



OPEN

# Modeling and estimation of physiochemical properties of cancer drugs using entropy measures

Qasem M. Tawhari<sup>1</sup>, Muhammad Naeem<sup>2</sup>, Abdul Rauf<sup>3</sup>, Muhammad Kamran Siddiqui<sup>4</sup> & Oladele Oyelakin<sup>5</sup>✉

Hyaluronic acid-paclitaxel conjugate is a nanoparticle-based drug delivery system that combines hyaluronic acid with paclitaxel, enhancing its solubility, stability, and targeting specificity. This conjugate shows promise in treating breast, lung, and ovarian cancers with reduced side effects. Entropy measures are used to predict physical and chemical properties of drugs. In this paper, we compute entropy measures for the hyaluronic acid-paclitaxel conjugate using the edge/connectivity partition approach. We establish a quantitative structure-property relationship using reverse entropy measures to predict physical properties of cancer drugs. Multiple linear, Ridge, Lasso, ElasticNet, and Support Vector regression models are employed using Python software. Our results show that reverse entropy measures exhibit high predictive capability for physical properties, based on the highest coefficient of determination and lowest mean squared error. We conclude that physical properties, including boiling point, enthalpy of vaporization, flash point, molar refractivity, molar volume, polarization, molecular weight, monoisotopic mass, topological polar surface area, and complexity, can be predicted using reverse entropy measures. We propose models for each relationship, including only the most significant models for estimating uncalculated physical properties.

**Keywords** Physiochemical characteristics, QSPR Analysis, drugs, Topological indices, Reverse entropy measures

The mathematical field of chemical graph theory focuses on the study of chemical graphs, which are mathematical structures representing pairwise interactions between entities<sup>1</sup>. A graph consists of two primary components: edges (bonds) and vertices (atoms), also referred to as nodes. In this representation, vertices denote atoms, while edges (bonds) represent the interactions between them<sup>2</sup>.

Mathematical chemistry, a subfield of theoretical chemistry, employs mathematical methodologies to investigate and understand the properties and dynamics of chemical compounds<sup>3,4</sup>. Topological indices are numerical values corresponding to the molecular structure of compounds. Various topological indices have been defined, including degree-based, eigenvalue-based, and distance-based indices<sup>5</sup>. Researchers have found applications for these indices in chemistry, pharmacy, and biology<sup>6</sup>.

The concept of topological indices was first introduced by Wiener<sup>7</sup>. Later, Milan Randić<sup>8</sup> proposed the Randić index. Amic et al.<sup>9</sup> and Bollobas et al.<sup>10</sup> subsequently suggested a generalization of the Randić index. In 2010, Trinajstić and Zhou introduced the sum connectivity index<sup>11</sup>. Zhong<sup>12</sup> defined the harmonic index in 2012. Nikolic et al.<sup>13</sup> presented a revised version of the second Zagreb index, while Fath-Tabar<sup>14</sup> introduced the third Zagreb index in 2011. In 2015, Munir et al.<sup>15</sup> converted degree-based indices into entropy measures. Ranjini et al.<sup>16</sup> proposed revised Zagreb indices in 2013. Notably, Shannon<sup>17</sup> laid the foundation for entropy metrics in 1948.

Researchers initiated investigations into the entropy value of network systems in the late 1950s, inspired by Shannon's influential work<sup>18</sup>. Rashevsky employs the concept of entropy measures to quantify the structural complexity of a graph. Shannon's entropy is employed to quantify the intricacy of the graph in this particular

<sup>1</sup>Department of Mathematics, College of Science, Jazan University, Jazan 45142, Saudi Arabia. <sup>2</sup>Department of Mathematics, National University of Sciences and Technology (NUST), Islamabad, Pakistan. <sup>3</sup>Department of Mathematics, Air University Multan Campus, Multan, Pakistan. <sup>4</sup>Department of Mathematics, Comsats University Islamabad, Lahore, Pakistan. <sup>5</sup>Chemistry Unit, School of Arts & Sciences, University of The Gambia, Serrekunda, The Gambia. ✉email: ooyelakin@utg.edu.gm

scenario. Mowshowitz<sup>19</sup> subsequently conducted an examination of the characteristics of graph entropy and conducted extensive measurements pertaining to his particular application. Entropy indicators based on graphs have been employed in diverse disciplines such as biology, chemistry, and computer science for the purpose of characterizing patterns<sup>20</sup>. The entropy proposed by Kornes<sup>21</sup> serves as another notable illustration.

Entropy measures have been developed using several graph invariances, including eigenvalues and connection information<sup>22</sup>, degree-based graph entropy<sup>23</sup>, and distance-based graph entropy<sup>24</sup>. Kulli introduced the concept of reverse degree-based indices<sup>25</sup>. Many researchers calculate the reverse degree-based indices for certain chemical structures. Wei Gao et al. calculated the indices for dendrimers<sup>26</sup>. Koam et al. calculated the indices for third type of chain hex-derived network<sup>27</sup>. Dongming Zhao et al. examine the polycyclic metal-organic network<sup>28</sup>. For more information about reverse degree-based indices calculations for certain chemical structures, see<sup>29–31</sup>.

Recently, Furtula et al.<sup>32</sup> did an analysis of the popular degree-based indices using the data of the octane isomer. They concluded that the symmetric division degree index is the descriptor. Recently, Rauf et al.<sup>33</sup> developed the QSPR between degree-based entropy and physical properties. They show that entropy measures are helpful in predicting physical properties. QSPR is a highly effective analytical methodology utilized to transform a given molecule into a series of numerical values that accurately depict its intrinsic chemical and physical characteristics. Several researchers have conducted studies on statistical linear, quadratic, and polynomial regression models to analyze the relationship between indices and physical properties<sup>34–36</sup>.

Motivated by these authors' works, we have developed the QSPR between the physical properties of cancer drugs and reverse degree-based entropy measures. We found that the entropy measures show the best correlation with physical properties. These results motivate and are helpful for researchers in predicting physical properties.

### Degree based entropy of a graph

Let  $Z = (V, E)$  be a finite, simple, and connected graph of order  $p$ , size  $q$  edges and  $\mathcal{F}$  be a real valued function. The **degree** of a vertex  $r$  is denoted by  $d(r)$  and is defined as the number of edges attached with vertex  $r$ . Kulli<sup>37</sup> introduced the **reverse degree**  $\Upsilon(r)$  defined as  $\Upsilon(r) = \Delta(Z) - d(r) + 1$ , where  $\Delta(Z)$  is the maximum degree of vertex among the vertices of a graph  $Z$ . For the undefined term of graph theory, see<sup>38</sup>. The entropy function of graph  $Z$  is defined as follows:

$$\mathcal{E}_{\mathcal{F}}(Z) = - \sum_{i=1}^p \frac{\mathcal{F}(r_i)}{\sum_{j=1}^p \mathcal{F}(r_j)} \log \left[ \frac{\mathcal{F}(r_i)}{\sum_{j=1}^p \mathcal{F}(r_j)} \right]. \quad (1)$$

Now, if  $r_i \in V$  and  $\mathcal{F}(r_i)$  is an information function that represent the degree of vertex  $r_i$ , denoted by  $\mathcal{F}(r_i) = \Upsilon(r_i)$ , then equation (1) becomes

$$\begin{aligned} \mathcal{E}_{\mathcal{F}}(Z) &= - \sum_{i=1}^p \frac{\Upsilon(r_i)}{\sum_{j=1}^p \Upsilon(r_j)} \log \left[ \frac{\Upsilon(r_i)}{\sum_{j=1}^p \Upsilon(r_j)} \right], \\ \mathcal{E}_{\mathcal{F}}(Z) &= \log \left( \sum_{j=1}^p \Upsilon(r_j) \right) - \frac{1}{(\sum_{j=1}^p \Upsilon(r_j))} \sum_{i=1}^p [\hbar(r_i) \log(\Upsilon(r_i))]. \end{aligned}$$

By simplifying the equation and using  $\sum_{j=1}^p \Upsilon(r_j) = 2q$ ,

$$\mathcal{E}_{\mathcal{F}}(Z) = \log(2q) - \frac{1}{2q} \log \left[ \prod_{i=1}^p \Upsilon(r_i)^{\Upsilon(r_i)} \right]. \quad (2)$$

### Edge weight-based entropy of graph

In 2014, Chen et al. introduced the concept of the entropy of an edge weight graph. For an edge weight graph  $P = (V(Z); E(Z) : \mathcal{F}(rs))$ , where  $E(Z)$  is the edge set,  $V(Z)$  is the vertex set, and  $\mathcal{F}(rs)$  is the edge weight of the edge  $rs$  in  $Z$ . The entropy is defined as

$$\mathcal{E}_{\mathcal{F}}(Z) = - \sum_{r' s' \in E(Z)} \frac{\mathcal{F}(r' s')}{\sum_{rs \in E(Z)} \mathcal{F}(rs)} \log \left[ \frac{\mathcal{F}(r' s')}{\sum_{rs \in E(Z)} \mathcal{F}(rs)} \right]. \quad (3)$$

By using Table 1 and simplifying the equation (3), we can derive the entropy measures written in Table 2.

### Methodology

In this section, we present the working methodology employed in this study. In “[Hyaluronic acid-paclitaxel conjugate](#)”, we compute the entropy measures for hyaluronic acid-paclitaxel conjugate to estimate their physical properties. The following steps are used to compute the entropy measures:

- We convert the hyaluronic acid-paclitaxel conjugate structure into a molecular graph by considering atoms as vertices and chemical bonds as edges.
- We partition the vertices and edges of the graph based on the reverse degree.

