e-ISSN 1983-4063 - *www.agro.ufg.br/pat* - Pesq. Agropec. Trop., Goiânia, v. 51, e66008, 2021

Research Article

Yield prediction of 'Prata Anã' and 'BRS Platina' banana plants by artificial neural networks[1]

Bruno Vinícius Castro Guimarães[2], Sérgio Luiz Rodrigues Donato[3], Ignacio Aspiazú[4], Alcinei Mistico Azevedo[5]

ABSTRACT

Prediction models may contribute to data analysis and decision-making in the management of a crop. This study aimed to evaluate the feasibility of predicting the yield of 'Prata-Anã' and 'BRS Platina' banana plants by means of artificial neural networks, as well as to determine the most important morphological descriptors for this purpose. The following characteristics were measured: plant height; perimeter of the pseudostem at the ground level, at 30 cm and 100 cm; number of live leaves at harvest; stalk mass, length and diameter; number of hands and fruits; bunches and hands masses; hands average mass; and ratio between the stalk and bunch masses. The data were submitted to artificial neural networks analysis using the R software. The best adjustments were obtained with two and three neurons at the intermediate layer, respectively for 'Prata-Anã' and 'BRS Platina'. These models presented the lowest mean square errors, which correspond to the higher proximity between the predicted and the real data, and, therefore, a higher efficiency of the networks in the yield prediction. By the coefficient of determination, the best adjustments were found for 'Prata-Anã' (R^2 = 0.99 for all the network compositions), while, for 'BRS Platina', the data adjustment enabled an R^2 with values between 0.97 and 1.00, approximately. Yield predictions for 'Prata-Anã' and 'BRS Platina' were obtained with high efficiency by using artificial neural networks.

KEYWORDS: *Musa* spp., mathematical models, rural planning.

RESUMO

Predição da produtividade de bananeiras 'Prata-Anã' e 'BRS Platina' por redes neurais artificiais

Modelos de predição podem contribuir para a análise de dados e tomada de decisões no manejo de uma cultura. Objetivou-se avaliar a viabilidade da predição de produtividade de bananeiras 'Prata-Anã' e 'BRS Platina', por meio de redes neurais artificiais, bem como determinar os descritores morfológicos mais importantes para este fim. Foram mensurados a altura de planta; perímetro do pseudocaule ao nível do solo, a 30 e 100 cm de altura; número de folhas vivas na colheita; massa, comprimento e diâmetro do engaço; número de pencas e de frutos; massa do cacho e das pencas; massa média das pencas; e relação entre a massa do engaço e do cacho. Os dados foram submetidos a análise por redes neurais artificiais, utilizando-se o software R. Os melhores ajustes foram obtidos com dois e três neurônios na camada intermediária, respectivamente, para 'Prata-Anã' e 'BRS Platina'. Esses modelos apresentaram os menores erros quadráticos médios, o que corresponde a maior proximidade entre os dados preditos e os reais, e, por conseguinte, maior eficiência das redes na predição da produtividade. Pelo coeficiente de determinação, verificaram-se os melhores ajustes para 'Prata-Anã' (R^2 = 0,99 para todas as composições de rede), enquanto, para 'BRS Platina', a adequação dos dados possibilitou R^2 com valores entre 0,97 e 1,00, aproximadamente. Previsões de produtividade para 'Prata-Anã' e 'BRS Platina' foram obtidas com alta eficiência por meio de redes neurais artificiais.

PALAVRAS-CHAVE: *Musa* spp., modelos matemáticos, planejamento rural.

INTRODUCTION

The future challenges of banana farming as a species of high importance in global food security will be shaped by its ability to maintain, or even increase, its productive potential and, at the same time, maximize its predictability for the most diverse consumer markets (Soares et al. 2014, Silva et al. 2019). Among the studies on models to describe phenomena and elucidate unknowns, most of them

[1] Received: Oct. 05, 2020. Accepted: Mar. 04, 2021. Published: Apr. 12, 2021. DOI: 10.1590/1983-40632021v5166008.
[2] Instituto Federal do Amazonas, São Gabriel da Cachoeira, AM, Brasil.
E-mail/ORCID: bvinicius20@yahoo.com.br/0000-0003-2585-8794.
[3] Instituto Federal Baiano, Guanambi, BA, Brasil. *E-mail/ORCID*: sergio.donato@ifbaiano.edu.br/0000-0002-7719-4662.
[4] Universidade Estadual de Montes Claros, Janaúba, MG, Brasil.
E-mail/ORCID: ignacio.aspiazu@unimontes.br/0000-0002-0042-3324.
[5] Universidade Federal de Minas Gerais, Montes Claros, MG, Brasil.
E-mail/ORCID: alcineimistico@hotmail.com/0000-0001-5196-0851.

attempt to correlate morphological and physiological data with the productive potential of genotypes at the end of the crop cycle (Guimarães et al. 2019).

The selection of robust tools has been a priority and, therefore, several studies have proposed alternatives to describe plant growth and behavior, especially with complex characteristics that are difficult to measure in the field (Azevedo et al. 2015, Guimarães et al. 2018, Gemici et al. 2019).

Even with the implementation of these methods, few studies involve the development of models for predicting yield based on phenotypic data obtained in early stages of banana development (Ogunsua et al. 2019). Prediction methods generally allow an early assessment of crop yield to improve agricultural planning and management, as well as the appropriate allocation of resources. Thus, it must be considered that the establishment of accurate prediction models ensures the early analysis of genotypes of greatest interest, enabling resource optimization (Soares et al. 2014).

The development of prediction models based on morphological characters using artificial neural networks (ANN) was able to predict the yield of cactus pear with high accuracy (Guimarães et al. 2018). Azevedo et al. (2015) also reported a significant mathematical adjustment when evaluating the potential of using estimators for indirect selection against flowering through six predictive characteristics in lettuce.

In corn, Soares et al. (2015) evaluated the performance of ANN in the prediction of yield based on morphological variables and evidenced a high predictive capacity due to the strong correlation between the estimated values and the real grain yield data obtained in field experiments. Similarly, Azevedo et al. (2017) estimated kale leaf area with high efficiency for genotype selection in breeding programs. Soares et al. (2014) developed a model to predict yield in 'Maçã' banana fruits using the 'BRS Tropical' (AAAB) hybrid by ANN and regression equations, and Guimarães et al. (2013) for 'Prata' bananas (AAB and AAAB) using regression equations.

Prediction models can contribute to data analysis and decision making management, in addition to optimizing time in assessments and resources allocated to research programs. However, there is a lack of models for predicting the yield of 'Prata' bananas, a group that includes the most planted cultivars in Brazil, by means of morphological characters and ANN. Therefore, this study aimed to evaluate the feasibility of predicting the yield of 'Prata-Anã' and 'BRS Platina' bananas using ANN, as well as to determine the most important morphological descriptors for this purpose.

