AGRONOMY SCIENCE AND BIOTECHNOLOGY

RESEARCH ARTICLE

Classification of soybean cultivars by means of artificial neural networks

João Victor Costa Carneiro da Paixão[1], Éder Matsuo[2,*], Ithalo Coelho de Sousa[3], Moysés Nascimento[4], Igor Silva Oliveira[1], Adalberto Filipe Macedo[1] and Gustavo Martins Santana[1]

[1]Agricultural Sciences Institute, Federal University of Viçosa, Rio Paranaíba Campus, Highway MG 230, Km7, PO Box 22, Rio Paranaiba, MG, Brazil, CEP 38810-000. [2]Technological and Exact Sciences Institute, Federal University of Viçosa, Rio Paranaíba Campus, Biostatistics Laboratory, Highway MG 230, Km7, PO Box 22, Rio Paranaiba, MG, Brazil, CEP 38810-000. [3]Department of Mathematics and Statistics, Federal University of Rondônia (UNIR), Ji-Paraná Campus, Ji-Paraná, RO, Brazil, CEP 76900-726. [4]Department of Statistic, Federal University of Viçosa, Bioagro, Computational Intelligence Laboratory and Statistical Learning (LICAE), Peter Henry Rolfs Avenue, s/n, Viçosa, MG, Brazil, CEP 36570-900. *Corresponding author, E-mail: edermatsuo@ufv.br

Check for updates

OPEN ACCESS

How to cite: Paixão, J. V. C. C., Matsuo, E., Souza, I. C., Nascimento, M., Oliveira, I. S., Macedo, A. F., & Santana, G. M. (2023). Classification of soybean cultivars by means of artificial neural networks. *Agronomy Science and Biotechnology*, 9, 1-11 https://doi.org/10.33158/ASB.r186.v9.2023

Received: March 14, 2023.
Accepted: March 23, 2023.
Published: May 31, 2023.

English by: André Luis Miyagaki

Copyright: © 2023 Agronomy Science and Biotechnology. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, since the original author and source are credited.

ABSTRACT

The cultivation of soy has an economic importance for the Brazilian agricultural scenario. The aim of this study was to establish a network architecture for the classification of soybean genotypes, by means of morphological characters measured in the juvenile phase of the plant, and finally to compare the results obtained through Artificial Neural Network (ANN) and Anderson Discriminant Analysis. The study analyzed plants of 10 conventional cultivars in the initial stages of development (V1, V2 and V3 stages). The experiment was carried out in a randomized block design with 5 replications, and the experimental unit was represented by 9 plants. The data were submitted to the Anderson Discriminant Analysis and multilayer Perceptron ANN, with 1 or 2 hidden layers. To analyze the homogeneity of the variance and covariance matrix, the Box's M-Test was adopted in the Program R, at 5% significance level. An input layer, one or two hidden layers, and an output layer formed the ANN architecture. The 5-fold cross validation was used to verify the efficiency of the discriminant functions and also in the ANN analysis. Subsequently, the apparent error rate (AER) was obtained. Box's M-Test indicated inhomogeneity in the variance and covariance matrices, which indicated the need to perform Anderson's Quadratic Discriminant Analysis. The ANNs presented lower apparent error rate when compared to the Anderson's Quadratic Discriminant Analysis and the artificial neural network with 1 hidden layer was sufficient to perform the classification of soybean cultivars.

Keywords: *Glycine max*, discriminant analysis, multilayer perceptron, neurophysiological character, soybean breeding, adaptability, phenotypic stability.

INTRODUCTION

The soybean crop (*Glycine max* (L.) Merr.) is of economic importance to the Brazilian scenario, for being a product that occupies the agricultural area of the country, approaching 40.7 million hectares planted in the 2021/22 harvest (Companhia Nacional de Abastecimento [CONAB], 2022). The prices of agricultural commodities have highlighted the importance of the sector and soybeans have shown to be a key product of the Brazilian nation. In 2020 an increase of 10.2% was recorded in comparison to 2019 and in 2021 a growth of 35.3% in the international agricultural market, configuring itself as an important product in the economic recovery in the post-pandemic scenario, especially for a developing country that stands out as a major exporter of the grain under discussion (CONAB, 2021).

What makes soy a valued *commodity* is the diversity of products that use it as raw material, ranging from animal to human nutrition. There are several products based on soy for human consumption, which added to the propagation of its health benefits make it even more important in the global agricultural scenario, culminating in greater adhesion and demand by the consumer as the nutritional qualities of the product are evidenced to consumers (Behrens et al., 2001).

