (R <sup>2</sup> = 0.246)	VEG	272.13x - 12.52; (R <sup>2</sup> = 0.235)
	Gl-cont	-0.2158x + 7.0035; (R <sup>2</sup> = 0.308)	L-GmnGr	-0.0798x + 15.68; (R <sup>2</sup> = 0.307)		WI	220.86x - 217.40; (R <sup>2</sup> = 0.243)	Wn-med	-0.6225x + 92.67; (R <sup>2</sup> = 0.218)
MC	Gl-cont	0.1424x + 85.00; (R <sup>2</sup> = 0.456)	RGBVI	-28.13x + 91.99; (R <sup>2</sup> = 0.389)	pH	WI	-0.8142x + 5.7736; (R <sup>2</sup> = 0.443)	Ov-GmnB	-0.0693x + 8.0592; (R <sup>2</sup> = 0.591)
	Ov-kuB	0.0006x + 86.96; (R <sup>2</sup> = 0.418)	Gl-cont	0.0718x + 84.89; (R <sup>2</sup> = 0.285)		RBDR	0.8142x + 6.5879; (R <sup>2</sup> = 0.443)	Ov-medG	-0.0957x + 18.36; (R <sup>2</sup> = 0.377)
	Ov-medG	0.1892x + 60.53; (R <sup>2</sup> = 0.235)	Ov-GmnB	0.0992x + 81.30; (R <sup>2</sup> = 0.229)	Proteins	RGDR	-5.0703x + 2.5177; (R <sup>2</sup> = 0.438)	Ov-mnB	-0.0414x + 7.312; (R <sup>2</sup> = 0.358)
Phenolics	Wn-mx	-2.1458x + 609.55; (R <sup>2</sup> = 0.341)	Ov-skG	167.97x + 588.64; (R <sup>2</sup> = 0.374)		Gl-cont	-0.0391x + 2.7014; (R <sup>2</sup> = 0.421)	Ov-kuB	0.0001x + 1.1598; (R <sup>2</sup> = 0.701)
	Wn-xpvh	-1.8216x + 500.79; (R <sup>2</sup> = 0.318)	Ov-kuB	0.0333x + 161.26; (R <sup>2</sup> = 0.316)		Ov-stdR	0.0265x + 1.5324; (R <sup>2</sup> = 0.257)	Ov-kuR	0.0018x + 0.9572; (R <sup>2</sup> = 0.690)
	Ov-kuR	0.7722x + 211.25; (R <sup>2</sup> = 0.289)	Wn-mxVD	-2.1042x + 317.09; (R <sup>2</sup> = 0.287)		Ov-stdB	0.0262x + 1.2445; (R <sup>2</sup> = 0.257)	WI	0.5206x + 0.4654; (R <sup>2</sup> = 0.510)
R-sugars	Ov-skG	-7.5294x - 6.3176; (R <sup>2</sup> = 0.458)	Ov-stdR	0.6574x - 0.7915; (R <sup>2</sup> = 0.376)	Tannins	Wn-mx	-0.0151x + 4.2023; (R <sup>2</sup> = 0.318)	Ov-skG	1.3642x + 4.6378; (R <sup>2</sup> = 0.313)
	Ov-stdR	0.3561x + 4.5304; (R <sup>2</sup> = 0.357)	RBRI	-8.1523x + 34.40; (R <sup>2</sup> = 0.303)		Wn-mxPH	-0.0218x + 3.1684; (R <sup>2</sup> = 0.270)	Wn-avVD	-0.0756x + 2.4145; (R <sup>2</sup> = 0.249)
	Ov-medR	-0.335x + 69.41; (R <sup>2</sup> = 0.314)	Ov-medG	1.3929x - 186.01; (R <sup>2</sup> = 0.292)		Ov-mnG	0.0834x - 8.4526; (R <sup>2</sup> = 0.252)	Wn-mxVD	-0.0171x + 2.4307; (R <sup>2</sup> = 0.239)
T-sugars	Ov-skG	-7.6094x - 5.4483; (R <sup>2</sup> = 0.417)	VEG	49.18x + 5.7874; (R <sup>2</sup> = 0.407)	TSS	Gl-cont	-0.0642x + 10.46; (R <sup>2</sup> = 0.344)	Gl-cont	-0.0509x + 10.42; (R <sup>2</sup> = 0.217)
	Ov-stdR	0.3869x + 5.0093; (R <sup>2</sup> = 0.376)	COMB1	6.2159x - 30.77; (R <sup>2</sup> = 0.405)		Ov-mnG	-0.0765x + 19.10; (R <sup>2</sup> = 0.343)	Ov-skG	-0.8118x + 7.7602; (R <sup>2</sup> = 0.205)
	Ov-medR	-0.3685x + 76.27; (R <sup>2</sup> = 0.338)	COMB2	4.3502x - 31.46; (R <sup>2</sup> = 0.405)		Ov-GmnG	-0.0439x + 14.61; (R <sup>2</sup> = 0.308)	Ov-medR	-0.0417x + 16.82; (R <sup>2</sup> = 0.203)
Vitamin C	Rn-avVD	63.05x - 26.61; (R <sup>2</sup> = 0.602)	Wn-Gmn	-3.7138x + 535.69; (R <sup>2</sup> = 0.442)					
	Wn-mn	-2.7542x + 361.94; (R <sup>2</sup> = 0.425)	Wn-mn	-3.7273x + 543.16; (R <sup>2</sup> = 0.437)					
	Wn-Gmn	-2.7725x + 357.73; (R <sup>2</sup> = 0.423)	Wn-mxVD	-3.7988x + 416.03; (R <sup>2</sup> = 0.425)					

