e-ISSN 1983-4063 - *www.agro.ufg.br/pat* - Pesq. Agropec. Trop., Goiânia, v. 55, e82842, 2025

Special Supplement: Pulses

Computer vision and supervised machine learning techniques for classifying seeds of chickpea cultivars[1]

José Victor Maurício de Jesus[2], Ivan David Briceño-Pinzon[2], Giulyana Isabele Silva Tavares[2], Lívia Karine Pereira[2], Daniel Lima da Silva[2], Heloisa Oliveira dos Santos[2], Raquel Maria de Oliveira Pires[2]

ABSTRACT

The morphological similarity among chickpea (*Cicer arietinum*) cultivars makes their correct identification challenging, compromising the varietal purity of the seeds. This study aimed to evaluate models for the classification of chickpea varieties using computer vision and supervised machine learning, analyzing the morphometric attributes of chickpea seeds extracted from digital images. In total, 21 color, shape and size attributes were determined from digital seed images of nine chickpea cultivars, corresponding to five kabuli and four desi cultivars. For the varietal classification, supervised learning models including Support Vector Machine, Multilayer Perceptron, Random Forest and k-Nearest Neighbors were employed. The evaluation of the models was performed through stratified k-fold cross-validation to determine the metrics performance for each model. The models with the best performances were Support Vector Machine and Random Forest, which showed high accuracy (95.37 and 94.26 %) and discriminatory capacity, according to the Matthews correlation coefficient (95.10 and 94.24 %), being considered adequate methods for the varietal differentiation of chickpea seeds.

KEYWORDS: *Cicer arietinum*, image analysis in seeds, cultivar identification, digital phenotyping.

RESUMO

Visão computacional e técnicas de aprendizado de máquina supervisionado na classificação de sementes de cultivares de grão-de-bico

A semelhança morfológica entre cultivares de grão-de-bico (*Cicer arietinum*) dificulta sua correta identificação, comprometendo a pureza varietal das sementes. Objetivou-se avaliar modelos para a classificação de variedades de grão-de-bico pelo uso de visão computacional e aprendizado de máquina supervisionado, analisando-se atributos morfométricos de sementes de grão-de-bico extraídos de imagens digitais. No total, 21 atributos de cor, forma e tamanho foram determinados a partir de imagens digitais de sementes de nove cultivares de grão-de-bico, correspondentes a cinco cultivares do tipo kabuli e quatro do tipo desi. Para a classificação varietal, foram utilizados os modelos de aprendizado supervisionado Support Vector Machine, Multilayer Perceptron, Random Forest e k-Nearest Neighbors. A avaliação dos modelos foi realizada por meio de validação cruzada k-fold estratificada, para determinar as métricas de desempenho para cada modelo. Os modelos com os melhores desempenhos foram Support Vector Machine e Random Forest, os quais apresentaram alta acurácia (95,37 e 94,26 %) e capacidade discriminatória, segundo o coeficiente de correlação de Matthews (95,10 e 94,24 %), podendo ser considerados métodos adequados na diferenciação varietal de sementes de grão-de-bico.

PALAVRAS-CHAVE: *Cicer arietinum*, análise de imagens em sementes, identificação de cultivares, fenotipagem digital.

INTRODUCTION

Chickpea (*Cicer arietinum*) is one of the most important food grain species in the world, ranking third among the most produced legumes (FAO 2025). Besides its agronomic significance, the crop is recognized for its high nutritional value, with high levels of proteins, fibers, carbohydrates and essential minerals such as iron, zinc and magnesium (Kumar et al. 2025).

The high genetic diversity of chickpea is widely recognized, being reported by the extensive collections conserved in germplasm banks worldwide (Upadhyaya et al. 2008, Piergiovanni 2022). Based on this diversity, it is possible to classify the species into two main groups, kabuli and desi, which

[1] Received: May 30, 2025. Accepted: July 24, 2025. Published: Sep. 04, 2025. DOI: 10.1590/1983-40632025v5582842.

[2] Universidade Federal de Lavras, Escola de Ciências Agrárias de Lavras, Departamento de Agricultura, Setor de Sementes, Lavras, MG, Brazil. *E-mail/ORCID*: vic.tor123mauricio@gmail.com/0000-0003-2458-0866; ivan.pinzon2@estudante.ufla.br/0000-0002-5896-3366; giulyanatavares@gmail.com/0000-0002-3709-4413; liviakarinep@gmail.com/0000-0002-3305-3243; ldaniel166@gmail.com/0009-0009-3039-3478; heloisa.osantos@ufla.br/0000-0003-1384-4969; raquelmopires@ufla.br/0000-0003-1369-4323.

Editor: Luis Carlos Cunha Junior

differ in terms of size, color, texture and chemical composition of the seeds (Gaur et al. 2010).

As with many crops, differentiating among chickpea cultivars presents significant challenges, especially when using seed morphology to identify materials (Behera et al. 2023). This complexity becomes even more evident because the genetic basis is inherently narrow among cultivars belonging to the same group (kabuli or desi), whose seeds exhibit considerable similarity in attributes including shape, color and size (Singh et al. 2022).

