

Artificial intelligence applied to the classification of greenish seeds and prediction of physiological quality in soybean

Inteligência artificial aplicada à classificação de sementes esverdeadas e predição de qualidade fisiológica em soja

Vasconcelos Romão de Oliveira^{1*}, Nara Oliveira Silva Souza¹, Márcio da Silva Arantes², Ricardo Carmona¹

ABSTRACT

The presence of greenish seeds represents an obstacle to the productive potential of soybean cultivation, causing significant impacts on the visual aspect and physiological quality of seeds. Traditionally, seeds are evaluated visually, a method that is subject to subjectivity and human error. This research proposes an innovative approach that integrates image analysis and artificial intelligence to develop a machine learning model capable of distinguishing greenish seeds from yellow ones based on color parameters. This study aims to enhance the accuracy of seed evaluation and expand understanding of the relationship between seed color tone and their physiological quality. The artificial intelligence was trained with 12,000 images captured and processed by the GroundEye® S800D. The methodology employed to train the system involved the use of a decision tree, utilizing the *sklearn.tree* library from Python. Each seed, after image capture, underwent a standard germination test. The normal seedlings were then reanalyzed using the GroundEye® S800D to determine their vigor through measurements of primary root and hypocotyl sizes. Yellow soybean seeds exhibit superior physiological quality compared to greenish ones, particularly in terms of germination and seedling growth. The hue angle (h) and luminosity (L) proved to be the most responsive criteria in the machine learning model, achieving an accuracy of 89.7%. The hue angle was demonstrated to be a robust predictor, correlating with higher germination rates in seeds with an angle less than 97.5°. The relationship between seed viability and hue angle was supported by a coefficient of determination (R^2) of 73%.

Index terms: Seeds quality; image analysis; machine learning model.

RESUMO

A presença de sementes esverdeadas representa um obstáculo ao potencial produtivo da cultura da soja, gerando impactos significativos no aspecto visual e na qualidade fisiológica. Tradicionalmente, as sementes são avaliadas visualmente, um método sujeito à subjetividade e erros humanos. Frente a isto, esta pesquisa propõe uma abordagem inovadora que integra análise de imagens e inteligência artificial para desenvolver um modelo de aprendizado de máquina capaz de distinguir sementes esverdeadas de amarelas com base em parâmetros de cor. Este estudo visa aprimorar a precisão na avaliação das sementes e expandir o entendimento sobre a relação entre a tonalidade da cor das sementes e sua qualidade fisiológica. A inteligência artificial foi treinada com 12.000 imagens capturadas e processadas pelo GroundEye® S800D. A metodologia empregada para treinar o sistema envolveu a utilização de uma árvore de decisão, utilizando a biblioteca *sklearn.tree* do Python. Cada semente, após a captura de imagem, foi submetida a um teste de germinação padrão. As plântulas normais foram então reanalisadas no GroundEye® S800D para determinar seu vigor através da mensuração do tamanho de raiz primária e hipocótilo. Sementes de soja amarelas demonstram qualidade fisiológica superior às esverdeadas, destacando-se em germinação e crescimento da plântula. O ângulo de tonalidade (h) e a luminosidade (L) mostraram-se os critérios mais responsivos ao modelo de aprendizado de máquina, alcançando uma acurácia de 89,7%. O ângulo de tonalidade demonstrou ser um preditor robusto, correlacionando-se com altas taxas de germinação em sementes com ângulo inferior a 97,5°. A relação entre a viabilidade das sementes e o ângulo de tonalidade foi apoiada por um coeficiente de determinação (R^2) de 73%.

Termos para indexação: Qualidade de sementes; análise de imagem; modelo de aprendizado de máquina.

Introduction

Soybeans (*Glycine max* (L.) Merrill) play a crucial role in the global economy, being one of the main products of the global agribusiness. They are one of the largest sources of plant protein and edible oil, vital for food chains worldwide (Singh & Shivakumar, 2010). Additionally, the Food and Agriculture Organization - FAO (2021) highlights soy as an essential component in the transition to biofuels, making it a key element in discussions on both food security and energy sustainability.

There is a technical consensus that high-quality seeds exhibit greater vigor, increased disease resistance, better field establishment, and higher productivity. On the other hand, greenish seeds pose a challenge to the maximum performance

Agricultural Sciences

Ciênc. Agrotec., 48:e003624, 2024
<http://dx.doi.org/10.1590/1413-7054202448003624>

Editor: Renato Paiva

¹Universidade de Brasília, Faculdade de Agronomia e Medicina Veterinária, Brasília, DF, Brasil

²Instituto SENAI de Inovação em Sistemas Embarcados, Florianópolis, SC, Brasil

Corresponding author: vasco.r.oliv@gmail.com

Received in February 20, 2024 and approved in May 10, 2024

desired in soy production. These seeds are characterized by high chlorophyll content within, resulting from physiological immaturity or adverse cultivation conditions. The presence of a green hue in the cotyledons indicates that the maturation process was prematurely interrupted, resulting in decreased quality and productivity of the seeds (Zorato et al., 2007; Arruda et al., 2016; Teixeira et al., 2020), and consequently impacting the profitability of the crops. Therefore, it is essential to implement quality control techniques capable of identifying and assessing the impact of greenish seeds on soybean cultivation, thereby ensuring the commercialization of products with high agronomic performance potential.

