ABRATES
Journal of Seed Science
ISSN 2317-1545
www.abrates.org.br/revista

ARTICLE

Journal of Seed Science, v.45, e202345038, 2023

Classification of lentil seed vigor based on seedling image analysis techniques and interactive machine learning

Marcelo Augusto Rocha Limão[1*], Denise Cunha Fernandes dos Santos Dias[1], Joyce de Oliveira Araújo[1], Júlia Martins Soares[1], Warley Marcos Nascimento[2], Laércio Junio da Silva[1]

http://dx.doi.org/10.1590/2317-1545v45277692

ABSTRACT: The search for techniques that allow for the rapid and accurate assessment of seed vigor, such as the Seedling Analysis System (SAPL®) and ILASTIK®, can be promising alternatives for seedling image analysis. The objective of this work was to classify the vigor of lentil seeds using seedling image analysis techniques and interactive machine learning. Seeds from seven lots were characterized for physiological potential through germination and vigor tests. For computerized seedling analysis, the seeds were subjected to seedling growth tests at 20 °C for three, four, five, and ten days, and then photographed using a digital camera. The images were processed using SAPL® software, yielding values for total length, root length, shoot length, and vigor, growth, and uniformity indices. ILASTIK® provided data on the percentage of vigorous seedlings, non-vigorous seedlings, and dead seeds. The total length of seedlings, root length, shoot length, and vigor indices determined at 4 days of germination by SAPL® allowed for the classification of lots in terms of vigor. Data obtained by ILASTIK® at 4 days of germination, used in machine learning studies, enable the development of models with high accuracy for seed vigor assessment.

Index terms: germination, *Lens culinaris* Medik, machine learning, physiological potential, software, vigor.

RESUMO: A busca por técnicas que possibilitem avaliar o vigor de sementes de forma rápida e assertiva como o Sistema de Análise de Plântulas (SAPL®) e o ILASTIK® podem ser alternativas promissoras para a análise de imagem de plântulas. O objetivo do trabalho foi realizar a classificação do vigor de sementes de lentilha utilizando técnicas de análise de imagem de plântulas e aprendizagem interativa de máquina. Sementes de sete lotes foram caracterizadas quanto ao potencial fisiológico pelos testes de germinação e vigor. Para a análise computadorizada de plântulas, as sementes foram submetidas ao teste de crescimento de plântulas a 20 °C por três, quatro, cinco e dez dias e, em seguida, fotografadas utilizando uma câmera digital. As imagens foram processadas pelo software SAPL®, obtendo-se valores de comprimento total, da raiz, parte aérea e índices de vigor, crescimento e uniformidade. Pelo ILASTIK®, foram obtidos dados de porcentagem de plântulas vigorosas, não vigorosas e sementes mortas. O comprimento total das plântulas, raiz, parte aérea e os índices de vigor determinados aos 4 dias de germinação pelo SAPL® permitem classificar os lotes quanto ao vigor. Os dados obtidos pelo ILASTIK®, aos 4 dias de germinação, utilizados nos estudos de aprendizagem de máquina, permitem o desenvolvimento de modelos com alta precisão para avaliação do vigor das sementes.

Termos para indexação: germinação, *Lens culinaris* Medik*, machine learning,* potencial fisiológico, software, vigor.

***Corresponding author**
marcelo.limao@ufv.br

Received: 08/17/2023.
Accepted: 10/20/2023.

[1]Universidade Federal de Viçosa (UFV), Departamento de Agronomia, Campus universitário, Av. PH Rolfs, 36570-900 – Viçosa, MG, Brasil.

[2]Embrapa Hortaliças, Caixa Postal 280 - Brasília, DF, Brasil.

INTRODUCTION

Physiological potential of seeds has traditionally been assessed through germination and vigor tests, which have limitations such as being destructive, time-consuming, and involving subjective interpretations (Ahmed et al., 2018; Elmasry et al., 2019; Xia et al., 2019).

In recent years, computerized image analysis of seedlings has been increasingly used to evaluate seed performance. Such techniques offer a rapid evaluation of seedlings, reliable results, and a database and images that can be stored and referenced as needed (Gomes-Junior, 2020). Several studies have demonstrated the efficiency of computerized analysis for assessing seed vigor in various crops such as common beans (Abud et al., 2022), chickpeas (Araújo et al., 2021), corn (Castan et al., 2018; Andriazzi et al., 2023), soybeans (Antunes-Neto et al., 2020), peanut (Barbosa et al., 2016), wheat (Brunes et al., 2016), *Brassica* spp. (Medeiros et al., 2022), among others.

Currently, several systems are available to assist in computerized seed and seedling analysis, with notable mentions including the Seed Vigor Imaging System (SVIS®), the Automated Seed Vigor Analysis System (Vigor-S®), GroundEye®, ImageJ, the Seedling Analysis System (SAPL®), and ILASTIK® (Sako et al., 2001; Collins, 2007; Castan et al., 2018; Medeiros and Pereira, 2018; Berg et al., 2019).

SAPL® stands out among these systems since it is a free software with low image acquisition costs that can be obtained using a digital camera or even smartphones. This system provides information on the total length of seedlings, primary root, and hypocotyl, as well as various indices such as vigor, growth, and seedling uniformity (Medeiros and Pereira, 2018). Several studies have confirmed its efficiency in evaluating physiological potential of seeds in crops such as soybeans (Medeiros e Pereira, 2018), *Moringa oleifera* (Pereira et al., 2020), bean (Medeiros et al., 2019) and chickpea (Araújo et al., 2021).

The ILASTIK® software employs an interactive machine learning (IML) to define IML models that enable users to classify and analyze images (Berg, 2019). This software can be applied in fields of seeds, allowing for classification of both seeds and seedlings obtained in germination tests, establishing different classes of vigor (Medeiros et al., 2020a). Medeiros et al. (2020a, b) found that ILASTIK® software was efficient in classifying vigor of soybean seeds based on their appearance, making it possible to identify damaged seeds and classify seedlings into different vigor levels.

In this context, this study aimed to classify vigor of lentil seeds using image analysis techniques and IML with the SAPL® and ILASTIK® software.

MATERIAL AND METHODS

The study was conducted at the Seed Research Laboratory of the Department of Agronomy, *Universidade Federal de Viçosa*, in Viçosa, MG, Brazil. Seven lots of lentil seeds from the Silvina cultivar were used. Initially, seeds from each lot were subjected to tests to characterize their initial quality:

Moisture content: two 20-seed replications were used, applying the oven method at 105 ± 3 °C for 24 hours. Results were expressed as a percentage (Brasil, 2009).

*Germination: f*our 50-seed replications were distributed on paper towels moistened with water at an amount equivalent to 2.5 times the weight of dry paper and kept in a germinator at 20 °C. Evaluations were performed on the fifth and tenth day after sowing (DAS), with results being expressed as percentage of normal seedlings on the tenth day (Brasil, 2009).

First germination count: this was performed along with the germination test, calculating the percentage of normal seedlings obtained on the fifth day after sowing (Brasil, 2009).

Seedling emergence (E) and emergence speed index (ESI): it was conducted in a controlled growth room using plastic trays containing a 1:1 mixture of soil and sand, with four 50-seed replications, sown at a depth of 1.0 cm. Daily counts were made, and emergence percentage was calculated by counting the total emerged seedlings, with cotyledons above the substrate surface, until the establishment of the stand. *ESI* was calculated following Maguire (1962).

