ISSN 1807-1929
Revista Brasileira de Engenharia Agrícola e Ambiental
Brazilian Journal of Agricultural and Environmental Engineering
v.28, n.9, e279246, 2024
Campina Grande, PB – http://www.agriambi.com.br – http://www.scielo.br/rbeaa

DOI: http://dx.doi.org/10.1590/1807-1929/agriambi.v28n9e279246

ORIGINAL ARTICLE

Leaf area estimation in *Coffea canephora* genotypes by neural networks and multiple regression[1]

Estimativa da área foliar de genótipos de *Coffea canephora* por meio de redes neurais e regressão múltipla

Edney L. da Vitória[2*], André O. Nardotto Júnior[2], Luis F. O. Ribeiro[2], Danielly Dubberstein[3] & Fábio L. Partelli[2]

[1] Research developed at Universidade Federal do Espírito Santo, São Mateus, ES, Brazil
[2] Universidade Federal do Espírito Santo/Centro Universitário Norte do Espírito Santo/Departamento de Ciências Agrárias e Biológicas, São Mateus, ES, Brazil
[3] Instituto Federal de Educação, Ciência e Tecnologia de Rondônia, Porto Velho, RO, Brazil

HIGHLIGHTS:
Backpropagation neural networks can be used to estimate leaf area of Coffea canephora.
The use of non-destructive methods is a viable alternative for determining leaf area of C. canephora.
Multiple regression can be replaced by artificial neural network models for estimating leaf area.

ABSTRACT: Leaf area data from coffee plants are important for studies and analyses of grain yield, physiology, adaptation to environmental conditions, and cultural management. The objective of this study was to predict leaf area of coffee plants using artificial neural networks and compare the efficiency of this methodology with multiple regression models. Forty-three genotypes of similar reproduction and age were evaluated, testing 14 types of multiple regression equations from combinations of leaf length and width. The backpropagation algorithm was used to develop multilayer perceptron neural networks; several combinations were tested between two activation functions of the intermediate layer (hidden layer) and the number of neurons in this layer. The best fitting results in the artificial neural network modeling were found with the sigmoid activation function and three neurons in the hidden layer (R^2 = 0.990 and RMSE = 2.855 in the training phase). Considering the errors (RMSE, MAE, and MAPE) and the coefficient of determination as criteria for best fit, the artificial neural network models better estimated the leaf area in the training and validation phases. Therefore, the artificial neural network methodology can be used as alternative for estimating leaf area of coffee plants.

Key words: statistical models, artificial intelligence, backpropagation, leaf length and width

RESUMO: Dados de área foliar de plantas de café são importantes para estudos e análises de produtividade, fisiologia, adaptação às condições ambientais e manejos culturais. O objetivo deste trabalho foi predizer a área foliar de plantas de café por meio de redes neurais artificiais e avaliar a eficiência dessa metodologia por meio de comparação com modelos de regressão múltipla. Foram avaliados 43 genótipos de reprodução e idade semelhantes e testados 14 tipos de equações de regressão múltipla a partir de combinações de comprimento e largura de folhas O algoritmo backpropagation foi utilizado para desenvolver redes neurais do tipo perceptron multicamadas, e foram testadas possíveis combinações entre duas funções de ativação da camada intermediária e o número de neurônios na camada intermediária. Na modelagem de redes neurais artificiais, os melhores resultados de ajuste foram encontrados com a função de ativação sigmoide e três neurônios na camada oculta (R^2 = 0,990; RMSE = 2,855 na fase de treinamento). Considerando os erros (RMSE, MAE e MAPE) e coeficientes de determinação como parâmetros estatísticos comparando os dois métodos utilizados, os modelos que utilizaram redes neurais artificiais apresentaram as melhores estimativas de área foliar nas fases de treinamento e validação. O método de redes neurais artificiais pode ser utilizado como alternativa ou modelo de apoio para estimativa de área foliar de cafeeiros.

Palavras-chave: modelos estatísticos, inteligência artificial, back propagation, comprimento foliar, largura

• Ref. 279246 – Received 05 Oct, 2023
* Corresponding author - E-mail: vitoria.edney@gmail.com
• Accepted 21 Apr, 2024 • Published 11 Jun, 2024
Editors: Ítalo Herbet Lucena Cavalcante & Walter Esfrain Pereira

Introduction

Leaf area is an important variable for studies evaluating agronomic aspects related to plant growth, development, and physiology, crop management and yield, efficacy of pesticide applications, water stress, and other crop-related aspects (Liu et al., 2021; Planas et al., 2022).

Different methods are available to measure leaf area, including direct and indirect methods (Teobaldelli et al., 2020). However, these methods are time-consuming, expensive, complex, and only suitable for some specific plant species (Kandiannan et al., 2009).

Mathematical and statistical approaches with modeling through linear and nonlinear regressions are often used to estimate leaf area based on leaf length, width, or their combination (Espindula et al., 2018; Lee et al., 2018; Dubberstein et al., 2019). Artificial intelligence techniques have recently been used to assist in decision-making, predictions, and estimates of morphological and physiological characteristics of agricultural crops (Azeem et al., 2020).

