Heliyon 9 (2023) e17834

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Artificial neural network approach for predicting the sesame (*Sesamum indicum* L.) leaf area: A non-destructive and accurate method

João Everthon da Silva Ribeiro *, Ester dos Santos Coêlho, Anna Kézia Soares de Oliveira, Antonio Gideilson Correia da Silva, Welder de Araújo Rangel Lopes, Pablo Henrique de Almeida Oliveira, Elania Freire da Silva, Aurélio Paes Barros Júnior, Lindomar Maria da Silveira

Federal Rural University of the Semi-Arid, Mossoró, Rio Grande do Norte, Brazil

A R T I C L E I N F O

Keywords:
Leaf length
Leaf width
Machine learning
Multilayer perceptrons
Sesamum indicum L.
Simple linear regression

A B S T R A C T

The estimative of the leaf area using a nondestructive method is paramount for successive evaluations in the same plant with precision and speed, not requiring high-cost equipment. Thus, the objective of this work was to construct models to estimate leaf area using artificial neural network models (ANN) and regression and to compare which model is the most effective model for predicting leaf area in sesame culture. A total of 11,000 leaves of four sesame cultivars were collected. Then, the length (L) and leaf width (W), and the actual leaf area (LA) were quantified. For the ANN model, the parameters of the length and width of the leaf were used as input variables of the network, with hidden layers and leaf area as the desired output parameter. For the linear regression models, leaf dimensions were considered independent variables, and the actual leaf area was the dependent variable. The criteria for choosing the best models were: the lowest root of the mean squared error (RMSE), mean absolute error (MAE), and absolute mean percentage error (MAPE), and higher coefficients of determination (R^2). Among the linear regression models, the equation $\hat{y} = 0.515 + 0.584 * LW$ was considered the most indicated to estimate the leaf area of the sesame. In modeling with ANNs, the best results were found for model 2-3-1, with two input variables (L and W), three hidden variables, and an output variable (LA). The ANN model was more accurate than the regression models, recording the lowest errors and higher R^2 in the training phase (RMSE: 0.0040; MAE: 0.0027; MAPE: 0.0587; and R^2: 0.9834) and in the test phase (RMSE: 0.0106; MAE: 0.0029; MAPE: 0.0611; and R^2: 0.9828). Thus, the ANN method is the most indicated and accurate for predicting the leaf area of the sesame.

1. Introduction

The leaf is the main photosynthetic organ of plants, responsible for important physiological and ecophysiological functions [1]. Several factors affect plant yield and growth, such as leaf size, leaf shape, leaf thickness, number of leaves per plant, and leaf area [2]. Among these, the leaf area is considered by some researchers as the most important parameter in evaluating the growth, development,

* Corresponding author.
E-mail address: j.everthon@hotmail.com (J.E. da Silva Ribeiro).

https://doi.org/10.1016/j.heliyon.2023.e17834
Received 12 June 2023; Received in revised form 21 June 2023; Accepted 28 June 2023
Available online 11 July 2023

productivity, and physiology of vegetables [3]. The estimation of leaf area is important for studies involving the interaction between plants and the environment, the composition of plant communities, evolution, and adaptation of plants [4–6]. In addition, leaf area is paramount in ecological, agronomic, and horticultural studies involving interception and absorption of light, respiration, transpiration, liquid CO_2 absorption, stomatal opening and closing, internal carbon concentration in responses to fertilizer use, irrigation methods and biotic and abiotic stresses in plants [7].

The leaf area can be measured using different methods, classified as direct and indirect, destructive and non-destructive methods [8]. Most direct methods destroy plant leaves, and planimeters, millimeter paper, scans, photographs, and leaf discs can be used to determine leaf area [9]. Thus, it is impossible to perform successive measurements on the same leaf or plant throughout the life cycle [10]. In addition to these methods, new equipment (portable or benchtop) and software were developed to determine the leaf area, such as laser optical scanners and software for analyzing digital images. These direct methods (destructive and high cost) are simple and precise; however, they require more time and workforce in the analyses [11]. On the other hand, indirect methods are low-cost, simple, fast, and accurate for the in situ estimation of the leaf area, being indicated for studies with multiple evaluations in the same individual without destroying the sample [12]. One of the indirect (non-destructive) methods most commonly used today is the estimation of leaf area using regression models and allometric equations, with leaf area being the dependent variable and dimensional parameters of leaf blade (e.g., length and width) the independent variables for these analyses [13].

Regression models using length and width were proposed to estimate the leaf area of many oilseed species such as macaúba palm [11], sunflower [14], walnut [4], canola [15,16], chia [17], cocoa [18], brazil chestnut [19], moringa [20], peanut [21] and basil [22]. In addition to linear dimensions, such as length and width, other studies have proposed a correction coefficient to correct a systematic bias in estimation and increase the accuracy of leaf area measurements in plants [23–25]. However, the estimation of leaf area through regression modeling presents limitations, where the best regression models depend on a good sampling amplitude and morphological aspects of leaves and are useful only when there is a linear or non-linear relationship between predictor variables and response variables. In situations with many predictors, it is impossible to observe a nonlinear and complex relationship between the variables of responses and predictor variables [26].

Currently, in contrast to traditional scientific methods, researchers are using and proposing other methodologies in modern research to determine the leaf area, such as artificial intelligence (AI), adaptive neuro-fuzzy inference system (ANFIS), and artificial neural network (ANN) [27]. ANNs are computational models based on biological neurons (artificial neurons) with the objective of processing and transmitting information [28]. The development of ANNs was based on principles of human brain functioning, and it is possible to obtain relevant results in estimating parameters compared to traditional methods, such as regression and correlation analyses [29]. Ann's overall structure consists of three layers: the input layer, the hidden layer, and the output layer. The input and hidden layers consist of artificial and input neurons that allow one to receive information from the external world and transmit the information received to a hidden layer without changing the input data [30].

Currently, ANNs are used in several areas worldwide due to their practicality and accuracy in predictive data analysis, including research in biological, ecological, and agronomic areas. In the last decade, ANNs have been increasingly employed in plant phenotyping community. They have been very effective in modeling complicated concepts, owing to their ability of distinguishing patterns and extracting regularities from data. Examples include variety identification in seeds [31] and in intact plants by using leaves [32]. The ANNs were applied in research with the identification of grain pests [33], modeling of dynamic responses of plant growth affected by climate change [34], estimation of corn grain yield [35], estimation of soil quality for crops [36], prediction of the area of harvest, yield, and production [37], prediction of greenhouse gas emissions [38], forecasts on the accumulation of heavy metals in crops [39] and determine seed germination [40]. In addition, non-destructive predictive models using ANN to estimate leaf area were proposed for species such as durian [13], cabbage [41], wedelia [42], tomato [43] and corn [44]. Thus, the use of ANN has been an important alternative to statistical methods, providing promising results that would be difficult to predict in biological and agricultural systems [45].

Sesamum indicum L. (sesame) is an oilseed belonging to the Pedaliaceae family, cultivated in tropical and subtropical regions of Africa, Asia, and South America; being considered the oldest oilseed in the world and known as the 'queen of oilseeds' due to the high economic importance in the growing regions [46,47]. In Brazil, culture represents an income alternative for small and large producers in semiarid regions, with a high commercial value of seeds and oil, being easy to productively manage and tolerant of the edaphoclimatic conditions of these regions [48]. Sesame seeds are rich in oil (46–64%), protein (15–25%) (mainly methionine), carbohydrates (20–25%), and some micronutrients such as lignans and phytosterol [49,50]. The oil fraction present in the seeds contains approximately 90% of unsaturated fatty acids, such as oleic and linoleic acids [49]. Essential oil and seeds of the species are widely used worldwide to produce various herbal products, food, and cosmetics [51,52].

