ISSN: 2317-0840

Application of Molecular Topology for Predicting the Antioxidant Activity of a Group of Phenolic Compounds

Fernando de Souza Bastos[1†], Diogo da Silva Machado[2], Jaime Barros Silva Filho[3], Maria Luiza Ferreira Delfim[4].

[1] *Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas, Departamento de Estatística, Viçosa - MG.*
[2] *Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas, Departamento de Matemática, Viçosa - MG.*
[3] *University of California, College of Natural & Agricultural Sciences, Department of Microbiology and Plant Pathology, Riverside – CA, Estados Unidos.*
[2] *Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas, Departamento de Química, Viçosa - MG.*

Abstract: *The study of compounds with antioxidant capabilities is of great interest to the scientific community, as it has implications in several areas, from Agricultural Sciences to Biological Sciences, including Food Engineering, Medicine, and Pharmacy. In applications related to human health, it is known that antioxidant activity can delay or inhibit oxidative damage to cells, reducing damage caused by free radicals, helping in the treatment, or even preventing or postponing the onset of various diseases. Among the compounds that have antioxidant properties, there are several classes of phenolic compounds, which include several compounds with different chemical structures. Despite their importance, identifying and predicting the antioxidant potential of phenolic compunds remains a significant challenge due to their structural diversity and the complexity of their mechanisms of action. In this work, based on the molecular branching of compounds and their intramolecular charge distributions, and using Molecular Topology, we propose a significant topological-mathematical model to evaluate the potential of candidate compounds to have an antioxidant function. The advantage of the model is that it allows for efficient predictive analysis, assisting in the identification of promising compounds more quickly and accurately, which can accelerate the development of new antioxidants with therapeutic applications.*

Keywords: *Antioxidant; Phenolic Compounds; Molecular Topology; Topological Indices.*

Introduction

Studies and research on compounds with antioxidant activities have gained prominence in recent years. This is because antioxidant activity delays or inhibits oxidative cell damage, decreasing the damage caused by free radicals, preventing or postponing the onset of various diseases and also helping in the treatment of diseases such as diabetes, inflammatory diseases, cardiovascular disorders, associated with oxidative stress, some types of cancer and aging (Soobrattee et al. (2005), Martins et al. (2016), Yasir et al. (2016), Yoo et al. (2017), Karunakaran et al. (2018), and Muller et al. (2019)).

Plants contain phenolic acids, which are used as a food base. Vegetable leaves, seeds, and fruit peels, in general, have a higher concentration of phenolic compounds. These phenolic acids are commonly found as amides, esters, or glycosides. The research and prediction of new antioxidant molecules are critical for public health. By inhibiting free radicals, antioxidants can help to reverse cell oxidation Kumar et al. (2019 2019).

Measuring the antioxidant activity/capacity of foods and biological compounds is therefore essential not only to ensure food quality but also to assess the efficiency of dietary antioxidants in the prevention and treatment of diseases related to oxidative stress. Organic solvent extraction can be used to identify phenolic compounds in plants Khoddami et al. (2013).

Among the compounds that have antioxidant properties, there are several classes of phenolic compounds, which are those that have one or more hydroxyl groups directly linked

[†]Autor correspondente: fernando.bastos@ufv.br

Manuscrito recebido em: 12/08/2024
Manuscrito revisado em: 12/12/2024
Manuscrito aceito em: 16/12/2024

to an aromatic ring. Such compounds represent a large group of molecules with a variety of functions related to plant growth, development, and defense. They include signaling molecules, pigments, and flavors that can attract or repel, as well as compounds that can protect the plant against insects, fungi, bacteria, and viruses.

With the growing recognition of their bioactive properties, research on the antioxidant behavior of phenolic compounds-which include thousands of molecules with diverse chemical structures—has increased significantly in recent decades. Furthermore, the methods and instruments used to measure antioxidant activity/capacity have made remarkable progress. According to Munteanu et al. (2021), there are several methods for assessing this ability, which generally fall into three distinct categories, namely, spectrometry, electrochemical assays, and chromatography. However, all are based on chemical reactions that demand time and financial resources.

Therefore, the implementation of mechanisms that can provide an early estimate of the potential of a molecule to exert an antioxidant function is an interesting research strategy. Molecular Topology based on Quantitative Structure-Activity Relations (QSAR) is a tool that has great potential in this context. The QSAR technique, developed by Hansch et al. (1964), combines Topology and Statistics concepts and its approach can be described as a statistical method of data analysis to develop models that can correctly predict certain biological activity or properties of compounds based on their structure chemistry. QSAR techniques apply descriptors based on molecular structures and use algorithms to correlate the obtained descriptors with the value of the target property of interest, for more information we suggest Cramer (2012).

The development of Molecular Topology is attributed to Randic (Randic, 1975) and Kier-Hall (Hall et al., 1978 1978; Kier, 2012) and is based, above all, on the premise that, in many cases, there is a close relationship between the structures of organic compounds and many of its chemical and biological properties. From this, a set of suitable numerical characterizations for the molecules of interest is obtained. Such characterizations are designated by topological indices, which allow, after statistical treatment (specifically, Linear Discriminant Analysis), the classification into groups, with greater and lesser probabilities of having a chemical function of interest. It should be noted that Molecular Topology has been used satisfactorily in obtaining new compounds for the production of drugs, cosmetics, and agrochemicals of great interest to society, see more details in Galvez et al. (1994b) and Mahmoudi et al. (2006) and Amigó et al. (2007).

To obtain the topological indices, the molecules to be considered are mathematically modeled as Topological Graphs, naturally taking their atoms as vertices and atomic bonds as edges. Thus, from the graphs, numerical matrices are obtained containing information about the molecules (such as the structural arrangement, global charge transfer and between pairs of atoms, atomic electronegativity, among others), which, through specific formulas, lead to the survey of topological indices.

