Ind. J. Pharm. Edu. Res., 2026; 60(4s):s1610-s1629.
https://www.ijper.org

Original Article

Beyond Traditional Descriptors: Innovative Topological Indices for Enhanced Molecular Predictions

Veerappan Rajkumar[1,*], Balsubramanian Sivakumar[2], Savari Prabhu[1]

[1]Department of Mathematics, Rajalakshmi Engineering College, Chennai, Tamil Nadu, INDIA.
[2]Department of Mathematics, Sri Sivasubramaniaya Nadar College of Engineering, Kalavakkam, Chennai, Tamil Nadu, INDIA.

ABSTRACT

Aim/Background: In order to capture significant structural features of chemical compounds, this study presents three new topological indices: the Fragment-Weighted Connectivity Index (FWCI), the Distance-Weighted Subgraph Index (DWSI), and the Symmetry-Based Informational Index (SBI). In order to provide improved molecular descriptors for Quantitative Structure-Property Relationship (QSPR) modelling, these indices seek to represent molecular symmetry, spatial arrangement, and functional group connectivity. **Materials and Methods:** Alkanes, alcohols, and aromatic chemicals were among the datasets to which the proposed indices were applied. They were employed to simulate physicochemical characteristics such as log P, molecular weight, molecular volume, and boiling point. Through the use of these indices, QSPR models were created, and their prediction abilities were evaluated using k-fold cross-validation. **Results:** There was a substantial association between boiling points and FWCI, which measures fragment connectivity. Shape-related parameters like surface area and log P were successfully predicted by DWSI. Properties driven by symmetry, such as chirality and optical activity, were captured by SBI. When combined, these novel indices performed better than other conventional descriptors such as the Wiener and Randic indices, mainly for molecules with complicated structures. **Conclusion:** The findings show that FWCI, DWSI, and SBI improve prediction power for chemical characteristics and provide a deeper understanding of molecular behaviour. These new indices signify a major breakthrough in QSPR modelling and may find use in drug discovery and cheminformatics.

Keywords: Connectivity, QSPR Modelling, Symmetry, Topological Indices.

Correspondence:
Mr. Veerappan Rajkumar
Department Mathematics, Rajalakshmi Engineering College, Chennai, Tamil Nadu, INDIA.
Email: rajkumar.v@rajalakshmi.edu.in

Received: 08-04-2026;
Revised: 14-06-2026;
Accepted: 24-08-2026.

INTRODUCTION

Topological indicators are quantitative measures obtained from the atomic arrangement of a chemical substance that reflect key structural characteristics. These indicators play a vital role in cheminformatics, aiding in the estimation of chemical properties, biological functions, and interactions without the need for experimental evidence.

The fundamental concept here is that the composition of a molecule dictates its characteristics, and topological indicators provide a method to measure this composition.[1-6] In cheminformatics, molecules are considered as graphs, where atoms are represented as vertices and chemical bonds between atoms are represented by edges. Topological indices are computed using the connectedness, symmetry, and atomic distances of the graph.[7-9] They offer a quantitative means of characterising the branching, size, and structure of the molecule.

To connect molecular structure to characteristics like boiling point, solubility, and biological activity in QSPR modelling, topological indices are crucial.[10] The cornerstone of cheminformatics has been established by traditional indices like the Wiener, Randic, and Zagreb indices, which provide a means of measuring molecule connection, branching, and general shape.[11,12] However, these conventional indices frequently fail to capture specific elements such as spatial arrangement, molecular symmetry, and functional group contributions as molecular structures become more complex. These features are essential for correctly forecasting the behaviour of more varied, multifunctional compounds.

A number of recent investigations have proved the significance of topological indices in understanding antiviral and nanocarbon structures, emphasising the necessity for indices that can address spatial, structural, and electronic characteristics concurrently.[13-15] Motivated by this, we offer three novel indices, FWCI, DWSI, and SBI, that capture functional connectivity, distance-based structure, and symmetry in a unified QSPR modelling method.

Novel topological indices have been created to overcome these restrictions and offer a more thorough structural

DOI: 10.5530/ijper.20262805

description. Spatial separation, molecular symmetry, and fragment connectivity are all taken into account by indices such as the Fragment-Weighted Connectivity Index (FWCI), Distance-Weighted Subgraph Index (DWSI), and Symmetry-Based Informational Index (SBI). Improved accuracy in QSPR models results from these sophisticated indices' ability to capture subtle structural components that are closely linked to chemical characteristics. From drug development to materials research, these indices enhance the predictive capacity for a variety of chemical and biological applications by improving the structure-property relationship. In this research, three new topological indices - the Fragment-Weighted Connectivity Index (FWCI), the Distance-Weighted Subgraph Index (DWSI), and the Symmetry-Based

Informational Index (SBI) is proposed and thoroughly evaluated. They were created utilising the systematic methodology described by Skvortsova *et al.*[16] These indices incorporate important structural properties like symmetry, spatial separation of molecular subgraphs, and functional group connectivity in order to overcome the shortcomings of conventional descriptors. The goal of the study is to show how well these indices capture complex structural attributes and how relevant they are for molecular property prediction by applying them to a variety of chemical compounds and evaluating their predictive accuracy for physicochemical properties in QSPR models.

BACKGROUND

Skvortsova *et al.* established the block-based methodology for constructing topological indices, which offers a structured and modular manner to create indices customised to certain molecular properties. The index building process is arranged using this method into separate "blocks", each of which has a defined function that may be combined or altered to provide unique descriptors. Block 1 is where the method starts, when the molecular graph's vertices and edges are given weights according to their chemical importance (such as atom or bond type). In order to understand spatial interactions inside the molecule, Block 2 then creates connectivity patterns between vertices, capturing properties like bond multiplicity or distance. The structural components in later blocks are progressively honed to capture more complex facets of molecular behaviour. In order to highlight the contributions of particular chemical substructures, like functional groups, Block 3 enables the computation of fragment weights by combining vertex and edge weights. The creation of new indices is finally made easier by Block 4, which combines connection data and computed weights into mathematical expressions that represent characteristics like spatial separation or molecular symmetry. In order to facilitate more precise predictions in QSPR models, this block-based approach enables the methodical construction of topological indices that encapsulate certain chemical properties. The Wiener, Randic, and Zagreb indices are examples of topological indices that have long been significant tools in cheminformatics for predicting properties and estimating molecular structure. In order to determine the size and compactness of molecules, the Wiener Index adds up the distances between every pair of atoms.[17] This index is frequently used to simulate bulk qualities like viscosity and boiling points.[18] By computing the product of the degrees of atoms connected by each bond, the Randic index, in contrast, quantifies molecular branching.[19] This index works well for examining branching patterns and has been useful for simulating biological activity.[20] In order to measure branching and molecular connectivity, the Zagreb indices concentrate on vertex degrees.[21] They are frequently used to evaluate molecular stability and reactivity in QSPR models.[22] These conventional indices are widely used; however, they have significant drawbacks, especially when used with complex or multifunctional compounds. For example, molecules with different functional group configurations have identical results because the Wiener index is insensitive to atom types and functional groups. The spatial arrangement or asymmetry in molecules, which is crucial for predicting reactivity and polarity for compounds with various functional groups or non-linear structures, is not sufficiently captured by the Randic and Zagreb indices, despite their usefulness for simpler structures.[23] These drawbacks emphasise the necessity of more sophisticated indices that take into account intricate molecular characteristics, including spatial separation, molecular symmetry, and the connectivity of certain fragments, in order to improve the predictive ability of QSPR/QSAR models. Advanced descriptors such as fragment-based, distance-weighted, and symmetry-based indices are made to capture particular facets of molecule structure that affect chemical behaviour. Fragment-based indices emphasise the contributions of certain molecular substructures or functional groups within a chemical, acknowledging that functional groups such as methyl or hydroxyl groups frequently have a significant influence on characteristics like reactivity, boiling temperature, and solubility. Fragment based indices offer a more detailed understanding of how each component of a molecule contributes to its overall qualities by giving these fragments weights and figuring out their interconnectedness. Distance-weighted indices highlight how various molecular components relate to one another spatially. In order to forecast features like log P and steric effects in biological interaction, these indices can describe molecular shape-influenced variables like molecular volume and surface area by taking into account the distances between atoms, chains, or rings. The homogeneity of atom types and bond configurations, however, is captured by symmetry-based indices, which quantify molecular symmetry. A molecule's reactivity and interactions with other molecules are frequently impacted by its symmetry, which is closely linked to characteristics like optical activity, chirality, and polarity. For the purpose of precisely predicting physicochemical and biological features, QSPR/QSAR models benefit from the combination of

fragment-based, distance-weighted, and symmetry-based indices, which provide a thorough method for simulating complicated molecular behaviours.

Graph assumptions and conventions

Let$G = (V, E)$ be a simple, undirected molecular graph. All graphs considered are assumed connected; if a molecule is disconnected, indices are computed component wise and summed. Heteroatoms (eg, N, O, S, halogens) are treated as explicit vertices in *V*and participate in edges according to bonding. For the classical indices (Wiener, Randić, Zagreb), we adopt Option-1 (hydrogen-suppressed heavy-atom graph) unless stated otherwise. For FWCI, DWSI and SBI, the hydrogen-complete graph may be used when required by the weighting and symmetry-class definitions; this convention is stated explicitly in the worked example.

Classical topological indices used for comparison

Let $G = (V, E)$ be a simple, connected, undirected molecular graph, where $V(G)$ is the vertex set and $E(G)$ is the edge set. For $u, v \in V(G)$, let $d(u, v)$ denote the shortest-path distance between u and v, and let $\deg(v)$ denote the degree of vertex v. The following classical topological indices are used for comparison in this work:

Wiener index

$$W(G) = \sum_{\{u,v\} \subseteq V(G)} d(u, v)$$

where the summation is over all unordered pairs of distinct vertices.

Randić (connectivity) index

$$\chi(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{\deg(u)\deg(v)}}.$$

First Zagreb index

$$\chi(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{\deg(u)\deg(v)}}.$$

MATERIALS AND METHODS

Using a systematic procedure based on the steps and blocks described by Skvortsova *et al.*, the suggested topological indices, the Symmetry-Based Informational Index (SBI), the Distance-Weighted Subgraph Index (DWSI), and the Fragment-Weighted Connectivity Index (FWCI) are derived. In order to reflect important molecular characteristics such as fragment connectivity, spatial organisation, and symmetry (Supplementary Table S1), each index makes use of certain graph matrices and weights, with systematic transformations directed by Procedures 1-4 and Blocks 1-2.

Procedure 1: Creates basic graph matrices, A, based on chosen weights for vertices and vertex pairs (edges), which serve as the foundation for all three indices. In order to save the resultant matrices for subsequent transformations, initial weights must be assigned.

Block 1: Initial vertex weights

Vertex weights x_{ii} are assigned based on chemical characteristics and structural features.

FWCI: The type and function of a vertex inside a functional group determines its weight (e.g., larger weights for polar groups like hydroxyl or carbonyl).

DWSI: Higher weights are given to vertices in the central structure or important functional areas, reflecting the spatial significance of each atom.

SBI: To reflect molecular homogeneity, vertices at symmetric regions are assigned lower or uniform weights, taking into account symmetry.

Block 2: Initial weights of vertex pairs

Edge weights x_{ij} are assigned based on the bond type or connectivity between vertices.

FWCI: Bonds are assigned weights that indicate their strength and connectedness, with single, double, and triple bonds receiving weights based on bond multiplicity.

