Revista Ciência Agronômica, v. 56, e202494450, 2025
Centro de Ciências Agrárias - Universidade Federal do Ceará, Fortaleza, CE
www.ccarevista.ufc.br

Scientific Article
ISSN 1806-6690

Predicting soybean seed germination using the tetrazolium test and computer intelligence[1]

Marcio Alves Fernandes[2], Izabela Cristina de Oliveira[3], Marcio Dias Pereira[4], Breno Zaratin Alves[2], Alan Mario Zuffo[5], Charline Zaratin Alves[2*]

ABSTRACT - Seed quality is critical to agricultural yield, and traditional testing can be time-consuming and subjective. Therefore, the use of machine learning can provide an efficient approach for predicting germination. The aim of this work was to investigate algorithms that, together with tetrazolium test data, lead to efficient prediction of soybean seed germination. The experiment was based on the collection and transcription of a database of thousand soybean seed analysis samples containing information on germination and tetrazolium tests (vigor and viability). The algorithms tested were REPTree, M5P, random forest, logistic regression, artificial neural networks and support vector machine, and the inputs tested were viability, vigor and vigor + viability (tetrazolium test) data. The data analysis used the correlation coefficient and mean absolute error as accuracy parameters of the algorithms. The results highlighted the support vector machine as the most effective algorithm for predicting germination, with the viability and vigor + viability inputs showing the best results. This study suggests that the integration of computational intelligence techniques with the tetrazolium test can make the assessment of soybean seed quality more efficient and contribute to fast and efficient decision making in agriculture.

Key words: Machine learning. Support vector machine. Seed viability. *Glycine max* (L.) Merrill.

DOI: 10.5935/1806-6690.20250064
Editor-in-Chief: Prof. Salvador Barros Torres - sbtorres@ufersa.edu.br
*Author for correspondence
Received for publication 07/11/2024; approved on 06/02/2025
[1]This study is part of the master's dissertation of the first author
[2]Federal University of Mato Grosso do Sul (UFMS), Chapadão do Sul-MS, Brazil, marciofernandes.agro@gmail.com (ORCID ID 0000-0002-8731-4437), breno.zaratin@ufms.br (ORCID ID 0009-0008-5064-0934), charline.alves@ufms.br (ORCID ID 0000-0001-6228-078X)
[3]State University of São Paulo (Unesp), Ilha Solteira-SP, Brazil, izabela.oliveira@unesp.br (ORCID 0000-0002-4666-801X)
[4]University of Rio Grande do Norte(UFRN), Macaíba-RN, Brazil, marcioagron@gmail.com (ORCID ID 0000-0002-0738-3687)
[5]State University of Maranhão (UEMA), Balsas-MA, Brazil, alan_zuffo@hotmail.com (ORCID ID 0000-0001-9704-5325)

INTRODUCTION

Soybean (*Glycine max* [L.] Merrill) is one of the world most important agricultural products and is a major source of protein and vegetable oil used in human and animal nutrition (Lin *et al*., 2023). Soybean plants are propagated by seeds, which must meet minimum quality standards on the basis of laboratory testing. The physiological quality of the seed is one of the pillars for achieving high yield in the field, as it is responsible for the establishment, growth and development of the crop (Marcos Filho, 2015).

Seeds require attributes that attest to their genetic, physical, physiological and sanitary qualities, such as varietal purity, humidity, germination and absence of pathogens, respectively (Krzyzanowski; França-Neto; Henning, 2018). To meet the minimum standards required for production and marketing in Brazil, soybeans must have a physical purity of 99% and germination of 80% (Brasil, 2009).

The germination test is often used to check the quality of seeds (Marcos Filho, 2015). However, the results of this test may be due to the requirements of the seeds, since it is carried out under controlled and favorable conditions. Thus, seeds that show good germination percentages in the laboratory may present different results in the field, making it necessary to use vigor tests to predict seed performance under uncontrolled conditions (Wendt *et al*., 2017).