Considering the aforementioned challenges, image analysis is a promising tool and has been consolidated as an important strategy for characterization, classification and identification of cultivars. Studies carried out demonstrate the potential of image analysis through computer vision and classification models in varietal differentiation based on morphometric characteristics in soybean seeds (Çetin 2022), common bean (Koklu & Ozkan 2020) and maize (Ali et al. 2020).

One of the primary challenges in the seed sector is the varietal distinction of the same species based on cultivars and the process of registration and certification of seeds through morphological descriptors (Ghaffari 2024). The use of machine learning offers a promising solution, allowing the identification and separation of mixed cultivars based on the extraction of characteristics through image analysis techniques, combined with advanced supervised learning algorithms, as an alternative in the fast, accurate, efficient and assertive identification and classification of seeds (Koklu et al. 2021).

Machine learning approaches for chickpea cultivar classification have been successfully implemented across multiple countries, including Turkey (Çetin et al. 2023, Varol et al. 2023, Kiliç & Yalçin 2025, Ulu et al. 2025), Pakistan (Ajaz & Hussain 2015), Iran (Pourdabani et al. 2019), India (Priyadarshi et al. 2023) and Ethiopia (Ayele & Tamiru 2020). These classification models offer substantial value for breeding programs and seed quality assessment (Çetin et al. 2023), particularly through their ability to standardize evaluation processes.

For chickpea, there is a limitation in studies that employ supervised learning methods and algorithms, with the restricted use of cultivars in the analyses and a predominance of materials belonging to the same genetic group, which compromise the representativeness and generalizability of the models (Ghamari 2012, Pourdarbani et al. 2019). Therefore, a more comprehensive and in-depth approach is needed to integrate a greater number of chickpea cultivars from different groups, in addition to using color, size and shape characteristics in the creation of classification models, which allow applicability in seed quality. Furthermore, it is noteworthy that there are no reports of the application of these models for chickpea cultivars duly registered in Brazil.

Thus, this study aimed to evaluate the integration of computer vision techniques used for the extraction of characteristics such as color, shape and size from digital images of seeds of chickpea cultivars with traditional machine learning algorithms, in order to analyze the efficiency of models for seed classification.

MATERIAL AND METHODS

The experiment was conducted at the Universidade Federal de Lavras, in Lavras, Minas Gerais state, Brazil, between April and May 2025.

Nine chickpea cultivars from the Empresa Brasileira de Pesquisa Agropecuária (Embrapa Hortaliças) were used (Figure 1), being produced under the same edaphoclimatic conditions in the experimental area. The cultivars BRS Kalifa, BRS

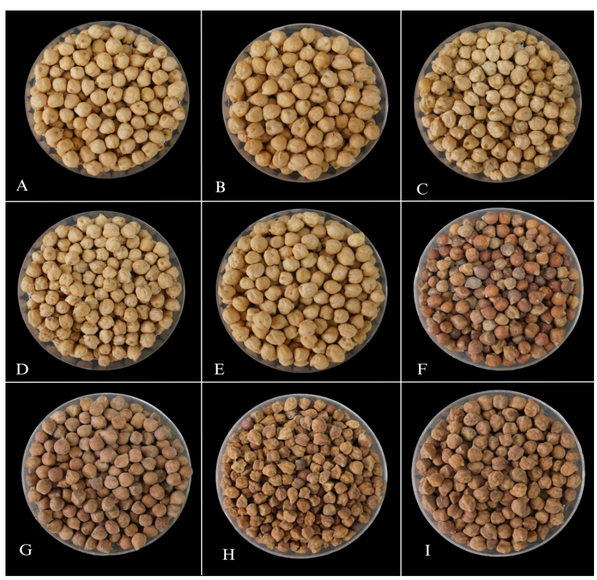

Figure 1. Chickpea cultivars used in the present study. Kabuli group: Aleppo (A); BRS Cícero (B); BRS Cristalino (C); BRS Kalifa (D); BRS Toro (E). Desi group: 10108 (F); 10209 (G); Hari (H); UPL-06 (I).

Cícero, BRS Toro, BRS Aleppo and BRS Cristalino correspond to seeds of the kabuli type, characterized by being whiter in color and larger in size, whereas the cultivars BRS Hari, 10108, 10209, Hari and UPL-06 of the desi type have dark integument and smaller size.

Images (.png format) of the seeds of each cultivar were acquired using a GroundEye S800® equipment, with spatial resolution of 5854 x 3884 pixels and color depth of 24 bits. This device has a closed capture module integrated by a high-resolution digital camera, a sliding opening structure integrated into the transparent acrylic tray to arrange the samples to be captured, and a lighting system that activates automatically at the time of capture (TBIT 2022).

The imaging dataset comprised a total of 4,301 seeds from nine distinct chickpea cultivars. Eight of these cultivars (Aleppo, Cristalino, Hari, Kalifa, Toro, UPL-06, 10108 and 10209) were represented by 500 seeds each. The ninth cultivar, Cícero, was represented by 301 seeds due to limited seed availability. In total, 28 individual images were acquired to capture this entire sample.

