Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 14** : 1–15.
ISSN-0037-8712
doi:10.5269/bspm.82194

Artificial Neural Networks Applied to Predict the Characteristics of Anticancer Medications Using K-Banhatti Indices

C. M. Veena[1], K. S. Onkarappa [2], M. C. Shanmukha[3,*], K. C. Savitha[4] and A. Usha[5]

ABSTRACT: Genetic alterations that cause abnormal and fast cell division give birth to cancer, which can be fatal. Anticancer medications are vital to treatment and successful drug development and discovery depend on an understanding of their physicochemical behaviour. Molecular topological indices-based Quantitative Structure-Property Relationship ($QSPR$) modelling provides a computionally effective substitute for traditional experimental investigation. In this work, we examine the predictive capability of K-Banahatti topological indices $B_1(G)$, $B_2(G)$, ${}^mB_1(G)$, ${}^mB_2(G)$ and $HB(G)$ with the physicochemical characteristics of specific anticancer drugs, such as boiling point, enthalpy and molar refraction. The correlation strength was first examined using linear regression models, which showed statistically significant connections. In particular, a strong linear association between ${}^mB_1(G)$ and molar refraction was highlighted. An Artificial Neural Network (ANN) regression framework was further constructed utilizing a multilayer perceptron architecture in order to capture complicated nonlinear dependencies beyond linear trends. The ANN's advantage in managing nonlinear interaction among descriptors was confirmed by its improved prediction performance, especially for molar refraction. Comparative investigation revealed that while ANN gives better predictive accuracy and generalization potential, linear regression improves interpretability. All things considered, the combination of $QSPR$ and ANN enhances our comprehension of the structure-property correlations of anticancer medications and offers a productive computational avenue to promote logical drug design on oncology.

Keywords: Anticancer drugs, topological indices, K-Banhatti indices, ANN.

Contents

1. Introduction

A large group of diseases that cause cell division rapidly and abnormally give rise to life-threatening disease called cancer. This spread to the neighbouring tissues and organs by disrupting the regular functions of the body. One of the major causes of deaths worldwide is cancer ($https://www.who.int/news-room/fact-sheets/detail/cancer$). In the year 2020, as per the studies of World Health Organisation, it was noted that 1 in 6 deaths were because of cancer. To overcome this, a lot of research is carried

2020 *Mathematics Subject Classification:* 05C07, 05C90.
Submitted February 25, 2026. Published April 30, 2026.

out to test new cancer treatments. Mutations is the main cause of cancer. It may be the changes to the DNA (deoxyribonucleic acid) in the cells or genetic mutations that could be inherited or may be because of environmental forces. The external causes, namely carcinogens may include, radiations and UV light, chemical carcinogens from cigarette smoke, alcohol, contaminated food, asbestos, air pollution, and even drinking water. Biological carcinogens include viruses, parasites, and bacteria. Studies reveal that 33 percent of cancer deaths may be caused due to the consumption of tobacco, obesity, less consumption of fruits and vegetables and not enough physical activity.

Cancer starts in a particular part of the body and may spread to the surrounding parts. They are named after the area it first affects based on the type of cell and the part of the body. For example, it may affect the lungs initially and may spread to liver and still it is called lung cancer. Other than these, there are various clinical terms in cancer such as, carcinoma, sarcoma, leukaemia, lymphoma, and myeloma and many more. The cancer that starts in the skin and spreads to other organs is carcinoma, while that affects the connective tissues like bones, cartilage and blood vessels is sarcoma. Cancer that affects bone marrow is leukaemia and cancers which disturbs the immune system are lymphoma and myeloma ($https://www.healthline.com/health/cancerearly-detection$).

As the cases of cancer are increasing, prevention is always better than cure. There are tests that can be done which reveals the signs of cancerous cells in the body at early stages. By detecting at earlier stages would benefit in lowering the mortality rate and the increase in the effectiveness of the medication. Some common cancer screenings may detect cervical cancer, prostate cancer, lung cancer, breast cancer, skin cancer and colorectal cancer. There is a high risk of cancer if there is a history in the family. It is always better to take early detection tests and be safer ($https://www.cancer.gov/about-cancer/understanding/what-is-cancer$).

Anticancer drugs are those that are successful in treating malignant or cancerous diseases. Alkylating agents, antimetabolites, natural products, and hormones are just a few of the many primary categories of anticancer drugs. Finding certain graph attributes that are invariant is always intriguing. The experimental method is exceedingly costly and time-consuming for the synthesis, examination, analysis, and finding of new anticancer lead. Soft computations derived from cheminformatics methods boost the probability of success and cut down on time and expense associated with lead structure discovery [1,2,3,4,5,6,7].

The rational drug design utilizing the $QSAR$ (Quantitative Structure-Activity Relationship) is the main use of cheminformatics in theoretical drug discovery research. Topological indices (TIs) have been the subject of many research articles over the past ten years [8,9,10,11,12,13,14]. In the study of $QSAR/QSPR$, molecular topological indices are crucial. Pharmaceutical drug design calls for knowledge of the physicochemical properties and biological activity of molecular graphs of compounds. The topological index, a widely used technique in chemical graph theory, can be used to predict these characteristics.

The topological indices are the numerical descriptors used to describe chemical systems that are derived from molecular graphs. For drug design and discovery research, the concept of $QSAR$ formulation based on various molecular descriptors such as physicochemical, constitutional and geometrical, electrostatic, and topological computed solely from the chemical structure have greater importance. The usage of $QSAR$ in the drug development process is depicted in Figure 1 as a scheme [15].