Vigor tests, including the tetrazolium test, characterize seeds into high-, medium- and low-vigor batches on the basis of the ability of the seeds to withstand different types of stress without altering their germination rate, membrane integrity or seedling performance in the field (Hao *et al*., 2020). The use of the tetrazolium test allows results to be obtained in less time than the germination test and represents the behavior of the seeds in a way that is proportional to what would happen to the seeds in the field under adverse conditions (Krzyzanowski; França-Neto; Henning, 2018).

However, the conventional tests used to determine seed quality are costly and time-consuming, taking days to provide information and requiring experience on the part of the evaluator. There is therefore a clear need for methods that provide information in a nondestructive, precise and accurate way, for example, by using computational intelligence techniques.

Computational intelligence use techniques, such as machine learning, makes it possible to measure various characteristics quickly and nondestructively through data processing (Guo *et al*., 2023). Prediction is one of the functions of machine learning algorithms, where the algorithm learns about the behavior of a given dataset during training and is then able to predict how a new dataset will behave (Karakatic; Podgorelec, 2016; Santana *et al*., 2022). However, each algorithm performs differently, so it is necessary to test different models to find the best model for prediction (Santana *et al*., 2023).

The hypothesis of this research was based on the possibility of predicting soybean germination using data from the tetrazolium test as input to machine learning models. The aim of this work was therefore to find the best algorithm and the best input for predicting soybean seed germination.

MATERIAL AND METHODS

Conducting the experiment

The experiment was carried out at the Seed Production and Technology Laboratory of the Federal University of Mato Grosso do Sul, Chapadão do Sul Campus (lat. 18º41'33"S; long. 52º40'45"W; alt. 790 m asl.). According to the Koppen's classification, the climate is tropical wet and dry or savanna (Aw), with an annual average temperature ranging from 13 to 28 °C, and annual average rainfall of 1.850 mm. The experiment was based on surveying and transcribing a database containing thousands of samples for soybean seed analysis, with information on germination and tetrazolium tests (vigor and viability). All the samples were obtained from the municipality of Chapadão do Sul - MS and the surrounding region.

Seed analyses

The germination test was carried out using four subsamples of 50 seeds, which were distributed on germitest paper previously moistened to 2.5 times their weight and then stored in a germinator at 25 °C (Brasil, 2009). The results are expressed as a percentage of normal seedlings.

The tetrazolium test was carried out on two subsamples of 50 seeds that were presoaked on germitest paper moistened to 2.5 times their weight and stored at 25 °C for 24 hours. After this period, the seeds were soaked in a 0.075% 2.3.5 triphenyltetrazolium chloride solution and placed in a B.O.D. in the dark at 40 °C for three hours. The seeds were then sectioned and evaluated according to França-Neto and Krzyzanowski (2019).

Statistical analyses

The data were subjected to machine learning analysis using viability (Viab), vigor (Vigor) and vigor + viability (Vigor+Viab) data as inputs and germination as the output variable. The data were divided into training and test sets within the software itself, following a ratio of 80:20, i.e., 80% of the data used for training and 20% for validation, ensuring that the models were properly trained and validated.

To maximize the quality of the data provided, the use of advanced cross-validation techniques ensured that the models achieved high accuracy and robustness. The stratified cross-validation method was used in the models with k-fold = 10 and ten repetitions (100 runs for each model). All model parameters were set according to the default configuration of Weka 3.8.5 software.

The machine learning models used were REPTree (Snousy *et al*., 2011) (RT), M5P (Blaifi *et al*., 2018), random forest (RF) (Belgiu; Dragut, 2016), logistic regression (LR) (Štepanovský *et al*., 2017), artificial neural network (ANN) (Egmont-Petersen; De Ridder; Handels, 2002), and support vector machine (SVM) (Nalepa; Kawulok, 2019). To assess the performance of the prediction models tested, the correlation coefficient (r) and mean absolute error (MAE) metrics were used.

Subsequently, an analysis of variance was carried out considering a 6×3 factorial scheme (models versus inputs) with five repetitions, using the Scott–Knott test at 5% probability. Boxplots were generated for each parameter (r and MAE) considering the models and inputs tested. On the basis of these statistics, the best ML technique and the best input for predicting soybean seed germination were identified. These analyses were carried out in R software using the ExpDes.pt and ggplot2 packages.

