

Article

Integrated Chemometric and Neural Network Analysis for the Differentiation of *Cucurbita maxima* and *Cucurbita moschata*

Milorad Miljić ^{1,*}, Biljana Lončar ², Biljana Kiprovska ³, Lato Pezo ⁴, Miloš Radosavljević ²,
Milenko Košutić ¹, Vesna Vasić ², Dragana Lukić ², Milena Rašeta ⁵ and Sanja Krstić ⁶

- ¹ Institute of Food Technology, University of Novi Sad, Bulevar Cara Lazara 1, 21000 Novi Sad, Serbia; milenko.kosutic@fins.uns.ac.rs
- ² Faculty of Technology Novi Sad, University of Novi Sad, Bulevar Cara Lazara 1, 21000 Novi Sad, Serbia; cbiljana@uns.ac.rs (B.L.); milosr@tf.uns.ac.rs (M.R.); vesna.vasic@uns.ac.rs (V.V.); dkukic@uns.ac.rs (D.L.)
- ³ Institute of Field and Vegetable Crops, National Institute of the Republic of Serbia, Maksima Gorkog 30, 21000 Novi Sad, Serbia; biljana.kiprovska@ifvcns.ns.ac.rs
- ⁴ Institute of General and Physical Chemistry, University of Belgrade, Studentski trg 12, 11000 Belgrade, Serbia; latopezo@yahoo.co.uk
- ⁵ Department of Chemistry, Biochemistry and Environmental Protection, Faculty of Sciences, University of Novi Sad, Trg Dositeja Obradovića 3, 21000 Novi Sad, Serbia; milena.raseta@dh.uns.ac.rs
- ⁶ Institute of Pharmaceutical Sciences, University of Graz, Beethovenstraße 8, 8010 Graz, Austria; sanja.krstic@uni-graz.at
- * Correspondence: milorad.miljic@fins.uns.ac.rs

Abstract

This study examines the compositional differentiation of two *Cucurbita* species, *C. maxima* Duchesne and *C. moschata* Duchesne, to identify chemical markers relevant for their nutritional and functional potential. Multivariate statistical analysis, including principal component analysis (PCA), was applied to chromatographic, chemical, and antioxidant descriptors to visualize patterns of variability among samples. Classification artificial neural network (cANN) models were used to explore the potential of machine learning for sample differentiation, using integrated lipidomic, carotenoid, phenolic, and liquid chromatographic datasets, providing a multidimensional biochemical characterization of *Cucurbita* samples, achieving good classification within the analyzed dataset, reflecting the model's capacity to describe the available data. The integration of chemometric and ANN approaches provides a framework for the compositional profiling and quality assessment of *Cucurbita* species, offering insights into their sustainable valorization as sources of bioactive compounds for food and nutraceutical applications while acknowledging the need for further validation on larger datasets.

Keywords: *Cucurbita maxima*; *Cucurbita moschata*; chemometrics; artificial neural networks; compositional profiling; multivariate analysis; bioactive compounds



Academic Editor: Ilaria Marotti

Received: 4 March 2026

Revised: 23 March 2026

Accepted: 25 March 2026

Published: 26 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Vegetables from the genus *Cucurbita* originate from Central and South America but are now intensively cultivated worldwide, with a total production of over 34 million tons according to FAOSTAT data for 2024, with China (62.1% production share) leading the way as the largest producer [1]. The genus *Cucurbita* belongs to the Cucurbitaceae family [2], and among the most important species for different industries are *Cucurbita maxima* Duchesne, *Cucurbita moschata* Duchesne, and *Cucurbita pepo* L. [3], with the focus of this work being on *C. maxima* and *C. moschata*.

The ripe fruits of these species are characterized by a high content of β -carotene [4–7], polyphenols [8–10], and dietary fibers [11,12]. Their consumption is associated with a reduced risk of cardiovascular disease, diabetes, and cancer [13–15], while the favorable nutritional composition of the pulp further suggests their importance in nutrition and processing [16,17].

The chemical complexity and bioactive richness of *Cucurbita* fruits present both opportunities and challenges for their nutritional and functional characterization. To address this, advanced data-driven approaches, including artificial intelligence (AI) and machine learning (ML), have increasingly been applied in the food sector to improve sample classification, quality assessment, and marker identification [18–21]. By leveraging collected chemical, chromatographic, and antioxidant data, models such as artificial neural networks (ANNs) can effectively distinguish food samples and identify key compositional markers, providing a powerful tool for valorization and quality control.

The classification artificial neural network (cANN) approach was selected in this study to address the challenge of analyzing complex, high-dimensional datasets generated by liquid chromatographic (LC), physicochemical, and antioxidant analyses of *Cucurbita* fruits. LC techniques generate extensive and highly reproducible datasets, providing precise numerical information for a wide range of samples [22]. Recent studies have demonstrated the effectiveness of cANNs in LC-based analyses, where machine learning algorithms are used to model relationships between chromatographic descriptors and to differentiate samples [23–25]. Similarly, physicochemical analytical methods produce large datasets with consistent quantitative descriptors suitable for advanced data analysis [26,27]. Owing to their ability to capture nonlinear relationships and complex interactions among variables, cANN models have increasingly been applied to such datasets, enabling the development of mathematical models that link compositional descriptors with different sample types and support reliable classification and differentiation [28,29].

Based on its proven efficacy in previous research [30], the cANN was selected in this study to identify chemical markers and compositional patterns that differentiate *Cucurbita maxima* and *Cucurbita moschata*, enabling robust sample classification and supporting the nutritional and functional valorization of these species. This approach allows us to capture nonlinear relationships among multiple descriptors that cannot be fully addressed by traditional statistical methods. This study aimed to develop a novel integrative approach for differentiating the tested accessions of two *Cucurbita* species, *C. maxima* and *C. moschata*, by combining previously published chemical data (e.g., high-performance liquid chromatography, HPLC; liquid chromatography–tandem mass spectrometry, LC-MS/MS; and antioxidant assay results) with a classification artificial neural network (cANN) model as an advanced analytical tool. Although individual chemical and antioxidant analyses of *Cucurbita* fruits have been reported, no previous study has integrated these datasets with machine learning models to systematically classify samples and identify key chemical markers. By addressing this methodological gap, the present study provides a framework for the robust compositional profiling, sample differentiation, and valorization of *Cucurbita* species in food and nutraceutical applications.

Previous studies have applied chemometric and machine learning methods to various food matrices, but the application of classification artificial neural networks (cANNs) for the integrated analysis of chromatographic, physicochemical, and antioxidant datasets in *Cucurbita* samples has received little attention. This study aims to explore the potential of the cANN as an advanced tool for the differentiation and classification of *Cucurbita* species based on multidimensional compositional data.

This study focused on evaluating the cANN for food classification in a different context. The specific objectives were to:

(1) Apply classical multivariate statistical analyses (cluster analysis—CA; principal component analysis—PCA) to differentiate the accessions of two *Cucurbita* species.

(2) Assess the ability of the cANN model to classify the two *Cucurbita* species accurately, using HPLC data, LC-MS/MS data, and antioxidant assay data.

2. Materials and Methods

2.1. Samples

Twenty *Cucurbita* accessions (Table 1), including eleven *C. maxima* and nine *C. moschata*, were obtained from the collection of the Institute of Field and Vegetable Crops, Novi Sad, Serbia. The samples were produced at the Department of Vegetable and Alternative Crops in Bački Petrovac (45°34' N; 19°67' E) under a conventional cultivation system.

Table 1. *Cucurbita* accessions included in this study.

<i>C. maxima</i>	<i>C. moschata</i>
Max 113 *	Mo 31 *
Max 118-1	Mo 39
Max 118-2	Mo 2
Max 117	Mo 29
Max 37	Mo 29-1
Hokk 54-1	Mo 29-2
Max 35	Mo 38
Max 65-1	Mo 32
Max 1	Mo F1 35 × 38
Max 105	-
Max 122	-

* The sample names (e.g., Max 113, Mo 31) represent official accession codes from the breeding collection of the Institute of Field and Vegetable Crops, Novi Sad, Serbia.

Each accession was cultivated on a separate plot, representing an independent experimental unit. Plants were grown under standard agronomic conditions typical for Serbian production, with a 5 m inter-row spacing. Mineral fertilization was applied, while irrigation was not required due to sufficient soil moisture, and no pesticides were used as no significant disease occurrence was observed. Weed control included two inter-row cultivations and manual removal within rows.

Sowing was conducted in early May, and fruits were harvested between late September and late October 2018. For each accession, 5–8 fruits were collected from the same plot and pooled to obtain a representative composite sample for further analyses.

This study was designed as a comparative compositional analysis of different accessions rather than a replicated agronomic field trial.

2.2. Analytical Methods

All analytical procedures were performed under controlled and standardized conditions, with minor modifications to previously published protocols to ensure consistency across all samples. All analyses were conducted on homogenized composite samples obtained for each accession, as described in Section 2.1. Depending on the analytical method, either fresh or lyophilized material was used, following the requirements of the respective protocols. For each accession, the same composite sample was used consistently across all analytical approaches (lipidomic, carotenoid, phenolic, and antioxidant analyses) to ensure the comparability of the obtained datasets. All measurements were performed in analytical triplicates, and the results are reported as the mean values used for subsequent chemometric and ANN analyses.

Lipophilic compounds were extracted from lyophilized *Cucurbita* samples (200 mg) using a solvent mixture of tert-butyl methyl ether and aqueous methanol (1:1, *v/v*), followed by ultrasound-assisted extraction and centrifugation. The extracts were analyzed using a UHPLC system (Vanquish, Thermo Scientific, Waltham, MA, USA) coupled to a Q Exactive™ Focus Hybrid Quadrupole-Orbitrap mass spectrometer equipped with a HESI-II probe (Thermo Scientific). Separation was achieved on a BEH C18 column (1.7 μm ; 2.1×100 mm) maintained at 40 °C under gradient elution. The mobile phases consisted of (A) water/methanol (95:5, *v/v*) and (B) isopropanol/methanol/water (65:30:5, *v/v/v*), both containing 5 mM ammonium formate and 0.1% formic acid. Data were acquired in both positive and negative ionization modes. Full MS scans were recorded at a resolution of 70,000 (*m/z* 200), while data-dependent MS/MS (Top 3) was performed at a resolution of 17,500. Quality control (QC) samples were periodically injected throughout the sequence. Detailed gradient conditions are provided in Table S1 (Supplementary Materials). Compound annotation was performed using MS-DIAL (version 4.38) and MS-Finder software (version 3.50), as described by Miljić et al. [31].

Carotenoids were extracted from fresh sample material using n-hexane:dichloromethane (1:1, *v/v*), followed by centrifugation and solvent evaporation. The residues were re-dissolved in methanol and analyzed using an Agilent 1200 series HPLC system (Agilent Technologies, Santa Clara, CA, USA) equipped with a diode array detector (DAD) and an Eclipse Plus C18 column (5.0 μm ; 3.0×250 mm). The mobile phases consisted of (A) acetone/water (75:25, *v/v*) and (B) acetone/methanol (75:25, *v/v*). The gradient program and additional chromatographic conditions are provided in Table S2 (Supplementary Materials). Detection was performed at 460 ± 4 nm, and full spectra were recorded in the range of 350–600 nm. Identification was based on retention times and UV-Vis spectral characteristics, and quantification was expressed as β -carotene equivalents using external calibration, according to the methods described by Miljić et al. [31] and previously established protocols [32,33].

Phenolic compounds were extracted from lyophilized samples using methanol and analyzed by LC-MS/MS using an Agilent 1200 series Rapid Resolution liquid chromatograph coupled with an Agilent 6410A (Agilent Technologies, Santa Clara, CA, USA) Triple Quadrupole mass spectrometer equipped with an electrospray ionization source. Data acquisition was performed in selected reaction monitoring (SRM) mode. Separation was achieved on a Zorbax Eclipse XDB-C18 column (1.8 μm ; 50×4.6 mm) maintained at 50 °C, with an injection volume of 5 μL . The composition of the mobile phases, gradient program, and SRM transitions are provided in Table S3 (Supplementary Materials). Quantification was carried out using external calibration curves prepared from serial dilutions of authentic standards, following the method described by Krstić et al. [34].

In vitro antioxidant capacity was determined using ABTS (2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)) and FRAP (Ferric Reducing Antioxidant Power) spectrophotometric assays. The results were expressed as mM Trolox equivalents (TXE) and mM ascorbic acid equivalents (AAE), respectively, based on calibration curves constructed with appropriate standards. All assays were performed according to previously described protocols [31], with all samples analyzed under identical experimental conditions.

