Revista Ciência Agronômica, v. 56, e202493547, 2025
Centro de Ciências Agrárias - Universidade Federal do Ceará, Fortaleza, CE
www.ccarevista.ufc.br

Scientific Article
ISSN 1806-6690

Near-infrared spectroscopy and accelerated ageing in evaluating the vigour of lentil seeds[1]

Marcelo Augusto Rocha Limão[2]*, Denise Cunha Fernandes dos Santos Dias[2], Warley Marcos Nascimento[3], Bruno Gomes de Noronha[2], Júlia Martins Soares[2], Laércio Junio da Silva[2]

ABSTRACT - The cultivation and consumption of lentils has been gaining importance in recent years due to the high nutritional value of the grain. Research related to seed technology is therefore important in providing high-quality seeds for the market. The aim of this study was to adapt the accelerated ageing test and evaluate the potential of FT-NIR spectroscopy for classifying lentil seeds based on physiological potential. Seven batches of lentil seeds were subjected to tests to characterise their physiological potential. The accelerated ageing test included the traditional method (100% RH) with an alternative method using saturated saline solution (76% RH) at 41 ºC for 24, 36, 48 and 72 hours. NIR spectra were also obtained, taking 200 spectral readings from each batch that were individually processed for 30 seconds. Following the spectral analysis, the seeds were submitted to tests of germination and accelerated ageing to validate the technique. The PLS-DA technique was employed, using 70% of the data for training and 30% for validation. Different pre-processing methods were used, including the standard normal variate (SNV), multiplicative scatter correction (MSC), and the 1st and 2nd Savitzky-Golay (SG) derivatives. It was concluded that the accelerated ageing test at 41 °C for 48 hours using the traditional method (100% RH) was the most efficient way of evaluating the physiological potential of the lentil seeds. The models derived from the FT-NIR spectral data were 99% accurate in predicting the class of physiological potential of the batches.

Key words: *Lens culinaris* Medik. Physiological quality. Deterioration. Artificial intelligence. FT-NIR.

DOI: 10.5935/1806-6690.20250034
Editor-in-Chief: Prof. Salvador Barros Torres - sbtorres@ufersa.edu.br
*Author for correspondence
Received for publication 11/06/2024; approved on 08/08/2024
[1]Article extracted from part of the dissertation of the lead author, presented to the Postgraduate Course in Phytotechnics – Federal University of Viçosa (UFV), Viçosa Campus, Viçosa-MG, Brazil
[2]Postgraduate Course in Phytotechnics, Federal University of Viçosa (UFV), Viçosa Campus, Viçosa-MG, Brazil, marceloliimao@gmail.com (ORCID ID 0000-0003-1265-1014), dcdias@ufv.br (ORCID ID 0000-0002-0596-2490), bruno.g.noronha@ufv.br (ORCID ID 0000-0002-3245-5312), julia.m.soares@ufv.br (ORCID ID 0000-0001-9432-6009), laercio.silva@ufv.br (0000-0001-7202-0420)
[3]Embrapa Hortaliças, Caixa Postal 280 – Brasília-DF, Brazil, warley.nascimento@embrapa.br (ORCID ID 0000-0002-6235-0917)

INTRODUCTION

Pulses, the edible dried seeds of legumes such as beans, peas, chickpeas, and lentils, play a prominent role in human nutrition due to their high protein value (Nascimento; Silva, 2019). The cultivation of lentils (*Lens culinaris* Medik.), which are among the five most important legumes in the world, is on the increase in Brazil (Nascimento, 2019). To consolidate this market, it is important to carry out research into technologies that allow the identification of seed batches with greater potential for storage and establishment in the field.

The accelerated ageing test has been shown to be efficient in this context, with results that correlate with seedling emergence in the field and storage potential (Marcos-Filho, 2020); while near-infrared spectroscopy, which is a faster and non-destructive method, is promising for selecting batches throughout the production process (Souza *et al*., 2023).

In the accelerated ageing test, the rate of seed deterioration increases as the seeds are exposed to conditions of high temperature and relative humidity (Marcos-Filho, 2020). Seeds with high physiological potential have a greater capacity to produce normal seedlings following such stress (Araújo *et al*., 2021).

The combination of temperature, relative humidity and period of exposure is an important factor in defining the appropriate methodology for the accelerated ageing test. The traditional method uses water to obtain a relative humidity of 100% (Marcos-Filho, 2020), while the alternative method uses a saturated saline solution (Jianhua; Mcdonald, 1996), for instance KCl (potassium chloride) or NaCl (sodium chloride), to obtain a relative humidity of 87% or 76%, respectively. This method results in slower seed hydration and is mainly used for small seeds; it can therefore be used to test lentil seeds, which are small and present rapid imbibition (Freitas; Nascimento, 2006). For the chickpea, the most appropriate method employed NaCl (76% RH) at 41 ºC/48 h (Araújo *et al*., 2021); for the soya bean (Marcos-Filho, 2020) and common bean (Bertolin; Sá; Moreira, 2011), the test should be conducted under 100% RH at 41 ºC/48 h, while wheat (Favoretto *et al*., 2024) should be tested under 100% RH at 43 °C/48 h.

Near-infrared (NIR) spectroscopy is a promising, non-destructive, high-performance technique that can be applied to seed quality. It is based on the absorption of electromagnetic radiation at wavelengths in the range of 780-2,500 nm, generating spectra related to the carbohydrate, protein, and lipid content (Xia *et al*., 2019).

From these spectra, it is possible to identify the wavelength peaks of chemical compounds that change according to the quality of the seeds (Medeiros *et al*., 2020). Considering that deterioration of the seed alters the biochemical composition and organisation of the cellular structures, this technique has been used to predict seed quality in various species, such as the chickpea (Ribeiro *et al*., 2020), soya bean (Soares *et al*., 2024), wheat (Xia *et al*., 2019), maize (Qiu *et al*., 2018) and brachiaria (Souza *et al*., 2023).

However, to obtain consistent results using FT-NIR, it is necessary to process the data correctly and select methods of classification that allow the relationship between the spectral data and the biochemical composition of the seeds to be modelled (Burns; Ciurczak, 2007).

In this respect, the aim of this study was to adapt the accelerated ageing test and evaluate the potential of FT-NIR spectroscopy to classify lentil seeds based on physiological potential.