Investments in research and studies that provide the classification of genotype through the evaluation/measurement of different plant characteristics is of fundamental importance for the genetic improvement of soybeans. Artificial neural networks (ANN) had their first form of work in 1949, being made from observations of neurophysiological character (Hebb, 1949). The biological nervous system served as a model in the development of artificial neural networks, being a simplification of biological neural networks. Its application can encompass several areas, such as medical, chemical, ecological, biological and automotive areas, having an extensive possibility of application directed to agricultural sciences, configuring itself as a tool that acts in the scope of universal function approximator, process control, pattern classification, data grouping, prediction systems, system optimization and associative memories (Silva, Spatti, & Flauzino, 2010).

The data analyzed by this system can be phenotypic in nature, whose values are expressed in measurements (plant height, epicotyl diameter, petiole length, among others). This method can also be used via images, and the validation of the trained ANN is able to distinguish and identify the tested soybean varieties using images (Khatchatourian et al., 2008). In addition, genotypic data can also be used for RNA analysis (Sousa et al., 2020).

Models were proposed via ANN for the projection of Brazilian soybean production, using data provided by CONAB of the last 41 years of planted areas, evidencing a 5% decrease in the 2017/2018 harvest in relation to the previous harvest (Abraham et al., 2019). The work done by Alves et al. (2019) followed the phenotypic scope using ANN as the primary research tool. The length of the hypocotyl was used as the subject of study, managing to identify soybean genotypes that had stability and predictability in their behavior, thus cooperating for Soybean Improvement (Alves et al., 2019). The adaptability of soy in certain environments is of great importance to achieve high yields. Oda et al. (2022) developed research based on the interaction between genotypes and environment, in which they evaluated the adaptability and phenotypic stability of the productivity of late-cycle crops, and several analyses were performed, including via ANN, classifying the cultivars analyzed in unfavorable adaptability, general or favorable, with high and low stability.

Thus, the aim was to establish a network architecture for soybean genotype classification and compare the obtained results through ANNs and Anderson Discriminant Analysis.

MATERIAL AND METHODS

The experiment was conducted in the experimental field of the Federal University of Viçosa - Rio Paranaíba Campus, in the city of Rio Paranaíba, state of Minas Gerais, in the greenhouse of the Biostatistics Laboratory of the Institute of Technology and Exact Sciences. Ten conventional soybean cultivars were analyzed, namely: BRS Tracajá, MG/BR-46 (Conquista), P98C81, TMG 803, MSOY 8757, BRSGO 8660, BRS 8381, TMG 801, BRSMG 68 [Vencedora] and BRSGO 7560.

The experiment was conducted according to a randomized block design with 5 repetitions (blocks) and each experimental unit was represented by 9 seedlings grown in three pots (3 plant per pots). The seeds were planted 2 cm deep and when the seedlings reached the VC development stage (Fehr & Caviness, 1977) thinning was performed, maintaining three plants per pot.

The evaluations occurred at the early stages of soybean development after the VC stage. In V1 and V2, by Fehr and Caviness (1977), the epicotyl length, plant height, hypocotyl diameter and epicotyl diameter were analyzed. At the V3 stage, by Fehr and Caviness (1977), an addition of variables were. The same ones previously analyzed in V1 and V2, and also the length of the petiole of the first trifoliolate leaf and the length of the rachis of the first trifoliate leaf. The measurements were made using a digital pachymeter for diameters and a millimeter ruler for lengths and heights. Phenotypic characteristics measured at early crop stages are potential to aid in distinguishing cultivars (Nogueira et al., 2008; Matsuo et al., 2012; Nogueira et al., 2019; Hanyu et al., 2020; Camargos et al. 2019; Alves et al. 2019).

The measurements were taken as each seedling reached the stage of development. The diameter of the hypocotyl was measured at one centimeter below the node of cotyledon insertion and that of the epicotyl was measured at the central position of the epicotyl (the central position between the cotyledon node and the node of insertion of the first trifoliolate leaf). The heights were measured at the last visible node of the main stem of the plant. The cultural treatments in the experiment were conducted according to Sediyama (2009). The air temperatures measured inside the greenhouse using a digital thermometer were 31.3 ºC (overall average), 16.1 ºC (average of minimum temperatures) and 44.6 ºC (average of maximum temperatures).