In most cases, there is a strong intraspecific variation in leaf shape. For more generic models, therefore, it is highly advised to encompass several cultivars, representing a wide variation in leaf shape [53]. For sesame culture, no studies are related to the estimation of leaf area using models. The modeling of agricultural species can provide information on phenological stages, planting conditions, and different crop environments. Thus, we formulated the hypothesis that the ANN model is the most indicated and precise regression model to estimate the leaf area of sesame cultivars. According to this perspective, the objectives of this study were to (i) construct models to estimate leaf area using linear measurements of leaves using ANN and regression analysis, (ii) to compare the most precise non-destructive method to estimate leaf area, and (iii) to evaluate the efficacy of the models in a set of grouped data to estimate the leaf area of the sesame independent of the cultivar.

2. Material and methods

2.1. Plant material and experimental conditions

The experiment was carried out in an experimental area of the Rafael Fernandes Experimental Farm in the district of Alagoinha, belonging to the Federal Rural University of the Semi-Arid, in the municipality of Mossoró, Rio Grande do Norte, Brazil (5°03′31.48″S, 37°23′47.28″W). The region's climate is dry and very hot, with a dry and rainy season classified as BSh [54]. The average annual temperature is approximately 27.8°C, and rainfall is around 555 mm annually. The region has soil classified as a Eutrophic Red-Yellow Argisol in the Brazilian Soil Classification System [55].

For the construction of the models, sesame cultivars were planted from July 2022 to September 2022, and the leaves were collected 60 days after planting (beginning of the flowering season). Each experimental plot consisted of a cultivar sowed in lines of 50 m, with a density of 3 plants per linear meter and spacing of 0.60 between rows, with a total area of 7.2 m^2 in each plot. The plants presented an average height of 1.70 m and a canopy diameter of approximately 0.50 m.

2.2. Plant sampling, image processing, and data analysis

Eleven thousand mature, fully-expanded leaves (n) were collected, free of pests and diseases and without damage caused by biotic or abiotic factors. The leaves were collected from four sesame cultivars (BRS Seda, CNPA G2, CNPA G3, and CNPA G4) (Fig. 1), and 2750 leaves were randomly collected in each cultivar. Leaves of different sizes and shapes were collected for greater data variability and generality of the models. After collection, the leaves were stored in plastic containers and kept in the shade to avoid water loss by transpiration, maintaining the degree of turbidity of the same.

The sheets were individually separated and then scanned in a scanner (Epson, model L3250, Tokyo, Japan) with a resolution of 800 × 800 dpi, and the images were processed, contrasted, and analyzed with the ImageJ (National Institutes of Health, USA) software in the public domain, according to the methodology of Ribeiro et al. (2018) [56]. For each measurement, reference scales were used using two rulers graduated in millimeters. In each leaf, the length (L) (distance between the end of the leaf tip to the petiole insertion point), width (W) (wider measure perpendicular to the central rib of the blade) was calculated, and the actual leaf area (LA), which was used as a reference for the construction of the models. The aspect ratio between length and width (L/W) was calculated, being a metric (without dimension) of the shape of the leaves. For the regression models, the data were analyzed with the software R® v.4.1.2 [57], and the artificial neural network model (ANN) was developed with the Software Matlab® v. 9.13. To convert the ANN model to a web page and Excel file, Tiberius data mining software v. 7.07 was used. The Student's t-test compared the leaf areas observed and estimated by the ANN and regression models for paired samples ($p < 0.01$).

2.3. ANN model

The development of MLP (Multi-Layer-Perceptron) networks was used by the Neural Network Toolbox of the Matlab software (version 9.13), a backpropagation algorithm with Levemberg-Marquadt optimization. The leaf's length (L) and width (W) were used as input variables of the network, with hidden layers and leaf area as the desired output parameter. The trial and error approach was used to find suitable neurons in the hidden layer of the network [58], and the learning rate was 0.04. The relative importance of the input variables (length and width) was analyzed to observe each leaf parameter's contribution to the ANN model. The final RNA model was selected based on the lowest root of the mean squared error (RMSE) (Eq. (1)) [59], mean absolute error (MAE) (Eq. (2)) [59], and absolute mean percentage error (MAPE) (Eq. (3)), and higher coefficients of determination (R^2) (Eq. (4)).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\widehat{y}i - yi)^2}{n}} \tag{1}$$

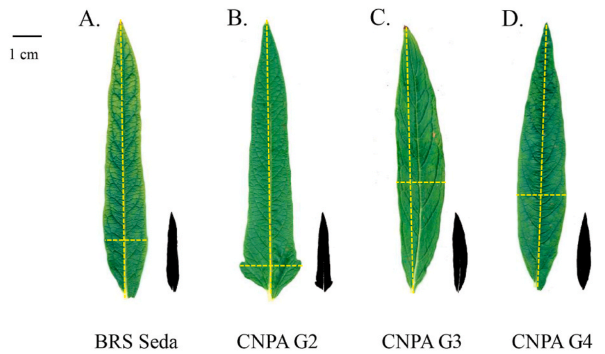

Fig. 1. Representative leaves of sesame cultivars: BRS Seda (A), CNPA G2 (B), CNPA G3 (C), and CNPA G4 (D). The vertical line represents the length of each sheet, and the dashed line represents the width of the booklet.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |\widehat{y}i - yi|}{n} \tag{2}$$

$$\mathrm{MAPE} = \frac{\sum_{i=1}^{n} \frac{|\widehat{yi - yi}|}{yi} * 100}{n} \tag{3}$$

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

2.4. Linear regression models

The leaf area was estimated using ten allometric equations constructed by the linear regression model ($\widehat{y} = \beta_0 + \beta_1 * x + \varepsilon_i$), in which the value of $\widehat{y}$ estimated the leaf area (LA) as a function of x, which corresponds to the linear dimensions of the leaves of the sesame cultivars. The leaf parameters used in the equations were: L, W, LW, LL, WW, L + W, $(L + W)^2$, (L^2W^2), $L^{0.5}$, and $W^{0.5}$). Were randomly collected 3300 independent leaves among the cultivars to validate the equations obtained with the regression models. Using more than one measure in regression models (e.g., length and width) may present colinearity between the parameters, resulting in low precision in estimating the coefficients of the models. Therefore, to verify whether or not there is accuracy in the regression coefficients, the degree of colinearity was evaluated by calculating the inflation factor of variance (VIF) (Eq. (5)) [60] and tolerance value (T) (Eq. (6)) [61]. If the VIF is less than 10 and the T is greater than 0.1, L and W do not have multilinearity and can be used in regression models for leaf area prediction [61]. The criteria for choosing the best regression models were the same as for choosing the ANN model (R^2, RMSE, MAE, and MAPE). The structure of this model is shown in Fig. 2.

$$\mathrm{VIF} = \frac{1}{1 - r^2} \tag{5}$$

$$T = \frac{1}{\mathrm{VIF}} \tag{6}$$

Where:
r is the correlation coefficient between L and W.