MATERIAL AND METHODS

The study was conducted at the Seed Research Laboratory of the Department of Agronomy at the Federal University of Viçosa, in Viçosa, Minas Gerais, Brazil. Seven batches of lentil seeds 'Silvina', were tested for moisture content (MC) in an oven at 105 °C for 24 hours (Brasil, 2009), with the results expressed as a percentage. The batches were then characterised for their physiological potential using the following tests:

Germination (G) and *first germination count (FGC)*: four replications of 50 seeds were spread on paper towels moistened with a volume of water equal to 2.5 times the weight of the dry paper. The towels were rolled up and placed in a germinator at 20 °C. For the first germination count, the percentage of normal seedlings on the fifth day after sowing was calculated, while percentage germination was determined 10 days after sowing (Brasil, 2009).

Seedling emergence (E) and *emergence speed index (ESI)*: conducted in a growth chamber with four replications of 50 seeds that were sown at a depth of 1.0 cm in plastic trays containing a mixture of soil and sand at a ratio of 1:2 and moistened to 60% of retention capacity (Brasil, 2009). The number of emerged seedlings was counted daily until the stand had completely stabilised; percentage emergence and the ESI were calculated as per Maguire (1962).

Seedling dry weight (DW): determined in seedlings obtained at the end of the emergence test described above. The cotyledons were removed, and the seedlings placed in an air circulation oven at 65 °C for 72 h. After drying, they were weighed (0.0001g), with the results expressed in mg seedling^{-1} (Krzysanowski *et al*., 2020).

Cold test (CT): carried out in a similar way to the germination test, covering the seeds with 60 mL of soil after sowing. The rolls were placed in plastic bags and kept in a biochemical oxygen demand (BOD) chamber at 10 °C for seven days. They were then removed from the plastic bags and kept in a germinator at 20 °C for a further five days to evaluate the percentage of normal seedlings (Cicero; Vieira, 2020).

Experimental design and statistical analysis

A completely randomised design (CRD) was used, with seven batches and four replications. The data were tested for a normal error distribution using the Shapiro-Wilk test and homogeneity of variance using Bartlett's test. They were then submitted to analysis of variance, and the mean values for each batch were compared by Tukey's test at a level of 5%.

Trial I - Adapting the accelerated ageing test to evaluate the physiological potential of lentil seeds

Two procedures were adopted, as described by Baalbaki *et al*. (2009) and Marcos-Filho (2020):

Traditional accelerated ageing – TAA: *using* 250 seeds from each treatment, distributed on a wire screen attached to a Gerbox plastic box containing 40 mL of water at the bottom. The boxes were closed to obtain an internal RH of 100% and kept in a BOD chamber at 41 °C for 24, 36, 48 and 72 h. The germination test was then carried out as described above to determine the percentage of normal seedlings five days after sowing.

Accelerated ageing with saturated saline solution – AASSS – conducted in a similar way to the TAA test, replacing the water with 40 mL of saturated NaCl solution (40 g NaCl/100 mL of water), in order to obtain a RH of 76% (Jianhua; Mcdonald, 1996). The results were expressed as percentage of normal seedlings on the fifth day after sowing.

Experimental design and statistical analysis

The trial was conducted in a completely randomised design (CRD) and analysed using a triple factorial scheme: 2 (ageing methods) x 7 (batches) x 4 (ageing periods). The data were tested for normal error distribution as described above, and then submitted to analysis of variance. The mean values obtained for each batch were compared by Tukey's test at a level of 5%, while the ageing methods were compared by F-test at a level of 5%. The ageing periods were submitted to regression analysis. Each of the analyses was processed using the R 4.1.1 statistical software.

Trial II - Near-infrared (NIR) spectroscopy and its relationship with the physiological potential of the seeds

Acquisition of FT-NIR spectra: seven batches of 200 randomly selected seeds were used, to give a total of 1400 seeds, which were individually placed in a circular metal support of the correct size and shape for lentils to prevent the light beams from passing through the sample, using a duration of 30 seconds for each seed. The spectral data from each seed were obtained using the Fourier-transform near infrared (FT-NIR) spectrometer (Thermo Scientific Antaris II). The reflectance spectra were expressed as log (1/R), where R is the reflectance. For each seed, 3111 points were collected per spectrum, at wavelengths ranging from 1,000 to 2,500 nm.

Germination: following acquisition of the spectra, the properly identified seeds were submitted to the germination test, as described in Trial 1. The percentage of vigorous seedlings (greater than 7 cm), non-vigorous seedlings (less than 7 cm) and dead seeds were determined eight days after sowing (Brasil, 2009).

Protein content of the seeds: determined as per Bradford (1976), using bovine serum albumin (BSA) as the standard. To obtain the crude enzyme extract and determine the protein, 0.1 g of the plant material was separated, and 2 mL of extraction medium, potassium phosphate buffer (0.1 M, pH 6.8), 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1% polyvinylpyrrolidone (PVPP) (w/v) were added (Chaffai; Marzouk; Ferjani, 2005). The mixture was centrifuged at 15,000 x g for 15 minutes at 4 °C to remove the oil layer from the supernatant. The protein content of the enzyme extracts was determined as per Bradford (1976) using BSA as the standard. Fifty µl of the enzyme extract was added to 1.5 ml of the Bradford reagent. After stirring for 20 minutes, the absorbance was read at a wavelength of 595 nm.

Preprocessing algorithms

The original spectral data were pre-processed using the following methods of scatter correction: Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), 1st Savitzky-Golay derivative (SG) and 2nd Savitzky-Golay derivative using a window size of 11 wavelength variables.

Development of the classification models

The batches were classified as of high physiological potential (HPP) - with percentage germination and vigorous seedlings greater than 95%, medium physiological potential (MPP) - with percentage germination and vigorous seedlings from 85% to 94%, and low physiological potential (LPP) – with a percentage of less than 84%. The number of batches per class included 600 spectra for HPP, 600 for MPP, and 200 for LPP.

Model validation

The models were built using 70% of the data for training and the remaining 30% for validation. In addition, five-fold cross-validation was performed for each model in the training set. The set with the highest hit rate was used to predict the performance of the batches in the validation set. The models were evaluated for accuracy and the *kappa* coefficient, i.e. by the percentage of batches correctly classified in the validation set.

