scientific reports

Check for updates

OPEN

Predicting tuberculosis drug properties using extended energy based topological indices via a python driven QSPR approach

Kiran Naz[1], Hafiz Muhammad Bilal[2], Muhammad Kamran Siddiqui[1], Sarfraz Ahmad[1] & Mustafa Ahmed Ali[3]✉

In the present work, the physicochemical characteristics of important anti-tuberculosis (TB) drugs such as isoniazid, pyrazinamide, ethambutol, ethionamide, linezolid, and levofloxacin are explored using extended energy-based topological indexes. Based on the molecules of the drugs, we calculate the extended energies of many widely recognized indexes such as Zagreb Second Index, Harmonic Index, Randic Index, Sombor Index, Reduced Sombor Index, and Average Sombor Index. All the calculations are done using Python, and the rigorous algorithmic implementation in the form of matrix formulation and computation of the eigenvalue is also given for reproducibility. We use the linear, quadratic, and logarithmic regression models to predict nine important physicochemical parameters: the boiling point, the melting point, the flash point, the molar refractivity, the polarizability, the molar volume, the molecular weight, the logarithm of the partition coefficient, and the surface area. Among the three models, the quadratic regression always yields the best predictability, as reflected in the largest coefficient of determination (R^2) as well as the minimum root mean square error (*RMSE*) values. Visual analyses such as heatmaps, scatter plot matrices, bar charts, and regression plots are employed to complement the numerical findings. Also, a rigorous discourse about model validity, model significance, and limitations is discussed. The entire source code and dataset are made available through GitHub to allow verification and transparency. The Python-based QSPR methodology, in addition to elucidating the high correlation of the topological descriptors with the properties of drugs, offers a drug design and optimization process in pharmaceutical research in an efficient way.

Keywords Zagreb Index, Harmonic Index, Randic Index, Sombor Indices, Extended energy, Adjacency matrix

The compounds are portrayed in terms of molecular graphs, with atoms represented in terms of vertices and bonds in terms of edges. Several descriptors of structure, represented in terms of topological indices, serve a mechanism for predicting behavior, reactivity, and stability. All these factors contribute to enhancing therapeutic effectiveness[1].

TB is an infectious disease produced by Mycobacterium tuberculosis and continues to be a worldwide medical problem[2]. It is most often a pulmonary disease but can extend to include other organs. Successful treatment for TB consists of a combination of antibiotics such as isoniazid[3], pyrazinamide[4], ethambutol, ethionamide, linezolid, and levofloxacin. All these drugs target a range of phases in infection and depend almost wholly on physicochemical factors for activity. Isoniazid stops reproduction of TB bacilli, and bacterial development is suppressed by pyrazinamide[4]. Ethambutol inhibits the growth of the bacterial cell wall, and ethionamide is used for multidrug-resistant therapy for TB. Linezolid and levofloxacin play a key role in overcoming resistant strains, with levofloxacin being preferred for its enhanced in-vitro activity in overcoming Mycobacterium tuberculosis.

The physicochemical characteristics of drugs play an important role in characterizing behavior, stability, and compatibility in the organism. Boiling point (BP), melting point (MP), flash point (FP), molecular refractivity (MR), polarity (P), molecular volume (MV), molecular weight (MW), log partition coefficient ($\log P$), and surface area (SA) are important in characterizing their pharmacokinetics and pharmacodynamics[5,6]. BP and MP

[1]Department of Mathematics, COMSATS University Islamabad, Lahore Campus, Islamabad, Pakistan. [2]Department of Mathematics, Lahore Garrison University, Lahore, Pakistan. [3]Department of Mathematics, Faculty of Science, Somali National University, Mogadishu, Somalia. ✉email: mustafa@snu.edu.so

have an impact on drugs' solubility and stability, and hence, in formulating and routes of administration. *FP* is important in terms of safety, characterizing flammability of a compound. MR and polarity convey information about molecule-molecule interactions, having an impact on absorption and receptor binding. *MV* and *MW* convey information about a drug's size and transport behavior[7]. Log P conveys information about lipophilicity, predicting membrane crossing behavior of a drug, and *SA* in drug-receptor interaction. All these together convey information about optimized drug design and delivery for therapeutic efficacy[8].