Each seed was analyzed individually, with the segmentation process carried out to obtain the region of interest, that is, the seed, whose objective was to obtain a suitable mask for each sample, in order to individualize each seed and ensure a noise-free background. For this, the color image in the RGB (Red, Green and Blue) color scale (Figure 2A) was transformed into the color space L*a*b, with L corresponding to luminosity, and a* and b* to the chromatic coordinates for red/green and yellow/blue, respectively. The channel b was chosen because it offers a better contrast between the background and the seeds (Figure 2B). Image thresholding was performed using the Otsu method (Otsu 1979) on the b* channel image, to obtain a binary mask, where the white color corresponded to the region of interest, the seeds, and the black color to the background (Figure 2C). With the mask, a new color image was generated (Figure 2D) with a homogeneous background and highlighting the regions of interest in the original RGB image (Silva et al. 2024). This process aimed to facilitate the extraction of color, shape and size characteristics.

From the RGB images obtained from the previous process, four color characteristics were extracted for each RGB color channel: minimum, maximum, average and variance, totalling twelve characteristics for color. In addition, ten size and shape characteristics (Xu et al. 2021) were obtained for each chickpea seed. The extracted parameters for shape and size are described in Table 1.

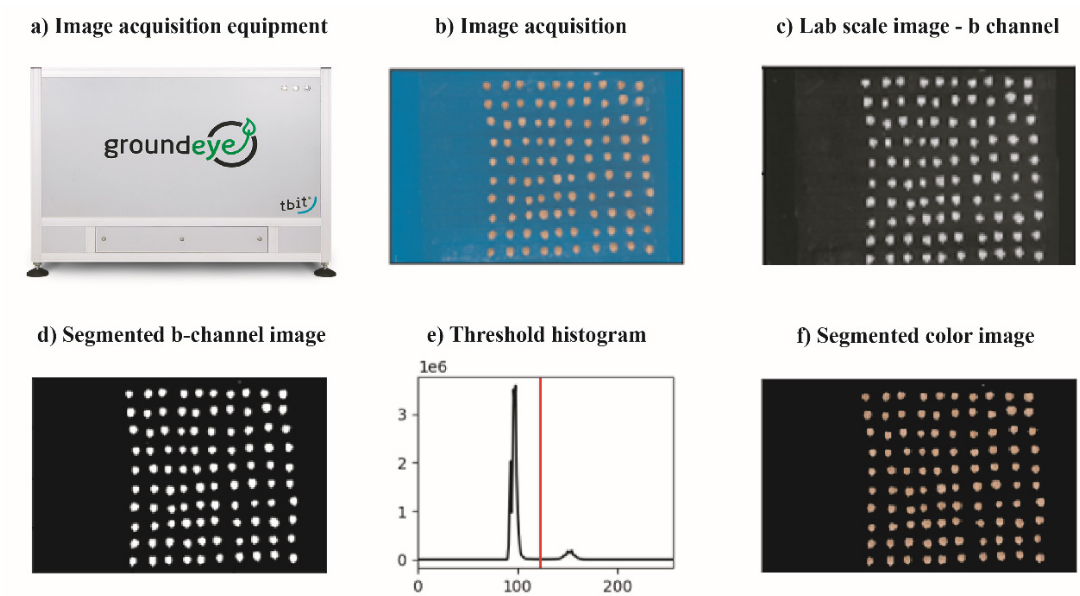

Figure 2. Stages of color image processing and segmentation. A) equipment used for image acquisition; B) original color image obtained by the equipment; C) image converted to channel b of the CIELAB color space, showing the contrast between the background and the seeds; D) image in binary pattern, resulting in the segmentation of the seeds (in white) on a black background; E) histogram of channel b, with a red line indicating the threshold value adopted for segmentation; F) final segmented image, preserving the original color of the seeds and with the background removed.

For the machine learning model development, the dataset was used including a total of 4,301 seeds with 22 attributes, being ten attributes of shape and size and twelve color parameters, implemented to create the classification models: Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forest (RF) and k-Nearest Neighbors (k-NN), whose performance was evaluated by cross-validation of k-fold stratified with k = 5. This procedure was adopted to ensure a good generalization capacity of the models, even when the number of seeds per cultivar was not the same in all the evaluated cultivars. For this purpose, the Scikit-learn library was used (Pederegosa et al. 2011).

All calculations were developed with the help of the Python 3 programming language. The inputs and selection of hyperparameters for each model were based on a systematic manual approach, where different combinations were tested iteratively and evaluated by cross-validation, using, for this, the Python library Scikit-learn. Table 2 presents the information on the selected hyperparameters.

The machine learning models for chickpea variety classification were evaluated according to the number of accurate estimates obtained in all evaluated forecasts. The matrix is used to describe the performance of each classification model, and is made up of four parameters: true positives, true negatives, false positives and false negatives. Table 3 shows the confusion matrix for multiclass assessments (n x n) (Koklu & Ozkan 2020).

Table 3. Confusion matrix for multiclass assessments.

True class	Predicted class				
	C_1	C_2	C_3	...	C_n
C_1	Vp_1	$F_{1,2}$	$F_{1,3}$	...	$F_{1,n}$
C_2	$F_{2,1}$	Vp_2	$F_{2,3}$	...	$F_{2,n}$
C_3	$F_{3,1}$	$F_{3,2}$	Vp_3	...	$F_{3,n}$
...	...	...	...	...	...
C_n	$F_{n,1}$	$F_{n,2}$	$F_{n,3}$	...	Vp_n

Table 1. Attributes of shape and size extracted from chickpea seeds.