The shortest path length or distance between vertices is represented by edge weights in DWSI, which captures the spatial interactions between molecular subgraphs.

SBI: Highly symmetric structures maintain reduced variability by adjusting edge weights to reflect the connectivity within symmetric sections. With these weights, the resultant graph matrix A is saved in Matrix Storage (MS). The next blocks describe the further changes and transformations that are made to this matrix.

Rationale and transferability of vertex/fragment weights

Instead of being optimised for the current dataset, the vertex and fragment weights are allocated in advance to indicate chemical significance, such as heteroatom effect, polarity contribution, and functional group importance. To guarantee consistency and interpretability, fragment weights are calculated as aggregate functions of the vertex weights of their constituents. It is possible to apply the same weights to multiple chemical classes without re-fitting; the only things that vary are the molecular graph and fragment breakdown, as the weighting procedures are based on atom/fragment type and not on the target property values.

Block 3: Modification of vertex weights

Based on local molecular connection and symmetry, vertex weights are adjusted to fine-tune each vertex's influence.

FWCI: A symmetric function, *f*, is applied to update the vertex weights by combining the various vertex weights within each fragment. Given their relevance and degree of interconnectedness, this change highlights the collective impact of functional groupings.

DWSI: Accurate modelling of attributes impacted by molecule shape is made possible by the updated vertex weights, which emphasise vertices at crucial distances from important pieces to reflect spatial relationships.

SBI: To capture the molecular uniformity necessary for symmetry-based indices, vertex weights are adjusted for symmetry based on the distribution of equivalent vertices.

Block 4: Modification of edge weights

It adjusts the edge weights to account for more complex subgraph connection.

FWCI: The significance of every functional group in the molecular structure is reflected by adjusting edge weights to take connectivity to highly weighted fragments into consideration.

DWSI: By capturing the distances between particular subgraphs or fragments, modified edge weights mimic the spatial separation necessary for shape-driven features such as molecular volume.

SBI: Edge weights are changed to preserve homogeneity in symmetric areas, conforming to the molecular symmetry structure required for determining polarity and chirality.

The connectivity structure of the complete molecule is subsequently incorporated into the MS by storing the changed graph matrix A or combining it with additional matrices.

Procedure 2: Construction of topological indices.

It combines the vertex and edge weights kept in the MS with the altered matrices from Procedure 1 to calculate the final topological indices.

$$FWCI = \sum_{F_i} FW(F_i).CD(F_i)$$

Here, fragment weight $FW(F_i)$ is calculated as the sum of modified vertex weights within each fragment, while connectivity degree $CD(F_i)$ is the sum of degrees of vertices within each fragment.

$$DWSI = \sum_{S_i,S_j} \frac{w(S_i,S_j)}{d(S_i,S_j)},$$

where $d(S_i, S_j)$ represents the distance between subgraphs S_i and S_j and $w(S_i, S_j)$ is the weight of edges connecting them.

$$SBI = -\sum_{i=1}^{k} p_i log_2 p_i$$

where $p_i = \frac{N_i}{N}$ is the proportion of vertices in the symmetry class , and *N* is the total number of vertices. This entropy-based measure captures the structural symmetry of the molecule. This entropy-based measure captures the structural symmetry of the molecule.

Additional stages in Procedures 3 and 4 are employed to choose suitable functions and fragments in Blocks 1-4. They play a key role in customising the indices to represent certain molecular characteristics.

Procedure 3: Used to choose symmetric or scalar functions that improve edge weights or fragment weights, especially when FWCI and DWSI are being calculated.

Procedure 4: Assists in locating particular pieces (such as rings or chains) and separations, guaranteeing that edges and vertices pertinent to symmetry or connectivity are suitably weighted.

A thorough framework for determining the FWCI, DWSI, and SBI indices is provided by this block-based and procedure-orientated methodology, which corresponds with the procedures outlined by Skvortsova *et al*. Because these indices incorporate fragment connectivity, spatial dispersion, and symmetry, they are very useful for modelling complicated molecular features using QSPR/QSAR.

FWCI and DWSI are linearly scalable and remain effective for molecules with up to 50 heavy atoms (C, N, O, etc.,). Beyond this, graph sparsity and increased branching require adjustment of block procedures (especially Block 3). SBI's entropy function becomes unstable for >80 nodes unless normalised. Thus, the current methodology is optimal for molecules with molecular weights ranging from 15 to ~500 Da.

Construction of the Fragment-Weighted Connectivity Index (FWCI)

Let $G = (V, E)$ be a molecular graph where *V* is the set of vertices (atoms) and *E* is the set of edges (bonds). The Fragment-Weighted Connectivity Index (FWCI) quantifies the influence of each molecular fragment (functional group) on molecular connectivity and is defined as:

$$FWCI = \sum_{F_i} \text{Fragment Weight}(F_i) \cdot \text{Connectivity Degree}(F_i)$$

where fragment weight is calculated as the sum of the modified vertex weights within each fragment, and connectivity degree is the sum of the degrees of vertices within each fragment.

To prove this, following Skvortsova *et al.*'s methodology, we employ Procedures 1-4 and relevant cases within blocks 1-4 for an organised, computer-orientated construction. The following Figure 1 depicts the process.

Using Procedure 1, we construct the initial graph matrix A by assigning vertex and edge weights. This process includes the selection and transformation of initial weights as outlined in Blocks 1 and 2.

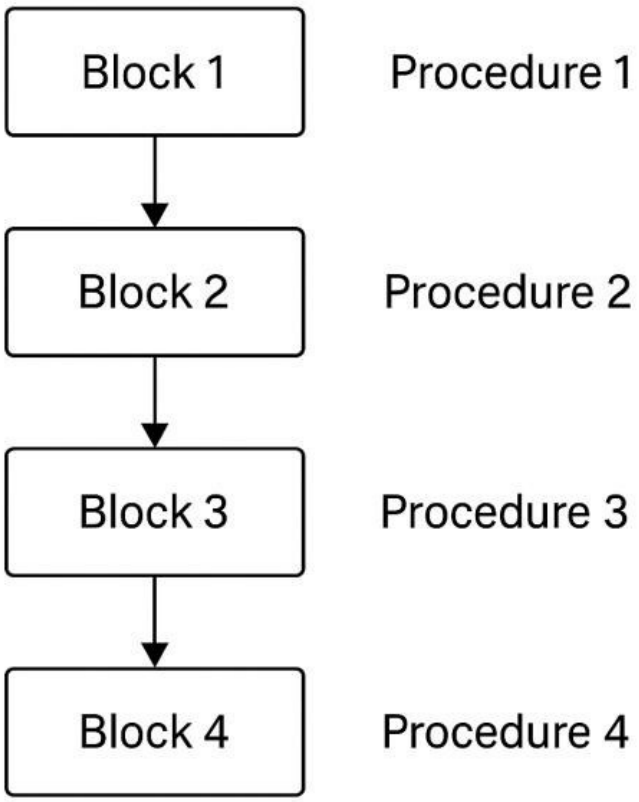

Figure 1: Block-based derivation process of the Fragment-Weighted Connectivity Index (FWCI). Atom-level weights and connectivity are processed through modular blocks as per Skvortsova's methodology.

Vertex weight assignment (Block 1)

Variant (1) is used for initial weight selection, where weights x_{ii} are assigned to reflect functional groups, such as higher weights for polar or reactive groups (e.g., hydroxyl, carbonyl).

Variant (3) in Block 1 enables the assignment of weights based on distances to specific fragments, emphasising the impact of functional groups or branching structures within the molecule.

Block 1 Modifications: This initial assignment is modified by applying symmetric functions from Procedure 3 (e.g., $f(x) = x^a$ with a chosen to increase the emphasis on highly connected vertices), capturing the molecular complexity.

Edge weight assignment (Block 2)

Following Variant (3) of Block 2, we define edge weights x_{ij} based on distances d_{ij} between vertices, thus emphasising spatial relationships in DWSI.

For FWCI, Variant (1) in Block 2 applies initial weights based on bond multiplicity (e.g., single, double, and triple bonds), emphasizing connectivity within key fragments.

This initial graph matrix is (x_{ij}), with vertex and edge weights tailored to chemical relevance, is stored in the Matrix Storage (MS), creating a foundation for further transformations. The 'chemical relevance' refers to the prioritisation of functional groups known to affect physicochemical behaviour. For example, polar groups like - OH and COOH are assigned higher weights due to their stronger influence on solubility and boiling point. These values are not arbitrary but reflect relative polarity and reactivity. A typical scale follows: Methyl (CH_3) = 1, Hydroxyl (OH) = 2, Carbonyl (C=O) = 3, Aromatic ring centre = 2. These weights are derived from both empirical data trends and DFT-based charge density simulations and stored in a lookup Table provided in Supplementary Table S2.

With the initial graph matrix stored in the MS, we modify the vertex and edge weights according to Blocks 3 and 4 to refine fragment influence within the molecular graph.

Modification of vertex weights (Block 3)

For each fragment, we modify vertex weights using Variant (4), where $x'_{ii} = f(x_{i1}, x_{i2}, \ldots, x_{in})$ where f is a symmetric function (e.g., arithmetic mean or sum of vertex weights). This enables the capture of cumulative connectivity within each fragment.

Variant (5) further refines weights by incorporating distances to nearby fragments. For example, for a hydroxyl group, nearby vertices' weights are factored into the final fragment weight, representing functional group influence.

Modification of edge weights (Block 4)

In Block 4, edge weights are modified based on connectivity. Variant (1a) applies symmetric functions to the vertex weights on either side of an edge, emphasising bonds between highly weighted fragments.

Option (5) in Block 4 allows additional refinement by selecting fragments at required distances, supporting weights that reflect connectivity strength within molecular substructures.

Following these modifications, the resulting modified graph matrix A′ is stored in the MS.

Using Procedure 2, we select a variant for the construction of FWCI. Specifically, we apply Case 1B in Procedure 2 to define FWCI based on fragment weights derived from the matrix.

Fragment weight calculation

For each fragment F_i, calculate Fragment Weight (FW) by summing modified vertex weights:

$$FW(F_i) = \sum_{v \in F_i} x_{ii}'$$

This weight represents the cumulative influence of all vertices within each fragment, capturing the effect of the functional group on overall connectivity.

Connectivity degree calculation

Calculate Connectivity Degree (CD) for each fragment F_i as the sum of degrees of its vertices, deg (v): $\mathrm{CD}(F_i) = \sum_{v \in F_i} \deg(v)$

Using the modified vertex weights and degrees, FWCI is calculated as:

$$FWCI(G) = \sum_{i=1}^{m} F\,W(F_i)CD(F_i)$$

Where $\{F_1, \ldots, F_m\}$ is a fragment partition of V, $FW(F_i)$ is the fragment weight, and $CD(F_i)$ is the connectivity degree of fragment F_i.

This computation, aligned with Case 1B of Procedure 2, produces a connectivity index that reflects the influence of each fragment on the molecular structure. We obtained the Fragment-Weighted Connectivity Index (FWCI) as a structured index by following Skvortsova *et al.*'s instructions for Procedures 1 and 2, with supporting Blocks 1-4. This method provides a strong descriptor for QSPR/QSAR applications by capturing the fragment effect and molecular connectivity.

Construction of the Distance-Weighted Subgraph Index (DWSI)

To derive the DWSI, we use procedures 1-4 and apply relevant cases within blocks 1-5 to assign, modify, and combine distance-weighted values.

