Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Harmonic Mean Cordial Labeling and Topological Descriptors for Graph Characterization and Machine Learning-Based Molecular Property Prediction

J. Gowri[1] , J. Jayapriya[2]

[1]Research scholar, Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India. E mail: dd.gowridaran@gmail.com
[2]Department of Mathematics, Sathyabama Institute of Science and Technology, Tamil Nadu, Chennai India. Email id: Priyanandam1975@rediffmail.com
Corresponding author : dd.gowridaran@gmail.com

Abstract
Graph labeling provides a rigorous mathematical framework for representing structural properties of graphs, while topological descriptors offer numerical measures of graph structure with important applications in chemical graph theory. Existing cordial and harmonic-mean-based labeling approaches primarily focus on establishing labeling properties for specific graph families, with limited integration of harmonic-mean cordial labeling and descriptor development for systematic molecular characterization and quantitative structure–property relationship (QSPR) analysis. This research develops a mathematically rigorous labeling framework and derives topology-sensitive descriptors for characterizing general graphs and molecular structures. First, fundamental graph-theoretic properties and cordial labeling conditions are formulated. A Harmonic Mean Cordial (HMC) labeling is then defined by incorporating harmonic-mean relationships between vertex-associated labels to construct induced edge labels. The existence conditions of HMC labeling are established for selected graph families through lemmas, propositions, and theorem-based proofs. Subsequently, HMC-based topological descriptors are formulated using vertex degrees and harmonic-mean label contributions. Their mathematical properties, bounds, and relationships with conventional degree-based descriptors are analytically investigated. Molecular structures are represented as molecular graphs. Finally, correlation analysis and multiple linear regression-based QSPR modeling are employed to examine the relationship between HMC descriptors and selected physicochemical properties. The analysis establishes HMC labeling conditions and descriptor formulations, while regression analyses evaluate their ability to characterize molecular structures and predict selected properties. Experimental results achieve 0.991 for molar volume, 0.988 for polarizability, 0.742 for PSA, and 0.993 for boiling point. The proposed approach connects graph structure with molecular properties and extends the application of graph-labeling theory to chemical graph theory.

Keywords: Harmonic Mean Cordial Labeling, Topological Descriptors, Quantitative Structure–Property Relationship, Graph Characterization, Molecular Property Prediction.

1. Introduction

Graph theory provides a rigorous mathematical framework for representing and analyzing relationships among interconnected objects through vertices and edges, enabling systematic investigation of structural properties such as connectivity, adjacency, degree distribution, paths, and cycles [1]. Graph labeling assigns numerical or symbolic values to vertices, edges, or both according to predefined rules, providing a mathematical mechanism for representing graph structures and examining relationships among their components [2]. Among different labeling techniques, cordial labeling establishes balanced conditions for vertex and induced edge labels, supporting the structural characterization of various graph families [3]. Harmonic mean-based labeling incorporates harmonic mean relationships into edge-label construction, providing an alternative approach for connecting vertex-associated labels with induced edge values [4]. The development of Harmonic Mean Cordial (HMC) labeling extends these concepts by integrating harmonic mean relationships with cordial labeling conditions and establishing systematic conditions for selected graph families [5]. Such labeling properties provide additional mathematical information for graph characterization and structural comparison [6]. Topological descriptors convert graph structural characteristics into numerical measures, enabling quantitative representation of connectivity, degree

relationships, adjacency patterns, and other topological features [7]. Degree-based descriptors utilize vertex degrees and their combinations to quantify structural variations among graphs and have significant applications in graph-theoretical analysis [8]. Harmonic mean-based descriptors further incorporate reciprocal degree relationships to provide complementary measures of graph structure and connectivity [9]. These concepts are particularly important in chemical graph theory, where molecular structures are represented as graphs with atoms as vertices and chemical bonds as edges, enabling mathematical analysis of molecular topology and connectivity [10]. Molecular topological descriptors provide numerical representations of molecular structures and facilitate quantitative comparison among different compounds [11]. The application in Quantitative Structure–Property Relationship (QSPR) analysis enables statistical relationships between molecular descriptors and physicochemical properties [12]. Correlation analysis evaluates descriptor–property relationships, while multiple linear regressions determine the contribution of descriptors to molecular property prediction. Thus, integrating HMC labeling with topology-sensitive descriptor development provides an interpretable mathematical framework for graph characterization, molecular structure representation, and physicochemical property prediction, extending harmonic mean-based graph concepts into chemical graph theory and QSPR applications [13]. The limitations include dependence on selected graph families, restricted molecular datasets, possible descriptor redundancy, linear QSPR assumptions, limited external validation, and reduced generalizability across diverse molecular structures and properties.

The QSPR analysis shows that the use of topology-based descriptors would represent the molecular structural aspects with the application of detour eccentricities in molecular representation and prediction of properties [14]. The use of degree-based topological indices provides good mathematical tools for molecular characterization. The possibility of using generalized multiplicative Zagreb indices in QSPR to connect molecular structures with their properties [15].

The main aim of this research is to develop Harmonic Mean Cordial (HMC) labelling and topology-sensitive descriptors for graph characterization and molecular structure analysis, and to assess their effectiveness in predicting physicochemical properties through QSPR modelling.The proposed QSPR framework predicts the physicochemical properties of four bioactive acids: ferulic acid, caffeic acid, p-hydroxybenzoic acid, and gallic acid, using HMC-based topological descriptors[19],[20].

2. PRELIMINARIES AND MATHEMATICAL FOUNDATIONS

The basic mathematical ideas needed for the creation of Harmonic Mean Cordial (HMC) labelling and the subsequent creation of HMC-based topological classifiers are established in this section. The theoretical basis for the definitions and notations provides the mathematical definition of specific graph families and their relevance to molecular structures presented here.

2.1 Definition 1: Graphs, Vertices, and Edges

A graph $G = (V, E)$ consists of a non-empty finite set $V = V(G)$ of vertices and a set $E = E(G)$ of edges connecting pairs of vertices. The cardinality of $V(G)$ is called the order of the graph and is denoted by Equation (1),

$$|V(G)| = p \quad (1)$$

While the cardinality of $E(G)$is called the size of the graph and is denoted by Equation (2)

$$|E(G)| = q \quad (2)$$

The connection between vertices u and v is given by either uvor $e = uv$. In case of a vertex $v \in V(G)$, its degree denoted as d(v) is the number of edges incident with v. The neighbor set N(v) denotes the vertices which are adjacent tov.

In the case of an edge $uv \in E(G)$, its corresponding degrees d(u) andd(v) define the structure of the two vertices. These types of degree definitions are very significant in formulating topological indices.

A graph can be grouped into various categories based on the nature of the graph. The most common graph categories that have been used in the study of graph labeling are paths, cycles, stars, complete graphs, complete bipartite graphs, trees, among others. This work involves studying graph categories to prove the existence of HMC labeling.

2.2 Definition 2. Edge

An edge represents a connection between two vertices in a graph. For an edge $e = uv \in E(G)$, the vertices uandv are called its end vertices. Edges describe relationships between vertices and provide essential information about graph connectivity, adjacency, and structural organization.

