Structure Property Modeling of Physicochemical Properties of Antipsychotic Drugs via Topological Descriptors, Multigraph Modeling with the Specific Treatment of Heteroatoms and Chemometric Methods

Ugasini Preetha P, M Suresh*

***Abstract*—The physicochemical properties for thirteen antipsychotic drugs: Chlorpromazine, Trifluoperazine, Thioridazine, Thiothixene, Haloperidol, Ziprasidone, Loxapine, Quetiapine, Aripprazole, Clozapine, Risperidone, Olanzapine and Sertindole were studied using some distance and degree based topological descriptors. The heterocyclic antipsychotic drugs are modeled as hydrogen suppressed molecular multigraph with the specific treatment for the heteroatoms using valence delta values δ^{v}. In the literature, the molecular connectivity index, originally for hydrocarbons, has been extended to molecules with heteroatoms. The delta value for the heteroatom considers its valence electrons $(\mathrm{Z}^{\mathrm{v}})$ minus the number of bonded hydrogen atoms (h), expressed as $\mathrm{Z}^{\mathrm{v}} - \mathrm{h}$. This modification accommodates heteroatoms in the index calculation on the recently introduced certain set of distance based leap Zagreb indices and degree-based indices. This research uses statistical techniques like quadratic, stepwise regression to link topological descriptors with QSPR models, improving the estimation of drug properties such as boiling point, melting point, enthalpy, flash point, molar refractivity, refractive index, complexity, molecular weight and refractivity. Validation involves comparing estimated values with actual drug properties. This study demonstrates both the similarities and distinctions among the chosen antipsychotic drugs by employing chemometric methods like principal component analysis (PCA) and cluster analysis (CA). These methods help in understanding the shared characteristics and variations present within the drugs being investigated.**

***Index Terms*—Antipsychotic drugs, distance based leap Zagreb indices, quadratic regression, stepwise regression, valence delta values, chemometric methods.**

I. Introduction

Schizophrenia is a complex mental disorder marked by distorted thoughts, perceptions, and emotions, often leading to hallucinations, delusions, and cognitive impairment. Treatment involves a combination of medication, therapy, and support to manage symptoms and improve daily functioning [1]. Boosting drug efficacy while reducing side effects remains pivotal. Parameters like solubility, stability, and toxicity play a vital role in drug design, relying on various chemical and physical drug characteristics. Recently, computational tools, including machine learning, have gained momentum in discovering schizophrenia drugs. These methods predict drug effectiveness and adverse effects, while computer simulations unravel disease mechanisms, aiding in new drug target identification. Lately, computational methods like quantitative structure-property relationship (QSPR) analysis have become influential tools within drug discovery and design. Very recently Das et al. [2] (2023) investigated the topological indices of Molnupiravir and its quantitative structure–activity relationship (QSPR) modeling with other antiviral drugs for COVID-19 treatment. Density Functional Theory(DFT) has been applied in constructing QSAR models for schizophrenia drugs, with utilizing multiple linear regression within the genetic function approximation method for model development [3]. Lately, various QSAR/QSPR models and clinical techniques have been employed to create powerful and effective medications for this condition [4], [5], [6] and [7].

This research article aims to utilize QSPR analysis for predicting both the effectiveness and potential side effects of schizophrenia disorder. Our study aims to offer deeper insights into the molecular mechanisms of antipsychotic drugs by using chemometric methods, thereby aiding the advancement of more potent medications for schizophrenia treatment. In the world of chemical graph theory, we engage in analyzing the structures of chemical graphs representing various chemical systems. From computational data, we extract the chemical properties, by using diverse topological indices associated with parameters such as degree, distance, and eccentricity of graphs, depicting the bonds within the compounds being investigated. Graph vertices symbolize atoms, while edges symbolize the connections or bonds formed between these atoms. Throughout this article we consider chemical structures that are converted into molecular multigraph where the heteroatoms are treated using valence delta values δ^{v}. An alternate approach for computing heteroatoms is rooted in earlier explorations within the literature on unsaturated molecules [8] and [9]. This method employed assessing the valence connectedness of atoms, assigning a δ value to carbon atoms within double bonds. For instance, in propene, this method assigns a δ of 3 to the central carbon atom.

Manuscript received February 14, 2024; revised August 27, 2024.

Ugasini Preetha P is a Research scholar in Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur 603203, Tamil Nadu, India (e-mail: up4755@srmist.edu.in).

M. Suresh is an Assistant Professor in Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur - 603203, Tamilnadu, India. (corresponding author; e-mail: sureshm@srmist.edu.in).

Kier et al. [10] employed a series of valence δ values initially designed for determining the connectivity index, χ, in molecules containing heteroatoms. The enhanced correlations observed between these values and the boiling point, as well as molar refraction across a diverse spectrum of molecules indicate the broad usefulness of these assigned values. We expanded our study to incorporate implications regarding certain distance-based leap Zagreb indices utilizing the equation $\delta^v = Z^v - h$. Specifically, we calculated valence δ values for heteroatoms present in antipsychotic drugs, considering various bonding scenarios. The use of the number of valence electrons together with the number of attached hydrogen atoms provides a strong connection between the structural characteristics expressed in the hydrogen suppressed molecular multigraph and the properties of antipsychotic drugs.

Naji et al. introduced and studied a new set of distance-based topological descriptors termed "leap Zagreb indices" in 2017 [11]. Since their introduction, these indices have attracted considerable attention among researchers, resulting in a rapid surge of related research articles. Recently Alsinai et al. [12] introduced the fourth leap Zagreb index, yielding significant outcomes. Zhu et al. [13] explored the third leap Zagreb index in tree structures, while Raza delved into leap Zagreb connection indices for various network models in [14] and computed them for benzenoid systems in a separate work [15].

The definitions for the first, second, and third leap Zagreb indices are as follows:

$$LM_1(\mathcal{G}) = \sum_{v \in V(\mathcal{G})} d_2(v)^2 \tag{1}$$

$$LM_2(\mathcal{G}) = \sum_{uv \in E(\mathcal{G})} d_2(u) d_2(v) \tag{2}$$

$$LM_3(\mathcal{G}) = \sum_{v \in V(\mathcal{G})} d(v) d_2(v) \tag{3}$$

Here, d(v) signifies the degree of a vertex v in graph $\mathcal{G}$, while $d_2(v)$ denotes the 2-degree of vertex v, representing the count of vertices at a distance of two from v within $\mathcal{G}$.

The leap eccentric connectivity index of G was introduced by Sharma et al. [16] and is defined as:

$$LEC(\mathcal{G}) = \sum_{v \in V(\mathcal{G})} d_2(v) e(v) \tag{4}$$

Let $\tau(v)$ represents the connection number of a vertex v within graph $\mathcal{G}$, indicating the 2-degree of vertex v (i.e., the count of vertices at a distance of two from vertex v).

In [17] authors introduced the reformulated leap Zagreb indices which are a new set of topological indices:

$$RZC_1(\mathcal{G}) = \sum_{e \in E(\mathcal{G})} \tau(e)^2 \tag{5}$$

where e = uv, $\tau(e) = \tau(u) + \tau(v) - 2$

$$RZC_2(\mathcal{G}) = \sum_{e\tilde{f}} \tau(e)\tau(f) \tag{6}$$

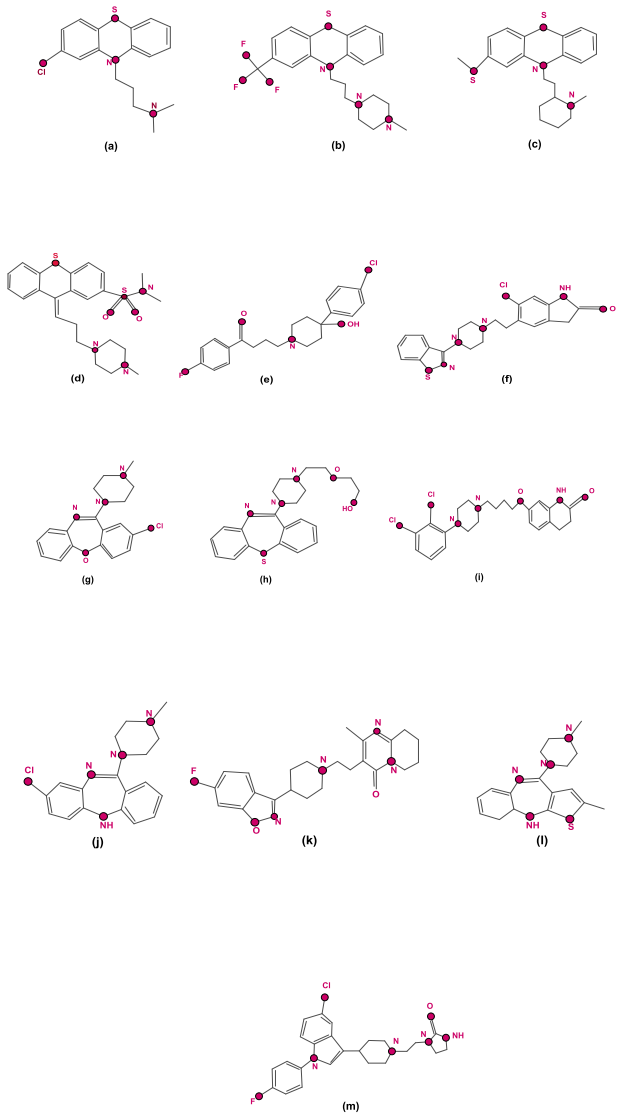

Fig. 1: 2D - Antipsychotic Chemical Structures of (a) Chlorpromazine (b) Trifluoperazine (c) Thioridazine (d) Thiothixene (e) Haloperidol (f) Ziprasidone (g) Loxapine, (h) Quetiapine (i) Aripprazole (j) Clozapine (k) Risperidone (l) Olanzapine and (m) Sertindole

where $e\tilde{f}$ denotes adjacent edges e and f in $\mathcal{G}$.

$$RZC_3(\mathcal{G}) = \sum_{e \in E(\mathcal{G})} deg(e)\tau(e) \tag{7}$$

where e = uv and deg(e) = deg(u)+deg(v)-2.