Name of Index	Notation	Formula	Formula expansion
Randić $\alpha = -1, \frac{-1}{2}, 1, \frac{1}{2}$	$R_\alpha(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{R_\alpha}(rs)$	$\sum_{rs \in E(Z)} (\Upsilon_r \times \Upsilon_s)^\alpha$
Atom bond connectivity	$ABC(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{ABC}(rs)$	$\sum_{rs \in E(Z)} \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}$
Geometric arithmetic	$GA(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{GA}(rs)$	$\sum_{rs \in E(Z)} \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s}$
First Zagreb	$M_1(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{M_1}(rs)$	$\sum_{rs \in E(Z)} (\Upsilon_r + \Upsilon_s)$
Second Zagreb	$M_2(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{M_2}(rs)$	$\sum_{rs \in E(Z)} (\Upsilon_r \times \Upsilon_s)$
Hyper Zagreb	$HM(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{HM}(rs)$	$\sum_{rs \in E(Z)} (\Upsilon_r + \Upsilon_s)^2$
Forgotten	$F(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_F(rs)$	$\sum_{rs \in E(Z)} [(\Upsilon_r)^2 + (\Upsilon_s)^2]$
Augmented Zagreb	$AZI(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{AZI}(rs)$	$\sum_{rs \in E(Z)} \left( \frac{\Upsilon_r \times \Upsilon_s}{\Upsilon_r + \Upsilon_s - 2} \right)^3$
Balaban	$J(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_J(rs)$	$\sum_{rs \in E(Z)} \left( \frac{q}{q-p+2} \times \frac{1}{\sqrt{\Upsilon_r \times \Upsilon_s}} \right)$
Redefined first Zagreb	$ReZG_1(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{ReZG_1}(rs)$	$\sum_{rs \in E(Z)} \frac{\Upsilon_r + \Upsilon_s}{\Upsilon_r \times \Upsilon_s}$
Redefined second Zagreb	$ReZG_2(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{ReZG_2}(rs)$	$\sum_{rs \in E(Z)} \frac{\Upsilon_r \times \Upsilon_s}{\Upsilon_r + \Upsilon_s}$
Redefined third Zagreb	$ReZG_3(Z)$	$\sum_{rs \in E(Z)} \mathcal{F}_{ReZG_3}(rs)$	$\sum_{rs \in E(Z)} (\Upsilon_r \times \Upsilon_s)(\Upsilon_r + \Upsilon_s)$

**Table 1.** Notation and formulas of the reverse degree based indices.

Name of entropies	Notation	Formula
Randić $\alpha = -1, \frac{-1}{2}, 1, \frac{1}{2}$	$\mathcal{E}_\alpha(Z)$	$\log(R_\alpha) - \frac{1}{R_\alpha} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{R_\alpha}(rs)]^{\mathcal{F}_{R_\alpha}(rs)} \right]$
Atom bond connectivity	$\mathcal{E}_{ABC}(Z)$	$\log(ABC) - \frac{1}{ABC} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{ABC}(rs)]^{\mathcal{F}_{ABC}(rs)} \right]$
Geometric arithmetic	$\mathcal{E}_{GA}(Z)$	$\log(GA) - \frac{1}{GA} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{GA}(rs)]^{\mathcal{F}_{GA}(rs)} \right]$
First Zagreb	$\mathcal{E}_{M_1}(Z)$	$\log(M_1) - \frac{1}{M_1} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{M_1}(rs)]^{\mathcal{F}_{M_1}(rs)} \right]$
Second Zagreb	$\mathcal{E}_{M_2}(Z)$	$\log(M_2) - \frac{1}{M_2} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{M_2}(rs)]^{\mathcal{F}_{M_2}(rs)} \right]$
Hyper Zagreb	$\mathcal{E}_{HM}(Z)$	$\log(HM) - \frac{1}{HM} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{HM}(rs)]^{\mathcal{F}_{HM}(rs)} \right]$
Forgotten	$\mathcal{E}_F(Z)$	$\log(F) - \frac{1}{F} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_F(rs)]^{\mathcal{F}_F(rs)} \right]$
Augmented Zagreb	$\mathcal{E}_{AZI}(Z)$	$\log(AZI) - \frac{1}{AZI} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{AZI}(rs)]^{\mathcal{F}_{AZI}(rs)} \right]$
Balaban	$\mathcal{E}_J(Z)$	$\log(J) - \frac{1}{J} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_J(rs)]^{\mathcal{F}_J(rs)} \right]$
Redefined first Zagreb	$\mathcal{E}_{ReZG_1}(Z)$	$\log(ReZG_1) - \frac{1}{ReZG_1} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{ReZG_1}(rs)]^{\mathcal{F}_{ReZG_1}(rs)} \right]$
Redefined second Zagreb	$\mathcal{E}_{ReZG_2}(Z)$	$\log(ReZG_2) - \frac{1}{ReZG_2} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{ReZG_2}(rs)]^{\mathcal{F}_{ReZG_2}(rs)} \right]$
Redefined third Zagreb	$\mathcal{E}_{ReZG_3}(Z)$	$\log(ReZG_3) - \frac{1}{ReZG_3} \log \left[ \prod_{rs \in E(Z)} [\mathcal{F}_{ReZG_3}(rs)]^{\mathcal{F}_{ReZG_3}(rs)} \right]$

**Table 2.** Notation and formulas of entropy measures.

- We compute the entropy measures and plot the graphical representation using Maple software.

In “Statistical analysis of entropy measures”, we develop the statistical analysis of entropy measures. The following steps are used for statistical analysis:

- We consider a specific class of drug.
- We take the chemical structures of the drug and convert them into molecular graphs by considering atoms as vertices and chemical bonds as edges.

- We use newGraph software to compute the adjacency matrix from the graph.
- We propose a Maple-based algorithm to compute the entropy measures based on the adjacency matrix.
- We obtain the physical properties of cancer drugs from <https://pubchem.ncbi.nlm.nih.gov/> and <https://www.chemspider.com/>.
- We develop the statistical analysis between entropy measures and the physical properties using Python.

### Hyaluronic acid-paclitaxel conjugate

Cancer is widely recognized as a prominent contributor to global mortality rates, with a persistent upward trend in fatality rates. The primary culprits behind these deaths are breast, stomach, lung, and colon cancers. Despite significant advancements in the field of cancer biology and the development of various therapeutic approaches to combat cancer, there persist challenges in effectively treating both primary and metastatic forms of the disease. Furthermore, it is important to acknowledge the presence of drawbacks in existing anticancer medications, as they often exhibit a lack of specificity and a high level of toxicity, thereby significantly impairing their effectiveness. Significant advancements have been made in the field of molecularly-targeted cancer treatment in recent years.

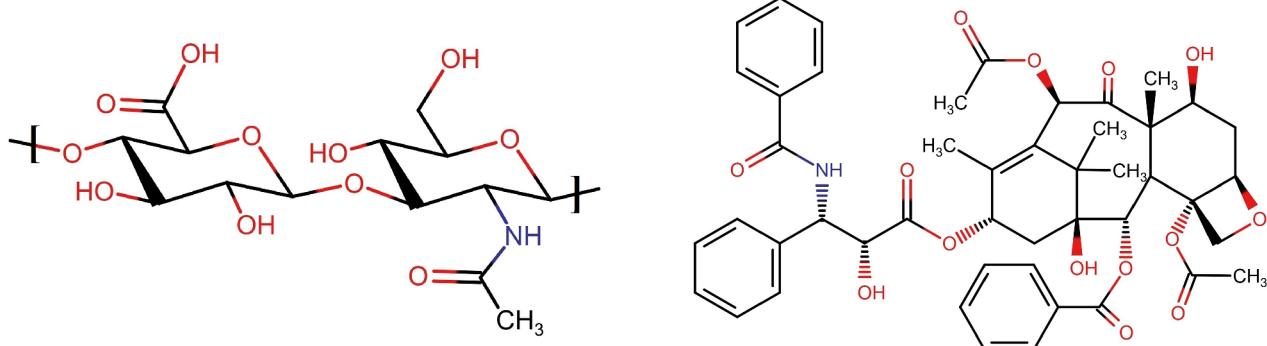
Hyaluronic acid (HA) is an endogenous compound. The compound in question is a polymer of glycosaminoglycan, consisting of a linear arrangement of D-glucuronic acid and N-acetyl-D-glucosamine units. These units are connected through alternating  $\beta$ -1,4- and  $\beta$ -1,3-glycosidic bonds. The primary structure of the disaccharide is considered to be energetically stable<sup>39</sup>. Hyaluronic acid (HA) exhibits considerable potential as a cancer therapeutic agent owing to its distinctive properties, including biodegradability, biocompatibility, non-toxicity, hydrophilicity, and non-immunogenicity. Moreover, the overexpression of HA receptors has been observed on numerous tumor cells, further supporting its promise as a cancer drug. Hyaluronic acid (HA) is currently being recognized as a promising platform for effectively targeting cells that overexpress CD44. The primary objective of utilizing HA in this context is to enhance the efficacy of anticancer treatments<sup>40-42</sup>. Hyaluronic acid (HA) exhibits promising characteristics as a drug carrier and drug-targeting agent. Paclitaxel (PTX) is a pharmacological agent that has demonstrated efficacy in the treatment of various malignancies such as bladder, lung, breast, esophageal, ovarian, and prostate cancers, among others<sup>43</sup>. Although the administration of PTX also faces certain limitations, such as its limited solubility and associated side effects, as well as the excipients commonly employed in its formulation. The initial proposal by Ringsdorf introduced a technique for the synthesis of polymeric macromolecule-drug conjugates. This method was specifically designed to facilitate the targeted delivery of small hydrophobic drug molecules to their intended sites of action<sup>44</sup>. The primary benefits of HA-PTX conjugates include enhanced water solubility and maintained activity. Moreover, these conjugates can be employed as targeted drug delivery systems to enhance the effectiveness of anti-tumor treatment<sup>45-47</sup>.