Experimental design and statistical analysis

A completely randomised design with seven batches and four replications was used to test for germination and protein content. The data were tested for a normal error distribution using the Shapiro-Wilk test and homogeneity of variance using Bartlett's test. They were then submitted to analysis of variance, and the mean values for each batch were compared by Tukey's test at a level of 5%.

The NIR spectra were assigned classes based on the physiological performance of the batch established by the tests for vigour. An exploratory analysis was carried out using the original spectra and the mean value of each class, and a classification model was generated using PLS-DA. The performance of the models was evaluated for accuracy and the *kappa* coefficient. In addition, the wavelength ranges that were most important in constructing the classification model were identified for each vigour class. PLS-DA was carried out using the caret package in the R software.

RESULTS AND DISCUSSION

The moisture content of the seeds from the different lentil batches was uniform, ranging from 13.1% to 13.4% (Table 1). This characteristic is important as standardising the seed moisture content is essential for consistent results (Marcos-Filho *et al*., 2020; Worma *et al*., 2019).

Seed germination in Batch 5 was higher than in the other batches, followed by Batches 1, 2, 4, 6 and 7, while Batch 3 stood out as having the lowest value (Table 1). The germination test allows seeds to express their maximum physiological potential, since it affords optimal environmental conditions (Queiróz *et al*., 2019). Using the first germination count, the highest percentages were obtained for Batches 6 and 5, followed by Batches 4 and 7, with lower values for Batches 1 and 3, albeit not differing from Batch 2. This test provides an indication of the speed of germination of the batches (Krzysanowski *et al*., 2020). As such, any difference in the performance of Batches 1, 2, 4, 6 and 7 that was not found by the germination test would be detected when the speed of germination was evaluated (Table 1).

The results for seedling emergence in relation to classifying the physiological potential of the batches were similar to those of the emergence speed index (ESI), with Batches 1, 4, 5 and 7 being superior, followed by Batches 2 and 6, considered intermediate, and Batch 3 showing the worst performance (Table 1). These tests allowed an even more detailed classification of the batches in terms of vigour.

The highest value for seedling dry weight was obtained for Batch 5 and the lowest for Batch 3, with intermediate values for Batches 4 and 6 (Table 1). Batches whose seedlings have a higher dry matter content present the greatest vigour, indicating more dry matter being transferred from the seed reserve tissue to the embryonic axis (Krzysanowski *et al*., 2020; Pedó *et al*., 2018). Batch 5 also performed better in the cold test together with Batch 6, while Batches 3 and 7 had the worst performance.

Table 1 - Characterisation of the physiological potential of seven batches of lentil seeds 'Silvina': moisture content (MC), germination (G), first germination count (FGC), emergence (E), emergence speed index (ESI), seedling dry weight (DW) and cold test (CT)

Characterisation of initial physiological quality							
Batch	MC (%)	G (%)	FGC (%)	E (%)	ESI (index)	DW (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 by F-test at 5% probability; F = calculated F value; CV = coefficient of variation. Mean values followed by the same letter in a column do not differ by Tukey's test at a level of 5%

In general, in each of the tests, Batch 5 was among those showing the best performance, while Batch 3 had the worst performance, with some variation for the batches of medium vigour, depending on the test. The use of batches with different levels of physiological potential is essential in studies that evaluate seed vigour, especially when the goal is to define new methodologies.

Trial I - Adjusting the accelerated ageing test for evaluating the physiological potential of lentil seeds

It can be seen that the TAA method (100% RH) grouped the batches into different levels of vigour for the four periods under evaluation (Table 2). After 24 hours, Batch 5 showed the highest vigour, but did not differ from Batch 1. Batch 3, on the other hand, was inferior to the other batches, with Batches 2, 4, 6 and 7 in an intermediate position, but also not differing from Batch 1. However, the clearest difference was seen after 36 and particularly 48 hours, after which Batch 5 was classified as superior, followed by Batches 1 and 7, which were superior to Batches 2, 4, 6 and 3, with Batch 3 highlighted as inferior to the others. After 72 hours, the values obtained were far lower than for the other periods, showing that deterioration was more drastic, and contributed to Batch 2 being equal to that with the lowest vigour (Batch 3). It was also found that in lentil seeds (Freitas; Nascimento, 2006; Marinke *et al*., 2019) and chickpea seeds (Araújo *et al*., 2021) the 72-hour ageing period gave very drastic results for seeds from different batches.

These results were similar to those obtained in the other tests used in the initial characterisation (Table 1) of the seed batches for each of the periods under test, allowing them to be separated into different levels of vigour. However, it should be noted that although the periods of 36 and 48 hours allowed the batches to be separated in terms of vigour, the classification obtained after 48 hours was clearer, allowing greater stratification; this period also has the advantage of being more practical than the 36-hour period due to the time required for evaluating the test, which would be outside the business hours of routine laboratories.

The accelerated ageing test makes it possible to evaluate the physiological potential of seeds based on the principle that the rate of seed deterioration increases when the seeds are subjected to conditions of high temperature and relative humidity (Baalbaki *et al*., 2009; Marcos-Filho; Kikuti; Lima, 2009). Therefore, unlike low-vigour batches, high-vigour batches are tolerant, and are able to germinate and produce normal seedlings when subjected to the test conditions.

With the AASSS, conducted at a lower relative humidity than that used in the TAA to reduce stress, the batches were separated into only three levels of vigour (Table 2), meaning it was less efficient than the traditional method. It should be noted that Batch 5, classified as more vigorous by the traditional method, did not differ significantly from Batches 1, 4 or 7 for most of the periods under test. A more detailed classification of the batches was only possible after 72 hours, with Batches 1 and 5 as the most vigorous and Batch 3 showing the worst performance, with the other batches presenting intermediate results (Table 2).

