Engenharia Agrícola

ISSN: 1809-4430 (on-line)

www.engenhariaagricola.org.br

Scientific Paper
Doi: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v43n4e20230121/2023

MACHINE LEARNING MODELS FOR PREDICTING MECHANICAL DAMAGE, VIGOR AND VIABILITY OF SOYBEAN SEEDS DURING STORAGE

Laila R. Cirqueira[1], Paulo C. Coradi[1,2,3*], Larissa P. R. Teodoro[1], Paulo E. Teodoro[1], Dágila M. Rodrigues[2,3]

[3*]Corresponding author. Laboratory of Postharvest (LAPOS), Federal University of Santa Maria, Campus Cachoeira do Sul, Passo D'Areia/Cachoeira do Sul - RS, Brazil.
E-mail: paulo.coradi@ufsm.br | ORCID ID: https://orcid.org/ 0000-0001-9150-2169

KEYWORDS

artificial intelligence, postharvest, storage packaging, storage temperature; storage time, tetrazolium test.

ABSTRACT

Artificial Intelligence has been widely applied in data prediction for better decision making and process optimization. In the post-harvest, the control of biotic and abiotic factors is fundamental for the conservation of seed quality. Meanwhile, the tetrazolium test has been used to evaluate seed quality, however, with several limitations that can lead to evaluation errors. Thus, machine learning models can be an alternative to predict the quality of soybean seeds, with gains in the speed of obtaining results in relation to laboratory analysis methods, making the processes more robust and with low operational cost. With this, the aim of this study was to identify the best machine learning model for predicting mechanical damage, vigor and viability of soybean seeds during storage, depending on different conditions (10, 15 and 25 ºC), packaging (with coating and uncoated) and storage times (0, 3, 6, 9 and 12 months). M5P decision tree (M5P) and Random Forest (RF) models showed the best performance for predicting seed vigor (r = 0.75 and MAE = 10.0), and viability (r = 0.85 and MAE = 5.1), and mechanical damage to seeds (r = 0.64 and MAE = 11.2). It was concluded that the Random Forest (RF) model was the one that best predicted the results of soybean seed quality, with a more simplified and agile analysis for the development of vigor and viability of soybean seeds in storage.

INTRODUCTION

The moisture content of the seeds, the temperature and relative humidity of the intergranular air and the storage environment are important variables to be monitored to preserve the quality of the seeds (Capilheira et al., 2019). However, variations in seed moisture content, shape, environment and storage time can influence the metabolic activity and physiological seeds quality (Mylona et al., 2012).

To reduce the metabolic activity of the seeds, it is suggested to control the temperature and relative humidity of the storage environment, so that the seeds remain in equilibrium moisture content with moisture content close to 12% (w.b.), considered a safe moisture (Ebone et al., 2020; Sarath et al., 2016). According to Oliveira et al. (2021), physical damage caused to seeds can cause reduced vigor, viability and even seed death (Rocha et al., 2017; Silva et al., 2022). For this, the tetrazolium test has been an important and efficient analysis to assess quality physical and mechanical damage, which interfere with seed vigor and viability (Rocha et al., 2017). However, the tetrazolium test has some limitations, including the need for advanced training and knowledge about seed science and technology to interpret the results, with the possibility of susceptible errors (Coradi et al., 2020). Seed quality analyzes often generate a quantity of information that makes a quick and effective short-term analysis impossible. Therefore, erroneous results may imply economic losses for seed processing units (André et al., 2022).

[1] Federal University of Mato Grosso do Sul, Campus de Chapadão do Sul/Chapadão do Sul - MS, Brazil.
[2] Department Agricultural Engineering, Rural Sciences Center, Federal University of Santa Maria/Santa Maria - RS, Brazil.
[3] Laboratory of Postharvest (LAPOS), Federal University of Santa Maria, Campus Cachoeira do Sul, Passo D'Areia/Cachoeira do Sul - RS, Brazil.

Area Editor: Gizele Ingrid Gadotti
Received in: 8-25-2023
Accepted in: 10-30-2023

Currently, the use of artificial cooling technologies in the conservation of stored seeds has been shown to be effective (Mylona & Magan 2011; Ferreira et al., 2017). The maintenance of seeds at low temperatures, associated with a controlled condition of air relative humidity can provide a favorable storage condition. However, the costs of refrigeration and displacement and waiting for the sowing of soybean seed lots after storage could still compromise the quality.

Thus, the use of technological tools for real-time monitoring of seed quality can help in prevention and decision-making about the ideal storage time for seed quality conservation (Souza et al., 2019; Jaques et al., 2022; Vo-Thanh et al., 2022). With this, over the years, Artificial Intelligence, more precisely Machine Learning models, are being introduced in the means of agricultural production, through the definition of standards so that the machine makes recommendations or takes decisions based on prediction with algorithms and a significant set of data, increasing the efficiency and optimization of processes (Baryshev et al., 2020; Helm et al., 2020, Benos et al., 2021; Lutz & Coradi, 2022).

According to Gadotti et at. (2022a), the seed sector faces several challenges when it comes to ensuring a quick and accurate decision making when working with large amounts of data on physiological quality of seed lots, which makes the process time-consuming and inefficient. Thus, artificial intelligence emerges as a new technological option in the seed sector to solve database problems in the post-harvest stages. In this context, the use of Machine Learning (ML) has offered capacity for processing, analyzing and interpreting data (Moreti et al., 2021). When properly modeled, ML techniques can offer responses in less time when compared to statistical regression models. Random Forests (FA) is an ML technique successfully used in yield forecasting and quality assessment (Ramos et al., 2020). This method proved to be an effective and easier-to-use method for predicting corn and wheat quality when compared to multiple linear regression models (Jeong et al., 2016). Artificial Neural Networks (ANN) is another method that can be trained from data related to corresponding inputs and outputs (Pazoki & Pazoki, 2011). ANNs are useful tools for analysis and interpretation of complex food security data, predictions of physical and chemical seed quality (Goyal, 2014).

During the last few years, research has investigated the results of using ML methods for classification within the context of agricultural problems, such as prediction of nitrogen content (Osco et al., 2019), soil correction, seed classification (Hussain & Ajaz, 2015), phosphorus reduction in wastewater (Kumar & Deswal, 2020), protein prediction in stored grains (Radhika & Rao, 2014; Lutz et al., 2022), grain quality stored (Liu et al., 2017), insects population in grain stored (Nyabako et al., 2020). Machine Learning models have been widely used to predict the quality of soybean (André et al., 2022) and corn (Xu et al., 2021) seeds, determine wheat yield (Baryshev et al., 2020), as well as evaluating the seed germination rate (Škrubej et al., 2015; Ropelewska & Piecko, 2022). Recent research has demonstrated the effectiveness of Maching Learning models in predicting the viability, vigor and germination speed of seeds from different crops (Medeiros et al., 2020; Pinheiro et al., 2021). Pereira et al. (2020), André et al. (2022), and Gadotti et al. (2022b) and achieved satisfactory results using Maching Learning algorithms, but the models that best performed the prediction of soybean quality were different, depending on processing and storage conditions.

To minimize the gaps caused by conventional seed analysis, dependent on personal interpretations, Machine Learning techniques can be a alternative to analyze the quality of stored soybean seeds, and can be used as a support tool for decision making about conditions, ways and storage times to maintain quality and reduce losses of soybean seeds. Thus, the aim of this study was to identify the best machine learning model for predicting mechanical damage, vigor and viability of soybean seeds during storage.

