Smart Agricultural Technology 12 (2025) 101091

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Integrating colorimetry and machine learning: an approach for optimizing fruit selection in *Licania tomentosa* seedling production

Douglas Martins Santana [a,*], Júlio César Altizani-Júnior [a], Francisco Guilhien Gomes-Junior [a], Durval Dourado-Neto [a], Renan Caldas Umburanas [a], Klaus Reichardt [a,b], Fábio Oliveira Diniz [c]

[a] *Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil*
[b] *Center for Nuclear Energy in Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil*
[c] *Federal Institute of Ceara, Boa Viagem, Ceara, Brazil*

ARTICLE INFO

Keywords:
RGB system
Epicarp coloration
Morphometric descriptors
Linear discriminant analysis

ABSTRACT

Licania tomentosa is a widely distributed species in Brazil, commonly used in urban landscaping and environmental restoration. Despite its potential, understanding the relationship between fruit maturation and seedling quality remains limited. This study aimed to evaluate the relationship between maturation stages - classified by epicarp coloration - and seedling performance through RGB colorimetric analysis, fruit morphometry, and the application of machine learning algorithms. Fruits were collected from mother trees and classified into four color stages based on the Munsell color chart. Digital images were analyzed to extract RGB values and morphometric parameters of the fruits using ImageJ® software. Subsequently, seedling emergence, biometric attributes, biomass accumulation, and the Dickson Quality Index (DQI) were evaluated. Yellow-Red fruits produced seedlings with higher emergence rates, greater shoot and root biomass accumulation, and higher DQI values, indicating greater seedling vigor. In contrast, Greenish Green-Yellow fruits resulted in less vigorous seedlings. The Red band was the main indicator of changes in the fruits. Morphometric parameters alone were insufficient to discriminate the maturation stages. Linear Discriminant Analysis correctly classified 90.48 % of the fruits according to their maturation stage. The integration of colorimetric data with machine learning proved to be an effective, non-destructive, and low-cost approach for optimizing seed selection. To enhance the predictive accuracy of the model it is recommended to expand the dataset under natural conditions and explore alternative color systems and complementary fruit traits.

1. Introduction

The family Chrysobalanaceae comprises 20 genera and more than 500 species of shrubs and trees, distributed across tropical and subtropical regions of the world [1,2]. Among these genera, *Licania* stands out as the most prominent, occurring from the United States to South America [3]. The species *Licania tomentosa* (Benth) Frisch, popularly known as oiti, is widely found in Brazil, with greater incidence in the Northeast Region [4]. The height of *L. tomentosa* ranges between 6 and 15 m, with a trunk diameter between 30 to 50 cm. Flowering occurs between the months of June and August, while fruiting is concentrated between January and March. Its dense crown, with high shading capacity, is one of the main factors justifying its widespread use in landscaping projects [5].

The species occupies a prominent position in the urban landscaping of squares and parks [6–8]; in the restoration of degraded areas [9]; and in the pharmaceutical industry [1]. It also has food value, with emphasis on the concentration of bioactive compounds such as polyphenols and flavonoids [4].

Despite its various applications, knowledge about the agronomic aspects of *L. tomentosa* is still limited. Studies on fruit and seed morphometry [10], seedling development [11] and management in different productive environments [12] are still incipient. One of the persistent challenges in seedling production is the absence of standardized criteria for fruit selection, a factor that directly affects vigor and uniformity. Since the species is multiplied by seeds contained in

* Corresponding author.
E-mail addresses: douglas.martinssantana@usp.br (D.M. Santana), altizani@usp.br (J.C. Altizani-Júnior), francisco1@usp.br (F.G. Gomes-Junior), ddourado@usp.br (D. Dourado-Neto), rumbu@usp.br (R.C. Umburanas), klaus@cena.usp.br (K. Reichardt), fabio.diniz@ifce.edu.br (F.O. Diniz).

https://doi.org/10.1016/j.atech.2025.101091
Received 7 April 2025; Received in revised form 2 June 2025; Accepted 8 June 2025
Available online 10 June 2025

drupe-type fruits, which are used for sowing without the need for depulping [13], the correct identification of the fruit maturation stage becomes essential to better understand this specie and to ensure the quality of the seeds.

In practice, the absence of standardization leads implies for seedling production. Samples composed of seeds from fruits at different stages of maturation result in seedlings with distinct physiological performance, which compromises uniformity and increases operational costs due to the need for replanting and longer cultivation time. Moreover, less vigorous seedlings exhibit lower survival and adaptation rates in the field, negatively affecting the outcomes of urban landscaping and environmental restoration programs [14].

The drupes of *L. tomentosa* have an elliptical shape, ranging from 12 to 16 cm in length. During maturation, the epicarp, which is smooth and thin, displays a coloration that varies from green (immature) to yellow/orange (mature). The mesocarp is fleshy, with a bittersweet flavor and yellow color, while the endocarp is membranous and light-colored. The seed, in turn, presents a high lipid content [10,15]. The fruit maturation process involves a series of morphological, physiological, and biochemical transformations [16]. Among the most evident visual indicators, the variation in epicarp coloration stands out, directly associated with chlorophyll degradation (green) and the synthesis of carotenoids (yellow) and anthocyanins (red) [17]. Such colorimetric changes can be used as relevant visual indicators to infer the fruit maturation stage [18]. However, human perception of color is subject to the interaction between brightness and intensity, which can result in subjective interpretations of the shades of a single color, compromising standardization and reproducibility of the process [19,20].

Computer vision systems, such as color analysis methods, are relevant, objective, and standardized alternatives [21], capable of overcoming the limitations inherent to human visual analysis and ensuring greater precision in fruit selection. Among the available techniques, the RGB (Red, Green, Blue) color system stands out as a widely used tool for fruit color analysis [22,19,23]. The RGB color channels operate at wavelengths of 700, 546, and 435 nm, respectively [21]. RGB usage requires simple algorithms, making it more cost-effective and employed for different types of chromatic analyses [24]. Its effectiveness has already been proven in detecting the maturation stage of strawberries [25], naval oranges [26], tomato [27], and Swingle citrumelo [19].

Additionally, recent advances have incorporated machine learning algorithms to classify maturation stages from digital images, enabling scalable, low-cost, and non-destructive approaches [20,28]. Linear Discriminant Analysis (LDA) is an algorithm used in situations where the analyzed objects present very similar visual characteristics, making it difficult for human vision to distinguish them. In such cases, LDA proves capable of identifying these similarities and accurately separating the objects [29].

LDA aims to find a linear combination of attributes that characterize the objects, separating them into classes. For this purpose, dispersion matrix analysis is used, which allows dimensionality reduction of the data without compromising the maximum separation between the classes [30]. LDA has demonstrated high performance in classifying six types of apples, achieving 98 % accuracy [31]. In the study conducted by Zulkifli et al., [32], LDA was the most efficient classifier among four evaluated methods, achieving 83.5 % accuracy in classifying different maturation stages of papaya.

Despite the relevant advances in these methodologies, to the authors' knowledge, there are still no studies that relate fruit coloration to the seedling performance of *L. tomentosa*. The objective of this study was to evaluate the relationship between the fruit maturation stage of *L. tomentosa* and seedling quality. To this end, we used data from colorimetric analysis, morphological characteristics of the fruits, and seedling performance. Based on this dataset, we applied machine learning algorithms as a strategy for fruit selection, with the aim of ensuring greater precision and standardization in the production process of high-quality seedlings of the species.