In this study, the physicochemical properties of these TB drugs were analyzed through the extended energies of several topological indices, including the Zagreb second index, Harmonic index, Randic index, Sombor index, reduced Sombor index, and average Sombor index. Linear, quadratic, and logarithmic regression models were applied to investigate the relationship between these indices and the drugs' physicochemical properties. The quadratic regression model emerged as the best fit, showing the highest R_v values and the lowest RMSE values, outperforming the other models. The correlation analysis revealed significant relationships between extended energies of indices and physicochemical descriptors of drugs. Various forms of visualization, such as heatmaps, scatter plot matrices, bar plots, and plots of a regression line, have been adopted in an effort to visualize such relationships in a better form. These findings illustrate that a quadratic model is the most reliable model for predicting physicochemical property of drugs for TB, and it can provide significant information about molecular descriptors of drugs. It can contribute positively in terms of enhancing drug design and optimization and in formulating effective drugs for treating TB.

Topological descriptors mean numerical descriptors representing molecular structure descriptors in terms of its graphical form, derived through its graphical form. In graphical form, atoms have been considered as vertices and bonds have been considered as edges[9]. These indices act as a bridge between molecular property and chemical structure, and useful information regarding reactivity, stability, and bioactivity of a compound can be derived through them. Some of the most prevalent types of topological indices include degree-based, distance-based, and connectivity indices, describing a specific molecular structure feature each one of them. With the use of these indices, one can make an estimation regarding boiling point, melting point, solubility, and toxicity, etc., and these values become an imperative for drug and chemical compound design and optimization. Mostly, degree based topological descriptors[10] are symbolized as:

$$TI\left(\Im\right) = \sum_{\varsigma_i\varsigma_j \in dir0o4Gamma(\Im)} \phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right)$$

Where, $\phi\left(y, z\right)$ is defined as mapping of *z*, *y* with the property $\phi\left(z, y\right) = \phi\left(y, z\right)$ and $\daleth\left(\varsigma\right)$ is the degree of the vertex $\wp$. Some well-known topological indices of these groups are as follows:

- Zagreb second descriptor $\phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right) = \daleth\left(\varsigma_i\right) \times \daleth\left(\varsigma_j\right)$,
- Harmonic descriptor $\phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right) = \frac{2}{\daleth(\varsigma_i)+\daleth\left(\varsigma_j\right)}$
- Randic descriptor $\phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right) = \frac{1}{\sqrt{\daleth(\varsigma*-11-_{i})\times\daleth\left(\varsigma_j\right)}}$
- Sombor descriptor $\phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right) = \sqrt{\daleth(\varsigma_i)^2 + \daleth(\varsigma_j)^2}$,
- Reduced Sombor descriptor $\phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right) = \sqrt{\left(\daleth\left(\varsigma_i\right) - 1\right)^2 + \left(\daleth\left(\varsigma_j\right) - 1\right)^2}$,
- Average Sombor descriptor $\phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right) = \sqrt{\left(\daleth\left(\varsigma_i\right) - \frac{2m}{n}\right)^2 + \left(\daleth\left(\varsigma_j\right) - \frac{2m}{n}\right)^2}$, where *n*, *m* are the total number of nodes and arcs.

A single node is a node with degree 1, it is associated to only one other node. Suppose this single node is symbolized as $\daleth\left(\varsigma_i\right)$ and its neighboring as $\daleth\left(\varsigma_j\right)$. Let $\daleth\left(\varsigma_j\right) = c$, then

$$\sqrt{\daleth(\varsigma_i)^2 + \daleth(\varsigma_j)^2} = \sqrt{(1)^2 + (c)^2}.$$

These mathematical expressions not only provide computational efficiency but also encapsulate fundamental structural features that influence key physicochemical properties[11]. Extended energies derived from indices such as the Zagreb second index, Harmonic index, Randic index, Sombor index, reduced Sombor index, and average Sombor index encode critical information about molecular symmetry, bond connectivity, and atomic distribution. These structural attributes exhibit strong correlations with physicochemical characteristics such as boiling point, melting point, molecular refractivity, polarity, and molecular weight[12]. By analyzing these indices, valuable insights into molecular behavior can be obtained, aiding in the prediction and optimization of drug properties for improved therapeutic applications.

