Communications in Mathematics and Applications
Vol. 16, No. 4, pp. 1143–1157, 2025
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v16i4.3378

http://www.rgnpublications.com

Research Article

Quantitative Structure-Property Relationship Analysis of Thiosemicarbazone Derivatives with Antimalarial Activity and Some Therapeutic Compounds Using Leap Indices

B. Roopa Sri[1] and Y. Lakshmi Naidu*[2]

[1] *Department of Mathematics and Computer Science, Sri Sathya Sai Institute of Higher Learning, Anantapur 515001, Andhra Pradesh, India*
[2] *Centre for Excellence in Mathematical Biology, Department of Mathematics and Computer Science, Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam 515134, Andhra Pradesh, India*
***Corresponding author:** ylakshminaidu@sssihl.edu.in

Received: August 16, 2025 **Revised:** September 30, 2025 **Accepted:** October 19, 2025

Abstract. In this article, we perform the QSPR analysis of two datasets of therapeutic compounds using topological indices. The least explored category of indices is chosen for this study. They are the leap indices, namely F-Sombor, modified F-Sombor, reduced Sombor, second F-Sombor, Sombor, modified Sombor, F, and F_1 leap indices. We have considered some thiosemicarbazone derivatives that are tested for their antimalarial activity in one dataset and some randomly chosen drugs in another dataset. We investigate the correlation between a few leap indices and some physicochemical properties of these two datasets separately using machine learning techniques.

Keywords. QSPR analysis, F-Sombor leap index, Second F-Sombor leap index, Modified F-Sombor leap index, Reduced Sombor leap index

Mathematics Subject Classification (2020). 05C90, 05C92, 92E10

1. Introduction

A topological index is a graph invariant decoded from a chemical graph, a simple graph G formed by considering atoms of the molecules as nodes or vertices and bonds between them as edges.

The connection-based or leap indices, a category of topological indices, are formulated using the connection number. Most of the leap indices are formulated by replacing the vertex degree in a degree-based *TI* with a connection number d_2. The connection number $d_2(u)$ is the number of vertices at a distance 2 from a vertex u in graph G. These leap indices were introduced by Randic [9] by replacing the degree of the vertex in Zagreb indices with its corresponding connection number as follows.

$$ZC_1(G) = \sum_{u \in V(G)} d_2(u)^2$$

and

$$ZC_2(G) = \sum_{uv \in E(G)} d_2(u) d_2(v).$$

These Zagreb leap indices were restudied by Ali and Trinajstić [1] to check their chemical significance,

$$ZC_1^*(G) = \sum_{uv \in E(G)} d_2(u) + d_2(v)$$

and

$$ZC_2^*(G) = \sum_{uv \in E(G)} d_G(u) d_2(u) + d_G(v) d_2(v).$$

Numerous studies on leap indices have shown a promising correlation with various molecular properties, such as entropy, standard enthalpy of vaporization, and boiling point, of chemical compounds. It has been proven that ZC_1 correlates well with the entropy and acentric factor of octane isomers compared to classical indices. In 2019, Maji and Ganesh [6] proved that the third leap Zagreb index correlated well with the physicochemical properties of octane isomers. In 2023, Ramane and Pise [8] showed that the boiling point of benzenoid hydrocarbons has a stronger association with the third Zagreb leap index compared to the other Zagreb leap indices. Gutman and Trinajstić [2] proved that the modified first Zagreb leap index appears in an approximation formula of total-ϕ electron energy of alternant hydrocarbons.

It has also been proven that these leap indices can be generalized as the *bond incident connection-number* (BIC) index:

$$BIC(G) = \sum_{0 \le a \le b \le n-2} y_{a,b}(G) * \phi_{a,b},$$

where $y_{a,b}(G)$ represents the number of edges in G incident on the vertices with connection numbers a and b, and $\phi_{a,b}$ represents a non-negative real-valued function that depends on a and b.

Malaria is a long-standing, mosquito-borne disease caused by Plasmodium species. According to the *World Health Organization* (WHO), in 2022, over 200 million cases and 608,000 deaths were recorded globally due to malaria[1]. The African region shares the highest burden among the 85 malaria-endemic countries. Malaria is primarily diagnosed in tropical and subtropical countries such as India and those in Africa, among others, and predominantly, pregnant women and children below 5 years are at greater risk. To control and eliminate malaria in these countries, the WHO Global Malaria Programme (GMP) was initiated under the guidance of the 'Global Technical Strategy for Malaria 2016-2030' in May 2015. Despite many antimalarial

[1]World Health Organisation, *World Malaria Report 2023*, (2023), URL: https://www.who.int/publications/i/item/9789240086173.

drugs and vaccines, the mortality rate has not been adequately controlled, and it is a tough challenge to eliminate it. Antimalarials include chloroquine phosphate, hydroxychloroquine, atovaquone-proguanil, primaquine phosphate, quinine sulfate plus doxycycline, tetracycline, and clindamycin, among others. An appropriate treatment course is determined by selecting the antimalarials depending on the severity and drug susceptibility of the infecting Plasmodium species. However, these antimalarials are known for adverse effects and contraindications like blurred vision, insomnia, hypersensitivity towards the drug, and anorexia. In some cases, these effects may persist even after discontinuation of the drug (Hill *et al*. [3]). Hence, there is an urgent need for global efforts to develop novel antimalarial drugs to achieve malaria elimination. To combat this, many novel compounds are being formulated and synthesized to test their antimalarial activity. Thiosemicarbazone derivatives are among those that have shown promising antimalarial activity.

In pharmaceutical sciences, especially in drug discovery and development, it is essential to study the physicochemical properties of a molecule to arrive at a new therapeutic compound. In general, it is time-consuming and expensive to study the physicochemical properties of millions of molecules to test their potential in treating a disease. Therefore, topological indices play a vital role in predicting the properties of millions of compounds computationally through QSPR analysis. This helps in identifying the desired compound without performing the experiments. Hence, we consider some leap indices to check their correlation with some physicochemical properties of therapeutic compounds.

2. Methods and Methodology

For this study, we considered the following leap indices.