2. Material and Methods

2.1. Fruit collection and material preparation

The study was conducted with intact and healthy fruits from ten matrix plants located at the Federal Institute of Piauí – Uruçuí Campus, Brazil (7° 16′ 42" S 44° 30′ 22" W, altitude of 359 m above the sea level). Manual collection occurred after spontaneous fruit drop. The region's climate is Aw-Köppen, tropical with abundant rainfall in the summer, with an average annual precipitation of 1069 mm and an average annual temperature of 27.2 °C. The fruits were classified according to epicarp coloration using the Munsell color chart for plant tissues (Gretag-Macbeth, New Winsor, NY, USA), which considers the parameters of hue, value, and chroma (Table 1).

2.2. Morpho-colorimetric characterization

For colorimetric analysis, photographs of 36 fruits of each coloration were captured with a digital camera (Nikon Coolpix S6200, 16 Mpx®), equipped with a charge-coupled device (CCD). Artificial lighting was provided by four fluorescent lamps (40W; 5250 K color temperature). The images, in Tagged Image File Format (TIFF) and 350 dpi resolution, were analyzed using ImageJ® software, version 1.54m [33]. The software allowed the extraction of pixel intensity values from the RGB color components (bands), and the grayscale. In addition to the average pixel intensity, which ranges from 0 to 255, was calculated using the ratio (R + G+ B)/3.

The original image of each fruit was subdivided into three 8-bit grayscale images, containing the R, G, and B components of the image pixels (bands). Furthermore, the software was also used to obtain morphological descriptors of the fruits, corresponding to perimeter, surface area, aspect ratio, roundness, and circularity (Fig. 1).

The perimeter represented the length of the two-dimensional contour of the fruit (cm). The surface area corresponded to the amount of two-dimensional space of the fruit, expressed in cm^2 after pixel conversion. The aspect ratio represented the ratio between the major and minor diameters of the ellipse that circumscribes the fruit, indicating the degree of elongation. Roundness reflected the inverse of the aspect ratio, calculated by the Eq. (1):

$$R = 4 \times \frac{\text{area}}{\pi \times \text{largest diameter}^2}. \quad (1)$$

Lower values indicate more circular fruits. Circularity corresponded to the degree of approximation of the fruit to a circular shape, calculated by Eq. (2):

$$C = 4\pi \times \frac{\text{area}}{\text{perimeter}^2} \quad (2)$$

Table 1
Epicarp color characterization of oiti (*L. tomentosa*) fruits through visual classification by Munsell color chart.

Epicarp color	Hue	Value and Chroma
Greenish Green-Yellow	7.5 GY	3/4* = low luminosity / moderate saturation
Yellow	5.0 Ye	6/8 = moderate luminosity / high saturation
Yellow Yellow-Red	7.5 yYR	6/10 = moderate luminosity / high saturation
Yellow-Red	5.0 YR	4/8 = low luminosity / high saturation

* The first number represents the value, which indicates the lightness of the color on a scale from 0 (black) to 10 (white); the second number represents the chroma, which indicates the intensity or saturation of the color, where higher values indicate purer and more intense colors.

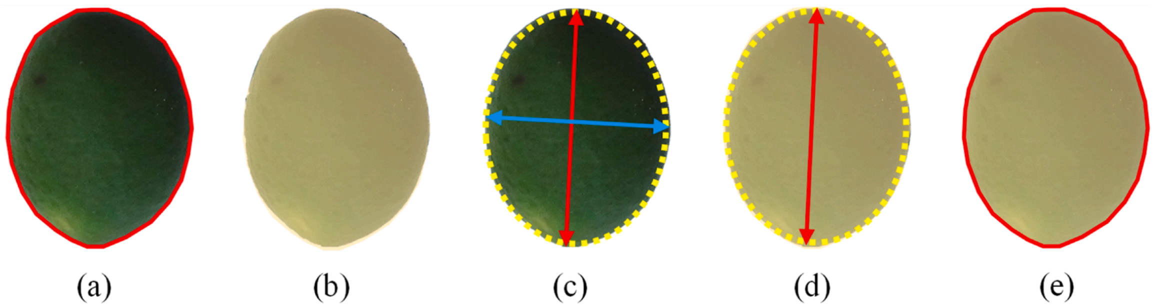

Fig. 1. Extracting morphological descriptors of oiti (*Licania tomentosa* Benth) fruits through image analysis: perimeter (a), surface area (b), aspect ratio (c), roundness (d), and circularity (e).

Values ranged from 0 (elongated shape) to 1 (perfect circle). Complementarily, the length (cm) and width (cm) of the fruits were also determined using ImageJ®.

2.3. Seedling production

After classification, 100 fruits of each epicarp coloration were sown at a depth of 3 cm in plastic boxes containing medium-textured sand (5 dm^3), arranged in a greenhouse (35 % shading). The substrate water

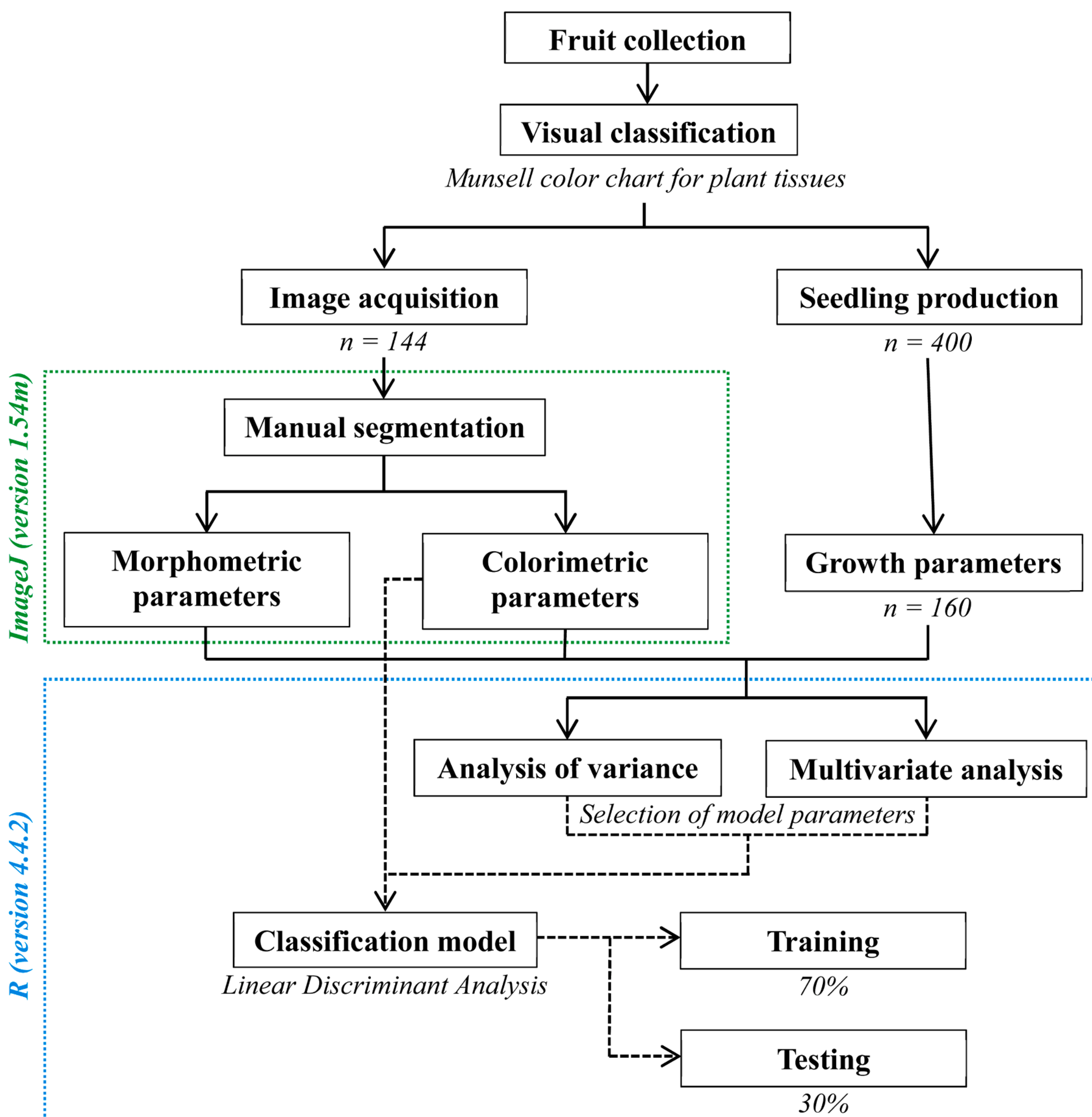

Fig. 2. Overview of the dataset acquisition and analysis workflow for model development.