The calculation method for valence connectivity indices is outlined in reference [10] and [18]:

$${}^m\chi^v = \sum_{i=1}^{N} \prod_{k=1}^{m+1} \left[\frac{1}{\delta_k^v}\right]^{\frac{1}{2}} \tag{8}$$

where $\delta_k^v = \frac{Z_k^v - H_k}{Z_k - Z_k^v - 1}$ is the valence connectivity for the k^{th} atom in the molecular graph, Z_k stands for the

total number of electrons in the k^{th} atom, Z_k^v represents the count of valence electrons in the k^{th} atom, H_k denotes the number of hydrogen atoms directly connected to the k^{th} non-hydrogen atom, and $m = 0$ signifies atomic valence connectivity indices (referred to as order-0).

Milan Randic invented the Randic index [19], which can be represented as

$$R(\mathcal{G}) = \sum_{uv \in E(\mathcal{G})} \frac{1}{\sqrt{d(u)d(v)}} \tag{9}$$

Trinajstić and Gutman [20] introduced a set of fundamental indices known as the first and second Zagreb indices, which are defined as,

$$\begin{aligned} M_1(\mathcal{G}) &= \sum_{uv \in E(\mathcal{G})} [d(u) + d(v)] \\ M_2(\mathcal{G}) &= \sum_{uv \in E(\mathcal{G})} [d(u) \times d(v)] \end{aligned} \tag{10}$$

Zhou and Trinajstić introduced the sum connectivity index, as described in [21], which is defined as,

$$SCI(\mathcal{G}) = \sum_{uv \in E(\mathcal{G})} \frac{1}{\sqrt{d(u) + d(v)}} \tag{11}$$

Vukičević et al. [22] defined the GA index as,

$$GA(\mathcal{G}) = \sum_{uv \in E(\mathcal{G})} \frac{2\sqrt{d(u)d(v)}}{d(u) + d(v)} \tag{12}$$

Fajtlowicz [23] introduced the harmonic index as,

$$H(\mathcal{G}) = \prod_{uv \in E(\mathcal{G})} \frac{2}{[d(u) + d(v)]} \tag{13}$$

Hyper Zagreb index was proposed by Shirdel et al. [24], which is defined as,

$$HM(\mathcal{G}) = \sum_{uv \in E(\mathcal{G})} [d(u) + d(v)]^2 \tag{14}$$

Forgotten topological index was proposed by Furtula et al. [25], which is defined as,

$$F(\mathcal{G}) = \sum_{uv \in E(\mathcal{G})} [d(u)^2 + d(v)^2] \tag{15}$$

In this study, we utilize certain degree and distance-based topological descriptors, as indicated in Equations 1 - 15, for various anti schizophrenic drugs to assess their physicochemical properties. Employing the above topological indices, we construct QSPR models using quadratic and stepwise regression for schizophrenia drugs. Additionally, we employ chemometric techniques like cluster analysis (CA) and principal component analysis (PCA) to compare and analyze the drugs comprehensively, considering theoretical topological descriptors to enhance the model's robustness.

II. METHODOLOGY

A. Specific Treatment of Heteroatoms

Kier-Hall's well-defined valence connectivity (δ_k^v) method [26]and [27] provides information on electric states and atomic orbitals of multiple bonds and heteroatoms in molecular structures, making it a potential contender for this investigation. However one notes that (δ_k^v) values can be directly applied to the recently introduced distance based leap Zagreb indices and as Kier et al. in [10], a collection of valence delta values was constructed to aid in the calculation of connectivity index, χ^v, for heteroatom-containing compounds. Improved correlations with boiling points and molar refraction for several compounds imply their universal application.

As we try to extend for certain antipsychotic drugs with the specific treatment for heteroatoms with the valence delta values (δ_k^v). A strong quadratic correlation is observed between the properties of antipsychotic drugs and structural features in the hydrogen-suppressed molecular multigraph. This correlation is obtained using the number of bonded hydrogen atoms and valence electrons. The quadratic regression model, which treats heteroatoms with valence delta values, provides a slightly better model compared to the linear regression model used by Zhang et al. [5], who employed a molecular simple graph. This molecular connectivity treatment helps in understanding the chemical nature of atoms and their properties. To assess the robustness of the estimators derived from the quadratic regression model, a stepwise regression approach was employed. Interestingly, the stepwise regression analysis yielded the same estimators as the original quadratic regression model, further validating the reliability and consistency of the results. Further CA and PCA helps us to visualize the drug similarity and efficiency.

Hence we see that, in a molecular multigraph heteroatoms are treated specifically with the (δ_k^v) values that is, the vertex degree d(v) of the i^{th} atom in the molecular multigraph is equal to the (δ_k^v) values.

$$d(v) = \delta_k^v$$

Table I shows the valence delta values for sulphur, nitrogen, chlorine, fluorine, oxygen, amino functional group (NH), and hydroxyl functional group (OH) is shown in various bonding circumstances seen in antipsychotic medications. Use of these values give rise to valence connectivity χ or χ^v and distance based leap Zagreb indices. These valence delta values are calculated from Equation 8 represented as (δ_k^v).

TABLE I: Valence Delta Values δ_k^v for Heteroatoms in the Schizophrenia (Antipsychotic) Drugs

Groups	$\delta_{\mathbf{k}}^{\mathbf{v}}$
S	0.667
N	5
Cl	0.778
F	7
O	6
NH	4
OH	5

B. Computation of Certain Degree and Distance Based Leap Zagreb Indices for Chlorpromazine Antipsychotic Drug

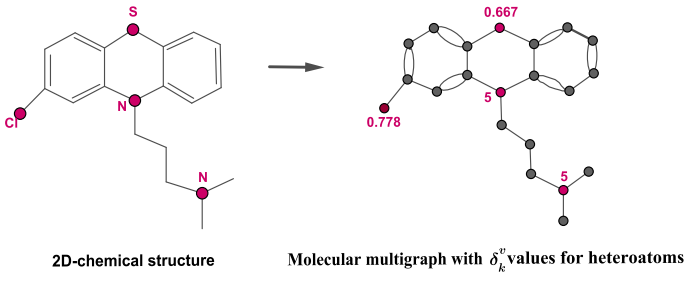

Fig. 2: 2D - Chemical and Molecular Structure of Chlorpromazine

Figure 2 shows the 2D - chemical structure and molecular multigraph with the valence delta values δ_k^v for heteroatoms in Chlorpromazine drug. We now calculate the certain degree based and distance based leap Zagreb indices for the Chlorpromazine drug.

Theorem 1. *Let $\mathcal{G}$ be the molecular multigraph of Chlorpromazine (C), then we have,*

1. $LM_1(\mathcal{G}) = 492.0502$, 2. $LM_2(\mathcal{G}) = 745.116$
3. $LM_3(\mathcal{G}) = 289.112$, 4. $LEC(\mathcal{G}) = 670.338$
5. $RZC_1(\mathcal{G}) = 2072.2791$, 6. $RZC_2(\mathcal{G}) = 3428.5849$
7. $RZC_3(\mathcal{G}) = 1078.2791$, 8. ${}^0\chi^v(\mathcal{G}) = 13.9154$
9. $R(\mathcal{G}) = 10.1075$, 10. $M_1(\mathcal{G}) = 189.112$
11. $M_2(\mathcal{G}) = 295.448$, 12. $SCI(\mathcal{G}) = 11.5283$
13. $GA(\mathcal{G}) = 27.3109$, 14. $H(\mathcal{G}) = 9.2677$
15. $HM(\mathcal{G}) = 1277.3911$, 16. $F(\mathcal{G}) = 686.4951$

Proof: By using definition Equations 1 - 15 and Tables II - VIII we calculate the above mentioned nine degree based indices and seven distance based leap Zagreb indices.

TABLE II: 2-Dist Deg and Eccentricity based Vertex Partition

Drug	0.667	0.778	5	4	6	7	3	2
C	1	1	2	3	6	3	2	3

TABLE III: 2-Dist Deg Edge Partition

Drug	**C**
(0.667,6)	2
(0.778,4)	1
(2,3)	2
(2,5)	2
(3,5)	2
(4,4)	2
(4,6)	4
(4,7)	1
(5,7)	2
(6,6)	5
(6,7)	4
(7,7)	2

TABLE IV: (Deg, 2-Dist Deg) Partition

Drug	**C**
(2,0.667)	1
(1,0.778)	1
(1,2)	2
(2,2)	1
(2,3)	2
(3,4)	2
(3,5)	2
(3,6)	4
(3,7)	1
(4,4)	1
(4,6)	2
(4,7)	2

TABLE V: 2-Dist Deg and Eccentricity based Vertex Partition

Drug	**C**
(0.667,8)	1
(0.778,9)	1
(2,6)	1
(2,9)	2
(3,5)	1
(3,7)	1
(4,8)	2
(4,9)	1
(5,5)	1
(5,8)	1
(6,7)	3
(6,8)	2
(6,9)	1
(7,6)	2
(7,7)	1

TABLE VI: Edge Partition for RZC_1

Drug	**C**
2.778	1
3	2
4.667	2
5	2
6	4
8	4
9	1
10	7
11	4
12	2

TABLE VII: Edge Partition for RZC_2

Edge e = uv with $(\tau(\mathbf{e}), \tau(\mathbf{f}))$	C
(2.778,8)	2
(2.778,9)	1
(3,3)	1
(3,6)	2
(5,5)	1
(5,6)	2
(6,8)	4
(6,10)	2
(8,9)	2
(8,10)	4
(9,12)	2
(10,10)	3
(10,11)	8
(10,12)	2
(11,11)	2
(11,12)	2
(4.667,4.667)	1
(4.667,10)	4
(4.667,11)	2

TABLE VIII: Edge Partition for RZC_3

Drug	C
(2,3)	2
(2,5)	2
(2.778,2.778)	1
(3,6)	2
(4,6)	2
(4,8)	2
(4,10)	1
(5,8)	2
(5,9)	1
(5,10)	5
(5,11)	2
(5,12)	2
(6,11)	2
(7,10)	1
(4.667,4.667)	2

1) $$LM_1(\mathcal{G}) = \sum_{v\in V(\mathcal{G})} d_2(v)^2$$
$$= 1(0.667)^2 + 1(0.778)^2 + 2(5)^2 + 3(4)^2 + 6(6)^2 + 3(7)^2 + 2(3)^2 + 3(2)^2 = 492.0502$$

2) $$LM_2(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})} d_2(u)d_2(v)$$
$$= 2(0.667\times 6) + 1(0.778\times 4) + 2(2\times 3) + 2(2\times 5) + 2(3\times 5) + 2(4\times 4) + 4(4\times 6) + 1(4\times 7) + 2(5\times 7) + 5(6\times 6) + 4(6\times 7) + 2(7\times 7) = 745.116$$