RESULTS AND DISCUSSION

The interaction between the machine learning (ML) algorithms and the inputs tested was not significant (Table 1). However, the inputs and machine learning algorithms were significant for the correlation coefficient (r) and mean absolute error (MAE), indicating that there are variations in the performance of the algorithms in relation to the input data used.

The use of machine learning allows for an effective approach to data processing, especially those with a high level of complexity, which process a large amount of information in a precise and accurate manner (Xiao *et al*., 2023). Thus, the application of computational intelligence techniques can accelerate decision-making in relation to seed germination, since the germination test for soybean takes eight days. The tetrazolium test takes less than 24 hours, and the use of its data associated with specific algorithms results in a prediction in minutes, promoting speed and efficiency in seed laboratory logistics.

In terms of the correlation coefficient (r), the machine learning algorithms M5P, RL, ANN and SVM yielded the best results, and they did not differ from each other (Figure 1). The use of the Viab and Vigor+Viab inputs resulted in results above 0.75 in terms of predicting germination (Figure 2). These results indicate that germination can be predicted using viability or vigor + viability obtained from the tetrazolium test, since the machine learning algorithms were efficient at predicting germination using these inputs.

Machine learning use, especially for predicting characteristics such as seed germination, depends on the performance of the algorithms tested. Each machine learning algorithm performs differently with the dataset, which highlights the need to test different algorithms to find the best predictor (Santana *et al*., 2023).

Tetrazolium test for soybean depends mainly on human observation and experience, which is subjective, unlike the use of machine learning, which makes a quick, accurate and unbiased assessment of the data presented (Lin *et al*., 2023). In this way, the interaction of conventional tests on soybean with machine learning mitigates human dependence and its interpretations, which can vary.

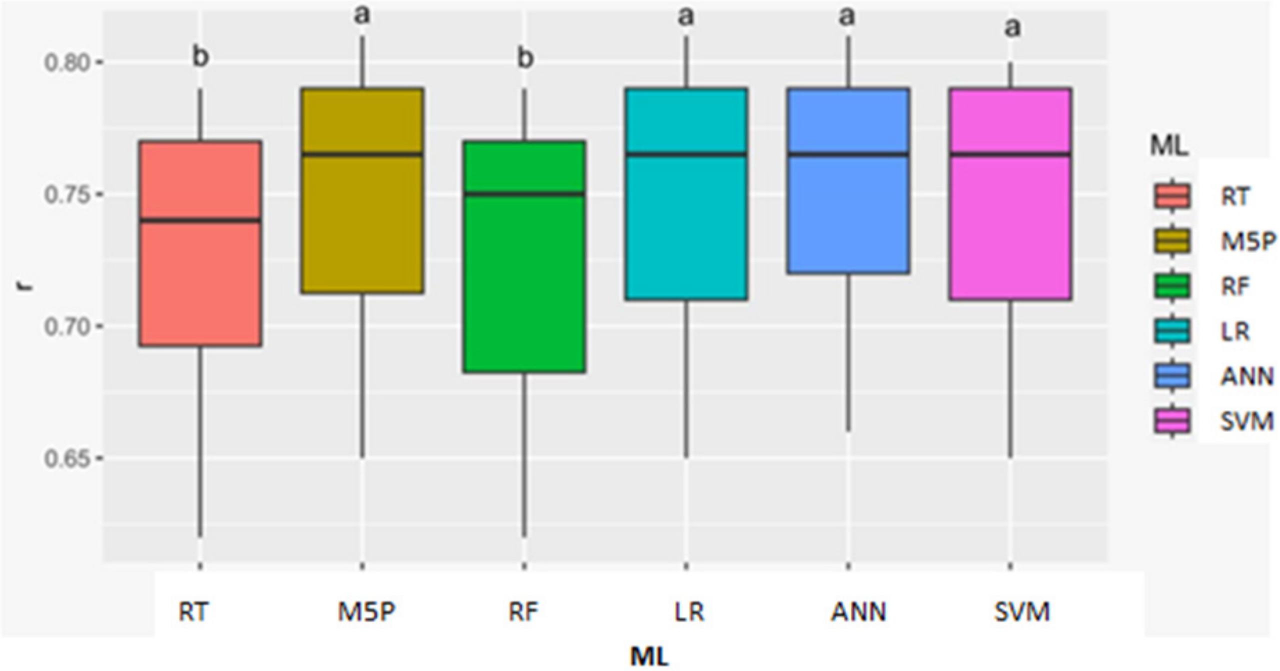
The MAE indicated that the SVM algorithm showed the best results in predicting germination, followed by the M5P and LR algorithms (Figure 3). The ANN algorithm was not suitable for predicting germination using data obtained from the tetrazolium test, as it had the highest absolute error compared with the other algorithms tested.