2.3. Correlation Analysis (Color Correlation Analysis)

Pearson correlation analysis was applied to evaluate linear relationships among the measured chromatographic, chemical, and antioxidant variables. The Pearson correlation coefficient (*r*) and the corresponding coefficient of determination (*r*²) were used to assess the strength of associations between variables and the proportion of variability explained by these relationships [35,36].

The strength of correlations between individual parameters was interpreted according to the guidelines proposed by Guilford [37] for the magnitude of the correlation coefficient.

No correction for multiple testing was applied to the pairwise correlation analyses, as the primary goal was the exploratory assessment of relationships among compositional and antioxidant descriptors. All correlation coefficients are reported with their associated *p*-values, and future studies with larger datasets should incorporate multiple testing corrections, such as Bonferroni or false discovery rate adjustments, to ensure robust statistical inference.

The color correlation diagram for the average values of the carotenoid and phenolic contents, antioxidant potential (ABTS and FRAP assays), and the terpenoid profile of the analyzed *Cucurbita* accessions was created using R software version 4.0.3 (64-bit). The graphical representation of the correlations between the analyzed samples was generated using the *corrplot* function with the “circle” method. Circle color indicates the direction and magnitude of the correlation coefficient (from negative—red to positive—blue), while circle size is proportional to the absolute value of the correlation, with larger circles representing stronger relationships and smaller circles indicating weaker associations.

2.4. Principal Component Analysis (PCA)

Principal component analysis (PCA) was applied to the chromatographic, chemical, and antioxidant descriptors measured in the *Cucurbita* samples to reduce data dimensionality and identify relationships among variables and samples [38–40]. Prior to analysis, all data were standardized to dimensionless values (z-scores) to ensure comparability between variables with different measurement scales [41]. All input variables, including liquid chromatographic, lipidomic, carotenoid, phenolic, and antioxidant parameters, were scaled using min-max normalization to map each variable to the [0, 1] range. PCA was performed using the correlation matrix, and principal components with eigenvalues greater than 1 were retained according to the Kaiser criterion. The resulting components enabled the visualization of sample grouping patterns and the assessment of the contribution of individual variables to the observed variability [42–44].

It should be noted that neither Centered Log-Ratio (CLR) nor Isometric Log-Ratio (ILR) transformation was applied to the compositional manuscript-derived data prior to PCA, correlation, or ANN analysis. Instead, standardized values were used for exploratory pattern recognition. This approach does not fully address compositional constraints, and future work with larger datasets should incorporate log-ratio transformations to improve the robustness and interpretability of multivariate analyses.

2.5. Cluster Analysis

Cluster analysis is used to group similar samples into the same cluster based on the variables that describe the analyzed samples. The data were normalized using the z-score technique. The complete linkage method was applied for clustering, while the differences between samples were represented using the Manhattan distance [44]. Cluster analysis was performed using the complete linkage method with Manhattan distance as the similarity measure. The complete linkage approach emphasizes separation between clusters, producing compact and well-differentiated groups, while Manhattan distance is suitable for multivariate datasets of standardized variables and is less sensitive to extreme values.

Both principal component analysis and cluster analysis were performed using StatSoft Statistica ver. 10 software (StatSoft Inc., Tulsa, OK, USA, 2010).

2.6. Artificial Neural Network (ANN) Models

Due to the algorithm requirements used in ANN computation, input and output data are typically normalized using Equation (2), where q_i represents the *i*-th measured value of

the carotenoid and phenolic contents, antioxidant potential, and the terpenoid profile of the analyzed *Cucurbita* accessions.

$$q_{i \text{ norm.}} = \frac{q_i - (q_i)}{(q_i) - (q_i)} \quad (1)$$

The neurons in the network are organized into two or more layers, where the first layer is the input layer, the last one is the output layer, and the layer(s) between them represent the hidden (inner) layer(s) (Figure 1).

In this paper, a multi-layer perceptron (MLP) structure was used, consisting of three layers: input, hidden, and output.

In the ANN model developed in this paper, used to predict the contents of carotenoids and phenolics, terpenoid profiles, and antioxidant potential of the analyzed *Cucurbita* accessions, hyperbolic tangent functions were applied as activation functions in both the hidden and output layers.

To estimate the required number of neurons in the hidden layer, Equation (2) was used, where x and y represent the number of neurons in the input and output layers, n denotes the number of neurons in the hidden layer, and m is the number of weight coefficients, i.e., the number of connections between layers in the neural network [45–47]:

$$m = n \times (x + 1) + y \times (n + 1) \quad (2)$$

The parameter m represents the number of samples used for network training divided by 10.

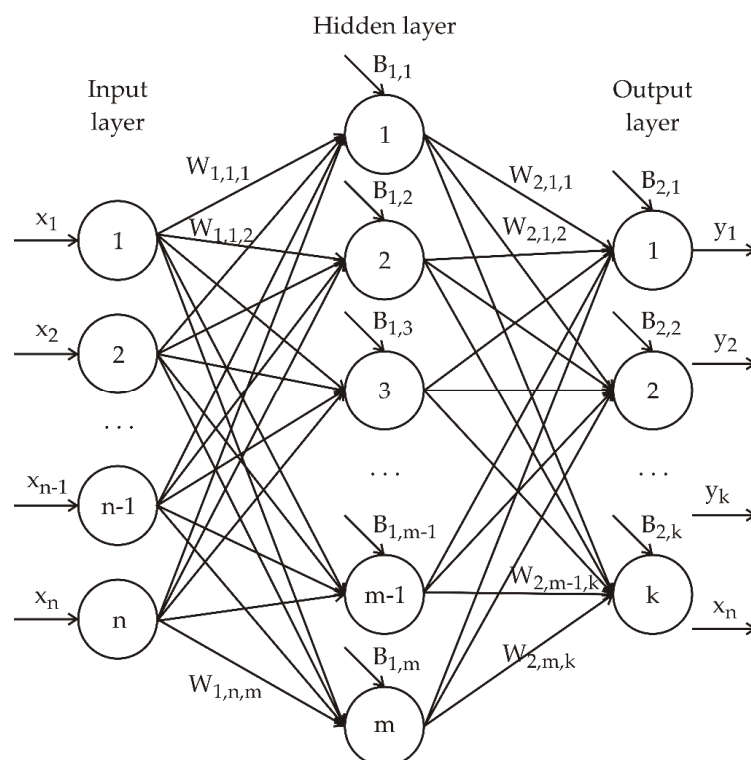


Figure 1. Schematic representation of three-layer neural network [46].

The resulting ANN model can be implemented through an algebraic system of equations, enabling the prediction of carotenoid, phenolic, and terpenoid contents, as well as antioxidant potential, in the analyzed *Cucurbita* accessions. This is achieved by substituting the corresponding weight coefficients and other parameters into the matrices within

Equation (3). The weight coefficients related to the hidden layer, together with computational errors, are grouped in matrices W_1 and B_1 , while those corresponding to the output layer are found in matrices W_2 and B_2 . If Y is the matrix of output variables, f_1 and f_2 are the transfer (activation) functions in the hidden and output layers, respectively, and X is the matrix of input variables, the artificial neural network (ANN) can be mathematically represented as follows [48]:

$$Y = f_1(W_2 \times f_2(W_1 \times X + B_1) + B_2) \quad (3)$$

The weight coefficients of the neurons' transfer functions, contained in matrices W_1 and W_2 , are determined during the network training process using optimization algorithms aimed at minimizing the differences between the values predicted by the network and the experimentally obtained results. Errors are quantified using the sum of squares (SOS) method, while optimization is performed through the BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm [49].

2.7. Classification Artificial Neural Network (cANN)

This study employed a multi-layer perceptron (MLP) model with input, hidden, and output layers, leveraging its well-established capability to approximate complex nonlinear relationships [50]. The classification artificial neural network (cANN) was trained using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, with both input and output data normalized to enhance model performance. The experimental dataset, comprising 20 *Cucurbita* accessions characterized by liquid chromatographic, lipidomic, carotenoid, phenolic, and antioxidant variables, was randomly divided into training (70%) and testing (30%) subsets. Additionally, five-fold cross-validation was applied to mitigate overfitting and provide a more robust evaluation of model performance.

Various network topologies were evaluated, including MLP 351-8-2 (for liquid chromatography data), MLP 10-12-2 (antioxidant data), and MLP 44-20-2 (lipidomic data) corresponding to the number of input neurons (matching the number of variables in each dataset), hidden neurons (selected based on trial and error and the heuristic formula $H = 2/3 I + O$, where H is the number of hidden neurons, I the number of input neurons, and O the number of output neurons), and two output neurons representing the *C. maxima* and *C. moschata* classes. The training process involved up to 100,000 epochs with randomly initialized weights and biases, a learning rate of 0.01, and early stopping based on validation error. Optimization focused on minimizing the mean squared error and achieving maximal classification accuracy.

The ANN achieved good classification on the dataset, reflecting its ability to describe the available observations rather than providing fully generalizable predictions. All analyses were performed using Statistica 10 software (StatSoft Inc., Tulsa, OK, USA, 2010), and network weights, biases, and hyperparameters were carefully documented to ensure reproducibility [51,52].

3. Results and Discussion

3.1. Color Correlation Analysis

Correlation analysis was performed using the Pearson correlation coefficients calculated from the mean values of the measured chromatographic, chemical, and antioxidant variables for the twenty *Cucurbita* accessions. The results were visualized using a color-coded correlation matrix to facilitate the interpretation of the strength and direction of relationships between variables.

A color-coded correlation diagram illustrating the relationships among all observed responses is shown in Figure 2. The correlation coefficients between two responses are represented by colors (blue for positive and red for negative correlations) and by the size of the circles (larger circles indicate stronger correlations).

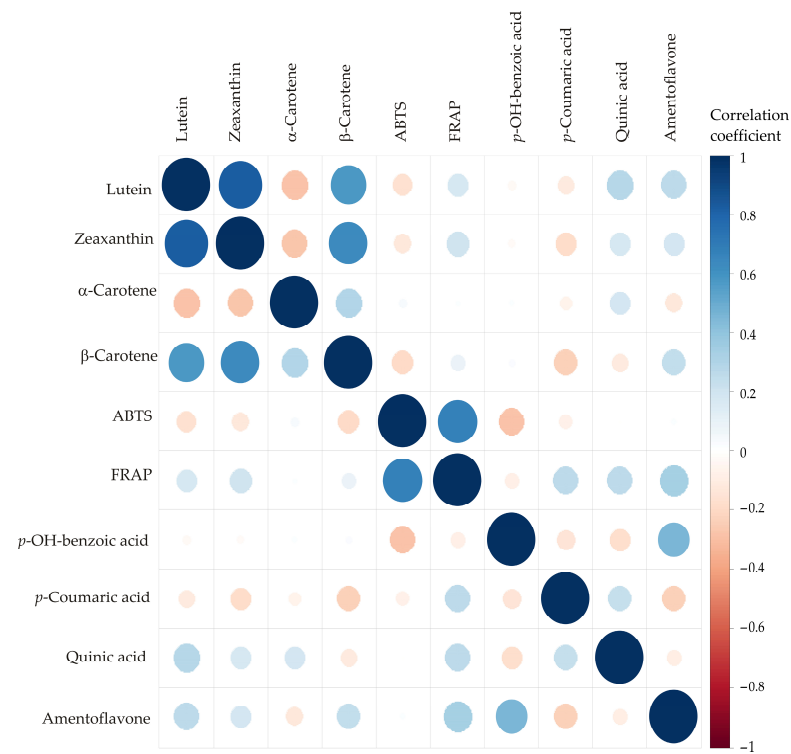


Figure 2. Correlation diagram between observed responses of carotenoid and phenolic compound contents and antioxidant activities (ABTS and FRAP).

In Figure 2, the correlation coefficients between different variables related to various *Cucurbita* samples are shown. As displayed, several correlations are significant: Lutein and zeaxanthin show a correlation of $r = 0.824$ ($p < 0.001$), indicating a very strong positive relationship between these variables. Zeaxanthin and β -carotene have a correlation of $r = 0.637$ ($p < 0.01$), also showing a significant positive correlation. Lutein and β -carotene exhibit a correlation of $r = 0.571$ ($p < 0.01$), which is significant but weak. Additionally, ABTS and FRAP show a statistically significant positive correlation of $r = 0.673$ ($p < 0.001$).