Artificial neural networks (ANNs) have been widely used to non-destructively estimate leaf area in several crops (Emamgholizadeh et al., 2015; Lee et al., 2018; Ercanlı et al., 2018; Azeem et al., 2020; Liu et al., 2021; Sá et al., 2022) based on leaf length and width data. In this sense, ANNs have yielded significant and accurate results for solving recurring problems connected to estimates involving complex nonlinear systems. ANNs can be used in agriculture systems for pest and disease control (Shah et al., 2020), root growth (Aji et al., 2020), and drought monitoring (Liu et al., 2020).

Studies on the use of ANNs to estimate leaf area have been carried out in several crops, such as wheat (Apolo-Apolo et al., 2020), *Capsicum annuum* L. (Lee et al., 2018), and peanuts (Qi et al., 2020); however, there are no studies on the use of this methodology for coffee crops. In this context, the objective of this study was to predict leaf area of coffee plants through artificial neural networks and compare its efficiency with multiple regression models.

Material and Methods

The experiment was conducted in a field containing 43 genotypes of *Coffea canephora* selected by local coffee growers (Table 1) in Nova Venécia, Espírito Santo, Brazil (18° 66’ 23” S, 40° 43’ 07” W, and altitude of 200 m). The climate of the region was classified as Aw, according to the Köppen classification, characterized by hot and humid summers and dry winters (Alvares et al., 2013), with a mean annual temperature of 23 °C.

The planting had been carried out with a spacing of 3.0 × 1.0 m, providing a density of 3,333 plants per hectare. Four-year-old coffee trees were evaluated in the experiment. The crop was irrigated using a drip irrigation system, and the plants were grown with four orthotropic stems.

Regarding the 43 genotypes, 42 of them were propagated by cuttings and one by seeds, using a randomized block experimental design with three replicates and seven plants per plot (Table 1).

Table 1. Identification of the 43 genotypes of *Coffea canephora* used in the experiment

Identification	Genotype	Identification	Genotype
1	Verdim R	23	Z18
2	B01	24	Z37
3	Bicudo	25	Z21
4	Alecrin	26	Z36
5	700	27	Ouro Negro
6	CH1	28	18
7	Imbigudinho	29	Tardio C
8	AD1	30	A1
9	Graudão HP	31	Cheique
10	Valcir	32	P2
11	Beira Rio 8	33	Emcapa 02
12	Late V	34	Emcapa 153
13	AP	35	P1
14	L80	36	LB1
15	Bamburral	37	122
16	Pirata	38	Verdim D
17	Peneirão	39	Seed
18	Z39	40	Emcapa 143
19	Z35	41	Ouro Negro 1
20	Z40	42	Ouro Negro 2
21	Z29	43	Clementino
22	Z38		

The third or fourth leaf of plagiotropic branches in the middle third of the plants were collected, resulting in 20 leaves for each genotype. The leaves were placed in identified bags and then sent to a laboratory for leaf measurements. Leaf dimensions were measured using a steel ruler with a precision of 0.5 mm. Leaf length was measured from the leaf tip to the leaf blade base on the petiole; leaf width was measured at the longest line perpendicular to the midrib of the leaf blade, considering the nearest millimeter. Leaf area was then measured using a leaf area meter (LI-3100, LI-COR, Lincoln, USA). These assessments were carried out in October 2016 and February 2017, totaling 1,720 leaf measurements.

Fourteen multiple regression equations were selected for estimating leaf area from combinations of the variables L (leaf length), W (leaf width), LW, L^2, W^2, L^2W^2, L^2W, and LW^2 (Table 2), as described in Encarli et al. (2018).

The use of few independent variables in the model can result in a low accuracy of the regression coefficient estimates; when low accuracy was found, factors connected to collinearity, i.e., the variance inflation factor (VIF) and tolerance values (T) were calculated. These measures are calculated based on the correlation coefficient. A VIF value less than 10 or T value higher than 0.10 represent no collinearity problems and no effect on the estimates using L, W, and their combinations.

Leaf blade length and width data were used as the input variable, while the measured leaf area data were used as the output variable; 70% (1,204 leaves) and 30% (516 leaves) of the data were used in the neural network training and validation processes, respectively. The input and output data were normalized to the range of 0.0 to 1.0 to improve network training efficiency, using Eq. 1.

$$V_{norm} = \frac{1+\left(V_{obs} - V_{est}\right)}{V_{max} - V_{min}} \quad (1)$$

Table 2. Types of multiple regression equations selected from combinations of the variables L (leaf length), W (leaf width), LW, L^2, W^2, L^2W^2, L^2W, and LW^2 for estimating leaf area of coffee (*Coffea canephora*) plants

Equation	Multiple regression models
1	$\hat{LA} = \beta_0 + \beta_1 L$
2	$\hat{LA} = \beta_0 + \beta_1 W$
3	$\hat{LA} = \beta_0 + \beta_1 L + \beta_3 W$
4	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 W$
5	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 W + \beta_3 W^2$
6	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 W + \beta_4 W^2$
7	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 W + \beta_4 LW$
8	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 W + \beta_4 WL$
9	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 W + \beta_3 W^2 + \beta_4 WL$
10	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_4 W + \beta_5 W^2 + \beta_6 WL$
11	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_4 W + \beta_5 W^2 + \beta_6 WL + \beta_7 L^2W$
12	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_4 W + \beta_5 W^2 + \beta_6 WL + \beta_7 LW^2$
13	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 W + \beta_4 W^2 + \beta_5 WL + \beta_6 L^2W + \beta_7 LW^2$
14	$\hat{LA} = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 W + \beta_4 W^2 + \beta_5 WL + \beta_6 L^2W + \beta_7 LW^2 + \beta_8 L^2W^2$

where:

V_{norm} - normalized value;
V_{obs} - observed value;
V_{max} - maximum value of the data sample; and,
V_{min} - minimum value of the data sample.