3) $$LM_3(\mathcal{G}) = \sum_{v\in V(\mathcal{G})} d(v)d_2(v)$$
$$= 1(2\times 0.667) + 1(1\times 0.778) + 2(1\times 2) + 1(2\times 2) + 2(2\times 3) + 2(3\times 4) + 1(4\times 4) + 2(3\times 5) + 4(3\times 6) + 2(4\times 6) + 1(3\times 7) + 2(4\times 7) = 289.112$$

4) $$LEC(\mathcal{G}) = \sum_{v\in V(\mathcal{G})} d_2(v)e(v)$$
$$= 1(0.667\times 8) + 1(0.778\times 9) + 2(2\times 9) + 1(2\times 6) + 1(3\times 7) + 1(3\times 5) + 2(4\times 8) + 1(4\times 9) + 1(6\times 9) + 2(6\times 8) + 3(6\times 7) + 1(7\times 7) + (7\times 6) + 1(5\times 5) + 1(5\times 8) = 670.338$$

5) $$RZC_1(\mathcal{G}) = \sum_{e\in E(\mathcal{G})} \tau(e)^2$$
$$= 1(2.778)^2 + 2(3)^2 + 2(4.667)^2 + 2(5)^2 + 4(6)^2 + 4(8)^2 + 1(9)^2 + 7(10)^2 + 4(11)^2 + 2(12)^2 = 2072.2791$$

6) $$RZC_2(\mathcal{G}) = \sum_{e\tilde{f}} \tau(e)\tau(f)$$
$$= 2(2.778\times 8) + 1(2.778\times 9) + 1(3\times 3) + 2(3\times 6) + 1(5\times 5) + 2(5\times 6) + 4(6\times 8) + 2(6\times 10) + 2(8\times 9) + 4(8\times 10) + 2(9\times 12) + 3(10\times 10) + 8(10\times 11) + 2(11\times 11) + 2(11\times 12) + 2(10\times 12) + 2(4.667\times 11) + 1(4.667\times 4.667) + 4(4.667\times 10) = 3428.5849$$

7) $$RZC_3(\mathcal{G}) = \sum_{e\in E(\mathcal{G})} deg(e)\tau(e)$$
$$= 2(2\times 3) + 2(2\times 5) + 1(2.778\times 2.778) + 2(3\times 6) + 2(4\times 6) + 2(4\times 8) + 1(4\times 10) + 2(5\times 8) + 1(5\times 9) + 5(5\times 10) + 2(5\times 11) + 2(5\times 12) + 2(6\times 11) + 1(7\times 10) + 2(4.667\times 4.667) = 1078.2791$$

8) $${}^{m}\chi^{v} = \sum_{i=1}^{N}\prod_{k=1}^{m+1}\left[\frac{1}{\delta_k^v}\right]^{\frac{1}{2}}$$
$$= 1\left(\frac{1}{0.778}\right)^{0.5} + 1\left(\frac{1}{0.667}\right)^{0.5} + 2\left(\frac{1}{1}\right)^{0.5} + 3\left(\frac{1}{2}\right)^{0.5} + 7\left(\frac{1}{3}\right)^{0.5} + 5\left(\frac{1}{4}\right)^{0.5} + 2\left(\frac{1}{5}\right)^{0.5}$$
$$= 13.9154$$

9) $$R(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})}\frac{1}{\sqrt{d(u)d(v)}}$$
$$= 2(1\times 5)^{-0.5} + 1(0.778\times 4)^{-0.5} + 2(0.667\times 4)^{-0.5} + 2(2\times 2)^{-0.5} + 5(3\times 3)^{-0.5} + 11(3\times 4)^{-0.5} + 2(4\times 4)^{-0.5} + 2(4\times 5)^{-0.5}$$
$$= 10.1075$$

10) $$M_1(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})}[d(u)+d(v)]$$
$$= 2(1+5) + 1(0.778+4) + 2(0.667+4) + 2(2+2) + 2(2+5) + 5(3+3) + 11(3+4) + 2(4+4) + 2(4+5)$$
$$= 189.112$$

11) $$M_2(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})}[d(u)\times d(v)]$$
$$= 2(1\times 5) + 1(0.778\times 4) + 2(0.667\times 4) + 2(2\times 2) + 2(2\times 5) + 5(3\times 3) + 11(3\times 4) + 2(4\times 4) + 2(4\times 5)$$
$$= 295.448$$

12) $$SCI(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})}\frac{1}{\sqrt{d(u)+d(v)}}$$
$$= 2\left(\frac{1}{\sqrt{1+5}}\right) + 1\left(\frac{1}{\sqrt{0.778+4}}\right) + 2\left(\frac{1}{\sqrt{0.667+4}}\right) + 2\left(\frac{1}{\sqrt{2+2}}\right) + 2\left(\frac{1}{\sqrt{2+5}}\right) + 2\left(\frac{1}{\sqrt{3+3}}\right) + 11\left(\frac{1}{\sqrt{3+4}}\right) + 2\left(\frac{1}{\sqrt{4+4}}\right) + 2\left(\frac{1}{\sqrt{4+5}}\right)$$
$$= 11.5283$$

13) $$GA(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})}\frac{2\sqrt{d(u)d(v)}}{d(u)+d(v)}$$
$$= 2\left(\frac{2\sqrt{1\times 5}}{1+5}\right) + 1\left(\frac{2\sqrt{0.778\times 4}}{0.778+4}\right) + 2\left(\frac{2\sqrt{0.667\times 4}}{0.667+4}\right) + 2\left(\frac{2\sqrt{2\times 2}}{2+2}\right) + 2\left(\frac{2\sqrt{2\times 5}}{2+5}\right) + 2\left(\frac{2\sqrt{3\times 3}}{3+3}\right) + 11\left(\frac{2\sqrt{3\times 4}}{3+4}\right) + 2\left(\frac{2\sqrt{4\times 4}}{4+4}\right) + 2\left(\frac{2\sqrt{4\times 5}}{4+5}\right)$$
$$= 27.3109$$

14) $$H(\mathcal{G}) = \prod_{uv\in E(\mathcal{G})}\frac{2}{[d(u)+d(v)]}$$
$$= 2\left(\frac{2}{1+5}\right) + 1\left(\frac{2}{0.778+4}\right) + 2\left(\frac{2}{0.667+4}\right) + 2\left(\frac{2}{2+2}\right) + 2\left(\frac{2}{2+5}\right) + 2\left(\frac{2}{3+3}\right) + 11\left(\frac{2}{3+4}\right) + 2\left(\frac{2}{4+4}\right) + 2\left(\frac{2}{4+5}\right)$$
$$= 9.2677$$

15) $$HM(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})}[d(u)+d(v)]^2$$
$$= 2(1+5)^2 + 1(0.778+4)^2 + 2(0.667+4)^2 + 2(2+2)^2 + 2(2+5)^2 + 5(3+3)^2 + 11(3+4)^2 + 2(4+4)^2 + 2(4+5)^2$$
$$= 1277.3911$$

16) $$F(\mathcal{G}) = \sum_{uv\in E(\mathcal{G})}[d(u)^2+d(v)^2$$
$$= 2(1^2+5^2) + 1(0.778^2+4^2) + 2(0.667^2+4^2) + 2(2^2+2^2) + 2(2^2+5^2) + 5(3^2+3^2) + 11(3^2+4^2) + 2(4^2+4^2) + 2(4^2+5^2)$$
$$= 686.4951$$

■

We show the numerical values of certain chosen distance based leap Zagreb indices and degree based indices for the other drugs are computed and presented in the Table IX.

III. Results and discussions

A. Quadratic Regression Analysis

This segment's main goal is to improve the quantitative structure-property relationship (QSPR) between the

TABLE IX: Distance and Degree based Indices for Schizophrenia (Antipsychotic) Drugs

Drugs	LM_1	LM_2	LM_3	LEC	RZC_1	RZC_2	RZC_3	${}^0\chi^v$	R	M_1	M_2	SCI	GA	H	HM	F
Chlor	492.050	745.116	289.112	670.338	2072.279	3428.585	1078.279	13.915	10.107	189.112	295.448	11.528	27.311	9.268	1277.391	686.495
Triflu	717.445	911.004	364.334	1394.337	2532.562	4462.135	1512.894	16.769	12.398	251.334	440.669	14.577	35.409	11.520	1888.228	1006.890
Thior	535.890	812.006	320.668	875.673	2225.120	3718.032	1204.785	16.705	12.293	210.668	325.671	14.148	31.035	12.194	1389.122	737.780
Thioth	808.890	1123.350	396.336	1447.341	3323.915	5882.764	1621.912	19.468	15.634	269.336	375.347	16.789	37.745	13.548	1774.253	1023.559
Halo	562.605	818.112	334.778	1575.226	2157.717	3391.894	1303.717	15.383	11.361	235.778	399.112	13.776	34.070	11.008	1675.829	877.605
Zipra	737.050	1077.005	411.112	1686.452	3056.057	5160.472	1786.894	17.022	12.918	277.112	471.115	15.475	38.267	12.136	2002.725	1060.495
Loxa	606.605	927.112	346.778	837.002	2623.717	4865.450	1508.717	13.753	10.174	233.778	409.112	12.603	31.920	9.743	1721.829	903.605
Queti	632.445	908.004	359.334	1410.671	2517.562	4407.802	1432.894	15.697	12.906	232.668	365.338	14.519	34.010	11.932	1590.456	859.780
Aripi	641.211	892.780	371.556	1874.230	2398.547	4114.568	1466.991	18.514	13.267	269.556	452.224	15.791	38.577	12.695	1921.659	1017.211
Cloza	582.605	919.112	342.778	824.780	2535.717	4435.894	1442.717	13.845	10.266	229.778	393.112	12.678	31.960	9.843	1649.829	863.605
Risp	785.000	1129.000	416.000	1926.000	3122.000	5271.000	1907.000	17.364	12.730	296.000	551.000	15.755	40.031	12.398	2296.000	1194.000
Olanza	553.445	845.003	316.334	758.336	2390.560	4074.462	1260.893	13.781	9.710	213.334	358.336	12.027	29.563	9.455	1527.562	810.890
Serti	768.605	1095.112	424.778	1771.114	2993.717	4935.450	1752.717	18.116	13.552	297.778	517.112	16.684	41.886	13.146	2167.829	1133.605