To establish the Anderson discriminant function, it was considered that for a population the observation vector should present a multivariate normal distribution and that it is essential to consider in the classification of observations the prior probability inherent to the various populations evaluated in a given study (Anderson, 1958). In this case, the equal probability was adopted for all populations, that is, a prior probability equal to 0.10 for each cultivar.

Assuming the homogeneity of the matrices of variances and covariances, a common variance and covariance matrix (Global) is used, and the Anderson Linear Discriminant Analysis is performed. For situations in which it is not possible to establish a common matrix of variances and covariances, because the matrices are heterogeneous (heteroscedasticity between them), the Anderson Quadratic Discriminant Analysis is applied (Cruz et al., 2011). Thus, to analyze the homogeneity of the matrix of variances and covariances, Box's M-test for Homogeneity of Covariance Matrices available in the *biotools* package of the R Core Team Program

(Silva, 2021), with a significance level of 5% was used. For the Anderson Discriminant Analysis, the original data were used, i.e., 45 informations per cultivar of the 14 variables which were analyzed/mensured.

Functions were generated (linear or quadratic combinations, depending on the matrix of variances and covariances) to perform the distinction between individuals (Cruz et al., 2011). Subsequently, to verify the efficiency of the discriminant functions in classifying the cultivars correctly, 5-*fold* cross validation was used and the apparent error rate (AER) for each *fold* was obtained.

The Artificial Neural Network (ANN) is a computer intelligence methodology based on models of biological nature, and is used for prediction, classification, and model adjustment issues (Nascimento & Cruz, 2018). Among the various ANN models, the model used in this study is the multilayer perceptron, which, according to Nascimento and Cruz (2018) has the ability to work with non-linearly separable problems. An input layer, one or two hidden layers, and an output layer formed the architecture of the ANN.

The input layer was composed of 14 neurons corresponding to the number of explanatory variables, the amount of neurons in hidden layer was composed of 1 to 20 neurons and the output layer composed of 10 neurons that are correspondent to the 10 cultivars. The neurons are interconnected with all neurons of the adjacent layers, and this connection is defined by synaptic weights; however, the neurons do not connect with other neurons that belong to the same ANN layer (Sousa et al., 2022). Besides the variation of the numbers of neurons in the hidden layers, it was tested six activation functions (linear, sigmoid, ReLU, hyperbolic tangent, ramp, and step) and three types of optimization (SGD, RMSprop, and Adam). In addition to these variations, the 5-*fold* cross-validation was used, totaling 36,000 ANNs per data set analysis.

To choose the best ANN, the apparent error rate (AER) was evaluated according to each analyzed architecture. For this analysis, the original data sets were used (45 information for each cultivar) and the augmented data were used as the basis for the data augmentation replication through the procedure "Data augmentation replication" of the Genes Program (Cruz, 2013) to obtain the simulated data sets with 200, 400, 600 and 800 information/population. Thus, the data used in the analyses were: Original (45 information/class), Amp_200 (200 information/class), Amp_400 (400 information/class), Amp_600 (600 information/class), and Amp_800 (800 information/class.

In order to visualize the dispersion of the cultivars in graphs, a graphical analysis was performed and using principal components considering the minimum number of eigenvalues that represented at least 80% of the total variation.

The analyses were performed in the Genes Program (Cruz, 2013) and the R Program (R Core Team, 2022).

RESULTS AND DISCUSSION

The Box's M-Test indicated non-homogeneity ($p<0.05$) in the matrices of variances and covariances. Thus, the Anderson Quadratic Discriminant Analysis was performed (Cruz et al., 2011). The analyses performed using the Anderson Quadratic Discriminant Analysis and ANN with 1 or 2 hidden layers provided different values of Apparent Error Rate (AER) (Table 1). The methodology that presented the highest AER was the Discriminant Analysis, reaching 0.59.

The analyses which used the Multilayer Perceptron ANN with 1 hidden layer resulted in different AER, according to the database which. Was used data and Amp_200 obtained 0.40 of AER, the subsequent augmentations, Amp_400 and Amp_600 also shared an equality between them, reaching 0.32. The ANN model with Amp_800 was the one that stood out for displaying the lowest rate, measuring 0.27 (Table 1).

Table 1. Network architecture, activation and optimization function, apparent error rate (AER) according to the techniques of the Anderson Quadratic Discriminant Analysis (ADQA) and artificial neural networks (ANN).