Atributte	Definition	Type
Area (A)	Total number of pixels within the region of interest that was targeted	Size
Convex area (CA)	Area that completely encloses all points of the general object	Size
Perimeter (P)	Total number of pixels counted at the boundary of the seed outline	Size
Length of the longest shaft (L)	Length of the row segments connecting the two points furthest from the seeds	Size
Length of the shortest shaft (I)	Shortest linear distance that crosses a seed, perpendicular to the major axis	Size
Equivalent diameter (Ed)	Diameter of a circle with the same area as the image of a seed	Size
Aspect ratio (Ar)	Relationship between the major axis and the minor axis. It describes how elongated or rounded the seeds are (Air = L/I)	Form
Circularity (Ci)	It shows how much the seed has a shape to a perfect circle with values close to 1.0 or is more elongated with values closer to 0.0	Form
Compactness (Cp)	Provides the area of the chickpea seed relative to the area of the circle with the same circumference	Form
Solidity (S)	Defined by the ratio of the object's area to its convex area. Indicates a more solid shape of the seed with values close to 1.0 and jagged edges with values close to 0.0	Form

Table 2. Detailed information about the hyperparameters of each supervised learning classification model.

Model	Hyperparameters
SVM	The regularization parameter was C = 1.0. The linear hyperplane function was used for data transformation, and the gamma value was '(n_features * X.var())'
MLP	Scaled dataset (z = x - μ/σ); MLP network structure: 22-10-9; 22 layers of inputs, 10 hidden layers and 9 layers of outputs. Activation function: hyperbolic tangent; learning rate: 0.01; momentum: 0.3; maximum number of iterations: 1,000
RF	100 decision trees were used, with the criterion for dividing the nodes being entropy, applying balanced weighting for all the attributes available in the divisions (max_features = none)
k-NN	Scaled dataset (z = x - μ/σ); number of neighbors: 5; Minkowski's distance metric with parameter p = 2, which is equivalent to the Euclidean distance.

SVM: Support Vector Machine; MLP: Multilayer Perceptron; RF: Random Forest; k-NN: k-Nearest Neighbors.

The performance of each used model was determined by the performance measures of accuracy, error rate, precision, recall, specificity, F1-score and the Matthews Correlation Coefficient (MCC) for multiclasses. The calculation and explanation of each measurement are described in Table 4.

RESULTS AND DISCUSSION

Through the analysis of the confusion matrix, it was possible to observe differences in the performance of the four classification models (SVM, MLP, RF and k-NN) among the nine chickpea cultivars (Figure 3). Confusion matrix is commonly used to interpret the estimated results of classification when using supervised learning models. Thus, they are essential, since a performance data, such as accuracy, is not enough to determine the superiority of one model over another (Xu et al. 2021).

When using the Support Vector Machine (SVM) learning model (Figure 3a), a high discriminatory capacity was verified, with minimal and proportionally distributed errors across cultivars, i.e., it presents a precise classification pattern with most observations concentrated on the main diagonal of the confusion matrix, corresponding to the true classes. By using different supervised learning models, Kiliç & Yalçin (2025), aiming to classify chickpea seed varieties of Turkish origin, observed that the SVM model showed a higher classification capacity (accuracy rate of 94 %), when compared to the k-NN and Naive Bayes models, with 89 and 86 %, respectively.

The Random Forest (RF) model (Figure 3d) demonstrated a high classification performance, similar to the SVM, but with a slight increase in incorrect classifications. Errors predominantly occurred among cultivars of the same group (kabuli and desi), suggesting that shared morphological traits pose challenges for varietal discrimination. Through this model, there is a tendency to confuse desi group cultivars, as observed among the cultivars Hari, UPL-06 and 10108, in which the model considered the visual characteristics less distinctive. However, Hong et al. (2015), when identifying seeds of rice cultivars, found that the RF-based algorithm was more efficient for classification. Therefore, despite slight divergences, the RF method has a consistent classification, revealed by the high accuracy, for most of the analyzed cultivars.

The Multilayer Perceptron (MLP) model (Figure 3b) exhibited intermediate classification performance, with a pronounced tendency to misclassify cultivars within the kabuli group. An example of incoherence can be observed in the classification of the Cristalino cultivar, which was mistakenly identified as Aleppo by the model. Gheeta & Shanthi (2020) and Ghaffari (2024) point out that MLP is a method that regularly employs

Tabela 4. Calculation formulas and explanations of the performance metrics.

