Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00367-6

Original Article

A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches

Hifza . [1, ✉], Hafiz Muhammad Atif [1]

1. *Department of Mathematics, University of Sialkot, Sialkot, PAK*

Received: June 19, 2026 | Review began: July 26, 2026 | Review ended: September 27, 2026 | Published: November 14, 2026

© **Copyright** 2026
This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Quantitative Structure-Property Relationship modeling is a valuable resource to predict drug-related physicochemical properties and supports rational drug development. Knee disorders are the most common musculoskeletal diseases which affect the knee joints, including arthritis and various inflammatory diseases. These disorders require effective pharmacological intervention to address their high prevalence and respective stability. In this study, distinct K-Banhatti indices were systematically evaluated for knee disorder drugs used to treat knee disorders. Their relationships with different physicochemical properties were studied to determine the property that could be predicted most accurately and the K-Banhatti index that was most suitable for predicting it. Regression models were applied and validated in Python using a dataset to develop a quantitative descriptor-property relationship. Among curvilinear regression models, cubic regression based on the product connectivity Banhatti index achieved the best performance with the lowest Root Mean Square Error (RMSE) = 1.707, Mean Absolute Error (MAE) = 1.347 and the largest R^2 = 0.963 value. Additionally, machine learning (ML) models [Artificial Neural Networks (ANN), Random Forest] were used to enhance the predictive accuracy of the models. All of the nonlinear models developed in this study significantly outperformed compared to the linear models across all evaluation metrics. Specifically, the ANN model has better accuracy with RMSE = 1.182, R^2= 0.982, and MAE = 0.885 and a smaller error distribution in comparison to other models. The models effectiveness were also demonstrated by a bar chart plot. It was observed that polarizability was identified as the best predicted property, based on its strong dependence on molecular descriptors. K-Banhatti indices combined with ensembled ML offer an effective approach for predicting physicochemical properties of musculoskeletal drugs. This computational framework enables and prioritizes promising drug candidates, minimizing unnecessary synthesis and experimental screening.

Categories: Ensemble Learning, Data Engineering, Machine Learning (ML)

Keywords: quantitative structured property relationship, random forest, artificial neural network (ann), k-banhatti indices, feature modelling

Introduction

The knee is the most complicated and biggest synovial hinge joint in the human body and is responsible for the movement. The author [1] explained the anatomical structure of the knee joint, namely, cartilage, bones, ligaments, tendons, bursae, and muscles and described each component's function. The National Library of Medicine published an extensive report [2] on knee injuries and disorders, specifically identifying osteoarthritis, rheumatoid arthritis, bursitis, ligament injuries, and meniscus injuries. The report identified how these knee injuries and disorders were responsible for knee pain, swelling, instability, and decreased ability to move and perform activities. A comprehensive review of the pharmacological therapies available [3] for the treatment of knee osteoarthritis was conducted. Their review included results from randomized controlled trials and meta-analyses, and supported the use of Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) for pain relief and intra-articular corticosteroids to improve short-term symptoms. A large-scale network meta-analysis [4] examined the effectiveness and safety of NSAIDs and opioids for treating knee and hip osteoarthritis. Their findings indicated that NSAIDs were superior to opioids and significantly improved knee and hip pain. A systematic review of knee osteoarthritis examined the clinical features, risk factors, and treatment options for this disease.

According to the study [5], the primary risk factors for developing knee osteoarthritis include aging, obesity, and repetitive stress to the knee joint and the importance of early diagnosis and using multiple treatment modalities to manage knee osteoarthritis is also discussed. Pharmacological therapies, surgical options, and physical therapy are the treatment options for knee osteoarthritis and stressed that need for a multidisciplinary approach to managing this disease to minimize long-term disability and improve the quality of life of individuals diagnosed with knee osteoarthritis. From 1990 to 2021, the global burden of knee osteoarthritis increased, with a faster rise among females, while population growth contributed over 70% of the overall burden [6].

Although computational methods have been shown to be effective at designing rationally targeted drugs, there is still a need to accurately predict the physicochemical properties of drugs to optimize their therapeutic efficacy. One approach to accomplish this, is by utilizing computational models such as Quantitative Structure-Property Relationship (QSPR). The important computational tools used for drug design, such as docking, Quantitative Structure-Activity Relationship (QSAR), and artificial intelligence can be integrated to improve prediction and lower costs. Tiwari also noted limitations related to data quality and the need for interpretable models [7]. Recent study created a foundation for chemical graph theory by defining the use of graph-theoretic descriptors to represent molecular structures. It also developed topological indices that are currently used in QSPR/QSAR modeling to predict physicochemical and biological properties. The book systematically connects graph invariants to molecular behavior and continues to be a fundamental reference in computational chemistry and drug design [8]. A recent study provided a comprehensive collection of references regarding topological indices and molecular descriptors commonly used in QSAR/QSPR modeling. The study describes how graph-theoretic parameters (i.e. connectivity indices, Wiener numbers, and eigenvalue-based descriptors) can predict physicochemical and biological properties of compounds. This book continues to be a foundational reference for researchers applying mathematical chemistry to drug design and property estimation [9]. Although the previous research has shown that topological modeling and QSPR can be useful for predicting physicochemical properties of molecules relevant to drug development, much of the research was focused on generalized drug datasets and did not examine disease-specific investigations. Therefore, despite the proven utility of many of the previous studies, the application of topological modeling and QSPR techniques to predict properties relevant to knee disorder therapeutics is underdeveloped. Furthermore, graph-theoretic descriptors can effectively model molecular behavior, supporting the continued use of topological indices in drug design and property estimation. Its approach enhanced predictive accuracy and provides a computational alternative to experimental profiling [10]. Ullah et al. evaluated the predictive power of K-Banhatti and Zagreb descriptors in modeling physicochemical properties of novel drug molecules. Their findings support the use of these topological indices in QSPR analysis for reliable property estimation and molecular screening [11]. These advanced models have shown improved predictive performance in various chemical and pharmaceutical applications. Nevertheless, many existing studies still restrict their analyses to basic regression techniques, limiting the accuracy and robustness of predictions. Kour and Sankar applied machine learning (ML) regression models to predict anti-cancer drug properties using topological indices within QSPR frameworks. Their study demonstrated that graph-based descriptors combined with ML algorithms enhance predictive accuracy and support efficient drug screening [12]. Shi et al. explored novel QSPR modeling techniques using topological indices to predict properties of cancer treatment compounds. Their study demonstrated improved accuracy and interpretability, supporting the role of graph-based descriptors in drug development [13]. Zhang et al. in their Machine Learning Study of Antidepressant Properties of Drugs (2026) used topological indices and ML as descriptors to predict the properties of antidepressant drugs. They showed that the accuracy of therapeutic properties estimation is significantly increased by combining the graph-theoretical descriptors with ML frameworks [14]. The eccentricity-based topological indices for type 2 antidiabetic drugs proposed by Prayia Karan et al. were verified to have strong predictive correlations with physicochemical properties [15]. Haider et al. developed a graph-based ML framework using topological indices to predict physicochemical properties of antiviral drugs. Their model demonstrated strong predictive performance, highlighting the synergy between molecular graph descriptors and ML algorithms in pharmaceutical modeling [16]. The K-Banhatti indices were introduced by Kulli.

The first and second K-Banhatti indices [17] of a graph G are defined as follows:

$$B_1(G) = \sum_{uv \in E(G)} \left[d_G(u) + d_G(e) \right] \tag{1}$$

$$B_2(G) = \sum_{uv \in E(G)} \left[d_G(u) d_G(e) \right] \tag{2}$$

The first and second K hyper Banhatti indices [18] of a graph G are defined as

$$HB_1(G) = \sum_{uv \in E(G)} \left[d_G(u) + d_G(e)\right]^2 \tag{3}$$

$$HB_2(G) = \sum_{uv \in E(G)} \left[d_G(u) \cdot d_G(e)\right]^2 \tag{4}$$

The sum connectivity Banhatti indices of Graph G is defined as

$$SB(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_G(u) + d_G(e)}} \tag{5}$$

The product connectivity Banhatti index [19] of a graph G is defined as

$$PB(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_G(u) d_G(e)}} \tag{6}$$

The modified first and second Banhatti indices [20] of a graph G are defined as

$${}^{m}B_1(G) = \sum_{ue} \frac{1}{d_G(u) + d_G(e)} \tag{7}$$

$${}^{m}B_2(G) = \sum_{ue} \frac{1}{d_G(u) d_G(e)} \tag{8}$$

The atom bond connectivity Banhatti index [21] of a graph G is defined as

$$ABCB(G) = \sum_{ue} \sqrt{\frac{d_G(u) + d_G(e) - 2}{d_G(u) \cdot d_G(e)}} \tag{9}$$

The geometric-arithmetic Banhatti index [22] of a graph G is defined as

$$GAB(G) = \sum_{ue} \frac{2\sqrt{d_G(u) \cdot d_G(e)}}{d_G(u) + d_G(e)} \tag{10}$$

The arithmetic-geometric Banhatti index of a graph G is defined as

$$AGB(G) = \sum_{ue} \frac{d_G(u) + d_G(e)}{2\sqrt{d_G(u) \cdot d_G(e)}} \tag{11}$$

The harmonic K-Banhatti index [21] of a graph G is defined as

$$H_b(G) = \sum_{ue} \frac{2}{d_G(u) + d_G(e)} \tag{12}$$

The inverse sum K-Banhatti of a graph G is defined as

$$IB(G) = \sum_{ue} \frac{d_G(u) \cdot d_G(e)}{d_G(u) + d_G(e)} \tag{13}$$

The majority of patients suffering from knee disorders have received some form of NSAID at some point in their disease process due to the ability of these drugs to significantly reduce inflammation and pain. Therefore, NSAIDs represent the main pharmacologic options for the treatment of knee disorders and osteoarthritis of the knee. Studies comparing NSAIDs such as Ibuprofen, Diclofenac, Meloxicam, Celecoxib, Naproxen, Etoricoxib, and Ketoprofen have consistently identified strong benefits in terms of efficacy for all of these drugs and there have been many studies where the efficacy of each of these drugs has been compared to one another. For example, Diclofenac and Etoricoxib were found to be two of the most effective NSAIDs studied in a large study comparing the efficacy of various NSAIDs [23]. In addition to the pharmacological evaluation of NSAIDs, there have been numerous QSPR/QSAR studies that model structural requirements for COX-2 selectivity, anti-inflammatory activity, and

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

ADMET properties (such as solubility, permeability, etc.) of analogs of Diclofenac, Ibuprofen, Naproxen, Meloxicam, and Celecoxib [24]. Other important classes of drugs include centrally acting medications and analgesics. For example, Duloxetine modulates chronic knee pain through central mechanisms while Acetaminophen only provides minimal pain relief. Tramadol, an opioid agonist, is typically reserved for use in patients who do not achieve adequate pain relief from other treatments, as it can produce serious side effects. QSPR/QSAR studies have modeled the analgesic potency and pharmacokinetics of derivatives of tramadol and duloxetine [25]. When knee pain is associated with muscle spasm, adjunctive therapies may help, including muscular relaxants such as Cyclobenzaprine, Methocarbamol, Tizanidine, Baclofen, and Thiocolchicoside. Topical medications such as Lidocaine, Capsaicin, and Diclofenac gel provide localized pain relief while intra-articular Corticosteroids, particularly Triamcinolone, temporarily relieve the symptoms of inflammatory flares. Febuxostat is critical for treating gout-induced knee pain while Glucosamine sulfate is a slowly-acting OA medication whose effectiveness is controversial [26]. Optimization of the chemical structure of both Glucosamine and Corticosteroid derivatives was supported by QSPR studies. Recent studies highlight persistent gaps in computational pharmacology. Despite advances in QSPR/QSAR, knee disorder drugs remain underexplored [27]. Many works still rely on limited descriptors, overlooking the predictive richness of advanced topological indices [28]. Comparative evaluations show that ML models outperform linear regression, yet such approaches are rarely applied to knee-related therapeutics [29,30].

Based on the literature review above, there is a major gap in predicting the properties of drugs used to treat knee disorders by QSPR analysis combined with ML. This study aims to bridge these gaps by developing a robust QSPR framework that integrates diverse topological indices with physicochemical descriptors. By applying advanced ML algorithms, it seeks to deliver accurate, disease-focused predictions for knee disorder drugs, aligning with recent calls for precision pharmacology and computationally guided therapeutics. Chemical structures of drugs used for the treatment of knee disorders are given in Figure *1*.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