Seedling dry mass: Eight 10-seed replications were distributed in a line drawn on the upper third of moistened germination paper as described for the germination test. Rolls were prepared and placed vertically in a germinator for ten days at 20 °C. The obtained seedlings had their cotyledons removed and were placed in a forced-air circulation oven at 65 °C for 72 hours. After this period, the samples were weighed on a precision balance (0.001 g). Results were expressed in mg per seedling (Krzyzanowski et al., 2020).

Cold test: Four 50-seed replications were sown on moistened paper towels as described for the germination test but kept at 10 °C for 24 hours before sowing. Then, 60 mL of soil was added over the seeds, creating rolls that were placed in plastic bags and kept in a BOD-type chamber at 10 °C for seven days. After this period, the plastic bags were removed, and the rolls were transferred to a germinator at 20 °C for five days to evaluate the percentage of normal seedlings (Cícero and Vieira, 2020).

Assay I - Evaluation of physiological potential of lentil seeds using SAPL® software

For computerized seedling analysis, seedling length test was conducted with eight 10-seed replications from each lot (eight images per lot) according to Krzyzanowski et al. (2020). Seeds were distributed in a line drawn on the upper third of germination paper, longitudinally. Rolls were prepared and placed vertically in a germinator for three, four, five, and ten days at 20 °C. At the end of each period, seedlings were transferred from the germination paper to a photographic base made of ethylene vinyl acetate (E.V.A.) sheet containing eleven cells of five centimeters in width. Subsequently, they were photographed using a Nikon digital camera, Coolpix P510 model, set to 16 Megapixels, with a shutter speed of 1/15 seconds and aperture of f/3.3. The camera was positioned at a height of 40 cm and an angle of 90° relative to the photographic base, using a copystand-type support.

Processing of images by SAPL® software: The images were stored and inserted into the selected default file of the SAPL® software. The indices provided by the software were defined by Sako et al. (2001). Subsequently, the reference value for the growth and uniformity index (70% and 30%, respectively) was selected for calculating the vigor index, resulting in the following equation: VI = (0.70 * C + 0.30 * U), where VI is the vigor index, C is the growth index, and U is the uniformity index. After recording the initial values, image processing of seedlings was performed for each lot replication. The software provided measurements of shoot length - SL (mm per seedling), root length - RL (mm per seedling), and total seedling length - TSL (mm per seedling). Subsequently, the seedling length values obtained by SAPL® were entered into the SeedCalc package of the R software (Silva et al., 2019) to generate more adjusted uniformity, growth, and vigor indices.

Assay II - Classification of lentil seedlings and interactive machine learning using ILASTIK® software

The images obtained as described above for the SAPL® software were processed in the ILASTIK® software. For segmentation, color and pixel classification tool was used, initially determining two segmentation classes, “seed or seedling” (region of interest) and “background” (discard region). To create probability maps, the pixels belonging to the regions of each segmented class were trained for seedling identification based on the colors defined in segmentation and illustrated in Figure 2. Then, the software was trained once more for recognition of three classes of seedlings: strong (with well-developed root and shoot, exceeding 5 cm); weak seedlings (with absence, underdevelopment, or deformation of some essential structures); and non-germinated seeds (Figure 1). The trained classifier was applied to all images, and probability maps were exported, generating data on the number of strong, weak seedlings, and non-germinated seeds. A brief summary of the main steps of the interactive method for classifying lentil seeds and seedlings is illustrated in Figure 2.

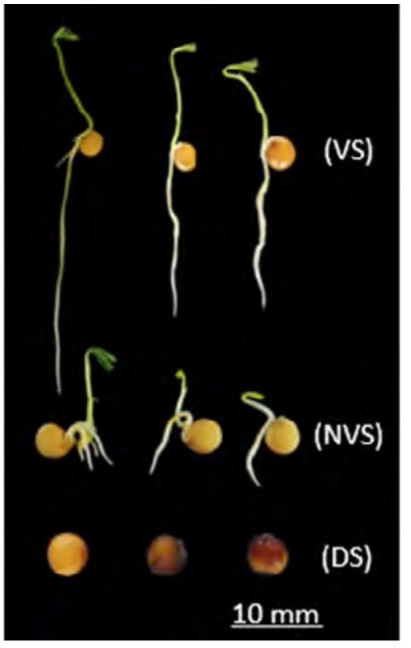

Figure 1. Standard for vigorous seedlings (VS), non-vigorous seedlings (NVS), and dead seeds (DS) of lentils for training in the ILASTIK® software.

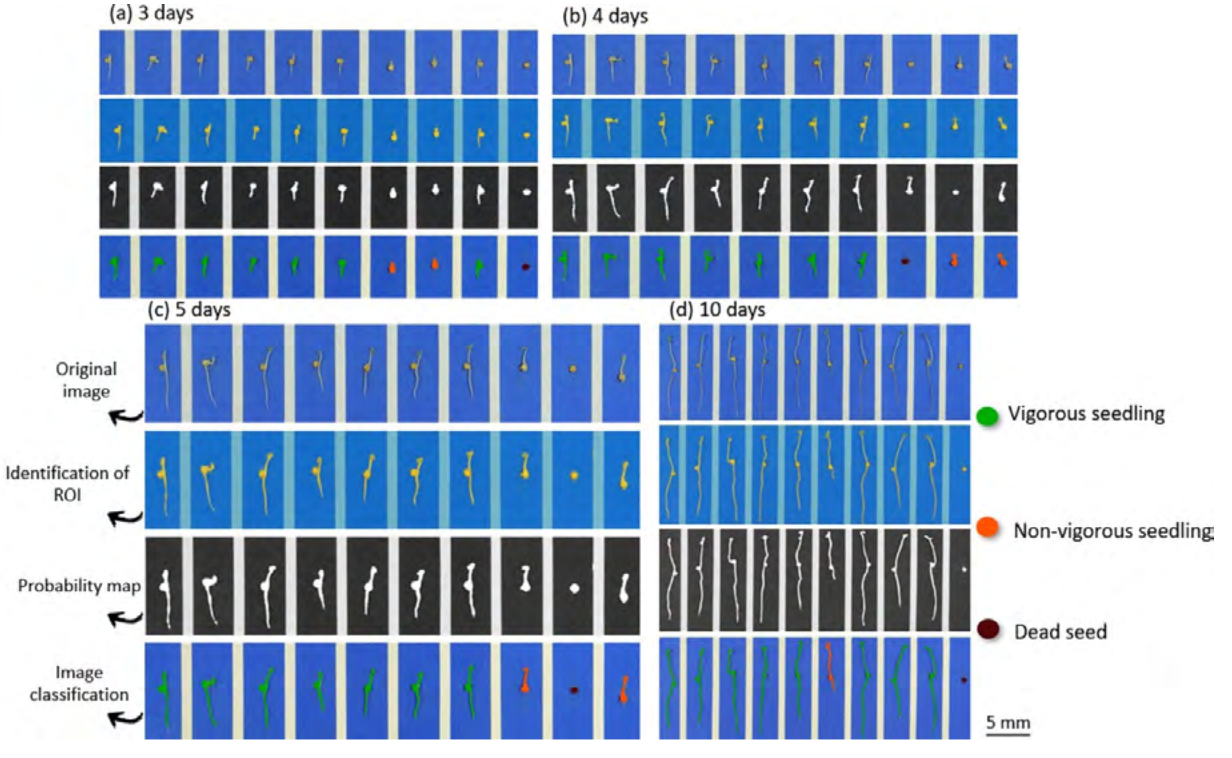

Figure 2. Representative schematic of the interactive machine learning training and classification stages for the physiological potential of different lentil lots at 3 (a), 4 (b), 5 (c), and 10 days (d) after sowing. Vigorous seedling (green color), non-vigorous seedling (orange color), dead seed (brown color). Initially, there are the original seedling images on the blue background (I). Then, rendering was applied for segmentation and improvement of the region of interest (ROI) to acquire the probability map (II). After rendering the ROI, the software identifies individual seedlings and seeds using the probability map (III), and finally, the prediction of the classification for each seed or seedling is made based on the training colors of the respective classifier groups (IV).