TAA therefore more efficiently classified the batches in terms of vigour from 36 hours onwards, but particularly after 48 hours, when separating the batches into levels of vigour became more stratified. This can

Table 2 - Germination (%) in batches of lentil seeds after different periods of traditional accelerated ageing (TAA) and accelerated ageing using saturated saline solution (AASSS)

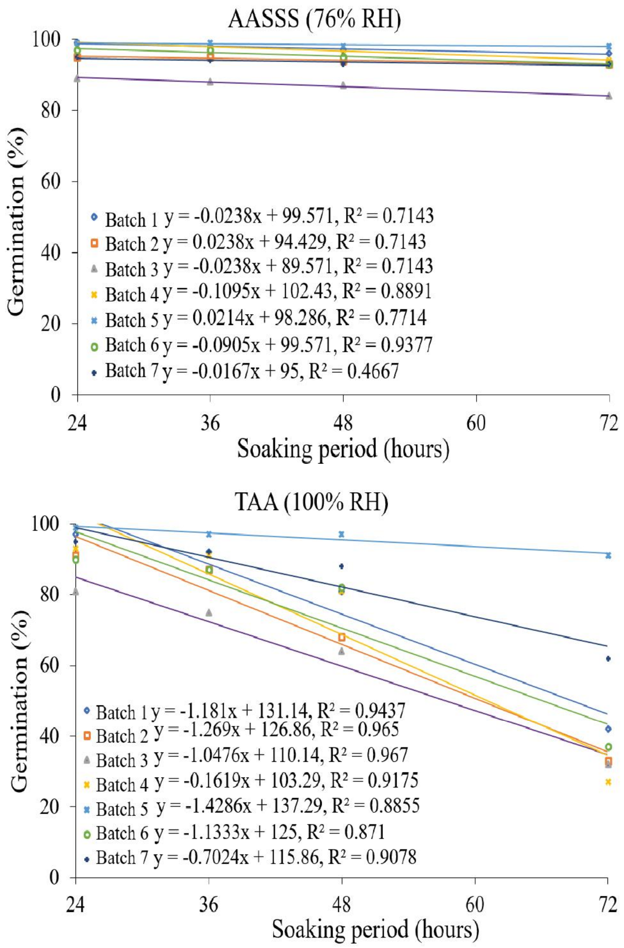
Batch	TAA (100% RH)				AASSS (76% RH)			
	24 h	36 h	48 h	72 h	24 h	36 h	48 h	72 h
1	97 abA	92 bB	85 bB	42 cB	99 aA	98 abA	97 abA	96 abA
2	91 bcB	87 cB	68 dB	33 deB	94 bA	95 bA	94 bA	93 bA
3	81 dB	75 dB	64 eB	32 eB	89 cA	88 cA	87 cA	84 cA
4	93 bcB	90 bcB	81 cB	37 dB	99 aA	98 abA	97 abA	94 bA
5	99 aB	97 aB	97 aB	91 aB	99 aA	99 aA	98 abA	98 aA
6	90 bcB	87 cB	82 cB	37 dB	97 abA	94 bA	93 bA	93 bA
7	95 bB	92 bB	88 bB	62 bB	98 aA	97 abA	96 abA	93 bA
CV (%)					2.1			

Mean values followed by the same uppercase letters in a row comparing methods within each period and by the same lowercase letters in a column do not differ by Tukey’s test at a level of 5%

be confirmed by comparing the different periods of accelerated ageing (Figure 1). For TAA, there was a marked linear reduction in the values obtained with the increase in ageing time, allowing greater stratification of the batches after 48 hours, with Batch 5 as the most vigorous followed by Batch 7, and Batch 3 with the lowest vigour. On the other hand, when using AASSS, separating the batches in terms of vigour showed poor stratification, with only Batch 3 standing out as having the worst performance in each of the periods under test. It can also be seen that there was no marked reduction in germination following AASSS, resulting in less moisture stress and less deterioration compared to TAA; this was also found by Araújo *et al*. (2021) in chickpea seeds.

The accelerated ageing test using the traditional method has been conducted at a temperature of 41 ºC for 48 hours in both the soya bean and common bean

Figure 1 - Germination percentage in seven batches of lentil seeds 'Silvina' following traditional accelerated ageing with water (TAA - 100% RH) and accelerated ageing with saturated saline solution (AASSS - 76% RH) for periods of 24, 36, 48 and 72 h



(Baalbaki *et al*., 2009). However, for chickpea seeds under the same conditions, there was a high incidence of microorganisms and seed death, indicating excessive deterioration mainly due to the high relative humidity (Araújo *et al*., 2021), so the alternative method at an RH of 76% for 48 hours is more appropriate. This procedure was also efficient in seeds of the green mung bean (Silva; Oliveira; Ferreira, 2018).

Trial II - Near-infrared (NIR) spectroscopy and its relationship with the physiological potential of the seeds

Germination data obtained with the seeds subjected to FT-NIR spectroscopy showed a significant difference between the batches, albeit Batches 1, 5 and 7 did not differ from each other and had a higher rate of germination than the others; these were followed by Batches 2, 4 and 6, with germination values that were similar to and higher than Batch 3, which had the worst performance (Figure 2a). Similar results were obtained for the seed protein content (Figure 2b). When evaluating the percentage of vigorous seedlings (Figure 2a), it was found that Batch 5 stood out from the other batches, with better performance and the lowest percentage of non-vigorous seedlings (Figure 2a) and dead seeds (Figure 2a). Batches 1, 2, 4, 6 and 7 were inferior to Batch 5, but superior to Batch 3, which maintained the worst performance found in the germination test, with a high proportion of dead seeds. When assessing seedling vigour, it was found that Batches 1, 5 and 7, which did not differ in terms of germination, differed in terms of vigour, which was greater in Batch 5. These results are consistent with those obtained with TAA after 48 hours (Figure 2c), where Batch 5 was classified as superior to the others, followed by Batches 1 and 7, with Batch 3 standing out as inferior.

High germination and seedling vigour are essential characteristics to ensure proper establishment in the field and, consequently, uniformity of the stand and plants (Finch-Savage; Bassel, 2016). Seedling vigour is an important characteristic that can be evaluated during the germination test without the need for additional tests or more complex methods that require more time to obtain the results.

In this study, germination, seedling vigour and the seed protein content were the characteristics used to test the efficiency of FT-NIR spectroscopy in classifying the batches according to their physiological potential, and act as efficiency indicators of the applicability of the models under test.

The original 1400 spectral readings of the seeds from the seven batches and the average for each class are shown in Figures 3A and 3B, respectively.