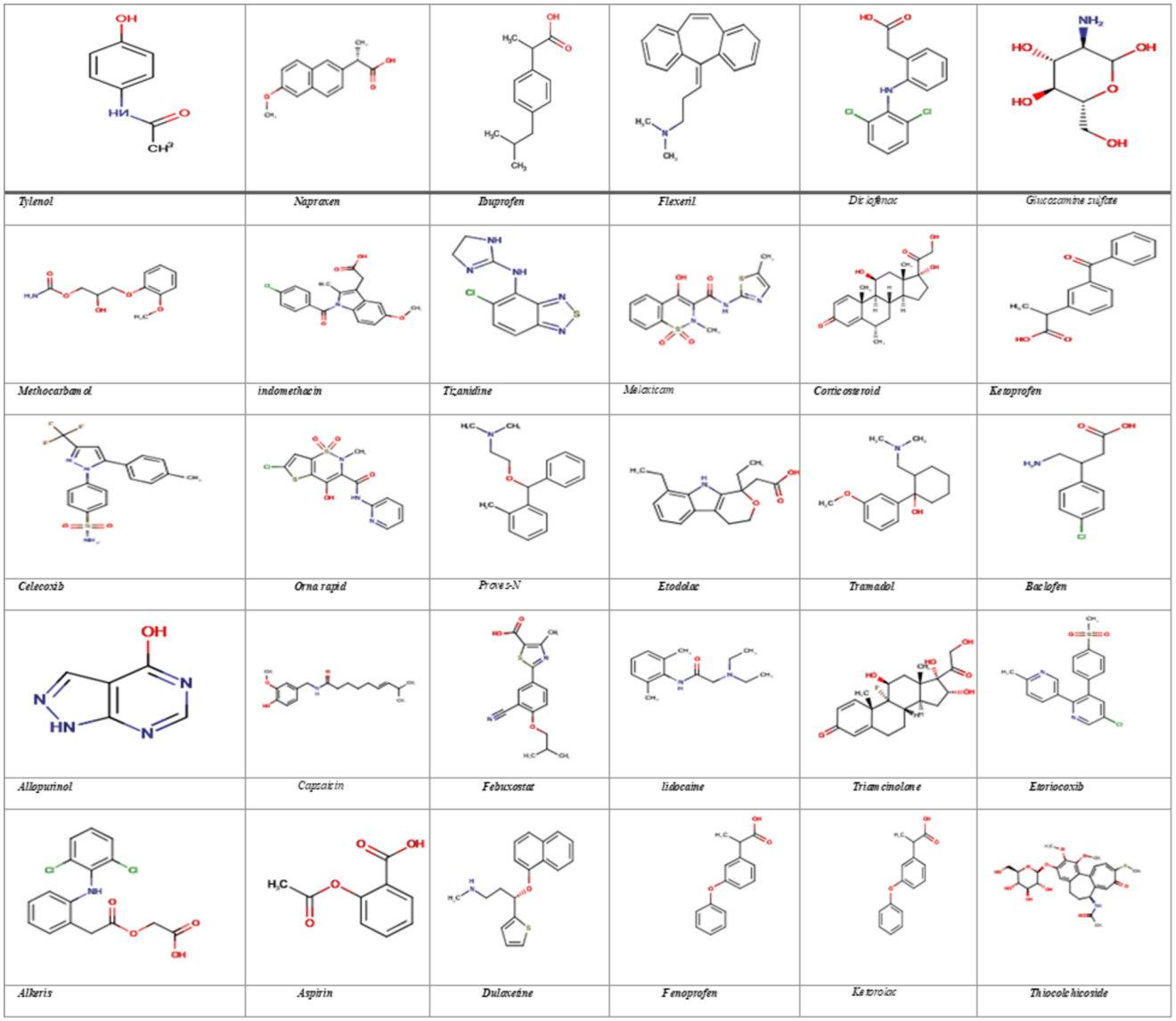


FIGURE 1: Chemical structure of drugs used for the treatment of knee disorder drugs

Physicochemical properties of drugs are listed in Table *1*. The physicochemical properties included Molecular weight (MW), Polarizability (Pol), Refractivity (Ref), Polar Surface Area (PSA), and Complexity (Com) for each drug name. All of these physicochemical properties are critical in understanding a drug's pharmacokinetic and pharmacodynamic properties, i.e., how a drug is absorbed, distributed, metabolized, and excreted.

Drugs	MW	Pol	Ref	PSA	Com
Tylenol	151.165	15.52	42.9	49.33	139
Ibuprofen	206.28	23.76	60.73	37.3	203
Diclofenac	296.149	27.93	75.46	49.33	304

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

Indomethacin	357.788	36.64	94.81	68.53	506
Meloxicam	351.401	34.25	88.62	99.6	628
Celecoxib	381.372	35.2	92.23	77.98	577
Naproxen	230.259	24.81	64.85	46.53	277
Flexeril	275.388	32.94	102.62	3.24	365
Methocarbamol	241.24	24.28	59.07	91.01	236
Tizanidine	253.711	23.97	64.77	62.2	300
Corticosteroid	374.47	40.77	103.04	94.83	754
Orna Rapid	371.81	33.3	87.9	99.6	634
Proves-N	269.381	31.83	84.97	12.47	260
Tramadol	263.381	30.51	78.27	32.7	282
Capsaicin	305.411	36.32	90.32	58.56	341
Lidocaine	234.338	27.77	73.93	32.34	228
Alkeris	354.18	32.76	86.32	75.63	411
Duloxetine	297.415	33.15	87.73	21.26	312
Ketorolac	255.269	26.67	70.19	26.67	376
Etodolac	287.353	31.94	81.16	62.32	399
Allopurinol	136.111	11.7	34.98	74.69	190
Febuxostat	316.375	33.88	93.93	83.21	448
Triamcinolone	394.433	39.8	99.38	115.06	807
Aspirin	180.158	17.1	44.45	63.6	212
Fenoprofen	242.269	25.3	68.18	46.53	271
Ketoprofen	254.28	26.56	72.52	54.37	331
Baclofen	213.661	21.13	54.83	63.32	191
Glucosamine Sulfate	179.171	16.87	37.58	116.17	155
Etoriocoxib	358.842	36.42	95.04	59.92	514
Thiocolchicoside	563.62	57.72	145.06	164.01	1010

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

TABLE 1: Physicochemical properties of knee disorder drugs

MW, Molecular Weight; Pol, Polarizability; Ref, Refractivity; PSA, Polar Surface Area; Com, Complexity

Materials And Methods

In this study, we investigated the combination of K-Banhatti indices with curvilinear regression and advanced ML models such as Random Forests (RF) and Artificial Neural Networks (ANN) to generate robust QSPR models that can predict physicochemical properties of knee disorder drugs. These indices are general structural measures, and their application in this study shows how graph-based descriptors can capture molecular connectivity and branching patterns that are important for drug property modeling.

From a methodological standpoint, the study first employed both linear and higher-order nonlinear relationships between the thirteen K-Banhatti topological indices and five physicochemical properties to determine the basic correlation between them and provide a basis for developing more advanced QSPR predictive models using ML algorithms.

Experimental details

Thirty suitable drugs for knee disorder were selected for QSPR study with five physicochemical properties namely; Molecular weight, Polarizability, Refractivity, Polar surface area and Complexity. These physicochemical properties and the chemical structures of the selected compounds are taken from PubChem, Chem Spider, and Drug Bank databases. All pre-processing and model building were performed in Python, and the following open-source libraries were employed; pandas, NumPy, scikit-learn, and Matplotlib. The well-known 13 topological indices are calculated for each drug given in Table *2*. The developed regression and ML models are assessed through using Root Mean Square Error (RMSE), Mean absolute Error (MAE), and coefficient of determination (R^2), the predictive accuracy of the models was evaluated allowing for comparison of the performance of each model.

Regression models

The linear and nonlinear regression analyses are used and their mathematical equations are given below

$$y = mx + c + \varepsilon \tag{14}$$

$$y = ax^2 + bx + c \tag{15}$$

$$y = ax^3 + bx^2 + cx + d \tag{16}$$

The statistical measure for determining how well a model predicts is the RMSE, which is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \bar{Y}_i\right)^2} \tag{17}$$

where:

Y_i= observed value,

$\bar{Y}_i$ = predicted value,

n = total number of observations.

The smallest values of RMSE, MAE, and MSE indicate good model performance since they indicate that the predicted values are closer to the observed values. To assist in making informed decisions regarding modeling strategies, it is beneficial to visually display the data and evaluate model performance using stacked finger plots, the coefficient of determination (R^2), and error

measures such as RMSE, MAE, and MSE.

Machine learning models

A data-based method called supervised ML is trained on labeled datasets to teach models the relationships between input characteristics and output values. Iterative optimization processes allow models to minimize the differences between the output generated by the model and the true value; therefore, supervised ML allows for predictive modeling. Examples of common supervised ML algorithms include RF, and ANNs, support vector regression, which may be suited to different types of data. Because they are capable of generalizing patterns learned from training data to novel, unseen cases, supervised learning is an effective tool for the design of drugs, image classification, and financial forecasting. The use of balanced datasets, careful feature selection, and proper cross-validation to avoid overfitting is important to achieve reliable results with supervised learning. In addition, supervised learning provides a systematic and replicable approach to making accurate predictions in many areas of both science and business.

Artificial neural network

ANNs are based on neurobiological concepts and represent a type of nonlinear, adaptive computer architecture. A combination of a stack of processing units (neurons) coupled by weighted synapses allows them to accomplish hierarchical feature extraction and sophisticated function approximations. An ANN develops latent representations that contain high-dimensional, nonlinear relationships between descriptors of the input and descriptors of the target, through iterative optimization techniques such as gradient descent and back-propagation. Due to their ability to generalize over multiple datasets, ANN have been particularly successful in developing predictive models of complex events in fields such as drug design, cheminformatics, and biomedical signal analysis. Variations of ANN such as deep neural networks, convolutional architectures, and recurrent models provide a broad base of support for contemporary ML applications through their ability to capture temporal dependencies, spatial correlations, and structured data.

The input features for the ANN model were selected topological descriptors namely, the 13 K-Banhatti indices. The ANN model consists of an input layer followed by two hidden layers, one with 64 neurons and the second with 32 neurons, both using the rectified linear unit (ReLU) activation function. The dataset is divided into an 80/20 train-test split.

Optimizer: Adam (learning rate = 0.001)

Loss function: MSE

Batch size:4

Epochs: 500

Random forest

RF regression was applied to an ensemble data set for better accuracy using decision trees. In the current study, RF regression was used to investigate structural-property correlations in a molecular data set. Each of the studied properties was modeled separately using a set of topological and physicochemical descriptors. The three key measures of accuracy and generalizability used to estimate the performance of the models included the coefficient of determination (R^2), RMSE, and the out of bag (OOB) score. The R^2 indicates the proportion of the experimental data variance explained by the predictions; the RMSE indicates the average magnitude of the prediction errors; and the OOB score is a form of internal cross-validation that evaluates how the model fits new, unseen bootstrap samples.

Hyperparameter

The RF regressor was implemented using scikit-learn. The model was trained with the following configuration:

Number of trees (n-estimators): 100

Bootstrap sampling: enabled (the out-of-bag score was calculated)

Random state: 42 (to ensure reproducibility)

This study was exploratory in nature; hence, an 80/20 train-test split or 5-fold cross-validation was applied.
Model reproducibility was assessed using OOB scores and standard error metrics (R^2, RMSE, MAE, MSE).
Future work will incorporate independent validation sets and k-fold cross-validation to strengthen reproducibility and

generalizability.

Results And Discussion

Theorem 1: Consider a graph of Allopurinol with an edge partitioning E(1,3) = 1, E(2,3) = 4, E(3,3) = 2, E(2,2) = 4, then the K-Banhatti indices of this structure are B1 = 112, B2 = 148, HB1 = 602, HB2 = 1212, SB = 9.9582, PB = 9.2364, MB1 = 4.5714, MB2 = 4.1111, ABCB = 15.1482, GAB = 21.8008, AGB = 22.2052, HB = 9.1429, IB = 29.5238 (Figure 2).

Proof:

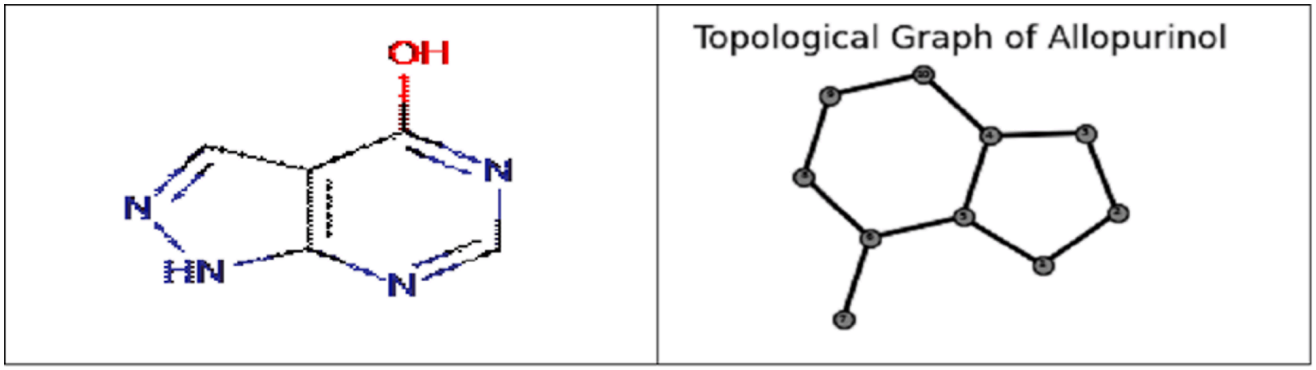


FIGURE 2: Molecular graph of allopurinol

$$B_1(G) = \sum_{u\in e}\left[d_G(u)+d_G(e)\right] + \sum_{v\in e}\left[d_G(v)+d_G(e)\right] = 1(3+5)+4(5+6)+2(14)+4(8) = 112$$

$$B_2(G) = \sum_{u\in e}\left[d_G(u)d_G(e)\right] + \sum_{v\in e}\left[d_G(v)d_G(e)\right] = 1(8)+4(15)+2(24)+32 = 148$$

$$HB_1(G) = \sum_{u\in e}\left[d_G(u)+d_G(e)\right]^2 + \sum_{v\in e}\left[d_G(v)+d_G(e)\right]^2 = 1(9+25)+4(25+36)+2(98)+4(32) = 602$$

$$HB_2(G) = \sum_{u\in e}\left[d_G(u)d_G(e)\right]^2 + \sum_{v\in e}\left[d_G(v)d_G(e)\right]^2 = 1(40)+4(117)+2(288)+4(32) = 1212$$

$$SB(G) = \sum_{u\in e}\frac{1}{\sqrt{d_G(u)+d_G(e)}} + \sum_{v\in e}\frac{1}{\sqrt{d_G(v)+d_G(e)}} = 1\left(\frac{1}{\sqrt{3}}+\frac{1}{\sqrt{5}}\right)+4\left(\frac{1}{\sqrt{5}}+\frac{1}{\sqrt{6}}\right)+2\left(\frac{2}{\sqrt{7}}\right)+4\left(\frac{2}{\sqrt{4}}\right) = 9.9582$$

.....

Similarly, we compute K-Banhatti topological indices of remaining drugs which are presented in Table 2.