Experimental design and statistical analysis: Initial characterization tests and assessment of physiological potential by SAPL® and ILASTIK® software were conducted in a completely randomized design (CRD) with seven lots and four replications. Data were tested for normal error distribution using the Shapiro-Wilk test and for variance homogeneity using the Bartlett test. Subsequently, data were subjected to analysis of variance for significance by the F-test and means obtained for each lot were compared by the Tukey test at a 5% probability level. Principal Component Analysis (PCA) was performed for all combinations of data obtained from the initial lot characterization tests and the data evaluated by SAPL® and ILASTIK® software. For all analyses, the statistical software R 4.1.1. was used (R Core Team, 2022).

RESULTS AND DISCUSSION

Seed moisture content from different lots of lentils was consistent, ranging from 13.1 to 13.4 (Table 1). This characteristic is important since moisture uniformity is essential for standardized evaluations and consistent results (Marcos-Filho, 2015; Worma et al., 2019).

Seed germination in lot 5 (99%) was higher than in the other lots, followed by lots 1 (95%), 2 (92%), 4 (94%), 6 (94%), and 7 (91%), while lot 3 (83%) stood out with the lowest germination percentage (Table 1), with 16 percentage points less when compared to lot 5. Germination test allows seeds to express their maximum physiological potential, as it provides optimal conditions of moisture, temperature, and light, resulting in important information about seed performance under these conditions (Queiróz et al., 2019). All lots exhibited germination rates higher than 80%, which is the minimum value established by regulations for the commercialization of lentil seeds in Brazil (Brasil, 2012).

When evaluating the first germination count, it is observed that lots 6 and 5 obtained the highest percentages (above 80%), followed by lots 4 and 7, with lots 1 and 3 having the lowest values, although not differing from lot 2. The first count is considered a complementary test to the germination test and is indicative of lot classification regarding vigor, expressing differences in germination speed among lots (Krzyzanowski et al., 2020). Therefore, differences in the performance of lots 1, 2, 4, 6, and 7, which were not observed in the germination test, could be detected when evaluating germination speed (Table 1).

Table 1. Initial quality characterization of seven lots of lentil cv. Silvina: moisture content (MC), germination (G), first germination count (FGC), seedling emergence (SE), emergence speed index (ESI), seedling dry weight (SDW), and cold test (CT).

Lot	MC (%)	G (%)	FGC (%)	SE (%)	ESI (index)	SDW (mg.seedling^{-1})	CT (%)
1	13.16	95 b	53 c	91 a	10.43 a	0.12 ab	91 ab
2	13.09	92 b	63 bc	88 b	9.66 b	0.11 ab	89 ab
3	13.19	83 c	58 c	72 c	8.08 c	0.08 c	74 c
4	13.39	94 b	70 b	91 a	10.84 a	0.10 b	92 ab
5	13.15	99 a	86 a	92 a	10.13 a	0.13 a	99 a
6	13.41	94 b	89 a	85 b	9.29 b	0.10 b	94 a
7	13.20	91 b	71 b	92 a	10.29 a	0.11 ab	87 b
F	-	34.29*	22.29*	8.9*	17.73*	88.39*	21.73*
CV (%)	-	1.82	8.09	5.44	4.45	8.5	3.67

* Significant at the 5% probability level; F = calculated F-value; CV = coefficient of variation. Means followed by the same letter in the column do not differ significantly from each other by the Tukey test at a 5% probability level.

Seedling emergence results in terms of lot physiological potential were similar to those of seedling emergence speed index (ESI). Lots 1, 4, 5, and 7 were superior, followed by lots 2 and 6, considered intermediate, and lot 3 had the poorest performance (Table 1). The seedling emergence tests, and ESI allowed for a more detailed classification of the lots in terms of vigor compared to the other tests conducted.

Regarding seedling dry mass, the best performance was achieved for lot 5, with the lowest value for lot 3 (Table 1). Lots with seedlings exhibiting higher dry mass contents have greater vigor, indicating a greater transfer of dry mass from seed reserve tissue to the embryonic axis (Pedó et al., 2018; Krzyzanowski et al., 2020). Lot 5 also showed superior performance in the cold test, along with lot 6, while lots 3 and 7 had lower performance. Thus, exposing seeds to a combination of low temperature and high humidity helped identify lots with different performance potentials (Cícero and Vieira, 2020).

In general, lot 5 consistently performed among the best, while lot 3 consistently exhibited the poorest performance in all the tests conducted, with some variations among the medium-vigor lots depending on the specific test. Using lots with different levels of physiological potential is fundamental in studies related to seed vigor assessment, especially when defining or testing new methodologies. It is important that these new methodologies yield vigor classifications for the lots that are similar to those obtained in already established tests for the species under study.

The results of the computerized analysis of seedlings using the SAPL® software revealed a significant difference in shoot length (SL) of seedlings from different lots starting on the third day of germination. Lot 5 exhibited the best performance, while lot 3 had the lowest shoot growth, with lots 1, 2, 4, 6, and 7 showing intermediate values (Table 2). However, it is worth noting that on the fourth, fifth, and tenth days of germination, there was even greater separation among the lots in terms of shoot length, with lots 5 and 3 remaining as the lots with the highest and lowest SL, respectively, as observed on the third day.

On the fourth day of germination, lots 1, 2, and 7 did not differ from each other and were superior to lots 4 and 6. By the fifth day, lots 1 and 7 were superior to lot 2 (Table 2). It is important to note that on the tenth day, it was also possible to classify all the lots into distinct vigor levels. However, considering the longer time required to obtain these results, the assessments made on the fourth and fifth days are more desirable as they provide faster information about the physiological potential of the seeds (Table 2).

Table 2. Shoot length (SL), primary root length (PRL), total seedling length (TSL), uniformity index (UI), growth index (GI), and vigor index (VI) obtained by SAPL® at three, four, five, and ten days after sowing for seven lots of lentil seeds.

Lot	SL	PRL	TSL	UI	GI	VI
	(mm. seedling^{-1})			Indexes		
			3 days			
1	4.73 b	9.1 b	13.83 b	430.96 b	620.66 b	517.09 ab
2	4.57 b	8.78 b	13.35 b	422.13 b	617.37 b	521.01 ab
3	3.59 c	8.09 b	11.67 c	385.66 c	537.41 c	448.69 c
4	4.61 b	8.29 b	12.90 b	420.52 b	613.92 b	554.79 ab
5	6.97 a	15.95 a	22.95 a	484.84 a	734.84 a	564.59 a
6	4.41 b	8.43 b	12.85 b	419.5 b	612.22 b	546.06 ab
7	4.63 b	8.59 b	13.22 b	424.61 b	620.54 b	531.89 ab
F	55.95*	141.87*	246.01*	46.29*	38.51*	33.52*
CV (%)	5.80	4.94	3.39	2.03	3.00	2.89

Continue...