The strong positive correlations observed among individual carotenoids in the present study are consistent with previous reports in *Cucurbita* and similar systems. For example, in a study by Abbas et al. [53], inter-carotenoid correlation coefficients ($r = 0.61$ – 0.88) indicated that carotenoids are synthesized together in the same way as a result of the coordinated biosynthesis of carotenoids. Furthermore, Akshaya et al. [54] found that there were also strong positive associations of total carotenoids with other secondary products like phenolics and antioxidant properties (FRAP and DPPH). This indicates that secondary products, such as carotenoid and antioxidant products, are closely connected in terms of their production and metabolism. Similarly, the work of Kulczynski and Gramza-Michalowska [55] reported strong associations of a number of metabolites including carotenoids, tocopherols, phenolic compounds, and certain minerals, suggesting that the carotenoid metabolic system acts as a single metabolic unit. Thus, these studies provide evidence that the correlation between individual carotenoids determined in this study is likely due to biosynthesis pathways being regulated together rather than merely chemical reactions between isolated events.

While the correlations presented in Figure 2 describe the relationships between antioxidant assays and selected carotenoids and phenolic compounds, the same correlation approach was applied to lipidomic data in order to examine interrelationships among terpenoid constituents. The results of this analysis are presented in Figure 3.

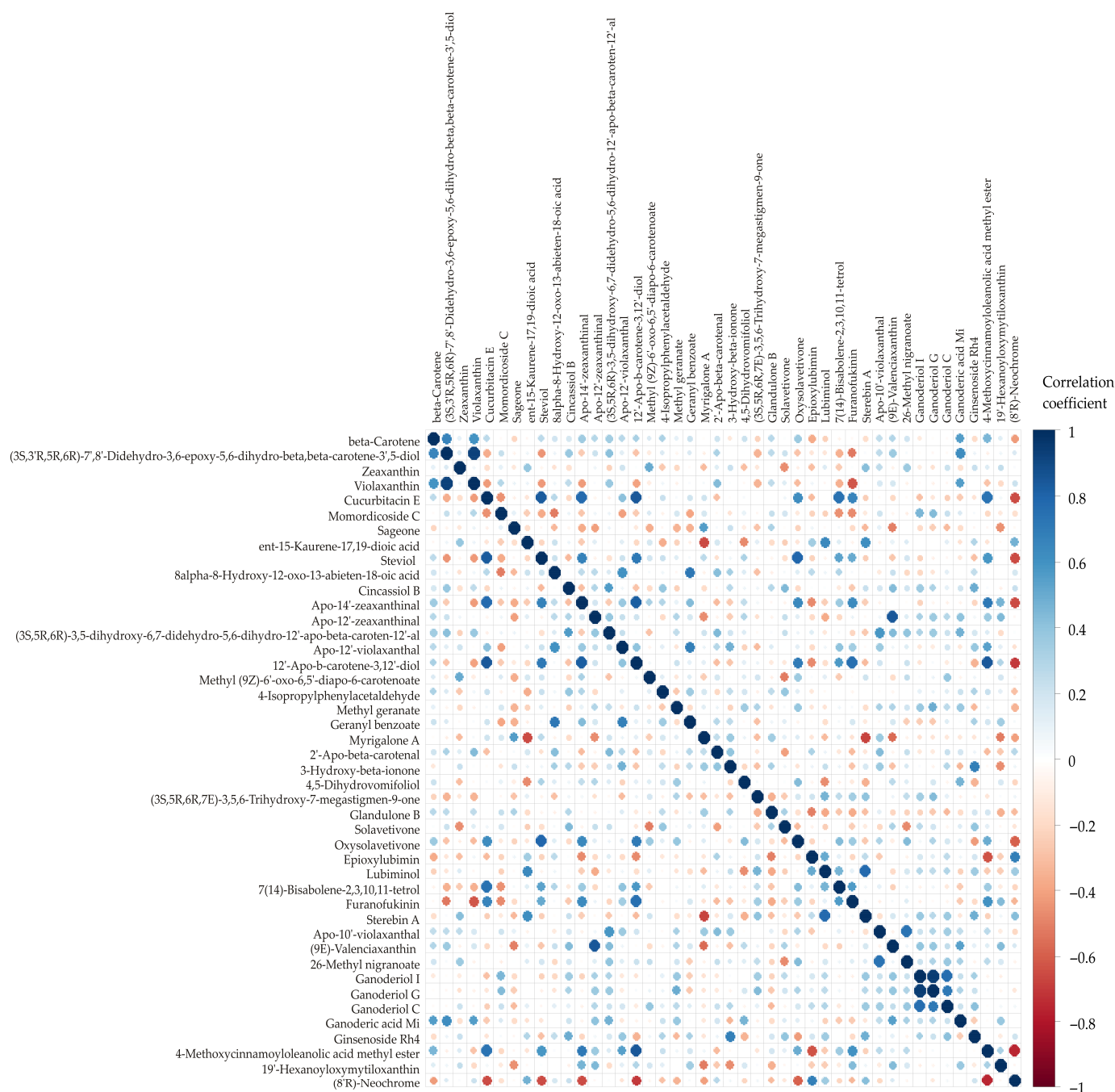


Figure 3. Correlation diagram between various compounds (lipidomic analysis) of different *Cucurbita* samples.

Lipidomic compound correlation analysis showed that β -carotene demonstrates strong positive associations with (3S,3'R,5R,6R)-7',8'-didehydro-3,6-epoxy-5,6-dihydro- β , β -carotene-3',5'-diol ($r = 0.930$, $p < 0.001$) and steviol ($r = 0.673$, $p < 0.001$). Furthermore, this diol relates positively to zeaxanthin ($r = 0.506$, $p < 0.05$) while simultaneously exhibiting inverse relationships with cucurbitacin E ($r = -0.458$, $p < 0.05$) and apo-12'-zeaxanthinal ($r = -0.472$, $p = 0.036$), indicating potential divergence in functional pathways. Zeaxan-

thin, in turn, shows negative associations with violaxanthin ($r = -0.621$, $p < 0.01$) and momordicoside C ($r = -0.514$, $p < 0.05$) yet aligns strongly with apo-14'-zeaxanthinal ($r = 0.829$, $p < 0.001$), suggesting pathway-specific biological interactions. Similarly, violaxanthin and apo-12'-zeaxanthinal form a pronounced positive linkage ($r = 0.854$, $p < 0.001$), while apo-14'-zeaxanthinal concurrently associates with apo-12'-zeaxanthinal ($r = 0.829$, $p < 0.001$) and (3S,5R,6R)-3,5-dihydroxy-6,7-didehydro-5,6-dihydro-12'-apo- β -caroten-12'-al ($r = 0.584$, $p < 0.01$). Further analysis reveals that steviol maintains robust positive relationships with cucurbitacin E ($r = 0.841$, $p < 0.001$), apo-12'-zeaxanthinal ($r = 0.585$, $p < 0.05$), (3S,5R,6R)-3,5-dihydroxy-6,7-didehydro-5,6-dihydro-12'-apo- β -caroten-12'-al ($r = 0.444$, $p < 0.050$), and apo-12'-violaxanthal ($r = 0.685$, $p < 0.001$), whereas contrasting tendencies are observed with momordicoside C ($r = -0.377$, $p < 0.05$) and ganoderiol G ($r = -0.655$, $p = 0.002$), implying potentially opposing roles within certain biochemical processes. In addition, cucurbitacin E correlates positively with apo-14'-zeaxanthinal ($r = 0.662$, $p < 0.001$) and steviol ($r = 0.841$, $p < 0.001$) while diverging from myrigalone A ($r = -0.678$, $p < 0.001$), which itself shows negative associations with apo-12'-violaxanthal ($r = -0.518$, $p < 0.05$) and apo-14'-zeaxanthinal ($r = -0.573$, $p < 0.01$). Moreover, ganoderiol G exhibits a very strong positive relationship with ganoderiol I ($r = 0.950$, $p < 0.001$) and ganoderiol C ($r = 0.717$, $p < 0.001$) yet opposes methyl (9Z)-6'-oxo-6,5'-diapo-6-carotenoate ($r = -0.516$, $p < 0.05$). Lubiminol further corresponds positively with epioxylubimin ($r = 0.772$, $p < 0.001$), and the latter also connects with solavetivone ($r = 0.508$, $p < 0.05$), reinforcing shared structural or biosynthetic relationships. Finally, (9E)-valencixanthin ($r = 0.549$, $p < 0.05$) and sterebin A ($r = 0.482$, $p < 0.05$) similarly demonstrate moderate positive associations within the broader terpenoid network.

The strong correlations observed among β -carotene, xanthophylls, and their oxygenated or apocarotenoid derivatives likely reflect coordinated regulation within the carotenoid biosynthetic pathway. Since these compounds originate from common isoprenoid precursors and are interconverted through specific enzymatic reactions (e.g., hydroxylation and epoxidation steps within the xanthophyll cycle), their co-accumulation or inverse relationships are chemically expected [56,57]. Therefore, the detected intra-group correlations primarily indicate metabolic linkage and pathway flux rather than independent biological effects.

From a nutritional perspective, correlations gain greater relevance when linked to antioxidant capacity. Although the present matrix revealed strong intra-group associations among carotenoids (e.g., $r = 0.930$ between β -carotene and its oxygenated derivative; $r = 0.829$ among apocarotenoids), previous studies have shown that antioxidant assays such as FRAP, DPPH, and ABTS more consistently correlate with phenolic content (r values often >0.80), highlighting the dominant contribution of phenolics to total redox activity [58,59]. Chemometric evaluations further suggest that strong metabolite clustering often reflects shared biosynthetic or regulatory origin rather than direct functional synergy [60,61]. Thus, while the observed correlations are chemically significant as indicators of coordinated secondary metabolism, their nutritional relevance should primarily be interpreted in relation to compounds directly associated with antioxidant performance.

3.2. Cluster Analysis (CA)

Cluster analysis (CA) was applied to examine the relationships among the observed responses and to group samples according to their similarity based on the analyzed parameters. Experimental data were used for the CA computation, and the obtained results were presented in the form of a dendrogram (Figure 4).

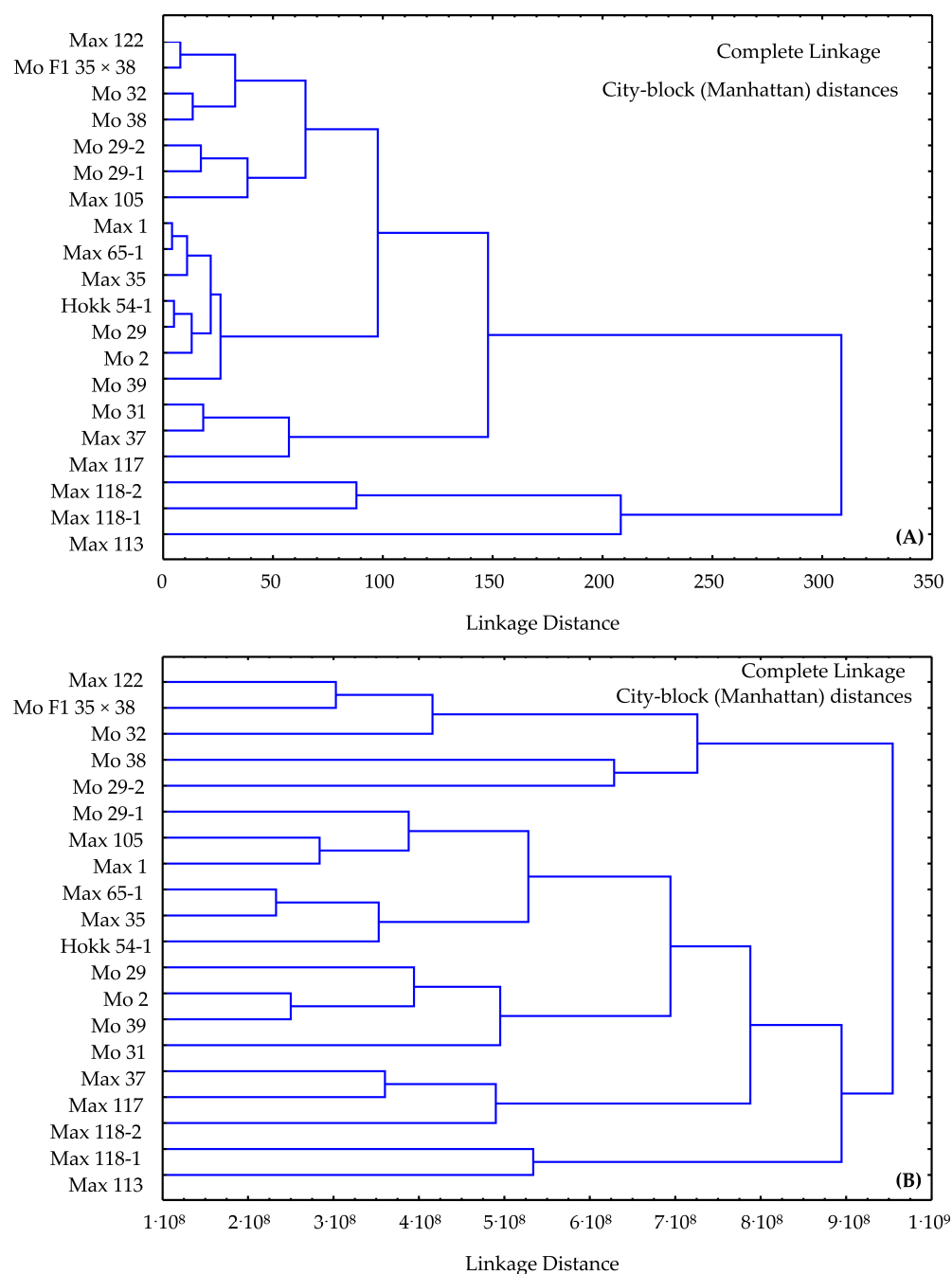


Figure 4. Dendrograms obtained by cluster analysis (CA) of tested *Cucurbita* samples based on (A) carotenoid and phenolic composition combined with antioxidant activity (ABTS and FRAP) and (B) terpenoid profile.