Drugs	B1	B2	HB1	HB2	SB	PB	MB1	MB2	ABC	GAB	AGB	H	IB
Tylenol	106	130	532	886	10.2065	9.7956	4.8	4.667	15.3137	21.6466	22.3676	9.6	25.8
Ibuprofen	150	194	796	1542	13.7674	13.181	6.4381	6.3333	20.7241	29.4508	30.5716	12.8762	36.3905
Diclofenac	202	264	1078	2120	18.2097	17.1261	8.4238	7.8889	27.5911	39.4672	40.5523	16.8476	49.3524
Indomethacin	288	400	1636	3604	24.1089	22.4827	11.0333	10.5556	36.1693	53.19	54.839	22.0667	70.2

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

Meloxicam	278	404	1662	4220	21.8513	19.9765	9.7532	8.7889	34.131	49.0619	51.0019	19.5063	67.3325
Celecoxib	314	452	1868	4604	24.1995	21.9061	10.6198	9.2333	38.6494	54.6773	57.4593	21.2397	75.5325
Naproxen	186	250	1022	2106	16.3251	15.4595	7.5905	7.5	24.0151	35.4672	36.5523	15.181	45.3524
Flexeril	232	304	1236	2456	20.9165	19.4728	9.6429	8.7222	31.7105	45.6015	46.4104	19.2957	57.0476
Methocarbamol	160	196	800	1392	16.1023	15.8216	7.7857	8.1111	22.9282	33.5487	34.4696	15.5714	39.0952
Tizanidine	186	248	1008	2032	16.136	14.7801	7.3238	6.3889	24.7627	35.6994	36.3048	14.6476	45.7524
Corticosteroids	388	620	2564	7964	27.0613	24.3059	11.8123	10.7514	42.3239	62.5521	65.5496	23.6246	93.3865
Orna Rapid	295	439	1807	4845	22.5382	20.4583	9.9893	8.9056	35.355	51.0239	53.0409	19.9786	71.4298
Proves-N	204	258	1050	1974	19.4742	18.5276	9.1571	8.6667	29.0862	41.5851	42.4297	18.3143	50.0857
Tramadol	189	259	1069	2513	16.1483	15.1012	7.3718	6.825	24.9291	35.2832	36.7768	14.7437	45.6258
Capsaicin	206	250	1020	1730	20.8378	20.4201	10.0524	10.3333	29.9184	43.4309	44.5921	20.1048	50.3619
Lidocaine	166	214	876	1760	15.9982	15.8476	7.7714	8.5556	22.0575	33.4508	34.5716	15.5429	40.3905
Alkeris	240	310	1266	2426	21.9452	20.7246	10.1905	9.6111	33.1671	47.3494	48.6748	20.381	58.619
Duloxetine	226	290	1176	2242	21.2514	20.1134	9.9905	9.5	31.0862	45.6994	46.3084	19.981	55.7524
Ketorolac	222	306	1248	2742	18.742	17.2597	8.5143	7.6667	28.7165	41.5235	42.4919	17.0286	54.3714
Etodolac	250	362	1486	3774	20.6023	19.4379	9.554	9.6236	29.8624	45.2978	46.7321	19.1079	60.7782
Allopurinol	112	148	602	1212	9.9583	9.2364	4.5714	4.1111	15.1482	21.8008	22.2052	9.1429	29.5238

Febuxostat	230	310	1270	2626	19.7257	18.4173	9.0095	8.4444	30.2154	43.2512	44.7768	18.019	55.9143
Triamcinolone	371	601	2495	8451	26.7388	24.7706	12.0126	11.8681	40.4168	60.4522	63.6707	24.0254	88.9052
Aspirin	128	164	672	1300	12.032	11.5824	5.6714	5.6111	17.9766	25.5686	26.4491	11.3429	31.1238
Fenoprofen	188	240	978	1824	17.4292	16.4334	8.1048	7.5556	26.3001	37.5652	38.4503	16.2095	46.0571
Ketoprofen	202	266	1086	2206	18.2547	17.2203	8.4762	8	27.5488	39.4871	40.5318	16.9524	49.381
Baclofen	136	172	704	1296	13.1023	12.8216	6.2857	6.6111	18.6856	27.5487	28.4696	12.5714	30.0952
Glucosamine Sulfate	126	176	724	1668	10.9728	10.7027	5.2095	5.6667	15.6493	23.4906	24.5305	10.419	30.4476
Etoricoxib	280	388	1594	3592	22.9032	20.9416	10.2385	9.0333	35.8678	51.0706	53.0064	20.477	67.9091
Thiocolchicoside	439	615	2517	5633	36.7389	34.6281	16.9762	17	53.7781	80.6973	83.3501	33.9524	106.881

TABLE 2: Calculated k-Banhatti topological indices

A combined 3D plot for each drug with their topological indices and physicochemical properties is displayed in Figure *3*.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

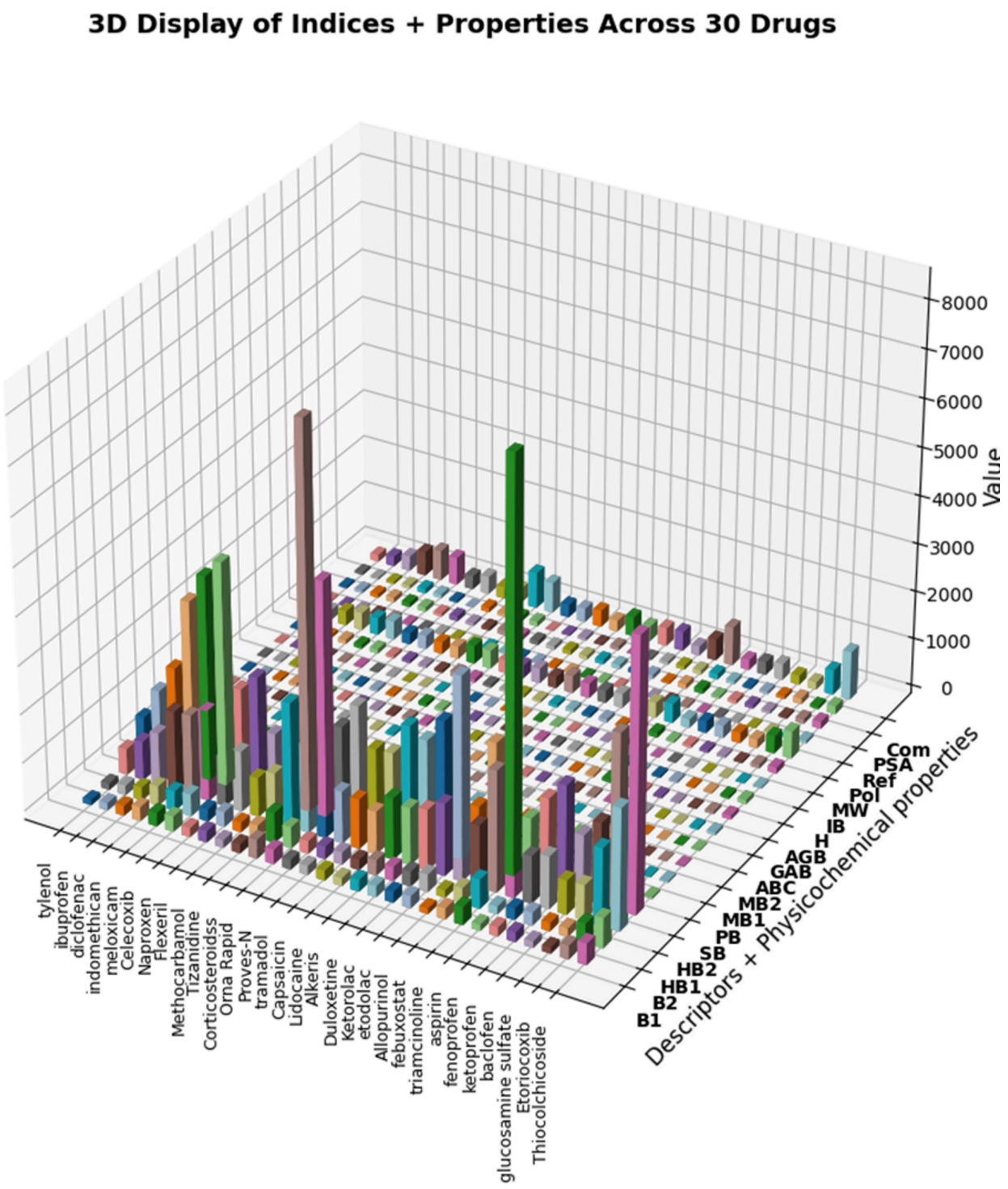


FIGURE 3: The 3D display of K-Banhatti indices and physicochemical properties across 30 drugs

This section outlines the results from regression analyses and ML predictions and compares the performance of classical statistical methods with that of more modern algorithms. The regression models (linear, quadratic, and cubic) were used to establish baseline correlations between the physicochemical properties and the topological indices calculated from the eccentricity of the atoms in the molecules, whereas the ML models (ANN and RF) were applied to the molecules to capture the complex nonlinear relationships and to increase the predictive accuracy of the models.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

The regression results are summarized into two parts: first, the strength of the association of descriptors with properties is determined and second, the results of the ML are discussed, mainly focusing on the comparative performance measures. This biperspective allows a complete assessment of predictive methods and highlights the innovativeness of the combination of mathematical graph theory with computational learning technology.

Linear results were verified using Table *3* based on the high degree of correlation (R > 0.975) between molecular weight, polarizability, refractivity, PSA, and drug properties. Polarizability was the most reliable index, with the highest values for the PB, MB1, and H indices. PSA, refractivity, and complexity were less correlated, indicating a minor contribution. A few drugs had very low correlation, especially HB2, which suggests that nonlinear regression may be appropriate in those cases and that linear descriptors alone will not adequately explain their behavior. The study demonstrates that PSA, Ref, and Com contributed less to the linear context, but MW, and P contributed greatly. Correlation values of R > 0.975 are emphasized in bold in Table *3*.

TI's	MW	Pol	Ref	PSA	Com
B1	0.948	0.9207	0.899	0.5128	0.9735
B2	0.8998	0.8642	0.8349	0.5419	0.9709
HB1	0.896	0.8599	0.8298	0.5442	0.9699
HB2	0.7629	0.7225	0.6818	0.5445	0.9043
SB	**0.9751**	0.9739	0.9591	0.4445	0.9266
PB	0.9705	**0.979**	0.9608	0.4418	0.9066
MB1	0.9675	**0.9783**	0.9618	0.4306	0.8995
MB2	0.9269	0.9576	0.9237	0.4652	0.8484
ABC	0.9746	0.9599	0.9505	0.4374	0.9401
GAB	0.9734	0.9632	0.9488	0.4592	0.94531
AGB	0.9739	0.961	0.94339	0.4727	0.9519
H	0.9675	**0.9783**	0.9618	0.4304	0.8995
IB	0.9489	0.9239	0.9035	0.5035	0.9710

TABLE 3: Correlation coefficients values obtained from linear regression

MW, Molecular Weight; Pol, Polarizability; Ref, Refractivity; PSA, Polar Surface Area; Com, Complexity

The stacked finger plots support this finding by demonstrating the relative contribution of each descriptor across drugs. As such, they provide evidence that MW and Pol contribute similarly and consistently across all drugs. The overall tabular and graphical results indicate that several models fit well, but the most reliable models were the PB, SB, and MB1 models as shown in Figure *4*.

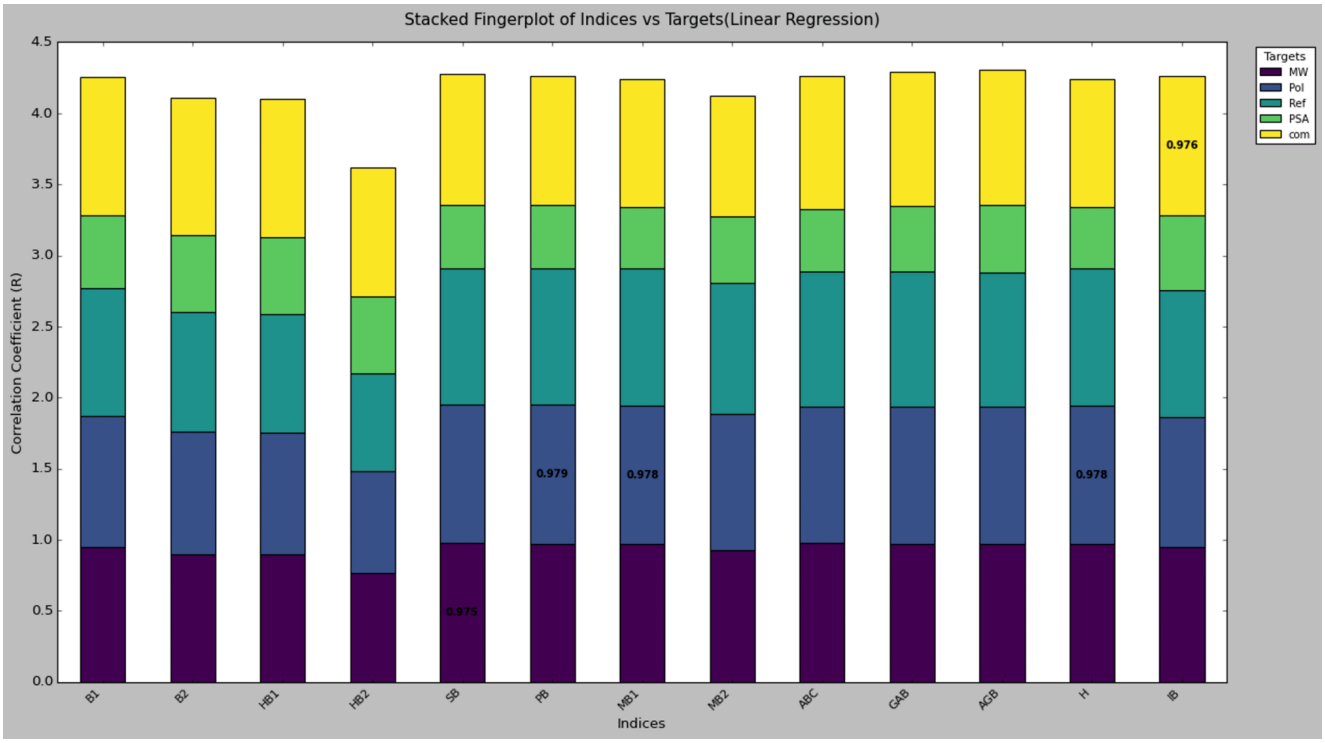


FIGURE 4: Stacked Finger plot of indices vs targets obtained from linear regression

As shown from Table *4*, correlation values R > 0.975 and > 0.957 demonstrate that many of the linear models are highly accurate in predicting drug properties. The PB index is the most reliable and robust of the indices tested, as demonstrated by its best correlation and lowest error values. Additionally, the robustness of MW, and Pol as descriptors is also supported by SB, PB, and MB1 models, which consistently performed well. However, the SB model showed a poorer generalization capability when compared to the other models, as indicated by higher error values and lower external predictivity values.