Short forms for the image-based independent parameters were used. They are combined to make the image-based independent properties, hence refer Table 1 to obtain the full form. Short forms: blue, cont - contrast, corr - correlation, G - green, Gl - gray level co-occurrence matrix, Gmn - geometric mean, Gr - gray, IDM - inverse difference moment, kurt - kurtosis, L - local binary pattern maximum, Ov - overall, PH - peak height, pvh - peak to valley height, rs - root mean square, R - red, Rn - roughness. sk - skewness, std - standard deviation, VD - valley depth, and Wn - waviness. COMB2, ExG, ExG2, GBVI, GCC, GRD, RBD, RBDR, RGD, RGDR, RGBVI, VDVI, VEG, and WI are available from Table 2. MC - moisture content; and TSS - total soluble solids.

and 819 combination. From the effort and complexity viewpoint, TSS is a relatively simple and quick measurement using a refractometer. Therefore, the recommendation of the “slice+819” combination, along with more sampling closer to the ultimate maturity stage, will effectively deliver the wet chemistry physicochemical properties through the simpler image-based computer vision methodology developed.

### 3.8. Wet Chemistry physicochemical properties Prediction Models Validation

A systematic validation strategy was employed using approximately 25 % percent of the reserved (test) data to assess the performance of developed models (based on 104 slices + 60 wholes = 164 images; Figure 6 and Section 3.7). The testing batch (unseen) of images of cashew apple samples (6 varieties; 2 sample types; 2 sample stages; 2–3 slice replications/variety; 2 whole replications/variety; 30 slices + 24 wholes; 54 total), were processed through the ImageJ plugin with the finalized inputs to generate the image-based parameters and the wet chemistry physicochemical properties were predicted using the topmost model (Table 7). Predicted wet chemistry parameters are compared with their actual values (Table 3), and the deviation expressed in terms of ratios (predicted/actual) are visualized as a paneled dodged bar graph across wet chemistry parameters, with maturity stages and cashew apple varieties (Figure 9).



**Figure 9.** Prediction and validation showing the ratio change of the predicted to actual physicochemical properties for cashew apple varieties from test image data. The gray dotted horizontal lines across the ratio of 1.0 indicate the ideal prediction.

Ideally, a ratio close to 1.0 indicates high predictive fidelity, while deviations highlight (<1.0) under- or (>1.0) over-estimations (horizontal dotted lines in Figure 9). Extreme values <0.75 or >1.25 depict poor prediction accuracy or model bias. For the 360 predictions (15 wet chemistry  $\times$  2 stages  $\times$  2 types  $\times$  6 varieties), the mean ratio for slice ( $1.081 \pm 0.553$ ) and whole ( $1.079 \pm 0.556$ ) were comparable, indicating similar prediction performance. Within the acceptable prediction ratio range of 0.75 to 1.25, a 66.7% for slice and 60.6% for whole samples demonstrate that a predominant number of predictions fall in this range, and slice performs better than whole samples. The under- and over-estimations by the models beyond this range were 14.4% and 18.9% for slice and 18.3% and 18.9% and 21.1% for whole samples, respectively.

Upon comparing the effects of maturity stages, the mean ratio of stage 819 (slice: 1.069; whole: 1.029) is closer to the ideal ratio of 1.0 than that of stage 817 (slice: 1.092; whole: 1.129) from 90 predictions each (Figure 9). This prediction result supports the idea that the ultimate maturity stage of 819 will be more favorable for the methodology, which was already anticipated from the analysis of the developed prediction models (Table 7).

Within the high prediction range of the ratio between 0.85 and 1.15, the wet chemistry physico-chemical properties of moisture content and TSS had a perfect count of 12 for both maturity stages and sample types. The subsequent five wet chemistry parameters are carbohydrates, pH, proteins, and total sugars, featuring in both stages and sample types (6–11 out of 12). This result suggests a strong model generalization for these traits. The least prediction performance was exhibited by phenolics, antioxidants, crude fat, flavonoids, tannins, and vitamin C (2–3 out of 12), both for slices and wholes. It is interesting to note that carotenoids can be predicted using slices (5/12) but cannot be predicted by whole samples (1/12) based on the analysis results.

### 3.8.1. Study's Limitations and Suggestions for Future Work

Several limitations were identified and will be encountered by users when reproducing this research in terms of experimentation and computer vision methodology development. Obtaining a suite of wet chemistry data for several varieties of cashew apple is expensive, time-intensive, and sometimes not affordable. The study utilized only six varieties at two maturity stages, resulting in the utilization of only six data points for each model developed (180 total). Therefore, this should be considered when applying the models. The study utilized only natural lighting during the image data acquisition, and this would have created some variation and affected the prediction accuracy. Developing new methodologies beyond those presented in the study requires a high level of computer vision and image analysis programming skills.

These limitations, in fact, present opportunities and guidance for future research in this knowledge domain. Expanding the dataset can assist in identifying and correcting outliers prior to training, thereby reducing the risk of skewed predictions. A larger dataset could also support more robust calibration across different varieties and development stages, ultimately enhancing the models' performance. Incorporating non-linear modeling approaches, along with increased data volume, may further enhance models' predictive performance. Color calibration of the input images will address and mitigate the variations induced by ambient lighting, potentially improving the computer vision analysis methodology. The current approach can be further optimized by integrating additional analytical techniques (e.g., textural features, color indices, other surface morphologies) as well as other relevant wet chemistry properties into the computer vision ImageJ plugin. The developed methodology can be applied to the evaluation of the wet chemistry physicochemical properties of other horticultural crops. As the results revealed that the optimal maturity stage was 819, it is feasible to develop an intermediate maturity stage, such as 818 (between stages 817 and 819), to expand the time window available for harvest, measurement, and analysis. Also, integration of the methodology with machine learning (ML) and artificial intelligence (AI) could enable universal, in-vitro tools for rapid fruit analysis and postharvest automation. These aspects will constitute a great potential for future research and can contribute to the enhancement of the overall performance of the computer vision methodology for indirect wet chemistry physicochemical properties measurements.