In 2022, Kulli *et al*. [5] introduced Sombor and modified Sombor leap indices, and calculated these indices for several chemical drugs.

$$SL(G)=\sum_{uv\in E(G)}\sqrt{d_2(u)^2+d_2(v)^2} \quad \text{and} \tag{2.1}$$

$$^{m}SL(G)=\sum_{uv\in E(G)}\frac{1}{\sqrt{d_2(u)^2+d_2(v)^2}}. \tag{2.2}$$

In 2018, Kulli [4] has also introduced F-leap and F_1-leap indices as follows:

$$FL(G)=\sum_{u\in V(G)} d_2(u)^3 \quad \text{and} \tag{2.3}$$

$$F_1L(G)=\sum_{uv\in E(G)} d_2(u)^2+d_2(v)^2. \tag{2.4}$$

We formulated the F-Sombor leap index, the modified F-Sombor leap index, and the reduced Sombor leap index in [12] as:

$$FSO_L(G)=\sum_{uv\in E(G)}\sqrt{d_2(u)^4+d_2(v)^4}, \tag{2.5}$$

$$^{m}FSO_L(G)=\sum_{uv\in E(G)}\frac{1}{\sqrt{d_2(u)^4+d_2(v)^4}} \quad \text{and} \tag{2.6}$$

$$^{red}SO_L(G)=\sum_{uv\in E(G)}\sqrt{(d_2(u)-1)^2+(d_2(v)-1)^2}. \tag{2.7}$$

In the same paper, we also introduced a new leap index called the second F-Sombor leap index:

$$^{2}FSO_L(G) = \sum_{uv\in E(G)} d_2(u)^4 \times d_2(v)^4. \tag{2.8}$$

Here, we have considered two datasets of therapeutic compounds: one consisting of sixteen thiosemicarbazone derivatives with antimalarial activity, and another comprising a set of 30 randomly chosen drugs.

Thiosemicarbazones belong to the class of Schiff bases obtained through the condensation of thiosemicarbazide with a suitable aldehyde or ketone. They have attracted researchers for their broad spectrum of antimalarial, antiviral, antibacterial, and other pharmacological activities (Shim *et al*. [11], and Shakya and Yadav [10]). Numerous studies have been conducted to investigate the biological activity of these derivatives against various diseases. Nevertheless, very few compounds have been tested for antimalarial activity (Matsa *et al*. [7]). The properties chosen for our study are *Boiling Point* (*BP*) and *Molecular Weight* (*MW*). The observed values of these properties for the molecules from a) to p) in Table 2 are taken from PubChem and ChemSpider.

We also consider a second dataset of 30 randomly selected therapeutic compounds used for the treatment of various diseases. The properties considered for the study are *Boiling Point* (*BP*), *Molar Volume* (*MV*), *Molar Refractivity* (*MR*), and *Polarizability* (*Polar*). Table 5 consists of the compounds and their corresponding experimental values of the selected properties taken from the database PubChem and ChemSpider. All the above-mentioned physicochemical parameters are highly associated with the molecular behaviour in biological systems. Therefore, studying these properties through *Quantitative Structure-Property Relationship* (QSPR) modeling helps predict the reactivity and stability of molecules using topological indices.

3. Main Results

3.1 QSPR Analysis of Thiosemicarbazone Derivatives

In this section, we consider sixteen thiosemicarbazone derivatives that have been studied for their antimalarial activity by Matsa *et al*. [7]. Using the SMILES strings of sixteen thiosemicarbazone derivatives listed in Table 1, all the selected leap indices are calculated and shown in Table 3.

Table 1. Chemical graphs of thiosemicarbazone derivatives with ChemSpider ID

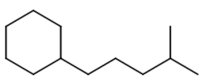
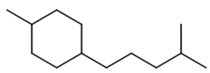
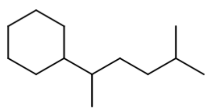
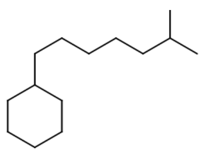
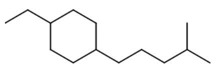
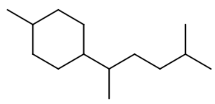
 a) 4649245	 b) 4976514	 c) 4698372
 d) 7850585	 e) 5261105	 f) 12624786

Table 1 (*continued*)

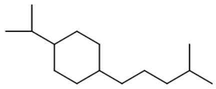
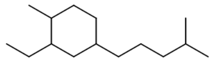
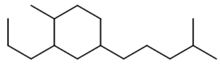
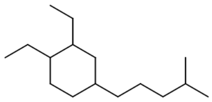
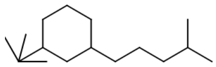
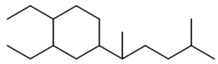
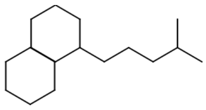
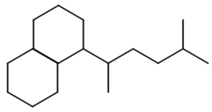
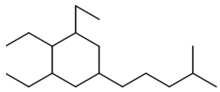
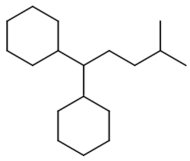
 g) 4976512	 h) 12234626	 i) 12232327
 j) 4976515	 k) 7855120	 l) 4576907
 m) 5278052	 n) 4540772	 o) 5258974
	 p) 227217	

Table 2. Experimental values of thiosemicarbazone derivatives

Molecule	*BP*	*MW*
a)	315.79	179.25
b)	329.01	193.28
c)	325.49	193.28
d)	345.02	205.29
e)	341.79	209.27
f)	361.51	209.27
g)	346.05	221.33
h)	376.72	225.27
i)	388.33	239.30
j)	365.54	239.30
k)	319.77	247.25
l)	373.98	253.33
m)	389.13	229.31
n)	397.57	243.34
o)	89.19	269.33
p)	399.22	255.35