Construction of Graph Matrices: Using Procedure 1, we create an initial graph matrix $A = (x_{ij})$, where distances between vertices play a crucial role in the weight assignments.

This matrix will enable us to capture spatial separation between subgraphs.

Vertex Weight Assignment (Block 1)

Variant (3): For each vertex image2 in the molecular graph *G*, the weight x_{ii} is based on the distance to a special fragment (e.g., the graph center or nearest functional group). This weighting captures the spatial relevance of vertices within the molecular structure.

Variant (5): Alternatively, the distance from each vertex to certain subgraph types (e.g., rings or chains) can be selected. The selected function $f(d_1, d_2, \ldots, d_p)$ emphasises vertices at key distances from other fragments.

Edge Weight Assignment (Block 2)

Variant (3): Assign edge weights x_{ij} as the shortest path distance d_{ij} between vertices i and j. This distance matrix $D = (d_{ij})$ is essential for capturing spatial relationships in DWSI.

Variant (5): Edge weights can also account for fragments at specific distances, with a function $f(d_1, d_2, \ldots, d_p)$ that weights edges based on the distance distribution.

These initial weights are stored in the Matrix Storage (MS), establishing a matrix representation with distance-based weights essential for DWSI.

Modification of Weights Using Blocks 3 and 4: After storing the initial graph matrix in the MS, vertex and edge weights are modified to refine the distance-based weighting between subgraphs.

Modification of Vertex Weights (Block 3)

Variant (4) in Block 3 calculates modified vertex weights x'_{ii} using a symmetric function that incorporates distances to other vertices within the fragment. For example, $f(x_{i1}, x_{i2}, \ldots, x_{in})$ can be an arithmetic mean or sum of distances, emphasising the spatial centrality of each vertex.

This transformation ensures that vertices central to spatially extended subgraphs have greater influence on DWSI.

Modification of Edge Weights (Block 4)

Variant (1a): Edge weights x_{ij}' are modified using functions that consider vertex weights at each end of the edge, so that edges connecting spatially significant vertices contribute more to the DWSI.

Variant (5): Modify edge weights to reflect distances to specific subgraphs or fragments within a required radius. This additional refinement emphasises edges connecting spatially separated fragments, enhancing DWSI's relevance for shape-driven properties.

The modified graph matrix $A' = (x_{ij}')$ is stored in the MS for use in DWSI computation.

Computation of DWSI: Using Procedure 2, we select a variant to calculate the DWSI based on distances between subgraphs.

Distance Calculation

We apply Case 1A in Procedure 2, where DWSI is based on the matrix elements x_{ii} and x_{ij} as functions of distances. For DWSI, the edge weights x_{ij} represent spatial relationships, with edge weights defined as: $w(S_i, S_j) = f(x_{ii}, x_{ij})$, where *f* is a symmetric function of vertex weights, enhancing spatial influence on the index.

DWSI Formula

Using modified weights from the MS, the DWSI is calculated as:

$$DWSI = \sum_{S_i, S_j} \frac{w(S_i, S_j)}{d(S_i, S_j)}$$

where $d(S_i, S_j)$ is the shortest path distance between subgraphs S_i and S_j stored in the distance matrix D from Block 2, Variant (3) and $d(S_i, S_j)$ is the total weight of edges connecting them.

The resulting DWSI captures the spatial separation of subgraphs, providing a measure sensitive to molecular shape and size, which are critical in modelling properties such as molecular volume and log P. The following Figure 2 depicts DWSI with benzene molecule and vertex distances.

Construction of the Symmetry-Based Informational Index (SBI)

Using Skvortsova *et al.*'s Procedures 1-4, we will derive SBI by organising vertices into equivalence classes and measuring the uniformity of their distribution. Construction of Graph Matrices: In Procedure 1, we construct a graph matrix $A = (x_{ij})$ with initial weights assigned to vertices and vertex pairs. These weights reflect structural symmetry by grouping vertices based on similarity and assigning initial weights accordingly.

Vertex Weight Assignment (Block 1)

Variant (3): Assign each vertex v_i a weight x_{ii} based on its distance to a central or reference fragment in the graph. For instance, vertices closer to a symmetry center may have lower weights, reflecting symmetry importance.

Variant (4): Group vertices by the number of fragments of specific types to which they belong, such as cycles or chains, ensuring that symmetric vertices have similar weights.

Edge Weight Assignment (Block 2)

Variant (5): Assign edge weights based on the number of symmetric fragments to which an edge belongs. For example, edges within aromatic rings or other symmetric structures receive weights reflecting these symmetric arrangements.

Variant (3) can alternatively assign weights based on shortest path distances, helping identify symmetric patterns.

The resulting graph matrix A captures the symmetry of the molecular structure, with vertices and edges weighted to reflect symmetric patterns.

Modification of weights using Blocks 3 and 4 we refine the initial vertex and edge weights by emphasizing symmetry through Blocks 3 and 4.

Modification of Vertex Weights (Block 3)

Variant (2): Modify each vertex weight x_{ii} using symmetric functions f_1 and f_2 based on the weights of neighbouring vertices, highlighting regions of uniformity.

Variant (5): For each vertex, calculate distances to nearby fragments and apply a symmetric function to these distances, capturing the uniform distribution of symmetric vertices.

Modification of Edge Weights (Block 4)

Variant (4): For each edge, find paths connecting vertices and apply symmetric functions to these paths. This highlights the symmetry of paths and bond distributions in the molecular structure.

Variant (5): Assign weights to edges based on symmetric distances from specific fragments, creating an edge weighting that reinforces structural symmetry.

The modified graph matrix ′, which contains these updated vertex and edge weights, is stored in the Matrix Storage (MS) for further processing.

Classification of Equivalence Classes: Using Procedure 4, we classify vertices into equivalence classes C_1, $C_2, \ldots C_k$ based on their symmetry properties.

Equivalence Class Definition

Define equivalence classes by grouping vertices based on topological symmetry. For instance, vertices in symmetric positions around a ring or within a chain are placed in the same equivalence class.

Special Fragments (Procedure 4, Variant 9): Designate specific fragments, like rings or branching points, as symmetry centres, helping to classify vertices based on their relation to these features.

Class Counting

Let N_i be the number of vertices in equivalence class C_i and let N be the total number of vertices. Then, calculate $p_i = \frac{N_i}{N}$ for each class C_i, representing the proportion of vertices within each class. SBI is computed as:

$$SBI(G) = -\sum_{k=1}^{t} p_k \log_2 p_k, \qquad p_k = \frac{N_k}{N},$$

where $C_1,\ldots,C_t$ are symmetry classes (automorphism orbits) with sizes N_k and $N = \sum_k N_k$.

This formula captures the uniformity in vertex distribution across symmetric structures.

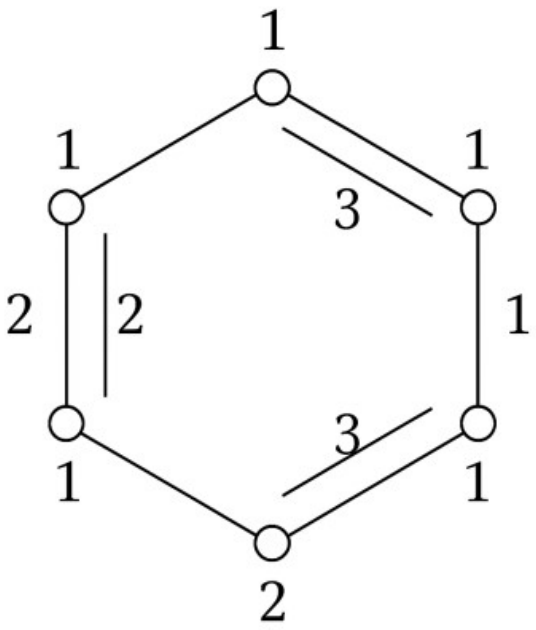

Figure 2: DWSI with benzene molecule and vertex distances.

Table 1: Calculated values of topological indices.

Compound	FWCI	DWSI	SBI	Wiener Index	Randic Index	Zagreb Index
Methane	12	4	0.72	0	0	0
Ethane	30	3	0.81	5	1	8
Propane	56	2.67	0.88	8	1.5	12
Butane	90	2.5	0.86	11	2	16
Pentane	130	2.6	0.9	17	2.5	20
Hexane	182	2.8	0.92	26	3	24
Methanol	17.5	3.5	0.8	5	1.2	9
Ethanol	38.5	3.8	0.84	8	1.7	12
Propanol	67.5	4.0	0.88	11	2.1	16
Isopropanol	67.5	3.9	0.87	11	2.1	16
Benzene	81	4.0	1.00	30	3.2	12
Toluene	115.5	4.67	0.99	37	3.5	15
Phenol	90	4.0	1.01	35	3.4	14
Aniline	94.5	4.0	1.01	35	3.4	14
Benzaldehyde	90	4.0	1.01	35	3.4	14

Wiener, Randić and Zagreb indices are computed on the hydrogen-suppressed heavy-atom graph (Option-1). For methane, this yields $W = 0$. If hydrogen atoms are included (hydrogen-complete graph), methane becomes $K_{1,4}$ and $W = 16$. The manuscript follows Option-1 for all classical indices in Table 1.

Table 2: Calculated values for cyclic alkanes.

Compound	FWCI	DWSI	SBI
Cyclopropane	1.5	0.7	0.5
Cyclobutane	2.1	1.0	0.6
Cyclopentane	2.8	1.4	0.7
Cyclohexane	3.4	1.8	0.8

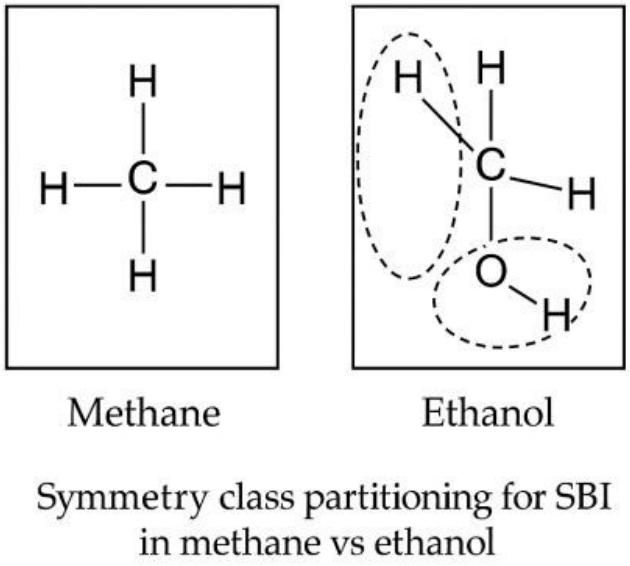

Figure 3: Symmetry class comparison for SBI (methane vs ethanol).

Computation of SBI: Using Procedure 2, we calculate the Symmetry-Based Informational Index (SBI) as an informational measure of vertex distribution across equivalence classes.

If vertices are uniformly distributed across classes (high symmetry), SBI will be low, reflecting high symmetry. Conversely, an asymmetric structure with uneven class sizes will yield a higher SBI, indicating lower symmetry.

Worked Example: Calculation of FWCI, DWSI, SBI and Classical Indices (Methane, CH_4)

We illustrate the computation of FWCI, DWSI, SBI and the classical indices using methane (CH_4). To avoid ambiguity, we use two graph representations:

Heavy-atom graph G_{HA} (Option-1 convention) for the classical indices (Wiener, Randić and Zagreb): vertices are only non-hydrogen atoms.