Model	R	R^2	MSE	RMSE	MAE	F-STAT	P value
MW = A + B(SB)	0.9751	0.950	364.366	19.088	16.304	541.846	0
Pol = -A + B(PB)	0.979	0.9585	3.261	1.806	1.370	646.943	0
Pol = A + B(MB1)	0.978	0.9571	3.372	1.8363	1.3901	624.847	0
Pol = A + B(H)	0.978	0.9571	3.371	1.836	1.389	624.92	0

TABLE 4: Statistical parameter metric obtained from linear regression

MSE, Mean Squared Error; RMSE, Root Mean Square Error; MAE, Mean Absolute Error

Strong nonlinear relationships are seen between Molecular descriptors and drug indices for most Quadratic results, Especially Pol and Com. The top values for Pol were found using MB1, PB and H, B1 and IB have the Largest Correlation Values with complexity while PSA, MW, and Ref marginally outperformed linear values. Overall, PB, MB1, IB, and B1 are identified as top performers by Quadratic modeling, which improves overall descriptor to index associations as verified in Table *5*.

TI's	MW	Pol	Ref	PSA	Com
B1	0.9488	0.9239	0.9093	0.7235	0.9802
B2	0.9148	0.8814	0.8683	0.6747	0.9712
HB1	0.9134	0.8796	0.8663	0.6693	0.97003
HB2	0.8703	0.8172	0.8008	0.5456	0.93702
SB	0.9754	0.9741	0.9620	0.6815	0.93486
PB	0.9705	0.9798	0.9646	0.6622	0.9124
MB1	0.9677	0.9789	0.9654	0.6544	0.9052
MB2	0.9279	0.9623	0.9326	0.6263	0.85002
ABC	0.976	0.9598	0.9519	0.7119	0.9551
GAB	0.9738	0.9634	0.9517	0.7013	0.9549
AGB	0.9741	0.9614	0.9477	0.7102	0.9614
H	0.9676	0.9789	0.9655	0.6544	0.9052
IB	0.9491	0.9257	0.9117	0.72316	0.97918

TABLE 5: Correlation coefficient value obtained from quadratic regression

MW, Molecular Weight; Pol, Polarizability; Ref, Refractivity; PSA, Polar Surface Area; Com, Complexity

The results were graphically represented by stacked bar plot given in Figure *5*, which shows that descriptor PB, B1, MB1, IB and H have the good correlations, where polarizability lead. The color-coded sections show that while PSA, MW and Ref have less influence, Com and Pol consistently contribute to initiative model performance. PB shows the best overall performance; MB1 and B1 are good substitutes, and IB should be taken vigilantly because of high error levels.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

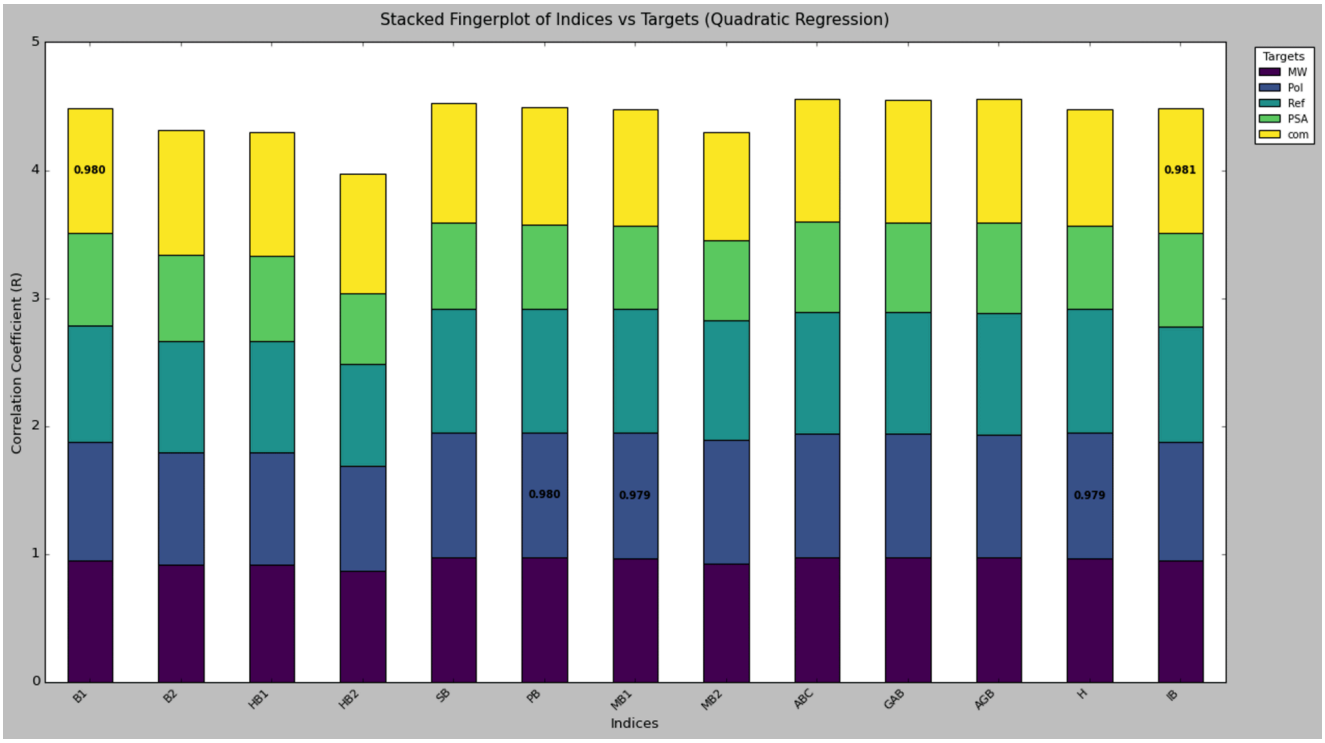


FIGURE 5: Stacked Finger plot of indices vs targets obtained from quadratic regression

With correlation values (R) above 0.978 and R^2 above 0.958, the quadratic regression analysis consistently performs well across the majority of models as given in Table *6*. This table reveals that PB is the most consistent index for predicting polarizability with R = 0.979, R^2 = 0.960, and the lowest error metrics (MSE = 3.142, RMSE = 1.772, MAE = 1.373). In spite of having good fits, IB and B1 display rather well correlations (R = 0.981 and 0.980), but their error values are significantly bigger, declining exterior predictivity and making them less generalizable.

Model	R	R^2	MSE	RMSE	MAE	F-STAT	P value
Com = A + B(B1) + C(B1)^2	0.980	0.960	1659.11	40.732	28.80	331.72	0
Pol = A + B(PB) + C(PB)^2	0.979	0.960	3.142	1.772	1.373	324.247	0
Pol = A + B(MB1) + C(MB1)^2	0.978	0.958	3.2715	1.808	1.401	310.945	0
Pol = A + B(H) + C(H)^2	0.979	0.958	3.271	1.808	1.401	310.96	0
Com = A + B(IB) + C(IB)^2	0.9791	0.9588	1747.99	41.809	29.728	314.712	0

TABLE 6: Statistical parameter metric obtained from quadratic regression

MSE, Mean Squared Error; RMSE, Root Mean Square Error; MAE, Mean Absolute Error

The cubic regression results confirm strong nonlinear relationships between molecular descriptors and drug indices with correlation values over 0.980 for most models; PB, MB1, and H have the greatest Polarizability (Pol) values and PB index has the greatest overall values (R = 0.9812). Similarly, IB and B1 lead property Com and IB has the greatest predictive alignment with the greatest correlation (R = 0.9812). Although its value for polarizability is still lower than those for MW, Pol, Ref, and Com, PSA continues to improve upon earlier models for IB (0.7400), ABC (0.7392), and AGB (0.7362). Overall, IB has the greatest correlation with Com and PB has the greatest correlation with Polarizability. These results, which consistently show that PB, MB1, B1, and IB are major contributors to cubic regression interactions between descriptors and indices, support the utility of cubic regression in capturing complex interactions between descriptors and indices (Table *7*).

TI's	MW	Pol	Ref	PSA	Com
B1	0.9603	0.9611	0.9479	0.7345	0.9812
B2	0.9166	0.9072	0.8950	0.7105	0.9726
HB1	0.9145	0.9024	0.8901	0.7113	0.9717
HB2	0.8737	0.8175	0.8031	0.7090	0.9498
SB	0.9754	0.9785	0.9676	0.7163	0.9527
PB	0.9706	0.9812	0.9669	0.6982	0.9338
MB1	0.9677	0.980	0.9670	0.6891	0.9264
MB2	0.9281	0.9631	0.9341	0.6471	0.8633
ABC	0.9777	0.9714	0.9646	0.7392	0.9647
GAB	0.9752	0.9740	0.9641	0.7299	0.9669
AGB	0.9766	0.952	0.9635	0.7362	0.9712
H	0.9677	0.980	0.9671	0.6890	0.92644
IB	0.9587	0.9585	0.9461	0.7345	0.98071

TABLE 7: Correlation coefficient value obtained from cubic regression

MW, Molecular Weight; Pol, Polarizability; Ref, Refractivity; PSA, Polar Surface Area; Com, Complexity

These results are supported by the stacked finger plot given in Figure *6*, which displays the strongest overall correlations for PB, MB1, B1, H, and IB; descriptor contributions are dominated by MW, Pol, and Com; however, PSA and Ref are less significant. In conclusion, IB leads in descriptor Com, but with a greater error, and PB is the best performing index for polarizability; since PB strikes a balance between precision and stability, it is the most reliable option among cubic models.

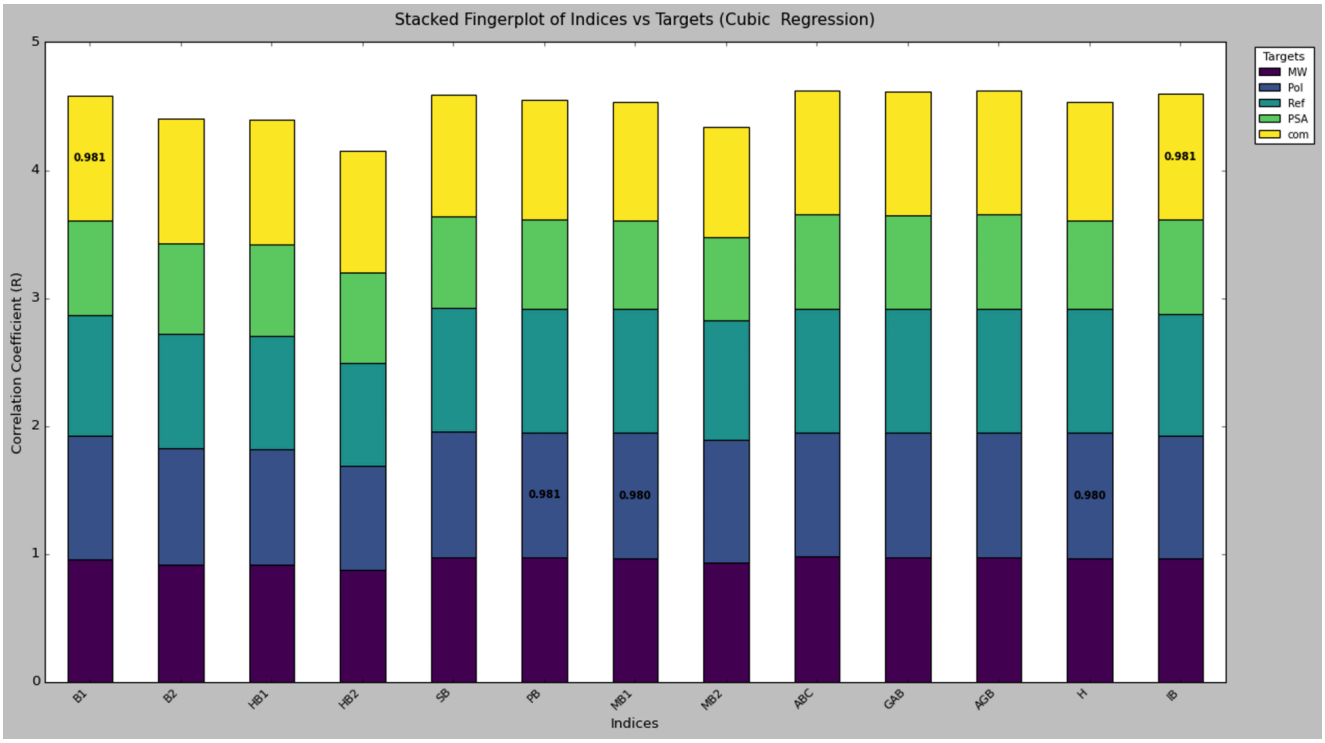


FIGURE 6: Stacked Finger plot of indices vs targets obtained from cubic regression

Overall, the stacked finger plot and cubic regression results provide evidence of strong nonlinear relationships between drug indices and descriptors. Most models exhibit excellent internal validity (>0.960) and high correlation (R > 0.980); the PB-based model (highlighted in bold) exhibits the best overall results and provides the best predictions of polarizability. With the lowest error metrics (MSE = 2.914, RMSE = 1.7071, MAE = 1.3473), the PB-based model emerges as the best overall model. Thus, PB is the most reliable index for predicting polarizability as shown in Table *8*.

Model	R	R^2	MSE	RMSE	MAE	F-STAT	P value
Com = A + B(B1) + $C(B1)^2$+$D(B1)^3$	0.981	0.962	1579.99	39.748	29.1983	224.071	0
Pol = A + B(PB) + $C(PB)^2$ + $D(PB)^3$	0.981	0.962	2.914	1.7071	1.3473	225.156	0
Pol = A + B(MB1) + $C(MB1)^2$ + $D(MB1)^3$	0.980	0.960	3.114	1.7649	1.3867	210.105	0
Pol = A + B(H) + $C(H)^2$ + $D(H)^3$	0.980	0.960	3.114	1.7648	1.3861	210.117	0

Com = A + B(IB) + $C(IB)^2$ + $D(IB)^3$	0.9807	0.9617	1620.91	40.2605	29.6717	218.183	0

TABLE 8: Statistical parameter metric obtained from cubic regression

MSE, Mean Squared Error; RMSE, Root Mean Square Error; MAE, Mean Absolute Error

Comparative result of linear, quadratic, cubic models

As given in Table *9* the top three of the best-performing linear, quadratic, and cubic models were statistically analyzed. Among the linear, quadratic, and cubic models, the cubic model was found to be the most effective, with the greatest R^2 and lowest MSE, MAE and RMSE for predicting the physicochemical property polarizability. Drug polarizability is the measure of how easy it is to alter a molecule's electron cloud via an external field. Polarizability can affect how drugs interact with a target and thus influence pharmacological activity and drug-target recognition through its effect on intermolecular interactions, solubility, and binding affinity.

