

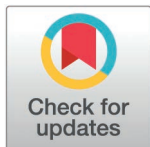
RESEARCH ARTICLE

Artificial intelligence-based modeling for accurate leaf area estimation in olive (*Olea europaea* L.) cultivars

Yazgan Tunç¹, Fatih Demirel^{2*}, Ali Khadivi^{3*}, Kadir Uğurtan Yılmaz⁴, Hüsni Demirsoy⁵, Bilal Cemek⁶, Hatice Gözel⁷, Daya Shankar Mishra⁸

1 Republic of Türkiye, Ministry of Agriculture and Forestry, General Directorate of Agricultural Research and Policies, Hatay Olive Research Institute Directorate, Hassa, Hatay, Türkiye, **2** Department of Agricultural Biotechnology, Faculty of Agriculture, Iğdır University, Iğdır, Türkiye, **3** Department of Horticultural Sciences, Faculty of Agriculture and Natural Resources, Arak University, Arak, Iran, **4** Department of Horticulture, Faculty of Agriculture, Kahramanmaraş Sütçü İmam University, Onikisubat, Kahramanmaraş, Türkiye, **5** Department of Horticulture, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Türkiye, **6** Department of Agricultural Structures and Irrigation, Faculty of Agriculture, Ondokuz Mayıs University, Atakum, Samsun, Türkiye, **7** Department of Horticulture, Faculty of Agriculture, Kilis 7 Aralık University, Kilis, Türkiye, **8** Department of Fruit Sciences, ICAR-CIAH RS Central Horticultural Experiment Station, Godhra, Gujarat, India

* fatih.demirel@igdir.edu.tr (FD); a-khadivi@araku.ac.ir (AK)



OPEN ACCESS

Citation: Tunç Y, Demirel F, Khadivi A, Yılmaz KU, Demirsoy H, Cemek B, et al. (2026) Artificial intelligence-based modeling for accurate leaf area estimation in olive (*Olea europaea* L.) cultivars. PLoS One 21(1): e0339865. <https://doi.org/10.1371/journal.pone.0339865>

Editor: Marzia Vergine, University of Salento Department of Biological and Environmental Sciences and Technologies: Università del Salento Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, ITALY

Received: October 14, 2025

Accepted: December 14, 2025

Published: January 2, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0339865>

Abstract

Estimating olive (*Olea europaea* L.) leaf area is an important aspect of monitoring plant health and evaluating growth processes in agriculture. Accurate estimation of leaf area allows for a better understanding of processes such as water and nutrient utilization, photosynthesis efficiency, respiration, and yield potential. This study aims to determine the most accurate, easy, and reliable leaf area estimation model using the geometric properties (length and width) of olive leaves. Additionally, the predictive performances of multiple linear regression (MLR) and artificial neural network (ANN) were compared. A total of 1320 leaf samples collected from 22 olive cultivars were used in the study. Leaf length and width were taken as input parameters, and both MLR and ANN models were developed for each cultivar. Both multiple linear regression (MLR) and artificial neural network (ANN) models demonstrated high predictive accuracy for olive leaf area estimation across 22 cultivars. The MLR models explained up to 96% of the variation in leaf area using leaf length (LL) and leaf width (LW), with low root mean square errors, indicating strong reliability. When cultivar identity was modeled as a categorical factor through dummy encoding, the model captured significant cultivar-specific effects without altering the overall predictive performance. The ANN models achieved slightly higher accuracy, with determination coefficients exceeding 0.99 and minimal prediction errors, confirming their superior ability to model nonlinear relationships. Across both approaches, leaf width contributed more strongly to leaf area than leaf length. Cultivar-specific differences were statistically significant for only a few genotypes, while most cultivars exhibited

Copyright: © 2026 Tunç et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the paper and its [Supporting information](#) files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

comparable patterns after adjustment for multiple testing. In conclusion, both MLR and ANN models demonstrated high accuracy in predicting olive leaf area, with ANN models showing slightly superior performance. However, MLR models also yielded highly reliable results, indicating that both approaches are viable for practical applications in olive cultivation. These predictive models can be effectively used for rapid, non-destructive phenotyping, growth monitoring, and precision management in olive breeding and production systems.

1. Introduction

The olive tree (*Olea europaea* L.) is one of the most widely cultivated fruit trees worldwide, particularly well-suited to the Mediterranean climate [1]. Various olive cultivars are grown depending on their intended use (edible olives and/or olive oil) and local microclimatic conditions [2,3].

With a cultivation history that spans over three thousand years, the olive tree is considered to be one of the most important and ancient fruit crops in the world [4]. It is especially well-suited to the Mediterranean environment, which is characterized by hot, dry summers and warm, rainy winters, which are favorable for its development. Olive is a significant crop in countries such as Spain, Italy, Greece, Türkiye, and Tunisia, as over ninety-five percent of the world's olive production is located in the Mediterranean basin [5]. Olive growth has spread beyond the Mediterranean area to other regions of the globe, including North and South America, Australia, and China, as a result of the economic and ecological advantages that olive farming offers [6]. An olive tree's capacity to flourish in dry and semi-arid settings, displaying a high tolerance to water shortages and poor soils, is one of the reasons why olive trees are so highly appreciated. Because of their extensive root systems, they can draw water from the deeper layers of the soil, which makes them an essential species for sustainable agriculture in regions that are prone to drought [7].

Estimation of leaf area (LA) in plants is considered a critical parameter in many scientific fields such as plant growth, agricultural productivity, ecosystem dynamics, and climate change studies [8–10]. Leaf area is a fundamental growth indicator that directly affects the photosynthetic capacity, transpiration rates, and biomass of the plant [11]. Leaf area is directly related to photosynthetic efficiency and growth rates. Leaf area width determines biomass increase by affecting the light capture capacity and carbon assimilation rates of the plant [12]. Plant parameters such as specific leaf area (SLA) and leaf mass area (LMA) provide information about water use, drought tolerance, and growth rate of the plant [13]. Leaf area is an important indicator for yield estimation and plant health monitoring in agricultural production. Leaf area width can lead to higher crop production by increasing photosynthetic efficiency. Furthermore, accurate calculation of leaf area index (LAI) in plants allows optimization of irrigation and fertilization strategies [14]. Plant water use and evaporation-transpiration processes depend on leaf area. Plants with large leaf areas may lose more water, while species with small leaf areas may be more tolerant of drought conditions. Thus,

leaf area measurements are used in planning agricultural irrigation programs and improving water use efficiency [15]. On the other hand, a decrease in leaf area may be a sign of disease, pest infestation, nutrient deficiency, or environmental stresses. Monitoring the change in leaf area over time is an important tool for the early detection of plant diseases and the development of intervention strategies [8]. In ecosystem studies, the leaf area index (LAI) is used to determine how much carbon plants can sequester from the atmosphere and how much biomass they can create through photosynthesis. In forest ecosystems, LAI calculations are a critical variable for modeling the carbon cycle and the effects of climate change [16,17]. Leaf area estimation is a vital variable in many areas such as plant growth, agricultural productivity, water management, plant health, ecosystem dynamics, and precision agriculture practices. Scientific studies show that accurate calculation of leaf area index improves agricultural management, contributes to climate change models, and is critical to understanding ecological processes [18].

In recent years, innovative technologies such as hand-held scanners and laser optical devices have become widely used in leaf area (LA) measurements. Although these advanced technologies offer high accuracy, they are generally preferred for more specialized research rather than basic experiments due to their complex structures and high costs [19]. In contrast, indirect methods offer significant advantages such as lower cost, speed, practicality, and less labor requirement compared with direct measurements. In addition, they are widely used in agricultural research and ecophysiological studies, enabling precise estimations without damaging the plant [20].

Indirect methods are quite valuable for investigating the empirical relationships between leaf area (LA) estimation and leaf dimensions. In particular, lamina dimensions such as leaf length (L) and leaf width (W) are widely used variables in estimating LA [21,22]. Various studies have shown that linear leaf measurements correlate strongly with LA. In this direction, studies have been conducted on leaf area estimation in many fruit trees, such as apple [23–25], pear [26], peach [27], apricot [28], grape [29], strawberry [30], cherry [31], citrus [32], chestnut [33], hazelnut [34], kiwi [35], black walnut [36], walnut [20], and mango [37].

Nowadays, there is a growing trend of utilizing machine learning (ML) techniques to address intricate issues. ML has been acquiring popularity at a rapid pace as a result of its capacity to resolve challenges in a variety of scientific research disciplines [38–41]. The importance of ML is derived from its capacity to analyze intricate datasets, reveal concealed patterns, and produce exact predictions. ML algorithms are a potent instrument in a variety of fields due to their superior predictive accuracy in comparison to traditional methods. Machine learning has become a fundamental component of contemporary problem-solving due to its extensive adaptability [42,43].

In recent years, ML's impact on agronomic applications has gained increasing attention. ML techniques have been successfully applied in various areas, including prediction of crop water content [44], estimation of crop nutrient concentration [45], modeling of crop yield [46], crop evapotranspiration estimation [46], and determination of soil moisture content [47]. To date, several ML-based, non-destructive models have been developed to estimate leaf area (LA) for different crop types. For example, artificial neural networks (ANNs) have proven to be successful in predicting LA for pepper [48], pear [26], tomato [49], durian [21], paprika [50], sugar beet [51], maize [52], cotton [53], and pepper [54].

Similarly, the Adaptive Neuro-Fuzzy Inference System (ANFIS) has also proven to be an effective method for LA estimation [55–57]. A comparative study by Sabouri et al. [58] demonstrated that ANN and ANFIS models outperformed traditional methods, such as multiple linear regression (MLR) and support vector regression (SVR), in LA estimation for plum genotypes. Specifically, the ANN and ANFIS models exhibited higher accuracy rates in leaf area estimation compared with conventional techniques.

This study aims to develop and compare LA estimation models for 22 widely cultivated olive cultivars in Türkiye using MLR and ANN approaches. Equations based on leaf length (LL), leaf width (LW), and the number of cultivars (CN) were formulated for both methods. Ultimately, this study aims to determine the most suitable approach for olive leaf area estimation and enhance its applicability in agricultural practices.

2. Materials and methods

2.1. Plant material and dataset

This study was conducted to develop and validate a leaf area estimation model for 22 olive cultivars grown in Türkiye (Republic of Türkiye, Ministry of Agriculture and Forestry, General Directorate of Agricultural Research and Policies, Hatay Olive Research Institute, Hassa Station) (Fig 1). The map was generated using ArcGIS software version 10.1 [59]. A total of 60 leaf samples were collected from each cultivar, and individual descriptive statistical characteristics are provided in S1 Table. All leaf samples were collected in January 2024. The leaves were collected from olive trees that were eight to ten years old. The dataset was split into training (70%) and testing (30%) subsets to ensure effective model development [60]. Table 1 presents the descriptive statistics for all olive cultivars, as well as the training and testing sets.

