scientific reports

Check for updates

OPEN

Applications of Sombor topological indices and entropy measures for QSPR modeling of anticancer drugs: a Python-based methodology

Yeliz Kara[1], Yeşim Sağlam Özkan[1], Ali Berkan Bektaş[1] & Micheal Arockiaraj[2]✉

The development of effective anticancer drugs remains a central objective in pharmaceutical research. In recent years, topological indices (TIs) have gained considerable attention for their ability to numerically represent molecular structures and support predictive modeling in cheminformatics. This study aims to explore the potential of recently introduced Sombor topological indices and their entropy-based extensions within the framework of quantitative structure–property relationship (QSPR) modeling. The study will focus specifically on anticancer compounds, utilizing graph theory and edge partition approach. A comprehensive Python-based computational framework was developed to compute the relevant topological descriptors and entropy measures. The calculated indices were then integrated with statistical regression and machine learning techniques to construct and evaluate QSPR models to predict characteristics such as boiling point, molar refractivity, heavy atom count, exact mass, flash point, and polarizability. A curated dataset of anticancer agents was employed to ensure data reliability and chemical diversity. Comparative regression analyses indicate that Sombor indices exhibit stronger predictive performance and higher statistical significance than their entropy-based counterparts. These findings highlight the promise of Sombor indices as reliable molecular descriptors for QSPR modeling and powerful tools in the cheminformatics-guided drug discovery process.

Keywords Anticancer drugs, Topological descriptors, Entropies, Physicochemical properties, Python algorithm, QSPR models

The development of effective anticancer drugs remains an important goal in the field of medicinal chemistry and pharmaceutical sciences. In recent years, the integration of chem-informatics and computational methods has significantly accelerated the drug discovery process, in particular by enabling the development of predictive modeling techniques[1]. In this context, quantitative structure-property relationship (QSPR) modeling has emerged as a powerful tool for establishing mathematical relationships between the chemical structures of compounds and their biological or physicochemical properties[2].

Topological indices (TIs), which play a central role in QSPR modeling, are numerical descriptors derived from the molecular graph of a compound. As they encapsulate fundamental structural information, these indices enable the prediction of molecular properties without the need for experimental procedures. The first theoretical QSPR approaches can be traced back to the late 1940s[2]. These approaches correlated biological activities and physicochemical properties with theoretical numerical indices derived from molecular structure. Many TIs, ranging from classical degree-based indices such as the Randić index[3] to the distance-based Wiener index[4] and Zagreb indices[5], have been successfully applied to the modeling of various molecular properties over the years[6–10]. The main advantage of these indices is their ability to establish strong correlations with both structural and physico-biological properties. In recent years, considerable research has been devoted to exploring and refining TIs due to their efficacy in QSPR analysis. The authors analyzed the molecular structures of drugs related to lung cancer treatment by computing various topological indices, including degree[11], neighborhood[12], and reverse-degree based indices[13]. The study in[14] aimed to construct a QSPR model for 14 tuberculosis drugs by employing Revan degree–based topological indices to predict key physicochemical properties. Similarly, the work in[15]

[1]Department of Mathematics, Faculty of Arts and Science, Bursa Uludag University, 16059 Bursa, Turkey. [2]Department of Mathematics, Loyola College, Chennai 600034, India. ✉email: marockiaraj@gmail.com

applied QSPR analysis to 19 prostate cancer drugs, computed various topological indices, and compared them across 13 physicochemical properties to assess their predictive performance. In[16,17], computational chemistry approaches were integrated with machine-learning techniques to investigate the relationships between diverse topological descriptors and the physicochemical characteristics of the examined compounds. Furthermore[18], reported the computation of topological descriptors for several colorectal cancer drugs and evaluated their utility in predicting four physicochemical properties through QSPR modelling.

Recently, a new generation of topological descriptors has been developed beyond these classical indices. In particular, the Sombor topological index, proposed by Gutman in 2021 [19], is defined based on the vertex degrees and provides a more balanced structural representation. The Sombor index provides a richer description in terms of topological information, especially when considering the edge structure of the molecular graph. The QSPR models based on ve-degree Sombor indices for predicting key properties of aromatic heterocyclic compounds were developed in[20]. The molecular structures of antiviral drugs were examined using graph theory and the edge-partition approach in[21]. The predictive performance of the Sombor index and its variants was evaluated using regression models developed for key PAHs in[22]. A theoretical investigation of Sombor indices is provided within the framework of chemical graph theory in[23].

Shannon entropy quantifies the degree of unpredictability or uncertainty within a data set, where higher entropy indicates greater complexity and randomness, while lower entropy reflects more order and predictability[24,25]. A comparative study of the two versatile framework topologies, BCT and DFT, is presented and an entropy-based structural characterization is provided using bond-wise scaled comparison in[26]. Entropy-based descriptors and degree-based topological indices were generated from molecular graph structures using edge partitioning and computed for anticancer drugs with a Python-based algorithm in[27]. Entropy-based measures, when combined with TIs, offer a powerful framework for quantifying the structural complexity, diversity, and information content of molecular graphs[28]. The incorporation of principles from information theory enables these descriptors to provide a probabilistic perspective on molecular symmetry and irregularity features that are frequently critical in determining chemical behavior and biological activity. This is particularly evident in the context of pharmaceutical compounds, where entropy measures complement traditional topological indices by capturing subtle variations in molecular structure that influence drug performance and efficacy[29].

The identification and development of anticancer pharmaceuticals remains a key challenge in pharmaceutical research due to the heterogeneity and complexity of cancer. Computational techniques such as QSPR modeling have become widespread in experimental drug screening. This is driven by the dual objectives of enhancing efficiency and reducing the time and costs associated with the process. These approaches depend on molecular descriptors, with TIs derived from graph theory being particularly significant.

The 30 pharmaceutical compounds examined in this study represent a broad therapeutic spectrum within the field of oncology. These pharmaceuticals encompass agents employed in the management of diverse solid tumors, including those affecting the breast, lung, prostate, and bladder. Moreover, they are extensively utilized in the management of various hematological malignancies (blood cancers), such as leukemia, T-cell lymphoma, and multiple myeloma. The portfolio under consideration includes a variety of pharmacological approaches to cancer treatment, including targeted therapies, alkylating agents, and chemotherapy-supportive agents. These medications constitute the foundation of personalized therapeutic strategies guided by cancer type, disease stage, and individual genetic variability. To quantitatively elucidate the relationship between the structural characteristics of these compounds and their anticancer activities, this study introduces a systematic, Python-based computational framework for QSPR modeling. Within this framework, Sombor topological indices and their entropy-based extensions are employed to capture the underlying molecular information. A dedicated Python program has been specifically developed to automate the computation of these indices and facilitate the modeling process. The statistical significance of the data is ascertained through the implementation of Python during the analytical and modeling procedures. This multifaceted approach has been demonstrated to enhance the accuracy of modeling, whilst also enabling a comparative analysis of the performance of different regression models. This work provides a novel comparative evaluation of Sombor and entropy-based topological indices, demonstrating their potential as reliable predictors in anticancer drug modeling.

Motivation and methodology

Cancer continues to represent a significant global health challenge, thus necessitating the development of efficient and cost-effective methodologies to facilitate a comprehensive understanding and optimization of the physicochemical properties of anticancer compounds. The employment of experimental techniques for the characterization of these properties is frequently accompanied by significant expenses and extended periods of time. This underscores the necessity for reliable computational alternatives. In this study, graph-theoretical modeling is employed within a QSPR framework to analyze molecular structures. This modeling is based on degree-dependent and entropy-related topological indices. This approach provides an interpretable and low-cost method for assessing drug properties and offers valuable insights that may facilitate the rational design and discovery of more effective anticancer agents. The structural framework of the methodology, along with the tools employed throughout the study, is illustrated in the flowchart provided in Figure 1.

In chemistry, TIs play a crucial role in the study of the structure and properties of chemical molecules. These indices are derived from the underlying molecular structure of the chemical molecule, which is represented as a graph. In this graphical representation, the atoms of a molecule are represented as vertices, while the chemical bonds connecting them are represented as edges. In this article, $\mathscr{G}$ is denoted as a molecular graph, with $\mathscr{G} = (V(\mathscr{G}), E(\mathscr{G}))$, where $V(\mathscr{G})$ represents the set of vertices (atoms) and $E(\mathscr{G})$ represents the set of edges (chemical bonds) in the graph. Any two vertices u and v of a graph $\mathscr{G}$ are said to be adjacent or neighboring vertices if there exists an edge $uv \in E(\mathscr{G})$ connecting them. The degree of a vertex $u \in V(\mathscr{G})$ is defined as the number of edges that are connected to vertex u, denoted by $d(u)$.

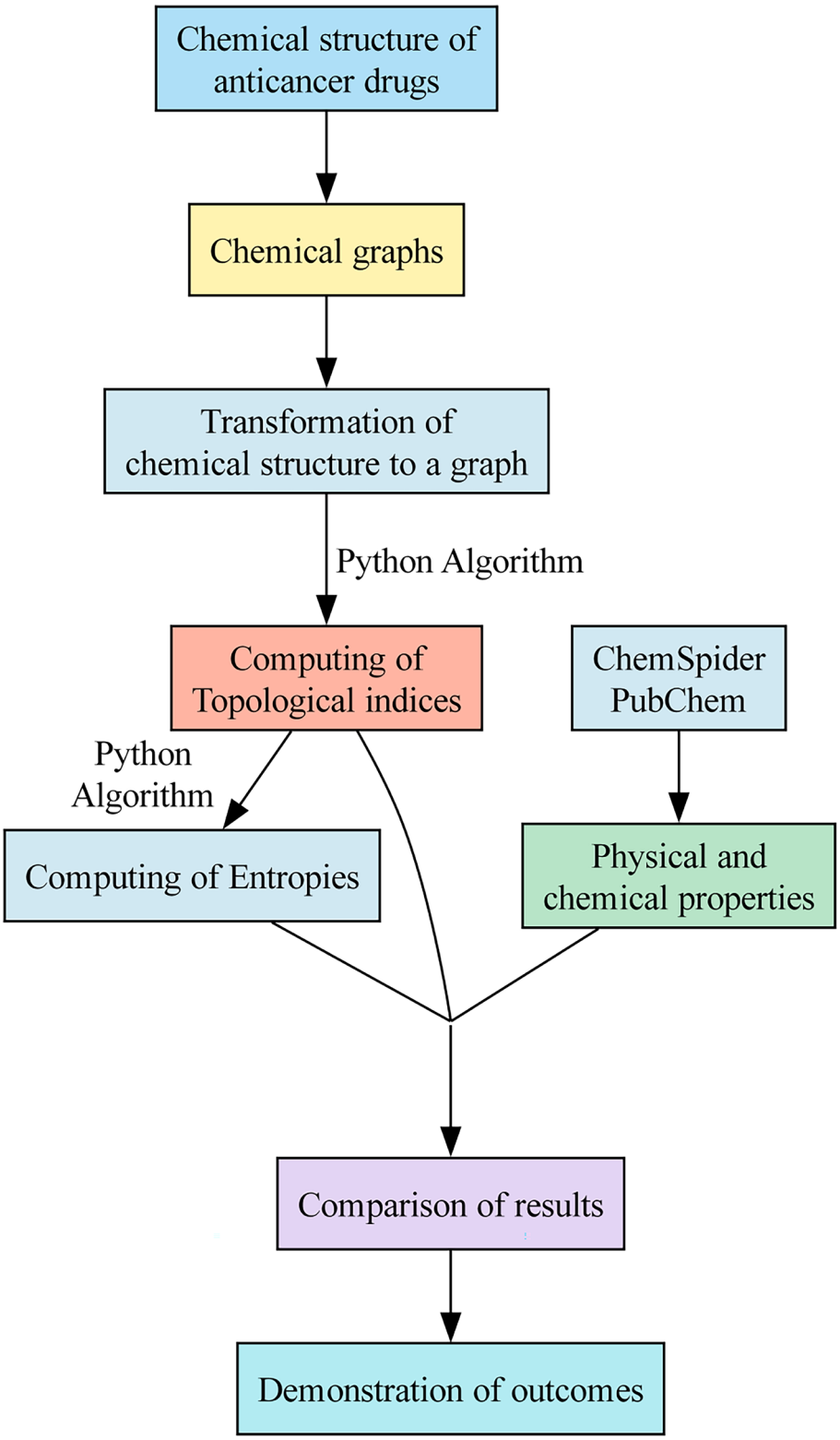

Fig. 1. Flow chart of methodology.

The general mathematical form of degree based topological index ($\mathcal{TI}$) with function ψ is defined as

$$\mathcal{TI}(\mathcal{G}) = \sum_{uv \in E(\mathcal{G})} \psi(d(u), d(v)) \tag{1}$$

where $\psi(d(u), d(v))$ is a real function of $d(u)$ and $d(v)$ with $\psi(d(u), d(v)) \geq 0$. The entropy measure based on the topological index function ψ is given by

$$\mathcal{I}_{\psi}(\mathcal{G}) = -\sum_{uv \in E(\mathcal{G})} \frac{\psi(d(u), d(v))}{\sum_{u_1 v_1 \in E(\mathcal{G})} \psi(d(u_1), d(v_1))} log \left(\frac{\psi(d(u), d(v))}{\sum_{u_1 v_1 \in E(\mathcal{G})} \psi(d(u_1), d(v_1))} \right). \tag{2}$$

The formulations of $\psi(d(u), d(v))$ for the degree-based TIs and entropies of a graph $\mathcal{G}$ are given in Table 1.

