scientific reports

Check for updates

OPEN

Comparative study of Sombor index and its various versions using regression models for top priority polycyclic aromatic hydrocarbons

B. Kirana[1,3], M. C. Shanmukha[2,3]✉ & A. Usha[4]

The aromatic compounds having structural configurations with two or more fused benzene rings are the polycyclic aromatic hydrocarbons (PAHs). Topological indices are valuable tools for studying the structure property relationships of PAHs and also helps in predicting various properties and activities. They find applications widely in computational chemistry, drug design and QSPR studies. This article focuses on analysing the potential predictive index for Sombor index (*SO*), elliptic Sombor index (*ESO*), Euler Sombor index (*EU*), reverse Sombor index (*RSO*), reverse elliptic Sombor index (*RESO*) and reverse Euler Sombor index (*REU*) using regression models for top priority 38 PAHs. From the study it is evident that, *SO* and *RSO* have proved to be potential predictive indices among the considered degree-based and reverse degree-based indices. The variation of best predictive index with minimal *RMSE* are plotted for linear, quadratic and cubic regression models for better understanding.

Keywords Vertex degree-based topological indices, PAHs, Physicochemical properties of PAHs, Regression models

A branch of mathematics, that applies the concept of graph theory to study molecular structures is chemical graph theory(CGT). Using CGT, PAHs are modelled into graphs where atoms are represented as vertices and the bond between the atoms are represented as edges. The connectivity of the atoms are analysed with respect to properties of these graphs where researchers gather information about the structure of PAHs molecules[1–5].

In general, CGT equips a dynamic framework for understanding the structure property relationships in PAHs molecule which is necessary for applications in the fields of design of new materials, monitoring environment and pharmaceuticals. A powerful tool to encode the structural information of PAH molecules in numerical representation is obtained by molecular descriptors. These molecular descriptors may be distance-based, graph spectral-based, and degree-based are used in QSPR/QSAR (quantitative structure property/activity relationship) studies to predict the physicochemical properties of PAH compounds[6–10].

A topological index (TI) is the numerical value associated with the molecular structure of a chemical compound. These indices provide an understanding into various physicochemical properties and biological activities of a molecule[11–18]. Polycyclic aromatic hydrocarbons(PAHs) are composed of multiple aromatic ring having significant applications in the field of pharmacology, materials science and environmental science[19–21]. In the literature, various TIs are studied to characterize the structural properties of PAHs. They include Wiener index (W) which was the oldest and first TI based on distances between pairs of vertices in a graph. This index contributes information about the molecular size and branching for PAHs.

Hyper-Wiener index (WW) is an extension of the Wiener index which is used in the study of large PAH molecules. The molecular connectivity and symmetry in PAHs are provided using Harary Index (H), since this index is based on average distance between pairs of vertices. The Balaban index (J) gives insights into the degree of molecular branching and symmetry of PAHs as this index is based on topological distance matrix. The information about branching and symmetry of PAHs is obtained by using Randic index (R)[22–25].

The main axioms of core chemistry postulates that there is an intricate link between the molecular structure with that of its physical properties. Topological indices are worthy in extracting appropriate details of the

[1]Department of Mathematics, KVG College of Engineering, Sullia 574327, India. [2]Department of Mathematics, PES Institute of Technology and Management, Shivamogga 577204, India. [3]Visvesveraya Technological University, Belagavi 590018, India. [4]Department of Mathematics, Alliance School of Sciences, Alliance University, Bangalore 562106, India. ✉email: mcshanmukha@gmail.com

construction of molecules, thereby proving its applications across diverse fields such as medicinal chemistry, pharmacy, materials science and etc.,[26–33].

In 2021, Gutman introduced Sombor index using Euclidean geometry which has become very popular within a short span of time for its contribution in chemistry and pharmacology. The general form of a vertex degree-based index is a function which is chosen such that, it satisfies symmetry property. The edge uv representation in 2-dimensional coordinate system is called the degree-point of the edge uv. The Euclidean distance between the degree-point (d_u, d_v) and the origin O is $\sqrt{d_u^2 + d_v^2}$ which is the definition of Sombor index[34].

In 2023, Gutman et al., introduced another version of Sombor index called elliptic Sombor index referring to the orbits of planets in the solar system which takes elliptic orbits with the Sun as focus point. In astronomy, the perimeter of an ellipse is of great importance from which elliptic Sombor index was derived[35].

In 2024, Gutman et al., showed that, in ellipse, the lengths of semi major and the semi minor axes are equal. The area of the ellipse was found to be $\pi\sqrt{\sqrt{(d_u^2 + d_v^2)}(d_u + d_v)}$. Leonard Euler found the approximate perimeter of the ellipse as $\pi\sqrt{2(d_u^2 + d_v^2)(d_u + d_v)^2}$. Using these relations, Euler Sombor index was proposed as $\sqrt{d_u^2 + d_v^2 + d_u.d_v}$. Algebraically, there is a geometric analogy of Sombor and Euler Sombor indices[36].

Numerous research work have been carried out on the above indices world wide. The information about graph degree could be relied on the new topological index introduced by Gutman in the year 2021, known as Sombor index. It has been proved that, it holds promise for decoding the thermodynamic behaviour of compounds. Hayat et al.,[37] proposed the minimum Sombor index of graphs while Sakandar et al.[38] employed valency-based indices in QSPR studies for monocarboxylic acids for physicochemical properties.

In a very short time of its existence, the Sombor index has attracted appreciable attention from both chemists as well as mathematicians. Redžepovi'c[39] studied about the alkane's entropy and enthalpy of vaporization by statistical techniques. The mathematical aspects of Sombor index is studied by Gutman et al.[40] giving more insights into the topic. The researchers are continuously studying about the Sombor index in which numerous articles are being published for which the extremal values of the index within graphs are considered as the foundation[41].

Using domination numbers for trees, Sun and Du[42,43] studied for the maximal Sombor index. Zhou et al.[44] used unicyclic graphs and classified the Sombor index with matching number. Li et al.[45] derived the extreme value of Sombor index for trees with fixed diameter. Réti et al.,[46] studied about the maximizing graphs for Sombor index using K-cyclic graphs where K takes the values from 1 to 5.

Narahari et al.[47] introduced a new vertex degree-based index known as reverse Sombor index for which mathematical properties are defined recently. Kulli[48] established some mathematical properties of reverse elliptic Sombor index for two families of dendrimer nanostars.

Carlos et al.[49] recently solved the extremal value problem of elliptic Sombor index with equal number of vertices over the set of chemical graphs and chemical trees. Shanmukha et al.[50] focussed on the chemical applicability of elliptic Sombor index using various benzenoid hydrocarbons through curvilinear regression models.

There has been a momentous progress in the study of the correlation capabilities of several families of graph theoretic descriptors. Gutman and Tosovi'c[51] initiated this study to assess the quality of degree-based indices which was measured by correlation with the physicochemical properties of octane isomers. It was followed by Malik et al.[52] to continue this study for benzenoid hydrocarbons for the characteristics that included total π -electronic energy.

Motivated by the above studies on Sombor index and its various versions, an attempt is made to study degree-based indices: Sombor index, elliptic Sombor index, Euler Sombor index and its reverse degree-based indices. To establish the potential index with respect to various physicochemical properties of top priority 38 PAHs using regression models is carried out.

