



Evaluation of biochemical profiles in walnut (*Juglans regia* L.) genotypes and predictive modeling using advanced machine learning algorithms

Fatih Temizyürek¹ · Muhammet Ali Gündeşli² · Yavuz Canbay³ · Hacı Osman Özatar¹ · Esra Gölcü⁴ · Vahdet Cemil Altun³ · Ömer Atagül¹ · Murat Güney⁵ · Ebru Kafkas⁴ · Remzi Ugur² · Mehmet Sütyemez⁶

Received: 11 April 2025 / Revised: 31 May 2025 / Accepted: 23 June 2025 / Published online: 4 July 2025
© The Author(s) 2025

Abstract

Biochemical parameters such as phenolic compounds, tocopherols and polyunsaturated fatty acids (omega-6 and omega-3 etc.) found in walnuts support cardiovascular health with their high antioxidant content, reduce inflammation and improve the blood lipid profile and show positive effects such as lowering LDL cholesterol, increasing HDL and reducing oxidative stress. This study investigates the application of machine learning algorithms to predict the biochemical properties of 10 walnut (*Juglans regia* L.) genotypes selected from the Kahramanmaraş region and three commonly cultivated walnut cultivars (Maraş-18, Chandler, and Franquette). Plant material was collected between 2020–2021, and analyses were conducted on biochemical components, including polyphenols, antioxidant capacity, protein, oil, fatty acids, sugars, and organic acids using HPLC, GC, and spectrophotometric methods. The highest total phenolic content was found in genotype 46KM-5 (402.6 mg GAE/100 g), and the highest antioxidant capacity in genotype 46KM-13 (73.77%). The 46KM-9 genotype exhibited the highest oil content (71.60%) and the highest total sugar content (3.27%), while sucrose was the dominant sugar. Regarding fatty acids, genotype 46KM-19 had the highest linoleic acid content (60.88%), and genotype 46KM-21 showed the highest oleic acid content (20.88%). Malic acid emerged as the dominant organic acid, with the highest content observed in genotype 46KM-9 (6.55%). In addition, various machine learning algorithms such as decision tree, support vector regression, K-nearest neighbor (KNN), gradient boosting, multi-layer perceptron, and linear regression (LR) have been used to assess the biochemical properties of genotypes using hyperparameter optimization. KNN algorithm generally provided high performance in prediction processes by exhibiting very low error rates. This study's results align with previous research, highlighting walnut genotype variation and nutritional benefits. Machine learning predicts key traits, aiding breeders and consumers in health-focused production and better genotype selection for improved cultivation.

Keywords *Juglans regia* · Walnut · Fatty acids · Biochemical content · Machine learning

✉ Muhammet Ali Gündeşli
maligun4646@gmail.com

¹ East Mediterranean Transitional Zone Agricultural Research Institute, Kahramanmaraş, Türkiye

² Department of Plant and Animal Production, Nurdagi Vocational High School, University of Gaziantep, 27310 Gaziantep, Türkiye

³ Department of Computer Engineering, Faculty of Engineering, University of Kahramanmaraş Sütçü İmam, 46050 Kahramanmaraş, Türkiye

⁴ Department of Horticulture, Faculty of Agriculture, University of Cukurova, 01330 Adana, Türkiye

⁵ Department of Horticulture, Faculty of Agriculture, University of Yozgat Bozok, 66900 Yozgat, Türkiye

⁶ Department of Horticulture, Faculty of Agriculture, University of Kahramanmaraş Sütçü İmam, 46040 Kahramanmaraş, Türkiye

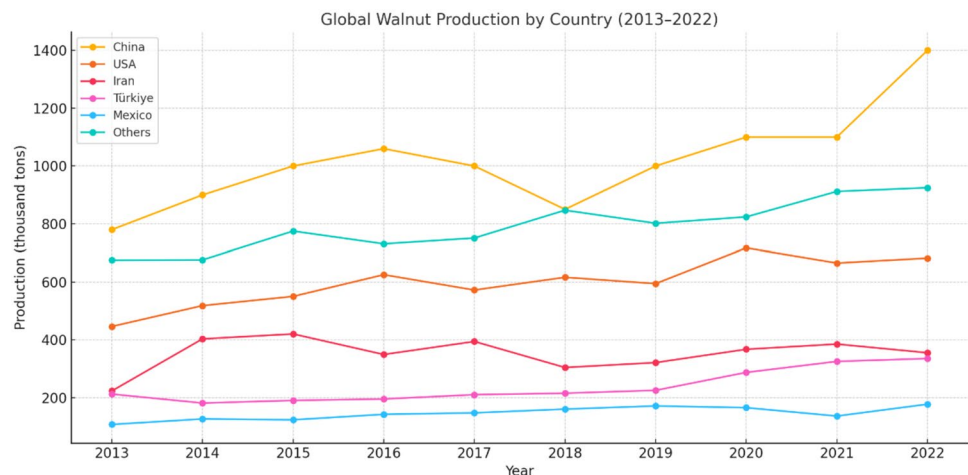
Introduction

Walnuts (*Juglans regia* L.) are a valuable nut known for their rich nutritional profile and health benefits. Anatolia, located at the intersection of the Iran-Turan, Mediterranean, and Euro-Siberian phytogeographical regions, is recognized as one of the world's most significant biodiversity hotspots [72]. It is not only rich in general plant diversity but is also considered one of the primary centers of origin and diversity for walnut (*Juglans regia* L.) [1, 52]. Among approximately 20 species within the genus *Juglans*, *J. regia* is the most economically important and widely cultivated species worldwide [16, 28, 80]. Known as the Anatolian or Persian walnut, *Juglans regia* L. is indigenous to Türkiye [67]. The diverse applications of walnuts have led to increased production in both the global and Turkish markets [25, 37, 54]. The world's walnut production stands at approximately 3.5 million tons. Türkiye, a key player in walnut cultivation, ranks fourth globally after China, the USA, and Iran, contributing around 9% of global production with 360,000 tons in 2023 [19]. In 2022, global walnut production reached 3.9 million tons. China ranked first with 1.4 million tons, followed by the United States in second place with 682 thousand tons, and Iran in third place with 355 thousand tons (Fig. 1). As of 2024, Turkey has a total of 17.5 million fruit-bearing walnut trees and 12.8 million non-bearing young walnut trees, indicating significant potential in walnut production. The presence of young orchards suggests that output is likely to increase in the coming years. The total walnut cultivation area spans 1,780,045 decares, with an annual production of 426,000 tons. The average yield per tree is calculated at 24 kg. (TÜİK, 2024). The world's walnut imports were 774 thousand tons in 2023. Turkey ranks first in world walnut imports with 99 thousand tons, while the UAE ranks second with 72 thousand tons and India ranks third with 62 thousand tons. Turkey carries out 13% of the world's walnut imports. In 2023, global walnut exports reached approximately 845

thousand tons. The United States led international exports with 223 thousand tons, followed closely by China with 218 thousand tons, and Chile with 143 thousand tons, placing them in the top three positions. The U.S. and China each represented around 26% of total global walnut exports, while Chile contributed about 17%, Mexico approximately 8%, and Türkiye held a 4% share in global export volumes. [19, 31].

Turkey, particularly its Eastern and Southeastern Anatolia regions, is recognized as an important center of genetic diversity for Persian walnut (*Juglans regia* L.) [28]. The Kahramanmaraş region in southeastern Türkiye offers favorable climatic and soil conditions for walnut cultivation. The region's specific temperature and precipitation patterns, soil characteristics, and altitude create an ideal environment for producing high-quality walnut cultivars [7, 27]. Recent research has focused on developing and characterizing new walnut cultivars in this region to meet growing demand and enhance the crop's economic value [69, 76], Koc et al. 2009, [27, 55]. Kahramanmaraş accounts for approximately 5.1% of Turkey's total fruit-bearing walnut trees and 6.1% of the country's total walnut cultivation area. However, the average yield per tree (23 kg) remains below the national average (24 kg). Despite its high production volume (20,898 tons), the region underperforms in productivity, which may be attributed to factors such as inadequate maintenance practices, suboptimal cultivar selection, insufficient irrigation, or the prevalence of aging orchards. Hakkâri, despite accounting for a relatively small share of Turkey's total walnut production, achieves the highest yield per tree (29 kg), significantly exceeding the national average (24 kg). This exceptional productivity may stem from the use of high-yielding cultivars, adoption of modern cultivation techniques, and/or favorable agroecological conditions. The region's performance highlights the potential of optimized practices in low-volume but high-efficiency production systems. Gaziantep contributes only 0.89% of Turkey's total

Fig. 1 . World walnut yield (kg/da)



walnut output (3812 tons), yet its yield per tree (26 kg) also surpasses the national average. This suggests intensive and efficient production within limited acreage, possibly due to targeted agronomic management or superior varietal selection TUIK, 2024).

Walnuts, a crucial component of the Mediterranean diet, are an important food source for many communities worldwide [30, 73]. The main factors that determine the nutritional value of walnuts are the variety and concentration of biochemical composition they contain. Therefore, Walnut (*Juglans regia* L.) is a strategic nut that contributes to human health in terms of both consumer health and commercial value, thanks to its high biochemical compound content. Critical parameters affecting the nutritional value and marketability of walnuts, almonds, and pistachios include their biochemical properties, fatty acid profiles, phenolic content, and antioxidant activity [24, 57]. In particular, components such as polyunsaturated fatty acids (especially linoleic and α -linolenic acid, omega-6, omega-3), tocopherols (especially γ -tocopherol), phenolic compounds and sterols increase the antioxidant capacity of walnuts and enable them to gain functional food qualities. In addition to its oil content, walnuts are also a valuable source of protein, contributing significantly to human nutrition [37]. The level of these compounds not only increases the nutritional value, but also directly affects the shelf life, flavor profile and marketability of the product, which supports health claims. In addition, biochemical content variations in different walnut genotypes can be optimized through selection and breeding studies to maximize the nutritional quality of commercial varieties. Therefore, quantification of bioactive components in walnuts is of critical importance in terms of developing high value-added products in the food industry.

In fact, biochemical composition differences between commercial walnut varieties are among the important criteria that determine the market adaptation capacity of genotypes; varieties with high γ -tocopherol and total phenolic content are in higher demand both in the domestic market and in export markets. Therefore, biochemical properties not only increase nutritional value but also play a critical role in determining commercial priorities in walnut breeding programs. Previous academic studies have emphasized the genetic and environmental factors influencing the biochemical composition of walnuts. Variations in fatty acid composition and tocopherol content among different walnut genotypes highlight the importance of genetic diversity in breeding programs [28]. Additionally, cultivation methods, post-harvest processes, and processing techniques can impact the nutritional quality of walnuts [2, 70, 78, 10, 18, 37, 77]. Unsaturated fatty acids, particularly omega-3 and omega-6, play, sugar and organic acids a significant role in human health by regulating lipid metabolism and reducing

inflammation. High polyunsaturated fatty acid content in walnuts has been shown to lower total plasma and LDL cholesterol levels, thereby preventing coronary heart disease. Tocopherols in walnuts are powerful antioxidants that protect cellular components from oxidative damage and contribute to overall health. Additionally, many biochemical compounds in walnuts exhibit significant antioxidant properties, enhancing their benefits against various human diseases, as supported by numerous epidemiological studies [50, 74, 13, 21, 26, 29, 64, 71, 75]. As a result, there are numerous epidemiological and clinical studies on the potential role of daily walnut consumption against many health problems, including heart disease and its predictability, reduction of lesions, cancer, outflow dysbiosis, and neurodegenerative infectious diseases (Lockyer et al. [47, 53, 34, 65]).