Methodology	Data	Number of hidden layer	Number of neurons in hidden layer	Activation functions[1]	Optimization function[2]	AER
ADQA	Original	N.A.[3]	N.A.	N.A.	N.A.	0.59
ANN	Original	1	12	Linear	adam	0.40
ANN	Original	2	17 e 20	Linear	sgd	0.38
ANN	Amp_200	1	10	ReLU	adam	0.40
ANN	Amp_200	2	11 e 9	ReLU	sgd	0.38
ANN	Amp_400	1	13	ReLU	sgd	0.32
ANN	Amp_400	2	17 e 17	Sigmoid	RMSprop	0.32
ANN	Amp_600	1	16	ReLU	sgd	0.32
ANN	Amp_600	2	10 e 15	ReLU	RMSprop	0.32
ANN	Amp_800	1	19	ReLU	sgd	0.27
ANN	Amp_800	2	17 e 15	ReLU	sgd	0.27

[1]Linear: linear function; ReLU: *Rectified Linear Unit*; and Sigmoid: logistic or sigmoid function. [2]Adam: *Adaptive Moment Estimation*; sgd: *Stochastic Gradient Descent*; e RMSprop: *Root Mean Square Prop*. [3]N.A.: does not apply.

All analyzed scenarios involving the use of the Multilayer Perceptron ANN in a 1 or 2 hidden layer configuration were superior (lower AER) in comparison to the results of the Discriminant Analysis, regardless of the utilized amplification. Therefore, there is a tendency that this technique can classify cultivars more accurately than the Anderson Quadratic Discriminant Analysis.

Among the ANN related analyses, there is a tendency for greater accuracy in the classification performed with data sets containing a larger number of individuals per class (800 individuals/class) regardless of whether there are one or two hidden layers.

The analyses involving the use of the Multilayer Perceptron ANN with 2 hidden layers achieved lower AER values when compared to Anderson's Quadratic Discriminant Analysis (Table 1). With the Original data and Amp_200 the AER reached 0.38. The values that were obtained were lower the higher the magnification of the data, i.e. Amp_400 and Amp_600 reached 0.32 and the amplification with Amp_800 had 0.27 of AER.

Comparing the two ANN scenarios that were used (one or two hidden layers) it was possible to verify some existing trends. In principle, both showed lower AER than Anderson's Quadratic Discriminant Analysis, whether with the original or the extended data (Amp_). As for the ANN, the procedure which was used differ substantially, either with 1 or 2 hidden layers, and in several procedures the values obtained were equal or close, and the smallest value obtained was in the Amp_800 amplification both with 1 or 2 hidden layers, and had similar AER values.

The evaluation of a discriminant function is fundamental to verify the reliability of the classification, since the field of genetic improvement needs reliable techniques. The reasons for a high AER in the Anderson's Quadratic Discriminant Analysis technique are covered in the literature. According to Sant'Anna et al. (2018), the reasons for this poor classification are related to the populations, which need to be sufficiently differentiated so that there is a proper distinction, and if there is a proper differentiation it is necessary to have a sufficient quantity and quality of the variables or descriptors analyzed for a correct classification.

The use of only one hidden layer was sufficient to perform the classification of soybean genotypes, as it showed values close or equal to those of two layers, being preferable because it has less complexity to perform the analysis. The number of hidden layers varies according to the necessity. There is the possibility to use three hidden layers in the ANN configuration, which is justified by the high similarity of the data.

Analyses involving neural networks demand intense data processing for the adjustment of a configuration with the number of layers that satisfactorily fulfills is purely empirical, on a trial and error basis (Braga et al., 2007). Based on this premise, it was found that for the present study a network with 1 hidden layer was sufficient.

When analyzing the data from the sum of folds of the correct and incorrect classification of cultivars by the Anderson Square Discriminant Analysis (Table 2) it was observed that for cultivars 9 (BRSMG 68 [Vencedora]), 3 (P98C81) and 2 (MG/BR-46 (Conquista)) were the ones with the lowest frequency of correct classifications, being that from the total of 45 individuals only 10, 12 and 14 individuals were correctly classified. By the analysis of the ANN Perceptron Multicolayer with 1 hidden layer, cultivars 3 (P98C81) and 7 (BRS 8381) had the lowest frequency of correct classifications, with, respectively, 21 and 22 correct classifications. As for multi-layer multilayer Perceptron ANN with two hidden layers, cultivars 9 (BRSMG 68 [Vencedora]), 3 (P98C81), and 2 (MG/BR-46 (Conquista)) had the lowest correct classifications, i.e., 23, 24, and 25.