Table 3. Calculated leap indices of thiosemicarbazone derivatives

Molecule	SL	^{m}SL	FL	F_1L	FSO_L	$^{m}FSO_L$	$^{2}FSO_L$	$^{red}SO_L$
a)	40.4413	3.7342	184	142	106.1945	1.5908	25492	24.2579
b)	46.0982	3.8521	230	170	125.9934	1.5712	38358	28.5006
c)	47.5932	3.8221	248	186	138.6904	1.5742	79843	30.0101
d)	46.0982	4.4413	200	158	117.5082	1.9444	26004	27.0864
e)	50.3856	4.1398	250	190	140.0667	1.6926	49944	31.5071
f)	53.2500	3.9390	294	214	158.4893	1.5545	92709	34.2527
g)	56.1877	4.2745	302	222	165.1601	1.7145	81590	35.8160
h)	57.7866	4.2331	332	238	177.7179	1.6706	141365	37.6379
i)	60.6549	4.7856	352	250	185.9295	2.2395	158071	39.0812
j)	62.3035	4.5676	370	264	198.3840	1.8379	203156	40.9783
k)	63.9777	4.3368	446	276	211.6272	1.6286	263668	42.5101
l)	69.3548	4.6670	434	308	229.5202	1.8350	288132	46.5904
m)	70.7333	4.5121	464	322	241.4157	1.7043	461714	47.7249
n)	77.9429	4.6344	552	372	279.1226	1.7209	763943	53.5357
o)	74.2457	5.0415	514	344	261.7326	2.0302	520050	50.4446
p)	76.2501	5.2538	480	338	251.4165	2.1102	466451	50.3238

3.1.1 Regression Analysis

We perform the simple linear and multiple linear regression analysis between all the leap indices and properties of the thiosemicarbazone derivatives.

Using the calculated values of indices from Table 3 and experimental properties, BP and MW, from Table 2, we first performed a simple linear regression analysis to determine whether a linear relationship existed between the indices and the properties. Their correlation matrix is given in Table 4.

Table 4. Correlation $|r|$ matrix of leap indices with BP and MW

$TI \rightarrow$	SL	^{m}SL	FL	F_1L	FSO_L	$^{m}FSO_L$	$^{red}SO_L$	$^{2}FSO_L$
BP	**0.8152**	0.8016	0.7371	0.7994	0.7855	0.6239	0.8018	0.7296
MW	**0.9144**	0.8869	0.8940	0.8904	0.8938	0.6047	0.9060	0.7259

The following are the corresponding best-fit linear regression models, involving a single leap index at a time, for each property of the chosen antimalarial drugs.

Best-fit Linear Regression Models:

$$BP = 241.8060(\pm 22.9011) + 1.9880(\pm 0.3775) \times SL; \tag{3.1}$$

$R^2: 0.6646$, $RMSE$: 16.1458, SE: 0.3774, F-value: 27.7371, p-value: 1.1918e-04.

$$MW = 109.5325(\pm 14.0782) + 1.9607(\pm 0.2321) \times SL; \tag{3.2}$$

$R^2: 0.8360$, $RMSE$: 9.9254, SE: 0.2320, F-value: 71.3891, p-value: 7.2094e-07.

From the parameters in equations (3.1) and (3.2), it is clear that the Sombor leap index, SL, exhibits the highest correlation with BP ($R^2 = 0.7$) and MW ($R^2 = 0.8$). The linear regression models related to each property involving each topological index are given below.

Simple Linear Regression Models for BP:

$$BP = 241.8060(\pm 22.9011) + 1.9880(\pm 0.3775) \times SL,$$

$$BP = 128.3399(\pm 46.4479) + 52.8324(\pm 10.5324) \times^{m} SL,$$

$$BP = 296.0836(\pm 16.5096) + 0.1817(\pm 0.0445) \times FL,$$

$$BP = 280.2366(\pm 16.6846) + 0.3206(\pm 0.0644) \times F_1L,$$

$$BP = 282.8776(\pm 16.9323) + 0.4142(\pm 0.0872) \times FSO_L,$$

$$BP = 206.3368(\pm 51.8619) + 86.6538(\pm 29.0126) \times^{m} FSO_L,$$

$$BP = 262.6388(\pm 19.9506) + 2.5181(\pm 0.5017) \times^{red} SO_L,$$

$$BP = 338.2137(\pm 7.5131) + 0.9635 \times 10^{-4}(\pm 0.2413 \times 10^{-4}) \times^{2} FSO_L.$$

Simple Linear Regression Models for MW:

$$MW = 109.5325(\pm 14.0782) + 1.9607(\pm 0.2321) \times SL,$$

$$MW = 0.7319(\pm 31.5603) + 51.3978(\pm 7.1565) \times^{m} SL,$$

$$MW = 157.9141(\pm 9.6243) + 0.1937(\pm 0.0259) \times FL,$$

$$MW = 147.9838(\pm 11.1159) + 0.3139(\pm 0.0429) \times F_1L,$$

$$MW = 148.9344(\pm 10.7898) + 0.4144(\pm 0.0556) \times FSO_L,$$

$$MW = 95.1574(\pm 46.4695) + 73.8594(\pm 25.9959) \times^{m} FSO_L,$$

$$MW = 129.3629(\pm 12.4271) + 2.5019(\pm 0.3125) \times^{red} SO_L,$$

$$MW = 207.0684(\pm 6.6446) + 0.8428 \times 10^{-4}(\pm 0.2134 \times 10^{-4}) \times^{2} FSO_L.$$

The plots of the actual and predicted values for each property, along with the best-fit line, are given in Figure 1. The red solid line represents the best-fit linear model of the predicted values, and the blue dots denote the observed values of the property. However, the R^2 value for BP does not exceed 0.8.