In this context, the present study aims to analyze the biochemical characteristics and fatty acid profiles of new walnut cultivar candidates developed in the Kahramanmaraş region of Türkiye. By employing advanced analytical techniques, this research seeks to provide a comprehensive evaluation of these new cultivars. It also examines the possible use of machine learning methods in predicting the biochemical content of genotypes. This methodology aims to develop more rapid and accurate prediction models than standard analytical approaches, significantly contributing to quality improvements and productivity enhancement in walnut cultivation. Due to the complexity and high dimensionality of walnut biochemical data, traditional statistical methods may often be insufficient to reveal nonlinear relationships between traits. Therefore, machine learning algorithms can provide a more practical solution by providing fast and accurate estimation of basic biochemical traits. Machine learning enables data-driven selection strategies by integrating genotype assessment and, as a result, contributes positively to the cultivation efficiency and variety development in walnut production.

Therefore, this study aimed to evaluate the biochemical composition, particularly the phenolic compounds, antioxidant capacity, and fatty acid profiles of promising walnut genotypes developed in the Kahramanmaraş region of Türkiye using machine learning-based modeling.

Materials and methods

Material

Experimental location description

Kahramanmaraş is located at 37°43' north latitude and 37°8' east longitude, at an elevation of approximately 900 m above

sea level. Kahramanmaraş is located in the Mediterranean climate region, it exhibits a transitional climate between the Mediterranean and Southeastern Anatolia regions. The Mediterranean climate is characterized by warm and rainy winters and hot and dry summers. Detailed weather data for the study years are given in Fig. 2.

Plant material

The walnut samples used in this study consist of 10 important walnut genotypes selected from Kahramanmaraş region. These genotypes were provided by the ‘Walnut Selection Breeding’ project supported by the General Directorate of Agricultural Research and Policies (TAGEM) carried out in Karamanmaraş. In addition, three important walnut cultivars Maraş-18, Chandler, and Franquette- that are widely grown were included in the study. The study was carried out in the central district of Kahramanmaraş province in 2022–2023 at an altitude of 500 m. All genotypes and varieties used in the study were grafted onto uniform seedling rootstocks to ensure consistency in root system effect. For each genotype and cultivar, samples were collected from three different trees, with 10 walnut fruits taken per tree, resulting in three replicates per genotype or cultivar. In total, 10 genotypes and 3 cultivars were included in the study, and the collected samples amounting to a total of 390 walnut fruits (13 genotypes/cultivars $\times$ 3 trees $\times$ 10 fruits) were used for the biochemical and physicochemical analyses. All sampled trees were 10 years old. Fruit maturity was determined based on visual and physical indicators, including full kernel development, hardening of the shell, and natural dehiscence of the green husk, which are considered standard markers of physiological ripeness in walnut. After harvesting, the samples were immediately transported to Çukurova University, Faculty of Agriculture, Department of Horticulture Laboratory and stored in cloth bags in a deep freezer at -80°C for detailed analysis.

Cultural practices, including irrigation, fertilization, plant protection, and pruning, were regularly performed in the orchards. A fertilization program was implemented with approximately 800 g of phosphorus and potassium and 1500 g of nitrogen per plant. Phosphorus and potassium were applied in a single dose in February, while nitrogen was applied in three equal portions over three months starting in March, with monthly intervals. Microelement applications were initiated at the end of April. Drip irrigation was conducted at approximately 15-day intervals from late May to late September.

Method

Preparation of samples for biochemical analysis

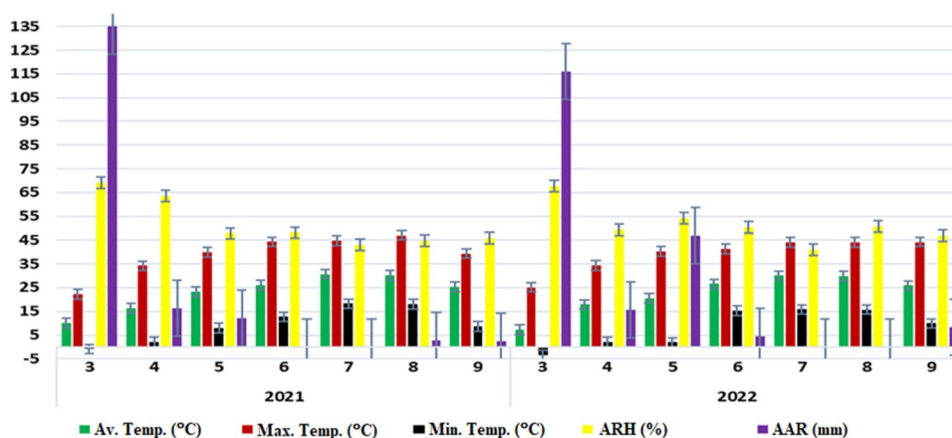
Following the pulverization of walnut samples using a grinder, 500 mg of the samples were placed in 15 mL tubes, and 2.5 mL of 80% methanol was added. The samples were thoroughly mixed and then centrifuged at 4°C at 4000 rpm for 10 min. After centrifugation, the samples were processed for biochemical parameter analysis[39].

Analyzing biochemical parameters

Analysis of total polyphenolics

Total phenolic content was measured using the modified Folin–Ciocalteu method as Spanos and Wrolstad (1990) described. Absorbance readings were taken at 765 nm using a spectrophotometer, and phenolic concentration was determined based on a calibration curve prepared with gallic acid. 50 μL aliquots were transferred from Falcon tubes to 2 mL Eppendorf tubes to prepare the samples. Then, 100 μL of Folin–Ciocalteu reagent and 1500 μL of distilled water were added and left to react for 10 min. Afterward, 50 μL of 20% Na_2CO_3 was added, and the samples were incubated in the dark for 2 h. Control samples and blanks were prepared

Fig. 2 Climatic data of the study area for 2022 and 2023, including average temperature, maximum temperature, minimum temperature, average relative humidity, and average annual rainfall



with ultrapure water, methanol, and the Folin–Ciocalteu reagent. Finally, 250 µL of each control, blank, and sample was pipetted onto the plate, and absorbance was measured at 765 nm using a Thermo Multi Scan Go spectrophotometer.

Analysis of total antioxidant capacity

Total antioxidant capacity was determined using the DPPH (2,2-diphenyl-1-picrylhydrazyl) radical scavenging assay as described by Brand-Williams et al. (1995). A 50 µL sample was transferred from Falcon tubes to 2 mL Eppendorf tubes, followed by the addition of 1950 µL of 0.06 mM DPPH solution under subdued light conditions. Each sample (250 µL) was loaded onto a plate for analysis. Controls and blanks were prepared using distilled water, methanol, and DPPH solution. Absorbance was measured at 515 nm using a Shimadzu spectrophotometer (Japan). The DPPH inhibition percentage (%) was calculated using the formula:

$$\begin{aligned} & \% \text{ DPPH inhibition} \\ &= \frac{(\text{control absorbance} - (\text{sample absorbance} - \text{blank absorbance}))}{\text{control absorbance}} \times 100 \end{aligned}$$

And the DPPH radical scavenging activity was calculated as:

$$\begin{aligned} & \% \text{ DPPH radical scavenging} \\ &= \frac{(\text{Control absorbance} - \text{sample absorbance})}{\text{control absorbance}} \times 100 \end{aligned}$$

Analysis of total oil content and oil extraction

The kernels were separated from the walnuts using a nutcracker and immediately homogenized using a laboratory blender to ensure uniform particle size before extraction. For oil extraction, 25 g of kernel powder was used, and oil was extracted using an automatic Soxhlet apparatus with hexane as the organic solvent. The extracted oil was methylated using boron trifluoride/methanol [32, 38]. The fat content was calculated as a percentage of the weight difference before and after extraction. The extracted fat was subsequently used for fatty acid analysis.

Analysis of total protein content

Kernel samples were dried uniformly at 30 °C for 24 h to prevent degradation of thermolabile compounds such as proteins and phenolics prior to biochemical analysis. Protein content was determined using the Kjeldahl method [36] on 0.25 g of ground walnut kernels. The percentage of protein was calculated by measuring the total nitrogen

content and applying a nitrogen-to-protein conversion factor ($K_c = 6.25, \% \text{ protein} = K_c * \% \text{ total N}$) [33].

Analysis of fatty acid composition

Fatty acid analysis was conducted using an Agilent GC equipped with a flame ionization detector (Agilent, 7820A). An SGE acid silicide salt tube (30 m length, 0.32 mm inner diameter, 0.25 µm film thickness, BP20 0.25 UM, USA) was used for the analysis. The oven temperature was initially set to 140°C for 5 min, then increased to 200°C at 4°C/min, and finally ramped to 220°C at 1°C/min. Standard fatty acid methyl esters (FAMES) containing 37 components (Supelco 37 FAME Mix, Sigma-Aldrich, Cat. No. 47885-U) were used as references to identify and quantify fatty acids.

Analysis of sugar content

Sugar content in walnut fruits was analyzed using a modified protocol based on the method by Bernardez et al. [9]. Briefly, 0.2 g of the fruit sample was combined with 2.5 mL of distilled water and incubated in an ultrasonic bath at 65°C for 30 min. The mixture was then centrifuged at 15,000 rpm for 20 min. After centrifugation, the supernatant was collected, and the process was repeated once. The supernatants were diluted to 5 mL with distilled water and filtered through a 0.45 µm nylon membrane. The filtered samples were then subjected to high-performance liquid chromatography (HPLC) for sugar content analysis. The individual sugars were separated using a CARBOsep COREGEL 87C column (7.8 × 100 mm) set at 75 °C, and detected with a refractive index detector connected to a chart recorder (Shimadzu, Kyoto, Japan). Deionized water (100%) was used as the mobile phase, and the flow rate was set at 0.3 mL/min. The sugars quantified from the fruit samples included glucose, fructose, sucrose, and total sugars.

Analysis of organic acid contents

Organic acids in walnut fruit samples were determined based on the method described by Bozan et al. [12], with slight modifications. For the extraction, 1 mL of homogenized fruit sample was mixed with 4 mL of 3% metaphosphoric acid. The mixture was placed in an ultrasonic water bath at 80 °C and sonicated for 15 min. After sonication, the samples were centrifuged at 5500 rpm for 15 min. The resulting supernatants were filtered through 0.45 µm nylon syringe filters (Whatman, 13 mm diameter) prior to HPLC analysis. The filtered extracts were analyzed using a high-performance liquid chromatography (HPLC) system (Shimadzu LC-20A VP, Kyoto, Japan) equipped with a UV detector (Shimadzu SPD-20A VP, Kyoto, Japan). Separation

was performed on an 87H column (300 mm × 7.8 mm I.D., 5 µm, Transgenomic). The mobile phase consisted of 0.05 mM sulfuric acid, with a flow rate of 0.8 mL/min. The column temperature was maintained at 40 °C, and the injection volume was set to 20 µL. Detection was carried out at 210 nm. Organic acids were identified by comparing their retention times and spectral data with those of analytical standards. Quantification was based on external calibration curves, and the results were expressed as mg per 100 g of sample.

Classical statistical analysis

The data obtained were subjected to variance analysis, and differences between means were determined using the LSD test. Differences among genotypes were tested using the least significant difference (LSD) method, and statistical significance was determined at the $p < 0.05$ level. The results were further evaluated using correlation, principal component, and heatmap analyses. All statistical analyses were performed using JMP Pro 14 (SAS Institute Inc., Cary, NC, USA).