When analyzing the cultivars with the best number of correct classifications cultivars 10 and 4 stood out by Discriminant Analysis, 1 and 10 by the ANN, with 1 hidden layer and 5 and 8 by the ANN with 2 hidden layers.

It was analized the results together with the scatter plot (Chart 1), obtained through the first two principal components. It was found that cultivars 1 (BRS Tracajá), 4 (TMG 803), 5 (MSOY 8757), 6 (BRSGO 8660), 8 (TMG 801), and 10 (BRSGO 7560) are graphically more distant from the others, which corroborates in the analysis of better separation and their correct classification (Figure 1).

Table 2. Number of correct (on the diagonal) and incorrect (off the diagonal) classifications for each cultivar by means of the Anderson Quadratic Discriminate Analysis, Artificial Neural Network (ANN) with 1 hidden layer and ANN with 2 hidden layers[1].

Anderson Quadratic Discriminate Analysis										
	1	2	3	4	5	6	7	8	9	10
1	*16*	1	8	5	2	4	1	7	1	0
2	4	*14*	2	0	4	2	2	10	3	4
3	2	2	*12*	5	2	6	4	7	2	3
4	1	0	6	*25*	0	6	1	3	1	2
5	2	2	4	0	*22*	2	6	2	4	1
6	0	3	2	4	3	*17*	5	9	1	1
7	3	2	6	0	7	4	*18*	3	1	1
8	3	5	6	4	1	5	2	*19*	0	0
9	3	10	2	5	3	1	4	1	*10*	6
10	0	3	3	1	1	0	0	4	4	*29*
ANN – 1 hidden layer[2]										
	1	2	3	4	5	6	7	8	9	10
1	*33*	1	0	2	1	1	2	4	1	0
2	0	*25*	1	1	3	4	1	3	4	3
3	1	4	*21*	8	2	2	1	4	1	1
4	3	0	3	*28*	0	5	0	0	3	3
5	1	2	1	0	*29*	2	4	0	6	0
6	2	2	4	3	1	*25*	5	3	0	0
7	2	3	2	2	6	2	*22*	2	3	1
8	3	4	4	1	1	4	0	*26*	0	2
9	0	4	3	1	4	0	3	0	*27*	3
10	0	3	1	4	1	0	0	0	2	*34*

ANN – 1 hidden layers [3]										
	1	2	3	4	5	6	7	8	9	10
1	*31*	1	1	2	1	1	3	4	1	0
2	0	*25*	1	1	3	3	3	1	4	4
3	0	2	*24*	6	1	2	2	5	1	2
4	4	1	2	*31*	0	3	0	0	2	2
5	1	2	0	0	*34*	1	5	0	2	0
6	2	2	4	2	1	*26*	4	3	1	0
7	2	1	1	2	6	2	*25*	4	1	1
8	2	3	1	1	1	3	1	*32*	0	1
9	0	4	6	1	6	0	4	0	*23*	1
10	0	2	1	4	1	0	1	0	6	*30*

[1]For each procedure, the values of correct and incorrect classifications refer to the sum of the 5 folds, totaling 45 classifications. [2]The ANN network which was used was multilayer perceptron with one hidden layer (12 neurons), activation function: Linear, and optimization function: adam. [3]The ANN network which was used was multilayer perceptron with two hidden layers (17 and 20 neurons, respectively), activation function: Linear, and optimization function: sgd. Cultivar (classes): 1 = BRS Tracajá, 2 = MG/BR-46 (Conquista), 3 = P98C81, 4 = TMG 803, 5 = MSOY 8757, 6 = BRSGO 8660, 7 = BRS 8381, 8 = TMG 801, 9 = BRSMG 68 [Vencedora] and 10 = BRSGO 7560.