Figure 1 shows the unit chemical structures of hyaluronic acid and paclitaxel. Figures 2 and 3 illustrates the molecular graph of hyaluronic acid-paclitaxel conjugate for  $s = 1$  and  $s = 2$ , respectively. If  $s < 1$ , then it means that hyaluronic acid-paclitaxel conjugate has degraded and as result breakdown into two or more parts (glucuronic acid, N-acetylglucosamine and paclitaxel). If  $s \geq 1$  then it means that the hyaluronic acid-paclitaxel conjugate polymerization has occurred equal to value of  $s$ . For example, if  $s = 2$  then degree of polymerization is 2.

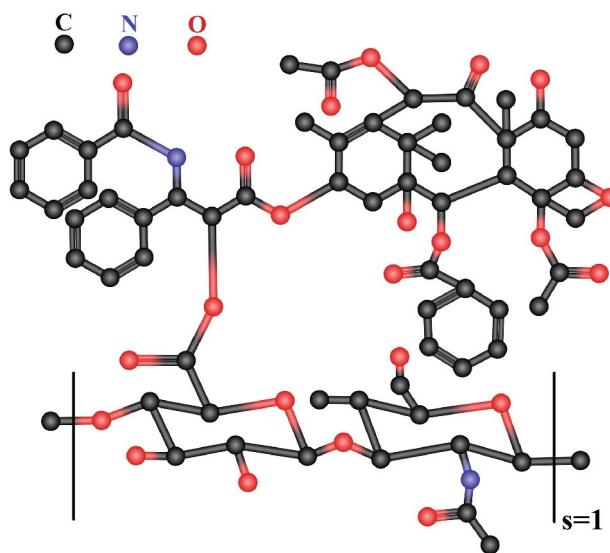
### Main results

Let HA be the molecular structure of hyaluronic acid-paclitaxel conjugate. We partitioned the edges, base on the reverse degree as list in Table 3.

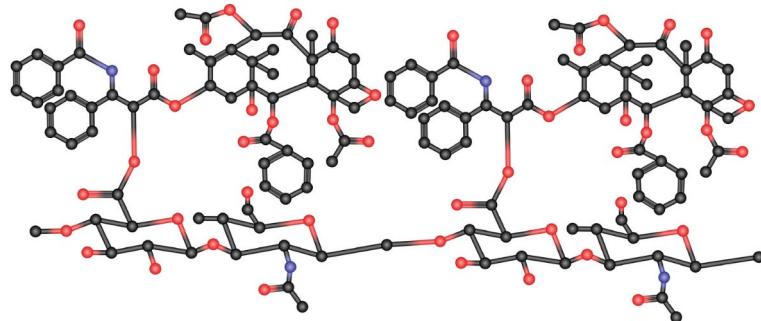
Using Table 3, we will calculate the following entropy measures.



**Fig. 1.** Unit structures of hyaluronic acid and paclitaxel.



**Fig. 2.** Molecular graph of hyaluronic acid-paclitaxel conjugate for  $s = 1$ .



**Fig. 3.** Molecular graph of hyaluronic acid-paclitaxel conjugate for  $s = 2$ .

Notations for edges	$(\Upsilon_r, \Upsilon_s)$	Frequency of edges
$E_1$	$(2, 2)$	$s$
$E_2$	$(2, 3)$	$7s$
$E_3$	$(3, 3)$	$19s - 1$
$E_4$	$(2, 4)$	$3s$
$E_5$	$(3, 4)$	$32s - 1$
$E_6$	$(4, 4)$	$13s + 1$
$E_7$	$(2, 5)$	$4s$
$E_8$	$(3, 5)$	$16s$
$E_9$	$(4, 5)$	$s + 1$

**Table 3.** Edge partition of HA based on reverse degree.

### Atom bond connectivity entropy

$$\begin{aligned}
 \mathcal{E}_{ABC}(HA) = & \log(ABC) - \frac{1}{(ABC)} \log \left[ \prod_{cd \in E_1} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \right. \\
 & \times \prod_{cd \in E_2} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \\
 & \times \prod_{cd \in E_3} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \times \prod_{cd \in E_4} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \\
 & \times \prod_{cd \in E_5} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \\
 & \times \prod_{cd \in E_6} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \times \prod_{cd \in E_7} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \\
 & \times \prod_{cd \in E_8} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \\
 & \times \left. \prod_{cd \in E_9} \left[ \sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}} \right]^{\sqrt{\frac{\Upsilon_r + \Upsilon_s - 2}{\Upsilon_r \times \Upsilon_s}}} \right], \\
 \mathcal{E}_{ABC}(HA) = & \log [62.6009s - 0.1082] - \frac{1}{[62.6009s - 0.1082]} \left[ \log \left[ (s) \left( \frac{1}{\sqrt{2}} \right)^{\frac{1}{\sqrt{2}}} \right] + \log \left[ (7s) \left( \sqrt{\frac{3}{6}} \right)^{\sqrt{\frac{3}{6}}} \right] \right. \\
 & + \log \left[ (19s - 1) \left( \sqrt{\frac{4}{9}} \right)^{\sqrt{\frac{4}{9}}} \right] + \log \left[ (3s) \left( \frac{1}{\sqrt{2}} \right)^{\frac{1}{\sqrt{2}}} \right] \\
 & + \log \left[ (32s - 1) \left( \sqrt{\frac{5}{12}} \right)^{\sqrt{\frac{5}{12}}} \right] + \log \left[ (13s + 1) \left( \sqrt{\frac{6}{16}} \right)^{\sqrt{\frac{6}{16}}} \right] \\
 & \left. + \log \left[ (4s) \left( \frac{1}{\sqrt{2}} \right)^{\frac{1}{\sqrt{2}}} \right] + \log \left[ (16s) \left( \sqrt{\frac{6}{15}} \right)^{\sqrt{\frac{6}{15}}} \right] + \log \left[ (s + 1) \left( \sqrt{\frac{7}{20}} \right)^{\sqrt{\frac{7}{20}}} \right] \right].
 \end{aligned}$$

### Geometric arithmetic entropy

$$\begin{aligned} \mathcal{E}_{GA}(HA) = & \log(GA) - \frac{1}{(GA)} \log \left[ \prod_{cd \in E_1} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \times \prod_{cd \in E_2} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \right. \\ & \times \prod_{cd \in E_3} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \times \prod_{cd \in E_4} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \times \prod_{cd \in E_5} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \\ & \times \prod_{cd \in E_6} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \times \prod_{cd \in E_7} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \\ & \times \prod_{cd \in E_8} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \times \prod_{cd \in E_9} \left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]^{\left[ \frac{2\sqrt{\Upsilon_r \times \Upsilon_s}}{\Upsilon_r + \Upsilon_s} \right]} \Big] \\ \mathcal{E}_{GA}(HA) = & \log [93.9919s + 1.9836] - \frac{1}{93.9919s + 1.9836} \left[ \log \left[ (s) \left[ \frac{2\sqrt{4}}{4} \right]^{\left( \frac{2\sqrt{4}}{4} \right)} \right] \right. \\ & + \log \left[ (7s) \left[ \frac{2\sqrt{6}}{5} \right]^{\left( \frac{2\sqrt{6}}{5} \right)} \right] + \log \left[ (19s - 1) \left[ \frac{2\sqrt{9}}{6} \right]^{\left( \frac{2\sqrt{9}}{6} \right)} \right] + \log \left[ (3s) \left[ \frac{2\sqrt{8}}{6} \right]^{\left( \frac{2\sqrt{8}}{6} \right)} \right] \\ & + \log \left[ (32s - 1) \left[ \frac{2\sqrt{12}}{7} \right]^{\left( \frac{2\sqrt{12}}{7} \right)} \right] + \log \left[ (13s + 1) \left[ \frac{2\sqrt{16}}{8} \right]^{\left( \frac{2\sqrt{16}}{8} \right)} \right] + \log \left[ (4s) \left[ \frac{2\sqrt{10}}{7} \right]^{\left( \frac{2\sqrt{10}}{7} \right)} \right] \\ & \left. + \log \left[ (16s) \left[ \frac{2\sqrt{15}}{8} \right]^{\left( \frac{2\sqrt{15}}{8} \right)} \right] + \log \left[ (s + 1) \left[ \frac{2\sqrt{20}}{9} \right]^{\left( \frac{2\sqrt{20}}{9} \right)} \right] \right]. \end{aligned}$$