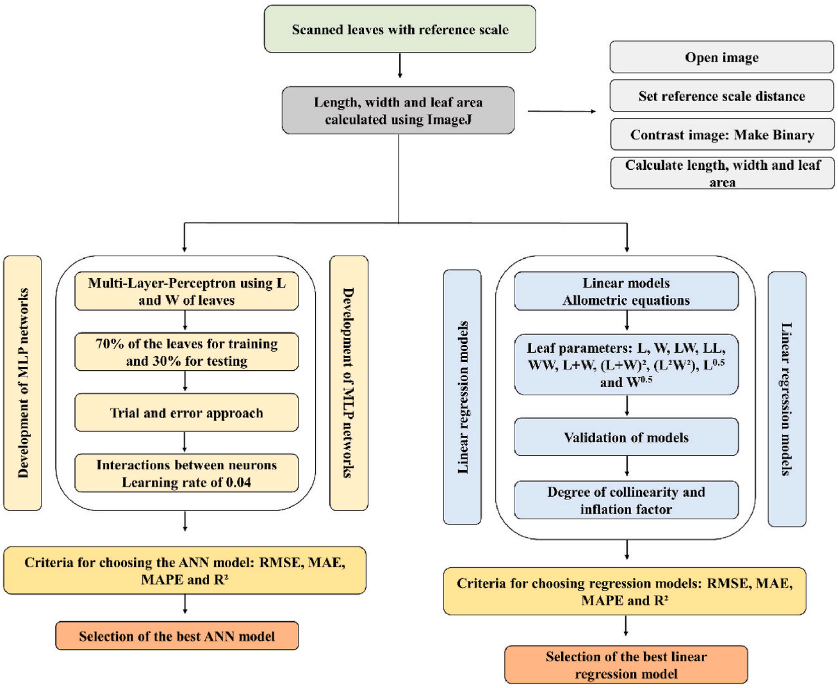

Fig. 2. Structure of models used to estimate sesame leaf area.

3. Results and discussion

3.1. Descriptive analysis of data

The maximum and minimum values, means, standard deviation, and coefficient of variation of the data of length, width, and leaf area were used to develop models that estimate the leaf area of the sesame, which are presented in Table 2. An amplitude of 12,495 cm and 6285 cm was observed for the length and width of the leaves and 37,353 cm^2 for the leaf area, where the highest coefficient of variation of the data was recorded (>76%) (Table 1). This wide variability of data is of great importance for studies with modeling for prediction of the leaf area of crops, providing models of greater representativeness to be used in plants with leaves of different shapes and dimensions in distinct phenological stages throughout the plant cycle [62,63]. Thus, the number of samples (11,000 leaves) used in this study is considered appropriate for constructing models for predicting leaf area as a function of the linear dimensions of the leaves.

3.2. ANN model

The ANN method was used as the first method for the construction of the model to estimate the leaf area of sesame. The structure of the model (2-3-1) to estimate leaf area is shown in Fig. 3. The complete data set (11,000 sheets) was randomly divided into two parts, 70% for training and 30% for the test. Thus, 7700 leaves were used for training and 3300 for the test. The summary of ANN parameters to determine leaf area is presented in Table 2. The maximum number of interactions between neurons was 3,000, using the logistic sigmoid activation function and the backpropagation algorithm for machine learning (Table 2).

ANN's results showed that RMSE values ranged from 0.0040 to 0.0106, MAPE between 0.0587 and 0.0611, MAE between 0.0027 and 0.0029, and R^2 between 0.9828 and 0.9834, respectively (Table 3). These results showed a high correlation between the actual leaf area and the leaf area estimated by the ANN model, with 98.34% in the training phase and 98.28% in the test phase (Table 3). The MAPE values were less than 10%, indicating that the ANN model had a high degree of precision for the prediction of the leaf area and good generalization capacity of the network for various formats and sizes of leaves [42,43,64]. According to Wang and Zhang (2012) [65], the efficiency of the models obtained with ANN depends on factors such as shape, size, and number of leaves of different cultivars, confirming that the high number of leaves in this study contributed to the accuracy of the model.

The total variability of the data in the training and test phases was greater than 98% (Table 3). Leaf length contributed 100% to the ANN model, while leaf width was 53.8% important (Fig. 4). The potential for employing a single leaf dimension has been exploited, though the associated reduction in prediction accuracy often outweighs the benefit of reduced labor. By using a single leaf dimension, the accuracy of LA estimation is compromised by the fact that changes in L and W are generally not proportional among replicate leaves, in combination with other changes in leaf shape [66]. This independent analysis of the importance of leaf parameters is essential for selecting input variables for studies with modeling using ANN [26].

The actual leaf area and leaf area estimated by the ANN model in the training phase (Fig. 5A) and the test phase (Fig. 5C) showed a high correlation between the data ($R^2 > 0.96$), with low dispersion of the residues, with no difference by the Student's t-test between

Table 1
Descriptive statistics of sesame leaf parameters (data from 11,000 leaves).

Parameters	Maximum	Minimum	Mean and standard deviation	Coefficient of variation (%)
BRS Seda				
Length (cm)	13.892	1.454	6.437 ± 2.351	36.52
Width (cm)	6.523	0.322	2.165 ± 1.184	54.68
Aspect ratio	31.901	1.142	3.678 ± 2.008	54.59
Leaf area (cm^2)	45.222	1.004	9.493 ± 7.142	75.23
CNPA G2				
Length (cm)	13.456	1.683	5.626 ± 2.079	36.95
Width (cm)	6.452	0.437	2.143 ± 1.078	50.27
Aspect ratio	21.006	1.182	3.114 ± 1.623	52.10
Leaf area (cm^2)	48.671	1.009	8.095 ± 6.205	76.65
CNPA G3				
Length (cm)	13.210	1.496	5.555 ± 1.917	34.51
Width (cm)	6.572	0.295	2.161 ± 1.157	53.56
Aspect ratio	23.559	1.125	3.169 ± 1.700	53.65
Leaf area (cm^2)	43.184	1.000	8.053 ± 6.287	78.07
CNPA G4				
Length (cm)	12.085	1.397	5.818 ± 2.097	36.04
Width (cm)	6.462	0.238	2.162 ± 1.135	52.50
Aspect ratio	32.639	1.142	3.307 ± 1.953	59.06
Leaf area (cm^2)	41.143	1.005	8.593 ± 6.503	75.67
Grouped data				
Length (cm)	13.892	1.397	5.860 ± 2.142	36.55
Width (cm)	6.523	0.238	2.157 ± 1.141	52.89
Aspect ratio	32.639	1.125	2.727 ± 1.821	54.86
Leaf area (cm^2)	38.217	0.864	8.564 ± 6.581	76.84

Table 2
ANN parameters to determine the leaf area of sesame.

Parameters	Value
Number of neurons in the input layer	2
Number of hidden layers	1
Number of hidden layer neurons	3
Number of output layers	1
Learning algorithm	Backpropagation algorithm
Learning rate	0.04
Maximum number of interactions	3000
Activation function	Logistic sigmoid

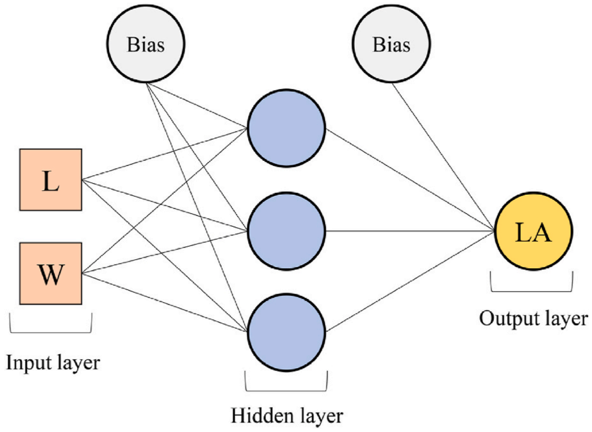

Fig. 3. ANN structure with the parameters of length (L) and width (W) (input layer) and leaf area (LA) (output layer).

Table 3
Root mean quadratic error (MSE), absolute mean percentage error (MAPE), mean absolute error (MAE), and coefficients of determination (R^2) for the training and test data of ANN.