The backpropagation algorithm was used to develop the multilayer perceptron (MLP) neural networks; the learning rate and the number of training times set in the Levenberg-Marquardt optimization algorithm were 0.2 and 1000, respectively. The best network configuration was established by testing possible combinations between two activation functions of the intermediate layer (hidden layer) and the number of neurons in this layer (1 to 10), as recommended by Emamgholizadeh et al. (2015). The activation functions used were the hyperbolic tangent and the sigmoid (Eq. 2 and 3, respectively).

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$f(x) = \tanh(x) \tag{3}$$

The efficiency of the trained networks was compared with the 14 multiple regression models tested by comparing the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), using Eqs. 4 to 7. The criteria for selecting the best models considered the highest R^2 values and the lowest errors.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{5}$$

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{6}$$

$$MAPE = \frac{1}{n} \times \sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{\hat{y}_i}\right| \tag{7}$$

where:

n - sample size;
y_i - observed values;
y_i - estimated values; and,
y - mean values.

The neural network modeling and multiple regression analysis were performed using the Neuranet package in R (R Core Team, 2023).

Results and Discussion

Descriptive statistical parameters and Pearson linear correlation coefficients for leaf length, width, and area of coffee plants are shown in Table 3. The results showed a range (difference between the maximum and minimum means of

Table 3. Descriptive statistics and Pearson correlation coefficients for leaf length, width, and area of *Coffea canephora* genotypes (n = 1,720)

Parameters	Leaf length (cm)	Leaf width (cm)	Leaf area (cm^2)
	Descriptive statistics		
Minimum	9.00	3.80	20.63
Maximum	19.00	9.70	101.61
Mean	13.97	5.79	51.08
Standard deviation	1.71	0.98	14.20
Coefficient of variation	12.24%	16.93%	27.80%
	Pearson correlation coefficients (r)		
Leaf length	1.000		
Leaf width	0.643*	1.000	
Leaf area	0.921*	0.823*	1.000
	VIF (T)		
Leaf length	-		
Leaf width	1.703 (0.587)	-	
Leaf area	6.579 (0.152)	3.096 (0.323)	-

*Significant at p ≤ 0.05 by the t-test; VIF - Variance inflation factor; Values in parentheses refer to standard deviations each variable) of 10.00, 5.90, and 80.98 cm^2 for leaf length, width, and area, respectively. The ranges of means and standard deviations for these variables showed a data variability (coefficients of variation) of 12.24, 16.93, and 27.80%, respectively. According to the methodological classification proposed by Costa et al. (2002), coefficient of variations between 10.00 and 30.00% are considered moderate. Similar results were found by Dubberstein et al. (2019), who reported coefficients ranging from 9.56 to 32.51% when evaluating leaf area of coffee plants.

Variability in leaf length, width, and area across different coffee genotypes is essential information for developing more accurate mathematical models for leaf area estimation, favoring the development of models for application to small, medium, and large leaves. Effects of leaf length and width variability on leaf area estimates have been found for coffee (Partelli et al., 2006), guava (Vitória et al., 2018), *Capsicum annuum* L. (Lee et al., 2018), and soybean (Sá et al., 2022) plants.

The independent variables leaf length and area had a moderate correlation (r = 0.643), whereas the correlations between leaf area and leaf length (r = 0.921) and between leaf area and leaf width (r = 0.823) were high, according to the classification proposed by Rousseau et al. (2018), which considers correlation coefficients above 0.7 as high and between 0.3 and 0.7 as moderate. Espindula et al. (2018) and Dubberstein et al. (2019) found similar trends for correlation coefficients between leaf length, width, and area of *C. canephora* plants. Studying the correlations between variables to be used in trained models provides important information to assess how variation in one variable affects another and how these effects impact the accuracy of model estimates (Dubberstein et al., 2019; Ercanlı et al., 2018).

The variance inflation factor (VIF) values ranged from 1.073 to 6.579 and the tolerance (T) values from 0.152 to 0.587. According to Gill (1986), a VIF value lower than 10.0 or T value higher than 0.10 denotes negligible collinearity between leaf length and width. Thus, both variables can be included in the model.

The results of the 14 multiple linear regression equations tested to estimate leaf area of *C. canephora* genotypes are shown in Table 4. The models had coefficients of determination (R^2) higher than 0.970, except for models 1, 2, and 3; the highest R^2 (0.973) was found for models 7, 8, 9, and 11. The lowest error values were used a an additional criterion for selecting the best model; the lowest RMSE and MAE (Figure 1D) were found for model 7, in which the variables L and W and the product of these variables resulted in the best leaf area estimate.

Model 7 (R^2 = 0.973; RMSE = 3.303) was selected and then for validated using a 35% subsample of the total collected data; the values were not used to define the model in the training process. The leaf area estimated by the selected model and its correlation with the observed leaf area are shown in Figure 1A. The leaf area estimated by the selected regression model had a high correlation (R^2 = 0.946) with the observed leaf area. The difference between observed and estimated leaf areas ranged from approximately -20 cm^2 and +20 cm^2, and more than 70% of relative errors ranged from -5 to 5%, as shown in Figures 1B and C.