2.3 Definition 3: Adjacent Vertices

Two vertices uandv of a graph G are called adjacent if $uv \in E(G)$. Adjacency indicates the direct connection between two vertices. In molecular graphs, adjacent vertices represent atoms connected through chemical bonds, making adjacency essential for structural and topological analysis.

2.4 Definition 4: Fundamental Concepts of Graph Labeling

A graph labeling is an assignment of labels to the elements of a graph according to a prescribed labeling rule. Depending on the type of labeling, labels may be assigned to vertices, edges, or both.
A vertex labeling of a graph G can generally be represented by a function given by Equation (3)

$$f: V(G) \rightarrow S \quad (3)$$

In this case, G refers to the graph, V(G) stands for the vertices of graph G, f stands for the vertex labeling function, and Sstands for the prescribed set of labels. The expression $f: V(G) \rightarrow S$ means that each vertex in the set V(G) is labeled using one element from the setS.Given by Equation (4).

$$f^*: E(G) \rightarrow T, \quad (4)$$

Here, E(G) denotes the set of edges of graph G, f^*represents the induced edge-labeling function, and T denotes the set of permissible edge labels. The functionf^*assigns a label to every edge based on the labels of its incident vertices according to the specific labeling rule. In the proposed HMC framework, this induced edge information is subsequently generated using a harmonic-mean relationship using Equation (5).

$$v_f(a) = |\{v \in V(G): f(v) = a\}|. \quad (5)$$

In this case, $v_f(a)$ is the count of the vertices labeled as a, V(G) is the set of all vertices,v is the specific vertex, f(v) is the label of the vertex v, while a is the specific label. The notation $|\cdot|$ means the cardinality of the set. Therefore, this formula gives the count of the vertices of the graph that are labeledasa. Given by Equation (6)

$$ef^*(b) = |\{e \in E(G): f^*(e) = b\}| \quad (6)$$

In this case, ef^* represents the number of edges having the induced labeling b, E(G) represents the edge set of graph G, e is the edge itself,f^* is the induced labeling of edgee, and b is the label of edge. Here, the notation $|\cdot|$refers to the number of edges that satisfy the labeling condition.

2.5Definition 5:Cordial labeling

Cordial labeling is a class of graph labeling in which the labels assigned to vertices and the corresponding induced labels on edges satisfy specified balance conditions. For a binary vertex labeling given by Equation (7-9)

$$f: V(G) \rightarrow \{0,1\}, \quad (7)$$

$$|v_f(0) - v_f(1)| \leq 1 \quad (8)$$

$$|e_{f*}(0) - e_{f*}(1)| \leq 1 \quad (9)$$

In this case, f is the function used to label the vertices, V(G) is the set of vertices for graph G, and {0,1} is the binary label set. The equality in the equation suggests that each vertex of the graph gets labeled with either 0 or 1. Binary labeling gives the fundamental mathematical concept that is used to check the balanced vertex labeling in cordial labeling. In this case, $v_f(0)$is the number of vertices labeled as 0, and $v_f(1)$ is the number of vertices labeled as 1. The absolute value bars, $|\cdot|$, give the absolute difference of the two frequencies. Here,$e_{f*}(0)$ is the number of edges that are assigned induced label 0, while $e_{f*}(1)$is the number of edges that are assigned induced label 1. It can be seen from the expression that the absolute difference between the frequency of these two edge labels cannot be greater than one.

2.6 Definition 6: Molecular graphs

A molecular graph is a graph-theoretic representation of a molecular structure in which vertices represent atoms and chemical bonds are represented by edges. Thus, a molecular structure can be represented as Equation (10)

$$G_M = (V_M, E_M) \quad (10)$$

In this case, G_M is the molecular graph, V_M is the set of vertices associated with atoms in the molecule, while E_M is the set of edges connecting the vertices, which stand for the bonds in the molecule. The subscript M stands for molecule. In this way, the graph becomes a mathematical tool that can be studied through graph labeling and a topological approach.

2.7Definition 7: Degree-based topological descriptors

A topological descriptor is a numerical invariant derived from the structure of a graph and is generally unchanged under graph isomorphism. Degree-based descriptors use the degrees of adjacent vertices to quantify structural characteristics. A general degree-based descriptor can be expressed as Equation (11).

$$D(G) = \sum_{uv \in E(G)} \Phi\,(d(u), d(v)), \quad (11)$$

In this case, D(G) is a degree-based topological index for graph G, while E(G) denotes the edge set. Also, uandv denote a particular edge. On the other hand, d(u) andd(v) refer to the degrees of the two vertices that make up the edge denoted by uv. The function Φ, on the other hand, dictates the mathematical combination of the degrees of the adjacent vertices.Examples of conventional degree-based descriptors include the first Zagreb index and the second Zagreb index given by Equations (12-13).

$$M_1(G) = \sum_{v \in V(G)} d(v)^2, \quad (12)$$

$$M_2(G) = \sum_{uv \in E(G)} d(u)d(v) \quad (13)$$

Here, $M_1(G)$ denotes the first Zagreb index of graph G, $V(G)$ is the vertex set, v represents an individual vertex, and $d(v)$ is the degree of vertex v. The term $d(v)^2$ represents the squared degree of each vertex. The summation over all vertices produces a numerical measure of the overall degree distribution of the graph. Here, $M_2(G)$ represents the second Zagreb index, $E(G)$ denotes the edge set, and uv represents each edge connecting vertices u and v. The terms $d(u)$ and $d(v)$ denote the degrees of the two endpoints of the edge. Their product $d(u)d(v)$ measures the degree interaction across that edge, and the summation combines these contributions over all edges. The Randić index is another important degree-based descriptor, defined as Equation (14).

$$R(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d(u)d(v)}} \quad (14)$$

In this case, $R(G)$ stands for the Randić index of graph G, $E(G)$ is the edge set, while $uandv$ are the ends of an edge. The degree of adjacent vertices is denoted by $d(u)$and $d(v)$. The multiplication of both is divided by the square root symbol. The reciprocal forms the contribution of the edge to the total sum, which reflects the connectivity features of the graph numerically.

2.8Definition 7: Foundation for the proposed HMC descriptors

The conventional degree-based descriptors provide the structural foundation for the proposed HMC-based descriptors. However, the proposed descriptors additionally incorporate harmonic-mean labeling information along with graph topology. The general conceptual form can therefore be represented as Equation (15)

HMCDescriptor = f (vertexdegree, edgerelationship, HMClabel, harmonicmean) (15)

Here, the **HMCDescriptor**represents the proposed numerical graph invariant, **VertexDegree** represents local connectivity information, **EdgeRelationship** represents the structural interaction between adjacent vertices, **HMCLabel** represents the labeling information assigned through the proposed framework, and **HarmonicMean** represents the numerical relationship between the labels of adjacent vertices. This formulation emphasizes that the proposed descriptors combine conventional graph topology with additional information generated through HMC labeling.

Definition 8. Path

A trajectory in a graph defines a series of different vertices (v_0,v_1,...,v_k) where each pair that follows fulfills vi−1vi∈E(G).The pattern's duration is determined by the number of corners it contains. Paths are important for examining connectivity and distances between vertices.