The study emphasizes a set of chosen anti-tuberculosis drugs, the use of which is mandatory in the control and treatment of Mycobacterium tuberculosis. These drugs, including such widely used substances as isoniazid, rifampicin, ethambutol, and pyrazinamide, are of essential importance in first-line anti-TB chemotherapy. Their molecular structures possess diverse chemical characteristics, affecting their physicochemical properties such as the boiling point, entropy, molar refractivity, and lipophilicity. In the work, a Quantitative Structure-Property Relationship (QSPR) model, that makes mathematical relations between the molecular structure of the drugs and their experimentally established properties based on extended energy-based topological indices, is used. QSPR modeling is a widely recognized method in the field of cheminformatics that may render predictions without the requirements of expensive experimental protocols. Utilizing graph-theoretical descriptors such as extended energy, the objective of the work is to study the effect of the structural characteristics of the TB drugs and to assist the rational design and optimization of anti-tuberculosis drugs.

Motivation

The advent of global drug-resistant tuberculosis is a major public health concern, prompting researchers to seek low-cost, yet efficient ways of comprehending and maximizing the physicochemical properties of anti-TB drugs. Conventional experimental methods of drug physicochemical property determination may be costly, time-consuming, and labor-intensive. Such a hurdle necessitates accurate and interpretable computational methods. Graph-theoretical modeling, particularly the utilization of extended energy-based topological indices, offers a potential alternative. Based on Quantitative Structure-Property Relationship (QSPR) models, the analysis of the structural features of molecules in this study proposes a low-cost yet efficient tool of assessing drug properties, a potential catalyst for the discovery of better TB drugs.

Methodology

In this part, we introduce the mathematical expressions of various graph-based descriptors, including the extended energies of indices such as the Zagreb second index, Harmonic index, Randic index, Sombor index, reduced Sombor index, and average Sombor index[13]. These descriptors establish relationships between atomic structure and molecular properties, which are essential for predicting physicochemical characteristics. Several types of matrices have been defined in the literature to represent molecular structures. Among these, the adjacency matrix[14], denoted as *Z*, plays a fundamental role. For a molecular graph $\Im$ with n vertices, the adjacency matrix *Z* is an $\Im$ $n \times n$ matrix, where its entries are defined as follows:

$$a_{i,j} = \begin{cases} 1, & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases} \tag{1}$$

Sarkar et al.[15]explained extended energy matrices for graph structures, by finding correlations with molecular characteristics. The n^{th} order general extended matrix Z_{TI} is symbolized as:

$$\alpha_{i,j} = \begin{cases} \phi\left(\daleth\left(\varsigma_i\right), \daleth\left(\varsigma_j\right)\right), & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases} \tag{2}$$

The extended energy of graph is stated as:

$$\Im_{TI}\left(\Im\right) = \sum_{i=1}^{n} |\chi_i|,$$

where, $\chi_1, \chi_2, \ldots, \chi_n$ are eigenvalues of matrix *Z*. The extended adjacency matrices[15] of the second Zagreb, Harmonic and Randic descriptors are explained as:

$$M_2 = \begin{cases} \daleth\left(\varsigma_i\right) \times \daleth\left(\varsigma_j\right), & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases}$$

$$H = \begin{cases} \frac{2}{\daleth(\varsigma_i) + \daleth\left(\varsigma_j\right)}, & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases}$$

$$R = \begin{cases} \frac{1}{\sqrt{\daleth(\varsigma_i) \times \daleth\left(\varsigma_j\right)}}, & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases}$$

