Volume 6, Issue 4, 062524 (1-11)

Fuzzy Optimization and Modeling Journal (FOMJ)

https://doi.org/10.57647/j.fomj.2025.0604.24

DFT-based fuzzy topological indices integrated with machine learning for accurate prediction of topical drug properties

Negar Kheirkhahan, Masoud Ghods*

Department of Applied Mathematics, Semnan University, Semnan, Iran.

***Corresponding author:** mghods@semnan.ac.ir

Original Research

Received:
30 September 2025
Revised:
10 December 2025
Accepted:
25 December 2025
Published in issue:
30 December 2025

Abstract:

In this study, we developed a novel framework to predict the physicochemical properties of topical drugs by integrating fuzzy topological indices with machine learning (ML) models. The chemical structures of the selected drugs were optimized using Gaussian software, and both bond lengths (edges) and vertex properties, corresponding to the atomic masses of each molecule, were fuzzified through Density Functional Theory (DFT). Fuzzy topological indices (FTIs) were then calculated to capture the relationships between molecular geometry, atomic composition, and topological features. Two machine learning algorithms, Linear Regression (LR) and optimized Support Vector Regression (SVR-Tuned), were employed for property prediction. The models were trained on the main dataset and validated on additional test drugs excluded from training, enabling a rigorous assessment of generalization, predictive accuracy, and the absence of overfitting. The results showed that the proposed fuzzy QSPR framework, combined with optimized, achieves high predictive performance, robustness, and strong generalization. This methodology provides an efficient computational tool for estimating molecular properties and can support the rational design of next-generation topical pharmaceutical agents.

Keywords: Machine Learning (ML); Fuzzy QSPR; Topical drugs; Fuzzy Topological Indices (FTIs)

1. Introduction

The concept of a fuzzy subset has been extensively reviewed in recent literature, highlighting core concepts, developments, and applications of membership functions and linguistic variables in modeling uncertainty and imprecision in complex systems [1]. Since then, fuzzy set theory has developed into a broad and active research area, influencing numerous fields such as logical systems, topological and algebraic structures, analytical methods, information sciences, intelligent systems, optimization and decision sciences, neural computing, and strategic planning [2]. As a natural generalization of classical graph theory, the notion of a fuzzy graph was proposed by Rosenfeld [3]. The fundamental components of fuzzy graphs were established through fuzzy analogues of classical graph-theoretic concepts. Bhattacharya [4] made significant contributions to this theory, and later, Mordeson and Peng [5] introduced several operations on fuzzy graphs. The study of complements of fuzzy graphs was carried out by Sunitha and Vijayakumar [6]. Meanwhile, fuzzy set theory and its extensions, including interval-valued fuzzy subsets, have been extensively reviewed in recent literature, highlighting core concepts, developments, and applications of linguistic variables and membership functions in modeling uncertainty [7]. The first formal definition of interval-valued fuzzy graphs was given by Hongmei and Lianhua [8]. Subsequently, Rao et al. [9] introduced novel concepts of interval-valued fuzzy graphs, including level graphs and advanced operations applicable to these structures, providing a more precise modeling framework for uncertainty in complex systems. Strong intuitionistic fuzzy graphs were later investigated by Akram and Davvaz [10]. Bipolar fuzzy graphs were initially formulated by Akram [11]. Later, Akram and Karunambigai [12] defined several important graph-theoretic parameters for bipolar fuzzy graphs, including length, distance, eccentricity, radius, and diameter, and also introduced the concept of self-centered bipolar fuzzy graphs. Furthermore, investigations into isomorphism and self-complementary fuzzy graphs were conducted by Nagoorgani and Malarvizhi [13, 14].

Several studies have applied fuzzy graph indices specifically for modeling chemical and molecular structures. For example, Hayat and Imran [15] calculated degree-based indices for naphthalene nanotubes, while Kalatgian et al. [16] examined several fuzzy graph indices in hydrocarbons. More recently, Mufti et al. [17] computed the first and second fuzzy Zagreb indices for linear hydrocarbons and multi-cyclic benzene derivatives. Several studies have explored fuzzy topological indices in various chemical structures to capture structural uncertainties and enhance QSPR modeling. For instance, Hasani and Ghods [18] investigated FTIs of linear and cyclic anthracene hydrocarbons, demonstrating the applicability of these indices in representing molecular structures and predicting their physicochemical properties. In addition, Islam and Pal [19] developed methods to calculate the second Zagreb index for fuzzy graphs and explored their applications in mathematical chemistry. However, these studies primarily focused on simple hydrocarbons or nanoscale structures, with limited application to complex pharmaceutical molecules. In the context of molecular modeling and QSPR analysis, uncertainty arises mainly from the continuous nature of molecular geometry, the approximations inherent in quantum chemical calculations such as DFT, and the gradual variation of physicochemical contributions of atoms and bonds, rather than from purely random noise. Therefore, this type of uncertainty is better characterized as vagueness or imprecision instead of stochastic variability. While stochastic and probabilistic approaches are effective when sufficient statistical data are available to estimate reliable probability distributions, such conditions are not always satisfied for DFT-optimized molecular descriptors and graph-based representations. Fuzzy set theory provides a flexible and mathematically consistent framework to model gradual transitions and partial memberships in vertex and edge attributes, which naturally aligns with the structural and topological nature of molecular graphs. Consequently, fuzzy modeling offers an interpretable and structurally meaningful approach for incorporating uncertainty into TIs and QSPR modeling, making it more suitable than purely stochastic alternatives for the present study. The main contribution of the present work lies in combining molecular optimization via Gaussian and GaussView with fuzzified vertex and edge characteristics to construct FTIs specifically tailored to topical drugs. This category of compounds has received limited attention within previous fuzzy QSPR studies. Unlike earlier investigations centered on hydrocarbons, nanotubes, or general fuzzy graph structures, our approach integrates DFT-based optimization, fuzzy modeling, and machine-learning-driven prediction of multiple physicochemical descriptors [20]. Additionally, evaluating extra drug candidates outside the training set provides a stringent assessment of the stability and generalization performance of the model, preventing overfitting. Consequently, the proposed framework delivers a precise, efficient, and broadly applicable computational methodology for analyzing and predicting pharmaceutical properties, marking a substantial improvement over existing techniques. To provide a clear overview of prior related studies and to highlight the distinct contributions of the present work, a summary Table 1 is presented. This table compares previous fuzzy graph-based studies in molecular modeling and QSPR analysis with the current study, emphasizing the novelty of integrating DFT-based molecular optimization, fuzzified vertex and edge properties, and machine-learning-driven prediction of multiple physicochemical descriptors.