TABLE X: Schizophrenia Drugs' Physicochemical Properties

Drugs	BP (°C)	MP (°C)	E (KJ/mol)	FP (°C)	MR (cm^3)	C	MW (g/mol)	R (cm^3)
Chlorpromazine	450.1	60	70.9	226	92.8	339	355.33	93.76
Trifluoperazine	506	242	77.6	259.8	108.2	510	480.4	110.98
Thioridazine	515.665	73	78.8	265.7	112.8	432	370.6	113.52
Thiothixene	599	114	89.2	316.1	126.5	711	443.62	137.85
Haloperidol	529	151.5	84.6	273.8	101	451	375.9	102.59
Ziprasidone	554.8	213	83.6	289.3	114.1	573	412.936	116.72
Loxapine	458.6	109	71.9	231.1	92.1	450	327.81	95.11
Quetiapine	556.5	172	88.2	290.4	110.2	496	383.51	114.09
Aripiprazole	646.2	139	95.3	344.6	120.3	559	448.4	124.34
Clozapine	489.2	183	75.5	249.6	93.7	446	326.8	97.36
Risperidone	572.4	170	85.8	300	111.7	731	410.5	111.7
Olanzapine	476	195	74	241.7	92.2	432	312.432	107.17
Sertindole	592.1	95	88.3	311.9	120.7	623	440.941	131.77

TABLE XI: The Values of Correlation Coefficient (R) acquired by Quadratic Regression Model between Topological Indices and Physicochemical Properties of Different Drugs Employed in the Remedy of Antipsychotic Drugs

TI	BP	MP	E	FP	MR	C	MW	R
LM_1	0.67	**0.63**	0.614	0.67	0.733	**0.944**	0.73	0.732
LM_2	0.584	0.59	0.507	0.584	0.63	0.926	0.494	0.644
LM_3	0.742	0.526	0.697	0.741	0.73	0.908	0.662	0.699
LEC	0.863	0.393	**0.871**	0.863	0.773	0.781	0.796	0.677
RZC_1	0.515	0.554	0.429	0.515	0.622	0.896	0.456	0.653
RZC_2	0.433	0.501	0.35	0.433	0.551	0.838	0.399	0.593
RZC_3	0.589	0.479	0.538	0.588	0.555	0.874	0.53	0.539
${}^0\chi^v$	0.892	0.333	0.833	0.892	0.982	0.815	**0.884**	0.931
R	0.867	0.247	0.848	0.867	0.979	0.803	0.854	0.923
M_1	0.774	0.577	0.72	0.774	0.714	0.908	0.69	0.68
M_2	0.561	0.603	0.512	0.561	0.433	0.737	0.528	0.385
SCI	**0.904**	0.477	0.863	**0.904**	0.958	0.895	0.826	0.928
GA	0.849	0.554	0.811	0.849	0.8	0.886	0.758	0.75
H	0.895	0.519	0.862	0.895	**0.99**	0.822	0.804	**0.957**
HM	0.64	0.571	0.587	0.639	0.539	0.83	0.612	0.512
F	0.687	0.568	0.629	0.687	0.622	0.901	0.67	0.597

topological indices and various physicochemical properties of antipsychotic (schizophrenic) drugs. These include first-generation drugs like chlorpromazine ($DB00477$), trifluoperazine ($DB00831$), thioridazine ($DB00679$), thiothixene ($DB01623$) and haloperidol ($DB00502$), as well as atypical (newer) drugs like ziprasidone ($DB00246$), loxapine ($DB00408$), quetiapine ($DB01224$), aripiprazole ($DB01238$), clozapine ($DB00363$), risperidone ($DB00734$), olanzapine ($DB00334$), and sertindole ($DB06144$) used in the treatment of schizophrenia patients, this study incorporates various topological indices to evaluate the efficacy of oral antipsychotic drugs. Our approach involves utilizing nine degree-based and seven distance-based leap Zagreb topological indices to model eight physicochemical characteristics of the drugs under consideration. These include their boiling point (BP), melting point (MP), enthalpy (E), flash point (FP), molar refractivity (MR), complexity (C), molecular weight (MW), and refractivity (R). The physicochemical property values of these medications are compiled from the ChemSpider chemical database and are presented in Table X.

Next, we will construct quadratic regression models to

TABLE XII: Optimistic Estimates of Physicochemical Properties Derived from Quadratic Regression Models

Models	TI	No. Eq	R	R^2	F	S.E	Adjusted - R^2
BP = $100.002 + 29.904(SCI) + 0.027(SCI)^2$	SCI	(A1)	0.904	0.818	22.405	27.630	0.781
MP = -1503.337+4.988(LM_1)-0.004$(LM_1)^2$	LM_1	(A2)	0.63	0.397	3.291	46.916	0.276
E = 55.087+0.028(LEC)-(5.490e-6)$(LEC)^2$	LEC	(A3)	0.871	0.759	15.717	4.085	0.71
FP = $13.937 + 18.128(SCI) + 0.015(SCI)^2$	SCI	(A4)	0.904	0.818	22.399	16.717	0.781
MR = $99.476 - 7.216(H) + 0.680(H)^2$	H	(A5)	0.990	0.981	252.899	1.824	0.977
C = 347.440-0.521(LM_1)+0.001$(LM_1)^2$	LM_1	(A6)	0.944	0.891	40.692	41.843	0.869
MW = -760.905+119.951$({}^0\chi^v)$-2.972$({}^0\chi^v)^2$	${}^0\chi^v$	(A7)	0.884	0.782	17.961	27.077	0.739
R = $328.700 - 48.097(H) + 2.511(H)^2$	H	(A8)	0.957	0.916	54.659	4.280	0.899

TABLE XIII: Comparison of the Results Obtained from Quadratic and Linear Regression Models

Physicochemical Properties	Our obtained results	Our quadratic model	Results obtained by Zhang [4]	Their linear regression model
	R values	Most relevant indices	R values	Most relevant indices
BP	0.904	SCI	0.915	H
MP	**0.63**	LM_1	**0.175**	M_2
E	0.871	LEC	0.885	H
FP	0.904	SCI	0.915	H
MR	**0.990**	H	**0.902**	RA
C	**0.944**	LM_1	**0.944**	M_2
MW	**0.884**	${}^0\chi^v$	**0.85**	ABC
R	**0.957**	H	**0.91**	M_2

analyze the relationship between the topological indices of these drugs and their physicochemical properties. Our choice of quadratic regression is based on its superior performance compared to linear and cubic regression models. Before proceeding, let's provide a brief overview of the quadratic regression model and highlight some statistical measures that will be computed. We will use the following equation in our regression analysis:

$$Y = c + u_1X + u_2X^2; N, R \qquad \text{(Quadratic equation)}$$

In the provided equation, where the dependent or response variable is denoted by Y, and the independent or predictor variable is denoted by X. Regression model constant and individual topological index coefficients, respectively, are denoted by the coefficients c and u_i (where i = 1, 2). Here, N corresponds to the total sample size and R denotes the correlation coefficient. Various statistical measures gauge the effectiveness of a regression model, including R^2, F-statistic (F), standard error (S.E.), and adjusted R^2. These metrics collectively provide insights into the model's goodness of fit and predictive accuracy.

Among the obtained values of R and R squared for each physicochemical property, the model with the highest R is considered the most optimistic regression model for that property. Consequently, the maximum R for each physicochemical property is emphasized in bold in Table XI. The Table XII presents the quadratic regression equations for each physicochemical property associated with the topological index having the highest R value. Additionally, the table includes the corresponding values of F, S.E, and adjusted - R^2.

Based on the quadratic regression models presented in Table XII for specific degree-based and distance-based topological indices, reveals noteworthy insights for these indices, as illustrated by equations **(A1) - (A8)** demonstrate a strong correlation with the physicochemical properties of drugs utilized in the treatment of schizophrenia.

We now present a comparison of the R values derived from our quadratic regression models for various physicochemical properties with those calculated in a prior study by Zhang et al. [5]. In their study, Zhang et al. utilized linear regression models to obtain R values for the same physicochemical properties. The comparative analysis is summarized and listed in Table XIII where the bolded values shows that our multigraph model overshadows the simple graph model. A set of valence delta values was devised to facilitate the calculation of certain degree and distance-based indices tailored for heteroatom-containing antipsychotic drugs. In our demonstrated quadratic regression models, the outcomes underscore notable enhancements in correlations achieved through our multigraph model, particularly in addressing the specific treatment for heteroatoms. For all the considered physicochemical properties of these drugs (except BP, E, and FP with minor variations), our results overshadow those reported by Zhang et al. in the article [5].

We assert that combining the number of valence electrons with the count of attached hydrogen atoms establishes a robust correlation between the structural features depicted in

the hydrogen-suppressed graph and the properties of drugs. Moreover, all the aforementioned implications are visually illustrated and depicted in Figure 3.

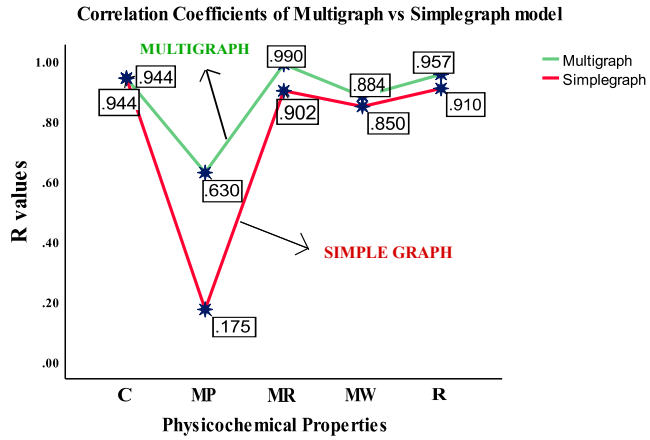

Fig. 3: Multigraph vs Simplegraph of R Values

The F-value is a statistical indicator employed in analysis of variance (ANOVA) to determine the significance of the overall model fit. Generally, an F-value greater than 1 is considered significant. A higher F-value implies a stronger fit of the regression model to the data, indicating that the independent variables collectively exert a substantial influence on the dependent variable. Consequently all our multigraph models exhibit $F > 1$. A higher value of R^2, closer to 1, along with a higher Adjusted-R^2 and a lower standard error (S.E.), indicate the effectiveness of a regression model in explaining the variability in the data. For a comprehensive understanding, refer articles [28] amd [29].

B. Stepwise Regression Analysis

Several linear QSPR models involving 16 topological descriptors were established and the strongest multivariable correlations were identified by the stepwise regression implemented in IBM SPSS Statistics 25 used to develop the linear model for the prediction of boiling point, melting point, enthalpy, flash point, molar refractivity, complexity, molecular weight and refractivity. Stepwise selection in regression is a method that combines aspects of forward selection and backward elimination to iteratively build or refine a regression model. This approach is aimed at identifying a subset of predictor variables that collectively provide the best fit for predicting the outcome variable.