One hundred neural networks were trained for each configuration; the highest R^2 and lowest RMSE for each configuration are shown in Table 5. Overall, the sigmoid activation function showed higher R^2 and lower RMSE than hyperbolic tangent. The sigmoid function with 3 to 8 neurons in the hidden layer presented the same high correlations in both the training (R^2 = 0.990) and validation (R^2 = 0.989) phases. Considering the lowest RMSE as the selection criterion, the neural network with sigmoid activation function showed the highest R^2 and lowest RMSE in the training (R^2 = 0.990; RMSE = 2.853) and validation (R^2 = 0.989 and RMSE = 2.862) phases.

The results for the efficiency of neural networks in estimating leaf area in coffee plants are shown in Figure 2. The estimated leaf area was highly correlated with that the observed leaf area in the training (R^2 = 0.944) and validation (R^2 = 0.946) phases, i.e., 94.4 and 94.6% of the variation in estimated leaf area values can be explained by the observed values in the training and validation phases, respectively (Figures 2A and B). The absolutes errors ranged from -20.0 to + 20.0 cm^2 in both phases (Figures 2C and D). More than 70% of the MAPE was between -5 and +5% in the training phase, and more than 75% was within the same range in the validation phase (Figures 2E and F).

The leaf area of the evaluated coffee genotypes showed a good regularity in length and width when estimated by the selected neural network (Figure 3A). Information on which input variable (leaf length or width) is more important for estimating leaf area is useful because it enables the estimation of leaf area considering only one of these variables, depending

Table 4. Multiple regression models for estimating leaf area (LA) of *Coffea canephora* genotypes

Id	Regression model	R^2	RMSE	F Value
1	$LA = -44.94 + 6.87^{*}L$	0.823	7.989	2,778.0
2	$LA = -25.98 + 13.30^{*}W$	0.921	5.527	7,209.3
3	$LA = -50.89 + 3.32^{ns}L + 9.59^{*}W$	0.970	3.419	10,448.5
4	$LA = -14.63 - 1.91^{*}L + 0.18^{ns}L^2 + 9.58^{*}W$	0.972	3.574	7,308.9
5	$LA = -33.97 + 3.42^{*}L + 3.46^{*}W + 0.49^{*}W^2$	0.972	3.371	7,235.1
6	$LA = -8.61 - 0.87^{*}L + 0.15^{ns}L^2 + 5.01^{*}W + 0.37^{*}W^2$	0.972	3.326	5,591.2
7	$LA = -9.63 + 0.42^{*}L + 2.26^{ns}W + 0.51^{*}LW$	0.973	3.303	7,492.6
8	$LA = -7.28 - 0.31^{*}L + 0.004^{ns}L^2 + 3.19^{*}W + 0.44^{*}LW$	0.973	3.314	5,620.7
9	$LA = -10.10 + 0.70^{ns}L + 1.76^{*}W + 0.09^{*}W^2 + 0.46^{*}LW$	0.973	3.322	5,619.8
10	$LA = -6.63 - 0.28^{ns}L + 0.06^{*}L^2 + 2.92^{*}W + 0.16^{*}W^2 + 0.32^{*}LW$	0.972	3.469	4,502.4
11	$LA = 39.94 - 6.99^{ns}L + 0.30^{*}L^2 - 5.46^{*}W + 0.16^{*}W^2 + 1.52^{*}LW - 0.04^{*}L^2W$	0.973	3.980	3,760.7
12	$LA = 42.14 - 3.42^{ns}L + 0.04^{*}L^2 - 15.05^{*}W + 1.66^{*}W^2 + 1.57^{*}LW - 0.10^{*}LW^2$	0.972	3.377	3,767.9
13	$LA = 43.43 - 3.78^{ns}L + 0.06L^{2\,ns} - 14.87^{*}W + 1.59^{*}W^2 + 1.60^{*}LW - 0.01^{*}L^2W - 0.09^{*}LW^2$	0.972	4.949	3,227.2
14	$LA = 50.72 - 4.83^{ns}L + 0.10^{ns}L^2 - 17.51^{*}W + 1.82^{*}W^2 + 1.98^{*}LW - 0.02^{ns}L^2W - 0.13^{*}LW^2 + 0.001^{*}L^2W^2$	0.973	6.283	2,821.6

L - Leaf length; W - Leaf width; LW – Multiplication of leaf length and width; RMSE - Root mean square error; * - Significant at $p < 0.05$; ^{ns} - Not significant