content was adjusted daily to maintain “pot capacity” [34]. The relative air humidity was around 65 %, and the average temperature was 24.8 °C, monitored with a digital thermo-hygrometer (Akso, model AK632) from seedling emergence to seedling evaluation. The percentage of emergence and the seedling emergence speed index – ESI [35] were recorded daily until stabilization, according to the Eq. (3):

$$ESI = \sum \frac{n_D}{D} \quad (3)$$

where n is the number of emerged seedlings on day D; and D is the number of days after sowing.

At 80 days after sowing (DAS), four replicates of ten seedlings per treatment were randomly collected for measurement of the following parameters: number of leaves (NL); shoot height and main root length (RL), using a graduated ruler (cm); stem diameter, measured using a caliper (cm). The shoot dry mass (SDM) and root system dry mass (RDM) were determined using an analytical balance (g seedling^{-1}), after drying in a forced-air oven at 65 °C for 72 h. Subsequently, the biomass partition for shoot and root system was determined. Finally, the Dickson Quality Index (DQI) was calculated, according to the Eq. (4) proposed by [36]:

$$DQI = \frac{\text{Total dry mass}}{\left(\frac{\text{Height}}{\text{Diameter}}\right) + \left(\frac{\text{SDM}}{\text{RDM}}\right)} \quad (4)$$

2.4. Classification of fruits based on colorimetric parameters

A maturation stage classification model was developed using traditional machine learning techniques based on Linear Discriminant Analysis (LDA). A database consisting of the colorimetric variables obtained from the evaluated fruits ($n = 144$) was established, which was randomly subdivided into two sets: 70 % for model training and 30 % for performance testing. The core methodological steps involved in the acquisition and analysis of the dataset, which underpinned the development of the model through a machine learning approach, are presented in Fig. 2.

2.5. Statistical analysis

A completely randomized experimental design was adopted, considering four categories of epicarp coloration and four replicates. After verifying the normality of the data using the Shapiro-Wilk test, analysis of variance (ANOVA) was performed. When significant, the Tukey test ($p \leq 0.05$) was used to assess statistical differences. Furthermore, the data was subjected to multivariate analysis using principal component analysis (PCA). The performance of the model was evaluated based on accuracy and Kappa metrics. The statistical implementation was performed using R software, version 4.4.2.

Fig. 3. Color characterization of oiti fruits (*L. tomentosa*) based on the Munsell color chart, decomposition into Red (R), Green (G), and Blue (B) bands, and grayscale representation. Scale bar corresponds to 1 cm.

3. Results and discussion

3.1. Colorimetric characterization of fruits

The fruits of *L. tomentosa* exhibited variations in epicarp coloration across ripening stages, identified based on the Munsell color chart. The original-colored images of the fruits and the decompositions into Red, Green, and Blue bands, as well as the grayscale representation are represented in Fig. 3.

The decomposition of the RGB color bands revealed distinct patterns capable of discriminating among fruit ripening stages (Fig. 4). The Red band demonstrated a significant increase in intensity, varying from 37.7 (GY) to 107.2 (yYR), being the main indicator of changes in the fruits. Similarly, the Green component also revealed an increase in average intensity from GY to yYR (47.9; 54.2; 65.8, respectively), although less markedly than the Red channel. Despite its relevance in distinguishing green and yellowish hues, the Green channel was not efficient in differentiating between the Ye and YR fruit ripening stages. Both the Red (62.4) and Green (36.2) channels showed a reduction in the YR pixel intensity. The mean of the RGB channels followed a similar trend, reflecting the cumulative increase in pixel intensity from GY (37.0) to yYR (63.0), followed by a decrease in YR (40.2).

Regardless of the ripening stage of *L. tomentosa* fruits, the Blue channel demonstrated the lowest intensities (25.5; 22.0; 16.1; 22.1 for GY, Ye, yYR, and YR, respectively), with low variability among treatments and a reduced incidence of cool tones. In contrast to the Red and Green channels, the Blue channel displayed an inverse pattern, with declining intensity from GY to yYR, followed by a slight increase in YR. Previous studies on Alphonso mango [37] and Swingle citrumelo [19] also indicated that this channel was insufficient for differentiating ripening stages.

The grayscale analysis revealed a progressive increase in brightness from the GY stage (47.7) to yYR (85.3), followed by a decline in YR (49.8). This pattern is associated with structural changes at the cellular level and the accumulation of pigments such as carotenoids and anthocyanins, which increase light absorption and reduce reflectance, thereby darkening the fruit's skin [25,38]. In our study, the use of the RGB system proved to be a practical and objective approach to characterize *L. tomentosa* fruit ripening stages, overcoming the subjective limitations of visual analysis. These findings corroborate with other studies [27,19,25], which also evidenced the high capacity of the RGB system to classify fruits in different ripening stages.

3.2. Morphometric characterization of fruits

The morphometric analysis of the fruits revealed discrete variations among the analyzed parameters, revealing relative stability in the fruit shape throughout maturation (Table 2). The variables perimeter (two-dimensional contour length) and surface area (two-dimensional space occupied by the fruit) demonstrated a significant trend of change as the fruits ripened, especially between GY and YR, evidencing that maturation is associated with a slight reduction in fruit size.

Parameters related to shape, such as aspect ratio and roundness, showed expected variations, as they are inversely proportional. The aspect ratio, which expresses fruit elongation, decreased from 1.52 ±

Table 2
Morphometric characterization of oiti (*L. tomentosa*) fruits harvested at different stages of ripening, characterized based on pericarp coloration using the Munsell color chart.

Parameters	Epicarp color of fruits				CV (%)
	GY	Ye	yYR	YR	
Perimeter (cm)	15.72 ± 1.19 a	14.33 ± 1.28 b	14.29 ± 1.31 b	14.37 ± 0.98 b	3.37
Surface area (cm^2)	17.97 ± 2.68 a	15.30 ± 2.47 b	15.37 ± 2.64 b	15.21 ± 2.09 b	7.43
Aspect ratio	1.52 ± 0.10 a	1.45 ± 0.13 ab	1.40 ± 0.11 b	1.44 ± 0.13 ab	3.08
Roundness	0.66 ± 0.05 b	0.70 ± 0.06 ab	0.72 ± 0.05 a	0.70 ± 0.06 ab	3.01
Circularity	0.91 ± 0.03 b	0.93 ± 0.03 a	0.94 ± 0.03 a	0.92 ± 0.03 ab	1.17
Length (cm)	5.89 ± 0.49 a	5.30 ± 0.59 b	5.21 ± 0.56 b	5.26 ± 0.43 b	3.15
Width (cm)	3.87 ± 0.33^{ns}	3.66 ± 0.27	3.73 ± 0.32	3.67 ± 0.30	4.74

*Means followed by the same letter do not differ statistically from each other according to the Tukey test ($p < 0.05$). Values are represented as mean ± standard deviation.
Greenish Green-Yellow (GY); Yellow (Ye); Yellow Yellow-Red (yYR); Yellow-Red (YR).