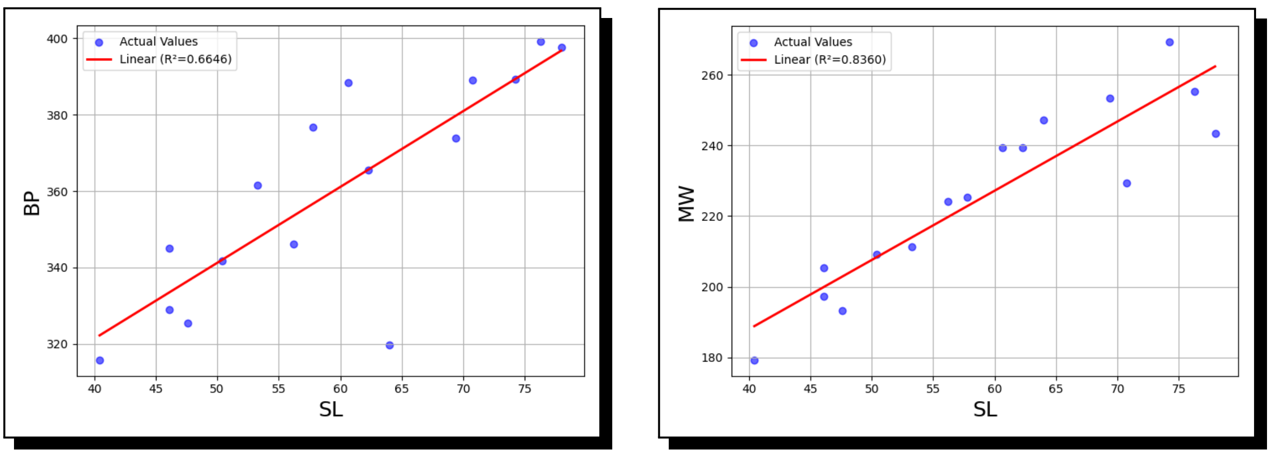


Figure 1. Correlation of SL with BP and MW: Leap indices

To enhance the performance of the leap indices in QSPR analysis, multiple linear regression was performed. The multiple linear regression model, which involved SL, ${}^{m}SL$, F_1L, FSO_L, ${}^{m}FSO_L$, and ${}^{red}SO_L$, yielded the following best-fit models for each property, with significant parameters, amongst all other combinations.

Best-fit Multiple Linear Regression Models for Each Property:

$$BP = 250.5704 - 10.5760 \times SL + 28.7308 \times {}^{m}SL + 5.7645 \times F_1L - 6.7191 \times FSO_L + 34.2894 \times {}^{m}FSO_L + 9.5191 \times {}^{red}SO_L; \tag{3.3}$$

R^2 : 0.8807, *RMSE*: 9.6287, *SE*: 12.8383, F-value: 11.0740, p-value: 1.0185e-03.

$$MW = 1.4368 - 8.6600 \times SL + 34.5205 \times {}^{m}SL - 3.4753 \times F_1L + 1.9411 \times FSO_L + 7.6250 \times {}^{m}FSO_L + 27.8774 \times {}^{red}SO_L; \tag{3.4}$$

R^2 : 0.9847, *RMSE*: 3.0303, *SE*: 4.0405, F-value: 96.6480, p-value: 1.1756e-06.

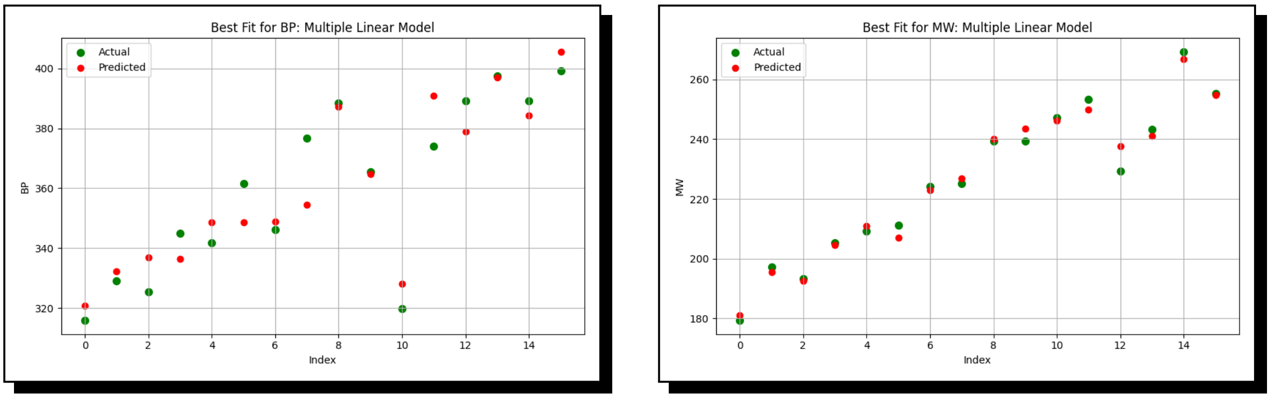


Figure 2. Plots for actual vs predicted values of *BP* and *MW* using (3.3) and (3.4): Thiosemicarbazone Derivatives

The multiple linear regression analysis involving six of the selected leap indices enhanced the $R^2(\%)$ from 70% and gave better results than simple linear regression models. Here, $R^2(\%)$ is observed to be 88% for *BP* and 98% for *MW*. The other statistical parameters in the model, including low *RMSE* values, a low *SE* value, a high F-value, and a p-value less than 0.05 for each property, indicate a significant correlation between the indices and the properties of thiosemicarbazone derivatives. The best-fit plots are depicted against the actual and predicted values of each property using the models (3.3) and (3.4), respectively. They are depicted in Figure 2. In each plot, the solid green dots represent the experimental values, and the red dots represent the predicted values of the corresponding property.

3.2 QSPR Analysis of Therapeutic Compounds

The leap indices calculated using SMILES string of the randomly chosen drugs from Table 5 are given in Table 6.