Assume that $\tau_1^{(1)}, \tau_2^{(1)}, \ldots, \tau_n^{(1)}$, $\tau_1^{(2)}, \tau_2^{(2)}, \ldots, \tau_n^{(2)}$ and $\tau_1^{(3)}, \tau_2^{(3)}, \ldots, \tau_n^{(3)}$ are the eigenvalues of second Zagreb, Harmonic and Randic descriptors. The second Zagreb, Harmonic and Randic energies are listed as:

$$EE_{M_2} = \sum_{i=1}^{n} \left|\tau_i^{(1)}\right|,$$

$$EE_H = \sum_{i=1}^{n} \left|\tau_i^{(2)}\right|,$$

$$EE_R = \sum_{i=1}^{n} \left|\tau_i^{(3)}\right|.$$

The Sombor, reduced Sombor and average Sombor descriptors are:

$$SO = \begin{cases} \sqrt{\daleth(\varsigma_i)^2 + \daleth(\varsigma_j)^2}, & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases}$$

$$SO_{red} = \begin{cases} \sqrt{(\daleth(\varsigma_i) - 1)^2 + (\daleth(\varsigma_j) - 1)^2}, & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases}$$

$$SO_{avg} = \begin{cases} \sqrt{(\daleth(\varsigma_i) - \frac{2m}{n})^2 + (\daleth(\varsigma_j) - \frac{2m}{n})^2}, & \text{for } \varsigma_i\varsigma_j \in \Gamma(\Im), \\ 0, & \text{for } \varsigma_i\varsigma_j \notin \Gamma(\Im). \end{cases}$$

Now, assume $\gamma_1^{(1)}, \gamma_2^{(1)}, \ldots, \gamma_n^{(1)}$, $\gamma_1^{(2)}, \gamma_2^{(2)}, \ldots, \gamma_n^{(2)}$ and $\gamma_1^{(3)}, \gamma_2^{(3)}, \ldots, \gamma_n^{(3)}$ are eigenvalues of Sombor descriptors. Then, the Sombor energies[16] are explained as:

$$EE_{SO} = \sum_{i=1}^{n} \left|\gamma_i^{(1)}\right|,$$

$$EE_{SO_{red}} = \sum_{i=1}^{n} \left|\gamma_i^{(2)}\right|,$$

$$EE_{SO_{avg}} = \sum_{i=1}^{n} \left|\gamma_i^{(3)}\right|.$$

The mathematical descriptors and definitions in this section present a consistent scheme for molecular property quantitation via graph-based indices[17]. With widespread application of energy matrices and eigenvalue calculation, such indices expose molecular connectivity and structure variation in a deeper level. By combining such descriptors, a complete analysis of molecular characteristics can be conducted, with an improvement in physicochemical property analysis of them[18]. Chemical graph theory practice, such an activity, is a significant contribution in predictive modeling in a variety of industries, including in chemistry, pharmacy, and materials science[19].

TB drug molecular descriptors were computed with RDKit, a widely used open-source cheminformatics package. PubChem-derived molecule structures were used for computation of the descriptors. Linear, quadratic, and logarithmic regressions for statistical modeling were conducted with Python and Scikit-Learn[20]. Standard R-squared (R^2) and Root Mean Squared Error (RMSE) measures were used for training and model evaluation for finding the best-fit model. Preprocessing, visualization, and correlation analysis were achieved with Pandas, NumPy, Matplotlib, and Seaborn. For reproducibility, all code and data have been released publicly on GitHub and archived with a DOI on Zenodo. Instructions for data access and repository links are provided in the 'Code Availability'. Energy-based topological indices have been meticulously investigated because of their interest in the analysis of molecular structure, as well as in predicting their properties. Graph energy based on the eigenvalues of the adjacency matrix was first conceptualized by Gutman, and that has been the cornerstone of energy-based indices[21]. A number of the extended versions of energy, such as Laplacian energy, Seidel energy, and Randi? energy, have subsequently been investigated for their predictability. Researchers in the form of Ili? and Stevanovi?[22], Das and Gutman[23], and Cavers et al.[24] particularly contributed toward the establishment and generalization of the indices. More recently, contributions by Chellali et al.[25] and Dehmer et al.[26] illustrate further the aptitude of spectral descriptors in the task of QSPR and QSAR model-building. These studies form the base of the research that is conducted in the current work using the Python-based approach by applying extended energy-based descriptors to the molecules of Tuberculosis drugs.

Dataset selection and justification

This data set consists of six FDA-approved tuberculosis (TB) medicines selected from PubChem based on their well-documented pharmacological relevance and previous experience with quantitative structure-property relationship (QSPR) studies. The data set has previously been employed in[29], where it performed well for predictive modeling. The drugs selected here represent structural and physicochemical variability relevant to TB drug design, allowing for meaningful inference regarding their behavior at a molecular level.