Indicator	Formula	Evaluation
Accuracy	$(Vp + Vn)/(Vp + Fp + Vn + Fn)$	Proportion of accurately estimated samples of the total number of samples
Error rate	$(Fp + Fn)/(Vp + Fp + Vn + Fn)$	Ratio between incorrectly estimated samples and total number of samples
Recall	$Vp/(Vp + Fn)$	Proportion of positive values classified as true
Specificity	$Vn/(Vn + Fp)$	Proportion of negative values classified as true
Precision	$Vp/(Vp + Fp)$	Ratio of correctly classified positive samples to estimated total positive samples. It is also called Positive Predictive Value
F1-score	$(2 * precision * recall)/(precision + recall)$	Harmonic mean sensitivity. It finds both false positives and false negatives
MCC	$(c * s - \sum_k^K pk * tk)/[\sqrt{(s^2 - \sum_k^K pk^2) * (s^2 - \sum_k^K tk^2)}]$	Performance of the model for the multiclass case, taking into account the confusion matrix, where C_{ij} is the value of the confusion matrix in row i, column j; $tk = \sum_i^K C_{ik}$ the total actual samples of class K; $pk = \sum_i^K C_{ik}$ the total of predictions as class k; $c = \sum_k^K C_{kk}$ the total correctly predicted samples; $s = \sum_i^K \sum_j^K C_{ij}$ the total number of samples

Vp: true positives; Vn: true negatives; Fp: false positives; Fn: false negatives; C: total actual samples of class K; S: total number of samples; PK: total predictions as class K; TK: total actual samples of class K; k: number of classes.

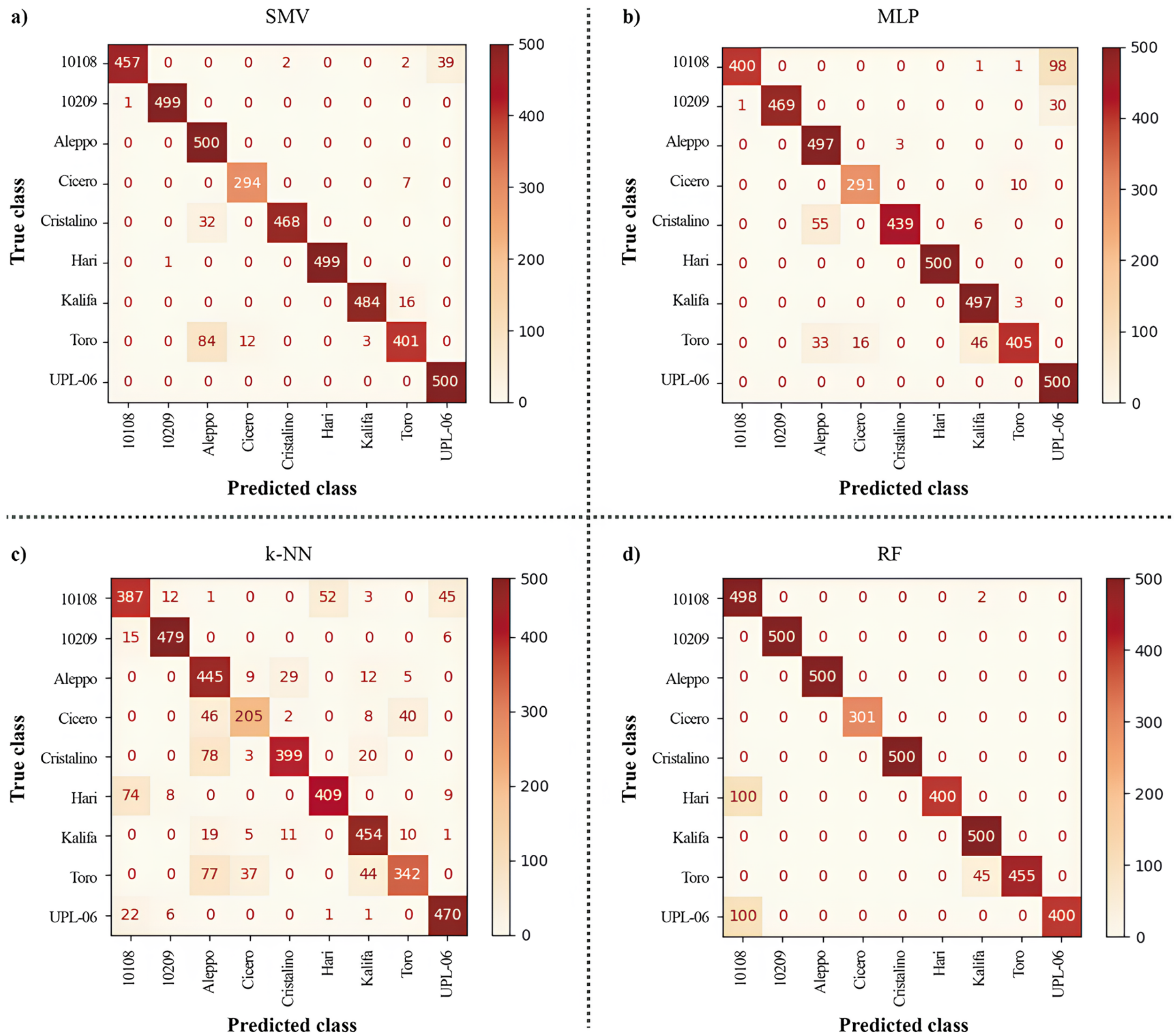

Figure 3. Confusion matrix for each classification model based on supervised learning. a) SVM: Support Vector Machine; b) MLP: Multilayer Perceptron; c) k-NN: k-Nearest Neighbors; d) RF: Random Forest. The numbers in the columns of real class and predicted class represent the cultivars: 1) 10108; 2) 10209; 3) Aleppo; 4) Cícero; 5) Cristalino; 6) Hari; 7) Kalifa; 8) Toro; 9) UPL-06.

few hidden neurons, which can result in inaccurate modeling and prediction, when compared to other models. In addition, it is noteworthy that cultivars of the kabuli group share a very similar morphology, which may have made it difficult to accurately discriminate the materials. Most kabuli cultivars have a light-colored seed coat, with very small variations in shade (Singh et al. 2022). Differences in diameter, length and width are also small and often overlap, and thus, the reduced morphometric variability can pose a challenge for the accurate discrimination in *C. arietinum* seeds (Pourdabani et al. 2019).