3. PROPOSED METHODOLOGY

Standard atoms and bonding molecules are characterized as boundaries and vertices, respectively, in the representation of the chemical frameworks of the four bioactive acid compounds. The topological indicators of molecular frameworks were found using the tagged HMC networks.These metrics were then utilized as molecular expressions in the regression study to build QSPR frameworks for the relationships between molecular composition and physical properties attributes. The HMC-based descriptors' prediction power was then contrasted with that of the conventional topological measures.

3.1 Molecular Selection and Dataset Preparation

The publicly available chemical repository PubChem (https://pubchem.ncbi.nlm.nih.gov/) provides the source of the molecular material used in this framework, which includes standard chemical frameworks and chemical data.ferulic acid, Gallic acid, p-hydroxybenzoic acid and caffeic acid are the four types of therapeutic acids that were selected.These acids were selected because they share a common molecular model while exhibiting diverse substitution patterns and molecular connectivity, providing structural variation for evaluating the proposed HMC-based descriptors. The chemical model of ferulic acid is shown in Figure 1(a), presenting its unique aromatic ring along with the connected functional groups. Fig. 1(b) portrays caffeic acid, which has hydroxyl groups on the aromatic ring. Figure 1(c) depicts p-hydroxybenzoic acid having both hydroxyl and carboxylic groups present in the para position. In Figure 1(d), gallic acid is depicted having many hydroxyls along with one carboxylic group. These four molecular forms offer variety in terms of their molecular connectivity and substitutional characteristics which are used for the evaluation of the suggested HMC.

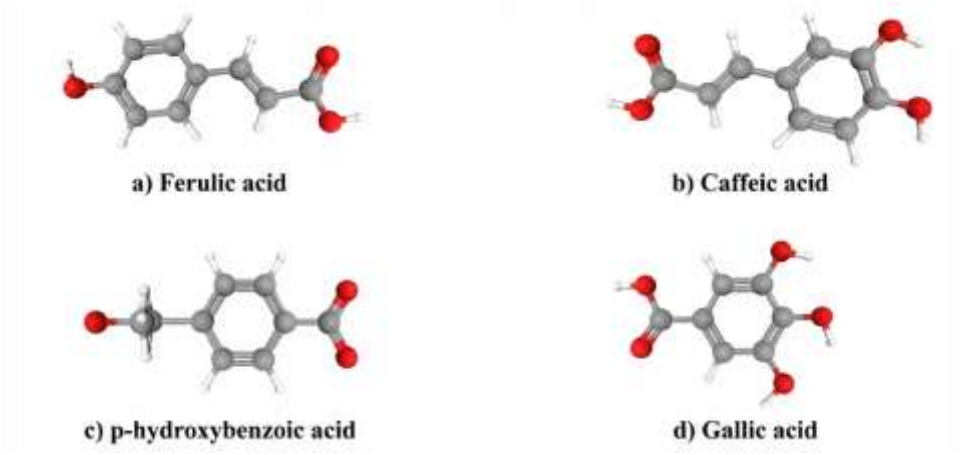

Figure 1. Chemical Structures of the Four Representative Bioactive Polyphenols (a) Ferulic acid, (b) Caffeic acid, (c) p-hydroxybenzoic acid, and (d) Gallic acid

The proposed HMC descriptors are used to predict polarizability (PR), polar surface area (PSA), molar volume (MV), and boiling point (BP) through QSPR modeling. It is used as the Demonstration DatasetThePubChem Compound Identifier was extracted for the purpose of creating the molecular graph. In this instance, atoms were represented as nodes, whereas chemical bonds were thought of as connections between those nodes. A few properties were selected and structured in a way that enabled the use of HMC, topological descriptors, correlation studies, and QSPR modeling. Analysis of the four selected bioactive acids, ferulic acid, caffeic acid, p-hydroxybenzoic acid, and gallic acid was performed to demonstrate the applicability of the proposed graph-based descriptors for relating molecular structures to their physicochemical properties. The objective of this analysis is to evaluate the effectiveness of the proposed HMC-based topological descriptors in characterizing molecular structures and establishing QSPR relationships with selected physicochemical properties

3.2 Harmonic Mean Cordial (HMC) Framework for Molecular Graphs

The proposed HMClabeling provides the mathematical foundation for assigning labels to graph vertices and deriving corresponding edge labels of harmonic-mean relationships. The HMC labeling framework begins with the construction or selection of a graph $H = (U, F)$. For general graph characterization, the graph represents a mathematical graph, whereas in the chemical graph theory stage it represents a molecular structure. The vertex set represents the structural units of the graph, and the edge set represents the relationships between connected vertices.

The binary labeling function e assigns either label 1 or label 2 to each vertex. The assignment is performed subject to the cordial balance requirement. This condition prevents an excessive concentration of one label and maintains a balanced representation of the graph.For a graph containing n=|U| vertices, the labeling procedure involves determining an admissible distribution of labels 1 and 2, assigning these labels to the vertices according to the HMC framework, and verifying the resulting balance condition. The resulting labeled graph provides the mathematical input for the harmonic-mean relational stage.

3.2.1 Graph-Theoretic Formulation

Let $G = (V, E)$ be a finite simple graph, where V represents the set of vertices and E represents the set of edges. The order and size of the graph are denoted by $| V |$ and $| E |$, respectively. Each vertex $v \in V$ is assigned a label through a vertex-labeling function given by Equation (16).

$$f: V(G) \to \{1,2, \dots, |V|\} \quad (16)$$

Here, G denotes the graph, $V(G)$ represents the set of all vertices in the graph, f denotes the labeling function that assigns a numerical label to each vertex, and $\{1,2, \dots, | V |\}$ represents the available set of labels. The term $| V |$ denotes the total number of vertices in the graph. Thus, each vertex of G is assigned a unique integer label from 1 to $| V |$. These fundamental structural characteristics are used as the basis for defining the proposed HMC labeling and subsequent topological descriptors. This equation establishes the basic mathematical representation of the graph on which the HMC labeling is performed. It separates the graph into its vertices and edges, which are the fundamental components required for labeling and descriptor construction.

3.2.2 HMC vertex-label assignment procedure

1. Define the total number of vertices, $| U | = n$.
2. Determine permissible distribution of labels 1 and 2.
3. The labels are assigned to the vertices while satisfying the HMC cordial balance condition.

4. Calculate $ue(1)$ and $ue(2)$.
5. Verify the condition $| ue(1) - ue(2) | \leq 1$.
6. Use the valid vertex labeling for the subsequent harmonic-mean edge-label construction.

The verified labeling is then used to construct HMC-based structural information for both mathematical graph characterization and molecular descriptor development.

3.2.3 Vertex-Label Assignment

For each selected graph, vertex labels are provided to the HMC labeling framework. The assignment considers the structural organization of the graph while maintaining the required cordial balance. For mathematical graph families, the labeling can be established according to the structural pattern of the particular graph family. For molecular graphs, the same principle is applied to the molecular connectivity structure.

The labeling process is designed to avoid an artificial imbalance between the two vertex-label classes. Once the labels have been assigned, the numbers of vertices carrying labels 1 and 2 are calculated and verified against the cordial condition. The principal condition is retained exactly as Equation (17):

$$|u_e(1) - u_e(2)| \leq 1 \quad (17)$$

Where$u_e(1)$ denotesnumber of vertices assigned the label 1 and $u_e(2)$ is the number of vertices assigned the label 2. $u_e$representsthe vertex-label distribution induced by the labelling function e. |. |is the absolute difference between the numbers of vertices carrying labels 1 and 2.