2. Methodology

In this study, a set of linear and polycyclic topical drugs was selected to calculate FTIs of molecular structures. The chemical structures were first drawn using ChemDraw and then converted into three-dimensional (3D) molecular models. These 3D structures were optimized using Gaussian and GaussView software, employing Density Functional Theory (DFT) to obtain stable molecular configurations. Optimized structural data, including bond lengths and atom types, were extracted for further analysis. Each molecule was represented as a molecular graph, with atoms as vertices and chemical bonds as edges. All computational analyses, including training and evaluation of machine learning models, were performed using Python (version 3.12.7), ensuring reproducibility and consistency of the results.

Table 1. Overview of prior fuzzy graph studies and novel contributions of this work.

Study	Method/indices	Key findings/contributions
Hayat and Imran	Degree-based indices for naphthalene nanotubes	Computed degree-based fuzzy graph indices for nanoscale hydrocarbon structures
Kalatgian et al.	Fuzzy graph indices in hydrocarbons	Evaluated multiple fuzzy graph indices for linear and cyclic hydrocarbons
Mufti et al.	First and second fuzzy Zagreb indices	Calculated fuzzy Zagreb indices for linear hydrocarbons and multi-cyclic benzene derivatives
Islam and Pal	Second Zagreb index for fuzzy graphs	Developed methods for computing second Zagreb index; applied in mathematical chemistry
Present work	DFT-based molecular optimization + Fuzzy topological indices (FTIs) + Machine learning prediction of multiple physicochemical properties	Integrates molecular optimization, fuzzified vertex and edge attributes, and ML-based QSPR prediction specifically for topical drugs; evaluates extra drug candidates outside training set for robust generalization

2.1 Fuzzy numbers Fuzzification of molecular vertices and edges

In this study, the fuzzification of molecular vertices (atomic masses) and edges (bond lengths) was performed using a normalization-based approach. First, the atomic mass of each vertex was divided by the maximum atomic mass in the molecule, producing membership degrees for all vertices in the interval [0,1]. Next, a scaling factor K was calculated as the ratio of the minimum normalized atomic mass to the maximum bond length. Each bond length was then multiplied by K to ensure all edge weights also lie within [0,1]. This procedure allows both vertex and edge attributes to reside within a consistent fuzzy range suitable for computing FTIs. Although this approach does not explicitly define classical trapezoidal fuzzy numbers, it can be interpreted as a simplified or degenerate form of trapezoidal membership functions, in which the core region is narrow and membership degrees vary linearly between 0 and 1. This choice was made because it provides a computationally efficient, interpretable, and consistent framework to represent the gradual uncertainty inherent in DFT-optimized molecular descriptors, such as bond lengths and atomic contributions, without requiring fully specified probability distributions as in stochastic methods. This methodology ensures a reliable, uniform, and practical fuzzy representation of molecular structures, suitable for QSPR modeling and machine-learning-driven property prediction. The fuzzy degree of a vertex was determined by summing the fuzzy weights of all incident edges, while atomic contributions were incorporated according to relative atomic masses in the vertex weights. To account for uncertainties in structural measurements, both vertex degrees and bond lengths were expressed as fuzzy values. FTIs were then calculated to facilitate molecular comparisons and to explore the relationships among molecular geometry, atomic composition, and topological characteristics. To predict the physicochemical properties of the topical drugs, two machine learning approaches were applied: SVR-Tuned and LR [21, 22]. The overall workflow involved extracting molecular descriptors, training the models, optimizing hyperparameters, and evaluating predictive performance.

- SVR-Tuned: This model is an optimized version of the basic SVR, with predictive performance improved through hyperparameter tuning. Grid Search with 5-fold cross-validation was applied to adjust key parameters, including C (regularization parameter), γ (RBF kernel coefficient), and ε(epsilon in the loss function). The ranges of values tested were:
- C:[0.1,1,10,100]
- ε:[0.01,0.1,0.2]
- kernel:[rbf,linear]

The optimal combination selected for this study was $C = 10$, $\varepsilon = 0.1$, and kernel = rbf, as summarized in Table 2. These parameters were chosen to minimize the prediction error on the validation set while ensuring robust generalization to unseen data.

Table 2. Final optimized parameters of the SVR model.

Model	C	γ (gamma)	ε	kernel
SVR-Tuned	10	auto	0.1	rbf

Note: γ was set to 'auto', allowing the algorithm to automatically use $1/n\_features$.

- LR: Used as a reference model. Linear regression is expressed as:

$$P = B + A(\text{FTIs}) \tag{1}$$

where P represents the predicted physicochemical property, B is the intercept, A is the regression coefficient, and FTI the TIs.

To validate the models, they were employed to comprehensively assess their predictive performance. The models were assessed through standard performance indicators such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the R^2 score. Additionally, four primary assessment measures (models 2 to 5) were utilized to examine both the accuracy and efficiency of each model simultaneously.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{3}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|^2 \tag{4}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of samples.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5}$$

where $\bar{y}_i$ represents the mean of the observed values.
The optimal model is characterized by an R^2 value that is very close to 1, indicating a strong capability to explain the variability present in the data. At the same time, lower values of error metrics such as MSE, MAE, and RMSE ideally approaching zero reflect reduced prediction errors and superior model performance. Consequently, a model that combines a high R^2 with minimal error values can be considered the most reliable and accurate for predictive purposes. Figure 1 illustrates a schematic representation of the methodology.