Models	Equations	R^2	MSE	RMSE	MAE
Linear	P = -A + B(PB)	0.9585	3.261	1.806	1.370
Quadratic	P = A + B(PB) + $C(PB)^2$	0.960	3.142	1.772	1.373
Cubic	P = A + B(PB) + $C(PB)^2$ + $D(PB)^3$	0.962	2.914	1.7071	1.3473

TABLE 9: Statistical parameter metric obtained from traditional regression

MSE, Mean Squared Error; RMSE, Root Mean Square Error; MAE, Mean Absolute Error

The bar graphs in Figure *7* show the difference in the predictions from the linear, quadratic and cubic models compared to the actual polarizability values of all medicines examined. As far as visual inspection goes, the cubic model (yellow bars) best represented the true data (dark blue) with very little variation for the majority of the medicines. Therefore, cubic regression provides the most accurate predictions. Indices like PB, MB1, IB and B1, used previously in table format to compare actual versus cubic models, show the strongest agreement between the actual and cubic models. On the other hand, the linear model (light blue) overestimates.

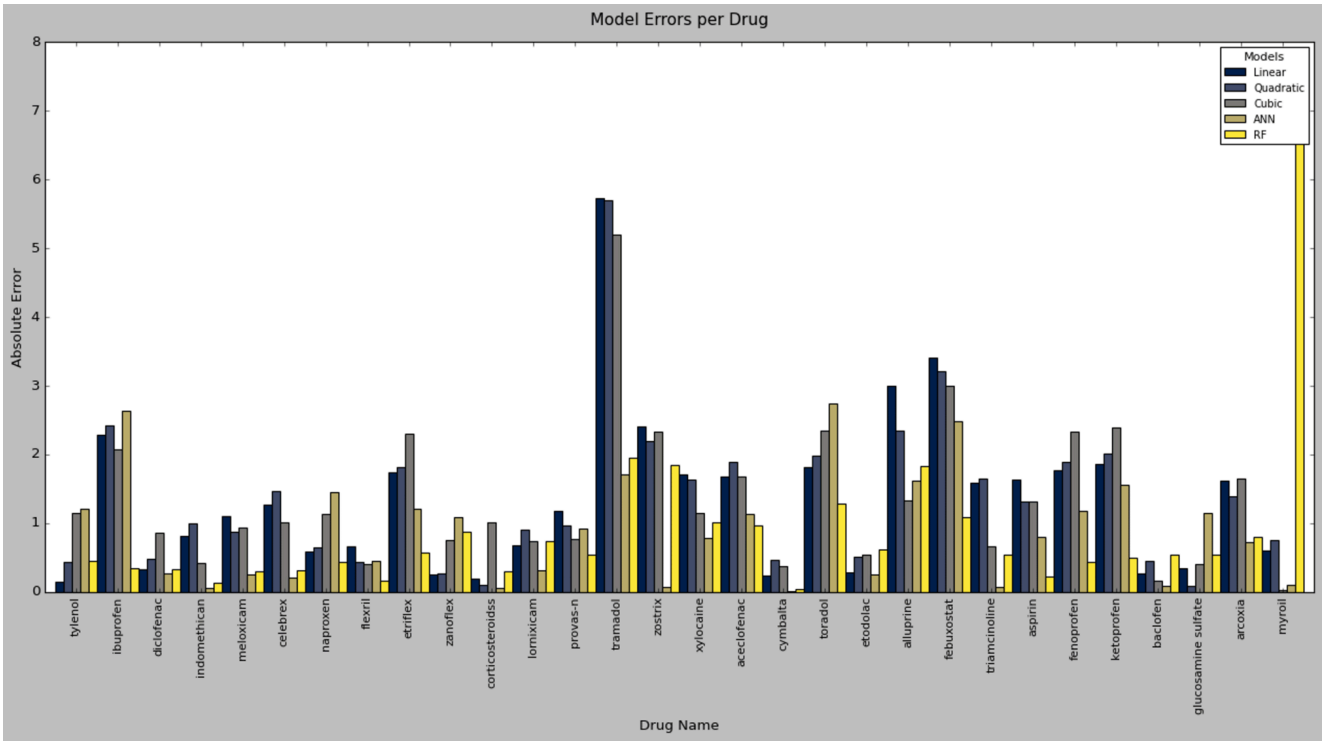


FIGURE 7: Bar chart plot of actual vs predicted polarizability obtained from linear, quadratic, cubic regression

and/or underestimates polarizability for many of the medicines that have complex behaviors. While the quadratic model (gray) performed better than the linear model, there are several instances where it did not perform as well as the cubic model. In general, the graphs demonstrate that Pol was the most accurately predicted property, and that cubic regression best describes the nonlinear nature of polarizability.

Machine learning results

All of the physicochemical variables analyzed by the ANN model demonstrated excellent prediction accuracy by 80\20 train-test spilt and 5-fold cross validation technique, as given in Table *10*. High correlation coefficients (R > 0.940) and determination coefficients were found for all of the physicochemical properties except for PSA, further supporting the model's reliability and low error rates. Of those properties studied, Polarizability shown to be the most predictive property (R = 0.969, R^2 = 0.933, RMSE = 1.480), which is consistent with the descriptors used. Although PSA showed negative correlation coefficient values relative to larger error magnitude values, Ref showed highly predictable values.

Property	R	R^2	RMSE	MAE	MSE
MW	0.948	0.518	28.784	25.265	828.498
Pol	0.969	0.933	1.480	1.181	2.190
Ref	0.931	0.852	6.520	5.716	42.506
PSA	0.358	-0.167	34.957	28.803	1222.007
Com	0.940	0.675	30.298	27.581	917.941

TABLE 10: Statistical parameter metric obtained from artificial neural network

MSE, Mean Squared Error; RMSE, Root Mean Square Error; MAE, Mean Absolute Error; MW, Molecular Weight; Pol, Polarizability; Ref, Refractivity; PSA, Polar Surface Area; Com, Complexity

Overall, the ANN framework was very successful in predicting the physicochemical properties with Pol being the most predictive property due to the clarity of descriptor linkage. The reason for this robust result is that polarizability is typically a smooth scaling quantity with respect to both molecule size and electronic dispersal. Thus, polarizability has a stronger association with the model descriptors than the other properties.

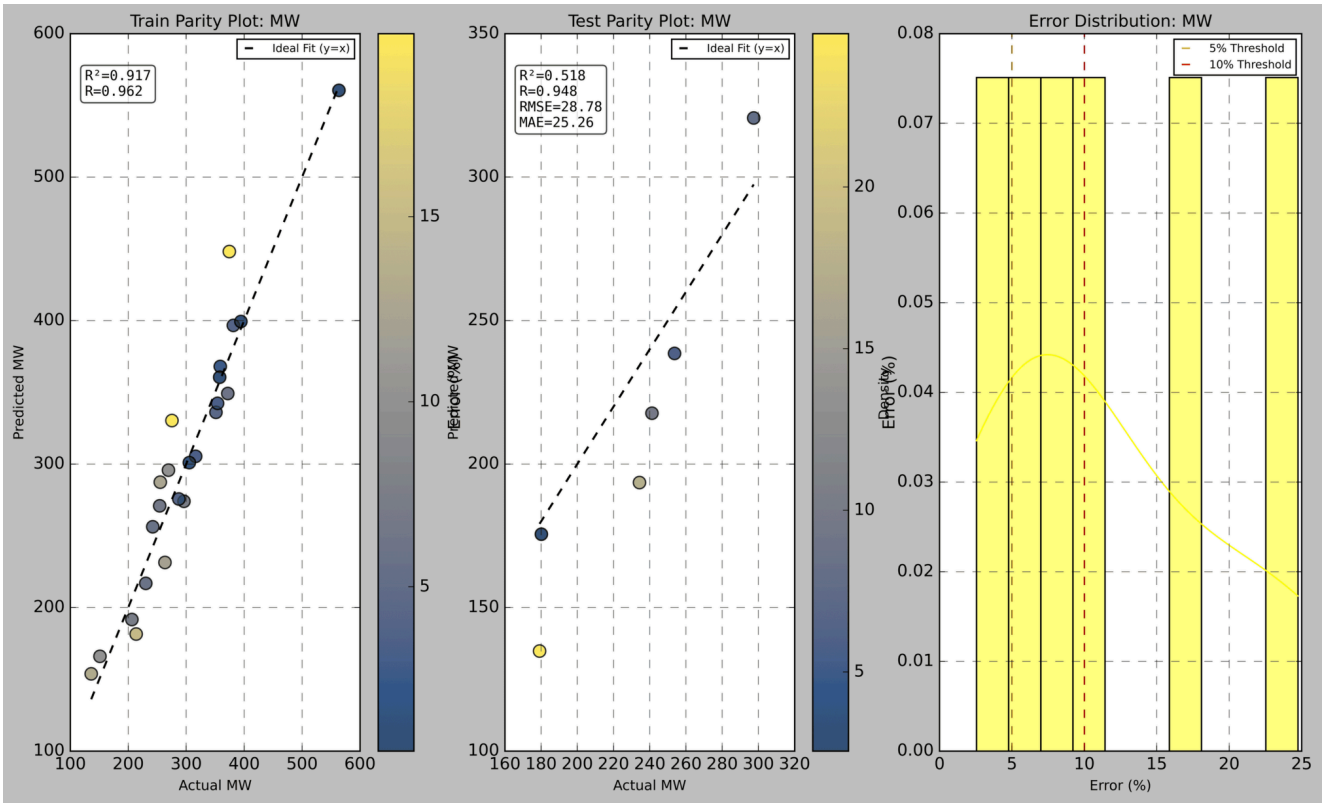


FIGURE 8: Parity and error distribution plots for MW

MW, Molecular Weight

In addition, the parity plot of the MW predictions shows a high degree of correspondence between the predicted MW and actual MW ($R = 0.948$, $R^2 = 0.518$). Additionally, the error distribution indicates that the majority of the data points lie below the 5% threshold. Therefore, collectively, these results demonstrate a high level of reliability in the MW predictions and low levels of variation in MW predictions as shown in Figure *8*.

Similarly, the parity plot of the predictions of polarizability (Pol) shows a high degree of correspondence between the predicted Pol and the actual Pol ($R = 0.969$, $R^2 = 0.933$) as shown in Figure *9*. Furthermore, the error distribution indicates that 27 predictions fell below the 5% threshold, indicating that the predictions were performed with high levels of reliability and accuracy. These results indicate that Pol is the most consistently predicted property, likely due to the fact that the model's descriptors provided a sufficient representation of the smooth dependence of Pol on molecule size and electronic dispersal.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

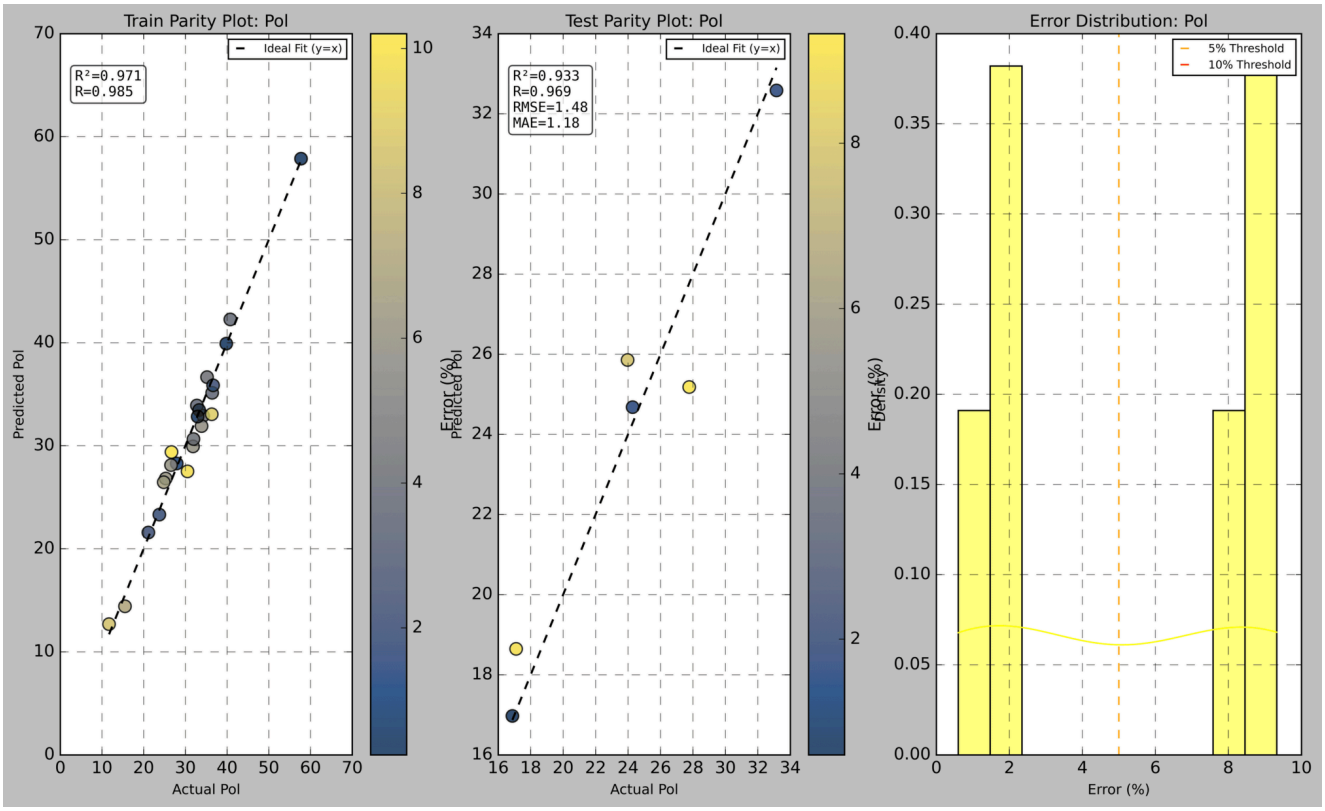


FIGURE 9: Parity and error distribution plots for polarizability

Finally, the parity plot of the predictions of refractivity (Ref) shows a high degree of correspondence between the predicted Ref and the actual Ref ($R = 0.931$, $R^2 = 0.852$). The error distribution supports the fact that the majority of the predictions fall below the 5% threshold, but the presence of a number of outliers indicates some potential for improvement in descriptor sensitivity. Overall, the model worked reasonably well for Ref verified from Figure *10*.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