MATERIAL AND METHODS

Description and experimental design

Initially, impurities and foreign matter were removed from the soybean seed lots, with the aid of an air machine and LC 160 sieve (Kepler Weber, Rio Grande do Sul, Brazil). Next, the soybean seeds were dried to levels of 12% (b.u.), using silos-dryers with radial air flow and a temperature of 40 °C (Silos Roma, Paraná, Brazil). After drying, the seeds were submitted to classification, using a spiral separator (Akyurek Technology, Mersin, Turkey) and an asymmetric table model SDS-80 (Silomax, Paraná, Brazil), in order to standardize the size and mass of the seeds. The experimental evaluations took place during the storage stage, in a completely randomized design (DIC), in a factorial design (3 x 2 x 5), with three storage temperatures (10, 15 and 25 ºC), two packages (with and without coating) and five storage times (0, 3, 6, 9 and 12 months) (Figure 1).

The seeds were stored in raffia bags with and without coating (polypropylene) in air-conditioned environments. The packages used were raffia bags measuring 20 cm (width) x 30 cm (height) x 0.25 cm (thickness), coated with high-density polypropylene. Seed mass temperature was monitored with digital thermo-hygrometers, Logbox model, RHT-LCD (Novus Electronic Products Company, Rio Grande do Sul, Brazil). Every three months of storage, soybean seeds were sampled for quality assessment.

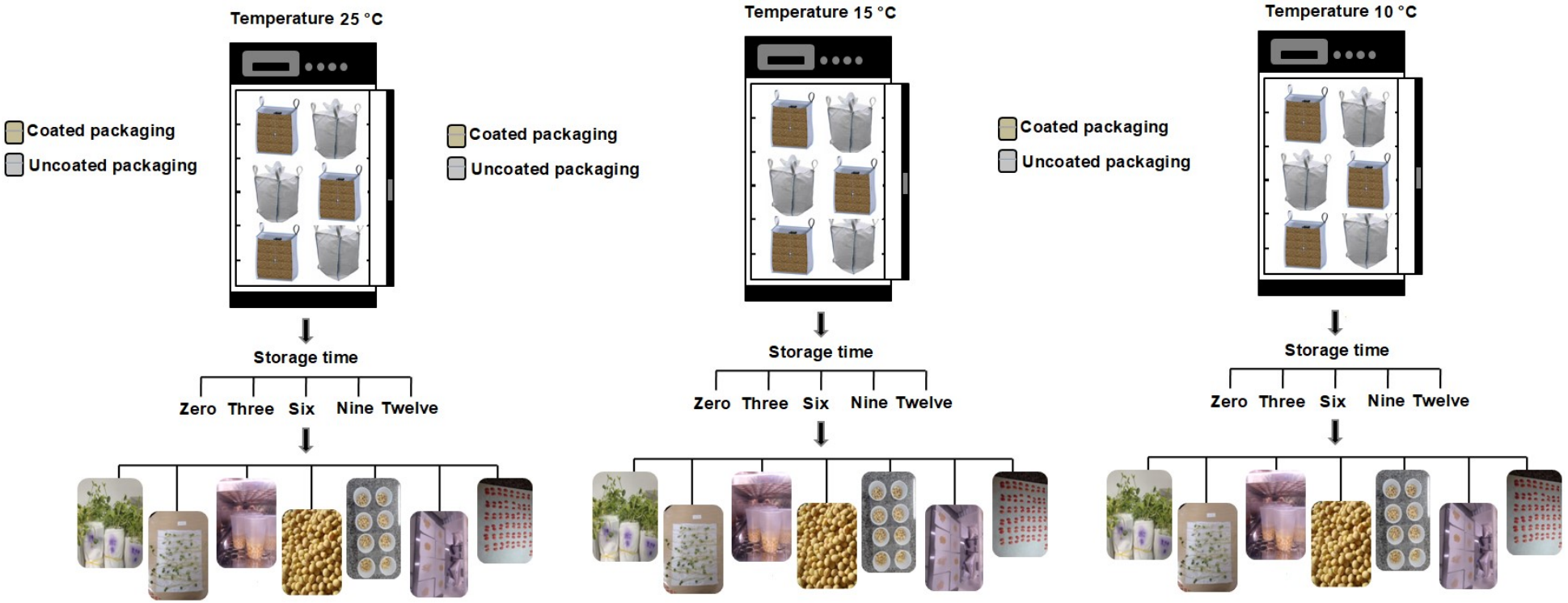

FIGURE 1. Experimental scheme and control in the storage of soybean seeds.

Soybean seed quality analysis

A forced air circulation oven with a capacity of 220 L (Tecnal, São Paulo, Brazil) and air temperature controlled at 105 °C ± 1 °C were used. Four seed samples of 50 g of seeds from each treatment were weighed on a scale model B13200H (Shimadzu, São Paulo - SP, Brazil) and placed in the greenhouse for 24 h. Afterwards, the samples were removed and placed in a desiccator (Tecnal, Piracicaba - SP, Brazil) with silica to cool the seeds. The seed water content (% w.b.) was determined by the difference in the initial and final weight of the seeds (Brazil, 2009). The apparent specific mass of the seeds was determined using a beaker with a known volume of one liter and a precision scale. The apparent specific mass of the seeds was calculated through the relation mass and volume of the sample. Four repetitions of the analysis were performed for each treatment (Brazil, 2009).

To determine the germination (viability) and vigor tests, four subsamples of 50 seeds from each experimental unit were used. The seeds were distributed in paper towel rolls (Germitest), moistened with distilled water in approximately 2.5 times the dry mass of the paper. After this process, the paper rolls with the seeds were placed in a Mangesdor model germinator (Tecnal, São Paulo, Brazil), conditioned to a temperature of 25 °C ± 2 °C. On the fifth day after the beginning of the tests, the vigor of the seeds was evaluated, and after the eighth day, the viability of the seeds was evaluated by the germination test, according to the Rules for Seed Analysis (Brazil, 2009). Viable seeds were those capable of producing normal seedlings. The soybean seed viability levels as a function of mechanical damage, humidity, vigor and viability are presented in Table 1, according to the recommendations by França-Neto & Krzyzanowski (2019).

TABLE 1. Soybean seed viability levels as a function of different mechanical damage and moisture content.

Analyzes	Seed viability levels
Mechanical damage 1 (DM1)	1 a 8
Mechanical damage 2 (DM2)	6 a 8
Moisture 1 (U1)	1 a 8
Moisture 2 (U2)	6 a 8
Vigor (VIG)	1 a 3
Viability (VIB)	1 a 6

Levels: 1 - viable and highest vigor, 2 - viable and high vigor, 3 - viable and medium vigor, 4 - viable and low vigor, 5 - viable and very low vigor, 6 - not viable, 7 - not viable, 8 – dead seed.

In the tetrazolium test, four subsamples of 50 seeds from each experimental unit were used. Seeds were pre-moistened on Germitest paper for 16 h at 25 °C and then immersed in a 0.075% tetrazolium solution, where they were kept for 3 h at 35 °C. After this period, the seeds were washed under running water and their vigor, viability and moisture damage (Kong et al., 2008; Mylona et al., 2012) were evaluated according to the methodology established by França-Neto & Krzyzanowski (2019).

Machine Learning analysis

Three supervised machine learning (ML) models for predicting mechanical damage, vigor and viability of soybean seeds during storage were tested: Artificial Neural Networks (ANN), M5P decision tree (M5P), and Random Forest (RF) (Figure 2). A conventional prediction using Multiple Linear Regression (MRL) was used as a control model. The input data for each model were: three different storage temperatures (SC) (10, 15 and 25 ºC), two packages (E) (coated and uncoated) and five storage times (ST) (0, 3, 6, 9 and 12 months) to predict the variables (output data) mechanical damage 1 (DM1), mechanical damage 2 (DM2), humidity 1 (U1), humidity 2 (U2), vigor (VIG) and viability (VIB).