QSPR models for boiling point (BP)

The best linear model for boiling point contains one topological descriptor, namely sum connectivity index (SCI). The model is presented below:

$$BP = 94.635 + 30.670(\pm 4.368)SCI \qquad \text{(Model 1)}$$

N = 13, R = 0.904, R^2 = 0.818, R^2_{adj} = 0.801, S.E = 26.3441, F = 49.291, sig = 0.000, Tolerance = 1, VIF = 1 and DW = 1.604.

Here $\pm$ 4.368 n the regression model equation indicates the standard error of the coefficient. This indicates the range within which the true value of the coefficient is expected to lie, given the statistical confidence level (usually 95 %).

QSPR models for melting point (MP)

While performing stepwise regression for melting point, no variables were entered into the equation since none of the independent variables met the criteria for entry into the regression model. This can happen for several reasons significance level might not reach the standard sig value set at 0.05, multicollinearity or the independent variable might not have significant predictive power to explain the variance in the dependent variable.

QSPR models for enthalpy (E)

The best linear model for enthalpy contains two topological descriptors, namely, leap eccentric connectivity (LEC) and second Zagreb index (M_2).

$$E = 63.021 + 0.014(\pm 0.003)LEC \qquad \text{(Model 1)}$$

N = 13, R = 0.865, R^2 = 0.749, R^2_{adj} = 0.726, S.E = 3.9711, F = 32.835, sig = 0.000, Tolerance = 1 and VIF = 1

$$E = 76.948 + 0.022(\pm 0.003)LEC - 0.057(\pm 0.020)M_2 \qquad \text{(Model 2)}$$

N = 13, R = 0.927, R^2 = 0.859, R^2_{adj} = 0.831, S.E = 3.1224, F = 30.451, sig = 0.000, Tolerance = 0.366, VIF = 2.734 and DW = 1.192.

QSPR models for flash point (FP)

The best linear model for flash point contains one topological descriptor, namely sum connectivity index (SCI). The model is presented below:

$$FP = 10.953 + 18.554(\pm 2.643)SCI \qquad \text{(Model 1)}$$

N = 13, R = 0.904, R^2 = 0.818, R^2_{adj} = 0.801, S.E = 15.9395, F = 49.278, sig = 0.000, Tolerance = 1, VIF = 1 and DW = 1.607.

QSPR models for molar refractivity (MR)

The best linear model for molar refractivity contains one topological descriptors, namely, harmonic (H).

$$MR = 14.632 + 8.101(\pm 0.414)H \qquad \text{(Model 1)}$$

N = 13, R = 0.986, R^2 = 0.972, R^2_{adj} = 0.970, S.E = 2.0867, F = 383.075, sig = 0.000, Tolerance = 1, VIF = 1 and DW = 2.481.

QSPR models for complexity (C)

The best linear model for complexity contains one topological descriptor, namely first leap Zagreb index (LM_1). The model is presented below:

$$C = -152.614 + 1.037(\pm 0.114)LM_1 \qquad \text{(Model 1)}$$

N = 13, R = 0.940, R^2 = 0.883, R^2_{adj} = 0.873, S.E = 41.2304, F = 83.120, sig = 0.000, Tolerance = 1, VIF =1 and DW = 1.920.

QSPR models for molecular weight (MW)

The best linear model for molecular weight contains one topological descriptor, namely valence connectivity index of order 0 ($^0\chi^v$). The model is presented below:

$$MW = 11.887 + 23.461(\pm 4.091)^0\chi^v \qquad \text{(Model 1)}$$

N = 13, R = 0.866, R^2 = 0.749, R^2_{adj} = 0.729, S.E = 27.6981, F = 32.886, sig = 0.000, Tolerance = 1, VIF =1 and DW = 2.038.

QSPR models for refractivity (R)

The best linear model for refractivity contains one topological descriptor, namely valence connectivity index of order 0 ($^0\chi^v$). The model is presented below:

$$MW = 9.549 + 6.337(\pm 0.827)^0\chi^v \qquad \text{(Model 1)}$$

N = 13, R = 0.918, R^2 = 0.842, R^2_{adj} = 0.828, S.E = 5.6002, F = 58.686, sig = 0.000, Tolerance = 1, VIF =1 and DW = 1.941.

DISCUSSION:

We studied the relationship between the topological indices and the physicochemical properties of 13 antispychotic drugs. In this section, to find the best model to predict the parameters mentioned, we use statistical parameters like VIF, tolerance, DW and residual to measure the significant estimators.

Multicollinearity is a statistical phenomenon that occurs when two or more independent variables in a regression model are highly correlated with each other. This can lead to unstable and unreliable estimates of the regression coefficients.

The VIF measures the degree to which the variance of a regression coefficient is inflated due to multicollinearity. A VIF value greater than 10 is generally considered to indicate the presence of multicollinearity. Conversely, a VIF value less than 1 may suggest that the variable is a linear combination of other independent variables and should be removed from the model. In all our final models the VIF lies within the acceptable range this indicates no multicollinearity issue.

The success of QSPR models relies on the accuracy of input data, the selection of appropriate molecular descriptors, and the use of robust statistical tools. All our above models produced a squared correlation coefficient close to 1, and the results of other statistical parameters are also very satisfactory.

For verification and validity of the regression models, we will focus on Durbin-Watson statistics and unstandardized predicted and residual values. The Durbin-Watson (DW) statistic helps evaluate the goodness of fit of a regression model. A DW value close to 2 suggests that the model effectively captures the relationship between the variables and that the residuals are independent. This implies that the model's predictions are reliable and not influenced by temporal dependencies in the data. DW statistic for all the models except for molar refractivity and molecular weight the DW values close to 2 which indicates that the model effectively captures the relationship between the variables.

The residual is the difference between the observed and predicted values. Comparison between predicted and observed values of all the above models of the antispychotic drugs is shown in Tables XIV to XX . Figures 4 - 10 show the linear correlation between the observed and the predicted residuals of the above optimal models for each physicochemical property.

NOTE: We observe that the results obtained from quadratic regression, the optimal estimator for the physicochemical properties of boiling point, flash point is the sum connectivity index (SCI) which has the same estimator while performing stepwise regression.

Similarly for the other physicochemical properties enthalpy, molecular refractivity, complexity and molecular weight except refractivity. This suggests that these predictors indicates that these predictors are considered significant in capturing the variation in the dependent variable, accounting for both linear and quadratic effects.

TABLE XIV: Comparison between Predicted and Observed Values of Boiling Point

Drugs	BP (OBS)	UB (PRED BP)	LB (PRED BP)	RES
Chlor	450.1	452.57	443.83	-2.47
Triflu	506	546.08	537.34	-40.08
Thior	515.665	532.92	524.19	-17.26
Thioth	599	613.92	605.19	-14.92
Halo	529	521.51	512.78	7.49
Zipra	554.8	573.62	564.89	-18.82
Loxa	458.6	485.53	476.80	-26.93
Queti	556.5	544.31	535.57	12.19
Aripi	646.2	583.31	574.58	62.89
Cloza	489.2	487.82	479.09	1.38
Risper	572.4	582.20	573.47	-9.80
Olanza	476	467.86	459.12	8.14
Serti	592.1	610.72	601.98	-18.62

TABLE XV: Comparison between Predicted and Observed Values of Enthalpy

Drugs	E (OBS)	UB (PRED E)	LB (PRED E)	RES
Chlor	70.9	74.88	74.83	-3.98
Triflu	77.6	82.53	82.48	-4.93
Thior	78.8	77.67	77.63	1.13
Thioth	89.2	87.42	87.37	1.78
Halo	84.6	88.88	88.83	-4.28
Zipra	83.6	87.22	87.17	-3.62
Loxa	71.9	72.07	72.02	-0.17
Queti	88.2	87.18	87.14	1.02
Aripi	95.3	92.43	92.38	2.87
Cloza	75.5	72.71	72.66	2.79
Risper	85.8	87.94	87.89	-2.14
Olanza	74	73.23	73.18	0.77
Serti	88.3	86.46	86.41	1.84

TABLE XVI: Comparison between Predicted and Observed Values of Flash Point

Drugs	FP (OBS)	UB (PRED FP)	LB (PRED FP)	RES
Chlor	226	227.49	222.20	-1.49
Triflu	259.8	284.06	278.77	-24.26
Thior	265.7	276.10	270.81	-10.40
Thioth	316.1	325.10	319.81	-9.00
Halo	273.8	269.20	263.91	4.60
Zipra	289.3	300.72	295.43	-11.42
Loxa	231.1	247.43	242.14	-16.33
Queti	290.4	282.99	277.70	7.41
Aripi	344.6	306.58	301.30	38.02
Cloza	249.6	248.82	243.53	0.78
Risper	300	305.91	300.62	-5.91
Olanza	241.7	236.74	231.45	4.96
Serti	311.9	323.16	317.87	-11.26

TABLE XVII: Comparison between Predicted and Observed Values of Molar Refractivity

Drugs	MR (OBS)	UB (PRED MR)	LB (PRED MR)	RES
Chlor	92.8	90.13	89.30	2.67
Triflu	108.2	108.37	107.54	-0.17
Thior	112.8	113.83	113.00	-1.03
Thioth	126.5	124.80	123.97	1.70
Halo	101	104.22	103.39	-3.22
Zipra	114.1	113.36	112.53	0.74
Loxa	92.1	93.98	93.15	-1.88
Queti	110.2	111.71	110.88	-1.51
Aripi	120.3	117.89	117.06	2.41
Cloza	93.7	94.79	93.96	-1.09
Risper	111.7	115.48	114.65	-3.78
Olanza	92.2	91.64	90.82	0.56
Serti	120.7	121.54	120.71	-0.84

TABLE XVIII: Comparison between Predicted and Observed Values of Complexity

Drugs	C (OBS)	UB (PRED C)	LB (PRED C)	RES
Chlor	339	357.76	357.53	-18.53
Triflu	510	591.49	591.26	-81.26
Thior	432	403.22	402.99	29.01
Thioth	711	686.32	686.09	24.91
Halo	451	430.92	430.69	20.31
Zipra	573	611.82	611.59	-38.59
Loxa	450	476.55	476.32	-26.32
Queti	496	503.35	503.12	-7.12
Aripi	559	512.44	512.21	46.79
Cloza	446	451.66	451.43	-5.43
Risper	731	661.55	661.32	69.68
Olanza	432	421.42	421.19	10.81
Serti	623	644.54	644.32	-21.32