While a larger data set would make for greater generalizability, one should bear in mind that what is most important for this research is correlation with drug properties via extended energy-based topological indices. Expanding the data set would mean additional experimental validation, which is beyond what this theoretical research can accommodate. Similar numbers of samples have been used for previous QSPR studies, which is a testament that a small data set can provide valid data if paired with rigorous statistical validation.

To establish our models' reliability, internal validation tools, including adjusted R-squared values and root mean square error (RMSE), were employed. These are effective measures for model predictability and accuracy. Although external validation on a second data set would further substantiate our data, currently, they are restricted due to a lack of TB drugs with experimentally validated physicochemical properties. However, trends from our research are consistent with published data, further confirming our methodology.

Main results and analysis for tuberculosis treatment drugs

In this section, we present a detail Table 1 representing extended energies of a variety of topological indices, including Zagreb second index, Harmonic index, Randic index, Sombor index, reduced Sombor index, and average Sombor index. All these indices serve as primitive descriptors, and a quantitative relation between

Drugs	EE_{M_2}	EE_{SO}	$EE_{SO_{red}}$	$EE_{SO_{avg}}$	EE_R	EE_H
Isoniazid	135.8332	67.5330	46.1748	19.4896	7.8364	7.5660
Pyrazinamide	122.6104	58.9596	40.6268	14.1648	39.6860	4.7826
Ethambutol	94.5792	66.0588	37.6668	21.6980	13.3054	12.7012
Ethionamide	138.26	67.6754	46.4028	22.4764	6.4488	6.0940
Linezolid	234.6297	124.8180	82.2286	77.7041	14.7616	12.3885
Levofloxacin	307.879	147.8148	109.7384	49.0022	13.9188	13.2632

Table 1. Computed extended energies of topological indices for tuberculosis treatment drugs.

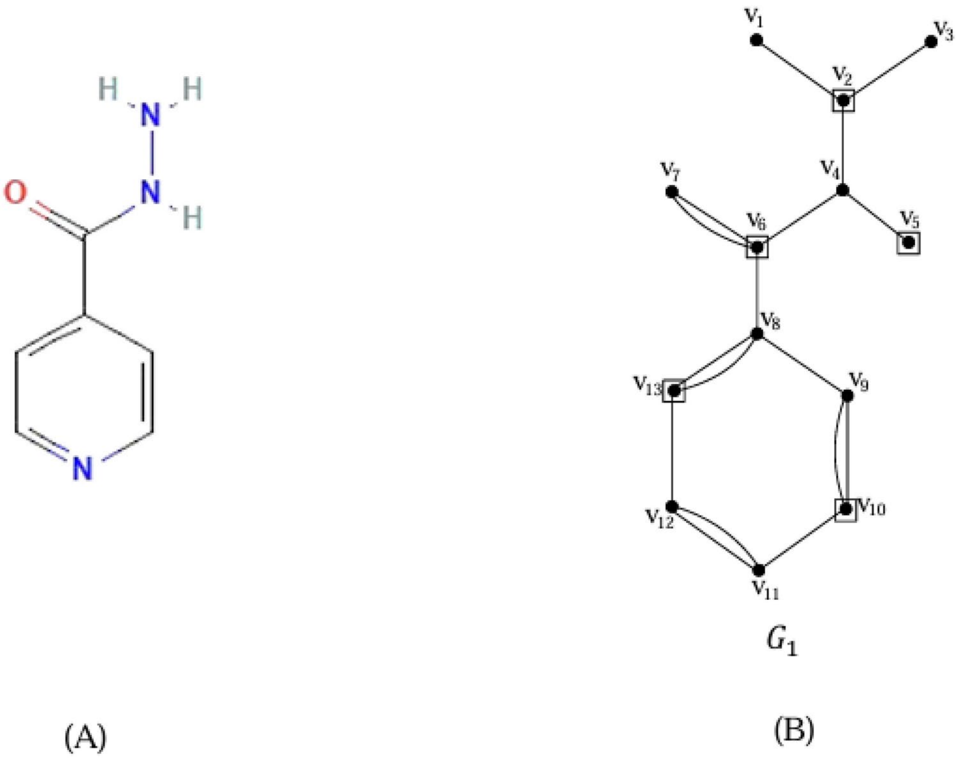


Fig. 1. (**A**) Chemical structure of isoniazid (**B**) Chemical graph of isoniazid.

