Prediction of 'Gigante' cactus pear yield by morphological characters and artificial neural networks

Bruno V. C. Guimarães[1], Sérgio L. R. Donato[2], Alcinei M. Azevedo[3],
Ignacio Aspiazú[4] & Ancilon A. e Silva Junior[2]

[1] Instituto Federal do Amazonas/Departamento de Ciências Agrárias. São Gabriel da Cachoeira, AM. E-mail: bvinicius20@yahoo.com.br - ORCID: 0000-0003-2585-8794 (Corresponding author)
[2] Instituto Federal Baiano/Departamento de Ciências Agrárias. Guanambi, BA. E-mail: sergio.donato@ifbaiano.edu.br - ORCID: 0000-0002-7719-4662; ancilon.silva@ifbaiano.edu.br - ORCID: 0000-0002-6134-6430
[3] Universidade Federal de Minas Gerais/Instituto de Ciências Agrárias. Montes Claros, MG. E-mail: alcineimistico@hotmail.com - ORCID: 0000-0001-5196-0851
[4] Universidade Estadual de Montes Claros/Departamento de Ciências Agrárias. Janaúba, MG. E-mail: ignacio.aspiazu@unimontes.br - ORCID: 0000-0002-0042-3324

A B S T R A C T

Estimating cactus pear yield is important for the planning of small and medium rural producers, especially in environments with adverse climatic conditions, such as the Brazilian semi-arid region. The objective of this study was to evaluate the potential of artificial neural networks (ANN) for predicting yield of 'Gigante' cactus pear, and determine the most important morphological characters for this prediction. The experiment was conducted in the Instituto Federal Baiano, Guanambi campus, Bahia, Brazil, in 2009 to 2011. The area used is located at 14° 13' 30" S and 42° 46' 53" W, and its altitude is 525 m. Six vegetative agronomic characters were evaluated in 500 plants in the third production cycle. The data were subjected to ANN analysis using the R software. Ten network architectures were trained 100 times to select the one with the lowest mean square error for the validation data. The networks with five neurons in the middle layer presented the best results. Neural networks with coefficient of determination (R^2) of 0.87 were adjusted for sample validation, assuring the generalization potential of the model. The morphological characters with the highest relative contribution to yield estimate were total cladode area, plant height, cladode thickness and cladode length, but all characters were important for predicting the cactus pear yield. Therefore, predicting the production of cactus pear with high precision using ANN and morphological characters is possible.

Predição da produtividade de palma forrageira 'Gigante' por caracteres morfológicos e redes neurais artificiais

R E S U M O

A estimativa da produtividade em palma forrageira é fundamental ao planejamento rural dos pequenos e médios produtores, sobretudo, em condições de adversidades climáticas como no Semiárido Brasileiro. Objetivou-se avaliar o potencial de Redes Neurais Artificiais (RNAs) na predição da produtividade em palma forrageira, e determinar os caracteres morfológicos mais importantes neste estudo. O experimento foi conduzido no Instituto Federal Baiano, Campus Guanambi, Bahia, Brasil, no período agrícola de 2009 a 2011. A região localiza-se nas seguintes coordenadas geográficas: latitude 14° 13' 30" Sul, longitude de 42° 46' 53" Oeste de Greenwich, altitude de 525 m. Avaliaram-se em 500 plantas seis caracteres agronômicos de natureza vegetativa no terceiro ciclo de produção. Os dados foram submetidos à análise no software R por RNAs. Dez arquiteturas de rede foram treinadas por 100 vezes, selecionando-se, ao final do treinamento, aquela com menor erro quadrático médio para os dados de validação. As redes com cinco neurônios na camada intermediária possibilitaram a máxima qualidade preditiva. Foram ajustadas redes neurais com coeficiente de determinação (R^2) de 87,21% para a amostra de validação, assegurando o potencial de generalização do modelo. Os caracteres morfológicos de maior contribuição relativa foram a área total do cladódio, altura da planta, espessura do cladódio e comprimento do cladódio; porém, todos são importantes na predição da produtividade. Logo, é possível predizer a produção de palma forrageira com alta precisão por meio de RNAs e caracteres morfológicos.