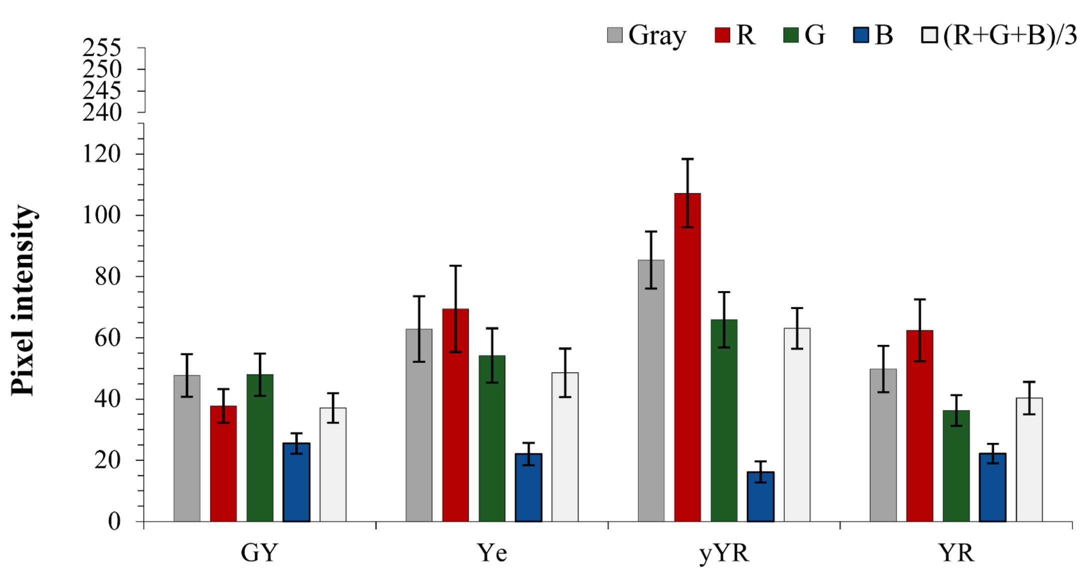

Fig. 4. Pixel intensity for each component of red (R), green (G), blue (B), average pixel value (R+G+B)/3 and gray scale for the color of oiti (*L. tomentosa*) fruits harvested at different stages of ripening.
Greenish Green-Yellow (GY); Yellow (Ye); Yellow Yellow-Red (yYR); Yellow-Red (YR).
Values are represented as mean ± standard deviation.

0.10 (GY) to 1.40 ± 0.11 (yYR), indicating less elongated fruits in yYR. In contrast, roundness increased from 0.66 ± 0.05 (GY) to 0.72 ± 0.05 (yYR), reflecting greater circularity in the latter. Circularity, which ranges from 0 (elongated fruits) to 1 (completely circular), increased slightly from GY to yYR, showing a slight reduction in the most mature stage (YR). The length showed a slight reduction throughout maturation, while the width remained stable, without significant variations.

Teixeira et al. [39], have already evidenced the stability in the length and diameter of *L. tomentosa* fruits. Similarly, Monteiro et al. [10] recorded average lengths (6.19 cm) and widths (3.3 cm) in fruits harvested in Jaboticabal-SP (Brazil), values compatible with those obtained in this study. These data confirm that, despite subtle visual and structural changes, the main dimensions of the fruits remain stable during ripening, making morphometric parameters ineffective to discriminate against the ripening stages of oiti fruits.

3.3. Seedling quality assessment

The maturation stages of the fruits significantly influenced the vigor of *L. tomentosa* seedlings (Table 3). Fruits in early maturation stages (GY) exhibited the lowest percentages of seedling emergence (58 %), seedling emergence speed index (ESI), main root length (MRL), shoot dry mass (SDM), and Dickson Quality Index (DQI). On the other hand, more mature fruits (YR) resulted in a higher percentage of seedling emergence (88 %) and superior results for ESI, SDM, and DQI, indicating superiority in the production of vigorous seedlings. The seedling emergence dynamics also varied among the maturation stages: GY fruits initiated the process at 43 DAS, stabilizing at 71 DAS, while YR fruits initiated at 38 and stabilized at 77 DAS.

L. tomentosa fruits from different maturation stages, also identified by epicarp colors, showed a high percentage of seedling emergence in sand, but without difference between dark green wrinkled fruits (86 %) and yellow fruits (95 %). Yellow fruits exhibited higher ESI, possibly due to the lower epicarp hardness compared to dark green fruits [11].

Seedlings from yYR and YR fruits showed about 40 % more leaves compared to those from gGY fruits (Table 3). There was no difference for seedling height and stem diameter. YR fruits produced seedlings with a higher average MRL (23.66 cm), although without difference in relation to Ye and yYR. Advanced fruit maturation may favor root growth, albeit with subtle variations.

The accumulated seedling biomass was also directly influenced by the fruit maturation stage. Seedlings from seeds from yYR and YR fruits showed higher RDM values, differing from GY and Ye, which exhibited

Table 3
Influence of fruit ripening stage on growth parameters of oiti (*L. tomentosa*) seedlings.

Parameters	Epicarp color of fruits				CV (%)
	GY	Ye	yYR	YR	
Seedling emergence (%)	58 c*	72 b	85 ab	88 a	9.7
Emergence speed index	0.25 b	0.26 b	0.39 a	0.40 a	10.8
Number of leaves	5 b	5 b	7 a	7 a	10.3
Height (cm)	29.74^{ns}	28.44	31.28	31.03	9.9
Diameter (cm)	4.65^{ns}	5.55	4.84	4.70	11.6
MRL (cm)	17.33 b	20.83 ab	21.15 ab	23.66 a	9.03
SDM (g seedling $^{-1}$)	6.29 d	9.06 c	13.26 b	18.58 a	8.3
RDM (g seedling^{-1})	2.30 b	2.63 b	3.32 a	4.44 a	11.2
SDM partition (%)	73.67^{ns}	77.46	79.91	80.65	5.46
RDM partition (%)	26.33^{ns}	22.54	20.09	19.35	10.2
DQI	0.94 c	1.36 b	1.58 b	2.13 a	11.4

* Means followed by the same letter do not differ statistically from each other according to the Tukey test ($p < 0.05$). Values are represented as mean ± standard deviation.

Greenish Green-Yellow (GY); Yellow (Ye); Yellow Yellow-Red (yYR); Yellow-Red (YR).

Main root length (MRL); Shoot dry mass (SDM); RDM = Root dry mass (RDM); Dickson Quality Index (DQI)

the lowest values. YR fruits showed superior SDM (18.58 g seedling^{-1}), while GY fruits produced seedlings with lower dry mass (6.29 g seedling^{-1}). This behavior reflects the greater availability of reserves accumulated in seeds from mature fruits, which were used to promote greater initial development. Similar results were observed in *Dovyalis hebecarpa* (Gardner) Warb. seeds, where seeds extracted from very mature fruits favored the increase of RDM and SDM [40].

No differences were observed between biomass partitioning in the shoot and root systems. In percentage terms, however, seedlings derived from mature fruits (YR) allocated a greater proportion of biomass to the shoot (80.65 %) compared to those derived from green fruits (GY), which presented 73.67 %. Conversely, GY fruits directed a higher amount of dry matter to the root system (26.33 %) compared to the 19.35 % observed in seedlings derived from YR fruits.