Despite the growing recognition of the bioactive properties of phenolic compounds and the significant advances in methods to measure antioxidant activity, identifying and predicting the antioxidant potential of these compounds remains a considerable challenge. This difficulty arises from the structural diversity of phenolic compounds, which encompasses thousands of molecules with distinct chemical properties, and the complexity of their mechanisms of action. Traditional methods for assessing antioxidant capacity, while effective, are time-consuming and require considerable financial and material resources, limiting their applicability in large-scale screening efforts.

Furthermore, the lack of efficient and cost-effective tools for early-stage prediction of antioxidant potential hinders the identification and development of new antioxidant molecules with therapeutic or industrial applications. This gap underscores the need for innovative methodologies that can streamline the evaluation process and provide reliable predictions based on molecular characteristics.

Utilizing Molecular Topology and Linear Discriminant Analysis, the objective of this study was to develop a predictive QSAR model to investigate the antioxidant activity of a group of phenolic compounds.

Materials and Methods

Compounds Analyzed and Their Antioxidant Activities

Phenolic compounds are the most abundant antioxidants in the human diet. They have a considerable structural diversity, characterized by the hydroxyl groups on aromatic rings. According to the number of phenol rings and the structural elements that bind rings to one another, such compounds are grouped and classified as simple phenols, phenolic acids, flavonoids, xanthones, stilbenes, and lignans. They are widely distributed in plants and play important roles in defense against oxidative stress Vuolo et al. (2019).

Phenolic compounds range from simple structures, such as phenolic acids, to complex polymers, such as lignins and tannins. Cheng et al. (2002) reports that this structural diversity is associated with different antioxidant mechanisms, such as free radical scavenging, hydrogen atom or electron donation, and the ability to chelate transition metals, which prevents oxidation chain reactions.

It is important to note that phenolic compounds combat oxidative stress in several ways:

- Soobrattee et al. (2005) highlights that the neutralization of Reactive Oxygen Species: They donate electrons or hydrogen atoms to neutralize free radicals such as the hydroxyl radical and superoxide anion.
- Transition Metal Chelation: By binding to metal ions such as iron and copper, phenolics prevent them from catalyzing the formation of free radicals via Fenton reactions Khokhar et al. (2003).
- Modulation of Cell Signaling Pathways: Some phenolics regulate the expression of antioxidant genes (Nguyen et al. (2003 2003)).

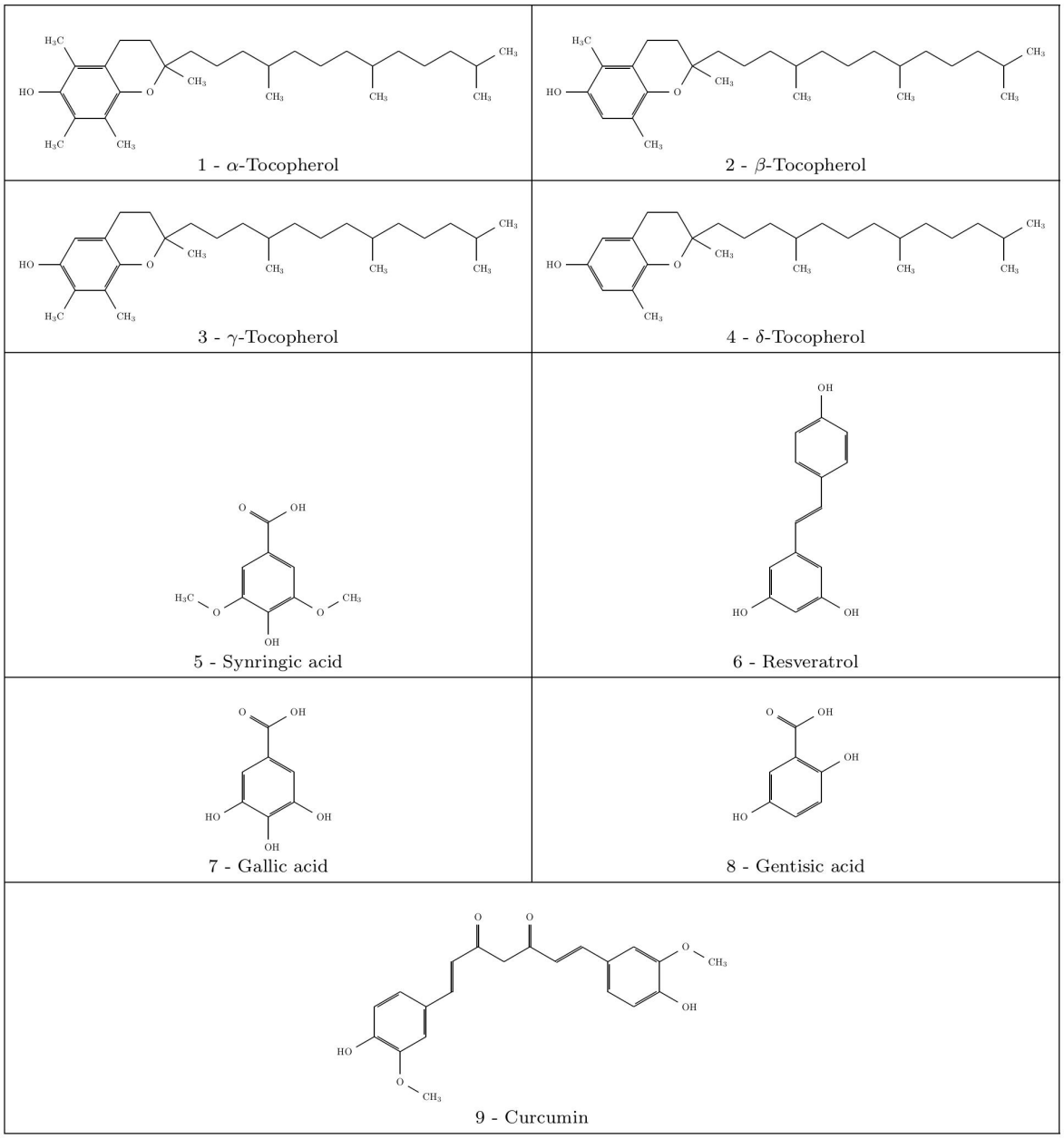
According to Soobrattee et al. (2005), the intake of phenolic compounds has been associated with a reduced risk of chronic diseases, such as cardiovascular, neurodegenerative and cancer. Compounds such as quercetin and gallic acid have high antioxidant activity, being more effective than synthetic antioxidants in some tests.