Table 2. Continuation.

Lot	SL	PRL	TSL	UI	GI	VI
	(mm. seedling^{-1})			Indexes		
			4 days			
1	9.08 b	18.34 b	27.42 b	452.71 ab	643.71 b	541.07 ab
2	9.10 b	18.14 b	27.24 b	431.32 ab	635.38 b	529.65 b
3	7.2 d	12.77 c	19.97 d	403.84 c	574.45 c	460.70 c
4	8.86 c	17.37 b	26.25 bc	427.53 ab	630.39 b	548.17 ab
5	11.74 a	26.48 a	38.22 a	510.84 a	751.10 a	565.67 a
6	7.93 cd	17.24 b	25.18 c	421.81 ab	622.15 b	546.67 ab
7	9.01 b	18.1 b	27.11 b	441.07 ab	647.92 b	555.13 ab
F	42.10*	153.69*	143.82*	39.51*	26.47*	11.56*
CV (%)	4.84	3.58	3.33	2.47	3.22	5.07
			5 days			
1	15.14 b	25.73 b	40.87 b	529.81 ab	753.62 b	588.70 ab
2	14.41 c	25.24 b	39.65 bc	511.75 ab	720.79 b	572.96 b
3	12.89 d	16.23 d	29.12 d	456.04 b	614.07 c	496.05 c
4	14.27 c	24.43 c	38.70 bc	517.96 ab	724.40 b	583.87 ab
5	19.95 a	34.10 a	54.08 a	554.90 a	822.67 a	634.91 a
6	13.33 cd	24.11 c	37.44 c	511.95 ab	712.17 b	584.56 ab
7	14.94 b	25.06 b	40.00 b	534.39 ab	728.29 b	596.59 ab
F	57.32*	255.75*	186.94*	11.33*	53.55*	19.96*
CV (%)	4.12	2.60	2.70	3.54	2.33	3.51
			10 days			
1	55.33 bc	62.50 c	117.84 b	667.68 a	755.92 b	709.91 ab
2	57.38 b	58.83 d	116.21 bc	631.88 ab	738.54 b	705.47 ab
3	38.59 e	46.32 e	84.91 d	610.33 b	663.14 c	596.42 c
4	54.29 c	60.46 d	114.76 c	624.85 ab	731.82 b	698.69 b
5	61.95 a	66.77 a	128.72 a	678.62 a	845.36 a	717.46 a
6	54.04 c	60.27 d	114.31 c	618.61 b	729.33 b	692.06 b
7	48.25 d	64.34 b	117.34 b	628.93 ab	750.52 b	704.83 ab
F	247.53*	316.02*	739.96*	31.92*	19.95*	29.25*
CV (%)	1.80	1.24	0.87	1.43	3.23	2.53

* Significant at the 5% probability level; F = calculated F-value; CV = coefficient of variation. Means followed by the same letter in the column do not differ significantly from each other by the Tukey test at a 5% probability level.

For the primary root length (RL) (Table 2), it can be observed that only lot 5 was superior to the others on the third day of germination, while the remaining lots did not differ from each other. On the fourth day, the best performance was observed for lot 5 and the poorest for lot 3. However, it was not possible to differentiate the other lots in terms of RL, which only occurred on the fifth and tenth days of germination. On the fifth day, a more detailed separation of lots 1, 2, 4, 6, and 7 in terms of RL is observed, similar to what was observed for SL, with lots 4 and 6 showing lower values compared to lots 1, 2, and 7. For these lots, on the tenth day, more significant differences in root growth are observed, with lower values for lots 2, 4, and 6, followed by lot 1, which is inferior to lot 7 (Table 2). Seeds with higher vigor produce seedlings with longer root lengths, a characteristic that is relevant for water and nutrient absorption from the soil (Krzyzanowski et al., 2020).

Regarding total seedling length (TSL), significant differences can be observed among the different lots on the third day, similar to what was observed for RL. Three classes were obtained, with lot 5 having the highest TSL and lot 3 having the lowest, while the other lots were in intermediate positions (Table 2). On the fourth, fifth, and tenth days of germination, the results were similar, with lot 5 performing the best, followed by lots 1 and 7, and lot 3 having the poorest performance. Lots 2, 4, and 6 exhibited intermediate performance. Considering that the speed in obtaining vigor test results is relevant for decision-making regarding lot management and disposition, it is recommended to perform assessments of seedling length (SL), root length (RL), and total seedling length (TSL) on the fourth day after sowing, as they allowed for the classification of lots into different levels of physiological potential.

Therefore, the use of SAPL® software enabled us to separate lots based on physiological potential as early as the fourth day after sowing, speeding up acquisition of information about seedling vigor, without the need to wait for the 10-day germination test period. The efficiency of this software in classifying soybean lots (Medeiros and Pereira, 2018) and chickpea lots (Araújo et al., 2021) based on physiological potential was confirmed when seedling analysis was also conducted on the fourth day after sowing. The application of SAPL® software for computerized seedling analysis based on growth data allowed for classification of different lots in terms of physiological potential, making it clear which lots performed best and worst in terms of their ability to transfer the seed reserves necessary for the full development of seedlings (Finch-Savage and Bassel, 2016; Araújo et al., 2021).

The results generated by SAPL® for the uniformity index (UI), growth index (GI), and vigor index (VI) on the third day of germination were similar to those of TSL in terms of classifying the physiological potential of lots. On the fourth and fifth days, lot 5, previously identified as superior, did not significantly differ from the lots classified as intermediate on the third day for UI and VI (Table 2). Unlike UI, which was not effective, Araújo et al. (2021) successfully classified chickpea seed lots based on physiological quality using GI and VI. However, on the tenth day of germination, GI and VI produced similar results, allowing for the classification of lots into three vigor levels.

Pereira et al. (2020) classified different lots of moringa (*Moringa oleifera*) based on seed vigor using data provided by the SAPL® software. This software has also proven to be efficient in assessing the physiological potential of bean (Medeiros et al., 2019) and crambe (Ribeiro et al., 2021) seeds, providing results that correlate with those obtained in other vigor tests such as root protrusion, germination, germination speed index, T50, and germination synchrony. Results related to the Growth Index (GI) and Vigor Index (VI) are crucial for estimating seed vigor and have shown greater sensitivity in characterizing seed lots in terms of physiological quality for various species (Pereira et al., 2020). The indices generated from seedling growth assessments are important for use in seed quality control programs (Silva et al., 2017).

The uniformity index (UI) is an important indicator to consider regarding seed quality. Non-uniformity in seedling emergence can lead to delays in the development and growth of plants in post-emergence phenological stages, resulting in non-uniformity during flowering and maturation, subsequently affecting the harvesting process (Marcos-Filho, 2015).

Table 2 highlights the greater stratification among lots in terms of physiological potential when evaluating TSL at 4, 5, and 10 days after sowing. These data show similarity with the results obtained in the initial physiological characterization tests (Table 1). This similarity is particularly evident for emergence variables, making these results significant. The introduction of new methodologies for the assessment of vigor that are faster, more sensitive, and efficient is important for decision-making regarding the management of lots in post-harvest and commercialization stages (Gomes-Junior et al., 2014; Medeiros and Pereira, 2018).