The fact that seeds contained in GY fruits originated seedlings with lower dry matter allocation to the shoot may be associated with the incomplete maturation of these seeds. In this regard, incomplete maturation may be corroborated by the lower vigor presented by the seedlings, as evidenced by variables such as seedling emergence, emergence speed index, number of leaves, MRL, SDM, RDM, and Dickson Quality Index (Table 3). Less vigorous seedlings, with lower initial growth of the root system compared to more vigorous ones, may, over time, allocate drier biomass to the shoot, thereby reducing the shoot-to-root ratio. Meanwhile, seeds derived from more mature fruits (YR) completed the maturation process, resulting in more vigorous seedlings, as demonstrated in Table 3. Gomes-Junior et al. [19] reported for Swingle citrumelo seeds extracted from Green fruits exhibited lower vigor, supported by variables such as first germination count, seedling emergence, seedling dry mass, and seed vigor index (SVIS), when compared to seeds from Greenish-Yellow and Yellow fruits.

The Dickson Quality Index (DQI) distinguished the performance of seedlings derived from fruits at different maturation stages. Seedlings obtained from GY fruits exhibited the lowest DQI (0.94), a result consistent with the lower accumulation of biomass in both the shoot and root systems. In contrast, the YR group presented the highest index value (2.13), despite showing the highest shoot-to-root dry mass ratio (SDM/RDM), indicating a disproportionate distribution of biomass favoring the aerial part. This result demonstrates that, although the DQI penalizes morphological imbalance, the substantial total biomass accumulation in this group was sufficient to offset this effect and raise the index value. Given that average height and collar diameter varied little between the evaluated groups, the increase in DQI was associated with the amount of accumulated biomass.

The DQI was calculated based on the ratio between the total dry mass and the sum of two quotients: the height-to-diameter ratio (robustness) and the SDM/RDM ratio. It is a dimensionless morphological variable that integrates multiple attributes into a single value and is used as a criterion for evaluating seedling quality. In this context, the higher the DQI value, the higher the seedling quality, reflecting biomass accumulation and partitioning, as well as robustness [41]. Thus fruit maturation plays an important role in the production of vigorous and high-quality seedlings. More mature fruits not only optimized emergence rates but also favored shoot biomass allocation and balanced seedling development.

3.4. Association between colorimetric parameters, morphological traits, and seedling performance

Principal Component Analysis (PCA) was employed as an integrative approach among colorimetric parameters, morphological traits, and seedling performance, allowing the identification of relevant patterns and the synthesis of interrelationships among the data obtained. PCA is a widely used statistical tool for reducing the dimensionality of a dataset while preserving most of the original variance [42]. In the present study, PCA explained 67.64 % of the total variance, with 48.34 % attributed to the first component (PC1) and 19.3 % to the second component (PC2)

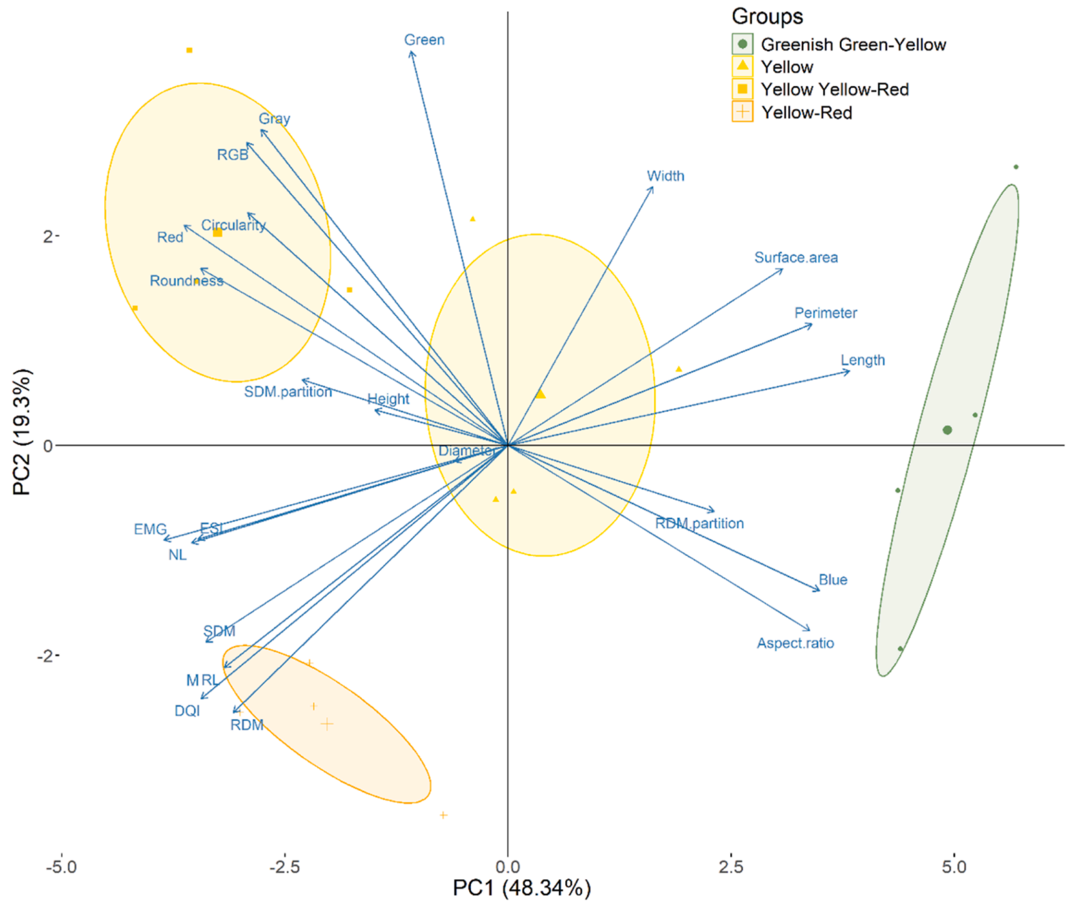

Fig. 5. Principal component analysis showing the distribution of oiti fruits (*L. tomentosa*) based on colorimetric, morphological, and physiological variables. Greenish Green-Yellow (GY); Yellow (Ye); Yellow Yellow-Red (yYR); Yellow-Red (YR); Seedling emergence (EMG); Emergence speed index (ESI); Number of leaves (NL); Main root length (MRL); Shoot dry mass (SDM); Root dry mass (RDM); Dickson Quality Index (DQI).

(Fig. 5). The morphological parameters of the fruits (width, surface area, perimeter, length, aspect ratio) and the Blue component were the main contributors to the variance explained by PC1. PC2 was influenced by colorimetric parameters (Green and Red channels, RGB and brightness – grayscale) and morphological traits (circularity and roundness).

The fruit groupings revealed clear separation patterns among maturation stages. Fruits classified as GY, positioned on the right, exhibited larger physical dimensions (perimeter, length, width, aspect ratio, and area) and higher intensity in the Blue channel. The fruits of the Ye group, located at the center of the graph, displayed intermediate characteristics, suggesting they represent a transitional stage between the other groups. The yYR group, located in the upper left corner, stood out for presenting the highest average pixel intensity values (RGB), especially in the Red channel, as well as greater brightness (Gray), circularity, and roundness. These characteristics indicate that yYR fruits already incorporate morphological and colorimetric attributes typical of mature fruits, although they are less associated with seedling performance variables. This suggests that such fruits have not yet reached a sufficient maturation stage for the production of vigorous seedlings. In contrast, the YR fruits, located in the lower left corner, showed an association with seedling performance variables such as root length (RL), shoot dry mass (SDM), root dry mass (RDM), and Dickson Quality Index (DQI). These results reinforced that the advanced maturation stage (YR) was the most suitable for the production of vigorous seedlings, as it combined mature fruit colorimetric attributes with superior performance of *L. tomentosa* seedlings.