TABLE XIX: Comparison between Predicted and Observed Values of Molecular Weight

Drugs	MW (OBS)	UB (PRED MW)	LB (PRED MW)	RES
Chlor	355.33	342.44	334.26	12.89
Triflu	480.4	409.40	401.21	71.00
Thior	370.6	407.89	399.71	-37.29
Thioth	443.62	472.72	464.53	-29.10
Halo	375.9	376.88	368.70	-0.98
Zipra	412.936	415.33	407.15	-2.40
Loxa	327.81	338.64	330.46	-10.83
Queti	383.51	384.25	376.07	-0.74
Aripi	448.4	450.33	442.15	-1.93
Cloza	326.8	340.80	332.62	-14.00
Risper	410.5	423.35	415.17	-12.85
Olanza	312.432	339.30	331.12	-26.87
Serti	440.941	441.00	432.81	-0.05

TABLE XX: Comparison between Predicted and Observed Values of Refractivity

Drugs	R (OBS)	UB (PRED R)	LB (PRED R)	RES
Chlor	93.76	98.56	96.90	-3.14
Triflu	110.98	116.64	114.99	-4.01
Thior	113.52	116.24	114.58	-1.06
Thioth	137.85	133.74	132.09	5.76
Halo	102.59	107.86	106.20	-3.61
Zipra	116.72	118.24	116.59	0.13
Loxa	95.11	97.53	95.87	-0.76
Queti	114.09	109.85	108.20	5.89
Aripi	124.34	127.70	126.05	-1.71
Cloza	97.36	98.11	96.46	0.90
Risper	111.7	120.41	118.76	-7.06
Olanza	107.17	97.71	96.05	11.12
Serti	131.77	125.18	123.52	8.25

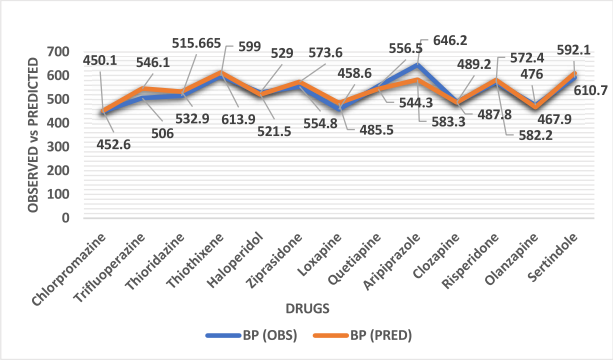

Fig. 4: Comparison between Predicted and Observed Residual Values of Boiling Point

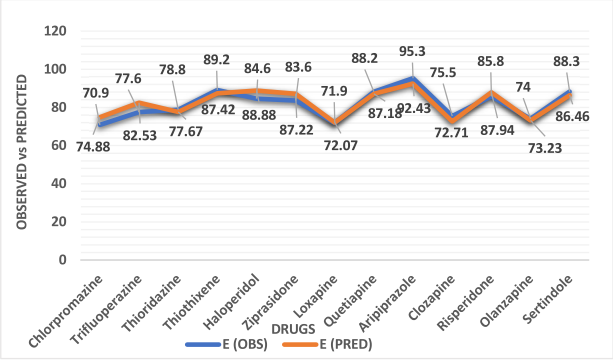

Fig. 5: Comparison between Predicted and Observed Residual Values of Enthalpy

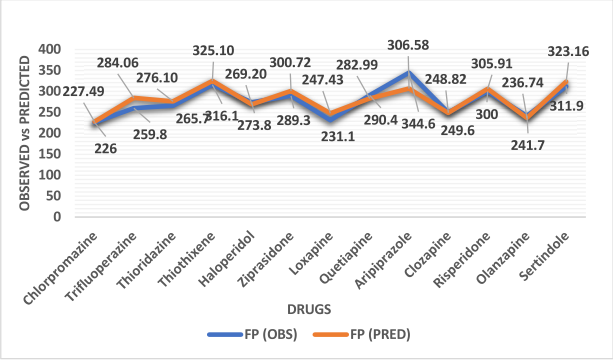

Fig. 6: Comparison between Predicted and Observed Residual Values of Flash Point

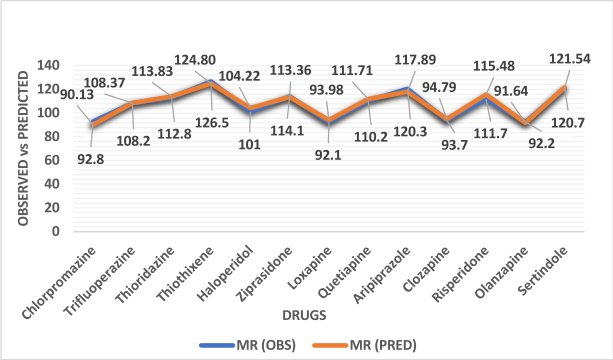

Fig. 7: Comparison between Predicted and Observed Residual Values of Molar Refractivity

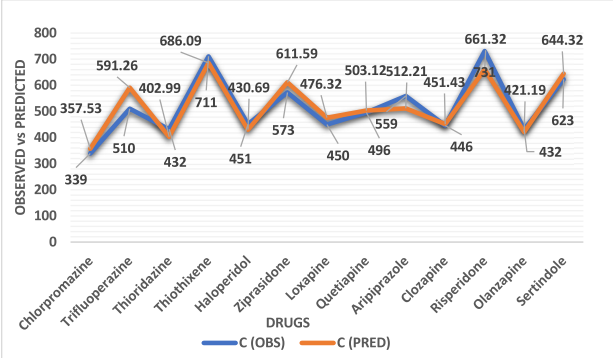

Fig. 8: Comparison between Predicted and Observed Residual Values of of Complexity

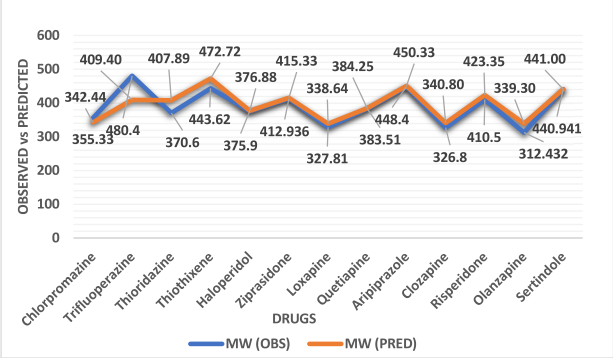

Fig. 9: Comparison between Predicted and Observed Residual Values of Molecular Weight

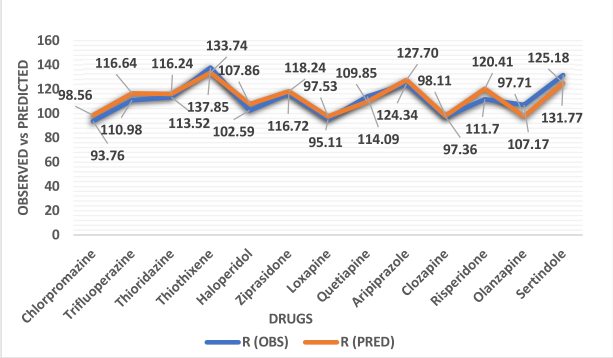

Fig. 10: Comparison between Predicted and Observed Residual Values of Refractivity

C. Multivariate Analysis

Multivariate analysis [30] is a statistical technique used to analyze data sets that involve the observation of multiple variables. Unlike univariate analysis, which considers only one variable at a time, multivariate analysis allows researchers to examine the relationships between two or more variables simultaneously. In the further parts of our study, we used the two chemometric methods, such as cluster analysis and principal component analysis. Chemometrics refers to a set of mathematical and statistical techniques used in chemistry and related fields for analyzing chemical data. These methods are applied to interpret and extract meaningful information from complex chemical datasets. Multivariate analysis were performed by means of software packages: OriginPro 2023b and IBM SPSS Statistics 25.

1) CLUSTER ANALYSIS (CA): This approach uses an unsupervised procedure to measure the distance or similarity of objects to be grouped. Based on how similar two objects are, they are grouped together into clusters. The first premise is that an object's proximity inside the space described by its variables indicates how similar it's characteristics are to each other. We employed Ward's method for combining clusters and used squared Euclidean distance as the measurement metric. The data presented in Table IX were utilized for analyzing the similarity of the drugs under study. Analysis of the data reveals variations in the topological index values based on the calculation formula and the structural characteristics of each drug used for descriptor calculations.

The initial analysis focused on comparing the physicochemical property values of these drugs obtained experimentally, depicted in Figure 11, with their corresponding theoretical values derived from topological indices, illustrated in Figures 12 and 13.

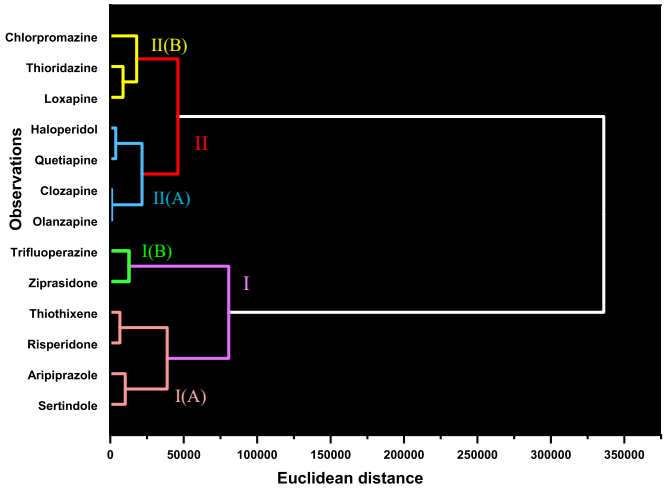

Fig. 11: Custer Analysis based on the Experimental Values of the Antipsychotic Drugs

A dendrogram of similarity analysis presented in Figures 11 and 12 shows that there are two visible clusters(primary) (color: red and purple) with its two sub clusters [color: purple → pink, green as sub cluster I(A) and I(B) and color:

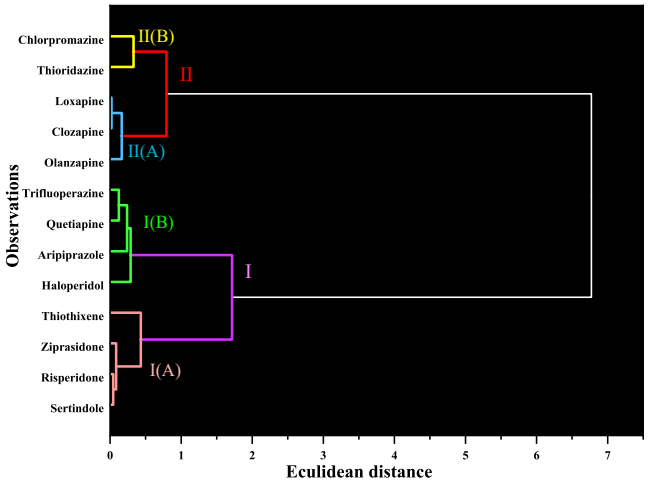

Fig. 12: Custer Analysis based on the Theoretical Values of the Antipsychotic Drugs

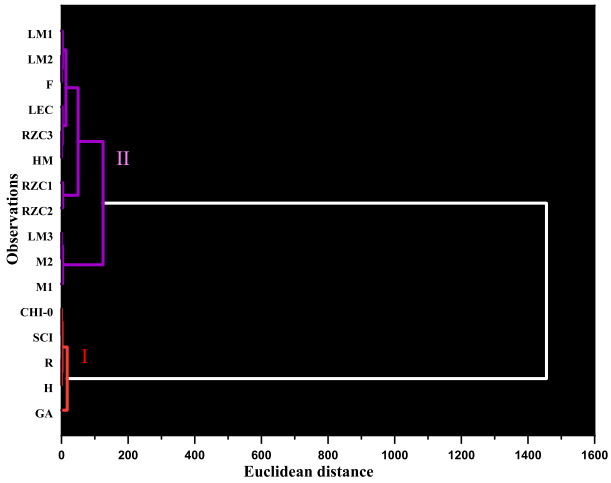

Fig. 13: Custer Analysis based on the Calculated Topological Indices

red → blue, yellow as sub cluster II(A) and II(B)] in the figure.

In employing the Ward's method utilizing Squared Euclidean distance as the metric, our similarity analysis between experimental and theoretical values reveals striking similarities. Notably, we observe that the primary clusters, labeled I and II, exhibit slight discrepancies in group compositions. Specifically, the drugs Haloperidol and Quetiapine, initially categorized within cluster I during similarity analysis with experimental values, transition to cluster II when analyzed against theoretical values.

Furthermore, an intriguing consistency emerges in the identification of the most representative observation, Quetiapine, and the least representative observation, Chlorpromazine, across both experimental and theoretical values cluster analyses.

The similarity observed among compounds within clusters stems from the computation of topological indices derived from their respective structures. Consequently, compounds grouped within clusters exhibit structural similarities, indicating that topological indices can effectively aid in categorizing first generation and second-generation antipsychotic drugs. However, it's crucial to note that the outcomes of this analysis are influenced not solely by the physicochemical properties of the compounds but also by their underlying chemical structures.

The subsequent similarity analysis (CA) allowed for a more precise comparison of the sixteen topological indices. Figure 13, the dendrogram illustrating the similarity of these indices, confirms that certain degree-based and distance-based indices such as two primary clusters, cluster I (color: red) and cluster II (color: purple) under which H index from cluster I has a least distance of 0.0449, ${}^0\chi^v$ index with a least distance of 0.18215 and from cluster II we have LM_1 index with a distance of 2.41391, this indicates that in the context of estimating the physicochemical properties of drugs using topological indices, these value signifies the similarity or closeness between the clusters being merged. A low least distance value signifies that the clusters being merged are cohesive and share common traits. Notably, our quadratic and stepwise regression models identifies H, ${}^0\chi^v$ and LM_1 as the most relevant indices for estimating physicochemical properties aligns with our dendrogram analysis, as these indices are grouped under similarity in clusters I and II respectively.

2) PRINCIPAL COMPONENT ANALYSIS (PCA): PCA is primarily used to reduce the dimensionality of the data while retaining as much variance as possible. Based on the results obtained in the preceding sections, we conducted a principal component analysis (PCA) in this study. Recently Wardecki et al. [31] (2023) evaluated topological indices' predictive power for physicochemical properties of bioactive substances, employing PCA and CA in their analysis and Ciura et al. [32] (2019) investigated the chromatographic behavior of antipsychotic drugs using quantitative structure–retention relationships using PCA and CA. In the year 2000's, Ražić et al. [33](2006) utilized PCA in their multivariate characterization of herbal drugs and rhizosphere soil samples based on metallic content. The data of this study underwent reduction employing the Kaiser criterion, resulting in three principal components chosen based on eigenvalues exceeding 1. These three components collectively account for 98.38 % of the system's variability.

Table IX presents the topological indices utilized in this study. Initially, all indices were subjected to a transformation involving the natural logarithm of the index value plus one. This adjustment was necessary due to potential variations in the scale of certain indices, which could differ by several orders of magnitude. Correlation coefficients showed significant enhancement when using a log-transformed dataset. Scree Plot of the eigenvalues of the principal componenet analysis is depicted in Figure 14 and the projection plot illustrating these components is displayed in Figure 15.

PCA score values reflect the transformation of the initial data projected onto the newly defined set of principal components. In the context of physicochemical properties

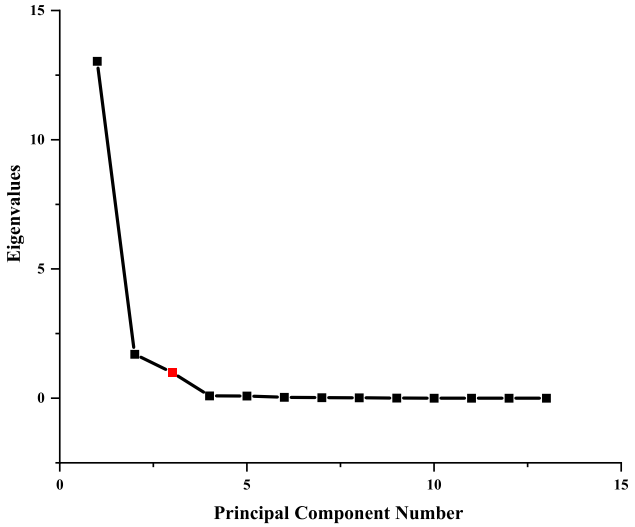

Fig. 14: Scree Plot of Eigenvalues of the PC's

of antipsychotic drugs, these scores essentially depict the performance of each drug with respect to the derived principal components. A higher score for a specific drug on a principal component suggests that the drug exhibits properties more closely aligned with that particular component. The results of the PCA are presented in Tables XXI, XXII and XXIII.

TABLE XXI: Eigenvalues, Percent of Variance and Cumulative Derived from PCA

No. of PC's	Eigenvalues	Variance explained %	Cumulative
1	13.03907	81.49%	81.49%
2	1.7007	10.63%	92.12%
3	1.00101	6.26%	98.38%

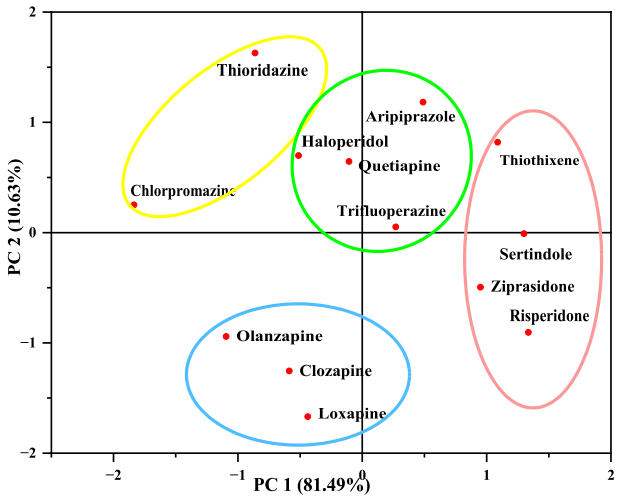

Fig. 15: Projection Plot of PCA Scores of Studied Drugs based on Log-Transformed TI's

Basically, the PCA confirmed the conclusions drawn from cluster analysis, quadratic and stepwise regression analysis. The clusters highlighted in Figure 15 provide further validation of the findings from the cluster analysis conducted on the entire dataset. Specifically, the sub clusters denoted as I(A), I(B) (colored in pink and green) and II(A), II(B) (colored in blue and yellow) consist of the same compounds identified in Figure 11. This observation underscores the effectiveness of PCA analysis, which condenses multiple data points describing the system into three principal components. Despite this reduction in dimensionality, the conclusions drawn from the analysis remain consistent.

TABLE XXII: PCA Scores of Antipsychotic Drugs

PC_1	PC_2	PC_3	Drugs
81.49%	**10.63%**	**6.26%**	
Scores			**Scores Labels**
-1.83534	0.25103	0.30666	Chlorpromazine
0.27019	0.05119	-0.47481	Trifluoperazine
-0.8619	1.62791	0.58853	Thioridazine
1.08917	0.81878	2.38593	Thiothixene
-0.51239	0.6981	-1.72986	Haloperidol
0.95062	-0.49491	0.16689	Ziprasidone
-0.43856	-1.66935	0.31534	Loxapine
-0.10657	0.64509	0.43803	Quetiapine
0.48912	1.18304	-1.32457	Aripiprazole
-0.58673	-1.25416	0.18421	Clozapine
1.33666	-0.90502	-0.56248	Risperidone
-1.09466	-0.94184	0.15466	Olanzapine
1.30038	-0.00986	-0.44853	Sertindole

TABLE XXIII: PCA Loadings of the Original Variables

No.	Parameter	PC loadings	
		PC_1	PC_2
1	LM_1	**0.26748**	-0.06531
2	LM_2	0.25654	-0.19729
3	LM_3	0.27376	-0.07175
4	LEC	0.24045	**0.20483**
5	RZC_1	0.24107	-0.21836
6	RZC_2	0.22131	-0.2707
7	RZC_3	0.26021	-0.24627
8	${}^0\chi^v$	0.22482	**0.42028**
9	R	0.22258	**0.41603**
10	M_1	0.27184	-0.05758
11	M_2	0.23452	-0.23757
12	SCI	0.25583	**0.28831**
13	GA	0.26982	0.07112
14	H	0.22897	**0.41401**
15	HM	0.25466	-0.19302
16	F	0.26565	-0.14762

Alternatively, when examining the PCA loading of the same descriptors that comprised subclusters (see Figure 12), a clear distinction can be made. Analysis of the data in Table XXIII reveals that the parameters ${}^0\chi^v$, H, R, SCI and LEC estimators exhibit stronger correlation with PC_2 than with PC_1, suggesting that PC_2 primarily represents a steric component. Bolded values represent the high positive loadings which coincides with this study of quadratic and stepwise regression results. Notably, this PCA analysis with high positive loadings confirms that the optimal estimators from Table XII.