Preprocessing is essential to optimise the accuracy and *kappa* coefficients of the model based on the FT-NIR spectra, since the spectra result in various combinations of tones and overtones, causing noise and masking overlapping results (Wu *et al*., 2009). Applying pre-processing techniques therefore reduces noise and optimises the choice of the model, to extract efficient and accurate classification results (Ribeiro *et al*., 2020). In this context, several scatter correction methods were

Figure 2 - Germination (G), vigorous seedlings (VS), non-vigorous seedlings (NVS), dead seeds (DS) (a), protein content (b) and TAA 48 h (c) of seeds from seven lentil batches. Mean values with different lowercase letters differ by Tukey's test ($p \leq 0.05$)

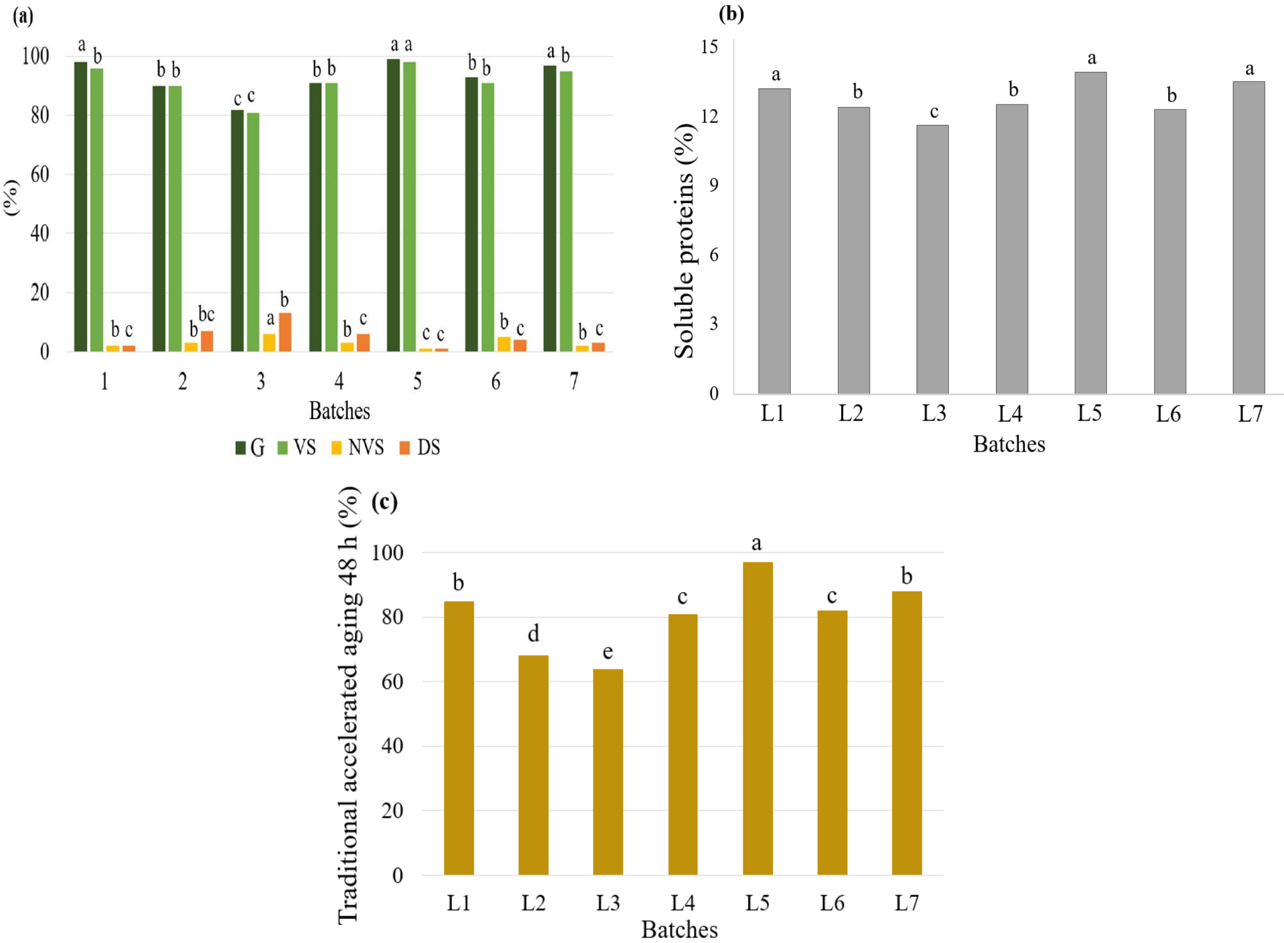
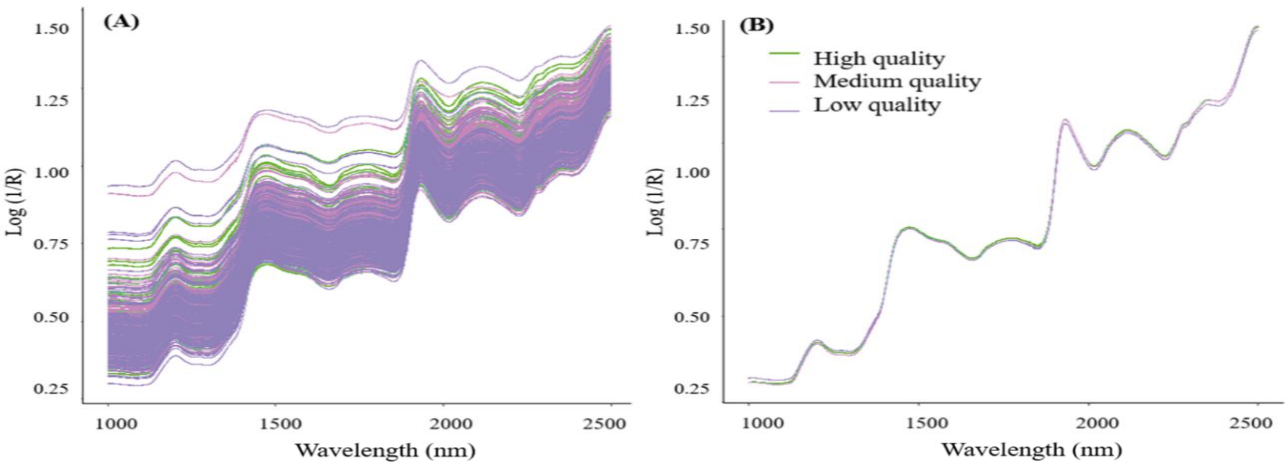


Figure 3 - Original spectra (A) and average for the original spectra (B) per vigour class of the lentil seeds



tested, such as multiplicative scatter correction (MSC) or standard normal variate (SNV) and derivative methods (Savitzky-Golay) that reduce peak overlap while smoothing out the signal (Agelet; Hurburgh, 2014). The most efficient and recommended pre-processing to obtain the calibration model was defined based on the ability to predict classes of physiological quality based on accuracy and the *kappa* coefficient.