MATERIAL AND METHODS

The experiment was conducted at the Instituto Federal Baiano, in Guanambi, Bahia state, Brazil (14º13'30"S, 42º46'53"W and altitude of 545 m, with average precipitation and annual temperature of 660 mm and 26 ºC, respectively). The data were collected from 2008 to 2011 and the statistical analyzes carried out in 2020.

The soil of the experimental area is originally classified as Latossolo Vermelho-Amarelo, medium texture (Santos et al. 2018), which corresponds to an Oxisol (USA 2014).

The studies were developed with cultivars of the 'Prata' type: 'Prata-Anã', triploid (AAB); and the hybrid 'BRS Platina', tetraploid (AAAB). For evaluating vegetative and yield characteristics, 98 'Prata-Anã' and 96 'BRS Platina' plants were randomly sampled for both cultivars, considered as replicates. These plants were originated from a field planted with micropropagated seedlings (spacing of 3.0 x 2.5 m), with conventional fixed sprinkler irrigation using under-canopy sprinklers. The planting and management recommendations followed Rodrigues et al. (2015).

The following characteristics were evaluated at the harvest time: plant height (cm), linear dimension between the soil surface and the leaf rosette, using an analog measuring tape; perimeter of the pseudostem at the ground level, at 30 cm and 100 cm, using a measuring tape; number of live leaves (unit), counting directly in the field, considering the leaves with more than 50 % of the leaf blade in green coloring as alive.

As yield descriptors, were considered the stalk mass (kg), length (cm) and diameter (mm), obtained, respectively, with the aid of a digital scale, analog measuring tape and mechanical caliper. The numbers of hands and fruits were counted at the harvest time, followed by determining the bunch (kg) and hands (kg) masses, with a digital scale. The average hands (kg) mass was evaluated by dividing the hands total mass by the number of hands, and the relationship between the masses of the stalk and the bunch by dividing the values of these characteristics.

For data normalization and network implementation, the data were subjected to analysis by artificial neural networks (ANN), using the R v. 3.5.1 software (R Core Team 2018) and its corresponding packages. For the ANN training phase, both for the input (plant height; perimeter of the pseudostem at the ground level, at 30 cm and 100 cm; number of leaves at harvest; stalk mass, length and diameter; number of hands and fruits; average hands mass; and relationship between the stalk and bunch masses) and output data (hands and bunch masses), normalizations were performed in the interval between zero and one, using the *normalizeData* function of the Stuttgart Neural Network Simulator (RSNNS) package (Bergmeir & Benítez 2012).

After normalization, and in order to certify the quality of the model, the data were randomly segmented into two samples: training and validation, using 80 % of the data for the first, which corresponded to approximately 78 and 77 basic units of ‘Prata-Anã’ and ‘BRS Platina’, respectively; while, for the second sample, 20 % of the data were employed for network validation, which represents approximately 20 and 19 basic units for the ‘Prata-Anã’ and ‘BRS Platina’ cultivars, in that order. Neural networks with multi-layers (Multi-Layer-Perceptron - MLP) were used, with the mlp function of the RSNNS package trained with back propagation algorithm and a learning rate of 0.1 (Bergmeir & Benítez 2012).

The training periods were arbitrated in 500 cycles with the logistic and linear sigmoid functions to activate the intermediate and output layers, concomitantly. In order to select the most appropriate network architecture, ten network formats were tested, with 1, 2, 3, ..., 9 and 10 neurons in the middle layer. Each model was tested 100 times, so that the randomly assigned free weights would enable the selection of the best result in response to the lowest mean of the mean square error in the validation sample (Soares et al. 2014).

Then, with the best network architecture selected in the first stage, 1,000 new trainings were carried out. This procedure is performed to maximize the efficiency of the process and reduce the computational effort, avoiding countless training sessions for each network composition. Additionally, the relative importance of the input data was obtained through the garson function (NeuralNetTools package) by the Garson (1991) method, in the R software, according to a study proposed by Guimarães et al. (2018).

Finally, the predictive capacity of the networks was tested by the regression analysis of the predicted yield with that observed in the validation data. With the regression model defined, the point of intersection at the origin of the Cartesian plane was fixed and the significance of the slope of the line was considered by the t test. We observed as null and alternative hypothesis the probabilities that the slope of the model is equal to or different from 1, respectively. In this situation, if the coefficient of determination is high and the slope of the line does not differ from 1, so the efficiency in the prediction process can be assumed.

RESULTS AND DISCUSSION

For the ‘Prata-Anã’ cultivar, it was observed that the yield characteristics associated directly with the bunch, such as the stalk mass and length, followed by the hands and bunch masses, showed the highest variation coefficients, with values of 27.93; 23.16; 22.88; and 22.95 %, respectively (Table 1). These same characteristics showed the highest variations for ‘BRS Platina’, as it follows: 21.72; 18.80; 22.46; and 23.35 %. The stalk diameter, number of hands and fruits were the characteristics with the least variation, with values of 11.13; 10.97; and 14.30 %, respectively, for ‘Prata-Anã’, and 9.76; 11.64; and 16.60 % for ‘BRS Platina’, in that order (Table 1).

Still based on the yield descriptors, intermediate variation coefficient values were found for the stalk and bunch mass ratio and for the average mass of hands, with variation coefficient values of 16.31 and 18.02 % for ‘Prata-Anã’ and 15.70 and 19.69 % for ‘BRS Platina’, respectively (Table 1). The vegetative variables perimeter of the pseudostem at the soil level, at 30 and 100 cm from the soil were similar between the cultivars, in terms of variability, with variation coefficient of 11.59-12.63 % for ‘Prata-Anã’ and 9.54-10.61 % for ‘BRS Platina’ (Table 1). On the other hand, the number of live leaves at harvest and the plant height showed, respectively, the highest and the lowest variations among the vegetative characteristics, with 16.21 and 10.35 % for ‘Prata-Anã’ and 9.81 % and 17.36 % for ‘BRS Platina’ (Table 1).

By the Pearson’s correlation (r), it was found that the descriptors of yield bunch mass and hand

Table 1. Descriptive analysis of vegetative and productive characters.