Thus the morphological parameters of the fruits are important markers of the variations that occur throughout maturation [16]; however, we observed that for *L. tomentosa*, these parameters exhibited low variation across maturation stages and were not efficient in differentiating between stages (as shown in section 3.2). In contrast, the colorimetric attributes (section 3.1) and seedling performance variables (section 3.3) were able to capture the differences between stages, discriminating them accurately. Thus, PCA efficiently synthesized the integrated contribution of the three categories of variables (colorimetry, morphology, and seedling performance). The dataset presented thus far reinforced the role of colorimetry as a predictive and objective tool. This provided methodological support for the subsequent application of LDA algorithm, aimed at classifying fruits according to maturation stages and standardizing the seedling production process.

3.5. *Classification model basED ON LINEAR DISCRIMINANT ANALYsis*

LDA is a statistical technique used for classification problems, which aims to find the directions in the multidimensional space that maximize the variance between classes and minimize the variance within the same class simultaneously [43]. In this study, LDA was applied to classify the maturation stages of *L. tomentosa* fruits based on colorimetric parameters. Both PCA and LDA used the same dataset. The model reduced the dimensionality of the original data and projected the variations into two main axes [Function 1: (7.805 × Gray) + (-23.889 × Red) + (-6.148 × Green) + (-2.428 × Blue) + (16.478 × RGB/3); Function 2: (-2.154 × Gray) + (-462.138 × Red) + (-219.901 × Green) + (-82.416 × Blue) + (605.674 × RGB/3)], which explained 85.37 % and 12.96 % of the variance, respectively (Fig. 6). Together, the axes accumulated 98.33 % of the total variance, reflecting that the model captured almost all the relevant information to differentiate the maturation stages.

In the scatterplot, each point represents a fruit classified by the

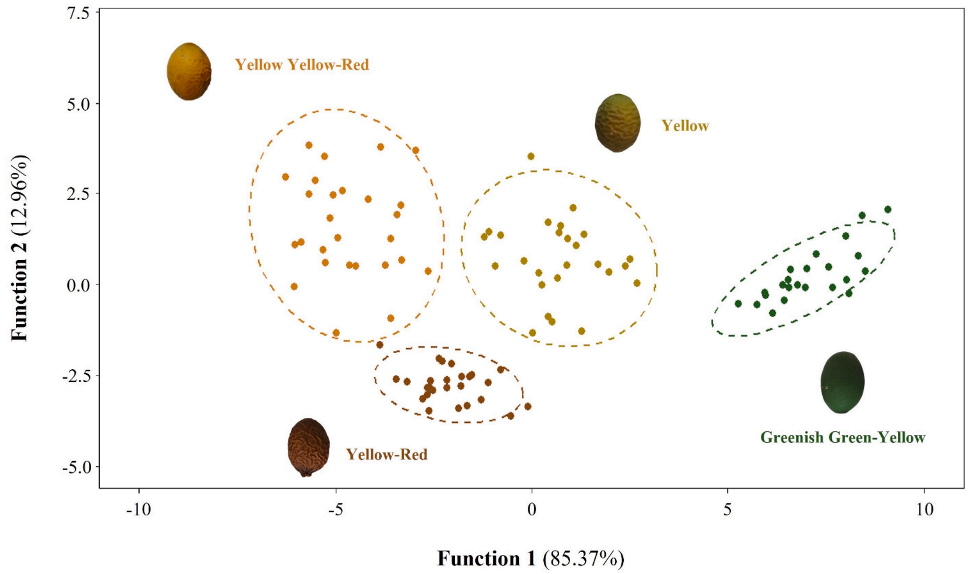

Fig. 6. Scatterplot of oiti fruit (*L. tomentosa*) ripening stages classified by Linear Discriminant Analysis based on colorimetric parameters. Greenish Green-Yellow (GY); Yellow (Ye); Yellow Yellow-Red (yYR); Yellow-Red (YR).

model, while the ellipses delimit the data concentration of each class. The arrangement of the classes by LDA (Fig. 6) corroborated the groupings observed in PCA (Fig. 5). The GY class was clearly separated on the right, while the yYR and YR classes were arranged on the left, which indicated that the model correctly identified the fruits belonging to these maturation stages. Meanwhile, the Ye class was highlighted in the center, which reflected the high similarity with the neighboring classes. This behavior, typical of transition stages, makes it difficult to differentiate these fruits, as observed in studies with cherry tomatoes in the half-ripe stage [42]. Avila et al. [44] stated that a portion of studies involving image analysis to estimate fruit maturity are unable to accurately distinguish intermediate maturation stages. This finding is also ratified by the confusion matrix (Table 4).

The confusion matrix was used to evaluate the performance of the LDA model, comparing the classes predicted by it with the actual classes (Table 4). The diagonal elements (in bold) indicate the correct classifications, while the other values represent classification errors. The GY, yYR, and YR classes were classified with 100 % accuracy, while the Ye class showed 60 % accuracy. About 40 % of the Ye fruits were erroneously assigned to other classes, with 20 % to GY; 10 % to yYR and 10 % to YR. This confirms the greater variability of colorimetric parameters for Ye fruits, making their precise separation difficult, as verified in strawberry fruits in intermediate stages [25].

The performance of the model was high, with an accuracy of 90.48 %, indicating that 9 out of 10 fruits were correctly classified. The Kappa index, of 87.17 %, reinforced the model's reliability by indicating a high

Table 4
Confusion matrix for the prediction of oiti (*L. tomentosa*) fruit ripening stages, characterized based on pericarp coloration using the Munsell color chart.

Classes	GY	Ye	yYR	YR
		%		
GY	**100.0**	20.0	0	0
Ye	0	**60.0**	0	0
yYR	0	10.0	**100.0**	0
YR	0	10.0	0	**100.0**
Accuracy (%)		90.48		
Kappa (%)		87.17		
p-value		$1.18*10^{-15}$		

Bold values indicate correct identification performance.
Greenish Green-Yellow (GY); Yellow (Ye); Yellow Yellow-Red (yYR); Yellow-Red (YR).

agreement between the predicted classes and the actual values, adjusting for chance. Furthermore, an extremely low p-value ($1.18*10^{-15}$) confirmed the separation between classes was statistically significant. Similar studies, employing LDA with RGB average intensity data, reported error rates of 17.5 % for mangoes in three maturation stages [24] and 33.8 % for four apple stages [17].

The results highlighted the efficiency of LDA in classifying the maturation stages of *L. tomentosa* fruits, especially for the GY, yYR, and YR classes. The lower accuracy for the Ye class emphasizes the need for complementary investigations, such as the texture and chemical composition of the fruits, to increase the model's precision.

4. Conclusions

For the first time, itt was proposed a robust and non-invasive methodology for the selection of *L. tomentosa* fruits for seedling production, combining image analysis and machine learning techniques to provide objective parameters for the classification of maturation stages.

The characterization of each maturation class in the Munsell color chart and its relationship with variations in epicarp coloration (RGB system) and seedling performance was evident. Immature fruits, characterized by a greenish coloration (Greenish Green-Yellow), exhibited lower percentages of seedling emergence, emergence speed, and seedling development. Conversely, fruits at more advanced maturation stages yielded superior seedling performance, indicating Yellow-Red fruits as the most suitable to produce high-quality seedlings. Overall, morphological parameters remained stable throughout maturation, rendering them ineffective for discriminating between the ripening stages of oiti fruits.

The utility of colorimetric analysis integrated with machine learning algorithms to optimize fruit selection processes. We specifically examined the predictive capacity of extracted colorimetric features for discerning fruit ripening stages, employing Linear Discriminant Analysis for model development and validation. The resultant classification model demonstrated robust performance across diverse maturation stages, exhibiting high accuracy, particularly for fruits categorized as Greenish-Green-Yellow, Yellow-Yellow-Red, and Yellow-Red. The proposed evaluation approach has potential to boost the rational and economically sustainable cultivation of *L. tomentosa* plants, by enabling the rapid and practical detection of propagative material quality, regardless of maturation time knowledge.