2.2 Topological index computation

In this subsection, we describe the computation of FTIs for the molecular graphs of the studied drugs. Starting from optimized three-dimensional molecular structures, each molecule is represented as a graph with atoms as vertices and chemical bonds as edges. The vertex and edge attributes are fuzzified to account for structural uncertainties, and the fuzzy degrees are used to calculate various fuzzy topological descriptors, such as the Zagreb, Randic, and Harmonic indices. This framework enables a systematic comparison

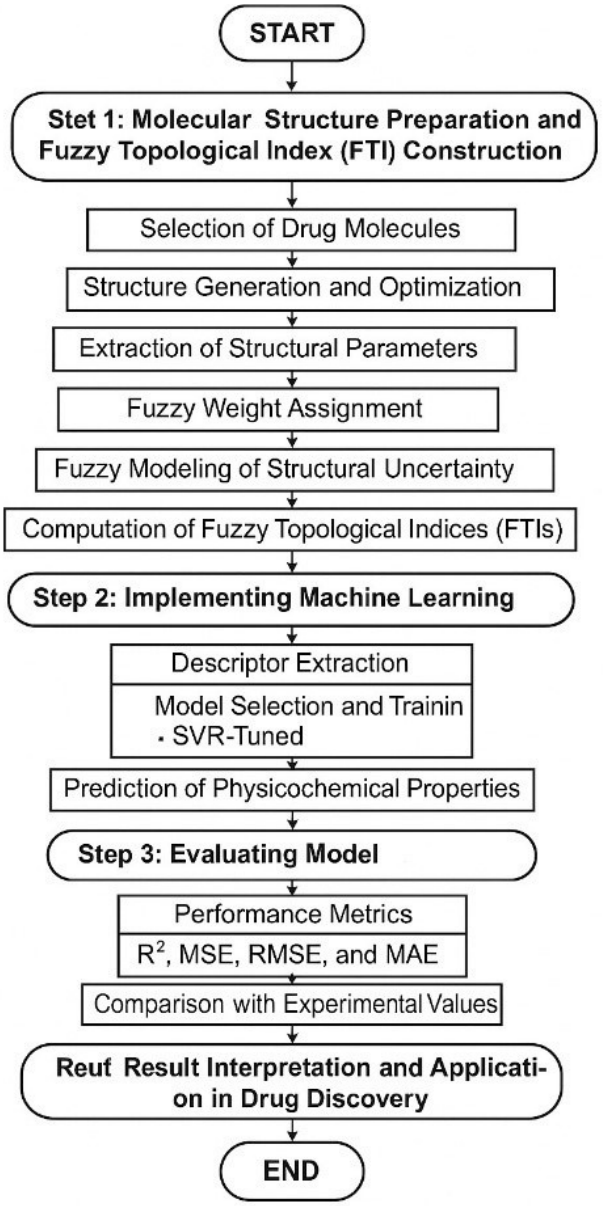

Figure 1. Schematic representation of the proposed workflow used for molecular modeling, fuzzy TIs computation, and ML analysis.

of molecular structures and provides the basis for predicting physicochemical properties.

Definition 1: Given a set X, a fuzzy subset σ of X is a function $\sigma : X \to [0,1]$ [23].

Definition 2: A fuzzy graph $G = (V, \sigma, \mu)$ is defined by a vertex set V and a fuzzy set σ on V, and a fuzzy set μ on E, such that for every edge $xy, \mu(xy) \leq \sigma(x) \Lambda \sigma(y)$.

Khalatian et al. [24] introduced the fuzzy graph indices, as summarized in Table 3.

Let S denote the molecular graph representing Salicylic Acid. In figure 2, the edge membership degrees of this molecular graph were optimized using Gaussian software, providing a refined representation of the molecular structure. As illustrated in figure 3, the initial molecular structure was transformed into a weighted graph, with edge weights further refined through Gaussian-based adjustments. Figure 4 presents the fuzzy molecular graph model, where both vertex membership degrees and edge membership values were fuzzified and optimized using Gaussian functions. In this model, the degree of each vertex is defined as the sum of the fuzzy membership degrees of all edges connected to that vertex, allowing for a more accurate computation of FTIs. Based on this fuzzy graph representation, all bond lengths and vertex membership degrees were fully fuzzified, and the corresponding FTIs were computed, with the results summarized in Table 4. The obtained results are as follows:

$$M(S) = \sum_{i=1}^{n} \sigma(u_i)[d(u_i)]^2 = 46.1694 \quad (6)$$

$$M^*(S) = \frac{1}{2}\left[\sum \sigma(u_i)d(u_i)\sigma(v_j)d(v_j)\right] = 51.5169 \quad (7)$$

$$R(S) = \frac{1}{2}\left[\sum \sigma(u_i)d(u_i)\sigma(v_j)d(v_j)\right]^{-\frac{1}{2}} = 4.6577 \quad (8)$$

$$H(S) = \frac{1}{2}\left[\sum \left[\frac{1}{\sigma(u_i)d(u_i)\sigma(v_j)d(v_j)}\right]\right] = 4.5182 \quad (9)$$

Building on the fuzzy indicators presented in Table 3, and following Definitions 1 and 2 as well as the procedures illustrated in figures 2-4, the FTIs for all drugs under study were computed. The final calculated values are summarized in Table 4.

Table 3. Topological descriptors fuzzy.

The names of FTIs	FTIs
Zagreb index of first in fuzzy graphs	$M(G) = \sum_{i=1}^{n} \sigma(u_l)\,[d(u_l)]^2$
Zagreb index of second kind in fuzzy graphs	$M^*(G) = \frac{1}{2}\left[\sum_{ij \in E(G)} \sigma(u_i)d(u_i)\sigma(v_j)d(v_j)\right]$
Randic index in fuzzy graphs	$R(G) = \frac{1}{2}\left[\sum_{i=1} n\sigma(u_i)d(u_i)\sigma(v_j)d(v_j)\right]^{-1/2}$
Harmonic index in fuzzy graphs	$H(G) = \frac{1}{2}\left[\sum_{ij \in E(G)} \left[\frac{1}{\sigma(u_i)d(u_i)+\sigma(v_j)d(v_j)}\right]\right]$

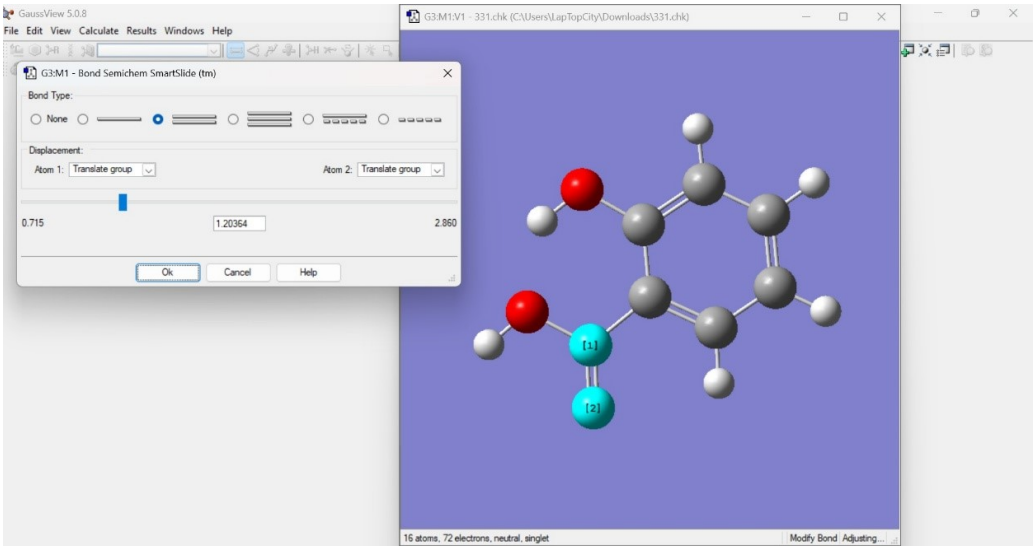

Figure 2. Three-dimensional optimized structure obtained using Gaussian.