Introduction

Estimating cactus pear (*Opuntia ficus indica*, Mill.) production is important for the planning of small and medium producers, especially in environments with adverse climatic conditions, such as the Brazilian semiarid region (BSA). This plant is an energetic source in the nutrition of ruminants (Aguiar et al., 2015a, 2015b). However, this crop needs technological tools to increase yield, since it can minimize risks of maintaining cattle herds in the dry season.

The development of prediction models for this forage palm is scarce and inexpensive in agronomic studies. Studies report components of the production in different species of forage *Opuntias*, but they lack information on estimation cladode production from the plant attributes, especially those measured in the pre-harvest phase of the first and second cladode (Padilha Junior et al., 2016).

The artificial neural networks (ANN) is a successful tool to describe, substantiate, and elucidate high-complexity issues in the field of modeling (Jana et al., 2012; Jana & Mohanty, 2012; Azevedo et al., 2015; Brasileiro et al., 2015; Soares et al., 2015; Aquino et al., 2016a, 2016b; Azevedo et al., 2017). Thus, the use of ANN in agronomic modeling for the cactus pear crop can be efficient for predicting yield.

ANN has better performance compared to other statistical modeling techniques, since it has universal fit of functions (Gianola et al., 2011), is non-parametric, admits data loss, and does not require much prior information about the phenomena to be modeled (Azevedo et al., 2015).

Thus, objective of this study was to evaluate the potential of artificial neural networks for predicting yield of 'Gigante' cactus pear, and determine the most important morphological characters for this prediction.

Material and Methods

The experiment was conducted in the Federal Institute for Education, Science, and Technology of Bahia, Guanambi campus, Brazil, from 2009 to 2011. The area used is located at 14° 13' 30" S and 42° 46' 53" W, and its altitude is 525 m. The soil of the area was classified as Entisol, and the region has annual average precipitation of 680 mm and annual average temperature of 26 °C (CODEVASF, 2017).

The evaluations were carried out in an uniformity test with the 'Gigante' cactus pear, at 930 days after planting (DAP), in the third production cycle. A blank test was conducted with spacing of 2.0 x 0.2 m, using a planting template, with 25,000 plants ha⁻¹. The soil were prepared with subsoiling, plowing and harrowing at 35, 25 and 20 cm depths, respectively. An organic fertilizer consisted of aged sheep manure was applied at a rate of 40 L m⁻¹. The planting was arranged in 10 rows with 50 plants each totaling 500 plants and an area of 200 m². This test does not present treatments and the crop practices were homogeneous in the whole area (Storck et al., 2011), thus constituting a blank or uniformity test.

The vegetative characters evaluated were: cladode length (CL; cm), cladode width (CW; cm), cladode thickness (CT;

mm), number of cladodes (NOC), plant height (PH; cm), cladode biomass (CB; Mg ha⁻¹ year⁻¹), cladode area (CA = CL x CW x 0.693, cm²), and total cladode area (TCA = ((CA x NOC)/10,000) x 2; m²) in the third production cycle.

The data were evaluated in the R software (R Development Core Team, 2012) using artificial neural networks (ANN). Both the input (CL, CW, CT, NOC, TCA and PH) and output data (PROD) were normalized into a 0 to 1 interval by the normalize data function of the package RSNNS to increase the efficiency in network training (Bergmeir & Benítez, 2012).

In the ANN analysis, 80% of the data (information of 400 plants) were intended for training the network, and 20% for the validation analysis (information of 100 plants). The samples from the training and validation plants were randomly chosen. The multilayer perceptron (MLP) neural model was used. The MLP function of the RSNNS package, with backpropagation algorithm and learning rate of 0.1, was applied to improve the MLP networks.

The maximum number of training was 500 and the activation functions for the intermediate and output layers fitted logistic and linear models, respectively. Ten architectures were tested, with 1, 2, 3, ..., 9 and 10 neurons in the middle layer to define the best network structure. The free parameters is generated randomly at the beginning of training and these initial values can influence the result of the training (Soares et al., 2014), thus, each ANN architecture was trained 100 times. The network that provided the best fit was selected considering the means of the mean square error (MSE) for the validation sample.