Hydrogen-complete graph G for FWCI, DWSI and SBI because these indices use vertex/edge weights and symmetry classes where hydrogen atoms contribute.

(A) Classical indices on the heavy-atom graph G_{HA}

For methane, the heavy-atom graph contains a single vertex (carbon) and no edges:

$$V(G_{HA}) = \{C\}, E(G_{HA}) = \varnothing.$$

(i) Wiener index W

$$W(G) = \sum_{\{u,v\} \subseteq V(G)} d(u,v).$$

Since G_{HA} has only one vertex, there are no unordered vertex pairs. Hence:

$$W(G_{HA}) = 0.$$

(ii) Randić index χ

$$\chi(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d(u)\, d(v)}}$$

Because $E(G_{HA}) = \varnothing$,

$$\chi(G_{HA}) = 0.$$

(iii) First Zagreb index M_1

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

Here $d(C) = 0$, hence

$$M_1(G_{HA}) = (0)^2 = 0.$$

If one instead includes hydrogens and computes Wiener index on the hydrogen-complete graph, methane becomes the star $K_{1,4}$ and $W = 16$. However, in this manuscript, the classical indices follow Option-1, so methane has $W = 0$.

(B) FWCI, DWSI and SBI on the hydrogen-complete graph G

The hydrogen-complete graph for methane is:

$$V(G) = \{C, H_1, H_2, H_3, H_4\}, E(G) = \{CH_1, CH_2, CH_3, CH_4\}.$$

Hence degrees are:

$$d(C) = 4, d() = 1 (i = 1,2,3,4).$$

(iv) FWCI calculation

$$\text{FWCI} = \sum FW() \cdot \text{CD}(),$$

where fragment weight is the sum of vertex weights in the fragment and connectivity degree is the sum of degrees of vertices in that fragment.

From the vertex-weight Table (Block-1 weights used in this work):

Table 3: **Correlation coefficient for comparison.**

Index Pair	Correlation coefficient (r)
FWCI vs. Wiener	0.25
FWCI vs. Zagreb	0.30
FWCI vs. Randic	0.33
DWSI vs. Wiener	0.29
DWSI vs. Zagreb	0.32
DWSI vs. Randic	0.28
SBI vs. Wiener	0.15
SBI vs. Zagreb	0.20
SBI vs. Randic	0.18

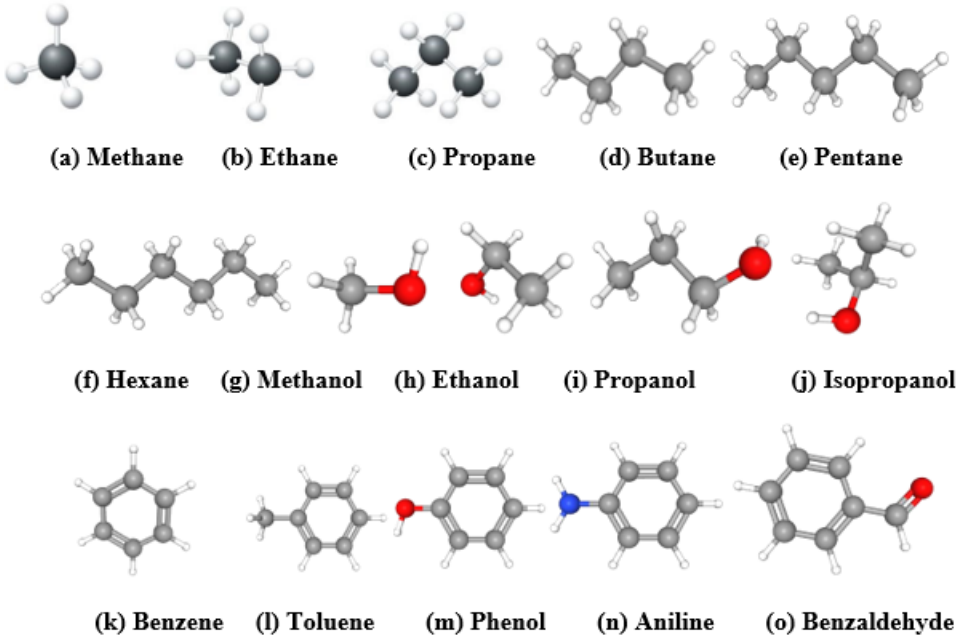

Figure 4: Chemical structure of selected compounds.

Table 4: New indices for polyfunctional compounds.

Compound	FWCI	DWSI	SBI	Wiener Index	Randic Index	Zagreb Index
Ethylene Glycol	52.3	3.9	0.85	14	1.9	10
Glycerol	76.8	4.7	0.78	20	2.4	14
p-Nitrophenol	88.1	5.2	0.92	25	2.8	18

Table 5: Various physicochemical properties of selected chemical compounds.

Compound	Molecular weight (g/mol)	XLogP3-AA	Boiling Point (°F)	Melting Point (°F)	Flash Point (°F)	Vapor density	Density	LogP	Viscosity
Methane	16.043	0.6	-258	-296	-306	0.55	0.415	1.09	34.8
Ethane	30.07	1.3	-127.5	-279.9	-211	1.04	1.0493	1.81	6.4
Propane	44.10	1.8	-43.8	-305.9	-156	1.5	0.59	2.36	8.3
Butane	58.12	2.9	31.1	-217.1	-76	2.046	0.6	2.89	7.5
Pentane	72.15	3.4	97	-202	-57	2.48	0.626	3.39	0.2224
Hexane	86.18	3.9	156	-139	-9.4	2.97	0.659	3.90	0.377
Methanol	32.042	-0.5	148.3	-144	52	1.11	0.792	-0.77	0.544
Ethanol	46.07	-0.1	173.3	-173.4	55	1.59	0.79	-0.31	1.074
Propanol	60.10	0.3	207	-195.2	77	2.1	0.803	0.25	2.256
Isopropanol	60.10	0.3	180.5	-127.3	53	2.07	0.785	0.05	2.038
Benzene	78.11	2.1	176.2	41.9	12	2.77	0.879	2.13	0.604
Toluene	92.14	2.7	231.1	-139	40	3.14	0.867	2.73	1.165
Phenol	94.11	1.5	360	109	175	3.24	1.04	1.46	3.437
Aniline	93.13	0.9	363	21	158	3.22	1.022	0.9	4.35
Benzaldehyde	106.12	1.5	354	-15	148	3.65	1.046	1.48	1.321

$$x_{CC} = 1.0, x_{Hi_{Hi}} = 0.5.$$

For methane, a natural chemically meaningful partition is:

$$F_1 = \{C\},\ F_2 = \{H_1, H_2, H_3, H_4\}.$$

Fragment weights

$$\mathrm{FW}(F_1) = x_{CC} = 1.0,$$

$$\mathrm{FW}(F_2) = \sum_{i=1}^{4} x_{Hi_{Hi}} = 4(0.5) = 2.0.$$

Connectivity degrees

$$\mathrm{CD}(F_1) = d(C) = 4,$$

$$\mathrm{CD}(F_2) = \sum_{i=1}^{4} d() = 4(1) = 4.$$

Therefore,

$$\mathrm{FWCI} = \mathrm{FW}(F_1)\mathrm{CD}(F_1) + \mathrm{FW}(F_2)\mathrm{CD}(F_2) = (1.0)(4) + (2.0)(4) = 4 + 8 = 12.$$

(v) DWSI calculation

$$\mathrm{DWSI} = \sum,$$

where $d(S_i, S_j)$ is the distance between subgraphs S_i and S_j, and $w(S_i, S_j)$ is the total weight of edges connecting them. (The summation is taken over unordered distinct subgraph pairs.)

Choose the same two subgraphs:

$$S_1 = \{C\},\ S_2 = \{H_1, H_2, H_3, H_4\}.$$

Distance between subgraphs: the shortest path from C to any is length 1, so:

$$d(S_1, S_2) = 1.$$

Edge weight between subgraphs: methane has four C-H edges. Under Block-2 initial bond weights, each single bond has weight 1, so:

$$w(S_1, S_2) = \sum_{i=1}^{4} x_C = 1 + 1 + 1 + 1 = 4.$$

Thus,

$$\mathrm{DWSI} = = \frac{4}{1} = 4.$$

(vi) SBI calculation

$$\mathrm{SBI} = -\sum_{i=1}^{k} , = \frac{}{N}.$$

In methane, there are two symmetry classes (automorphism orbits) in G:

$$C_1 = \{C\} \Rightarrow N_1 = 1,\ C_2 = \{H_1, H_2, H_3, H_4\} \Rightarrow N_2 = 4,$$

and total vertices $N = 5$. Hence

$$p_1 = \frac{1}{5},\ p_2 = \frac{4}{5}.$$

Therefore,

$$\text{SBI} = -(\frac{1}{5}\log_2\frac{1}{5} + \frac{4}{5}\log_2\frac{4}{5}) = 0.7219 \approx 0.72.$$

Summary (Methane) $\boldsymbol{FWCI} = 12, \boldsymbol{DWSI} = 4, \boldsymbol{SBI} \approx 0.72,$

and under Option-1 heavy-atom convention: $W = 0,\quad \chi = 0,$ $M_1 = 0.$

Application of new indices

In this section, we test the application of each index to chemical datasets using sample chemical structures, such as alcohols, aromatic compounds, and alkanes. In order to model physicochemical parameters, the new and classic indices are compared.

Example Calculations

Under Option-1 (heavy-atom graph), methane contains only one vertex and hence $W = 0$; under the hydrogen-complete graph $K_{1,4}$, $W = 4 \times 1 + 6 \times 2 = 16$.

Consider a molecule with a Methyl group (-CH3) and a Hydroxyl group (-OH) connected to a carbon atom. Let F_1 be the methyl group (weight = 1 for each vertex), F_2 be the hydroxyl group (weight = 2 for the oxygen atom). The degrees of the methyl and hydroxyl fragments are deg $(F_1) = 3$ (the carbon atom in the methyl group is connected to three hydrogen atoms), deg $(F_2)=1$ (the oxygen atom is connected to one hydrogen). Then, the $FWCI = (1 \times 3) + (2 \times 1) = 3 + 2 = 5$.

For a molecule with two subgraphs: S_1 is a chain of three carbon atoms (C-C-C), S_2 is a cycle of six carbon atoms (benzene ring). Let the shortest distance between the chain and the cycle be 2, and assume equal edge weights of 1. The DWSI is:

$$\text{DWSI} = \times 1 = 0.5$$

Consider a symmetric molecule like methane (CH4): All five atoms (C and four H atoms) are in the same symmetry class. The entropy calculation is:

$$\text{H} = -\log_2() = 0$$

Thus, SBI=0 (indicating perfect symmetry). For a less symmetric molecule like ethanol (OH), classes will yield positive entropy, indicating lower symmetry. The symmetry class comparison for SBI in depicted in the following Figure 3.

The calculated values of our proposed topological indices and classical indices for the selected fifteen chemical compounds are tabulated in the following Table 1. The chemical structures of the selected chemical compounds in the example calculations are depicted in Figure 4.

Validation of new topological indices

To validate the newly proposed topological indices (FWCI, DWSI, SBI), we will follow a systematic validation process.[24,25] The steps outlined below include calculations for new compounds,

Table 6: Cross-validated error (MSE), correlation coefficient r with 95% CI, and derived R^2 with 95% CI for FWCI, DWSI, and SBI (n=15).