With their diverse structures and multiple antioxidant mechanisms, phenolic compounds play a pivotal role in mitigating oxidative stress, with significant implications for both human health and industrial applications. Charts 1, 2, and 3 present the phenolic compounds analyzed in this study, selected for their reported antioxidant activity and structural variability. The topological indices, described in subsection *Topological Load Indexes*, were calculated for the compounds listed in Charts 1 and 2. These indices served as molecular descriptor vectors to estimate the discriminant function, facilitating the differentiation of compounds with high antioxidant activity from those with low or negligible activity.

The choice of topological indices was guided by their sensitivity to specific molecular features, such as connectivity and branching, which are critical for the antioxidant mechanisms of phenolic compounds. Chart 1 presents the compounds with experimentally confirmed antioxidant capacity, highlighting their structural variations. In contrast, Chart 2 includes compounds with similar structural frameworks but minimal or no measurable antioxidant activity, serving as a comparative set. Chart 3 consolidates additional relevant properties and indices for all compounds studied, facilitating a comprehensive analysis.

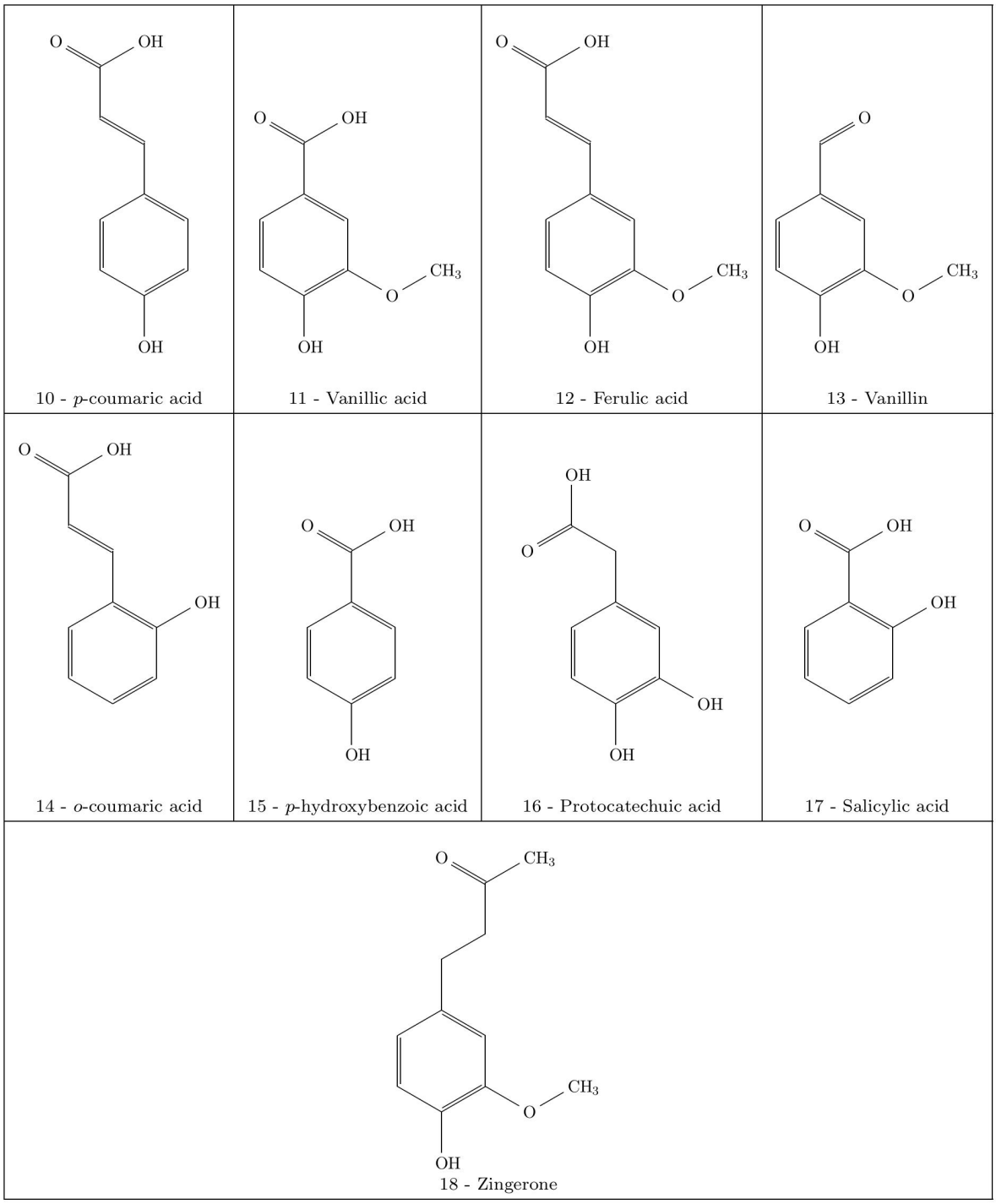
In Chart 3, we present a group of five phenolic compounds with confirmed antioxidant capacity. These compounds were selected as the test group to validate the discriminant function developed in this study.

Chart 1: Structural formulas of phenolic compounds used in the present study, classified with non-active antioxidant capacity.