Statement specifying permissions. For this study, we acquired permission to study olive issued by the Agricultural and Forestry Ministry of the Republic of Türkiye.

Statement on experimental research and field studies on plants. The either cultivated or wild-growing plants sampled comply with relevant institutional, national, and international guidelines and domestic legislation of Türkiye.

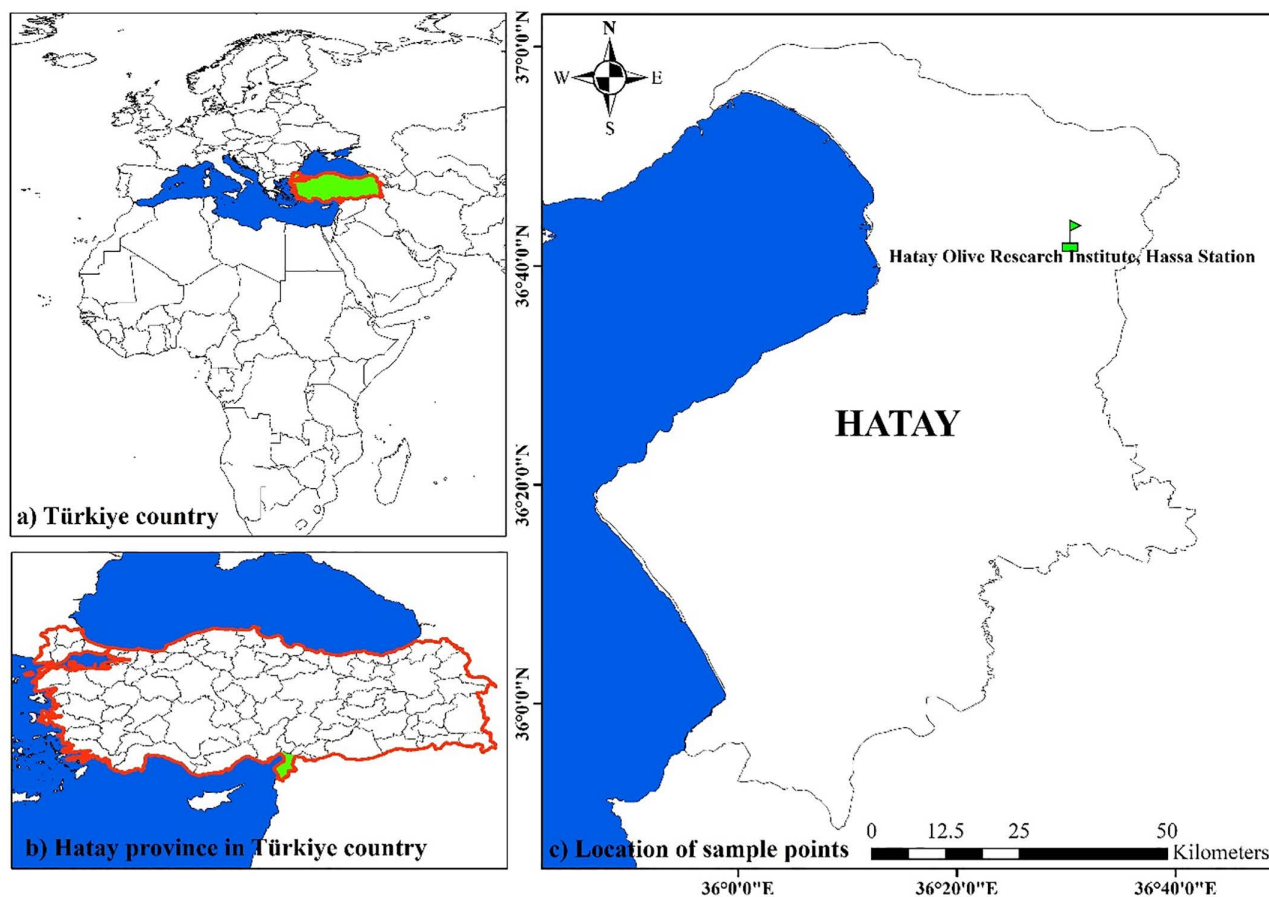


Fig 1. Geographic location of the studied olive cultivars. The above map image is original and was generated by the third author, Ali Khadivi, using ArcGIS software version 10.1 (<http://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f>).

<https://doi.org/10.1371/journal.pone.0339865.g001>

Table 1. Descriptive statistics for leaves of olive cultivars used in this study.

	Training (n=880)			Testing (n=440)			All (n=1320)		
	LL (mm)	LW (mm)	LA (cm ²)	LL (mm)	LW (mm)	LA (cm ²)	LL (mm)	LW (mm)	LA (cm ²)
Min.	12.02	4.53	0.56	12.52	5.31	0.56	12.02	4.53	0.56
Max.	93.81	29.78	14.58	68.94	27.37	9.12	93.81	29.78	14.58
Avg.	49.40	13.72	5.30	34.29	10.18	2.72	44.83	12.65	4.52
SE	0.53	0.13	0.09	0.54	0.15	0.07	0.45	0.11	0.07
SD	16.07	4.09	2.75	10.78	3.08	1.44	16.23	4.15	2.70
Cv	0.33	0.30	0.52	0.31	0.30	0.53	0.36	0.33	0.60
Skewness	-0.49	0.54	-0.17	-0.36	4.09	2.75	-0.30	0.64	0.29
Kurtosis	0.27	0.59	0.63	0.25	1.43	1.34	0.45	0.75	0.91

Min minimum, Max maximum, Avg average, SE standard error, SD standard deviation, Cv coefficient of variation, LA leaf area, LW leaf width, LL leaf length.

<https://doi.org/10.1371/journal.pone.0339865.t001>

2.2. Leaf measurements

Each leaf sample was placed on an A4 sheet and photocopied at a 1:1 scale to preserve the actual size. The actual leaf area (LA) was determined using a Placom digital planimeter (Sokkisha Planimeter Inc., Model KP-90) from the photocopied images. Additionally, leaf width (LW) and leaf length (LL) were measured for model construction. Leaf width (mm) was recorded at the widest part of the lamina from edge to edge, while leaf length (mm) was measured from the tip of the lamina to the petiole junction along the midrib. All measurements were taken to the nearest 0.1 cm for accuracy.

2.3. K-Fold cross-validation

All input and output variables were normalized to the range of 0–1 before training to ensure model convergence (see [S1 Note](#) for normalization equation [Eq. 1](#)).

K-fold cross-validation is a widely used technique for model validation, where the dataset is divided into K subsets (folds) to ensure robust evaluation across different partitions. In this study, 5-fold cross-validation was applied, with the dataset split into training (70%) and testing (30%) subsets ([Fig 2](#)). This approach helps prevent overfitting and improves the generalization of the model [\[61\]](#).

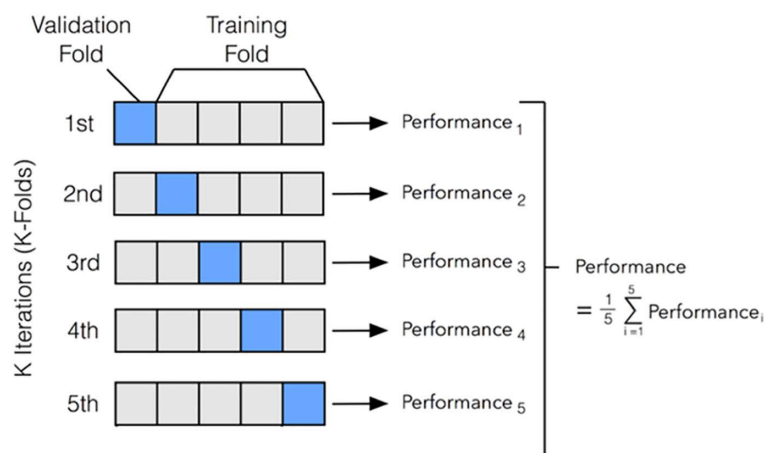


Fig 2. K-fold cross-validation [\[61\]](#).

<https://doi.org/10.1371/journal.pone.0339865.g002>

2.4. Multiple linear regression (MLR)

Multiple linear regression (MLR) was employed to develop an analytical model for leaf area estimation using observational data from the 22 olive cultivars. Microsoft Office Excel 2015 (Microsoft Corporation, Redmond, WA, ABD) (<https://www.microsoft.com>) was used for regression analysis, incorporating different combinations of independent variables, primarily leaf length (LL) and leaf width (LW). The regression model was refined iteratively until the sum of squared deviations was minimized.

MLR is one of the most commonly used methods for leaf area estimation. A stepwise regression approach was utilized to determine the best-fit model. The independent variables were selected sequentially using backward stepwise, linear, and quadratic methods. The general form of the regression formula is given as [Eq. 1](#).

$$y = b_0 + b_1 X_1 + b_2 X_2 \quad (1)$$

Where Y represents the dependent variable (leaf area, LA), b_0 is the intercept, b_1 and b_2 are regression coefficients, and X_1 and X_2 correspond to leaf width (LW) and leaf length (LL), respectively.

When two or more independent variables in the model (LL and LW) exhibit a high degree of correlation, collinearity may occur. Collinearity introduces redundancy in the information about the response variable, potentially diminishing the model's accuracy. To assess collinearity, variance inflation factors (VIFs) were computed for each predictor variable [\[62\]](#). If the VIF values are less than 10, then there is no collinearity problem in the model. The VIF was calculated using [Eq. 2](#).

$$[VIF]_j = 1 / (1 - R_j^2) \quad (2)$$

Where represents the coefficient of determination for the regression model that includes all predictors except the j th predictor.

The cultivar number (CN) variable was defined as a categorical factor. To account for the fixed effects of cultivars, $K - 1$ dummy (one-hot) indicator variables were created using treatment contrast coding, with 'Gemlik' selected as the reference cultivar. Accordingly, the general linear model was expressed as:

$$LA = \beta_0 + \beta_1 LL + \beta_2 LW + \sum (\gamma_g \times D_g) \quad (3)$$

Where D_g represents the dummy variable corresponding to each cultivar (e.g., D_{AYV} for 'Ayvalık', D_{DDOM} for 'Domat', D_{MEM} for 'Memecik', etc.). By using this approach, the categorical nature of the cultivar factor was modeled correctly. The estimated dummy coefficients and their statistical significance levels for each cultivar are presented in [S2 Table](#). The Table lists the estimated coefficients (γ_g), standard errors, and p-values for each cultivar dummy variable in [Eq. 2](#). Positive coefficients indicate a higher mean leaf area than the reference cultivar ('Gemlik') at fixed LL and LW values, whereas negative coefficients indicate lower mean leaf area. This categorical treatment of cultivar identity using dummy (one-hot) encoding followed the procedure described by James et al. [\[63\]](#).