Certain limitations persists, as our analyses relied on a relatively small sample size, and all image acquisitions occurred under controlled environmental conditions, which may limit the generalizability of the findings to real-world field scenarios. To improve the applicability and robustness of the proposed approach, future research should incorporate fruit samples derived from diverse mother plant populations and conduct image acquisition under natural field conditions. Furthermore, exploring alternative color systems beyond RGB, such as HSV (Hue, Saturation, Value), CMYK (Cyan, Magenta, Yellow, Key/Black), and CIELab (Luminosity, Red-Green axis, Yellow-Blue axis) may confer superior sensitivity to discrete fruit ripening stages.

Complementary assessments of fruit traits - including chemical composition, firmness, weight, and texture - are also crucial, as they could provide more accurate indicators of ripening and support the refinement of predictive algorithms. These advancements would not only strengthen the reliability of the model but also lay the foundation for developing practical tools, including mobile applications for in-field use by producers. Moreover, the integration of these data holds the potential to the establishment of a comprehensive database for mapping and characterizing *L. tomentosa* populations, thereby supporting more efficient seedling production strategies and advancing our knowledge of the species' biodiversity.

CRediT authorship contribution statement

Douglas Martins Santana: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Júlio César Altizani-Júnior:** Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis. **Francisco Guilhien Gomes-Junior:** Writing – review & editing, Visualization, Validation, Supervision, Investigation, Formal analysis. **Durval Dourado-Neto:** Writing – review & editing. **Renan Caldas Umburanas:** Writing – review & editing, Visualization. **Klaus Reichardt:** Writing – review & editing, Visualization. **Fábio Oliveira Diniz:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

To the *Instituto Federal do Piauí* (IFPI) and *the Escola Superior de Agricultura "Luiz de Queiroz"* (Esalq-USP).

Funding

This work was supported by the Fundação de Estudos Agrários "Luiz de Queiroz" (FEALQ).

Data availability

The data that supports the fundings of this work are available from the corresponding author on reasonable request.

References

[1] F. Carnevale Neto, A.C. Pilon, V.S. Bolzani, I. Castro-Gamboa, Chrysobalanaceae: secondary metabolites, ethnopharmacology and pharmacological potential, Phytochem. Rev. 12 (2013) 121–146, https://doi.org/10.1007/s11101-012-9259-z.

[2] N.A. Jud, C.W. Nelson, F. Herrera, Fruits and wood of *Parinari* from the early Miocene of Panama and the fossil record of Chrysobalanaceae, Am. J. Bot. 103 (2016) 277–289, https://doi.org/10.3732/ajb.1500425.

[3] C.A. Sothers, G.T. Prance, M.W. Chase, Towards a monophyletic Licania: a new generic classification of the polyphyletic Neotropical genus Licania (Chrysobalanaceae), Kew. Bull. 71 (2016) 58, https://doi.org/10.1007/s12225-016-9664-3.

[4] I.P. Pessoa, J.J. Lopes Neto, T.S. Almeida, D.F. Farias, L.R. Vieira, J.L. Medeiros, A. A. Boligon, A. Peijnenburg, I. Castelar, A.F.U. Carvalho, Polyphenol composition, antioxidant activity and cytotoxicity of seeds from two underexploited wild licania species: L. rigida and L. tomentosa, Molecules 21 (2016) 1755, https://doi.org/10.3390/molecules21121755.

[5] M.F. Moro, C. Westerkamp, F.S. De Araújo, How much importance is given to native plants in cities' treescape? A case study in Fortaleza, Brazil, Urban. For. Urban. Green. 13 (2014) 365–374, https://doi.org/10.1016/j.ufug.2014.01.005.

[6] P.L. Alves, K.T.M. Formiga, M.A.B. Traldi, Rainfall interception capacity of tree species used in urban afforestation, Urban. Ecosyst. 21 (2018) 697–706, https://doi.org/10.1007/s11252-018-0753-y.

[7] W.K.D. Freitas, L.M.S. Magalhães, C.A.A.D. Santana, E.R. Pereira Junior, L.D.C.M. D. Souza, R.A.B. Toledo, B.R. Garção, Tree composition of urban public squares located in the Atlantic forest of Brazil: a systematic review, Urban. For. Urban. Green. 48 (2020) 126555, https://doi.org/10.1016/j.ufug.2019.126555.

[8] R.O.D. Santos, R.N. Soares, B.M.D.S.E. Silva, Compositional similarity of urban green areas in Southeastern Brazil, Floresta Ambient. 26 (2019) e20170747, https://doi.org/10.1590/2179-8087.074717.

[9] D.O. Lisboa, M.A. Silva, D.B. Pinho, O.L. Pereira, G.Q. Furtado, Diversity of pathogenic and endophytic Colletotrichum isolates from Licania tomentosa in Brazil, For. Pathol. 48 (2018) e12448, https://doi.org/10.1111/efp.12448.

[10] K.L. Monteiro, C. Oliveira, B.M.S. Silva, F.V. Môro, D.A. Carvalho, Caracterização morfológica de frutos, de sementes e do desenvolvimento pós-seminal de Licania tomentosa (Benth.) Fritsch, Ciênc. Rural. 42 (2012) 90–97.

[11] J.J.R. Silva, M.I.M. Silva, G.A. Azeredo, V.C. Souza, Emergência de plântulas e armazenamento de endocarpos de *Licania tomentosa* (Benth.) Fritsch, Nucleus 15 (2018) 93–102, https://doi.org/10.3738/1982.2278.2717.

[12] V.C.D. Alves, A.G. Silva, Edilson Costa, T. Dantas, F.F.S. Binotti, M.B. Martins, E. P. Vendruscolo, G.H.C. Vieira, Influence of different protected environments and reflector material on cultivated Oiti (*Licania tomentosa* [Benth.] Fritsch) seedlings, Arboric. Urban For. 49 (2023) 159–169, https://doi.org/10.48044/jauf.2023.015.

[13] H. Lorenzi, Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil, 5th ed., Plantarum, Nova Odessa, São Paulo, 1992.

[14] S.C. Grossnickle, Why seedlings survive: influence of plant attributes, New For. 43 (2012) 711–738, https://doi.org/10.1007/s11056-012-9336-6.

[15] R.S.R. Moreira-Araújo, N.V.A. Barros, R.G.C.L. Porto, A.C.A.S. Brandão, A. Lima, R. Fett, Bioactive compounds and antioxidant activity three fruit species from the Brazilian Cerrado, Rev. Bras. Frutic. 41 (2019), https://doi.org/10.1590/0100-29452019011 e-011.

[16] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, A. Albarelli, Fruit ripeness classification: a survey, Artif. Intell. Agric. 7 (2023) 44–57, https://doi.org/10.1016/j.aiia.2023.02.004.

[17] C. Garrido-Novell, D. Pérez-Marin, J.M. Amigo, J. Fernández-Novales, J. E. Guerrero, A. Garrido-Varo, Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras, J. Food Eng. 113 (2012) 281–288, https://doi.org/10.1016/j.jfoodeng.2012.05.038.

[18] A.S. Oliveira, T.T. Almeida, R.G.E. Reis, Morfologia interna e externa de diásporos de *Tectona grandis* L. e sua relação com a qualidade de sementes, Ciênc. Florest 33 (2023) e71755, https://doi.org/10.5902/1980509871755.