	PH	PPSL	PP30	PP100	NLH	SM	SL	SD	NH	NFR	RSBM	ABM	BM	HM
‘Prata Anã’														
MIN	250.00	74.00	65.00	48.00	9.00	1.13	20.00	38.00	9.00	123.00	7.48	1.34	13.38	12.06
MED	335.24	112.77	98.80	78.01	13.89	2.84	38.61	73.28	10.94	182.88	10.64	2.17	26.62	23.79
MAX	418.00	138.00	122.00	96.00	19.00	5.20	74.00	86.00	13.00	238.00	15.72	3.32	40.10	36.49
SD	34.69	14.25	11.45	9.14	2.25	0.79	8.94	8.15	1.20	26.15	1.74	0.39	6.09	5.46
CV (%)	10.35	12.63	11.59	11.72	16.21	27.93	23.16	11.13	10.97	14.30	16.31	18.02	22.88	22.95
‘BRS Platina’														
	PH	PPSL	PP30	PP100	NLH	SM	SL	SD	NH	NFR	RSBM	ABM	BM	HM
MIN	250.00	74.00	51.00	52.00	6.00	1.58	28.00	42.00	7.00	77.00	5.69	1.33	10.88	9.28
MEAN	348.14	107.78	92.93	72.33	10.84	2.56	43.08	68.51	8.34	128.61	10.74	2.59	24.14	21.59
MAX	462.00	131.00	114.00	91.00	16.00	3.77	73.00	81.00	12.00	191.00	16.63	4.78	40.58	38.27
SD	34.16	10.29	9.39	7.67	1.88	0.56	8.10	6.69	0.97	21.35	1.69	0.51	5.42	5.04
CV (%)	9.81	9.54	10.11	10.61	17.36	21.72	18.80	9.76	11.64	16.60	15.70	19.69	22.46	23.35

PH: plant height; PPSL: perimeter of the pseudostem at the soil level; PP30: perimeter of the pseudostem at 30 cm; PP100: perimeter of the pseudostem at 100 cm; NLH: number of leaves at harvest; SM: stalk mass; SL: stalk length; SD: stalk diameter; NH: number of hands; NFR: number of fruits; RSBM: ratio between the stalk and bunch masses; ABM: average bunches mass; BM: bunch mass; HM: hands mass; MIN: minimum values; MEAN: mean values; MAX: maximum values; SD: standard deviation; CV: coefficient of variation.

mass were positively associated ($p \leq 0.05$) with the characteristics average hand mass and stalk mass for the evaluated cultivars. However, as the evaluating of these characteristics destroys the plant, the measurement is limited and of low practical application (Guimarães et al. 2013).

On the other hand, the morpho-agronomic characters perimeter of the pseudostem (soil, 30 and 100 cm) and number of hands and fruits, besides being easily obtained in the field by direct measurement and counting, showed significant correlation coefficients with parameters related to yield, with r values equal to 0.76; 0.70; 0.69; 0.70; and 0.74, respectively, for ‘Prata-Anã’; and 0.75; 0.64; 0.74; 0.60; and 0.74, in that order, for ‘BRS Platina’ (Table 2). Additionally, it is worth considering that the perimeter of the pseudostem can be measured at the time of flowering, as, from the change from the vegetative to the reproductive stage, the perimeter of the pseudostem no longer undergoes variation and/or an increase in its dimension, what allows the prediction of yield up to four months in advance using this feature (Guimarães et al. 2013).

These results differ from the values found by Soares et al. (2012), who worked with the Tropical cultivar and observed a correlation between yield and perimeter of the pseudostem and number of fruits per hand, with values of r equal to 0.025 and 0.471, respectively. However, the associations between agronomic descriptors (r) do not allow inferring about cause and effect relationships involving a set of characteristics, what makes it necessary to use more robust tools for the unfolding of prediction inferences. In this context, ANN are highly efficient in predicting various phenomena (Azevedo et al. 2015, Aquino et al. 2016a, Aquino et al. 2016b, Azevedo et al. 2017, Guimarães et al. 2018).

The prediction capacity of the model can be linked to the maximum and exhaustive number of trainings in the network, in favor of the most robust, adequate and accurate architecture (Soares et al. 2014). For this purpose, ten ANN models were evaluated, with one to ten neurons in the intermediate layer, to predict the yield for the two banana cultivars (Figure 1). The performance of the network architectures was measured by the mean quadratic error (Figures 1A and 1C) and by the coefficient of determination (R^2) (Figures 1B and 1D). The combinations with two and three neurons in the intermediate layer provided the best network structures for the ‘Prata-Anã’ (Figures 1A and 1B) and ‘BRS Platina’ (Figures 1C and 1D) cultivars, respectively.

The lowest mean square error values (Figures 1A and 1C) represent the highest proximity between the predicted and the actual data, and, therefore, the highest efficiency of the networks in predicting yield. For the R^2, adequate results were verified for the two cultivars; however, ‘Prata-Anã’ showed the best adjustments - with R^2 of approximately 0.99 for all the network compositions (Figure 1B). For ‘BRS Platina’, the data adequacy enabled a high-magnitude

Table 2. Pearson’s correlation estimates between vegetative and productive traits for ‘Prata-Anã’ and ‘BRS Platina’ bananas.