2.5. Multi-layer perceptron (MLP)

Multi-layer perceptron (MLP) is one of the most widely used artificial neural network (ANN) models. MLP consists of a network of interconnected units (neurons or nodes). These nodes operate as functions where the sum of incoming inputs is modified through a nonlinear transfer or activation function [\[64,65\]](#). Connections are weighted, and each node generates an output signal.

In this study, different learning algorithms of artificial neural networks (ANNs) were tested for leaf area estimation, including Levenberg-Marquardt (LM), scaled conjugate gradient (SCG), and resilient backpropagation (RP) [\[66\]](#). The

activation functions used in the model were Tansig (hyperbolic tangent sigmoid) between the input and hidden layers and Purelin (linear) between the hidden and output layers.

In Model 1, leaf width (LW) and leaf length (LL) were used as input variables, while leaf area (LA) was estimated as the output variable. In Model 2, LW, LL, and cultivar number (CN) were used as input variables, while LA was estimated as the output variable (Fig 3).

2.6. Model training and performance evaluation

The ANN models were trained using a backpropagation algorithm with a predefined number of epochs to ensure convergence. The performance of the models was evaluated using standard statistical metrics [67–69], including mean bias error (MBE), root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). These metrics were used to compare ANN models with MLR, determining which model provided the most accurate leaf area estimation. Analyses were performed using MATLAB R2010b (<http://www.mathworks.com/>) software (The MathWorks Inc., Natick, MA). The MATLAB codes are provided as [S1 Code](#).

To statistically compare the predictive performances of the models, a paired-sample t-test was performed between the observed, MLR-predicted, and ANN-predicted leaf area (LA) values for each olive cultivar as well as for the overall dataset (S3 Table). The analysis was used to test whether the prediction errors (residuals) of the ANN and MLR models differed significantly. For all cultivars, the test evaluated three paired groups: (i) observed vs. MLR, (ii) observed vs. ANN, and (iii) MLR vs. ANN. Statistical significance was determined at $p < 0.05$.

To further assess model generalization and potential overfitting, training loss convergence curves were recorded for both ANN architectures (2–3–1 and 3–4–1), as shown in S1 and S2 Figs. Each network was trained using the Levenberg–Marquardt optimization algorithm with a learning rate of 0.01 and an epoch range of 100–500. Early stopping was applied empirically at the iteration corresponding to the minimum root mean square error (RMSE), ensuring that the model did not overfit the training data.

2.7. ANN model development

The architectures and hyperparameter configurations of the ANN models (2–3–1 and 3–4–1) are summarized in S4 Table, including the number of hidden neurons, activation functions, learning algorithm, and epoch settings.

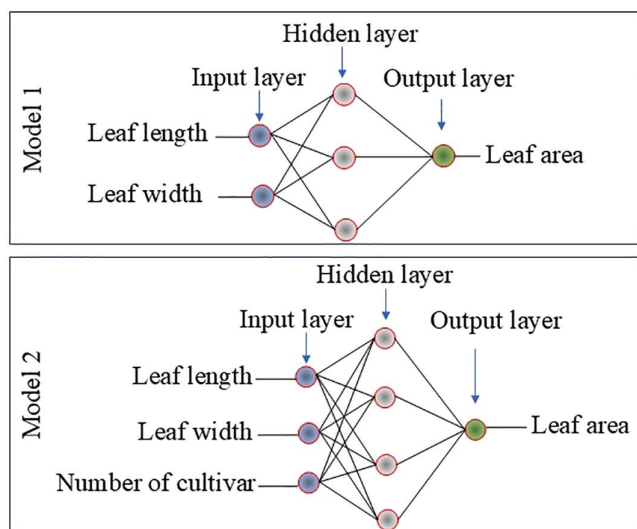


Fig 3. Architecture of the multi-layer perceptron applied to estimate leaf area.

<https://doi.org/10.1371/journal.pone.0339865.g003>

3. Results

Descriptive statistical data for leaf area (LA), leaf length (LL), and leaf width (LW) of the olive cultivars, along with the training and test datasets, are presented in [Table 1](#). In terms of leaf length (LL), the average values in the training dataset (49.40 mm) were higher than those in the test dataset (34.29 mm). This showed that the training dataset covered a wider range of leaf lengths. In terms of leaf width (LW), the mean values in the training dataset (13.72 mm) were higher than those in the test dataset (10.18 mm), indicating that the training dataset covered a wider range of leaf widths. For leaf area (LA), the average LA values in the test dataset (2.72 cm²) were lower than those in the training dataset (5.30 cm²). This indicated that smaller leaves were prevalent in the test dataset. The coefficient of variation (CV) value of leaf area was higher for the test dataset (0.53), while the leaf length value was higher for the training dataset (0.33). This indicated that leaf area exhibited more variability for the test dataset, while leaf length exhibited more variability for the training dataset. The coefficient of variation (0.30) value of leaf width was the same for the test and train datasets.

Leaf length (LL) was skewed to the left in both data sets (−0.49 for train and −0.36 for test), indicating slightly longer leaves than shorter ones. Leaf width (LW) had a highly skewed distribution in the test set (0.30), indicating some vast leaves compared with the rest. Leaf area (LA) was more symmetrical in the training set (−0.17) but skewed to the right in the test set, indicating that most leaves were small for the test set (2.75), with a few large ones, skewing the distribution to the right.

All LL, LW, and LA values were below 3, meaning that none of the distributions had strong peaks or extreme outliers. Leaf width (1.43) and leaf area (1.34) in the test set had higher kurtosis than in the train set (0.59 and 0.63, respectively), indicating more central clustering but still not extreme. Lower kurtosis values indicate a more even spread of the data rather than a high concentration near the mean.

3.1. Multiple linear regression (MLR) models

Before developing the LA model for different olive cultivars, collinearity between L and W was evaluated. Since the VIF values were less than 10, there was no collinearity between L and W. Therefore, these variables can be included in the LA models. The best MLR models for different olive cultivars are presented in [S5 Table](#). For each olive cultivar, leaf area (LA) was estimated using leaf length (LL) and leaf width (LW) as independent variables.

In the training set, R² values generally ranged between 0.91 and 0.99, indicating that the models explained 91–99% of the variability in the training data. The R² values in the test set were generally between 0.92 and 0.99, showing a performance close to that of the training set. This indicates that the models have a high generalization ability. The MAE values in the training set typically ranged between 0.07 and 0.48, while in the test set, they varied between 0.11 and 1.52, demonstrating that the models had minor errors. The RMSE values in the training set generally ranged between 0.10 and 0.58, whereas in the test set, RMSE values were between 0.14 and 0.90. This suggests that the models kept even significant errors at a reasonable level.

When evaluating the model performance metrics (R², MAE, RMSE), all models yielded highly successful results. The training and test sets showed overall consistency, demonstrating strong generalization capability in the models. From a total of 22 olive cultivars, the Saurani cultivar had the best-performing model in the train set (R²: 0.99, MAE: 0.08, RMSE: 0.12), while the Girit Zeytini cultivar had the top-performing models in the test set (R²: 0.99, MAE: 0.14, RMSE: 0.21).

The coefficients in the models were statistically significant, indicating that leaf length and leaf width were significant factors in increasing leaf area. LL coefficients varied between 0.077 and 0.138, while LW coefficients were between 0.207 and 0.478. LW coefficients were higher than LL coefficients, suggesting that leaf width contributed more to leaf area.

[Table 2](#) presents the general multiple linear regression (MLR) models developed for olive cultivars. The first model included only leaf length (LL) and leaf width (LW) as predictors, while the second model incorporated cultivar-specific dummy variables (D_i) representing each olive cultivar (e.g., D_Ayv for 'Ayvalik', D_Dom for 'Domat', D_Mem for

Table 2. Multiple linear regression general leaf area models were developed for olive cultivars.

Model	Train (n=880)				Test (n=440)			
	R2	MBE	MAE	RMSE	R2	MBE	MAE	RMSE
LA= -4.67+0.105LL+0.347LW	0.96	-0.01	0.41	0.56	0.96	0.14	0.40	0.50
LA= -4.58+0.10LL+0.35LW+ γ AyvDAyv+ γ DomD-Dom+ γ MemDMem+ + γ usluDuslu	0.96	0.13	0.40	0.58	0.96	0.21	0.39	0.49

LA leaf area (cm²), LW leaf width (cm), LL leaf length (cm), R2 coefficient of determination, MBE mean bias error, MAE mean absolute error, RMSE root mean square error.

<https://doi.org/10.1371/journal.pone.0339865.t002>

‘Memecik’, etc.), with corresponding coefficients (γ_i) estimating the specific effect of each cultivar on leaf area. This approach allowed the model to account for inter-cultivar variability more explicitly than a single numeric “cultivar number” (CN) parameter. Both models achieved identical R² values (0.96) for the training and test datasets, indicating strong predictive power and high generalization ability. Inclusion of dummy variables slightly adjusted the bias (MBE) and error (MAE, RMSE) metrics but did not markedly improve overall performance, suggesting that LL and LW remain the primary determinants of olive leaf area.

Fig 4 shows the linear regression between the predicted values and the actual values for Model 1 and Model 2. In Model 1, the coefficient of determination (R²) values were 0.9567 for the training dataset and 0.9613 for the test dataset. The regression equations show that the estimated values closely followed the measured values, indicating that LL and LW alone provided a strong predictive capability. When Model 2 included CN as an additional input, the R² values improved slightly to 0.9565 for training and 0.9623 for testing. Although the improvement was minimal, the inclusion of cultivar numbers helped refine the model’s predictive accuracy.

Overall, the MLR models demonstrated a strong capability to estimate leaf area based on leaf dimensions. The slight difference between the two models suggests that while cultivar number contributes to prediction accuracy, leaf length and leaf width remain the determinants of leaf area estimation.