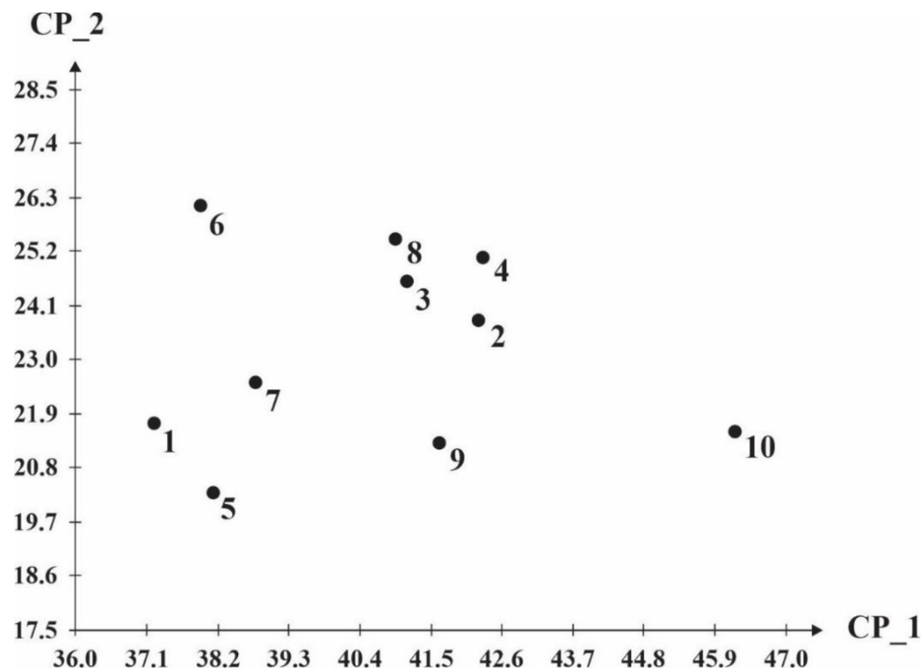

Figure 1. Graphic dispersion of 10 conventional soybean cultivars based on the first two principal components (CP1 and CP2). The first two eigenvalues explained 81.97% of the total variation. Cultivar 1 = BRS Tracajá, 2 = MG/BR-46 (Conquista), 3 = P98C81, 4 = TMG 803, 5 = MSOY 8757, 6 = BRSGO 8660, 7 = BRS 8381, 8 = TMG 801, 9 = BRSMG 68 [Vencedora] and 10 = BRSGO 7560.

In all techniques correct and incorrect classifications were possible, but with variable AER. For the correct classification of the 10 cultivars of the present study with 14 variables, ANNs were superior to the Anderson Quadratic Discriminant Analysis. And, it was found that the lowest AER presented magnitude of 0.27 with ANN.

CONCLUSIONS

The use of the Multilayer Perceptron ANN has greater efficiency when compared with the Anderson Quadratic Discriminant Analysis for the correct classification of conventional soybean cultivars by phenotypic characteristics measured in the early stages of the crop.

The artificial neural network with 1 hidden layer was sufficient to perform the classification of soybean cultivars by phenotypic characteristics measured in the early stages of the crop.

ACKNOWLEDGEMENTS

The authors thank the National Council for Scientific and Technological Development (CNPq) for the financial support.

REFERENCES

Abraham, E. R., Reis, J. G. M., Toloi, R. C., Souza, A. E., & Colossetti, A. P. (2019). Estimativa da produção da soja brasileira utilizando redes neurais artificiais. *Agrarian*, *12*(44), 261-271. doi: 10.30612/agrarian.v12i44.9209

Alves, G. F.; Nogueira, J. P. G.; Machado Junior, R.; Ferreira, S.C.; Nascimento, M.; & Matsuo, E. (2019). Stability of the hypocotyl length of soybean cultivars using neural networks and traditional methods. *Ciência Rural*, *49*(3), e20180300, doi: 10.1590/0103-8478cr20180300

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. New York: John Wiley & Sons.

Behrens, J. H., Roig, S. M., & Silva, M. A. A. P. (2001) Aspectos de Funcionalidade, de Rotulagem e de Aceitação de Extrato Hidrossolúvel de Soja Fermentado e Culturas Lácteas Probióticas. *Boletim SBCTA*, *34*(2), 99-106.

Braga, A. P., Ludermir, T. B., & Carvalho, A. C. P. L. F. (2007). *Redes neurais artificiais: teoria e aplicações.* Rio de Janeiro: LTC.

Camargos, T. V. C., Campos, N. S, Alves, G. F., Ferreira, S. C., & Matsuo, E. (2019). The effect of soil volume, plant density and sowing depth on soybean seedlings characters. *Agronomy Science and Biotechnology*, 5(2), 47. doi: 10.33158/ASB.2019v5i2p47

CONAB - Companhia Nacional de Abastecimento. (2021). *Soja. Perspectivas para a agropecuária*. Safra 2021-2022, edição grãos verão, v. 9, p. 71-84. Brasília, DF: CONAB.