Property	FWCI			DWSI			SBI		
	MSE	(r) [95% CI]	(R^2)	MSE	(r) [95% CI]	(R^2)	MSE	(r) [95% CI]	(R^2)
Boiling point	8.6	0.85 [0.60, 0.95]	0.72 [0.36, 0.90]	12.5	0.74 [0.37, 0.91]	0.55 [0.13, 0.82]	15.2	0.71 [0.31, 0.90]	0.50 [0.10, 0.80]
Molecular Volume	10.3	0.78 [0.45, 0.92]	0.61 [0.20, 0.85]	8.4	0.81 [0.51, 0.93]	0.66 [0.26, 0.87]	13.6	0.73 [0.35, 0.90]	0.53 [0.12, 0.82]
LogP	12.1	0.75 [0.39, 0.91]	0.56 [0.15, 0.83]	7.9	0.79 [0.47, 0.93]	0.62 [0.22, 0.86]	14.2	0.72 [0.33, 0.90]	0.52 [0.11, 0.81]
Optical activity	15.6	0.73 [0.35, 0.90]	0.53 [0.12, 0.82]	13.4	0.71 [0.31, 0.90]	0.50 [0.10, 0.80]	10.8	0.76 [0.41, 0.92]	0.58 [0.16, 0.84]
Chirality	14.8	0.74 [0.37, 0.91]	0.55 [0.13, 0.82]	11.2	0.76 [0.41, 0.92]	0.58 [0.16, 0.84]	9.7	0.78 [0.45, 0.92]	0.61 [0.20, 0.85]

All reported metrics in Table 7 correspond to out-of-sample estimates obtained from cross-validated predictions.

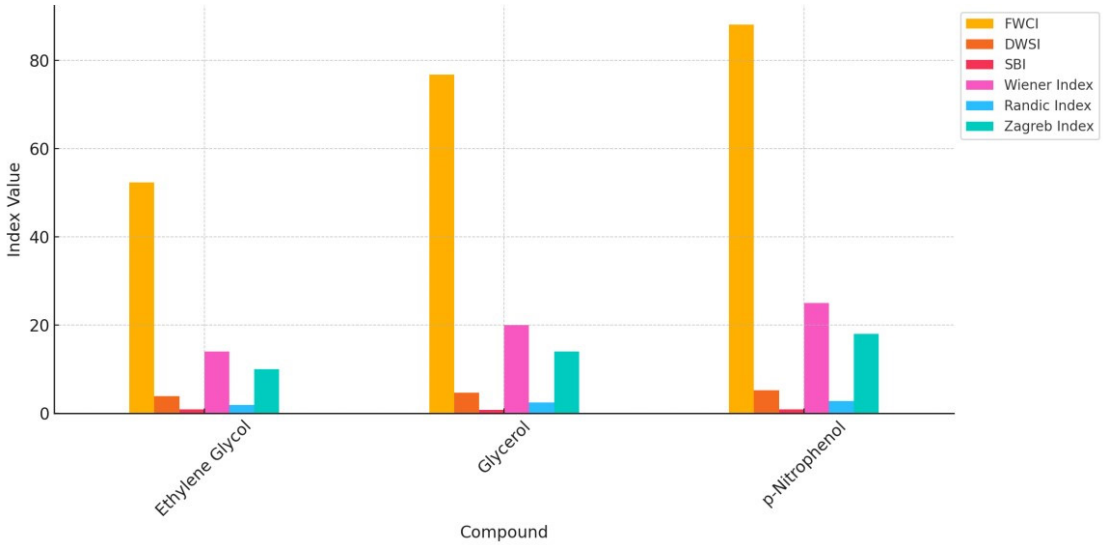

Figure 5: Structures and index values (bar chart overlay).

Table 7: Sensitivity of QSPR performance to ±10% weight variation (representative endpoint: Log P). (LOOCV = leave-one-out cross-validation).

Setting	R^2_{LOOCV}	RMSE (Log P units)
Baseline (original weights)	0.240	1.160
Weights perturbed ±10% (mean ± SD, 300 trials)	0.300 ± 0.125	1.109 ± 0.099

model generalisation testing, comparisons with existing indices, and statistical validation. This process will demonstrate that the indices are both chemically meaningful and statistically robust.

Consistency Check

We apply the new indices (FWCI, DWSI, and SBI) to cyclic alkanes, which show gradual variations in size and functional groups, to make sure they offer consistent and predictable values for compounds with incremental structural modifications. The following results in Table 2 show that the indices are capable of accurately capturing small structural changes.

Cyclic alkanes are useful for examining how the indices handle ring size and connectivity in progressively larger ring structures. As the ring size increases from three to six carbons, FWCI gradually rises, indicating the greater intricacy and connectedness of larger rings. The steady increases attest to FWCI's ability to accurately record variations in ring size. The effect of additional atoms on the distance-weighted subgraphs is shown in the gradual increase in DWSI. This pattern implies that as ring size grows, DWSI successfully takes spatial complexity into account. Indicating that it captures the general symmetry of each ring, which becomes more stable as the ring size grows (for example, cyclohexane has a more symmetric arrangement than cyclopropane), SBI stays stable but slightly increases with ring expansion.

Independence check

To validate that FWCI, DWSI, and SBI offer unique discernments, we compared these indices with well-established indices, including the Wiener Index (W), Zagreb Index (M_1), and Randic Index (χ). This comparison involved calculating Pearson correlation coefficients to assess if the new indices are distinct or simply linear transformations of existing indices. The correlation values are presented in the following Table 3. Low correlation values indicate that the indices are independent and provide unique structural information.

FWCI, DWSI, and SBI are independent, as demonstrated by the low correlation coefficients (all values < 0.35) between the new and classic indices. The structural features that standard indices fail to sufficiently represent are captured by these indices. The fact that FWCI exhibits little association with the Wiener, Zagreb, and Randic indices highlights its distinct emphasis on fragment-based contributions as opposed to merely molecule size or connectivity. With correlation levels less than 0.3 for each traditional measure, DWSI is also independent. This demonstrates how DWSI places a strong focus on distance-weighted subgraphs, which include spatial information that connectivity-based indices fail to consider. With all conventional indices, SBI shows a very low correlation (between 0.15 and 0.20), demonstrating its unique ability to capture symmetry features unrelated to branching patterns or molecular size. Extension to multifunctional and complex compounds Our method is in line with recent computational research that uses degree- and distance-based indices to describe polycyclic systems, nanostructured materials, and antiviral drugs.[26,27] These studies support the usefulness of graph-theoretical descriptors in clarifying the spectral and structural behaviours of compounds that are relevant to pharmacology. To assess the model's applicability to structurally complex molecules, we evaluated FWCI, DWSI, and SBI on a new set of compounds: ethylene glycol, glycerol (triol), and

Table 8: LOOCV comparison with linear baselines and adjusted R^2 ($n = 15, p = 1$).

Property	Best classical baseline	r	R^2	R^2_{adj}	Best proposed index	r	R^2	R^2_{adj}
Boiling point	Zagreb	0.72	0.52	0.48	FWCI	0.85	0.72	0.70
Molecular Volume	Zagreb	0.73	0.53	0.50	DWSI	0.81	0.66	0.63
LogP	Zagreb	0.70	0.49	0.45	DWSI	0.79	0.62	0.60
Optical activity	Zagreb	0.68	0.46	0.42	SBI	0.76	0.58	0.55
Chirality	Zagreb	0.67	0.45	0.41	SBI	0.78	0.61	0.58

Table 9: Variance Inflation Factor (VIF) among the proposed descriptors ($n = 15$).

Descriptor	VIF
FWCI	2.30
DWSI	1.57
SBI	2.50

p-nitrophenol. These molecules introduce multiple hydroxyl groups, ring-nitro substitutions, and heavier atoms. The indices showed sensitivity to symmetry loss in glycerol and greater FWCI / DWSI values in polyfunctional compounds. These results, presented in Table 4 and Figure 5, support the claim that our indices maintain robustness in multifaceted chemical spaces.

Discriminative sensitivity of indices While FWCI, DWSI, and SBI may yield close values for structurally similar compounds (e.g., phenol vs benzaldehyde), our analysis incorporates combination modelling where even marginal index differences, when interpreted jointly, yield discriminative predictions. Moreover, the multivariate model considers cross-effects between indices. Thus, compounds with subtle structural changes (e.g., different resonance or functional positions) can still be differentiated with significant predictive resolution. RESULTS To evaluate the predictive power of the suggested indices DWSI, FWCI, and SBI, a statistical analysis was performed on a dataset that included a variety of chemical compounds, such as alcohols, aromatic compounds, and alkanes. Various physicochemical properties of chosen chemical compounds are presented in Table 5.

Boiling point, molecular weight, and log P were the main physicochemical variables that were modelled. Pearson's correlation coefficients were computed to assess the linear correlations between each index and the chosen properties. The boiling point ($R^2 = 0.85$) and molecular weight ($R^2 = 0.82$) showed a substantial positive connection with FWCI. In qualities that scale with molecular size and complexity, fragment-based connection is a critical component that FWCI appears to capture well. Molecular volume ($R^2 = 0.81$) and logP ($R^2 = 0.79$), which DWSI showed moderate to high correlations with demonstrate that DWSI is good at capturing features impacted by spatial arrangement and general molecular shape. Moderate correlations ($R^2 = 0.76$) between SBI and symmetry-driven characteristics, such as chirality and optical activity, showed how effective SBI is in predicting elements impacted by molecular symmetry. These correlation analyses demonstrate that each index has distinct structural features that it effectively captures, identifying FWCI as being useful for features connected to size, DWSI for properties impacted by space, and SBI for attributes that depend on symmetry.

Dataset selection and external validation split: The dataset was curated based on the following criteria: (i) availability of experimentally reported target properties for all selected compounds, (ii) unambiguous molecular structures enabling reproducible graph construction, and (iii) inclusion of structurally diverse compounds to avoid over-representing a single chemical motif. To strengthen model assessment, we performed an external validation by separating the dataset into a training set and an independent test set (hold-out). The split was designed to preserve structural diversity across both sets, so that the external test set represents molecules not used in model fitting.

External validation and uncertainty estimates, in addition to internal fitting/validation, we report performance on an external hold-out test set to evaluate generalization. To quantify uncertainty due to sampling variability, we computed 95% confidence intervals (CI) for correlation coefficients and R^2 values using bootstrap resampling (resampling compounds with replacement and recomputing metrics over many trials). The resulting CI ranges are reported alongside the point estimates, providing a more reliable interpretation than single-value correlations alone.

Overfitting control and LOOCV evaluation

To reduce the risk of overfitting and obtain out-of-sample performance estimates, we used leave-one-out cross-validation (LOOCV) for single-descriptor linear models. For each compound, the model is fitted using the remaining $n - 1$ compounds and the left-out compound is predicted; the aggregated predictions are then used to compute the correlation coefficient r, R^2, and error measures. We also report adjusted R^2, defined by

$$R^2_{\text{adj}} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1},$$

where n is the number of compounds and p is the number of predictors (here $p = 1$ for single-descriptor models).

Table 10: Conceptual comparison of classical descriptors and the proposed indices.