Among the evaluated models, the k-Nearest Neighbors (k-NN) (Figure 3c) exhibited the lowest performance in discriminating the chickpea cultivars, proving to be less effective, with very similar characteristics regardless of the group to which they belong. In common bean varieties, this behavior was also observed. Li et al. (2025) found that the k-NN model had a low classificatory capacity. Despite the increase in accuracy (51.59 %), after performing a pre-processing, the value found was considered low. The authors argue that this performance was due to the high number of variables in the data set, which undermines the classifications obtained through this model. In addition, it is important to highlight that the low performance of k-NN may also be associated with the high similarity between the cultivars in both groups and the adjustment of hyperparameters, and further studies that apply different k values and distance metrics are necessary.

In view of the variations in efficiency observed among the presented models, a set of factors is taken into account, which allow inferences about their performance. As presented in the confusion matrices (Figure 3), the results of the performance values (Table 5) are based on the number of positive and negative instances that are classified from the confusion matrix (Hossin & Sulaiman 2015). The performance metrics corroborated that SVM was the most accurate model (95.37 %), when compared to the other models. Unlike the SVM, the k-NN model showed accuracy of 83.46 %, highlighting a greater difficulty in differentiating cultivars, in relation to the others.

This pattern aligns with previous findings across diverse crops. Hu et al. (2020) found that the SVM model showed a high accuracy (87.5-91.67 %) in discriminating legume seeds. For corn, the accuracy rate of 98.2 % was found when this model was explored to identify corn varieties (Yang et al. 2015). For the RF model, Hong et al. (2015) observed that the accuracy of the classification can reach 90.54 % in different rice cultivars.

The proportion of incorrect classifications, considered as the error rate (Table 5), was low for the SVM (4.63 %) and RF (5.74 %) models, whereas k-NN presented the highest percentage value (16.54 %). In contrast, accuracy, which analyzes the reliability of positive ratings, was the highest in SVM (97.15 %). This result showed that the SVM model had a lower occurrence of false positives, when compared to the other tested algorithms.

Table 5. Results of the performance values evaluated for each model: Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forest (RF) and k-Nearest Neighbors (k-NN).

Performance	SVM	MLP	RF	k-NN
Accuracy (%)	95.37	92.95	94.26	83.46
Error rate (%)	4.63	7.05	5.74	16.54
Precision (%)	97.15	96.006	93.34	84.16
Recall	95.37	92.95	94.26	83.46
Specificity (%)	99.42	89.05	99.27	88.59
F1-score (%)	95.15	91.79	93.34	81.77
MCC (%)*	95.10	92.57	94.24	82.00

* MCC: Matthews correlation coefficient.

The RF model also showed a good accuracy (93.34 %), whereas k-NN (84.16 %) presented a greater difficulty in classifying the divergences among the cultivars.The recall metric (Table 5), which measures each model's capacity to correctly identify positive class instances, exhibited values closely aligned with the accuracy trends across all evaluated supervised learning models. This consistency suggests a stable model performance in both precision and sensitivity aspects.

Considering that the F1-score (Table 5) is directly related to recall and accuracy (Alabi et al. 2020), the SVM (95.15 %) and RF (93.34 %) models showed a higher performance, whereas k-NN (81.77 %) presented a lower performance. In this bias, it is noteworthy that a high percentage of F1-score ensures that it does not make serious mistakes, allowing cultivars of interest not to go unnoticed (Khatri et al. 2022).

Finally, reinforcing the superiority of the SVM and RF systems, the Matthews Correlation Coefficient (MCC) (Table 5) confirmed the superiority of performance of these models, with values of 95.10 and 94.24 %, respectively, standing out as the most balanced and reliable model for the classification task applied to the seeds of chickpea cultivars used in this study. The reason for this is related to the MCC's ability to evaluate the models, due to the correct prediction of both the class with the highest number of data and the one with a smaller number, what leads to a much more reliable performance, as it is considering the imbalance among the evaluated classes, in this case cultivars (Chicco & Jurman 2020). However, it is necessary to recognize that the performance measures were able to differentiate the efficiency of the classification models applied to chickpea seeds, taking into account both successes and errors in the prediction of the corresponding cultivar.

Computer vision and machine learning techniques enable automated phenotypic evaluations to help to reduce subjectivity, promote the selection of materials faster and more objectively, reducing the need for manual evaluations. In this way, the mentioned technologies are important auxiliary and/or complementary tools for the Brazilian legislation, facilitating procedures required for registration in the National Registry of Cultivars (Brasil 2022) and according to the Plant Variety Protection Law (Brasil 1997), as well as in internal and external quality control programs, ensuring quality standards and serving as support for the Rules for Seed Analysis (Brasil 2009). Therefore, the use of technologies aimed at image analysis and artificial intelligence showed promise by offering a greater accuracy in varietal identification and distinction.