The principal component analysis shown in Figure 3 explains 92.8% of the total data variability. It shows that lots 5, 1, and 7 are clustered in the negative axes of the first component (PC 1), associated with variables characterizing initial

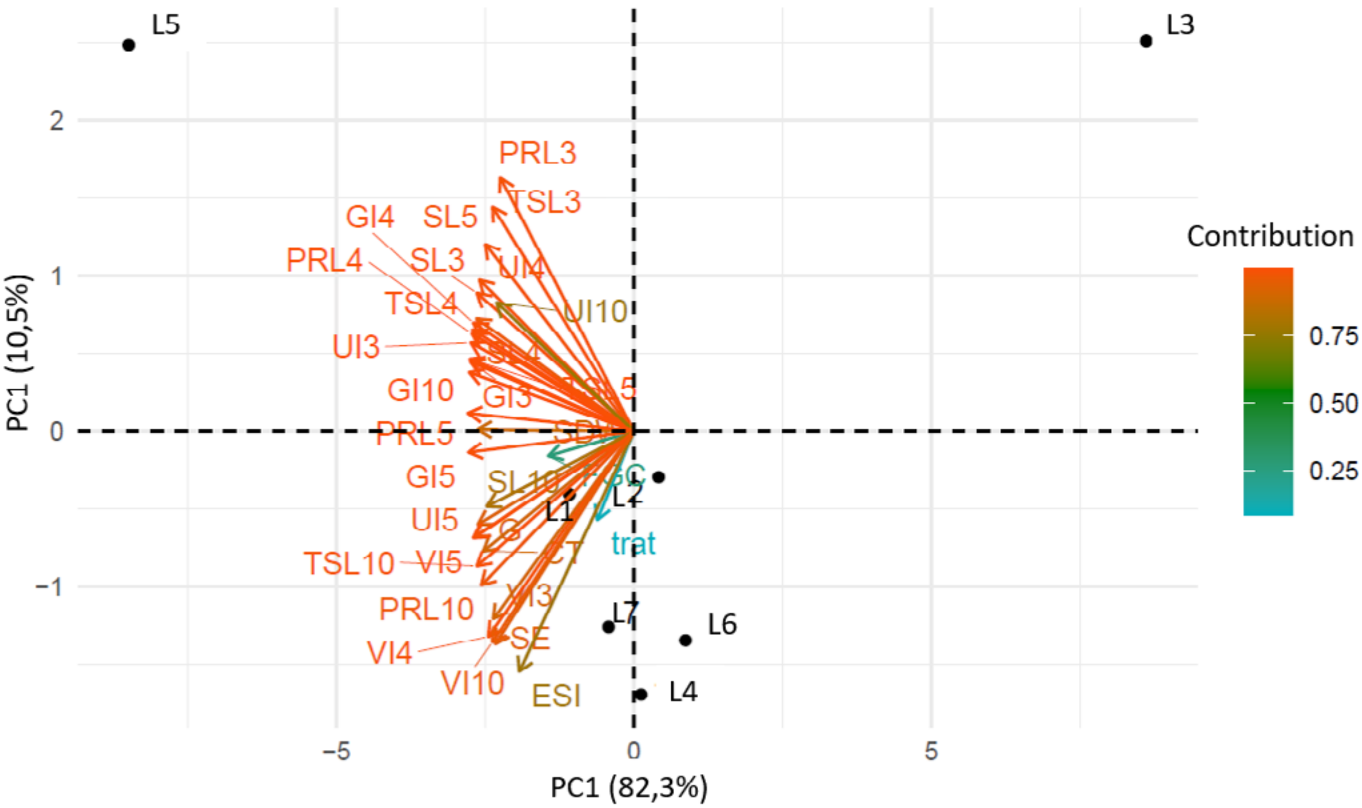

Figure 3. Principal Component Analysis (PCA) for the variables characterizing the initial seed quality, including germination (G), first germination count (FGC), seedling emergence (SE), emergence speed index (ESI), seedling dry weight (SDW), and cold test (CT), and data generated by SAPL® software, shoot length (SL), primary root length (PRL), total seedling length (TSL), uniformity index (UI), growth index (GI), and vigor index (VI) at the 3rd, 4th, 5th, and 10th day after sowing.

quality and seedling length data. Following them are lots 2, 4, and 6, which fall in the intermediate range. However, it is apparent that lot 3 exhibits a negative correlation with physiological potential variables, highlighting it as the least vigorous. This demonstrates that the application of technologies based on seedling image analysis using the SAPL® software is efficient in estimating the vigor of lentil seeds, providing rapid, efficient, and reliable results starting from the fourth day after sowing.

Figure 4 illustrates the probability map of lots generating vigorous seedlings, non-vigorous seedlings, and dead seeds at 3, 4, 5, and 10 days after sowing. More vigorous lots have a higher probability of producing seedlings with more developed shoot and primary roots, and a lower likelihood of generating weak seedlings and a lower proportion of dead seeds. It is evident that the seedling images obtained on the third day of germination did not allow for the classification of lots in terms of vigor. Instead, they only enabled the differentiation of lots with low vigor based on the probability of producing abnormal seedlings and dead seeds, as seen with lot 3.

However, greater reliability and sensitivity in distinguishing between classes were achieved starting from the fourth day of germination, with results similar to those obtained on the fifth day, indicating accuracy in assessing seed vigor. Thus, the ILASTIK® software allowed for the classification of lot 5 as the most vigorous, based on a higher probability of producing vigorous seedlings and a lower probability of generating non-vigorous seedlings and dead seeds. On the other hand, lot 3 exhibited a lower probability of producing vigorous seedlings and a higher probability of generating non-vigorous seedlings and dead seeds due to its lower physiological potential, as also confirmed in the tests in Tables 1 and 2. In general, lots 1, 4, and 7 showed intermediate vigor, followed by lots 2 and 6 (Figure 4).