The Sombor index, a geometric approach to degree-based TIs, was introduced by Gutman[19]. The fundamental concept of this innovative index is to associate a pair (X, Υ) with an edge uv, where $X = max\{d(u), d(v)\}$ and $\Upsilon = min\{d(u), d(v)\}$. This is to say that each edge uv is mapped to a specific point (X, Υ) in a two-dimensional metric space R^2, where the distance from the origin is defined as $d(X, \Upsilon) = \sqrt{X^2 + \Upsilon^2}$. The Sombor index $SO(\mathcal{G})$ is obtained by summing up the values of $d(X, \Upsilon)$ for all edges in a graph $\mathcal{G}$. In the aforementioned paper, Gutman proposed a reduced version of the Sombor index, which was termed the reduced Sombor index. In consideration of the Sombor and reduced Sombor indices, Kulli and Gutman[30] proposed a modified version of these indices. Motivated by the work of Sombor indices, Kulli commenced the first and second Banhatti–Sombor indices[31,32].

In 2024, Gutman et al. introduced an alternative version of the Sombor index, termed the "elliptic Sombor index." This index refers to the orbits of planets in the solar system and incorporates elliptic orbits with the Sun as the focus point. In the field of astronomy, the perimeter of an ellipse is of paramount importance. The elliptic Sombor index, a crucial metric in the study of celestial mechanics, was derived from this fundamental principle[33]. A new type of Sombor index, i.e, the Euler Sombor index was proposed and discussed by Gutman[34]and Tang et al[35]. This index in question is derived from an approximate expression of the circumference of an ellipse. Consequently, it can be regarded as a geometry-based invariant.

In order to assess the practical implications of the theoretically defined Sombor indices, drug molecules used in the treatment of various types of cancer have been included in the analysis. In this context, the pharmaceutical properties of the compounds examined in the study are presented below.

In this study, a total of 30 drugs have been discussed. Afatinib and Alpelisib have been shown to be effective in solid tumours, including those of the lung and breast cancer types[36,37]. Conversely, Anastrozole has been identified as particularly effective in cases of hormone receptor-positive breast cancer. Belinostat is a widely used drug in the treatment of haematological cancers, such as T-cell lymphoma. Bortezomib, Lenalidomide and

Descriptor Name	Index	Formula	Entropy	Formula
Sombor index[19]	SO	$\sqrt{d(u)^2 + d(v)^2}$	$\mathcal{I}_{SO}$	$-\sum_{uv \in E(\mathcal{G})} \frac{\sqrt{d(u)^2+d(v)^2}}{SO} \log \frac{\sqrt{d(u)^2+d(v)^2}}{SO}$
Reduced Sombor index[19]	RSO	$\sqrt{(d(u)-1)^2 + (d(v)-1)^2}$	$\mathcal{I}_{RSO}$	$-\sum_{uv \in E(\mathcal{G})} \frac{\sqrt{(d(u)-1)^2+(d(v)-1)^2}}{RSO} \log \frac{\sqrt{(d(u)-1)^2+(d(v)-1)^2}}{RSO}$
Modified Sombor index[30]	mSO	$\frac{1}{\sqrt{d(u)^2+d(v)^2}}$	$\mathcal{I}_{mSO}$	$-\sum_{uv \in E(\mathcal{G})} \frac{\frac{1}{\sqrt{d(u)^2+d(v)^2}}}{mSO} \log \frac{\frac{1}{\sqrt{d(u)^2+d(v)^2}}}{mSO}$
Reduced Modified Sombor index[30]	$mRSO$	$\frac{1}{\sqrt{(d(u)-1)^2+(d(v)-1)^2}}$	$\mathcal{I}_{mRSO}$	$-\sum_{uv \in E(\mathcal{G})} \frac{\frac{1}{\sqrt{(d(u)-1)^2+(d(v)-1)^2}}}{mRSO} \log \frac{\frac{1}{\sqrt{(d(u)-1)^2+(d(v)-1)^2}}}{mRSO}$
First Banhatti-Sombor index[31]	BSO_1	$\sqrt{\frac{1}{d(u)^2} + \frac{1}{d(v)^2}}$	$\mathcal{I}_{BSO_1}$	$-\sum_{uv \in E(\mathcal{G})} \frac{\sqrt{\frac{1}{d(u)^2} + \frac{1}{d(v)^2}}}{BSO_1} \log \frac{\sqrt{\frac{1}{d(u)^2} + \frac{1}{d(v)^2}}}{BSO_1}$
Second Banhatti-Sombor index[32]	BSO_2	$\frac{1}{\sqrt{\frac{1}{d(u)^2} + \frac{1}{d(v)^2}}}$	$\mathcal{I}_{BSO_2}$	$-\sum_{uv \in E(\mathcal{G})} \frac{\frac{1}{\sqrt{\frac{1}{d(u)^2} + \frac{1}{d(v)^2}}}}{BSO_2} \log \frac{\frac{1}{\sqrt{\frac{1}{d(u)^2} + \frac{1}{d(v)^2}}}}{BSO_2}$
Elliptic Sombor index[33]	ESO	$(d(u) + d(v))\sqrt{d(u)^2 + d(v)^2}$	$\mathcal{I}_{ESO}$	$-\sum_{uv \in E(\mathcal{G})} \frac{(d(u)+d(v))\sqrt{d(u)^2+d(v)^2}}{ESO} \log \frac{(d(u)+d(v))\sqrt{d(u)^2+d(v)^2}}{ESO}$
Euler Sombor index[34,35]	$EUSO$	$\sqrt{d(u)^2 + d(v)^2 + d(u)d(v)}$	$\mathcal{I}_{EUSO}$	$-\sum_{uv \in E(\mathcal{G})} \frac{\sqrt{d(u)^2+d(v)^2+d(u)d(v)}}{EUSO} \log \frac{\sqrt{d(u)^2+d(v)^2+d(u)d(v)}}{EUSO}$

Table 1. Mathematical formulations of Sombor degree-based TIs.

Pomalidomide are used in the treatment of multiple myeloma[38]. Alkylating agents such as Busulfan, Carmustine and Lomustine are particularly used in the treatment of leukaemia and brain tumours. In the treatment of chronic myeloid leukemia (CML), Dasatinib and Nilotinib are commonly used, whereas Daunorubicin and Mitomycin C are employed against a range of tumors, including leukemia and bladder cancer. Erdafitinib is used to treat advanced bladder cancer, while Flutamide and Orgovyx are used to treat prostate cancer[39]. Futibatinib and Repotrectinib act as targeted therapy agents, while Granisetron is a pharmaceutical agent used to prevent chemotherapy-induced nausea and vomiting. Melphalan is used to treat epithelial ovarian cancer. Brutinib, Zanubrutinib, Prednisone, and Midostaurin have demonstrated efficacy in the treatment of lymphoma and certain types of leukemia. Olaparib is used for managing BRCA-mutated ovarian and breast cancers, while Olutasidenib is employed in treating IDH1 mutation–positive acute myeloid leukemia (AML). Plerixafor is indicated for stem cell mobilization, and the treatment of T-cell lymphoma involves the use of Pralatrexate[40].

In the context of oncological treatments, Ribociclib has been utilized for the management of hormone receptor-positive and HER2-negative breast cancer. These medications form the basis of personalised treatment approaches that are informed by cancer type, disease stage, and individual genetic variations. It is imperative to acknowledge that the information pertaining to pharmaceuticals is intended solely for educational purposes and should not be construed as a substitute for professional medical counsel[41]. The molecular structures of the thirty cancer drugs discussed in this study, illustrated in Figures 2–3, were drawn using Chemcraft visualizer [Chemcraft - graphical software for visualization of quantum chemistry computations. Version 1.8, build 682. https://www.chemcraftprog.com].

Topological descriptors calculation

In this section, we present the mathematical calculations for the representative molecular graph $\mathscr{G}$ of Belinostat, whose edge set is partitioned based on the vertex degrees of its 22 vertices, as shown in Figure 4. The subsets $E_{i,j} = \{uv \in E(\mathscr{G}) : d(u) = i,\ d(v) = j\}$ denote the equivalence classes of $E(\mathscr{G})$ based on the terminal degrees. For Belinostat, we have $|E_{1,2}| = 1$, $|E_{1,3}| = 1$, $|E_{1,4}| = 2$, $|E_{2,2}| = 7$, $|E_{2,3}| = 10$, $|E_{2,4}| = 1$, and $|E_{3,4}| = 1$. By applying equations in Table 1 for topological indices, the following results are obtained:

$$\begin{aligned} SO(\mathscr{G}) &= 1\sqrt{1^2+2^2} + 1\sqrt{1^2+3^2} + 2\sqrt{1^2+4^2} + 7\sqrt{2^2+2^2} + 10\sqrt{2^2+3^2} \\ &\quad + 1\sqrt{2^2+4^2} + 1\sqrt{3^2+4^2} \approx 78.971 \end{aligned} \tag{3}$$

$$\begin{aligned} RSO(\mathscr{G}) &= 1\left(\sqrt{(1-1)^2+(2-1)^2}\right) + 1\left(\sqrt{(1-1)^2+(3-1)^2}\right) + 2\left(\sqrt{(1-1)^2+(4-1)^2}\right) \\ &\quad + 7\left(\sqrt{(2-1)^2+(2-1)^2}\right) + 10\left(\sqrt{(2-1)^2+(3-1)^2}\right) + 1\left(\sqrt{(2-1)^2+(4-1)^2}\right) \\ &\quad + 1\left(\sqrt{(3-1)^2+(4-1)^2}\right) \approx 48.028 \end{aligned} \tag{4}$$

$$\begin{aligned} mSO(\mathscr{G}) &= 1\left(\frac{1}{\sqrt{1^2+2^2}}\right) + 1\left(\frac{1}{\sqrt{1^2+3^2}}\right) + 2\left(\frac{1}{\sqrt{1^2+4^2}}\right) + 7\left(\frac{1}{\sqrt{2^2+2^2}}\right) + 10\left(\frac{1}{\sqrt{2^2+3^2}}\right) \\ &\quad + 1\left(\frac{1}{\sqrt{2^2+4^2}}\right) + 1\left(\frac{1}{\sqrt{3^2+4^2}}\right) \approx 6.9205 \end{aligned} \tag{5}$$

$$\begin{aligned} mRSO(\mathscr{G}) &= 1\left(\frac{1}{\sqrt{(1-1)^2+(2-1)^2}}\right) + 1\left(\frac{1}{\sqrt{(1-1)^2+(3-1)^2}}\right) + 2\left(\frac{1}{\sqrt{(1-1)^2+(4-1)^2}}\right) \\ &\quad + 7\left(\frac{1}{\sqrt{(2-1)^2+(2-1)^2}}\right) + 10\left(\frac{1}{\sqrt{(2-1)^2+(3-1)^2}}\right) + 1\left(\frac{1}{\sqrt{(2-1)^2+(4-1)^2}}\right) \\ &\quad + 1\left(\frac{1}{\sqrt{(3-1)^2+(4-1)^2}}\right) \approx 12.1821 \end{aligned} \tag{6}$$

$$\begin{aligned} BSO_1(\mathscr{G}) &= 1\left(\sqrt{\frac{1}{1^2}+\frac{1}{2^2}}\right) + 1\left(\sqrt{\frac{1}{1^2}+\frac{1}{3^2}}\right) + 2\left(\sqrt{\frac{1}{1^2}+\frac{1}{4^2}}\right) + 7\left(\sqrt{\frac{1}{2^2}+\frac{1}{2^2}}\right) + 10\left(\sqrt{\frac{1}{2^2}+\frac{1}{3^2}}\right) \\ &\quad + 1\left(\sqrt{\frac{1}{2^2}+\frac{1}{4^2}}\right) + 1\left(\sqrt{\frac{1}{3^2}+\frac{1}{4^2}}\right) \approx 16.1684 \end{aligned} \tag{7}$$

$$\begin{aligned} BSO_2(\mathscr{G}) &= 1\left(\frac{1}{\sqrt{\frac{1}{1^2}+\frac{1}{2^2}}}\right) + 1\left(\frac{1}{\sqrt{\frac{1}{1^2}+\frac{1}{3^2}}}\right) + 2\left(\frac{1}{\sqrt{\frac{1}{1^2}+\frac{1}{4^2}}}\right) + 7\left(\frac{1}{\sqrt{\frac{1}{2^2}+\frac{1}{2^2}}}\right) + 10\left(\frac{1}{\sqrt{\frac{1}{2^2}+\frac{1}{3^2}}}\right) \\ &\quad + 1\left(\frac{1}{\sqrt{\frac{1}{2^2}+\frac{1}{4^2}}}\right) + 1\left(\frac{1}{\sqrt{\frac{1}{3^2}+\frac{1}{4^2}}}\right) \approx 34.5128 \end{aligned} \tag{8}$$

$$\begin{aligned} ESO(\mathscr{G}) &= 1\left((1+2)\sqrt{1^2+2^2}\right) + 1\left((1+3)\sqrt{1^2+3^2}\right) + 2\left((1+4)\sqrt{1^2+4^2}\right) + 7\left((2+2)\sqrt{2^2+2^2}\right) \\ &\quad + 10\left((2+3)\sqrt{2^2+3^2}\right) + 1\left((2+4)\sqrt{2^2+4^2}\right) + 1\left((3+4)\sqrt{3^2+4^2}\right) \approx 381.895 \end{aligned} \tag{9}$$