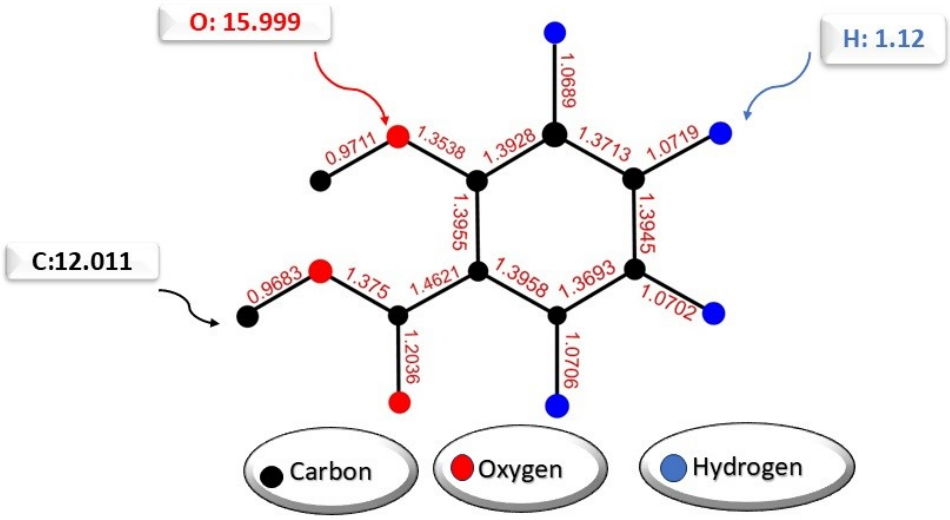

Figure 3. Molecular graph representation of salicylic acid with Gaussian-based edge weight refinement.

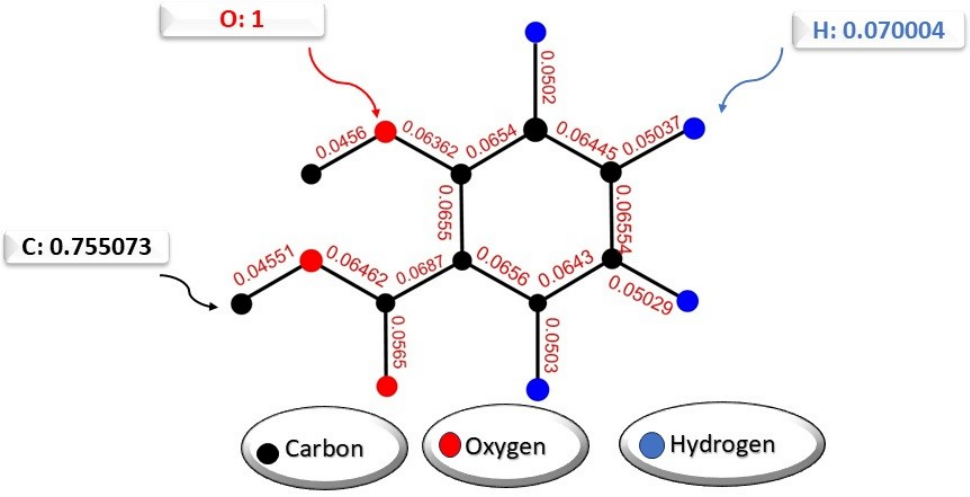

Figure 4. A Gaussian-optimized fuzzy molecular graph model for salicylic acid with joint vertex and edge membership adjustment.

Table 4. Calculated fuzzy metrics for all investigated drugs..

Drugs	$M(G)$	$M^*(G)$	$R(G)$	$H(G)$
Salicylic Acid	46.1694	51.5169	4.6577	4.5182
Hydrocortisone	155.0661	199.2966	12.1862	11.4235
Triamcilone	167.8361	220.9906	12.9178	12.0579
Betamethasone	167.8361	220.2596	12.8926	12.1338
Tazarotene	123.6714	143.1462	11.585	11.2353
Acitertin	111.674	125.1378	11.2617	10.6306
Clobetasol	168.5372	220.1454	12.9519	11.9928
Apremilast	166.6607	199.12	15.0867	14.3132
Methotrexate	168.4534	192.5788	15.5833	14.7927
Calcipotriene	164.3477	197.5796	14.2053	13.6656
Calcitriol	159.9032	187.7958	13.9816	13.1564
Pimecrolimus	287.6756	341.0703	26.4558	25.1042
Tacrolimus	292.3798	344.6282	26.9118	25.6188
Cyclasporin	413.1054	477.5758	39.2824	36.3505

3. Performance assessment and analysis of ML in QSPR research

In this section, the predictive performance of two ML, SVR-Tuned and LR, is examined for three key physicochemical properties, namely Polarizability (PO), Molecular Weight (MW), and Molar Refractivity (MR), across 10 pharmaceutical compounds. The corresponding results are presented in Table 5. All computational predictions were performed using Python and relevant machine learning libraries. The models' performance was quantitatively assessed using the evaluation metrics R^2, MSE, MAE, and RMSE, with the results provided in Tables 6 to 9. Furthermore, figure 5 provides a comparative graphical analysis of the two algorithms based on these performance metrics. Additionally, in Algorithm 1, a comprehensive description of the applied algorithms is provided to enhance reproducibility and support a deeper understanding of the modeling approaches. In these tables, higher R^2 means greater model accuracy, and lower MSE, MAE, and RMSE indicate less error and better performance in predicting drug properties.