Neural networks can be applied in graph theory using topological indices to analyse and predict properties of molecules represented as graphs [16]. Neural networks learn complex patterns and relationships between topological indices and properties. Recent advancements in machine learning have generated considerable interest in applying neural networks to predict properties of complex system. A kind of neural networks called graph neural networks is made specifically to handle graph structured data. By combining nearby nodes and edges, graph neural networks update node representations. In this method nodes update their representations by exchanging information with their connected vertices. A neural network is a computational model that draws inspiration from the form and operation of biological networks.

Artificial Neural Network (ANN) is particularly useful in $QSPR$ studies as it can capture nonlinear relationships between molecular descriptor and physicochemical properties, providing improved prediction accuracy over conventional linear models [17,18]. Artificial neural networks ($ANNs$) are inspired by the functioning of the human brain. $ANNs$ have become powerful computational models that can extract

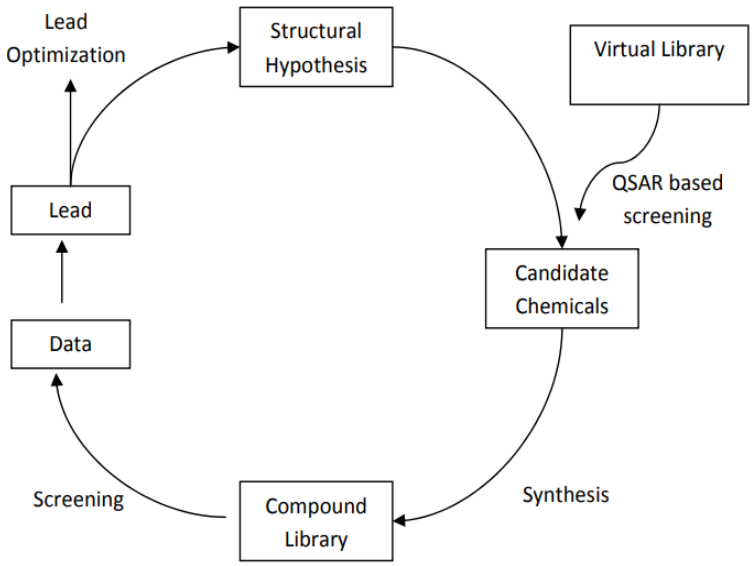


Figure 1: $QSAR$ and drug design using numerical descriptors.

complex nonlinear correlations from data. Neural networks, in contrast to traditional regression models, are capable of processing. They are especially useful for QSPR analysis because they reveal hidden patterns in high dimensional descriptor spaces. The predictive ability of $QSPR$ model has improved recently when molecular graph theoretic descriptors, particularly topological indices are integrated with neural network topologies.

To identify the predictive ability of the physicochemical properties of various anticancer drugs, some topological indices have already been computed in the literature [19,20,21]. However, for the sake of studying anticancer drugs, only vertex-degree-based topological indices are taken into account. In this study a brand new class of topological indices that are defined by the vertex and edge degrees are investigated.

In this paper, the molecular graph $G = (V(G), E(G))$, which are simple and connected, and the vertex set $V(G)$, an edge set $E(G)$ are considered. The cardinality of the set $V(G)$ and $E(G)$ are known as the order, and the size of the graph G respectively. The number of edges occurring to a vertex $x \in V(G)$ is called the vertex degree, denoted by d_x. The degree of an edge e denoted by d_e, and defined by, $d_e = d_x + d_y - 2$ where, $e = xy$. Using $x \sim e$ for the vertex x and an edge e are adjacent in the graph G [22,23,24,25].

In [26,27], Kulli proposed various novel degree-based TIs viz., the first K-Banhatti index ($B_1(G)$), the second K-Banhatti index ($B_2(G)$), the modified first K-Banhatti index (${}^mB_1(G)$), the modified second K-Banhatti index (${}^mB_2(G)$), and the harmonic K-Banhatti index ($HB(G)$) of a graph G is defined as

$$B_1(G) = \sum_{x \sim e} (d_x + d_e), \qquad B_2(G) = \sum_{x \sim e} (d_x \cdot d_e),$$

$${}^mB_1(G) = \sum_{x \sim e} \left(\frac{1}{d_x + d_e}\right), \qquad {}^mB_2(G) = \sum_{x \sim e} \left(\frac{1}{d_x \cdot d_e}\right),$$

$$\text{and} \quad HB(G) = \sum_{x \sim e} \left(\frac{2}{d_x + d_e}\right)$$

respectively. And $x \sim e$ represents the vertex x, and an edge e are incident in the graph G. In [28,29] the values of K-Banhatti indices for some graph operations are computed. In this article, we examine the

predictive capability of anticancer drug properties based on K-Banahatti topological indices. The study evaluates and compares the performance of conventional linear regression models with that of ANN to identify the most effective approach for modeling these relationships

2. Linear Regression Model

In this study, a linear regression analysis was employed to investigate the relationships between the physiochemical properties of anticancer drugs through K-Banhatti indices. This focussed analysis is aimed to determine the influence of variations in the selected topological index with its properties.

The linear regression model used for this analysis is expressed as follows

$$Y = \beta_0 + \beta_1 X$$

Where Y represents the physiochemical property of interest, X is the single topological index chosen for the study, β_0 is the intercept, β_1 is the coefficient associated with the topological index.

The coefficient β_1 was estimated to quantify the strength and direction of the relationship between the physiochemical property and the selected topological index.

3. Statistical Tests

3.1. Model Significance

To assess the overall significance of the linear regression model, an analysis of variance (ANOVA) was conducted. The ANOVA test evaluated whether the model, which includes the topological index as a predictor, provided a better fit than an intercept model.