The dendrogram obtained by cluster analysis illustrates similarity relationships among the analyzed *Cucurbita* samples based on the selected distance metric and linkage method. The clustering should be interpreted as an exploratory visualization of sample similarity rather than as a statistically validated cluster structure.

The CA dendrogram (Figure 4A), based on carotenoid and phenolic contents and calculated using City-block (Manhattan) distance, revealed distinct clusters while reducing outlier influence. Highly similar accessions (e.g., Mo 32-Mo 38; Mo 29-2-Mo 29-1; Max 122-Mo F1 35 × 38; Max 118-1-Max 118-2) merged at low linkage distances, whereas larger groups (e.g., Max 1, Max 65-1, Max 35, Hokk 54-1, Mo 29, Mo 2, Mo 39, Mo 31) showed

overall similarity with some heterogeneity. A gradient of similarity was observed in the Max 37-Max 117-Max 118-Max 113 cluster.

Terpenoid-based clustering (Figure 4B) separated samples into two main branches, one more homogeneous (e.g., Max 118-1, Max 113, Max 37, Max 117, Max 118-2) and the other more heterogeneous, including several subclusters (e.g., Max 122-Mo F1 35 × 38; Mo 32-Mo 38-Mo 29-2), indicating metabolic diversity.

The clustering aligns with strong metabolite correlations, such as ganoderiol G-ganoderiol I ($r = 0.950$, $p < 0.001$) and β -carotene-its diol ($r = 0.930$, $p < 0.001$). Similar chemometric relationships were reported by Mitic et al. [59] (CUPRAC-TPC $r = 0.91$; FRAP-TPC $r = 0.77$; FRAP-CUPRAC $r = 0.87$). The species-dependent variability described by Stryjecka et al. [62] (DPPH up to 11.6-fold; FRAP up to 4.65-fold) and Kostecka-Gugała et al. [3] (TPC up to 70.8 mg CAE·100 g⁻¹ FW with strong variability) supports the observed grouping, also reflected in correlations such as steviol-cucurbitacin E ($r = 0.841$), violaxanthin-apo-12'-zeaxanthinal ($r = 0.854$), and apo-14'-zeaxanthinal-apo-12'-zeaxanthinal ($r = 0.829$).

At the biosynthetic level, Chen et al. [63] linked metabolite accumulation to phenylpropanoid pathway gene expression, explaining strong associations among related compounds (e.g., lubiminol-epioxylubimin, $r = 0.772$; apo-14'-zeaxanthinal-apo-12'-zeaxanthinal, $r = 0.829$). Overall, high correlations (up to $r = 0.950$) indicate coordinated pathway-level regulation rather than random variation.

3.3. Principal Component Analysis (PCA)

The proximity of points representing the observed samples on the PCA biplot indicates sample similarity based on experimental values. When the principal components account for a larger portion of the total variance, the measured values according to experimental data will be more similar for samples that are geometrically closer on the PCA plot. PCA was applied to investigate the relationships between the observed responses (carotenoids, phenolic compounds, ABTS and FRAP assays) and to group the samples according to their similarity based on the observed parameters in the factor plane. The PCA transformation was performed using experimental data for the analyzed *Cucurbita* samples based on the contents of carotenoids and phenolic compounds and antioxidant values (ABTS and FRAP), as well as the lipidomic results.

The PCA biplots (Figure 5) reveal several major directions of variability in the dataset, associated with carotenoid-, antioxidant-, and phenolic-related descriptors. These patterns reflect the multidimensional structure of the measured compositional data and illustrate how samples are distributed in the principal component space. However, given the study design and the diversity of accessions included, the observed variability cannot be attributed to specific causal factors such as species, genotype, environmental conditions, or fruit maturity.

The results of PCA for the analyzed *Cucurbita* samples, based on carotenoid and phenolic compound contents and antioxidant values (ABTS and FRAP), are presented in Figure 5. The first five factor coordinates represent 85.5% of the total variability, which is considered adequate to describe the overall variability of the observed responses, as the cumulative variance exceeds 70% (Figure 5A). The calculation of the first principal component (PC1) was the most positively influenced by the contents of lutein (30.35%, based on correlation), zeaxanthin (30.78%), and β -carotene (21.48%). The antioxidant responses ABTS (25.83%) and FRAP (34.31%) contributed the most positively to the second principal component (PC2). The third principal component (PC3) was the most positively influenced by the contents of amentoflavone (30.80%), *p*-hydroxybenzoic acid (15.60%), and quinic acid (10.69%), while *p*-coumaric acid showed a negative influence of -6.93% .

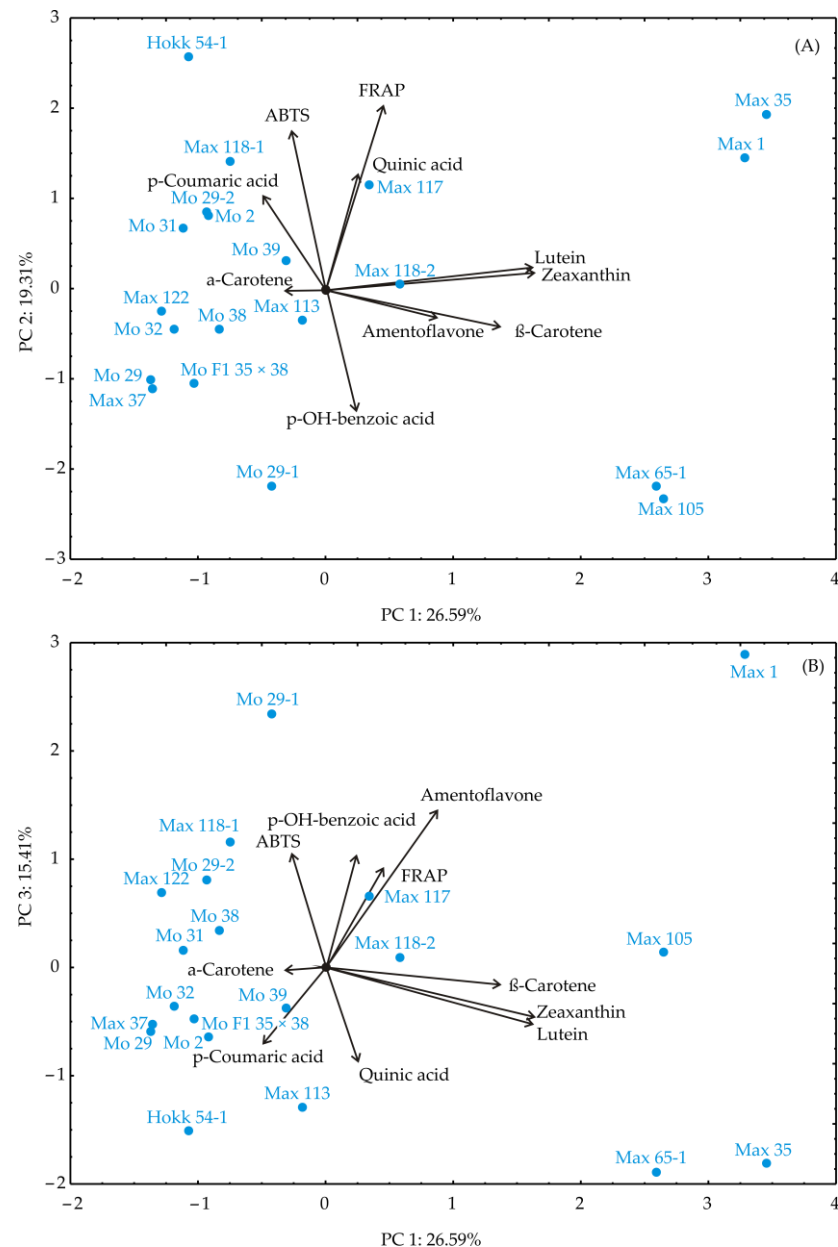


Figure 5. PCA biplots of the tested *Cucurbita* samples based on carotenoid and phenolic compound contents together with antioxidant parameters (ABTS and FRAP): (A) first factor plane (PC1–PC2) and (B) first factor plane (PC1–PC3).

The proximity of samples in the factor plane indicates their similarity, while the closeness of a sample to a vector denotes a higher content of the corresponding compound in that sample. PCA reveals three distinct patterns among the analyzed *Cucurbita* accessions. Samples Max 1, Max 35, Max 65-1 and Max 105 are clearly positioned along the positive side of the PC1 axis, reflecting a profile dominated by carotenoids. In contrast, Max 117 and Hokk 54-1 exhibit higher PC2 values and are positioned near the FRAP and ABTS vectors, suggesting a stronger association with antioxidant activity. Sample Mo 29-1, and to a lesser extent Max 117, shows a greater contribution of the PC3 component, which is associated with phenolic compounds, particularly amentoflavone and *p*-hydroxybenzoic acid. The remaining samples are distributed centrally within the factor plane, indicating the absence of the pronounced dominance of any specific group of compounds.

The results of the PCA for the analyzed *Cucurbita* samples based on their lipidomic profiles are presented in Figure 6A,B, which show the arrangement of variables (loadings)

and samples in the factor plane for the first two principal components (PC1 and PC2) and the first and third principal components (PC1 and PC3), respectively.