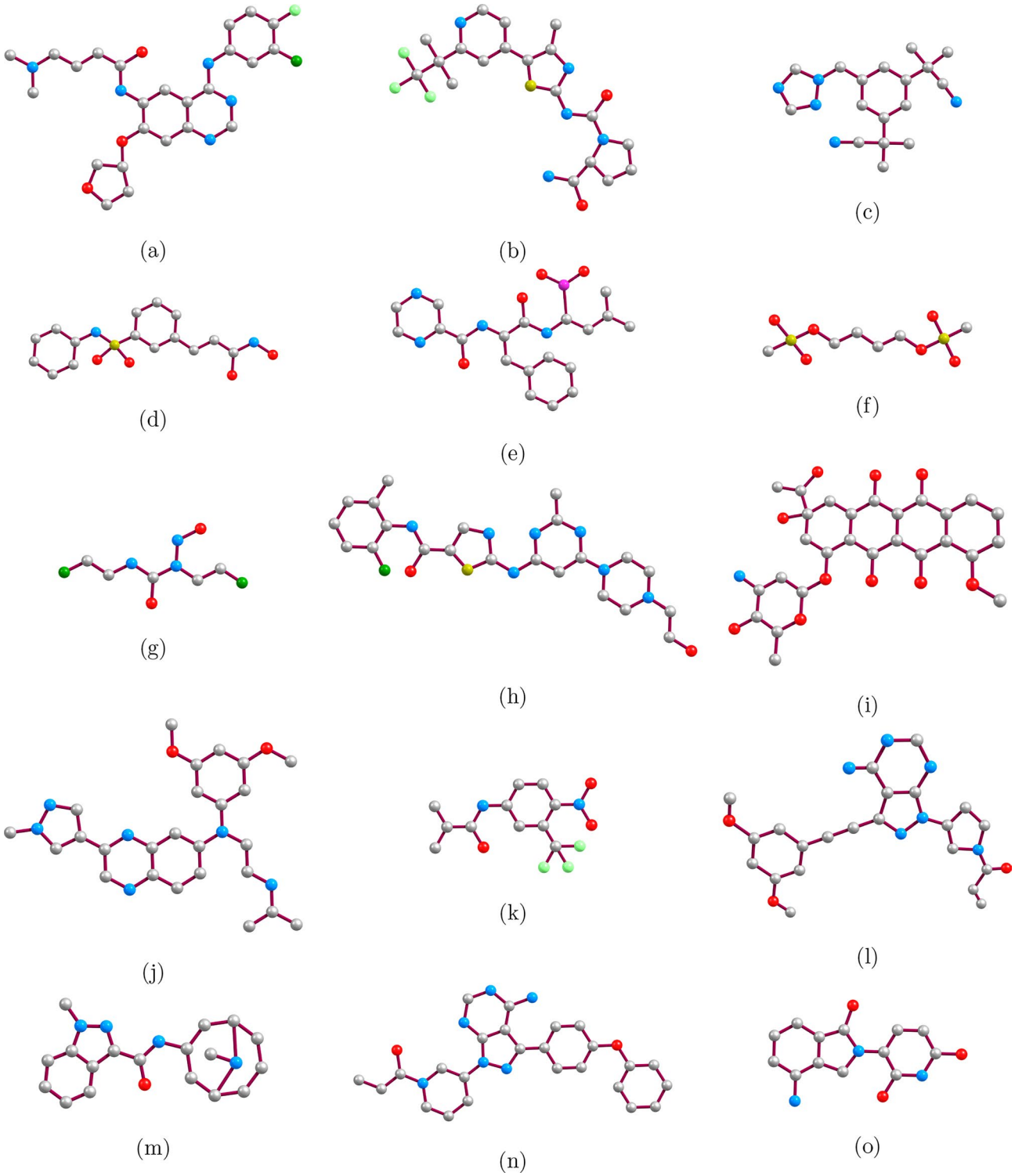

Fig. 2. Chemical structures: (**a**) Afatinib, (**b**) Alpelisib, (**c**) Anastrozole, (**d**) Belinostat, (**e**) Bortezomib, (**f**) Busulfan, (**g**) Carmustine, (**h**) Dasatinib, (**i**) Daunorubicin, (**j**) Erdafitinib, (**k**) Flutamide, (**l**) Futibatinib, (**m**) Granisetron, (**n**) Ibrutinib, (**o**) Lenalidomide.

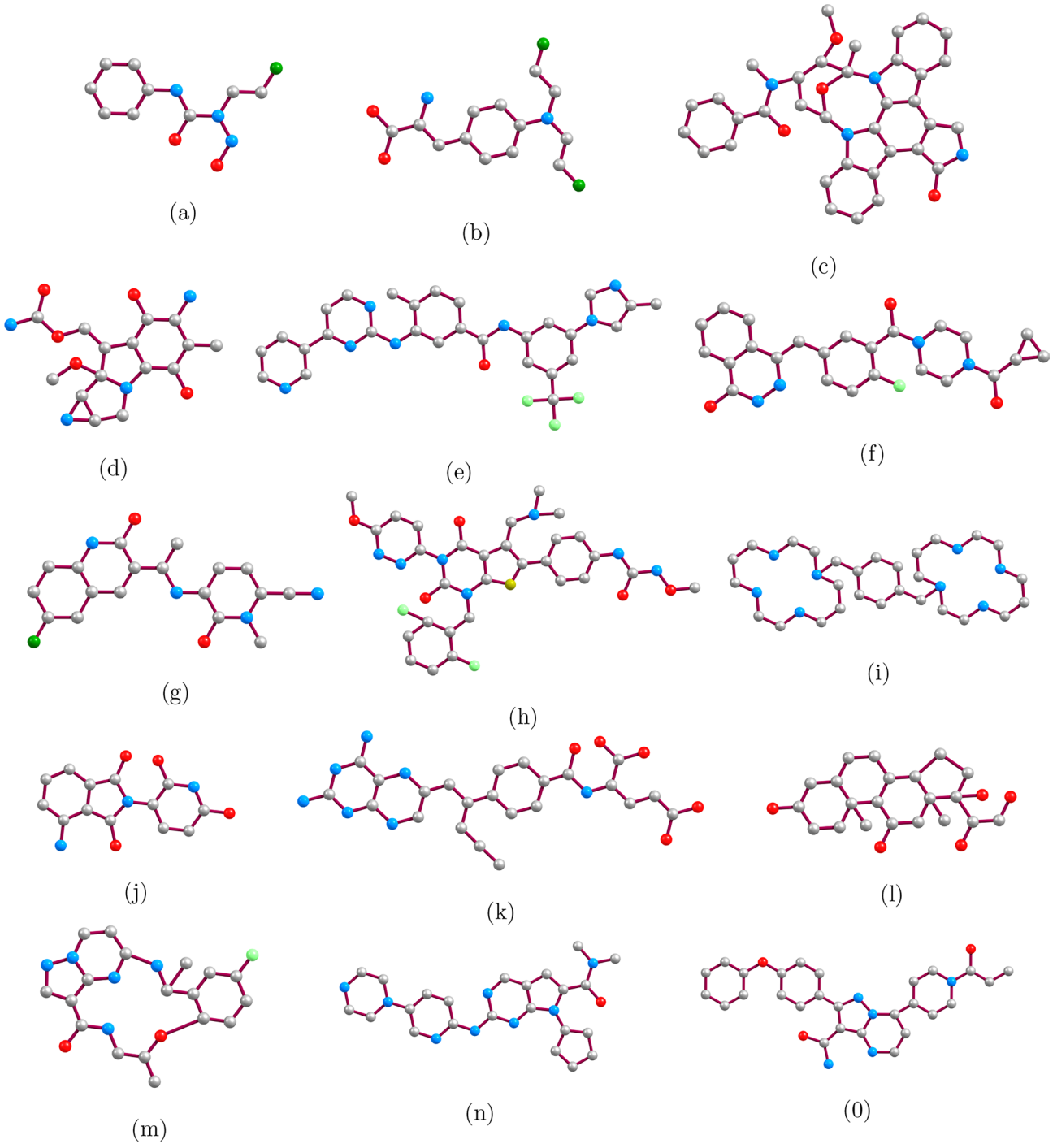

Fig. 3. Chemical structures (continued): (**a**) Lomustine, (**b**) Melphalan, (**c**) Midostaurin, (**d**) Mitomycin C, (**e**) Nilotinib, (**f**) Olaparib, (**g**) Olutasidenib, (**h**) Orgovyx, (**i**) Plerixafor, (**j**) Pomalidomide, (**k**) Pralatrexate, (**l**) Prednisone, (**m**) Repotrectinib, (**n**) Ribociclib, (**o**) Zanubrutinib.

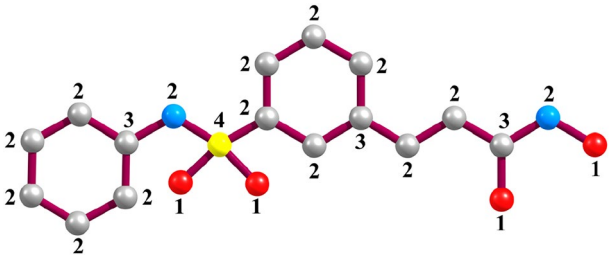

Fig. 4. Graph for Belinostat.

Drugs	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(2, 2)	(2, 3)	(2, 4)	(3, 3)	(3, 4)	(4, 4)
Afatinib	0	0	5	0	8	20	0	4	0	0
Alpelisib	0	0	4	5	4	12	0	5	1	1
Anastrozole	0	2	0	4	3	10	2	0	2	0
Belinostat	0	1	1	2	7	10	1	0	1	0
Bortezomib	0	0	6	0	8	12	0	3	0	0
Busulfan	0	0	0	6	5	0	2	0	0	0
Carmustine	0	3	1	0	3	3	0	1	0	0
Dasatinib	0	1	4	0	6	21	0	4	0	0
Daunorubicin	0	1	9	1	2	11	2	15	1	0
Erdafitinib	0	2	3	0	5	22	0	4	0	0
Flutamide	0	0	5	3	1	6	0	3	1	0
Futibatinib	0	3	2	0	4	19	0	6	0	0
Granisetron	0	0	3	0	5	12	0	6	0	0
Ibrutinib	0	1	2	0	10	17	0	7	0	0
Lenalidomide	0	0	4	0	3	8	0	6	0	0
Lomustine	0	2	1	0	5	6	0	1	0	0
Melphalan	0	2	3	0	4	8	0	2	0	0
Midostaurin	0	1	3	1	11	12	1	20	2	0
Mitomycin C	0	1	6	0	1	6	1	9	3	0
Nilotinib	0	0	3	3	8	24	0	4	1	0
Olaparib	0	0	4	0	8	16	0	8	0	0
Olutasidenib	0	1	5	0	2	13	0	6	0	0
Orgovyx	0	2	7	0	7	20	0	12	0	0
Plerixafor	0	0	0	0	26	12	0	0	0	0
Pomalidomide	0	0	5	0	3	6	0	8	0	0
Pralatrexate	0	1	7	0	5	19	0	5	0	0
Prednisone	0	1	3	3	3	8	3	3	4	1
Repotrectinib	0	0	4	0	4	16	0	5	0	0
Ribociclib	0	0	3	0	10	16	0	7	0	0
Zanubrutinib	0	1	3	0	10	17	0	8	0	0

Table 2. Edge partitions of anticancer drugs.

$$EUSO(\mathscr{G}) = 1\left(\sqrt{1^2+2^2+1(2)}\right) + 1\left(\sqrt{1^2+3^2+1(3)}\right) + 2\left(\sqrt{1^2+4^2+1(4)}\right) + 7\left(\sqrt{2^2+2^2+2(2)}\right) + 10\left(\sqrt{2^2+3^2+2(3)}\right) + 1\left(\sqrt{2^2+4^2+2(4)}\right) + 1\left(\sqrt{3^2+4^2+3(4)}\right) \approx 94.6284 \tag{10}$$

Table 2 provides a concise overview of the edge partitions of the molecular graphs for the anticancer drugs under investigation. The edges are partitioned according to the degrees of their terminal vertices (e.g., edges of type (1,2), (2,3), etc.). This classification forms the basis for computing the Sombor-like topological indices and their entropy-related descriptors through mathematical simplification, as illustrated for the structure of Belinostat. Such calculations for the 30 structures considered in this study are both time- and space-consuming; however, within the Python framework, these partitions were automatically generated for each compound and subsequently utilized in all index and entropy calculations.

As illustrated in Algorithm 1, the degree-based topological indices are subsequently computed through the utilization of a Python script that has been developed for the purpose of this study. The program commences by importing the requisite libraries, including networkx and pandas for the purpose of numerical calculations. Please refer to Tables 3 and 4 for the calculated values of the topological indices and entropies, respectively.

Modeling structure property through topological methods

QSPR analysis is a computational method for predicting a chemical compound's properties based on its molecular structure. The selected drugs' physicochemical properties and associated TIs have been elucidated using the QSPR modeling approach. This study investigated six physicochemical properties of the anticancer drugs in question: boiling point (BP), molar refractivity (MR), heavy atom count (HAC), exact mass (EM), flash point (FP), and polarizability (P). These properties play a crucial role in determining how a drug behaves, interacts, and performs within biological systems, influencing absorption, distribution, metabolism, and excretion, and ultimately affecting overall drug efficacy and safety. As shown in Table 5, the chemical and physical properties of the drugs under investigation have been obtained from the following sources: PubChem[42] and ChemSpider[43].