This article mainly concentrates on

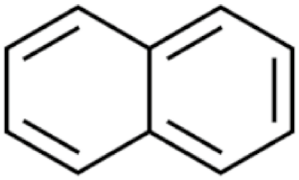
- Identifying the potential vertex degree-based topological index in the considered indices with respect to physicochemical properties of top priority 38 PAH's: Sombor index, elliptic Sombor index, Euler Sombor index and its Reverse degree-based indices.
- To check the potential index, statistical analysis is carried out using regression models.
- We employ RMSE measure to find the minimal error between the set of actual values and the predicted values.
- Based on the obtained RMSE values in this work, we opt for minimal RMSE value which signifies minimal error between the actual and predicted values.
- For better understanding of statistical analysis, a scatter diagram is depicted for linear regression model and is extended for quadratic and cubic regression models with minimal RMSE to notice the variation.

Methodology

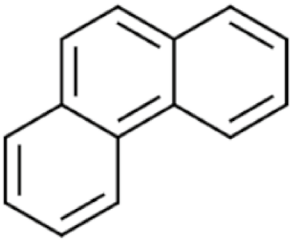
G=(V, E) is a simple graph with V as vertex set and E as edge set. For a vertex u belonging to V, d_u indicates the degree of the vertex u[53,54]. In this work, the top priority 38 PAHs (Fig. 1) are modeled as molecular graphs for which 6 vertex degree-based topological indices are computed.

The considered degree-based indices such as Sombor index, elliptic Sombor index and Euler Sombor index[34–36] are defined as follows

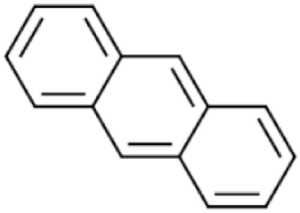
Benzene



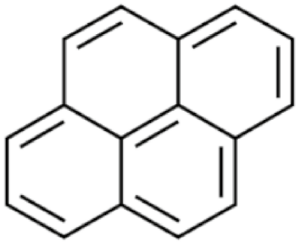
Naphthalene



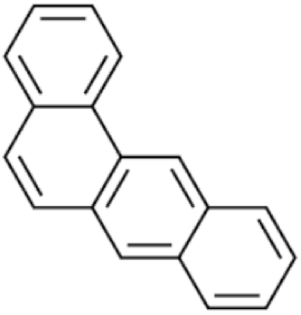
Phenanthrene



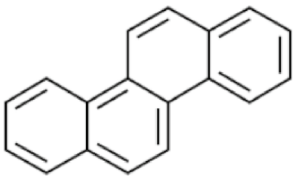
Anthracene



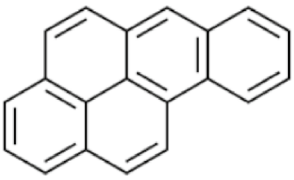
Pyrene



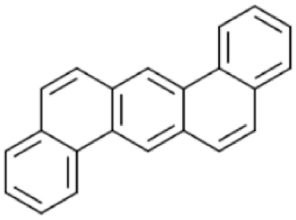
Benzo[a]anthracene



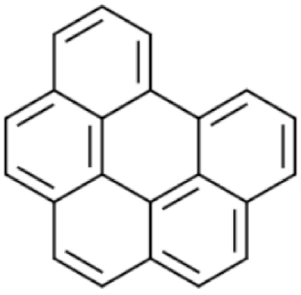
Chrysene



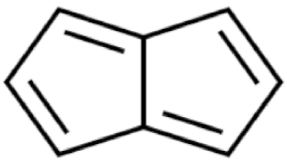
Benzo[a]pyrene



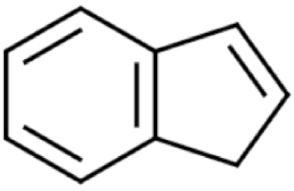
Dibenz[a,h]anthracene



Benzo[g,h,i]perylene

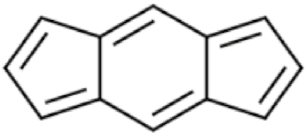


Pentalene

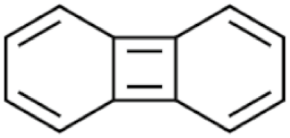


Indene

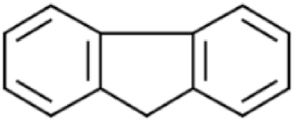
Figure 1. Molecular structures of the top priority PAHs.



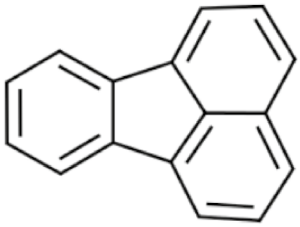
s-indacene



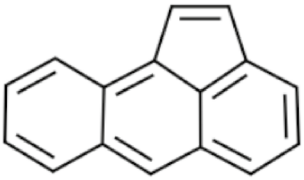
Biphenylene



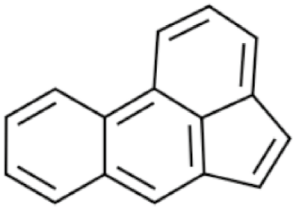
Fluorene



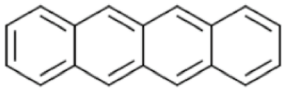
Fluoranthene



Aceanthrylene

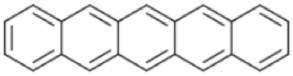


Acephenanthrylene

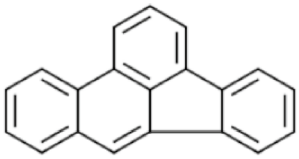


Naphthacene

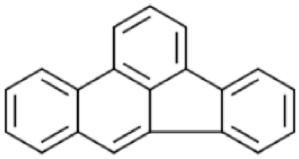
Corannulene



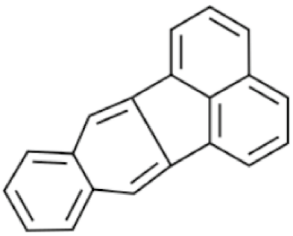
Pentacene



Ovalene

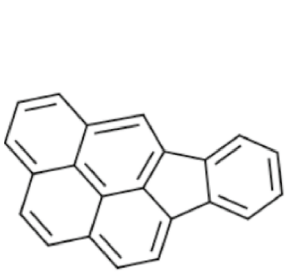


Benzo[b]fluoranthene

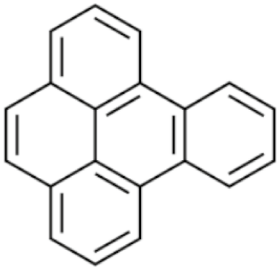


Benzo[k]fluoranthene

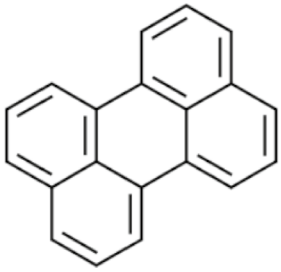
Figure 1. (continued)