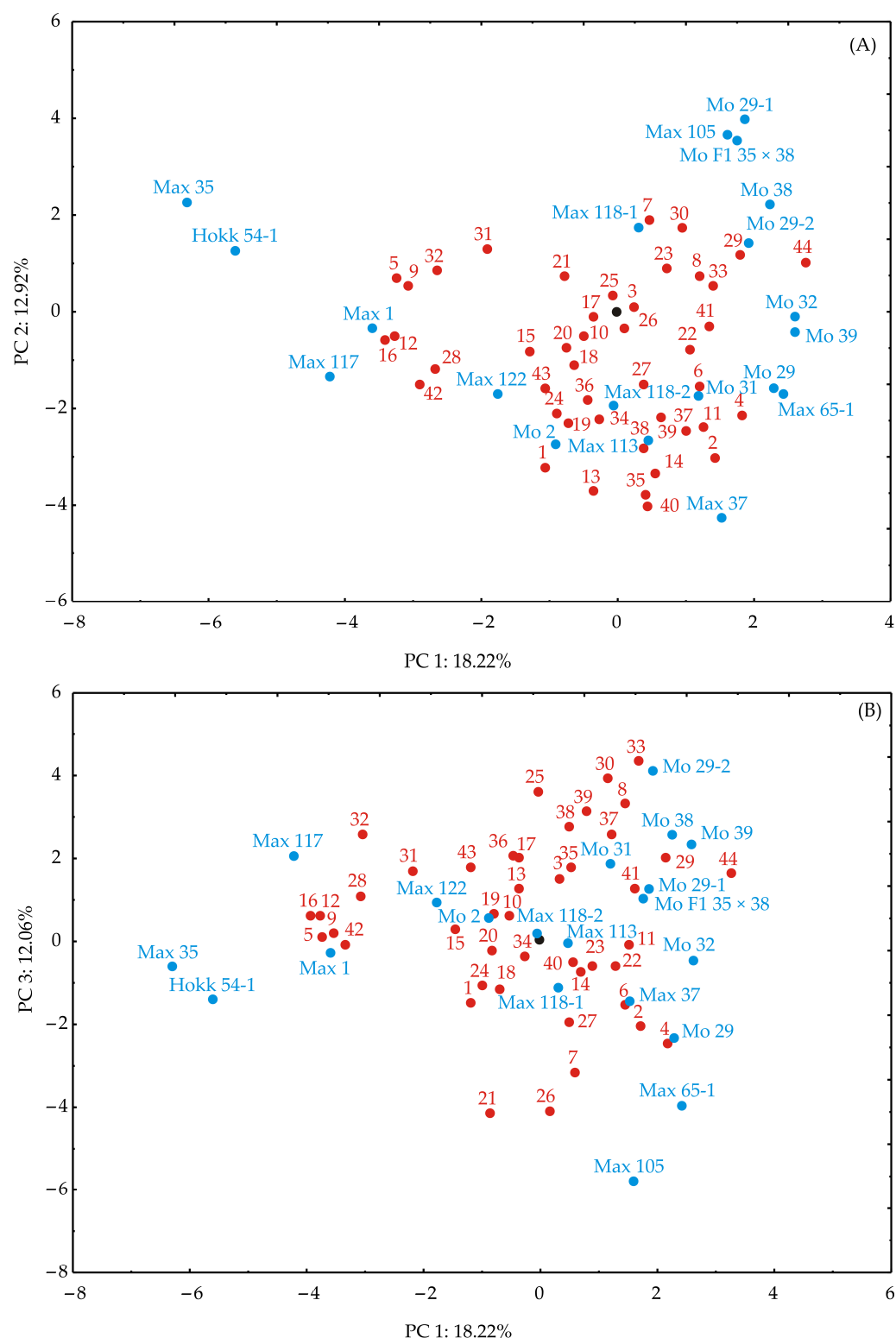


Figure 6. PCA biplots of the tested *Cucurbita* samples (blue) based on lipidomic profile data (Table S4 contains details on lipidomic compounds presented in red): (A) first factor plane (PC1–PC2) and (B) first factor plane (PC1–PC3).

The first principal component (PC1) was the most positively influenced by the content of (8'R)-neochrome (7.29%), while the most negative influences were observed for 12'-apo-

β -carotene-3,12'-diol (−10.86%), apo-14'-zeaxanthinal (−10.04%), cucurbitacin E (−9.92%), and steviol (−8.84%). The second principal component (PC2) was the most negatively influenced by the contents of ganoderic acid Mi (−10.73%), (9E)-valenciananthin (−9.48%), β -carotene (−6.82%), and (3S,3'R,5R,6R)-7',8'-didehydro-3,6-epoxy-5,6-dihydro- β , β -carotene-3',5-diol (−5.98%). The third principal component (PC3) was the most positively influenced by the contents of sterebin A (10.06%), lubiminol (8.26%), and (3S,5R,6R,7E)-3,5,6-trihydroxy-7-megastigmen-9-one (7.01%), while myrigalone A and glandulon B had the strongest negative effects (−8.88% and −8.64%, respectively). In Figure 6A,B, the samples are shown in blue, while the variables (compounds determined by lipidomic analyses) are shown in red. The proximity of sample points indicates their similarity, whereas the proximity of variable points indicates a higher concentration of that compound in the respective sample.

This separation does not represent merely a statistical differentiation of the samples, but it also suggests distinct biosynthetic “weighting” among the investigated accessions. The contrast between intact carotenoids and apocarotenoids indicates differences in carotenoid turnover, since apocarotenoids originate from the oxidative cleavage of carotenoids, either via reactive oxygen species and/or enzymatic cleavage mediated by carotenoid cleavage dioxygenases (CCDs) [64]. In line with the conserved carotenoid/xanthophyll network described in plants [56], and the genetically and physiologically modulated balance within xanthophyll-related processes (including zeaxanthin-violaxanthin interconversion and carotenoid turnover) [57], the present pattern supports the interpretation that some genotypes maintain higher proportions of intact carotenoid structures, whereas others show a shift toward oxidative transformation and apocarotenoid accumulation. In addition, the variability in cucurbitacin E (a triterpenoid typical for Cucurbitaceae) alongside carotenoid-derived features further supports genotype-dependent differences in terpenoid pathway regulation [56,57]. Collectively, these results point to the pronounced diversity of terpenoid-related metabolism within *Cucurbita* accessions, with potential implications for bioactivity, pigment stability, and nutritional relevance.

3.4. Artificial Neural Network Analysis

The cANN modeling pipeline was implemented as a multi-layer perceptron (MLP) with input, hidden, and output layers, trained using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. Target classes were numerically encoded for classification, and input and output data were normalized. For the lipidomic dataset, the output layer used identity or tanh activation to accommodate continuous normalized values, whereas for liquid chromatographic data, tanh was applied in the output layer to capture both positive and negative normalized measurements. For antioxidant data combined with phenolic and carotenoid compositional data, softmax activation was used in the output layer to provide probabilistic class assignment for categorical outcomes. Hidden layers consistently employed tanh activation to model nonlinear relationships. Loss functions were selected according to data type, including logistic or entropy-based functions. No explicit regularization techniques (e.g., weight decay or dropout) or early stopping were applied; instead, model robustness was achieved through empirical hyperparameter tuning, exploring different hidden neuron numbers (5–20 for smaller datasets, larger networks for combined datasets) and iterative training up to 100,000 epochs, selecting networks that minimized validation error. The choice of tanh in output layers for continuous variables was motivated by its ability to represent both positive and negative normalized values, while softmax ensured appropriate classification outputs for categorical data.

The MLP 351-8-2 classification neural network (for identifying *Cucurbita* accessions and species using lipidomic data), comprising 351 input neurons, 8 hidden neurons, and

2 output neurons, achieved 100.00% accuracy on the training set, 100.00% accuracy on the test set, and 66.67% on the validation test set, demonstrating strong classification performance with good generalization (Table S5). The network was trained using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm with cross-entropy loss, ensuring effective optimization for multi-class classification. The identity activation function in the hidden layer enhances the model’s ability to capture nonlinear relationships, while the softmax activation in the output layer converts predictions into probability distributions across the accessions of the two *Cucurbita* species.

The softmax function was used in the calculation as a specialized activation function for the classification neural network. This function performs a normalized exponential (adding the outputs up to 1.000), making the MLP network to be modified for class probability estimation, in combination with the cross-entropy error function.

The minimal performance drop between training and testing suggests that the model is well-fitted without significant overfitting, suggesting the effectiveness of the cANN approach for differentiating *Cucurbita* accessions and species based on lipidomic data.

The classification summary of the MLP 351-8-2 model indicates strong performance in distinguishing between the Max and Mo accessions and species, with 90.91% and 100% accuracy, respectively. Out of 11 Max samples, 10 were correctly identified, while 1 was misclassified (9.09% error rate). All nine Mo samples were correctly classified, demonstrating high reliability for this species (Figure 7).

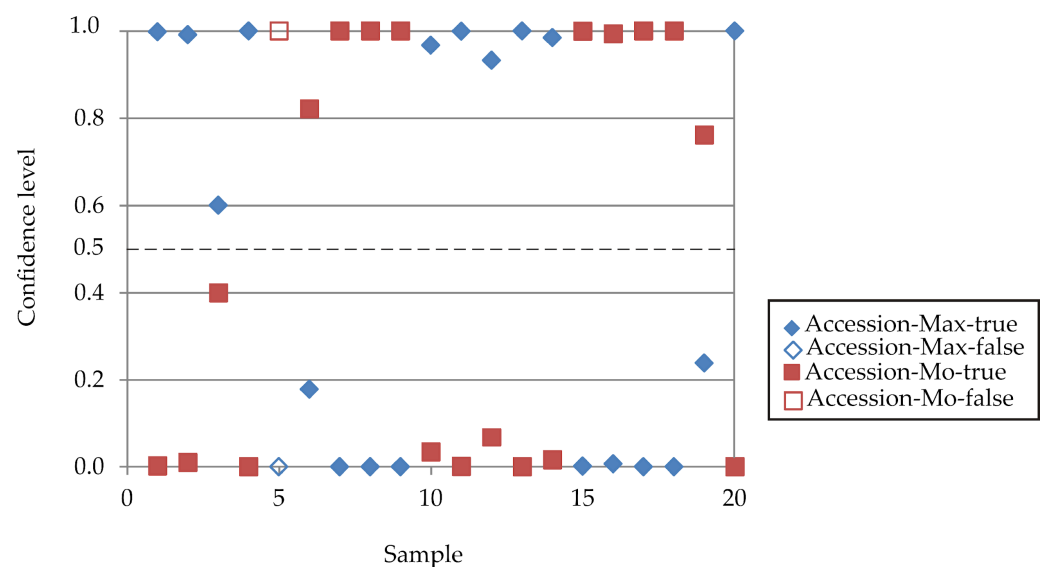


Figure 7. Classification results of MLP 351-8-2 artificial neural network model for identification of Max and Mo *Cucurbita* accessions based on liquid chromatographic data.

The confidence levels from the ANN classification model for classifying tested *Cucurbita* accessions and species (using the MLP 351-8-2 network) reveal varied performance across samples. For the majority of the Max-targeted samples, the model shows high confidence, with values close to 1 (e.g., samples Max 113 and Max 117). However, sample Max 37 shows lower confidence for the correct species, indicating some uncertainty in classification. For Mo-targeted samples, the model also demonstrates high confidence, with values nearing 1 for the correct classification, suggesting good classification performance.

The classification artificial neural network (cANN) model, MLP 10-12-2, designed to classify *Cucurbita* samples using antioxidant data combined with phenolic and carotenoid compositional data, exhibited good classification performance with 100.00% accuracy on both the training and test sets. This indicates an exceptionally well-fitted model with

no observable misclassifications. The network architecture consists of 10 input neurons corresponding to the parameters lutein, zeaxanthin, α -carotene, β -carotene, ABTS, FRAP, *p*-hydroxybenzoic acid, *p*-coumaric acid, quinic acid, and amentoflavone; 12 hidden neurons; and 2 output neurons (representing the two *Cucurbita* species). The model was trained using the BFGS 2 algorithm with the entropy error function, which effectively minimized classification errors. The tangent hyperbolicus (Tanh) activation function in the hidden layer allows the model to capture complex nonlinear relationships between antioxidant data and *Cucurbita* samples, while the softmax activation function in the output layer ensures well-separated class predictions. The good classification performance suggests that the selected antioxidant descriptors provide highly discriminative information for distinguishing between the two *Cucurbita* species, suggesting the robustness of the cANN approach in food classification tasks.

The classification summary of the MLP 10-12-2 model demonstrates 100% accuracy in distinguishing all two *Cucurbita* species (Max and Mo) using antioxidant data together with phenolic and carotenoid profiles. All samples were correctly classified, with no misclassifications observed across training and testing sets. This good classification performance suggests that the selected parameters effectively differentiate between the *Cucurbita* species, and the MLP 10-12-2 network, trained with the BFGS 2 algorithm, successfully captures the underlying patterns in the dataset (Figure 8).

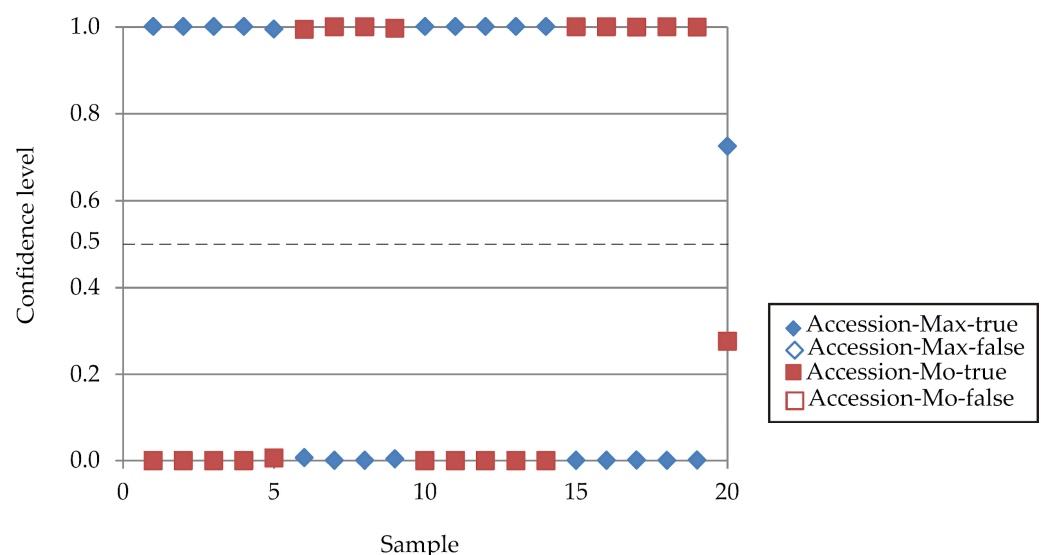


Figure 8. Classification results of MLP 10-12-2 artificial neural network model for discrimination of Max and Mo *Cucurbita* samples based on antioxidant data combined with phenolic and carotenoid compositional data.