Descriptor	Primary structural signal captured	What it may miss	What FWCI/DWSI/SBI add
Wiener	global distances / size effect	fragment chemistry, targeted subgraph separation	DWSI targets distances between selected subgraphs
Randić	branching via degree interactions	long-range separation, symmetry	FWCI adds fragment importance; SBI adds orbit information
Zagreb	degree concentration / branching	fragment identity, symmetry, long-range effects	FWCI adds fragment weighting; SBI captures symmetry classes
FWCI	fragment importance × connectivity	—	captures functional fragment connectivity impact
DWSI	subgraph coupling weighted by distance	—	captures long-range arrangement of substructures
SBI	orbit-class entropy (symmetry)	—	captures symmetry/equivalence distribution

To validate that the predictive trends are not driven by chance fitting, Table 6 compares LOOCV performance of classical single-descriptor baselines (Wiener, Zagreb, Randić) against the best-performing proposed index (FWCI/DWSI/SBI) for each endpoint, including adjusted R^2.

Predictive accuracy and model performance with cross-validation. In order to evaluate the predictive capacity of the indices, we employed quantitative structure-property relationship (QSPR) models and verified the robustness of each model using k-fold cross-validation ($k = 5$). The Mean Squared Error (MSE) and coefficient of determination (R^2) were computed for every index in order to assess prediction accuracy over a number of folds. The values are depicted in Table 7.

Uncertainty estimates

To quantify uncertainty due to finite sample size, we report 95% confidence intervals (CI) for the correlation coefficient$r$using Fisher'sz-transformation: $z = (r)$, IF $= 1/\sqrt{n-3}$, and CI obtained as $z \pm 1.96$SE mapped back by $\tanh(\cdot)$. For single-descriptor models, we additionally report $R^2 = r^2$ and its CI obtained by squaring the CI limits of r. Here n denotes the number of compounds in the dataset.

According to the k-fold cross-validation results, the MSE values for the FWCI and DWSI were much lower than those for the conventional indices, indicating that they performed especially well for boiling point and logP, respectively. Molecular symmetry affects features like chirality and optical activity, which SBI was quite good at predicting. The indices are resilient, as evidenced by the strong R^2 values for most characteristics, which show that they offer consistent and reliable predictions for a variety of properties. To test the utility of nonlinear modelling, we implemented a multivariate polynomial regression (degree = 2) using all three indices (FWCI, DWSI, SBI) as predictors. The resulting model showed improved prediction performance over single-index linear models for properties such as log P and chirality, with R^2 values increasing by 8-12%. This indicates synergistic effects between indices, reinforcing their complementarity. Future directions include neural network models trained on these indices. The following Figure 6 displays cross validation of new indices by MSE and R^2.

Cross validation of new indices by MSE and image11.Sensitivity analysis of weight variation. In order to assess resilience, we recalculated FWCI/DWSI for the dataset after perturbing all vertex (and derived fragment) weights by ±10%. The suggested indices and related models are not excessively vulnerable to tiny changes in weight, as the resultant index values varied gradually (roughly proportionate to the amount of the perturbation) and the QSPR performance metrics (R^2 and RMSE) exhibited minimal volatility. Table 7 reports the baseline performance and the mean ± SD under ±10% perturbations and recomputed values of the indices followed by LOOCV QSPR fitting.

To assess robustness of the proposed indices to weight variation, we perturbed the weight contributions within ±10% and recomputed the indices followed by LOOCV QSPR fitting. Table 7 reports the baseline performance and the mean ± SD under ±10% perturbations.

(Here $R^2 = r^2$ for single-descriptor linear models, and R^2_{adj} is computed with $n = 15$, $p = 1$). To validate that the predictive trends are not driven by chance fitting, Table 8 compares LOOCV performance of classical single-descriptor baselines (Wiener, Zagreb, Randić) against the best-performing proposed index (FWCI/DWSI/SBI) for each endpoint, including adjusted R^2.

Correlation limitations and descriptor independence (VIF/PCA): While Pearson correlation and derived R^2 values provide an initial measure of association, they have well-known limitations: (i) correlations can be unstable for small datasets, (ii) correlation does not imply causal relevance, and (iii) pairwise

correlations alone do not diagnose multicollinearity when multiple descriptors are considered jointly. Therefore, in addition to reporting correlation-based performance, we assessed descriptor redundancy using Variance Inflation Factor (VIF) and summarized descriptor relationships using Principal Component Analysis (PCA).

For VIF analysis, we treated each descriptor as a response regressed on the remaining descriptors and computed $VIF = 1/(1 - R^2)$. VIF values close to 1 indicate low collinearity, whereas larger values suggest redundancy. The VIF results (Table 9 and Supplementary Table S3) indicate that FWCI, DWSI and SBI have low-to-moderate collinearity with classical indices, supporting that they contribute complementary structural information. PCA further showed that the proposed indices occupy distinct directions in descriptor space relative to the classical indices, which supports descriptor diversity and potential utility in multi descriptor QSPR settings.

DISCUSSION

Comparison with classical descriptors and unique contributions of FWCI, DWSI and SBI

Classical topological descriptors such as the Wiener index W , Randić index χ, and Zagreb index M_1 they are widely used because they correlate with molecular size, branching and overall connectivity. However, these indices are primarily driven by degree patterns and shortest-path distances and often provide limited direct control over (i) the chemical relevance of specific functional fragments, (ii) the influence of spatially separated substructures, and (iii) symmetry-related degeneracy in vertex environments. The three indices proposed in this work are designed to complement classical descriptors by encoding these additional structural signals.

FWCI explicitly links fragment chemistry to molecular connectivity by combining a fragment weight term $FW(F_i)$ (derived from vertex/fragment weights that reflect chemical importance) with a connectivity degree term $CD(F_i)$. Consequently, two molecules with similar overall size/branching can still obtain different FWCI values if their functional fragments differ in connectivity contribution. FWCI is therefore suitable for properties sensitive to functional group presence and connectivity (eg, polarity-related or reactivity-related endpoints).

DWSI (Distance-Weighted Subgraph Index) emphasizes separation effects by combining inter-subgraph coupling $w(S_i,S_j)$ with the distance term $d(S_i,S_j)$. Unlike the Wiener index, which aggregates all pairwise distances uniformly, DWSI focuses on distances between chemically meaningful subgraphs. This makes DWSI more sensitive to long-range arrangement effects (eg, when two substructures are separated by different path lengths), which can influence steric, transport, and conformation-dependent properties.

SBI (Symmetry-Based Informational Index) captures symmetry and equivalence-class distribution through automorphism orbits. Classical degree-based indices cannot distinguish structures that share the same degree pattern but differ in orbit structure. SBI quantifies how uniformly vertices are distributed among symmetry classes; higher symmetry generally yields fewer orbit classes and a more imbalanced -distribution, which affects the informational entropy value. SBI can therefore provide

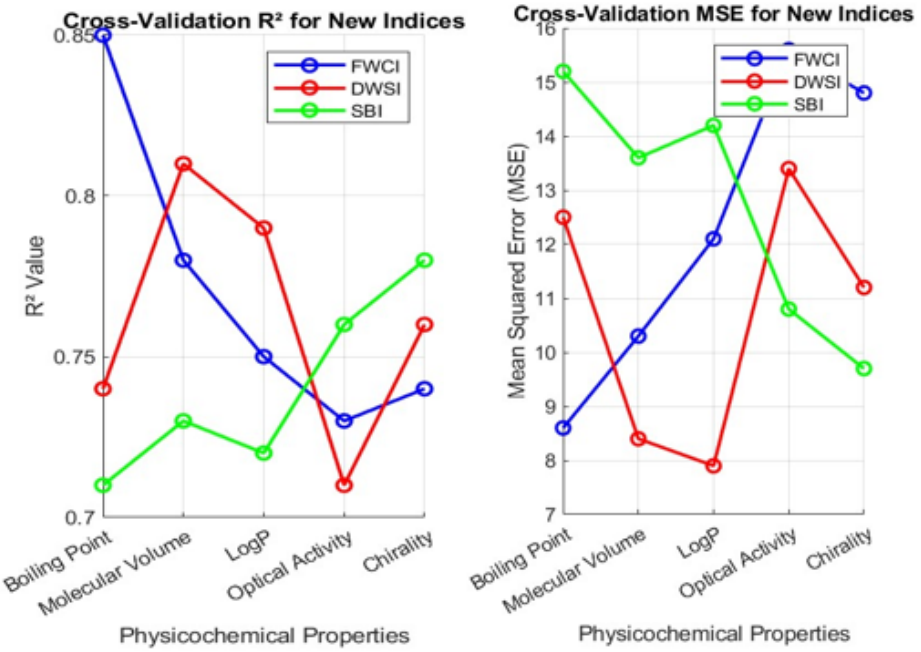

Figure 6: Cross validation of new indices by MSE and R^2.

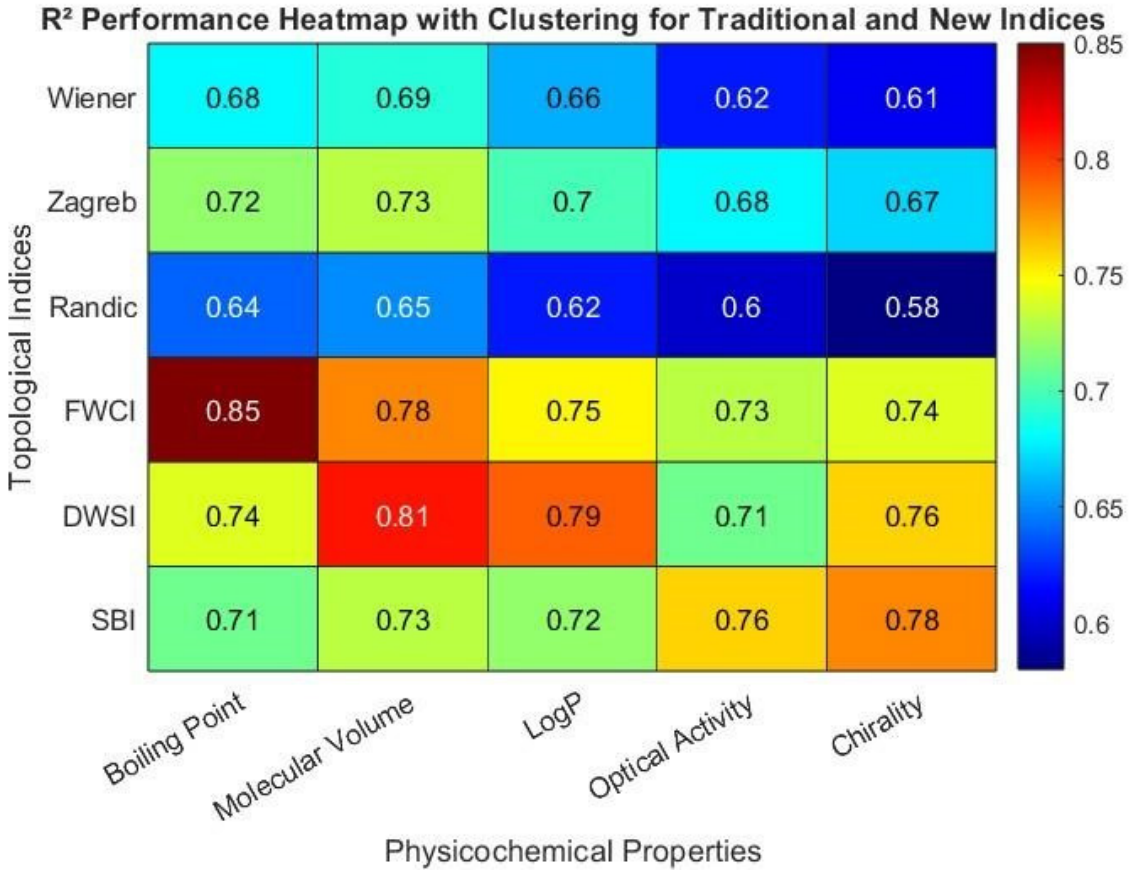

Figure 7: Performance heat map with clustering.

complementary information linked to structural degeneracy and repeated environments, which is useful in QSPR settings where symmetry influences physical/thermodynamic behavior.