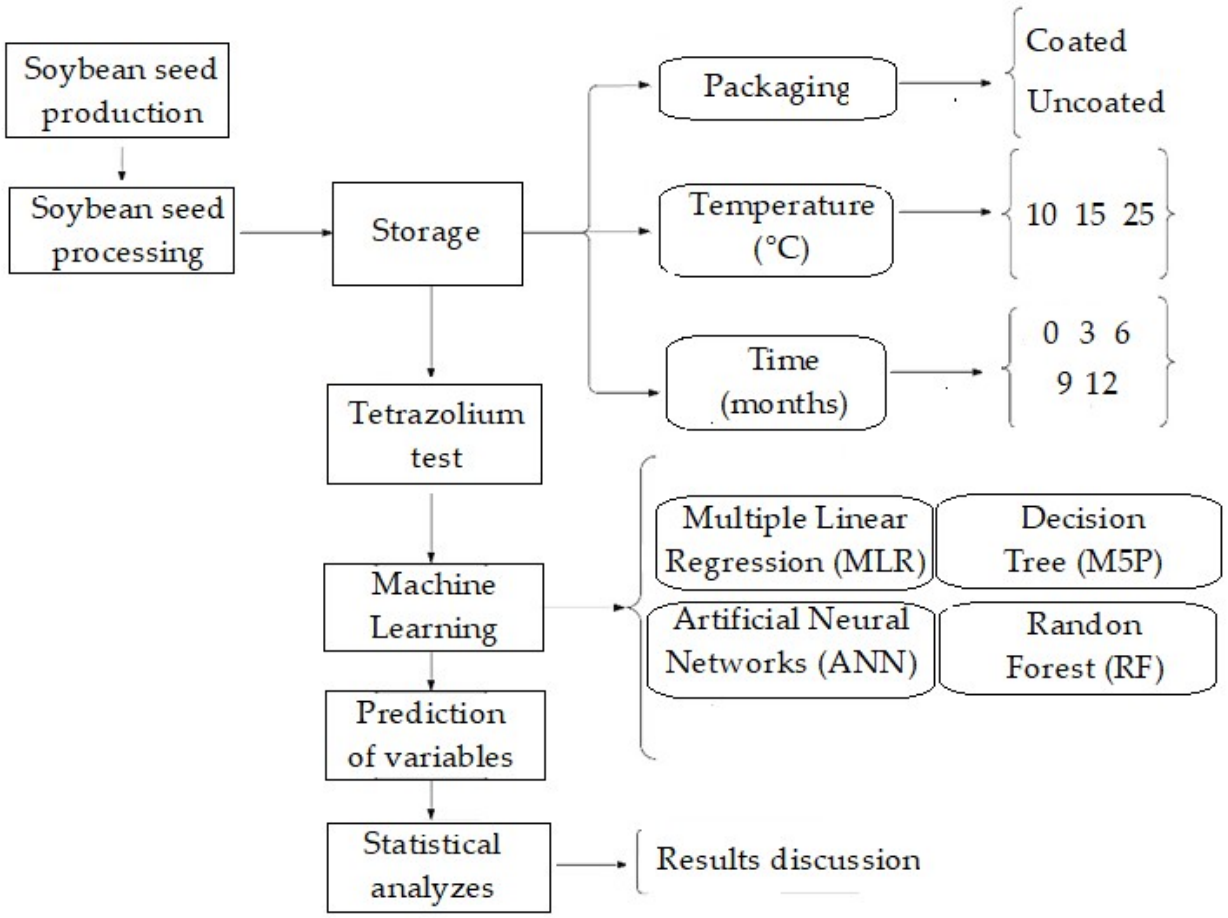

FIGURE 2. Flowchart representing the storage conditions and the application of Machine Learning models to predict the quality of soybean seeds.

ML analyses were performed using the default software configuration for all models tested (Bouckaert et al., 2008). The ANN tested consists of a single hidden layer formed by a number of neurons that is equal to the number of attributes plus the number of classes, all divided by 2 (Egmont-Petersen et al., 2002). The ANN adopted a learning rate of 0.3 and a momentum rate of 0.2, and used the backpropagation algorithm to learn a Multilayer Perceptron to predict the variables. M5P model is a reconstruction of Quinlan's M5 algorithm (Quinlan, 1993) based on the conventional decision tree with the addition of a linear regression function to the leaf nodes (Blaifi et al., 2018). In M5P algorithm, the pruning procedure was adopted and the minimum number of instances to allow at a leaf node adopted was equal to 4. RF model produces several prediction trees for the same data set and use a voting scheme among all learned trees to predict new values (Belgiu & Drăguţ, 2016). RF was built using a number of trees equal to 100, number of execution slots (threads) to use for constructing the ensemble equal to 1, and adopting the default settings of the Weka software for the remaining hyperparameters.

All ML analysis were performed on an Intel® CoreTM i5-3317U CPU with 4 Gb of RAM using a stratified random cross-validation with k-fold = 10 and 10 repetitions (100 runs). In k-fold cross-validation, we divide the input data into subsets of data called k-folds. The ML model is trained on all but not in a fold (k-1) and then evaluates the model on the dataset that was not used for training. Thus, in a *k*-fold cross-validation, *k* -1 subsets are used for training and 1 subset for validation. This procedure is repeated *k* times (here, ten times) to use all possible combinations of training and validation sets. From this approach, all data points are predicted and validated, while still keeping a separate training set. This strategy has been widely adopted in ML analyses to avoiding overfitting or biased learning (Granholm et al., 2012; André et al., 2022; Ramos et al., 2020; Gava et al., 2022; Baio et al., 2023; Santana et al., 2023a; Santana et al., 2023b). Additionally, the cross-validation strategy provides the prediction results of one variable for each fold (k), making it possible to have repetitions for the model's accuracy values both in classification and regression studies, such as correlation between observed *versus* predicted (André et al., 2022; Baio et al., 2023; Santana et al., 2023b). This makes it possible to carry out statistical tests to compare or group means for the different techniques, to be described in detail in the next subtopic, which provides an accurate recommendation about the best ML models.

Statistical analyzes

Pearson's correlation analysis was performed between monitored and predicted variables. These analyzes were performed with the aid of the Rbio Software, following the procedures recommended by Bhering (2017). Subsequently, the correlation coefficient (r) and the mean apparent error (MAE) for predicting the

quality of stored soybean seeds were obtained. Then, analysis of variance was performed, adopting a completely randomized design (CRD). For the CRD, Machine Learning models (ANNs, M5P and RF) plus multiple linear regression (MLR) were considered for comparison. 10 repetitions were provided for each model. For grouping the means of r and MAE, the Scott-Knott test at 5% probability was adopted. Afterwards, boxplot graphs were generated for the accuracy parameters (r and MAE) for each output variable. The Rbio software (Team, 2018) and the R software (Team, 2018) with the ExpDes.pt and ggplot2 packages were used for the analyzes.

RESULTS AND DISCUSSION

Pearson's correlation analysis

In Figure 3 are the correlations between the monitored and predicted variables, where vigor (VIG) x viability (VIB) had a high and positive correlation and a high and negative correlation with humidity (U2), while the other variables showed correlations weak. Pearson's correlations showed that the effects of damage caused by moisture altered seed vigor and viability, depending on conditions, packaging and storage time.