The best network architecture selected were subjected to 1,000 new trainings. The strategy of selecting the best network architecture and carry out several trainings for this configuration aims to reduce computational effort by avoiding performing training for each architecture. In addition, the relative importance of the input characteristics was obtained using Garson' method (Garson, 1991) by the Garson function (Neural Net Tools package).

The predictive capacity in the training of the networks was tested using a regression analysis on the predicted yield found with the selected network for the sample validation. The intercession point was fixed at the origin of the Cartesian plane for a practical interpretation of the regression model. The t test was used for the angular coefficient of the line to verify if it is equal to or different from 1. Thus, the efficiency in the prediction process is based on the value 1 for the line coefficient and the high coefficient of determination of the model.

Results and Discussion

A high variation was found for the yield per plant, with estimates of 0.60 to 34.70 kg plant⁻¹ of fresh biomass, and coefficient of variation of 52.24% (Table 1). Regarding the morphological data, the highest coefficients of variation were found for the number of cladodes (NOC) (37.08) and total cladode area (TCA) (41.24%). The cladode length and width had the greatest morphological uniformity, with coefficients of variation of 7.07 and 8.39%, respectively. However, cladode thickness (CT) presented high phenotypic variability (Table 1).

Table 1. Descriptive analysis of the cactus pear cladode length (CL), width (CW), thickness (CT), number (NOC), total area (TCA) and fresh biomass yield (YIELD), and plant height (PH)

Parameters	CL	CW	CT	NOC	PH	TCA	YIELD
	(cm)		(mm)	(un)	(m)	(m²)	(kg plant⁻¹)
Minimum	19.56	10.11	4.87	3.00	0.46	0.13	0.60
Average	29.74	14.54	13.56	20.88	1.18	1.29	12.27
Maximum	36.35	19.44	24.65	48.00	1.89	2.94	34.70
Std. Deviation	2.10	1.22	4.29	7.74	0.22	0.53	6.41
Coefficient of variation (%)	7.07	8.39	31.61	37.08	18.86	41.24	52.24
	Pearson's correlation						
CW (cm)	0.74 **	-	-	-	-	-	-
CT (mm)	0.07 ns	0.21 **	-	-	-	-	-
NOC (und)	0.19 **	0.16 **	0.08 ns	-	-	-	-
PH (m)	0.49 **	0.45 **	-0.13 **	0.56 **	-	-	-
TCA (m²)	0.44 **	0.43 **	0.12 **	0.94 **	0.64 **	-	-
YIELD (kg plant⁻¹)	0.39 **	0.39 **	0.26 **	0.81 **	0.56 **	0.86 **	-

ns Non-significant ($p \geq 0.05$); * Significant at a 5% level of probability ($0.01 \leq p < 0.05$); ** Significant at a 1% level of probability ($p < 0.01$)

According to Azevedo et al. (2015), the high variability of morphological characters is essential for the generalization in the ANN. Thus, trained ANN can also be used in other prediction models, similar to those in the present study.

Ten network architectures were evaluated for the ANN. A significant number of training sessions was performed, which showed the adjustment of the network with two neurons in the middle layer presenting the largest mean square errors (MSE) (Figure 1A) and the lowest coefficient of determination (R^2). Soares et al. (2014) point out the importance of intensive network training in the search for the most appropriate architecture.

The smallest mean square error (MSEs) was obtained when five neurons were tested in the intermediate layer (Figure 1A). The smaller the MSEs, the greater the proximity of the predicted MSE by ANN to the real values; thus, a low indicates a high efficiency of the networks. Regarding the coefficient of determination (R^2), satisfactory results were obtained for the

Figure 1. Estimates of mean square error (A) and coefficient of determination (B) considering different numbers of neurons in the intermediate layer

average of five neurons in the intermediate layer, with fitting of the data of 78.57% (Figure 1B). The deviations presented represent the lower and higher limits from the 95% confidence level obtained by bootstrap with 10,000 simulations.