In this study, the best results were obtained by applying the 2nd Savitzky-Golay derivative. Although the original spectra resulted in an accuracy of 86% and *kappa* of 78% for validation, which is already a very good result, maximum values of 99% for accuracy and *kappa* were reached using the 2nd Savitzky-Golay derivative, further validating and optimising the classification model. Landis and Koch (1977) state that *kappa* values close to 1.0 establish an almost perfect agreement, ensuring that pre-processing was efficient in separating the different seed classes (Table 3).

According to results obtained in studies with seeds of the chickpea (Ribeiro *et al*., 2020), soya bean (Soares *et al*., 2024), common bean (Yang *et al*., 2015), wheat (Xia *et al*., 2019) and brachiaria (Souza *et al*., 2023), pre-processing is essential for optimising accuracy and the results of the *kappa* coefficient, and extracting accurate and efficient classification information.

The PLS-DA classification model with pre-processed spectra by means of the 2nd Savitzky-Golay derivative was confirmed as the most efficient for classifying vigour in lentil seeds (Table 3). The use of PLS-DA with the 2nd Savitzky-Golay derivative was also efficient in the seeds of brachiaria, resulting in the best classification models (Souza *et al*., 2023). The *kappa* coefficient is important in determining the ideal classification model as it indicates the agreement and reliability of the model (Cohen, 1968). Ribeiro *et al*. (2020) found that PLS-DA models developed using FT-NIR spectral data were efficient in identifying differences in the quality of chickpea seeds subjected to the pre-harvest application of different herbicides.

Figure 4 shows the confusion matrix relating to the efficiency and performance of the classification models from the FT-NIR spectra by means of the 2nd Savitzky-Golay derivative. The model based on the second-order derivative using the Savitzky-Golay filter proved to be more efficient, with a hit rate for high-, medium- and low-vigour seeds of 99%, 98% and 98%, respectively, in the validation. It should be noted that the rows and columns of the confusion matrix display the distribution of the seeds in their actual and predicted classes, respectively. The confusion matrix of the data after processing with the 2nd Savitzky-Golay derivative shows that there were virtually no prediction errors for the model. Only a small error distribution can be seen for the medium vigour class when compared to the high and low vigour classes in the validation, and these errors may result from the additive or multiplicative effects of the spectra.

Figure 5 shows the peaks of importance in the wavelength ranges corresponding to the functional groups with the highest contribution to classifying the seed batches into high, medium and low physiological potential. In the PLS-DA classification, the most representative ranges were 1000 - 1250 nm, 1400 nm, 1900 nm, 2275 - 2400 nm and 2500 nm.

These peaks are related to the functional groups and to the water, protein, oil and carbohydrate content, which are broad and overlap throughout the spectral range. For example, the most important wavelengths at the peaks of 1060 - 1250 and 1400 nm are related to the functional groups N-H, C-H, and O-H, and are associated with the protein and carbohydrate content of the seeds (Ambrose *et al*., 2016; Mukasa *et al*., 2019). Although the peaks at 1900 nm were less marked

Table 3 - Values for accuracy and the *kappa* coefficient in the training and validation (test) of the different pre-processing methods, obtained with the classification models via PLS-DA for batches of lentil seeds

Pre-processing	Calibration (n = 980)		Validation (n = 420)	
	Accuracy	Kappa	Accuracy	Kappa
Original spectra	0.8652	0.7762	0.8732	0.7890
SNV	0.8948	0.8264	0.8995	0.8343
MSC	0.8846	0.8089	0.8828	0.8062
1st SG derivative	0.9796	0.9667	0.9713	0.9530
2nd SG derivative	0.9999	0.9999	0.9856	0.9765

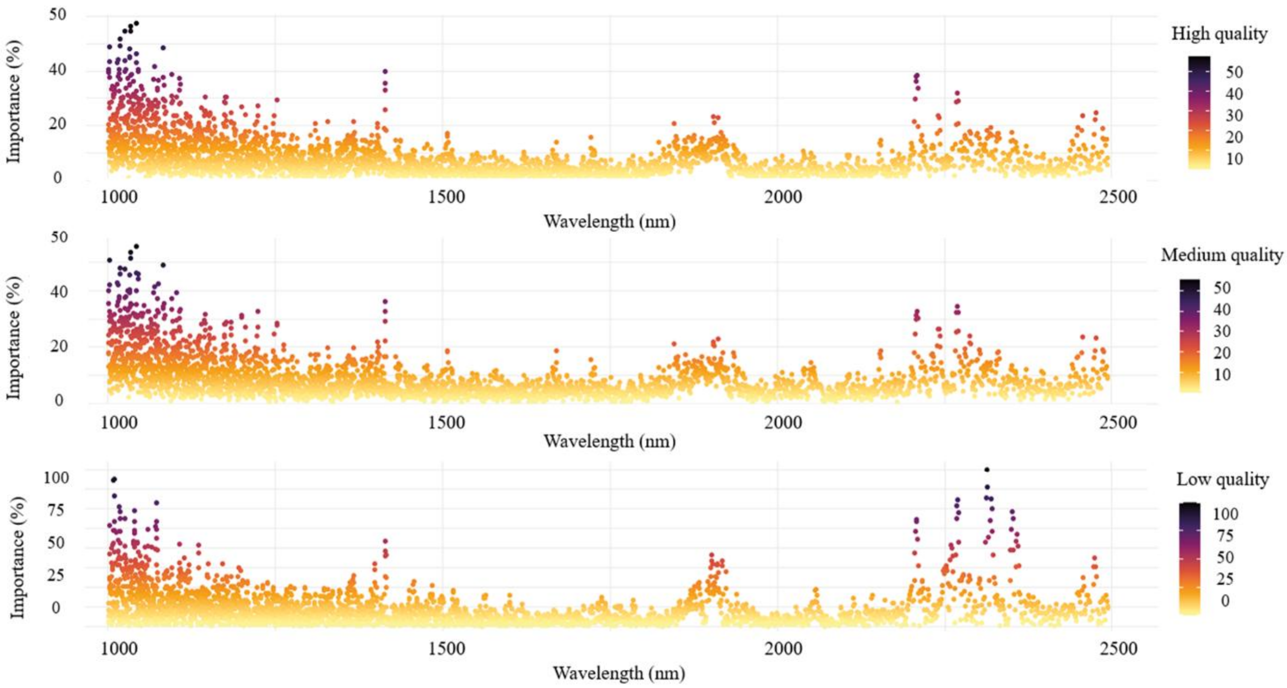
n = number of spectra used for training and/or testing the classification model. SNV = Standard Normal Variate; MSC = Multiplicative Scatter Correction; SG = Savitzky-Golay in relation to the importance of the model, they belong to the O-H functional group associated with carbohydrates. However, the importance of the model is more pronounced in the 2275 - 2400 nm range, and only weakly 2500 nm, which can be attributed to the C-H+C-C and C-H+C-H groups that make up the components of polysaccharides and proteins and have a weak association with lipid absorbance (Figure 5) (Ambrose *et al*., 2016; He *et al*., 2019; Shenk; Workman; Westerhaus, 2001; Xu *et al*., 2020). These peaks therefore correspond to the differences in the biochemical composition of the seeds of the various classes, which are directly related to seed vigour. It can be concluded that there was a correlation between the results for seedling vigour (Figure 2a), protein content (Figure 2b) and accelerated ageing (Figure 2c), and classifying the batches by FT-NIR (Figures 4 and 5).