Table 5. Experimental values of 30 therapeutic compounds

Number	Drugs	*BP*	*MV*	*MR*	*Polar*
1	Quetiapine	556.5	301.1	110.2	43.7
2	Haloperidol	529.0	303.3	101.0	40.0
3	Metformin	172.5	100.8	33.4	13.2
4	Pioglitazone	575.4	282.8	98.2	38.9

Table 5 (*continued*)

Number	Drugs	*BP*	*MV*	*MR*	*Polar*
5	Chloroquine	460.6	287.9	97.4	38.6
6	Lumefantrine	642.5	422.3	151.0	59.9
7	Pyrimethamine	368.4	180.2	67.1	26.6
8	Hydroxychloroquine	516.7	285.4	99.0	39.2
9	Leflunomide	289.3	194.1	61.0	40.6
10	Baricitinib	707.2	238.1	98.2	38.9
11	Upadacitinib	–	243.0	91.6	36.3
12	Fluticasone propionate	568.3	377.0	121.1	48.0
13	Dacomitinib	665.7	349.5	129.5	51.3
14	Paclitaxel	957.1	610.6	219.3	86.9
15	Ispinesib	708.0	425.2	149.2	59.2
16	Donepezil	527.9	332.5	110.4	43.8
17	Rivastigmine	316.2	241.2	73.1	29.0
18	Linezolid	585.5	259.0	83.0	32.9
19	Moxifloxacin	636.4	285.0	101.8	40.4
20	Pyrazinamide	273.3	87.7	31.9	12.6
21	Rifampin	1004.4	611.7	213.1	84.5
22	Doxorubicin	810.3	336.6	131.5	52.1
23	Carmustine	309.6	146.4	46.6	18.5
24	Afatinib	676.9	352.0	131.2	52.0
25	Gefitinib	586.8	337.8	118.8	47.1
26	Gemcitabine	482.7	142.3	52.1	20.6
27	Capecitabine	517.0	240.5	82.3	32.6
28	Exemestane	453.7	260.6	85.8	34.0
29	Letrozole	563.5	234.5	87.1	34.5
30	Fulvestrant	674.8	505.1	154.0	61.1

Table 6. Calculated leap indices of 30 therapeutic compounds

Number	SL	^{m}SL	FL	F_1L	FSO_L	$^{m}FSO_L$	$^{2}FSO_L$	$^{red}SO_L$
1	133.9438	7.6223	962	654	474.3170	3.1734	1327400	92.1066
2	124.2156	6.5869	834	574	420.3022	2.2170	578810	85.1712
3	26.4632	2.5988	140	94	73.6464	1.1585	24128	15.9061
4	114.1417	6.6108	702	496	360.6787	2.2985	298782	76.5830
5	96.3539	5.9764	636	436	324.6954	2.2317	575395	64.8669
6	177.4524	9.7302	1534	942	690.2053	4.2975	4472126	124.8857
7	82.3985	4.3318	656	412	313.5154	1.5251	926098	57.7867
8	99.3314	6.5108	656	448	333.9209	2.7837	580726	66.4593
9	90.9920	4.5632	670	430	327.4680	1.4402	346474	63.4895

Table 6 (*continued*)

Number	SL	^mSL	FL	F_1L	FSO_L	$^mFSO_L$	$^2FSO_L$	$^{red}SO_L$
10	144.2099	6.4594	1176	761	559.7966	2.3593	1690426	104.2093
11	151.6072	6.4792	1362	826	614.1663	1.9722	3272857	110.9553
12	226.5174	7.4844	2870	1586	1212.4511	2.7118	24806563	176.1913
13	159.9646	8.6303	1106	752	548.4109	2.9965	1043463	109.9744
14	361.6029	14.7839	3758	2177	1633.1143	4.9229	24221784	268.6955
15	187.9511	9.7551	1540	980	725.1099	3.8692	3405849	132.5816
16	144.1618	7.1954	1058	710	515.6952	2.4702	1134408	101.1726
17	74.6639	4.6132	456	326	241.8069	1.6884	242338	49.8977
18	117.8005	6.2941	870	574	426.0818	2.2304	868055	82.1559
19	168.5122	7.7868	1600	968	721.1852	3.1452	4905441	123.3409
20	32.9969	2.5925	166	128	95.5327	1.0798	49184	20.5024
21	321.0333	13.5516	2834	1772	1315.4132	4.2342	8847628	234.7719
22	237.6974	8.5233	2384	1410	1054.1217	2.4576	7550754	178.8317
23	36.3887	4.0389	174	138	105.7632	2.4609	54836	21.6739
24	164.4766	8.9577	1170	778	567.6972	3.1322	1118919	113.4177
25	151.2300	8.1526	1034	712	517.7998	2.8446	993271	103.9390
26	98.2051	4.0616	898	548	413.3332	1.2856	2050163	72.2310
27	117.6617	6.5843	888	584	431.8704	2.7415	1241091	81.7945
28	146.8878	4.5849	1564	916	686.2308	1.2279	5538499	112.4079
29	107.2972	5.8590	766	520	383.4535	2.0995	1177704	74.1121
30	213.3161	10.0351	1828	1136	837.3056	3.4616	5200345	152.1297

Now, we perform simple linear regression involving one topological index and also involving a combination of two topological indices with a fitting parameter. First, we perform simple regression analysis between leap indices in Table 6 and properties in Table 5 to estimate their correlation. The correlation matrix with $|r|$ values is given in Table 7. From $|r|$ values in bold, it is clear that ^mSL exhibits the highest correlation for all properties, with $|r|$ ranging from 0.92 to 0.98. The index $^2FSO_L$ exhibits the least correlation with all properties.

Table 7. Correlation $|r|$ matrix of leap indices

$TI \rightarrow$	SL	^mSL	FL	F_1L	FSO_L	$^mFSO_L$	$^{red}SO_L$	$^2FSO_L$
BP	0.9122	**0.9210**	0.8130	0.8525	0.8424	0.8077	0.8960	0.5158
MV	0.9228	**0.9671**	0.8305	0.8648	0.8560	0.8803	0.9024	0.6026
MR	0.9402	**0.9861**	0.8415	0.8781	0.8684	0.8977	0.9192	0.5924
$Polar$	0.9338	**0.9742**	0.8353	0.8709	0.8621	0.8753	0.9132	0.5850

The robust linear models for each property are as follows:

$$BP = 119.9732(\pm 38.1776) + 61.8998(\pm 5.0377) \times^m SL; \tag{3.5}$$

R^2 : 0.8483, $RMSE$: 72.7147, SE: 5.0377, F-value: 150.9757, p-value: 1.44e-12.

$$MV = -7.8546(\pm 16.7876) + 43.8145(\pm 2.2152) \times^m SL; \tag{3.6}$$

R^2 : 0.9354, $RMSE$: 31.9744, SE: 2.2152, F-value: 391.2042, p-value: 1.35e-17.

$$MR = -5.8924(\pm 3.8549) + 15.6947(\pm 0.5087) \times {}^m SL; \tag{3.7}$$

R^2 : 0.9724, $RMSE$: 7.3422, SE: 0.5087, F-value: 951.9882, p-value: 1.37e-22.

$$Polar = -0.5132(\pm 2.0394) + 6.0427(\pm 0.2691) \times {}^m SL; \tag{3.8}$$

R^2 : 0.9492, $RMSE$: 3.8843, SE: 0.2691, F-value: 504.2063, p-value: 5.31e-19.