Criteria	Training data	Test data
RMSE	0.0040	0.0106
MAPE	0.0587	0.0611
MAE	0.0027	0.0029
R^2	0.9834	0.9828

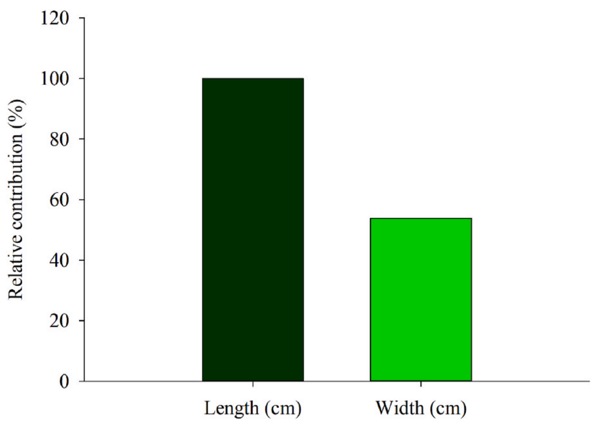

Fig. 4. The relative importance of input variables (length and width) for the ANN model.

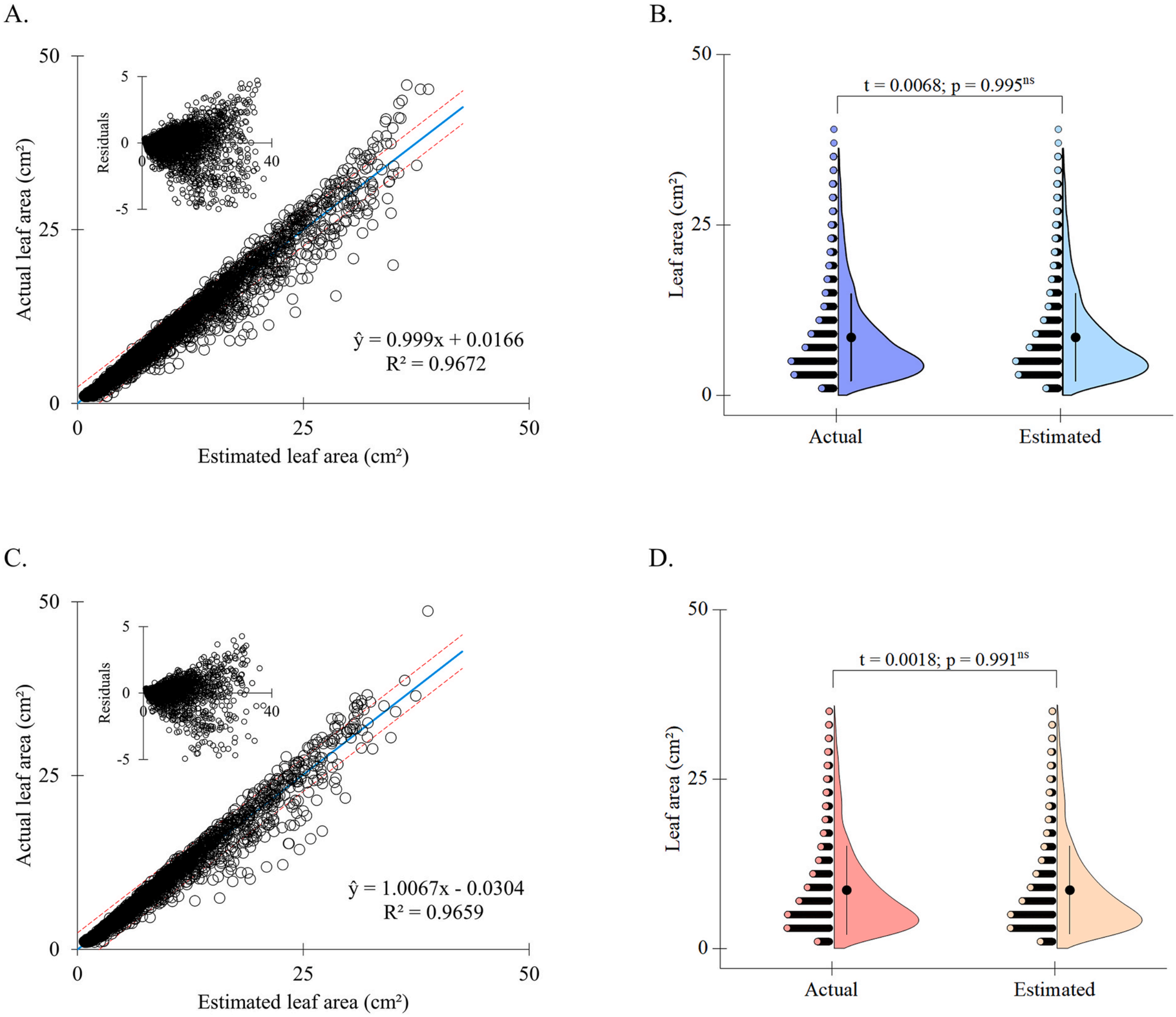

Fig. 5. Relationship between actual and estimated leaf area using data from training (A) and ANN test (C), and comparison between actual (B and D) and estimated leaf area using Student's t-test ($p < 0.01$). The dispersion analysis of the residuals is presented in the insert.

the actual and estimated values by ANN (Fig. 5B and D). To help researchers predict sesame leaf area using leaf length and width as input data, the ANN model was converted into a Web page (compatible with major browsers) (Fig. 6) and an Excel file (see supplementary material) to assist in measurements quickly and accurately. In this way, the proposed ANN model is high computational processing power and low-cost tool with a significant advantage, which is its real-time application for the prediction of leaf area in the sesame crop, being able to use electronic devices, such as computers and smartphones, to calculate the leaf area through the available web page.

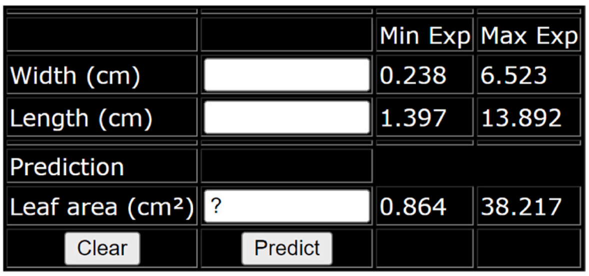

Fig. 6. Webpage interface for predicting sesame leaf area by ANN is based on the leaf length and width measurements.

3.3. Linear regression models

Before the definition of each model, the analysis was performed to calibrate the models to observe if there is a collinearity between the data of length and width of the leaves. The VIF ranged between 2.1862 and 9.6246, and the Values of T ranged from 0.1070 to 0.4574, respectively. Thus, the VIF values were <10, and T was >0.10, indicating that the parameters of length and width of the leaves can be included in the regression models, with no collinearity between the data. Models were constructed through regression analysis using different combinations between the linear dimensions of the leaves (length and width), as shown in Table 4. Based on the selection criteria for choosing the best regression model (RMSE, MAE, MAPE, and R^2) to estimate the leaf area of the sesame, it was observed that model #3 ($\hat{y} = 0.515 + 0.584 * \mathrm{LW}$) was constructed using the product between length and width (LW), is the most indicated and precise, recording RMSE of 1.2115, MAE of 0.7023, MAPE of 0.0838, and R^2 of 0.9560 (Table 4, Fig. 7). In addition to this model, the other models can also be used to estimate the leaf area of the sesame, except for models #1, #4, and #9, which showed little adjustment about the line, with coefficients of determination (R^2) lower than 0.60 (Table 4, Fig. 7). Regression models that use only one leaf dimension (length or width) can be an alternative to simplify the analyses; however, these models may cause a loss of precision in estimating the leaf area [67].