Although significant advancements have been made in the evaluation of seed quality, with the standardization of methodologies for analyzing physical, physiological, sanitary, and genetic parameters, most laboratory assessments are still performed subjectively. This subjective nature is largely due to the reliance on the experience of technicians responsible for the visual inspections of the seeds. In response to this challenge, advanced techniques have been developed, providing a more accurate and objective assessment of seed quality (Patel et al., 2012; Xia et al., 2019), driven by the revolution of artificial intelligence (AI) applied to agriculture.

Undoubtedly, AI is redefining boundaries in various fields of activity, standing out for its ability to process and analyze data efficiently, especially through techniques of computer vision and machine learning (Russell & Norvig, 2016). According to Bishop (2006), while computer vision facilitates the capture and processing of images, machine learning is dedicated to the development of algorithms and models that learn patterns and make predictions based on data. The combined application of these two techniques is capable of identifying, categorizing, and even predicting the quality of seeds from different crops, as several studies have demonstrated, especially in the last decade (Škrubej et al., 2015; Momin et al., 2017; Mahajan et al., 2018; Medeiros et al., 2020; Barros et al., 2021; Thakur et al., 2022).

Image analysis, by evaluating various characteristics of seeds such as size, shape, color, and texture, is capable of detecting defects and irregularities often overlooked by conventional visual assessment methods (Bauriegel et al., 2011). Moreover, this approach can offer a significant advantage in terms of operational efficiency and cost, due to its faster processing compared to manual methods.

In this study, the main objective was to create a classification model to distinguish greenish soybean seeds from yellow ones through image analysis, using machine learning techniques to construct a model capable of distinguishing the color of the seeds. Through this work, the goal was not only to improve the accuracy in seed evaluation but also to expand knowledge on how the color tone of the seeds can affect physiological quality.

Material and Methods

Study description

The experiments were conducted at the Seed Analysis Laboratory of Corteva Agriscience do Brasil Ltda, a multinational agricultural company located in Brasília, Distrito Federal, Brazil.

The seeds were produced and processed in the municipality of Lagoa da Confusão in the state of Tocantins, during the 2021/2021 growing season. The harvest from the production fields was carried out in October 2021, in a mechanized manner and with technical-agronomic monitoring. Although the cultivation occurred during the sanitary vacuum, the region had legal authorization to produce seeds during this period. The sanitary vacuum is a strategic pause during the off-season, during which living soy plants are not allowed in Brazilian agricultural fields as a preventive measure against Asian rust, a disease caused by the fungus *Phakopsora pachyrhizi* (Godoy et al., 2017).

One of the main challenges faced by soybean seed producers in the Tocantins region is the high prevalence of greenish seeds, a phenomenon extensively documented in the literature (Pádua et al., 2007; Arruda et al., 2016; Ferrari et al., 2023). This problem is often attributed to environmental stress conditions, such as water deficit or high temperatures (França-Neto et al., 2016), which are typical characteristics of this region. Thus, the expectation of a high incidence of greenish seeds was the determining factor for choosing lots produced from this origin. Additionally, the fact that the seeds had undergone natural environmental stress made them an ideal subject for the purpose of the study.

Ten batches of seeds were selected, each weighing 1 kilogram, belonging to two distinct cultivars, coded as A and B. For each cultivar, the ten batches were combined into a single composite sample, resulting in a total of 10 kilograms per cultivar. This quantity was adequate to ensure a sufficient source for the manual and individual selection of the seeds analyzed in this study, allowing for a comparison between greenish seeds and yellow seeds.

The samples were stored under controlled environmental conditions during the testing period, where the temperature was maintained below 13 °C and the relative humidity of the air did not exceed 60%. The segregation, identification, and storage occurred in a manner to ensure the correct traceability of the samples.

Image analysis through GroundEye®

The GroundEye® is a technological innovation developed and patented by the Brazilian company Tbit Tecnologia e Sistemas, launched in 2012 under the name Seed Analysis System - SAS® and later renamed GroundEye. The GroundEye S800D® is one of the models offered as a solution for the analysis of images of seeds, seedlings, and leaves. It is an advanced system that combines hardware and software, equipped with two high-resolution cameras that capture images of the top and bottom of each object analyzed.

The system is capable of providing detailed information about the color, shape, geometry, and texture of the analyzed objects.

An experienced analyst selected 3,000 seeds from each cultivar (A and B) through visual inspection, totaling 6,000 units. Within each cultivar, a distinct group of 1,500 “yellow seeds” and another 1,500 “greenish seeds” was separated. Thus, each experimental group consisted of a set of 3,000 seeds. A seed was considered yellow if it showed no traces of chlorophyll pigmentation, while greenish seeds were identified when such evidence was noticed, even if only partially. Each seed was assigned a unique numerical identifier.