Table 1 - Summary of the analysis of variance for the accuracy variables: correlation coefficient (r) and mean absolute error (MAE) as a function of machine learning algorithms and inputs

SV	DF	r	MAE
inputs	2	0.129395*	38.072*
ML	5	0.003918*	14.8684*
inputs*ML	10	0.000656^{ns}	0.3241^{ns}
Resídue	162	0.000677	0.211012
CV (%)		3.5	5.8

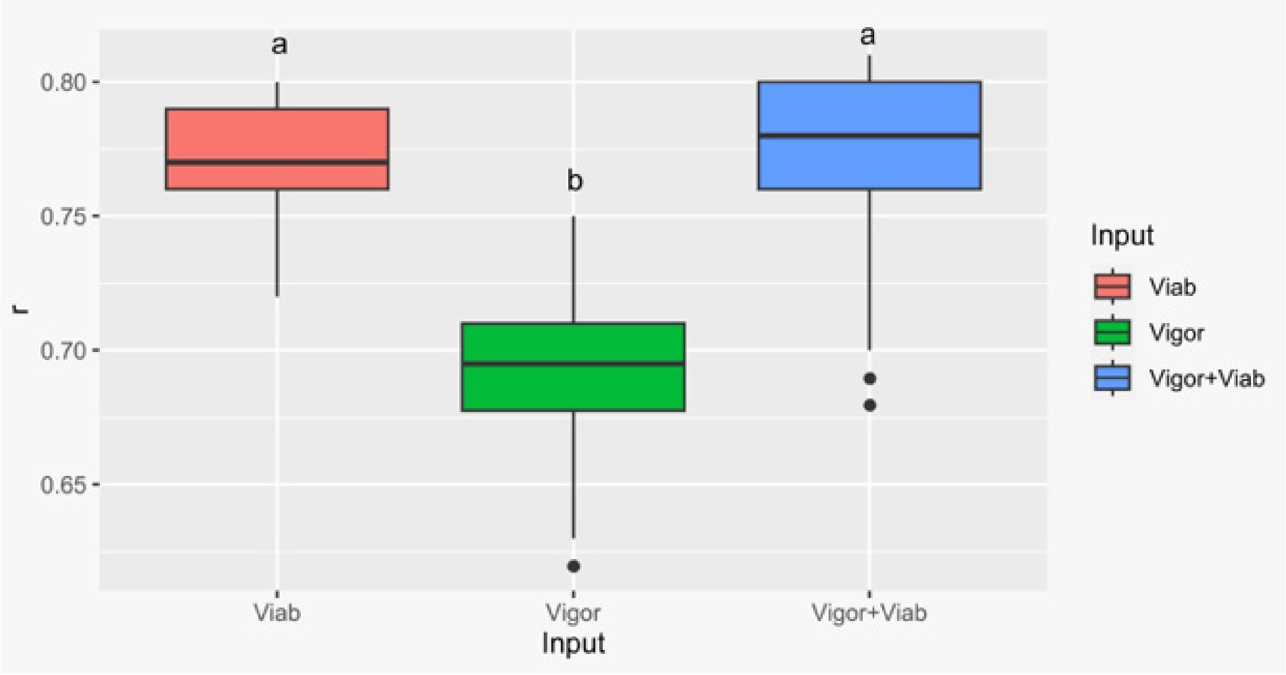
* significant at 5% by the F test; SV: sources of variation; DF: degrees of freedom; ML: machine learning; CV: coefficient of variation