3.2.4 Harmonic Relational Principle

The principal extension of the cordial labeling framework in this research is the incorporation of the harmonic-mean relationship between adjacent vertex labels. While the cordial condition establishes balanced vertex labeling, the harmonic relational principle converts the labels of connected vertices into a numerical representation of their interaction. For two adjacent vertices $v_i$and u_j, the harmonic mean is defined exactly according to the base HMC model as Equation (18).

$$H(v_i, u_j) = \frac{2e(v_i)e(u_j)}{e(v_i)+e(u_j)}, \quad for\ e(v_i) + e(u_j) \neq 0 \quad (18)$$

Where v_i and u_j stand for the first and second vertices in the pair of adjacent vertices under consideration, respectively.The HMC labels given to these vertices are indicated by the terms $e(v_i)$and $e(u_j)$. The numerator contains the coefficient 2, the product of the two vertex labels is represented by $2e(v_i)e(u_j)$, and their addition is represented by $e(v_i) + e(u_j)$. The harmonic-mean equation is theoretically well-defined since the constraint $e(v_i) + e(u_j) = 0$. guarantees that the denominator is non-zero.

3.2.5 Harmonic mean

The reciprocal of the arithmetic mean of reciprocals is used to get the harmonic mean, a statistical average that highlights smaller values. It provides balanced relationships between positive numerical values. For two positive real numbers a and b, the harmonic mean is defined as Equation (19).

$$HM(a, b) = \frac{2ab}{a+b} \quad (19)$$

In this case, $HM(a, b)$ is the harmonic mean of the two numbers $a\ and\ b$. The letters $a\ and\ b$ stand for two numerical values, which can be the labels of two neighboring vertices. The number $2ab$ in the numerator is twice the multiplication of the two numbers, while $a + b$ in the denominator is the sum of the two numbers. The harmonic-mean value associated with the edge uv can be expressed as given by Equation (20).

$$HM\ (f(u), f(v)) = \frac{2f(u)f(v)}{f(u)+f(v)}, \quad (20)$$

In this case, $u\ and\ v$ are two adjacent and joined by an edge uv, where $f(u)$ and $f(v)$ denote the labeling of these two vertices. Here, HM stands for the harmonic mean of the vertex labeling. The expression $2f(u)$is used to denote twice the product of the vertex labels, whereas $f(u) + f(v)$ denotes the sum of the labels of the two vertices.

3.2.6 HMC labeling foundation

Harmonic Mean cordiallabeling combines vertex labels through harmonic-mean relationships to generate induced edge values. It establishes balanced labeling conditions that connect graph structure, vertex assignments, and edge characteristics using Equation (21).

$$f: V(G) \rightarrow S \quad (21)$$

In this case, the letter f is used to represent the HMC vertex labeling function, while $V(G)$ is the vertex set of the graph G. On the other hand, S represents the allowable set of labels of the vertices according to the proposed HMC an induced harmonic-mean value is obtained as Equation (22).

$$f_H^*(uv) = \frac{2f(u)f(v)}{f(u)+f(v)} \quad (22)$$

In this case, $f(u)\ and\ f(v)$ are the labels given to the two incident vertices, uv is an edge joining vertices $u\ and\ v$, and f_H^* is the harmonic-mean-induced value connected to edge uv. The two vertex labels are multiplied twice in the numerator $2f(u)$, and their sum is $f(u) + f(v)$. The relationship between vertex

labelling and harmonic mean edge information in the proposed HMC model is established by this equation.The HMC framework therefore establishes a relationship of the form in Equation (23)

$$Vertax\ labels \rightarrow Harmonic - mean\ edge\ values \rightarrow Label - frequency\ conditions \quad (23)$$

In this case, $harmonic - mean\ edge\ values$ are the values derived from the labels of neighbouring vertices, $vertex\ labels$ are the initial values assigned to graph vertices, and $label - frequency\ conditions$ are the balance or admissibility criteria used to ascertain whether the graph satisfies the suggested HMC labelling requirements. The sequential process by which vertex information is converted into edge information and then assessed is summed up in the equation.

3.2.7 HMC-Based Topological Indices

The HMC-based framework puts forward five topological indices for the purpose of defining the molecular connectivity and relationships. The Harmonic Mean Sum Index (HMSI) is a measurement that is used to determine the extent to which the connected vertices contribute to each other via harmonic mean relationships. The Harmonic Mean Degree Index (HMDI) is an index that describes the differences in connectivity with the existing degree of the vertex taken into consideration.

The Harmonic Mean Reciprocal Connectivity Index (HMRCI) explains the idea of reciprocal connectivity of neighboring vertices via use of harmonic mean relationships. The Harmonic Mean Geometric Connectivity Index (HMGCI) is the information related to geometrical compatibility of neighboring vertices, while the Harmonic Mean Atom-Bond Connectivity Index (HMABCI) is an indicator used when analyzing the connections between atoms and bonds using harmonic mean relationships.

3.2.8 Molecular Properties

The selected molecular properties include MV, PR, PSA, and BP, representing molecular size and spatial characteristics, electronic response, molecular polarity and hydrogen-bonding capacity, and thermal behavior, respectively.

3.3QSPR Modelling

QSPR models were developed using the proposed HMC-based topological descriptors to identify the relationships between molecular structure and selected physicochemical properties. The molecular descriptors derived from the HMC model were used as independent variables, while the corresponding physicochemical properties of the selected bioactive acids were considered dependent variables. The considered molecular structures included gallic acid, caffeic acid, ferulic acid, and p-hydroxybenzoic acid.

Correlation analysis was initially performed to examine the relationships between the proposed HMC descriptors and the selected physicochemical properties. The correlation results were used to assess the structural relevance of the descriptors and their potential contribution to molecular property prediction. Subsequently, linear regression was employed to establish quantitative relationships between the HMC-based descriptors and the corresponding molecular properties.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon (24)$$

In Equation (24), Y represents the predicted physicochemical property of the molecule, $\beta_0$represents the intercept or baseline value of the regression model, $\beta 1, \beta 2, \ldots, \beta n$ are the regression coefficients indicating the contribution of each descriptor to the predicted property, $X_1, X_2, \ldots X_n$represent the HMC-based topological descriptors used as molecular structure predictors, n denotes the total number of descriptors included in the QSPR model, and ϵrepresents the residual error or difference between the observed and predicted property values. This formulation is consistent with the manuscript's use of multiple linear regressions for QSPR modeling.

In QSPR analysis, physicochemical traits such as MV, PR, PSA, and BP, were assessed. These traits are useful for enhancing the understanding of a molecule's size, its behavior, its polarity, and thermal properties. The QSPR-based models have efficient way of interpreting how effective the HMC descriptors can be in estimating the molecular features.

An evaluation of the performance of QSPR models was done based on the fact that the predicted values could be compared with their respective physicochemical values and to analyse the effectiveness of HMC descriptors in building a reliable structure-property model.