The determination of the optimal number of neurons in the intermediate layer is important. According to Soares et al. (2015), the best relationship between the number of training samples and the number of intermediate connections must be considered in the selection of the prediction model; and the latter should be higher than 2 to reach the least average relative error of validation. Similar prediction studies confirm that the addition of neurons per layer does not always favor the performance of the model (Soares et al., 2014; Azevedo et al., 2015, Aquino et al., 2016a). According to Silva et al. (2010), the continuous addition of neurons in the network, in the training phase, allows memorization of the studied data, but does not identify the probable associations between the data inserted in the input and output layers - a technical condition called overfitting. The network structure with five neurons in the middle layer is shown in Figure 2A.

According to the relative importance of the responses obtained by Garson's method (Garson, 1991), the total cladode area (TCA) was the most important (Figure 2B), with a relative contribution to yield of 25.07%. The highest expression of the TCA correlates with the highest coefficient of correlation with yield, 0.86 (Table 1). Cladode width (CW) had the lowest relative contribution to yield (12.33%).

Neural networks with R^2 of 0.87 were fitted for the validation sample (Figure 3A). The high value of R^2, and the non-significance of the angular coefficient of the line (Ho: b = 1), proves the prediction efficiency and generalization of the model. Gianola et al. (2011) stated that good results obtained by artificial neural networks could be explained by the adequate adjustment to the nonlinear systems. Moreover, according to Aquino et al. (2016a), the ANN consider numerous explanatory variables concomitantly in the model that does not always present good results by other statistical models, such as multiple linear regression. Several researchers also confirmed the efficiency of artificial intelligence found in this study (Azevedo et al., 2015; Brasileiro et al., 2015; Soares et al., 2015; Aquino et al., 2016a, 2016b).
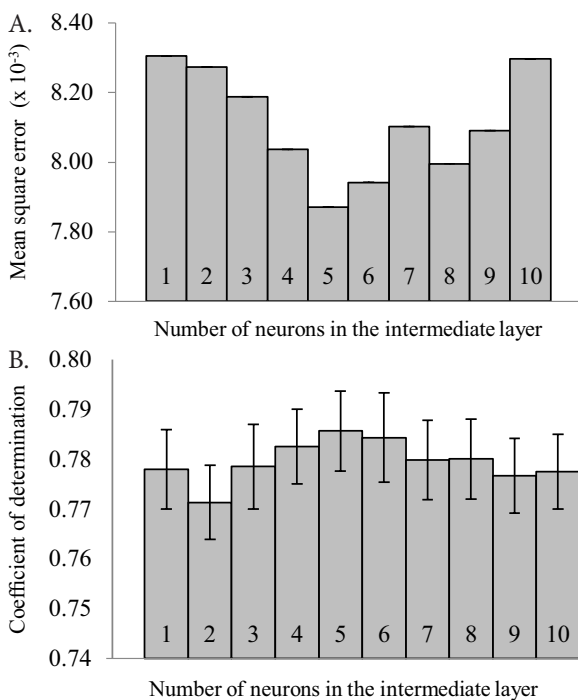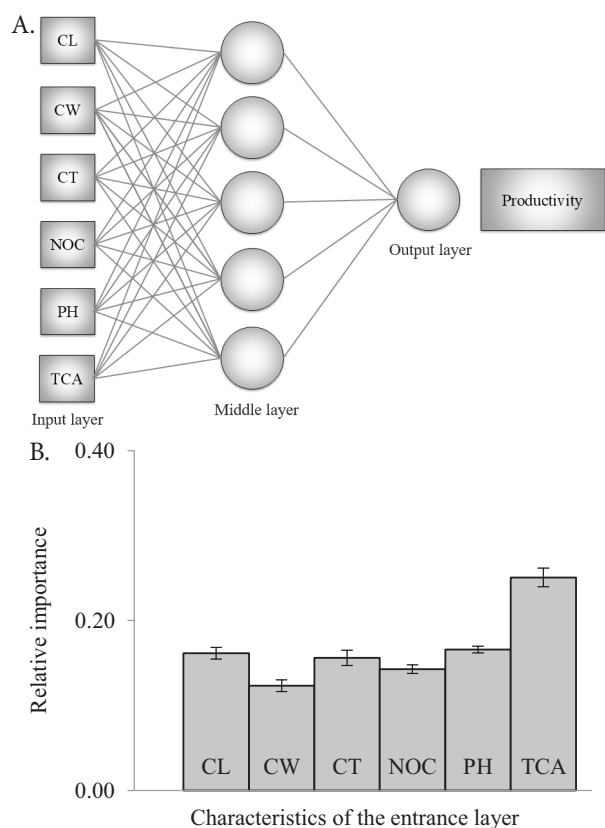
Figure 2. Topology of the best adjusted network (A) and relative contribution to yield (B), obtained by Garson's method (Garson, 1991), of the agronomic parameters cladode length (CL), cladode width (CW), cladode thickness (CT), number of cladode (NOC), plant height (PH), and total cladode area (TCA), presented in the input layer for the prediction of cactus pear yield by artificial neural networks
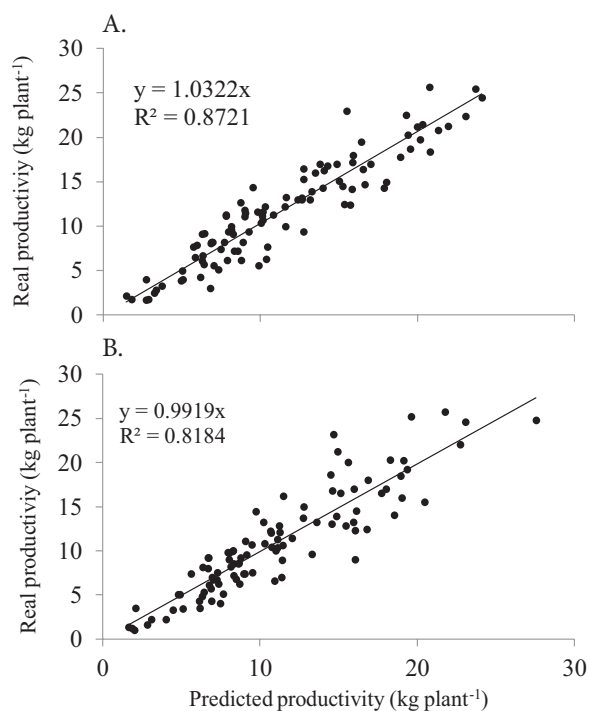