Figure 1 - Boxplot comparing means for the correlation coefficient (r) as a function of the algorithms tested for predicting soybean seed germination



RT: REPTree, M5P: M5P, RF: Random Forest, LR: Logistic Regression, ANN: Artificial Neural Network, SVM: Support Vector Machine

Figure 2 - Boxplot comparing means for the correlation coefficient (r) as a function of the inputs tested for predicting soybean seed germination



The Viab and Vigor+Viab inputs (Figure 4) had the best performance in terms of the MAE, indicating that when the viability and vigor + viability of soybean seeds were used as inputs, there was less error in predicting germination. Viability reflects the theoretical ability of a seed to germinate, whereas vigor refers to the ability to germinate when environmental conditions, especially temperature and humidity, are unfavorable

(Xia *et al*., 2019). Therefore, predicting germination using the tetrazolium test is a viable alternative for reducing the time needed to obtain results from seeds.

The SVM algorithm is an efficient machine learning method for predicting plant characteristics, as it is a versatile algorithm that can be used for linear

Figure 3 - Boxplot comparing means for mean absolute error (MAE) as a function of the algorithms tested for predicting soybean seed germination RT: REPTree, M5P: M5P, RF: Random Forest, LR: Logistic Regression, ANN: Artificial Neural Network, SVM: Support Vector Machine