After manual selection, each seed was individually positioned on a clear acrylic tray inside the GroundEye® S800D equipment for image capture. These images were acquired using the camera settings standardized by the manufacturer to ensure consistency in lighting and positioning. Figure 1 displays a representation of the yellow and greenish seeds captured and processed by the device. The background of the image is blue to contrast the colors of the seeds. The image files are originally in PNG format, with an average size of 17KB.

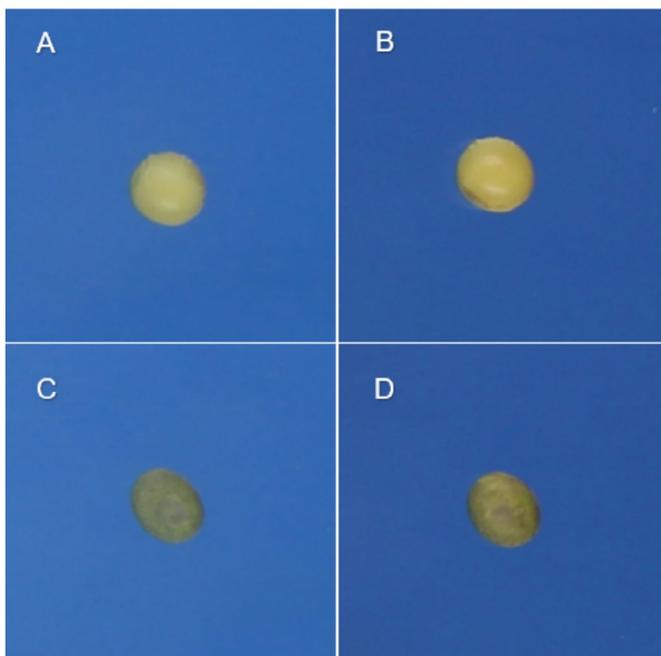


Figure 1: Images of soybean seeds captured by the GroundEye® S800D. (A) and (B) illustrate yellow seeds; (C) and (D) show greenish seeds.

The decision to analyze each seed individually was made with the goal of building a robust dataset for the machine learning model, as well as enabling a direct correlation with physiological quality. Thus, a total of 3,000 analyses were processed for the “greenish seeds” set, following an identical procedure for the “yellow seeds” set. Since each captured seed

generated two images, one from above and one from below, a total of 12,000 data points were obtained for training and validating the classification algorithm. The consideration of each side of the seed as an individual data point for evaluation was based on the common analytical practice for this type of damage, where, regardless of its position on the seed, the presence of greenish tones is recorded.

Although the GroundEye® is capable of evaluating a wide range of colorimetric parameters, this investigation specifically focused on the CIE Lab* and LCh* color spaces. The CIE Lab* color space is designed to provide an objective representation of colors based on human perception, divided into three dimensions: L* (lightness), a* (green to red variation), and b* (blue to yellow variation). On the other hand, the LCh* color space is derived from CIE Lab* through a polar transformation, retaining L* for lightness, while C* denotes chroma, and h* indicates the hue angle. The values for L*, a*, and b* were directly obtained by the GroundEye® software, while the values for C* and h* were calculated from the available data. The formulas for calculating chroma (C*) and hue angle (h*) follow the definitions established by Robertson (1977), described as follows.

Chroma (C*) represents the intensity or purity of color and is calculated as the radial distance in the ab plane, according to Equation 1.

$$Chroma = \sqrt{a^2 + b^2} \quad (1)$$

The hue angle (h*) corresponds to the angle in the ab plane that describes the color in terms of hue. To calculate the hue angle, Equation 2 is used. The function atan2 is the arc tangent of two arguments, returning the angle in the ab plane between the line passing through the coordinates (a*, b*) and the a* axis. The hue angle is expressed in degrees, with values ranging from 0° to 360° (2).

$$Hue\ angle = \tan^{-1} \frac{b^*}{a^*} \quad (2)$$

Figure 2 illustrates these representations and facilitates the understanding of the relationship between the components of each space, offering a view of the color distribution in the perceptible human spectrum.

Validation model

A dataset containing 12,000 images of soybean seeds was compiled, with the images equally divided into two categories: “greenish seed” and “yellow seed.” Using the GroundEye® software, chromatic information pertaining to lightness (L*), green to red variation (a*), and blue to yellow variation (b*) was extracted from each image. Additionally, the values for chroma (C*) and hue angle (h*), which were calculated mathematically, were incorporated as analytical variables.