Machine learning algorithms

Decision tree regression (DT) is a hierarchical regression algorithm composed of nodes and branches. Nodes with outgoing edges are referred to as internal nodes, whereas nodes with the predicted target values are termed leaves. This algorithm partitions the data into two by allocating input variables to branches and target variables to leaves [59]. Gradient boosting regression (GB) is an ensemble learning approach that iteratively combines multiple tree models. This method progressively trains weak predictors, allowing each model to learn from the errors of its predecessor, so creating a robust model. This tree-based approach seeks to systematically reduce margin errors to create a more resilient model [20, 56]. The K nearest neighbour regression (KNN) approach is a non-linear and efficient regression method that predicts outputs by analyzing the nearest K neighbors. The K value, a hyperparameter that determines the model's flexibility, significantly influences the model's performance [49].

Linear regression (LR) is a core element of machine learning, characterized by its simplicity, quick learning capacity, and considerable explanatory strength. This method assumes a linear relationship between dependent and independent variables [41]. The multi-layer perceptron (MLP) is a technique utilizing an artificial neural network design. It comprises an input layer, many hidden layers, and an output layer. It is capable of learning non-linear correlations. Each layer consists of unique nodes connected to the

following layer and contains neurons with nonlinear activation functions. This approach is generally trained using back-propagation techniques, yielding effective outcomes in challenging tasks [35]. Support vector regression (SVR) aims to forecast the target variable by optimizing the distance of a point from the hyperplane. If linear separation of the plane is impossible, linear divisions are executed by converting the data to a non-linear domain. Thus, it operates within a multi-dimensional polynomial framework. This model reduces the marginal error [6] [66].

Model setup and design

All the experiments were conducted on a Windows OS using Python 3.12 within the Jupyter Notebook environment. The implementation utilized standard libraries such as NumPy and Pandas for data preprocessing and manipulation. All machine learning models were executed on a system equipped with a 3.4 GHz CPU and 16 GB RAM.

This work outlines a workflow consisting of various stages. The stages in this workflow consist of data preprocessing, feature selection, model training, hyperparameter optimization, model testing and evaluations. The dataset is first normalized for effective comparison and reducing the effect of different scales. Then, it is divided into two training and testing. A k-fold cross-validation method was used to provide a reliable machine learning model ($k=5$). This method involves splitting the data into k subsets. Each subset is used for testing, while the remaining $k - 1$ subsets are used for training. Hyperparameter optimization has been performed using a method called Optuna. This approach makes a significant contribution by making effective and scalable hyperparameter optimization. Decision trees, support vector machine, K-nearest neighbors, gradient boosting, multi-layer perceptron, and linear regression were used for predictions. The effectiveness of these models has been evaluated using the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). We selected a single feature in each assessment as the dependent variable and the others as independent variables.

Results and discussion

Quantification of total phenolics and assessment of antioxidant capacity

Significant differences were observed among the tested walnut genotypes and cultivars regarding total phenolic content (Table 1, $p < 0.001$). The total phenolic content ranged from 138.51 mg to 402.64 mg GAE/100 g, indicating

Table 1 Total phenolics, antioxidant capacity, total protein content, and total lipid of the walnut cultivars and genotypes in the kernels of ten walnut genotypes

Genotypes/ cultivars	Total Phenolics (mgGAE/100g)	DPPH Radical (%)	Total Protein (%)	Total Fat (%)
46KM-5	402.64 a	63.51 g	15.41 lm	51.75 g
46KM-8	376.99 b	67.49 ef	14.49 h	64.19 d
46KM-9	348.38 c	72.64 b	17.97 d	71.60 a
46KM-13	293.43 d	73.77 a	20.05 a	50.24 g
46KM-15	254.71 f	67.96 de	14.07 h	65.00 cd
46KM-16	216.86 h	63.50 g	18.20 cd	64.16 d
46KM-17	216.15 h	70.48 b	17.96 d	51.50 g
46KM-18	191.52 k	69.33 c	14.42 h	54.92 f
46KM-19	138.51 l	66.36 ef	15.38 g	67.90 b
46KM-21	202.12 j	62.57 g	19.01 b	66.43 bc
Franquette	240.36 g	64.76 fg	18.60 bc	50.37 g
Chandler	210.19 i	59.70 g	15.98 f	50.34 g
Maraş-18	264.14 e	68.46 cd	13.23 i	60.77 e
LSD _{0.05}	5.80**	1.90**	17.02 e	2.11**

Values followed by different letters in the same column are significantly different according to LSD test at $p < 0.05$. ** indicates significant differences at the 1% level ($p < 0.01$)

considerable biochemical variability across the material studied. The highest phenolic concentration was recorded in genotype 46KM-5 (402.64 mg GAE/100 g), followed by 46KM-8 (376.99 mg GAE/100 g) and 46KM-9 (348.38 mg GAE/100 g), while the lowest was observed in genotype 46KM-19 (138.51 mg GAE/100 g). Among the standard cultivars, Chandler and Franquette displayed relatively low values of 210.19 mg and 240.36 mg GAE/100 g, respectively, falling within the mid-to-lower range of the studied material. Overall, the results show that genotypes were statistically distinguishable in terms of phenolic content, with most genotypes forming separate significance groups. Notably, only 46KM-16 and 46KM-17 were statistically similar in phenolic concentration.

Our findings are consistent with previous studies reporting moderate phenolic content in walnuts grown under Mediterranean conditions, such as Gündeşli et al. [27], who reported a range of 237.51–412.96 mg GAE/100 g. However, some studies have reported considerably higher values, including Pycia et al. [62] with 540.08–1067.81 mg GAE/100 g and Persic et al. (2018) with 750.67–1245.64 mg GAE/100 g. These discrepancies may arise due to differences in genotype, altitude, extraction methods, or analytical techniques. Consequently, we have emphasized comparisons with studies conducted in similar agroecological regions for better contextual relevance.

Likewise, significant variation was observed in antioxidant activity (DPPH radical scavenging capacity), ranging from 59.70% to 73.77%. The highest antioxidant activity

was found in genotype 46KM-13 (73.77%), followed by 46KM-9 (72.64%) and 46KM-17 (70.48%), while Chandler exhibited the lowest value (59.70%). Several genotypes, including 46KM-5, 46KM-16, 46KM-21, and the Franquette and 46KM-19 accessions, also displayed relatively lower antioxidant activity. Our results are in agreement with the findings of Gündeşli et al. [27], who reported DPPH values ranging from 54.65% to 71.99%, and Okatan et al. [54], who recorded values between 61.37% and 75.41%. The results further reveal that antioxidant capacity generally increased with phenolic content. This relationship supports the well-established notion that phenolic compounds significantly contribute to radical scavenging capacity in walnuts. Given the dual importance of high phenolic content and antioxidant capacity in terms of nutritional quality and breeding potential, genotypes such as 46KM-5, 46KM-9, and 46KM-13 stand out as promising candidates for further evaluation and potential commercial cultivation.

Quantification of protein and oil levels

The total protein content of the evaluated walnut genotypes and cultivars ranged from 13.23% (Maraş-18) to 20.05% (46KM-13), indicating a considerable variation among the materials studied ($p < 0.01$). The high protein concentration in 46KM-13 highlights its potential as a protein-rich genotype, which may offer nutritional advantages in human diets and functional food applications. In contrast, cultivars such as Maraş-18 and 46KM-15 exhibited relatively low protein levels. The average protein content across all genotypes was 16.21%, aligning with values reported in earlier studies (e.g., Savage, 2001; [5, 17, 37], Liu et al., 2020), which generally range between 14 and 19%. This consistency confirms the genetic stability of protein traits across diverse walnut backgrounds.

Regarding oil content, the genotypes also displayed significant variation ($p < 0.01$), ranging from 50.24% (46KM-13) to 71.60% (46KM-9). Genotype 46KM-9 was identified as the highest oil-yielding type, followed by 46KM-19 (67.90%) and 46KM-21 (66.43%). On the other end of the spectrum, standard cultivars such as Chandler (50.34%) and Franquette (50.37%) showed comparatively lower oil levels. Six of the studied genotypes surpassed 60% oil content, which is considered desirable for both nutritional and industrial applications. These findings are largely consistent with global reports that cite typical walnut oil content ranging between 52 and 72% (Mitrović et al., 1995; [11, 22, 37, 38, 62]).

The nutritional relevance of both protein and oil content in walnuts cannot be overstated. Protein-rich genotypes may be favored in breeding programs aimed at developing high-protein functional varieties, particularly in health-oriented

food products. Meanwhile, high-oil genotypes are valuable not only for caloric density but also for their unsaturated fatty acid profiles, contributing to cardiovascular health and oxidative stability. The observed variability offers crucial insights for genotype selection in breeding efforts targeting both nutritional enhancement and commercial adaptability.

Quantification of fatty acids

Table 2 presents the ratios of saturated fatty acids and total saturated fatty acids for various walnut cultivars and selected genotypes, highlighting statistically significant differences ($P < 0.01$). Notable variations were observed in the saturated fatty acid ratios among the walnut genotypes and cultivars, underscoring significant differences ($P < 0.001$) between them. However, no statistically significant differences were found among groups sharing the same letter designation. The highest myristic acid content, at 0.07%, was detected in genotypes 46KM-5 and 46KM-18, while genotype 46KM-21 exhibited the lowest myristic acid content at 0.02%. Genotypes 46KM-13, 46KM-16, and Chandler displayed moderate levels of myristic acid, ranging from 0.05% to 0.06%, while genotype 46KM-9 exhibited a lower content of 0.03% (Table 2). The highest palmitic acid content, 7.80%, was observed in genotype 46KM-18, whereas genotype 46KM-5 had the lowest palmitic acid content at 6.49%. Genotype 46KM-21 exhibited a palmitic acid content of 7.19%, while Franquette and Chandler genotypes displayed contents of 7.53% and 7.48%, respectively. Stearic acid content also varied among the genotypes, with 46KM-19

having the highest value (3.46%) and 46KM-18 the lowest (2.77%). Chandler (3.22%) and Maraş-18 (3.16%) also showed relatively high stearic acid levels (Table 2). These findings comprehensively understand the saturated fatty acid profiles of walnut genotypes and cultivars, highlighting significant variability.

Consuming high levels of saturated fatty acids is associated with potential health risks. Therefore, walnut genotypes and cultivars with lower saturated fatty acid levels may offer health advantages [11]. The results of this study indicate that genotypes 46KM-16 and 46KM-18 have higher levels of saturated fatty acids, while 46KM-9 and Maraş-18 exhibit lower levels. Consequently, walnut growers should prioritize genotypes with lower saturated fatty acid content for cultivation. Previous studies conducted in various countries, including Turkey, have consistently reported low levels of myristic acid in walnut cultivars, often below 0.05%. Palmitic acid is typically found in the range of 5–9%, and stearic acid ranges from 1 to 4%, as noted in several studies. Additionally, total saturated fatty acids in walnuts are commonly reported to range between 10–20% [43, 46, 48, 58], Tsao et al., 2007, [4], [63], [8], Şimşek 2016, [25, 27, 37, 68].