This study also evaluated cultivar-specific effects on leaf area by re-specifying the cultivar number (CN) as a categorical factor through dummy (one-hot) encoding, using ‘Gemlik’ as the reference cultivar. In this revised model, cultivar-specific intercept differences in leaf area were estimated after controlling for leaf length (LL) and leaf width (LW). Following the Holm–Bonferroni correction, statistically significant positive coefficients were determined for ‘Domat’, ‘Edincik Su’, ‘Elmacık’, and ‘Sarı Haşebi’, indicating that these cultivars had higher mean leaf area than ‘Gemlik’ at fixed LL and LW values. Conversely, significantly negative coefficients were observed for ‘Çelebi’, ‘Nizip Yağlık’, and ‘Sarı Ulak’, indicating a lower mean leaf area compared with ‘Gemlik’. For the remaining cultivars, no statistically significant differences were detected once multiplicity was adjusted. These findings demonstrated that the cultivar effect on leaf area was primarily concentrated in a limited number of genotypes, whereas variations among other cultivars were not statistically supported. The estimated dummy coefficients and their significance levels are presented in S6 Table.

3.2. Artificial neural network (ANN) models

All models were analyzed starting with 100 iterations, increasing by 100 until reaching 500 iterations. The best results for all 22 models were obtained between 200 and 300 iterations. In all models, tansig was used as the activation function between the input and hidden layers, while purelin was used between the hidden and output layers. The training algorithm was Levenberg–Marquardt (LM). The Levenberg–Marquardt algorithm demonstrated stable convergence without overfitting, as shown by the RMSE reduction curves in S1 and S2 Figs.

For the general model created using the combined data of all cultivars based on length and width as inputs, the same activation function and training algorithm were used. The dataset was split into 70% for training and 30% for testing,

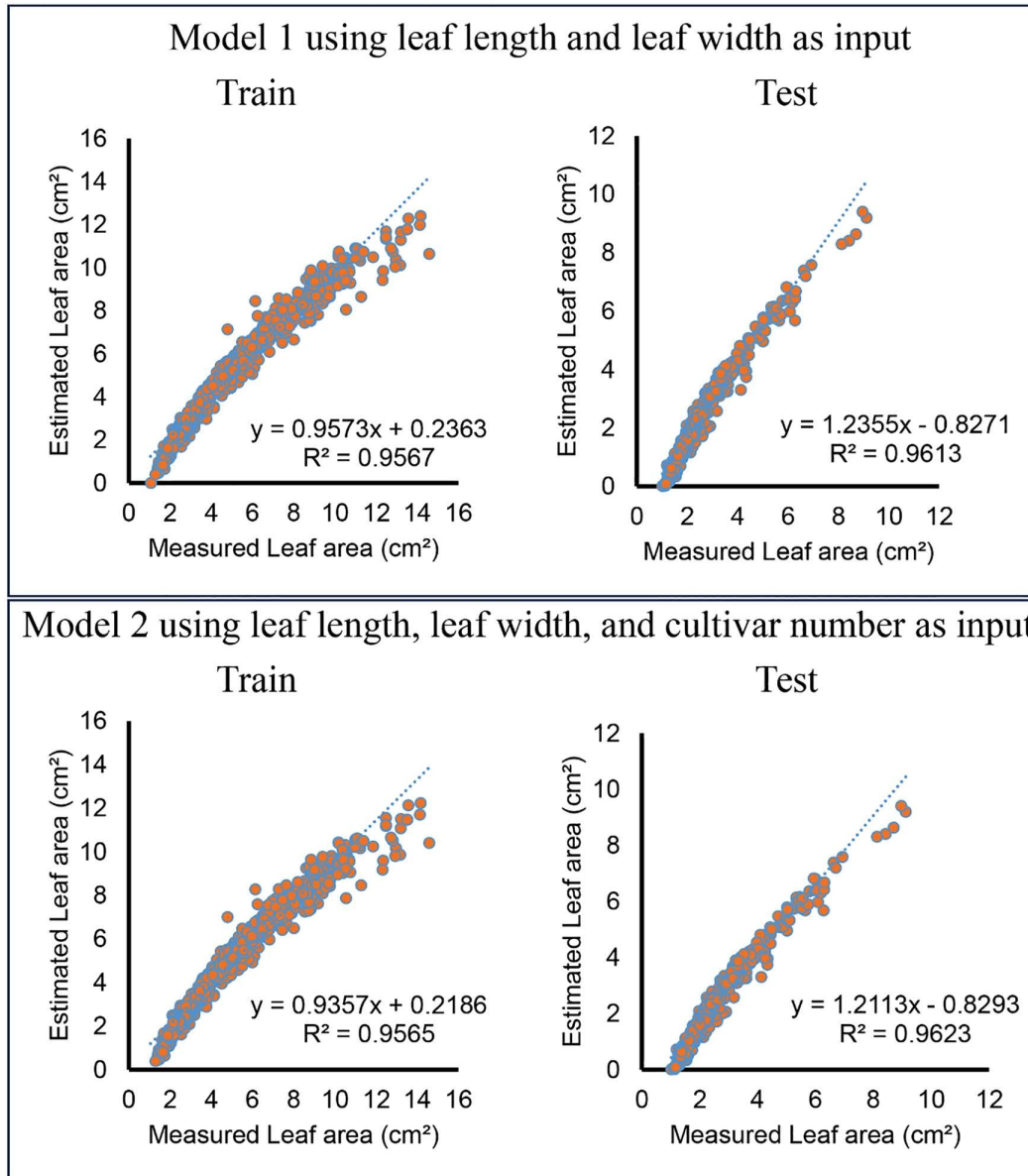


Fig 4. The relationship between measured and estimated leaf area (LA) for the multiple linear regression (MLR) model in olive cultivars.

<https://doi.org/10.1371/journal.pone.0339865.g004>

with 880 samples for training and 440 for testing (Table 3). Another general model was developed by adding the cultivar number (CN) as an additional input, resulting in a three-input model with four neurons in the hidden layer. Different neuron numbers were tested in the hidden layers, and the best predictions were obtained using 2-3-1 and 3-4-1 network architectures. Additionally, different activation functions and training algorithms were tested, but the best results for all models were consistently achieved with tansig and purelin activation functions and the LM training algorithm.

The ANN models demonstrated high predictive performance for leaf area estimation across different olive cultivars. Two general models were developed: the first used leaf length (LL) and leaf width (LW) as input variables, while the second model incorporated cultivar number (CN) as an additional input (Table 4). For the first model, the ANN achieved an R^2

Table 3. Estimation of the performance of the ANN in both training and testing phases for olive cultivar.

Cultivars	CN	Train				Test			
		MBE	MAE	RMSE	R2	MBE	MAE	RMSE	R2
'Arbequina'	1	0.050	0.140	0.019	0.998	-0.043	0.051	0.077	0.990
'Ayvalık'	2	0.001	0.173	0.028	0.990	0.034	0.083	0.119	0.989
'Çelebi'	3	-0.347	0.358	0.103	0.985	-0.217	0.222	0.266	0.982
'Domat'	4	1.045	1.045	0.775	0.989	0.324	0.324	0.384	0.977
'Edincik Su'	5	0.461	0.501	0.193	0.994	0.236	0.297	0.374	0.949
'Elmacık'	6	0.086	0.103	0.009	0.960	0.996	1.485	0.819	0.970
'Frantoio'	7	-0.214	0.360	0.100	0.982	-0.219	0.222	0.268	0.994
'Gemlik'	8	-0.060	0.159	0.022	0.992	-0.072	0.087	0.121	0.981
'Gemlik-21'	9	0.037	0.179	0.028	0.991	0.042	0.149	0.199	0.975
'Girit Zeytini'	10	-0.085	0.098	0.007	0.988	-0.093	0.093	0.118	0.997
'Halhalı'	11	0.061	0.136	0.013	0.990	0.069	0.069	0.085	0.991
'Karamani'	12	-0.237	0.246	0.047	0.992	-0.208	0.209	0.244	0.967
'Kilis Yağlık'	13	-0.340	0.373	0.130	0.988	-0.128	0.138	0.172	0.961
'Manzanilla'	14	-0.183	0.224	0.035	0.992	-0.192	0.195	0.225	0.969
'Memecik'	15	-0.030	0.174	0.026	0.983	-0.009	0.068	0.075	0.992
'Nizip Yağlık'	16	-0.335	0.430	0.258	0.912	-0.100	0.127	0.176	0.979
'Sarı Haşebi'	17	0.143	0.156	0.027	0.996	0.012	0.093	0.117	0.931
'Sarı Ulak'	18	-0.254	0.286	0.052	0.990	-0.217	0.220	0.257	0.995
'Sarı Yaprak'	19	0.011	0.251	0.046	0.985	0.016	0.074	0.088	0.995
'Saurani'	20	0.003	0.059	0.003	0.992	0.066	0.066	0.075	0.984
'Taşan Yüreği'	21	0.164	0.323	0.079	0.988	0.080	0.169	0.229	0.992
'Uslu'	22	0.013	0.246	0.053	0.984	-0.130	0.134	0.171	0.981

CN number of cultivar, R2 coefficient of determination, MBE mean bias error, MAE mean absolute error, RMSE root mean square error.

<https://doi.org/10.1371/journal.pone.0339865.t003>

Table 4. Estimation of the performance of ANN in both training and testing phases for olive cultivars.

All cultivars	Train				Test			
	MBE	MAE	RMSE	R2	MBE	MAE	RMSE	R2
Model 1	0.001	0.275	0.432	0.975	-0.052	0.166	0.227	0.975
Model 2	-0.001	0.167	0.245	0.992	0.019	0.142	0.040	0.995

Model 1: leaf length and leaf width as input; Model 2: leaf length, leaf width, and number of cultivars as input.

<https://doi.org/10.1371/journal.pone.0339865.t004>

value of 0.975 in both the training and test sets, with an RMSE of 0.432 and 0.227, respectively. When the cultivar number was included as an input, model performance slightly improved, with the training R² increasing to 0.992 and RMSE reducing to 0.245. Among individual cultivars, the best-performing ANN models in the test set were observed for Girit Zeytini (R²: 0.997, RMSE: 0.118, MAE: 0.093) and Arbequina (R²: 0.990, RMSE: 0.077, MAE: 0.051), indicating highly reliable predictions (Table 3). However, the ANN models performed relatively less effectively for certain cultivars, such as Elmacık in the test set, where RMSE was 0.819, and MAE was 1.485, suggesting potential challenges in predicting leaf area for this cultivar (Table 3).

Overall, the ANN models exhibited strong predictive accuracy, with low MBE values, indicating minimal systematic bias. The inclusion of the cultivar number as an additional predictor slightly improved model performance, highlighting the potential influence of cultivar-specific traits on leaf area estimation.

3.3. ANN weight analysis

The ANN model's weight distributions between the input, hidden, and output layers are presented in [Tables 6–8](#). The weight values indicate the significance of each feature in determining the final output.