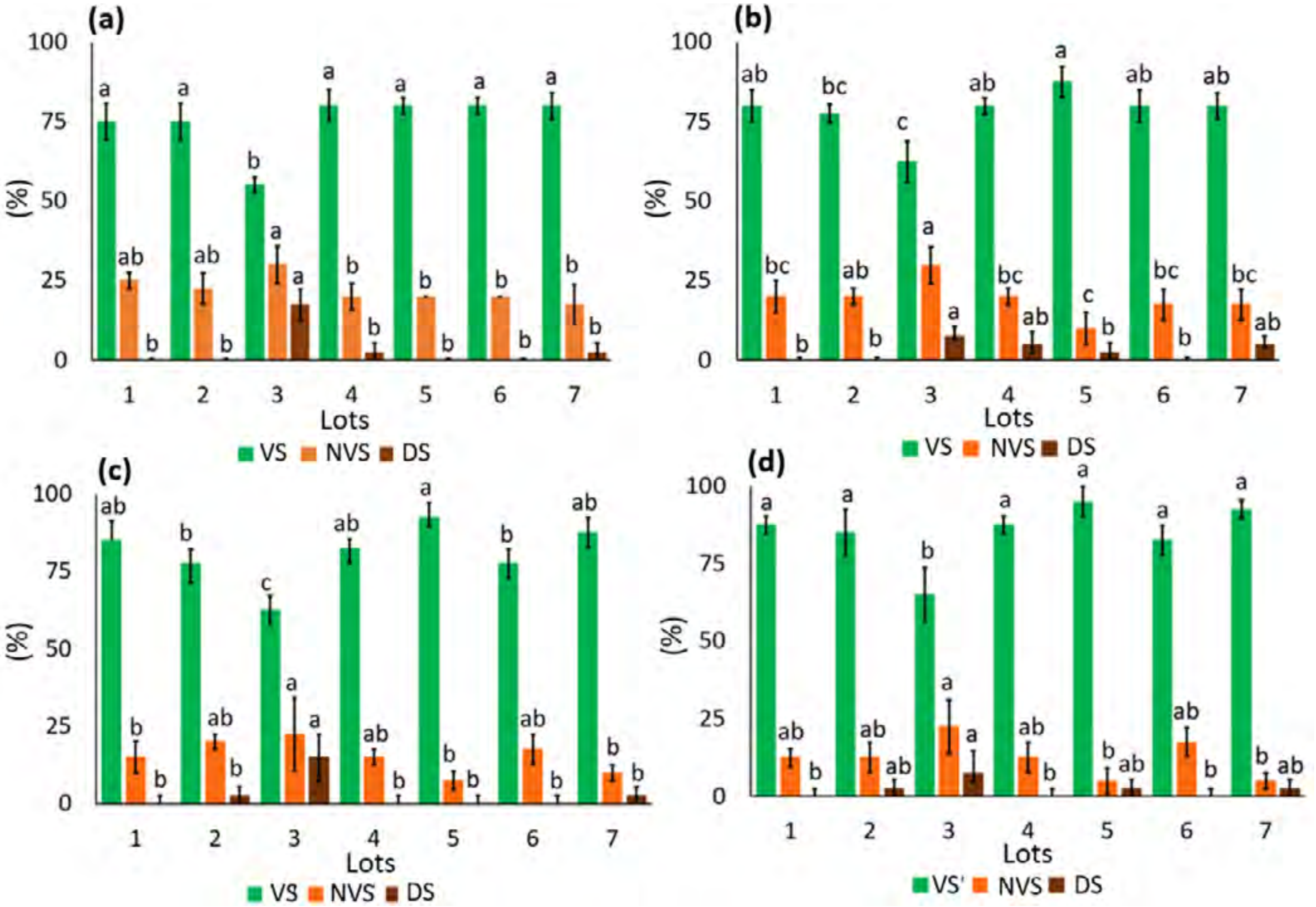

Figure 4. Percentage of vigorous seedlings (VS), non-vigorous seedlings (NVS), and dead seeds (DS) on the third (a), fourth (b), fifth (c), and tenth (d) day after sowing, according to data obtained by ILASTIK® software. The bars represent a 95% confidence interval.

By the tenth day, there was clustering of high and medium-vigor lots that did not differ from each other (lots 1, 2, 4, 5, 6, and 7). This can be attributed to the longer germination time, giving less-developed seedlings at 4 and 5 days a chance to reach the growth pattern of the others by the tenth day (Figure 4). Consequently, on the tenth day, it was not possible to obtain relevant information about the initial performance of the seeds, especially germination and seedling growth speed. It is likely that non-vigorous seedlings in the initial germination phase were able to achieve a growth pattern similar to that of vigorous seedlings after a longer germination period. The lower initial vigor of seedlings is a relevant characteristic to consider for field establishment, as less vigorous or slower-growing seedlings are more susceptible to stress conditions, if they occur, which could compromise the final stand. However, ILASTIK® allowed for the identification of lot 3 as the least vigorous and with a higher probability of producing non-vigorous seedlings and dead seeds compared to the others (Figure 4).

The application of technologies through open-source software like ILASTIK® indicates valuable solutions for various branches of scientific research involving image analysis, offering flexibility and reproducibility of results (Berg et al., 2019; Dietz et al., 2020). The use of image analysis is driving increasingly significant scientific and technological advancements in seed quality assessment. Medeiros et al. (2020a) assessed the application of machine learning with data generated by ILASTIK® software for classifying the physiological quality of soybean seeds and seedlings. They concluded that this tool is accurate, as it allowed for the identification of seed damage and the classification of seedlings in terms of vigor.

Van Tol et al. (2018), while testing software based on CellProfiler (open-source software without prior training) like ILASTIK® and MeioSeed for counting fluorescent seeds in cross-frequency analysis in Arabidopsis thaliana, found that both were efficient in facilitating the screening of new abiotic and biotic modulators of crossover frequency.

The developed models can be further improved and trained within the ILASTIK® software. This can be achieved by increasing number of images, expanding range of lots and treatments, and using all required tools available in the software. These tools include rapid classifier features, probabilistic models, and graphical interfaces that are user-friendly, allowing for quick interactive training and the extraction of precise information from bioimages (Berg et al., 2019).

The application of open-source software like SALP® and ILASTIK® for image analysis opens numerous possibilities in the field of seed quality research, allowing for the separation of seed lots into different vigor levels. Medeiros et al. (2020a) emphasized that freely accessible software like SVIS, Vigor-S, SALP®, and ImageJ also enable the evaluation of seed quality through image analysis, including seedling length data and other parameters. In this context, ILASTIK® deserves recognition as another free and efficient alternative. While it may have been less tested for various species, it has proven suitable for assessing the physiological potential of lentil seeds.

The Principal Component Analysis (PCA) explained 89.6% of the data variability, with all lots concentrated on the negative axes of Component 1 (PC1), which is related to variables representing the best physiological potential results and the probability of producing vigorous seedlings. On the other hand, it is observed that lot 3 is concentrated on the opposite side, in the positive scores of Component 1, which are directly associated with non-vigorous seedlings and dead seeds (Figure 5).

Therefore, it can be affirmed that the parameters obtained from the classification of images generated by both computerized systems, such as SAPL® and ILASTIK®, can be used in quality control programs for these seeds to classify lots according to their physiological potential.