Figure 3. Graphical representation of the quality of artificial neural networks in predicting cactus pear yield for the validation sample considering vegetative characters (A) and excluding characteristic of lower relative contribution to yield (B)

The efficiency of using of MLP type ANN to obtain agronomic estimates was confirmed by several researchers (Binoti et al., 2013; Miguel et al., 2016). According to Vendruscolo et al. (2017), ANN models present statistical indicators with errors of estimation lower than 10%, thus ensuring the prediction of the phenomena.

New ANN's were adjusted excluding the characteristic of lower relative contribution to yield in this study (CW) obtaining a coefficient of determination of 0.82 (Figure 3B). This lower estimate, when compared to the adjusted networks with all the characteristics (Figure 2B), indicates that its exclusion for yield prediction is not feasible, even though the CW had a smaller contribution to yield.

Therefore, cactus pear yield can be estimated by agronomic characteristics measured in the field (CL, CW, CT, NOC, PH, and TCA). The results found in this study show that the application of the ANN models allows the prediction of cactus pear yield, and is an efficient and strategic tool in the decision making of its production, especially for agricultural planning for periods of scarcity or low feed availability for animal nutrition.

Conclusions

1. Predictions of forage palm yield are obtained with high efficiency through multi-layer perceptron-type artificial neural networks.

2. The morphological characters with the greatest relative contribution to predicting forage palm yield are total cladode area, plant height, cladode thickness and cladode length.