Our research findings are consistent with the general trends observed in these previous studies. Myristic acid levels were found to be low across all cultivars and genotypes. Palmitic acid levels in most cultivars and genotypes were either average or slightly higher than the reported range, indicating consistency with the broader literature. While stearic acid levels varied among genotypes and cultivars, they generally aligned with the expected outcomes. The variations in total saturated fatty acid content, ranging from 9 to 11%, are consistent with the values reported in previous walnut studies, supporting the reliability and comparability of our results with existing research.

The ratios of unsaturated fatty acids, including total unsaturated fatty acids, for various walnut cultivars and selected genotypes are presented in Table 3. Statistically significant differences are indicated ($P < 0.01$), with notable variations in both polyunsaturated fatty acid (PUFA) and monounsaturated fatty acid (MUFA) ratios among the genotypes and cultivars ($P < 0.001$). However, no statistically significant differences existed among groups sharing the same letter designation.

Linoleic acid and linolenic acid are the two main components of PUFAs. The highest linoleic acid content was found in genotype 46KM-19 at 60.88%, while genotype 46KM-8 exhibited the lowest content at 58.39%. The highest total PUFA content, 80.21%, was observed in genotype 46KM-9. Other genotypes, including 46KM-16, Chandler, and Maraş-18, also displayed high total PUFA contents (Table 3). Linoleic acid is commonly the most abundant fatty acid in

Table 2 Saturated fatty acids content of walnut (*Juglans regia* L.) walnut genotypes

Cultivars and genotypes	Saturated fatty acids (%)			
	Myristic acid C14:0	Palmitic acid C16:0	Stearic acid C18:0	Σ Saturated FA
46KM-5	0.07 a	6.49 h	3.15 cd	10.33
46KM-8	0.04 cde	7.74 ab	3.43 ab	9.93
46KM-9	0.03 de	7.26 c-f	3.10 cde	8.37
46KM-13	0.05 abc	7.32 cde	3.15 cde	9.29
46KM-15	0.05 a-d	6.90 fg	3.42 ab	9.34
46KM-16	0.06 abc	6.93 efg	3.05 cde	11.86
46KM-17	0.05 a-d	7.38 bcd	3.15 cd	9.77
46KM-18	0.07 a	7.80 a	2.77 f	10.39
46KM-19	0.06 ab	6.64 gh	3.46 a	9.46
46KM-21	0.02 e	7.19 c-f	2.94 def	9.58
Franquette	0.04 bcd	7.53 abc	2.88 ef	10.11
Chandler	0.06 abc	7.48 abc	3.22 abc	9.31
Maraş-18	0.05 a-d	7.01 d-g	3.16 bcd	8.77
LSD _{0.05}	0.02**	0.41**	0.26**	0.74

Values followed by different letters in the same column are significantly different according to LSD test at $p < 0.05$. ** indicates significant differences at the 1% level ($p < 0.01$)

Table 3 Unsaturated fatty acids content of walnut (*Juglans regia* L.) walnut genotypes

Cultivars and genotypes	Unsaturated fatty acids (%)					
	Polyunsaturated fatty acids			Monounsaturated fatty acids		
	Linoleic acid	Linolenic acid	Total PUFA	Oleic acid	Palmitoleic acid	Total MUFA
46KM-5	60.25 abc	9.53 bcd	72.47	19.23 b-e	0.25 ab	19.48
46KM-8	58.39 e	9.46 bcd	75.83	20.01 ab	0.15 ef	20.16
46KM-9	60.47 ab	9.24 d	80.21	18.53 def	0.12 fg	18.65
46KM-13	59.34 cde	10.16 ab	69.93	18.37 ef	0.21 cd	18.58
46KM-15	59.56 bcd	10.26 a	72.99	18.36 ef	0.20 cd	18.56
46KM-16	59.15 de	9.63 a-d	77.15	19.37 bcd	0.22 abc	19.59
46KM-17	59.52 bcd	9.99 abc	76.36	18.91 c-f	0.18 de	19.09
46KM-18	59.74 bcd	9.09 de	71.15	19.04 cde	0.26 a	19.30
46KM-19	60.88 a	9.22 d	72.99	18.00 f	0.24 ab	18.24
46KM-21	59.06 de	8.37 e	73.96	20.88 a	0.08 g	20.96
Franquette	59.51 bcd	9.16 d	72.23	19.75 bc	0.18 de	19.93
Chandler	59.36 cde	9.15 d	76.14	19.17 b-e	0.22 bc	19.39
Maraş-18	59.79 bcd	9.33 cd	76.32	19.07 cde	0.19 cd	19.26
LSD _{0.05}	0.96**	0.71**	70.75	0.90**	0.03**	0.64

Values followed by different letters in the same column are significantly different according to LSD test at $p < 0.05$. ** indicates significant differences at the 1% level ($p < 0.01$)

walnuts, typically found in the 50–65% range, according to numerous studies. Research from various countries, including Turkey, has reported linoleic acid contents ranging from 57.0% to 62.4% in walnut cultivars and genotypes.

In our study, genotype 46KM-19 (60.88%) exhibited an upper-range linoleic acid content, consistent with typically high levels reported in walnuts (Tsamouris et al., 2002; [5, 43], 51, 4, 11, 22, 27, 46, 60, 61, 62]. Previous studies have indicated that linolenic acid typically ranges from 8 to 15% [5, 43, 48], 50, 62, 8, 14, 37]. In Turkey, studies have reported linolenic acid content in walnut cultivars ranging from 10.0 to 15.0% [3, 17, 58]. Şimşek, 2016, 11, 25, 27, 37, 38, 55, 79]. According to our results, genotype 46KM-15 (10.26%) falls at the lower end of the range reported in the literature, suggesting that some walnut cultivars may contain this fatty acid at lower levels.

Oleic acid and palmitoleic acid are the main components of MUFAs. The highest oleic acid content was found in genotype 46KM-21 at 20.88%, while genotype 46KM-19 exhibited the lowest at 18.00%. Genotype 46KM-13 had the highest total MUFA content at 20.78%. Other genotypes, including 46KM-18, Franquette, and Maraş-18, also showed high total MUFA contents (Table 3). In various studies, oleic acid content has been reported to range from 12 to 28% (Beceanu et al., 2000; Savage, 2001; [5, 61, 63], Persic et al., 2018, 62].

Studies conducted in Turkey have reported that oleic acid content in Turkish walnut cultivars ranges from 14.0 to 24.5% [15, 44, 17], Beyazit and Sumbul, 2012, 27, 37, 38, 54]. Genotype 46KM-21 fits within this range, indicating that it is a cultivar rich in oleic acid. Regarding palmitoleic acid, genotype 46KM-18 falls within the range reported in

the literature, representing a typical level for this fatty acid, which is commonly found at lower concentrations.

Our results generally align with the existing data on walnut fatty acid profiles reported in the literature. Notably, cultivars such as 46KM-9 and 46KM-13 stand out with their high PUFA and MUFA contents. The data are consistent with unsaturated fatty acid profiles previously documented in walnut studies, supporting the reliability and comparability of our findings with earlier research. Due to their nutritional potential, these cultivars may be considered valuable genotypes for future studies (Table 3).

Quantification of carbohydrate composition

Tree nuts, especially walnut, almond, hazelnut, chestnut, and pistachio, are significant in human nutrition. The carbohydrates, or essential sugars, in these fruits are among the key components determining fruit quality [23]. Sugar content varies depending on genetic makeup, environmental factors, and cultivation conditions. These sugar components directly impact the taste and aroma of the fruit, and in walnuts, sugar content is particularly important as a commercial quality parameter. Studies on the sugar content of walnuts are limited in the literature. On the other hand, tree nuts' fructose, glucose, and sucrose contents vary depending on species, cultivar, genotype, ecological conditions, and climate.

This study investigated the carbohydrate (sugar) content of different walnut (*Juglans regia* L.) genotypes and cultivars, and the results are presented in Table 4. Statistically significant differences ($P < 0.05$) were observed among the walnut genotypes and cultivars regarding sucrose, glucose, fructose, and total sugar contents. Among the genotypes

Table 4 Carbohydrate content of walnut genotypes

Cultivars and genotypes	Carbohydrate (sugar) (%)			
	Sucrose	Glucose	Fructose	Total sugar
46KM-5	2.57 ab	0.23 f	0.23 f	3.03 ab
46KM-8	1.73 ef	0.31 de	0.31 cde	2.35 d
46KM-9	2.37 a-d	0.35 cd	0.30 cde	3.02 ab
46KM-13	2.55 ab	0.33 de	0.27 ef	3.15 ab
46KM-15	1.65 f	0.44 a	0.39 a	2.48 cd
46KM-16	1.97 c-f	0.3 abc	0.35 a-d	2.62 bcd
46KM-17	2.23 b-e	0.36 bcd	0.32 b-e	2.91 abc
46KM-18	2.47 abc	0.41 ab	0.37 abc	3.25 a
46KM-19	1.84 def	0.41 a	0.38 ab	2.63 bcd
46KM-21	2.45 abc	0.42 a	0.39 a	3.26 a
Franquette	2.79 a	0.29 e	0.30 de	3.38 a
Chandler	2.11 b-f	0.31 de	0.27 ef	2.69 bcd
Maraş-18	2.38 abc	0.34 cd	0.29 def	3.01 ab
LSD _{0.05}	0.30**	0.05**	0.04**	0.27**

Values followed by different letters in the same column are significantly different according to LSD test at $p < 0.05$. ** indicates significant differences at the 1% level ($p < 0.01$)

studied, the 46KM-21 genotype had the highest total sugar content (3.27%), while the lowest sugar content was observed in the 46KM-8 genotype (2.35%). Sucrose was the dominant sugar among the genotypes. The Franquette cultivar exhibited the highest sucrose content (2.79%), whereas the 46KM-15 genotype had the lowest (1.65%). In terms of glucose and fructose, the 46KM-19 and 46KM-21 genotypes showed the highest levels of these sugars. These findings indicate significant variations in the sugar profiles among walnut genotypes, which are primarily influenced by genetic factors (Table 4). The results are consistent with previous studies on tree nuts, highlighting the substantial variation in sugar content among walnut genotypes. Kazankaya et al. [40] conducted a study analyzing the sugar content of walnuts, almonds, and hazelnuts and found that fructose and sucrose were the dominant sugars in these nuts. Their research reported fructose levels ranging from 1.13 to 5.04%, glucose from 1.01 to 4.25%, and sucrose from 2.58 to 4.74% in different cultivars of tree nuts. These findings align with our study, where sucrose was the most prominent sugar across walnut genotypes, followed by glucose and fructose. Koç et al. [45] analyzed the sugar profiles of various walnut genotypes and identified malic acid and sucrose as the dominant components. Our study similarly found sucrose to be the predominant sugar in walnut genotypes, supporting previous research findings. Additionally, the effect of altitude on sugar metabolism was explored by Fuentealba et al. [21], who reported that the sugar content in walnuts increased at higher altitudes. In our study, the variations in sugar content among genotypes may also be influenced by environmental factors such as altitude and cultivation practices, which affect sugar metabolism and accumulation in walnut fruits. Overall, this research's results align with previous studies'

Table 5 Organic acids (%) of walnut genotypes in Türkiye

Genotypes	Organic acids (%)				
	Oxalic acid	Citric acid	Malic acid	Succinic acid	Total acid
46KM-5	0.24 a	1.52 b	4.67 cd	3.70 c	10.14 b
46KM-8	0.22 ab	1.38 c	3.80 f	0.81 h	6.22 f
46KM-9	0.21 abc	1.55 b	6.55 a	3.33 d	11.66 a
46KM-13	0.17 cde	1.66 a	4.82 bc	2.80 e	9.45 c
46KM-15	0.19 bcd	1.05 d	4.95 b	3.45 cd	9.66 c
46KM-16	0.24 a	1.50 b	4.47 d	1.09 g	7.30 e
46KM-17	0.12 f	0.09 e	1.37 i	1.74 f	4.15 h
46KM-18	0.17 de	0.91 e	3.78 ef	2.55 e	7.38 e
46KM-19	0.12 f	0.71 f	3.60 f	0.96 gh	5.41 g
46KM-21	0.05 g	1.10 d	2.8 h	1.12 g	5.18 g
Franquette	0.15 ef	0.92 e	3.80 ef	4.37 a	9.26 c
Chandler	0.22 ab	1.11 d	3.12 g	3.56 cd	8.03 d
Maraş-18	0.17 f	1.13 d	4.00 e	4.05 b	9.37 c
LSD _{0.05}	0.04**	0.09**	0.20**	0.26**	0.43**

Values followed by different letters in the same column are significantly different according to LSD test at $p < 0.05$. ** indicates significant differences at the 1% level ($p < 0.01$)

findings on the sugar content of tree nuts, demonstrating that walnut genotypes show significant differences in sugar profiles. The dominance of sucrose as the primary sugar in walnuts suggests its importance for flavor and commercial quality. Identifying genotypes with higher sugar content can have important commercial implications, especially for consumer preferences and market value.