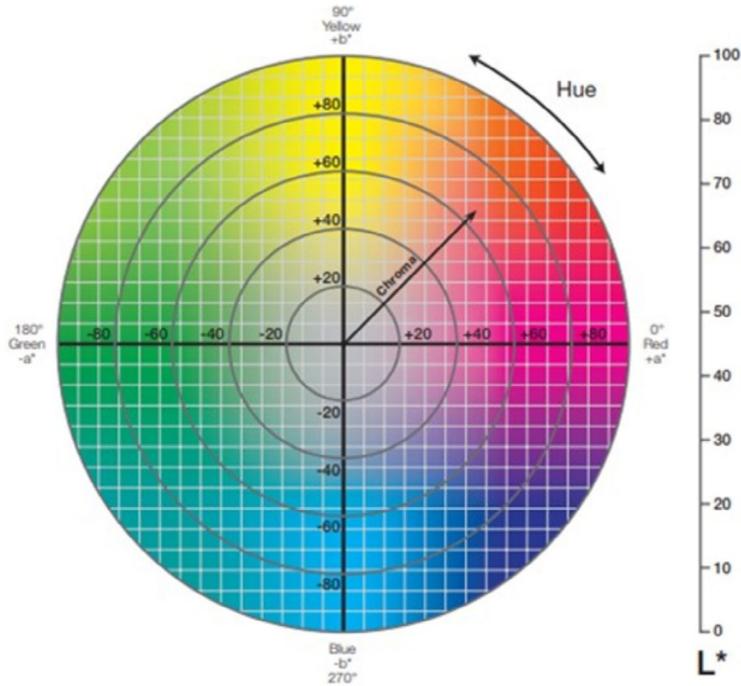


Figure 2: Theoretical CIE Lab* color space (Hunter, 1958).

A decision tree was adopted for the classification of the seeds. This type of predictive model segments the dataset into increasingly smaller subsets while constructing a decision-making model at each node. This method was chosen for its ability to formulate clear binary questions—crucial for distinguishing between the “greenish seed” and “yellow seed” categories based on visual characteristics captured in the images.

The implementation was carried out using the sklearn. tree library from Python, provided by the Python Software Foundation (2021). The DecisionTreeClassifier model was employed to develop a decision tree that learns from the chromatic characteristics extracted from the images (L^* , a^* , b^* , C^* , and h^*). This tree was trained to differentiate between greenish and yellow seeds, using these attributes to perform accurate classifications based on the visual distinctions between the two seed categories. This process involved training the model on the dataset, divided into 66% (7,919 images) for training and 34% (3,041 images) for testing, allowing the model to learn and then validate its ability to generalize on unseen data.

To evaluate the efficacy of the model, a confusion matrix was used, which helped visualize and quantify the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These rates were used to calculate the performance indicators of the model obtained by the decision tree, namely: accuracy, precision, sensitivity, and specificity. Accuracy refers to the proportion of predictions made by the model that proved correct, Equation 3. Precision indicates the proportion of positive identifications made by the model that were indeed

correct, Equation 4. Sensitivity represents the proportion of real positives that the model correctly identified, Equation 5. Specificity corresponds to the proportion of real negatives that the model correctly identified, Equation 6.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3)$$

$$Precision = \frac{TP}{(TP + FP)} \times 100 \quad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (6)$$

Additionally, the Gini index was used as a criterion for feature selection in constructing the decision tree, providing a measure of impurity that optimized the partitioning of categories across the nodes. This index helps to determine how each feature contributes to the homogeneity of the nodes and leaves in the tree, aiming to minimize the probability of misclassification. Using the Gini index in this context enhances the tree’s ability to clearly distinguish between the different categories based on the most relevant features extracted from the seed images.

Germination analysis

The 6,000 seeds previously selected by the analyst, comprising 3,000 greenish and 3,000 yellow seeds, were individually subjected to a germination test, seed by seed. Each seed was placed on a moistened germitest paper, with the amount of water equivalent to 2.5 times the mass of the dry substrate. In each paper roll, 25 seeds were evenly distributed, and each was given a unique identifier to ensure accurate traceability. After sowing, the paper rolls were placed in a germinator with a controlled temperature of 25 ± 2 °C. On the fifth day after sowing, the seedlings were evaluated and classified as normal, abnormal, and dead, according to the criteria established by the Rules for Seed Analysis (Brasil, 2009). A single evaluation was conducted, given that the same seedling was also subjected to a vigor test through the GroundEye® subsequently. Moreover, in seed laboratories, it is common practice to perform the germination reading in a single analysis on the fifth day.

According to the Rules for Seed Analysis, normal seedlings are those that have all essential structures — both root systems and aerial parts — fully developed, demonstrating potential to continue their development and produce normal plants under favorable conditions. In contrast, abnormal seedlings are those that, even under ideal conditions, do not show potential to continue development and produce normal plants. Dead seeds, on the other hand, are those that show no signs of germination initiation at the end of the test; these often appear softened and attacked by microorganisms (Brasil, 2009).

The data related to germination were correlated with the hue angle to evaluate the relationship between these two parameters. For this, the results were considered independently of their classification category (yellow or greenish). This analysis aimed to identify possible patterns or trends indicating how variations in hue angle might influence the germination rate of seeds.

Vigor via GroundEye®

Seed vigor refers to their ability to germinate and establish quickly under adverse cultivation conditions. Only normal seeds, which demonstrate appropriate structural development, possess the necessary potential to express vigor and establish effectively in the field. Thus, in this research, vigor was measured only in normal seedlings, hence abnormal seedlings and dead seeds were disregarded, reducing the sample size compared to the 6,000 analyses performed for the germination test.