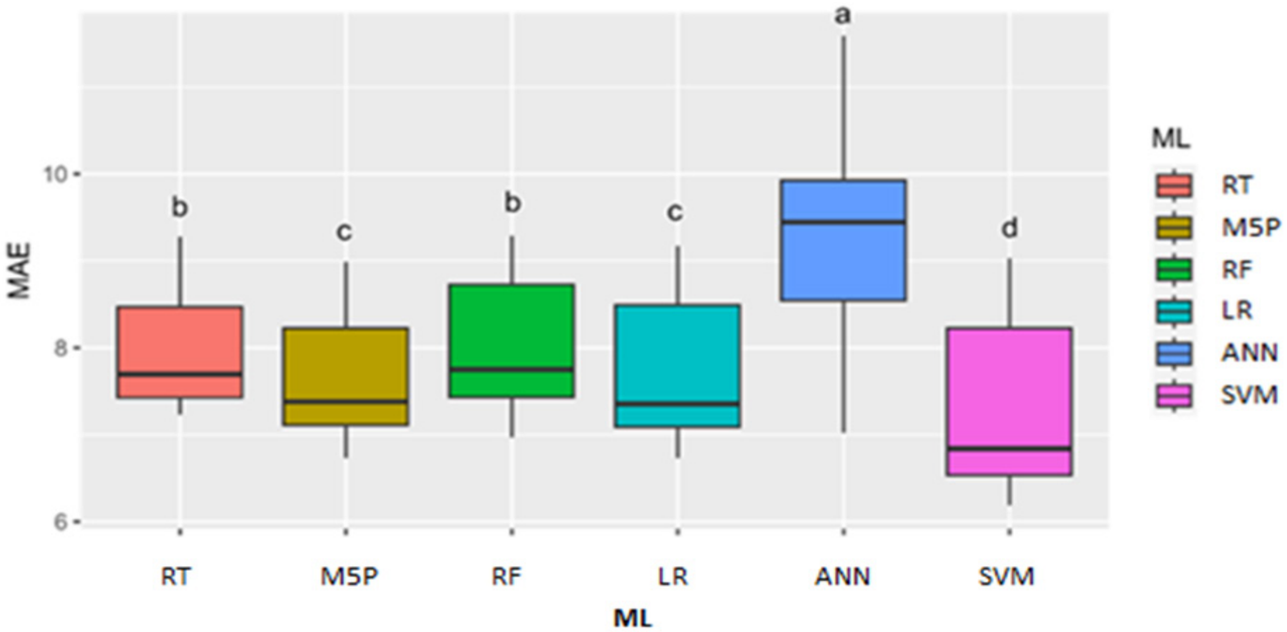
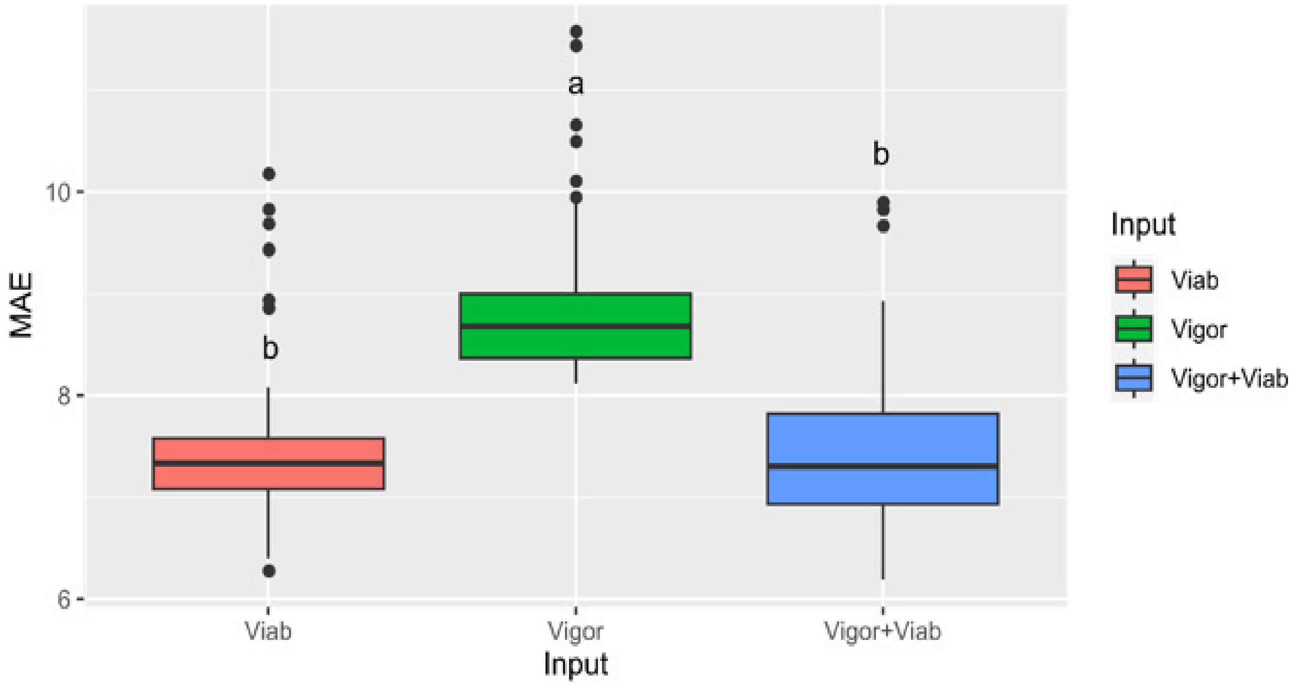


Figure 4 - Boxplot comparing means for mean absolute error (MAE) as a function of the inputs tested for predicting soybean seed germination



and nonlinear regression tasks since it is based on the principle of finding the best boundary that separates the data (Sarkar *et al*., 2023). The use of SVMs has enabled the classification of soybean seeds in relation to color histograms (Jitanan; Chimlek, 2019) and the physical characteristics of beans (Koklu; Ozkan, 2020). The SVM algorithm has already been used to classify the physical characteristics of soybeans and other legumes.

The findings of this work show that the use of machine learning can be used to predict the germination of soybean seeds on the basis of the viability and vigor information obtained from the tetrazolium test. The SVM algorithm had the highest r value and the lowest MAE value, demonstrating its superiority in prediction over the other algorithms.