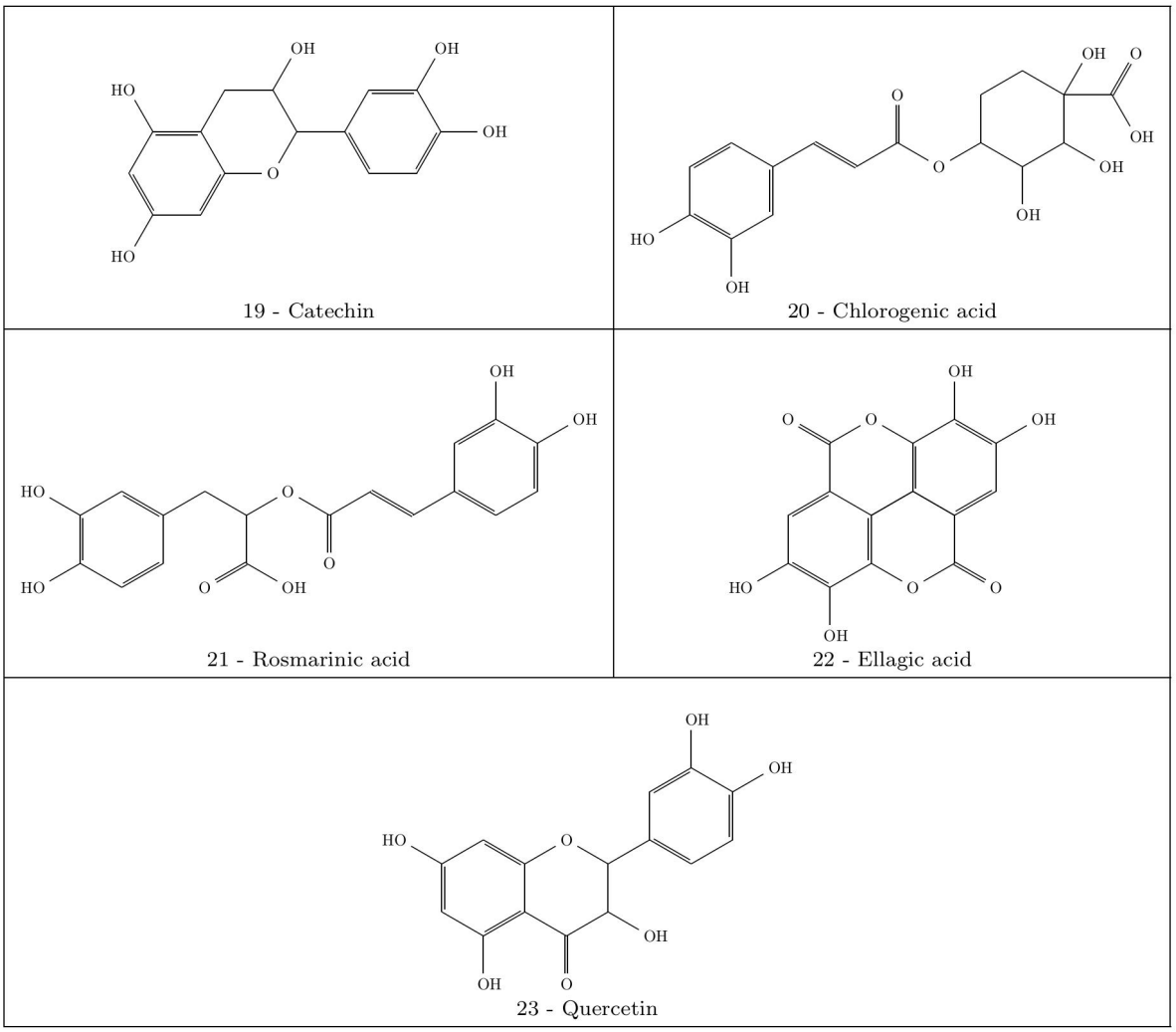
Source: from the authors (2024).

Chart 2: Structural formulas of phenolic compounds used in the present study, classified with non-active antioxidant capacity.



Source: from the authors (2024).

Chart 3: Structural formulas of phenolic compounds used in the present study as a test group, all with antioxidant capacity.



Source: from the authors (2024).

Calculation of Topological Indices

Topological indices are numerical functions of molecular graphs and are considered important molecular descriptors. According to Vasilyev et al. (2014), in addition to being graph invariant, topological indices do not consider information about molecular geometry, such as bond lengths, bond angles, or twist angles, but instead, encode information about adjacencies of atoms and branches within a molecule. Also, according to the same authors, since the computation of topological indices uses fewer resources than the computation of those molecular descriptors that also take into account molecular geometry, topological indices have gained considerable popularity and many new topological indices have been proposed and studied in the literature specialized in recent years.

We assume that the following aspects of molecules are relevant to the investigation of the occurrence of antioxidant activity: the type of molecular branching and the distribution of intramolecular charge. Therefore, in this work, we use two types of topological indices: the *Randic index* χ^1 and the *topological load indices* G_4, J_2 and J_5.

All the topological indices used in this work were obtained with the aid of an appropriate python algorithm that is available in the supplementary material that accompanies this article.

Randic Index

According to Gutman et al. (2018), among the several hundred descriptors of molecular structures based on graphs (see Todeschini et al. (2008)), the Randic index is certainly the most widely applied in chemistry and pharmacology. This index characterizes the branching of the molecular graph and was introduced by Milan Randic (Randic, 1975). If $E(G)$ denotes the set of edges of the molecular graph G, then the Randic index of G is defined by

$$\chi^1(G) = \sum_{e_{ij} \in E(G)} (deg_i \cdot deg_j)^{-1/2}, \tag{1}$$

where, deg_i and deg_j are the degrees of vertices i and j, respectively.