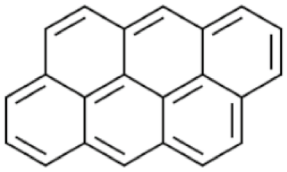
Indeno[1,2,3-cd]pyrene



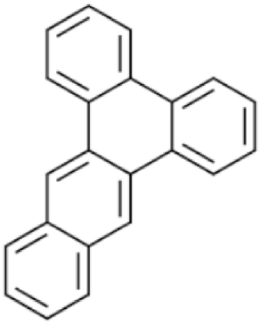
Benzo[e]pyrene



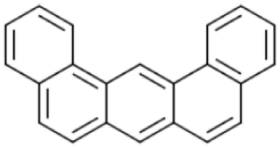
Perylene



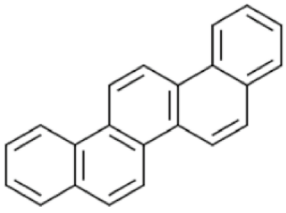
Anthanthrene



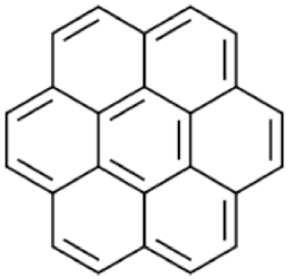
Dibenz[a,c]anthracene



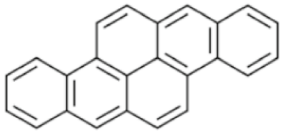
Dibenz[a,j]anthracene



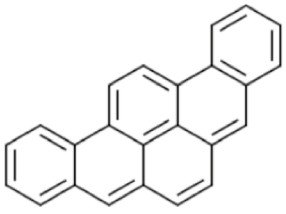
Picene



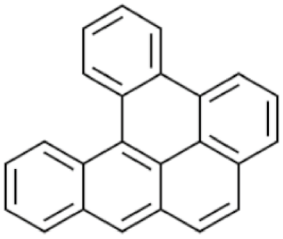
Coronene



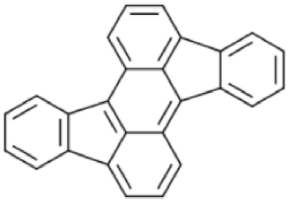
Dibenzo[a,h]pyrene



Dibenzo[a,i]pyrene

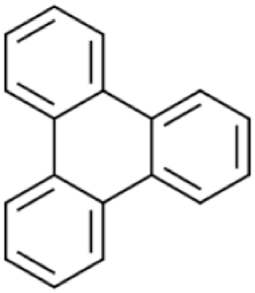


Dibenzo[a,l]pyrene

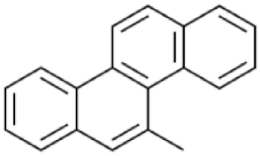


Rubicene

Figure 1. (continued)



isochrysene



5-Methylchrysene

Figure 1. (continued)

$$SO(G) = \sum_{uv \in E(G)} \sqrt{d_u^2 + d_v^2}\,, ESO(G) = \sum_{uv \in E(G)} (d_u + d_v)(\sqrt{d_u^2 + d_v^2})$$
$$EU(G) = \sum_{uv \in E(G)} \sqrt{d_u^2 + d_v^2 + d_u d_v}.$$

Reverse degree-based indices such as reverse Sombor index, reverse elliptic Sombor index[47,48] are defined as follows

$$RSO(G) = \sum_{uv \in E(G)} \sqrt{c_u^2 + c_v^2},$$
$$RESO(G) = \sum_{uv \in E(G)} (c_u + c_v)(\sqrt{c_u^2 + c_v^2}).$$

An attempt is made to define reverse Euler Sombor index and is defined as

$$REU(G) = \sum_{uv \in E(G)} \sqrt{c_u^2 + c_v^2 + c_u c_v}.$$

where $c_u = \Delta - d_u + 1$ for any vertex $u \in E(G)$ and Δ is the maximum vertex degree of the graph G.

Results and discussions

The evaluation of regression models is conducted as follows,

$$y = a + b_1 x_1; \quad n,\ r,\ F\ (Linear) \tag{1}$$

$$y = a + b_1 x_2 + b_2 x_2^2; \quad n,\ r,\ F\ (Quadratic) \tag{2}$$

$$y = a + b_1 x_3 + b_2 x_3^2 + b_3 x_3^3; \quad n,\ r,\ F\ (Cubic) \tag{3}$$

Here, y is the dependent variable. a being the regression constant and b_i where$(i = 1, 2, 3)$ are the regression coefficients and $x_i (i = 1, 2, 3)$ are the independent variables. Samples are used for regression equation, r being the correlation coefficient, SE is the standard error of the estimates and F is the Fisher's statistic.

RMSE plays a vital role to understand the behaviour of statistical models to evaluate the accuracy of regression models and is used to measure the difference between the actual values and the predicted values. It is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{n}}$$

Here, n denotes the number of data points. y_i is the actual value for the i^{th} data point, $\hat{y}_i$ is the predicted value for the i^{th} data point.

Regression models

In this study, 8 physicochemical properties (Table 1) of PAHs are considered such as molecular weight($MW\ g/mol$), melting point($MP\ ^\circ C$), boiling point($BP\ ^\circ C$), molar refractivity($MR\ cm^3$), polarizability

Name	MW	MP	BP	MR	PO	MV	FP	C
Benzene	78.11	5	78.8	26.3	10.4	89.4	−11.1	15.5
Naphthalene	128.17	81	221.5	44.1	17.5	123.5	78.9	80.6
Phenanthrene	178.23	100	337.4	61.9	24.6	157.7	146.6	174
Anthracene	178.23	217	337.4	61.9	24.6	157.7	146.6	154
Pyrene	202.25	150	404	72.5	28.7	162	168.8	217
Benzo[a]anthracene	228.3	161	436.7	79.8	31.6	191.8	209.1	294
Chrysene	228.3	255	448	79.8	31.6	191.8	209.1	264
Benzo[a]pyrene	252.3	176	495	90.3	35.8	196.1	228.6	372
Dibenz[a,h]anthracene	278.3	266	524.7	97.6	38.7	225.9	264.5	361
Benzo[g,h,i]perylene	276.3	273	501	100.8	40	200.4	247.2	411
Pentalene	102	–	308.9	34.2	13.6	96.1	97.2	174
Indene	116.16	−2	181.6	38	15.1	111.8	58.9	124
s-indacene	152.19	–	525.3	50	19.8	129.9	214.4	341
Biphenylene	152.19	115	469.9	50	19.8	129.9	187.2	339
Fluorene	166.22	115	293.6	53.8	21.3	148.3	133.1	165
Fluoranthene	202.25	110	375	72.5	28.7	162	168.4	243
Aceanthrylene	154.21	94	279	51.7	20.5	134.9	135.3	155
Acephenanthrylene	202.25	140	405.7	69.1	27.4	162.3	188.6	303
Naphthacene	228.3	350	436.7	79.8	31.6	53.5	209.1	236
Corannulene	250.3	269	438	93.5	37.1	170.6	210.1	303
Pentacene	278.3	257	524.7	97.6	38.7	225.9	264.5	325
Ovalene	398.5	257	524.7	97.6	38.7	225.9	264.5	696
Benzo[b]fluoranthene	252.3	166	467.5	90.3	35.8	196.1	228.6	372
Benzo[k]fluoranthene	252.3	217	480	90.3	35.8	196.1	228.6	338
Indeno[1,2,3-cd]pyrene	276.3	163.6	497.1	100.8	40	200.4	247.2	453
Benzo[e]pyrene	252.3	177.5	467.5	90.3	35.8	196.1	228.6	336
Perylene	252.3	276	467.5	90.3	35.8	196.1	228.6	304
Anthanthrene	276.3	261	497.1	100.8	40	200.4	247.2	411
Dibenz[a,c]anthracene	278.3	206	518	97.6	38.7	225.9	264.5	361
Dibenz[a,j]anthracene	278.3	196	524.7	97.6	38.7	225.9	264.5	363
Picene	278.3	367	519	97.6	38.7	225.9	264.5	361
Coronene	300.4	440	525.6	111.4	44.1	204.7	265.2	376
Dibenzo[a,h]pyrene	302.4	308	552.3	108.1	42.9	230.2	282	436
Dibenzo[a,i]pyrene	302.4	281.5	552.3	108.1	42.9	230.2	282	436
Dibenzo[a,l]pyrene	302.4	162.4	552.3	108.1	42.9	230.2	282	480
Rubicene	326.4	306	579	118.7	47	234.5	298.8	514
isochrysene	228.3	199	425	79.8	31.6	191.8	209.1	217
5-Methylchrysene	242.3	118	449.4	84.6	33.5	208.1	217.8	320