Immediately after the individual evaluation of seedlings in the germination test, all those classified as normal were photographed with the assistance of the GroundEye® system for vigor analysis. A total of 2,880 normal seedlings from the “yellow seed” group and another 2,070 normal seedlings from the “greenish seed” group were analyzed. Figure 3 shows an image representation of a seedling processed via the GroundEye® software.



Figure 3: Image of normal soybean seedling captured by GroundEye® S800D.

For the photographic record, each seedling was positioned on the equipment tray, allowing for image acquisition and the measurement of the primary root and hypocotyl lengths. The total length of the seedling was determined by adding the size of the primary root to that of the hypocotyl, and the results were presented in centimeters. Barros et al. (2021) validated the effectiveness of the automated measurement of seedling and root lengths through the GroundEye® system, equating it with manual measurement.

Statistical analysis

In the study conducted to evaluate germination, each of the 6,000 seeds, divided between greenish and yellow, was considered a single sample unit. Subsequently, in the vigor analysis, each normal seedling was also treated as a unique sample unit, totaling 4,950 normal seedlings, divided into the same two treatments as the germination.

Statistical analysis was conducted through an analysis of variance (ANOVA) to identify significant differences between the treatments. After this, the Tukey test was applied for multiple comparisons between the group means, using a significance level of 5%. The statistical software used was Assisat, version 7.7 (Silva & Azevedo, 2016).

Results and Discussion

New solutions in computer vision, combined with artificial intelligence algorithms, show great potential to revolutionize the analysis of biological images. They have the ability to minimize

subjectivity and optimize the analysis process, resulting in more efficient and precise quality control in agriculture. With these technologies, it is possible to discern patterns in biological images that may be difficult to distinguish with the naked eye, allowing for a more objective and accurate evaluation of seed quality (Medeiros et al., 2020).

In this context, this study demonstrated that the chromatic data obtained through image analysis and applied in machine learning modeling enabled effective classification between seed classes. Moreover, the findings of this research highlight that the physiological quality of greenish seeds is inferior to that of yellow seeds. Uniquely, the relationship between the hue angle and the viability of soybean seeds was explored, establishing it as a significant predictor for this purpose.

Machine learning model: Decision tree

In the context of computer vision, the Lab* color space is often preferred because its composition is designed to approximate the way humans perceive colors. This means that differences in the Lab* color space correlate more directly and intuitively with the differences that humans actually perceive, compared to other color spaces, such as the usual RGB (Kanan & Cottrell, 2012). Lin et al. (2019), agreeing with Kanan and Cottrell (2012), argued that the visual perception of the components (red, green, and blue) of the RGB image of soybeans is extremely close. However, when the color space conversion from RGB to Lab* is performed, the components of the Lab* color space display a significant visual difference between the three distinct color channels, allowing for the formation of well-defined features.

These findings were confirmed in this study, where variables from the CIE Lab* color space were superior in the task of

segregating greenish seeds from yellow ones. The hue angle (h) and lightness (L) were the most responsive discriminative criteria to the learning model (Figure 4). The model results indicated that a hue angle of less than or equal to 97.885° had a strong association with the assessment of seeds as “yellow”. On the other hand, hue angle values above this threshold were predominantly associated with the classification of seeds as “greenish”.

Furthermore, as shown in Figure 4, the exclusive use of the hue angle (h) was not sufficient to effectively differentiate the two classes of seeds. However, by incorporating lightness (L) as a second attribute, a significant improvement in classification was observed, evidenced by the reduction in the Gini index. “Nodes” with reduced Gini indices in decision trees generally perform better in classifying new data, as they group data subsets more homogeneously (Hastie et al., 2009; James et al., 2023).

Table 1 presents a confusion matrix, where the samples were divided into two categories: yellow and greenish. In this study, 1870 samples were correctly classified as yellow, 1791 samples as greenish, 170 yellow samples were mistakenly classified as greenish, and 250 greenish samples were mistakenly classified as yellow. This demonstrates the effectiveness of integrating lightness (L) with hue angle (h) in improving the accuracy of seed classification.

Table 1: Confusion matrix for soybean seed classification as yellow or greenish.

	n = 4081 yellow	Real class	
		yellow	greenish
Predicted class	yellow	1870	170
	greenish	250	1791