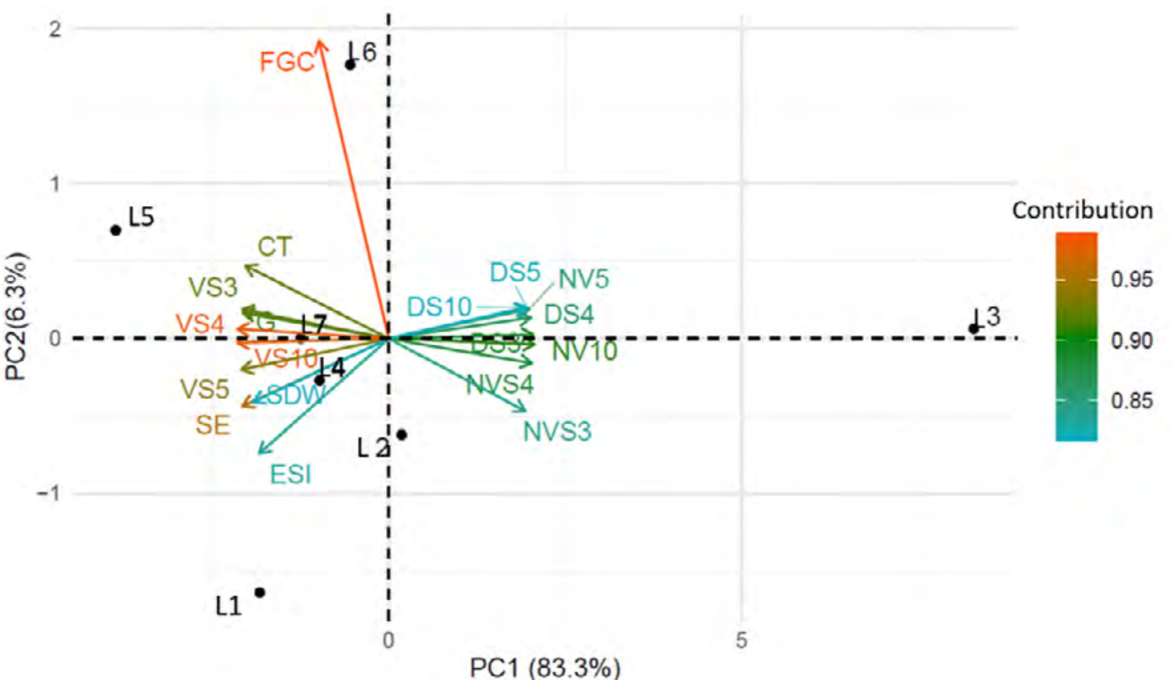

Figure 5. Principal component analysis (PCA) for the variables characterizing seed quality, including germination (G), first germination count (FGC), emergence (E), emergence speed index (ESI), seedling dry weight (SDW), and cold test (CT), and data generated by ILASTIK® software, comprising vigorous seedlings (VS), non-vigorous seedlings (NVS), or dead seeds (DS) on the 3rd, 4th, 5th, and 10th day after sowing.

The results of this study indicate a range of practical applications for this methodology, which can be implemented in seed analysis laboratories, scientific research organizations, and quality control programs. Assessments of seed physiological potential and vigor are essential for the development of modern agriculture, as these methods provide quick and precise results for the classification of seeds in terms of vigor.

CONCLUSIONS

At 4 days of germination, the total length of seedlings, primary root, and shoot, along with the vigor indices determined by SAPL®, can classify lentil lots for vigor. The data from ILASTIK® at 4 days, used in machine learning, can develop highly accurate models for seed vigor assessment, yielding results comparable to conventional methods.

ACKNOWLEDGMENTS

We would like to express our gratitude to the *Universidade Federal de Viçosa* (UFV), the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES; Finance Code: 001), and *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (FAPEMIG).

REFERENCES

ABUD, H.F.; MESQUITA, C.M.S.; SARMENTO, E.C.S.; MELO, R.S.; LIMA, K.A.P.; SILVA, A.K.F. Análise de imagens de sementes e plântulas de *Vigna radiata* L. *Revista Ciência Agronômica*, v.53, e20207303, 2022. https://doi.org/10.5935/1806-6690.20220012

AHMED, M.R.; YASMIN, J.; COLLINS, W.; CHO, B.K. X-Ray CT image analysis for morphology of muskmelon seed in relation to germination. *Biosystems Engineering*. 175, p.183–193, 2018. https://doi.org/10.1016/j.biosystemseng.2018.09.015

ANDRIAZZI, C.V.G.; ROCHA, D.K.; CUSTÓDIO, C.C. Determination of the physiological quality of corn seeds by infrared equipment. *Journal of Seed Science*, v.45, e202345002, 2023. https://doi.org/10.1590/2317-1545v45265346

ANTUNES-NETO, A.; MATIAS, F.I.; PASSOS, A.M.A.; ROCHA, N.M. Determinação de sementes de soja esverdeadas por meio de análise de imagens. In: *Embrapa Milho e Sorgo*. Sete Lagoas, MG. 2020, p.21. (Boletim de Pesquisa e Desenvolvimento, Documento 212) http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1126547

ARAÚJO, J.O.; DIAS, D.C.F.S.; MEDEIROS, A.D.; SILVA, L.J.; NASCIMENTO, W.M. Chickpea seed vigor evaluated by computerized seedling analysis. *Semina: Ciências Agrárias*, v.42, n.1, p.71-86, 2021a. https://doi.org/10.5433/1679-0359.2021v42n1p71

BARBOSA, R.M.; VIEIRA, B.G.T.L.; GOMES-JUNIOR, F.G.; VIEIRA, R.D. Image analysis and peanut seeds performance during the production process. *Científica*, v.44, n.3, p.412-420, 2016. http://dx.doi.org/10.15361/1984-5529.2016v44n3p412-420

BERG, S.; KUTRA, D.; KROEGER, T.; STRAEHLE, C.N.; KAUSLER, B.X.; HAUBOLD, C.; SCHIEGG, M.; ALES, J.; BEIER, T.; RUDY, M.; EREN, K.; CERVANTES, J.I.; XU, B.; BEUTTENMUELLER, F.; WOLNY, A.; ZHANG, C.; KOETHE, U.; HAMPRECHT, F.A.; KRESHUK, A. Ilastik: interactive machine learning for (bio) image analysis. *Nature Methods*, v.16, n.12, p.1226-1232. https://doi.org/10.1038/s41592-019-0582-9

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. *Regras para Análise de Sementes*. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Brasília: MAPA/ACS, 2009. 399p. https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise__sementes.pdf

BRASIL. Sistema de Consulta à Legislação - SISLEGIS. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. *PORTARIA Nº 111*, 2012.

BRUNES, A.P.; ARAÚJO, A.D.; DIAS, L.K.; VILLELA, F.A.; AUMONDE, T.Z. Seedling length in wheat determined by image processing using mathematical tools. *Revista Ciência Agronômica*, v.47, n.2, p.374-379, 2016. https://doi.org/10.5935/1806-6690.20160044

CASTAN, D.O.C.; GOMES-JUNIOR, F.G.; MARCOS-FILHO, J. Vigor-S, a new system for evaluating the physiological potential of maize seeds. *Scientia Agricola,* v.5, n.2, p.167-172, 2018. http://dx.doi.org/10.1590/1678-992X-2016-0401

CICERO, S.M.; VIEIRA, R.D. Teste de frio. In: KRZYZANOWSKI, F. C.; VIEIRA, R. D.; MARCOS-FILHO, J.; FRANÇA NETO, J. B. (Eds.). *Vigor de sementes: conceitos e testes*. Londrina, PR: ABRATES. 2020. 280-307p.

COLLINS, T.J. ImageJ for microscopy. *BioTechniques*, v.43, n.1, p.25–30, 2007. https://doi.org/10.2144/000112517

DIETZ, C.; RUEDEN, C.T.; HELFRICH, S.; DOBSON, E.T.A.; HORN, M.; EGLINGER, J.; EVANS, E.L.; MCLEE, D.T.; NOVITSKAYA, T.; RICKE, W.A.; SHERER, N.M.; ZIJLSTRA, A.; BERTHOOLD, M.R.; ELICEIRI, K.W. Integration of the ImageJ ecosystem in the KNIME analytics platform. *Frontiers of Computer Science,* v.2, n.8, 2020. https://doi.org/10.3389/fcomp.2020.00008

ELMASRY, G.; MANDOUR, N.; WAGNER, M.H.; DEMILLY, D., VERDIER, J., BELIN, E., ROUSSEAU, D. Utilization of computer vision and multispectral imaging techniques for classification of cowpea (*Vigna unguiculata*) seeds. *Plant Methods,* v.15, n.24, 2019. https://doi.org/10.1186/s13007-019-0411-2

FINCH-SAVAGE, W.E.; BASSEL, G.W. Seed vigour and crop establishment: extending performance beyond adaptation. *Journal of Experimental Botany*, v.67, n.3, p.567–591, 2016. https://doi.org/10.1093/jxb/erv490

GOMES-JUNIOR, F.G. Análise computadorizada de imagens de plântulas. In: KRZYZANOWSKI, F. C.; VIEIRA, R. D.; FRANÇA-NETO, J. B.; MARCOS-FILHO, J. (Eds.). *Vigor de sementes: conceitos e testes*. Londrina, PR: ABRATES, 2020. 139-181p.