3.4. Hidden layer weights (LL and LW as input)

[Table 5](#) shows the weight values between the input layer and hidden layer for all olive cultivars. The connections between neurons varied significantly, with W1i values ranging from 0.523 to 0.0078 and W2i values ranging from 3.211 to 0.0232, suggesting that LW had a more substantial influence than LL. Bias values ranged from −6.979 to 16.960, which influenced the activation function shifts.

3.5. Output layer weights (LL and LW as input)

[Table 6](#) presents the weight values between the hidden layer and the output layer. The final leaf area prediction was computed using a weighted sum of the hidden layer outputs, where weights varied significantly among neurons, particularly with Wi values ranging from 0.545 to 30.613. The bias value was 16.657, contributing to network adjustments.

The hidden layer outputs were calculated using the tansig activation function, defined as [Eqs. 4–6](#).

$$[\text{Output}]_{11} = 2 / (1 + \exp (-2 \times ((0.523 \times \text{LL}) + (3.211 \times \text{LW}) - 6.979))) - 1 \quad (4)$$

$$[\text{Output}]_{12} = 2 / (1 + \exp (-2 \times ((-0.255 \times \text{LL}) + (0.249 \times \text{LW}) + 16.960))) - 1 \quad (5)$$

$$[\text{Output}]_{13} = 2 / (1 + \exp (-2 \times ((0.0078 \times \text{LL}) + (0.0232 \times \text{LW}) - 1.6677))) - 1 \quad (6)$$

The final output was obtained using the purelin activation function ([Eq. 7](#)).

$$\text{Output} = (10.565 \times [\text{Output}]_{11}) + (0.545 \times [\text{Output}]_{12}) + (30.613 \times [\text{Output}]_{13}) + 16.657 \quad (7)$$

3.6. Hidden layer weights (LL, LW, and CN as input)

[Table 7](#) presents the weight values when the cultivar number (CN) was included as an additional input. The weights indicate a more complex network structure, with W1i values ranging from 0.0159 to 0.0058, suggesting relatively minor contributions from LL. The bias values varied significantly (−1.628 to 5.309), altering activation function outputs.

Table 5. The weight values between the input layer and the hidden layer for olive cultivars.

Weights	Number of neurons in the hidden layer (i)		
	1	2	3
W1i	0.523	−0.255	0.0078
W2i	3.211	0.249	0.0232
Bias1	−6.979	16.960	−1.6677

<https://doi.org/10.1371/journal.pone.0339865.t005>

Table 6. The weight values between the output layer and the hidden layer for olive cultivars.

Weights	Number of neurons in the hidden layer (i)			
	1	2	3	Bias 2
Wi	10.565	0.545	30.613	16.657

<https://doi.org/10.1371/journal.pone.0339865.t006>

Table 7. The weight values between the input layer and the hidden layer for olive cultivars.

Weights	Number of neurons in the hidden layer (i)			
	1	2	3	4
W1i	0.0159	0.640	0.0096	0.0058
W2i	0.0512	0.290	0.1615	0.16884
W3i	-0.0017	-0.307	-0.0125	-0.01383
Bias1	-1.628	5.309	-4.627	-4.55271

<https://doi.org/10.1371/journal.pone.0339865.t007>

The complete weight matrices for the ANN model incorporating dummy-encoded cultivar variables (reference: 'Gemlik') are provided in [S6 Table](#).

3.7. Output layer weights (LL, LW, and CN as input)

[Table 8](#) displays the weight values between the hidden and output layers when CN was included. The output was computed using:

The hidden layer outputs were calculated using the tansig activation function, defined as [Eqs. 8–11](#).

$$[\text{Output}]_{11} = 2 / (1 + \exp(-2 \times ((0.0159 \times \text{LL}) + (0.0512 \times \text{LW}) + (-0.0017 \times \text{CN}) - 1.628))) - 1 \quad (8)$$

$$[\text{Output}]_{12} = 2 / (1 + \exp(-2 \times ((0.640 \times \text{LL}) + (0.290 \times \text{LW}) + (-0.307 \times \text{CN}) - 5.309))) - 1 \quad (9)$$

$$[\text{Output}]_{13} = 2 / (1 + \exp(-2 \times ((0.0096 \times \text{LL}) + (0.1615 \times \text{LW}) + (-0.0125 \times \text{CN}) - 4.627))) - 1 \quad (10)$$

$$[\text{Output}]_{14} = 2 / (1 + \exp(-2 \times ((0.0058 \times \text{LL}) + (0.16884 \times \text{LW}) + (-0.01383 \times \text{CN}) - 4.55271))) - 1 \quad (11)$$

The final output was obtained using the purelin activation function ([Eq. 12](#)).

$$\text{Output} = (0.774 \times [\text{Output}]_{11}) + (-0.329 \times [\text{Output}]_{12}) + (-0.537 \times [\text{Output}]_{13}) + (-2.591 \times [\text{Output}]_{14}) + 1.868 \quad (12)$$

Overall, the ANN model demonstrated that LW had a more substantial effect on leaf area prediction than LL, and the inclusion of cultivar number slightly improved performance, as reflected in weight distributions and bias shifts. Since Model 2 also accounts for cultivar numbers, it provides a more detailed prediction, but the performance difference between the two models is quite slight. If the cultivar is known, Model 2 should be preferred; if the cultivar is unknown, Model 1 would be the more practical choice.

The detailed mathematical formulation of the forward pass and network equations is provided in [S2 Note](#).

[Fig 5](#) presents the relationship between measured and estimated leaf area values for the ANN models. The scatter plots illustrate the model's performance for both training and testing datasets. In the model utilizing LL and LW as inputs,

Table 8. The weight values between the output layer and the hidden layer for olive cultivars.

Weights	Number of neurons in the hidden layer (i)				
	1	2	3	4	Bias 2
Wi	0.774	-0.329	-0.537	-2.591	1.868

<https://doi.org/10.1371/journal.pone.0339865.t008>

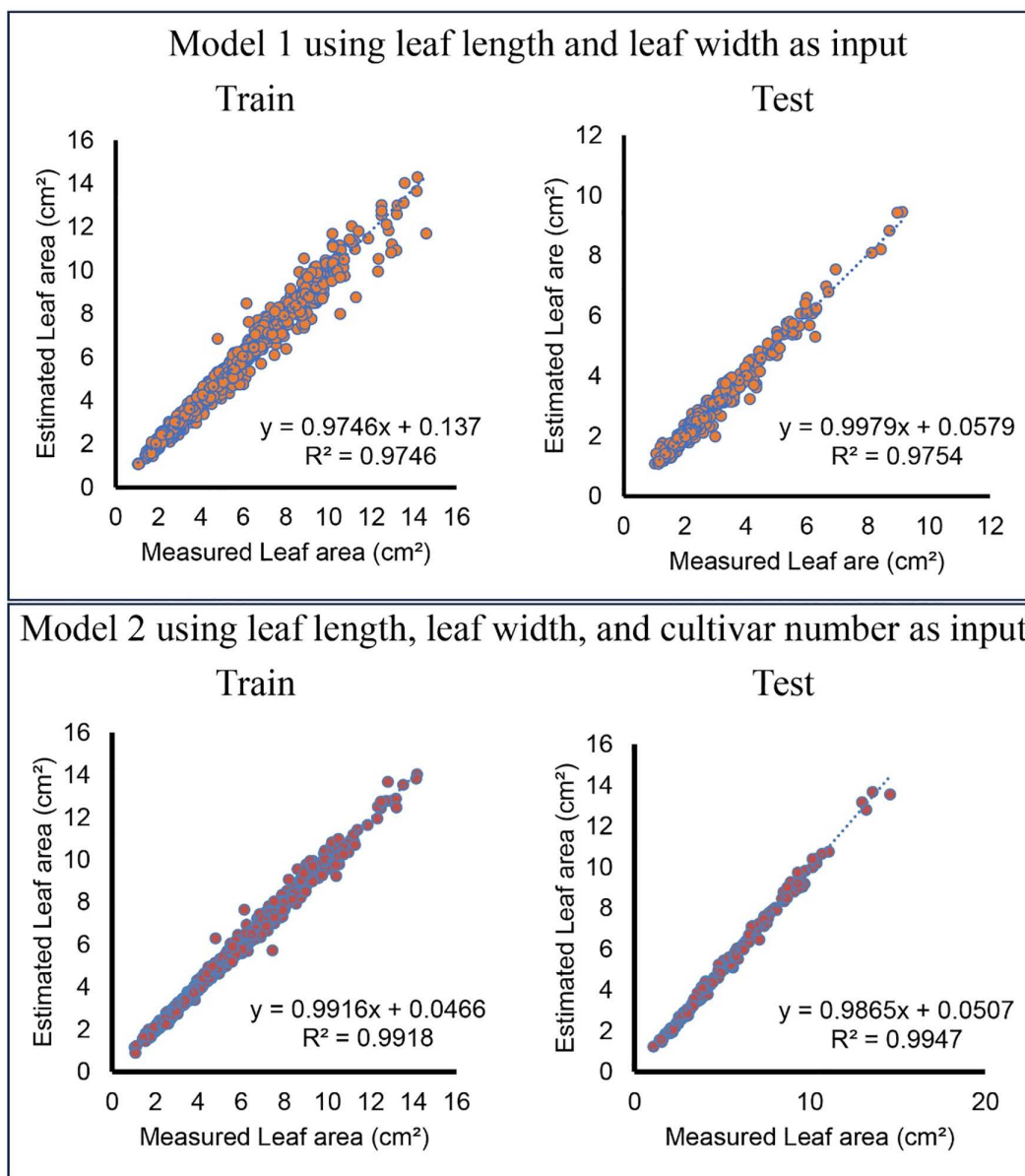


Fig 5. The relationship between measured and estimated leaf area for the artificial neural networks (ANN) model in olive cultivars.

<https://doi.org/10.1371/journal.pone.0339865.g005>

the coefficient of determination (R^2) values were 0.9746 for the training set and 0.9754 for the test set, indicating strong agreement between estimated and measured values. When cultivar number (CN) was included as an additional input, the R^2 values improved slightly to 0.9918 for training and 0.9947 for testing, demonstrating that adding cultivar-specific information marginally enhanced the model's predictive power. The regression equations in the scatter plots showed a relationship close to 1:1 between the measured and predicted values, with slopes close to 1 and low intercepts. This indicates that the ANN model generalizes well across different olive cultivars and minimizes systematic biases. In general, the high R^2 values across all configurations indicate that the ANN model effectively captures the relationship between leaf

dimensions and leaf area. The slight improvement observed with the inclusion of CN indicated that cultivar-specific differences played a minor role in the estimation of leaf area.