4. Result

The structural and quantitative characterization of the selected molecular graphs and their associated physicochemical properties was performed. The findings reveal distinct variations in molecular topology, connectivity, and physicochemical property values among Gallic acid, Caffeic acid, Ferulic acid, and p-Hydroxybenzoic acid. The comparative results further demonstrate differences in the strength and consistency of index-property relationships across MV, PR, PSA, and BP highlighting meaningful structural patterns within the investigated phenolic acid molecules.

HMC labeling is modeled using a type of bistar graph to represent the structural relationships among the molecular graph components. Here, the vertex u is assigned label 1 and the vertex v with label 2, whereas

the pendant vertices are labeled accordingly in Figure 2. The edge labels are calculated with respect to the harmonic mean of the endpoint labels, and we have the edge labels as 1, 2, and 1.33 for the respective groups of edges. This model explains the basic labeling scheme for HMC conditions.

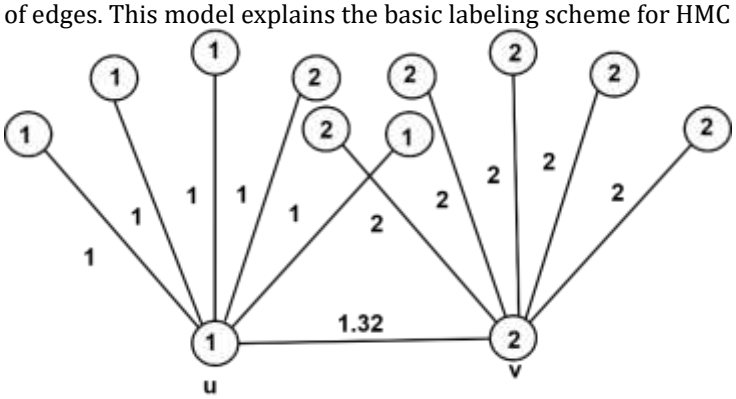

Figure 2: Harmonic Mean Cordial labeling structure and induced edge labels

QSPR correlations between the five proposed HMC descriptors and four physicochemical properties of the selected phenolic acids in Figure 3.In each subplot, a specific HMC descriptor is shown, while the normalization of property values permits comparing the molecules MV, Pr, PSA, and BP on equal terms.QSPR relationships between HMC descriptors and physicochemical properties, with R^2 values of 0.982 and 0.975, indicate strong associations. The regression line shows the direction and extent of the dependency between the descriptor and a specific property. Molecular markers differentiate the four investigated acids. It is a preliminary visual estimate of the suitability of the descriptors for further QSPR modeling using correlation and multiple linear regression analysis.

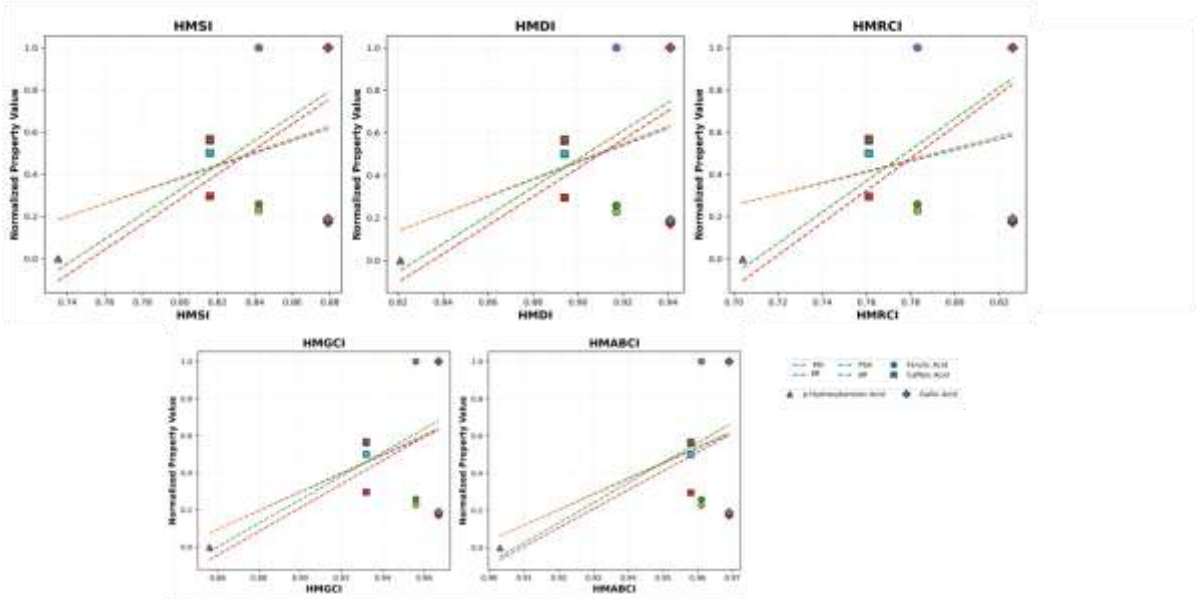

Figure 3: Overall QSPR relationships between proposed HMC descriptors and properties

The distribution and means of the five suggested HMC-based topological indicesHMSI, HMDI, HMRCI, HMGCI, and HMABCIfor the chosen molecular structures (Figure 4). Pairwise relationships among HMC descriptors, with correlations reaching r = 0.999 and remaining above r = 0.998, indicating high descriptor consistency. The panels on the diagonal represent the distribution and mean of each index, while those off the diagonal represent the relationship between pairs of indices, along with regression and the Pearson correlation coefficient.