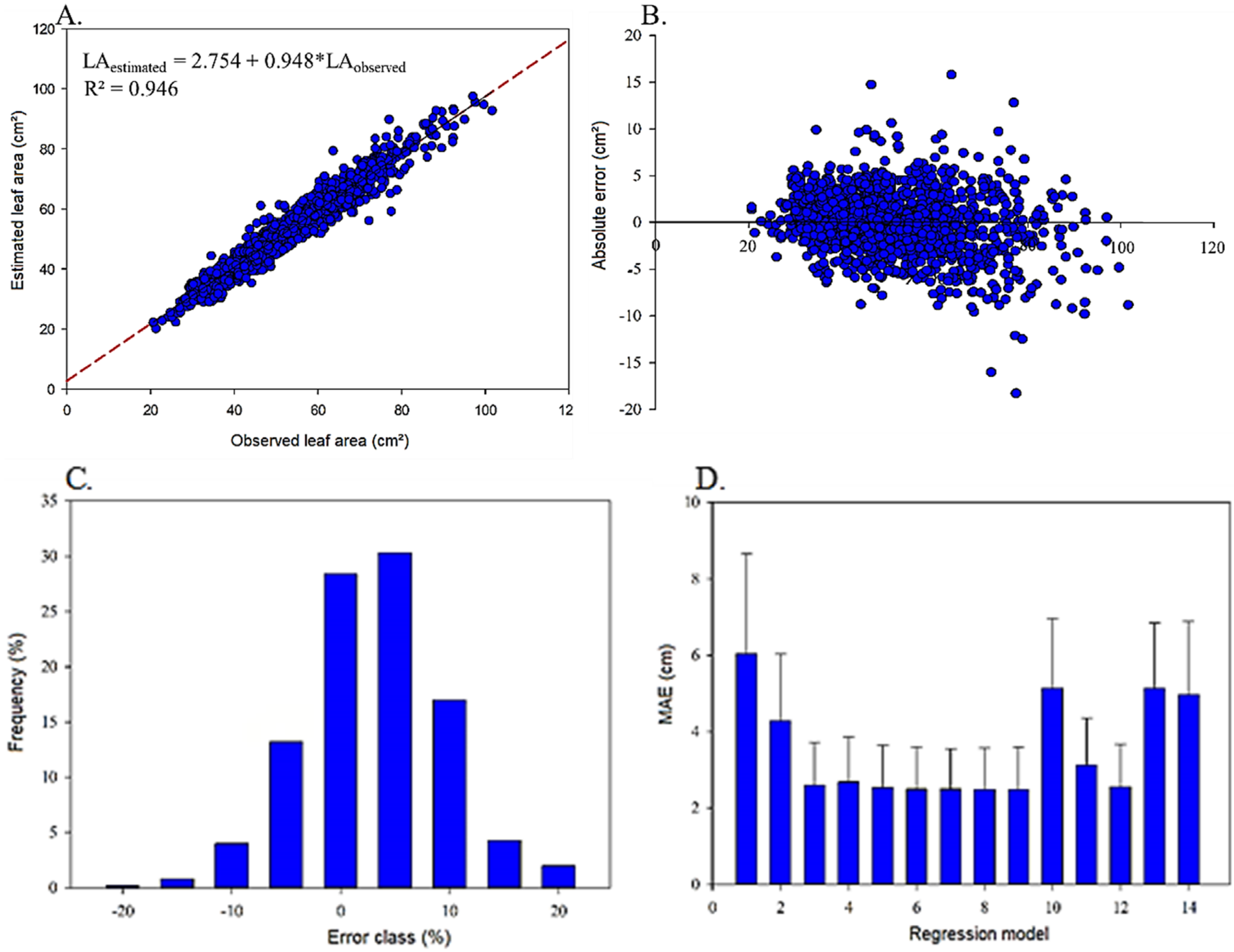


Figure 1. Leaf area estimated by the selected model and its correlation with the observed leaf area. Correlation between observed and estimated leaf area (A); distribution of absolute errors between observed and estimated values (B); distribution of error frequency (C); mean absolute error (D)

Table 5. Root mean square error (RMSE) and coefficients of determination (R^2) for 100 neural networks trained for each network configuration

Activation function	Number of neurons in the hidden layer	Training		Validation	
		RMSE	R²	RMSE	R²
Sigmoid	1	5.559	0.967	5.853	0.958
	2	4.771	0.978	4.831	0.974
	3	2.853	0.990	2.862	0.989
	4	2.866	0.990	2.874	0.989
	5	2.968	0.990	2.980	0.989
	6	2.909	0.990	2.945	0.989
	7	2.895	0.990	2.921	0.989
	8	2.919	0.990	2.945	0.989
	9	3.289	0.988	3.349	0.986
	10	3.170	0.988	3.219	0.986
Hyperbolic tangent	1	56.680	0.923	55.222	0.915
	2	52.680	0.870	55.221	0.785
	3	52740	0.820	55.665	0.797
	4	52.890	0.693	55.222	0.656
	5	56.648	0.804	55.225	0.777
	6	55.611	0.931	55.150	0.920
	7	56.672	0.935	55.213	0.927
	8	56.684	0.879	55.225	0.870
	9	56.684	0.740	55.225	0.697
	10	56.674	0.710	55.217	0.671

on its contributions (Dubberstein et al., 2019; Ercanlı et al., 2018). The method of Garson (1991) enables the obtaining the percentage of contribution of input variables to leaf area estimation. Leaf length contributed more (56.8%) to leaf area estimation through neural network than leaf width (Figure 3B). This higher contribution was expected, as leaf length had