In this section, a rigorous evaluation of linear, quadratic and cubic regression models is conducted. In this analysis, the independent variables are the previously calculated TIs, and the dependent variables are the properties of the pharmaceutical compounds under investigation. $\mathcal{Y} = \alpha + \beta\ (\mathcal{TI})$, $\mathcal{Y} = \alpha + \beta\ (\mathcal{TI}) + \gamma\ (\mathcal{TI})^2$ and $\mathcal{Y} = \alpha + \beta\ (\mathcal{TI}) + \gamma\ (\mathcal{TI})^2 + \delta\ (\mathcal{TI})^3$ define the linear, quadratic and cubic regression relationship, respectively, with $\mathcal{Y}$ describing the physical property of the drug, $\mathcal{TI}$ representing the topological index, and α, β, γ and δ serving as regression parameters.

A comprehensive array of statistical evaluations has been conducted to assess the models, incorporating correlation-based and error-based performance measures, including the correlation coefficient (*R*), the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE), the standard error (SE) of the predictions, the significance level (*p*-value), and Fisher's statistical measure (*F*). The predictive capability of the model is evaluated using the RMSE metric, with the optimal model exhibiting the minimum error. RMSE is calculated as the square root of the mean of the squares of the differences between the predicted parameter value and the actual parameter value, divided by the sample size. The correlation coefficient gets closer to 1 when the results of experiments and theoretical predictions match each other. Furthermore, R-squared (R^2), also referred to as the coefficient of determination, is a statistical metric that quantifies the extent to which the independent variables explain the variability in the dependent variable. It represents the extent to which the independent variables account for the variance in the dependent variable. The MAE is a metric that quantifies the average absolute deviation of the model's predictions from the actual values, thereby indicating the overall magnitude of prediction errors. The MPE is a metric that calculates the average prediction bias in percentage terms, thus indicating whether the model systematically overestimates or underestimates the target values. The SE is a quantitative metric that quantifies the dispersion of the predictions around the mean, thereby providing an indication of the level of uncertainty associated with the model's estimates.

Tables 6, 7 and 8 show the relationship between the experimental properties and different topological indices using a linear, quadratic and cubic regression models, respectively, where the highest correlation values are shown in bold. Notably, BSO_2 exhibits a strong positive correlation with BP and FP. The correlation coefficients of 0.99 (linear\quadratic) and 0.9904 (cubic) indicate a very strong positive relationship between *mSO* and both MR. The correlation coefficients of 0.9899, 0.99, 0.9903 indicate a very strong positive relationship between *mSO* and both P in linear, quadratic and cubic models, respectively. Similarly, BSO_1 shows a significant positive correlation with HAC (R = 0.999). Furthermore, BSO_1 shows a strong positive correlation with EM in all models, as evidenced by its high R-value of 0.989, 0.9911, and 0.9923, respectively. Subsequently, the utilization of linear, quadratic and cubic regressions facilitate the establishment of best mathematical models encompassing the physicochemical properties and topological indices. The best model was determined based on the R^2, RMSE, and p-values. The performance metrics of best regression models, as detailed in Table 9, demonstrate their efficacy and accuracy in predicting molecular properties.

The linear, quadratic and cubic models explain 84.8% of the variation in BP and BSO_2, as indicated by $R^2 = 0.848$. It accounts for 98% of the variation in MR and *mSO* ($R^2 = 0.980$) in linear and quadratic models. The cubic model explains 98.1% of the variation in MR and *mSO*, as indicated by $R^2 = 0.981$. The R^2 value of 0.998 demonstrates that the models provides an almost complete explanation (99.8%) of the variations in HAC and BSO_1. Similarly, 97.8% of the variation in EM and BSO_1 is explained by the linear model ($R^2 = 0.978$), while 98.2% and 98.5% of the variation are captured by the quadratic and cubic models ($R^2 = 0.982$ and $R^2 = 0.985$, respectively). For all three models, 85.7% of the variation in FP and BSO_2 is explained ($R^2 = 0.857$). Furthermore, the linear and quadratic models show a remarkable ability to explain 98% of the variation in P for *mSO* ($R^2 = 0.980$), while the cubic model accounts for 98.1% of the variation ($R^2 = 0.981$). As indicated by the statistical parameters displayed in Table 9, the ensuing conclusions can be derived concerning the linear, quadratic and cubic regression models for the indices in question. For all of the properties that were examined, the SE, MPE, and MAE values of the linear, quadratic, and cubic models were very close to each other. This indicates that the three regression forms demonstrate similar levels of predictive accuracy. For BP$\sim BSO_2$, the cubic model provides slightly lower SE and MAE values, although the differences are minimal. A similar pattern is observed for MR$\sim mSO$, HAC$\sim BSO_1$, EM$\sim BSO_1$, FP$\sim BSO_2$, and P$\sim mSO$, where the cubic model consistently yields the smallest SE and MAE values, but the improvements over the linear and quadratic models remain marginal. The error-based metrics demonstrate that all three models demonstrate high accuracy across all properties. The cubic models offer slightly better predictive stability and lower error levels, although the gains in performance are negligible.

```
    import json
    from math import log, sqrt
    import pandas as pd
    import networkx as nx
    drug_list = sorted(["Afatinib",  "Anastrozole", "Alpelisib", "Busulfan",    "
        ↪ Daunorubicin", "Dasatinib",
"Erdafitinib", "Melphalan", "Mitomycin", "Nilotinib", "Olaparib",
    "Orgovyx",  "Plerixafor",   "Prednisone", "Zanubrutinib", "Belinostat", "Bortezomib"
        ↪ ,
    "Carmustine",   "Flutamide",    "Futibatinib", "Granisetron", "Ibrutinib", "
        ↪ Lenalidomide",
    "Lomustine",    "Midostaurin",  "Olutasidenib", "Pomalidomide", "Pralatrexate",
     "Repotrectinib", "Ribociclib"])
    with open("edge_list.json", "r", encoding="utf-8") as file:
        data = json.load(file)
    def calculate_indices(G):
    indices = { "Sombor": 0, "Red. Sombor": 0,  "Mod. Sombor": 0, "Red. Mod. Sombor": 0,
        ↪
        "BSO1": 0, "BSO2": 0, "Elliptic Sombor": 0, "Euler Sombor": 0   }
    for x, y in G.edges():
    dx, dy = G.degree(x), G.degree(y)
    indices["Sombor"] += sqrt(dx**2 + dy**2)
    indices["Red. Sombor"] += sqrt((dx - 1)**2 + (dy - 1)**2)
    indices["Mod. Sombor"] += sqrt(1 / (dx**2 + dy**2))
    indices["Red. Mod. Sombor"] += sqrt(1 / ((dx - 1)**2 + (dy - 1)**2))
    indices["BSO1"] += sqrt(dx**2 + dy**2) / (dx * dy)
    indices["BSO2"] += (dx * dy) / sqrt(dx**2 + dy**2)
    indices["Elliptic Sombor"] += (dx + dy) * sqrt(dx**2 + dy**2)
    indices["Euler Sombor"] += sqrt(dx**2 + dy**2 + dx * dy)
        return indices
    def calculate_entropy(G, indices):
    entropy = {f"ent_{key}": 0 for key in indices}
    for x, y in G.edges():
    dx, dy = G.degree(x), G.degree(y)
    elements = {    "Sombor": sqrt(dx**2 + dy**2),
        "Red. Sombor": sqrt((dx - 1)**2 + (dy - 1)**2),
        "Mod. Sombor": sqrt(1 / (dx**2 + dy**2)),
        "Red. Mod. Sombor": sqrt(1 / ((dx - 1)**2 + (dy - 1)**2)),
            "BSO1": sqrt(dx**2 + dy**2) / (dx * dy),
        "BSO2": (dx * dy) / sqrt(dx**2 + dy**2),
        "Elliptic Sombor": (dx + dy) * sqrt(dx**2 + dy**2),
        "Euler Sombor": sqrt(dx**2 + dy**2 + dx * dy)   }
    for key, value in indices.items():
    if value != 0:
    element = elements[key]
    entropy[f"ent_{key}"] += -1 * (element / value)
    * log(element / value)
    return entropy
```

Algorithm 1. Python code to compute several topological indices (TIs) and their entropies.

The best predictors estimated across all regression models for the considered physicochemical properties (BP, MR, HAC, EM, FP and P) are visualized in Figures 5, 6 and 7, while Figure 8 provides a collective comparison of linear, quadratic, and cubic regression curves to highlight differences in model behavior across polynomial degrees.

Drugs	*SO*	*RSO*	*mSO*	*mRSO*	BSO_1	BSO_2	*ESO*	*EUSO*
Afatinib	127.52	77.349	10.899	18.515	24.831	57.824	616.134	153.703
Alpelisib	119.715	77.48	8.775	14.142	22.537	50.107	622.796	142.49
Anastrozole	84.45	54.139	6.546	11.114	16.441	34.931	433.763	100.352
Belinostat	78.971	48.028	6.92	12.182	16.168	34.513	381.895	94.628
Bortezomib	97.596	58.632	8.761	15.084	20.607	43.339	459.105	117.241
Busulfan	47.825	31.396	3.67	6.168	10.838	16.47	233.927	55.399
Carmustine	33.415	18.779	3.786	7.317	8.804	14.988	146.254	40.208
Dasatinib	124.543	75.756	10.601	18.048	24.082	56.606	605.593	150.174
Daunorubicin	157.721	101.782	11.477	18.38	28.266	69.334	832.602	189.162
Erdafitinib	124.394	75.578	10.655	18.288	24.04	56.801	606.366	150.109
Flutamide	70.37	45.921	5.234	8.228	14.506	27.817	355.94	83.064
Futibatinib	118.308	72.113	10.072	17.447	22.537	54.583	585.94	143.001
Granisetron	92.351	56.874	7.459	12.523	16.737	42.614	463.584	111.621
Ibrutinib	127.838	76.954	10.98	19.149	23.813	60.073	629.806	154.972
Lenalidomide	75.435	47.102	5.959	9.82	13.974	34.078	381.495	90.863
Lomustine	47.652	27.316	4.878	9.072	10.903	21.915	216.256	57.567
Melphalan	62.602	37.202	5.948	10.613	13.977	27.847	291.752	75.228
Midostaurin	189.55	119.331	14.193	23.92	31.121	89.251	1012.005	229.837
Mitomycin C	103.327	67.265	7.307	11.721	17.807	46.066	563.019	124.202
Nilotinib	152.987	94.899	12.304	20.582	28.636	67.894	759.793	183.758
Olaparib	126.906	77.718	10.417	17.641	23.259	58.705	633.197	153.447
Olutasidenib	96.032	59.868	7.755	12.849	18.443	42.828	479.677	115.444
Orgovyx	169.43	104.562	13.958	23.637	32.24	77.067	847.181	204.311
Plerixafor	116.806	63.602	12.521	23.751	25.596	56.739	510.489	142.373
Pomalidomide	79.871	50.286	6.192	10.133	14.769	35.941	409	96.143
Pralatrexate	128.233	78.698	10.877	18.3	25.807	56.831	621.627	154.005
Prednisone	113.223	74.768	7.758	12.553	19.413	48.365	626.786	135.196
Repotrectinib	102.865	63.576	8.295	13.752	19.017	46.684	511.575	124.002
Ribociclib	125.158	75.718	10.572	18.201	23.148	58.463	617.719	151.573
Zanubrutinib	135.243	81.783	11.532	20.002	25.338	63.143	667.911	163.774

Table 3. TIs of anticancer drugs.

In the following part, the relationship between entropies and the physicochemical properties of the drugs under investigation is examined. In Tables 10, 11, and 12, where the highest correlation values are shown in bold, linear, quadratic, and cubic regression analyses have been used to calculate the correlation coefficients between the entropy values and the physicochemical characteristics of the drugs, respectively. In Table 10, a high R-value indicates a strong positive correlation between $\mathcal{I}_{BSO_1}$ and MR, HAC, EM and P, with values of 0.9552, 0.9715, 0.9373 and 0.9550, respectively. Similarly, $\mathcal{I}_{BSO_2}$ exhibits significant positive associations with FP, as reflected by its high R-value of 0.9550. Furthermore, $\mathcal{I}_{ESO}$ shows a strong positive correlation with BP, with R-value of 0.9103. In Tables 11 and 12, a high R-value demonstrates a strong positive relationship between $\mathcal{I}_{BSO_1}$ and MR as well as P. Likewise, $\mathcal{I}_{ESO}$ displays a strong positive correlation with BP, HAC, and FP. Moreover, $\mathcal{I}_{mRSO}$ reveals a significant positive association with EM.

We now implement the linear, quadratic, and cubic regression models to establish the best mathematical relationships encompassing physicochemical properties and entropies. As demonstrated in Table 13, the performance metrics of the best regression models indicate their effectiveness and accuracy in predicting molecular properties. However, a comparison of Tables 9 and 13 shows that the correlations obtained using topological indices for each physicochemical property are more meaningful than those obtained using entropies. Therefore, the comparative analysis will be performed using topological indices in the following section.