The best regression model to estimate the leaf area of the sesame showed high adjustment of the data about the straight obtained ($R^2 = 0.9560$), evidencing a residual homogeneity and low dispersion of the data (Fig. 8). The leaf area estimated using the chosen model strongly correlated with the actual leaf area (determined by the destructive method), with a coefficient of determination (R^2) of 0.9648 (Fig. 9A), confirming that there is a significant relationship between the actual and estimated leaf area (Fig. 9B). Thus, the proposed regression model can be used to estimate the leaf area of the sesame using the product between leaf length and width (LW). For other species, models were also proposed to estimate leaf area using leaf product (LW) as an independent variable [20,21,68–70].

3.4. Comparison of ANNs and regression models

When comparing the methods presented in the present study, it was found that the ANN model presented superiority over the regression models. Based on the criteria, it was observed that the ANN model recorded higher values of R^2 (>0.98) and lower errors (RMSE: ≤0.0106; MAE: ≤0.0611; MAPE: ≤0.0029) compared to regression models in the training phase and the test phase, indicating that the ANN model is more accurate than regression models. Analyzing these methods (regression and ANN) with combinations of linear parameters of leaves and input and output variables, researchers confirmed that ANN modeling was the most accurate, reliable, and efficient method than linear regression models, being possible to explain the nonlinear and complex relationships between input and output values [71,72].

The actual leaf area presented irregularities as a function of leaf length and width (Fig. 10A and B), a fact that may be associated with different leaf morphotypes of sesame cultivars, hindering measurements and may cause underestimated or overestimated measurements. On the other hand, using the ANN model provided a greater regularity for the leaf area estimated according to these parameters, according to Fig. 10C and D.

Thus, through ANN, researchers can estimate the sesame's leaf area using only the leaves linear dimensions (length and width). This method is an efficient and practical alternative to predict the leaf area of sesame culture throughout the plant life cycle without the need for high-cost equipment and quick evaluations. However, the method proposed through the ANN model cannot be used in sesame cultivars with leaf morphotypes different from the cultivars of the present study, which is a possible limitation of this model.

On a commercial scale, capital investment is initially required for adopting the employed approach [73]. Nevertheless, the wide-ranging large-scale commercial applications can provide high returns through considerable process enhancement and cost reduction improvements.

4. Conclusions

Based on the dataset, the ANN model had the lowest errors and higher R^2 than the regression models. However, this study had limitations. We identified concerns for estimating the leaf area of other sesame cultivars with different leaf morphotypes from those presented in the present study. Despite these limitations, this model can accurately and quickly estimate the leaf area without needing high-cost equipment. For future studies, this study will provide a non-destructive method to successfully predict the sesame leaf area simply and easily through ANN modeling using only the length and width of the leaves.

Author contribution statement

Conceived and designed the experiments; João Everthon da Silva Ribeiro, Anna Kézia Soares de Oliveira, Antonio Gideilson Correia da Silva

Performed the experiments; Anna Kézia Soares de Oliveira, Welder de Araújo Rangel Lopes, Aurélio Paes Barros Júnior, Lindomar Maria da Silveira

Analyzed and interpreted the data; João Everthon da Silva Ribeiro, Ester dos Santos Coêlho, Elania Freire da Silva

Contributed reagents, materials, analysis tools or data; Antonio Gideilson Correia da Silva, Elania Freire da Silva, Aurélio Paes Barros Júnior, Lindomar Maria da Silveira

Wrote the paper.João Everthon da Silva Ribeiro, Ester dos Santos Coêlho, Anna Kézia Soares de Oliveira

Table 4
Statistical models, regression coefficients (β_0 and β_1), coefficient of determination (R^2), mean square error (MSE), and equations for estimating leaf area of sesame cultivars as a function of linear leaf dimensions (pooled data).

Model code	Model	Coefficients		Equation
		β_0	β_1	
#1	$\hat{y} = \beta_0 + \beta_1 * L + \varepsilon_i$	- 5.029	2.319	$\hat{y} = -5.029 + 2.319 * L$
#2	$\hat{y} = \beta_0 + \beta_1 * W + \varepsilon_i$	- 2.430	5.095	$\hat{y} = -2.430 + 5.095 * W$
#3	$\hat{y} = \beta_0 + \beta_1 * LW + \varepsilon_i$	0.515	0.584	$\hat{y} = 0.515 + 0.584 * LW$
#4	$\hat{y} = \beta_0 + \beta_1 * LL + \varepsilon_i$	1.596	0.179	$\hat{y} = 1.596 + 0.179 * LL$
#5	$\hat{y} = \beta_0 + \beta_1 * WW + \varepsilon_i$	3.144	0.909	$\hat{y} = 3.144 + 0.909 * W W$
#6	$\hat{y} = \beta_0 + \beta_1 * (L + W) + \varepsilon_i$	- 8.429	2.119	$\hat{y} = -8.429 + 2.119 * (L + W)$
#7	$\hat{y} = \beta_0 + \beta_1 * (L + W)^2 + \varepsilon_i$	- 0.103	0.119	$\hat{y} = -0.103 + 0.119 * (L + W)^2$
#8	$\hat{y} = \beta_0 + \beta_1 * (L^2W^2) + \varepsilon_i$	5.120	0.011	$\hat{y} = 5.120 + 0.011 * (L^2W^2)$
#9	$\hat{y} = \beta_0 + \beta_1 * (L^{0.5}) + \varepsilon_i$	- 17.460	10.940	$\hat{y} = -17.460 + 10.940 * (L^{0.5})$
#10	$\hat{y} = \beta_0 + \beta_1 * (W^{0.5}) + \varepsilon_i$	−2.430	10.190	$\hat{y} = -2.430 + 10.190 * (W^{0.5})$

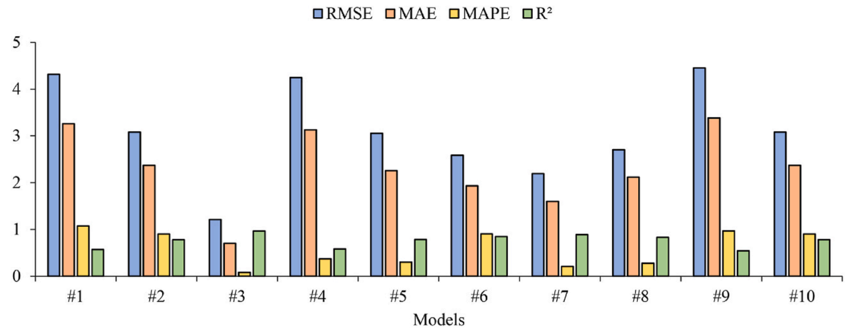

Fig. 7. Root of the mean squared error (RMSE), coefficients of determination (R^2), mean absolute error (MAE) (Eq. (3)), and absolute mean percentage error (MAPE) (Eq. (4)) of sesame leaf area as a function of linear leaf dimensions (length and width). The description of the models is presented in Table 4.