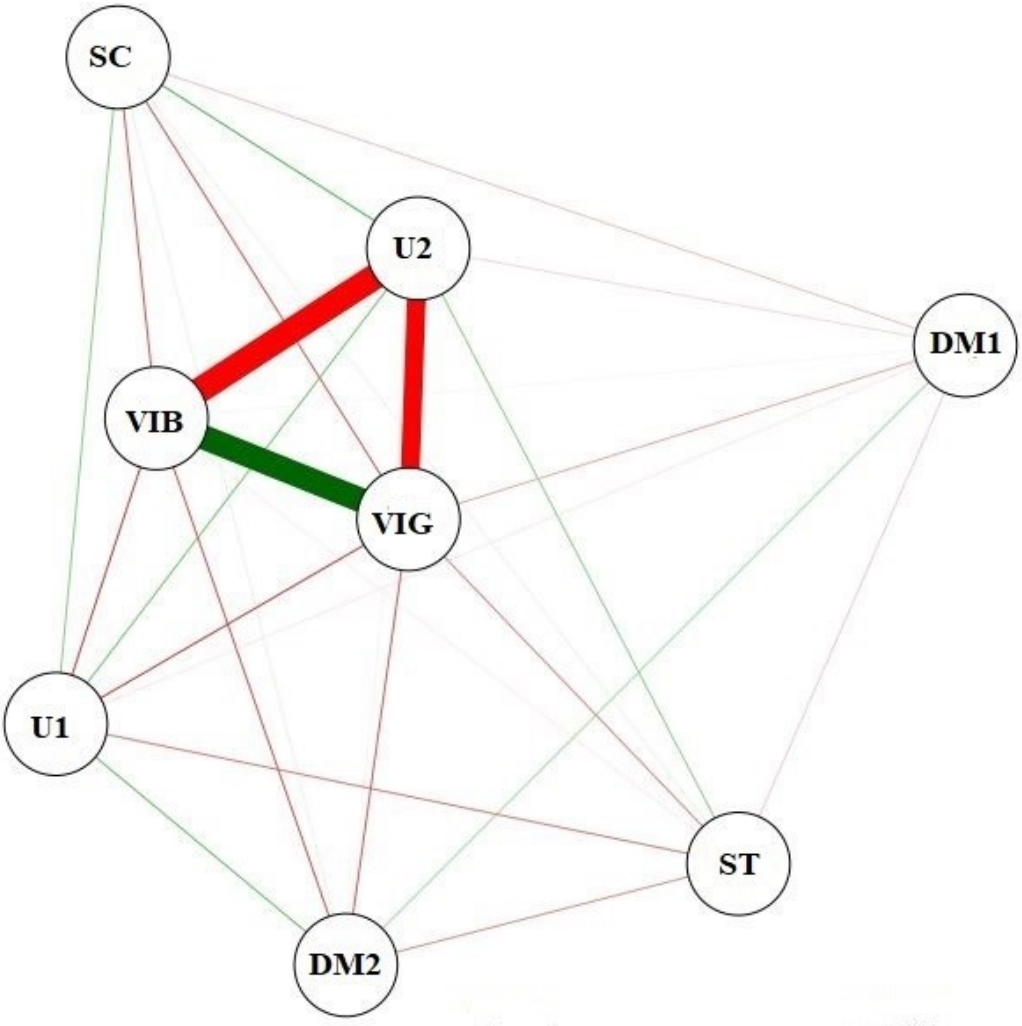

FIGURE 3. Pearson correlation network between storage and quality variables of soybean seeds.

Machine Learning models

Table 2 shows the *p*-value results for Pearson's correlation coefficient (r) and the mean absolute errors (MAE), resulting from the prediction of the variables analyzed by the ANN, MLR, M5P and RF models. Significant interactions (p-value<0.05) were observed for variables U1, U2, VIG and VIB.

TABLE 2. *p*-value of analysis of variance for r and MAE between observed and estimated values of mechanical damage 1 (DM1), mechanical damage 2 (DM2), moisture 1 (U1), moisture 2 (U2), vigor (VIG) and viability (VIB) of soybean seeds stored by Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), M5P decision tree Algorithm (M5P) and Random Forest (RF) models.

Variables	r	MAE
DM1	0.417	0.008
DM2	0.995	0.703
U1	0.002	0.126
U2	0.000	0.000
VIG	0.000	0.000
VIB	0.000	0.000

Mechanical damage 1 (DM1), mechanical damage 2 (DM2), moisture 1 (U1), moisture 2 (U2), vigor (VIG), and viability (VIB), Pearson's correlation coefficient (r) and mean absolute error (MAE).

Mechanical damage assessment

In the evaluation of mechanical damage (DM1), at viability levels from 1 to 8 (Figure 4), it was found that the seeds lost physiological potential (Neve et al., 2016) confirmed by Oliveira et al. (2021), when they found that the mechanical damage caused by storage affected the germination of soybean seeds. The researchers observed that the progressive increase in mechanical damage to the tegument also reduced the vigor of the soybean seeds.

The tetrazolium test assessed mechanical damage to the seeds, but the subjective analysis varied in its results (Pereira et al., 2020). To predict the results, Machine Learning models were used. Thus, it was observed that Pearson's correlation coefficient (r = 0.64) did not present statistical differences between them, while the mean absolute error (MAE = 11.2) indicated that the Randon Forest (RF) model was the one that statistically differed from the others (Figure 4).

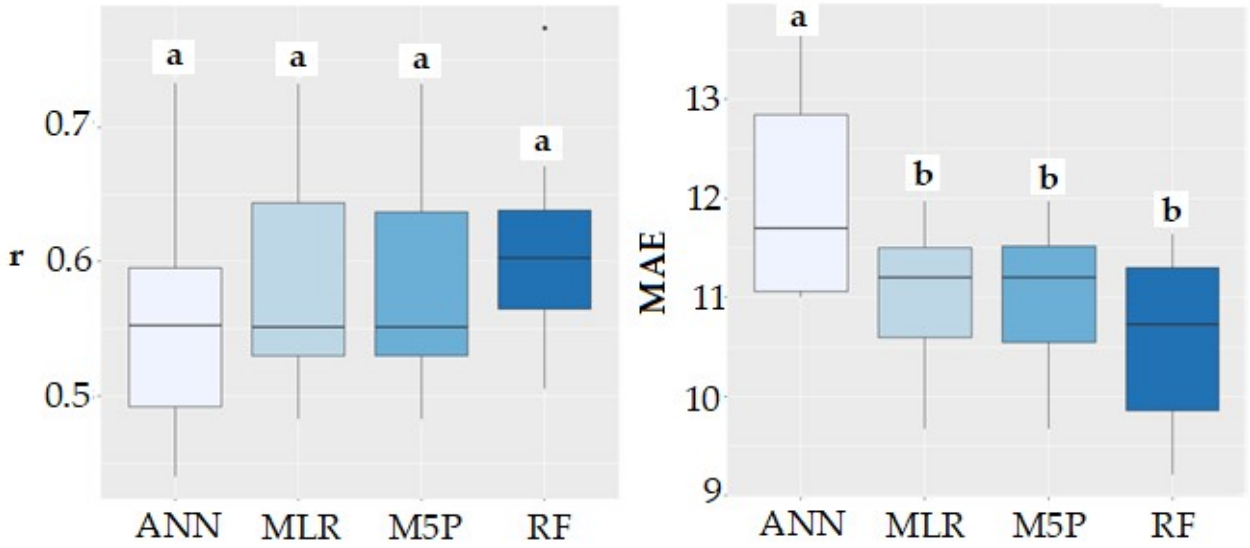

FIGURE 4. Boxplots of mechanical damage assessment (DM1) for Pearson's correlation coefficient (r) and mean absolute error (MAE) for different Machine Learning (ML) models. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), M5P decision tree Algorithm (M5P) and Random Forest (RF). Averages followed by the small letter to compare the results on the line by the Scott-Knott test at 5% probability.