3.1 Performance Evaluation of ML Models via Error Profiles and Residual analysis

This subsection provides a detailed analysis of prediction errors and residuals to complement the overall performance metrics. Error and residual analyses were conducted across multiple physicochemical properties to assess model stability, robustness, and predictive reliability. As illustrated in figure 6, the LR model demonstrates more consistent performance, yielding smaller and more compact errors. In contrast, the tuned SVR model shows greater variability in its errors, which may reflect the characteristics of the dataset or indicate that further parameter tuning could

Table 5. Prediction of PO, MW, and MR using two regression models.

Drugs	Actual PO	SVR-Tuned	LR	Actual MW	SVR-Tuned	LR	Actual MR	SVR-Tuned	LR
Salicylic Acid	13.9	14.5337	14.2262	138.12	108.3316	95.10236	35.1	40.5008	34.4419
Hydrocortisone	37.9	37.8813	37.8907	351.5	340.3909	346.4733	95.6	95.5451	95.5771
Triamcilone	38.5	38.5698	38.4967	362.5	362.3000	354.3304	97	97.3146	97.0922
Betamethasone	39.7	39.7243	39.7159	392.5	395.8752	391.9830	100.2	100.1001	100.1457
Tazarotene	40.2	40.2106	40.2010	394.4	394.6289	396.5995	101.4	101.5006	101.4148
Clobetasol	41	40.9900	40.9945	410.9	397.6078	412.2676	103.5	103.5997	103.4748
Methotrexate	47.2	47.1898	47.2005	412.6	429.0141	422.4595	119	118.8996	118.9810
Calcipotriene	48.4	48.3896	48.4007	416.6	417.5649	436.6613	122.2	120.3830	122.2127
Calcitriol	49.3	48.8681	49.1156	454.4	454.1999	455.8843	124.4	122.5596	124.5499
Cyclasporin	130.4	130.4098	130.3997	1202.6	1202.400	1201.4021	328.8	328.6996	328.8013

Table 6. R^2-based assessment of ML models for drug property prediction.

Models	PO	MW	MR
SVR-Tuned	0.9999	0.9978	0.9993
LR	0.9999	0.9964	0.9999

Table 7. MSE-based assessment of ML models for drug property prediction.

Models	PO	MW	MR
SVR-Tuned	0.0594	146.9359	3.6009
LR	0.0140	245.2808	0.04689

Table 8. RMSE-based assessment of ML models for drug property prediction.

Models	PO	MW	MR
SVR-Tuned	0.2438	12.1217	1.8976
LR	0.1186	15.6614	0.2165

Table 9. MAE-based assessment of ML techniques for drug property prediction.

Models	PO	MW	MR
SVR-Tuned	0.1229	7.5772	0.9928
LR	0.0547	9.2901	0.1050

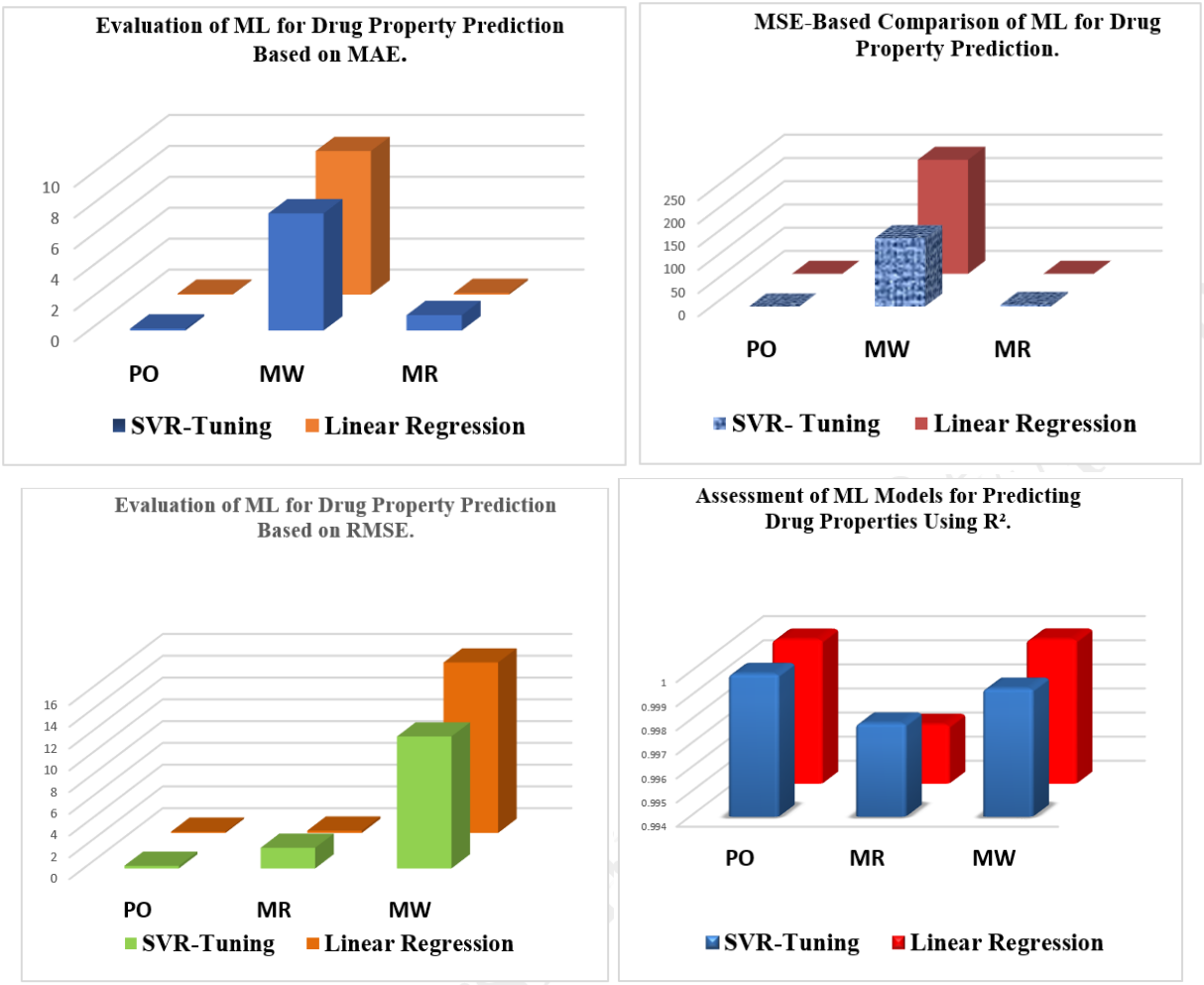

Figure 5. For estimating schizophrenia drug physicochemical profiles using R^2, MSE, MAE, and RMSE.