Topological Load Indexes

Topological Charge Indices were introduced in the literature by Galvez et al. (1994a) and have the ability to describe molecular charge distribution. In fact, given a molecular graph G, let $A(e)$ be its adjacency matrix (modified), with the relative electronegativity of each atom in the main diagonal entries and D^* the distance matrix inverse square of G, with entries on the main diagonal taking on a value of zero. Consider $M = [m_{ij}]$ the square matrix of order N (where N is the number of vertices of G) defined by

$$M = A(e) \times D^* \tag{2}$$

and take, for each i, j, with $1 \leq i, j \leq N$ the *charge term* CT_{ij} defined by

$$CT_{ij} = m_{ij} - m_{ji}. \tag{3}$$

The (valency) *topological load indices*, G_k, with $1 \leq k \leq N-1$ is defined by

$$G_k = \sum_{i,j=i+1}^{i=N-1,j=N} \delta(k, d_{ij}) \cdot \mid CT_{ij} \mid, \tag{4}$$

where δ is the Kronecker delta function ($\delta(p, q) = 1$, if $p = q$ and 0 otherwise) and d_{ij} denote the entries of the matrix of topological distance. Remember that in the main diagonal entries of the matrix $A(e)$ the relative electronegativity of an element Q can be calculated by the formula

$$Re^-(Q) = \lambda \cdot (e^-(Q) - e^-(C)), \tag{5}$$

where $e^-(Q)$ and $e^-(C)$ denote, respectively, the Pauling electronegativity of the element Q and the carbon atom C. The value λ constitutes the conversion factor. In this work, we consider two values for λ: 2.2 (using hydrogen normalization = 0.77) and 3.28 (using chlorine normalization = 2). For this reason, to distinguish the two respective situations, we will use the symbols $G_k^{2.2}$ and $G_k^{3.28}$ to indicate the considered conversion factor.

On the other hand, the index J_k is defined by,

$$J_k = \frac{G_k}{N-1}, \quad k = 1, \cdots, N-1 \tag{6}$$

which measures the average value of charge transfer for each chemical bond in the substance.

Discriminant Analysis

Discriminant Analysis is a statistical method used to classify observations into predefined groups based on a set of predictor variables. The goal is to derive a discriminant function that best separates the groups by maximizing the between-group variance relative to the within-group variance. This technique is particularly useful when the groups are known a priori, and the objective is to assign new observations to one of these groups.

The discriminant function is a linear combination of the predictor variables that provides the best separation between two or more groups. The function is expressed as:

$$DF = w_1x_1 + w_2x_2 + \cdots + w_nx_n + c, \tag{7}$$

where w_i are the coefficients (weights) for the predictor variables x_i, and c is a constant. These coefficients are calculated to maximize the ratio of the variance between groups to the variance within groups.

Once the discriminant function is established, it can be used to classify new observations. For a two-group problem, let the discriminant scores for the two groups be denoted by D_1 and D_2. The classification rule is defined as:

$$\text{Assign to Group 1 if } D \geq D_c, \quad \text{Assign to Group 2 if } D < D_c, \tag{8}$$

where D_c is the classification threshold, typically chosen as the midpoint between the group centroids:

$$D_c = \frac{\mu_1 + \mu_2}{2}, \tag{9}$$

with μ_1 and μ_2 being the mean discriminant scores of Group 1 and Group 2, respectively.

For multi-group problems, the classification involves calculating the discriminant scores for each group and assigning the observation to the group with the highest score.

In this study, Linear Discriminant Analysis (LDA) was applied to classify phenolic compounds into groups with and without antioxidant activity. The predictor variables were the topological indices derived from the molecular structures of the compounds. Fisher's discriminant function was used to compute the scores for each compound, and the classification rule was applied to assign compounds to their respective groups. This approach not only provided a quantitative method for distinguishing between active and non-active compounds but also facilitated the identification of molecular features contributing to antioxidant activity.

To enhance the robustness of the Discriminant Analysis, we calculated a set of topological indices for each phenolic compound. These indices quantify molecular features such as connectivity, branching, and electron distribution, capturing structural aspects relevant to antioxidant activity. For classification, the experimental data consisted of antioxidant activity determined by DPPH free radical scavenging and lipid peroxidation inhibitory effects expressed as the concentration of 50% lipid peroxidation inhibition ($IC_{50}/\ mmol\ , L^{-}1$).

QSAR Algorithms: Linear Discriminant Analysis

To detect the presence of the antioxidant function in phenolic compounds, the QSAR prediction model by Linear LDA was established, a branch of multivariate statistics used in problems of discrimination and classification of categories or objects, which appears in - Fisher's seminal work Fisher (1936), and is currently a widespread topic Johnson et al. (1992) and Khattree et al. (2000). In addition, several works use Discriminant Analysis in establishing QSAR models, see Cronin et al. (1994), Contrera et al. (2005), Konovalov et al. (2008), Ayoub et al. (2018a), and Lu et al. (2022).

Although Discriminant Analysis cannot provide concrete predicted values of the antioxidant effect, it can determine the likelihood of classifying compounds as active or inactive for antioxidant activity and thus aid in the discovery and development of efficient antioxidants.

Therefore, we obtained a discriminant function (DF) that allows a classification between compounds with active and inactive antioxidant activity. This function linearly depends on the topological indices considered, which are assumed to be usable in distinguishing compounds.

In the set of phenolic compounds used to obtain the discriminant function, we separated each compound into two groups: the first formed by compounds with proven antioxidant activity and the second by compounds without antioxidant activity. The linear discriminant analysis (LDA) was applied using the statistical software R Core Team (2024).

Results and Discussion

To enhance the robustness of the Discriminant Analysis, a set of topological indices was calculated for each phenolic compound in both the training and test groups. These indices, presented in Table 1, include features such as connectivity, branching, and electron distribution, and were used as predictive variables to establish a discriminant function (DF) capable of classifying compounds as antioxidant or non-antioxidant.

Table 1: Topological indices related to the phenolic compounds presented in Charts 1 and 2, divided by group, as used in the Discriminant Analysis.