Pearson’s correlations														
‘Prata-Anã’														
	PH	PPSL	PP30	PP100	NLH	SM	SL	SD	NH	NFR	RSBM	ABM	BM	HM
PH	1.00	0.72**	0.68**	0.69**	0.26*	0.57**	0.33**	0.39**	0.58**	0.58**	0.12^{ns}	0.48**	0.64**	0.63**
PPSL	0.72**	1.00	0.94**	0.89**	0.41**	0.68**	0.36**	0.58**	0.76**	0.79**	0.11^{ns}	0.51**	0.76**	0.75**
PP30	0.68**	0.94**	1.00	0.91**	0.39**	0.64**	0.30**	0.57**	0.73**	0.76**	0.11^{ns}	0.46**	0.70**	0.69**
PP100	0.69**	0.89**	0.91**	1.00	0.32**	0.65**	0.32**	0.56**	0.67**	0.72**	0.16^{ns}	0.47**	0.69**	0.67**
NLH	0.26*	0.41**	0.39**	0.32**	1.00	0.29**	0.09^{ns}	0.29**	0.35**	0.37**	0.01^{ns}	0.24*	0.35**	0.35**
SM	0.57**	0.68**	0.64**	0.65**	0.29**	1.00	0.50**	0.78**	0.71**	0.74**	0.57**	0.57**	0.82**	0.77**
SL	0.33**	0.36**	0.30**	0.32^{ns}	0.09^{ns}	0.50**	1.00	0.15^{ns}	0.24*	0.26*	0.43**	0.28**	0.35**	0.32**
SD	0.39**	0.58**	0.57**	0.56**	0.29**	0.78**	0.15^{ns}	1.00	0.63**	0.64**	0.38**	0.48**	0.69**	0.66**
NH	0.58**	0.76**	0.73**	0.67**	0.35**	0.71**	0.24*	0.63**	1.00	0.95**	0.22*	0.28**	0.70**	0.68**
NFR	0.58**	0.79**	0.76**	0.72**	0.37**	0.74**	0.26**	0.64**	0.95**	1.00	0.22*	0.36**	0.74**	0.72**
RSBM	0.12^{ns}	0.11^{ns}	0.11^{ns}	0.16^{ns}	0.01^{ns}	0.57**	0.43**	0.38**	0.22*	0.22*	1.00	-0.22*	0.02^{ns}	-0.06^{ns}
ABM	0.48**	0.51**	0.46**	0.47**	0.24	0.57**	0.28**	0.48**	0.28**	0.36**	-0.22*	1.00	0.87**	0.88**
BM	0.64**	0.76**	0.70**	0.69**	0.35**	0.82**	0.35**	0.69**	0.70**	0.74**	0.02^{ns}	0.87**	1.00	1.00**
HM	0.63**	0.75**	0.69**	0.67**	0.35**	0.77**	0.32**	0.66**	0.68**	0.72**	-0.06^{ns}	0.88**	1.00**	1.00
‘BRS Platina’														
	PH	PPSL	PP30	PP100	NLH	SM	SL	SD	NH	NFR	RSBM	ABM	BM	HM
PH	1.00	0.59**	0.55**	0.65**	0.30**	0.45**	0.03^{ns}	0.47**	0.29**	0.41**	-0.12*	0.41**	0.49**	0.48**
PPSL	0.59**	1.00	0.78**	0.84**	0.56**	0.66**	0.07^{ns}	0.72**	0.57**	0.71**	-0.20**	0.58**	0.75**	0.73**
PP30	0.55**	0.78**	1.00	0.80**	0.48**	0.51**	-0.11*	0.74**	0.51**	0.66**	-0.22**	0.49**	0.64**	0.63**
PP100	0.65**	0.84**	0.80**	1.00	0.52**	0.66**	0.02^{ns}	0.68**	0.62**	0.74**	-0.18**	0.54**	0.74**	0.73**
NLH	0.30**	0.56**	0.48**	0.52**	1.00	0.35**	-0.17*	0.47**	0.51**	0.59**	-0.31**	0.39**	0.53**	0.53**
SM	0.45**	0.66**	0.51**	0.66**	0.35**	1.00	0.45**	0.66**	0.39**	0.53**	0.31**	0.57**	0.71**	0.65**
SL	0.03^{ns}	0.07^{ns}	-0.11*	0.02^{ns}	-0.17*	0.45**	1.00	-0.01^{ns}	-0.18*	-0.14*	0.51**	0.13*	0.08^{ns}	0.04^{ns}
SD	0.47**	0.72**	0.74**	0.68**	0.47**	0.66**	-0.01^{ns}	1.00	0.58**	0.68**	-0.11*	0.49**	0.68**	0.66**
NH	0.29*	0.57**	0.51**	0.62**	0.51**	0.39**	-0.18*	0.58	1.00	0.90**	-0.32**	0.21**	0.60**	0.60**
NFR	0.41**	0.71**	0.66**	0.74**	0.59**	0.53**	-0.14*	0.68	0.90**	1.00	-0.36**	0.41**	0.74**	0.73**
RSBM	-0.12*	-0.20*	-0.22*	-0.18*	-0.31**	0.31**	0.51**	-0.11	-0.33**	-0.36**	1.00	-0.44**	-0.42**	-0.49**
ABM	0.41**	0.58**	0.49**	0.54**	0.39**	0.57**	0.13*	0.49**	0.21*	0.41**	-0.44**	1.00	0.90**	0.90**
BM	0.49**	0.75**	0.64**	0.74**	0.53**	0.71**	0.08^{ns}	0.68**	0.60**	0.74**	-0.42**	0.90**	1.00	1.00**
HM	0.48**	0.73**	0.64**	0.73**	0.53**	0.65**	0.04^{ns}	0.66**	0.60**	0.73**	-0.49**	0.90**	1.00**	1.00

PH: plant height (cm); PPSL: perimeter of the pseudostem at the soil level; PP30: perimeter of the pseudostem at 30 cm; PP100: perimeter of the pseudostem at 100 cm; NLH: number of leaves at harvest (unit); SM: stalk mass (kg); SL: stalk length (cm); SD: stalk diameter; NH: number of hands; NFR: number of fruits; RSBM: ratio between the stalk and bunch masses; ABM: average bunch masses; BM: bunch mass; HM: hands mass. **, * and ^{ns}: significant at 1 %, significant at 5 % and not significant ($p > 0.05$), respectively, by the t test.

R^2; however, with values between 0.97 and 1.00, approximately (Figure 1D).

For ‘Cavendish’ bananas, Ogunsua et al. (2019) selected the best adjustments for the prediction of damage to fruits and hands masses with 18; 16; 8; and 12 neurons in the hidden layers, in which the lowest mean predictive error was recorded with 0.811; 0.581; 0.412; and 0.450 and R^2 of 0.76; 0.96; 0.86; and 0.88, respectively, with four variables in the input layer - amount of thermal energy transfer, maturity stage; minimum temperature and maximum temperature. In the study, the authors affirm that the practical application of these models will allow commercial banana producers to estimate the fruit yield, skin maturity, relationship between the fruit pulp and skin, as well as injuries from sunburn and damage caused by thrips.

For Soares et al. (2015), the best adaptation of neurons in the intermediate layer, pre-established during the training phase, should consider an arrangement with more than two neurons with the progressive addition in the adjustable layer, in order to select the best architecture, in response to the lowest mean square error in the validation sample. However, this process is not linear and does not always favor the performance of the model, as it varies according to the phenomenon or is unknown to be elucidated (Soares et al. 2014, Azevedo et al. 2015, Aquino et al. 2016a, Aquino et al. 2016b). Silva et al. (2010) add that increasing neurons in ANN ensures