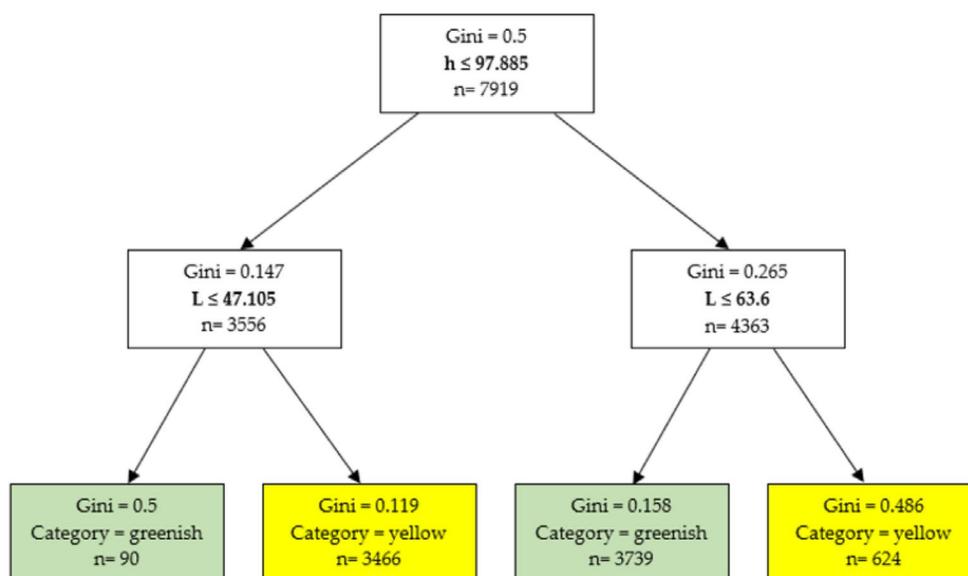


Figure 4: Decision tree for classification of soybean seeds as yellow or greenish, according to hue angle (h) and luminosity (L).

Based on the metrics presented in Table 2, it can be inferred that the results obtained from the decision tree model indicate good performance in distinguishing between soybean seeds classified as “yellow” and “greenish.” The analyzed model demonstrated a sensitivity of 88.2%, indicating that it was able to correctly identify 88.2% of the true cases of yellow seeds. The specificity achieved was 91.3%, reflecting the model’s ability to accurately classify the samples that are not yellow seeds. Moreover, the precision for predictions of the yellow seed class was highlighted at 91.7%. The overall accuracy of the model was 89.8%, showcasing the accuracy in predictions in most cases. This performance suggests that the decision tree model, enhanced by the use of both hue angle (h) and lightness (L) as features, effectively supports robust classification within the context of seed quality analysis.

Table 2: Performance metrics for the decision tree.

Sensibility	Specificity	Precision	Accuracy
----- % -----			
88.2	91.3	91.7	89.8

The efficacy in distinguishing between the classes of yellow and greenish seeds, evidenced by high rates of sensitivity, specificity, precision, and accuracy, underscores the efficiency of the classification model developed in this study. The applicability of this model is supported by other researchers who have also employed machine learning in classifying soybean seeds and seedlings with varying levels of assertiveness (Momin et al., 2017; Lin et al., 2019; Andrade et al., 2024). The varied degrees of success in AI research indicate that validation models need continuous refinement to achieve higher accuracy. This implies that they must be fed an increasing amount of data for enhancement, thereby improving their capacity for generalization and accuracy in predictions.

Thilakarathne et al. (2018) analyzed black cumin (*Nigella sativa*) seeds from different origins and observed a significant change in lightness values after decortication. Although the cultivars were visually indistinguishable, color analysis contributed to a clear separation between them. Park et al. (2023) concluded that integrating near-infrared spectroscopy (SWIR) with machine learning-based classification techniques proves effective in detecting specific artificial adulterations in red pepper (*Capsicum annum*).

The relevance of these previous studies is emphasized by the current investigation, which highlights how combining hue angle and lightness in the classification of greenish seeds introduces valuable objectivity to the process. These parameters allow for a quantitative and unbiased assessment of color, essential for mitigating the subjectivity inherent in human perception. Therefore, the simultaneous use of these two measures enables an objective

methodology for seed classification, enhancing the accuracy and reproducibility of the classification process, and demonstrating its potential applicability in broader contexts beyond soy.

Physiological quality

Seeds classified as yellow exhibited a higher average germination rate of 96%, and also produced seedlings with greater initial development (Table 3) compared to the greenish ones. This suggests that yellow seeds not only germinate better but also produce more vigorous seedlings, which can improve crop performance during the early stages. The greenish color is somehow related to a decrease in the physiological quality of soybean seeds.

Table 3: Germination and seedling growth of yellow and greenish soybean seeds.

Color	Germination (%)	Primary root (cm)	Hypocotyl (cm)	Seedling length (cm)
Yellow	96 a	13.91 a	7.88 a	21.72 a
Greenish	69 b	10.95 b	6.87 b	17.82 b

Means followed by the same letter in each column do not differ according to the Tukey test ($p \leq 0.05$).

The results obtained in this study align with the existing literature, where it was observed that greenish coloring in seeds can negatively impact physiological quality, reducing both germination and seed vigor. For example, Arruda et al. (2016) and Teixeira et al. (2020) had already identified that the presence of greenish seeds could impair the physiological quality of soybean seeds depending on their incidence rate in lots.

This study complements and expands previous conclusions by incorporating the traditional germination test and introducing image analysis as an additional tool for evaluating seed vigor. Furthermore, the investigation into the influence of the hue angle on the assessment of seed viability introduced a new and relevant approach to the theme of seed quality prediction, and therefore, deserves to be rigorously deepened through further scientific research.