Figure 4 - Confusion matrix of the classification of lentil seed batches into vigour classes by means of the PLS-DA model using spectral data after processing with the 2nd Savitzky-Golay derivative

2nd Savitzky-Golay Derivative

(%)	Calibration (n = 980)			Validation (n = 420)		
	High	Medium	Low	High	Medium	Low
High	99.86	0.10	0.00	98.89	2.22	0.00
Medium	0.14	99.90	0.10	1.11	97.78	1.67
Low	0.00	0.00	99.90	0.00	0.00	98.33

Figure 5 - Importance of wavelength variables for the PLS-DA classification of vigour levels in lentil seeds



In general, the higher seed vigour or lower level of deterioration may be related to the higher protein content, as seen in Figure 2b (Delarmelino-Ferraresi; Villela; Aumonde, 2014; Henning *et al*., 2010; Wilcox; Cavines, 1992), greater seedling growth (Figure 2a) and better performance in the accelerated ageing test (Figure 2c), as can be seen in Batch 5, followed by Batches 1 and 7. During the seed deterioration process, tolerance to stress conditions is reduced, particularly high temperature and RH, with a reduction in protein content and synthesis, as well as denaturation of the seeds, which contributes to a reduction in vigour (Marcos-Filho, 2020). The protein content of the seeds has been related to vigour and field emergence in the common bean (Ehrhardt-Brocardo; Coelho, 2022), soya bean (Henning *et al*., 2010) and rice (Bortolotto *et al*., 2008).

It should be noted that the process of predicting and classifying seeds according to their physiological potential using near-infrared spectroscopy represents an important technological advance for the seed sector, as it affords quick, non-destructive and accurate results. It is therefore an interesting method, with potential for use in programmes of seed quality control, allowing a company to speed up the decision-making process in terms of batch disposition.

CONCLUSIONS

The accelerated ageing test using the traditional method (100% RH) at 41 °C for 48 hours is efficient for evaluating the vigour of lentil seeds. Near-infrared spectroscopy is a promising tool for classifying seed batches based on physiological potential, with results that can be related to accelerated ageing and seedling vigour, especially when combined with PLS-DA models, with 99% accuracy when submitted to the 2nd Savitzky-Golay derivative.

ACKNOWLEDGEMENTS

The authors wish to thank the Universidade Federal de Viçosa (UFV), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES: Finance Code 001) and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

REFERENCES

AGELET, L. E.; HURBURGH, C. R. Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. **Talanta**, v. 121, p. 288-299, 2014.

AMBROSE, A. *et al.* High-speed measurement of corn seed viability using hyperspectral imaging. **Infrared Physics & Technology**, v. 75, p. 173-179, 2016.

ARAÚJO, J. O. *et al.* Accelerated aging test and antioxidant enzyme activity to assess chickpea seed vigor. **Journal of Seed Science**, v. 43, e202143038, 2021.

BAALBAKI, R *et al*. **Seed vigor testing handbook**. Ithaca: AOSA, 2009. 345 p.

BERTOLIN, D. C.; SÁ, M. E.; MOREIRA, E. R. Parâmetros do teste de envelhecimento acelerado para determinação do vigor de sementes de feijão. **Revista Brasileira de Sementes**, v. 35, n. 1, p. 104-112, 2011.

BORTOLOTTO, R. P. *et al*. Teor de proteína e qualidade fisiológica de sementes de arroz. **Bragantia**, v. 67, n. 2, p. 513-520, 2008.

BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, p. 248-254, 1976.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. **Regras para análise de sementes**. Brasília, DF: Mapa/ACS, 2009. 399 p.

BURNS, D. A.; CIURCZAK, E. W. (ed.). **Handbook of near-infrared analysis**. 3rd ed. Boca Raton: CRC Press, 2007. 834 p.

CHAFFAI, R. R.; MARZOUK, B.; FERJANI, E. E. Aluminum mediates compositional alterations of polar lipid classes in maize seedlings. **Phytochemistry**, v. 66, n. 16, p. 903-1912, 2005.

CICERO, S. M.; VIEIRA, R. D. Testes de frio. *In*: KRZYZANOWSKI, F. C. *et al.* **Vigor de sementes**: conceitos e testes. Londrina: ABRATES, 2020. p. 280-307.

COHEN, J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. **Psychological Bulletin**, v. 70, n. 4, p. 213-220, 1968.