In the future, the implementation of learning models in the classification of cultivars, through computer vision technologies, should be directed to the improvement and validation of these methodologies, with the objective of refining the classification accuracy. In addition, for each model, it is of fundamental importance to implement techniques to identify the most important attributes to facilitate the process of choosing the algorithm.

CONCLUSIONS

1. The Support Vector Machine and Random Forest models are the most recommended for the classification of chickpea cultivars, combining high accuracy, precision and recall, as well as high specificity and high Matthews Correlation Coefficient;
2. The extraction of morphometric traits from digital images and machine learning have proven to be effective tools for evaluating phenotypic traits and seed distinction of chickpea cultivars, offering a significant potential for automation in the seed sector;
3. Machine learning models are promising and can be useful for the seed industry, helping in the process of faster decision-making, as well as identification, certification and registration of cultivars, under a scientific criterion.

ACKNOWLEDGMENTS

The authors would like to thank the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Universidade Federal de Lavras (UFLa), for their financial and structural support. Finally, the authors thank the Empresa Brasileira de Pesquisa Agropecuária (Embrapa Hortaliças) and Warley Marcos Nascimento, for providing the seeds for the experiment.

REFERENCES

ALABI, R. O.; ELMUSRATI, M.; SAWAZAKI-CALONE, I.; KOWALSKI, L. P.; HAGLUND, C.; COLETTA, R. D.; MÄKITIE, A. A.; SALO, T.; ALMANGUSH, A.; LEIVO, I. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *International Journal of Medical Informatics*, v. 136, e104068, 2020.

ALI, A.; QADRI, S.; MASHWANI, W. K.; BRAHIM BELHAOUARI, S.; NAEEM, S.; RAFIQUE, S.; JAMAL, F.; CHESNEAU, C.; ANAM, S. Machine learning approach for the classification of corn seed using hybrid features. *International Journal of Food Properties*, v. 23, n. 1, p. 1110-1124, 2020.

AYELE, N. A.; TAMIRU, H. K. Developing classification model for chickpea types using machine learning algorithms. *International Journal of Innovative Technology and Exploring Engineering*, v. 10, n. 1, p. 5-11, 2020.

AJAZ, R. H.; HUSSAIN, L. Seed classification using machine learning techniques. *Journal of Multidisciplinary Engineering Science and Technology*, v. 2, n. 5, p. 1098-1102, 2015.

BEHERA, K.; BABBAR, A.; VYSHNAVI, R. G.; PATEL, T.; PRAJAPATI, S. S. Exploring the chickpea genotypes through morphological characterization for improved breeding. *International Journal of Plant & Soil Science*, v. 35, n. 18, p. 551-563, 2023.

BRASIL. Decreto-lei nº 9456 de 25 de abril de 1997. Lei de proteção de cultivares. *Diário Oficial da República Federativa do Brasil*, Brasília, DF, 28 abr. 1997. Seção 1, p. 8241-8246.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. *Regras para análise de sementes*. Brasília, DF: MAPA, 2009.

BRASIL. Portaria nº 502, de 19 de outubro de 2022. Estabelece as normas para a inscrição de cultivares e de espécies no Registro Nacional de Cultivares - RNC. *Diário Oficial da República Federativa do Brasil*, Poder Executivo, Brasília, DF, 20 out. 2022. Seção 1, p. 3.

ÇETIN, N. Machine learning for varietal binary classification of soybean (*Glycine max* (L.) Merrill) seeds based on shape and size attributes. *Food Analytical Methods*, v. 15, n. 8, p. 2260-2273, 2022.

ÇETIN, N.; OZAKTAN, H.; UZUN, S.; UZUN, O.; CIFTCI, C. Y. Machine learning based mass prediction and discrimination of chickpea (*Cicer arietinum* L.) cultivars. *Euphytica*, v. 219, e20, 2023.

CHICCO, D.; JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, v. 21, e6, 2020.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). *Faostat*: food and agriculture data. Rome: FAO, 2025.

GAUR, P. M.; TRIPATHI, S.; GOWDA, C. L.; RANGA RAO, G. V.; SHARMA, H. C.; PANDE, S.; SHARMA, M. *Chickpea seed production manual*. Andhra Pradesh: International Crops Research Institute for the Semi-Arid Tropics, 2010.

GEETHA, M. C. S.; SHANTHI, I. E. Forecasting the crop yield production in Trichy district using fuzzy c-means algorithm and Multilayer Perceptron (MLP). *International Journal of Knowledge and Systems Science*, v. 11, n. 3, p. 83-98, 2020.

GHAFFARI, A. Precision seed certification through machine learning. *Technology in Agronomy*, v. 4, e19, 2024.

GHAMARI, S. Classification of chickpea seeds using supervised and unsupervised artificial neural networks. *African Journal of Agricultural Research*, v. 7, n. 21, p. 3193-3201, 2012.

HONG, P. T. T.; HAI, T. T. T.; LAN, L. T.; HOANG, V. T.; THUY, N. T. Identification of seeds of different rice varieties using image processing and computer vision techniques. *Vietnam Agricultural Science Journal*, v. 13, n. 6, p. 1036-1042, 2015.