The hue angle, identified in this study as an effective discriminative criterion in the task of classifying greenish seeds, also proved to be an important predictor of physiological quality. Regardless of whether the seeds are yellow or greenish, they achieve their highest germination rates when the average hue angle is up to 97.5° (Figure 5). Within this range, an average germination rate of 95% was observed, compared to a germination rate of 71% when the hue angle exceeded 97.5°.

It is also observed through the linear regression equation that as the hue angle increases, the viability of soybean seeds tends to decrease (Figure 5). The coefficient of determination suggests that about 73% of the variation in soybean seed viability can

be explained by the variation in the hue angle. This indicates a considerable relationship between these two variables, suggesting that the hue angle may be a good indicator of soybean seed viability. The limitations of our study include the need to explore other variables that may influence seed viability, which underscores the importance of further studies and the adoption of multivariate analyses.

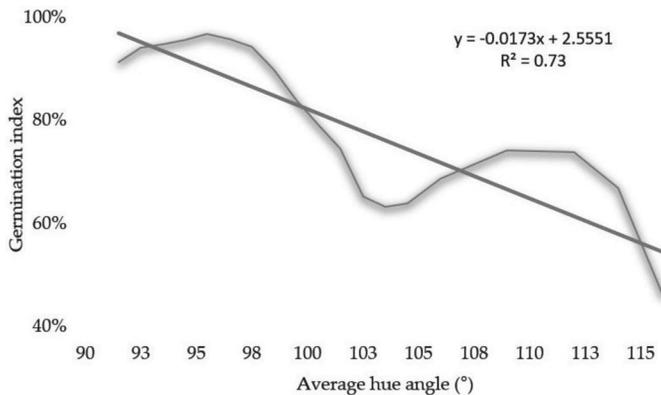


Figure 5: Germination index by average hue angle.

The hue angle, for both yellow and greenish seeds, is situated within the same spectrum of chromatic variation, indicating that both colors are close variants within a continuous spectrum of colors. In the chromaticity circle, the color yellow is positioned at an angle of 90°, as illustrated in Figure 2. This positioning suggests that the distinction between yellow and greenish seeds can be subtle, based on small variations in hue. This proximity in the color spectrum, in turn, may contribute to subjectivity in analysis, making the task of distinguishing the seeds prone to error, especially when conducted manually or by untrained evaluators. Therefore, selecting appropriate tools and techniques to discern these nuances is essential to ensure accurate classification and minimize ambiguity.

Various studies have demonstrated that image analysis is an effective tool for assessing the physical characteristics of seeds from various crops, providing detailed characterization that includes aspects such as color, geometry, texture, and size. For example, Lima et al. (2018) worked with seeds of paricarana (*Bowdichia virgilioides* Kunth), characterizing them according to variations in seed coat coloring. Similarly, Xavier et al. (2019) differentiated species of *Amaranthus* spp., based on geometric characteristics and color predominance. These findings corroborate the discoveries of this work, highlighting the broad field of application that image analysis offers for quality control assessments, as a non-destructive, objective, precise, and rapid technique.

In recent years, there has been a significant advancement in the resources available to explore and validate mathematical models in various research areas. Despite this technological evolution, the application of these techniques in seed analysis,

especially using colors as predictors of quality, is still notably underrepresented in the scientific literature. The proposal to incorporate chromatic characteristics in the evaluation of seed quality represents an innovative methodology, which stands out not only for its practical applicability but also for its ability to amplify results when combined with emerging technologies such as AI and machine learning.

This new paradigm promises not only to improve the accuracy of seed analysis but also to make the process more efficient. However, it is important to note that these new technologies do not completely eliminate the need for human intervention. In fact, they can serve as a valuable tool to assist experts in decision-making, complementing, but not replacing, human knowledge and experience.

Conclusions

Yellow soybean seeds exhibit superior physiological quality compared to greenish ones, particularly in terms of germination and seedling growth. The hue angle (h) and luminosity (L) proved to be the most responsive criteria in the machine learning model, achieving an accuracy of 89.7%. The hue angle was demonstrated to be a robust predictor, correlating with higher germination rates in seeds with an angle less than 97.5°. The relationship between seed viability and hue angle was supported by a coefficient of determination (R^2) of 73%.

Author Contribution

Conceptual Idea: Oliveira, V. R.; Methodology design: Oliveira, V. R.; Souza, N. O. S.; Arantes, M. S.; Data collection: Oliveira, V. R.; Data analysis and interpretation: Oliveira, V. R.; Souza, N. O. S.; Carmona, R; Arantes, M. S.; and Writing and editing: Oliveira, V. R.; Souza, N. O. S.; Carmona, R.