Compounds no.	χ^1	$G_4^{3.28}$	$J_2^{3.28}$	$J_5^{3.28}$	$G_4^{2.2}$	$J_2^{2.2}$	$J_5^{2.2}$
Antioxidant Group Training							
1	13.72456	3.59624	0.33965	0.05315	3.41601	0.33924	0.05059
2	13.19576	3.04046	0.3174	0.05595	2.92031	0.31262	0.05329
3	13.19576	3.07157	0.3174	0.05666	2.95142	0.31262	0.05401
4	12.77311	2.51578	0.26956	0.05174	2.45571	0.2603	0.049
5	5.35317	1.55516	0.41686	0.06137	1.14092	0.4076	0.04784
6	5.63077	0.59578	0.26229	0.01284	0.53571	0.3079	0.00993
7	4.37239	0.67669	0.46465	0.02008	0.59762	0.51897	0.01968
8	3.84159	0.84513	0.42525	0.0125	0.73126	0.43543	0.0125
9	9.83297	2.81071	0.5237	0.06254	2.5695	0.51444	0.04775
Non-antioxidant Group Training							
10	4.0856	0.57801	0.54821	0.04045	0.51793	0.52525	0.02297
11	4.40225	1.18803	0.4169	0.03334	1.01364	0.40595	0.03293
12	5.06293	1.14313	0.50661	0.05282	1.102253	0.49734	0.03508
13	3.99562	1.17424	0.34116	0.0303	1.05364	0.35315	0.025
14	3.89175	0.76889	0.42424	0.03879	0.76889	0.45645	0.03237
15	3.42492	0.71179	0.44781	0.01066	0.53157	0.41975	0.01016
16	4.29869	1.01402	0.45592	0.0397	0.83379	0.47666	0.02371
17	3.4309	0.74291	0.45847	0.00694	0.56268	0.44444	0.00694
18	5.34935	1.43179	0.47242	0.06691	1.37126	0.46315	0.05508
Test Group							
19	8.1157	1.84068	0.31414	0.04914	1.84651	0.33093	0.03719
20	9.89595	2.43829	0.52694	0.03863	1.83233	0.43631	0.02781
21	9.40614	2.59738	0.50404	0.06857	2.11678	0.50306	0.04338
22	8.02602	3.50632	0.47763	0.05734	2.66436	0.5331	0.04339
23	7.76622	2.52803	0.33478	0.04599	246.000	0.40183	0.04106

Source: from the authors (2024).

The discriminant function derived in this study is defined as:

$$\mathbf{DF} = -31.023674 + 4.4419931\,\chi^{\mathbf{1}} + 41.2828013\,\mathbf{G_4^{3.28}} - 151.5908411\,\mathbf{J_2^{3.28}} - \\ - 270.9518622\,\mathbf{J_5^{3.28}} - 52.570601\,\mathbf{G_4^{2.2}} + 180.1409685\,\mathbf{J_2^{2.2}} + 310.4418446\,\mathbf{J_5^{2.2}}$$

this function combines the topological indices into a single metric, DF, which determines the classification of each compound. Compounds are classified as antioxidant if $DF \geq 31.02367$; otherwise, they are classified as non-antioxidant.

The discriminant function was first applied to the training group, which consisted of 18 compounds (9 antioxidants and 9 non-antioxidants). The classification results showed an overall accuracy of 94.4%, with 89% of the antioxidant group and 100% of the non-antioxidant group correctly classified.

The model was then validated using a test group of five phenolic compounds. In this independent evaluation, the model achieved 80% accuracy, with only one antioxidant compound, Catechin, misclassified as non-antioxidant ($DF < 31.02367$). Table 1 presents the classification results for the test group.

Table 2: Classification of compounds after application of the discriminant function.

	Compounds	Discriminant Function	Classification
1	Catechin	25.19	Non−antioxidant
2	Chlorogenic acid	45.18	Antioxidant
3	Rosmarinic acid	46.83	Antioxidant
4	Ellagic acid	61.90	Antioxidant
5	Quercetin	31.46	Antioxidant
6	α-Tocopherol	40.77	Antioxidant
7	β-Tocopherol	40.20	Antioxidant
8	γ-Tocopherol	39.88	Antioxidant
9	δ-Tocopherol	38.72	Antioxidant
10	Curcumin	35.79	Antioxidant
11	Resveratrol	36.75	Antioxidant
12	Gallic acid	39.66	Antioxidant
13	Gentisic acid	27.98	Non−Antioxidant
14	Synringic acid	36.46	Antioxidant
15	*p*-coumaric acid	22.47	Non−Antioxidant
16	Vanillic acid	26.43	Non−Antioxidant
17	Ferulic acid	21.11	Non−Antioxidant
18	Vanillin	22.29	Non−Antioxidant
19	*o*-coumaric acid	26.06	Non−Antioxidant
20	*p*-hydroxybenzoic acid	24.65	Non−Antioxidant
21	Protocatechuic acid	30.48	Non−Antioxidant
22	Salicylic acid	27.17	Non−Antioxidant
23	Zingerone	21.57	Non−Antioxidant

Source: from the authors (2024).

These results demonstrate that the discriminant function effectively separates antioxidant from non-antioxidant compounds based on their topological indices. The high accuracy in the training group and the good performance in the test group indicate the robustness of the proposed QSAR model. The misclassification of Catechin, a known antioxidant, highlights potential areas for refinement, such as incorporating additional molecular descriptors or exploring non-linear discriminant functions.

The predictive power of the model underscores the potential of combining Molecular Topology and Discriminant Analysis for evaluating antioxidant activity in phenolic compounds. This approach offers a cost-effective and efficient alternative to experimental methods, facilitating the early identification of promising antioxidant candidates for therapeutic and industrial applications. Tables 1 and 2 provide a detailed breakdown of the indices and classification results.

Comparison with the Literature on QSAR for Phenolics

Previous studies confirm that topological indices are powerful tools for predicting biological activities of phenolic compounds. For instance, Chen et al. (2015) utilized a quantitative structure–activity relationship approach was used to build a multiple linear regression model describing the dependence of antioxidant activity on structure of compounds, using features exclusively related to their topology. Our work align with this the approach of these authors, highlighting the relevance of global indices like $G_4^{3.28}$ and $J_5^{3.28}$ for capturing antioxidant activity patterns.