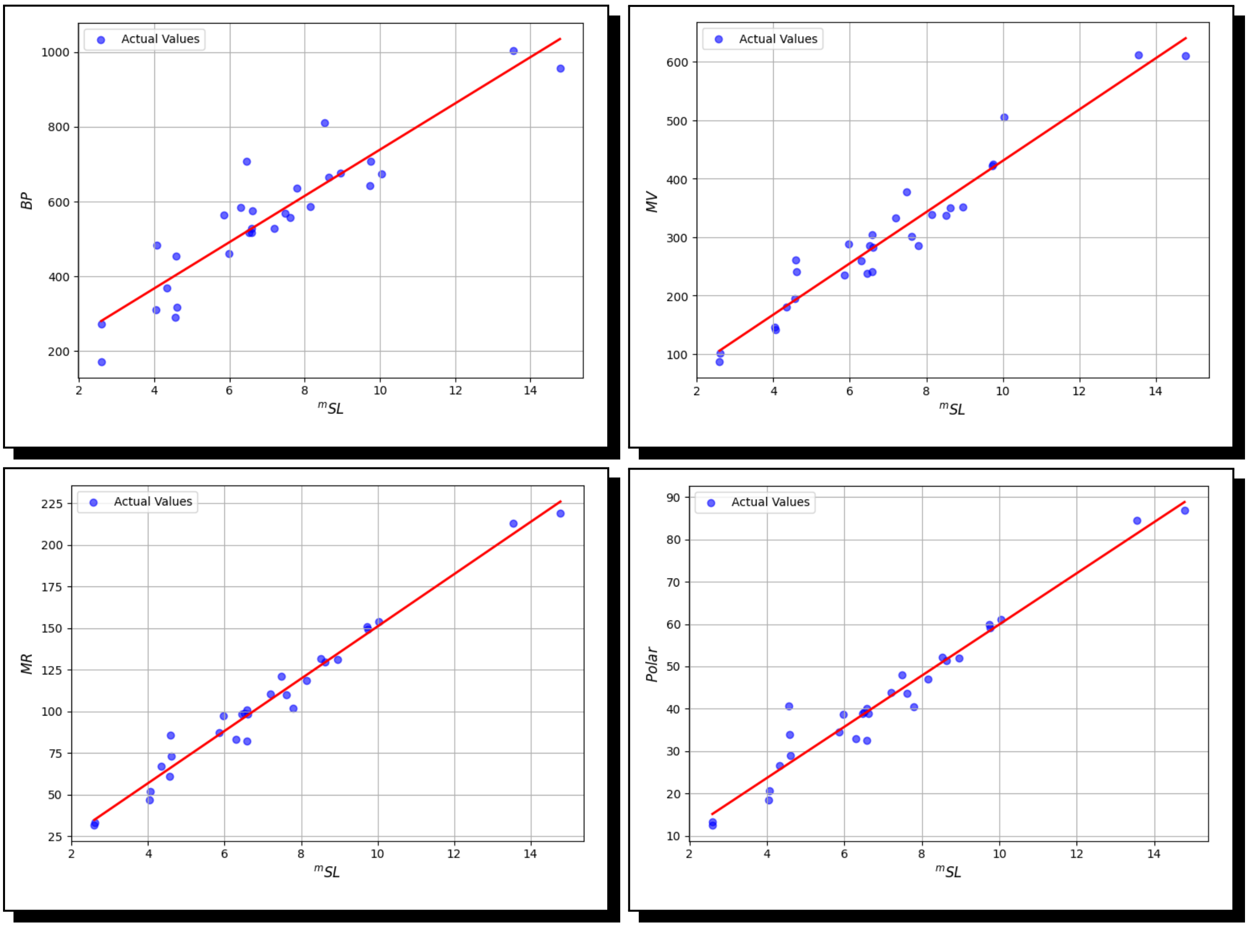


Figure 3. Correlation of ${}^m SL$ with BP, MV, MR, and $Polar$: Therapeutic compounds

Figure 3 depicts the plots of the best-fit models of all properties. All the statistical parameters— $R^2 > 0.9$, low $RMSE$ and SE, high F-value, and p-value < 0.05 —make models (3.6), (3.7), and (3.8), robust for the properties MV, MR, and Polarizability, respectively.

The following linear regression models were developed for each property using individual leap indices.

Simple Linear Regression Models for BP:

$$BP = 230.7847(\pm 31.7555) + 2.2712(\pm 0.1962) \times SL,$$

$$BP = 119.9732(\pm 38.1776) + 61.8998(\pm 5.0377) \times {}^m SL,$$

$$BP = 337.8585(\pm 36.6754) + 0.1814(\pm 0.0250) \times FL,$$

$$BP = 303.1033(\pm 35.2909) + 0.3344(\pm 0.0394) \times F_1 L,$$

$$BP = 308.5638(\pm 36.1334) + 0.4406(\pm 0.0542) \times FSO_L,$$

$$BP = 152.2316(\pm 60.5892) + 157.2405(\pm 22.0838) \times {}^{m}FSO_L,$$
$$BP = 256.9372(\pm 32.7004) + 2.9328(\pm 0.2795) \times {}^{red}SO_L,$$
$$BP = 499.3410(\pm 35.7786) + 0.1572 \times 10^{-4}(\pm 0.0502 \times 10^{-4}) \times {}^{2}FSO_L.$$