References

- Andrade, D. B. et al. (2024). Artificial intelligence tools and a diagrammatic scale for evaluating the quality of coating in treated soybean seeds. *Neural Computing and Applications*, 36(6):3101-3106.
- Arruda, M. H. M. et al. (2016). Qualidade fisiológica de lotes de sementes de soja com diferentes percentuais de sementes esverdeadas. *Magistra*, 28(2):194-200.
- Barros, B. E. et al. (2021). Image analysis for the evaluation of soybean seeds vigor. *Acta Agronômica*, 70(3):311-316.
- Bauriegel, E. et al. (2011). Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Computers and Electronics in Agriculture*, 75(2):304-312.

- Bishop, C. M. (2006). *Pattern recognition and machine learning (Information science and statistics)*. New York, United States: Springer-Verlag, 738p.
- Brasil. (2009). Ministério da Agricultura, Pecuária e Abastecimento. *Regras para análise de sementes*. Departamento Nacional de Produção Vegetal: Brasília, DF, Brasil, 399p.
- Food and Agriculture Organization - FAO. (2021). *World food and agriculture - Statistical yearbook 2021*. Italy, Roma, 368p.
- Ferrari, J. M. et al. (2023). Comportamento da qualidade fisiológica de sementes de soja esverdeadas durante o armazenamento. *Revista Caribeña de Ciências Sociales*, 12(7):3055-3060.
- França-Neto, J. B. et al. (2016). *Tecnologia da produção de semente de soja de alta qualidade*. Londrina: Embrapa Soja, 82p. (Embrapa Soja. Documentos, 380).
- Godoy, C. V. et al. (2017). *Boas práticas para o enfrentamento da ferrugem-asiática da soja*. Londrina: Embrapa Soja, 5 p. (Embrapa Soja. Comunicado técnico, 92).
- Hastie, T. et al. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: springer, 745p.
- Hunter, R. S. (1958). Photoelectric color difference meter. *Josa*, 48(12):985-995.
- James, G. et al. (2023). *An introduction to statistical learning: With applications in python*. New York, NY: Springer Nature. 75p.
- Kanan, C., & Cottrell, G. W. (2012). Color-to-grayscale: does the method matter in image recognition?. *PLoS One*, 7(1):e29740.
- Lima, J. M. E. et al. (2018). Técnicas de análise de imagem para caracterização da qualidade de sementes de paricarana (*Bowdichia virgilioides* Kunth). *Ciência Florestal*, 28(3):1202-1216.
- Lin, P. et al. (2019). Rapidly and exactly determining postharvest dry soybean seed quality based on machine vision technology. *Scientific Reports*, 9(1):17143.
- Mahajan, S. et al. (2018). Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (*Glycine max*). *Journal of food science and technology*, 55(10):3949-3959.
- Medeiros, A. D. et al. (2020). Interactive machine learning for soybean seed and seedling quality classification. *Scientific Reports*, 10(1):11267.
- Momin, M. A. et al. (2017). Machine vision based soybean quality evaluation. *Computers and Electronics in Agriculture*, 140:452-460.
- Pádua, G. P. et al. (2007). Tolerance level of green seed in soybean seed lots after storage. *Revista Brasileira de Sementes*, 29(3):128-138.
- Park, J. et al. (2023). Detection of red pepper powder adulteration with allura red and red pepper seeds using hyperspectral imaging. *Foods*, 12(18):3471.
- Patel, K. K. et al. (2012). Machine vision system: A tool for quality inspection of food and agricultural products. *Journal of food Science and Technology*, 49:123-141.
- Python Software Foundation. (2021). *Python*. Versão 3.10 [s.l.]. Available in: <<https://www.python.org/>>.
- Robertson, A. R. (1977). The CIE 1976 color-difference formulae. *Color Research & Application*, 2(1):7-11.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson.
- Silva, F. A. S., & Azevedo, C. A. V. (2016). The assistat software version 7.7 and its use in the analysis of experimental data. *African Journal of Agricultural Research*, 11(39):3733-3740.
- Singh, G., & Shivakumar, B. G. (2010). The role of soybean in agriculture. In G. Singh. *The soybean: Botany, production and uses*. CAB International, Oxfordshire, UK, (pp. 24-47).
- Škrubej, U. et al. (2015). Assessment of germination rate of the tomato seeds using image processing and machine learning. *European Journal of Horticultural Science*, 80(2):68-75.
- Teixeira, S. B. et al. (2020). Green soybean seeds: Effect on physiological quality. *Ciência Rural*, 50(2):e2018063.
- Thakur, P. S. et al. (2022). Deep transfer learning based photonics sensor for assessment of seed-quality. *Computers and Electronics in Agriculture*, 196:106891.
- Thilakarathne, R. C. N. et al. (2018). Morphological characteristics of black cumin (*Nigella sativa*) seeds. *Chemistry Research Journal*, 3(3):40-45.
- Xavier, J. B. et al. (2019). Caracterização morfológica, química e fisiológica de sementes de *Amaranthus* spp. *Journal of Seed Science*, 41(4):478-487.
- Xia, Y. et al. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, 1:35-47.
- Zorato, M. F. et al. (2007). Presença de sementes esverdeadas em soja e seus efeitos sobre seu potencial fisiológico. *Revista Brasileira de Sementes*, 29(1):11-19.