### First zagreb entropy

$$\begin{aligned} \mathcal{E}_{M_1}(HA) = & \log(M_1) - \frac{1}{(M_1)} \log \left[ \prod_{cd \in E_1} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_2} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \\ & \times \left[ \prod_{cd \in E_3} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \\ & \times \left[ \prod_{cd \in E_4} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_5} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_6} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \\ & \times \left[ \prod_{cd \in E_7} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_8} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_9} [\Upsilon_r + \Upsilon_s]^{(\Upsilon_r + \Upsilon_s)} \right] \\ \mathcal{E}_{M_1}(HA) = & \log(664s + 4) - \frac{1}{664s + 4} \left[ \log [(s) \times (4)^4] + \log [(7s) \times (5)^5] + \log [(19s - 1) \times (6)^6] \right. \\ & + \log [(3s) \times (6)^6] + \log [(32s - 1) \times (7)^7] + \log [(13s + 1) \times (8)^8] \\ & \left. + \log [(4s) \times (7)^7] + \log [(16s) \times (8)^8] + \log [(s + 1) \times (9)^9] \right]. \end{aligned}$$

### Second zagreb entropy

$$\begin{aligned}\mathcal{E}_{M_2}(HA) &= \log(M_2) - \frac{1}{(M_2)} \log \left[ \prod_{cd \in E_1} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_2} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \\ &\quad \times \left[ \prod_{cd \in E_3} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \\ &\quad \times \left[ \prod_{cd \in E_4} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_5} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_6} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \\ &\quad \times \left[ \prod_{cd \in E_7} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_8} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \times \left[ \prod_{cd \in E_9} [\Upsilon_r \times \Upsilon_s]^{(\Upsilon_r \times \Upsilon_s)} \right] \\ \mathcal{E}_{M_2}(HA) &= \log(113s + 15) - \frac{1}{113s + 15} [\log [(s) \times (4)^4] + \log [(7s) \times (6)^6] + \log [(19s - 1) \times (9)^9] \\ &\quad + \log [(3s) \times (8)^8] + \log [(32s - 1) \times (12)^{12}] + \log [(13s + 1) \times (16)^{16}] \\ &\quad + \log [(4s) \times (10)^{10}] + \log [(16s) \times (15)^{15}] + \log [(s + 1) \times (20)^{20}]].\end{aligned}$$

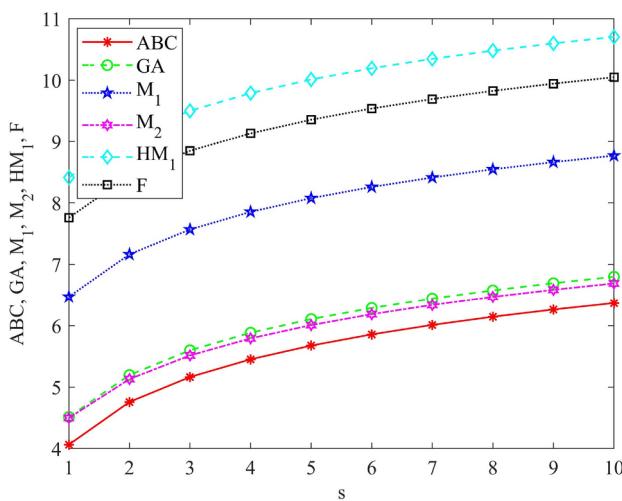
### Hyper first zagreb entropy

$$\begin{aligned}\mathcal{E}_{HM_1}(HA) &= \log(HM_1) - \frac{1}{(HM_1)} \log \left[ \prod_{cd \in E_1} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_2} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \\ &\quad \times \left[ \prod_{cd \in E_3} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \\ &\quad \times \left[ \prod_{cd \in E_4} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_5} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_6} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \\ &\quad \times \left[ \prod_{cd \in E_7} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_8} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_9} [(\Upsilon_r + \Upsilon_s)^2]^{(\Upsilon_r + \Upsilon_s)^2} \right] \\ \mathcal{E}_{HM_1}(HA) &= \log(4684s + 60) - \frac{1}{4684s + 60} [\log [(s) \times (16)^{16}] + \log [(7s) \times (25)^{25}] \log [(19s - 1) \times (36)^{36}] \\ &\quad + \log [(3s) \times (36)^{36}] + \log [(32s - 1) \times (49)^{49}] + \log [(13s + 1) \times (64)^{64}] \\ &\quad + \log [(4s) \times (49)^{49}] + \log [(16s) \times (64)^{64}] + \log [(s + 1) \times (81)^{81}]].\end{aligned}$$

### Forgotten entropy

$$\begin{aligned}\mathcal{E}_F(HA) &= \log(F) - \frac{1}{(F)} \log \left[ \prod_{cd \in E_1} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \\ &\quad \times \left[ \prod_{cd \in E_2} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_3} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \\ &\quad \times \left[ \prod_{cd \in E_4} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_5} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \\ &\quad \times \left[ \prod_{cd \in E_6} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_7} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \\ &\quad \times \left[ \prod_{cd \in E_8} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \times \left[ \prod_{cd \in E_9} [(\Upsilon_r)^2 + (\Upsilon_s)^2]^{(\Upsilon_r)^2 + (\Upsilon_s)^2} \right] \\ \mathcal{E}_F(HA) &= \log(2418s + 30) - \frac{1}{2418s + 30} [\log [(s) \times (8)^8] + \log [(7s) \times (13)^{13}] \\ &\quad + \log [(19s - 1) \times (18)^{18}] + \log [(3s) \times (20)^{20}] + \log [(32s - 1) \times (25)^{25}] \\ &\quad + \log [(13s + 1) \times (32)^{32}] + \log [(4s) \times (29)^{29}] + \log [(16s) \times (34)^{34}] \\ &\quad + \log [(s + 1) \times (41)^{41}]].\end{aligned}$$

Graphical representations and numerical values of entropy measures are shown in Fig. 4 and Table 4.



**Fig. 4.** Graphical representation of  $\mathcal{E}_{ABC}$ ,  $\mathcal{E}_{GA}$ ,  $\mathcal{E}_{M_1}$ ,  $\mathcal{E}_{M_2}$ ,  $\mathcal{E}_{HM_1}$  and  $\mathcal{E}_F$ .

[s]	$\mathcal{E}_{ABC}$	$\mathcal{E}_{GA}$	$\mathcal{E}_{M_1}$	$\mathcal{E}_{M_2}$	$\mathcal{E}_{HM_1}$	$\mathcal{E}_F$
[1]	4.0652	4.5147	6.4724	4.4986	8.4133	7.7584
[2]	4.7592	5.1975	7.1625	5.1314	9.1001	8.4454
[3]	5.165	5.5994	7.567	5.5159	9.5034	8.8488
[4]	5.4528	5.8854	7.8542	5.7929	9.7901	9.1355
[5]	5.676	6.1075	8.077	6.0096	10.0126	9.358
[6]	5.8584	6.2891	8.2591	6.1876	10.1945	9.5399
[7]	6.0126	6.4427	8.4131	6.3387	10.3483	9.6938
[8]	6.1461	6.5759	8.5466	6.4699	10.4816	9.8271
[9]	6.2639	6.6934	8.6643	6.5859	10.5992	9.9447
[10]	6.3693	6.7985	8.7696	6.6898	10.7044	10.0499

**Table 4.** Numerical representation of the reverse entropy measures.

### Statistical analysis of entropy measures

Quantitative structure-property relationship (QSPR) studies have emerged as a vital tool in predicting physical properties of molecules using topological indices. These indices, derived from molecular graphs, encode structural information that correlates with physical properties of drugs.

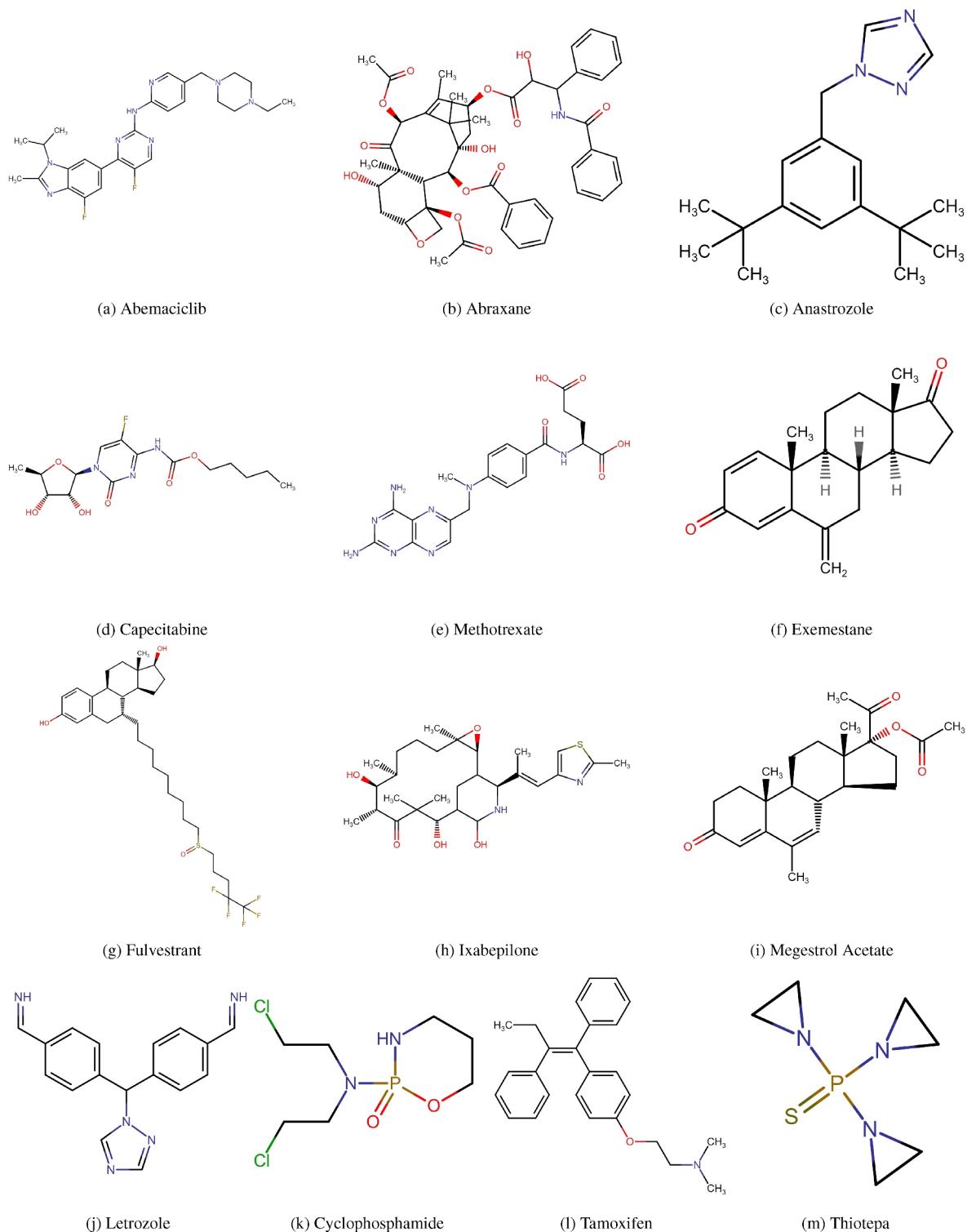
The QSPR between the physical characteristics of cancer medications and their entropy measures is being developed in this area. The cancer medications shown in Fig. 5. To compute the entropy measures, first we convert the chemical structure of drugs into molecular graph given in Fig. 6. The entropy measures value given in Table 5 and the physical properties in Table 6.