## 4. Conclusions

This study successfully developed a vision-based methodology for assessing 15 physicochemical properties of cashew apples, which are conventionally evaluated through wet chemistry protocols, across six diverse varieties and two maturity stages from the images of slices and whole samples. From the wet chemistry evaluation of physicochemical properties, significant changes were observed between maturity stages 817 and 819 across all varieties, which reflect variety-dependent ripening dynamics and support the use of biochemical markers for quality assessment. The alternative methodology of a developed user-coded and user-friendly computer vision ImageJ plugin, employing five image-based methodologies (color grid, surface morphology, GLCM, LBP, and color indices), analyzed the input images of slices and whole cashew apple samples, in batch processing mode, and generated 144 image-based parameters for further analysis (correlation, linear models, predictions). Optimal input configuration for the plugin was color grid:  $7 \times 7$ ; waveness and roughness cutoff radius: 35 pixels; GLCM sampling size: 1 pixel; and GLCM angle:  $270^\circ$ . The plugin execution was rapid, clocking a CPU time of about  $2.87 \pm 0.79$  s/image (Intel® Core™ i9 13950HX 5.5 GHz, 32 GB RAM).

The correlation between the image-based parameters (independent) with wet chemistry (dependent) traits, exhibited stronger correlations for slice-based image variables ( $r = -0.86$  to  $0.87$ ;  $\bar{r} = 0.268$ ) compared to that of whole fruit images ( $r = -0.72$  to  $0.73$ ;  $\bar{r} = 0.252$ ) indicating enhanced sensitivity to internal quality features. Strong inter-correlations among wet chemistry properties ( $r = -0.85$  to  $0.96$ ;  $\bar{r} = 0.443$ ) indicate that reliably predicted parameters can effectively serve as proxies for those with weaker image associations, thereby reinforcing the potential of image-based analysis and biochemical surrogates for non-destructive quality assessment. Surface morphology (52 % & 38 %), color indices (25.3 % & 14.7 %), and color grid (14 % & 32.7 %) features consistently ranked as top predictors across physicochemical properties, with slice and whole fruit images revealing distinct yet complementary patterns. The color indices developed in the study contributed significantly, featuring among the top-ranking 150 parameters (6.0 % & 2.0 %), and among the 22 color-based indices (40.9 % & 13.6 %). This ranking framework highlights the potential of easy-to-evaluate RGB descriptors for their non-destructive estimation.

The developed models for the slice at the maturity stage 819 ( $R^2 > 0.5$ ) covered 73.3 % (11 out of 15) wet chemistry physicochemical properties and successfully modeled acidity, antioxidant, carbohydrates, carotenoids, crude fat, flavonoids, pH, phenolics, proteins, tannins, and vitamin C. In contrast, the wet chemistry properties covered by whole samples were 33.3 % (5 out of 15; acidity, carbohydrates, carotenoids, proteins, and vitamin C). The combination “slice+819”, along with more sampling closer to the ultimate maturity stage, is recommended to effectively deliver the wet chemistry properties through the developed simpler image-based computer vision methodology.

From the prediction and validation of the developed model using the unseen image data, 66.7 % for slice and 60.6 % for whole samples belonged to the prediction ratio range of 0.75 to 1.25, demonstrating model validity. Models of the slice samples performed better than the whole samples. The wet chemistry properties predicted by the models in the decreasing order of performance were moisture content, TSS, carbohydrates, pH, proteins, and total sugars; and those with the least prediction performance were phenolics, antioxidants, crude fat, flavonoids, tannins, and vitamin C. Suggestions for future research include a larger dataset across different varieties and maturity stages; exploring non-linear modeling approaches; performing color calibration of the input images; incorporating additional wet chemistry and analytical techniques; developing an intermediate maturity stage; extending this methodology to other horticultural crops; and integrating it with modern ML and AI techniques.

**Author Contributions:** Conceptualization, M.J.G., C.I., and J.N.; methodology, M.J.G., C.I., and J.N.; formal analysis, M.J.G., C.I., J.N., A.J., H.T., S.S., A.M., J.D.A., and P.K.; investigation, M.J.G., C.I., J.N., A.J., H.T., S.S., A.M.; J.D.A., and P.K.; resources, M.J.G., J.N., J.D.A., P.K., and C.I.; data curation, M.J.G., and C.I.; writing-original draft preparation, M.J.G., J.N., and C.I.; writing-review and editing, M.J.G., J.N., C.I., A.J., H.T., S.S., A.M., P.K., J.D.A., and S.R.; visualization, M.J.G., C.I., and J.N.; supervision, M.J.G., J.N., and C.I.; project administration, J.N.,

J.D.A., M.J.G., and P.K.; funding acquisition, J.N., and J.D.A., All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Indian Council for Agricultural Research, Project ID.No: IXX20122, Project title: Enhancing the shelf life of cashew apple to increase the market potential and development of functional foods.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The image data used in the manuscript is available from Mendeley Data, V1, at <https://doi.org/10.17632/67kn43rj5c.1> (accessed on 18 October 2025). This dataset is cited in the manuscript as [39]. This dataset contains three files, namely, (1) Raw input images, (2) Wet chemistry data of cashew apples, and (3) R script used for analysis in the study.