CONAB - Companhia Nacional de Abastecimento. (2022). *Acompanhamento da Safra Brasileira de Grãos.* Safra 2021/22, v. 9, n. 6, sexto levantamento, março 2022. Brasília, DF: CONAB.

Cruz, C. D. (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum. Agronomy*, *35*(3), 271-276. doi: 10.4025/actasciagron.v35i3.21251

Cruz, C. D., Ferreira, F. M., & Pessoni, L. A. (2011). *Biometria aplicada ao estudo de diversidade genética*. Visconde do Rio Branco, MG: Suprema.

Fehr, W.R., & Caviness, C.E. (1977) *Stages of soybean development*. Ames: Iowa State University of Science and Technology.

Hanyu, J., Costa, S. C., Cecon, P. R., & Matsuo, E. (2020). Genetic parameters estimate and characters analysis in phenotypic phase of soybean during two evaluation periods. *Agronomy Science and Biotechnology*, *6*, 1-12. doi: 10.33158/ASB.r104.v6.2020

Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory.* New York, USA: Wiley.

Khatchatourian, O., & Padilha, F. R. R. (2008). Reconhecimento de variedades de soja por meio do processamento de imagens digitais usando redes neurais artificiais. *Engenharia Agrícola*, *28*(4), 759-769. doi: 10.1590/S0100-69162008000400016

Matsuo, É., Sediyama, T., Cruz, C. D., Oliveira, R. C. T., & Cadore, L. R. (2012). Estimates of the genetic parameters, optimum sample size and conversion of quantitative data in multiple categories for soybean genotypes. *Acta Scientiarum. Agronomy*, *34*(3), p. 265-273. doi: 10.4025/actasciagron.v34i3.14015

Nascimento, M., & Cruz, C.D.; (2018). *Inteligência Computacional Aplicada ao Melhoramento Genético*. Viçosa, MG: Editora UFV.

Nogueira, A. P. O., Sediyama, T., Cruz, C. D., Reis, M. S.; Pereira, D. G., & Jangarelli, M. (2008). Novas características para diferenciação de cultivares de soja pela análise discriminante. *Ciência Rural*, *38*(9), 2427-2433. doi: 10.1590/S0103-84782008005000025

Nogueira, J. P. G., Campos, N. S., Camargos, R. V. C., Bomtempo, G. L., Ferreira, S. C., Cecon, P. R., Matsuo, E. (2019). Análises do diâmetro do hipocótilo em plantas de soja. *Nucleus*, *16*(1), 373-384. doi: 10.3738/1982.2278.3551

Oda, M. C., Sediyama, T., Cruz, C. D., Nascimento, M., & Matsuo, E. (2021). Adaptability and yield stability of soybean genotypes by mean Eberhart and Russell methods, artificial neural networks and centroid. *Agronomy Science and Biotechnology*, 8, 1-13. doi:10.33158/ASB.r142.v8.2022

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Sant´Anna, I. C., Carneiro, V. Q., Ferreira, R. A. D. C., Silva, G. N., Nascimento, M., & Cruz, C. D. (2018). *RNA-Aplicação em Estudos Classificatórios.* In: Cruz, C. D., & Nascimento, M. (Eds.). Inteligência computacional aplicada ao melhoramento genético. p. 189-213. Viçosa, MG: Editora UFV.

Sediyama, T. (2009). *Tecnologias de produção e usos da soja*. Londrina, PR: Editora Mecenas.

Silva, A.R. (2021). *Biotools: Tools for Biometry and Applied Statistics in Agricultural Science. R package version 4.2.* Disponível em: <https://cran.r-project.org/package=biotools>

Silva, I. N., Spatti, D. H., & Flauzino. R. A. (2010). *Redes neurais artificiais para engenharia e ciências aplicadas - curso prático.* São Paulo, SP: Artliber Editora.

Sousa, I. C., Nascimento M., Sant'Anna, I. C., Caixeta, E. T., Azevedo, C. F., Cruz, C. D., Silva, F. L., Alkimin, E. R., Nascimentos, A. C. C., & Serrão, N. V. L. (2022). Marker effects and heritability estimates using additive-dominance genomic architectures via artificial neural networks in *Coffea canephora*. *PLoS ONE*, *17*(1): e0262055. doi: 10.1371/journal.pone.0262055