3.2. Coefficient Significance

The coefficient (β_1) in the linear regression model was tested for significance using a t-test. This test determined whether the selected topological index had a statistically significant impact on the physiochemical property.

4. Data Analysis

4.1. Data Preprocessing

Prior to the linear regression analysis, the dataset underwent data preprocessing, which included handling missing data, identifying and addressing outliers, and ensuring that all variables met the assumptions of linear regression.

4.2. Model Fitting

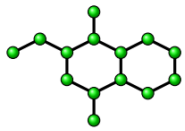
The linear regression model was fitted to the preprocessed data using a standard least squares approach. The coefficient (β_1) was estimated, and model diagnostics (e.g., residuals analysis) were performed to assess the model's goodness of fit.

4.3. Interpretation

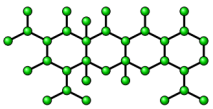
The coefficient (β_1) obtained from the linear regression analysis was interpreted in the context of the physiochemical property and the specific topological index used. The direction and magnitude of β_1 indicated how changes in the topological index influenced the target property.

4.4. Reporting

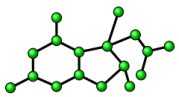
The results of the linear regression analysis, including the coefficient estimate (β_1), p-value, and confidence interval, were reported in the paper. Visual aids, such as regression plots, were utilized to facilitate the interpretation of the findings.



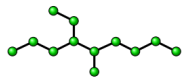
(a) Caulibugulone E



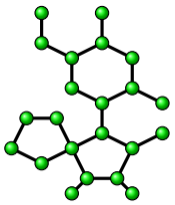
(b) Minocycline



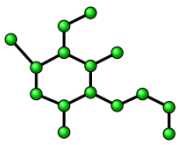
(c) Convolutamydine A



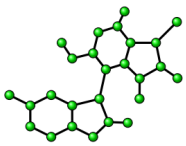
(d) Carmustine



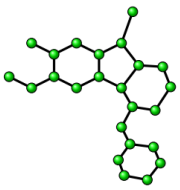
(e) Amathaspiramide E



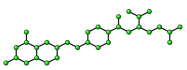
(f) Convolutamine F



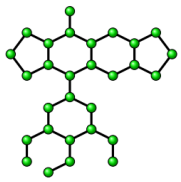
(g) Aspidostomide F



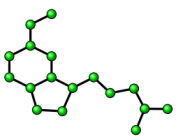
(h) Pterocelin B



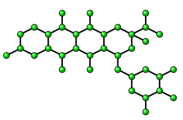
(i) Aminopterin



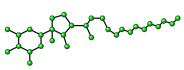
(j) Podophyllotoxin



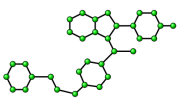
(k) Melatonin



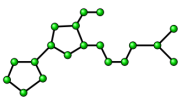
(l) Daunorubicin



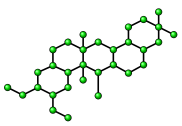
(m) Convolutamide A



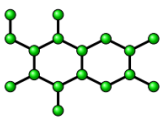
(n) Raloxifene



(o) Tambjamine K



(p) Deguelin



(q) Perfragilin A

Figure 2: Molecular graphs of various anticancer drugs.

Table 1: The physicochemical properties of anticancer drugs [30].

Sl.No.	*Drugs*	*BP*	*MP*	*E*	*MR*
1	Amathaspiramide E	572.7	209.72	90.3	89.4
2	Aminopterin	782.27	344.45	-	114
3	Aspidostomide E	798.8	-	116.2	116
4	Carmustine	309.6	120.99	63.8	46.6
5	Caulibugulone E	373	129.46	62	52.2
6	Convolutamide A	629.9	-	97.9	130.1
7	Convolutamine F	387.7	128.67	63.7	73.8
8	Convolutamydine A	504.9	199.2	81.6	68.2
9	Daunorubicin	770	208.5	117.6	130
10	Deguelin	560.1	213.39	84.3	105.1
11	Melatonin	512.8	182.51	78.4	67.6
12	Minocycline	803.3	326.3	122.5	116
13	Perfragilin A	431.5	187.62	68.7	63.6
14	Podophyllotoxin	597.9	235.86	93.6	104.3
15	Pterocellin B	521.6	199.88	79.5	87.4
16	Raloxifene	728.2	289.58	110.1	136.6
17	Tambjamine K	391.7	-	64.1	76.6

Table 2: The values of topological indices of anticancer drugs.

Sl.No.	*Drugs*	$B1(G)$	$B2(G)$	$mB1(G)$	$mB2(G)$	$HB(G)$
1	Amathaspiramide E	226	345	7.494	6.6305	14.8588
2	Aminopterin	364	479	14.6762	13.6111	29.3524
3	Aspidostomide E	320	467	13.1423	10.4444	26.2846
4	Carmustine	100	113	5.669	6.9166	11.338
5	Caulibugulone E	156	214	6.3762	6.3888	12.7524
6	Convolutamide A	312	401	13.1885	13.2042	26.377
7	Convolutamine F	150	202	6.8761	4.0777	13.7522
8	Convolutamydine A	210	316	6.6282	5.6708	13.2564
9	Daunorubicin	489	719	17.2254	16.1888	34.4508
10	Deguelin	435	616	12.8635	12.1833	25.727
11	Melatonin	178	228	7.8381	7.7778	15.6762
12	Minocycline	460	734	14.0933	13.0292	28.1865
13	Perfragilin A	198	286	7.8333	7.3889	15.6666
14	Podophyllotoxin	373	535	13.7429	13.3889	27.4858
15	Pterocellin B	272	369	10.6452	9.75	21.2905
16	Raloxifene	388	517	15.2476	13.3889	30.4952
17	Tambjamine K	166	190	7.3381	7.0714	14.6762