Quantification of organic acid content

Organic acids play a crucial role in determining the taste, aroma, and overall quality of the fruit, and they show substantial variation between different genotypes. Genetic factors, environmental conditions, and cultivation practices influence the organic acid content. This study analyzed the organic acid content of different walnut (*Juglans regia* L.) genotypes and cultivars, and the results are presented in Table 5. Statistically significant differences ($P < 0.05$) were observed among the walnut genotypes and cultivars regarding oxalic acid, citric acid, malic acid, and succinic acid contents. The 46KM-9 genotype exhibited the highest total organic acid content (11.66%), primarily due to its elevated malic acid level (6.55%), which was also the highest among all genotypes. In contrast, the 46KM-17 genotype had the lowest total acid content (4.15%) and malic acid level (1.37%). Malic acid was the predominant organic acid across all genotypes. Significant differences were observed among genotypes for oxalic, citric, and succinic acid contents. Notably, the Franquette cultivar had the highest succinic acid content (4.37%), while the 46KM-21 genotype showed one of the lowest total acid contents (5.18%). There are not enough studies in the literature on the organic acid

content of walnuts. The results obtained are consistent with previous research on tree nuts, highlighting significant variations in organic acid content among walnut genotypes. Kazankaya et al. [40] studied the organic acid content of walnuts, almonds, and hazelnuts and reported malic acid as the dominant organic acid. Their research found that malic acid levels ranged between 3.00% and 7.33%, which aligns with the malic acid values observed in our study. Koç et al. [45] analyzed the organic acid profiles of different walnut genotypes and identified malic acid and succinic acid as the dominant components. Similarly, in this study, malic acid emerged as the most prominent organic acid, while succinic acid levels varied among genotypes. Erdoğan et al. [18] determined that the oxalic acid content in various walnut genotypes ranged from 5.51 to 7.56%, succinic acid from 5.67 to 9.64%, and malic acid from 0.23 to 0.66%. In a study conducted by Koç et al. [45] on different walnut genotypes, it was found that the contents of oxalic, succinic, malic, and fumaric acids ranged from 8.39 to 12.08 mg, 2.86 to 5.32 mg, 0.26 to 3.00 mg, and 0.26 to 0.58 mg, respectively. It has been determined that our results are consistent with the findings obtained by other researchers.

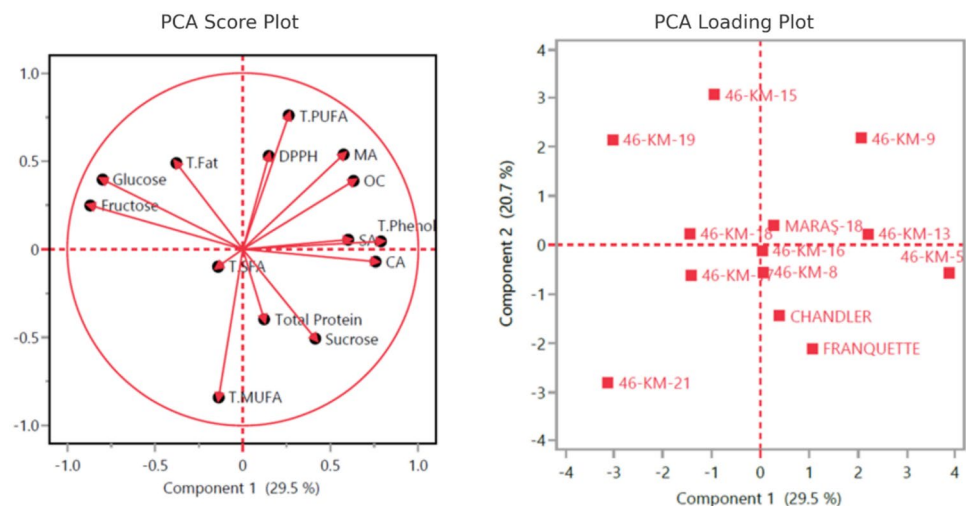
The walnut genotypes' phenolic contents, protein and fat ratios, and fatty acid compositions were analyzed using principal component analysis (PCA) (Fig. 3). A heat map analysis was also applied to cluster genotypes and reveal the major traits contributing to variance. PCA, a multivariate statistical approach, allows genotype classification based on measured biochemical characteristics.

In this study, the first two principal components, shown on the x and y axes, accounted for 29.5% and 20.7% of the total variance, respectively, with a combined explanation of 50.2%. The biplot revealed that sucrose, total protein, total phenolics, and citric acid (CA) contributed most to the first principal component (PC1), while total PUFA, DPPH, and malic acid (MA) were the strongest contributors to the second principal component (PC2).

The biplot revealed distinct groupings among the 13 walnut genotypes based on traits such as sugars, organic acids, phenolic content, antioxidant activity (DPPH), total fat, protein, and fatty acid composition (SFA, MUFA, PUFA). PC1 was primarily influenced by sucrose, total protein, total phenolics, and citric acid, while PC2 was mainly driven by PUFA, DPPH, and malic acid. These axes reflected key biochemical trade-offs between antioxidant content, carbohydrates, and lipid metabolism. Genotype 46KM-5 was separated from other genotypes due to its high levels of sucrose, protein, and phenolic content, while genotype 46KM-9 stood out for its association with DPPH activity and total PUFA content. The genotypes MARAŞ-18, 46KM-16, and 46KM-18 clustered closely, indicating biochemical similarity. These patterns were consistent with the clustering seen in the heat map analysis.

In addition to the main findings, PCA results provided deeper insights into the variability and grouping of walnut genotypes. Genotypes located in the negative region of PC1 and PC2, such as 46KM-21, Chandler, and Franquette, were characterized by lower levels of most biochemical components, suggesting their relatively limited nutritional potential compared to other accessions. On the other hand, genotypes such as 46KM-15 and 46KM-19 were positioned in the upper quadrants, indicating moderate contributions from organic acids and antioxidant activity. The close clustering of several local genotypes, including MARAŞ-18 and 46KM-16, reflects the presence of genetic or environmental uniformity, possibly due to shared cultivation conditions in the Kahramanmaraş region. Variables like glucose and fructose, positioned in the negative region of both components, were associated with a different set of genotypes, highlighting metabolic differences that could be relevant for flavor and consumer preference. Overall, the PCA plot effectively visualized the multidimensional relationships among traits and genotypes, supporting its application in genotype selection, breeding, and nutritional characterization (Fig. 3).

Fig. 3 PCA of walnut genotypes according to biochemical, sugar, organic acids, and fatty acids



Machine learning perspective and regression applications

Various regression models were used on the dataset used to estimate the biochemical characteristics of genotypes in experimental studies. The dataset comprises total phenolic content, DPPH radical inhibition, DPPH radical scavenging activity, protein rate, total fat, oleic acid, linoleic acid, linolenic acid, palmitic acid, palmitoleic acid, stearic acid, myristic acid, sucrose, glucose, fructose, citric acid, malic acid, succinic acid, and oxalic acid.

The results obtained show that the use of hyperparameter optimization techniques such as Optuna and the integration of reliable algorithms such as k-fold cross-validation are of vital importance for machine learning models. The results also showed that machine learning approaches have a high potential in analyzing complex biochemical data and can also produce important outputs (Table 6). These algorithms can be effectively used in this domain.

Table 6 demonstrates the overall results for the conducted experiments. The evaluations for the results are presented below.

The KNN method had the lowest error rates for all measures in comparison to DT, GB, LR, MLP, and SVR for Total Phenol. The KNN method demonstrates superior accuracy in estimating the associated feature. Furthermore, it is evident that GB and SVR had comparable performance, however LR demonstrated the greatest MSE and MAE results, indicating that LR is an inadequate choice for predicting this feature. Upon evaluating the DPPH_Inh traits, it was noted that the KNN algorithm exhibited the most effectiveness, demonstrating the lowest error rates, followed by the GB method. Nevertheless, the results indicated that the MLP and LR algorithms exhibited higher error rates. This suggests that these algorithms may necessitate more hyperparameter adjustment or feature engineering to improve the outcomes.

The predictive results for DPPH_Rdcl demonstrate that DT ranks the top and GB ranks second in terms of performance ratios. However, the SVR method has the highest error rates, indicating its insufficient capacity to capture the complexity of this feature. In predicting the Protein Rate feature, LR emerged as the most successful model, exhibiting the lowest error rates across all criteria. KNN and SVR demonstrated comparable outcomes, however DT and GB exhibited substantial error rates. The predictive outcomes for Total Fat indicated that KNN and DT yielded the most favorable results in terms of MSE, RMSE, MAE, and MAPE metrics. Furthermore, GB and MLP yielded exceptional outcomes, whereas SVR and LR exhibited relatively elevated error rates.

The results for the Oleic Acid trait demonstrated that SVR exhibited the lowest error rate, followed by MLP in second place. Although GB and LR displayed satisfactory outcomes, KNN revealed the greatest error rates. KNN is ineffective for precise predictions. MLP emerged as the most successful strategy with the lowest error rates for predicting Linoleic Acid. This instance demonstrates that the MLP algorithm effectively identifies patterns within the data. DT and LR demonstrated satisfactory performances, however KNN and GB exhibited the highest error rates.

The predictive outcomes for Linolenic Acid demonstrated that the MLP model was the most effective, with the lowest error values. Furthermore, the GB and KNN algorithms demonstrated satisfactory performance; however, the LR exhibited elevated error rates, making it inappropriate for predicting this feature. GB and KNN emerged as the optimal models, exhibiting the lowest error rates for predicting Palmitic Acid values. According to these algorithms, the SVR and MLP models yielded comparable results, however the LR exhibited higher error rates. The findings for Palmitoleic Acid indicated that LR and KNN surpassed the other models, providing the lowest error rates. Furthermore, MLP and DT emerged as the least effective algorithms, presenting elevated error rates, thereby making them less suitable for this trait.