The confidence levels from the ANN classification model (using the 1. MLP 10-12-2 network) for classifying *Cucurbita* accessions show varying degrees of certainty for the correct classifications. The confidence level results obtained from the MLP 10-12-2 artificial neural network model demonstrate a clear distinction between the two *Cucurbita* species, Max and Mo. For the majority of samples, the model exhibited extremely high classification confidence, with output probabilities approaching 1.0 for the predicted class and near-zero for the alternative, indicating strong model certainty and discriminative power. Notably, samples such as 1, 3, 7, and 11–14 were classified with near-absolute confidence, while only sample 20 showed moderate classification ambiguity, with a confidence level of 72.4% for Max and 27.6% for Mo. These results suggest that the ANN effectively differentiates between the two *Cucurbita* species, with only minimal uncertainty observed in borderline cases.

The cANN model, MLP 44-20-2, was developed to classify two *Cucurbita* species based on chemical composition data. The network achieved a training performance of 100.00% and a good test performance of 100.00%. The architecture consists of 44 input neurons (corresponding to chemical descriptors), 20 hidden neurons, and 2 output neurons (representing the two *Cucurbita* species). The model was trained using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (iteration 8) with the sum of squares (SOS) error function, optimizing classification accuracy. The identity activation function in the hidden layer effectively captured complex relationships within the chemical dataset, while the tanh activation function in the output layer ensured probabilistic classification across the two species. The model's ability to achieve good classification on the test set highlights the strong discriminatory power of chemical data in distinguishing between the two tested *Cucurbita* species, suggesting the effectiveness of the cANN approach for this classification task.

The classification summary of the MLP 44-20-2 model reveals good accuracy for the Max and Mo *Cucurbita* species, with 100% correct classification for both. Specifically, out of 11 Max samples, all were correctly classified, and out of 9 Mo samples, all were also correctly classified. The 100% accuracy for both Max and Mo indicates that the model effectively discriminates these species based on the compound data (Figure 9).

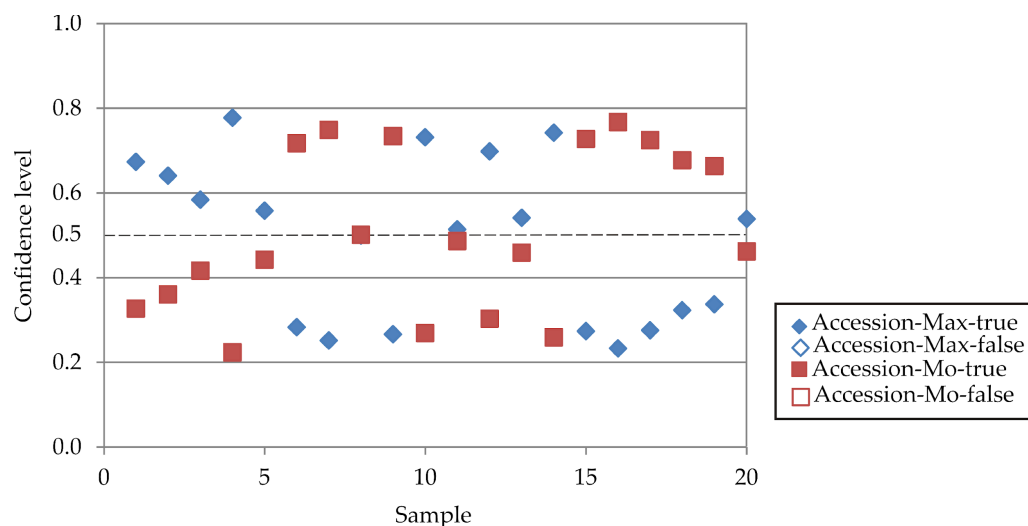


Figure 9. Classification performance of MLP 44-20-2 ANN model for identification of Max and Mo *Cucurbita* species using chemical composition descriptors.

In the classification of analyzed *Cucurbita* samples based on compound profiles using the MLP 44-20-2 neural network, the confidence levels indicate a generally moderate distinction between the two groups, Max and Mo. Unlike the previous model with sharper separability, this configuration exhibited more balanced probability distributions across the two classes. Several samples, such as 1–5 and 10–14, showed a modest preference for the Max class (e.g., sample 4: 77.7% Max, 22.3% Mo), while others such as samples 6, 7, 9, and 15–17 were more confidently aligned with the Mo class, showing confidence levels exceeding 70%. Notably, sample 8 yielded nearly equal confidence values (49.9% Max, 50.1% Mo), indicating classification ambiguity. Overall, the ANN demonstrated the ability to detect class-related patterns, although with less certainty compared to the previous model, potentially reflecting more overlapping chemical characteristics between the two groups.

The LC-MS/MS data, antioxidant activities, and chemical data used to train the cANN models play a crucial role in determining the accuracy of the identification of the two analyzed *Cucurbita* species. High-quality data, free from errors, outliers, and

inconsistencies, contributes to a more accurate model. Generally, a larger dataset provides more representative patterns for the model to learn from, resulting in better generalization and higher accuracy. In the context of the provided cANN model results, it could be mentioned that the accuracy of the model may benefit from an increase in the number of samples, allowing the model to capture more diverse characteristics of various types of *Cucurbita* species.

The obtained results reveal the reliability of the cANN model for identifying the two *Cucurbita* species determined via liquid chromatographic data, antioxidant parameters, and chemical data analysis.

The sample size of *Cucurbita* is limited, but this preliminary study demonstrates the feasibility of combining liquid chromatographic (LC-MS/MS), chemical composition, and antioxidant data with machine learning for species differentiation. Despite the small dataset, the ANN models reliably distinguish between *C. maxima* and *C. moschata*, with the high classification accuracy of the MLP architecture suggesting their ability to model complex nonlinear relationships in multivariate biological systems, relationships that are often difficult to capture with classical linear methods. A similar strategy was employed by Oliveira et al. [65] to predict the chromatographic retention of terpenoids using molecular descriptors, accelerating compound replication and facilitating the identification of novel bioactive substances. The ANN modeling of antioxidant data (ABTS and FRAP) aligns with previous studies on blueberries [66] and *Cucurbita* [67], effectively describing correlations between phenolic and carotenoid profiles and antioxidant activity.

The ANN analysis demonstrated that the applied models were able to distinguish the analyzed samples based on the available compositional descriptors. However, given the relatively small sample size and the large number of input variables, the observed classification performance should be interpreted with caution. The obtained accuracy primarily reflects the model's ability to describe the specific dataset used in this study, rather than providing definitive evidence of broadly generalizable species discrimination.

The cANN models developed in this study demonstrated high predictive performance on the available dataset, with some models reaching near-perfect accuracy across training metrics. However, it should be noted that the total number of samples was limited ($n = 20$), and the validation set in the first ANN consisted of only three samples. This small sample size substantially limits the statistical power and generalizability of the models, and the observed perfect or near-perfect performance likely reflects overfitting to the training data. While these results illustrate the potential of cANN modeling for classification based on LC-MS/MS, lipidomic, and antioxidant data, they should be interpreted with caution. Future studies with larger sample sizes are necessary to validate the robustness and generalizability of the models.

This study represents a preliminary, proof-of-concept application of an ANN for classifying *Cucurbita* accessions based on chemical and antioxidant descriptors. Due to the limited sample size and single-season experimental design, external validation was not performed; instead, model performance was assessed using a five-fold cross-validation scheme. Future studies involving larger, multi-season datasets, along with external validation, will be required to suggest the predictive robustness and generalizability of the proposed approach.

In contemporary food chemistry research, ANN models have demonstrated superiority over classical regression and response surface methodology (RSM) approaches in modeling complex systems [68], and they increasingly represent a standard framework for the prediction and optimization of properties in complex food matrices. Their widespread application in quality modeling and formulation of functional products further suggests

that artificial neural networks have become a standard tool within modern food chemometrics [69].

Future research should focus on expanding the developed ANN model to include a larger number of genotypes and species within the genus *Cucurbita*, encompassing additional genotypes of *C. maxima* and *C. moschata*, as well as other economically important species such as *C. pepo* and *C. ficifolia*. Expanding the dataset would enable more robust external validation and improve the model's generalization capacity. The integration of additional analytical approaches, such as LC-MS/MS metabolomic profiling or volatile compound analysis, could further enhance the model's discriminative power and contribute to a deeper understanding of chemotaxonomic differences within the genus.

3.5. Study Limitations

This study is preliminary and exploratory, and several limitations should be considered when interpreting the results. First, the sample size was limited ($n = 20$ accessions), which restricts statistical power and may increase the risk of overfitting in the ANN models. Second, no independent external validation dataset was available; model performance was assessed using internal train–test splits and five-fold cross-validation, which provides only a preliminary estimate of predictive robustness. Third, all accessions were grown in a single location and year, minimizing environmental variability but potentially confounding accession-specific chemical profiles with local growing conditions. Finally, uncertainty estimates for correlations, PCA loadings, and ANN predictions were not included due to the small dataset. Together, these limitations indicate that the results should be interpreted with caution, and future studies with larger, multi-location datasets and external validation are necessary to suggest the observed trends and improve model generalizability.

3.6. Future Studies

This preliminary study provides insight into the potential of combining liquid chromatographic, lipidomic, carotenoid, phenolic, and antioxidant data with classification artificial neural networks (cANNs) for exploring chemical variability among *Cucurbita* accessions. Future research should focus on several key aspects to build on these findings. First, the sample size should be expanded to include a larger number of accessions, covering multiple species, varieties, and geographic origins, to improve the robustness and generalizability of the models. Second, multi-year and multi-location experiments are recommended to assess environmental influences and separate them from genetic or biochemical effects. Third, additional performance metrics for ANN models, including sensitivity, specificity, precision, recall, and confidence intervals, should be calculated to provide a more rigorous evaluation of predictive performance. Fourth, the models should be validated on independent datasets and applied to commercial or processed products to evaluate practical applicability in quality control and food authentication. Advanced statistical approaches, such as log-ratio transformations for compositional data and bootstrap resampling to assess the stability of PCA and dendrogram patterns, could enhance interpretability and reproducibility. Together, these approaches will enable a more comprehensive understanding of chemical diversity in *Cucurbita* and support the development of robust predictive models for both research and industrial applications.

4. Conclusions

A comprehensive chemometric and machine learning approach was applied to differentiate *C. maxima* and *C. moschata* based on integrated compositional data, including chromatographic, chemical, and antioxidant parameters. Correlation analysis revealed significant associations among carotenoids, phenolics, and antioxidant responses. Cluster

analysis (CA) and principal component analysis (PCA) successfully grouped the samples, highlighting clear multivariate patterns and species-related similarities.

The classification artificial neural network (cANN) models, developed using HPLC, LC–MS/MS, and antioxidant data, achieved high classification accuracy, supporting the robustness of the applied approach. The integration of multivariate statistics and neural network modeling provided a consistent framework for differentiation and compositional characterization among accessions.

All applied analytical approaches consistently indicated that the analyzed samples exhibit well-defined compositional profiles driven by coordinated patterns of carotenoids, terpenoids, phenolic compounds, and associated antioxidant activity. The observed differences reflect distinct compositional characteristics between the analyzed species.