Algorithm 1. Aggregated kernel graph-cut.

Step 1: Molecular Structure Preparation and Fuzzy Topological Index (FTIs) Construction

1. Selection of Drug Molecules
2. Structure Generation and Optimization
3. Extraction of Structural Parameters
4. Fuzzy Weight Assignment
5. Fuzzy Modeling of Structural Uncertainty
6. Computation of Fuzzy Topological Indices (FTIs)

Step 2: Application of Machine Learning Models

7. Descriptor Extraction
8. Model Selection and Training
   - Linear Regression (LR)
   - (SVR-Tuned)
9. Prediction of Physicochemical Properties
10. Software and Frameworks.

Step 3: Evaluating Model Performance

11. Performance Metrics

For each model, the following metrics were computed:

R^2, MSE, RMSE, MAE.

12. Comparison with Experimental Values
13. Graphical Performance Analysis

Step 4: Preventing Overfitting

14. Training vs. Testing Evaluation

Model performance was examined on both training and testing subsets to detect possible overfitting.

15. Comparative Visualization

Comparative plots were generated to visualize the consistency of predictions between training and test datasets.

Step 5: Analysis and Implications for Drug Discovery

16. Evaluation of Accuracy and Reliability
17. Implications for Drug Development
18. Generalizability for Real-World Applications.

enhance performance, rather than a limitation of the SVR method itself. Figure 7 shows residual plots, indicating that both the SVR-Tuned and LR models perform comparably. Most residuals for both models are close to zero, with only a few larger deviations observed. Overall, neither model displays a clear superiority, suggesting similar stability and error distribution. As seen in figure 8, both models perform almost equally, although LR exhibits slightly smaller errors and reduced dispersion, while the tuned SVR shows somewhat larger and more scattered errors despite achieving comparable predictions at many points.

4. Assessment of algorithm performance using test data

In this section, the initial training utilized data from ten drug types to develop two ML in Python, which were then employed to predict their physicochemical properties. Subsequently, four additional drug samples were included as a test set, and their properties were estimated using the same models. Predictions for three key properties PO, MW,

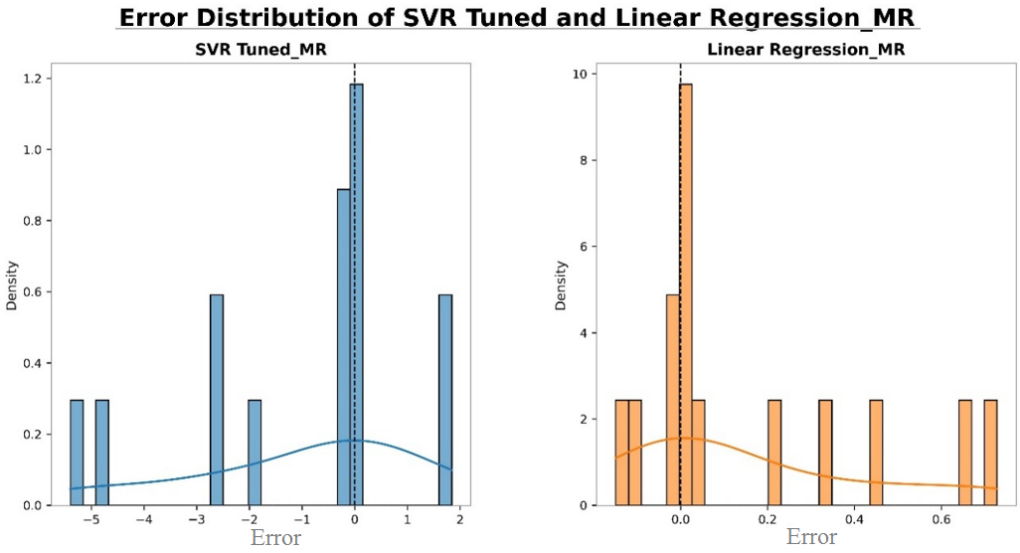

Figure 6. Prediction error profiles of the models in predicting MR illustrated. Histograms and KDE plots depict the variability and accuracy of the predictions, enabling clear comparisons.

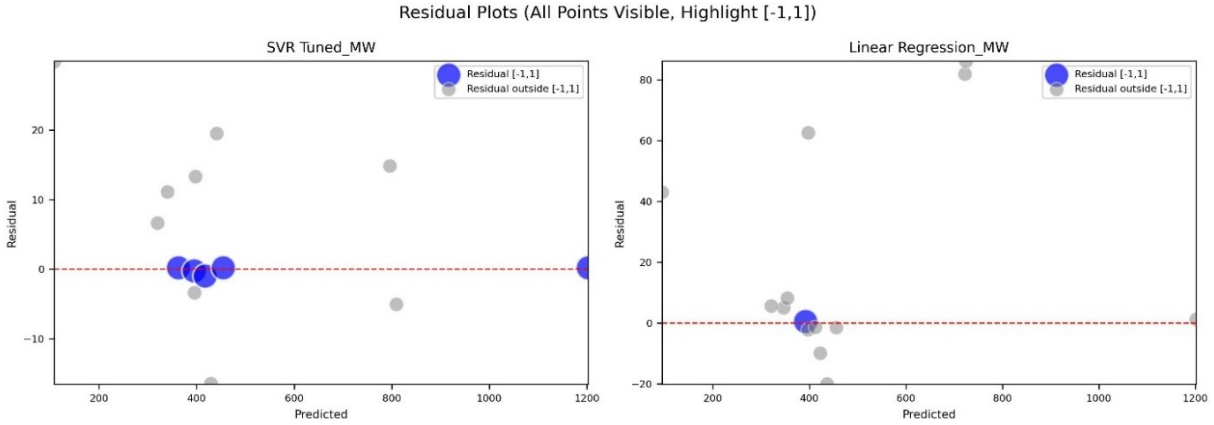

Figure 7. Analysis of residual distributions across models for MW prediction.