A comparative analysis approach

In the case of linear regression models, the findings of this study indicate that BP and FP can be effectively predicted using BSO_2. The *mSO* index successfully predicts P and MR, while the BSO_1 index proves effective in predicting HAC and EM. We compared our models with existing models for drugs used in cancer treatment, cited in references[44–46]. The following observations are revealed: In the study referenced in[44], it was proposed that, upon consideration of ten drugs and boiling point, the highest correlation for Detour index (D) was identified to be

$$BP = 176.8677 + 0.4304[D], \quad R^2 = 0.5348, \quad RMSE = 75.5004. \tag{11}$$

Drugs	$\mathcal{I}_{SO}$	$\mathcal{I}_{RSO}$	$\mathcal{I}_{mSO}$	$\mathcal{I}_{mRSO}$	$\mathcal{I}_{BSO_1}$	$\mathcal{I}_{BSO_2}$	$\mathcal{I}_{ESO}$	$\mathcal{I}_{EUSO}$
Afatinib	3.603	3.591	3.603	3.586	3.583	3.59	3.579	3.603
Alpelisib	3.453	3.433	3.453	3.431	3.416	3.419	3.408	3.452
Anastrozole	3.116	3.083	3.111	3.059	3.092	3.098	3.066	3.117
Belinostat	3.119	3.087	3.119	3.084	3.106	3.111	3.076	3.121
Bortezomib	3.359	3.344	3.359	3.341	3.334	3.34	3.33	3.358
Busulfan	2.547	2.507	2.545	2.492	2.539	2.537	2.514	2.552
Carmustine	2.376	2.337	2.376	2.336	2.354	2.356	2.308	2.374
Dasatinib	3.575	3.561	3.573	3.552	3.554	3.562	3.549	3.574
Daunorubicin	3.726	3.713	3.724	3.702	3.679	3.693	3.687	3.723
Erdafitinib	3.574	3.558	3.571	3.543	3.552	3.562	3.545	3.573
Flutamide	2.934	2.922	2.935	2.921	2.891	2.89	2.898	2.932
Futibatinib	3.513	3.494	3.509	3.473	3.49	3.501	3.477	3.512
Granisetron	3.249	3.235	3.249	3.229	3.227	3.235	3.219	3.248
Ibrutinib	3.599	3.578	3.598	3.57	3.586	3.592	3.564	3.599
Lenalidomide	3.035	3.023	3.035	3.018	3.001	3.012	3.002	3.033
Lomustine	2.693	2.664	2.691	2.657	2.674	2.682	2.648	2.692
Melphalan	2.93	2.907	2.928	2.893	2.904	2.911	2.888	2.929
Midostaurin	3.916	3.892	3.914	3.879	3.894	3.904	3.872	3.915
Mitomycin C	3.28	3.263	3.277	3.248	3.23	3.247	3.227	3.276
Nilotinib	3.752	3.736	3.752	3.732	3.733	3.739	3.727	3.753
Olaparib	3.574	3.558	3.573	3.553	3.554	3.561	3.543	3.573
Olutasidenib	3.286	3.275	3.284	3.262	3.251	3.262	3.254	3.284
Orgovyx	3.859	3.842	3.857	3.829	3.828	3.839	3.822	3.857
Plerixafor	3.631	3.613	3.632	3.618	3.635	3.635	3.612	3.632
Pomalidomide	3.081	3.068	3.08	3.063	3.04	3.051	3.043	3.078
Pralatrexate	3.602	3.59	3.601	3.581	3.572	3.581	3.573	3.601
Prednisone	3.347	3.321	3.344	3.304	3.309	3.317	3.285	3.345
Repotrectinib	3.36	3.35	3.36	3.346	3.336	3.344	3.336	3.359
Ribociclib	3.573	3.555	3.573	3.551	3.559	3.564	3.542	3.573
Zanubrutinib	3.652	3.632	3.65	3.623	3.635	3.641	3.616	3.651

Table 4. Entropies of anticancer drugs.

In[45], thirteen drugs are considered and the highest correlation for the harmonic index (H) was identified to be

$$BP = 377.349 + 12.647[H], \quad R^2 = 0.526, \quad SE = 135.48656. \tag{12}$$

As mentioned below, the highest correlation for boiling point is obtained in the linear model that we have suggested with thirty drugs and by involving the BSO_2 index.

$$BP = 265.648 + 7.323[BSO_2], \quad R^2 = 0.848, \quad RMSE = 47.688, \quad SE = 50.4155. \tag{13}$$

In the study referenced in[45], it was proposed that, upon consideration of thirteen drugs and flash point, the highest correlation for the harmonic index (H) was identified to be

$$FP = 187.144 + 7.500[H], \quad R^2 = 0.596, SE = 45.90422. \tag{14}$$

In the study referenced in[46], it was proposed that, upon consideration of ten drugs and flash point, the highest correlation for Randic index (R) was identified to be

$$FP = 29.580 + 16.484[R], \quad R^2 = 0.760, \quad SE = 82.16599. \tag{15}$$

The highest correlation for flash point was obtained in the linear model that was suggested with thirty drugs and the BSO_2 index, as mentioned below:

$$FP = 113.959 + 4.469[BSO_2], \quad R^2 = 0.857, \quad RMSE = 28.082, \quad SE = 29.688. \tag{16}$$

This study extends the dataset and the methodological advancements beyond previous studies in[44–46]. In[44] and[46], several physicochemical properties of ten anticancer drugs were predicted using topological indices through linear regression and its derivatives. Similarly, in[45], employed nine indices to predict five physicochemical

Drugs	BP (°C)	MR (cm^3)	HAC	EM (Da)	FP (°C)	P (cm^3)
Afatinib	676.9	131.2	34	485.16	363.2	52
Alpelisib		106.7	30	441.14		42.3
Anastrozole	469.7	90	22	293.16	237.9	35.7
Belinostat		83	22	318.06		32.9
Bortezomib		103	28	384.19		40.8
Busulfan	464	50.9	14	246.02	234.4	20.2
Carmustine	309.6	46.6	12	213	141	18.5
Dasatinib		132	33	487.15		52.3
Daunorubicin	770	130	38	527.17	419.5	51.5
Erdafitinib	662.3	129.6	33	446.24	354.4	51.4
Flutamide	400.3	61.3	19	276.07	195.9	24.3
Futibatinib	733.8	115.8	31	418.17	397.6	45.9
Granisetron	532	89.8	23	312.19	275.6	35.6
Ibrutinib	715	126.1	33	440.19	386.2	50
Lenalidomide	614	66.5	19	259.09	325.1	26.3
Lomustine		57.8	15	233.09		22.9
Melphalan	473.1	78.8	19	304.07	239.9	31.2
Midostaurin		160.3	43	570.22		63.5
Mitomycin C	581.8	80.8	24	334.12	305.6	32
Nilotinib		141.8	39	529.18		56.2
Olaparib		116.9	32	434.17		46.3
Olutasidenib	603.3	93.4	25	354.08	318.7	37
Orgovyx		158.2	44	623.17		62.7
Plerixafor	657.5	151.5	36	502.44	361.8	60.1
Pomalidomide	582.9	66.6	20	273.07	306.3	26.4
Pralatrexate		125.9	35	477.17		49.9
Prednisone	573.7	94.1	26	358.17	314.8	37.3
Repotrectinib		93.2	26	355.14		37
Ribociclib	730.8	123.4	32	434.25	395.8	48.9
Zanubrutinib	713.4	133.1	35	471.22	385.2	52.8

Table 5. Chemical and physical properties of drugs.

	BP	MR	HAC	EM	FP	P
SO	0.908	0.932	0.9746	0.9502	0.9131	0.9317
RSO	0.8811	0.8904	0.9478	0.9221	0.8855	0.89
mSO	0.8781	**0.99**	0.9912	0.9756	0.885	**0.9899**
mRSO	0.83	0.9877	0.9687	0.9563	0.8387	0.9877
BSO_1	0.8866	0.974	**0.999**	**0.989**	0.8928	0.9739
BSO_2	**0.9208**	0.9453	0.9761	0.948	**0.9257**	0.945
ESO	0.8768	0.8789	0.9364	0.9077	0.8821	0.8785
EUSO	0.9113	0.9351	0.9757	0.9507	0.9163	0.9348

Table 6. Correlation between experimental properties and TIs (linear regression model).

properties of thirteen anticancer compounds using a linear regression approach. In contrast, the present study extends these previous frameworks by incorporating a larger and more diverse dataset of anticancer agents and by employing higher-order regression models, including quadratic and cubic forms. This comprehensive framework enables more accurate and robust prediction of physicochemical properties and contributes to the development of more reliable QSPR models for anticancer compounds.

The comparisons of Tables 6, 7, 8 and 10, 11, 12 show that the correlations obtained using topological indices for each physicochemical property are higher than those obtained using entropies. Consequently, in the comparative analysis, the most effective predictive topological indices for each property are utilized in this section. Tables 14, 15, and 16 present comparisons of actual and predicted values for the most effective predictive models within the linear, quadratic, and cubic regression frameworks, respectively. The results further reinforce the robustness and predictive reliability of the QSPR approach proposed in this study.

	BP	MR	HAC	EM	FP	P
SO	0.9094	0.9337	0.9752	0.9504	0.915	0.9334
RSO	0.886	0.8934	0.949	0.9222	0.8912	0.893
mSO	0.8941	**0.99**	0.9912	0.977	0.9002	**0.99**
mRSO	0.8767	0.989	0.971	0.9564	0.881	0.989
BSO_1	0.8926	0.9745	**0.999**	**0.9911**	0.8994	0.9743
BSO_2	**0.9208**	0.947	0.977	0.948	**0.9258**	0.9468
ESO	0.8823	0.8843	0.9397	0.9082	0.8883	0.884
EUSO	0.9123	0.9368	0.9763	0.9508	0.9178	0.9365

Table 7. Correlation between experimental properties and TIs (quadratic regression model).

	BP	MR	HAC	EM	FP	P
SO	0.9094	0.94	0.9796	0.9626	0.915	0.9398
RSO	0.8861	0.8977	0.9531	0.9332	0.8915	0.8974
mSO	0.9041	**0.9904**	0.9913	0.9777	0.9076	**0.9903**
mRSO	0.8913	0.9891	0.9713	0.9582	0.8915	0.9891
BSO_1	0.8931	0.9768	**0.999**	**0.9923**	0.8997	0.9767
BSO_2	**0.9209**	0.9544	0.9826	0.9634	**0.9258**	0.9542
ESO	0.8823	0.8876	0.9439	0.9206	0.8884	0.8872
EUSO	0.9124	0.9432	0.9809	0.9635	0.9178	0.943

Table 8. Correlation between experimental properties and TIs (cubic regression model).

Property∼TI	Models	Regression Equations	R^2	RMSE	F-test	p-value	SE	MPE	MAE
BP∼ BSO_2	Linear	$265.648 + 7.323(BSO_2)$	0.848	47.688	94.759	2.29×10^{-8}	10.94	−0.01	39.258
	Quadratic	$258.827 + 7.704(BSO_2) - 0.005(BSO_2)^2$	0.848	47.676	44.619	2.85×10^{-7}	10.938	−0.009	39.136
	Cubic	$241.673 + 9.292(BSO_2) - 0.047(BSO_2)^2 + 0.00034(BSO_2)^3$	0.848	47.66	27.909	2.20×10^{-6}	10.934	−0.009	39.121
MR∼ mSO	Linear	$6.071 + 11.101(mSO)$	0.980	4.492	1376.008	2.37×10^{-25}	0.82	−0.003	3.569
	Quadratic	$2.929 + 11.897(mSO) - 0.045(mSO)^2$	0.980	4.476	668.273	1.01×10^{-23}	0.817	−0.002	3.501
	Cubic	$20.024 + 5.115(mSO) + 0.771(mSO)^2 - 0.030(mSO)^3$	0.981	4.399	444.487	1.90×10^{-22}	0.803	−0.003	3.515
HAC∼ BSO_1	Linear	$-0.555 + 1.390(BSO_1)$	0.998	0.381	13329.799	4.73×10^{-39}	0.07	−0.00009	0.306
	Quadratic	$-0.640 + 1.399(BSO_1) - 0.000(BSO_1)^2$	0.998	0.381	6430.358	6.86×10^{-37}	0.07	−0.00005	0.305
	Cubic	$1.775 + 0.995(BSO_1) + 0.021(BSO_1)^2 - 0.00034(BSO_1)^3$	0.998	0.37	4377.383	2.93×10^{-35}	0.068	−0.0002	0.288
EM∼ BSO_1	Linear	$31.870 + 17.558(BSO_1)$	0.978	15.732	1247.238	9.14×10^{-25}	2.872	−0.00022	13.273
	Quadratic	$98.748 + 10.388(BSO_1) + 0.176(BSO_1)^2$	0.982	14.11	750.815	2.16×10^{-24}	2.576	−0.001	10.997
	Cubic	$232.928 - 12.074(BSO_1) + 1.332(BSO_1)^2 - 0.019(BSO_1)^3$	0.985	13.178	553.923	1.14×10^{-23}	2.406	−0.001	10.744
FP∼ BSO_2	Linear	$113.959 + 4.469(BSO_2)$	0.857	28.082	101.759	1.36×10^{-8}	6.443	−0.014	22.771
	Quadratic	$105.443 + 4.944(BSO_2) - 0.006(BSO_2)^2$	0.857	28.051	48.013	1.73×10^{-7}	6.435	−0.013	22.618
	Cubic	$108.329 + 4.677(BSO_2) + 0.001(BSO_2)^2 - 0.00006(BSO_2)^3$	0.857	28.05	30.01	1.39×10^{-6}	6.435	−0.013	22.62
P∼ mSO	Linear	$2.408 + 4.400(mSO)$	0.980	1.788	1364.087	2.68×10^{-25}	0.326	−0.003	1.42
	Quadratic	$1.172 + 4.713(mSO) - 0.018(mSO)^2$	0.980	1.782	662.371	1.14×10^{-23}	0.325	−0.002	1.393
	Cubic	$8.138 + 1.949(mSO) + 0.315(mSO)^2 - 0.012(mSO)^3$	0.981	1.75	441.323	2.08×10^{-22}	0.319	−0.003	1.399

Table 9. Regression analysis for TIs based on best predictors.