Given the limited sample size, lack of external validation, and single-location growth conditions, these findings should be interpreted with caution. Further studies including larger and more diverse datasets are required to confirm the robustness and generalizability of the proposed approach.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agriculture16070733/s1>, Table S1: Mobile phase composition and UHPLC gradient program for lipidomic analysis (A: water/methanol (95:5, v/v) + 5 mM ammonium formate + 0.1% formic acid; B: 2-Propanol/methanol/water (65:30:5, v/v/v) + 5 mM ammonium formate + 0.1% formic acid); Table S2: HPLC-DAD gradient for carotenoid analysis (A: acetone/water (75:25, v/v); B: acetone/methanol (75:25, v/v)); Table S3: Mobile phase gradient for phenolic compound analysis by HPLC–MS/MS (A: 0.05% formic acid in water; B: HPLC-grade methanol); Table S4: Normalized peak intensities (a.u., values divided by 107) of lipidomic compounds identified in *Cucurbita* accessions by UHPLC-Orbitrap analysis; Table S5: Performance of MLP 351-8-2 classification neural network; Table S6: Performance of MLP 10-12-2 classification neural network; Table S7: Performance of MLP 44-20-2 classification neural network.

Author Contributions: Conceptualization, M.M., B.L. and L.P.; methodology, M.M.; software, B.L. and L.P.; validation, M.M., B.L., S.K., L.P. and M.R. (Milena Rašeta); formal analysis, M.M. and S.K.; investigation, M.M.; resources, M.M. and B.K.; data curation, B.L. and L.P.; writing—original draft preparation, M.M., B.L. and L.P.; writing—review and editing, M.M., B.L., B.K., L.P., S.K., M.R. (Miloš Radosavljević), M.R. (Milena Rašeta), V.V. and D.L.; visualization, M.M. and B.L.; supervision, B.L., S.K., L.P. and M.R. (Milena Rašeta); funding acquisition, M.M., B.L., B.K., L.P., S.K., M.R. (Miloš Radosavljević), V.V. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant numbers: 451-03-33/2026-03/200222 (M.M., M.K.); 451-03-33/2026-03/200134 (B.L., M.R. (Miloš Radosavljević), V.V., D.K.), 451-03-33/2026-03/200051 (L.P.) and 451-03-33/2026-03/200032 (B.K.)).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors gratefully acknowledge that the *Cucurbita maxima* and *Cucurbita moschata* fruit flesh samples were kindly provided by Milka Brdar-Jokanović, Institute of Field and Vegetable Crops, National Institute of the Republic of Serbia, Novi Sad, Serbia. The accessions are part of the Institute's extensive breeding collection, established by the distinguished *Cucurbita* breeder, the late Berényi János. These accessions were selected from the collection based on multi-year evaluations of fruit nutritional quality and agronomic performance.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AAE	Ascorbic Acid Equivalents
ABTS	2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
AI	Artificial Intelligence
ANN	Artificial Neural Network
BFGS	Broyden–Fletcher–Goldfarb–Shanno
CA	Cluster Analysis
cANN	Classification Artificial Neural Network
CCD	Carotenoid Cleavage Dioxygenase
CUPRAC	Cupric Reducing Antioxidant Capacity
DPPH	2,2-Diphenyl-1-picrylhydrazyl
FRAP	Ferric Reducing Antioxidant Power
FW	Fresh Weight
HCA	Hierarchical Cluster Analysis
HPLC	High-Performance Liquid Chromatography
HPLC-DAD	High-Performance Liquid Chromatography with Diode Array Detection
LC-MS/MS	Liquid Chromatography–Tandem Mass Spectrometry
ML	Machine Learning
MLP	Multi-Layer Perceptron
PCA	Principal Component Analysis
PC	Principal Component
RSM	Response Surface Methodology
SOS	Sum of Squares
SRM	Selected Reaction Monitoring
Tanh	Hyperbolic Tangent
TXE	Trolox Equivalents
UHPLC	Ultra-High-Performance Liquid Chromatography
UV-Vis	Ultraviolet–Visible

References

1. FAO. FAOSTAT Data: Pumpkins, Squash and Gourds (Item Code 0394), Production Quantity by Country; Food and Agriculture Organization of the United Nations: Rome, Italy, 2024.
2. Kaur, S.; Panghal, A.; Garg, M.K.; Mann, S.; Khatkar, S.K.; Sharma, P.; Chhikara, N. Functional and nutraceutical properties of pumpkin—A review. *Nutr. Food Sci.* **2019**, *50*, 384–401. [\[CrossRef\]](#)
3. Kostecka-Gugała, A.; Kruczek, M.; Ledwożyw-Smołeń, I.; Kaszycki, P. Antioxidants and health-beneficial nutrients in fruits of eighteen *Cucurbita* cultivars: Analysis of diversity and dietary implications. *Molecules* **2020**, *25*, 1792. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Sook, J.; Jane, B.; Quan, Z.; Neidlinger, T.R. Extraction and chromatography of carotenoids from pumpkin. *J. Chromatogr. A* **2005**, *1073*, 371–375. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Provesi, J.G.; Amante, E.R. Carotenoids in pumpkin and impact of processing treatments and storage. In *Processing and Impact on Active Components in Food*; Elsevier: Amsterdam, The Netherlands, 2015. [\[CrossRef\]](#)
6. Bergantin, C.; Maietti, A.; Tedeschi, P.; Font, G.; Manyes, L.; Marchetti, N. HPLC-UV/Vis-APCI-MS/MS determination of major carotenoids and their bioaccessibility from “delica” (*Cucurbita maxima*) and “violina” (*Cucurbita moschata*) pumpkins as food traceability markers. *Molecules* **2018**, *23*, 2791. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Kamiloglu, S.; Koc Alibasoglu, E.; Acoglu Celik, B.; Celik, M.A.; Bekar, E.; Unal, T.T.; Kertis, B.; Akpinar Bayizit, A.; Yolci Omeroglu, P.; Copur, O.U. Bioaccessibility of carotenoids and polyphenols in organic butternut squash (*Cucurbita moschata*): Impact of industrial freezing process. *Foods* **2024**, *13*, 239. [\[CrossRef\]](#)
8. Zhou, C.L.; Liu, W.; Zhao, J.; Yuan, C.; Song, Y.; Chen, D.; Ni, Y.Y.; Li, Q.H. The effect of high hydrostatic pressure on the microbiological quality and physical-chemical characteristics of Pumpkin (*Cucurbita maxima* Duch.) during refrigerated storage. *Innov. Food Sci. Emerg. Technol.* **2014**, *21*, 24–34. [\[CrossRef\]](#)
9. Enneb, S.; Drine, S.; Bagues, M.; Triki, T.; Boussora, F.; Guasmi, F.; Nagaz, K.; Ferchichi, A. Phytochemical profiles and nutritional composition of squash (*Cucurbita moschata* D.) from Tunisia. *S. Afr. J. Bot.* **2020**, *130*, 165–171. [\[CrossRef\]](#)

10. Jahan, F.; Islam, M.B.; Moulick, S.P.; Al Bashera, M.; Hasan, M.S.; Tasnim, N.; Saha, T.; Bobby, F.; Waliullah, M.; Saha, A.K.; et al. Nutritional characterization and antioxidant properties of various edible portions of *Cucurbita maxima*: A potential source of nutraceuticals. *Heliyon* **2023**, *9*, e16628. [\[CrossRef\]](#)
11. Chanwitheesuk, A.; Teerawutgulrag, A.; Rakariyatham, N. Screening of antioxidant activity and antioxidant compounds of some edible plants of Thailand. *Food Chem.* **2005**, *92*, 491–497. [\[CrossRef\]](#)
12. Saavedra, M.J.; Aires, A.; Dias, C.; Almeida, J.A.; De Vasconcelos, M.C.B.M.; Santos, P.; Rosa, E.A. Evaluation of the potential of squash pumpkin by-products (seeds and shell) as sources of antioxidant and bioactive compounds. *J. Food Sci. Technol.* **2015**, *52*, 1008–10015. [\[CrossRef\]](#)
13. Hamer, M.; Chida, Y. Intake of fruit, vegetables, and antioxidants and risk of type 2 diabetes: Systematic review and meta-analysis. *J. Hypertens.* **2007**, *25*, 2361–2369. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Aune, D.; Keum, N.N.; Giovannucci, E.; Fadnes, L.T.; Boffetta, P.; Greenwood, D.C.; Tonstad, S.; Vatten, L.J.; Riboli, E.; Norat, T. Dietary intake and blood concentrations of antioxidants and the risk of cardiovascular disease, total cancer, and all-cause mortality: A systematic review and dose-response meta-analysis of prospective studies. *Am. J. Clin. Nutr.* **2018**, *108*, 1069–1091. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Abou-Elella, F.; Mourad, R. Anticancer and anti-oxidant potentials of ethanolic extracts of *Phoenix dactylifera*, *Musa acuminata* and *Cucurbita maxima*. *Res. J. Pharm. Biol. Chem. Sci.* **2015**, *6*, 710–720.
16. Amin, M.Z.; Islam, T.; Uddin, M.R.; Uddin, M.J.; Rahman, M.M.; Satter, M.A. Comparative study on nutrient contents in the different parts of indigenous and hybrid varieties of pumpkin (*Cucurbita maxima* Linn.). *Heliyon* **2019**, *5*, e02462. [\[CrossRef\]](#)
17. Batool, M.; Modassar, M.; Nawaz, A.; Roobab, U.; Manzoor, M.F.; Farooq, U.; Nadeem, H.R.; Nadeem, M.; Kanwal, R.; Abdelgawad, H.; et al. Nutritional Value, Phytochemical Potential, and Therapeutic Benefits of Pumpkin (*Cucurbita* sp.). *Plants* **2022**, *11*, 1394. [\[CrossRef\]](#)
18. Bidyalakshmi, T.; Jyoti, B.; Mansuri, S.M.; Srivastava, A.; Mohapatra, D.; Kalnar, Y.B.; Narsaiah, K.; Indore, N. Application of Artificial Intelligence in Food Processing: Current Status and Future Prospects. *Food Eng. Rev.* **2025**, *17*, 27–54. [\[CrossRef\]](#)
19. Hassoun, A.; Jagtap, S.; Garcia-Garcia, G.; Trollman, H.; Pateiro, M.; Lorenzo, J.M.; Trif, M.; Rusu, A.V.; Aadil, R.M.; Šimat, V.; et al. Food quality 4.0: From traditional approaches to digitalized automated analysis. *J. Food Eng.* **2023**, *337*, 111216. [\[CrossRef\]](#)
20. Bhagya Raj, G.V.S.; Dash, K.K. Comprehensive study on applications of artificial neural network in food process modeling. *Crit. Rev. Food Sci. Nutr.* **2022**, *62*, 2756–2783. [\[CrossRef\]](#)
21. Liu, M.; Wang, M.; Wang, J.; Li, D. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sens. Actuators B Chem.* **2013**, *177*, 970–980. [\[CrossRef\]](#)
22. Debska, B.; Guzowska-Świder, B. Application of artificial neural network in food classification. *Anal. Chim. Acta* **2011**, *705*, 283–291. [\[CrossRef\]](#)
23. Abba, S.I.; Usman, A.G.; Işık, S. Simulation for response surface in the HPLC optimization method development using artificial intelligence models: A data-driven approach. *Chemom. Intell. Lab. Syst.* **2020**, *201*, 104007. [\[CrossRef\]](#)
24. Singh, Y.R.; Shah, D.B.; Maheshwari, D.G.; Shah, J.S.; Shah, S. Advances in AI-Driven Retention Prediction for Different Chromatographic Techniques: Unraveling the Complexity. *Crit. Rev. Anal. Chem.* **2024**, *54*, 3559–3569. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Minh, L.A.C.; dos Santo, W.N.L. Heterogeneous ensemble learning applied to UV-VIS identification of multi-class pesticides by high-performance liquid chromatography with diode array detector (HPLC/DAD). *Chemom. Intell. Lab. Syst.* **2025**, *261*, 105385. [\[CrossRef\]](#)
26. Kumar, R.D.; Julie, E.G.; Robinson, Y.H.; Vimal, S.; Seo, S. Recognition of food type and calorie estimation using neural network. *J. Supercomput.* **2021**, *77*, 8172–8193. [\[CrossRef\]](#)
27. Zhang, Y.; Deng, L.; Zhu, H.; Wang, W.; Ren, Z.; Zhou, Q.; Lu, S.; Sun, S.; Zhu, Z.; Gorris, J.M.; et al. Deep learning in food category recognition. *Inf. Fusion* **2023**, *98*, 101859. [\[CrossRef\]](#)
28. de Araújo Gomes, A.; Azcarate, S.M.; Špánik, I.; Khvalbota, L.; Goicoechea, H.C. Pattern recognition techniques in food quality and authenticity: A guide on how to process multivariate data in food analysis. *Trends Anal. Chem.* **2023**, *164*, 117105. [\[CrossRef\]](#)
29. Tseng, Y.J.; Chuang, P.J.; Appell, M. When Machine Learning and Deep Learning Come to the Big Data in Food Chemistry. *ACS Omega* **2023**, *8*, 15854–15864. [\[CrossRef\]](#)
30. Xiao, Z.; Wang, J.; Han, L.; Guo, S.; Cui, Q. Application of Machine Vision System in Food Detection. *Front. Nutr.* **2022**, *9*, 888245. [\[CrossRef\]](#)
31. Miljić, M.; Rocchetti, G.; Krstić, S.; Mišan, A.; Brdar-Jokanović, M.; Marcheggiani, F.; Martinelli, E.; Lucini, L.; Damiani, E. Comparative in vitro antioxidant capacity and terpenoid profiling of pumpkin fruit pulps from a Serbian *Cucurbita maxima* and *Cucurbita moschata* breeding collection. *Antioxidants* **2021**, *10*, 1580. [\[CrossRef\]](#)
32. Morais, H.; Ramos, A.C.; Cserhádi, T.; Forgács, E. Effects of fluorescent light and vacuum packaging on the rate of decomposition of pigments in paprika (*Capsicum annuum*) powder determined by reversed-phase high-performance liquid chromatography. *J. Chromatogr. A* **2001**, *936*, 139–144. [\[CrossRef\]](#)