To propose the relationship between entropy measures and the physical properties of cancer drug, multiple linear (MLR), Ridge (RR), Lasso (LR), ElasticNet (ENR) and Support Vector regression (SVR) are used. We calculate the only correlation coefficient, coefficient of determination and mean squared error to decide that which model predict the desire physical property listed in Tables 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17.

By using Multiple Linear, Ridge, Lasso, Elastic Net, and Support Vector regression, we proposed the models against each physical property for boiling point (BP), enthalpy of vaporization (EoV), flash point (FP), molar refractivity (MR), molar volume (MV), polarization (P), molecular weight (MW), monoisotopic mass (MM), topological polar surface area (TPSA) and complexity (C). Here we are writing only those model that shows the most significant relationship.

$$\text{Lasso Regression : } \text{BP} = 550.1200 + (165.7870)E_{ReZG_3}, \quad (4)$$

$$\begin{aligned} \text{Ridge Regression : } \text{EoV} = & 86.0600 + (2.1374)E_{R_1} + (0.2140)E_{R_{-1}} \\ & + (1.1891)E_{R_{1/2}} + (0.9586)E_{R_{-1/2}} + (0.9586)E_J \\ & + (1.0045)E_{GA} + (0.8850)E_{ABC} + (1.3779)E_{M_1} + (2.1374)E_{M_2} \\ & + (2.9152)E_F + (2.3871)E_{HM} \\ & + (1.3474)E_{AZI} + (0.6050)E_{ReZG_1} + (1.1075)E_{ReZG_2} + (3.8297)E_{ReZG_3}, \end{aligned} \quad (5)$$

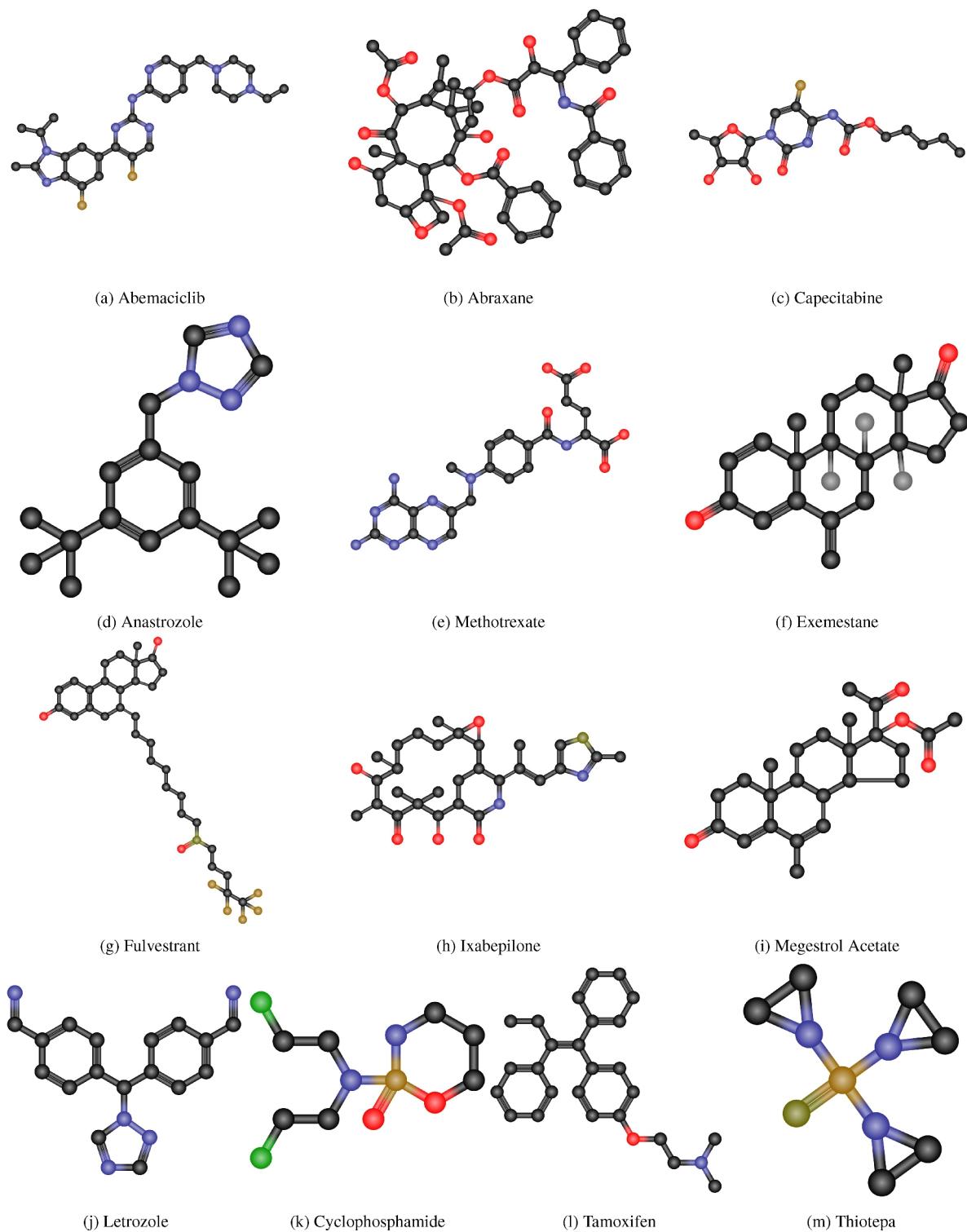
**Fig. 5.** Chemical structures of cancer drugs.

$$\text{Lasso Regression : } \text{FP} = 261.7250 - (28.6916)E_{R-1} - (4.4866)E_{R-1/2} - (110.9448)E_{GA} - (60.1245)E_{ABC} - (42.4353)E_{ReZG_1} + (337.8187)E_{ReZG_3}, \quad (6)$$

$$\text{Lasso Regression : } \text{MR} = 108.4800 + (41.7586)E_{ReZG_3}, \quad (7)$$

$$\text{Lasso Regression : } \text{MV} = 319.0400 + (-360.6798)E_{R-1} + (479.8992)E_{ReZG_3}, \quad (8)$$

$$\text{Lasso Regression : } \text{P} = 43.000 + (15.9447)E_{ReZG_3}, \quad (9)$$



**Fig. 6.** Molecular graphs of cancer drugs.

Drugs	AZI	$M_1$	$M_2$	$m M_2$	H	$ReZG_3$	SDD	I	F
Abemaciclib	98.1406	58	63	3.1111	6	298	30	13.6667	140
Abraxane	592.718	361	458	13.1181	28.2357	2580	165.333	84.5976	1023
Anastrozole	139.701	94	108	3.6944	7.819	558	47.8333	21.3643	258
Capecitabine	213.219	128	149	6	11.8333	760	63.3333	30.0833	330
Cyclophosphamide	112.194	64	72	3.4167	6.4857	368	32.6667	14.9143	166
Exemestane	240.867	156	208	4.9861	10.5714	1308	74.75	34.8452	500
Fulvestrant	340.536	212	251	8.6181	18.0548	1346	105	48.8952	584
Ixabepilone	273.146	186	220	7.4444	15.1571	1188	95	42.0095	524
Letrozole	202.172	112	131	5.0833	10.6333	646	51	27.2333	276
Megestrol Acetate	253.422	164	209	5.8125	12.2262	1218	78.1667	37.3905	488
Methotrexate	279.359	172	200	7.6111	15.5	1014	84.3333	40.4167	444
Tamoxifen	244.313	138	158	6.4444	13.4	768	65	33.3667	336
Thiotepa	115.842	68	88	2.25	5.1571	500	29.5	16.1429	194