5. Results & Discussions

5.1. The linear regression models for the K-Banhatti indices and physicochemical properties of anticancer drugs

In this section, we study the predictive potential of some physicochemical properties of anticancer drugs (Figure 2) for K-Banhatti indices. The four physicochemical properties (boiling point (BP), melting point (MP), enthalpy (E), and molar refraction (MR)) of anticancer drugs are tabulated in Table 1 and corresponding value of the topological indices are tabulated in the Table 2 for anticancer

drugs. The regression model used is given below,

$$PAD = \alpha\ (TI) + \beta \tag{5.1}$$

where PAD is a place-holder standing for the physical property of interest for an anticancer drug, TI is a topological index, α is a regression coefficient, and β is a constant.

Throughout this paper, the notations used are N for the population, S_e for standard error of the estimate, F for F-values, and p be the significance of probability value. Table 3 to Table 7 displays the statistical parameters of K-Banhatti indices and physical properties of anticancer drugs. Using equation (5.1), the linear models for the K-Banhatti indices in the study are obtained as below,

1. The linear regression models for the first K-Banhatti index (B_1)

$$\begin{aligned} BP &= 1.152\ B_1 + 243.9, \\ MP &= 0.3936\ B_1 + 100.128, \\ E &= 0.1443\ B_1 + 47.1632, \\ MR &= 0.20688\ B_1 + 34.4175. \end{aligned}$$

2. The linear regression models for the second K-Banhatti index (B_2)

$$\begin{aligned} BP &= 0.7394\ B_2 + 276.3810, \\ MP &= 0.2511\ B_2 + 110.8029, \\ E &= 0.0936\ B_2 + 50.5401, \\ MR &= 0.1263\ B_2 + 42.7704. \end{aligned}$$

3. The linear regression models for the modified first K-Banhatti index ($^mB_1$)

$$\begin{aligned} BP &= 37.8939\ {}^mB_1 + 165.9875, \\ MP &= 12.9615\ {}^mB_1 + 76.2906, \\ E &= 4.7920\ {}^mB_1 + 37.36, \\ MR &= 7.0415\ {}^mB_1 + 17.8724. \end{aligned}$$

4. The linear regression models for the modified second K-Banhatti index ($^mB_2$)

$$\begin{aligned} BP &= 36.3527\ {}^mB_2 + 211.8249, \\ MP &= 12.9470\ {}^mB_2 + 86.4471, \\ E &= 4.6476\ {}^mB_2 + 42.5550, \\ MR &= 6.8901\ {}^mB_2 + 25.0634. \end{aligned}$$

5. The linear regression models for the Harmonic K-Banhatti index (HB)

$$\begin{aligned} BP &= 18.9127\ HB + 166.8605, \\ MP &= 6.4688\ HB + 76.6013, \\ E &= 2.3913\ HB + 37.4829, \\ MR &= 3.515\ HB + 18.0225. \end{aligned}$$

Table 3: Statistical specifications for the linear model of first K-Banhatti index (B_1).

Physical Prop.	N	α	β	r	F	p	*Indicator*	S_e
BP	17	1.152	243.9	0.856	41.0586	0.0000	significant	86.763
MP	14	0.3936	100.128	0.7244	14.6840	0.0024	significant	47.8218
E	16	0.1443	47.1632	0.8616	36.7279	0.0000	significant	11.3090
MR	17	0.2069	34.4175	0.8738	48.448	0.0000	significant	14.3376

Table 4: Statistical specifications for the linear model of second K-Banhatti index (B_2).

Physical Prop.	N	α	β	r	F	p	*Indicator*	S_e
BB	17	0.7394	276.3810	0.8490	38.7211	0.0000	significant	88.6306
MP	14	0.2511	110.8029	0.7244	13.2466	0.0033	significant	49.1643
E	16	0.0936	50.5401	0.8616	40.3484	0.0000	significant	10.9259
MR	17	0.1263	42.7704	0.8251	31.9813	0.0005	significant	16.668

Table 5: Statistical specifications for the linear model of modified first K-Banhatti index ($^mB_1$).

Physical Prop.	N	α	β	r	F	p	*Indicator*	S_e
BP	17	37.8939	165.9875	0.8843	53.8148	0.0000	significant	78.3096
MP	14	12.9615	76.2906	0.7530	15.7118	0.0018	significant	46.9266
E	16	4.7920	37.3653	0.8673	42.4968	0.0000	significant	10.7061
MR	17	7.0415	17.8724	0.9347	103.765	0.0000	significant	10.4795

Table 6: Statistical specifications for the linear model of modified second K-Banhatti index ($^mB_2$).

Physical Prop.	N	α	β	r	F	p	*Indicator*	S_e
BP	17	36.3527	211.8249	0.8038	27.3925	0.0001	significant	99.7726
MP	14	12.9470	86.4471	0.7164	12.6548	0.0039	significant	49.7509
E	16	4.6476	42.5550	0.7978	24.5058	0.0002	significant	12.9804
MR	17	6.8901	25.0634	0.8666	45.2526	0.0000	significant	14.7128

Table 7: Statistical specifications for the linear model of Harmonic K-Banhatti index (HB).