In comparison with the study by Spiegel et al. (2020), which developed QSAR models for antioxidants in polymeric materials, our approach offers a simpler discriminant function but achieves similar classification accuracy. Both approaches emphasize the importance of structural descriptors in modeling antioxidant activity. Additionally, our method focuses on phenolic compounds relevant to dietary sources, complementing their work on synthetic polymers.

The study by Nagarajan et al. (2020) applied QSAR models to identify antioxidants in natural extracts, achieving a predictive accuracy of approximately 85%. While their approach incorporated non-linear models and more diverse descriptors, our linear discriminant function provides comparable performance with a more interpretable structure, suggesting that topological indices alone can be sufficient for high-accuracy predictions in specific contexts.

The work of Laganà Vinci et al. (2024), which utilized chromatographic data for QSAR modeling, demonstrates the potential of combining experimental and computational methods for detailed profiling of phenolic antioxidants. Compared to our study, their model offers higher resolution in identifying subtle differences between structurally similar compounds. However, our approach is advantageous for large-scale screening, as it avoids the need for experimental chromatographic data.

The specific QSAR model developed by Ayoub et al. (2018b) for proteasome inhibitors from *Olea europaea* and *Ficus carica* highlights the utility of tailored models for unique compound classes. Their focus on specialized descriptors for proteasome activity from plant source is analogous to our use of topological indices for antioxidant activity, emphasizing the adaptability of QSAR methodologies to different biochemical properties.

The misclassification of Catechin as non-antioxidant can be attributed to limitations in the linearity of the model or the insufficiency of descriptive variables specific to highly active compounds. This aligns with observations by Soobrattee et al. (2005), who highlighted that compounds like flavonoids exhibit more complex antioxidant mechanisms, such as metal chelation and modulation of cellular signaling pathways, which may not be fully captured by purely topological metrics.

The use of topological indices in Discriminant Analysis proves promising for the initial screening of compounds with antioxidant potential. This approach can accelerate the discovery of new antioxidants for applications in food, pharmaceuticals, and cosmetics, reducing the time and costs associated with traditional experimental methods.

Future works include:

- Expanding the dataset to include structurally diverse compounds to improve model generalization.

- Investigating non-linear approaches, such as Quadratic Discriminant Analysis or Neural Networks, to capture more complex interactions.
- Integrate additional physicochemical properties, such as polarizability and acidity constants, to complement topological indices.

These improvements could help overcome current limitations and provide a more robust approach to QSAR modeling of phenolic compounds.

Conclusions

In this study, Discriminant Analysis was employed to establish a QSAR model incorporating topological indices to predict the antioxidant activity of 23 known phenolic compounds. The model demonstrated high accuracy, correctly classifying 94.4% of the training set and 80% of the test set. The use of topological indices, particularly sensitive to molecular branching (χ^1) and intramolecular charge distributions (G_4, J_2, and J_5), proved to be a robust approach for capturing structural features relevant to antioxidant activity.

The proposed methodology provides a cost-effective and efficient alternative to experimental antioxidant assays, enabling the preliminary screening of potential antioxidant compounds. By identifying promising candidates computationally, this approach can save time and financial resources, accelerating the development of new antioxidants for applications in food, pharmaceuticals, and cosmetics.

However, the study also highlights certain limitations, such as the misclassification of Catechin, which may reflect the need for additional descriptors or non-linear modeling techniques to account for complex antioxidant mechanisms. Future work should focus on expanding the dataset to include a more diverse range of phenolic compounds, integrating additional physicochemical properties, and exploring advanced computational techniques, such as machine learning models, to improve prediction accuracy and generalizability.

In conclusion, the combination of Discriminant Analysis and Molecular Topology demonstrated strong potential for QSAR modeling of phenolic compounds. This approach represents a valuable tool for researchers, providing early-stage insights into the antioxidant properties of molecular compounds and paving the way for further advancements in the field of computational chemistry and bioactivity prediction.

Acknowledgment

The second author extends gratitude to the University of California, Riverside, for its hospitality.

References

Amigó, José Manuel et al. (2007). “Topología molecular”. In: Boletín de la Sociedad Española de Matemática Aplicada. N. 39 (2007).

Ayoub, Lahmadi et al. (2018a). “A specific QSAR model for proteasome inhibitors from Olea europaea and Ficus carica”. In: Bioinformation 14.7, p. 384.

Ayoub, Lahmadi et al. (July 2018b). “A specific QSAR model for proteasome inhibitors from Olea europaea and Ficus carica”. In: Bioinformation 14, pp. 384–392.

Chen, Yu-Zhen et al. (2015). "Structure-Thermodynamics-Antioxidant Activity Relationships of Selected Natural Phenolic Acids and Derivatives: An Experimental and Theoretical Evaluation". In: PLoS ONE 10. doi: 10.1371/journal.pone.0121276.

Cheng, Zhiyong et al. (2002). "Study on the multiple mechanisms underlying the reaction between hydroxyl radical and phenolic compounds by qualitative structure and activity relationship." In: Bioorganic & medicinal chemistry 10 12, pp. 4067–73. doi: 10.1016/S0968-0896(02)00267-5.

Contrera, Joseph F et al. (2005). "QSAR modeling of carcinogenic risk using discriminant analysis and topological molecular descriptors". In: Current Drug Discovery Technologies 2.2, pp. 55–67.

Cramer, Richard D (2012). "The inevitable QSAR renaissance". In: Journal of computer-aided molecular design 26.1, pp. 35–38.

Cronin, MTD and DA Basketter (1994). "Multivariate QSAR analysis of a skin sensitization database". In: SAR and QSAR in Environmental Research 2.3, pp. 159–179.