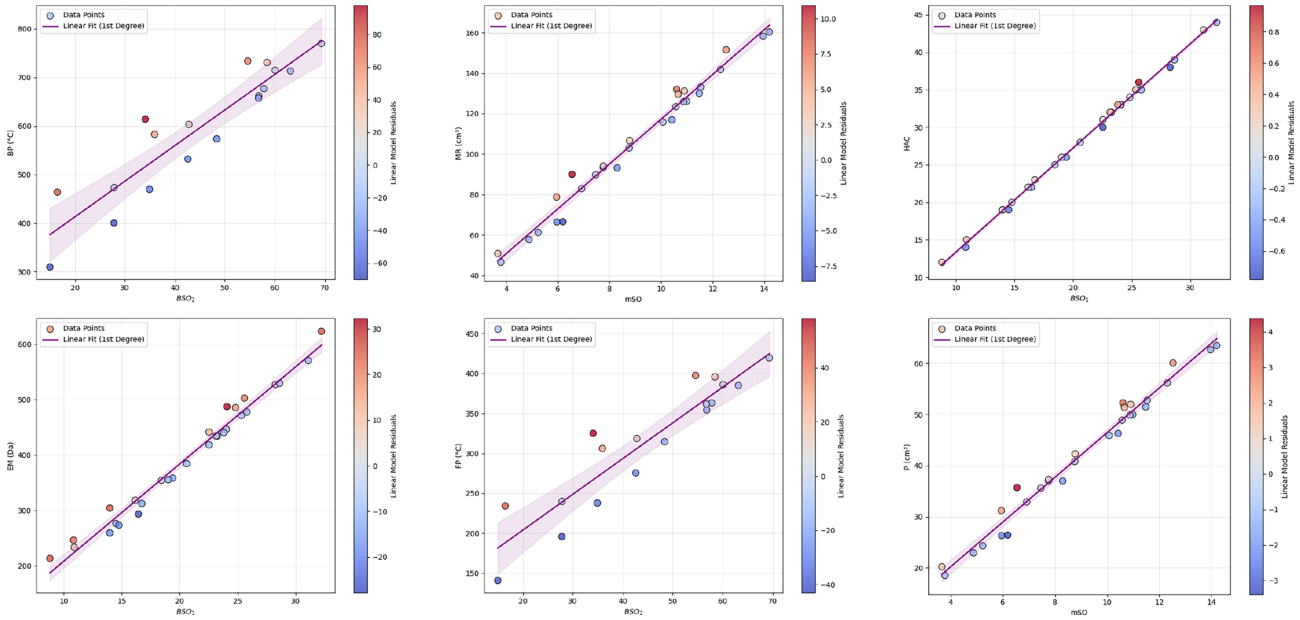

Fig. 5. The best predictors are based on the linear regression analyses.

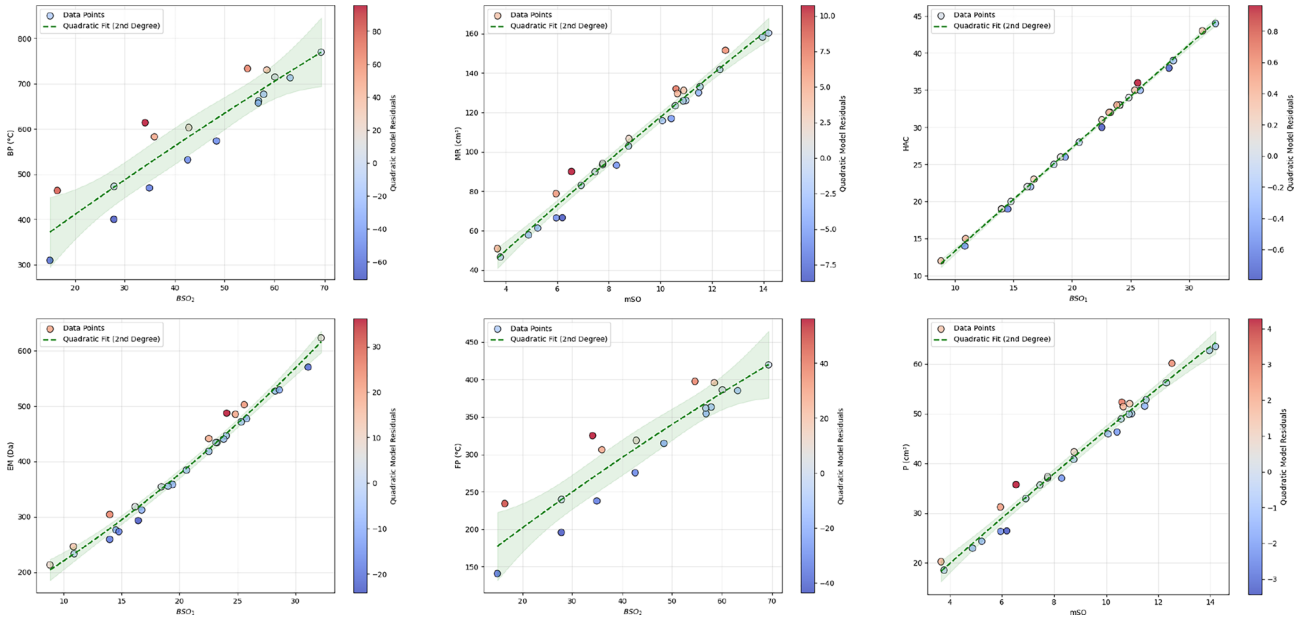

Fig. 6. The best predictors are based on the quadratic regression analyses.

Limitations and future work

The findings of this study emphasize the significance of Sombor TIs and their associated entropies in modeling the physicochemical properties of thirty selected anticancer drugs. Although the proposed models exhibit good predictive accuracy, this study is not without limitations. While a larger dataset could improve the generalizability of the findings, the primary objective was to investigate the correlations between drug properties and degree-based topological indices, along with their entropy extensions. The extension of this approach to predict biological activity or toxicity would require more comprehensive data and advanced descriptors. Despite the limited nature of the dataset, analogous sample sizes have been employed in prior QSPR studies, provided they are supported by rigorous statistical validation. The reliability of the model was confirmed through the utilization of internal validation metrics, including R^2, RMSE, and p-values. The trends obtained are consistent with those reported in the extant literature, thereby further substantiating the robustness of the methodology.

Future research directions may include the incorporation of machine learning algorithms for the enhancement of predictive capabilities, extending beyond conventional regression models to capture more

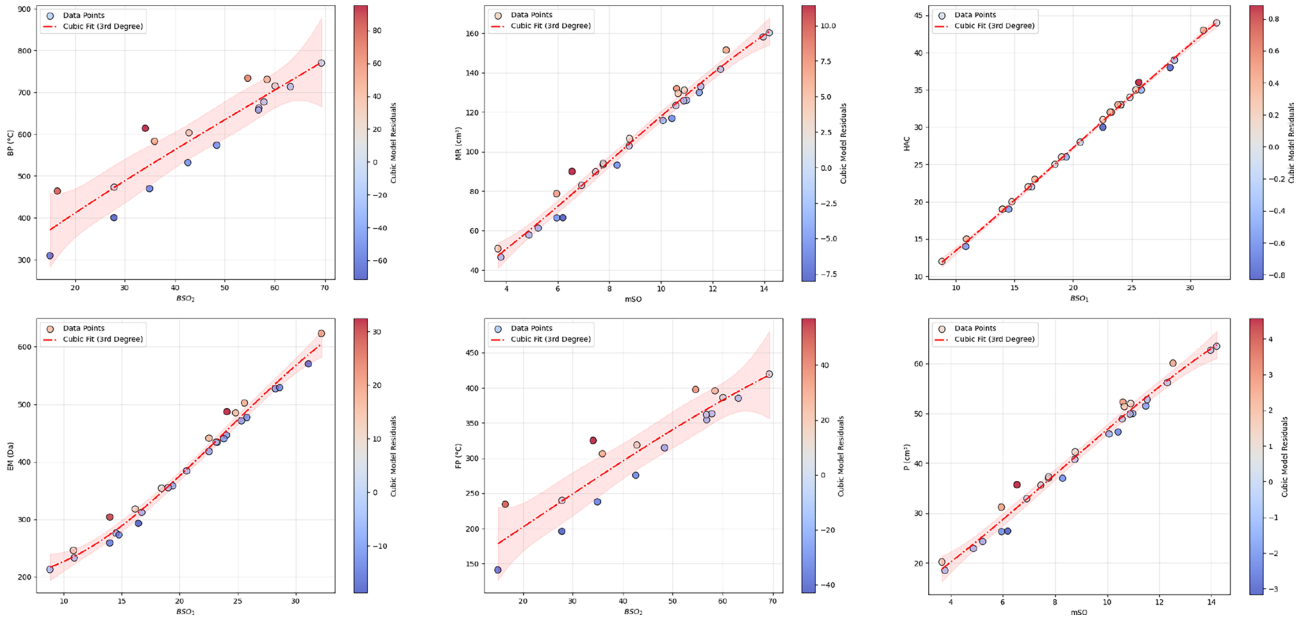

Fig. 7. The best predictors are based on the cubic regression analyses.

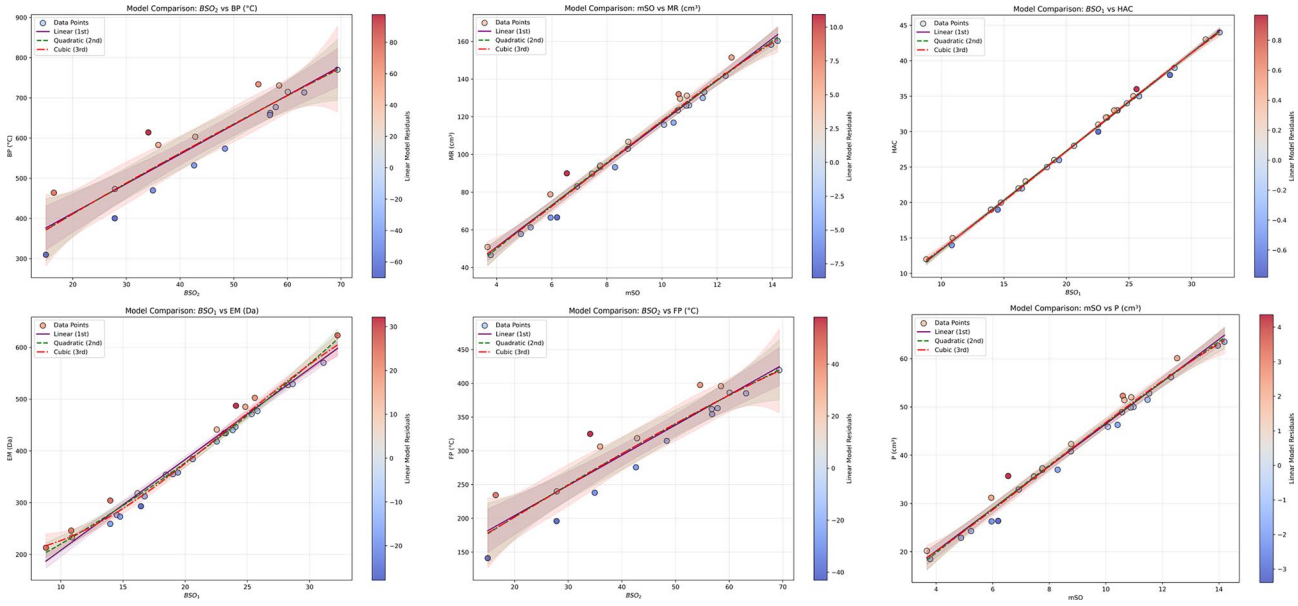

Fig. 8. Collective comparison of linear, quadratic, and cubic regression models.

intricate correlations within the data, thereby facilitating a deeper and more sophisticated understanding of the complex interplay between cancer cells and their therapeutic response to novel pharmacological agents.

Conclusion

This research makes a substantial contribution to the fields of computational chemistry, QSPR modeling, and molecular design. The findings highlight the effectiveness of degree-based topological indices in elucidating the complex nature of molecular interactions. The developed models utilize topological indices derived from hydrogen-depleted molecular graphs of anticancer drugs to quantitatively characterize molecular structure and connectivity, as well as to predict key physicochemical properties of the compounds. These predicted properties are closely associated with a drug's biological activity, stability, and interactions with cellular mechanisms, making them highly valuable for the early identification of promising drug candidates. Furthermore, the proposed models provide a scalable and generalizable framework for analyzing novel molecules, thereby laying a solid foundation for the discovery of more effective and personalized cancer therapies.

	BP	MR	HAC	EM	FP	P
$\mathcal{I}_{SO}$	0.908	0.949	0.9697	0.9349	0.9142	0.9487
$\mathcal{I}_{RSO}$	0.9078	0.9457	0.9675	0.9322	0.9138	0.9455
$\mathcal{I}_{mSO}$	0.9079	0.949	0.9698	0.935	0.9141	0.9487
$\mathcal{I}_{mRSO}$	0.907	0.9459	0.9675	0.9326	0.913	0.9456
$\mathcal{I}_{BSO_1}$	0.9084	**0.9552**	**0.9715**	**0.9373**	0.9148	**0.9550**
$\mathcal{I}_{BSO_2}$	0.9099	0.9538	0.9707	0.9361	**0.9161**	0.9536
$\mathcal{I}_{ESO}$	**0.9103**	0.9492	0.9687	0.9341	0.9160	0.949
$\mathcal{I}_{EUSO}$	0.9083	0.9496	0.97	0.9353	0.9145	0.9494

Table 10. Correlation between experimental properties and entropies (linear regression model).