molecular structure and physicochemical property is derived through them. By comparing these values, one can understand in a deeper manner the structural feature of TB drugs and its role in altering thermodynamic property. In the below-presented table, a detail depiction of these calculated indices is represented, and a deeper analysis of molecular behavior prediction can be performed through them. The molecular structures of the selected anti-tuberculosis drugs isoniazid, pyrazinamide, ethambutol, ethionamide, linezolid, and levofloxacin are illustrated in Figs. 1, 2, 3, 4, 5, and in Fig. 6. These structures were sketched using ChemSketch and served as the basis for calculating the extended energy-based topological indices used in this study.

To further explore relations between extended energies of extended topological indices, a scatter plot matrix is represented. In a pairwise analysis, extended energies for drugs for treating tuberculosis can be represented in a visualization, and through it, one can reveal concealed trends and relations between them. Examining such a scatter plot, such as in Fig. 7, one can reveal trends in molecular structure variation and its effect, possibly, on physicochemical property values. With such a graphical visualization, one can gain a deeper understanding of how extended topological indices act together and contribute towards characterizing drugs for treating TB. For example, in the case of Isoniazid, the eigenvalues are calculated using the extended matrix in MATLAB. The values are 26.7296, 16.0189, 10.4357, 9.0000, 4.6809, 1.0515, 0.0000, 1.0515, 4.6809, 9.0000, 10.4357, 16.0189, and 26.7296, with the sum of these eigenvalues being 135.8332. The extended energies for the remaining cases can be calculated on the same pattern.

Table 2 presents six drugs for treating TB, i.e., isoniazid, pyrazinamide, ethambutol, ethionamide, linezolid, and levofloxacin, and its physicochemical characters including boiling point (*BP*), melting point (*MP*), flash point (*FP*), molecular refractivity (*MR*), polarity (*P*), molecular volume (*MV*), molecular weight (*MW*), log partition coefficient ($\log P$), and surface area (*SA*). All such mentioned characteristics have a significant role in describing molecular behavior and character of drugs. All such factors impact solubility, bio-availability,

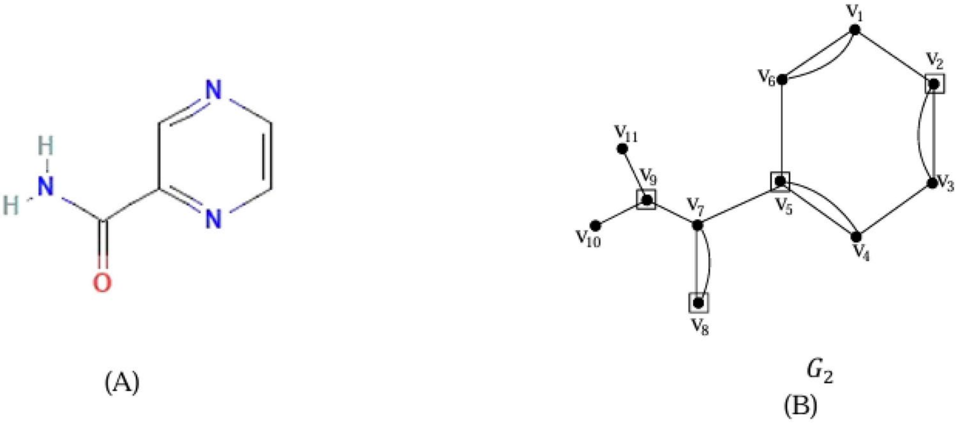


Fig. 2. (**A**) Chemical structure of pyrazinamide (**B**) Chemical graph of pyrazinamide.