Physical Prop.	N	α	β	r	F	p	*Indicator*	S_e
BP	17	18.9127	166.8605	0.8835	53.3644	0.0000	significant	78.5672
MP	14	6.4688	76.6013	0.7523	15.6476	0.0019	significant	46.9811
E	16	2.3913	37.4829	0.8663	42.1260	0.0000	significant	10.7515
MR	17	3.515	18.0225	0.9340	102.5357	0.0000	significant	10.5341

Table 8: Correlation coefficient (r) between physicochemical properties of anticancer drugs with K-Banhatti indices.

PAD	$B_1(G)$	$B_2(G)$	$^mB_1(G)$	$^mB_2(G)$	$HB(G)$
BP	0.8566	0.8490	**0.8843**	0.8038	0.8835
MP	0.7418	0.7244	**0.7530**	0.7164	0.7523
E	0.8509	0.8616	**0.8673**	0.7978	0.8663
MR	0.8738	0.8251	**0.9347**	0.86663	0.9340

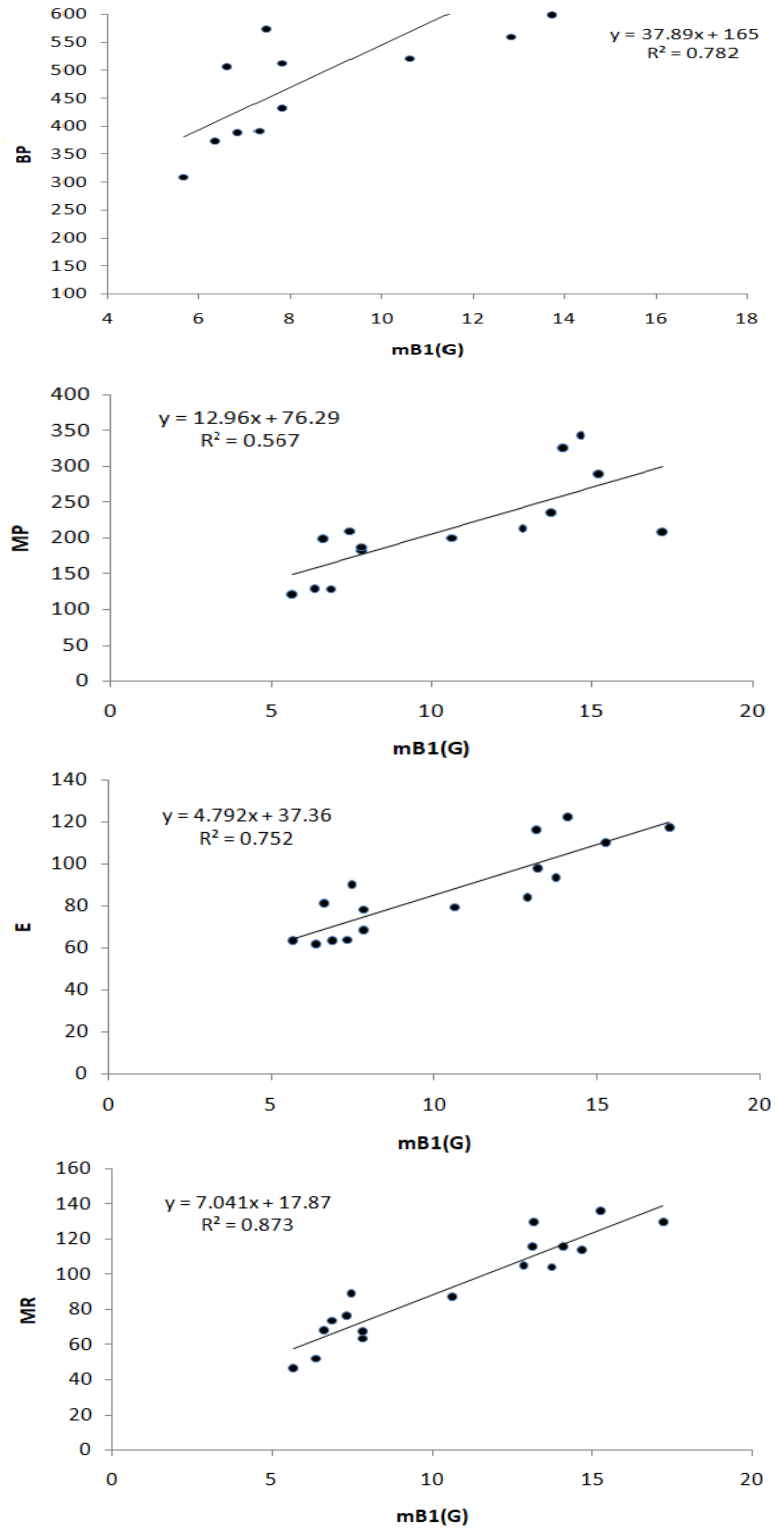


Figure 3: The correlation between BP, MP, E, MR with modified first K-Banhatti index($^{m}B_1(G)$).

5.2. Neural Network regression model for the K-Banhatti indices and physicochemical properties of anticancer drugs

The physical properties of complex molecular compounds are influenced by structural descriptors in a nonlinear manner, therefore, artificial neural networks (ANNs) are used in QSPR analysis to capture these complex relationships. Molecular structure property interactions are frequently nonlinear, multimodal and interconnected, whereas traditional linear regression only captures straight line correlations (Table 8 & Figure 3). By modifying the inputs through hidden layers, artificial neural networks ($ANNs$) are specifically made to learn such nonlinear patterns, allowing the model to approximate complicated functions that cannot be represented by linear equations. Thus, the use of ANN in this study serves two purposes. To assess if the K-Banhatti topological indices contain nonlinear structure information applicable to physicochemical property prediction, ANN first functions as a supplementary modeling tool. Second, it offers a strong predictive framework that may identify relationships between topological characteristics that linear regression could miss. In this study, the ANN outperformed regression in prediction accuracy by successfully identifying a stronger nonlinear relationship for molar refraction (MR). This shows that when the property of interest depends on higher order or nonlinear combinations of chemical characteristics, ANN is very useful. Overall, by making it possible to methodically investigate and test both linear and nonlinear structure property connections, ANN improves the dependability of $QSPR$ modeling.