Table 1. Physicochemical properties of top priority 38 PAHs.

($PO\ \ 10^{-24}cm^3$), molar volume($MV\ \ cm^3$), flash point($FP\ \ ^\circ C$), complexity (C) for which coefficient of correlation are calculated using the computed values of TIs (Table 2).

From Table 3, it is obvious that of all the 3 degree-based indices considered in the study *SO* has high correlation with 7 (MW, BP, MR, PO, MV, FP, C) properties out of 8 properties considered while *ESO* has high correlation with the property *MP*. From Table 4, it is obvious that of all the 3 reverse degree-based indices considered in the study *RSO* has high correlation with 7 (MW, MP, BP, MR, PO, FP, C) properties out of 8 properties considered while *RESO* has high correlation with the property *MV*.

Linear regression model: Degree-based TIs are studied using linear regression models for Eq. (1), are as follows

Drug name/topological index	SO	ESO	EU	RSO	RESO	REU
Benzene	16.970	20.785	67.882	16.940	67.881	20.784
Naphthalene	35.635	43.416	165.447	27.328	97.542	33.099
Phenanthrene	54.159	65.99	263.727	37.457	127.928	45.318
Anthracene	54.299	66.047	263.013	37.686	127.202	45.414
Pyrene	67.027	81.635	339.379	41.929	135.687	50.61
Benzo[a]anthracene	72.684	88.621	361.293	47.815	157.588	57.633
Chrysene	72.684	88.564	362.007	47.586	158.314	57.538
Benzo[a]pyrene	85.552	88.678	360.578	52.058	166.074	62.829
Dibenz[a,h]anthracene	76.927	93.76	387.462	57.944	187.975	69.853
Benzo[g,h,i]perylene	98.279	119.798	514.026	56.3	174.559	68.025
Pentalene	29.978	36.488	142.82	21.672	74.914	26.171
Indene	32.807	39.952	154.134	24.500	86.228	29.635
s-indacene	48.643	59.12	240.386	32.03	104.575	38.486
Biphenylene	48.363	59.004	241.814	31.571	106.027	38.295
Fluorene	51.331	62.526	252.413	34.629	116.614	41.854
Fluoranthene	66.887	81.578	340.094	41.7	136.413	50.514
Aceanthrylene	67.027	81.635	339.379	41.929	135.687	50.61
Acephenanthrylene	67.027	81.635	339.379	41.929	135.687	50.61
Naphthacene	72.964	88.678	394.519	48.045	156.862	57.729
Corannulene	88.380	107.673	465.943	49.229	149.103	59.365
Pentacene	95.234	115.668	476.171	60.639	193.231	72.689
Ovalene	155.268	189.251	839.978	79.616	228.949	96.023
Benzo[b]fluoranthene	85.412	104.152	438.373	51.828	166.8	62.734
Benzo[k]fluoranthene	85.412	104.152	438.374	52.058	166.074	62.829
Indeno[1,2,3-cd]pyrene	98.279	119.798	514.026	56.3	174.559	68.025
Benzo[e]pyrene	85.412	104.152	438.373	51.828	166.8	62.734
Perylene	85.412	104.152	438.373	51.828	166.8	62.734
Anthanthrene	98.419	119.855	513.312	56.530	173.833	68.121
Dibenz[a,c]anthracene	91.209	111.137	460.286	57.715	188.701	69.757
Dibenz[a,j]anthracene	76.927	93.759	387.462	57.944	187.975	69.853
Picene	91.209	111.137	460.286	57.715	188.701	69.757
Coronene	111.147	135.443	589.679	60.773	182.318	73.317
Dibenzo[a,h]pyrene	104.077	126.783	535.939	62.187	196.46	75.049
Dibenzo[a,i]pyrene	99.834	121.587	510.484	60.772	193.632	73.317
Dibenzo[a,l]pyrene	103.937	126.726	536.653	61.957	197.186	74.953
Rubicene	116.53	142.256	613.748	65.971	206.397	80.054
isochrysene	72.546	88.506	362.73	47.356	206.397	80.054
5-Methylchrysene	77.264	93.901	388.807	49.335	162.481	59.412

Table 2. Computed PAHs values of Sombor index, elliptic Sombor index, Euler Sombor index and its reverse degree-based indices.

Index	MW	MP	BP	MR	PO	MV	FP	C
SO	**0.9713**	0.7009	**0.7978**	**0.903**	**0.9029**	**0.7786**	**0.8595**	**0.8994**
ESO	0.9606	**0.7041**	0.7801	0.886	0.886	0.7441	0.8425	0.894
EU	0.9684	0.7039	0.7916	0.8988	0.8987	0.775	0.775	0.8944

Table 3. The correlation coefficient r between degree-based TIs and physicochemical properties of PAHs. Significant values are in bold.