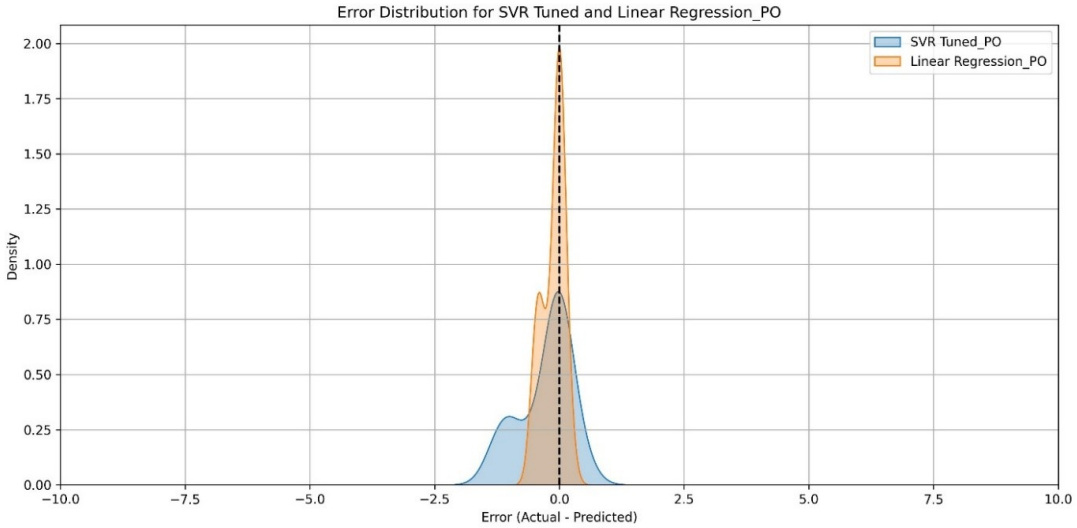

Figure 8. Prediction error profiles of various models in predicting PO.

and MR are summarized in Table 10. Model performance on the test samples is further illustrated in Figures 9 and 10, highlighting predictive accuracy and reliability. The results indicate that the models maintained consistent performance on the test set, comparable to that on the training data, demonstrating high predictive accuracy, strong generalization to unseen samples, and no evidence of overfitting. As illustrated in Figures 9 and 10, the developed models demonstrate high predictive accuracy in estimating actual values. Two different algorithms were employed to evaluate performance; for clarity, only the PO property is presented here, while predictions for other properties followed a similar pattern. The proximity of the data points blue representing the whole dataset and red representing the test set to the ideal line $y = x$ reflects the models' strong predictive capability. The close alignment between training and test results further confirms that the models generalize effectively to unseen samples, indicating excellent transferability. Additionally, the minimal deviations from the ideal line suggest low prediction errors and demonstrate that the models effectively learned the relevant physicochemical properties. The consistent performance across all investigated properties underscores the robustness and reliability of this modeling approach in accurately predicting actual values.

Algorithm 10. Prediction of PO, MW, and MR properties of test data using the algorithms.

Drugs	Actual PO	SVR-Tuned	LR	Actual MW	SVR-Tuned	LR	Actual MR	SVR-Tuned	LR
Acitertin	40.3	40.5645	40.4810	326.4	319.7720	320.7907	101.8	103.8053	101.3444
Apremilast	45.9	46.8617	46.3607	460.5	441.0300	397.8389	115.9	120.6015	115.1719
Pimecrolimus	84.5	85.6234	84.9218	804	809.0119	722.022	213	215.5231	212.7760
Tacrolimus	84.9	86.1123	85.3711	810.4	795.5735	724.2747	214	216.6839	213.6677

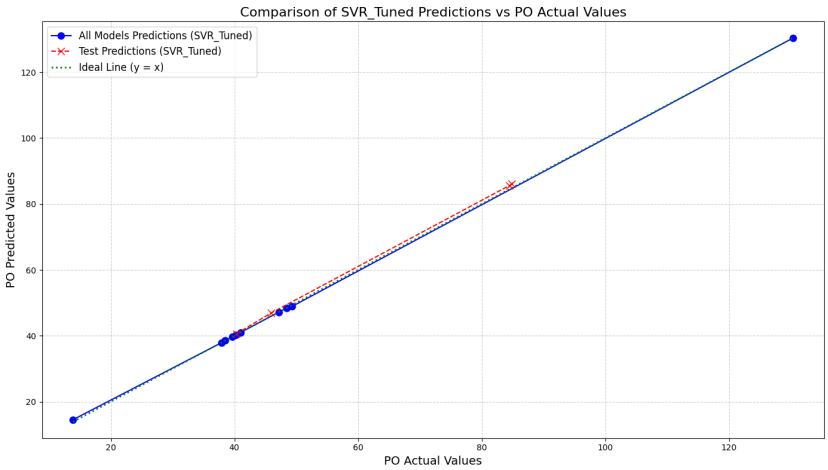

Figure 9. SVR-Tuned model performance compared to actual PO data.

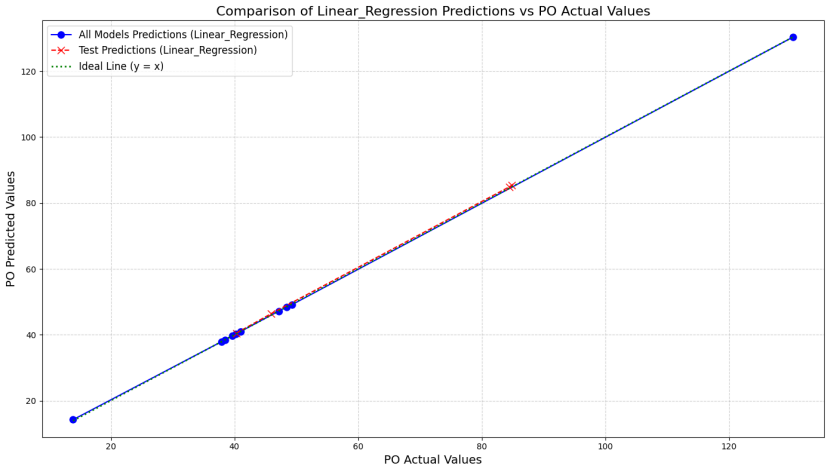

Figure 10. Agreement between SVR-Tuned predictions and observed MR values.