GOMES-JUNIOR, F.G.; CHAMMA, H.M.C.P.; CICERO, S.M. Automated image analysis of seedlings for vigor evaluation of common bean seeds. *Acta Scientiarum. Agronomy*, v.36, n.2, p.195-200, 2014. https://doi.org/10.4025/actasciagron.v36i2.21957

KRZYZANOWSKI, F.C.; FRANÇA-NETO, J.B.; GOMES-JUNIOR, F.G.; NAKAGAWA, J. Testes de vigor baseados em desempenho de plântulas. In: KRZYZANOWSKI, F. C.; VIEIRA, R. D.; MARCOS-FILHO, J.; FRANÇA NETO, J. B. (Eds.). *Vigor de sementes: conceitos e testes*. Londrina, PR: ABRATES, 2020. 80-126p.

MAGUIRE, J.D. Speed of germination - Aid in selection and evaluation for seedling emergence and vigor. *Crop Science*, v.2, n.2, p.176, 1962. https://acsess.onlinelibrary.wiley.com/doi/abs/10.2135/cropsci1962.0011183X000200020033x

MARCOS-FILHO, J. *Fisiologia de sementes de plantas cultivadas*. Piracicaba: FEALQ, 2015. 660 p.

MEDEIROS, A.D.; PEREIRA, M.D. SAPL®: a free software for determining the physiological potential in soybean seeds. *Pesquisa Agropecuária Tropical*, v.48, n.(3), p.222-228, 2018. https://doi.org/10.1590/1983-40632018v4852340

MEDEIROS, A.D.; PINHEIRO, D.T.; XAVIER, W.A.; SILVA, L.J.; DIAS, D. C. F. S. Classificação de qualidade de sementes de pinhão manso utilizando imagens radiográficas e aprendizado de máquina. *Industrial Crops and Production*, v.146, e112162, 2020b. https://doi.org/10.1016/j.indcrop.2020.112162

MEDEIROS, A.D.; CAPOBIANGO, N.P.; SILVA, J.M.; SILVA, L.J.; SILVA, C.B.; DIAS, D.C.F.S. Interactive machine learning for soybean seed and seedling quality classification. *Scientific Reports*, v.10, n.1, 11267, 2020a. https://doi.org/10.1038/s41598-020-68273-y

MEDEIROS, A.D.; SILVA, L.J.; CAPOBIANGO, N.P.; FIALHO, C.A.; DIAS, D.C.F.S. Assessing the physiological quality of common bean seeds using the Vigor-S® system and its relation to the accelerated aging test. *Journal of Seed Science*, v.41, n.2, p.187-195, 2019. https://doi.org/10.1590/2317-1545v41n2211401

MEDEIROS, M. L.S.; CRUZ-TIRADO, J.P.; LIMA, A.F.; SOUZA-NETTO, J.M.; RIBEIRO, A.P.B.; BASSEGIO, D.; GODOY, H.T.; BARBIN, D.F. Assessment oil composition and species discrimination of *Brassicas* seeds based on hyperspectral imaging and portable near infrared (NIR) spectroscopy tools and chemometrics. *Journal of Food Composition and Analysis*, v.107, 104403, 2022. https://doi.org/10.1016/j.jfca.2022.104403

PEDÓ, T.; MARTINAZZO, E.G.; BACARIN, M.A.; ANTUNES, I.F.; KOCH, F.; MONTEIRO, M.A.; PIMENTEL, J.R.; TROYJACK, C.; VILLELA, F.A.; AUMONDE, T.Z. Crescimento de plantas e vigor de sementes de feijão em resposta à aplicação exógena de ácido giberélico. *Revista de Ciências Agrárias*, v.41, n.3, p.757-770, 2018. https://doi.org/10.19084/RCA17169

PEREIRA, M.D.; REIS, J.A.; FERRARI, C.S.; VALE, A.M.P.G. Processamento digital de imagens de plântulas na avaliação do vigor de sementes de *Moringa oleifera* Lam. *Ciência Florestal*, v.30, n.2, p.291, 2020. https://doi.org/10.5902/1980509825750

QUEIRÓZ, T.N.; VALIGUZSKI, A.L.; BRAGA, C.S.; SOUZA, S.A.M.; DA ROCHA, A.M. Avaliação da qualidade fisiológica de sementes de variedades tradicionais de milho. *Revista da Universidade Vale do Rio Verde***,** v.17, n.1, 2019. http://dx.doi.org/10.5892/ruvrd.v17i1.5130

R CORE TEAM. *R. A Language and Environment for Statistical Computing*, 2022.

RIBEIRO, A.S.; NERI, T.F.S.; MEDEIROS, A.D.; MILAGRES, C.C.; SILVA, L.J. Use of radiographic images for rapid and non-destructive assessment of crambe seed quality. *Journal of Seed Science*, v.43, p.43-55, 2021. https://doi.org/10.1590/2317-1545v43239136

SAKO, Y.; MCDONALD, M.B.; FUJIMURA, K.; EVANS, A.F.; BENNETT, M.A. A system for automated seed vigour assessment. *Seed Science and Technology*, v.29, n.3, p.625-636, 2001. Recuperado em https://www.eurofinsus.com/media/162083/seed-vigor-imaging-system.pdf

SILVA, L.J.D.; MEDEIROS, A.D.D.; OLIVEIRA, A.M.S. SeedCalc, a new automated R software tool for germination and seedling length data processing. *Journal of Seed Science*, v.41, n.2, p.250-257, 2019. https://doi.org/10.1590/2317-1545v42n2217267

SILVA, P.P.; BARROS, A.C.S.A.; MARCOS-FILHO, J.; GOMES-JUNIOR, F.G.; NASCIMENTO, W.M. Assessment of squash seed vigor using computerized image analysis. *Journal of Seed Science*, v.39, n.2, p.159-165, 2017. https://doi.org/10.1590/2317-1545v39n2171177

VAN TOL, N.; ROLLOOS, M.; VAN LOON, P.; ZAAL, B.J.V.D. MeioSeed: a cellprofiler-based program to count fluorescent seeds for crossover frequency analysis in *Arabidopsis thaliana*. *Plant Methods*, v.14, n.1, p.14-32, 2018. https://doi.org/10.1186/s13007-018-0298-3

WORMA, M.; SEGATTO, C.; STEFEN, C.; BUBA, G.P.; LEOLATO, L.S. Qualidade fisiológica de sementes de milho produzidas com adubação biológica e bioestimulante em diferentes preparos de solo. *Revista Engenharia na Agricultura - Reveng*, v.27, n.3, p.187-194, 2019. https://doi.org/10.13083/reveng.v27i3.893

XIA, Y.; XU, Y.; LI, J.; ZHANG, C.; FAN, S. Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, v.1, p.35–47, 2019. https://doi.org/10.1016/j.aiia.2019.05.001