Acknowledgements

Literature Cited

Aguiar, M. do S. M. A.; Silva, F. F. da; Donato, S. L. R.; Rodrigues, E. S. de O.; Costa, L. T.; Mateus, R. G.; Souza, D. R. de; Silva, V. L. da. Palma forrageira em dietas de novilhas leiteiras confinadas: Desempenho e viabilidade econômica. Semina: Ciências Agrárias, v.36, p.1013-1030, 2015a. https://doi.org/10.5433/1679-0359.2015v36n2p1013

Aguiar, M. do S. M. A.; Silva, F. F. da; Donato, S. L. R.; Schio, A. R.; Souza, D. D. de; Meneses, M. de A.; Lédo, A. A. Síntese de proteína microbiana e concentração de ureia em novilhas leiteiras alimentadas com palma forrageira *Opuntia*. Semina: Ciências Agrárias, v.36, p.999-1012, 2015b. https://doi.org/10.5433/1679-0359.2015v36n2p999

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, p.268-274, 2016a. https://doi.org/10.1590/1678-4499.467

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, p.824-833, 2016b. https://doi.org/10.1590/S0100-204X2016000700005

Azevedo, A. M.; Andrade Júnior, V. C. de; Pedrosa, C. E.; Oliveira, C. M. de; 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, p.387-393, 2015. https://doi.org/10.1590/1678-4499.0088

Azevedo, A. M.; Andrade Júnior, V. C. de; 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, p.14-19, 2017. https://doi.org/10.1590/s0102-053620170103

Bergmeir, C.; Benítez, J. M. Neural networks in R using the Stuttgart Neural Network Simulator: RSNNS. Journal of Statistical Software, v.46, p.1-26. 2012. https://doi.org/10.18637/jss.v046.i07

Binoti, D. H. B.; Binoti, M. L. M.; Leite, H. G.; Silva, A. Redução de custos em inventário de povoamentos equiâneos. Agrária, v.8, p.125-129, 2013. https://doi.org/10.5039/agraria.v8i1a2209

Brasileiro, B. P.; Marinho, C. D.; Costa, P. M. de A.; Cruz, C. D.; Peternelli, L. A.; Barbosa, M. H. P. Selection in sugarcane families with artificial neural networks. Crop Breeding and Applied Biotechnology, v.15, p.72-78, 2015. https://doi.org/10.1590/1984-70332015v15n2a14

CODEVASF - Companhia de Desenvolvimento dos Vales do São Francisco e do Parnaíba. Disponível em: <http://www.codevasf.gov.br/search?SearchableText=temperatura>. Acesso em: 06 Nov. 2017.

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

Gianola, D.; Okut, H.; Weigel, K. A.; Rosa, G. J. M. Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat. BMC Genetics, v.12, p.1-14, 2011. https://doi.org/10.1186/1471-2156-12-87

Jana, R. B.; Mohanty, B. P. A. Comparative study of multiple approaches to soil hydraulic parameter scaling applied at the hillslope scale. Water Resources Research, v.48, p.1-16, 2012. https://doi.org/10.1029/2010WR010185

Jana, R. B.; Mohanty, B. P.; Sheng, Z. Upscaling soil hydraulic parameters in the Picacho Mountain region using Bayesian Neural Networks. Transactions of the ASABE, v.55, p.463-473, 2012. https://doi.org/10.13031/2013.41396

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, p.1914-1923, 2016. https://doi.org/10.5897/AJAR2016.11015

Padilha Junior, M. C.; Donato, S. L. R.; Silva, J. A. da; Donato, P. E. R.; Souza, E. S. Características morfométricas e rendimento da palma forrageira 'Gigante' sob diferentes adubações e configurações de plantio. Revista Verde de Agroecologia e Desenvolvimento Sustentável, v.11, p.67-72, 2016. https://doi.org/10.18378/rvads.v11i1.3710

R Development Core Team. The R Project for Statistical Computing. 2012. Disponível em: <https://www.r-project.org/>. Acesso em: 10 Mai. 2012.

Silva, I. N.; Spatti, D. H.; Flauzino, R. A. Redes neurais artificiais: Para engenharia e ciências aplicadas. 1.ed. São Paulo: Artliber. 2010. 399p.

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, p.1987-1993, 2015. http://dx.doi.org/10.1590/0103-8478cr20141524

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, p.84–90, 2014. https://doi.org/10.1590/0103-8478cr20141524

Storck, L.; Garcia, D. C.; Lopes, S. J.; Estefanel, V. Experimentação Vegetal. 3.ed. Santa Maria: UFSM, 2011. 200p.

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, p.52-58, 2017. https://doi.org/10.5935/2318-7670.v05n01a09