The use of viability + vigor information as input to the models resulted in the highest r values and the lowest MAEs, making them suitable for predicting germination. Future prospects are based on the possibility of creating an equation that can be used to develop software that allows tetrazolium data to be input and germination data to be output, contributing to faster decision-making in relation to batches of soybean seeds.

CONCLUSIONS

1. The soybean seed germination can be effectively predicted using the tetrazolium test combined with computer intelligence techniques;
2. Among the various algorithms tested, SVM stood out, especially when the viability and vigor + viability inputs were used.

REFERENCES

BELGIU, M.; DRĂGUŢ, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24-31, 2016.

BLAIFI, S. *et al*. M5P model tree based fast fuzzy maximum power point tracker. **Solar Energy**, v. 163, p. 405-424, 2018.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Regras para análise de sementes**. Brasília: Secretaria Nacional de Defesa Agropecuária, 2009. 398 p.

EGMONT-PETERSEN, M.; DE RIDDER, D.; HANDELS, H. Image processing with neural networks: a review. **Pattern Recognition**, v. 35, n. 10, p. 2279-2301, 2002.

FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C. Tetrazolium: an important test for physiological seed quality evaluation. **Journal of Seed Science**, v. 41, p. 359-366, 2019.

GUO, Z. *et al*. Application of visible-near-infrared hyperspectral imaging technology coupled with wavelength selection algorithm for rapid determination of moisture content of soybean seeds. **Journal of Food Composition and Analysis**, v. 116, e105048, 2023.

HAO, Q. *et al*. Evaluation of seed vigor in soybean germplasms from different eco-regions. **Oil Crop Science**, v. 5, n. 1, p. 22-25, 2020.

JITANAN, S.; CHIMLEK, P. Quality grading of soybean seeds using image analysis. **International Journal of Electrical and Computer Engineering (IJECE)**, v. 9, n. 5, p. 3495-3503, 2019.

KARAKATIČ, S.; PODGORELEC, V. Improved classification with allocation method and multiple classifiers. **Information Fusion**, v. 31, p. 26-42, 2016.

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.

KRZYZANOWSKI, F. C.; FRANÇA-NETO, J. B.; HENNING, A. A. **A alta qualidade da semente de soja**: fator importante para a produção da cultura. Londrina: Embrapa, 2018. (Circular técnica, 136).

LIN, W. *et al*. Online classification of soybean seeds based on deep learning. **Engineering Applications of Artificial Intelligence**, v. 123, e106434, 2023.

MARCOS FILHO, J. Seed vigor testing: an overview of the past, present and future perspective. **Scientia Agricola**, v. 72, n. 4, p. 363-374, 2015.

NALEPA, J.; KAWULOK, M. Selecting training sets for support vector machines: a review. **Artificial Intelligence Review**, v. 52, n. 2, p. 857-900, 2019.

SANTANA, D. C. *et al*. Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. **Remote Sensing Applications: Society and Environment**, v. 29, e100919, 2023.

SANTANA, D. C. *et al*. High-throughput phenotyping allows the selection of soybean genotypes for earliness and high grain yield. **Plant Methods**, v. 18, n. 1, p. 13, 2022.

SARKAR, S. *et al*. Soybean seed composition prediction from standing crops using PlanetScope satellite imagery and machine learning. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 204, p. 257-274, 2023.

SNOUSY, M. B. A. *et al*. Suite of decision tree-based classification algorithms on cancer gene expression data. **Egyptian Informatics Journal**, v. 12, n. 2, p. 73-82, 2011.

ŠTEPANOVSKÝ, M. *et al*. Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. **Forensic Science International**, v. 279, p. 72-82, 2017.

WENDT, L. *et al*. Relação entre testes de vigor com a emergência a campo em sementes de soja. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, v. 12, n. 2, p. 166-171, 2017.

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

XIAO, J. *et al*. Enhancing assessment of corn growth performance using unmanned aerial vehicles (UAVs) and deep learning. **Measurement**, v. 214, e112764, 2023.