[19] F.G. Gomes-Junior, N. Arruda, J. Marcos-Filho, Swingle citrumelo seed vigor and storability associated with fruit maturity classes based on RGB parameters, Sci. Agric. 74 (2017) 357–363, https://doi.org/10.1590/1678-992x-2016-0173.

[20] P.B. Pathare, U.L. Opara, F.A.J. Al-Said, Colour measurement and analysis in fresh and processed foods: a review, Food Bioproc. Technol. 6 (2013) 36–60, https://doi.org/10.1007/s11947-012-0867-9.

[21] M. Palumbo, M. Cefola, B. Pace, G. Attolico, G. Colelli, Computer vision system based on conventional imaging for non-destructively evaluating quality attributes in fresh and packaged fruit and vegetables, Postharvest Biol. Technol. 200 (2023) 112332, https://doi.org/10.1016/j.postharvbio.2023.112332.

[22] W. Castro, J. Oblitas, M. De-La-Torre, C. Cotrina, K. Bazan, H. Avila-George, Classification of cape gooseberry fruit according to its level of ripeness using machine learning techniques and different color spaces, IEEe Access. 7 (2019) 27389–27400, https://doi.org/10.1109/ACCESS.2019.2898223.

[23] A. Vidal, P. Talens, J.M. Prats-Montalbán, S. Cubero, F. Albert, J. Blasco, In-line estimation of the standard colour index of citrus fruits using a computer vision system developed for a mobile platform, Food Bioprocess. Technol. 6 (2013) 3412–3419, https://doi.org/10.1007/s11947-012-1015-2.

[24] V. Eyarkai Nambi, K. Thangavel, S. Shahir, V. Thirupathi, Comparison of various RGB image features for nondestructive prediction of ripening quality of "alphonso" mangoes for easy adoptability in machine vision applications: a multivariate approach, J. Food Qual. 39 (2016) 816–825, https://doi.org/10.1111/jfq.12245.

[25] S. Karki, J.K. Basak, B. Paudel, N.C. Deb, N.E. Kim, J. Kook, M.Y. Kang, H.T. Kim, Classification of strawberry ripeness stages using machine learning algorithms and colour spaces, Hortic. Env. Biotechnol. 65 (2024) 337–354, https://doi.org/10.1007/s13580-023-00559-2.

[26] H. Galal, S. Elsayed, O. Elsherbiny, A. Allam, M. Farouk, Using RGB imaging, optimized three-band spectral indices, and a decision tree model to assess orange fruit quality, Agriculture 12 (2022) 1558, https://doi.org/10.3390/agriculture12101558.

[27] T.B. Bello, A.G. Costa, T.R. Silva, J.L. Paes, M.V.M. Oliveira, Tomato quality based on colorimetric characteristics of digital images, Rev. Bras. Eng. Agríc. Ambient 24 (2020) 567–572, https://doi.org/10.1590/1807-1929/agriambi.v24n8p567-572.
[28] P. Wan, A. Toudeshki, H. Tan, R. Ehsani, A methodology for fresh tomato maturity detection using computer vision, Comput. Electron. Agric. 146 (2018) 43–50, https://doi.org/10.1016/j.compag.2018.01.011.
[29] W. Shao, Y. Li, S. Diao, J. Jiang, R. Dong, Rapid classification of Chinese quince (Chaenomeles speciosa Nakai) fruit provenance by near-infrared spectroscopy and multivariate calibration, Anal. Bioanal. Chem. 409 (2017) 115–120, https://doi.org/10.1007/s00216-016-9944-7.
[30] A. Tharwat, T. Gaber, A. Ibrahim, A.E. Hassanien, Linear discriminant analysis: a detailed tutorial, AIC 30 (2017) 169–190, https://doi.org/10.3233/AIC-170729.
[31] A. Sucipto, A.K. Zyen, B.B. Wahono, T. Tamrin, H. Mulyo, R.R. Ali, Linear discriminant analysis for apples fruit variety based on color feature extraction, in: 2021 International Seminar on Application for Technology of Information and Communication (iSemantic). Presented at the 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), IEEE, Semarangin, Indonesia, 2021, pp. 184–189, https://doi.org/10.1109/isemantic52711.2021.9573200.
[32] N. Zulkifli, N. Hashim, H.H. Harith, M.F. Mohamad Shukery, D.I. Onwude, Prediction of the ripening stages of papayas using discriminant analysis and support vector machine algorithms, J. Sci. Food Agric. 102 (2022) 3266–3276, https://doi.org/10.1002/jsfa.11669.
[33] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (2012) 671–675, https://doi.org/10.1038/nmeth.2089.
[34] P.C. Teixeira, G.K. Donagemma, A. Fontana, W.G. Teixeira, Manual De Métodos De Análise De Solos, 3rd ed., Embrapa, Brasília, 2017.
[35] J.D. Maguire, Speed of Germination—Aid In selection and evaluation for seedling emergence and vigor, Crop. Sci. 2 (1962) 176–177, https://doi.org/10.2135/cropsci1962.0011183X000200020033x.
[36] A. Dickson, A.L. Leaf, J.F. Hosner, Quality appraisal of white spruce and white pine seedlings stock in nurseries, For. Chron. 36 (1960) 10–13, https://doi.org/10.5558/tfc36010-1.
[37] R.P. Salunkhe, A.A. Patil, Image processing for mango ripening stage detection: RGB and HSV method, in: 2015 Third International Conference on Image Information Processing (ICIIP). Presented at the 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, India, IEEE, 2015, pp. 362–365, https://doi.org/10.1109/ICIIP.2015.7414796.
[38] W.W. Zhang, S.Q. Zhao, L.C. Zhang, Y. Xing, W.S. Jia, Changes in the cell wall during fruit development and ripening in Fragaria vesca, Plant Physiol. Biochem. 154 (2020) 54–65, https://doi.org/10.1016/j.plaphy.2020.05.028.
[39] L. Teixeira, L. de, G.A. Silva, J.B.M. Macedo, M.G. Almeida, M.F.V. Moura, Physicochemical characterization and phenolic profile of oiti fruits (Licania tomentosa Benth Fristch), Jnfrt 2 (2019) 7–12, https://doi.org/10.30881/jnfrt.00015.
[40] F. Villa, D.F. Silva, M.C.C. Rotili, N.F.M. Herzog, M.M. Malavasi, Seed physiological quality and harvest point of dovyalis fruits, Pesqui. Agropecu. Trop 49 (2019) e54520, https://doi.org/10.1590/1983-40632019v4954520.
[41] K.H. Lin, C.W. Wu, Y.S. Chang, Applying dickson quality index, chlorophyll fluorescence, and leaf area index for assessing plant quality of pentas lanceolata, Not. Bot. Horti. Agrobo. 47 (2018) 169–176, https://doi.org/10.15835/nbha47111312.
[42] Y. Zhang, X. Yin, X. Zou, J. Zhao, On-line sorting maturity of cherry tomato bymachine vision. Computer and Computing Technologies in Agriculture II, Volume 3, IFIP Advances in Information and Communication Technology, Springer US, Boston, MA, 2009, pp. 2223–2229, https://doi.org/10.1007/978-1-4419-0213-9_74.
[43] T. Amoriello, R. Ciccoritti, M. Paliotta, K. Carbone, Classification and prediction of early-to-late ripening apricot quality using spectroscopic techniques combined with chemometric tools, Sci. Hortic. 240 (2018) 310–317, https://doi.org/10.1016/j.scienta.2018.06.031.
[44] F. Avila, M. Mora, M. Oyarce, A. Zuñiga, C. Fredes, A method to construct fruit maturity color scales based on support machines for regression: Application to olives and grape seeds, J. Food Eng. 162 (2015) 9–17, https://doi.org/10.1016/j.jfoodeng.2015.03.035.