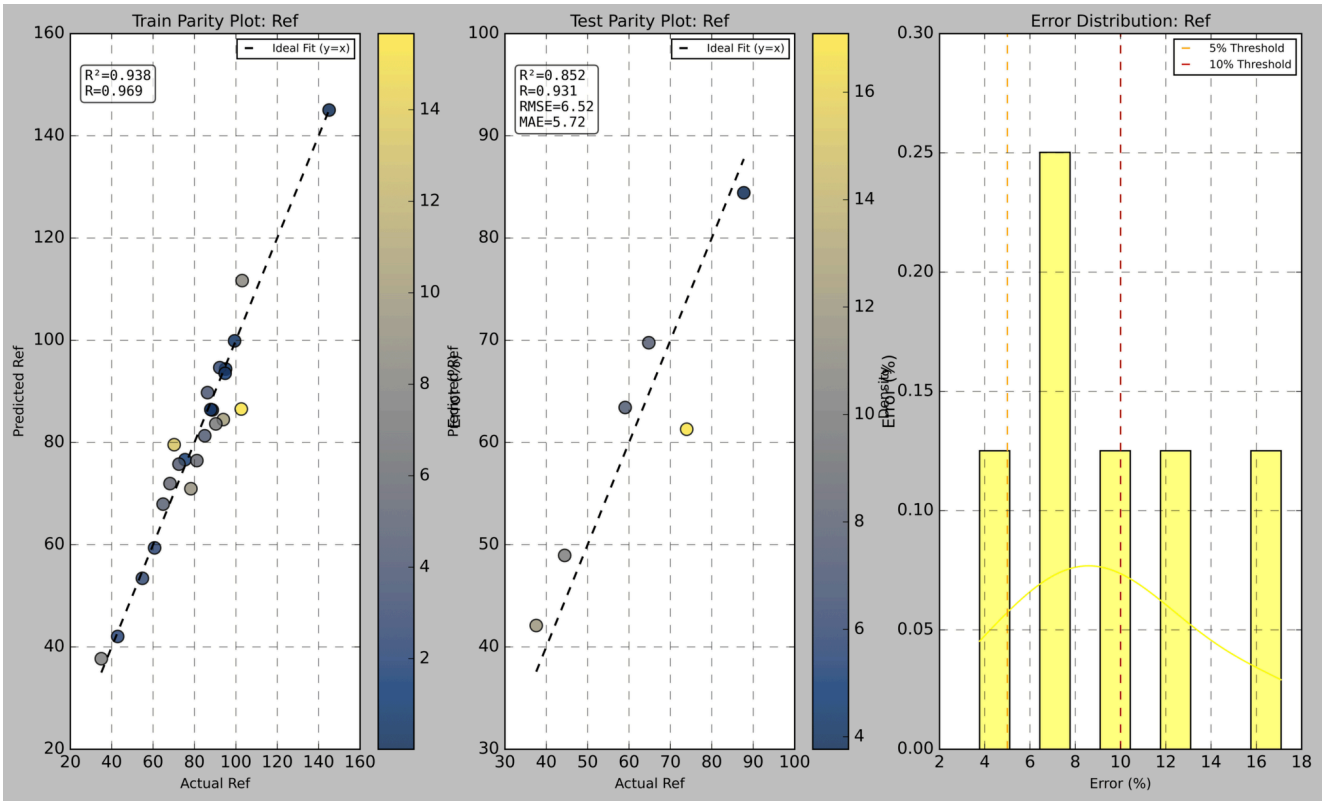


FIGURE 10: Parity and error distribution plots for refractivity

From Figure *11*, the ANN model produced moderate success in predicting PSA, with a lower correlation (R = 0.358, R^2 = -0.167) and higher error metrics (RMSE = 34.96, MAE = 28.88) than observed for MW and Pol. The parity plot revealed a poor fit to the ideal line. The error distribution confirmed a wide spread of error values, with the majority exceeding the 5% threshold. Collectively, these results indicated that descriptors currently available do not effectively capture the structural complexity of PSA, and therefore PSA represents the least reliably predicted property in the dataset.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

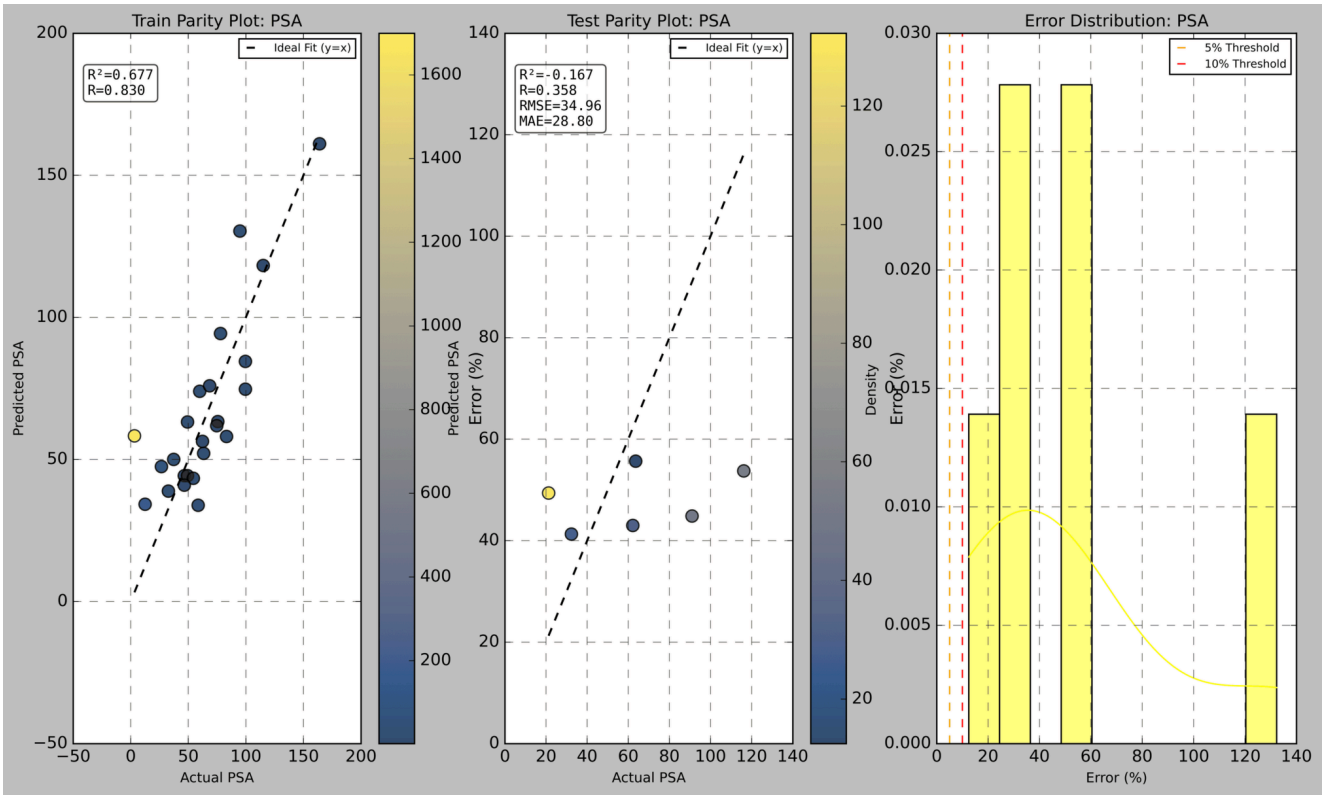


FIGURE 11: Parity and error distribution plots for polar surface area

The strong correlation between the predicted Com attribute and the actual values (R = 0.940, R^2 = 0.675), combined with the high degree of clustering around the line of best fit in the parity plot, illustrates the ability to predict the Com attribute well. The larger errors in the error distribution are likely due to the wider range of numeric values associated with the Com attribute. Also, the slightly larger errors are supported by the larger RMSE and MAE (30.30 and 25.58), but the majority of the predictions were still within the 10% threshold. Overall, this represents good performance from the model as verified from Figure *12*.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

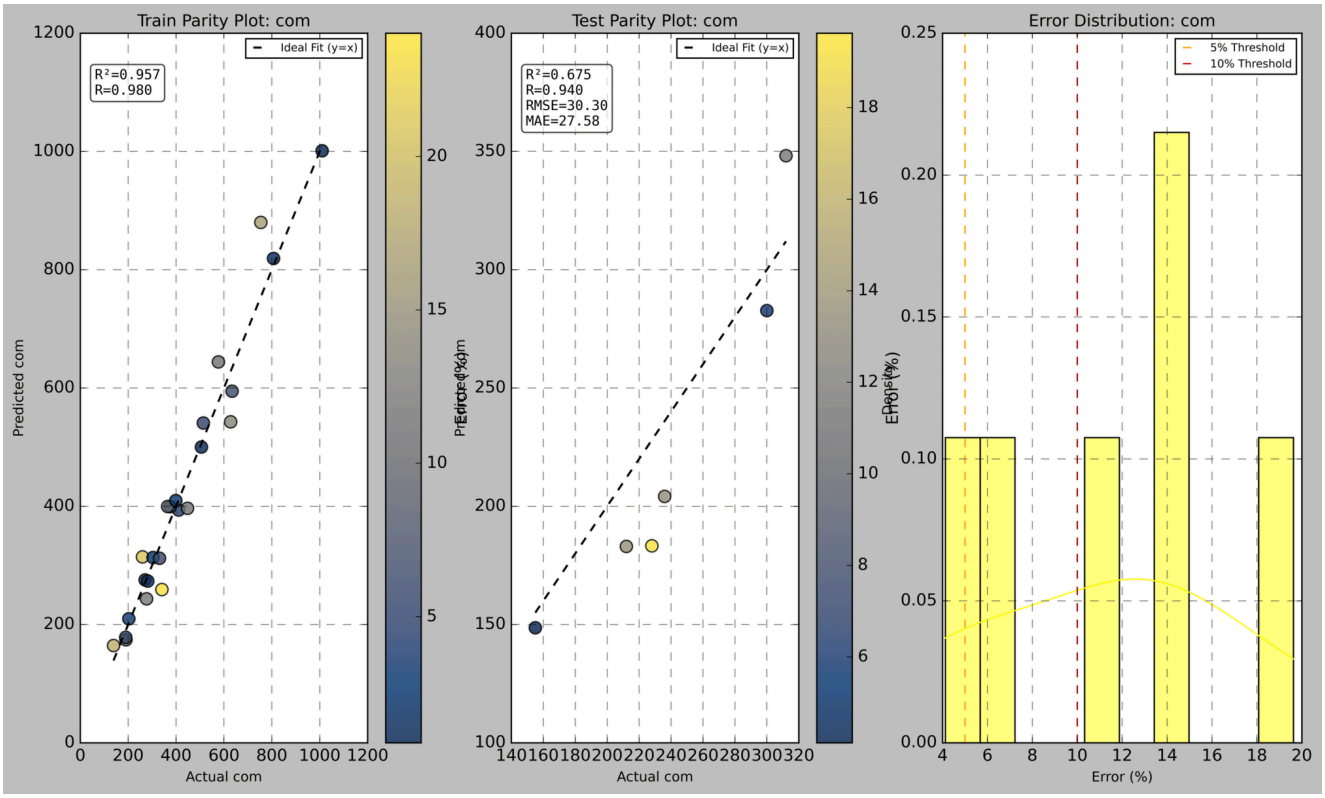


FIGURE 12: Parity and error distribution plots for complexity

The correlation heatmap of actual vs predicted physicochemical properties of drugs used for treating knee disorders is shown in Figure *13*. This correlation heatmap shows how thoroughly the model's predictions align with the actual values. MW, Pol, Ref, and com have very strong correlations, while PSA shows weaker correlation, indicating that the model requires more training.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

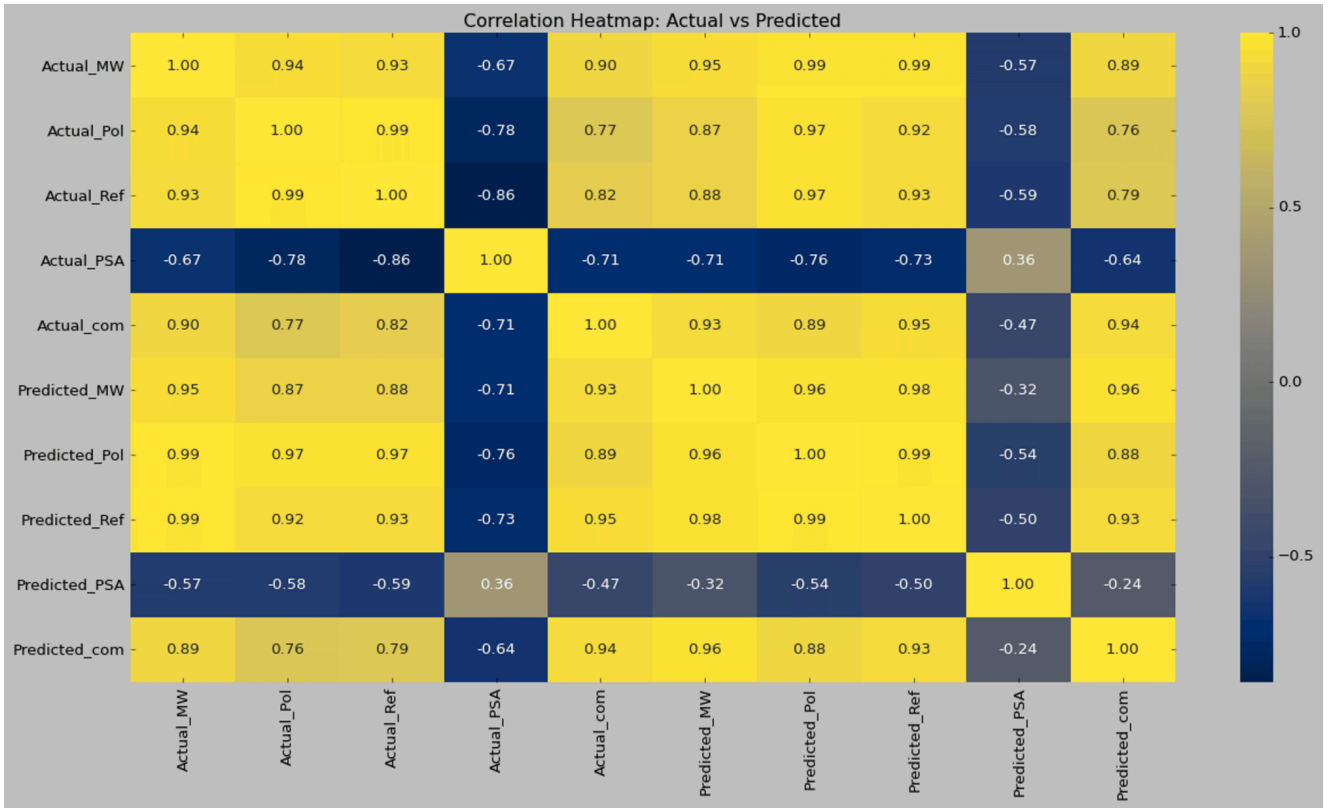


FIGURE 13: Correlation heatmap between actual and predicted physicochemical properties

To further enhance the predictive performance, Random Forest was applied with 5-fold cross validation, with the High correlation coefficients ($R > 0.973$) and high determination coefficients ($R^2 \geq 0.828$) indicated high fitness and little error for the RF model over the vast majority of the characteristics investigated (Table *11*). The best predictor of Pol had the smallest RMSE (1.711) and MAE (1.377); Complexity retained a $R^2 = 0.510$ even with larger error magnitudes, suggesting robustness across a broader value range. Due to its consistent behavior with respect to descriptors, Pol emerged as the most accurately predicted feature. Generally, Random Forest provided reliable predictions. Figures *14*-*18* show the parity, Error distribution and SHAP bar plot for MW, Pol, Ref, PSA, and Com.