**Table 5.** Entropy measures values for cancer drugs.

Drugs	BP	EoV	FP	MR	MV	P	MW	MM	PSA	HAC	C
Abemaciclib	689.3	101	370.7	140.4	382.3	55.7	506.6	506.27	75	37	738
Abraxane	957.1	146	532.6	219.3	610.6	86.9	853.9	853.3	221	62	1790
Anastrozole	469.7	73.2	237.9	90	270.3	35.7	293.4	293.16	78.3	22	456
Capecitabine	517.0	81.2	87	82.3	240.5	32.6	359.4	359.1	121	25	582
Cyclophosphamide	336.1	57.9	157.1	58.1	195.7	23	261.1	260.02	41.6	14	212
Exemestane	453.7	71.3	169	85.8	260.6	34	296.4	296.17	34.1	22	653
Fulvestrant	674.8	104.1	361.9	154	505.1	61.1	606.8	606.31	76.7	41	854
Ixabepilone	697.8	107.3	375.8	140.1	451.6	55.5	506.7	506.2	140	35	817
Letrozole	563.5	84.7	294.6	87.1	234.5	34.5	285.3	285.1	78.3	22	420
Megestrol Acetate	507.1	77.7	77.7	106.4	333.4	42.2	384.5	384.23	60.4	28	821
Methotrexate	561.3	84.1	284.15	119	295.7	47.2	454.4	454.17	211	33	704
Tamoxifen	482.3	74.7	140	118.9	118.9	47.1	371.5	371.2	12.5	28	463
Thiotepa	270.2	50.8	117.2	49.1	125.8	19.5	189.2	189.04	41.1	11	194

**Table 6.** The physical properties of cancer drugs.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.477	0.228	100534.7691
Ridge regression	0.999	0.997	3402.2063
Lasso regression	0.999	0.998	3349.7946
ElasticNet regression	0.999	0.997	3517.0989
Support vector regression	0.993	0.986	8921.9773

**Table 7.** Regression model for BP.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.559	0.313	1892.3409
Ridge regression	0.999	0.998	48.2737
Lasso regression	0.999	0.998	49.7108
ElasticNet regression	0.999	0.998	51.3297
Support vector regression	0.992	0.984	139.9596

**Table 8.** Regression model for EoV.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.346	0.120	92689.8022
Ridge regression	0.965	0.931	15647.8733
Lasso regression	0.968	0.938	15416.2368
ElasticNet regression	0.966	0.934	15801.9674
Support vector regression	0.908	0.824	20067.9223

**Table 9.** Regression model for FP.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.684	0.467	3201.1490
Ridge regression	0.902	0.813	253.6365
Lasso regression	0.908	0.824	245.7076
ElasticNet regression	0.908	0.825	250.6917
Support vector regression	0.899	0.808	1256.6219

**Table 10.** Regression model for MR.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	0.219	0.048	69903.6016
Ridge regression	0.707	0.500	11182.2656
Lasso regression	1.000	1.000	10708.8748
ElasticNet regression	0.690	0.476	11344.4313
Support vector regression	0.704	0.495	13154.8234

**Table 11.** Regression model for MV.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.678	0.459	498.1708
Ridge regression	0.904	0.817	39.9896
Lasso regression	0.910	0.828	39.0514
ElasticNet regression	0.910	0.829	39.6940
Support vector regression	0.890	0.791	193.7721

**Table 12.** Regression model for P.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.417	0.174	63525.0926
Ridge regression	1.000	1.000	885.8908
Lasso regression	0.915	0.838	1903.2440
ElasticNet regression	1.000	0.999	897.8231
Support vector regression	1.000	0.999	12622.7928

**Table 13.** Regression model for MW.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.416	0.173	63042.1280
Ridge regression	1.000	1.000	887.2696
Lasso regression	0.917	0.840	1898.1879
ElasticNet regression	1.000	0.999	899.1999
Support vector regression	1.000	0.999	12267.7786

**Table 14.** Regression model for MM.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	0.998	0.996	8327.3621
Ridge regression	0.681	0.464	3427.4200
Lasso regression	0.681	0.463	3450.3741
ElasticNet regression	0.711	0.506	3384.5901
Support vector regression	0.733	0.537	1560.9485

**Table 15.** Regression model for TPSA.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	-0.338	0.114	200.2285
Ridge regression	0.997	0.995	8.9858
Lasso regression	0.998	0.996	9.0149
ElasticNet regression	0.998	0.996	9.2303
Support vector regression	0.998	0.995	65.6067

**Table 16.** Regression model for HAC.

Model	Pearson R	Coef. of determination	Mean squared error
Multiple linear regression	0.799	0.638	37851.3148
Ridge regression	0.380	0.145	20700.5452
Lasso regression	0.996	0.993	11751.3839
ElasticNet regression	0.355	0.126	21292.8623
Support vector regression	0.475	0.225	26543.2514

**Table 17.** Regression model for C.

$$\begin{aligned}
\text{Ridge Regression : MW} = & 410.6550 + (17.6518)E_{R_1} + (-10.7141)E_{R_{-1}} + (9.0215)E_{R_{1/2}} \\
& + (5.4112)E_{R_{-1/2}} + (5.4112)E_J \\
& + (7.7493)E_{GA} + (6.4299)E_{ABC} + (10.7622)E_{M_1} \\
& + (17.6518)E_{M_2} + (23.6332)E_F + (20.3607)E_{HM} \\
& + (4.2815)E_{AZI} + (1.5745)E_{ReZG_1} + (7.6497)E_{ReZG_2} + (36.4554)E_{ReZG_3},
\end{aligned} \tag{10}$$

$$\begin{aligned}
\text{Ridge Regression : MM} = & 410.2570 + (17.6365)E_{R_1} + (-10.6745)E_{R_{-1}} \\
& + (9.0273)E_{R_{1/2}} + (5.4242)E_{R_{-1/2}} + (5.4242)E_J \\
& + (7.7591)E_{GA} + (6.4450)E_{ABC} + (10.7603)E_{M_1} \\
& + (17.6365)E_{M_2} + (23.5873)E_F + (20.3302)E_{HM} \\
& + (4.3202)E_{AZI} + (1.6057)E_{ReZG_1} + (7.6624)E_{ReZG_2} + (36.3930)E_{ReZG_3},
\end{aligned} \tag{11}$$

$$\begin{aligned}
\text{Multi. Linear Reg. : TPSA} = & 104.3100 + (37034.9815)E_{R_1} + (9668.6404)E_{R_{-1}} \\
& + (16874.7756)E_{R_{1/2}} - (54374.9634)E_{R_{-1/2}} \\
& - (54375.3403)E_J + (17354.7896)E_{GA} + (24258.0159)E_{ABC} + (45163.9464)E_{M_1} \\
& + (37034.9815)E_{M_2} + (-5794.9893)E_F + (-4960.5334)E_{HM} + (213.1249)E_{AZI} \\
& + (11719.2140)E_{ReZG_1} + (-55383.7579)E_{ReZG_2} + (-24398.5301)E_{ReZG_3},
\end{aligned} \tag{12}$$

$$\begin{aligned}
\text{Ridge Regression : HAC} = & 28.7000 + (1.1989)E_{R_1} + (-0.4720)E_{R_{-1}} \\
& + (0.7580)E_{R_{1/2}} + (0.5440)E_{R_{-1/2}} + (0.5440)E_J \\
& + (0.7124)E_{GA} + (0.6437)E_{ABC} + (0.8500)E_{M_1} \\
& + (1.1989)E_{M_2} + (1.5385)E_F + (1.3437)E_{HM} \\
& + (0.5232)E_{AZI} + (0.3693)E_{ReZG_1} + (0.6917)E_{ReZG_2} + (2.2080)E_{ReZG_3},
\end{aligned} \tag{13}$$

$$\text{Lasso Regression : HAC} = 28.7000 + (11.8539)E_{ReZG_3}, \tag{14}$$

$$\begin{aligned}
\text{ElasticNet Regression : HAC} = & 28.7000 + (0.8651)E_{R_1} \\
& + (0.5162)E_{R_{-1}} + (0.7719)E_{R_{1/2}} + (0.7276)E_{R_{-1/2}} + (0.7277)E_J \\
& + (0.7611)E_{GA} + (0.7466)E_{ABC} + (0.7915)E_{M_1} \\
& + (0.8652)E_{M_2} + (0.9375)E_F + (0.8967)E_{HM} \\
& + (0.7131)E_{AZI} + (0.6888)E_{ReZG_1} + (0.7564)E_{ReZG_2} + (1.0758)E_{ReZG_3},
\end{aligned} \tag{15}$$

$$\text{Lasso Regression : C} = 668.2000 + (-1529.9283)E_{R_{-1}} + (619.6337)E_{R_{1/2}} + (515.4111)E_{ReZG_3}. \tag{16}$$

## Discussion

Entropy measures are used to predict the physical and chemical properties of drugs or chemical compounds. In “[Hyaluronic acid-paclitaxel conjugate](#)”, we computed the reverse degree-based entropy measures for hyaluronic acid-paclitaxel conjugate for  $s \geq 1$ . Table 4 shows the numerical comparisons of reverse degree-based entropy measures for small values of  $s$  for hyaluronic acid-paclitaxel conjugate. Figure 4 demonstrates that all entropy measures exhibit an upward trend as the value of  $s$  increases. These results will be helpful to the pharmaceutical industry.