DELARMELINO-FERRARESI, L. M.; VILLELA, F. A.; AUMONDE, T. Z. Desempenho fisiológico e composição química de sementes de soja. **Revista Brasileira de Ciências Agrárias**, v. 9, n. 1, p. 14-18, 2014.

EHRHARDT-BROCARDO, N. C. M.; COELHO, C. M. M. Mobilization of seed storage proteins is crucial to high vigor in common bean seeds. **Ciência Rural**, v. 52, n. 2, e20200894, 2022.

FAVORETTO, M. de M. G. *et al*. Standardization of the accelerated aging test for evaluation of wheat seed vigor. **Journal of Seed Science**, v. 46, e202446007, 2024.

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.

FREITAS, R. A.; NASCIMENTO, W. M. Teste de envelhecimento acelerado em sementes de lentilha. **Revista Brasileira de Sementes**, v. 28, n. 3, p. 59-63, 2006.

HE, X. *et al.* Rapid and nondestructive measurement of rice seed vitality of different years using near-infrared hyperspectral imaging. **Molecules**, v. 24, n. 12, e2227, 2019.

HENNING, F. A. *et al*. Composição química e mobilização de reservas em sementes de soja de alto e baixo vigor. **Bragantia**, v. 69, n. 3, p. 727-734, 2010.

JIANHUA, Z.; MCDONALD, M. B. The saturated salt accelerated aging test for small seeds crops. **Seed Science and Technology**, v. 25, n. 1, p. 123-131, 1996.

KRZYZANOWSKI, F. C. *et al*. Testes de vigor baseados em desempenho de plântulas. *In*: KRZYZANOWSKI, F. C. *et al*. **Vigor de sementes**: conceitos e testes. Londrina: ABRATES, 2020. p. 80-126.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.

MAGUIRE, J. D. Speed of germination-and in selection and evaluation for seeding emergence and vigor. **Crop Science**, v. 2, n. 2, p. 176-177, 1962.

MARCOS-FILHO, J. Testes de envelhecimento acelerado. *In*: KRZYZANOWSKI, F. C. *et al*. **Vigor de sementes**: conceitos e testes. Londrina: ABRATES, 2020. p. 182-244.

MARCOS-FILHO, J.; KIKUTI, A. L. P.; LIMA, L. B. Métodos para avaliação do vigor de sementes de soja, incluindo análise computadorizada de imagens. **Revista Brasileira de Sementes**, v. 31, n. 1, p. 102-112, 2009.

MARINKE, L. S. *et al*. Vigor of lentil seeds evaluated by the tests of accelerated aging and controlled deterioration. **Brazilian Journal of Development**, v. 5, n. 12, p. 30846-30858, 2019.

MEDEIROS, A. D. *et al*. Classificação de qualidade de sementes de pinhão manso utilizando imagens radiográficas e aprendizado de máquina. **Industrial Crops and Production**, v. 146, e112162, 2020.

MUKASA, P. *et al*. Determination of viability of Retinispora (*Hinoki cypress*) seeds using FT-NIR spectroscopy. **Infrared Physical Technology**, v. 98, p. 62-68, 2019.

NASCIMENTO, W. M. **Lentilhas**: muito além do Réveillon. 2019. Disponível em: https://www.embrapa.br/busca-de-noticias/-/noticia/49159157/artigo---lentilhas-muito-alem-do-reveillon. Acesso em: 3 mar. 2024.

NASCIMENTO, W. M.; SILVA, P. P. Grão-de-bico: nova aposta do agronegócio brasileiro. **Seed News**, ano 23, n. 3, maio 2019. Disponível em: https://seednews.com.br/artigos/2969-grao-de-bico-nova-aposta-do-agronegocio-brasileiro-edicao-maio-2019. Acesso em: 6 out. 2023.

PEDÓ, T. *et al*. 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.

QIU, G. *et al*. Single-kernel FT-NIR spectroscopy for detecting supersweet corn (*Zea mays* L. Saccharata sturt) seed viability with multivariate data analysis. **Sensors**, v. 18, n. 4, e1010, 2018.

QUEIRÓZ, T. N. *et al*. Avaliação da qualidade fisiológica de sementes de variedades tradicionais de milho. **Revista da Universidade Vale do Rio Verde**, v. 17, n. 1, p. 1-8, 2019.

RIBEIRO, J. P. O. *et al*. FT-NIR and linear discriminant analysis to classify chickpea seeds produced with harvest aid chemicals. **Food Chemistry**, v. 341, e128324, 2020.

SHENK, J. S.; WORKMAN, J.; WESTERHAUS, M. O. Application of NIR spectroscopy to agricultural product. *In*: SHENK, J. S.; WORKMAN, J. J.; WESTERHAUS, M. O. **Handbook of near-infrared anaalysis**. Boca Raton: CRC Press, 2001. p. 347-386.

SILVA, E. C.; OLIVEIRA, L. A. B.; FERREIRA, N. C. F. Adequação da metodologia do teste de envelhecimento acelerado para sementes de feijão mungo-verde. **Scientia Agraria Paranaensis**, v. 17, n. 4, p. 451-455, 2018.

SOARES, J. M. *et al*. Classification of the physiological potential of soybean seed batches using infrared spectroscopy and chemometric methods. **Journal of Seed Science**, v. 46, e202446009, 2024.

SOUZA, L. R. *et al*. Near infrared spectroscopy and seedling image analysis to evaluate the physiological potential of *Urochloa decumbens* (Stapf) R.D. Webster seeds. **Journal of Seed Science**, v. 45, e202345032, 2023.

WILCOX, J. R.; CAVINES, J. F. Normal and low lenolenic acid soybean strains: response to planting date. **Crop Science**, v. 32, p. 1248-1251, 1992.

WORMA, M. *et al*. 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.

WU, D. *et al*. Determination of a linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination. **Analytica Chimica Acta**, v. 634, p. 166-171, 2009.

XIA, Y. *et al*. Recent advances in emerging techniques for non-destructive detection of seed viability: a review. **Artificial Intelligence in Agriculture**, v. 1, p. 35-47, 2019.

XU, R. *et al*. Use of near-infrared spectroscopy for the rapid evaluation of soybean [*Glycine max* (L.) Merri.] water soluble protein content. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 224, e117400, 2020.

YANG, X. *et al*. Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. **Sensors**, v. 15, n. 7, p. 15578-94, 2015.