Property	R	R^2	RMSE	MAE	OOB
MW	0.969	0.937	10.387	8.415	0.749
Pol	0.973	0.911	1.711	1.377	0.747
Ref	0.969	0.828	7.020	6.249	0.721
PSA	0.692	-0.060	33.312	27.596	0.342
Com	0.855	0.510	37.224	31.354	0.878

TABLE 11: Statistical parameter metric obtained from Random Forest regression

RMSE, Root Mean Square Error; MAE, Mean Absolute Error; OOB, Out of Bag; MW, Molecular Weight; Pol, Polarizability; Ref, Refractivity; PSA, Polar Surface Area; Com, Complexity

The RF model predicts Molecular Weight (MW) reasonably well as given in Figure *14* with $R^2 = 0.937$. The parity plot demonstrates a good overall fit, but there is evidence of some outliers. The error distribution also suggests that while most predictions lie within reasonable limits, there are a few predictions that are beyond acceptable limits. SHAP analysis identified ABC, IB, and GAB as the top contributors to MW, suggesting that these descriptors have the greatest influence on its prediction. Although the model demonstrates good predictive performance, further refinement of the feature space or model parameters may improve its reliability.

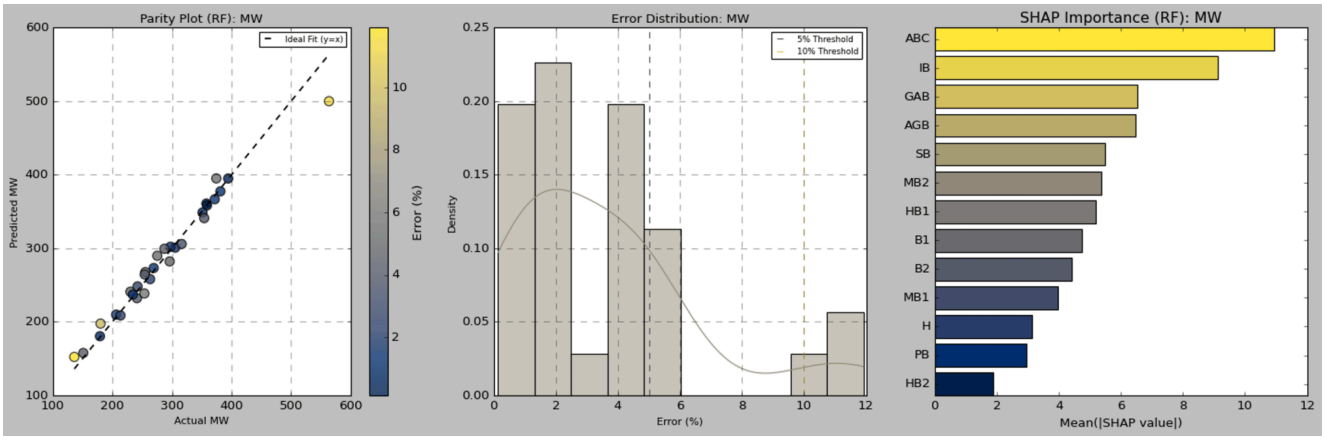


FIGURE 14: Parity, error distribution, and SHAP bar plots for molecular weight

SHAP, SHapley Additive exPlanations

The RF model that predicts Polarizability performed extremely well, with a $R^2 = 0.911$ and $R = 0.973$, which demonstrate almost complete agreement between the predicted and actual values. The error measurements (RMSE = 1.711, MAE = 1.377) confirm very small deviations from the actual values. As seen in the Figure *15* error distribution, virtually all the predictions fell below the 5% threshold except 1. Based on SHAP analysis, the top contributing features to Pol include MB1, MB2 and ABC, which suggest that Pol can be described through descriptors related to bonding and electronic structure. Therefore, the model is typically very reliable and interpretable when predicting Pol.

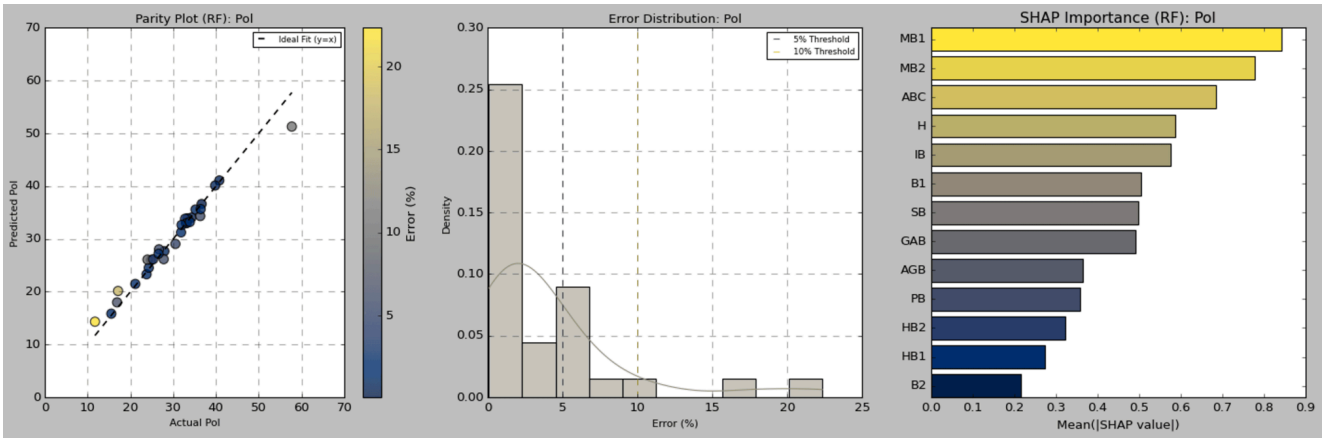


FIGURE 15: Parity, error distribution, and SHAP bar plots for polarizability

SHAP, SHapley Additive exPlanations

The Random Forest model predicting Refractivity shown in Figure *16* demonstrated strong performance, with $R^2 = 0.828$ and very low error measures RMSE = 7.8020, MAE = 6.249. Virtually all of the predictions fell within reasonable error limits (only 2 exceeded 10%). The parity plot demonstrates excellent agreement with the actual values. In addition, the error distribution is very tightly

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

grouped about the mean. Based on SHAP analysis, the features ABC, MB2, and MB1 contributed most to Ref, indicating that descriptors associated with atomic and bonding properties capture Ref well. Overall, the model is reliable and interpretable for Ref prediction, with no obvious outliers.

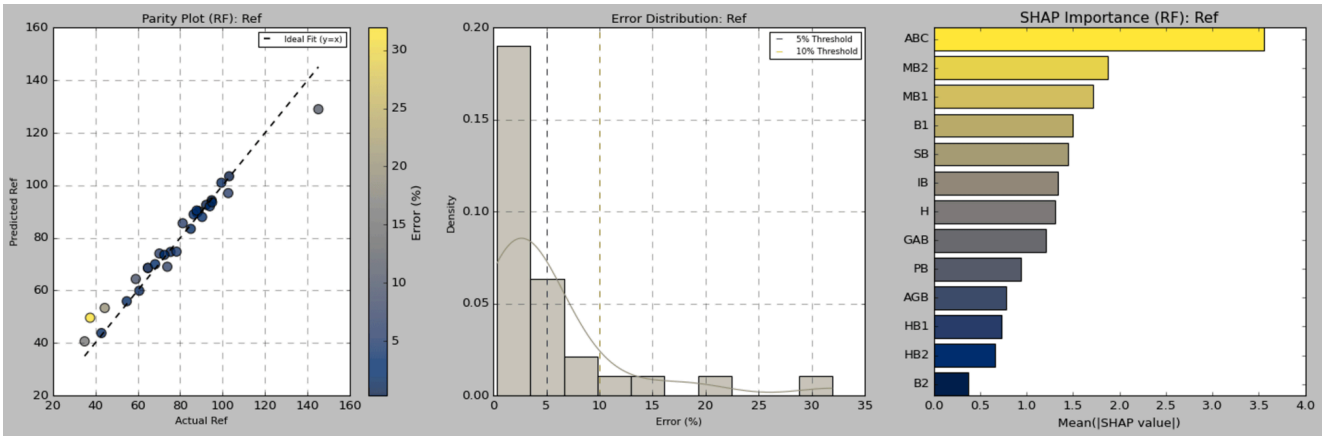


FIGURE 16: Parity, error distribution, and SHAP bar plots for refractivity

SHAP, SHapley Additive exPlanations

The model for forecasting PSA performs moderately bad as indicated by R^2 = -0.060 and large error values (RMSE = 33.312, MAE = 27.596) (as shown in Figure *17*) . The error distribution indicates a broad spread of error values. Based on SHAP analysis, the highest importance features are HB1, B2, and HB2, which suggest that complex structural descriptors affect PSA. Due to PSA's sensitivity to very small chemical differences, it remains the least accurate property forecasted in total.

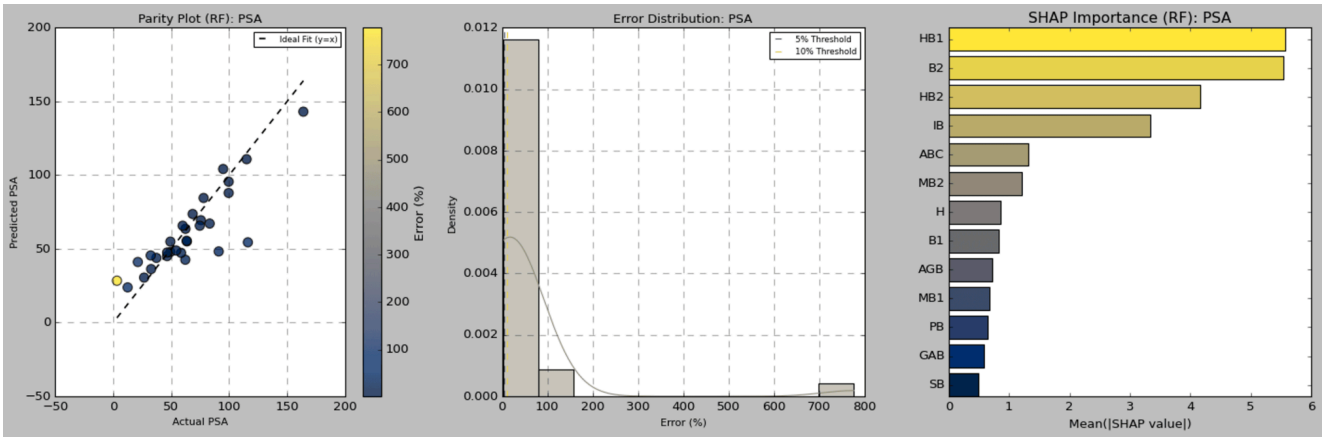


FIGURE 17: Parity, error distribution, and SHAP bar plots for polar surface area

SHAP, SHapley Additive exPlanations

The model for forecasting Complexity is performing well as indicated by R^2 = 0.510 and R = 0.855 which present in Figure *18*. Only 3 out of 30 forecasts fall outside of the 10% error bounds. The error metrics are relatively low (RMSE = 37.224, MAE = 31.354). The error distribution is tightly clustered, and the parity plot shows close agreement between modeled and actual values. As indicated by SHAP analysis, the highest importance features for forecasting Com are HB1, IB, and B2 which suggests that branching and hydrogen bonding descriptors appropriately describe Com given in Figure *18*. The model for forecasting Complexity is typically very reliable and interpretable.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

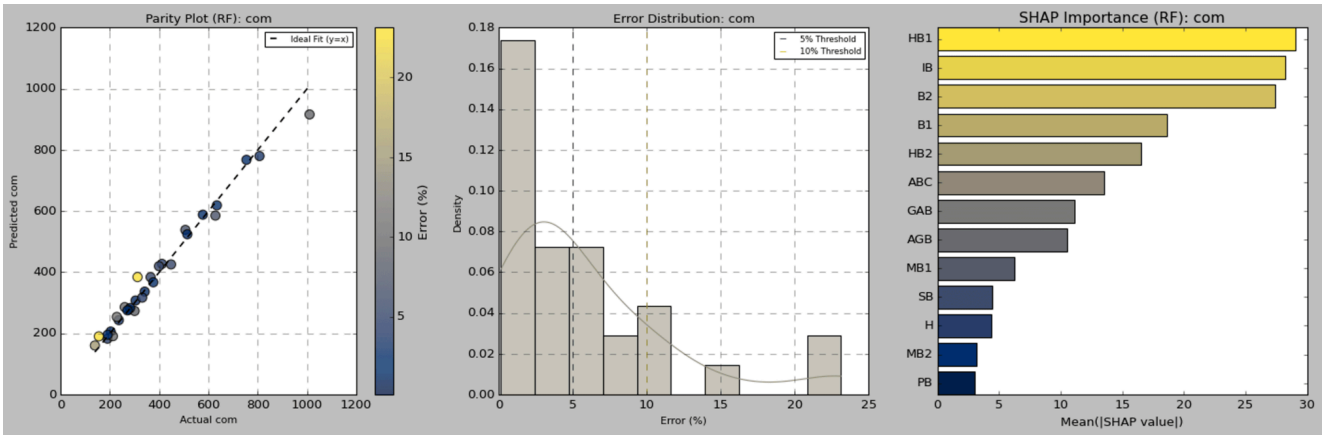


FIGURE 18: Parity, error distribution, and SHAP bar plots for complexity

SHAP, SHapley Additive exPlanations

Historical comparative analysis of linear and nonlinear model performance in knee disorders drug property prediction

In comparison to linear, quadratic, and cubic regression models, the ANN and RF models demonstrated significantly better performance with an RMSE values of 1.182 and an MAE of 0.885, respectively, and narrower, more concentrated error distributions (Table *12*). The ANN model was found to provide the most accurate and consistent predictions across the entire dataset and was therefore determined to be the best model for predicting drug property polarizability.