The results were similar in the analysis of mechanical damage (DM2), for viability levels between 6 and 8, when there were no statistical differences between the prediction models. Furthermore, the r values were very low (r = 0.21) (Figure 5). Still, RF model (r = 0.26) fitted the data. The application of Machine Learning models helped to improve the accuracy of predicting soybean seed vigor (Batarseh et al., 2021; Coradi et al., 2022). According to a study carried out by Ropelewska & Piecko (2022), the Artificial Neural Networks (ANN) method had high accuracy to determine seed viability.

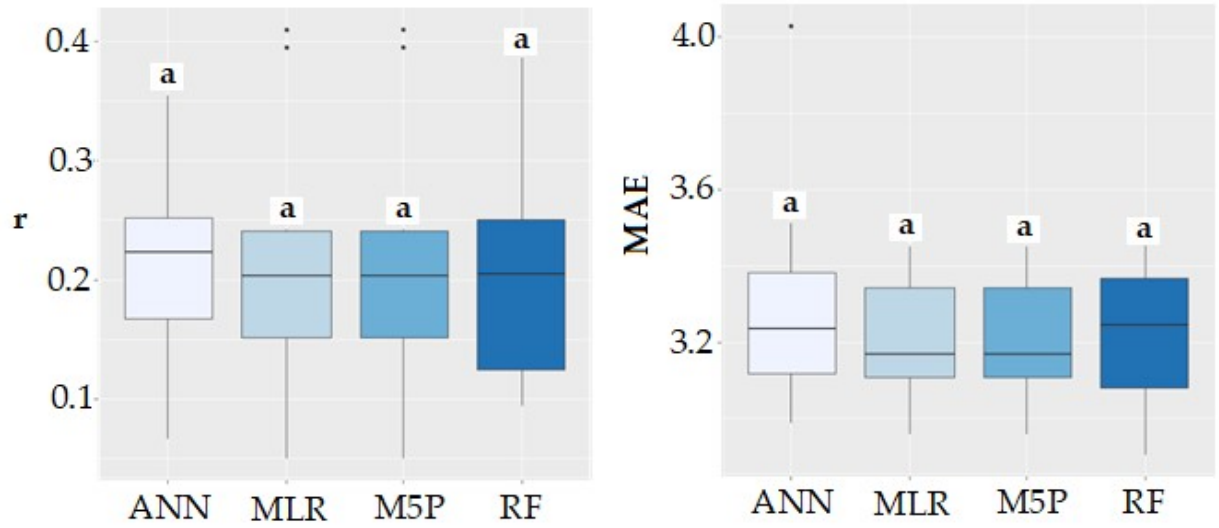

FIGURE 5. Boxplots of mechanical damage assessment (DM2) for Pearson's correlation coefficient (r) and mean absolute error (MAE) for different Machine Learning (ML) models. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), M5P decision tree algorithm (M5P) and Random Forest (RF). Averages followed by the small letter to compare the results on the line by the Scott-Knott test at 5% probability.

This variation also occurred in this study with soybean seeds, however M5P decision tree and Random Forest (RF) algorithms were more accurate due to the smaller statistical errors (r and MAE), considering the different conditions of temperature, packaging and storage time. Škrubej et al. (2015) evaluated a computer vision system, based on image processing and machine learning techniques, for automatic evaluation of seed germination rate. The results indicated that the artificial neural network model performed better than other models. However, Xu et al. (2021) evaluated the use of machine vision and machine learning models to develop a rapid seed detection method based on variety purity, where the results indicated that the support vector machine model was the most accurate.

Moisture evaluation

In the moisture analysis 1 (Figure 6), the M5P, RF and ANN models had averages of r, ranging from 0.45 and 0.41, while the MLR model reached an r = 0.32.

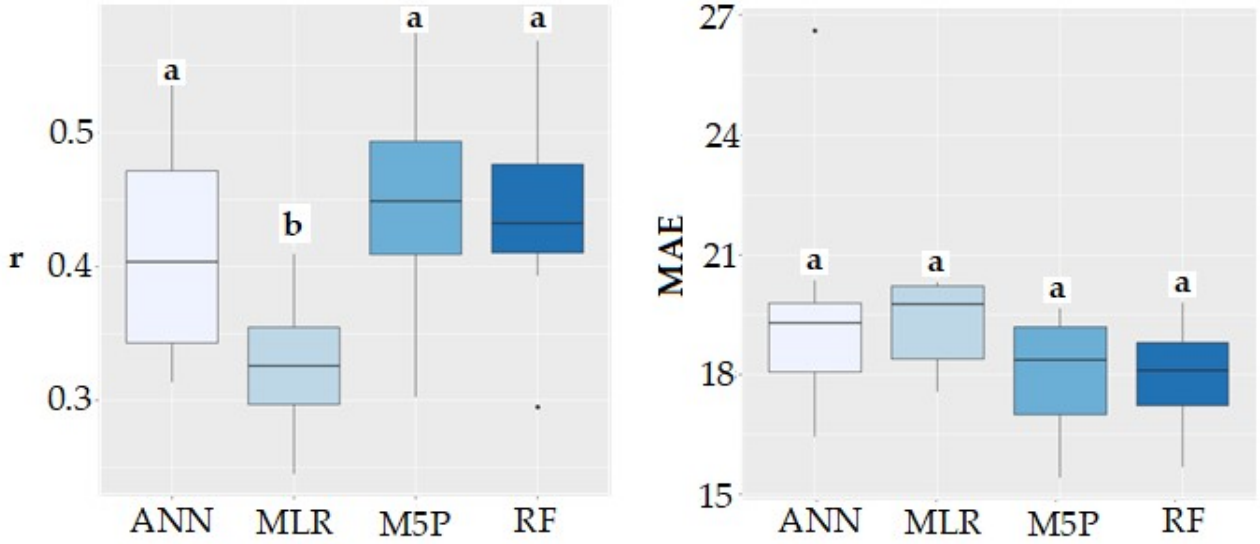

FIGURE 6. Moisture evaluation 1 (U1) boxplots for Pearson's correlation coefficient (r) and mean absolute error (MAE) for different Machine Learning (ML) models. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), M5P decision tree (M5P) and Random Forest (RF). Averages followed by the small letter to compare the results on the line by the Scott-Knott test at 5% probability.

The MAE results showed no statistically significant differences between the models, ranging from 19.53 to 18.04. For the variable moisture 2 (Figure 7) the r values were satisfactory for the RF, M5P and ANN models, not presenting statistical differences between them and varying between 0.87 and 0.85. For the MAE, the average results were 3.37 for the RF, M5P and ANN models, while for the MLR model the MAE value reached 6.17. Thus, the RF, M5P and ANN methods were superior and statistically equal according to Pereira et al. (2020) and Coradi et al. (2022).