Index	MW	MP	BP	MR	PO	MV	FP	C
RSO	**0.9915**	**0.7298**	**0.8402**	**0.9446**	**0.9446**	0.8418	**0.9114**	**0.8727**
RESO	0.956	0.7078	0.8176	0.9267	0.9266	**0.8506**	0.8974	0.7938
REU	0.9664	0.7106	0.8157	0.924	0.9212	0.8332	0.8909	0.8261

Table 4. The correlation coefficient r between reverse degree-based TIs and Physicochemical properties of PAHs. Significant values are in bold.

r^2	F	SE	RMSE	Significant
0.943	599.376	16.437	15.999	0.000
0.496	33.435	69.709	67.745	0.000
0.636	63.022	68.764	66.930	0.000
0.815	158.945	10.358	10.081	0.000
0.815	158.921	4.107	3.997	0.000
0.606	35.405	28.973	28.2	0.000
0.739	101.832	35.678	34.727	0.000
0.809	152.383	57.756	56.215	0.000

Table 5. Various Statistical parameters of linear regression models.

$$
\begin{aligned}
MW &= 41.778 + 2.463(SO),\\
MP &= 8.472 + 0.476(ESO),\\
BP &= 178.784 + 3.342(SO),\\
MR &= 19.11 + 0.799(SO),\\
PO &= 7.67 + 0.317(SO),\\
MV &= 7.824 + 1.32(SO),\\
FP &= 36.234 + 2.204(SO),\\
C &= -26.6 + 4.364(SO).
\end{aligned}
$$

Linear regression model

Reverse degree-based TIs are studied using linear regression models for equation (1), are as follows

$$
\begin{aligned}
MW &= 5.061(RSO) - 11.62,\\
MP &= 5.597(RSO) - 75.626,\\
BP &= 7.082(RSO) + 95.856,\\
MR &= 1.683(RSO) - 0.184,\\
PO &= 0.667(RSO) - 0.074,\\
MV &= 0.999(RESO) + 23.564,\\
FP &= 4.703(RSO) - 19.988,\\
C &= 8.522(RSO) - 99.813.
\end{aligned}
$$

Conclusion

This work concentrates on analysing the potential predictive index for *SO*, *ESO*, *EU*, *RSO*, *RESO* and *REU* using regression models for top priority 38 PAHs. From the results, it is evident that Sombor index shows high correlation with considered physical properties compared to that of newly introduced elliptic Sombor and Euler Sombor indices. For best predictive index the minimal RMSE value is considered. From the analysis, it is clear that *SO* is the best predictive index with minimal RMSE from the considered degree-based indices (Tables 5, 6). *RSO* is the best predictive index with minimal RMSE from the considered reverse degree-based indices (Tables 7, 8). The variation of best predictive indices with minimal RMSE are plotted for linear, quadratic and cubic regression models for better understanding (Figs. 2, 3). This study may be useful for the researchers who wish to study further about PAHs and also about the applications of the considered indices.

Name	*MW*	*MP*	*BP*	*MR*	*PO*	*MV*	*FP*	*C*
Benzene	83.581	40.786	235.49	32.675	12.954	100.226	73.633	47.461
Naphthalene	129.559	87.231	297.861	47.595	18.869	124.866	114.768	128.919
Phenanthrene	175.193	134.016	359.764	62.403	24.74	149.322	155.595	209.767
Anthracene	175.537	133.676	360.231	62.515	24.784	149.506	155.903	210.378
Pyrene	206.888	170.03	402.76	72.688	28.818	166.307	183.952	265.921
Benzo[a]anthracene	220.824	180.461	421.663	77.21	30.611	173.775	196.419	290.61
Chrysene	220.824	180.801	421.663	77.21	30.611	173.775	196.419	290.61
Benzo[a]pyrene	252.522	180.121	464.662	87.496	34.689	190.763	224.778	346.769
Dibenz[a,h]anthracene	231.276	192.918	435.841	80.601	31.955	179.377	205.77	309.127
Benzo[g,h,i]perylene	283.875	253.168	507.194	97.67	38.722	207.565	252.829	402.316
Pentalene	115.624	–	278.958	43.073	17.076	117.398	102.301	104.231
Indene	122.593	81.845	288.411	45.334	17.973	121.133	108.536	116.577
s-indacene	161.602	–	341.328	57.993	22.991	142.038	143.436	185.689
Biphenylene	160.913	123.584	340.392	57.769	22.903	141.669	142.819	184.467
Fluorene	168.224	128.63	350.31	60.141	23.843	145.587	149.36	197.42
Fluoranthene	206.544	170.369	402.292	72.576	28.773	166.123	183.643	265.31
Aceanthrylene	206.888	170.03	402.76	72.688	28.818	166.307	183.952	265.921
Acephenanthrylene	206.888	170.03	402.76	72.688	28.818	166.307	183.952	265.921
Naphthacene	221.513	196.278	422.599	77.434	30.699	174.145	197.036	291.832
Corannulene	259.488	230.279	474.112	89.756	35.585	194.496	231.011	359.111
Pentacene	276.372	235.148	497.015	95.235	37.757	203.544	246.116	389.023
Ovalene	424.257	408.334	697.624	143.223	56.783	282.796	378.423	651.025
Benzo[b]fluoranthene	252.177	217.155	464.194	87.384	34.644	190.578	224.47	346.158
Benzo[k]fluoranthene	252.177	217.155	464.194	87.384	34.644	190.578	224.47	346.158
Indeno[1,2,3-cd]pyrene	283.875	253.168	507.194	97.67	38.722	207.565	252.829	402.316
Benzo[e]pyrene	252.177	217.155	464.194	87.384	34.644	190.578	224.47	346.158
Perylene	252.177	217.155	464.194	87.384	34.644	190.578	224.47	346.158
Anthanthrene	284.22	252.828	507.662	97.782	38.767	207.75	253.138	402.927
Dibenz[a,c]anthracene	266.457	227.586	483.566	92.018	36.482	198.231	237.246	371.457
Dibenz[a,j]anthracene	231.276	192.918	435.841	80.601	31.955	179.377	205.77	309.127
Picene	266.457	227.586	483.566	92.018	36.482	198.23	237.246	371.457
Coronene	315.574	289.182	550.193	107.956	42.801	224.552	281.189	458.475
Dibenzo[a,h]pyrene	298.156	263.599	526.565	102.303	40.56	215.218	265.605	427.616
Dibenzo[a,i]pyrene	287.704	251.482	512.387	98.912	39.215	209.617	256.254	409.098
Dibenzo[a,l]pyrene	297.811	263.939	526.097	102.192	40.515	215.033	265.297	427.005
Rubicene	328.832	300.64	568.178	112.258	44.506	231.657	293.05	481.964
isochrysene	220.484	181.145	421.202	77.099	30.567	173.593	196.115	290.007
5-Methylchrysene	232.106	193.559	436.967	80.871	32.062	179.822	206.513	310.598

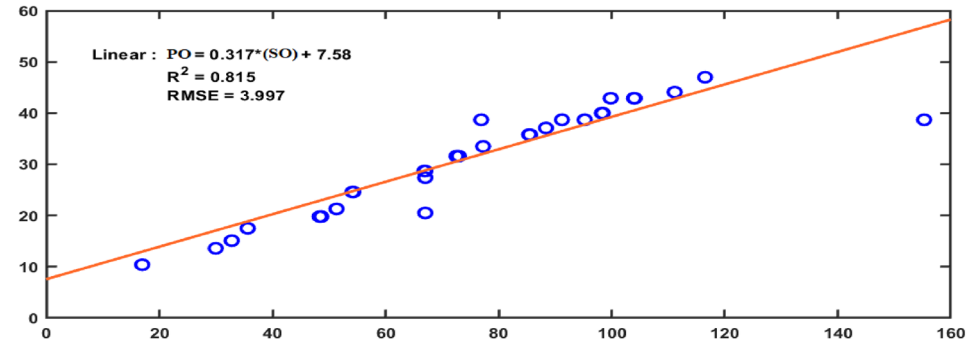
Table 6. Predicted values of physicochemical properties of PAHs from linear regression model with minimal RMSE:$PO = 7.67 + 0.317(SO)$.