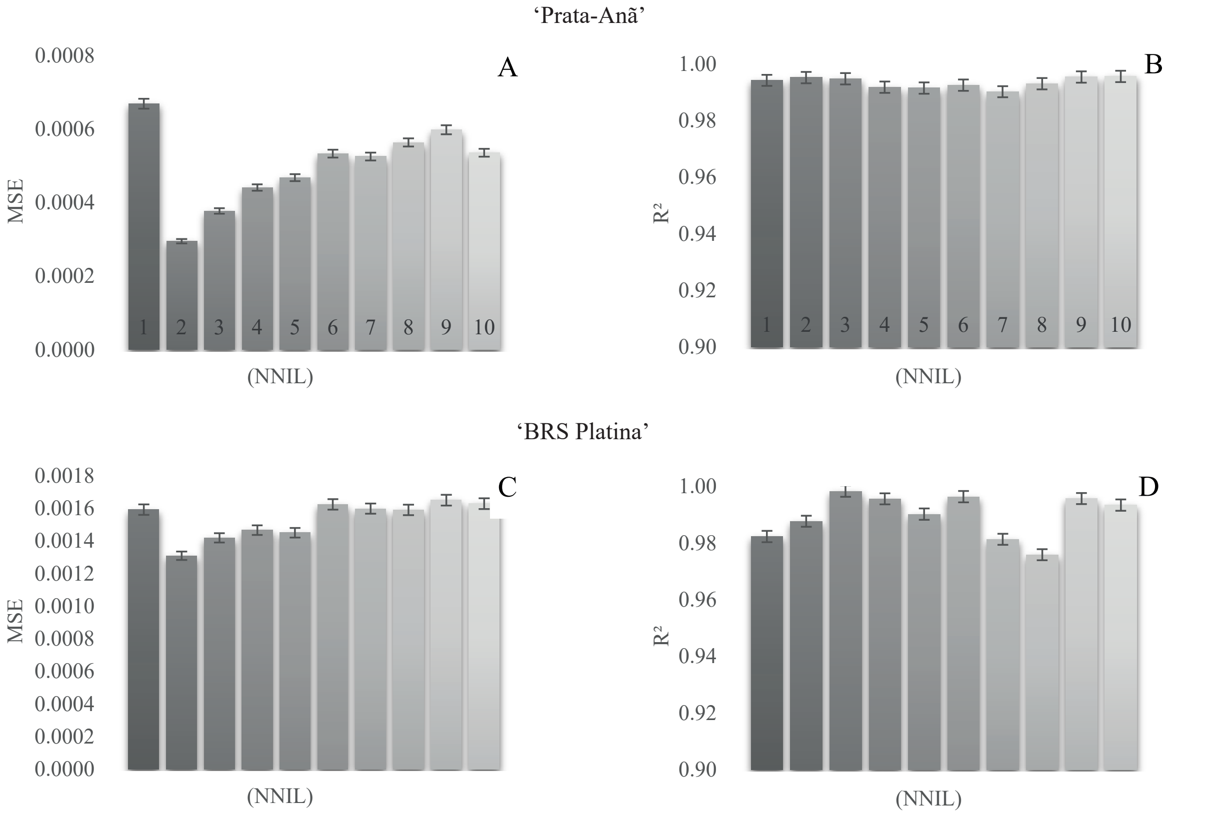

Figure 1. Estimates of mean square error (MSE) (A and C) and coefficient of determination (R^2) (B and D) obtained considering different numbers of neurons in the intermediate layer (NNIL) for the 'Prata-Anã' and 'BRS Platina' banana cultivars. Deviations refer to the lower and upper limits to the 95 % confidence level obtained by bootstrap, with 10,000 simulations.

the memorization of the studied data, but does not decode the interrelationships between the input and output layers - an operation called *overfitting*.

Observing the relative importance of the predictive characteristics, weighted by the Garson method (Garson 1991), the perimeter of the pseudostem at 100 cm in height is more relevant (Figures 2B and 2D), with a contribution of 25 and 16 % in the yield for 'Prata-Anã' and 'BRS Platina', respectively. The relative importance of this characteristic also showed a linear association with yield, as determined by the coefficient of correlation (Table 2), which, in this case, showed significant values. Among the vegetative characteristics, the plant height and number of live leaves at the harvest showed the smallest relative contributions to estimate the yield of the two cultivars (Figures 2B and 1D). On the other hand, the characteristics stalk length, diameter and mass expressed the lowest values of relative importance (Figure 2).

Figure 2 shows the parameters of the model with the best performance to predict yield in the 'Prata-Anã' (Figures 1A and 1B) and 'BRS Platina' (Figures 1C and 1D) cultivars, involving the main production components. To test the efficiency of the prediction model, the estimated yield was compared with the yield observed through the input variables destined to the validation sample. Thus, the ANN model was successful due to the similarity between observed and predicted data, as tested for 'Prata-Anã' and 'BRS Platina', with R^2 values greater than 0.99 for both cultivars (Figure 2A).

The analysis of the correlation and regression coefficients and the relative information of the evaluated characteristics made it possible to construct the models of yield prediction for 'Prata-Anã' and 'BRS Platina' (Figure 2). The selection of models and the proof of their efficiency for generalization of ANN occurred in response to the R^2 and the non-significance ($p > 0.05$) of the t test for the null hypothesis of the slope of the line (Ho: b = 1).

Neural networks with R^2 of 0.70 and 0.65 were adjusted for 'Prata-Anã' (Figures 3A, 3B and 3C) and 'BRS Platina' (Figures 3D, 3E and 3F), respectively.