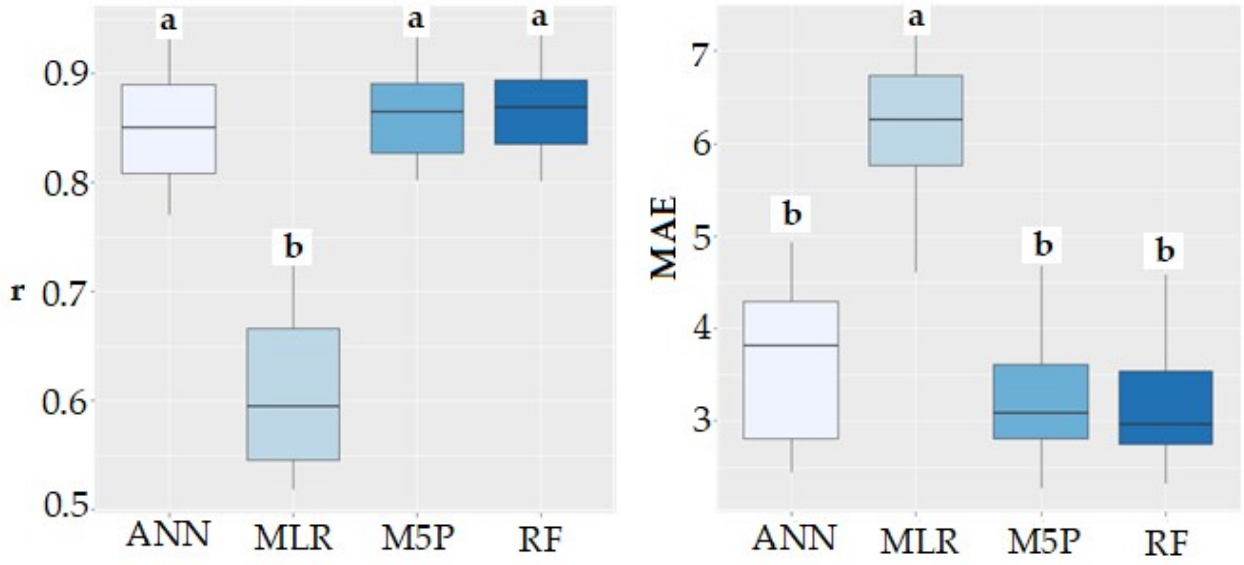

FIGURE 7. Moisture evaluation 2 (U2) boxplots for Pearson's correlation coefficient (r) and mean absolute error (MAE) for different Machine Learning (ML) models. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), M5P decision tree (M5P) and Random Forest (RF). Averages followed by the small letter to compare the results on the line by the Scott-Knott test at 5% probability.

Vigor analysis

In the vigor analysis, the M5P and RF models achieved the best results of r (0.75), while the ANN model obtained an r = 0.74 and the RLM model an r = 0.52. MAE results were lower (10.0) for the M5P and RF models, not statistically different (Figure 8). André et al. (2022) analyzed the performance of Machine Learning algorithms based on variables monitored during seed conditioning and storage time to predict the physical and physiological quality of stored soybean seeds. Among the results, the authors observed that germination had the best results in the ANN, REPTree, M5P and RF models. In the analysis of soybean seed viability, considering the average absolute error of the variable, it was verified that the RF and M5P models presented the smallest errors (Figure 9).

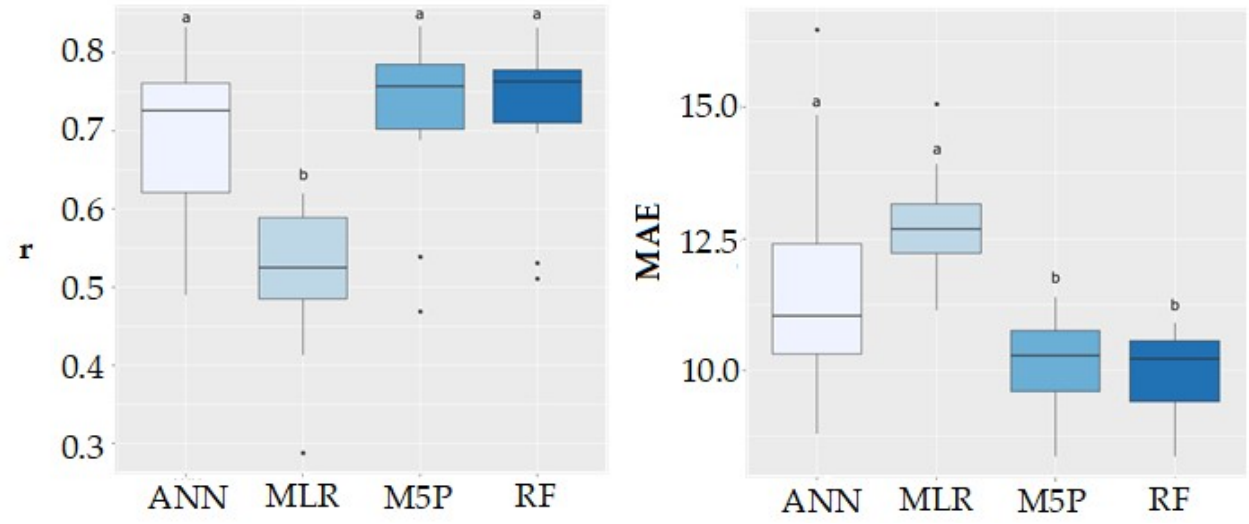

FIGURE 8. Vigor evaluation (VIG) boxplots for Pearson's correlation coefficient (r) and mean absolute error (MAE) for different Machine Learning (ML) models. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), M5P decision tree (M5P) and Random Forest (RF). Averages followed by the small letter to compare the results on the line by the Scott-Knott test at 5% probability.

Viability assessment

The amount of information generated by the high number of analyzes could be better predicted and interpreted by the RF model. Similar results were found by Gadotti et al. (2022b), who evaluated the quality of soybean seeds in different soybean cultivars using machine learning techniques. According to the authors, the random forest model obtained the highest precision, in agreement with the results obtained in this study. Confirming the efficiency of the Randon Forest model, Ropelewska & Piecko (2022) evaluated discriminant models to predict seed quality based on texture parameters of the outer surface of seeds, calculated from images converted into individual color channels. In all cases, the seeds were better and more accurately discriminated by the Randon Forest model.

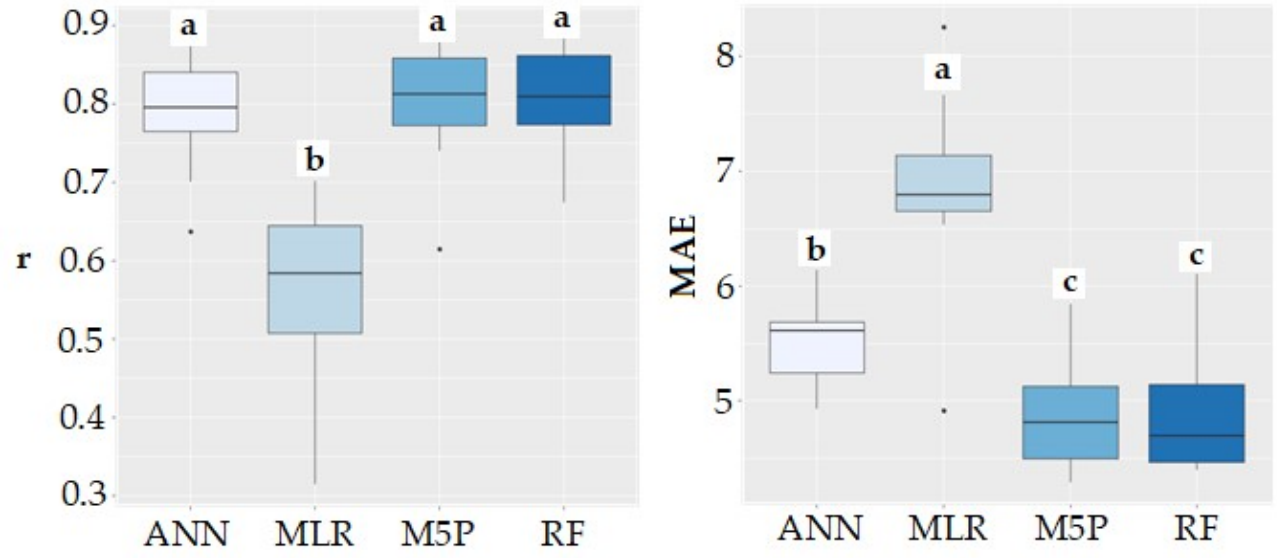

FIGURE 9. Viability assessment (VIB) boxplots for Pearson's correlation coefficient (r) and mean absolute error (MAE) for different Machine Learning (ML) models. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), M5P decision tree (M5P) and Random Forest (RF). Averages followed by the small letter to compare the results on the line by the Scott-Knott test at 5% probability.