Case Processing Summary

		N	Percent
Sample	Training	9	69.2%
	Testing	4	30.8%
Valid		13	100.0%
Excluded		4	
Total		17	

(a)

The case processing summary presents the distribution of data used in model development and validation (Figure 4). Out of the total 17 anticancer drugs observations initially considered and 13 records were identified as valid and included for analysis while four records were excluded because of incomplete or inconsistent data or out of range. SPSS software randomly assigned 9 cases (69.2 %) to the training set and 4 cases (30.8 %) to the testing set, the model performance could be evaluated on both learned and unseen data.

The Network information table provides an overview of the network architecture utilized for the prediction model. The input covariates were five molecular descriptors: $B_1(G)$, $B_2(G)$, ${}^mB_1(G)$, ${}^mB_2(G)$ and $HB(G)$. Before training, all input variables are standardized to provide uniform scaling and improve the numerical stability of the model. The hyperbolic tangent activation function (tanh produces the output values from $(-1\ to\ +1)$ was used to a single hidden layer of four neurons that made up the neural network. Complex interactions between the chemical descriptors and the output variables can be captured

Network Information

Input Layer	Covariates	1	$B_1(G)$
		2	$B_2(G)$
		3	${}^mB_1(G)$
		4	${}^mB_2(G)$
		5	$HB(G)$
	Number of Units[a]		5
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		4
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	BP
		2	MP
		3	E
		4	MR
	Number of Units		4
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

(b)

Figure 4: (a) Case Processing Summary, (b) Network Information.

by this non-linear function. The output layers consisted of four dependent variables BP, MP, E and MR that represented the anticipated physicochemical characteristics made up the output layer. Before training, the other variables were also standardized. In order to maintain a straight linear link between the predicted values and the actual scaled output, the network employed an identity activation function at the output layer. Model performance was optimized using the sum of squares error function, which evaluates the squared differences between the expected and actual values. This design ensures a reasonable balance between model complexity and generalization ability, enabling effective property prediction based on the selected descriptors.

The network architecture (Figure 5) consists of three major components the input layer, hidden layer and output layer. The input layer contains five neurons corresponding to five K-Banahatti indices, namely $B_1(G)$, $B_2(G)$, ${}^mB_1(G)$, ${}^mB_2(G)$ and $HB(G)$, along with an input bias node that helps adjust the activation level. These descriptors supply structural information of anticancer drugs to the model. The information is fully propagated to the hidden layer, which consists of four neurons ($H(1:1)$ to $H(1:4)$) and an associated bias node. Each hidden unit applies a hyperbolic tangent activation function, enabling the learning of nonlinear correlations between indices and physicochemical properties. From the hidden layers, signals move toward four output neurons representing Boiling point (BP), Melting point (MP), Enthalpy (E) and Molar refraction (MR).

The connection lines between layers represent learned synaptic weights. Grey lines denote positive weights and blue lines indicate negative weights. In regression and neural networks positive or negative effect is normal both are important. A negative relationship may actually help increase prediction ac-