In the prediction of Stearic Acid, the MLP exhibited the lowest error rates, with LR and KNN ranking second and third, respectively. In contrast, GB and DT demonstrated significant errors, indicating their inadequacy in predicting this feature. The findings for Myristic Acid indicated that all models had small error values due to the minor variation in this feature. Nevertheless, KNN yielded superior outcomes compared to the other models, exhibiting the lowest values for MSE, RMSE, and MAE. KNN demonstrated exceptional effectiveness in predicting the Sucrose feature with no error values. GB and MLP had elevated error rates, reflecting their restricted effectiveness for this characteristic.

The most precise Glucose prediction was attained by LR, which had the lowest error rates, following by MLP. KNN and DT demonstrated elevated error rates, hence constraining their predictive efficiency for this feature. DT and SVR are the most effective algorithms for precise Fructose prediction. Although these algorithms have the lowest values for MSE, RMSE, MAE, and MAPE, GB and LR showed higher error rates. Total Sugar was accurately forecasted by GB and SVR with negligible error rates. Although KNN showed robust performance, DT faced challenges in adequately representing this trait.

The prediction findings for Oxalic Acid indicated that KNN and DT produced the lowest errors, making them the most accurate predictive methods. LR exhibited robust performance; however, SVR displayed higher error rates. Citric

Table 6 Regression results of predicting various targets using different machine learning algorithms

Targets	Regression algorithms	MSE	RMSE	MAE	MAPE (%)
Total phenol	DT	2895.62847	53.40193	46.53686	14.20678
	GB	2594.43530	48.84923	40.18638	13.37343
	KNN	457.55039	19.92047	11.69039	4.16990
	LR	4880.02777	63.29504	52.50235	19.33827
	MLP	3338.12394	52.25551	39.09781	11.97547
	SVR	2664.60490	49.22936	40.83798	13.70892
Dpph-Inh	DT	6.37967	2.49356	2.04012	3.00802
	GB	3.18091	1.72819	1.27221	1.84483
	KNN	2.68104	1.54271	1.06224	1.57186
	LR	40.33702	6.60922	4.15281	6.18993
	MLP	37.80272	6.00411	5.25952	7.62650
	SVR	6.88587	2.54401	1.97977	2.90204
Dpph-Rdcl	DT	5.06339	2.16368	1.77706	2.63990
	GB	198.00505	7.60597	3.90647	2.99582
	KNN	1607.23680	38.14375	26.19351	28.35468
	LR	803.67026	26.85504	16.77779	17.63091
	MLP	1800.12651	36.39627	28.40371	43.66760
	SVR	1805.09780	38.00204	19.47204	14.63459
Protein rate	DT	774.02688	25.67936	19.69702	23.32464
	GB	1473.57471	34.04543	16.63014	11.82800
	KNN	0.60690	0.74866	0.63655	4.17907
	LR	0.30565	0.50118	0.36128	2.39056
	MLP	0.78628	0.82069	0.57439	3.72710
	SVR	0.64736	0.69810	0.52851	3.06573
Total fat	DT	0.78525	0.71582	0.53629	3.22442
	GB	1.33630	1.24855	0.91902	6.32308
	KNN	0.64218	0.72785	0.56625	3.34279
	LR	2.61468	1.69383	1.26269	2.00718
	MLP	1.79068	1.29030	1.01383	1.77331
	SVR	6.99614	2.48011	1.85555	3.25782
Oleic acid	DT	7.64980	2.41324	1.83688	3.06261
	GB	3.64797	1.80637	1.35711	2.43892
	KNN	14.47919	4.06065	3.55365	5.87835
	LR	3.13746	1.72134	1.32728	2.36174
	MLP	0.61978	0.73172	0.58430	2.98698
	SVR	0.52491	0.68961	0.56821	2.93350
Linoleic acid	DT	0.51546	0.67684	0.52286	2.68079
	GB	0.70069	0.77381	0.59859	3.20283
	KNN	0.80273	0.82660	0.66959	3.41254
	LR	0.50599	0.63373	0.55643	3.00599
	MLP	0.29248	0.51230	0.39512	2.04194
	SVR	0.93796	0.74423	0.74909	1.19803
Linolenic acid	DT	0.72787	0.78859	0.64543	1.07486
	GB	0.65047	0.74743	0.64800	1.08216
	KNN	0.62843	0.74615	0.58731	0.98448
	LR	0.77627	0.82897	0.68634	1.15296
	MLP	0.59412	0.71073	0.61499	1.04581
	SVR	0.76416	0.80455	0.67460	1.12979
Palmitic acid	DT	0.46559	0.69215	0.59600	6.21344
	GB	0.34156	0.56190	0.45256	4.95326
	KNN	0.34232	0.54294	0.44247	4.83061
	LR	0.59055	0.76307	0.62336	6.66571
	MLP	0.44441	0.63290	0.52725	5.65023
	SVR	0.36886	0.60388	0.48153	5.33567

Table 6 (continued)

Targets	Regression algorithms	MSE	RMSE	MAE	MAPE (%)
Palmitoleic acid	DT	0.29907	0.52202	0.41748	4.53167
	GB	0.27852	0.51254	0.43719	5.12848
	KNN	0.22687	0.45605	0.37854	5.47661
	LR	0.11037	0.30002	0.23287	3.43293
	MLP	0.34180	0.52506	0.42129	6.14608
	SVR	0.27289	0.48209	0.38515	5.72606
Stearic acid	DT	0.23464	0.45170	0.36913	5.36876
	GB	0.15885	0.35770	0.29377	4.35132
	KNN	0.00017	0.01272	0.01093	28.29065
	LR	0.00016	0.01218	0.00987	26.63444
	MLP	0.00009	0.00927	0.00734	19.79813
	SVR	0.00030	0.01477	0.01150	25.66130
Myristic acid	DT	0.00014	0.01166	0.00955	26.26817
	GB	0.00014	0.01174	0.00896	24.49668
	KNN	0.00011	0.01037	0.00789	21.33083
	LR	0.09178	0.27345	0.24589	7.46948
	MLP	0.05937	0.23505	0.19381	6.34855
	SVR	0.05762	0.23095	0.17191	5.68540
Sucrose	DT	0.00312	0.05528	0.04358	0.04392
	GB	0.24555	0.48960	0.42936	0.43674
	KNN	0.00000	0.00000	0.00000	0.00000
	LR	0.00082	0.02605	0.02247	0.02399
	MLP	0.18083	0.42500	0.34890	0.36519
	SVR	0.00000	0.00073	0.00063	0.00064
Glucose	DT	0.02106	0.13472	0.11588	14.88517
	GB	0.00171	0.04016	0.03135	5.61736
	KNN	0.24555	0.48960	0.42936	71.98556
	LR	0.00000	0.00000	0.00000	0.00000
	MLP	0.00015	0.01097	0.00901	1.15674
	SVR	0.18180	0.40434	0.35472	57.03708
Fructose	DT	0.00000	0.00072	0.00058	0.08801
	GB	0.02170	0.14301	0.11702	5.31057
	KNN	0.01301	0.11024	0.08618	4.06620
	LR	0.09492	0.29315	0.23854	10.99284
	MLP	0.00037	0.01747	0.01280	0.59026
	SVR	0.00030	0.01658	0.01352	0.61872
Total sugar	DT	0.08017	0.26691	0.22048	12.15508
	GB	0.00012	0.01002	0.00511	0.23349
	KNN	0.00109	0.03194	0.02759	8.66232
	LR	0.00111	0.03300	0.02668	8.39303
	MLP	0.00140	0.03669	0.02971	9.19610
	SVR	0.00023	0.01384	0.01045	3.08302
Oxalic acid	DT	0.00050	0.02133	0.01811	5.32511
	GB	0.00147	0.03931	0.02991	9.33414
	KNN	0.00010	0.00921	0.00541	1.68513
	LR	0.00091	0.02911	0.02499	8.40644
	MLP	0.00095	0.02977	0.02453	8.19373
	SVR	0.00187	0.04171	0.03443	11.72266
Citric acid	DT	0.00026	0.01529	0.01241	4.02880
	GB	0.00046	0.02117	0.01717	5.77018
	KNN	0.00180	0.03868	0.03232	11.37901
	LR	0.00019	0.01239	0.00766	2.52883
	MLP	0.02027	0.14093	0.11951	4.10100
	SVR	0.01568	0.12452	0.10322	3.62856

Table 6 (continued)

Targets	Regression algorithms	MSE	RMSE	MAE	MAPE (%)
Malic acid	DT	0.08942	0.29443	0.23273	8.23080
	GB	0.00037	0.01749	0.01283	0.44601
	KNN	0.00047	0.02080	0.01616	0.55471
	LR	0.06975	0.27292	0.20106	6.96288
	MLP	0.00008	0.00762	0.00418	0.14465
	SVR	0.00230	0.03928	0.03992	33.20356
Succinic acid	DT	0.00235	0.04811	0.03883	28.96098
	GB	0.00116	0.03334	0.02677	20.01455
	KNN	0.00245	0.04607	0.03655	20.64135
	LR	0.00144	0.03744	0.03267	24.69640
	MLP	0.00186	0.04679	0.03162	30.09702
	SVR	0.00134	0.03488	0.02905	23.16816

Acid was accurately predicted by LR and DT with negligible error rates. While GB exhibited robust performance, MLP faced challenges in precisely capturing this characteristic. The findings for Malic Acid demonstrated that MLP was the most precise model. DT and LR showed higher errors, indicating their constrained capability for this characteristic. The prediction results for Succinic Acid demonstrated that GB was the superior model, providing lower error rates, followed by SVR. DT and KNN reported higher error rates, highlighting potential limitations in forecasting this trait.

Conclusion

This study highlights the considerable variation in the biochemical composition of walnut genotypes and cultivars. The findings indicate that specific genotypes, such as 46KM-5, 46KM-9, and 46KM-13, exhibit superior biochemical profiles regarding phenolic content, antioxidant capacity, oil, sugar, and fatty acid content. These results suggest that selecting and cultivating these genotypes could improve walnut quality and meet consumer preferences. The study also confirms that environmental factors, such as altitude, play a role in shaping the biochemical profiles of walnuts. Future research should explore the relationship between these factors and walnut quality to optimize cultivation practices for enhanced nutritional benefits. Furthermore, integrating these high-performing genotypes into national walnut germplasm conservation programs will strengthen their long-term availability and support their adoption as commercial cultivars under changing climatic and market conditions. Ultimately, this study paves the way for targeted walnut breeding strategies, optimizing genotypes for enhanced health benefits and commercial viability in global markets. This study illustrates the potential use of machine learning for analyzing the biochemical traits of various walnut kinds. DT, SVR, KNN, GB, MLP, and LR were used for the machine learning algorithms in the experiments. Furthermore, the k-fold

cross-validation method was used for reliable results and to eliminate the problem of overfitting. Also, employing the Optuna method for hyperparameter optimization improved model accuracies by determining optimal configurations for each algorithm. The results presented that model performances vary based on the target trait. It can be seen that certain algorithms improve in predicting specific biochemical traits. SVR and GB algorithms significantly enhance the performance of detecting nonlinear relationships, but simple algorithms like LR effectively identify linear patterns. This work is highlighted as a supplementary resource in biochemical research in agriculture. In addition, applying sophisticated, comprehensive models and complex relationships across walnut genotypes can be revealed alongside machine learning and biochemical efforts. Machine learning algorithms played a critical role by enabling accurate, fast, and cost-effective predictions of these complex biochemical traits. The results obtained show that current technological approaches such as machine learning can make the analysis of large biochemical data sets easier and thus more effective in determining high-quality walnut genotypes. As a result, walnut breeders, researchers and producers who want to optimize nutritional value and market potential can benefit from better decision-making mechanisms. Integrating machine learning into agricultural biochemistry has great potential for genotype selection and product improvement efforts. Machine learning is an effective tool for agricultural applications. For instance, it may help breeders to identify high-performance genotypes in early stage which reduces time and costs of experimental trials, to select genotypes best suited for specific regions and real time estimation of walnut quality traits during processing or storage phases.