## Appendix A

### Appendix A.1 Physicochemical properties

#### (i) Moisture, protein, and fat

Moisture content in cashew apples was determined by following the ISO 712:1998 method with a gravimetric principle using a hot air oven. The protein was estimated by following the Kjeldahl method as described by the AOAC method (920.151) [68], where a Kjeldahl apparatus (DX VATS, Pelican, Kelplusclassic, 2015) was used for digestion, distillation, and titration. For crude fat estimation Soxhlet apparatus (KSEA-1B, KEMI) was used, and samples were subjected to fat extraction using n-hexane. The extraction was carried out for approximately 3 h, and the percentage of fat was calculated as per the AOAC method (920.39).

#### (ii) Total soluble solids, pH, and titratable acidity determination

Total soluble solids (TSS) were analyzed based on the AOAC method (932.12) using a portable hand refractometer corrected for 20 °C. The AOAC method (981.12) was also used for determining the pH of the fruit. The AOAC method (942.15) was followed for assessing the titratable acidity (TA) of cashew apples, and results were expressed as percentage acidity in terms of citric acid.

#### (iii) Total carbohydrate

Total carbohydrate was estimated spectrophotometrically using the anthrone method [69], where the sample was hydrolyzed using 2.5N HCl for 3 h in a boiling water bath. After cooling to room temperature, it was neutralized with Na<sub>2</sub>CO<sub>3</sub> until the effervescence ceased. The solution was centrifuged after adding 10 mL of distilled water. The supernatant was collected, and the volume was made up to 100 mL with distilled water. Total carbohydrate was analyzed by adding 4 mL of anthrone reagent to a mixture of 0.1 mL aliquot and 0.9 mL distilled water. The mixture was incubated in a boiling water bath for 8 min and cooled to room temperature. The green to dark green color developed was measured using a spectrophotometer at 630 nm. With spectrophotometry methods, the sample values are measured indirectly with reference to a standard solution of known concentration. A standard curve was prepared by plotting the absorbance values of standard glucose solutions. The concentration of carbohydrate is calculated as:

$$\frac{\text{Absorbance of sample} - y\text{-intercept of standard curve}}{\text{Slope of standard curve}}$$

#### (iv) Total sugar

Total sugar content in cashew apple samples was determined using the Hedge and Hofreiter [70] spectrophotometric method. Sugars were extracted by mixing the samples with 80% hot ethanol and were refluxed for 1–2 h. The mixture was filtered, and the ethanol was removed using a rotary evaporator. The volume of the filtrate was made up to 100 mL. An aliquot (0.1 mL) was mixed with 0.9 mL distilled water and 5 mL anthrone reagent. The solution was then incubated at 100 °C for

11 min and the absorbance was read at 630 nm. D-glucose was used as a standard, and the results were expressed as a percentage on a fresh weight basis.

(v) Reducing sugar

Reducing sugar was estimated using the dinitro salicylic acid (DNSA) method [71]. Briefly reducing sugars were extracted from the samples as described for total sugars using 80% hot ethanol. A 3 mL DNSA reagent was added to 1 mL aliquot and incubated in boiling water bath for 16 min. The mixture was then cooled to room temperature, and the volume was made to 25 mL with distilled water. The absorbance of the sample was read at 575 nm using a spectrophotometer. The reducing sugar content was calculated from the calibration curve of standard D-glucose, and the results were expressed as a percentage on a wet basis.

(vi) Vitamin C (ascorbic acid) estimation

Vitamin C was determined using the 2,6-dichlorophenolindophenol (DCPIP) method (AOAC Method, 967.21). The samples were extracted using 3% metaphosphoric acid to obtain a solution of vitamin C. The extract was titrated with a standardized DCPIP solution until a pale pink color appeared. The results were expressed as mg/100 g of ascorbic acid on a wet basis.

*Appendix A.2 Biochemical Properties*

For the estimation of biochemical properties, polyphenolics were extracted as described by Ramful et al. [72] with some modifications. Approximately 1 g of pulp was extracted using 10 mL of 80% ethanol. The extract was sonicated for 30 min at 25 °C and then centrifuged at 10,000 Xg for 10 min at 4 °C and the extract was decanted. The residue was extracted twice following the aforementioned procedure. The combined extracts were adjusted to 30 mL and were filtered with a 0.4 µm nylon filter. The filtered extract was stored at –20 °C and used for the determination of total phenol, total flavonoids, antioxidant activity, and tannins.

(i) Total phenolic content (TPC)

TPC was estimated spectrophotometrically using the Folin Ciocalteu reagent (FCR) [73]. The aliquot (0.1 mL) was mixed with 2.9 mL of deionized water, 0.5 mL of Folin Ciocalteu reagent, and 2.0 mL of 20% Na<sub>2</sub>CO<sub>3</sub> solution. The mixture was allowed to stand for 20 min and absorption was measured at 750 nm against a reagent blank in a UV-Vis spectrophotometer. Results were expressed as gallic acid equivalent (mg GAE/100 g fresh weight [FW]).

(ii) Total flavonoids content (TFC)

TFC in cashew apple was measured spectrophotometrically using the method of Zhishen et al. [74]. A known volume (0.3 mL) of sample extract in acidified ethanol (2.1 mL) was taken, and mixed with 0.3 mL of 5% w/v NaNO<sub>2</sub>. After 5 min, 0.3 mL of 10% w/v AlCl<sub>3</sub> was added to it followed by addition of 2 mL of NaOH (1M) after 6 min. The solution was thoroughly mixed, resulting in a pink to yellow color. The absorbance was read at 510 nm against a reagent blank using 80% aqueous ethanol instead of the sample in a UV-Vis spectrophotometer. The results were expressed as quercetin equivalent (µg QE/g, wet basis).