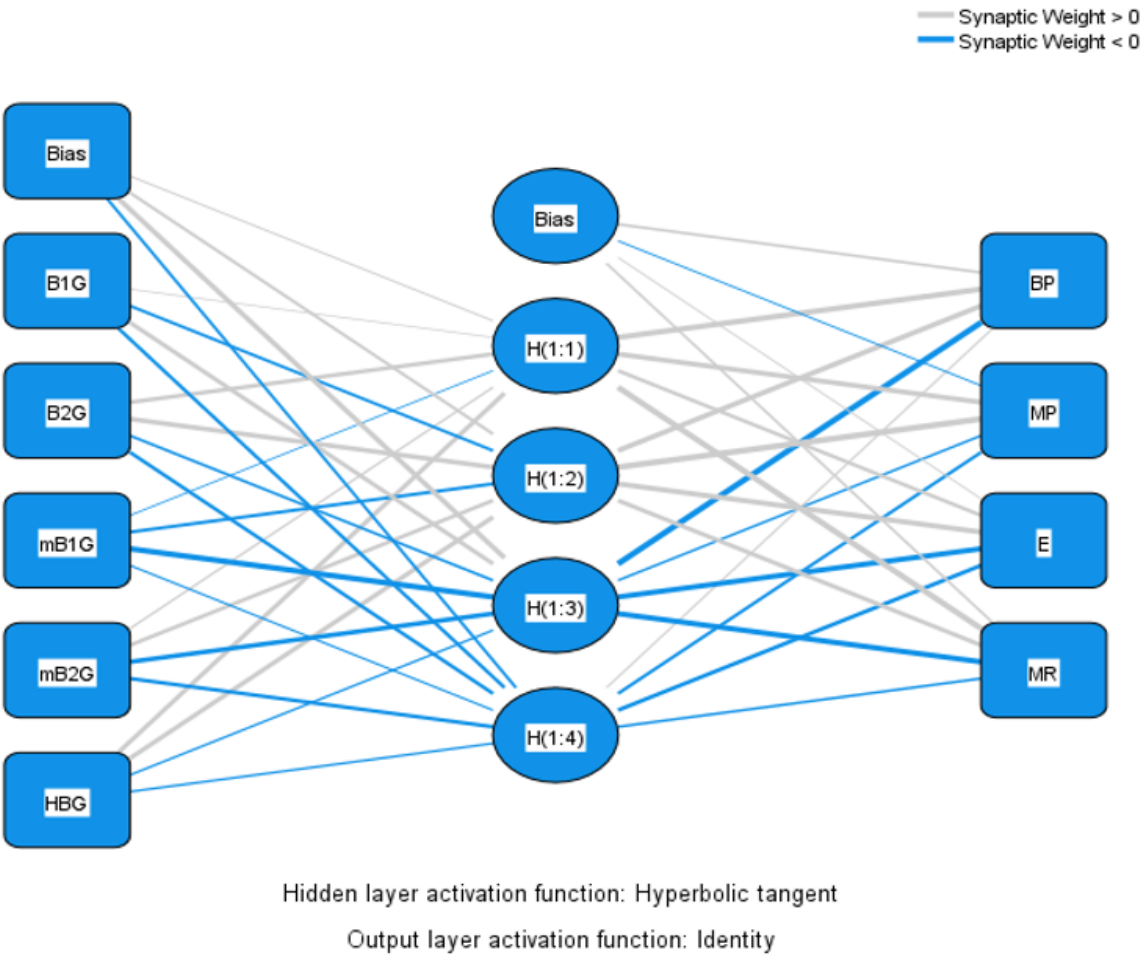


Figure 5: Neural network predicted values of BP, MP, E, MR.

curacy depending on how the property behaves with respect to that descriptor. The thickness of the lines reflects the magnitude of influence, thicker lines represent stronger contributions. From the connection distribution, it is evident that all indices influence all hidden neurons in a fully connected manner. Additionally, MR exhibits several strong connections, which aligns with its higher prediction accuracy observed during training. In the ANN model, the connection between ${}^{m}B_1(G)$ and MR also shows a relatively stronger synaptic weight (represented by a thicker and dark-blue line).

The Multilayer perception model (Figure 6 & Table 9) performed well during training, giving a low sum of squares error 3.822 and an average overall relative error of 0.239. Among the output variables, Molar refraction (MR) showed the least error 0.155, at the same time Melting point (MP) and Enthalpy (E) exhibited slightly higher deviations. The training automatically stopped after one iteration without further improvement, indicating stable convergence. On the testing data, the model showed higher SSE 7.161 and relative error 0.392, which is expected due to unseen observations. Molar refraction (MR) again showed the lowest error 0.122, confirming its strong predictive performance. Overall, the MLP model showed good learning and acceptable learning and generalization, performing best in predicting MR.

Model Summary

Training	Sum of Squares Error		3.822
	Average Overall Relative Error		0.239
	Relative Error for Scale Dependents	BP	0.242
		MP	0.280
		E	0.279
		MR	0.155
	Stopping Rule Used		1 consecutive step(s) with no decrease in error[a]
	Training Time		0:00:00.00
Testing	Sum of Squares Error		7.161
	Average Overall Relative Error		0.392
	Relative Error for Scale Dependents	BP	0.362
		MP	0.442
		E	0.545
		MR	0.122
a. Error computations are based on the testing sample			

Figure 6: Model Summary.

Table 9: Analysis of MLP Neural network and linear regression models.

Property	Linear regression	Training relative error	Testing relative error	Better performing method
Boiling point (BP)	r = 0.8843 for ${}^{m}B_1(G)$	0.242	0.362	Linear regression (slightly stronger correlation)
Melting point (MP)	r = 0.7530 for ${}^{m}B_1(G)$	0.28	0.442	Linear regression (better stability)
Enthalpy (E)	r = 0.8673 for ${}^{m}B_1(G)$	0.279	0.545	Linear regression (significantly better)
Molar refraction (MR)	r = 0.9347 for ${}^{m}B_1(G)$	0.155	0.122	ANN best predictive accuracy

Conclusion

Linear regression (LR) and Artificial neural networks (ANN) complement each other because they capture different types of relationships between molecular descriptors and physicochemical properties. Linear regression is ideal for identifying simple, interpretable, linear trends, allowing to determine which topological indices show the strongest direct correlations with each property. ANN, on the other hand, is capable of learning complex, nonlinear patterns that cannot be captured by linear models. When used together, LR provides clarity and interpretability, while ANN adds predictive flexibility and the ability to model hidden interactions among descriptors. In this study, LR revealed and the modified first K-Banhatti index ${}^{m}B_1(G)$ has a strong linear relationship with molar refraction (MR), whereas ANN further improved prediction accuracy by capturing additional nonlinear effects. Thus, combining LR and ANN offers a more comprehensive understanding of structure-property relationships, enhances model consistency and validates the stability of the selected descriptors. In linear regression, ${}^{m}B_1(G)$ exhibited the highest correlation with MR (Table 8), confirming its strong direct influence. In the ANN model (Table 9), the connection between ${}^{m}B_1(G)$ and MR also shows a relatively stronger synaptic weight, indicating that the neural network learned this strong dependency as well. Although the sign of the weight is negative, the magnitude highlights that ${}^{m}B_1(G)$ remains a dominant predictive factor for MR. Hence both regression and ANN consistently emphasize the significance of ${}^{m}B_1(G)$ in predicting MR.

Conflicts of Interest

The authors declare that they have no competing interests.