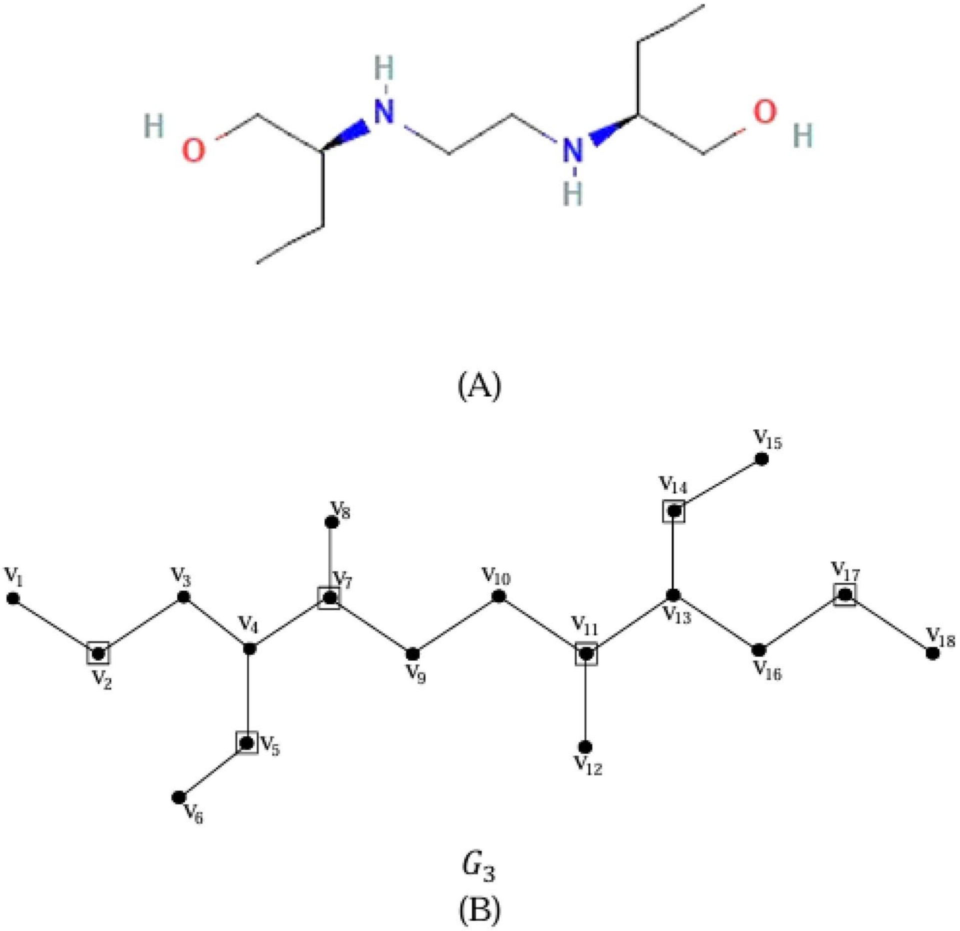


Fig. 3. (**A**) Chemical structure of ethambutol (**B**) Chemical graph of ethambutol.

and compatibility with biological processes of drugs. Comparison with other drugs is significant in providing information regarding drugs' character and efficacy in treating tuberculosis.

The box plot of physicochemical characteristics of drugs for treating TB in Fig. 8 is a graphical representation of distribution and variation in significant molecular descriptors, including boiling point (*BP*), melting point (*MP*), flash point (*FP*), molecular refractivity (*MR*), polarity (*P*), molecular volume (*MV*), molecular weight (*MW*), log partition coefficient $(\log P)$, and surface area (*SA*). In each plot, a range of interquartile range is represented in a form of a central box, depicting 50% of the data, and a dash in form of a horizontal line in a box representing value of a median. Horizontal lines extending outwards denote minimum and maximum values in