Fisher, Ronald A (1936). "The use of multiple measurements in taxonomic problems". In: Annals of eugenics 7.2, pp. 179–188.

Galvez, Jorge et al. (1994a). "Charge indexes. New topological descriptors". In: Journal of Chemical Information and Computer Sciences 34.3, pp. 520–525.

Galvez, Jorge et al. (1994b). "Topological approach to analgesia". In: Journal of Chemical Information and Computer Sciences 34.5, pp. 1198–1203.

Gutman, Ivan, Boris Furtula, and Vladimir Katanicc (2018). "Randic index and information". In: AKCE International Journal of Graphs and Combinatorics 15.3, pp. 307–312. issn: 0972-8600.

Hall, Lowell H and Lemont B Kier (1978). "Molecular connectivity and substructure analysis". In: Journal of pharmaceutical sciences 67.12, pp. 1743–1747.

Hansch, Corwin and Toshio Fujita (1964). "$p - \sigma - \pi$ Analysis. A Method for the Correlation of Biological Activity and Chemical Structure". In: Journal of the American Chemical Society 86.8, pp. 1616–1626.

Johnson, Richard A, Dean W Wichern, et al. (1992). "Applied multivariate statistical analysis". In: New Jersey 405.

Karunakaran, Thiruventhan et al. (2018). "Nitric oxide inhibitory and anti-Bacillus activity of phenolic compounds and plant extracts from Mesua species". In: Revista Brasileira de Farmacognosia 28, pp. 231–234.

Khattree, Ravindra and Dayanand N Naik (2000). "Multivariate data reduction and discrimination". In: SAS Institute, Cary, North Carolina.

Khoddami, Ali, Meredith A Wilkes, and Thomas H Roberts (2013). "Techniques for analysis of plant phenolic compounds". In: Molecules 18.2, pp. 2328–2375.

Khokhar, Santosh and Richard.K. Owusu Apenten (2003). "Iron binding characteristics of phenolic compounds: some tentative structure–activity relations". In: Food Chemistry 81.1, pp. 133–140. doi: https://doi.org/10.1016/S0308-8146(02)00394-1.

Kier, Lemont (2012). Molecular connectivity in chemistry and drug research. Vol. 14. Elsevier.

Konovalov, Dmitry A et al. (2008). "Robust cross-validation of linear regression QSAR models". In: Journal of chemical information and modeling 48.10, pp. 2081–2094.

Kumar, Naresh and Nidhi Goel (2019). "Phenolic acids: Natural versatile molecules with promising therapeutic applications". In: Biotechnology Reports 24, e00370.

Lagana Vinci, Roberto et al. (2024). "Prediction of retention data of phenolic compounds by quantitative structure retention relationship models under reverse-phase liquid chromatography". In: Journal of Chromatography A 1730, p. 465146.

Lu, Ang et al. (2022). "QSAR study of phenolic compounds and their anti-DPPH radical activity by discriminant analysis". In: Scientific Reports 12.1, pp. 1–9.

Mahmoudi, Nassira et al. (2006). "Identification of new antimalarial drugs by linear discriminant analysis and topological virtual screening". In: Journal of Antimicrobial Chemotherapy 57.3, pp. 489–497.

Martins, Natalia, Lillian Barros, and Isabel CFR Ferreira (2016). "In vivo antioxidant activity of phenolic compounds: Facts and gaps". In: Trends in Food Science & Technology 48, pp. 1–12.

Muller, Ashley G et al. (2019). "Delivery of natural phenolic compounds for the potential treatment of lung cancer". In: DARU Journal of Pharmaceutical Sciences 27.1, pp. 433–449.

Munteanu, Irina Georgiana and Constantin Apetrei (2021). "Analytical methods used in determining antioxidant activity: A review". In: International Journal of Molecular Sciences 22.7, p. 3380.

Nagarajan, Subhalakshmi et al. (2020). "Antioxidant Activity of Synthetic Polymers of Phenolic Compounds". In: Polymers 12.8. doi: 10.3390/polym12081646.

Nguyen, Truyen, Philip J. Sherratt, and C. B. Pickett (2003). "Regulatory mechanisms controlling gene expression mediated by the antioxidant response element." In: Annual review of pharmacology and toxicology 43, pp. 233–60.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. url: https://www.R-project.org/.

Randic, Milan (1975). "Characterization of molecular branching". In: Journal of the American Chemical Society 97.23, pp. 6609–6615.

Soobrattee, Muhammed Asin et al. (2005). "Phenolics as potential antioxidant therapeutic agents: mechanism and actions". In: Mutation Research/Fundamental and Molecular mechanisms of mutagenesis 579.1-2, pp. 200–213.

Spiegel, Maciej et al. (2020). "Antioxidant activity of selected phenolic acids–ferric reducing antioxidant power assay and QSAR analysis of the structural features". In: Molecules 25.13, p. 3088.

Todeschini, Roberto and Viviana Consonni (2008). Handbook of molecular descriptors. John Wiley & Sons.

Vasilyev, Alexander and Dragan Stevanovic (2014). "MathChem: a Python package for calculating topological indices". In: MATCH Commun. Math. Comput. Chem 71, pp. 657–680.

Vuolo, Milena Morandi, Verena Silva Lima, and Maario Roberto Maroostica Junior (2019). "Chapter 2 - Phenolic Compounds: Structure, Classification, and Antioxidant Power". In: Bioactive Compounds. Ed. by Maira Rubi Segura Campos. Woodhead Publishing, pp. 33–50.

Yasir, Muhammad, Bushra Sultana, and Matthew Amicucci (2016). "Biological activities of phenolic compounds extracted from Amaranthaceae plants and their LC/ESI-MS/MS profiling". In: Journal of Functional Foods 26, pp. 645–656.

Yoo, Sae-Rom et al. (2017). "Simultaneous determination and anti-inflammatory effects of four phenolic compounds in Dendrobii Herba". In: Natural Product Research 31.24, pp. 2923–2926.