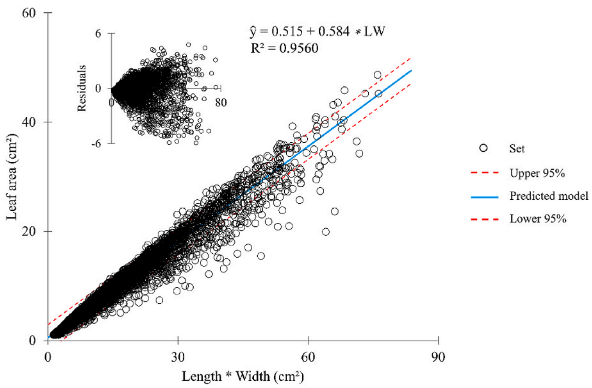

Fig. 8. Relationship between the observed leaf area and the product between length and width of sesame leaves (grouped data), from the model LA = 0.515 + 0.584*LW. The dispersion analysis of the residuals is presented in the insert.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Joao Everthon da Silva Ribeiro reports administrative support, article publishing charges, equipment, drugs, or supplies, statistical analysis, travel, and writing assistance were provided by Federal Rural University of the Semi-Arid. There was no financial

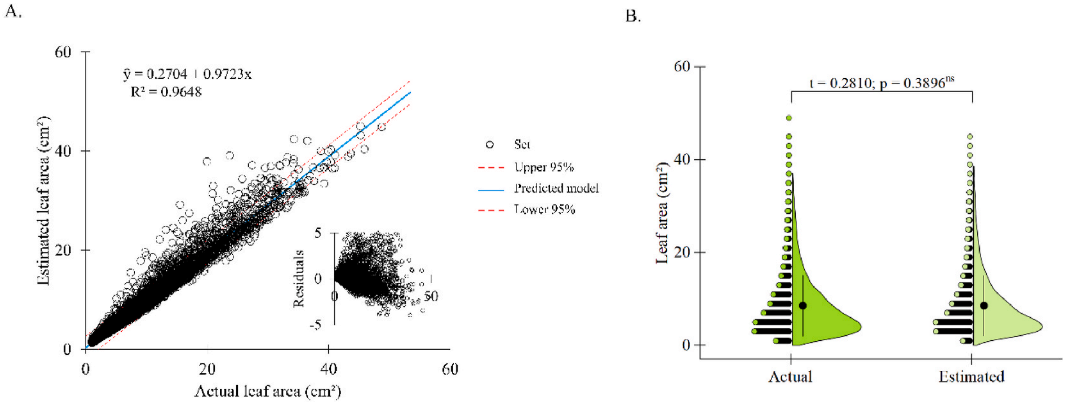

Fig. 9. (A) Relationship between leaf area observed and estimated by the model (LA = 0.515 + 0.584*LW) as a function of the product between length and width. The dispersion analysis of the residuals is presented in the insert. (B) Observed and estimated leaf area compared by Student's t-test ($p < 0.01$).

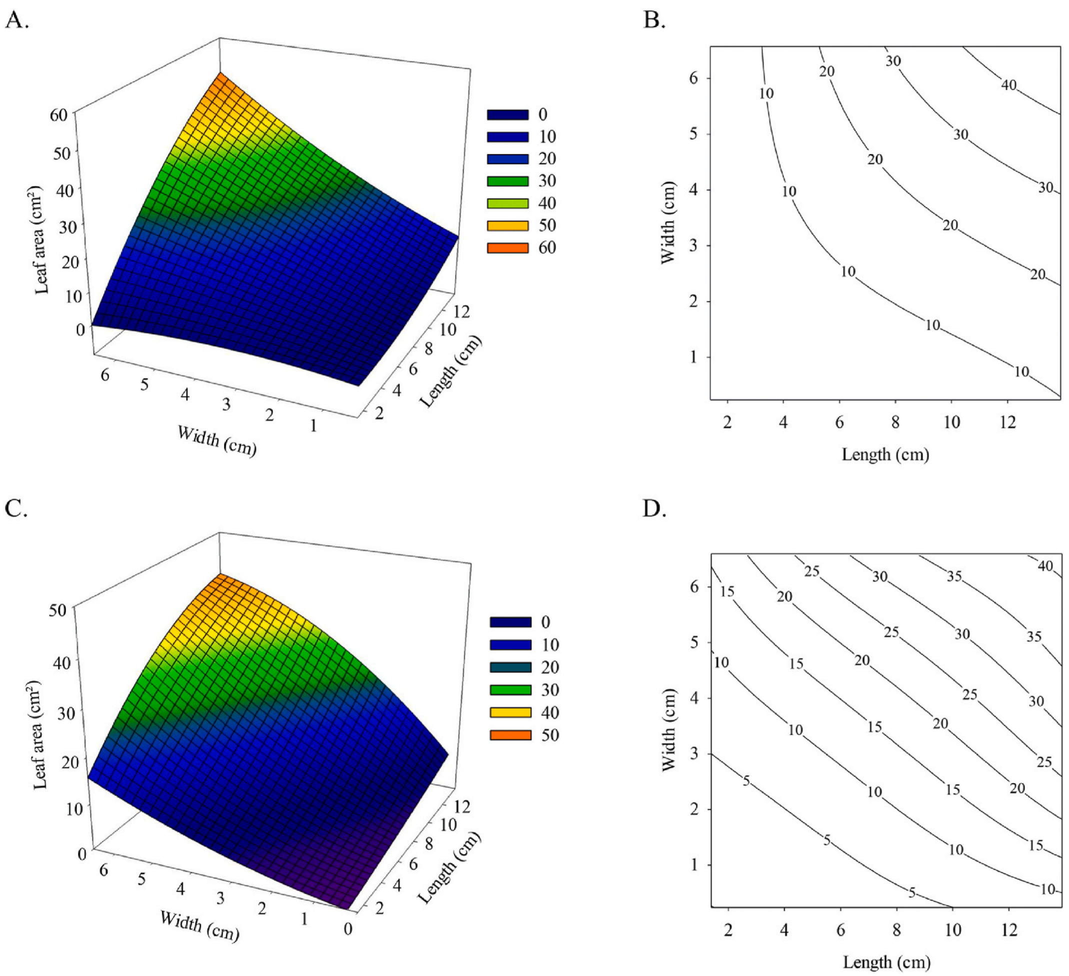

Fig. 10. Plots of the surface response of actual (A and B) and RNA estimated leaf area (C and D) versus length and width of sesame leaves (data from 11,000 leaves).

interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e17834.