Author contributions FT, MAG, YC, HOÖ, EG, VCA, ÖA, MG, EK, RU, and MS conceptualized the experiment. FT, MAG, YC, HOÖ, EK, ÖA, and MG designed methodologies and analyzed data. RU, HOÖ, VCA, and MG performed laboratory analyses and contributed to data interpretation. HOÖ, and MAG assisted in experiment setup and data collection. All authors contributed to statistical analysis, result validation, and manuscript editing.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). We thank the General Directorate of Agricultural Research and Policies (TAGEM/BBAD/Ü/20/A1/P3/1728) for their financial support for this article.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akça Y, Şen SM (2001) A study on the genetic variability and selection of superior walnut (*Juglans regia* L.) trees within seedling population of around Van Lake. *Acta Hort* 544:119–124. <https://doi.org/10.17660/ActaHortic.2001.544.15>
- Akça Y, Şen SM (2001) Study on the selection of superior walnut trees in Hizan (Bitlis) populations. *Acta Hort* 544:115–118. <https://doi.org/10.17660/ActaHortic.2001.544.14>
- Akça Y, Kara H, Yazıcıgil Z, Öztekin Y, Özgen M, Sütyemez M, Kalyoncu L (2006) Fatty acid and dietary fibre content of walnut (*Juglans regia* L.) varieties grown in Turkey. *Asian J Chem* 18(2):1361–1365. <https://asianpubs.org/index.php/ajchem/article/view/21156/21105>
- Ali M, Ullah A, Ullah H, Khan F, Ibrahim SM, Ali L, Ahmad S (2010) Fruit properties and nutritional composition of some walnut cultivars grown in Pakistan. *Pak J Nutr* 9(3):240–244. <https://doi.org/10.3923/pjn.2010.240.244>
- Amaral JS, Casal S, Pereira JA, Seabra RM, Oliveira BP (2003) Determination of sterol and fatty acid compositions, oxidative stability, and nutritional value of six walnut (*Juglans regia* L.) cultivars grown in Portugal. *J Agric Food Chem* 51(26):7698–7702. <https://doi.org/10.1021/jf030451d>
- Apriyadi MR, Ermatita RDP (2023) Hyperparameter optimization of support vector regression algorithm using metaheuristic algorithm for student performance prediction. *Int J Adv Comput Sci Appl*. <https://doi.org/10.14569/IJACSA.2023.0140218>
- Aydemir Z, Özcan A, Aytekin M (2025) Production and marketing of walnut in Çağlayanerit District of Kahramanmaraş Province. *Black Sea J Agric* 8(2):194–204. <https://doi.org/10.47115/b.sagriculture.1624433>
- Bayazit S, Çalışkan O, Kılıç D (2020) Yükseltinin Chandler ceviz çeşidinde meyve kalite özelliklerine etkisi. *Gaziosmanpaşa Bilimsel Araş Derg* 9(2):124–132. <https://dergipark.org.tr/tr/download/article-file/1164078>
- Bernardez MM, Miguelez JM, Queijeiro JG (2004) HPLC determination of sugars in varieties of chestnut fruits from Galicia (Spain). *J Food Compos Anal* 17:63–67. [https://doi.org/10.1016/S0889-1575\(03\)00093-0](https://doi.org/10.1016/S0889-1575(03)00093-0)
- Beyhan Ö, Gözlekçi Ş, Gündoğdu M, Ercişli S (2016) Physicochemical and antioxidant characteristics in fruits of walnut (*Juglans regia* L.) genotypes from inner Anatolia. *Not Bot Horti Agrobo*. <https://doi.org/10.15835/nbha44210304>
- Beyhan O, Özcan A, Özcan H, Kafkas NE, Kafkas S, Sütyemez M, Ercişli S (2017) Fat, fatty acids and tocopherol content of several walnut genotypes. *Not Bot Horti Agrobo*. <https://doi.org/10.15835/nbha45210932>
- Bozan B, Başer KHC, Kara S (1997) Quantitative determination of naphthaquinones of *Arnebia densiflora* (Nordm.) Ledeb by an improved high-performance liquid chromatographic method. *J Chromatogr A* 782(1):133–136
- Bujdosó G, Konya E, Berki M, Nagy-Gasztonyi M, Bartha-Szuegyi K, Marton B, Adányi N (2016) Fatty acid composition, oxidative stability, and antioxidant properties of some Hungarian and other Persian walnut cultivars. *Turk J Agric For* 40(2):160–168. <https://doi.org/10.3906/tar-1501-20>
- Büyüksolak Z, Aşkın M, Kahramanoğlu I, Okatan V (2020) Effects of altitude on the pomological characteristics and chemical properties of Chandler walnuts: a case study in Uşak province. *Acta Agrobot*. <https://doi.org/10.5586/aa.7333>
- Çağlırım N (2003) Biochemical and physical properties of some walnut genotypes (*Juglans regia* L.). *Food Nahrung* 47(1):28–32. <https://doi.org/10.1002/food.200390004>
- Demir B, Sayıncı B, Cetin N, Yaman M, Cömlek R, Aydın Y, Sütyemez M (2018) Elliptic Fourier based analysis and multivariate approaches for size and shape distinctions of walnut (*Juglans regia* L.) cultivars. *Grasas Aceites* 69(4):1–12. <https://doi.org/10.3989/gya.0104181>
- Dogan M, Akgül A (2005) Fatty acid composition of some walnut (*Juglans regia* L.) cultivars from East Anatolia. *Grasas Aceites* 56(4):328–331. <https://doi.org/10.3989/gya.2005.v56.i4.101>
- Erdoğan Ü, Argin S, Turan M, Çakmakçı R, Olgun M (2021) Biochemical and bioactive content in fruits of walnut (*Juglans regia* L.) genotypes from Turkey. *Fresen Environ Bull* 30(6A):6713–6727. <https://hdl.handle.net/20.500.12428/3611>
- FAO (2024) Food and Agriculture Organization of the United Nations. *Crops data*. <http://www.fao.org/faostat/en/#data>
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232
- Fuentealba C, Hernández I, Saa S, Toledo L, Burdiles P, Chirinos R, Pedreschi R (2017) Colour and in vitro quality attributes of walnuts from different growing conditions correlate with key precursors of primary and secondary metabolism. *Food Chem* 232:664–672. <https://doi.org/10.1016/j.foodchem.2017.04.029>
- Gao P, Jin J, Liu R, Jin Q, Wang X (2018) Chemical compositions of walnut (*Juglans regia* L.) oils from different cultivated regions in China. *J Am Oil Chem Soc* 95(7):825–834. <https://doi.org/10.1002/aocs.12097>
- Guney M, Gundesli MA, Guney M, Kafkas S, Kafkas NE (2023) Exploring carbohydrate concentration fluctuations in pistachio (*Pistacia vera* L. cv. Uzun) for deeper insights into alternate bearing patterns. *Sustainability* 15(21):15300
- Guney M, Gundesli MA, Guney M, Kafkas NE, Kafkas S, Ergün Z (2024) Quantitative variation of phenolic compounds in different tissues of pistachio (*Pistacia vera* L. cv. Uzun) associated with alternate bearing. *Turk J Agric For* 48(4):539–556
- Gülsoy E, Pehlivan M, Şimşek M (2019) Determination of fatty acids, α -tocopherol, β -carotene, minerals, and some pomological properties of walnut genotypes selected from Aras Valley (Eastern Turkey). *Iran J Chem Chem Eng*. <https://doi.org/10.30492/IJCC.2019.31203>