	BP	MR	HAC	EM	FP	P
$\mathcal{I}_{SO}$	0.9128	0.9741	0.9952	0.9792	0.9187	0.974
$\mathcal{I}_{RSO}$	0.9137	0.9732	0.9951	0.9796	0.9193	0.973
$\mathcal{I}_{mSO}$	0.9126	0.9742	0.9952	0.9794	0.9186	0.974
$\mathcal{I}_{mRSO}$	0.9123	0.9737	0.9951	**0.9802**	0.9181	0.9735
$\mathcal{I}_{BSO_1}$	0.911	**0.9793**	0.9947	0.9783	0.9173	**0.9792**
$\mathcal{I}_{BSO_2}$	0.913	0.9781	0.9946	0.978	0.9191	0.978
$\mathcal{I}_{ESO}$	**0.9143**	0.9765	**0.9954**	0.9801	**0.9199**	0.9763
$\mathcal{I}_{EUSO}$	0.9127	0.9747	0.9952	0.9793	0.9187	0.9745

Table 11. Correlation between experimental properties and entropies (quadratic regression model).

	BP	MR	HAC	EM	FP	P
$\mathcal{I}_{SO}$	0.9153	0.9745	0.9952	0.9792	0.9205	0.9744
$\mathcal{I}_{RSO}$	0.9156	0.9734	0.9952	0.9797	0.9207	0.9732
$\mathcal{I}_{mSO}$	0.9149	0.9746	0.9953	0.9794	0.9202	0.9745
$\mathcal{I}_{mRSO}$	0.9132	0.9738	0.9953	**0.9803**	0.9187	0.9736
$\mathcal{I}_{BSO_1}$	0.913	**0.9796**	0.9948	0.9783	0.9188	**0.9795**
$\mathcal{I}_{BSO_2}$	0.9152	0.9784	0.9947	0.978	0.9207	0.9783
$\mathcal{I}_{ESO}$	**0.9161**	0.9766	**0.9956**	0.9802	**0.9211**	0.9765
$\mathcal{I}_{EUSO}$	0.9153	0.9751	0.9953	0.9793	0.9206	0.975

Table 12. Correlation between experimental properties and entropies (cubic regression model).

Property$\sim \mathcal{I}_{TI}$	Models	Regression Equations	R^2	RMSE	F-test	p-value	SE	MPE	MAE
BP$\sim \mathcal{I}_{ESO}$	Linear	$-359.745 + 295.132(\mathcal{I}_{ESO})$	0.829	50.627	82.16	6.40×10^{-8}	11.615	−0.008	42.083
	Quadratic	$248.307 - 108.530(\mathcal{I}_{ESO}) + 65.798(\mathcal{I}_{ESO})^2$	0.836	49.528	40.759	5.25×10^{-7}	11.362	−0.009	41.275
	Cubic	$-3718.492 + 3947.831(\mathcal{I}_{ESO}) - 1297.071(\mathcal{I}_{ESO})^2 + 150.648(\mathcal{I}_{ESO})^3$	0.839	49.028	26.098	3.34×10^{-6}	11.248	−0.009	40.708
MR$\sim \mathcal{I}_{BSO_1}$	Linear	$-166.030 + 81.553(\mathcal{I}_{BSO_1})$	0.912	9.412	291.868	2.42×10^{-16}	1.718	−0.002	7.729
	Quadratic	$246.893 - 183.330(\mathcal{I}_{BSO_1}) + 41.797(\mathcal{I}_{BSO_1})^2$	0.959	6.435	316.406	1.82×10^{-19}	1.175	−0.004	4.696
	Cubic	$545.895 - 476.522(\mathcal{I}_{BSO_1}) + 136.343(\mathcal{I}_{BSO_1})^2 - 10.040(\mathcal{I}_{BSO_1})^3$	0.96	6.397	205.674	3.17×10^{-18}	1.168	−0.004	4.755
HAC$\sim \mathcal{I}_{BSO_1}$	Linear	$-43.942 + 21.698(\mathcal{I}_{BSO_1})$	0.944	1.974	469.631	4.89×10^{-19}	0.36	0.003	1.533
HAC$\sim \mathcal{I}_{ESO}$	Quadratic	$65.787 - 48.751(\mathcal{I}_{ESO}) + 11.148(\mathcal{I}_{ESO})^2$	0.991	0.796	1462.179	3.01×10^{-28}	0.145	−0.001	0.663
	Cubic	$-2.910 + 19.355(\mathcal{I}_{ESO}) - 11.033(\mathcal{I}_{ESO})^2 + 2.377(\mathcal{I}_{ESO})^3$	0.991	0.777	986.073	6.95×10^{-27}	0.142	−0.001	0.623
EM$\sim \mathcal{I}_{BSO_1}$	Linear	$-492.960 + 267.074(\mathcal{I}_{BSO_1})$	0.879	37.007	202.463	2.42×10^{-14}	6.756	−0.004	31.053
EM $\sim \mathcal{I}_{mRSO}$	Quadratic	$1371.757 - 931.093(\mathcal{I}_{mRSO}) + 189.373(\mathcal{I}_{mRSO})^2$	0.961	21.01	331.244	1.01×10^{-19}	3.836	−0.003	17.492
	Cubic	$914.052 - 479.734(\mathcal{I}_{mRSO}) + 43.112(\mathcal{I}_{mRSO})^2 + 15.597(\mathcal{I}_{mRSO})^3$	0.961	20.981	213.252	2.02×10^{-18}	3.831	−0.003	17.404
FP$\sim \mathcal{I}_{BSO_2}$	Linear	$-275.920 + 181.563(\mathcal{I}_{BSO_2})$	0.839	29.753	88.799	3.67×10^{-8}	6.826	−0.012	24.407
FP$\sim \mathcal{I}_{ESO}$	Quadratic	$92.301 - 59.096(\mathcal{I}_{ESO}) + 39.021(\mathcal{I}_{ESO})^2$	0.846	29.114	43.996	3.14×10^{-7}	6.679	−0.013	23.833
	Cubic	$-1950.495 + 2029.823(\mathcal{I}_{ESO}) - 662.821(\mathcal{I}_{ESO})^2 + 77.580(\mathcal{I}_{ESO})^3$	0.849	28.889	28.006	2.15×10^{-6}	6.628	−0.012	23.374
P$\sim \mathcal{I}_{BSO_1}$	Linear	$-65.790 + 32.319(\mathcal{I}_{BSO_1})$	0.912	3.738	290.547	2.56×10^{-16}	0.683	−0.002	3.07
	Quadratic	$98.047 - 72.780(\mathcal{I}_{BSO_1}) + 16.584(\mathcal{I}_{BSO_1})^2$	0.959	2.559	314.253	1.99×10^{-19}	0.467	−0.004	1.866
	Cubic	$220.946 - 193.290(\mathcal{I}_{BSO_1}) + 55.445(\mathcal{I}_{BSO_1})^2 - 4.127(\mathcal{I}_{BSO_1})^3$	0.959	2.543	204.45	3.41×10^{-18}	0.464	−0.004	1.89

Table 13. Regression analysis for entropies of TIs based on best predictors.

Properties	Boiling point		Molar refractivity		Heavy atom count		Exact mass		Flash point		Polarizability	
Drugs	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Afatinib	676.9	689.111	131.2	127.059	34	33.965	485.16	467.84	363.2	372.369	52	50.361
Alpelisib		632.597	106.7	103.481	30	30.775	441.14	427.563		337.883	42.3	41.016
Anastrozole	469.7	521.458	90	78.737	22	22.301	293.16	320.532	237.9	270.062	35.7	31.209
Belinostat		518.397	83	82.889	22	21.921	318.06	315.739		268.194	32.9	32.854
Bortezomib		583.033	103	103.326	28	28.092	384.19	393.677		307.637	40.8	40.954
Busulfan	464	386.263	50.9	46.811	14	14.512	246.02	222.158	234.4	187.562	20.2	18.555
Carmustine	309.6	375.41	46.6	48.099	12	11.684	213	186.446	141	180.939	18.5	19.066
Dasatinib		680.191	132	123.751	33	32.923	487.15	454.689		366.926	52.3	49.05
Daunorubicin	770	773.402	130	133.476	38	38.74	527.17	528.15	419.5	423.806	51.5	52.904
Erdafitinib	662.3	681.619	129.6	124.351	33	32.865	446.24	453.952	354.4	367.798	51.4	49.287
Flutamide	400.3	469.36	61.3	64.173	19	19.611	276.07	286.559	195.9	238.271	24.3	25.436
Futibatinib	733.8	665.376	115.8	117.879	31	30.775	418.17	427.563	397.6	357.885	45.9	46.722
Granisetron	532	577.723	89.8	88.872	23	22.712	312.19	325.729	275.6	304.397	35.6	35.226
Ibrutinib	715	705.581	126.1	127.959	33	32.549	440.19	449.966	386.2	382.42	50	50.717
Lenalidomide	614	515.212	66.5	72.221	19	18.871	259.09	277.218	325.1	266.25	26.3	28.626
Lomustine		426.138	57.8	60.221	15	14.602	233.09	223.299		211.895	22.9	23.87
Melphalan	473.1	469.58	78.8	72.099	19	18.875	304.07	277.271	239.9	238.405	31.2	28.578
Midostaurin		919.26	160.3	163.626	43	42.709	570.22	578.277		512.814	63.5	64.854
Mitomycin	581.8	603.003	80.8	87.185	24	24.2	334.12	344.516	305.6	319.824	32	34.557
Nilotinib		762.856	141.8	142.656	39	39.254	529.18	534.646		417.371	56.2	56.543
Olaparib		695.562	116.9	121.709	32	31.779	434.17	440.239		376.306	46.3	48.24
Olutasidenib	603.3	579.29	93.4	92.158	25	25.084	354.08	355.683	318.7	305.353	37	36.528
Orgovyx		830.033	158.2	161.017	44	44.264	623.17	597.923		458.364	62.7	63.82
Plerixafor	657.5	681.165	151.5	145.065	36	35.028	502.44	481.271	361.8	367.52	60.1	57.497
Pomalidomide	582.9	528.855	66.6	74.807	20	19.976	273.074	291.176	306.3	274.576	26.4	29.651
Pralatrexate		681.838	125.9	126.815	35	35.321	477.17	484.976		367.932	49.9	50.264
Prednisone	573.7	619.839	94.1	92.191	26	26.433	358.17	372.713	314.8	330.098	37.3	36.541
Repotrectinib		607.529	93.2	98.153	26	25.882	355.14	365.761		322.586	37	38.904
Ribociclib	730.8	693.79	123.4	123.429	32	31.625	434.25	438.291	395.8	375.225	48.9	48.922
Zanubrutinib	713.4	728.063	133.1	134.086	35	34.669	471.22	476.741	385.2	396.139	52.8	53.146

Table 14. Comparing exact and predicted values using linear regression models for TIs of best predictors.

Properties	Boiling point		Molar refractivity		Heavy atom count		Exact mass		Flash point		Polarizability	
Drugs	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Afatinib	676.9	689.045	131.2	127.226	34	33.968	485.16	464.985	363.2	372.288	52	50.427
Alpelisib		633.394	106.7	103.845	30	30.782	441.14	422.07		338.878	42.3	41.159
Anastrozole	469.7	522.367	90	78.869	22	22.305	293.16	317.014	237.9	271.196	35.7	31.261
Belinostat		519.279	83	83.091	22	21.925	318.06	312.614		269.295	32.9	32.934
Bortezomib		584.139	103	103.689	28	28.1	384.19	387.397		309.018	40.8	41.097
Busulfan	464	384.473	50.9	45.981	14	14.499	246.02	231.965	234.4	185.328	20.2	18.229
Carmustine	309.6	373.269	46.6	47.322	12	11.662	213	203.819	141	178.266	18.5	18.76
Dasatinib		680.298	132	123.97	33	32.928	487.15	450.77		367.06	52.3	49.136
Daunorubicin	770	771.04	130	133.518	38	38.734	527.17	532.7	419.5	420.859	51.5	52.921
Erdafitinib	662.3	681.699	129.6	124.561	33	32.87	446.24	449.979	354.4	367.898	51.4	49.37
Flutamide	400.3	469.597	61.3	63.958	19	19.611	276.07	286.395	195.9	238.566	24.3	25.352
Futibatinib	733.8	665.739	115.8	118.171	31	30.782	418.17	422.07	397.6	358.339	45.9	46.837
Granisetron	532	578.838	89.8	89.153	23	22.717	312.19	321.813	275.6	305.789	35.6	35.336
Ibrutinib	715	705.162	126.1	128.11	33	32.555	440.19	445.713	386.2	381.897	50	50.777
Lenalidomide	614	516.064	66.5	72.217	19	18.87	259.09	278.208	325.1	267.314	26.3	28.625
Lomustine		425.468	57.8	59.886	15	14.59	233.09	232.888		211.058	22.9	23.738
Melphalan	473.1	469.821	78.8	72.092	19	18.874	304.07	278.254	239.9	238.705	31.2	28.575
Midostaurin		910.069	160.3	162.682	43	42.691	570.22	592.136		501.341	63.5	64.482
Mitomycin	581.8	604.035	80.8	87.446	24	24.206	334.12	339.42	305.6	321.112	32	34.66
Nilotinib		760.848	141.8	142.469	39	39.247	529.18	540.242		414.865	56.2	56.469
Olaparib		695.364	116.9	121.956	32	31.785	434.17	435.378		376.059	46.3	48.337
Olutasidenib	603.3	580.403	93.4	92.471	25	25.091	354.08	350.076	318.7	306.743	37	36.651
Orgovyx		825.449	158.2	160.185	44	44.241	623.17	616.212		452.643	62.7	63.492
Plerixafor	657.5	681.254	151.5	144.807	36	35.03	502.44	479.707	361.8	367.631	60.1	57.396
Pomalidomide	582.9	529.821	66.6	74.861	20	19.977	273.074	290.48	306.3	275.782	26.4	29.673
Pralatrexate		681.915	125.9	126.986	35	35.323	477.17	483.804		368.027	49.9	50.331
Prednisone	573.7	620.756	94.1	92.505	26	26.44	358.17	366.602	314.8	331.242	37.3	36.665
Repotrectinib		608.535	93.2	98.504	26	25.89	355.14	359.815		323.841	37	39.042
Ribociclib	730.8	693.629	123.4	123.653	32	31.631	434.25	433.32	395.8	375.024	48.9	49.01
Zanubrutinib	713.4	727.087	133.1	134.116	35	34.672	471.22	474.719	385.2	394.921	52.8	53.157

Table 15. Comparing exact and predicted values using quadratic regression models for TIs of best predictors.