References

[1] A.C. White, A. Rogers, M. Rees, C.P. Osborne, How can we make plants grow faster? A source–sink perspective on growth rate, J. Exp. Bot. 67 (1) (2016) 31–45.
[2] W. Huang, D.A. Ratkowsky, C. Hui, P. Wang, J. Su, P. Shi, Leaf fresh weight versus dry weight: which is better for describing the scaling relationship between leaf biomass and leaf area for broad-leaved plants? Forests 10 (3) (2019) 256.
[3] H. Fang, F. Baret, S. Plummer, G. Schaepman-Strub, An overview of global leaf area index (LAI): methods, products, validation, and applications, Rev. Geophys. 57 (3) (2019) 739–799.
[4] I. Keramatlou, M. Sharifani, H. Sabouri, M. Alizadeh, B. Kamkar, A simple linear model for leaf area estimation in Persian walnut (Juglans regia L.), Sci. Hortic. 184 (2015) 36–39.
[5] S. Díaz, J. Kattge, J.H.C. Cornelissen, I.J. Wright, S. Lavorel, S. Dray, et al., The global spectrum of plant form and function, Nature 529 (7585) (2016) 167–171.
[6] S. Munitz, A. Schwartz, Y. Netzer, Water consumption, crop coefficient and leaf area relations of a Vitis vinifera cv.'Cabernet Sauvignon'vineyard, Agric. Water Manag. 219 (2019) 86–94.
[7] JEdS. Ribeiro, EdS. Coêlho, F.R.A. Figueiredo, W.E. Pereira, MBd Albuquerque, M.F. Melo, Morphophysiological aspects of Erythroxylum pauferrense Plowman seedlings submitted to shading, Ciência Florest. 32 (2022) 395–416.
[8] W. Zhang, Digital image processing method for estimating leaf length and width tested using kiwifruit leaves (Actinidia chinensis Planch), PLoS One 15 (7) (2020), e0235499.
[9] I.A. Hernández-Fernandéz, A. Jarma-Orozco, M.F. Pompelli, Allometric models for non-destructive leaf area measurement of stevia: an in depth and complete analysis, Hortic. Bras. 39 (2021) 205–215.
[10] JEdS. Ribeiro, EdS. Coêlho, F.R.A. Figueiredo, SdF. Lopes, MBd Albuquerque, Estimation of leaf area of Erythroxylum citrifolium from linear leaf dimensions, Biosci. J. (2019) 1923–1931.
[11] C.S. Mota, H.G. Leite, M.A.O. Cano, Equações para estimar área foliar de folíolos de Acrocomia aculeta, Pesquisa Florestal Brasileira 34 (79) (2014) 217–224.
[12] J.N.B. Santos, A. Jarma-Orozco, W.C. Antunes, K.R. Mendes, J.M. Figueiroa, L.M. Pessoa, et al., New approaches to predict leaf area in woody tree species from the Atlantic Rainforest, Brazil, Austral Ecol. 46 (4) (2021) 613–626.
[13] M.K. Kumar, R.S. Kumar, V. Sankar, T. Sakthivel, G. Karunakaran, et al., Non-destructive estimation of leaf area of durian (Durio zibethinus)–An artificial neural network approach, Sci. Hortic. 219 (2017) 319–325.
[14] A. Ghadami Firouzabadi, M. Raeini-Sarjaz, A. Shahnazari, H. Zareabyaneh, Non-destructive estimation of sunflower leaf area and leaf area index under different water regime managements, Arch. Agron Soil Sci. 61 (10) (2015) 1357–1367.
[15] A. Cargnelutti Filho, M. Toebe, B.M. Alves, C. Burin, J.A. Kleinpaul, Estimação da área foliar de canola por dimensões foliares, Bragantia 74 (2015) 139–148.
[16] FdL. Tartaglia, E.Z. Righi, Ld Rocha, L.H. Loose, I.C. Maldaner, A.B. Heldwein, Non-destructive models for leaf area determination in canola, Rev. Bras. Eng. Agrícola Ambient. 20 (2016) 551–556.
[17] P.C.H. Goergen, I. Lago, N.T. Schwab, A.F. Alves, CPdO. Freitas, V.S. Selli, Allometric relationship and leaf area modeling estimation on chia by non-destructive method, Rev. Bras. Eng. Agrícola Ambient. 25 (2021) 305–311.
[18] J.C.S. Salazar, L.M. Melgarejo, E.H.D. Bautista, J.A. Di Rienzo, F. Casanoves, Non-destructive estimation of the leaf weight and leaf area in cacao (Theobroma cacao L.), Sci. Hortic. 229 (2018) 19–24.
[19] L. Bouvié, A.C. da Silva, D.R. Borella, C.C. da Silva, M. Pizzatto, Área do limbo foliar da castanheira-do-Brasil por medidas lineares, Nativa 8 (3) (2020) 380–388.
[20] A.P.S. Macário, RLdS. Ferraz, PdS. Costa, JFd Brito Neto, ASd Melo, J. Dantas Neto, Allometric models for estimating Moringa oleifera leaflets area, Cienc. E Agrotecnol 44 (2020).
[21] JEdS. Ribeiro, EdS. Coêlho, PHdA. Oliveira, WdAR. Lopes, EFd Silva, AKSd Oliveira, et al., Allometric models to estimate peanuts leaflets area by non-destructive method, Bragantia (2022) 81.
[22] JEdS. Ribeiro, J.S. Nóbrega, EdS. Coêlho, T.J. Dias, M.F. Melo, Estimating leaf area of basil cultivars through linear dimensions of leaves, Rev. Ceres 69 (2022) 139–147.
[23] S.K. Sarker, N. Das, M.Q. Chowdhury, M.M. Haque, Developing allometric equations for estimating leaf area and leaf biomass of Artocarpus chaplasha in Raghunandan Hill Reserve, Bangladesh, South. For. a J. For. Sci. 75 (1) (2013) 51–57.
[24] J.C.C. dos Santos, R.N. Costa, D.M.R. Silva, A.A. de Souza, F. de Barros Prado Moura, J.M. da Silva Junior, et al., Use of allometric models to estimate leaf area in Hymenaea courbaril L, Theoretical and Experimental Plant Physiology 28 (2016) 357–369.
[25] M. Toebe, F.J. Soldateli, RRd Souza, A.C. Mello, A. Segatto, Leaf area estimation of Burley tobacco, Ciência Rural. (2020) 51.
[26] M. Abdipour, S.H.R. Ramazani, M. Younessi-Hmazekhanlu, M. Niazian, Modeling oil content of sesame (Sesamum indicum L.) using artificial neural network and multiple linear regression approaches, J. Am. Oil Chem. Soc. 95 (3) (2018) 283–297.
[27] G.N. Silva, R.S. Tomaz, IdC. Sant'Anna, M. Nascimento, L.L. Bhering, C.D. Cruz, Neural networks for predicting breeding values and genetic gains, Sci. Agric. 71 (2014) 494–498.
[28] I. Pantic, J. Paunovic, J. Cumic, S. Valjarevic, G.A. Petroianu, P.R. Corridon, Artificial neural networks in contemporary toxicology research, Chem. Biol. Interact. (2022), 110269.
[29] A. Lavecchia, Deep learning in drug discovery: opportunities, challenges and future prospects, Drug Discov. Today 24 (10) (2019) 2017–2032.
[30] M.R.N. Rad, H.R. Fanaei, M.R.P. Rad, Application of Artificial Neural Networks to predict the final fruit weight and random forest to select important variables in native population of melon (Cucumis melo L.), Sci. Hortic. 181 (2015) 108–112.
[31] A. Taheri-Garavand, A. Nasiri, D. Fanourakis, S. Fatahi, M. Omid, N. Nikoloudakis, Automated in situ seed variety identification via deep learning: a case study in chickpea, Plants 10 (7) (2021) 1406.
[32] A. Nasiri, A. Taheri-Garavand, D. Fanourakis, Y.-D. Zhang, N. Nikoloudakis, Automated grapevine cultivar identification via leaf imaging and deep convolutional neural networks: a proof-of-concept study employing primary iranian varieties, Plants 10 (8) (2021) 1628.
[33] P. Boniecki, K. Koszela, K. Świerczyński, J. Skwarcz, M. Zaborowicz, J. Przybył, Neural visual detection of grain weevil (Sitophilus granarius L.), Agriculture 10 (1) (2020) 25.
[34] G.K. Aji, K. Hatou, T. Morimoto, Modeling the dynamic response of plant growth to root zone temperature in hydroponic chili pepper plant using neural networks, Agriculture 10 (6) (2020) 234.
[35] H. García-Martínez, H. Flores-Magdaleno, R. Ascencio-Hernández, A. Khalil-Gardezi, L. Tijerina-Chávez, O.R. Mancilla-Villa, et al., Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and RGB images acquired with unmanned aerial vehicles, Agriculture 10 (7) (2020) 277.