r^2	*F*	*SE*	*RMSE*	*Significant*
0.983	2083.791	8.999	8.759	0.000
0.533	38.74	67.119	65.228	0.000
0.706	86.415	61.846	60.196	0.000
0.892	298.432	7.908	7.697	0.000
0.892	298.44	3.135	3.051	0.000
0.723	94.182	24.278	23.630	0.000
0.831	176.53	28.732	27.965	0.000
0.762	114.998	64.511	62.7902	0.000

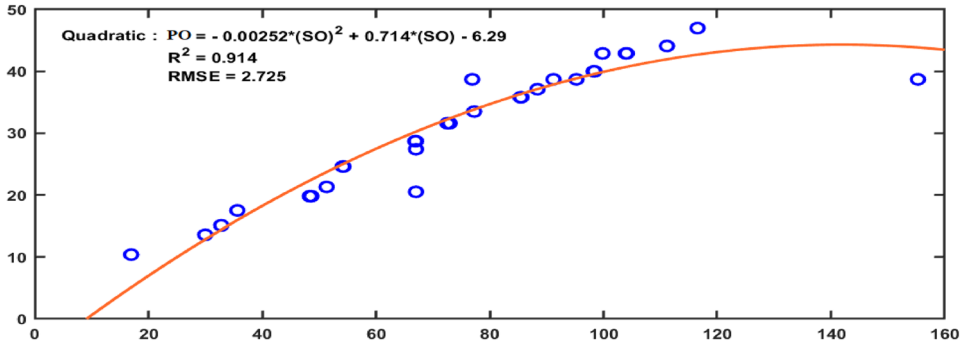
Table 7. Various statistical parameters of linear regression model.

Name	*MW*	*MP*	*BP*	*MR*	*PO*	*MV*	*FP*	*C*
Benzene	74.106	19.185	215.831	28.325	11.228	91.381	59.676	44.552
Naphthalene	126.676	77.326	289.403	45.807	18.16	121.014	108.528	133.08
Phenanthrene	177.935	134.018	361.14	62.853	24.918	151.371	156.161	219.402
Anthracene	179.094	135.3	362.762	63.239	25.071	150.646	157.238	221.353
Pyrene	200.566	159.047	392.813	70.379	27.902	159.123	177.192	257.513
Benzo[a]anthracene	230.353	191.991	434.499	80.285	31.83	181.004	204.872	307.675
Chrysene	229.194	190.709	432.878	79.9	31.677	181.729	203.795	305.723
Benzo[a]pyrene	251.825	215.739	464.55	87.426	34.661	189.482	224.826	343.834
Dibenz[a,h]anthracene	281.612	248.682	506.237	97.331	38.588	211.362	252.506	393.996
Benzo[g,h,i]perylene	273.292	239.481	494.593	94.565	37.491	197.959	244.775	379.985
Pentalene	98.053	–	249.345	36.288	14.386	98.407	81.929	84.879
Indene	112.364	61.498	269.374	41.048	16.273	109.71	95.228	108.98
s-indacene	150.471	–	322.704	53.72	21.297	128.04	130.64	173.152
Biphenylene	148.148	101.074	319.453	52.94	20.991	129.491	128.481	169.24
Fluorene	163.623	118.19	341.111	58.094	23.031	140.068	142.862	195.301
Fluoranthene	199.407	157.766	391.191	69.994	27.75	159.848	176.115	255.561
Aceanthrylene	200.566	159.047	392.813	70.379	27.902	159.123	177.192	257.513
Acephenanthrylene	200.566	159.047	392.813	70.379	27.902	159.123	177.192	257.513
Naphthacene	231.517	193.278	436.128	80.672	31.983	180.278	205.954	309.635
Corannulene	237.508	199.905	444.514	82.665	32.773	172.52	211.522	319.725
Pentacene	295.25	263.766	525.324	101.867	40.387	216.613	265.18	416.963
Ovalene	391.285	369.979	659.726	133.804	53.049	252.298	354.423	578.689
Benzo[b]fluoranthene	250.661	214.451	462.921	87.039	34.508	190.207	223.744	341.874
Benzo[k]fluoranthene	251.825	215.739	464.55	87.426	34.661	189.482	224.826	343.834
Indeno[1,2,3-cd]pyrene	273.292	239.481	494.593	94.565	37.491	197.959	244.775	379.985
Benzo[e]pyrene	250.661	214.451	462.921	87.039	34.508	190.207	223.744	341.874
Perylene	250.661	214.451	462.921	87.039	34.508	190.207	223.744	341.874
Anthanthrene	274.456	240.768	496.222	94.952	37.645	197.233	245.856	381.945
Dibenz[a,c]anthracene	280.453	247.401	504.615	96.946	38.436	212.087	251.429	392.044
Dibenz[a,j]anthracene	281.612	248.682	506.237	97.331	38.588	211.362	252.506	393.996
Picene	280.453	247.401	504.615	96.946	38.436	212.087	251.429	392.044
Coronene	295.928	264.516	526.273	102.092	40.476	205.71	265.81	418.105
Dibenzo[a,h]pyrene	303.084	272.43	536.287	104.472	41.42	219.839	272.46	430.155
Dibenzo[a,i]pyrene	295.923	264.51	526.266	102.091	40.475	217.014	265.805	418.097
Dibenzo[a,l]pyrene	301.92	271.143	534.658	104.085	41.266	220.564	271.378	428.195
Rubicene	322.233	293.609	563.087	110.84	43.944	229.767	290.255	462.403
isochrysene	228.03	189.422	431.249	79.513	31.524	229.767	202.714	303.763
5-Methylchrysene	238.045	200.498	445.265	82.843	32.844	185.892	212.02	320.628

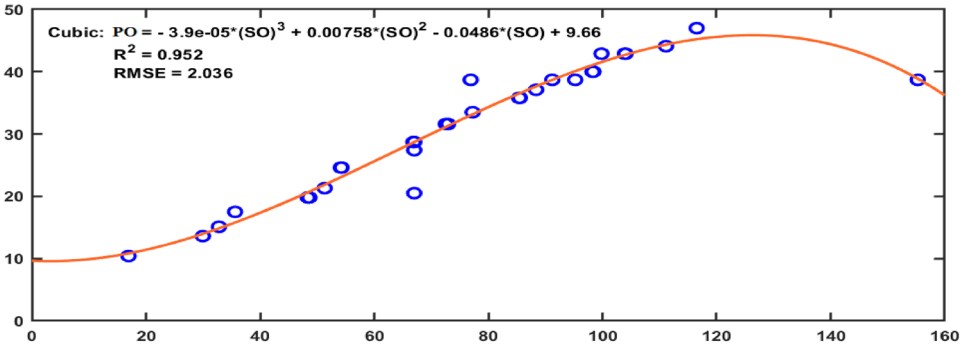
Table 8. Predicted values of physicochemical properties of PAHs from linear regression model with minimal RMSE:$PO = 0.667(RSO) - 0.074$.