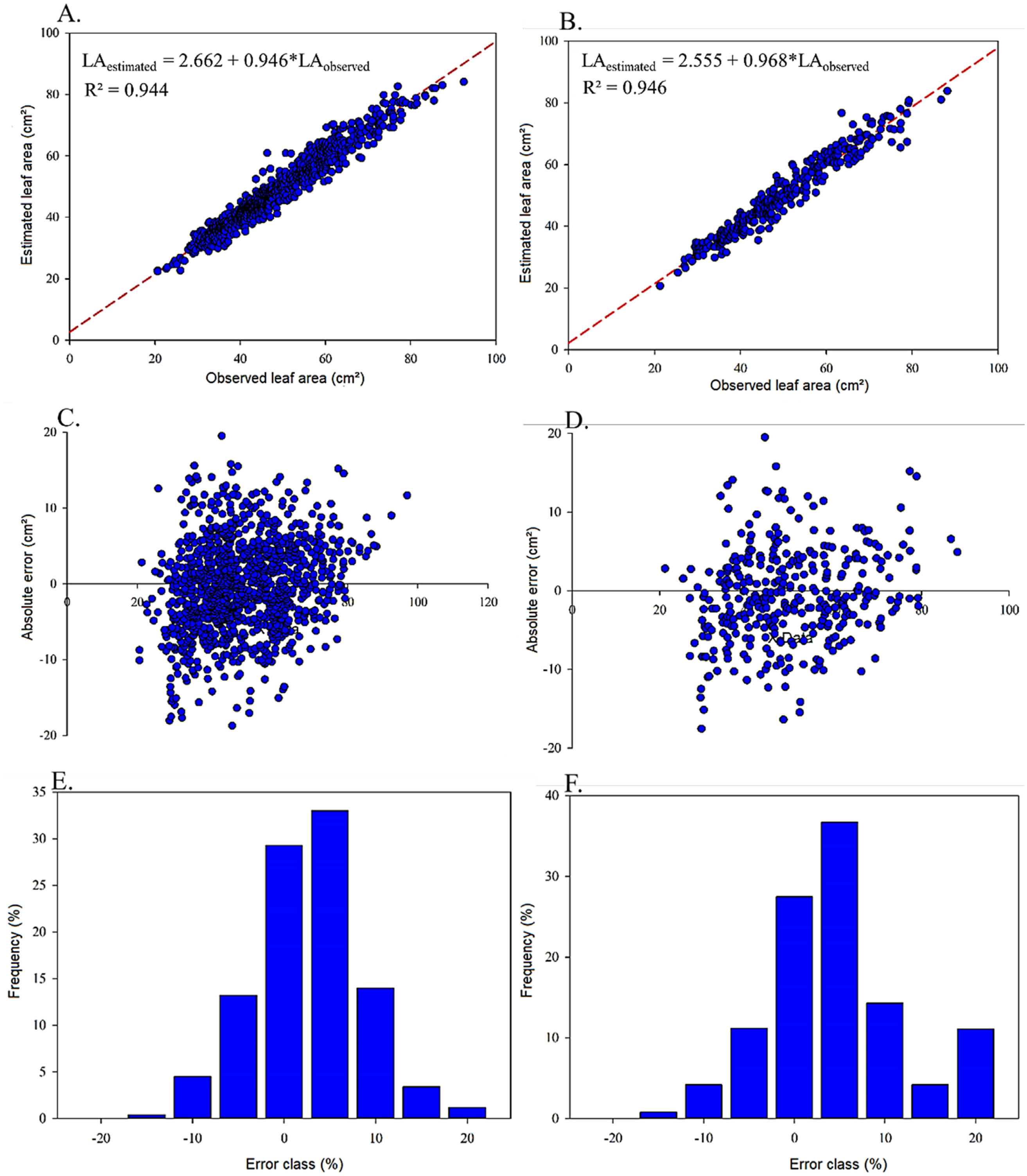


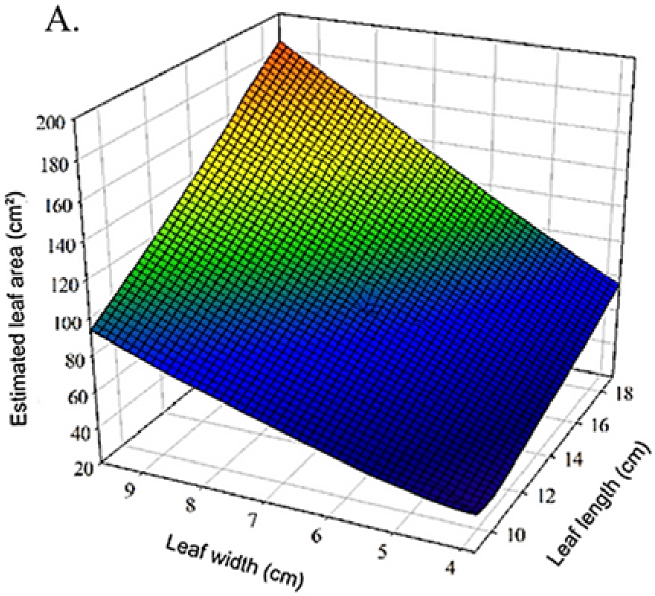
Figure 2. Results of the efficiency in estimating coffee leaf area by neural networks. Correlation between estimated and observed areas in the training (A) and validation (B) phases; distribution of absolute errors between observed and estimated values in the training (C) and validation (D) phases; distribution of error frequency in the training (E) and validation (F) phases

higher correlation with leaf area (r = 0.921) than leaf width (r = 0.823) (Table 3).

The comparison between the selected multiple regression model (model 7) and the trained and validated neural network model is shown in Table 6.

The coefficient of determination of the artificial neural network model was higher than that found for the multiple linear regression model in both the testing and validation phases. The same trend was found for the root mean square error. Artificial neural networks were more flexible when using nonlinear models between the input variables (leaf length and width) and the output variable (leaf area), which is essential for explaining the variability found. Therefore, the comparison with the multiple linear regression model showed the superior performance of the artificial neural network in estimating leaf area of the evaluated coffee genotypes.