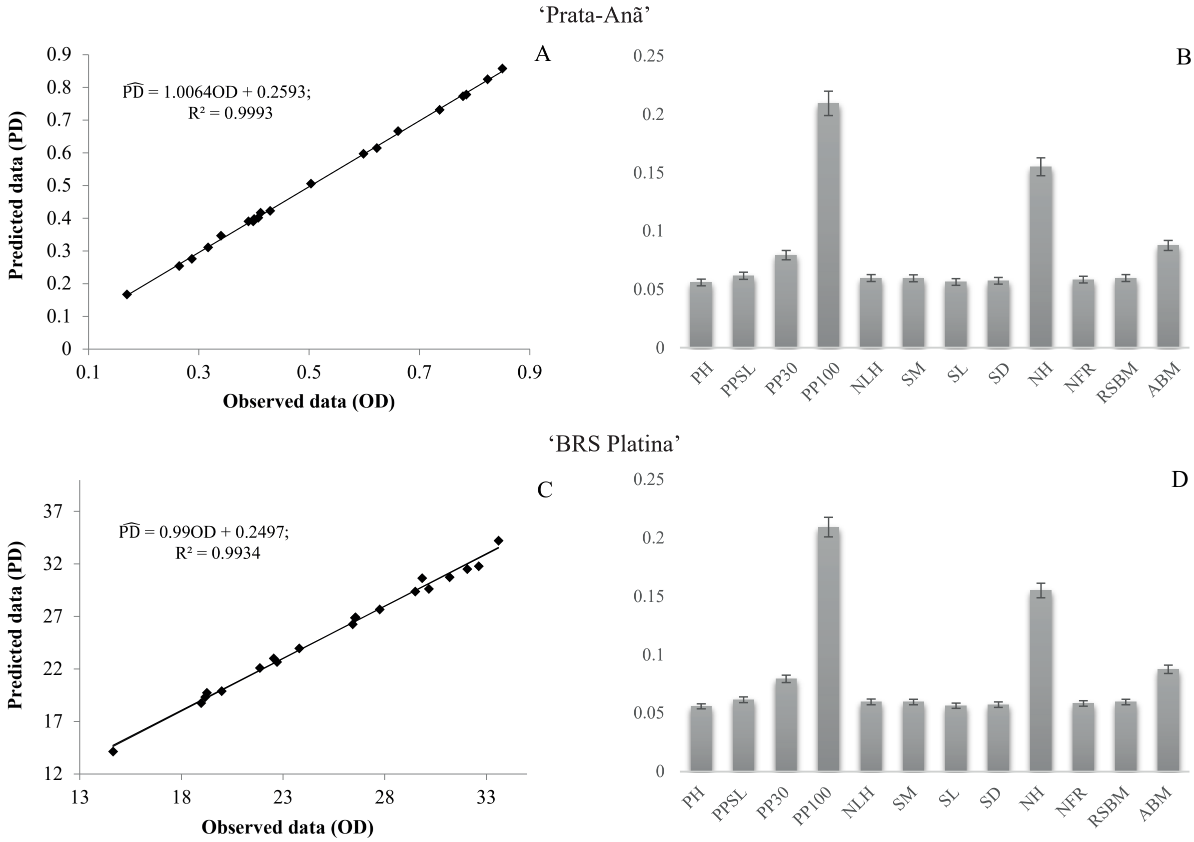

Figure 2. Graphical representation of the yield prediction quality of ‘Prata-Anã’ and ‘BRS Platina’ bananas by artificial neural networks for the validation sample considering all the evaluated characters (A) and relative information (B) obtained by the Garson (1991) method for the agronomic parameters. PH: plant height; PPSL, PP30 and PP100: perimeter of the pseudostem at the soil level, at 30 cm and 100 cm, respectively; NLH: number of live leaves at harvest; SM, SL and SD: stalk mass, length and diameter, respectively; NH and NFR: number of hands and fruits, respectively; RSBM: ratio between the stalk and bunch masses; ABM: average bunch mass (kg plant^{-1}). Deviations refer to 95 % confidence intervals, obtained by the BCa bootstrap, with 10,000 simulations.

In these models, were considered characteristics that are easy to measure in the field, and that can be measured at the time the inflorescence is emerging, about 120-150 days before the harvest for cultivars of the ‘Prata’ type (Donato et al. 2016), what facilitates scheduling the harvest and forecasting delivery by the producer (Guimarães et al. 2013), such as the perimeter of the pseudostem at 30 and 100 cm and the number of hands for both cultivars.

In order to obtain an estimate for the ‘Prata-Anã’ bananas yield with higher predictive capacity, higher accuracy and in line with the practicality of obtaining them in the field, were tested, with data from the validation sample, the morphological components associated with the characteristics with higher relative information, higher correlations and based only on the easily measured characteristics for ‘Prata-Anã’ and ‘BRS Platina’.

However, the best adjustments were associated with characteristics with the highest correlation coefficients and the largest relative information, with R^2 values equal to 0.99 for both cultivars (Figure 2). Thus, for the first model, 12 characteristics were used (Figure 2); while, for the second one, a reduced model was considered with only three predictive characteristics that are easy to measure in the field, with R^2 values equal to 0.70 and 0.65, respectively, for ‘Prata-Anã’ (Figures 3A, 3B and 3C) and ‘BRS Platina’ (Figures 3D, 3E and 3F).

Considering the models with variables that are easy to measure in the field, an expressive adjustment was observed to predict yield, with R^2 of 0.85 and 0.79 for ‘Prata-Anã’ (Figures 4D, 4E and 4F) and ‘BRS Platina’ (Figures 4D, 4E and 4F), respectively, using the characteristics plant height, perimeter of the pseudostem at the soil level, stalk diameter

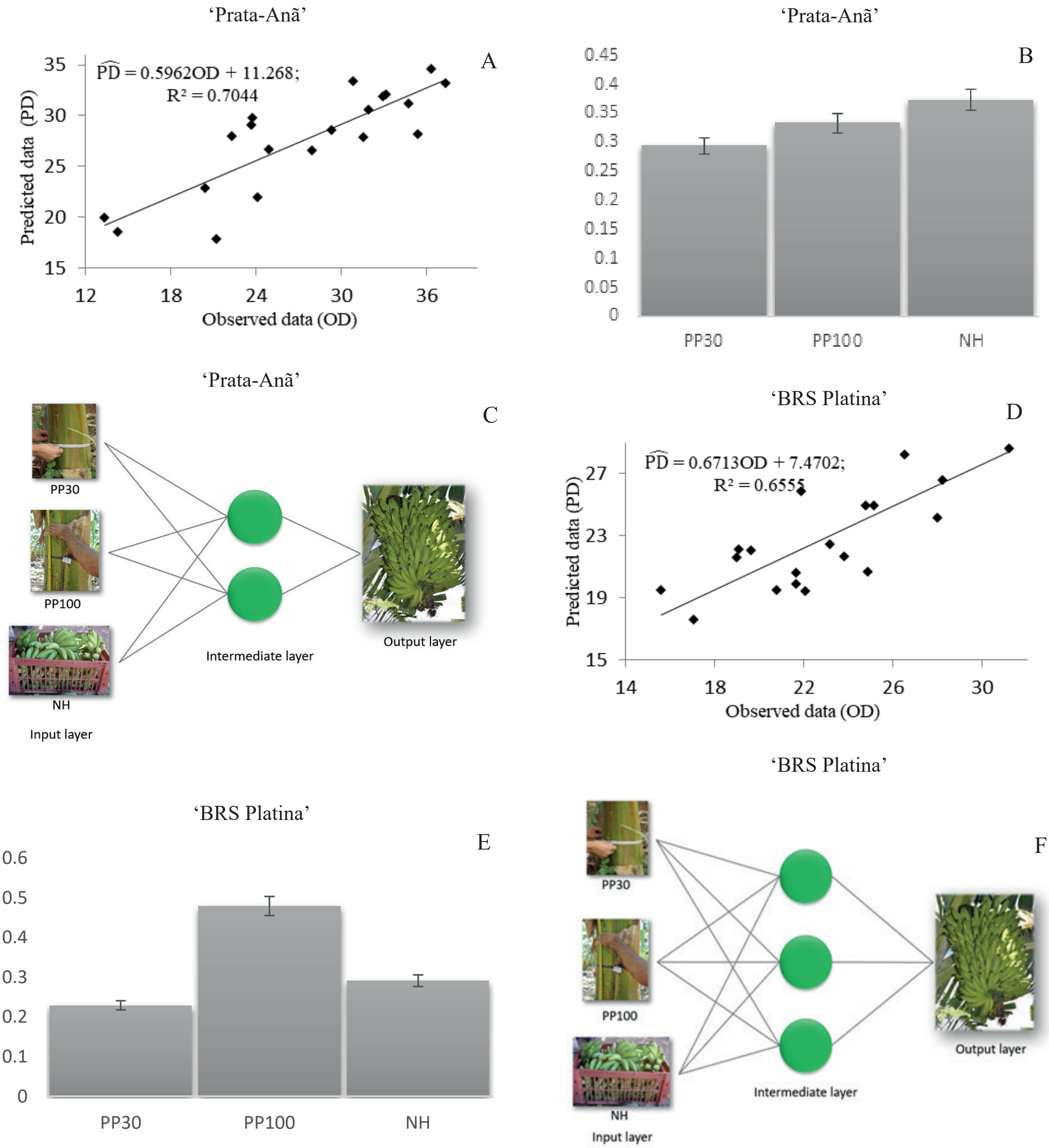

Figure 3. Prediction of 'Prata-Anã' and 'BRS Platina' bananas yield (kg plant^{-1}) by means of agronomic characteristics (PP30 and PP100: pseudostem perimeter at 30 and 100 cm from the ground, respectively; NH: number of hands) (A), relative information (B) obtained by the Garson (1991) method and topology of the network (C) with two and three neurons in the intermediate layer, for the respective genotypes, by artificial neural networks. Deviations refer to 95 % confidence intervals, obtained by the BCa bootstrap, with 10,000 simulations.

and number of fruits. The model estimates provide significant results through the relationship between observed and predicted data for the 'Prata-Anã' (Figure 4A) and 'BRS Platina' (Figure 4D) cultivars.