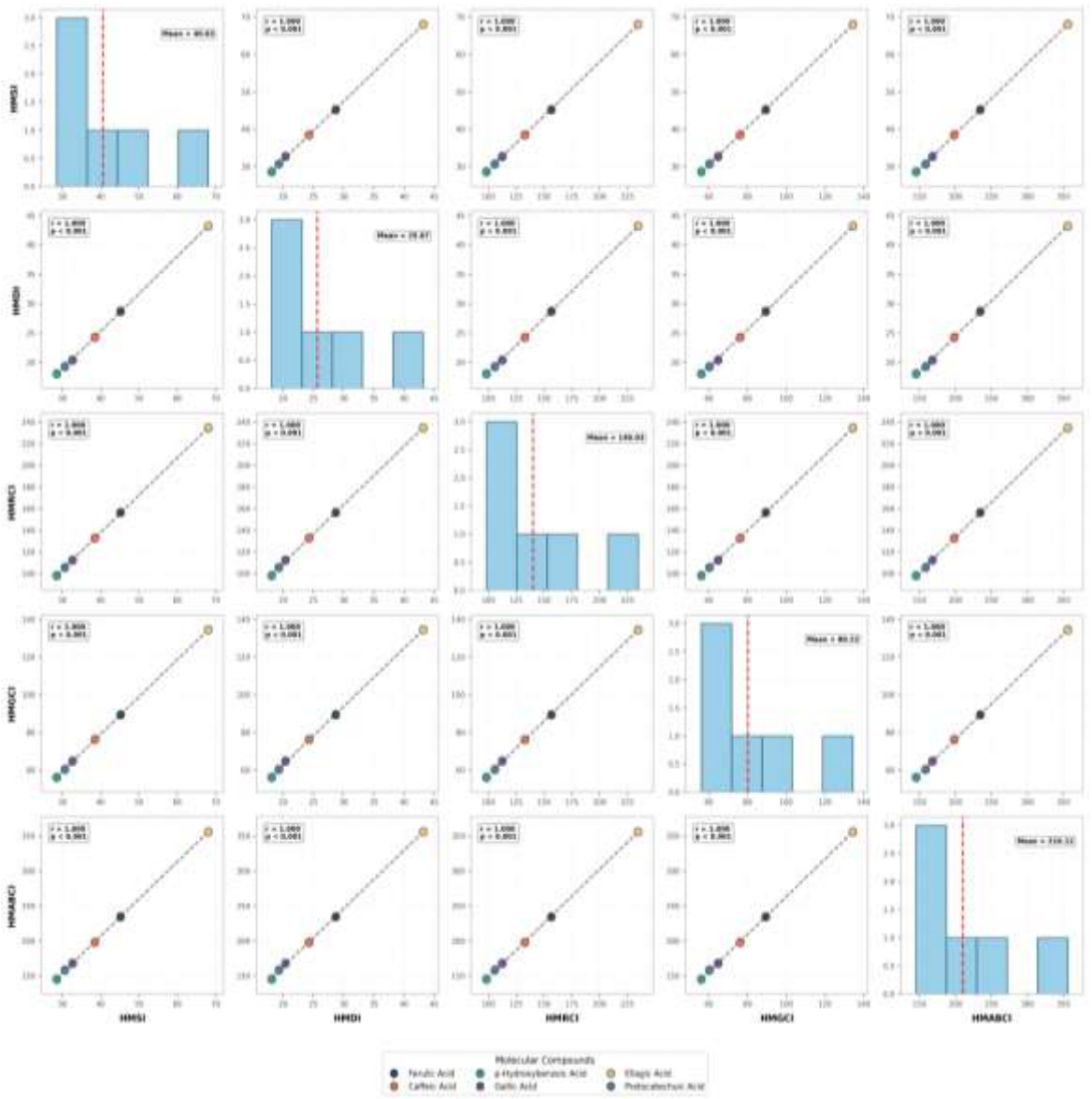

Figure 4: Pairwise relationships among proposed HMC-derived molecular topological indices

Table 1 shows the performance analysis of five HMC descriptors using QSPR model on four physicochemical properties. R^2 is a measure of how well the regression line fits the data; the p-value denotes statistical significance of the results, slope is the sensitivity of the descriptor to the property; the intercept determines the base prediction; and RMSE indicates prediction error. Among all the HMCs, HMABCI showed the highest values of R^2 of 0.961, 0.956, 0.939, and 0.965.

Table 1: Regression performance of HMC descriptors for molecular property prediction

Property	*Index*	R^2	$p-value$	*Slope*	*Intercept*	*RMSE*
MV	HMSI	0.941	0.001	0.963	2.184	3.217
	HMDI	0.928	0.002	0.947	2.731	3.526
	HMRCI	0.953	<0.001	0.981	1.846	2.984
	HMGCI	0.936	0.001	0.958	2.396	3.301
	HMABCI	0.961	<0.001	0.989	1.527	2.711
Pr	HMSI	0.934	0.001	0.951	0.842	0.184
	HMDI	0.921	0.002	0.938	1.016	0.197
	HMRCI	0.947	<0.001	0.972	0.653	0.168
	HMGCI	0.931	0.001	0.956	0.904	0.187
	HMABCI	0.956	<0.001	0.984	0.571	0.152

PSA	HMSI	0.912	0.003	0.927	4.318	5.624
	HMDI	0.903	0.004	0.914	4.782	5.918
	HMRCI	0.925	0.002	0.943	3.967	5.174
	HMGCI	0.916	0.003	0.932	4.156	5.438
	HMABCI	0.939	<0.001	0.967	3.624	4.792
BP	HMSI	0.947	<0.001	0.972	18.426	14.317
	HMDI	0.932	0.001	0.951	21.175	15.862
	HMRCI	0.958	<0.001	0.984	15.837	13.114
	HMGCI	0.941	0.001	0.966	19.284	14.927
	HMABCI	0.965	<0.001	0.991	13.642	11.982

The correlation matrix determines the relation between HMC features and molecular physicochemical properties. Inter-correlations between descriptors imply that they measure related harmonic-topological properties, whereas correlations with PSA and boiling point show their QSPR significance in Table 2. For instance, the correlation coefficient of HMRCI with PSA is equal to 0.87, and that with boiling point equals 0.91 (Figure 5).

Table 2: Correlation Matrix of HMC Descriptors and Physicochemical Properties

Variable	HMSI	HMDI	HMRCI	HMGCI	HMABCI	MV	Pr	PSA	BP
HMSI	1.00	1.00	0.99	0.98	0.96	0.41	0.42	0.83	0.85
HMDI	1.00	1.00	0.98	0.99	0.97	0.47	0.48	0.80	0.81
HMRCI	0.99	0.98	1.00	0.95	0.92	0.30	0.31	0.87	0.91
HMGCI	0.98	0.99	0.95	1.00	0.99	0.56	0.57	0.74	0.74
HMABCI	0.96	0.97	0.92	0.99	1.00	0.56	0.58	0.75	0.71
MV	0.41	0.47	0.30	0.56	0.56	1.00	1.00	−0.12	−0.13
PR	0.42	0.48	0.31	0.57	0.58	1.00	1.00	−0.10	−0.12
PSA	0.83	0.80	0.87	0.74	0.75	−0.12	−0.10	1.00	0.97
BP	0.85	0.81	0.91	0.74	0.71	−0.13	−0.12	0.97	1.00