33. Kevrešan, Z.S.; Mastilović, J.S.; Mandić, A.I.; Torbica, A.M. Effect of different ripening conditions on pigments of pepper for paprika production at green stage of maturity. *J. Agric. Food Chem.* **2013**, *61*, 9125–9130. [[CrossRef](#)] [[PubMed](#)]
34. Krstić, S.; Miljić, M.; Antić-Stanković, J.; Božić, D.D.; Krivokuća, M.J.; Pirković, A. Pumpkin pulp extracts from a Serbian *Cucurbita maxima* Breeding Collection: Phenol profile and in vitro bioactivity. *Food Chem. Adv.* **2023**, *3*, 100395. [[CrossRef](#)]
35. Tasić, A.; Pezo, L.; Lončar, B.; Pešić, M.; Tešić, Ž.; Kalaba, M. Assessing the Impact of Botanical Origins, Harvest Years, and Geographical Variability on the Physicochemical Quality of Serbian Honey. *Foods* **2024**, *13*, 1530. [[CrossRef](#)] [[PubMed](#)]
36. Delicado, P.; Smrekar, M. Measuring non-linear dependence for two random variables distributed along a curve. *Stat. Comput.* **2009**, *19*, 255–269. [[CrossRef](#)]
37. Guilford, J.P. *Fundamental Statistics in Psychology and Education*; McGraw-Hill: New York, NY, USA, 1956. [[CrossRef](#)]
38. Aćimović, M.; Zeremski, T.; Šovljanski, O.; Lončar, B.; Pezo, L.; Zheljazkov, V.D.; Pezo, M.; Šuput, D.; Kurunci, Z. Seasonal Variations in Essential Oil Composition of *Immortelle* Cultivated in Serbia. *Horticulturae* **2022**, *8*, 1183. [[CrossRef](#)]
39. Gemperline, P. *Practical Guide to Chemometrics*; CRC Press: Boca Raton, FL, USA, 2006. [[CrossRef](#)]
40. Granato, D.; Santos, J.S.; Escher, G.B.; Ferreira, B.L.; Maggio, R.M. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends Food Sci. Technol.* **2018**, *72*, 83–90. [[CrossRef](#)]
41. Cumming, J.A.; Wooff, D.A. Dimension reduction via principal variables. *Comput. Stat. Data Anal.* **2007**, *52*, 550–565. [[CrossRef](#)]
42. Bouwmans, T.; Zahzah, E.H. Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance. *Comput. Vis. Image Underst.* **2014**, *122*, 22–34. [[CrossRef](#)]
43. Daffertshofer, A.; Lamothe, C.J.C.; Meijer, O.G.; Beek, P.J. PCA in studying coordination and variability: A tutorial. *Clin. Biomech.* **2004**, *19*, 415–428. [[CrossRef](#)]
44. Lužaić, T. Possibilities and Limitations of Cold-Pressed Oil and Cake Production Obtained from Selected New Sunflower Hybrids. Ph.D. Thesis, Faculty of Technology, University of Novi Sad, Novi Sad, Serbia, 2021.
45. Lončar, B.; Cvetanović Kljakić, A.; Arsenijević, J.; Petronijević, M.; Panić, S.; Đogo Mračević, S.; Ražić, S. Valorization of Lemon, Apple, and Tangerine Peels and Onion Skins—Artificial Neural Networks Approach. *Separations* **2025**, *13*, 9. [[CrossRef](#)]
46. Thawornwong, S.; Enke, D. The adaptive selection of financial and economic variables for use with artificial neural networks. *Neurocomputing* **2004**, *56*, 205–232. [[CrossRef](#)]
47. Lončar, B.; Pezo, L.; Ilić, M.; Kanurić, K.; Vukić, D.; Degenek, J.; Vukić, V. Modeling and optimization of herb-fortified fresh kombucha cheese: An artificial neural network approach for enhancing quality characteristics. *Foods* **2024**, *13*, 548. [[CrossRef](#)] [[PubMed](#)]
48. Kollo, T.; von Rosen, D. *Advanced Multivariate Statistics with Matrices*; Springer: Dordrecht, The Netherlands, 2005. [[CrossRef](#)]
49. Aitkin, M.; Foxall, R. Statistical modelling of artificial neural networks using the multi-layer perceptron. *Stat. Comput.* **2003**, *13*, 227–239. [[CrossRef](#)]
50. Voća, N.; Pezo, L.; Jukić, Ž.; Lončar, B.; Šuput, D.; Krička, T. Estimation of the storage properties of rapeseeds using an artificial neural network. *Ind. Crops Prod.* **2022**, *187*, 115358. [[CrossRef](#)]
51. Kojić, P.; Omorjan, R. Predicting hydrodynamic parameters and volumetric gas-liquid mass transfer coefficient in an external-loop airlift reactor by support vector regression. *Chem. Eng. Res. Des.* **2017**, *125*, 398–407. [[CrossRef](#)]
52. Šovljanski, O.; Lončar, B.; Pezo, L.; Savelijć, A.; Tomić, A.; Brunet, S.; Filipović, V.; Filipović, J.; Čanadanović-Brunet, J.; Četković, G.; et al. Unlocking the Potential of the ANN Optimization in Sweet Potato Varieties Drying Processes. *Foods* **2023**, *13*, 134. [[CrossRef](#)]
53. Abbas, H.M.K.; Huang, H.X.; Huang, W.J.; Xue, S.D.; Yan, S.J.; Wu, T.Q.; Li, J.X.; Zhong, Y.J. Evaluation of Metabolites and Antioxidant Activity in Pumpkin Species. *Nat. Prod. Commun.* **2020**, *15*, 1934578X2092098. [[CrossRef](#)]
54. Akshaya, H.R.; Namita, N.; Singh, K.P.; Saha, S.; Panwar, S.; Bharadwaj, C. Determination and correlation of carotenoid pigments and their antioxidant activities in marigold (*Tagetes* sp.) flowers. *Indian J. Agric. Sci.* **2017**, *87*, 390–396. [[CrossRef](#)]
55. Kulczyński, B.; Gramza-Michałowska, A. The Profile of Carotenoids and Other Bioactive Molecules in Various Pumpkin Fruits (*Cucurbita maxima* Duchesne) Cultivars. *Molecules* **2019**, *24*, 3212. [[CrossRef](#)]
56. Sathasivam, R.; Radhakrishnan, R.; Kim, J.K.; Park, S.U. An update on biosynthesis and regulation of carotenoids in plants. *S. Afr. J. Bot.* **2021**, *140*, 290–302. [[CrossRef](#)]
57. Demmig-Adams, B.; Hodges, A.K.; Polutchnko, S.K.; Adams, W.W. Zeaxanthin and Other Carotenoids: Roles in Abiotic Stress Defense with Implications for Biotic Defense. *Plants* **2025**, *14*, 2703. [[CrossRef](#)] [[PubMed](#)]
58. Singh, J.; Singh, V.; Shukla, S.; Rai, A.K. Phenolic Content and Antioxidant Capacity of Selected *Cucurbita* Fruits Extracted with Different Solvents. *J. Nutr. Food Sci.* **2016**, *6*, 565. [[CrossRef](#)]
59. Mitic, V.; Nikolic, J.; Dimitrijevic, M.; Stankov-Jovanovic, V.; Stojanovic, G. Application of principal component and hierarchical cluster analysis to classify different edible *Cucurbita* based on in vitro antioxidant activity. *J. Serbian Chem. Soc.* **2025**, *90*, 639–660. [[CrossRef](#)]

60. Mashitoa, F.M.; Shoko, T.; Shai, J.L.; Slabbert, R.M.; Sivakumar, D. Changes in Phenolic Metabolites and Biological Activities of Pumpkin Leaves (*Cucurbita moschata* Duchesne ex Poir.) During Blanching. *Front. Nutr.* **2021**, *8*, 641939. [[CrossRef](#)]
61. Pinna, N.; Ben Abbou, S.; Ianni, F.; Angeles Flores, G.; Pietercelie, A.; Perretti, G.; Blasi, F.; Angelini, P.; Cossignani, L. Phenolic compounds from pumpkin pulp: Extraction optimization and biological properties. *Food Chem. X* **2024**, *23*, 101628. [[CrossRef](#)]
62. Stryjecka, M.; Krochmal-Marczak, B.; Cebulak, T.; Kieltyka-Dadasiewicz, A. Assessment of Phenolic Acid Content and Antioxidant Properties of the Pulp of Five Pumpkin Species Cultivated in Southeastern Poland. *Int. J. Mol. Sci.* **2023**, *24*, 8621. [[CrossRef](#)]
63. Chen, C.; Xu, L.; Xia, W.; Qi, C.; Liu, J.; Hua, B.; Zhang, Z.; Qi, X.; Miao, M. Metabolomics and Transcription Profiling of Pumpkin Fruit Reveals Enhanced Bioactive Flavonoids and Coumarins in Giant Pumpkin (*Cucurbita maxima*). *J. Agric. Food Chem.* **2023**, *71*, 10459–10469. [[CrossRef](#)]
64. Carnà, M.; Korwin Krukowski, P.; Tosato, E.; D'Alessandro, S. Apocarotenoids as Stress Signaling Molecules in Plants. *Agriculture* **2025**, *15*, 926. [[CrossRef](#)]
65. Oliveira, T.B.; Gobbo-Neto, L.; Schmidt, T.J.; Da Costa, F.B. Study of chromatographic retention of natural terpenoids by chemoinformatic tools. *J. Chem. Inf. Model.* **2015**, *55*, 26–38. [[CrossRef](#)]
66. Guiné, R.P.F.; Matos, S.; Gonçalves, F.J.; Costa, D.; Mendes, M. Evaluation of phenolic compounds and antioxidant activity of blueberries and modelization by artificial neural networks. *Int. J. Fruit Sci.* **2018**, *18*, 199–214. [[CrossRef](#)]
67. Stajčić, S.M.; Pezo, L.L.; Četković, G.S.; Čanadanović-Brunet, J.M.; Mandić, A.I.; Tumbas Šaponjac, V.T.; Vulić, J.J.; Travičić, V.N.; Belović, M.M. Antioxidant activity according to bioactive compounds content in dried pumpkin waste. *J. Serbian Chem. Soc.* **2024**, *89*, 13–27. [[CrossRef](#)]
68. Simić, V.M.; Rajković, K.M.; Stojičević, S.S.; Veličković, D.T.; Nikolić, N.; Lazić, M.L.; Karabegović, I.T. Optimization of microwave-assisted extraction of total polyphenolic compounds from chokeberries by response surface methodology and artificial neural network. *Sep. Purif. Technol.* **2016**, *160*, 89–97. [[CrossRef](#)]
69. Lončar, B.; Pezo, L.; Knežević, V.; Nićetin, M.; Filipović, J.; Petković, M.; Filipović, V. Enhancing Cookie Formulations with Combined Dehydrated Peach: A Machine Learning Approach for Technological Quality Assessment and Optimization. *Foods* **2024**, *13*, 782. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.