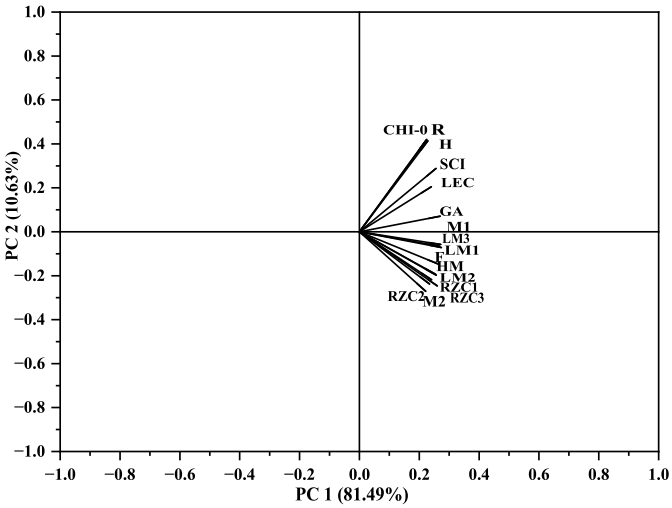

Fig. 16: Loading Plot

Features with significant positive or negative loadings essentially dictate the factor's influence. Establishing a rule regarding the minimum amount of interpretable loadings is not feasible. These factors are associated with the estimators (parameters) of the drugs examined in the study. If we set our threshold for coefficients at 0.5, only three factors were identified with high positive loading in PC_2, while all others exhibited negative loadings. The corresponding loading plot, visualizing this, is presented in Figure 16.

IV. Conclusion

In conclusion, this study demonstrates the applicability and significance of employing distance and degree-based topological descriptors, alongside chemometric methods like principal component analysis (PCA) and cluster analysis (CA), in understanding the physicochemical properties of antipsychotic drugs. Moreover the optimal estimators obtained from quadratic and stepwise regression confirmed that the further analysis using chemometric methods like cluster analysis and principal component analysis also predicts the same estimators. This suggests that these predictors are robust in explaining the physicochemical properties of the antispychotic drugs. The extension of traditional molecular connectivity indices to accommodate heteroatoms has notably enhanced the accuracy of estimating various drug properties. The findings not only reveal both similarities and distinctions among the investigated antipsychotic drugs but also provide insights into their structural characteristics and potential pharmacological behaviors. This research contributes to the advancement of quantitative structure-property relationship (QSPR) modeling and offers valuable insights for drug design and optimization processes. Future research could explore the application of these methodologies to a broader range of pharmaceutical compounds and further refine the predictive models to enhance drug development efforts. Additionally, the incorporation of additional statistical techniques and the exploration of novel descriptors could potentially enhance the accuracy and applicability of QSPR models in drug discovery and development.

References

[1] McDonald, C. and Murphy, K.C., "The new genetics of schizophrenia", *Psychiatric Clinics*, vol. 26, no. 1, pp. 41-63, 2003.

[2] Das, S., Rai, S. and Kumar, V., "On topological indices of Molnupiravir and its QSPR modelling with some other antiviral drugs to treat COVID-19 patients", *Journal of Mathematical Chemistry*, pp. 1-44, 2023.

[3] Olasupo, S.B., Uzairu, A., Shallangwa, G.A. and Uba, S., "Chemoinformatic studies on some inhibitors of dopamine transporter and the receptor targeting schizophrenia for developing novel antipsychotic agents", *Heliyon*, vol. 6, no. 7, pp. e04464, 2020.

[4] Olasupo, S.B., Uzairu, A., Shallangwa, G.A. and Uba, S., "Computer-aided drug design and in silico pharmacokinetics predictions of some potential antipsychotic agents", *Scientific African*, vol. 12, pp. e00734, 2021.

[5] Zhang, X., Saif, M.J., Idrees, N., Kanwal, S., Parveen, S. and Saeed, F., "QSPR analysis of drugs for treatment of schizophrenia using topological indices", *ACS omega*, vol. 8, no. 44, pp. 41417-41426, 2023.

[6] Miyamoto, S., Duncan, G.E., Marx, C.E. and Lieberman, J.A., "Treatments for schizophrenia: a critical review of pharmacology and mechanisms of action of antipsychotic drugs", *Molecular Psychiatry*, vol. 10, no. 1, pp. 79-104, 2005.

[7] Granger, K.T., Sand, M., Caswell, S., Lizarraga-Valderrama, L.R., Barnett, J.H. and Moran, P.M., "A new era for schizophrenia drug development–Lessons for the future", *Drug Discovery Today*, vol. 28, no. 7, pp. 103603, 2023.

[8] Murray, W.J., Hall, L.H. and Kier, L.B., "Molecular connectivity III: Relationship to partition coefficients", *Journal of Pharmaceutical Sciences*, vol. 64, no. 12, pp. 1978-1981, 1975.

[9] Kier, L.B. and Hall, L.H., "Derivation and significance of valence molecular connectivity", *Journal of Pharmaceutical Sciences*, vol. 70, no. 6, pp. 583-589, 1981.

[10] Kier, L.B. and Hall, L.H., "Molecular connectivity VII: specific treatment of heteroatoms", *Journal of Pharmaceutical Sciences*, vol. 65, no. 12, pp. 1806-1809, 1976.

[11] Gutman, I., Naji, A.M. and Soner, N.D., "On leap Zagreb indices of graphs", *Communications in Combinatorics and Optimization*, vol. 2, no. 2, pp. 99-117, 2017.

[12] Alsinai, A., Saleh, A., Ahmed, H., Mishra, L.N. and Soner, N.D., "On fourth leap Zagreb index of graphs", *Discrete Mathematics, Algorithms and Applications*, vol. 15, no. 2, pp. 2250077, 2023.

[13] Zhu, J.M., Dehgardi, N. and Li, X., "The third leap Zagreb index for trees", *Journal of Chemistry*, vol. 2019, no. 1, pp. 9296401, 2019.

[14] Raza, Z., "Leap Zagreb connection numbers for some networks models", *Indonesian Journal of Chemistry*, vol. 20, no. 6, pp. 1407-1413, 2020.

[15] Raza, Z., "Zagreb connection indices for some benzenoid systems", *Polycyclic Aromatic Compounds*, vol. 42, no. 4, pp.1814-1827, 2022.

[16] Sharma, V., Goswami, R. and Madan, A.K., "Eccentric connectivity index: a novel highly discriminating topological descriptor for structure– property and structure–activity studies",

Journal of Chemical Information and Computer Sciences, vol. 37, no. 2, pp. 273-282, 1997.

[17] Ravi, V., Chidambaram, N., Colakoğlu, *Ö*., Ahmed, H., Jaganathan, S. and Jaganathan, J., "QSPR study and distance-based new topological descriptors of some drugs used in the COVID-19 treatment", *Journal of Mathematics*, vol. 2023, no. 1, pp. 1187415, 2023.

[18] Ren, B., "Novel atomic-level-based AI topological descriptors: Application to QSPR/QSAR modeling", *Journal of Chemical Information and Computer Sciences*, vol. 42, no.4, pp. 858-868, 2002.

[19] Randic, M., "Characterization of molecular branching", *Journal of the American Chemical Society*, vol. 97, no. 23, pp.6609-6615, 1975.

[20] Gutman, I. and Trinajstić, N., "Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons", *Chemical Physics Letters*, vol. 17, no. 4, pp. 535-538, 1972.

[21] Zhou, B. and Trinajstić, N., "On general sum-connectivity index", *Journal of Mathematical Chemistry*, vol. 47, no. 1, pp. 210-218, 2010.

[22] Vukičević, D. and Furtula, B., "Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges", *Journal of Mathematical Chemistry*, vol. 46, no. 4, pp. 1369-1376, 2009.

[23] Fajtlowicz, S., "On conjectures of Graffiti", *Annals of Discrete Mathematics*, vol. 38, pp. 113-118, 1988.

[24] Shirdel, G.H., Rezapour, H. and Sayadi, A.M., "The hyper-Zagreb index of graph operations", *Iranian Journal of Mathematical Chemistry*, vol.4, no. 2, pp. 213-220, 2013.

[25] Furtula, B., Gutman, I. and Dehmer, M., "On structure-sensitivity of degree-based topological indices", *Applied Mathematics and Computation*, vol. 219, no. 17, pp. 8973-8978, 2013.

[26] Kier, L, "Molecular connectivity in chemistry and drug research", *Academic Press*, 1976, https://doi.org/10.1002/jps.2600660852.

[27] Kier, L. B., & Hall, L. H., "Molecular connectivity in structure-activity analysis", *John Wiley and Sons, New York*, 1986.

[28] Consonni, V., Ballabio, D. and Todeschini, R., "Comments on the definition of the Q^2 parameter for QSAR validation", *Journal of Chemical Information and Modeling*, vol. 49, no. 7, pp. 1669-1678, 2009.

[29] DEVORE, J.L., "PROBABILITY AND STATISTICS: For Engineering and the Sciences", *Cengage*, 2012.

[30] Backhaus, K., Erichson, B., Gensler, S., Weiber, R. and Weiber, T., "Multivariate analysis", *Springer*, 2021.

[31] Wardecki, D., Dolowy, M. and Bober-Majnusz, K., "Evaluation of the Usefulness of Topological Indices for Predicting Selected Physicochemical Properties of Bioactive Substances with Anti-Androgenic and Hypouricemic Activity", *Molecules*, vol. 28, no. 15, pp. 5822, 2023.

[32] Ciura, K., Rutecka, A., Szewczyk, A., Kawczak, P., Bączek, T. and Nowakowska, J., "Study of the chromatographic behavior of selected antipsychotic drugs on RP-TLC based on quantitative structure–retention relationships", *Journal of the Iranian Chemical Society*, vol. 16, pp.1019-1027, 2019.

[33] Ražić, S.S., Dogo, S.M. and Slavković, L.J., "Multivariate characterization of herbal drugs and rhizosphere soil samples according to their metallic content", *Microchemical Journal*, vol. 84, no. 1-2, pp.93-101, 2006.