Considering the errors (RMSE, MAE, and MAPE) and coefficient of determination ($R^2 = 0.989$), the artificial neural



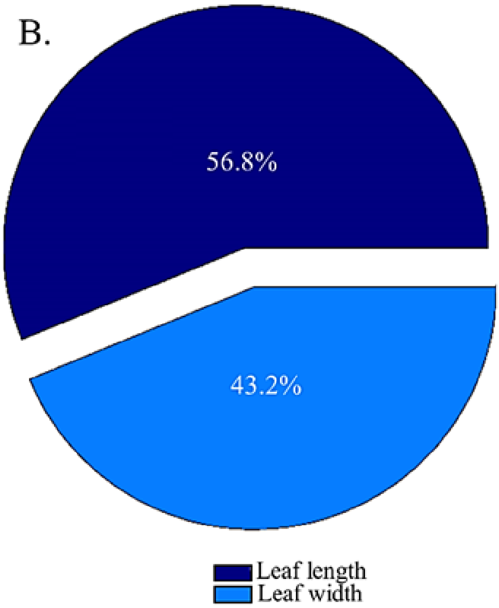


Figure 3. Leaf area estimation in coffee plants based on leaf width and length (A) by neural network, and percentage of contribution of leaf width and length to leaf area estimation (B)

Table 6. Comparison of the coefficient of determination (R^2) and root mean square error (RMSE) between the multiple regression model (model 7) and the neural network model

		Artificial neural network	
		Training	Validation
R^2	0.973	0.990	0.989
RMSE	3.303	2.853	2.862

network model better estimated leaf area in the training and validation phases (Tables 4, 5, and 6). This better performance can be attributed to the large amount of data from different coffee genotypes used for the estimations and to the use of nonlinear correlations for input and output data by artificial neural networks.

The efficiency of leaf area estimation by artificial neural networks depends on variations in leaf length and width, i.e., the greater the variability, the higher accuracy of the estimation. The variability in leaf length and width was assured by the high number of coffee genotypes evaluated in the present study.

The efficiency of artificial neural networks in predicting leaf area depends on the variability in leaf shape data (length and width data) used in the training phase (Wang et al., 2017). Thus, the use of a high number of genotypes contributes to make the trained network generalist for different leaf shapes and sizes.

Conclusions

1. Artificial neural networks are more efficient to estimate leaf area of *Coffea canephora* plants than multiple regression models.
2. A network architecture with three neurons in the hidden layer provides better leaf area estimation.
3. The proposed artificial neural network model is simple and fast, requiring only leaf length and leaf blade width data, allowing repeated, non-destructive measurements on the same leaves.

Contribution of authors: Edney L. da Vitória: elaboração da primeira versão do manuscrito, revisão da literatura, pesquisa, análise dos dados, orientação e implementação dos modelos computacionais administração dos recursos e revisão da versão final do manuscrito. André O. N. Orlandi: preparation of the first version of the manuscript implementation of the computer models. Luis F. O. Ribeiro: preparation of the manuscript, literature review and final revision of the manuscript. Danielly Dubberstein: collection, acquisition of data in the experimental field, literature review and final revision of the manuscript. Fábio L. Partelli: guidance of the research project, responsible for data curation, administration and acquisition of funding, revision of the final version of the manuscript.

Supplementary documents: There are no supplementary sources.

Conflict of interest: The authors declare no conflict of interest.

Financing statement: This study was partially funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and by the Fundação de Amparo à Pesquisa do Espírito Santo (FAPES), Brazil.

Literature Cited

Aji, G. K.; Hatou, K.; Morimoto, T. Modeling the dynamic response of plant growth to root zone temperature in hydroponic chili pepper plant using neural networks. Agriculture, v.10, e234, 2020. https://doi.org/10.3390/agriculture10060234

Alvares, C. A.; Stape, J. L.; Sentelhas, P. C.; Gonçalves, J. L. de M.; Sparovek, G. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift, v.22, p.711-728, 2013. https://doi.org/10.1127/0941-2948/2013/0507

Apolo-Apolo, O. E.; Pérez-Ruiz, M.; Martínez-Guanter, J.; Egea, G. A mixed data-based deep neural network to estimate leaf area index in wheat breeding trials. Agronomy, v.10, e175, 2020. https://doi.org/10.3390/agronomy10020175

Azeem, A.; Javed, Q.; Sun, J.; Du, D. Artificial neural networking to estimate the leaf area for invasive plant Wedelia trilobata. Nordic Journal of Botany, v.38, p.1-8, 2020. https://doi.org/10.1111/njb.02768

Costa, N. H. D. C.; Seraphin, J. C.; Zimmermann, F. J. P. A new method of variation coefficient classification for upland rice crop. Pesquisa Agropecuária Brasileira, v.37, p.243-249, 2002. https://doi.org/10.1590/s0100-204x2002000300003

Dubberstein, D.; Martins, L. D.; Ferreira, A.; Guilhen, J. H.; Ramalho, J. C.; Partelli, F. L. Equations for estimation of the foliar area of *Coffea canephora* genotypes. Genetics and Molecular Research, v.18, p.1-12, 2019. https://doi.org/10.4238/gmr18486