Similarly to the present study, the high performance of ANN in obtaining agronomic estimates has been confirmed in several studies (Binoti et al. 2013, Soares et al. 2014, Soares et al. 2015, Aquino et al. 2016a, Aquino et al. 2016b, Miguel et al. 2016, Gemici et al. 2019, Vitor et al. 2019). Among the reasons associated with the efficiency of the ANN models, there is the average predictive error of less

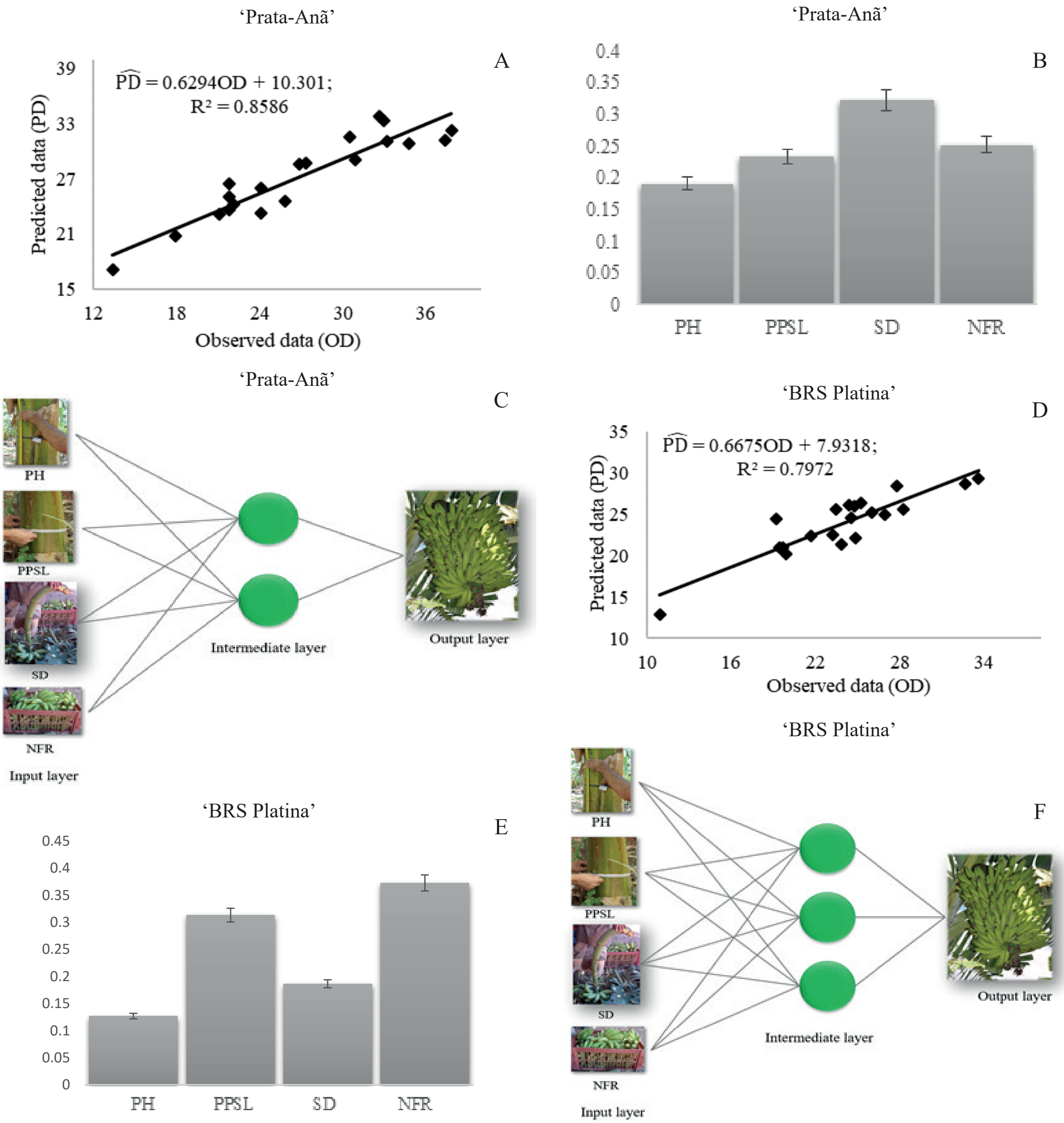

Figure 4. Models for predicting the ‘Prata-Anã’ bananas yield (kg plant^{-1}) by artificial neural networks, considering the characteristics with the highest relative information (A), the highest correlations (B) and the models based only on the variables of easy measurement in the field (C), for the ‘Prata-Anã’ and ‘BRS Platina’ cultivars. PH: plant height; PPSL: perimeter of the pseudostem at the soil level; SD: stalk diameter; NFR: number of fruits.

than 10 %, which ensures the predictability of the phenomena (Vendruscolo et al. 2017).

The predictive capacity of mathematical models based on morphological characteristics by ANN have been attested in several crops, such as Tropical banana (Soares et al. 2014), corn (Soares et al. 2015) and cactus pear (Guimarães et al. 2018), with agreement indexes (relationship between the estimated and observed values) similar to the adjusted models for ‘Prata-Anã’ and ‘BRS Platina’ (Figure 1), equal to 0.91; 1.0; and 0.87, respectively.

The development of models to predict yield based on phenotypic data obtained in early stages allows the early assessment of yield descriptors in crops, what favors decision-making and agricultural management, as well as the appropriate allocation of resources. Thus, it is worth mentioning that the prediction estimators developed in this study allow to ensure, with high accuracy (0.99), the strategic planning of the commercialization and export activities of 'Prata-Anã' and 'BRS Platina' up to 120 days before the harvest.

CONCLUSIONS

1. The yield prediction of 'Prata-Anã' and 'BRS Platina' banana plants are obtained with high efficiency by means of multilayer perceptron artificial neural networks;
2. The morphological characteristics easy to measure in the field with the highest relative information were associated with the characteristics plant height, perimeter of the pseudostem at the soil level, stalk diameter and number of fruits per bunch for 'Prata-Anã' and 'BRS Platina';
3. Characteristics with the highest relative information and highest correlation coefficients enabled the best adjustments for predicting the yield of 'Prata-Anã'; therefore, they must be adopted to compose the predictive models, in order to make the rural planning more efficient.

REFERENCES

AQUINO, C. F.; SALOMÃO, L. C. C.; AZEVEDO, A. M. Fenotipagem de alta eficiência para vitamina A em banana utilizando redes neurais artificiais e dados colorimétricos. *Bragantia*, v. 75, n. 3, p. 268-274, 2016a.