(iii) Antioxidant activity using DPPH assay

DPPH (2,2-Diphenyl-1-picrylhydrazyl) scavenging assay was performed as described by Brand-Williams et al. [75]. A 3.9 mL aliquot of a 0.0634 mmol/L of DPPH solution in methanol (70%) was added to 0.1 mL of each extract and vigorously shaken. Change in the absorbance of the sample extract was measured at 517 nm for 30 min. The percentage inhibition of DPPH by the test sample and known solutions was calculated by the following formula, and results were expressed in terms of Trolox equivalent per gram fresh weight (µmol TE/g FW).

$$\text{Inhibition, \%} = \left( \frac{A_0 - A}{A_0} \right) \times 100$$

where,  $A_0$  was the initial absorbance of DPPH solution at 517 nm and  $A$  was the final absorbance of the sample extract added with DPPH solution. Methanol (70 %) was used as a blank.

#### (iv) Tannin content

Tannin content was measured using Folin's Denis method [76]. An ethanolic extract (1 mL) was added to 5 mL of Folin-Denis reagent and 10 mL of saturated sodium carbonate solution. The absorbance of the mixture was measured at 700 nm after 30 min incubation in the dark. Tannic acid was used as the standard, and distilled water as the blank. The results were expressed as mg TaE (tannic acid equivalent)/100 g on a wet basis.

#### (v) Carotene estimation

Carotene in cashew apples was estimated by following the method described by Rodriguez [77]. Exactly 5 g of the sample was repeatedly extracted using acetone in a pestle and mortar until the residue became colorless. The acetone extract was collected in a separating funnel containing 20 mL of petroleum ether and mixed gently. Next, 20 mL of a 5 % sodium sulfate solution was added, and the mixture was gently shaken. The top petroleum ether layer was separated, and the lower aqueous phase was extracted again with another 20 mL of petroleum ether. Both petroleum ether extracts were combined and rinsed with a small amount of distilled water. Then, 10 g of anhydrous sodium sulfate was added to the petroleum ether extract, which was allowed to stand for 30 min. The petroleum ether extract was decanted into a 100 mL volumetric flask, and the volume was made up with petroleum ether. The absorbance in the spectrophotometer was measured at 452 nm using petroleum ether as a blank. The contents of carotene and lycopene were expressed as mg per 100 g wet basis.

## References

1. Maciel, M.L.; Hansen, T.J.; Aldinger, S.B.; Labows, J.N. Flavor chemistry of cashew apple juice. *Journal of Agricultural and Food Chemistry* **1986**, *34*, 923–927.
2. Das, I.; Arora, A. Post-harvest processing technology for cashew apple—A review. *Journal of Food Engineering* **2017**, *194*, 87–98.
3. Pushpalatha, P.; Sobhana, A.; Mini, C. Processing and product diversification in cashew apple. *Advances in Cashew Production Technology* **2015**, pp. 109–116.
4. Sri Lakshmi, C.; Prema, A. Resource-use efficiency in raw cashew nut production in Kerala, India. *Asian Journal of Agricultural Extension, Economics and Sociology* **2025**, *43*, 92–98.
5. FAOSTAT crops and livestock database, 2025.
6. Ganesh, S.; Kannan, M.; Jawaharlal, M. Cashew industry-an outlook. In Proceedings of the I International Symposium on Cashew Nut 1080, 2011, pp. 89–95.
7. Anoopkumar, A.; Gopinath, C.; Annadurai, S.; Abdullah, S.; Tarafdar, A.; Hazeena, S.H.; Rajasekharan, R.; Kuriakose, L.L.; Aneesh, E.M.; de Souza Vandenberghe, L.P.; et al. Biotechnological valorisation of cashew apple: Prospects and challenges in synthesising wide spectrum of products with market value. *Bioresource Technology Reports* **2024**, *25*, 101742.
8. Gnagne, A.A.G.B.; Soro, D.; Ouattara, Y.; Kouli, E.; Koffi, E. A literature review of cashew apple processing. *African Journal of Food, Agriculture, Nutrition and Development* **2023**, *23*, 22452–22469.
9. Trevisan, M.; Pfundstein, B.; Haubner, R.; Würtele, G.; Spiegelhalder, B.; Bartsch, H.; Owen, R. Characterization of alkyl phenols in cashew (*Anacardium occidentale*) products and assay of their antioxidant capacity. *Food and Chemical toxicology* **2006**, *44*, 188–197.
10. Honorato, T.L.; Rabelo, M.C.; Gonçalves, L.R.B.; Pinto, G.A.S.; Rodrigues, S. Fermentation of cashew apple juice to produce high added value products. *World Journal of Microbiology and Biotechnology* **2007**, *23*, 1409–1415.
11. dos Santos Lima, F.C.; da Silva, F.L.H.; Gomes, J.P.; da Silva Neto, J.M. Chemical composition of the cashew apple bagasse and potential use for ethanol production. *Advances in Chemical Engineering and Science* **2012**, *2*, 519–523.