HOSSIN, H.; SULAIMAN, M. N. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, v. 5, n. 2, p. 1-11, 2015.

HU, X.; YANG, L.; ZHANG, Z. Non-destructive identification of single hard seed via multispectral imaging analysis in six legume species. *Plant Methods*, v. 16, n. 1, e116, 2020.

KHATRI, A.; AGRAWAL, S.; CHATTERJEE, J. M. Wheat seed classification: utilizing ensemble machine learning approach. *Scientific Programming*, v. 2022, e2626868, 2022.

KILIÇ, I. B.; YALÇIN, N. A novel hybrid methodology based on transfer learning, machine learning, and relief for chickpea seed variety classification. *Applied Sciences*, v. 13, n. 3, e1334, 2025.

KOKLU, M.; OZKAN, I. A. Multiclass classification of dry beans using computer vision and machine learning techniques. *Computers and Electronics in Agriculture*, v. 174, e105507, 2020.

KOKLU, M.; SARIGIL, S.; OZBEK, O. The use of machine learning methods in classification of pumpkin seeds (*Cucurbita pepo* L.). *Genetic Resources and Crop Evolution*, v. 68, n. 7, p. 2713-2726, 2021.

KUMAR, N.; HONG, S.; ZHU, Y.; GARAY, A.; YANG, J.; HENDERSON, D.; ZHANG, X.; XU, Y.; LI, Y. Comprehensive review of chickpea (*Cicer arietinum*): nutritional significance, health benefits, techno-functionalities, and food applications. *Comprehensive Reviews in Food Science and Food Safety*, v. 24, e70152, 2025.

LI, S.; SUN, L.; JIN, X.; FENG, G.; ZHANG, L.; BAI, H.; WANG, Z. Research on variety identification of common bean seeds based on hyperspectral and deep learning. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 326, e125212, 2025.

OTSU, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, v. 9, n. 1, p. 62-66, 1979.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLASS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; HESNAY, R. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, v. 12, n. 85, p. 2825-2830, 2011.

PIERGIOVANNI, A. R. *Ex situ* conservation of plant genetic resources: an overview of chickpea (*Cicer arietinum* L.) and lentil (*Lens culinaris* Medik.) worldwide collections. *Diversity*, v. 14, e941, 2022.

POURDARBANI, R.; SABZI, S.; GARCÍA-AMICIS, V. M.; GARCÍA-MATEOS, G.; MOLINA-MARTÍNEZ, J. M.; RUIZ-CANALES, A. Automatic classification of chickpea varieties using computer vision techniques. *Agronomy*, v. 9, e672, 2019.

PRIYADARSHI, M. B.; SHARMA, A.; CHATURVEDI, K. K.; BHARDWAJ, R.; LAL, S. B.; FAROOQI, M. S.; KUMAR, S.; MISHRA, D. C.; SINGH, M. Comparing various machine learning algorithms for sugar prediction in chickpea using near-infrared spectroscopy. *Legume Research*, v. 46, n. 2, p. 251-256, 2023.

SILVA, L. C. D.; CARDOSO, E. da S.; MENCALHA, J.; GOMES, D. A.; MIGUEL, J. A. de C.; CARDOSO, J. V. C.; SANTOS, H. O. dos; CARNEIRO, V. Q. Computer vision for assessment the seed coat color of carioca common beans. *Agronomy Journal*, v. 116, n. 5, p. 2141-2150, 2024.

SINGH, R. K.; SINGH, C.; AMBIKA, C. B. S.; MAHTO, R. K.; PATIAL, R.; GUPTA, A.; GAHLAUT, V.; HAMWIEH, A.; UPADHYAYA, H. D.; KUMAR, R. Exploring chickpea germplasm diversity for broadening the genetic base utilizing genomic resources. *Frontiers in Genetics*, v. 13, e905771, 2022.

TBIT. *Sistema GroundEye*®. 2022. Available at: https://www.tbit.com.br/sistema-groundeye/. Access on: May 30, 2025.

ULU, B.; OZAKTAN, H.; ÇETIN, N.; JAHANBAKHSHI, A.; ULU, B.; UZUN, S.; UZUN, O. Smart chickpea seed classification system based on machine vision and deep learning along with a review of challenges, limitations and future trends. *Smart Agricultural Technology*, v. 12, e101093, 2025.

UPADHYAYA, H. D.; DWIVEDI, S. L.; BAUM, M.; VARSHNEY, R. K.; UDUPA, S. M.; GOWDA, C. L.; HOISINGTON, D.; SINGH, S. Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biology*, v. 8, e106, 2008.

VAROL, İ. S.; ÇETIN, N.; KIRNAK, H. Evaluation of image processing technique on quality properties of chickpea seeds (*Cicer arietinum* L.) using machine learning algorithms. *Journal of Agricultural Sciences*, v. 29, n. 2, p. 427-442, 2023.

XU, P.; YANG, R.; ZENG, T.; ZHANG, J.; ZHANG, Y.; TAN, Q. Varietal classification of maize seeds using computer vision and machine learning techniques. *Journal of Food Process Engineering*, v. 44, n. 11, e13846, 2021.

YANG, X.; HONG, H.; YOU, Z.; CHENG, F. Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors*, v. 15, n. 7, p. 15578-15594, 2015.