4. Discussion

This study aimed to compare the performance of artificial neural networks (ANN) and multiple linear regression (MLR) models in predicting leaf area for different olive cultivars. Based on the obtained performance metrics (R^2 , MAE, RMSE), both models produced successful results, but the ANN models exhibited higher accuracy.

Among the cultivars, 'Domat', 'Elmacık', and 'Sarı Haşebi' exhibited significantly higher mean leaf area values compared with the reference cultivar 'Gemlik' after adjustment for multiple testing. These differences likely reflect intrinsic morphological and physiological variations among olive genotypes. 'Domat', which is typically cultivated in humid and fertile regions of the Aegean Basin, is known for its vigorous growth habit and larger canopy structure, traits commonly associated with broader leaves and enhanced photosynthetic surface area. Similarly, 'Elmacık' is characterized by thick, elliptic leaves with high specific leaf area, which may confer improved photosynthetic efficiency under moderate water availability. In contrast, cultivars such as 'Çelebi' and 'Nizip Yağlık', adapted to drier environments, tend to possess smaller, narrower leaves as a xeromorphic adaptation to minimize transpiration. These findings indicate that the observed cultivar-specific deviations are not only statistical but also biologically consistent with the ecological and morphological diversity within *Olea europaea* germplasm. Comparable patterns have been reported in previous studies linking leaf morphological traits to environmental adaptation and genotypic variation in olives [2,3,7].

The overall success of the ANN model, attributed to its capacity to learn nonlinear relationships, demonstrated that it provides a more effective approach for leaf area estimation ($R^2=0.995$). In contrast, the coefficient of determination (R^2) for the MLR models was calculated as 0.96, both for the model using LL and LW as input parameters and for the model incorporating LL, LW, and CN as inputs. The inclusion of the CN variable in the ANN model improved performance and enhanced generalizability, enabling more comprehensive predictions. Overall, ANN models outperformed MLR models in both the training and testing phases.

Yuan et al. [70] compared different regression methods, including ANN, random forest (RF), and support vector machine (SVM) models, for predicting the leaf area index (LAI) in soybean crops. Their study concluded that ANN provided the highest precision, particularly for single growth phases, outperforming SVM and traditional statistical approaches. Similarly, Kumar et al. [71] developed an ANN model incorporating leaf length and width as inputs for durian leaf area estimation. They compared ANN with regression models and found that ANN yielded significantly better results, achieving an R^2 of 0.96 in the training phase and 0.94 in the testing phase, outperforming linear regression methods. These findings align with our study, where ANN-based models demonstrated a strong correlation between estimated and actual leaf area across different olive cultivars.

Shabani et al. [72] further validated ANN's capability by testing it on 61 plant species with varying leaf morphologies. Their study revealed that ANN models performed consistently well across different species, reinforcing their versatility and robustness. The research also highlighted the importance of selecting input parameters, showing that pre-processing methods, such as MinMax normalization, significantly improved ANN model accuracy. These findings suggest that ANN models are adaptable to different plant species and growth conditions, making them ideal for agricultural applications.

The findings of this study align with previous research on similar topics. Kumar et al. [71], Küçükönder et al. [73], and Ahmadian-Moghadam [54] also demonstrated that ANN models achieve high accuracy in leaf area estimation. However, Shabani et al. [72] emphasized the need to develop cultivar-specific models, as generalized ANN models may underperform for certain plant species.

Additionally, Kallel et al. [74] indicated that ANN models can be effectively utilized for leaf area index (LAI) estimation, proving beneficial for monitoring olive tree health. Similarly, Mhanna [75] reported that mathematical models developed for estimating leaf area in the 'Khoderi' olive cultivar in Syria achieved an R^2 value of 0.962, and the same approach was

successfully applied to the French 'Picholine' cultivar. This further supports the idea that cultivar-specific models yield more accurate predictions.

Koubouris et al. [76] highlighted that using both leaf length (LL) and width (LW) together for leaf area estimation produces better results than using either variable alone. Their modeling study on 10 different olive cultivars showed R^2 values ranging from 0.71 to 0.92. Similarly, in this study, using both LL and LW in ANN and MLR models improved prediction accuracy. In MLR models, leaf width (LW) was found to be more influential in predicting leaf area.

Öztürk et al. [26] confirmed that ANN models outperformed MLR models in predicting the leaf area (LA) of pear cultivars. They found that ANN models provided higher accuracy in both training and testing phases, with R^2 values reaching 0.991 in training and 0.985 in testing, while MLR models showed slightly lower performance with R^2 values of 0.987 in training and 0.979 in testing. These findings further validate the conclusions of our study, which demonstrated that ANN models offer a more reliable and accurate approach to LA estimation compared with traditional regression-based models.

In a study by Asriani [77], the performance of artificial neural networks (ANN) in leaf area estimation was examined using the backpropagation algorithm. Leaf length and width were used as input variables, while leaf area was the output variable. Their results showed that an ANN model with a 2-50-1 architecture achieved 99.99% accuracy in predicting leaf area for seven plant species.

Kishore et al. [78] developed a non-destructive model for estimating apple (*Malus domestica*) leaf area, using linear leaf measurements such as leaf length (L) and width (W). The study tested multiple regression models and found that the most accurate model was $LA = 1.120301 + 0.615(L \times W)$. This model achieved an R^2 of 0.98, demonstrating high accuracy in predicting leaf area across multiple apple cultivars. Kumar and Sharma [79] developed a leaf area estimation model for *Picrorhiza kurroa*, an important medicinal plant. They explored multiple linear, quadratic, exponential, and logarithmic regression models using leaf length (L) and width (W) as independent variables. The most accurate model was found to be $LA = 0.333 + 0.603(L \times W)$, which achieved an R^2 of 0.955 in model calibration and 0.9053 in validation, demonstrating high accuracy and robustness. These findings further support the idea that empirical models using leaf dimensions can provide reliable non-destructive leaf area estimation. However, our study, along with others incorporating artificial neural networks (ANNs) [26,71], suggests that machine learning-based approaches offer an even more precise and adaptable solution, particularly for complex and heterogeneous datasets.

Monitoring plant health and evaluating growth processes are crucial for enhancing agricultural productivity and achieving sustainable production goals [8]. In this context, leaf area estimation is a key parameter for assessing water and nutrient use, evaluating photosynthetic efficiency, and analyzing yield potential. Accurately and reliably estimating leaf area in economically significant crops such as olives contributes to a better understanding of plant physiology and supports strategic decision-making in agricultural practices.

The findings of this study indicate that ANN models offer a strong alternative for leaf area estimation, providing more accurate results compared with MLR models. However, it is concluded that cultivar-specific analyses and the integration of additional variables play a critical role in improving model performance.

The results show that ANN produces lower error without creating systematic bias compared to MLR with the same inputs. Effect sizes indicate small-to-medium/medium-level improvements in practical terms and quantitatively confirm an accuracy advantage consistent with high R^2 values.

5. Conclusions

In this study, multiple linear regression (MLR) and artificial neural network (ANN) models were developed to predict leaf area (LA) for 22 different olive cultivars. While leaf length (LL) and leaf width (LW) were used for individual predictions, cultivar number (CN) was included in general predictions. The developed models enabled non-destructive and cost-effective leaf area estimation, eliminating the need for expensive equipment. Deterministic models that can be easily used in accession were formulated.

Leaf area is a critical parameter for plant growth and productivity. Various methods exist for LA measurement, including regression models, grid counting, gravimetric analysis, planimetry, image processing, and adaptive neuro-fuzzy inference systems. In this study, ANN models successfully produced results using a 2-3-1 network architecture for individual predictions and a 3-4-1 network architecture for general predictions. Comparative analyses revealed that ANN models outperformed MLR models in LA prediction. ANNs demonstrated greater flexibility and adaptability to different olive cultivars, making them a more effective approach. However, MLR models remain a practical and reliable method due to their ease of use. Predicting olive leaf area is crucial for monitoring plant health and productivity in agriculture. AI-based methods can analyze satellite imagery data to assess plant health and stress levels. LA estimation can help optimize production planning and prevent fraud in the olive industry by providing early yield predictions.

In conclusion, AI-based methods are effective and innovative tools for predicting olive leaf area. The integration of artificial neural networks (ANN) with remote sensing technologies could further enhance agricultural productivity and support the development of sustainable farming practices. The findings of this study confirm that ANN models provide a strong alternative for leaf area estimation, offering more accurate results than MLR models. However, it is concluded that cultivar-specific analyses and the integration of additional variables play a critical role in improving model performance. Practically, the developed models can be applied as rapid, non-destructive, and cost-efficient tools for phenotypic characterization, growth monitoring, and precision management in olive breeding and cultivation systems. Future research should aim to validate these models under field conditions, expand datasets with more diverse genotypes and environmental factors, and explore advanced algorithms such as convolutional or ensemble learning methods to further improve predictive accuracy and generalization capability.

Supporting information

S1 Table. Descriptive statistics for individual olive cultivars used in this study.

(DOCX)

S2 Table. Dummy (one-hot) coefficients for cultivars (reference: 'Gemlik') in the multiple linear regression model.

(DOCX)

S3 Table. Paired-sample t-test results for comparison of LA estimation methods.

(DOCX)

S4 Table. Hyper-parameter and functions selected during the training phase.

(DOCX)

S5 Table. Multiple linear regression leaf area models were developed for olive cultivar.

(DOCX)

S6 Table. Weight values between the input and hidden layers of the ANN model incorporating dummy (one-hot) encoded cultivar variables (reference: 'Gemlik').

(DOCX)

S1 Note. Normalization of input and output data.

(DOCX)

S2 Note. Mathematical formulation of the ANN model.

(DOCX)

S1 Fig. Iteration RMSE change according to 2-3-1 network structure.

(DOCX)

S2 Fig. Iteration RMSE change according to 3-4-1 network structure.

(DOCX)

S1 Code. The MATLAB codes used.

(DOCX)

Author contributions

Conceptualization: Ali Khadivi, Hüsnü Demirsoy, Hatice Gözel.

Data curation: Kadir Uğurtan Yılmaz, Hüsnü Demirsoy, Hatice Gözel, Daya Shankar Mishra.

Formal analysis: Yazgan Tunç, Fatih Demirel, Kadir Uğurtan Yılmaz, Hüsnü Demirsoy, Bilal Cemek.

Investigation: Yazgan Tunç, Fatih Demirel, Ali Khadivi, Kadir Uğurtan Yılmaz, Hüsnü Demirsoy, Bilal Cemek.

Methodology: Yazgan Tunç, Ali Khadivi, Bilal Cemek, Daya Shankar Mishra.

Project administration: Hatice Gözel, Daya Shankar Mishra.