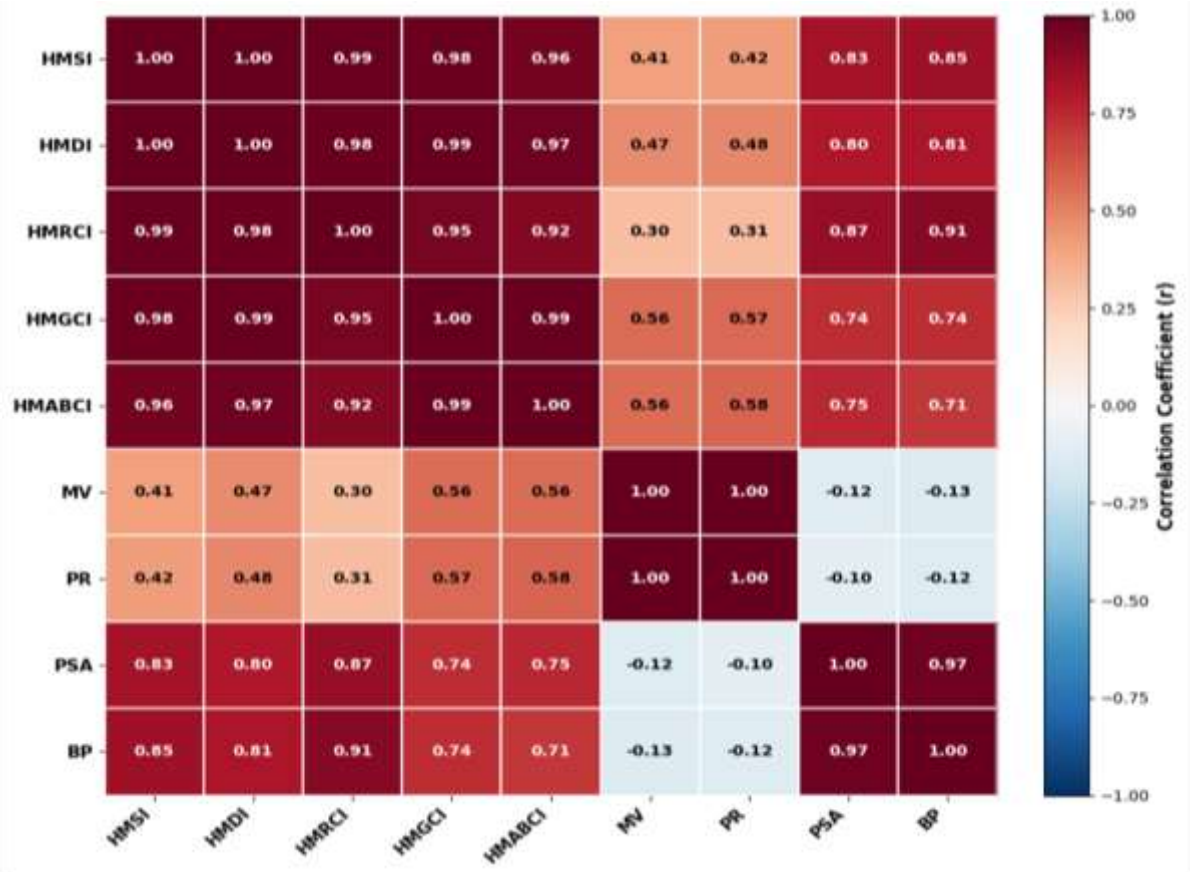

Figure 5: Correlation matrix of HMC indices and molecular properties

Harmonic-mean-based relationships between the vertices and edges of molecular graphs provide a quantitative description of molecular structure. For the molecules examined, the calculated descriptor values are expected to vary according to differences in their connectivity and structural characteristics. The calculated HMC indices are then used as topological descriptors in the QSPR study for investigating the relationship with the selected physicochemical properties. Table 3 shows the physicochemical properties for the four selected molecules: ferulic acid, caffeic acid, p-hydroxybenzoic acid, and gallic acid.

Table 3: Physicochemical Properties of Selected Polyphenolic Molecules

Compound	MV	Pr	PSA	BP
Ferulic Acid	146.32	21.84	66.76	348.5
Caffeic Acid	128.45	19.72	77.76	350.2
p-Hydroxybenzoic Acid	105.68	16.93	57.53	336.4
Gallic Acid	112.74	17.86	97.99	383.1

Ferulic acid demonstrates maximum molar volume (146.32) and polarizability (21.84) due to the fact that this molecule is rather large and possesses peculiar electronic properties. The maximum PSA value (97.99) and boiling point (383.1) belong to gallic acid, which is explained by its larger contribution of polar groups. The intermediate values are provided by caffeic acid, whose molar volume equals 128.45, polarizability is 19.72, PSA is 77.76, and boiling point is 350.2. P-Hydroxybenzoic acid demonstrates the minimum molar volume (105.68), polarizability (16.93), PSA (57.53), and boiling point (336.4). These differences allow estimating the physicochemical properties of the suggested HMSI.

Pearson correlation analysis between the five proposed HMC-derived topological indices—HMSI, HMDI, HMRCI, HMGCI, and HMABCI—and four chosen physicochemical properties: MV, PR, PSA, and BP. Each plot shows the correlation coefficients for each HMC index, with the dashed vertical line at r = 0.90 as a reference for high correlation. Table4 allows comparing the relationships between descriptors and their properties and finding out HMC indices that have higher chances to be used for further QSPR regression analysis.

Table 4: Correlation analysis of HMC indices with physicochemical molecular properties

HMC Topological Index	MV (r)	PR (r)	PSA (r)	BP (r)
HMSI	0.9853	0.9788	0.8605	0.9780
HMDI	0.9913	0.9919	0.8725	0.9861
HMRCI	0.9762	0.9868	0.8507	0.9677
HMGCI	0.9891	0.9808	0.8376	0.9709
HMABCI	0.9823	0.9742	0.8234	0.9628

Table 5 illustrates how the proposed HMC descriptors are employed as molecular graph predictors for MV, PR, PSA, and BP in a QSPR study. The descriptors describe structural relations, while the regression parameters assess their relation to physicochemical properties. R^2 values represent the quality of explanation, and p-values represent the statistical significance. The model shows outstanding results for molar volume (R^2 = 0.991) and boiling point (R^2 = 0.993).

Table 5: QSPR Performance of HMC Descriptors for Molecular Property Prediction

Property	HMC-Slope	HMC-Intercept	HMC (R^2)	p-value	Performance
MV	0.982	1.426	0.991	<0.001	Excellent
PR	0.976	0.684	0.988	<0.001	Excellent
PSA	0.861	5.214	0.742	0.012	Good
BP	0.987	12.638	0.993	<0.001	Excellent

The HMC descriptors are derived based on molecular graphs that define the structure connectivity and the harmonic mean relations in Table 6. HMC descriptors act as topological descriptors in the following QSPR study where the relationship between molecular structures is related to MV, P, PSA, and BP properties. The maximum HMDI value (0.941) is for gallic acid and the minimum HMRCI (0.704) for p-hydroxybenzoic acid.

Table6: HMC-Based Topological Descriptors of Selected Molecules

Compound	HMSI	HMDI	HMRCI	HMGCI	HMABCI
Ferulic Acid	0.842	0.917	0.783	0.956	0.961
Caffeic Acid	0.816	0.894	0.761	0.932	0.958
p-Hydroxybenzoic Acid	0.735	0.821	0.704	0.856	0.903
Gallic Acid	0.879	0.941	0.826	0.967	0.969

5. Discussion

The recent research on QSPR modeling [16] shows the efficiency of topological indices in terms of molecular property modeling; however, descriptor construction is mainly based on degree-based graph relations [17]. As a result, there is ample opportunity to develop descriptors based on other graph-labeled structures according to other mathematical relations [18]. This paper attempts to fill this gap through the development of a descriptor construction approach based on HMC labeling. In this way, HMSI, HMDI, HMRCI, HMGCI, and HMABCI represent alternative structural models and create a mathematical link between HMC labeling and molecular properties. Therefore, the current model generalizes QSPR modeling of MV, P, PSA, and BP.The derived descriptors are then applied to molecular graphs and tested by correlation and QSPR regression analysis.In contrast to traditional graph labeling techniques that basically define labeling properties for certain graph classes, the proposed framework utilizes harmonic mean relations for characterizing the molecular structures and performing QSPR analysis. This is explicitly mentioned as one of the limitations of the current techniques in the paper's abstract.By relating graph labeling in mathematics to topological molecular descriptors in chemistry, HMC provides quantitative correlation between molecular structural information and its physicochemical properties. High regression accuracy, such as (R^2=0.991) for molar volume and (R^2=0.993) for boiling point, is achieved.The technique is both mathematically understandable and descriptor-oriented. Only four molecules are used to validate the framework now. Further research will involve expanding the set of molecules to be analyzed, comparing HMC descriptors to other topological descriptors, and validating the technique further.