The identification of the best model for predicting the quality of stored soybean seeds provided a reduction in errors that could commonly occur in laboratory analyzes based on visual estimates. In addition, the application of Machine Learning models stands out due to the speed of obtaining results in relation to routinely used analysis methods, making processes more robust and of low operational cost. With this, the use of intelligent algorithms can be an auxiliary computational tool in making decisions about seed storage time, contributing to the conservation of quality and reduction of losses. For future researches, it is suggested to expand the storage conditions, taking into account some variables, such as the relative humidity of the storage air, as well as the equilibrium moisture content condition of the seeds to predict and justify the physiological alterations of the seeds.

CONCLUSIONS

The combination of input variables satisfactorily predicted the quality of soybean seeds under different types of packaging, conditions and storage time. The packaging effect was suppressed by low temperatures and storage time, allowing the same results to be achieved but with a smaller number of input variables. Thus, Machine Learning techniques outperformed the proposed control model (multiple linear regression). In addition, M5P decision tree (M5P) and Random Forest (RF) models showed the best performance for predicting seed vigor (r = 0.75 and MAE = 10.0), and viability (r = 0.85 and MAE = 5.1), and mechanical damage to seeds (r = 0.64 and MAE = 11.2). It was concluded that the Random Forest (RF) model was the one that best predicted the results of soybean seed quality, with a more simplified and agile analysis for the development of vigor and viability of soybean seeds in storage, enabling a better handling of overfitting problems.

REFERENCES

André GS, Coradi PC, Teodoro LPR, Teodoro PE (2022) Predicting the quality of soybean seeds stored in different environments and packaging using machine learning. Scientific Reports 12(1):1-13. https://doi.org/10.1038/s41598-022-12863-5

Baio FHR, Santana DC, Teodoro LPR, Oliveira ICd, Gava R, de Oliveira JLG, Silva Junior CAd, Teodoro PE, Shiratsuchi LS (2023) Maize Yield Prediction with Machine Learning, Spectral Variables and Irrigation Management. Remote Sensing 15(1):79. https://doi.org/10.3390/rs15010079

Baryshev DD, Barysheva NN, Pronin SP, Nikol'skii OK (2020) Comparison of machine learning methods for solving the problem of wheat seeds classification by yield properties. Russian Agricultural Sciences 46(4):410-417. https://doi.org/10.3103/S1068367420040047

Batarseh FA, Freeman L, Huang CH (2021) A survey on artificial intelligence assurance. Journal of Big Data 8(1):1-30. https://doi.org/10.1186/s40537-021-00445-7

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: a review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114:24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D (2021) Machine Learning in agriculture: a comprehensive updated review. Sensors 21(11):3758. https://doi.org/10.3390/s21113758

Bhering LL (2017) Rbio: a tool for biometric and statistical analysis using the R platform. Crop Breeding and Applied Biotechnology 17:187-190. https://doi.org/10.1590/1984-70332017v17n2s29

Blaifi SA, Moulahoum S, Benkercha R, Taghezouit B, Saim A (2018) M5P model tree based fast fuzzy maximum power point tracker. Solar Energy 163:405-424. https://doi.org/10.1016/j.solener.2018.01.071

Bouckaert RR, Frank E, Hall M, Kirkby R, Reutemann P, Seewald A, Scuse D (2008) Weka manual for version 3-6-0. University of Waikato, Hamilton, New Zealand, 2. Available: https://www.academia.edu/download/57056677/WekaManual-3-6-2.pdf

Brazil (2009) Ministry of agriculture, Livestock and supply. Normative instruction No. 06, of February 16, 2009. Official gazette of the federative republic of Brazil. Brasília, Brazil: Executive Branch.

Capilheira AF, Cavalcante JA, Gadotti GI, Bezerra BR, Hornke NF, Villela FA (2019) Storage of soybean seeds: packaging and modified atmosphere technology. Revista Brasileira de Engenharia Agrícola e Ambiental 23:876-882. https://doi.org/10.1590/1807-1929/agriambi.v23n11p876-882

Coradi PC, Lima RE, Alves CZ, Teodoro PE, Cândido ACDS (2020) Evaluation of coatings for application in raffia big bags in conditioned storage of soybean cultivars in seed processing units. Plos One 15(11):e0242522. https://doi.org/10.1371/journal.pone.0242522

Coradi PC, Lutz É, Bilhalva NS, Jaques LBA, Leal MM, Teodoro LPR (2022) Prototype wireless sensor network and Internet of Things platform for real-time monitoring of intergranular equilibrium moisture content and predict the quality corn stored in silos bags. Expert Systems With Applications 208:118242. https://doi.org/10.1016/j.eswa.2022.118242

Ebone LA, Caverzan A, Tagliari A, Chiomento JLT, Silveira DC, Chavarria G (2020) Soybean seed vigor: uniformity and growth as key factors to improve yield. Agronomy 10:1–15. https://doi.org/10.3390/agronomy10040545

Egmont-Petersen M, de Ridder D, Handels H (2002) Image processing with neural networks-a review. Pattern Recognition 35(10):2279-2301. https://doi.org/10.1016/S0031-3203(01)00178-9

Ferreira FC, Villela FA, Meneghello GE, Soares VN (2017) Cooling of soybean seeds and physiological quality during storage. Journal of Seed Science 39:385–392. https://doi.org/10.1590/2317-1545v39n4177535

França-Neto JDB, Krzyzanowski FC (2019) Tetrazolium: an important test for physiological seed quality evaluation. Journal of Seed Science 41:359-366. https://doi.org/10.1590/2317-1545v41n3223104

Gadotti GI, Moraes NA, Silva JGD, Pinheiro RDM, Monteiro RD (2022a) Prediction of ranking of lots of corn seeds by artificial intelligence. Engenharia Agrícola 42(4):e20210005. https://doi.org/10.1590/1809-4430-Eng.Agric.v42n4e20210005/2022

Gadotti GI, Ascoli CA, Bernardy R, Monteiro RD, Pinheiro RDM (2022b) Machine Learning for soybean seeds lots classification. Engenharia Agrícola 42(nepe):20210101. https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210101/2022

Gava R, Santana DC, Cotrim MF, Rossi FS, Teodoro LPR, da Silva Junior CA, et al. (2022) Soybean cultivars identification using remotely sensed image and machine learning models. Sustainability 14:7125. Available: https://www.mdpi.com/2071-1050/14/12/7125

Goyal S (2014) Artificial neural networks in fruits: a comprehensive review. International Journal of Image, Graphics and Signal 6:53-63. https://www.mecs-press.org/ijigsp/ijigsp-v6-n5/IJIGSP-V6-N5-7.pdf

Granholm V, Noble WS, Käll L (2012) A cross-validation scheme for machine learning algorithms in shotgun proteomics. BMC Bioinformatics 13 (Suppl 16): S3. https://doi.org/10.1186/1471-2105-13-S16-S3

Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, Ramkumar PN (2020) Machine learning and artificial intelligence: definitions, applications, and future directions. Current Reviews in Musculoskeletal Medicine 13(1):69-76. https://doi.org/10.1007/s12178-020-09600-8

Hussain L, Ajaz R (2015) Seed classification using machine learning techniques. Journal of Engineering Science and Technology 2:1098-1102.