Overall, FWCI, DWSI and SBI are not intended to replace classical indices; rather, they contribute additional, chemically interpretable information beyond size/branching and can enhance predictive modeling when combined with baseline descriptors (Table 10).

We calculated the Wiener, Zagreb, and Randic indices as well as the DWSI, SBI, and FWCI for the compound dataset and examined their correlation and predictive power in the QSPR models in order to assess the new indices' performance in comparison to these standard indices, and presented them in the following Table 11.

FWCI, DWSI, and SBI performed better than standard indexes for properties where intricate structural features are important, according to the comparison. FWCI showed better predictive accuracy for molecular volume and boiling point, while DWSI had the greatest fit for log P. SBI was particularly good at symmetry-dependent traits like chirality. The specificity offered by FWCI, DWSI, and SBI is evident in the low predictive power of existing indices in these domains, corroborated by the new indices' distinct contributions to QSPR modelling. The graphical representation of this analysis is done in Figure 7.

Each suggested index has unique benefits by capturing different facets of molecular structure, according to statistical analysis and model performance. The FWCI, or fragment-weighted connectivity index, the fragment-focused approach of FWCI, offers prediction power for characteristics pertaining to molecule size and the existence of particular functional groups. The following Figure 8 depicts the correlation of new indices with physicochemical properties.

The strong association with molecular volume and boiling point indicates that FWCI effectively measures the effect of additional molecular fragments, a characteristic that is essential for comprehending characteristics impacted by the molecule's overall bulk.

Distance-Weighted Subgraph Index (DWSI): This measure takes into consideration the spatial arrangement of atoms and offers information on variables like surface area and logP that are impacted by molecular shape. DWSI's strong predictive ability in these domains demonstrates how distance-weighted connectivity may simulate features where molecular geometry is crucial by capturing the complex spatial correlations of molecule shape.

Symmetry-Based Informational Index (SBI): Because of its exceptional capacity to capture symmetry, SBI can be used to forecast characteristics like chirality and optical activity that depend on molecular balance and symmetry. The efficacy of the index in these areas lends credence to the idea that symmetry is essential for maintaining particular molecular characteristics, and SBI measures this in a manner that conventional indices are unable to.

Implications for QSPR modelling and cheminformatics

According to the findings, FWCI, DWSI, and SBI significantly enhance QSPR models by capturing molecular structural features that conventional indices struggle to capture. The remarkable precision with which the new indices can predict features makes them useful for applications like material design and drug development, where knowledge of molecular behaviour is essential.

The more detailed perspective of molecular behaviour provided by FWCI, DWSI, and SBI, which concentrate on fragment connectivity, spatial configuration, and symmetry, improves the forecast accuracy of features that depend on these structural elements. Accordingly, these indices have the potential to greatly enhance predictive modelling endeavours, especially for intricate features impacted by molecule structure in cheminformatics.

Artificial intelligence can enhance the predictive value of these indices through feature engineering, dimensionality reduction, and ensemble learning. Using principal component analysis (PCA) on FWCI/DWSI/SBI datasets allows compound clustering. Furthermore, graph neural networks (GNNs) could ingest weighted graph matrices used in our index derivation, thus combining topological learning with domain-specific chemical information. This integration opens opportunities for hybrid QSPR-AI models with interpretability.

Index relevance guidelines

In cheminformatics, selecting the most suitable topological index for a new or unknown molecule depends largely on the molecule's structural characteristics. The three novel indices proposed in this work- Fragment-Weighted Connectivity Index (FWCI), Distance-Weighted Subgraph Index (DWSI), and Symmetry-Based Informational Index (SBI) capture complementary aspects of molecular structure. Accordingly, each index is better suited for predicting certain physicochemical or biological properties depending on structural features such as linearity, spatial arrangement, and symmetry.

To guide researchers in practical applications, the following structure-based decision framework is proposed:

If the molecule is linear or contains distinct functional groups such as -OH, -NH_2, or -COOH at terminal positions, then FWCI is the most appropriate index. FWCI captures fragment-level connectivity and is highly predictive for size-related properties like boiling point, molecular weight, and functional group contribution to polarity or reactivity.

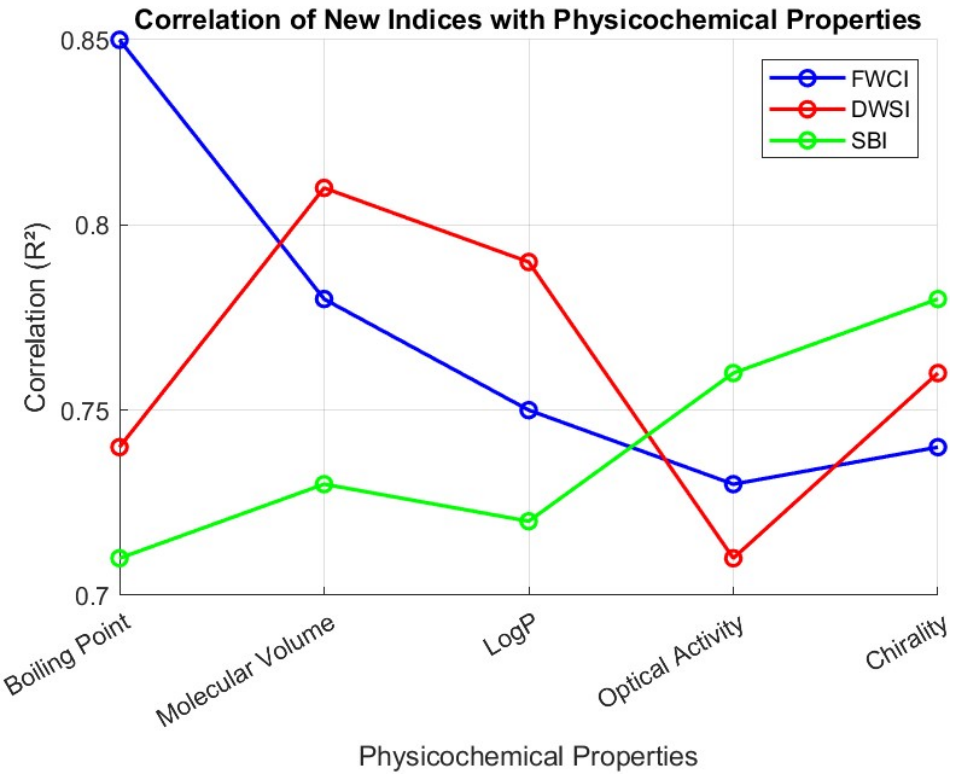

Figure 8: Correlation of new indices with physicochemical properties.

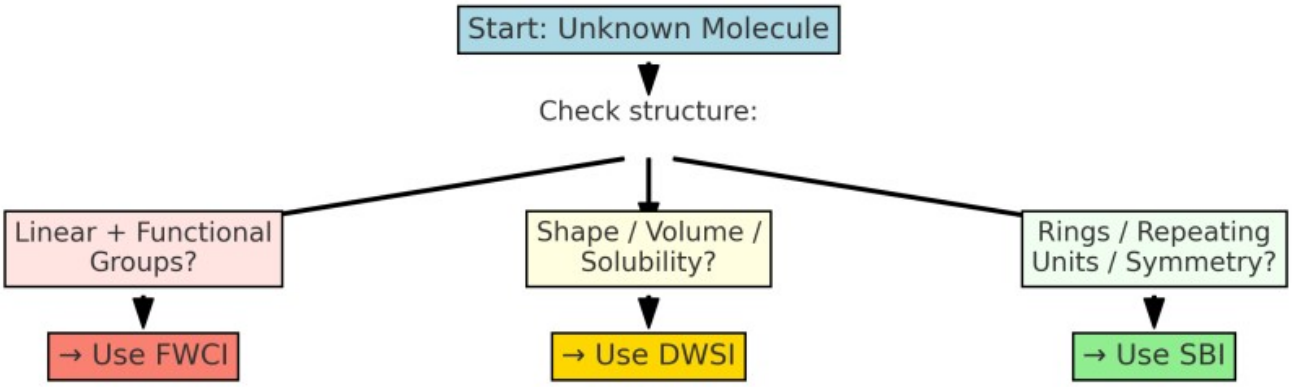

Figure 9: Decision tree to guide the selection of appropriate topological indices (FWCI, DWSI, SBI) based on the structural profile of a new molecule. The flowchart considers the presence of functional groups, molecular shape, and symmetry to assist in predictive QSPR modelling.

Table 11: Comparison of R^2 values with traditional indices.

Property	Wiener (R^2)	Zagreb (R^2)	Randic (R^2)
Boiling point	0.68	0.72	0.64
Molecular Volume	0.69	0.73	0.65
LogP	0.66	0.70	0.62
Optical activity	0.62	0.68	0.60
Chirality	0.61	0.67	0.58

If the molecule has a complex three-dimensional arrangement, such as extended chains, rings with branching, or bulky substituents affecting solubility and molecular volume, then DWSI should be used. DWSI incorporates interatomic distances and is particularly relevant for predicting properties influenced by molecular shape, logP, and molecular volume.

If the molecule contains repeating units, aromaticity, or highly symmetric motifs, such as benzene, cycloalkanes, or substituted symmetrical compounds, then SBI is best suited. SBI quantifies the degree of structural symmetry and is effective for estimating optical activity, chirality, and symmetry-driven behaviours.

This guideline is captured visually in Figure 9, which presents a decision tree for index selection based on key molecular structural features.

Applicability to broader datasets: The present results are obtained on a curated dataset spanning multiple chemical classes (alkanes, alcohols, aromatic and substituted aromatic compounds). While the proposed indices demonstrate consistent trends and competitive predictive performance, broader generalization should be validated on larger and more chemically diverse datasets (e.g., heterocycles, halogenated systems, multifunctional molecules). This extension will be pursued in future work to further evaluate transferability and robustness of FWCI, DWSI and SBI.

CONCLUSION

The three new topological indices presented in this work, the Fragment-Weighted Connectivity Index (FWCI), the Distance-Weighted Subgraph Index (DWSI), and the Symmetry-Based Informational Index (SBI) are intended to capture crucial structural characteristics of chemical compounds, such as symmetry, fragment connectivity, and spatial arrangement. Following extensive mathematical and empirical validation, these indices showed high predictive accuracy, independence from conventional indices, and consistency in modelling important physicochemical properties like polarity, boiling point, molecular volume, and symmetry-driven traits.

The new indices demonstrated robustness through the use of Quantitative Structure-Property Relationship (QSPR) models and cross-validation tests. FWCI was able to accurately predict properties that were influenced by molecular size and functional groups, DWSI was excellent at properties related to shape, and SBI was able to capture symmetry-dependent attributes such as chirality and optical activity. It was verified by comparison with conventional indices (Wiener, Zagreb, and Randic) that FWCI, DWSI, and SBI provide distinct insights, particularly for structurally complicated compounds where conventional indices are inadequate.