References

1. H. Meyer, *Theorie der Alkoholnarkose*, Arch. Exp. Pathol. Pharmakol. 42, 109–118, (1899).

2. E. Overton, *Studien Uber die Narcose*, Fischer, Jena, Germany, (1901).

3. A. Albert, *Selective toxicity: the physicochemical bases of therapy*, 7th ed., Chapman and Hall, London, p. 33, (1945).

4. P. H. Bell and R. O. Roblin Jr., *Studies in chemotherapy. VII. A theory of the relation of structure to activity of sulfanilamide type compounds*, J. Am. Chem. Soc. 64, 2905–2917, (1942).

5. L. P. Hammett, *Physical Organic Chemistry*, 2nd ed., McGraw-Hill, New York, (1970).

6. B. Figuerola and C. Avila, *The Phylum Bryozoa as a Promising Source of Anticancer Drugs*, Marine Drugs 17(8), 477–487, (2019).

7. G. Genovese et al., *Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence*, N. Engl. J. Med. 371(26), 2477–2487, (2014).

8. D. Afzal et al., *On computation of recently defined degree-based topological indices of some families of convex polytopes via M-polynomial*, Complexity, 1–11, (2021).

9. Y. H. Pan et al., *Topological study of polycyclic silicon carbide structure*, Polycycl. Aromat. Compd. 43(2), 1056–1067, (2023).

10. B. Furtula and I. Gutman, *A forgotten topological index*, J. Math. Chem. 53(4), 1184–1190, (2015).

11. I. Redžepović, *Chemical applicability of Sombor indices*, J. Serb. Chem. Soc., (2021).

12. I. Gutman et al., *On atom-bond connectivity index and its chemical applicability.*

13. B. Basavanagoud and P. Shruti, *Chemical applicability of Gourava and hyper-Gourava indices*, Nanosyst. Phys. Chem. Math. 12(2), 142–150, (2021).

14. M. Chellali et al., *On ve-degrees and ev-degrees in graphs*, Discrete Math. 340(2), 31–38, (2017).

15. S. Nandi and M. C. Bagchi, *Importance of Kier-Hall topological indices in the QSAR of anticancer drug design*, Curr. Comput.-Aided Drug Des. 8(2), 159–170, (2012).

16. M. Javaid and J. Cao, *Computing topological indices of probabilistic neural network*, Neural Comput. Appl. 30(12), 3869–3876, (2018).

17. J. Jaén-Oltra et al., *Artificial neural network applied to prediction of fluorquinolone antibacterial activity by topological methods*, J. Med. Chem. 43(6), 1143–1148, (2000).

18. M. Imran et al., *Topological properties of cellular neural networks*, J. Intell. Fuzzy Syst. 37(3), 3605–3614, (2019).

19. W. Gao et al., *Topological indices study of molecular structure in anticancer drugs*, J. Chem., 1–8, (2016).

20. S. A. K. Kirmani et al., *On ve-degree and ev-degree topological properties of hyaluronic acid-anticancer drug conjugates with QSPR*, J. Chem., 1–23, (2021).

21. A. Altassan et al., *Inverse Sum Indeg Index (Energy) with Applications to Anticancer Drugs*, Mathematics 10(24), 4749, (2022).

22. F. Chaudhry et al., *M-Polynomial and Degree-Based Topological Indices of Tadpole Graph*, J. Discrete Math. Sci. Cryptogr. 24(7), 2059–2072, (2021).

23. I. Gutman and N. Trinajstić, *Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons*, Chem. Phys. Lett. 17(4), 535–538, (1972).

24. N. Trinajstić, *Chemical graph theory*, Routledge, (2018).

25. A. Aslam et al., *Computing Certain Topological Indices of the Line Graphs of Subdivision Graphs of Some Rooted Product Graphs*, Mathematics 7, 393–401, (2019).

26. V. R. Kulli, *New K-Banhatti Topological Indices*, Int. J. Fuzzy Math. Arch. 12(1), 29–37, (2017).

27. V. R. Kulli, *On K-Banhatti Indices of Graphs*, J. Comput. Math. Sci. 7(4), 213–221, (2016).

28. M. E. Nazari et al., *M-Polynomial of Some Operations of Path and K-Banhatti Indices*, Math. Stat. Eng. Appl. 71(3s3), 38–55, (2022).

29. K. J. Gowtham, *The first and second K-Banhatti indices of some graph operations*, Int. J. Nonlinear Anal. Appl. 14(1), 3135–3144, (2023).

30. M. C. Shanmukha et al., *Degree based topological indices on anticancer drugs with QSPR analysis*, Heliyon 6, 1–9, (2020).

C. M. Veena,
Department of Mathematics,
GM Institute of Technology, Davanagere, 577006
and Affiliated to Visvesvaraya Technological University, Belagavi, 590018, India.
E-mail address: veenacm@gmit.ac.in

and

K. S. Onkarappa,
Department of Mathematics,
GM Institute of Technology, Davanagere, 577006
and Affiliated to Visvesvaraya Technological University, Belagavi, 590018, India.
E-mail address: onkarappaks@gmit.ac.in

and

K. C. Savitha,
Department of Mathematics,
GM Institute of Technology, Davanagere, 577006
and Affiliated to Visvesvaraya Technological University, Belagavi, 590018, India.
E-mail address: savithakc@gmit.ac.in

and

M. C. Shanmukha,
Department of Mathematics,
PES Institute of Technology and Management, Shivamogga, 577204
and Affiliated to Visvesvaraya Technological University, Belagavi, 590018, India.
E-mail address: shanmukhamc@pestrust.edu.in

and

A. Usha,
Department of Mathematics, Alliance University Alliance College of Engineering & Design,
Bangalore, 562106, India.
E-mail address: usha.arcot@alliance.edu.in