AQUINO, C. F.; SALOMÃO, L. C. C.; AZEVEDO, A. M. Qualidade pós-colheita de banana 'Maçã' tratada com ácido giberélico avaliada por redes neurais artificiais. *Pesquisa Agropecuária Brasileira*, v. 51, n. 7, p. 824-833, 2016b.

AZEVEDO, A. M.; ANDRADE JÚNIOR, V. C.; PEDROSA, C. E.; OLIVEIRA, C. M.; DORNAS, M. F. S.; CRUZ, C. D.; VALADARES, N .R. Application of artificial neural networks in indirect selection: a case study on the breeding of lettuce. *Bragantia*, v. 74, n. 1, p. 1-7, 2015.

AZEVEDO, A. M.; ANDRADE JÚNIOR, V. C.; SOUSA JÚNIOR, A. S.; SANTOS, A. A.; CRUZ, C. D.; PEREIRA, S. L.; OLIVEIRA, A. J. M. Eficiência da estimação da área foliar de couve por meio de redes neurais artificiais. *Horticultura Brasileira*, v. 35, n. 1, p. 14-19, 2017.

BERGMEIR, C.; BENÍTEZ, J. M. Neural networks in R using the Stuttgart Neural Network Simulator: RSNNS. *Journal of Statistical Software*, v. 46, n. 1, p. 1-26, 2012.

BINOTI, D. H. B.; BINOTI, M. L. M. da S.; LEITE, H. G.; SILVA, A. Redução de custos em inventário de povoamentos equiâneos. *Revista Brasileira de Ciências Agrárias*, v. 8, n. 1, p. 125-129, 2013.

DONATO, S. L. R.; COELHO, E. F.; MARQUES, P. R. R.; ARANTES, A. de M. Considerações ecológicas, fisiológicas e de manejo. *In*: FERREIRA, C. F.; SILVA, S. de O. e; AMORIM, E. P.; SANTOS-SEREJO, J. A. dos. *O agronegócio da banana*. Brasília, DF: Embrapa, 2016. p. 45-110.

GARSON, G. D. Interpreting neural network connection weights. *International Journal of Artificial Intelligence and Expert Systems*, v. 6, n. 1, p. 47-51, 1991.

GEMICI, E.; YUCEDAG, C.; OZEL, H. B.; IMREN, E. Predicting cone production in clonal seed orchard of anatolian black pine with artificial neural network. *Applied Ecology and Environmental Research*, v. 17, n. 2, p. 2267-2273, 2019.

GUIMARÃES, B. V. C.; DONATO, S. L. R.; MAIA, V. M.; ASPIAZÚ, I.; RODRIGUES, M. G. V.; MARQUES, P. R. R. M. Simple and multiple linear regressions for harvest prediction of Prata type bananas. *African Journal of Agricultural Research*, v. 8, n. 48, p. 6300-6308, 2013.

GUIMARÃES, B. V. C.; DONATO, S. L. R.; AZEVEDO, A. M.; ASPIAZÚ, I.; SILVA JUNIOR, A. A. e. Prediction of "Gigante" cactus pear yield by morphological characters and artificial neural networks. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 22, n. 5, p. 315-319, 2018.

GUIMARÃES, B. V. C.; DONATO, S. L. R.; ASPIAZÚ, I.; AZEVEDO, A. M.; CARVALHO, A. J. de. Size of plots for experiments with cactus pear cv. Gigante. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 23, n. 5, p. 347-351, 2019.

MIGUEL, E. P.; MOTA, F. C. M.; TÉO, S. J.; NASCIMENTO, R. G. M.; LEAL, F. A.; PEREIRA, R. S.; REZENDE, A. V. Artificial intelligence tools in predicting the volume of trees within a forest stand. *African Journal of Agricultural Research*, v. 11, n. 21, p. 1914-1923, 2016.

OGUNSUA, J.; SAENGRAYAP, R.; HABIB ULLAH, H.; CHAIWONG, S. Application of artificial neural networks in the prediction of fruit damages and hand weight in Cavendish banana. *Journal of Food Science and Agricultural Technology*, v. 5, n. 1, p. 140-145, 2019.

RODRIGUES, M. G. V.; DIAS, M. S. C.; DONATO, S. L. R. Implantação e manejo do bananal. *Informe Agropecuário*, v. 36, n. 288, p. 27-44, 2015.

SANTOS, H. G.; JACOMINE, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.;

ALMEIDA, J. A.; ARAÚJO FILHO, J. C.; OLIVEIRA, J. B.; CUNHA, T. J. F. *Sistema brasileiro de classificação de solos*. 5. ed. rev. ampl. Brasília, DF: Embrapa, 2018.

SOARES, F. C.; ROBAINA, A. D.; PEITER, M. X.; RUSSI, J. L. Predição da produtividade da cultura do milho utilizando rede neural artificial. *Ciência Rural*, v. 45, n. 11, p. 1987-1993, 2015.

SOARES, J. D. R.; PASQUAL, M.; RODRIGUES, F. A.; LACERDA, W. S.; DONATO, S. L. R.; SILVA, S. e O.; PAIXÃO, C. A. Correlation between morphological characters and estimated bunch weight of the Tropical banana cultivar. *African Journal of Biotechnology*, v. 11, n. 47, p. 10682-10687, 2012.

SOARES, J. D. R.; PASQUAL, M.; LACERDA, W. S.; SILVA, S. O.; DONATO, S. L. R. Comparison of techniques used in the prediction of yield in banana plants. *Scientia Horticulturae*, v. 167, n. 1, p. 84-90, 2014.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes neurais artificiais para engenharia e ciências aplicadas*: curso prático. São Paulo: Artliber, 2010.

SILVA, W. R.; VALE, L. S. R.; PEREIRA, D. R. M. Performance of banana cultivars under Ceres - GO edaphoclimatic conditions. *Revista de Ciências Agrárias*, v. 62, n. 1, p. 1-6, 2019.

R CORE TEAM. *R*: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2018.

USA. Soil Survey Staff. *Keys to soil taxonomy*. 12. ed. Washington, DC: United States Department of Agriculture, 2014.

VITOR, A. B.; DINIZ, R. P.; MORGANTE, C. V.; ANTONIO, R. P.; OLIVEIRA, E. J. de. Early prediction models for cassava root yield in different water regimes. *Field Crops Research*, v. 239, n. 1, p. 149-158, 2019.

VENDRUSCOLO, D. G. S.; CHAVES, A. G. S.; MEDEIROS, R. A.; SILVA, R. S. da; SOUZA, H. S.; DRESCHER, R.; LEITE, H. G. Estimativa da altura de árvores de *Tectona grandis* L. f. utilizando regressão e redes neurais artificiais. *Nativa*, v. 5, n. 1, p. 52-58, 2017.