12. Dedehou, E.; Dossou, J.; Anihouvi, V.; Soumanou, M.M. A review of cashew (*Anacardium occidentale* L.) apple: Effects of processing techniques, properties and quality of juice. *African Journal of Biotechnology* **2016**, *15*, 2637–2648.
13. Pascal, A.D.C.; Virginie, G.; Diane, B.F.T.; Estelle, K.R.; Félicien, A.; Valentin, W.D.; Dominique, S.K.C. Nutritional profile and chemical composition of juices of two cashew apple's varieties of Benin. *Chemistry Journal* **2018**, *4*, 91–96.
14. Reina, L.J.C.; Durán-Aranguren, D.D.; Forero-Rojas, L.F.; Tarapuez-Viveros, L.F.; Durán-Sequeda, D.; Carazzone, C.; Sierra, R. Chemical composition and bioactive compounds of cashew (*Anacardium occidentale*) apple juice and bagasse from Colombian varieties. *Heliyon* **2022**, *8*.
15. Akyereko, Y.G.; Yeboah, G.B.; Wireko-Manu, F.D.; Alemawor, F.; Mills-Robertson, F.; Odoom, W. Nutritional value and health benefits of cashew apple. *JFA reports* **2023**, *3*, 110–118.
16. Prasertsri, P.; Roengrit, T.; Kanpetta, Y.; Tong-Un, T.; Muchimapura, S.; Wattanathorn, J.; Leelayuwat, N. Cashew apple juice supplementation enhanced fat utilization during high-intensity exercise in trained and untrained men. *Journal of the International Society of Sports Nutrition* **2013**, *10*, 13.
17. Aidoo, R.; Kwofie, E.M.; Ngadi, M.O. Circularity of cashew apples: examining the product-process pathways, techno-functional, nutritional/phytomolecular qualities for food applications. *ACS Food Science & Technology* **2022**, *2*, 1051–1066.
18. da Silveira Vasconcelos, M.; Gomes-Rochette, N.F.; de Oliveira, M.L.M.; Nunes-Pinheiro, D.C.S.; Tome, A.R.; Maia de Sousa, F.Y.; Pinheiro, F.G.M.; Moura, C.F.H.; Miranda, M.R.A.; Mota, E.F.; et al. Anti-inflammatory and wound healing potential of cashew apple juice (*Anacardium occidentale* L.) in mice. *Experimental Biology and Medicine* **2015**, *240*, 1648–1655.
19. Salehi, B.; Gültekin-Özgüven, M.; Kirkin, C.; Özçelik, B.; Morais-Braga, M.F.B.; Carneiro, J.N.P.; Bezerra, C.F.; Silva, T.G.d.; Coutinho, H.D.M.; Amina, B.; et al. Antioxidant, antimicrobial, and anticancer effects of anacardium plants: an ethnopharmacological perspective. *Frontiers in Endocrinology* **2020**, *11*, 295.
20. Attri, B. Effect of initial sugar concentration on the physico-chemical characteristics and sensory qualities of cashew apple wine. *Natural product radiance* **2009**, *8*, 374–379.
21. Kouassi, E.K.A.; Soro, Y.; Vaca-Garcia, C.; Yao, K.B. Chemical composition and specific lipids profile of the cashew apple bagasse. *Rasayan Journal of Chemistry* **2018**, *11*, 386–391.
22. Okpanachi, U.; Attah, S.; Shaahu, D. A comparative study between vitamins and amino acid profile of sun-dried red and yellow cashew pulp. *Int. J. An. Biol* **2015**, *1*, 23.
23. Runjala, S.; Kella, L. Cashew apple (*Anacardium occidentale* L.) therapeutic benefits, processing and product development: An overview. *The Pharma Innovation* **2017**, *6*, 260.
24. Sivaprakasam, S.; Chinnaiyan, U.; Varatharajan, S.; Uthayasuryan, E.; Chandrasekaran, M.; Paramasivam, S.; Ramasamy, P. Unveiling the physicochemical attributes and antioxidant potential of cashew apple juice varieties. *Applied Fruit Science* **2025**, *67*, 74.
25. Kiatti, D.d.; Serrapica, F.; Musco, N.; Di Palo, R.; Calabrò, S. Potential use of tropical and subtropical fruits by-products in pig diet: In vitro two-step evaluation. *Animals* **2025**, *15*, 1454.
26. Preethi, P.; Rajkumar, A.; Shamsudheen, M.; Nayak, M. Prospects of cashew apple—A compilation report. *Technical Bulletin* **2019**, *2*, 1–28.
27. Codjia, F.S.; Dabadé, D.S.; Agbobatinkpo, P.; Collombel, I.; Achir, N.; Azokpota, P.; Dossou, J. Fermented cashew apple beverages: Current state of knowledge and prospects. *Beverages* **2025**, *11*, 49.
28. Mtashobya, L.A.; Mgeni, S.T.; Emmanuel, J.K. Potential contributions of cashew apple juice to nutrition, medicine and bioethanol generation: An outlook. *Natural Product Communications* **2025**, *20*, 1934578X251357414.
29. Levy, H.M.; Rodrigues, T.H.S.; de Albuquerque, T.L.; Sant'Ana, H.B.; Feitosa, F.X.; Rocha, M.V.P. Biosynthesis of biolubricant catalyzed by lipase B from *Candida antarctica* immobilized on nanocomposite of cashew apple bagasse lignin. *Renewable Energy* **2025**, *243*, 122452.
30. Osei, E.D.; Amotoe-Bondzie, A.; Ataa Pokuah, A.; Laar, W.S.; Afoakwah, N.A.; Ivanišová, E. Cashew apple pomace: Chemical composition and applications in functional food product development—A review. *Food Science & Nutrition* **2025**, *13*, e70185.
31. Cormier, R.; Michodjehoun, S.; Fulcrand, B. Clarification of cashew apple juice and commercial applications. *Oxfarm Quebec, Benin, West Africa* **2008**, pp. 1–9.
32. Ogunjobi, M.; Ogunwolu, S. Physicochemical and sensory properties of cassava flour biscuits supplemented with cashew apple powder. *Journal of Food Technology* **2010**, *8*, 24–29.