Simple Linear Regression Models for MV:

$$MV = 79.0333(\pm 20.1279) + 1.5486(\pm 0.1244) \times SL,$$
$$MV = -7.8546(\pm 16.7876) + 43.8145(\pm 2.2152) \times {}^{m}SL,$$
$$MV = 150.5766(\pm 23.6474) + 0.1249(\pm 0.0161) \times FL,$$
$$MV = 127.8735(\pm 22.8535) + 0.2287(\pm 0.0256) \times F_1L,$$
$$MV = 131.3122(\pm 23.3678) + 0.3018(\pm 0.0351) \times FSO_L,$$
$$MV = 4.1575(\pm 32.8636) + 115.5093(\pm 11.9783) \times {}^{m}FSO_L,$$
$$MV = 97.7668(\pm 21.3885) + 1.9910(\pm 0.1829) \times {}^{red}SO_L,$$
$$MV = 256.1302(\pm 22.4657) + 0.1238 \times 10^{-4}(\pm 0.0315 \times 10^{-4}) \times {}^{2}FSO_L.$$

Simple Linear Regression Models for MR:

$$MR = 25.2884(\pm 6.2518) + 0.5544(\pm 0.0386) \times SL,$$
$$MR = -5.8924(\pm 3.8549) + 15.6947(\pm 0.5087) \times {}^{m}SL,$$
$$MR = 51.1906(\pm 8.0569) + 0.0445(\pm 0.0055) \times FL,$$
$$MR = 42.9816(\pm 7.6509) + 0.0816(\pm 0.0086) \times F_1L,$$
$$MR = 44.2584(\pm 7.8732) + 0.1076(\pm 0.0118) \times FSO_L,$$
$$MR = -1.6158(\pm 10.7202) + 41.3866(\pm 3.9073) \times {}^{m}FSO_L,$$
$$MR = 32.0115(\pm 6.8671) + 0.7125(\pm 0.0587) \times {}^{red}SO_L,$$
$$MR = 89.2447(\pm 7.9678) + 0.0427 \times 10^{-4}(\pm 0.0112 \times 10^{-4}) \times {}^{2}FSO_L.$$

Simple Linear Regression Models for Polar:

$$Polar = 11.3290(\pm 2.5585) + 0.2146(\pm 0.0158) \times SL,$$
$$Polar = -0.5132(\pm 2.0394) + 6.0427(\pm 0.2691) \times {}^{m}SL,$$
$$Polar = 21.3674(\pm 3.1953) + 0.0172(\pm 0.0022) \times FL,$$
$$Polar = 18.2103(\pm 3.0619) + 0.0315(\pm 0.0034) \times F_1L,$$
$$Polar = 18.6846(\pm 3.1361) + 0.0416(\pm 0.0047) \times FSO_L,$$
$$Polar = 1.6722(\pm 4.5866) + 15.7248(\pm 1.6717) \times {}^{m}FSO_L,$$
$$Polar = 13.9259(\pm 2.7702) + 0.2758(\pm 0.0237) \times {}^{red}SO_L,$$
$$Polar = 36.1186(\pm 3.1258) + 0.0164 \times 10^{-4}(\pm 0.0043 \times 10^{-4}) \times {}^{2}FSO_L.$$

To enhance the R^2 value of BP in (3.5), a combination of two leap indices was used with a fitting parameter λ ranging from -20 to $+20$. The model that we considered here is:

$$Property = A + B(TI_1 + \lambda TI_2), \tag{3.9}$$

where A and B are the coefficients and TI_i represents a topological index.

We got the robust model (3.10) with significant parameters for the combination ($FSO_L + \lambda^{red}SO_L$) as the independent variable in the simple linear regression model (3.9). The best-fit model for BP is:

$$BP = 175.1011 - 1.1747(FSO_L - 8.68^{red}SO_L); \tag{3.10}$$

R^2 : 0.9022, $RMSE$: 58.3892, F-value: 249.0198, p-value: 3.7471e-15.

The change in $|r|$ with the change in λ from -20 to $+20$ for BP is depicted in Figure 4. The red dotted line represents the maximum $R^2 = 0.90$ obtained at $\lambda = -8.68$.

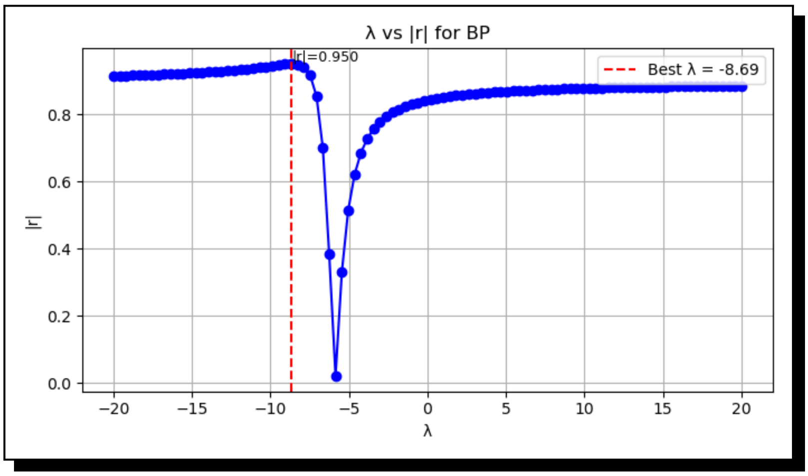


Figure 4. Plot for λ vs $|r|$: BP

The best-fit regression model for BP is depicted in Figure 5. The blue circles represent the actual values of BP, and the red line is the best-fit model.