In “[Statistical analysis of entropy measures](#)”, we propose a statistical analysis of reverse degree-based entropy measures using the physical properties of cancer drugs. We find that the reverse degree-based entropy measures show a significant relationship with the physical properties. We employ Multiple Linear, Ridge, Lasso, Elastic Net, and Support Vector regression to examine the relationship between entropy measures and physical properties. All the computed results are listed in Tables 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17. Additionally, we propose the model for each relationship and include only the most significant models that will be used to estimate those physical properties that have not yet been calculated. We examine the following relationships:

- Table 7 presents a comparative analysis of various regression models, highlighting their differences in predictive capabilities. The results indicate that Multiple Linear Regression exhibits the weakest performance, characterized by a relatively low coefficient of determination ( $R^2 = 0.228$ ) and the highest mean squared error (MSE = 100534.7691), suggesting a poorer fit and higher prediction errors. In contrast, Ridge and Lasso Regression demonstrate significant improvements, with  $R^2$  values of 0.997 and 0.997, respectively. These models also exhibit substantially lower MSEs, indicating their superior ability to capture the relationship between entropy measures and the boiling point. Notably, ElasticNet Regression achieves a high  $R^2$  of 0.998 and the lowest MSE among the models (3349.7946), striking a balance between prediction accuracy and generalization. Support Vector Regression (SVR) yields the highest MSE (8921.9773), indicating low predictive power, although its  $R^2 = 0.986$  is slightly lower than the other models, implying a potential trade-off in explaining data variability. Overall, Lasso Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for boiling point. The Lasso Regression model for boiling point is presented in Eq. (4).
- Table 8 presents a comparative analysis of regression models, highlighting their differences in predictive capabilities. The results indicate that Multiple Linear Regression exhibits the weakest performance, characterized by a relatively low coefficient of determination and the highest mean squared error. This means the model is a

poorer fit and has higher prediction errors. In contrast, Ridge, Lasso and ElasticNet Regression demonstrate significant improvements, with the same  $R^2$  value of 0.998 and lower MSEs of 48.2737, 49.7108 and 51.3297, respectively. Ridge Regression achieves a high  $R^2$  and the lowest MSE among the models, striking a balance between prediction accuracy and generalization. Support Vector Regression yields that  $R^2 = 0.984$  is slightly lowest than the other models and the highest MSE (139.9596), indicating the model is a poorer fit and has higher prediction errors. Overall, Ridge Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for enthalpy of vaporization. Ridge Regression model for enthalpy of vaporization is presented in Eq. (5).

- Table 9 presents a comparative analysis of regression models for flash point. The results indicate that Multiple Linear Regression exhibits the weakest performance, characterized by a relatively low coefficient of determination and the highest mean squared error. In contrast, Lasso Regression demonstrates significant improvements, with  $R^2$  value of 0.938 and lower MSE of 25416.2368. On the other hand, Ridge, ElasticNet, and Support Vector Regression yield that  $R^2$  values are slightly lower than the Lasso Regression and the highest MSEs, indicating the models are poorer fit and have higher prediction errors. Overall, Lasso Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for flash point. The Lasso Regression model for flash point is presented in Eq. (6).
- Table 10 presents a comparative analysis of regression models for molar refractivity. The results indicate that Multiple Linear Regression exhibits the weakest performance, characterized by a relatively low coefficient of determination and the highest mean squared error. In contrast, Ridge and Lasso Regression demonstrate significant improvements, with  $R^2$  values of 0.813 and 0.824, respectively and with lower MSEs. Notably, ElasticNet Regression achieves a high  $R^2$  of 0.825 and the lowest MSE among the models (250.6917), striking a balance between prediction accuracy and generalization. Support Vector Regression (SVR) yields the highest MSE (1256.6219), indicating low predictions, although its  $R^2$  is slightly lower than the other models, implying a potential trade-off in explaining data variability. Overall, ElasticNet is preferred for maximizing explanatory power while maintaining good prediction accuracy for molar refractivity. ElasticNet Regression model for molar refractivity is presented in Eq. (7).
- Table 11 presents a comparative analysis of regression models for molar volume. The results indicate that Multiple Linear, Ridge, ElasticNet and vector Support Regression exhibit the weakest performances, characterized by a relatively low coefficient of determination and the highest mean squared errors. In contrast, Lasso Regression demonstrates significant improvements, with a  $R^2$  value of 1.000 and with lower MSE among the models, striking a balance between prediction accuracy and generalization. Overall, Lasso Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for molar volume. The Lasso Regression model for molar volume is presented in Eq. (8).
- Table 12 presents a comparative analysis of regression models for polarization. The results indicate that Multiple Linear and Support Vector Regression exhibits the weakest performances, characterized by a relatively low coefficient of determination and the highest mean squared error. In contrast, Ridge, Lasso and ElasticNet Regression demonstrate significant improvements, with  $R^2$  values of 0.817, 0.828 and 0.829, and lower MSEs of 39.9896, 39.0514 and 39.6940, respectively. Lasso Regression achieves the lowest MSE among the models, striking a balance between prediction accuracy and generalization. Overall, Lasso Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for polarization and no multicollinearity. Ridge Regression model for polarization is presented in Eq. (9).
- Table 13 presents a comparative analysis of regression models for molecular weight. The results indicate that Multiple Linear, Lasso, and Vector Support Regression exhibits the weakest performances, characterized by the highest mean squared errors. In contrast, Ridge Regression and ElasticNet demonstrate significant improvements, with  $R^2$  values of 1.000 and 0.999 with lower MSEs 885.8908 and 897.8231 among the models, striking a balance between prediction accuracy and generalization. Overall, Ridge Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for molecular weight. Ridge Regression model for molecular weight is presented in Eq. (10).
- Table 14 presents a comparative analysis of regression models for monoisotopic mass. The results indicate that Multiple Linear, Lasso, and Vector Support Regression exhibits the weakest performances, characterized by the highest mean squared errors. In contrast, Ridge and ElasticNet Regression demonstrate significant improvements, with  $R^2$  values of 1.000 and 0.999 with lower MSEs 887.2696 and 899.1999 among the models, striking a balance between prediction accuracy and generalization. Overall, Ridge Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for monoisotopic mass. Ridge Regression model for monoisotopic mass is presented in Eq. (11).
- Table 15 presents a comparative analysis of regression models for topological polar surface area. The results indicate that Multiple Linear Regression exhibits the highest performance, characterized by a relatively high coefficient of determination ( $R^2 = 996$ ). In contrast, Ridge, Lasso, ElasticNet and Support Vector Regression demonstrate the lowest performance, characterized by a relatively low coefficient of determination. Overall, Multiple Linear Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for topological polar surface area. Multiple Linear Regression model for topological polar surface area is presented in Eq. (12).
- Table 16 presents a comparative analysis of regression models for heavy atom count. The results indicate that Multiple Linear and Support Vector Regression exhibits the weakest performances, characterized by a relatively low coefficient of determination and the highest mean squared error. In contrast, Ridge, Lasso and ElasticNet Regression demonstrate significant improvements, with  $R^2$  values of 0.995, 0.996 and 0.996, and lower MSEs of 8.9858, 9.0149 and 9.2303, respectively. Lasso Regression achieves the highest  $R^2$  value among the models, striking a balance between prediction accuracy and generalization. Overall, Lasso Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for heavy atom

- count. Models for heavy atom count are presented in Eqs. (13)–(15). However, the best model is the Lasso Regression model (Eq. 15), as it does not exhibit multicollinearity.
- Table 17 presents a comparative analysis of regression models for complexity. The results indicate that Multiple Linear, Ridge, ElasticNet and Vector Support Regression exhibits the weakest performances, characterized by low coefficient of determination and highest mean squared errors. In contrast, Lasso Regression demonstrates significant improvements, with  $R^2$  value of 0.993 with lower MSE 11751.3839 among the models, striking a balance between prediction accuracy and generalization. Overall, Lasso Regression is preferred for maximizing explanatory power while maintaining good prediction accuracy for complexity. The Lasso Regression model for complexity is presented in Eq. (16).

Based on the above results and their explanations, we can conclude that the physical properties, such as boiling point, enthalpy of vaporization, flash point, molar refractivity, molar volume, polarization, molecular weight, monoisotopic mass, topological polar surface area, and complexity, can be predicted using the reverse entropy measures.

## Conclusion

Entropy measures are utilized to predict physical and chemical properties of drugs. In this study, we computed entropy measures for the hyaluronic acid-paclitaxel conjugate. The results exhibited numerical values that demonstrated the effectiveness of entropy measures. The utilization of reverse degree-based entropy measures proved valuable in quantitative structure-property relationship (QSPR) investigations as predictive measures. This investigation focused on assessing the predictive ability of entropy measures by analyzing the physical properties of cancer drugs. The obtained findings demonstrated a robust positive correlation between the boiling point, enthalpy of vaporization, flash point, molar refractivity, molar volume, polarization, molecular weight, monoisotopic mass, topological polar surface area, complexity, and entropy measures. Our analysis determined that the entropy measures  $\mathcal{E}_{R_1}$ ,  $\mathcal{E}_{R_{-1}}$ ,  $\mathcal{E}_{R_{\frac{1}{2}}}$ ,  $\mathcal{E}_{R_{-\frac{1}{2}}}$ ,  $\mathcal{E}_{ABC}$ ,  $\mathcal{E}_{GA}$ ,  $\mathcal{E}_{M_1}$ ,  $\mathcal{E}_{M_2}$ ,  $\mathcal{E}_{AZI}$ ,  $\mathcal{E}_{HM_1}$ ,  $\mathcal{E}_F$ ,  $\mathcal{E}_J$ ,  $\mathcal{E}_{ReZG_1}$ ,  $\mathcal{E}_{ReZG_2}$ , and  $\mathcal{E}_{ReZG_3}$  can be used for predicting physical properties. We developed a predictive model for each relationship and selected only the most significant models to estimate physical properties that have not yet been calculated.