5. Conclusion

In this study, we developed a comprehensive computational framework for predicting the physicochemical properties of pharmaceutical compounds by integrating FTIs with advanced machine learning algorithms. The framework incorporates optimized vertex and edge membership degrees for each compound, calculated using Gaussian software, to effectively capture structural uncertainties inherent in molecular graphs. Two predictive models were implemented: LR and SVR-Tuned. The results demonstrated excellent predictive performance across all evaluated properties. LR provided the highest accuracy for PO and, MR achieving R^2 values of 0.9999 and 0.9999, respectively. In contrast, SVR-Tuned outperformed LR for MW with an R^2 of 0.9979. Error metrics further confirmed these observations: For MW, SVR-Tuned achieved lower MAE (7.58) and RMSE (12.12) than LR, whereas for MR, LR exhibited superior accuracy with MAE $=$ 0.105 and RMSE $=$ 0.217. These findings emphasize that model selection should be tailored to the specific property being predicted, highlighting the complementary strengths of LR and SVR-Tuned Overall, combining optimized fuzzy membership degrees with FTIs provides a robust and interpretable computational representation of molecular structures. This integration, together with machine learning models, allows for precise, property-specific predictions, facilitating rational drug design, reducing experimental costs, and accelerating the drug development process. The proposed framework also demonstrates strong generalization capability, as both models maintained stable performance on external test datasets, indicating robustness to unseen compounds. Furthermore, the methodology offers flexibility for incorporating additional molecular descriptors, exploring alternative fuzzy membership functions, and extending to larger, more chemically diverse datasets, thereby enhancing predictive accuracy and applicability.
For future research, several directions can be considered:

1. Exploring new or alternative FTIs and membership functions to further improve predictive accuracy.
2. Expanding the framework to larger and more chemically diverse datasets to assess its generalization capabilities.
3. Integrating additional molecular descriptors to enhance prediction of complex properties.
4. Investigating the application of advanced or hybrid

machine learning models to optimize property-specific performance.

5. Evaluating the framework for predicting experimental properties of novel compounds, facilitating virtual screening and rational drug design.

In conclusion, the present study not only establishes an effective computational strategy for pharmaceutical property prediction but also provides a scalable, adaptable, and extensible platform for future research in drug design, virtual screening, and QSPR modeling.

- The code for predicting physicochemical properties using the LR and SVR-Tuned algorithms is provided at the following link. https://doi.org/10.6084/m9.figshare.30813773

Authors contributions
All authors contributed equally to the conception, design, execution, and writing of this work. All authors read and approved the final manuscript.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Saatchi. Fuzzy logic concepts, developments and implementation. *Information*, 15(10):656, 2024.

[2] M. N. M. Kumari and R. Chandrasekhar. Isomorphism on interval-valued fuzzy graphs. *IOSR Journal of Mathematics*, 12(1):24–31, 2016.

[3] A. Rosenfeld. Fuzzy graphs. in l. a. zadeh, k. s. fu, and m. shimura, editors, fuzzy sets and their applications. Academic Press, New York: 77–95, 1975.

[4] P. Bhattacharya. Some remarks on fuzzy graphs. *Pattern Recognition Letters*, 6:297–302, 1987.

[5] J. N. Mordeson and C.-S. Peng. Operations on fuzzy graphs. *Information Sciences*, 79(3-4):159–170, 1994.

[6] M. S. Sunitha and A. Vijayakumar. Complement of a fuzzy graph. *Indian Journal of Pure and Applied Mathematics*, 33(9):1451–1464, 2002.

[7] K. Kumari. Fuzzy sets and fuzzy logic: A review of concepts, trends, and applications. *International Journal of Physics and Mathematics*, 7(2):155–161, 2025.

[8] J. Hongmei and W. Lianhua. Interval-valued fuzzy subsemigroups and subgroups associated by interval-valued fuzzy graphs. *IEEE WRI Global Congress on Intelligent Systems*, page 484–487, 2009.

[9] Y. Rao, S. Lei, A. A. Talebi, and M. Mojahedfar. A novel concept of level graph in interval-valued fuzzy graphs with application. *Symmetry*, 15(12), 2023.

[10] M. Akram and B. Davvaz. Strong intuitionistic fuzzy graphs. *Filomat*, 26(1):177–196, 2012.

[11] M. Akram. Bipolar fuzzy graphs. *Information Sciences*, 181(24): 5548–5564, 2011.

[12] M. Akram and M. G. Karunambigai. Metric in bipolar fuzzy graphs. *World Applied Sciences Journal*, 14(12):1920–1927, 2011.

[13] A. Nagoorgani and J. Malarvizhi. Isomorphism properties on strong fuzzy graphs. *International Journal of Algorithms, Computation and Mathematics*, 2(1):39–47, 2009.

[14] A. Nagoor Gani and J. Malarvizhi. Isomorphism on fuzzy graphs. *International Journal of Computer Mathematics Sciences*, 2(4): 200–206, 2008.

[15] S. Hayat and M. Imran. Degree-based indices for naphthalene nanotubes. *Computational and Theoretical Chemistry*, 1072:106–115, 2015.

[16] S. Kalatgian et al. Fuzzy graph indices in hydrocarbons. *Computational and Theoretical Chemistry*, 1175:113–122, 2020.

[17] Z. S. Mufti, E. Fatima, R. Anjum, F. Tchier, M. Saleem, et al. Computing first and second fuzzy zagreb indices of linear and multiacyclic hydrocarbons. *Journal of Function Spaces*, page 6281592, 2022.

[18] M. Hasani and M. Ghods. Investigating fuzzy topological indices of linear and cyclic anthracene hydrocarbon. *Fuzzy Optimization and Modeling Journal (FOMJ)*, 5(4):1–19, 2024.

[19] S. R. Islam and M. Pal. Second zagreb index for fuzzy graphs and its application in mathematical chemistry. *Iranian Journal of Fuzzy Systems*, 20(1):119–136, 2023.

[20] M. Roozbeh, S. Babaie-Kafaki, and M. Manavi. A heuristic algorithm to combat outliers and multicollinearity in regression model analysis. *Iranian Journal of Numerical Analysis and Optimization*, 12(1):173–186, 2022.

[21] M. Roozbeh and M. Maanavi. Mammalian eye gene expression using support vector regression to evaluate a strategy for detecting human eye disease. *Iranian Journal of Health Sciences*, 10:47–58, 2022.

[22] X. Shi, S. Kosari, M. Ghods, and N. Kheirkhahan. Innovative approaches in qspr modelling using topological indices for the development of cancer treatments. *PLoS One*, 20:e0317507, 2025.

[23] A. Verma, S. Mondal, N. De, and A. Pal. Topological properties of bismuth tri-iodide using neighborhood m-polynomial. *International Journal of Mathematics Trends and Technology*, 65(10):83–94, 2019.

[24] S. Kalathian, S. Ramalingam, S. Raman, and N. Srinivasan. Some topological indices in fuzzy graphs. *Journal of Intelligent & Fuzzy Systems*, 39(5):6033–6046, 2020.