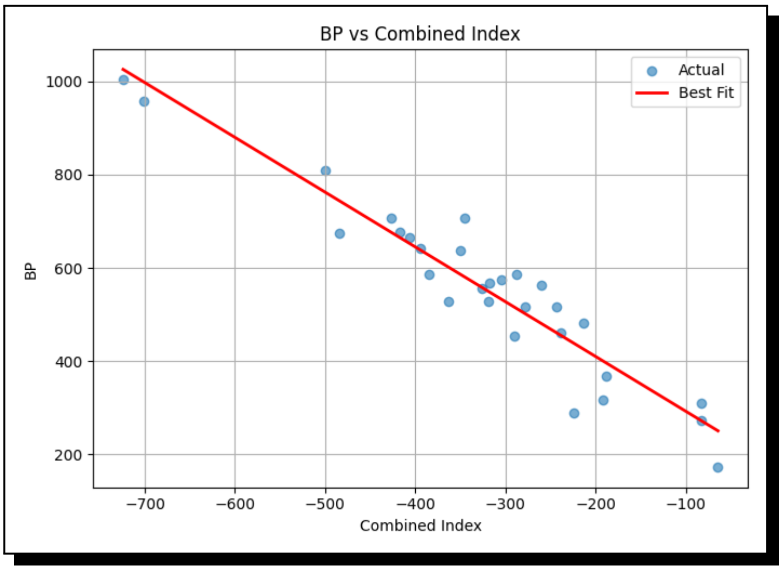


Figure 5. Correlation of $(FSO_L + \lambda^{red}SO_L)$ with BP

4. Conclusion

From the analysis in Section 3.1 on thiosemicarbazone derivatives, multiple linear regression models involving Sombor, modified Sombor, F_1, F-Sombor, modified F-Sombor, and reduced Sombor leap indices gave better results compared to simple linear models in terms of significant R^2, low $RMSE$, low SE, high F-value, and low p-value for boiling point and molecular weight of thiosemicarbazone derivatives. From the plots related to multiple linear models in Figure 2, it is very clear that the predicted values of BP and MW are not so far from the experimental values. The highlighted index for the simple linear regression model is the Sombor leap index, SL, for both properties. In Section 3.2 related to randomly chosen therapeutic drugs, the modified Sombor leap index, ^{m}SL, is the highlighted index for simple linear regression of all the properties. However, for the boiling point, the combination of FSO_L and $^{red}SO_L$ produced a higher correlation than the model involving ^{m}SL. The best-fit plot, Figure 5, for this model demonstrates its accuracy in accurately predicting the experimental boiling points of these drugs, which closely follow the best-fit line.

From both the sections, it is very clear that the leap indices have shown better correlation with the properties of the dataset having a larger number of compounds compared to the dataset of thiosemicarbazone derivatives. This concludes that the potential scalability of leap indices is higher on larger datasets than on smaller datasets. All the leap indices are calculated for all 46 compounds simultaneously using RDKit and Python programming software, and the time and cost required to calculate them were negligible. This minimal computational cost indicates high computational efficiency of the leap indices for molecular datasets of moderate size and complexity. From this, we conclude that leap indices considered in this study can be used in predicting various physicochemical properties of millions of other therapeutic compounds.

5. Future Scope

This study is exclusively limited to performing QSPR analysis on two datasets consisting of 16 and 30 therapeutic compounds in each. To demonstrate that the obtained QSPR model is both accurate and predictive, validation is essential. Internal validation techniques, such as k-fold cross-validation, Leave-One-Out cross-validation, or external validation, can be used to assess the predictive power of the model obtained. Hence, incorporating such methods can be the future study to ensure that the obtained model is statistically sound and generalizable. These models can be applied to predict the physicochemical properties of larger datasets instead of expensive and time-consuming laboratory experiments.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] A. Ali and N. Trinajstić, A novel/old modification of the first Zagreb index, *Molecular Informatics* **37**(6-7) (2018), 1800008, DOI: 10.1002/minf.201800008.

[2] I. Gutman and N. Trinajstić, Graph theory and molecular orbitals. Total ϕ-electron energy of alternant hydrocarbons, *Chemical Physics Letters* **17**(4) (1972), 535 – 538, DOI: 10.1016/0009-2614(72)85099-1.

[3] S. R. Hill, R. K. Thakur and G. K. Sharma, *Antimalarial Medications*, StatPearls Publishing, Treasure Island, Florida, (2023).

[4] V. R. Kulli, On F-leap indices and F-leap polynomials of some graphs, *International Journal of Mathematical Archive* **9**(12) (2018), 41 – 49, URL: https://ijma.info/index.php/ijma/article/view/5897.

[5] V. R. Kulli, N. Harish and B. Chaluvaraju, Sombor leap indices of some chemical drugs, *Research Review – International Journal of Multidisciplinary* **7**(10) (2022), 158 – 166, DOI: 10.31305/rrijm.2022.v07.i10.019.

[6] D. Maji and G. Ganesh, A novel graph invariant: The third leap Zagreb index under several graph operations, *Discrete Mathematics, Algorithms and Applications* **11**(5) (2019), 1950054, DOI: 10.1142/S179383091950054X.

[7] R. Matsa, P. Makam, M. Kaushik, S. L. Hoti and T. Kannan, Thiosemicarbazone derivatives: Design, synthesis and in vitro antimalarial activity studies, *European Journal of Pharmaceutical Sciences* **137** (2019), 104986, DOI: 10.1016/j.ejps.2019.104986.

[8] H. S. Ramane and K. S. Pise, New results on leap Zagreb indices, *Annals of Mathematics and Computer Science* **15** (2023), 20 – 30, URL: https://annalsmcs.org/index.php/amcs/article/view/169.

[9] M. Randic, Characterization of molecular branching, *Journal of the American Chemical Society* **97**(23) (1975), 6609 – 6615, DOI: 10.1021/ja00856a001.

[10] B. Shakya and P. N. Yadav, Thiosemicarbazones as potent anticancer agents and their modes of action, *Mini-Reviews in Medicinal Chemistry* **20**(8) (2020), 638 – 661, DOI: 10.2174/1389557519666191029130310.

[11] J. Shim, N. Jyothi and N. A. Farook, Biological applications of thiosemicarbazones and their metal complexes, *Asian Journal of Chemistry* **25**(10) (2013), 5838 – 5840, DOI: 10.14233/ajchem.2013.OH105.

[12] B. R. Sri and Y. L. Naidu, On Zagreb connection indices of γ graphyne and its zigzag nanoribbon, *International Journal of Mathematics and Physics* **15**(1) (2024), 83 – 90, DOI: 10.26577/ijmph.2024v15i1a10.