## Data availability

All data generated or analysed during this study are included in this published article.

Received: 6 October 2024; Accepted: 21 January 2025

Published online: 08 February 2025

## References

1. Sorgun, Sezer, Küçük, H. & Birgin, K. Some distance-based topological indices of certain polysaccharides. *J. Mol. Structure* **1250**, 131716 (2022).
2. Redžepović, Izudin & Furtula, B. On relationships of eigenvalue-based topological molecular descriptors. *Acta Chimica Slovenica* **67**(1), 312–318 (2020).
3. Gao, W., W. Wang, & M.R. Farahani. Topological indices study of molecular structure in anticancer drugs. *J. Chem.* **2016** (2016).
4. Gao, W., Farahani, M. R. & Shi, L. Forgotten topological index of some drug structures. *Acta medica mediterranea*. **32**(1), 579–585 (2016).
5. Antalan, J.R.M., Francis Joseph, H. C., & Edwin D.I. A note on some distance-based topological indices of circulant network  $C_n$  (1, a). *J. Discrete Math. Sci. Cryptogr.* 1–19 (2021).
6. González-Díaz, H., & C.R. Munteanu. Topological indices for medicinal chemistry, biology, parasitology, neurological and social networks. Kerala: Transworld Research Network (2010).
7. Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Society* **69**(1), 17–20 (1947).
8. Randić, Milan. Characterization of molecular branching. *J. Am. Chem. Society* **97**(23), 6609–6615 (1975).
9. Amić, Dragan, Bešlo, D., Lukić, B., Nikolić, S. & Trinajstić, N. The vertex-connectivity index revisited. *Journal of chemical information and computer sciences* **38**(5), 819–822 (1998).
10. Bollobás, B. & Erdos, Paul. Graphs of extremal weights. *Ars Combinatoria* **50**, 225–233 (1998).
11. Zhou, B. & Trinajstić, N. On general sum-connectivity index. *J. Math. Chem.* **47**(1), 210–218 (2010).
12. Zhong, L. The harmonic index for graphs. *Appl. Math. Lett.* **25**(3), 561–566 (2012).
13. Gutman, I., Furtula, B., Vukicevic, Z. K. & Popivoda, G. On Zagreb indices and coindices. *MATCH Commun. Math. Computer Chem.* **74**(1), 5–16 (2015).
14. Fath-Tabar, G. H. Old and new Zagreb indices of graphs. *MATCH Commun. Math. Computer Chem.* **65**(1), 79–84 (2011).
15. Munir, Mobeen, Nazeer, W., Rafique, S. & Kang, S. M. M-polynomial and related topological indices of nanostar dendrimers. *Symmetry* **8**(9), 97 (2016).
16. Ranjini, P. S., Lokesha, V., Bindusree, A. R. & Phani Raju, M. New bounds on Zagreb indices and the Zagreb co-indices. *Boletim da Sociedade Paranaense de Matemática* **31**(1), 51–55 (2013).
17. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948).
18. Shannon, C. E., & W. Weaver. The mathematical theory of communication. (University of Illinois Press, 1949).
19. Mowshowitz, A. Entropy and the complexity of graphs: I. An index of the relative complexity of a graph. *Bull. Math. Biophys.* **30**(1), 175–204 (1968).
20. Mowshowitz, Abbe. Entropy and the complexity of graphs: II. The information content of digraphs and infinite graphs. *Bull. Math. Biophys.* **30**(2), 225–240 (1968).
21. Körner, János. Coding of an information source having ambiguous alphabet and the entropy of graphs. in *6th Prague conference on information theory*, 411–425 (1973).
22. Dehmer, Matthias M., Barbarini, N. N., Varmuza, K. K. & Gruber, A. A. Novel topological descriptors for analyzing biological networks. *BMC Struct. Biol.* **10**(1), 1–17 (2010).
23. Chen, Zengqiang, Dehmer, M., Emmert-Streib, F. & Shi, Y. Entropy bounds for dendrimers. *Appl. Math. Comput.* **242**, 462–472 (2014).
24. Dehmer, Matthias & Mowshowitz, A. A history of graph entropy measures. *Inform. Sci.* **181**(1), 57–78 (2011).

25. Kulli, V. R. Reverse Zagreb and reverse hyper-Zagreb indices and their polynomials of rhombus silicate networks. *Ann. Pure Appl. Math.* **16**(1), 47–51 (2018).
26. Gao, Wei, Younas, M., Farooq, A., Virk, A. U. R. & Nazeer, W. Some reverse degree-based topological indices and polynomials of dendrimers. *Mathematics* **6**(10), 214 (2018).
27. Koam, A.N.A., A. Ahmad, & A.A. Qummer. On the study of reverse degree-based topological properties for the third type of chain hex-derived network. *J. Math.* **2021** (2021).
28. Zhao, D., Y.-M. Chu, M.K. Siddiqui, K. Ali, M. Nasir, M.T. Younas, & M. Cancan. On reverse degree based topological indices of polycyclic metal organic network. *Polycyclic Aromatic Compounds*. 1–18 (2021).
29. Kwun, Young Cheol, Virk, A. U. R., Rafaqat, M., Rehman, M. U. & Nazeer, W. Some reversed degree-based topological indices for graphene. *J. Discrete Math. Sci. Cryptogr.* **22**(7), 1305–1314 (2019).
30. Koam, A. N. A., Ahmad, A. & Ahmad, Y. Computation of reverse degree-based topological indices of hex-derived networks. *AIMS Math.* **6**(10), 11330–11345 (2021).
31. Jung, Chahn Yong, Gondal, M. A., Ahmad, N. & Kang, S. M. Reverse degree based indices of some nanotubes. *J. Discrete Math. Sci. Cryptogr.* **22**(7), 1289–1294 (2019).
32. Furtula, B., Kinkar, C. & I. Gutman. Comparative analysis of symmetric division deg index as potentially useful molecular descriptor. *Int. J. Quant. Chem.* **118**(17), e25659 (2018).
33. Rauf, A., M. Naeem, & A. Aslam. Quantitative structure-property relationship of edge weighted and degree-based entropy of benzene derivatives. *Int. J. Quant. Chem.* e26839.
34. Zhong, J.-F., A. Rauf, M. Naeem, J. Rahman, & A. Aslam. Quantitative structure-property relationships (QSPR) of valency based topological indices with COVID-19 drugs and application. *Arab. J. Chem.* 103240 (2021).
35. Rauf, A., M. Naeem, & S. U. Bukhari. Quantitative structure-property relationship of Ev-degree and Ve-degree based topological indices with physico-chemical properties of benzene derivatives and application. *Int. J. Quant. Chem.* e26851.
36. Lucić, Bono, Lukovits, I., Nikolić, S. & Trinajstić, N. Distance-related indexes in the quantitative structure-property relationship modeling. *J. Chem. Inform. Comput. Sci.* **41**(3), 527–535 (2001).
37. Smith, M. T. Advances in understanding benzene health effects and susceptibility. *Annu. Rev. Public Health.* **31**, 133–148 (2010).
38. Hayat, Sakander & Khan, S. Quality testing of spectrum-based valency descriptors for polycyclic aromatic hydrocarbons with applications. *J. Mol. Struct.* **1228**, 129789 (2021).
39. Streitwieser, A., & C.M.H. Smith. Dictionary of Pi-electron Calculations. (1965).
40. Nikolić, Sonja, Trinajstić, N. & Baučić, I. Comparison between the vertex-and edge-connectivity indices for benzenoid hydrocarbons. *J. Chem. Inform. Computer Sci.* **38**(1), 42–46 (1998).
41. Basak, S.C., G.D. Grunwald, & G.J. Niemi. Use of graph-theoretic and geometrical molecular descriptors in structure-activity relationships. in *From chemical topology to three-dimensional geometry*, 73–116. (Springer, 2002).
42. Zhang, Xiaodong et al. CD44-targeted facile enzymatic activatable chitosan nanoparticles for efficient antitumor therapy and reversal of multidrug resistance. *Biomacromolecules* **19**(3), 883–895 (2018).
43. Markman, M. & Mekhail, T. M. Paclitaxel in cancer therapy. *Expert Opinion Pharmacother.* **3**(6), 755–766 (2002).
44. Armstrong, H.E. XXVIII.-An explanation of the laws which govern substitution in the case of benzenoid compounds. *J. Chem. Society Trans.* **51**, 258–268 (1887).
45. Kosheleva, O., & V. Kreinovich. Kekulé's benzene structure: A case study of teaching usefulness of symmetry. (2014).
46. Dewar, James. On the oxidation of phenyl alcohol, and a mechanical arrangement adapted to illustrate structure in the nonsaturated hydrocarbons. *Proc. R. Society Edinburgh* **6**, 82–86 (1869).
47. Kekulé, Auguste. Sur la constitution des substances aromatiques. *Bulletin mensuel de la Société Chimique de Paris* **3**, 98 (1865).

## Acknowledgements

The authors gratefully acknowledge the funding of the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia, through Project Number: GSSRD-24

## Declarations

### Competing interest

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to O.O.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025