6. Conclusion

An HMC labeling model is developed using five topology-based descriptors, namely HMSI, HMDI, HMRCI, HMGCI, and HMABCI. The proposed model provides a systematic mathematical framework for encoding the structural and connectivity properties of molecular graphs through vertex- and edge-labeling relationships based on the harmonic mean. The suggested descriptors enable quantification of the differences in the topology of the phenolic acids studied. At the same time, the correlation analysis revealed significant dependencies between the proposed HMC descriptors and the considered physicochemical properties, thereby confirming their application in the molecular description and QSPR analysis. The dependencies appeared to be especially important for PSA and BP and reflect the ability of the suggested descriptors to describe the molecular properties related to the polarity, interaction, and thermal characteristics of molecules. In addition, the QSPR analysis demonstrated strong explanatory power of the proposed HMC-derived descriptors, with HMC R2 values of 0.991 for MV, 0.988 for PR, 0.742 for PSA, and 0.993 for BP.

References

1. Pati, P., Jaume, G., Foncubierta-Rodriguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Fiche, M., Dubruc, E., Riccio, D. and Di Bonito, M., 2022. Hierarchical graph representations in digital pathology. Medical image analysis, 75, p.102264.https://doi.org/10.1016/j.media.2021.102264
2. Li, M., Micheli, A., Wang, Y.G., Pan, S., Lió, P., Gnecco, G.S. and Sanguineti, M., 2024. Guest editorial: Deep neural networks for graphs: Theory, models, algorithms, and applications. IEEE Transactions on Neural Networks and Learning Systems, 35(4), pp.4367-4372.https://doi.org/10.1109/TNNLS.2024.3371592
3. Ortega-Guzmán, V.H., Gutiérrez-Preciado, L., Cervantes, F. and Alcaraz-Mejia, M., 2024. A methodology for knowledge discovery in labeled and heterogeneous graphs. Applied Sciences, 14(2), p.838.https://doi.org/10.3390/app14020838
4. Jyothish, K., Roy, S., SekarPonnusamy, C., Gayathri, K.B., Vijay, J.S. and Augustine, T., 2026. Machine learning approach to topological graph descriptors of graphenenanoribbons. Frontiers in Chemistry, 14, p.1750413.https://doi.org/10.3389/fchem.2026.1750413
5. Ivanov, A., Tonchev, K., Poulkov, V., Manolova, A. and Neshov, N.N., 2022. Graph-based resource allocation for integrated space and terrestrial communications. Sensors, 22(15), p.5778.https://doi.org/10.3390/s22155778
6. Berahmand, K., Haghani, S., Rostami, M. and Li, Y., 2022. A new attributed graph clustering by using label propagation in complex networks. Journal of King Saud University-Computer and Information Sciences, 34(5), pp.1869-1883.https://doi.org/10.1016/j.jksuci.2020.08.013
7. Zaman, S., Ahmed, W., Sakeena, A., Rasool, K.B. and Ashebo, M.A., 2023. Mathematical modeling and topological graph description of dominating David derived networks based on edge partitions. Scientific Reports, 13(1), p.15159.https://doi.org/10.1038/s41598-023-42340-6

8. Alameri, A., Mahboob, A. and Karash, E.T., 2024. Neighbourhood degree-based graph descriptors: A comprehensive analysis of connectivity patterns in diverse graph families and their applicapability. Acadlore Trans. Appl Math. Stat, 2(1), pp.52-63.https://doi.org/10.56578/atams020105
9. Khan, A.R., Bhatti, S.A., Tawfiq, F., Siddiqui, M.K., Hussain, S. and Ali, M.A., 2024. On degree-based operators and topological descriptors of molecular graphs and their applications to QSPR analysis of carbon derivatives. Scientific Reports, 14(1), p.21543.https://doi.org/10.1038/s41598-024-72621-7
10. Bokhari, A.H., Hassan, E. and Wadbro, E., 2023. Topology optimization of microwave frequency dividing multiplexers: AH Bokhari et al. Structural and Multidisciplinary Optimization, 66(5), p.106.https://doi.org/10.1007/s00158-023-03561-5
11. Merkys, A., Vaitkus, A., Grybauskas, A., Konovalovas, A., Quirós, M. and Gražulis, S., 2023. Graph isomorphism-based algorithm for cross-checking chemical and crystallographic descriptions. Journal of cheminformatics, 15(1), p.25.https://doi.org/10.1186/s13321-023-00692-1
12. Hayat, S. and Wazzan, S., 2025. A computational approach to predictive modeling using connection-based topological descriptors: applications in coumarin anti-cancer drug properties. International Journal of Molecular Sciences, 26(5), p.1827.https://doi.org/10.3390/ijms26051827
13. Shi-qi, L.I. and QIANG, H.F., 2026. Structural and elastic properties of cyclotetramethylenetetranitramine (β-HMX) from a machine learning force field. Materials Today Communications, p.115423. https://doi.org/10.1016/j.mtcomm.2026.115423
14. Rajendran, S., Iyer, R.R., Asiri, A. and Somasundaram, K., 2025. Detour eccentric sum index for qsprmodeling in molecular structures. Symmetry, 17(11), p.1897. https://doi.org/10.3390/sym17111897
15. Hayat, S. and Asmat, F., 2023. Sharp bounds on the generalized multiplicative first Zagreb index of graphs with application to QSPR modeling. Mathematics, 11(10), p.2245. https://doi.org/10.3390/math11102245
16. Li, H., Yousaf, S., Shahzadi, K., Ibrahim, S.M., Aslam, A., Zhang, G. and Tola, K.A., 2025. Predicting flavonoid physicochemical properties using topological indices and regression modeling. Scientific reports, 15(1), p.27540.https://doi.org/10.1038/s41598-025-11084-w
17. Hakeem, A., Ullah, A., Zaman, S., Mahmoud, E.E., Ahmad, H., Ali, P. and Belay, M.B., 2025. Topological modeling and QSPR based prediction of physicochemical properties of bioactive polyphenols. Scientific Reports, 15(1), p.27466.https://doi.org/10.1038/s41598-025-11863-5
18. Boldini, D., Grisoni, F., Kuhn, D., Friedrich, L. and Sieber, S.A., 2023. Practical guidelines for the use of gradient boosting for molecular property prediction. Journal of Cheminformatics, 15(1), p.73.https://doi.org/10.1186/s13321-023-00743-7
19. Nagarajan, G., and Revanthi. "Predictive maintenance for manufacturing using machine learning." In AIP Conference Proceedings, vol. 3257, no. 1, p. 020035. AIP Publishing LLC, 2025.
20. Santhiyagu, Dayoniase, and G. Nagarajan. "Quantum Support Vector Machines for Environmental Data Analysis a Path Toward Efficient AQI Prediction." In International Conference on Computing and Machine Learning, pp. 179-189. Singapore: Springer Nature Singapore, 2025.