Jaques LBA, Coradi PC, Müller A, Rodrigues HE, Teodoro LPR, Teodoro PE, Steinhaus JI (2022) Portable-mechanical-sampler system for real-time monitoring and predicting soybean quality in the bulk transport. IEEE Transactions on Instrumentation and Measurement 71:1-12. https://doi.org/10.1109/TIM.2022.3204078

Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE (2016) Random forests for global and regional crop yield predictions. Plos One 11(6):e0156571. https://doi.org/10.1371/journal.pone.0156571

Kong F, Chang SKC, Liu Z, Wilson LA (2008) Changes of soybean quality during storage as related to soymilk and tofu making. Journal of Food Science 73(3):S134-S144. https://doi.org/10.1111/j.1750-3841.2007.00652.x

Kumar S, Deswal S (2020) Estimation of phosphorus reduction from wastewater by artificial neural network, Random Forest and M5P model tree approaches. India Pollut 6:427–438.

Liu X, Li B, Shen D, Cao J, Mao B (2017) Analysis of grain storage loss based on decision tree algorithm. Procedia Computer Science 122:130–137. https://doi.org/10.1016/j.procs.2017.11.351

Lutz É, Coradi PC (2022) Applications of new technologies for monitoring and predicting grains quality stored: Sensors, Internet of Things, and Artificial Intelligence. Measurement 188:110609. https://doi.org/10.1016/j.measurement.2021.110609

Lutz É, Coradi PC, Jaques LBA, Carneiro LO, Teodoro LPR, Teodoro PE, de Souza GAC (2022) Real-time equilibrium moisture content monitoring to predict grain quality of corn stored in silo and raffia bags. Journal of Food Process Engineering e14076. https://doi.org/10.1111/jfpe.14076

Medeiros RA, Farias VSO, de Oliveira TMQ, Silva Junior AF, Lima ARN, Pereira MTL, Ataide JSP (2020) Drying behavior of melon (Cucumis Melo L.) seeds in thin layer using empirical models. Brazil Journal Development 6(8):64001-64009. https://doi.org/10.34117/bjdv6n8-725

Moreti MP, Oliveira T, Sartori R, Caetano W (2021) Artificial intelligence in agribusiness and the challenges for the protection of intellectual property. Prospect Not 14(60).

Mylona K, Magan N (2011) *Fusarium langsethiae*: Storage environment influences dry matter losses and T2 and HT-2 toxin contamination of oats. Journal of Stored Prod Research 47:321–327. https://doi.org/10.1016/j.jspr.2011.05.002

Mylona K, Sulyok M, Magan N (2012) Relationship between environmental factors, dry matter loss and mycotoxin levels in stored wheat and maize infected with *Fusarium* species. Food Additives & Contaminants: Part A 29:1118-1128. https://doi.org/10.1080/19440049.2012.672340

Neve JM, Oliveira JA, Silva HPD, Reis RDG, Zuchi J, Vieira AR (2016) Quality of soybean seeds with high mechanical damage index after processing and storage. Revista Brasileira de Engenharia Agrícola e Ambiental 20:1025-1030. https://doi.org/10.1590/1807-1929/agriambi.v20n11p1025-1030

Nyabako T, Mvumi BM, Stathers T, Mlambo S, Mubayiwa M (2020) Predicting *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) populations and associated grain damage in smallholder farmers' maize stores: A machine learning approach. Journal of Stored Products Research 87:01592. https://doi.org/10.1016/j.jspr.2020.101592

Oliveira GRFD, Cicero SM, Krzyzanowski FC, Gomes-Junior FG, Batista TB, França-Neto JDB (2021) Treatment of soybean seeds with mechanical damage: effects on their physiological potential. Journal of Seed Science 43. https://doi.org/10.1590/2317-1545v43247404

Osco LP, Paula A, Ramos M, Pereira DR, Akemi E, Moriya S, Matsubara ET (2019) Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. Remote Sensing 11(24):2925–2942. https://doi.org/10.3390/rs11242925

Pazoki A, Pazoki Z (2011) Classification system for rain fed wheat grain cultivars using artificial neural network. African Journal of Biotechnology 10. https://doi.org/10.5897/AJB11.488

Pereira DF, Bugatti PH, Lopes FM, Souza ALSM, Saito PTM (2020) Assessing active learning strategies to improve the quality control of the soybean seed vigor. IEEE Transactions on Industrial Electronics 68(2):1675-1683. https://doi.org/10.1109/TIE.2020.2969106

Pinheiro RM, Gadotti GI, Monteiro RDCM, Bernardy R (2021) Inteligência artificial na agricultura com aplicabilidade no setor sementeiro. Diversitas Journal 6(3):2996-3012. https://doi.org/10.48017/Diversitas_Journal-v6i3-1857

Quinlan JR (1993) C4. 5: Programming for machine learning. Morgan Kauffmann, pp.38:49.

Radhika V, Rao V (2014) Computational approaches for the classification of seed storage proteins. Journal of Food Science and Technology 52:4246-4255.

Ramos APM, Osco LP, Furuya DEG, Gonçalves WN, Cordeiro DC, Pereira LRT, Junior CAS, Silva GFC, LI J, Baio FHR, Junior JM, Teodoro PE, Pistori H (2020) A Random Forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices. Computer Electronics in Agriculture 178. https://doi.org/10.1016/j.compag.2020.105791

Rocha GC, Neto AR, Cruz SJS, Campos GWB, Castro ACO, Simon GA (2017) Physiological quality of treated and stored soybean seeds. Multi-Science Journal 4:50-65.

Ropelewska E, Piecko J (2022) Discrimination of tomato seeds belonging to different cultivars using machine learning. European Food Research and Technology 248(3):685-705. https://doi.org/10.1007/s00217-021-03920-w

Santana DC, Teodoro LPR, Baio FHR, dos Santos RG, Coradi PC, Biduski B, et al. (2023a) Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. Remote Sensing Applications: Society and Environment 100919. https://doi.org/10.1016/j.rsase.2023.100919

Santana DC, Santos RGd, da Silva PHN, Pistori H, Teodoro LPR, Poersch NL, de Azevedo GB, de Oliveira Sousa Azevedo GT, da Silva Junior CA, Teodoro PE (2023b) Machine learning methods for woody volume prediction in eucalyptus. Sustainability 15(14):10968. https://doi.org/10.3390/su151410968

Sarath KLL, Goneli ALD, Filho CPH, Masetto TE, Oba GC (2016) Physiological potential of peanut seeds submitted to drying and storage. Journal of Seed Science 38:233–240. https://doi.org/10.1590/2317-1545v38n3165008

Silva AM, Figueiredo JC, Tunes LV, Gadotti GI, Rodrigues DB, Capilheira AF (2022) Chickpea seed storage in different packagings, environments and periods. Revista Brasileira de Engenharia Agrícola e Ambiental 26:649-654. https://doi.org/10.1590/1807-1929/agriambi.v26n9p649-654

Škrubej U, Rozman Č, Stajnko D (2015) Assessment of germination rate of the tomato seeds using image processing and machine learning. European Journal of Horticultural Science 80(2):68-75. http://dx.doi.org/10.17660/eJHS.2015/80.2.4

Souza RS, Lopes JLB, Geyer CFR, João LDRS, Cardozo AA, Yamin AC, Gadotti GI, Barbosa JLV (2019) Continuous monitoring seed testing equipaments using internet of things. Computers and Electronics in Agriculture 158:122-132. https://doi.org/10.1016/j.compag.2019.01.024

Team RCR (2018) A language and environment for statistical computing.

Vo-Thanh H, Amar MN, Lee KK (2022) Robust machine learning models of carbon dioxide trapping indexes at geological storage sites. Fuel 316: 123391. https://doi.org/10.1016/j.fuel.2022.123391

Xu P, Yang R, Zeng T, Zhang J, Zhang Y, Tan Q (2021) Varietal classification of maize seeds using computer vision and machine learning techniques. Journal of Food Process Engineering 44(11):e13846. https://doi.org/10.1111/jfpe.13846