Properties	Boiling point		Molar refractivity		Heavy atom count		Exact mass		Flash point		Polarizability	
Drugs	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Afatinib	676.9	688.087	131.2	127.958	34	34.05	485.16	469.523	363.2	372.449	52	50.725
Alpelisib		632.381	106.7	103.719	30	30.824	441.14	424.388		339.048	42.3	41.108
Anastrozole	469.7	523.513	90	78.015	22	22.212	293.16	311.814	237.9	271.003	35.7	30.913
Belinostat		520.475	83	82.262	22	21.83	318.06	307.304		269.094	32.9	32.596
Bortezomib		584	103	103.556	28	28.092	384.19	386.956		309.042	40.8	41.043
Busulfan	464	383.497	50.9	47.68	14	14.551	246.02	234.87	234.4	185.492	20.2	18.921
Carmustine	309.6	371.54	46.6	48.793	12	11.903	213	217.196	141	178.557	18.5	19.359
Dasatinib		679.22	132	124.639	33	33.001	487.15	454.784		367.241	52.3	49.409
Daunorubicin	770	774.218	130	134.277	38	38.786	527.17	535.609	419.5	420.324	51.5	53.23
Erdafitinib	662.3	680.637	129.6	125.244	33	32.942	446.24	453.957	354.4	368.076	51.4	49.648
Flutamide	400.3	471.16	61.3	63.559	19	19.519	276.07	281.308	195.9	238.303	24.3	25.189
Futibatinib	733.8	664.565	115.8	118.665	31	30.824	418.17	424.388	397.6	358.536	45.9	47.039
Granisetron	532	578.816	89.8	88.45	23	22.627	312.19	316.774	275.6	305.792	35.6	35.05
Ibrutinib	715	704.559	126.1	128.853	33	32.623	440.19	449.485	386.2	381.998	50	51.08
Lenalidomide	614	517.31	66.5	71.449	19	18.787	259.09	273.571	325.1	267.105	26.3	28.312
Lomustine		426.345	57.8	59.793	15	14.638	233.09	235.539		210.91	22.9	23.701
Melphalan	473.1	471.384	78.8	71.327	19	18.791	304.07	273.614	239.9	238.442	31.2	28.264
Midostaurin		940.272	160.3	160.901	43	42.586	570.22	586.298		496.26	63.5	63.757
Mitomycin	581.8	603.485	80.8	86.699	24	24.132	334.12	335.264	305.6	321.205	32	34.355
Nilotinib		763.154	141.8	142.986	39	39.287	529.18	542.46		414.477	56.2	56.68
Olaparib		694.523	116.9	122.573	32	31.843	434.17	438.574		376.2	46.3	48.589
Olutasidenib	603.3	580.346	93.4	91.872	25	25.029	354.08	346.629	318.7	306.752	37	36.407
Orgovyx		835.508	158.2	158.852	44	44.03	623.17	604.502		450.951	62.7	62.949
Plerixafor	657.5	680.187	151.5	145.196	36	35.117	502.44	484.513	361.8	367.811	60.1	57.554
Pomalidomide	582.9	530.838	66.6	74.038	20	19.883	273.074	285.243	306.3	275.611	26.4	29.337
Pralatrexate		680.856	125.9	127.714	35	35.41	477.17	488.629		368.205	49.9	50.628
Prednisone	573.7	619.914	94.1	91.907	26	26.401	358.17	364.423	314.8	331.384	37.3	36.421
Repotrectinib		607.9	93.2	98.14	26	25.841	355.14	357.096		323.948	37	38.894
Ribociclib	730.8	692.753	123.4	124.315	32	31.686	434.25	436.39	395.8	375.171	48.9	49.28
Zanubrutinib	713.4	727.29	133.1	134.869	35	34.757	471.22	479.468	385.2	394.887	52.8	53.464

Table 16. Comparing exact and predicted values using cubic regression models for TIs of best predictors.

Data availability

All data used in the research are included in the manuscript.

Received: 28 August 2025; Accepted: 13 December 2025
Published online: 23 December 2025

References

1. Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.* **3**(2), 140–146. https://doi.org/10.1038/nrd1301 (2004).
2. Todeschini, R. & Consonni, V. *Molecular Descriptors for Chemoinformatics (Vols. 1–2)* (Wiley-VCH, 2009). https://doi.org/10.1002/9783527628766
3. Randic, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **97**(23), 6609–6615. https://doi.org/10.1021/ja00856a001 (1975).
4. Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **69**(1), 17–20. https://doi.org/10.1021/ja01193a005 (1947).
5. Gutman, N. & Trinajstič, N. *Chem. Phys. Lett.* **174**, 535 (1972).
6. Trinajstič, N. *Chemical Graph Theory* 2nd edn. (CRC Press, 1992).
7. Kumar, V. & Das, S. On structure sensitivity and chemical applicability of some novel degree-based topological indices. *MATCH Commun. Math. Comput. Chem.* **92**, 165–203 (2024).
8. Zaman, S. et al. QSPR analysis of some novel drugs used in blood cancer treatment via degree based topological indices and regression models. *Polycycl. Aromat. Compd.* **44**(4), 2458–2474 (2024).
9. Arockiaraj, M. et al. QSPR analysis of distance-based structural indices for drugs in tuberculosis treatment. *Heliyon* **10**, e23981 (2024).
10. Hayat, S. et al. Predictive potential of eigenvalues-based graphical indices for determining thermodynamic properties of polycyclic aromatic hydrocarbons with applications to polyacenes. *Comput. Mater. Sci.* **238**, 112944 (2024).
11. Kara, Y. et al. Computational insights and predictive models for lung cancer molecular structures. *Chem. Pap.* **79**, 1869–1878 (2025).
12. Sağlam Özkan, Y. & Kara, Y. Topological coindices and QSPR analysis for some potential drugs used in lung cancer treatment via CoM and CoNM-polynomials. *Phys. Scr.* **99**, 105058 (2024).
13. Arockiaraj, M. et al. Comparative study of degree, neighborhood and reverse degree based indices for drugs in lung cancer treatment through QSPR analysis. *Scientific Reports* **15**(1), 3639 (2025).
14. Shenoy, B. G. et al. Statistical analysis of Revan topological indices for drugs used in treatment of tuberculosis. *Physica Scripta* **100**(3), 035203 (2025).
15. Balasubramaniyan, D. et al. Estimating physico-chemical properties of drugs for prostate cancer using degree-based and neighbourhood degree-based topological descriptors. *Physica Scripta* **99**(6), 065233 (2024).
16. Ahmed, W. et al. Topological and statistical regression study of chemical structures using graph-theoretic descriptors: applications to cancer therapeutics. *Chemical Papers*, 1-23 (2025).
17. Kara, Y. et al. Data-driven QSPR analysis of anti-cancer drugs using python-based topological techniques. *Journal of the Indian Chemical Society*, 101993 (2025).
18. Nasir, S. Topological descriptors of colorectal cancer drugs and characterizing physical properties via QSPR analysis. *International Journal of Analytical Chemistry* **1**, 5512172 (2025).
19. Gutman, I. Geometric approach to degree-based topological indices: Sombor indices. *MATCH Communications in Mathematical and in Computer Chemistry* **85**(1), 105–118 (2021).
20. Rauf, A. et al. Quantitative structure-property relationship (QSPR) modeling for evaluating fluorescence attributes across various aromatic heterocyclic compounds with ve-degree-based Sombor indices. *Chem. Pap.* **78**(11), 6343–6354 (2024).
21. Samiei, Z. & Movahedi, F. Investigating graph invariants for predicting properties of chemical structures of antiviral drugs. *Polycycl. Aromat. Compd.* **44**(10), 6696–6713 (2024).
22. Kirana, B. et al. Comparative study of Sombor index and its various versions using regression models for top priority polycyclic aromatic hydrocarbons. *Scientific Reports* **14**(1), 19841 (2024).
23. Gutman, I. et al. On Sombor indices of chemical graphs. *Symmetry* **13**(8), 1406 (2021).
24. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948).
25. Estrada, E. *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, 2011).
26. Arockiaraj, M. et al. Entropy structural characterization of zeolites BCT and DFT with bond-wise scaled comparison. *Sci. Rep.* **13**, 10874 (2023).
27. Ashraf, T. et al. Molecular insights into anticancer drugs through predictive mathematical modelling: a QSPR perspective. *Journal of Micromechanics and Molecular Physics* **10**(03), 35–52 (2025).
28. Dehmer, M. Information processing in complex networks: Graph entropy and information functionals. *Appl. Math. Comput.* **201**(1–2), 82–94 (2008).
29. Bonchev, D. & Trinajstič, N. Information theory, distance matrix, and molecular branching. *J. Chem. Phys.* **67**(10), 4517–4533. https://doi.org/10.1063/1.435385 (1977).
30. Kulli, V. R. & Gutman, I. Computation of Sombor indices of certain networks. *SSRG Int. J. Appl. Chem.* **8**(1), 1–5 (2021).
31. Kulli, V. R. On Banhatti-Sombor indices. *SSRG Int. J. Appl. Chem.* **8**(1), 21–25 (2021).
32. Kulli, V. R. On second Banhatti-Sombor indices. *Int. J. Math. Arch.* **12**(5), 11–16 (2021).
33. Gutman, I. et al. Geometric approach to vertexdegree-based topological indices-Elliptic Sombor index, theory and application. *Int. J. Quantum Chem.* **124**(2), e27346 (2024).
34. Gutman, I. Relating sombor and Euler indices. *Vojnoteh. Glas.* **72**, 1–12 (2024).
35. Tang, Z. et al. The Euler Sombor index of a graph. *Int. J. Quantum Chem.* **124**, e27387 (2024).
36. Zhang, Y. et al. Efficacy of afatinib in the treatment of patients with non-small cell lung cancer and head and neck squamous cell carcinoma: A systematic review and meta-analysis. *Front. Oncol.* **12**, 803593. https://doi.org/10.3389/fonc.2022.803593 (2022).
37. Slamon, D. J. et al. Alpelisib monotherapy for PI3K-altered, pretreated advanced breast cancer: A phase II study. *Cancer Discov.* **12**(9), 2058–2069. https://doi.org/10.1158/2159-8290.CD-21-1696 (2022).
38. Richardson, P. G. et al. Pomalidomide, bortezomib, and dexamethasone for patients with relapsed or refractory multiple myeloma previously treated with lenalidomide (OPTIMISMM): A randomised, open-label, phase 3 trial. *Lancet Oncol.* **20**(6), 781–794 (2019).
39. Montazeri, K. & Bellmunt, J. Erdafitinib for the treatment of metastatic bladder cancer. *Expert Rev. Clin. Pharmacol.* **13**(1), 1–6 (2020).
40. Hong, J. Y. et al. Pralatrexate in patients with recurrent or refractory peripheral T-cell lymphomas: A multicenter retrospective analysis. *Scientific Reports* **9**(1), 20302 (2019).
41. National Cancer Institute. https://cancer.gov/about-cancer/treatment/drugs
42. PubChem. An open chemistry database at the National Institutes of Health (NIH). https://pubchem.ncbi.nlm.nih.gov

43. Chemspider. Search and share chemistry. http://www.chemspider.com
44. Kour, S. & Sankar J, R. Machine learning regression models for predicting anti-cancer drug properties: Insights from topological indices in QSPR analysis. *Contemporary Mathematics*, 6515–6526 (2024).
45. Nasir, S. et al. Topological indices of novel drugs used in blood cancer treatment and its QSPR modeling. *AIMS Mathematics* **7**(7), 11829–11850 (2022).
46. Bokhary, S. A. U. H. et al. On topological indices and QSPR analysis of drugs used for the treatment of breast cancer. *Polycycl. Aromat. Compd.* **42**(9), 6233–6253. https://doi.org/10.1080/10406638.2021.1977353 (2022).

Author contributions
Methodology, Y.K., Y.S.O., A.B.B., and M.A.; validation, Y.K., and Y.S.O.; formal analysis, Y.K., Y.S.O., A.B.B., and M.A.; investigation, Y.K., Y.S.O., A.B.B., and M.A.; resources, Y.K., Y.S.O., and A.B.B.; visualization, Y.K., Y.S.O., A.B.B., and M.A.; supervision, Y.K. All authors have read and agreed to the current version of the manuscript.

Funding
This study was supported by Bursa Uludag University Research Projects Coordination Office under the Grant Number FGA-2025-2068. The authors extend their gratitude to the BUU BAP Unit for their invaluable support.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.A.

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.