[36] M. Zhu, S. Liu, Z. Xia, G. Wang, Y. Hu, Z. Liu, Crop growth stage GPP-driven spectral model for evaluation of cultivated land quality using GA-BPNN, Agriculture 10 (8) (2020) 318.
[37] E.R. Abraham, J.G. Mendes dos Reis, O. Vendrametto, PLd Oliveira Costa Neto, R. Carlo Toloi, AEd Souza, et al., Time series prediction with artificial neural networks: an analysis using Brazilian soybean production, Agriculture 10 (10) (2020) 475.
[38] G. Vlontzos, P.M. Pardalos, Assess and prognosticate green house gas emissions from agricultural production of EU countries, by implementing, DEA Window analysis and artificial neural networks, Renew. Sustain. Energy Rev. 76 (2017) 155–162.
[39] L. Rossi, M. Bagheri, W. Zhang, Z. Chen, J.G. Burken, X. Ma, Using artificial neural network to investigate physiological changes and cerium oxide nanoparticles and cadmium uptake by Brassica napus plants, Environ. Pollut. 246 (2019) 381–389.
[40] S. Emamgholizadeh, M. Parsaeian, M. Baradaran, Seed yield prediction of sesame using artificial neural network, Eur. J. Agron. 68 (2015) 89–96.
[41] A.M. Azevedo, V.C. Andrade, A.S. Sousa, A.A. Santos, C.D. Cruz, S.L. Pereira, et al., Eficiência da estimação da área foliar de couve por meio de redes neurais artificiais, Hortic. Bras. 35 (2017) 14–19.
[42] A. Azeem, Q. Javed, J. Sun, D. Du, Artificial neural networking to estimate the leaf area for invasive plant Wedelia trilobata, Nord. J. Bot. 38 (6) (2020).
[43] H. Küçükönder, S. Boyaci, A. Akyüz, A modeling study with an artificial neural network: developing estimationmodels for the tomato plant leaf area, Turk. J. Agric. For. 40 (2) (2016) 203–212.
[44] M.S. Odabas, E. Ergun, F. Oner, Artificial neural network approach for the predicition of the corn (Zea mays L.) leaf area, Bulgarian Journal of Agricultural Science 19 (4) (2013) 766–769.
[45] I.A. Samborska, V. Alexandrov, L. Sieczko, B. Kornatowska, V. Goltsev, M.D. Cetner, et al., Artificial neural networks and their application in biological and agricultural research, J NanoPhotoBioSciences 2 (2014) 14–30.
[46] N. Pathak, A.K. Rai, R. Kumari, K.V. Bhat, Value addition in sesame: a perspective on bioactive components for enhancing utility and profitability, Phcog. Rev. 8 (16) (2014) 147.
[47] M. Andargie, M. Vinas, A. Rathgeb, E. Möller, P. Karlovsky, Lignans of sesame (Sesamum indicum L.): a comprehensive review, Molecules 26 (4) (2021) 883.
[48] RTd Silva, ABd Oliveira, MdFdQ. Lopes, MdA. Guimarães, A.S. Dutra, Physiological quality of sesame seeds produced from plants subjected to water stress, Rev. Cienc. Agron. 47 (2016) 643–648.
[49] L. Elhanafi, M. Houhou, C. Rais, I. Mansouri, L. Elghadraoui, H. Greche, Impact of excessive nitrogen fertilization on the biochemical quality, phenolic compounds, and antioxidant power of Sesamum indicum L seeds, J. Food Qual. 2019 (2019).
[50] S.S. Dhaliwal, V. Sharma, A.K. Shukla, V. Verma, S.K. Behera, P.S. Sandhu, et al., Assessment of agroeconomic indicators of Sesamum indicum L. as influenced by application of boron at different levels and plant growth stages, Molecules 26 (21) (2021) 6699.
[51] K. Vanisha, A.K. Atwal, S.S. Dhaliwal, S.K. Banga, Assessment of diverse sesame (Sesamum indicum L.) germplasm for mineral composition, J. Plant Sci. Res. 29 (1) (2013) 29.
[52] E. Yol, R. Toker, M. Golukcu, B. Uzun, Oil content and fatty acid characteristics in Mediterranean sesame core collection, Crop Sci. 55 (5) (2015) 2177–2185.
[53] D. Fanourakis, F. Kazakos, P.A. Nektarios, Allometric individual leaf area estimation in chrysanthemum, Agronomy 11 (4) (2021) 795.
[54] C.A. Alvares, J.L. Stape, P.C. Sentelhas, JLdM. Gonçalves, G. Sparovek, Köppen's climate classification map for Brazil, Meteorol. Z. 22 (6) (2013) 711–728.
[55] Embrapa, Sistema brasileiro de classificação de solos, vol. 3, Centro Nacional de Pesquisa de Solos, Rio de Janeiro, 2013.
[56] JEdS. Ribeiro, A.J.S. Barbosa, MBd Albuquerque, Leaf area estimate of Erythroxylum simonis Plowman by linear dimensions, Floresta e Ambiente 25 (2018).
[57] R.C.R. Team, The R Project for Statistical Computing, 2023.
[58] F.E. Ahmed, Artificial neural networks for diagnosis and survival prediction in colon cancer, Mol. Cancer 4 (1) (2005) 1–12.
[59] P.H.M. Janssen, P.S.C. Heuberger, Calibration of process-oriented models, Ecol. Model. 83 (1–2) (1995) 55–66.
[60] D.W. Marquaridt, Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation, Technometrics 12 (3) (1970) 591–612.
[61] J.L. Gill, Outliers, residuals, and influence in multiple regression, J. Anim. Breed. Genet. 103 (1–5) (1986) 161–175.
[62] G.A. Dalmago, C.A.M. Bianchi, S. Kovaleski, E. Fochesatto, Evaluation of mathematical equations for estimating leaf area in rapeseed, Rev. Cienc. Agron. 50 (2019) 420–430.
[63] A. Cargnelutti Filho, R.V. Pezzini, I.M.M. Neu, G.E. Dumke, Estimation of buckwheat leaf area by leaf dimensions, Semina Ci agr (2021) 1529–1548.
[64] A. Taner, Y.B. Öztekin, A. Tekgüler, H. Sauk, H. Duran, Classification of varieties of grain species by artificial neural networks, Agronomy 8 (7) (2018) 123.
[65] Z. Wang, L. Zhang, Leaf shape alters the coefficients of leaf area estimation models for Saussurea stoliczkai in central Tibet, Photosynthetica 50 (2012) 337–342.
[66] G. Koubouris, D. Bouranis, E. Vogiatzis, A.R. Nejad, H. Giday, G. Tsaniklidis, et al., Leaf area estimation by considering leaf dimensions in olive tree, Sci. Hortic. 240 (2018) 440–445.
[67] M.F. Pompelli, W.C. Antunes, D.T.R.G. Ferreira, P.G.S. Cavalcante, H.C.L. Wanderley-Filho, L. Endres, Allometric models for non-destructive leaf area estimation of Jatropha curcas, Biomass Bioenergy 36 (2012) 77–85.
[68] G.L. Richter, A. Zanon Júnior, N.A. Streck, J.V.C. Guedes, B. Kräulich, TSMd Rocha, et al., Estimativa da área de folhas de cultivares antigas e modernas de soja por método não destrutivo, Bragantia 73 (2014) 416–425.
[69] J.O. Carvalho, M. Toebe, F.L. Tartaglia, C.T. Bandeira, A.L. Tambara, Leaf area estimation from linear measurements in different ages of Crotalaria juncea plants, An. Acad. Bras. Cienc. 89 (2017) 1851–1868.
[70] A. Shabani, A.R. Sepaskhah, Leaf area estimation by a simple and non-destructive method, Iran Agric. Res. 36 (2) (2017) 101–105.
[71] M.A. Vazquez-Cruz, R. Luna-Rubio, L.M. Contreras-Medina, I. Torres-Pacheco, R.G. Guevara-Gonzalez, Estimating the response of tomato (Solanum lycopersicum) leaf area to changes in climate and salicylic acid applications by means of artificial neural networks, Biosyst. Eng. 112 (4) (2012) 319–327.
[72] O.M. Adisa, Botai Jo, A.M. Adeola, A. Hassen, C.M. Botai, D. Darkey, et al., Application of artificial neural network for predicting maize production in South Africa, Sustainability 11 (4) (2019) 1145.
[73] A. Taheri-Garavand, H. Mumivand, D. Fanourakis, S. Fatahi, S. Taghipour, An artificial neural network approach for non-invasive estimation of essential oil content and composition through considering drying processing factors: a case study in Mentha aquatica, Ind. Crop. Prod. 171 (2021), 113985.