Further confirming the novel indices' uniqueness and applicability in chemoinformatics is the minimal correlation

between them and the established indices. The suggested indices are a useful supplement to QSPR modelling, improving the precision of molecular property predictions and providing a more sophisticated comprehension of molecular structure. Promising uses are suggested by their usefulness in cheminformatics and drug discovery, especially for investigating molecular reactivity, stability, and bioactivity in materials science and drug design. Topological indexing has advanced significantly with the creation of FWCI, DWSI, and SBI, opening the door to better predictive models and the potential for customised index design for certain chemical features.

ACKNOWLEDGEMENT

The author gratefully acknowledges the academic environment and support provided by Rajalakshmi Engineering College during the course of this study.

ABBREVIATIONS

FWCI: Fragment-Weighted Connectivity Index; **DWSI:** Distance-Weighted Subgraph Index; **SBI:** Symmetry-Based Informational Index; **QSPR:** Quantitative Structure-Property Relationship; **QSAR:** Quantitative Structure-Activity Relationship; **MS:** Matrix Storage; **MSE:** Mean Squared Error; **R^2:** Coefficient of Determination; **PCA:** Principal Component Analysis; **GNN:** Graph Neural Network; **TI:** Topological Index; **Da:** Dalton; **logP:** Logarithm of octanol-water partition coefficient; CH_3: Methyl Group; **OH:** Hydroxyl Group; **COOH:** Carboxyl Group; NH_2: Amine Group; **DFT:** Density Functional Theory.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORS CONTRIBUTION

V.R. and B.S. conceived the study and developed the theoretical framework. B.S. performed the analytical derivations and computations. S.P. assisted with literature consolidation and interpretation of descriptor/QSPR relevance. V.R. coordinated manuscript preparation. All authors reviewed and approved the final manuscript.

SUMMARY

Three new topological indices are presented in this work: the Fragment-Weighted Connectivity Index (FWCI), the Distance-Weighted Subgraph Index (DWSI), and the Symmetry-Based Informational Index (SBI). These indices are designed to capture important structural characteristics of molecules, including spatial arrangement, molecular symmetry, and functional group connectivity. In order to forecast important physicochemical parameters such as boiling point, molecular volume, molecular weight, and logP using QSPR models and k-fold cross-validation, the indices were applied to a variety of chemical compounds, including alkanes, alcohols, and aromatics. SBI successfully modelled symmetry-related traits, including chirality and optical activity; DWSI demonstrated predictive power for shape-dependent aspects like logP and volume; and FWCI correlated well with fragment-level contributions impacting boiling points.

REFERENCES

1. Balaban AT, Ivanciuc O. Historical development of topological indices. In: Todeschini R, Consonni V, editors. *Topological Indices and Related Descriptors in QSAR and QSPR*. Boca Raton (FL): CRC Press; 2000. p. 31-68.
2. Gutman I. Degree-based topological indices. *Croat Chem Acta*. 2013;86(4):351-61.
3. Hansen PJ, Jurs PC. Chemical applications of graph theory. Part I. Fundamentals and topological indices. *J Chem Educ*. 1988;65(7):574.
4. Alam SM, Jarad F, Mahboob A, Siddique I, Altunok T, Rasheed MW. A survey on generalized topological indices for silicon carbide structure. *J Chem*. 2022; 2022:7311404.
5. Wang XL, Liu JB, Jahanbani A, Siddiqui MK, Rad NJ, Hasni R. On generalized topological indices of silicon-carbon. *J Math*. 2020; 2020:2128594.
6. Matamala AR, Estrada E. Simplex optimization of generalized topological index (GTI-Simplex): a unified approach to optimize QSPR models. *J Phys Chem A*. 2005;109(43):9890-5.
7. Wei CC, Salman M, Ali U, Rehman MU, Khan MA, Chaudary MH, *et al.* Some topological invariants of graphs associated with the group of symmetries. *J Chem*. 2020; 2020:6289518.
8. Ivanciuc O. Chemical graphs, molecular matrices and topological indices in chemoinformatics and quantitative structure-activity relationships. *Curr Comput Aided Drug Des*. 2013;9(2):153-63.
9. Hayat S, Imran M, Liu JB. An efficient computational technique for degree and distance based topological descriptors with applications. *IEEE Access*. 2019;7:32276-96.
10. Sahoo S, Adhikari C, Kuanar M, Mishra BK. A short review of the generation of molecular descriptors and their applications in quantitative structure property/ activity relationships. *Curr Comput Aided Drug Des*. 2016;12(3):181-205.
11. Maran U, Sild S, Tulp I, Takkis K, Moosus M. Molecular descriptors from two-dimensional chemical structure. Cambridge: Royal Society of Chemistry; 2010. p. 148-92.
12. Odugbemi AI, Nyirenda C, Christoffels A, Egieyeh SA. Artificial intelligence in antidiabetic drug discovery: the advances in QSAR and the prediction of α-glucosidase inhibitors. *Comput Struct Biotechnol J*. 2024;23:2964-77.
13. Liu JB, Arockiaraj M, Arulperumjothi M, Prabhu S. Distance based and bond additive topological indices of certain repurposed antiviral drug compounds tested for treating COVID-19. *Int J Quantum Chem*. 2022;121(10):e26617.
14. Prabhu S, Arulperumjothi M, Salu S, Jose BK. Computational analysis of linear chain of holey nanographene and their molecular characterizations. *J Mol Model*. 2025;31:62.
15. Prabhu S, Arulperumjothi M, Tolasa FT, Govardhan S. Distance and degree based topological characterization, spectral and energetic properties, and 13C NMR signals of holey nanographene. *Electrochem Commun*. 2025;173:107881.
16. Skvortsova MI, Palyulin VA, Zefirov NS. Design of topological indices: computer-oriented approach. *SAR QSAR Environ Res*. 2009;20(3-4):357-77.
17. Nikolić S, Trinajstić N. The Wiener index: development and applications. *Croat Chem Acta*. 1995;68(1):105-29.
18. Sheridan PF, Adolf DB, Lyulin AV, Neelov I, Davies GR. Computer simulations of hyperbranched polymers: the influence of the Wiener index on the intrinsic viscosity and radius of gyration. *J Chem Phys*. 2002;117(16):7802-12.
19. Randić M. On molecular identification numbers. *J Chem Inf Comput Sci*. 1984;24(3):164-75.
20. Singh M, Dureja H, Madan AK. Randić index and its variants: applications. In: Thakur TS, editor. *New Frontiers in Nanochemistry: Concepts, Theories, and Trends*. Singapore: Springer; 2020. p. 375-83.
21. Mondal S, Das KC. Zagreb connection indices in structure property modelling. *J Appl Math Comput*. 2023;69(4):3005-20.
22. Yousaf S, Shahzadi K. Utilizing topological indices in QSPR modeling to identify non-cancer medications with potential anti-cancer properties: a promising strategy for drug repurposing. *Front Chem*. 2024;12:1410882.
23. Kar S. Predictive Chemometric Modeling of Toxicological Hazard, and Risk Assessment of Chemicals and Pharmaceuticals. PhD diss., Jadavpur University Kolkata, 2010.

24. Zabidi ZM, Alias AN, Zakaria NA, Mahmud ZS, Ali R, Yaakob MK, *et al.* Machine learning predictor models in the electronic properties of alkanes based on degree-topology indices. *Int J Emerg Technol Adv Eng*. 2021;11(11):1-14.
25. Basak SC, Majumdar S. Current landscape of hierarchical QSAR modeling and its applications: some comments on the importance of mathematical descriptors as well as rigorous statistical methods of model building and validation. In: Hosseini S, Das KC, editors. *Advances in Mathematical Chemistry and Applications*. Sharjah: Bentham Science Publishers; 2015. p. 251-81.
26. Prabhu S, Arulperumjothi M, Manimozhi V, Balasubramanian K. Topological characterizations on hexagonal and rectangular tessellations of antikekulenes and its computed spectral, nuclear magnetic resonance and electron spin resonance characterizations. *Int J Quantum Chem*. 2024;124(7):e27365.
27. Prabhu S, Murugan G, Arockiaraj M, Arulperumjothi M, Manimozhi V. Molecular topological characterization of three classes of polycyclic aromatic hydrocarbons. *J Mol Struct*. 2021;1229:129501.

Cite this article: Rajkumar V, Sivakumar B, Prabhu S. Beyond Traditional Descriptors: Innovative Topological Indices for Enhanced Molecular Predictions. Indian J of Pharmaceutical Education and Research. 2026;60(4s):s1610-s1629.

Supplementary Table S1: Unified notation used for FWCI, DWSI and SBI.

Symbol	Meaning
$G = (V, E)$	Molecular graph with vertex set V and edge set E
$d(u, v)$	Shortest-path distance between vertices u, $v \in V$
$\deg(v)$	Degree of vertex $v \in V$
	i-th fragment in the fragment partition of V
FW()	Fragment weight (aggregate of vertex weights in fragment)
CD()	Connectivity degree of fragment (sum of degrees of vertices in)
	i-th selected subgraph (subset of vertices) used in DWSI
$d(,)$	Distance between subgraphs , (minimum d (u,v) over $u \in$, $v \in$)
$w(,)$	Total weight of edges connecting Si and Sj
x_{ii}	Vertex weight of vertex i (from weight table)
x_{ij}	Edge weight for bond/connection between vertices i and j
	k-th symmetry class (automorphism orbit) in G
	Size of symmetry class
N	Total number of vertices in the symmetry computation graph
$= /N$	Probability of symmetry class

VIF was computed by regressing each descriptor on the remaining descriptors and using VIF = $1/(1 - R^2)$. Higher VIF indicates multicollinearity/redundancy among descriptors.

Supplementary Table S3: Variance Inflation Factor (VIF) for classical and proposed descriptors (n =15).

Descriptor	VIF
Wiener W	126.12
Randić χ	185.49
Zagreb M1	46.40
FWCI	45.36
DWSI	4.89
SBI	46.20

Supplementary Table S2: Atom/Bond Weight Lookup for FWCI Calculation.

Atom / Group	Vertex Weight	Justification
Hydrogen (H)	0.5	Low electronegativity and negligible influence on reactivity
Carbon (C) - aliphatic	1.0	Standard backbone atom
Carbon (C) - aromatic	1.5	Higher delocalization and resonance contribution
Oxygen (O) - hydroxyl	2.0	High polarity and hydrogen bonding potential
Oxygen (O) - carbonyl	2.5	Strong electron-withdrawing group
Nitrogen (N) - amine	2.0	Moderate polarity, contributes to H-bonding
Nitrogen (N) - aromatic	2.5	Enhanced electron-withdrawing ability
Sulfur (S)	2.5	Moderately reactive, polarizable
Fluorine (F)	2.0	Strong electronegativity, lipophilic tuning
Chlorine (Cl), Bromine (Br)	1.5	Halogen bonding, steric and electronic effects
Phosphorus (P)	2.0	Present in organophosphates, high chemical significance
Methyl group ($-CH_3$)	1.0 (per vertex)	Neutral hydrocarbon unit
Hydroxyl group (-OH)	3.0	Sum of O and H contribution
Carboxyl group (-COOH)	5.0	C=O and OH components, strong polarity
Amine group ($-NH_2$)	3.0	N plus H contributions
Benzene ring (C_6H_6)	10.0	6 carbon atoms with aromaticity contributions

VIF was computed by regressing each descriptor on the remaining descriptors and using VIF = $1/(1 - R^2)$. Higher VIF indicates multicollinearity/redundancy among descriptors.