(a)

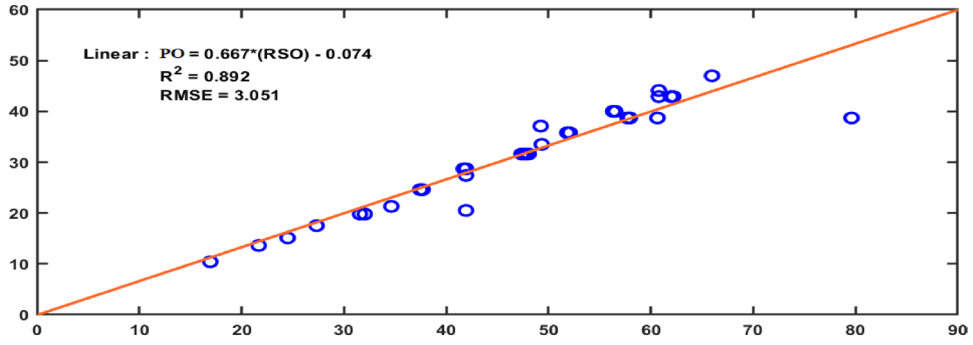


(b)

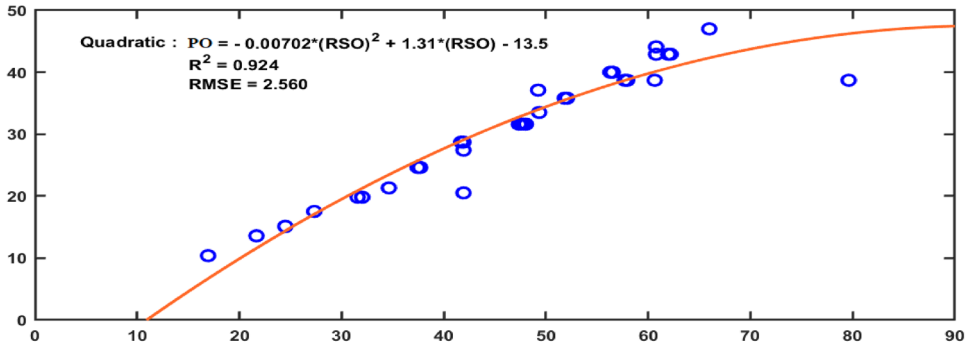


(c)

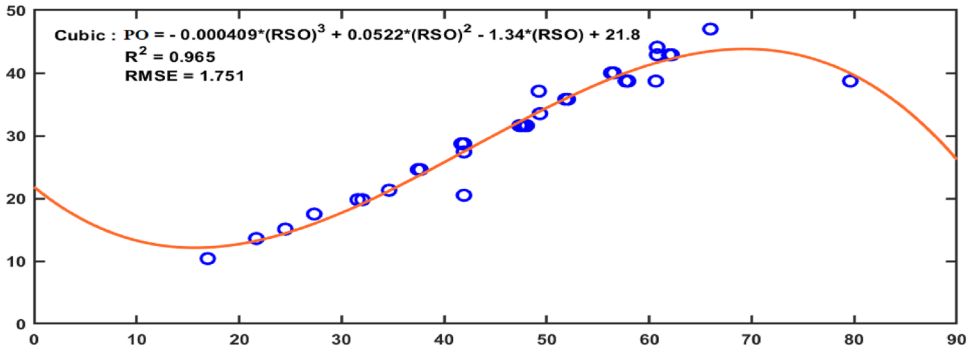
Figure 2. Scatter diagrams of property PO with Sombor index: linear, quadratic, cubic.



(a)



(b)



(c)

Figure 3. Scatter diagrams of property PO with reverse Sombor index: linear, quadratic, cubic.

Data availability

The data used to support the working are cited within the text as references.

Received: 6 May 2024; Accepted: 5 August 2024
Published online: 27 August 2024

References

1. Trinajstic, N. *Chemical Graph Theory* (CRC Press, 1992).
2. Farahani, M. R. Some connectivity indices of polycyclic aromatic hydrocarbons (PAHs). *Adv. Mater. Corros.* **1**, 65–69 (2013).
3. Ahmed, W., Zaman, S. & Khalid, F. Analyzing polycyclic aromatic hydrocarbons using topological indices and QSPR analysis to reveal molecular complexity. *J. Micromech. Mol. Phys.* **9**, 25 (2024).