33. Mohammadzadeh, Z.; Shojaeiyan, A.; Mahfeli, M.; Ayyari, M.; Tohidfar, M.; Mokhtassi-Bidgoli, A.; Atighi, M.R. Predictive modeling of CIELAB color parameters in okra accessions based on phytochemical composition and antioxidant activity: A non-destructive ImageJ and RSM approach. *LWT* **2025**, *228*, 118080. <https://doi.org/https://doi.org/10.1016/j.lwt.2025.118080>.
34. Jinorose, M.; Prachayawarakorn, S.; Soponronnarit, S. A novel image-analysis based approach to evaluate some physicochemical and cooking properties of rice kernels. *Journal of Food Engineering* **2014**, *124*, 184–190. <https://doi.org/https://doi.org/10.1016/j.jfoodeng.2013.08.009>.
35. Sabzi, S.; Nadimi, M.; Abbaspour-Gilandeh, Y.; Paliwal, J. Non-destructive estimation of physicochemical properties and detection of ripeness level of apples using machine vision. *International Journal of Fruit Science* **2022**, *22*, 628–645.
36. Cárdenas-Pérez, S.; Chanona-Pérez, J.; Méndez-Méndez, J.V.; Calderón-Domínguez, G.; López-Santiago, R.; Perea-Flores, M.J.; Arzate-Vázquez, I. Evaluation of the ripening stages of apple (Golden Delicious) by means of computer vision system. *Biosystems Engineering* **2017**, *159*, 46–58.
37. Abalone, R.; Cassinera, A.; Gaston, A.; Lara, M. Some physical properties of amaranth seeds. *Biosystems Engineering* **2004**, *89*, 109–117.
38. Adiga, J.D.; Muralidhara, B.M.; Preethi, P.; Savadi, S. Phenological growth stages of the cashew tree (*Anacardium occidentale* L.) according to the extended BBCH scale. *Annals of Applied Biology* **2019**, *175*, 246–252.
39. Gupta, M.J.; Nishad, J.; Igathinathane, C. Assessment of Physicochemical Properties of Cashew Apple through Computer Vision — Original Images, Data, and R Script. Mendeley Data, V1, <https://doi.org/10.17632/67kn43rj5c.1>, 2025.
40. Pawlus, P.; Reizer, R.; Wiczorowski, M. Functional importance of surface texture parameters. *Materials* **2021**, *14*, 5326.
41. Chinga, G. Waveness Roughness. ImageJ Plugins <https://imagej.net/ij/plugins/waveness-roughness.html>, 2018. Accessed: 2025-09-09.
42. Haralick, R.M.; Shanmugam, K. Computer classification of reservoir sandstones. *IEEE Transactions on Geoscience Electronics* **1973**, *11*, 171–177.
43. Mohanaiah, P.; Sathyanarayana, P.; GuruKumar, L. Image texture feature extraction using GLCM approach. *International journal of scientific and research publications* **2013**, *3*, 1–5.
44. Song, Y.; Sa, J.; Luo, Y.; Zhang, Z. A comprehensively improved local binary pattern framework for texture classification. *Multimedia Tools and Applications* **2025**, *84*, 22043–22068.
45. Woebbecke, D.M.; Meyer, G.E.; Von Bargaen, K.; Mortensen, D.A. Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* **1995**, *38*, 259–269.
46. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 79–87.
47. Yuan, H.; Zhang, J.; Zhang, H.; Xu, W.; Peng, J.; Wang, X.; Chen, P.; Li, P.; Lu, F.; Yan, J.; et al. Monitoring Autumn Phenology in Understory Plants with a Fine-Resolution Camera. *Remote Sensing* **2025**, *17*, 1025.
48. Quille-Mamani, J.; Porrás-Jorge, R.; Saravia-Navarro, D.; Herrera, J.; Chavez-Galarza, J.; Arbizu, C.I. Prediction of biometric variables through multispectral images obtained from UAV in beans (*Phaseolus vulgaris* L.) during ripening stage **2021**.
49. Beniaich, A.; Silva, M.L.N.; Avalos, F.A.P.; de Menezes, M.D.; Cândido, B.M. Determination of vegetation cover index under different soil management systems of cover plants by using an unmanned aerial vehicle with an onboard digital photographic camera. *Semina: Ciências Agrárias* **2019**, *40*, 49–66.
50. Ju, Z.; Liu, C.; Yuan, Y.; Wang, Y.; Liu, G. Coloration potential, anthocyanin accumulation, and enzyme activity in fruit of commercial apple cultivars and their F1 progeny. *Scientia Horticulturae* **1999**, *79*, 39–50.
51. Guijarro, M.; Pajares, G.; Riomoros, I.; Herrera, P.; Burgos-Artizzu, X.; Ribeiro, A. Automatic segmentation of relevant textures in agricultural images. *Computers and Electronics in Agriculture* **2011**, *75*, 75–83.
52. Meyer, G.E.; Neto, J.C. Verification of color vegetation indices for automated crop imaging applications. *Computers and electronics in agriculture* **2008**, *63*, 282–293.
53. Strati, V.; Albéri, M.; Barbagli, A.; Boncompagni, S.; Casoli, L.; Chiarelli, E.; Colla, R.; Colonna, T.; Elek, N.I.; Galli, G.; et al. Advancing Grapevine Disease Detection Through Airborne Imaging: A Pilot Study in Emilia-Romagna (Italy). *Remote Sensing* **2025**, *17*, 2465.