Emamgholizadeh, S.; Parsaeian, M.; Baradaran, M. Seed yield prediction of sesame using artificial neural network. European Journal of Agronomy, v.68, p.89-96, 2015. https://doi.org/10.1016/j.eja.2015.04.010

Ercanlı, İ.; Günlü, A.; Şenyurt, M.; Keleş, S. Artificial neural network models predicting the leaf area index: A case study in pure even-aged crimean pine forests from Turkey. Forest Ecosystems, v.5, p.1-12, 2018. https://doi.org/10.1186/s40663-018-0149-8

Espindula, M. C.; dos Passos, A. M. A.; Araújo, L. F. B.; Marcolan, A. L.; Partelli, F. L.; Ramalho, A. R. Indirect estimation of leaf area in genotypes of "Conilon" coffee (*Coffea canephora* Pierre ex A. Froehner). Australian Journal of Crop Science, v.12, p.990-994, 2018. https://doi.org/10.21475/ajcs.18.12.06.PNE1090

Garson, G. D. Interpreting neural-network connection weights. Journal AI Expert, v.6, p.47-51, 1991.

Gill, J. L. Outliers, residuals, and influence in multiple regression. Journal of Animal Breeding and Genetics, v.103, p.161-175, 1986.

Kandiannan, K.; Parthasarathy, U.; Krishnamurthy, K. S.; Thankamani, C. K.; Srinivasan, V. Modeling individual leaf area of ginger (*Zingiber officinale* Roscoe) using leaf length and width. Scientia Horticulturae, v.120, p.532-537, 2009. https://doi.org/10.1016/j.scienta.2008.11.037

Lee, J.; Moon, T.; Nam, D. S.; Park, K. S.; Son, J. E. Estimation of leaf area in paprika based on leaf length, leaf width, and node number using regression models and an artificial neural network. Horticultural Science and Technology, v.36, p.183-192, 2018. https://doi.org/10.12972/kjhst.20180019

Liu, X.; Zhu, X.; Zhang. Q.; Yang, T.; Pan, Y.; Sun, P. A remote sensing and artificial neural network-based integrated agricultural drought index: Index development and applications. Catena, v.186, e104394, 2020. https://doi.org/10.1016/j.catena.2019.104394

Liu, S.; Jin, X.; Nie, C.; Wang, S.; Yu, X.; Cheng, M.; Shao, M.; Wang, Z.; Tuohuti, N.; Bai, Y.; Liu, Y. Estimating leaf area index using unmanned aerial vehicle data: Shallow vs. Deep machine learning algorithms. Plant Physiology, v.187, p.1551-1576, 2021. https://doi.org/10.1093/plphys/kiab322

Partelli, F. L.; Vieira, H. D.; Detmann, E.; Campostrini, E. Estimativa da área foliar do cafeeiro conilon a partir do comprimento da folha. Revista Ceres, v.53, p.204-210, 2006.

Planas, S.; Román, C.; Sanz, R.; Rosell-Polo, J. R. Bases for pesticide dose expression and adjustment in 3D crops and comparison of decision support systems. Science of the Total Environment, v.806, e150357, 2022. https://doi.org/10.1016/j.scitotenv.2021.150357

Qi, H.; Zhu, B.; Wu, Z.; Liang, Y.; Li, J.; Wang, L.; Chen, T.; Lan, Y.; Zhang, L. Estimation of peanut leaf area index from unmanned aerial vehicle multispectral images. Sensors, v.20, e6732, 2020. https://doi.org/10.3390/s20236732

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available on: <https://www.r-project.org> Accessed on: Aug. 2023.

Rousseau, R.; Egghe, L.; Guns, R. Becoming metric wise. 1ed. Cambridge: Elsevier, 2018.

Sá, L. G. De; Juliano, C.; Albuquerque, B.; Valadares, N. R.; Brito, G.; Mota, A. N., Clara, A.; Fernandes, G.; De, A. M. Area estimation of soybean leaves of different shapes with artificial neural networks. Acta Scientiarum. Agronomy, v.44, p.1-9, 2022. https://doi.org/https://doi.org/10.4025/actasciagron.v44i1.54787

Shah, T.M.; Nasika, D.P.B.; Otterpohl, R. Plant and weed identifier robot as an agroecological tool using artificial neural networks for image identification. Agriculture, v.11, e222, 2020. https://doi.org/10.3390/agriculture11030222

Teobaldelli, M.; Rouphael, Y.; Gonnella, M.; Buttaro, D.; Rivera, C. M.; Muganu, M.; Basile, B. Developing a fast and accurate model to estimate allometrically the total shoot leaf area in grapevines. Scientia Horticulturae, v.259, e108794, 2020. https://doi.org/10.1016/j.scienta.2019.108794

Vitória, E. L.; Freitas, I. L.J.; Locatelli, T.; Lacerda, E.G.; Valle, J. M.; Pereira, R. C.; Almeida, P. F. P.; Vitória, R. Z.; Simon, C.P.; Fernandes, A. A. Mathematical models for leaf area estimates of guava. Journal of Agricultural Science, v.10, p.272-278, 2018. https://doi.org/10.5539/jas.v10n12p272

Wang, G.; Sun, Y.; Wang, J. Automatic image-based plant disease severity estimation using deep learning. Comput. Computational Intelligence and Neuroscience, v.1, e2917536, 2017. https://doi.org/10.1155/2017/2917536