4. Öztürk Sözen, E., Eryaşar, E. & Çakmak, Ş. Szeged-like topological descriptors and COM-polynomials for graphs of some Alzheimer's agents. *Mol. Phys.* **122**, e2305853 (2024).
5. Rasheed, M. W., Mahboob, A. & Hanif, I. An estimation of physicochemical properties of heart attack treatment medicines by using molecular descriptor's. *S. Afr. J. Chem. Eng.* **45**, 20–29 (2023).
6. Hayat, S., Khan, A., Ali, K. & Liu, J. B. Structure-property modeling for thermodynamic properties of benzenoid hydrocarbons by temperature-based topological indices. *Ain Shams Eng. J.* **15**(3), 102586 (2024).
7. Hayat, S., Alanazi, S. J. & Liu, J. B. Two novel temperature-based topological indices with strong potential to predict physicochemical properties of polycyclic aromatic hydrocarbons with applications to silicon carbide nanotubes. *Phys. Scr.* **99**, 055027 (2024).
8. Prabhu, S., Murugan, G., Arockiaraj, M., Arulperumjothi, M. & Manimozhi, V. Molecular topological characterization of three classes of polycyclic aromatic hydrocarbons. *J. Mol. Struct.* **1229**, 129501 (2021).
9. Zhao, D. *et al.* On molecular descriptors of polycyclic aromatic hydrocarbon. *Polycyclic Aromat. Compd.* **42**(6), 3422–3433 (2022).
10. Lokesha, V., Deepika, T. & Cangul, I. N. Symmetric division Deg and inverse Sum Indeg indices of polycyclic aromatic hydrocarbons (PAHs) and polyhex nanotubes. *Southeast Asian Bull. Math.* **41**(5), 707 (2017).
11. Zhou, H. *et al.* On QSPR analysis of molecular descriptor and thermodynamic features of narcotic drugs. *Polycycl. Aromat. Compd.* **44**, 1–21 (2023).
12. Öztürk Sözen, E. & Eryaşar, E. An algebraic approach to calculate some topological coindices and QSPR analysis of some novel drugs used in the treatment of breast cancer. *Polycycl. Aromat. Compd.* **44**(4), 2226–2243 (2024).
13. Furtula, B. & Gutman, I. A forgotten topological index. *J. Math. Chem.* **53**(4), 1184–1190 (2015).
14. Estrada, E. & Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **8**(13), 1573–1588 (2001).
15. Hosoya, H. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **44**(9), 2332–2339 (1971).
16. Mondal, S., De, N. & Pal, A. Topological properties of Graphene using some novel neighborhood degree-based topological indices. *Int. J. Math. Ind.* **11**(01), 1950006 (2019).
17. Mahboob, A., Rasheed, M. W., Amin, L. & Hanif, I. A study of novel molecular descriptors and quantitative structure-property relationship analysis of blood cancer drugs. *Eur. Phys. J. Plus* **138**(9), 856 (2023).
18. Öztürk Sözen, E. & Eryaşar, E. Graph theoretic approach for calculation of new Banhatti indices VIA recent algebraic polynomials with a chemical application. *Mol. Phys.* **121**(24), e2245072 (2023).
19. Sahoo, B. M., Ravi Kumar, B. V., Banik, B. K. & Borah, P. Polyaromatic hydrocarbons (PAHs): structures, synthesis and their biological profile. *Curr. Org. Synth.* **17**(8), 625–640 (2020).
20. Huang, M. & Penning, T. M. Processing contaminants: Polycyclic aromatic hydrocarbons (PAHs). *Encycl. Food Saf.* **2**, 416–423 (2014).
21. Adeniji, A. O., Okoh, O. O. & Okoh, A. I. Analytical methods for polycyclic aromatic hydrocarbons and their global trend of distribution in water and sediment: A review. In *Recent Insights in Petroleum Science and Engineering* (ed. Zoveidavianpoor, M.) (Intech, 2018).
22. Li, X. *et al.* Computing a closed formula of the Wiener index of the Polycyclic Aromatic Hydrocarbons PAHk by using the cut method. *J. Comput. Theor. Nanosci.* **14**(7), 3636–3640 (2017).
23. Rezaei, M., Jamil, M. K. & Foruzanfar, Z. On the terminal wiener indices of polycyclic aromatic hydrocarbons PAHs. *Int. J. Pure Appl. Math.* **113**(1), 49–57 (2017).
24. Farahani, M. R. Hosoya, Schultz, modified Schultz polynomials and their topological indices of benzene molecules: First members of polycyclic aromatic hydrocarbons (PAHs). *Int. J. Theor. Chem.* **1**(2), 09–16 (2013).
25. Randić, M. Aromaticity of polycyclic conjugated hydrocarbons. *Chem. Rev.* **103**(9), 3449–3606 (2003).
26. Huang, R., Mahboob, A., Rasheed, M. W., Alam, S. M. & Siddiqui, M. K. On molecular modeling and QSPR analysis of lyme disease medicines via topological indices. *Eur. Phys. J. Plus* **138**(3), 243 (2023).
27. Öztürk Sözen, E. & Eryaşar, E. QSPR analysis of some drug candidates investigated for COVID-19 via new topological coindices. *Polycyc. Aromat. Compd.* **44**(2), 1291–1308 (2024).
28. Roy, Kunal. Topological descriptors in drug design and modeling studies. *Mol. Divers.* **8**, 321–323 (2004).
29. Gao, W., Wang, W. & Farahani, M. R. Topological indices study of molecular structure in anticancer drugs. *J. Chem.* **2016**, 3216327 (2016).
30. Gao, W., Siddiqui, M. K., Imran, M., Jamil, M. K. & Farahani, M. R. Forgotten topological index of chemical structure in drugs. *Saudi Pharm. J.* **24**(3), 258–264 (2016).
31. Aslam, A., Bashir, Y., Ahmad, S. & Gao, W. On topological indices of certain dendrimer structures. *Zeitschrift für Naturforschung A* **72**(6), 559–566 (2017).
32. Zaman, S., Jalani, M., Ullah, A., Ahmad, W. & Saeedi, G. Mathematical analysis and molecular descriptors of two novel metal-organic models with chemical applications. *Sci. Rep.* **13**(1), 5314 (2023).
33. Nadeem, M. F., Azeem, M. & Siddiqui, H. M. A. Comparative study of zagreb indices for capped, semi-capped, and uncapped carbon nanotubes. *Polycycl. Aromat. Compd.* **42**(6), 3545–3562 (2022).
34. Gutman, I., Furtula, B., & Elphick, C. (2014). Three new/old vertex-degree-based topological indices. MATCH communications in mathematical and in computer chemistry.
35. Gutman, I., Furtula, B. & Oz, M. S. Geometric approach to vertex-degree-based topological indices-Elliptic Sombor index, theory and application. *Int. J. Quantum Chem.* **124**(2), e27346 (2024).
36. Ivan, G. Relating Sombor and Euler indices. *Vojnotehnički glasnik* **72**(1), 1–12 (2024).
37. Hayat, S., Arshad, M., & Das, K. C. (2022). On the Sombor index of graphs with given connectivity and number of bridges. arXiv preprint arXiv:2208.09993.
38. Hayat, S., Arshad, M. & Khan, A. Graphs with given connectivity and their minimum Sombor index having applications to QSPR studies of monocarboxylic acids. *Heliyon* **10**(1), e23392 (2024).
39. Redžepović, I. & Redžepović, I. Chemical applicability of Sombor indices: Survey. *J. Serb. Chem. Soc.* **86**(5), 445–457 (2021).
40. Gutman, I. Geometric approach to degree-based topological indices: Sombor indices. *MATCH Commun. Math. Comput. Chem* **86**(1), 11–16 (2021).
41. Deng, H., Tang, Z. & Wu, R. Molecular trees with extremal values of Sombor indices. *Int. J. Quantum Chem.* **121**(11), e26622 (2021).
42. Das, K. C. & Shang, Y. Some extremal graphs with respect to Sombor index. *Mathematics* **9**(11), 1202 (2021).
43. Sun, X. & Du, J. On Sombor index of trees with fixed domination number. *Appl. Math. Comput.* **421**, 126946 (2022).
44. Zhou, T., Lin, Z., & Miao, L. (2021). The Sombor index of trees and unicyclic graphs with given maximum degree. arXiv preprint arXiv:2103.07947.
45. Li, S., Wang, Z. & Zhang, M. On the extremal Sombor index of trees with a given diameter. *Appl. Math. Comput.* **416**, 126731 (2022).
46. Réti, T., Došlic, T. & Ali, A. On the Sombor index of graphs. *Contrib. Math* **3**, 11–18 (2021).
47. Swamy, N. N., Manohar, T., Sooryanarayana, B. & Gutman, I. Reverse sombor index. *Bull. Int. Math. Virt. Inst.* **12**(2), 267–272 (2022).
48. Kulli, V. R. Reverse elliptic Sombor and modified reverse elliptic Sombor indices. *Int. J. Math. Arch.* **15**(1), 1–7 (2024).

49. Espinal, C., Gutman, I., & Rada, J. (2024). Elliptic Sombor index of chemical graphs. Communications in Combinatorics and Optimization.
50. Shanmukha, M. C., Usha, A., Kulli, V. R. & Shilpa, K. C. Chemical applicability and curvilinear regression models of vertex-degree-based topological index: Elliptic Sombor index. *Int. J. Quantum Chem.* **124**(9), e27376 (2024).
51. Gutman, I. & Tošović, J. Testing the quality of molecular structure descriptors. Vertex-degree-based topological indices. *J. Serb. Chem. Soc.* **78**(6), 805–810 (2013).
52. Malik, M. Y. H., Binyamin, M. A. & Hayat, S. Correlation ability of degree-based topological indices for physicochemical properties of polycyclic aromatic hydrocarbons with applications. *Polycycl. Aromat. Compd.* **42**(9), 6267–6281 (2022).
53. Harary, F. *Graph Theory* (Addison-Wesely, 1969).
54. Kulli, V. R. *College Graph Theory* (Vishwa International Publication, 2012).

Author contributions

B.K.—Conceptualization, Article writing. M.C.S.—Conceptualization, Methodology, Article writing, Formal analysis, Resources, Data curation, Investigation. A.U.—Review of the manuscript, Formal analysis,, Suggestions given for correction of manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.