26. Gündeşli MA (2020) Determination of sugar, total phenol content, and antioxidant activity of various parts of 'Uzun' pistachio cultivar (*Pistacia vera* L.). *Int J Agric Environ Food Sci* 4(1):62–69. <https://doi.org/10.31015/jaefs.2020.1.8>
27. Gündeşli MA, Uğur R, Yaman M (2023) The effects of altitude on fruit characteristics, nutrient chemicals, and biochemical properties of walnut fruits (*Juglans regia* L.). *Horticulturae* 9(10):1086. <https://doi.org/10.3390/horticulturae9101086>
28. Güney M, Kafkas S, Keles H, Zarifikhosroshahi M, Gündeşli MA, Ercisli S, Necas T, Bujdosó G (2021) Genetic diversity among some walnut (*Juglans regia* L.) genotypes by SSR markers. *Sustainability* 13:6830. <https://doi.org/10.3390/su13126830>
29. Haddad EH, Gaban-Chong N, Oda K, Sabaté J (2014) Effect of a walnut meal on postprandial oxidative stress and antioxidants in healthy individuals. *Nutr J* 13:4. <https://doi.org/10.1186/1475-2891-13-4>
30. Hayes D, Angove MJ, Tucci J, Dennis C (2016) Walnuts (*Juglans regia* L.): Chemical composition and research in human health. *Crit Rev Food Sci Nutr* 56(8):1231–1241. <https://doi.org/10.1080/10408398.2012.760516>
31. ITC, 2024. International Trade Center, TradeMap Veri tabanı, <https://www.intracen.org/>
32. Ichihara KI, Shibahara A, Yamamoto K, Nakayama T (1996) An improved method for rapid analysis of the fatty acids of glycerolipids. *Lipids* 31(5):535–539. <https://doi.org/10.1007/BF02522648>
33. James CS (1995) Analytical chemistry of foods. Springer, Berlin. <https://doi.org/10.1007/978-1->
34. Jarai D, Koller A (2024) Walnut consumption may contribute to healthy cardiovascular/endothelial function by maintaining membrane integrity. *Life* 14:1426. <https://doi.org/10.3390/life14111426>
35. Jin H, Kim YG, Jin Z, Rushchitc AA, Al-Shati AS (2022) Optimization and analysis of bioenergy production using machine learning modeling: Multi-layer perceptron, Gaussian processes regression, K-nearest neighbors, and artificial neural network models. *Energy Rep* 8:13979–13996. <https://doi.org/10.1016/j.egy.2022.10.334>
36. Kacar B (1984) Bitki Besleme Uygulama Kılavuzu. *Ankara Univ Ziraat Fak Yayınları* 900, Uygulama Kılavuzları:214, Ankara.
37. Kafkas E, Attar SH, Gundesli MA, Ozcan A, Ergun M (2020) Phenolic and fatty acid profile, and protein content of different walnut cultivars and genotypes (*Juglans regia* L.) grown in the USA. *Int J Fruit Sci* 20(sup3):S1711–S1720. <https://doi.org/10.1080/15538362.2020.1830014>
38. Kafkas E, Burgut A, Ozcan H, Ozcan A, Sütyemez M, Kafkas S, Türemis N (2017) Fatty acid, total phenol and tocopherol profiles of some walnut cultivars: a comparative study. *Food Nutr Sci* 8(12):1074. <https://doi.org/10.4236/fns.2017.812079>
39. Karen Slinkard K, Singleton VL (1977) Total phenol analysis: automation and comparison with manual methods. *Am J Enol Vitic.* 1977(28):49–55. <https://doi.org/10.5344/ajev.1977.28.1.49>
40. Kazankaya A, Doğan A, Piral KY, Encü T (2017) Bitlis yöresi ümitvar ceviz (*Juglans regia* L.) tiplerinin belirlenmesi. *Yüzüncü Yıl Univ Tar Bil Derg* 27:172–182
41. Kim SJ, Bae SJ, Jang MW (2022) Linear regression machine learning algorithms for estimating reference evapotranspiration using limited climate data. *Sustainability* 14(18):11674. <https://doi.org/10.3390/su141811674>
42. Koc A, Keles H, Ercisli S (2019) Some pomological properties of promising seed propagated walnut genotypes from inner Turkey. *Not Bot Horti Agrobo* 47(4):1094–1099. <https://doi.org/10.15835/nbha47411600>
43. Kornsteiner M, Wagner KH, Elmadfa I (2006) Tocopherols and total phenolics in 10 different nut types. *Food Chem* 98(2):381–387. <https://doi.org/10.1016/j.foodchem.2005.07.033>
44. Koyuncu MA, Koyuncu F, Yıldırım Akıncı F, Dilmaç Ünal T, Vural E (2005) Gelincik (Isparta) doğal ceviz genotiplerinin yan dal verimliliği ve meyve özelliklerinin belirlenmesi. *Bahçe Ceviz* 34(1):73–82. <https://dergipark.org.tr/tr/pub/bahce/issue/3349/46318>
45. Koç A, Keleş H, Ercişli S (2019) Some pomological properties of promising seed propagated walnut genotypes from inner Turkey. *Not Bot Horti Agrobo* 47(4):1094–1099
46. Li L, Tsao R, Yang R, Kramer JK, Hernandez M (2007) Fatty acid profiles, tocopherol contents, and antioxidant activities of heartnut (*Juglans ailanthifolia* var. cordiformis) and Persian walnut (*Juglans regia* L.). *J Agric Food Chem* 55(4):1164–1169. <https://doi.org/10.1021/jf062322d>
47. Lockyer S, de la Hunty AE, Steenson S, Spiro A, Stanner SA (2022) Walnut consumption and health outcomes with public health relevance—a systematic review of cohort studies and randomized controlled trials published from 2017 to present. *Nutr Rev* 81(1):26–54. <https://doi.org/10.1093/nutrit/nuac040>
48. Maguire LS, O'Sullivan SM, Galvin K, O'Connor TP, O'Brien NM (2004) Fatty acid profile, tocopherol, squalene, and phyto-sterol content of walnuts, almonds, peanuts, hazelnuts, and the macadamia nut. *Int J Food Sci Nutr* 55(3):171–178. <https://doi.org/10.1080/09637480410001725175>
49. Mailagaha Kumbure M, Luukka P (2022) A generalized fuzzy k-nearest neighbor regression model based on Minkowski distance. *Granul Comput* 7(3):657–671. <https://doi.org/10.1007/s41066-021-00288-w>
50. Martínez ML, Maestri DM (2008) Oil chemical variation in walnut (*Juglans regia* L.) genotypes grown in Argentina. *Eur J Lipid Sci Technol* 110(12):1183–1189. <https://doi.org/10.1002/ejlt.200800121>
51. Martínez ML, Mattea MA, Maestri DM (2006) Varietal and crop year effects on lipid composition of walnut (*Juglans regia* L.) genotypes. *J Am Oil Chem Soc* 83(9):791–796. <https://doi.org/10.1007/s11746-006-5016-z>
52. Nadeem MA (2021) Deciphering the genetic diversity and population structure of Turkish bread wheat germplasm using iPBS-retrotransposon markers. *Mol Biol Rep* 48:6739–6748. <https://doi.org/10.1007/s11033-021-06670-w>
53. Ni ZJ, Zhang YG, Chen SX, Thakur K, Wang S, Zhang JG, Wei ZJ (2021) Exploration of walnut components and their association with health effects. *Crit Rev Food Sci Nutr* 62(19):5113–5129. <https://doi.org/10.1080/10408398.2021.1881439>
54. Okatan V, Bulduk I, Kaki B, Gündeşli M, Usanmaz S, Alas T, SeyedHajizadeh H (2021) Identification and quantification of biochemical composition and antioxidant activity of walnut pollens. *Pak J Bot.* [https://doi.org/10.30848/PJB2021-6\(44\)](https://doi.org/10.30848/PJB2021-6(44))
55. Okatan V, Gündeşli MA, Kafkas NE, Attar ŞH, Kahramanoğlu İ, Usanmaz S, Aşkın MA (2022) Phenolic compounds, antioxidant activity, fatty acids, and volatile profiles of 18 different walnut (*Juglans regia* L.) cultivars and genotypes. *Erw-Obstbau* 64(2):247–260. <https://doi.org/10.1007/s10341-021-00633-y>
56. Otchere DA, Ganat TOA, Ojero JO, Tackie-Otoo BN, Taki MY (2022) Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *J Petrol Sci Eng* 208:109244. <https://doi.org/10.1016/j.petrol.2021.109244>
57. Özcan H, Güney M, Karıcı H, Tevfik H, Kafkas S, Acar İ, Kafkas E (2024) Exploring nut quality traits in almond varieties: a comparative study of local and foreign cultivars and their F1 hybrid offsprings. *Euphytica* 220(7):96
58. Ozkan G, Koyuncu MA (2005) Physical and chemical composition of some walnut (*Juglans regia* L.) genotypes grown in Turkey. *Grasas Aceites* 56(2):141–146. <https://doi.org/10.3989/gya.2005.v56.i2.122>

59. Pekel E (2020) Estimation of soil moisture using decision tree regression. *Theor Appl Climatol* 139(3):1111–1119. <https://doi.org/10.1007/s00704-019-03048-8>
60. Pereira JA, Oliveira I, Sousa A, Ferreira IC, Bento A, Estevinho L (2008) Bioactive properties and chemical composition of six walnut (*Juglans regia* L.) cultivars. *Food Chem Toxicol* 46(6):2103–2111. <https://doi.org/10.1016/j.fct.2008.02.002>
61. Popa VM, Hadaruga N, Gruia A, Raba DN, Moldovan C, Poiana AM (2011) Gas chromatography from the GC-MS analysis for the walnut grown in Romania. *J Agroaliment Process Technol* 17(3):261–265
62. Pycia K, Kapusta I, Jaworska G, Jankowska A (2019) Antioxidant properties, profile of polyphenolic compounds, and tocopherol content in various walnut (*Juglans regia* L.) varieties. *Eur Food Res Technol* 245:607–616. <https://doi.org/10.1007/s00217-018-3184-3>
63. Rabrenović B, Dimić E, Maksimović M, Šobajić S, Gajić-Krstajić L (2011) Determination of fatty acid and tocopherol compositions and the oxidative stability of walnut (*Juglans regia* L.) cultivars grown in Serbia. *Czech J Food Sci* 29(1):74–78. <https://doi.org/10.17221/180/2010-CJFS>
64. Rao G, Sui J, Zhang J (2016) Metabolomics reveals significant variations in metabolites and correlations regarding the maturation of walnuts. *Biol Open* 5:829–836
65. Rashki M, Ghasemzadeh Rahbardi M, Boskabady MH (2024) Nutritional advantages of walnut (*Juglans regia* L.) for cardiovascular diseases: a comprehensive review. *Food Sci Nutr* 13(1):e4526. <https://doi.org/10.1002/fsn3.4526>
66. Rodríguez-Pérez R, Bajorath J (2022) Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *J Comput Aided Mol Des* 36(5):355–362. <https://doi.org/10.1007/s10822-022-00442-9>
67. Şen SM (1986) Ceviz yetiştiriciliği. Eser Matbaası, Samsun.
68. Sarikhani S, Vahdati K, Ligterink W (2021) Biochemical properties of superior Persian walnut genotypes originated from south-west Iran. *Int J Hort Sci Technol* 8(1):13–24. <https://doi.org/10.22059/ijhst.2020.309363.392>
69. Simsek M (2010a) Determination of walnut genotypes with high fruit-bearing and quality in Dicle, Hani, Egil, and Kocaköy townships. *GOU J Agric Fac* 27(1):85–93. <https://search.trdizin.gov.tr/tr/yayin/detay/114278>
70. Simsek M, Osmanoglu A (2010b) Mazıdağı (Mardin) yöresindeki doğal cevizlerin (*Juglans regia* L.) seleksiyonu. *Yüzüncü Yıl Univ J Agric Sci* 20(2):131–137. <https://dergipark.org.tr/tr/download/article-file/204789>
71. Slatnar A, Mikulic-Petkovsek M, Stampar F, Veberic R, Solar A (2015) Identification and quantification of phenolic compounds in kernels, oil, and bagasse pellets of common walnut (*Juglans regia* L.). *Food Res Int* 67:255–263. <https://doi.org/10.1016/j.foodres.2014.11.016>
72. Sümbül A, Yıldız E, Nadeem MA (2023) Elucidating the genetic variations among Turkish grape varieties using morphological and molecular markers. *Genet Res Crop Evol* 70(5):1349–1361. <https://doi.org/10.1007/s10722-022-01503-6>
73. Tapia MI, Sanchez-Margado JR, Garcia-Parra J, Ramirez R, Hernandez T, Gonzales-Games D (2013) Comparative study of the nutritional and bioactive compounds content of four walnut (*Juglans regia* L.) cultivars. *J Food Compos Anal* 31:232–237. <https://doi.org/10.1016/j.jfca.2013.06.004>
74. Torabian S, Haddad E, Rajaram S, Banta J, Sabate J (2009) Acute effect of nut consumption on plasma total polyphenols, antioxidant capacity, and lipid peroxidation. *J Hum Nutr Diet* 22(1):64–71. <https://doi.org/10.1111/j.1365-277X.2008.00923.x>
75. Trandafir I, Cosmulescu S, Nour V (2017) Phenolic profile and antioxidant capacity of walnut extract as influenced by the extraction method and solvent. *Int J Food Eng* 13:1556–3758. <https://doi.org/10.1515/ijfe-2015-0284>
76. Ünver H, Sakar E (2014) Türkiye’de ceviz yetiştiriciliğinin durumu ve yapılan seleksiyon çalışmaları. *Harran Tarım Gıda Bil Derg* 15(3):61–59. <https://dergipark.org.tr/tr/pub/harranziraat/issue/18436/194142>
77. Yalçın B, Yıldırım A, Yıldırım F, Çelik C (2021) Bazı ceviz çeşitlerinin Isparta ekolojisinde biyokimyasal ve mineral madde içerikleri. *Adnan Menderes Univ Ziraat Fak Derg* 18(2):285–291. <https://doi.org/10.25308/aduziraat.993686>
78. Yıldız K, Akça Y, Ünver H, Oğuz Hİ (2017) Seçilmiş ceviz genotiplerine ait bazı meyve özelliklerinin değerlendirilmesi. *J Agric Fac Gaz Univ* 34(1):164–169. <https://doi.org/10.13002/jafag1040>
79. Yılmaz S, Akça Y (2017) Determination of biochemical properties and fatty acid composition of new walnut (*Juglans regia* L.) genotypes. *J Agric Fac Gaz Univ* 34(2):74–80. <https://doi.org/10.13002/jafag4203>
80. Zhao H, Bai H, Jing Y, Li W, Yin S, Zhou H (2017) A pair of taxifolin-3-O-arabinofuranoside isomers from *Juglans regia* L. *Nat Prod Res* 31(8):945–950. <https://doi.org/10.1080/14786419.2016.1258554>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.