Models	R^2	RMSE	MAE	MSE
Linear	0.959	1.806	1.370	3.262
Quadratic	0.960	1.773	1.373	3.143
Cubic	0.963	1.707	1.347	2.914
ANN	0.933	1.480	1.181	2.190
RF	0.911	1.711	1.377	2.928

TABLE 12: Statistical parameter metric obtained for traditional regression and machine learning

RMSE, Root Mean Square Error; MAE, Mean Absolute Error; MSE, Mean Squared Error

Figure *19* shows a bar chart comparing the absolute prediction errors of five different models (Linear Regression, Quadratic Regression, Cubic Regression, ANN, and RF) for a number of drugs. Each model's performance is represented by a group of five bars, with each bar representing the performance for each drug. As can be observed from the figure, the ANN and RF models generally provided the smallest error bars for the majority of drugs.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

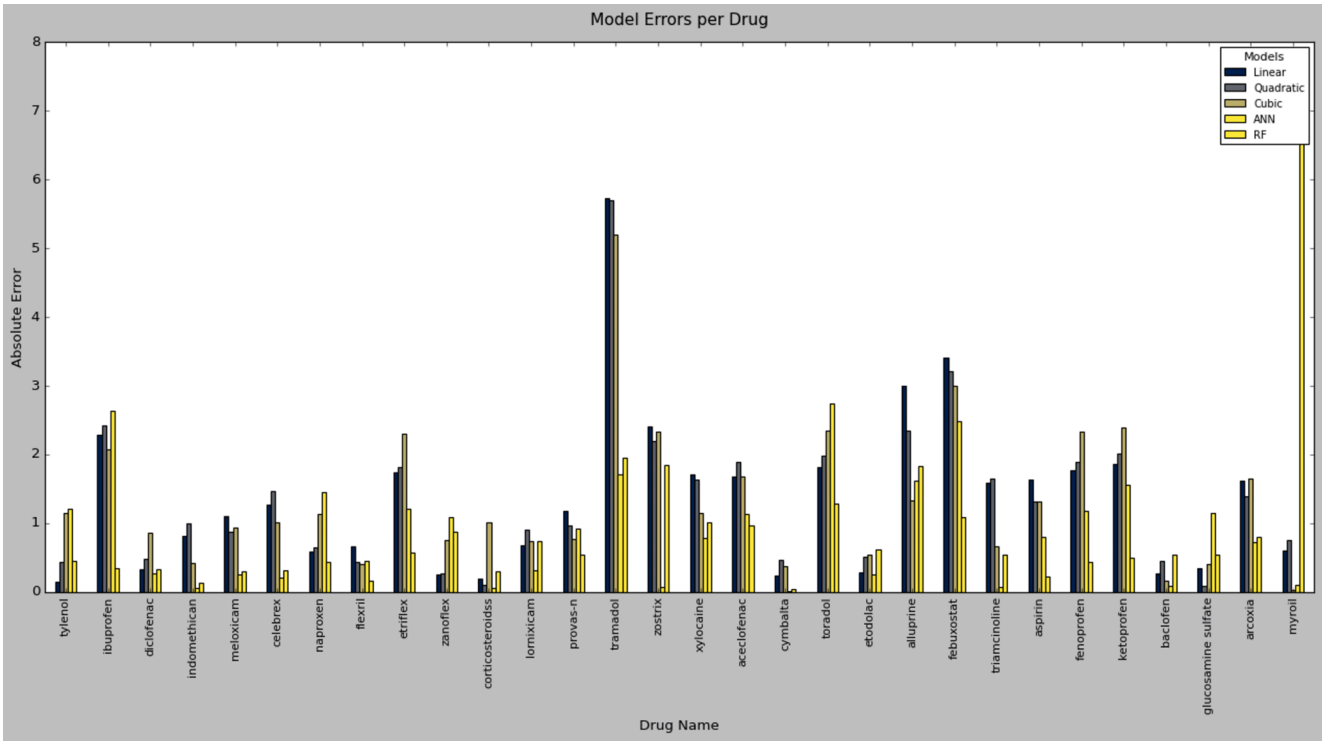


FIGURE 19: Bar chart plot of traditional regression and ML models

ML, Machine Learning

This indicates that these two models provided the most accurate predictions for the majority of drugs. Conversely, the Linear and Polynomial Regression Models often had larger error bars when modeling drugs with complex molecular structures. Overall, the ANN model demonstrated the highest level of predictability for the drug set studied.

Conclusions

This study uses the 13 K-Banhatti indices as useful molecular descriptors and demonstrates the superiority of nonlinear, interpretable ML models, including ANN and RF, for predicting physicochemical properties related to knee disorder drugs. Among the regression models, the cubic regression model utilizing the product Banhatti index was the best-performing classical model for predicting polarizability, with lower values of both RMSE and MAE than those of linear and quadratic regression models. Among the ML models, ANN demonstrated superior performance for polarizability, offering potential applications in drug screening and optimization, while RF achieved stronger results for MW. Overall, ANN was consistently more accurate than the regression and RF models, achieving an RMSE of 1.480, R^2 of 0.933, and MSE of 2.190 in predicting polarizability. Comparative visualizations, including bar charts of predictions made by each model, further support the reliability of ANN predictions.

The findings also indicate that the accuracy of drug property prediction depends on the relationships between the properties and molecular descriptors. High consistency was observed in predicting polarizability, which is highly dependent on molecular structure and its effects on electron binding and affinity. Therefore, these results demonstrate the methodological benefits of combining rare topological descriptors with nonlinear ML techniques for predictive pharmacology. This approach may provide a basis for optimizing molecular descriptors, developing rationally designed drugs, and improving therapeutic strategies for musculoskeletal disorders.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

Concept and design: Hifza .

Drafting of the manuscript: Hifza .

Acquisition, analysis, or interpretation of data: Hafiz Muhammad Atif

Critical review of the manuscript for important intellectual content: Hafiz Muhammad Atif

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following:

- **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work.
- **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work.
- **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. Wilson C: Knee joint anatomy: Structure, function & injuries. Knee Pain Explained. 2025, Accessed: March 25, 2025: https://www.knee-pain-explained.com/knee-joint-anatomy.html.
2. Knee injuries and disorders. MedlinePlus. (2020). Accessed: July 1, 2020: https://medlineplus.gov/kneeinjuriesanddisorders.html.
3. Primorac D, Molnar V, Matišić V, et al.: Comprehensive review of knee osteoarthritis pharmacological treatment and the latest professional societies' guidelines. Pharmaceuticals. 2021, 14:205. 10.3390/ph14030205
4. da Costa BR, Pereira TV, Saadat P, et al.: Effectiveness and safety of non-steroidal anti-inflammatory drugs and opioid treatment for knee and hip osteoarthritis: network meta-analysis. BMJ. 2021, 375:n2321. 10.1136/bmj.n2321
5. Ouyang Y, Dai M: Global, regional, and national burden of knee osteoarthritis: findings from the Global Burden of Disease study 2021 and projections to 2045. Journal of Orthopaedic Surgery and Research. 2025, 20:766. 10.1186/s13018-025-06140-0
6. Li M, Xia Q, Nie Q, Ding L, Huang Z, Jiang Z: Burden of knee osteoarthritis in China and globally: 1990-2045. BMC Musculoskeletal Disorders. 2025, 26:582. 10.1186/s12891-025-08858-8
7. Zhang Y, Luo M, Wu P, Wu S, Lee TY, Bai C: Application of computational biology and artificial intelligence in drug design. International Journal of Molecular Sciences. 2022, 23:13568. 10.3390/ijms232113568
8. Tiwari A, Singh S: Computational approaches in drug designing. Bioinformatics. 2022, 207-217. 10.1016/B978-0-323-89775-4.00010-9
9. Trinajstić N: Chemical Graph Theory. CRC Press, Boca Raton; 1992. 10.1201/9781315139111
10. Klein DJ: Topological Indices and Related Descriptors in QSAR and QSPR Edited by James Devillers & Alexandru T. Balaban. Gordon and Breach Science Publishers: Singapore. 1999. 811 pp. 90-5699-239-2. $198.00. Journal of Chemical Information and Computer Sciences. 2002, 42:1507. 10.1021/ci010441h
11. Hakeem A, Ullah A, Zaman S, Mahmoud EE, Ahmad H, Ali P, Belay MB: Topological modeling and QSPR based prediction of physicochemical properties of bioactive polyphenols. Scientific Reports. 2025, 15:27466. 10.1038/s41598-025-11863-5
12. Ullah A, Jabeen S, Zaman S, Hamraz A, Meherban S: Predictive potential of K-Banhatti and Zagreb type molecular descriptors in structure-property relationship analysis of some novel drug molecules. Journal of the Chinese Chemical Society. 2024, 71:250-276. 10.1002/jccs.202300450
13. Kour S, Sankar R: Machine learning regression models for predicting anti-cancer drug properties: Insights from topological indices in QSPR analysis. Contemporary Mathematics. 2024, 5:6515-6526. 10.37256/cm.5420245826

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6

14. Zhang G, Noureen S, Azam S, Abdalla MEM, Dafaalla ME, Aslam A, Tola KA: Leveraging topological indices and machine learning for advanced prediction of antidepressant drug properties. Scientific Reports. 2026, 16:3537. 10.1038/s41598-025-33532-3
15. Priya KS, Arokia Lancy A: Eccentricity-based topological indices of type II drugs used for diabetes and their QSPR modeling. International Journal of Drug Delivery Technology. 2026, 16:1033-1046. 10.25258/ijddt.16.18s.116
16. Shi X, Kosari S, Ghods M, Kheirkhahan N: Innovative approaches in QSPR modelling using topological indices for the development of cancer treatments. PLOS ONE. 2025, 20:e0317507. 10.1371/journal.pone.0317507
17. Haider I, Li M, Jamil MK, Li A: A graph-based machine learning framework for predicting physicochemical properties of antiviral drugs via topological indices. Journal of Chemical Information and Modeling. 2025, 65:12038-12057. 10.1021/acs.jcim.5c01125
18. Kulli VR: On K Banhatti indices of graphs. Journal of Computer and Mathematical Sciences. 2016, 7:213-218.
19. Kulli VR: On K hyper-Banhatti indices and coindices of graphs. International Research Journal of Pure Algebra. 2016, 6:300-304.
20. Kulli VR, Chaluvaraju B, Boregowda HS: The product connectivity Banhatti index of a graph. Discussiones Mathematicae Graph Theory. 2019, 39:505-517. 10.7151/dmgt.2098
21. Kulli VR: New K Banhatti topological indices. International Journal of Fuzzy Mathematical Archive. 2017, 12:29-37.
22. Kulli VR: Computing Banhatti indices of networks. International Journal of Advances in Mathematics. 2018, 2018:31-40.
23. Kulli VR: ABC Banhatti and augmented Banhatti indices of chemical networks. Journal of Chemistry and Chemical Sciences. 2018, 8:1018-1025.
24. Kulli VR: A new Banhatti geometric-arithmetic index. International Journal of Mathematical Archive. 2017, 8:112-115.
25. Wang ZY, Shi SY, Li SJ, Chen F, Chen H, Lin HZ, Lin JM: Efficacy and safety of duloxetine on osteoarthritis knee pain: a meta-analysis of randomized controlled trials. Pain Medicine. 2015, 16:1373-1385. 10.1111/pme.12800
26. Che X, Song R, Chen B, Dong X: Targeting CXCR1/2: The medicinal potential as cancer immunotherapy agents, antagonists research highlights and challenges ahead. European Journal of Medicinal Chemistry. 2020, 185:111853.
27. Kauffman GW, Jurs PC: QSAR and k-nearest neighbor classification analysis of selective cyclooxygenase-2 inhibitors using topologically based numerical descriptors. Journal of Chemical Information and Computer Sciences. 2001, 41:1553-1560. 10.1021/ci010073h
28. Huang RR, Azam S, Aslam A, Noureen S: Topological indices and QSPR modeling of gonalgia-associated drug molecules via M-polynomials. The European Physical Journal E. 2025, 48:63. 10.1140/epje/s10189-025-00529-9
29. Darlami J, Sharma S: The role of physicochemical and topological parameters in drug design. Frontiers in Drug Discovery. 2024, 4:1424402. 10.3389/fddsv.2024.1424402
30. Alzahrani E, Hanif MF: A machine learning framework for QSPR modeling of drug-like compounds using graph invariants. AIMS Mathematics. 2025, 10:24651-24690. 10.3934/math.20251093

How to cite this article:
. H, Atif (November 14, 2026) A Computational Framework for Drug Property Prediction Using K Banhatti Topological Descriptors and Hybrid Machine Learning Modeling Approaches. Cureus J Comput Sci 3 : es44389-026-00367-6. DOI https://doi.org/10.7759/s44389-026-00367-6