54. Jiménez-Muñoz, J.C.; Sobrino, J.A.; Plaza, A.; Guanter, L.; Moreno, J.; Martínez, P. Comparison between fractional vegetation cover retrievals from vegetation indices and spectral mixture analysis: Case study of PROBA/CHRIS data over an agricultural area. *Sensors* **2009**, *9*, 768–793.
55. Banerjee, K.; Dutta, S.; Das, B.; Roy, D.; Sen, S.; Mandal, B.P.; Chatterjee, A. Crop type discrimination through low cost proximal RGB imaging and multivariate analysis. *Arabian Journal of Geosciences* **2025**, *18*, 31.
56. Woebbecke, D.M.; Meyer, G.E.; Von Bargen, K.; Mortensen, D.A. Plant species identification, size, and enumeration using machine vision techniques on near-binary images. In Proceedings of the Optics in Agriculture and Forestry. SPIE, 1993, Vol. 1836, pp. 208–219.
57. Hussain, S.; Teshome, F.T.; Tulu, B.B.; Awoke, G.W.; Hailegnaw, N.S.; Bayabil, H.K. Leaf area index (LAI) prediction using machine learning and UAV based vegetation indices. *European Journal of Agronomy* **2025**, *168*, 127557.
58. Ritchie, G.; Sullivan, D.; Vencill, W.; Bednarz, C.; Hook, J. Sensitivities of normalized difference vegetation index and a green/red ratio index to cotton ground cover fraction. *Crop science* **2010**, *50*, 1000–1010.
59. Aynalem, H.M.; Righetti, T.L.; Reed, B.M. Non-destructive evaluation of in vitro-stored plants: a comparison of visual and image analysis. *In Vitro Cellular & Developmental Biology-Plant* **2006**, *42*, 562–567.
60. Gamon, J.A.; Surfus, J.S. Assessing leaf pigment content and activity with a reflectometer. *The New Phytologist* **1999**, *143*, 105–117.
61. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87.
62. Xiaoqin, W.; Miaomiao, W.; Shaoqiang, W.; Yundong, W. Extraction of vegetation information from visible unmanned aerial vehicle images. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*.
63. Song, H.; Asghari, M.; Zahedipour-Sheshglani, P.; Aljanabi, S.M.A.K.; Diao, E.; Xiang, X.; Qian, S.; Liang, X. Impact of exogenous cinnamic acid on some quality indices, phenolic compounds, main phenolic biosynthesis enzymes and senescence rate of strawberry fruit. *Food Chemistry* **2025**, *487*, 144726.
64. Hasanzadeh, B.; Abbaspour-Gilandeh, Y.; Soltani-Nazarloo, A.; Hernández-Hernández, M.; Gallardo-Bernal, I.; Hernández-Hernández, J.L. Non-destructive detection of fruit quality parameters using hyperspectral imaging, multiple regression analysis and artificial intelligence. *Horticulturae* **2022**, *8*, 598.
65. Paulo, D.J.; Neves, C.M.; Wessel, D.F.; Neves, J.C. WildFruIP: Estimating Fruit Physicochemical Parameters from Images Captured in the Wild. In Proceedings of the Iberoamerican Congress on Pattern Recognition, 2023, pp. 314–326.
66. Andrés-Bello, A.; Barreto-Palacios, V.; García-Segovia, P.; Mir-Bel, J.; Martínez-Monzó, J. Effect of pH on color and texture of food products. *Food Engineering Reviews* **2013**, *5*, 158–170.
67. Akkaya, E. Characterization of Physicochemical, Colour and Textural Properties of Turkish Type Cheeses. *Dicle Üniversitesi Veteriner Fakültesi Dergisi* **2024**, *17*, 137–142.
68. of Official Analytical Chemists, T.A. *Official methods of analysis of the Association of Official Analytical Chemists*; Vol. 11, The Association, 2000.
69. Sadasivam, S.; Manikam, A. *Biochemical Methods*. Third Print. *New Age International, New Delhi* **2010**.
70. Hedge, J.; Hofreiter, B. *Methods in Carbohydrate Chemistry*, Vol. 17. *Whistler, RL and BeMiller, JN Ed* **1962**, *17*, 420.
71. Miller, G.L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical chemistry* **1959**, *31*, 426–428.
72. Ramful, D.; Tarnus, E.; Aruoma, O.I.; Bourdon, E.; Bahorun, T. Polyphenol composition, vitamin C content and antioxidant capacity of Mauritian citrus fruit pulps. *Food research international* **2011**, *44*, 2088–2099.
73. Singleton, V.L.; Orthofer, R.; Lamuela-Raventós, R.M. [14] Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent. In *Methods in enzymology*; Elsevier, 1999; Vol. 299, pp. 152–178.
74. Zhishen, J.; Mengcheng, T.; Jianming, W. The determination of flavonoid contents in mulberry and their scavenging effects on superoxide radicals. *Food chemistry* **1999**, *64*, 555–559.
75. Brand-Williams, W.; Cuvelier, M.E.; Berset, C. Use of a free radical method to evaluate antioxidant activity. *LWT-Food science and Technology* **1995**, *28*, 25–30.
76. Smit, C.; Joslyn, M.; Lukton, A. Determination of tannins and related polyphenols in foods. *Analytical Chemistry* **1955**, *27*, 1159–1162.
77. Rodriguez-Amaya, D.B.; Raymundo, L.; Lee, T.; Simpson, K.; Chichester, C. Carotenoid pigment changes in ripening Momordica charantia fruits. *Annals of Botany* **1976**, *40*, 615–624.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.