Supervision: Fatih Demirel.

Validation: Kadir Uğurtan Yılmaz.

Writing – original draft: Fatih Demirel, Ali Khadivi, Kadir Uğurtan Yılmaz, Hüsnü Demirsoy, Hatice Gözel, Daya Shankar Mishra.

Writing – review & editing: Yazgan Tunç, Ali Khadivi, Kadir Uğurtan Yılmaz, Hüsnü Demirsoy, Bilal Cemek.

References

1. Palaz EB, Demirel S, Popescu GC, Demirel F, Uğur R, Yaman M, et al. Refinement of surface sterilization protocol for in vitro olive (*Olea europaea* L.) shoot proliferation and optimizing by machine learning techniques. *Hortic Environ Biotechnol*. 2025;66(4):813–28. <https://doi.org/10.1007/s13580-025-00685-z>
2. Sebastiani L, Busconi M. Recent developments in olive (*Olea europaea* L.) genetics and genomics: applications in taxonomy, varietal identification, traceability and breeding. *Plant Cell Rep*. 2017;36(9):1345–60. <https://doi.org/10.1007/s00299-017-2145-9> PMID: [28434019](https://pubmed.ncbi.nlm.nih.gov/28434019/)
3. Rotondi A, Fabbri A, Ganino T, Beghè D, Magli M, Morrone L. Genetic and Landscape Characterization of Ancient Crops: The Olive Tree, a Case Study in Northern Italy. *Innovations in Landscape Research*. Springer International Publishing; 2021. pp. 457–77. https://doi.org/10.1007/978-3-030-67448-9_22
4. Wang J, Zhang D, Farooqi TJA, Ma L, Deng Y, Jia Z. The olive (*Olea europaea* L.) industry in China: its status, opportunities and challenges. *Agroforest Syst*. 2017;93(2):395–417. <https://doi.org/10.1007/s10457-017-0129-y>
5. Uzundumlu AS, Ateş T. Olive production forecasts in the leader countries for 2023–2027. *Appl Fruit Sci*. 2025;67(2). <https://doi.org/10.1007/s10341-025-01268-z>
6. Loumou A, Giourga C. Olive groves: “The life and identity of the Mediterranean”. *Agric Hum Values*. 2003;20(1):87–95. <https://doi.org/10.1023/a:1022444005336>
7. Connor DJ. Adaptation of olive (*Olea europaea* L.) to water-limited environments. *Austr J Agric Res*. 2005;56(11):1181–9. <https://doi.org/10.1071/ar05169>
8. Ewert F. Modelling plant responses to elevated CO₂: how important is leaf area index? *Ann Bot*. 2004;93(6):619–27. <https://doi.org/10.1093/aob/mch101> PMID: [15102613](https://pubmed.ncbi.nlm.nih.gov/15102613/)
9. Bakó K, Rácz C, Dövényi-Nagy T, Molnár K, Dobos A. Advancements in leaf area index estimation for maize using modeling and remote sensing techniques: a review. *Agronomy*. 2025;15(3):519. <https://doi.org/10.3390/agronomy15030519>
10. Demirsoy H. Leaf area estimation in some species of fruit tree by using models as a non-destructive method. *Fruits*. 2009;64(1):45–51. <https://doi.org/10.1051/fruits/2008049>
11. Pandey SK, Singh H. A simple, cost-effective method for leaf area estimation. *J Bot*. 2011;2011:1–6. <https://doi.org/10.1155/2011/658240>
12. Poorter H, Niinemets Ü, Poorter L, Wright IJ, Villar R. Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytol*. 2009;182(3):565–88. <https://doi.org/10.1111/j.1469-8137.2009.02830.x> PMID: [19434804](https://pubmed.ncbi.nlm.nih.gov/19434804/)
13. Feilhauer H, Schmid T, Faude U, Sánchez-Carrillo S, Cirujano S. Are remotely sensed traits suitable for ecological analysis? A case study of long-term drought effects on leaf mass per area of wetland vegetation. *Ecol Indic*. 2018;88:232–40. <https://doi.org/10.1016/j.ecolind.2018.01.012>

14. Patil P, Biradar P, U. Bhagawathi A, S. Hejjegar I. A review on leaf area index of horticulture crops and its importance. *Int J Curr Microbiol App Sci*. 2018;7(04):505–13. <https://doi.org/10.20546/ijcmas.2018.704.059>
15. Gower ST, Kucharik CJ, Norman JM. Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sens Environ*. 1999;70(1):29–51. [https://doi.org/10.1016/s0034-4257\(99\)00056-5](https://doi.org/10.1016/s0034-4257(99)00056-5)
16. Nemani RR, Running SW. Testing a theoretical climate-soil-leaf area hydrologic equilibrium of forests using satellite data and ecosystem simulation. *Agric For Meteorol*. 1989;44(3–4):245–60. [https://doi.org/10.1016/0168-1923\(89\)90020-8](https://doi.org/10.1016/0168-1923(89)90020-8)
17. Yang J, Medlyn BE, De Kauwe MG, Duursma RA. Applying the concept of ecohydrological equilibrium to predict steady state leaf area index. *J Adv Model Earth Syst*. 2018;10(8):1740–58. <https://doi.org/10.1029/2017ms001169>
18. Fang H, Baret F, Plummer S, Schaepman-Strub G. An overview of global leaf area index (LAI): methods, products, validation, and applications. *Rev Geophys*. 2019;57(3):739–99. <https://doi.org/10.1029/2018rg000608>
19. Kandianan K, Parthasarathy U, Krishnamurthy KS, Thankamani CK, Srinivasan V. Modeling individual leaf area of ginger (*Zingiber officinale* Roscoe) using leaf length and width. *Sci Hortic*. 2009;120(4):532–7. <https://doi.org/10.1016/j.scienta.2008.11.037>
20. Keramatlou I, Sharifani M, Sabouri H, Alizadeh M, Kamkar B. A simple linear model for leaf area estimation in Persian walnut (*Juglans regia* L.). *Sci Hortic*. 2015;184:36–9. <https://doi.org/10.1016/j.scienta.2014.12.017>
21. Sankar V, Sakthivel T, Karunakaran G, Tripathi PC. Non-destructive estimation of leaf area of durian (*Durio zibethinus*)—an artificial neural network approach. *Sci Hort*. 2017;219:319–25.
22. Ebrahimi N, Salihi AR, Alipour S, Mozafari SH, Aliyar J, Darwish I. Prediction of distribution of dry matter and leaf area of faba bean (*Vicia faba*) using nonlinear regression models. *Agric Res*. 2024;13(2):381–9. <https://doi.org/10.1007/s40003-024-00700-2>
23. Sala F, Arsene G-G, Iordănescu O, Boldea M. Leaf area constant model in optimizing foliar area measurement in plants: A case study in apple tree. *Sci Hortic*. 2015;193:218–24. <https://doi.org/10.1016/j.scienta.2015.07.008>
24. Soysal D. A non-destructive leaf area prediction model and some physical leaf properties in apples. *App Fruit Sci*. 2024;66(4):1181–6. <https://doi.org/10.1007/s10341-024-01103-x>
25. Demirsoy H, Küçüktopçu E, Doğan DE. Novel Machine Learning Approaches for Accurate Leaf Area Estimation in Apples. *Appl Fruit Sci*. 2025;67(2). <https://doi.org/10.1007/s10341-025-01292-z>
26. Öztürk A, Cemek B, Demirsoy H, Küçüktopcu E. Modelling of the leaf area for various pear cultivars using neuro computing approaches. *Span J Agric Res*. 2019;17(4):e0206. <https://doi.org/10.5424/sjar/2019174-14675>
27. Demirsoy H, Demirsoy L, Uzun S, Ersoy B. Non-destructive leaf area estimation in peach. *Europ J Hort Sci*. 2004;69(4):144–6.
28. Cirillo C, Pannico A, Basile B, Rivera CM, Giaccone M, Colla G, et al. A simple and accurate allometric model to predict single leaf area of twenty-one European apricot cultivars. *Euro J Hortic Sci*. 2017;82(2):65–71. <https://doi.org/10.17660/ejhs.2017/82.2.1>
29. Tsialtas JT, Koundouras S, Zioziou E. Leaf area estimation by simple measurements and evaluation of leaf area prediction models in Cabernet-Sauvignon grapevine leaves. *Photosynt*. 2008;46(3):452–6. <https://doi.org/10.1007/s11099-008-0077-x>
30. Demirsoy H, Demirsoy L, Öztürk A. Improved model for the non-destructive estimation of strawberry leaf area. *Fruits*. 2005;60(1):69–73. <https://doi.org/10.1051/fruits:2005014>
31. Demirsoy H, Lang GA. Short communication.validation of a leaf area estimation model for sweet cherry. *Span J Agric Res*. 2010;8(3):830–2. <https://doi.org/10.5424/sjar/2010083-1285>
32. Budiarto R, Poerwanto R, Santosa E, Efendi D, Agusta A. Short Communication: Allometric model to estimate bifoliate leaf area and weight of kaffir lime (*Citrus hystrix*). *Biodiversitas*. 2021;22(5). <https://doi.org/10.13057/biodiv/d220545>
33. Serdar Ü, Demirsoy H. Non-destructive leaf area estimation in chestnut. *Sci Hortic*. 2006;108(2):227–30. <https://doi.org/10.1016/j.scienta.2006.01.025>
34. Cristofori V, Roupheal Y, Gyves EM, Bignami C. A simple model for estimating leaf area of hazelnut from linear measurements. *Sci Hortic*. 2007;113(2):221–5. <https://doi.org/10.1016/j.scienta.2007.02.006>
35. Mendoza-de Gyves E, Roupheal Y, Cristofori V, Mira FR. A non-destructive, simple and accurate model for estimating the individual leaf area of kiwi (*Actinidia deliciosa*). *Fruits*. 2007;62(3):171–6. <https://doi.org/10.1051/fruits:2007012>
36. Zellers CE, Saunders MR, Morrissey RC, Shields JM, Bailey BG, Dyer J, et al. Development of allometric leaf area models for intensively managed black walnut (*Juglans nigra* L.). *Ann For Sci*. 2012;69(8):907–13. <https://doi.org/10.1007/s13595-012-0215-2>
37. Ghoreishi M, Hossini Y, Maftoon M. Simple models for predicting leaf area of mango (*Mangifera indica* L.). *J Biol Earth Sci*. 2012;2(2):45–53.
38. Cemek B, Arslan H, Küçüktopcu E, Simsek H. Comparative analysis of machine learning techniques for estimating groundwater deuterium and oxygen-18 isotopes. *Stoch Environ Res Risk Assess*. 2022;36(12):4271–85. <https://doi.org/10.1007/s00477-022-02262-7>
39. Küçüktopcu E. Comparative analysis of data-driven techniques to predict heating and cooling energy requirements of poultry buildings. *Buildings*. 2023;13(1):142. <https://doi.org/10.3390/buildings13010142>
40. Tunca E, Köksal ES, Çetin Taner S. Calibrating UAV thermal sensors using machine learning methods for improved accuracy in agricultural applications. *Infrared Phys Technol*. 2023;133:104804. <https://doi.org/10.1016/j.infrared.2023.104804>
41. Yıldırım D, Küçüktopcu E, Cemek B, Simsek H. Comparison of machine learning techniques and spatial distribution of daily reference evapotranspiration in Türkiye. *Appl Water Sci*. 2023;13(4). <https://doi.org/10.1007/s13201-023-01912-7>

42. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349(6245):255–60. <https://doi.org/10.1126/science.aaa8415> PMID: 26185243
43. Attri I, Awasthi LK, Sharma TP. Machine learning in agriculture: a review of crop management applications. *Multimed Tools Appl*. 2023;83(5):12875–915. <https://doi.org/10.1007/s11042-023-16105-2>
44. Tunca E, Köksal ES, Öztürk E, Akay H, Çetin Taner S. Accurate estimation of sorghum crop water content under different water stress levels using machine learning and hyperspectral data. *Environ Monit Assess*. 2023;195(7). <https://doi.org/10.1007/s10661-023-11536-8>
45. Ahsan M, Eshkabilov S, Cemek B, Küçüktopcu E, Lee CW, Simsek H. Deep learning models to determine nutrient concentration in hydroponically grown lettuce cultivars (*Lactuca sativa* L.). *Sustainability*. 2021;14(1):416. <https://doi.org/10.3390/su14010416>
46. van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: a systematic literature review. *Comput Electron Agric*. 2020;177:105709. <https://doi.org/10.1016/j.compag.2020.105709>
47. Bertalan L, Holb I, Pataki A, Négyesi G, Szabó G, Kupásné Szalóki A, et al. UAV-based multispectral and thermal cameras to predict soil water content – a machine learning approach. *Comput Electron Agric*. 2022;200:107262. <https://doi.org/10.1016/j.compag.2022.107262>
48. Cemek B, Ünlükara A, Kurunç A, Küçüktopcu E. Leaf area modeling of bell pepper (*Capsicum annuum* L.) grown under different stress conditions by soft computing approaches. *Comput Electron Agric*. 2020;174:105514. <https://doi.org/10.1016/j.compag.2020.105514>
49. Vazquez-Cruz MA, Luna-Rubio R, Contreras-Medina LM, Torres-Pacheco I, Guevara-Gonzalez RG. 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*. 2012;112(4):319–27. <https://doi.org/10.1016/j.biosystemseng.2012.05.003>
50. Lee J, Moon T, Nam DS, Park KS, Son JE. Estimation of Leaf Area in Paprika Based on Leaf Length, Leaf Width, and Node Number Using Regression Models and an Artificial Neural Network. *Hort Sci Technol*. 2018;36(2):183–92. <https://doi.org/10.12972/kjhst.20180019>
51. Kıymaz S, Karadavut U, Ertek A. A comparison of artificial neural networks and some nonlinear models of leaf area estimation of sugar beet at different nitrogen levels. *Turkish J Agri Nat Sci*. 2018;5(3):303–309. <https://doi.org/10.30910/turkjans.448371>
52. Odabas MS, Ergun E, Oner F. Artificial neural network approach for the prediction of the corn (*Zea mays* L.) leaf area. *Bulgarian J Agri Sci*. 2013;19(4):766–9.
53. Aboukarima A, Elsoury H, Menyawi M. Artificial neural network model for the prediction of the cotton crop leaf area. *Int J Plant Soil Sci*. 2015;8(4):1–13. <https://doi.org/10.9734/ijpss/2015/19686>
54. Ahmadian–Moghadam H. Prediction of pepper (*Capsicum annuum* L.) leaf area using group method of data handling–type neural networks. *Int J Agri Sci*. 2012;2:993–9.
55. Amiri MJ, Shabani A. Application of an Adaptive Neural-Based Fuzzy Inference System Model for Predicting Leaf Area. *Commun Soil Sci Plant Anal*. 2017;48(14):1669–83. <https://doi.org/10.1080/00103624.2017.1373801>
56. Sabouri H, Sajadi SJ, Jafarzadeh MR, Rezaei M, Ghaffari S, Bakhtiari S. Image processing and prediction of leaf area in cereals: a comparison of artificial neural networks, an adaptive neuro-fuzzy inference system, and regression methods. *Crop Sci*. 2021;61(2):1013–29. <https://doi.org/10.1002/csc2.20373>
57. Rahimikhoob H, Delshad M, Habibi R. Leaf area estimation in lettuce: Comparison of artificial intelligence-based methods with image analysis technique. *Measurement*. 2023;222:113636. <https://doi.org/10.1016/j.measurement.2023.113636>
58. Sabouri A, Bakhshipour A, Poornoori M, Abouzari A. Application of image processing and soft computing strategies for non-destructive estimation of plum leaf area. *PLoS One*. 2022;17(7):e0271201. <https://doi.org/10.1371/journal.pone.0271201> PMID: 35816484
59. ESRI. “Topographic” [basemap]. Scale Not Given. ArcGIS Desktop (Version 10.1). “World Topographic Map”. 2012. [cited 2025 Jun 20] <http://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f>
60. Benlioğlu B, Demirel F, Türkoğlu A, Haliloğlu K, Özaktan H, Kujawa S, et al. Insights into Drought Tolerance of Tetraploid Wheat Genotypes in the Germination Stage Using Machine Learning Algorithms. *Agriculture*. 2024;14(2):206. <https://doi.org/10.3390/agriculture14020206>
61. Kamble VH, Dale MP. Machine learning approach for longitudinal face recognition of children. *Machine Learning for Biometrics*. Elsevier; 2022. pp. 1–27. <https://doi.org/10.1016/b978-0-323-85209-8.00011-0>
62. Mansfield ER, Helms BP. Detecting Multicollinearity. *Am Stat*. 1982;36(3a):158–60. <https://doi.org/10.1080/00031305.1982.10482818>
63. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. New York: Springer; 2013.
64. Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmosph Environ*. 1998;32(14–15):2627–36. [https://doi.org/10.1016/s1352-2310\(97\)00447-0](https://doi.org/10.1016/s1352-2310(97)00447-0)
65. Haykin S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall; 1999.
66. Ansari HR, Zarei MJ, Sabbaghi S, Keshavarz P. A new comprehensive model for relative viscosity of various nanofluids using feed-forward back-propagation MLP neural networks. *Int Commun Heat Mass Transfer*. 2018;91:158–64. <https://doi.org/10.1016/j.icheatmasstransfer.2017.12.012>
67. Türkoğlu A, Bolouri P, Haliloğlu K, Eren B, Demirel F, Işık Mİ, et al. Modeling Callus Induction and Regeneration in Hypocotyl Explant of Fodder Pea (*Pisum sativum* var. *arvense* L.) using machine learning algorithm method. *Agronomy*. 2023;13(11):2835. <https://doi.org/10.3390/agronomy13112835>

68. Demirel F, Eren B, Yilmaz A, Türkoğlu A, Haliloğlu K, Niedbala G, et al. Prediction of grain yield in wheat by CHAID and MARS algorithms analyses. *Agronomy*. 2023;13(6):1438. <https://doi.org/10.3390/agronomy13061438>
69. Demirel F, Uğur R, Popescu GC, Demirel S, Popescu M. Usage of machine learning algorithms for establishing an effective protocol for the in vitro micropropagation ability of black chokeberry (*Aronia melanocarpa* (Michx.) Elliott). *Horticulturae*. 2023;9(10):1112. <https://doi.org/10.3390/horticulturae9101112>
70. Yuan H, Yang G, Li C, Wang Y, Liu J, Yu H, et al. Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: analysis of RF, ANN, and SVM regression models. *Remote Sensing*. 2017;9(4):309. <https://doi.org/10.3390/rs9040309>
71. Kumar K, Kumar S, Sankar V, Sakthivel T, Karunakaran G, Tripathi P. Non-destructive estimation of leaf area of durian (*Durio zibethinus*)—An artificial neural network approach. *Sci Hort*. 2017;219:319–325. <https://doi.org/10.1016/j.scienta.2017.03.028>
72. Shabani A, Ghaffary KA, Sepaskhah AR, Kamgar-Haghighi AA. Using the artificial neural network to estimate leaf area. *Sci Hortic*. 2017;216:103–10. <https://doi.org/10.1016/j.scienta.2016.12.032>
73. Küçükönder H, Boyacı S, Akyüz A. A modeling study with an artificial neural network: developing estimation models for the tomato plant leaf area. *Turk J Agric For*. 2016;40:203–12. <https://doi.org/10.3906/tar-1408-28>
74. Kallel A, Makhloufi A, Ben Ali A. Olive tree health monitoring approach using satellite images and based on Artificial Intelligence: Satellite image for Olive tree health monitoring. In *Proceedings of the Federated Africa and Middle East Conference on Software Engineering*. 2022;June;18–25. <https://doi.org/10.1145/3531056.3531070>
75. Mhanna MA, Douay F, Rajab M. Identifying the Optimal Pollinizer for 'Dermlali' Olive Cultivar under Syrian Coast Conditions. *Int J Fruit Sci*. 2020;20(sup3):S1984–93. <https://doi.org/10.1080/15538362.2020.1844112>
76. Koubouris G, Bouranis D, Vogiatzis E, Nejad AR, Giday H, Tsaniklidis G, et al. Leaf area estimation by considering leaf dimensions in olive tree. *Sci Hortic*. 2018;240:440–5. <https://doi.org/10.1016/j.scienta.2018.06.034>
77. Asriani E. Application of artificial neural network with backpropagation algorithm for estimating leaf area. *IOP Conf Ser: Earth Environ Sci*. 2020;599(1):012046. <https://doi.org/10.1088/1755-1315/599/1/012046>
78. Kishore DK, Pramanick KK, Verma JK, Singh R. Non-destructive estimation of apple (*Malus domestica* Borkh.) leaf area. *The J Hort Sci Biotechnol*. 2012;87(4):388–90. <https://doi.org/10.1080/14620316.2012.11512881>
79. Kumar R, Sharma S. Simulation and validation of leaf area prediction model for *Picrorhiza kurroa* an endangered medicinal plant of Western Himalaya. *J Medic Plant Res*. 2013;7:1467–73.