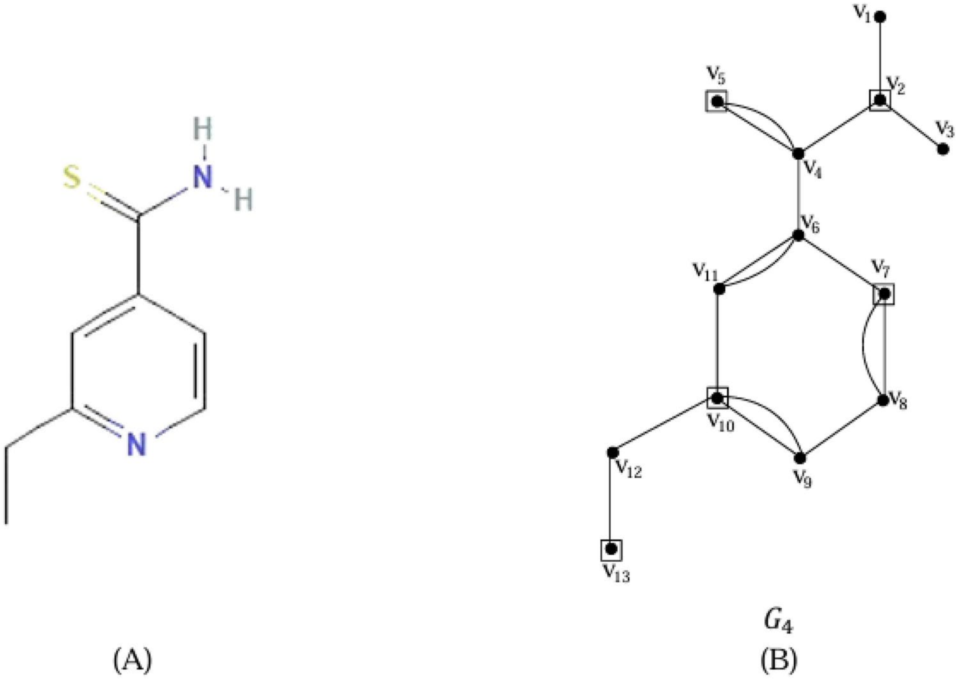


Fig. 4. (**A**) Chemical structure of ethionamide (**B**) Chemical graph of ethionamide.

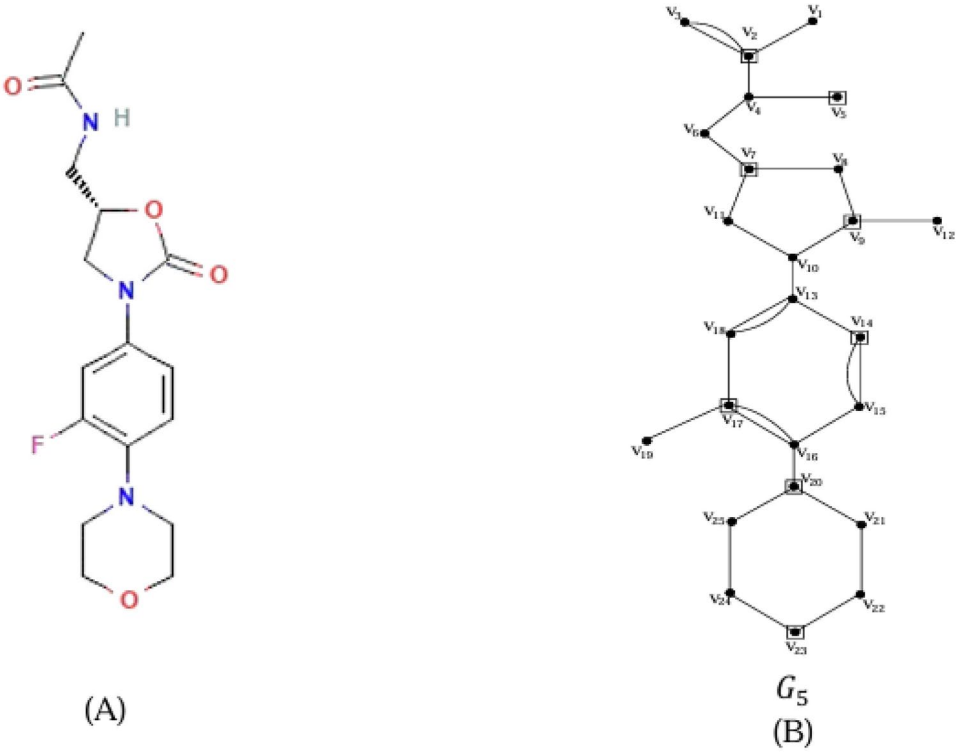


Fig. 5. (**A**) Chemical structure of linezolid (**B**) Chemical graph of linezolid.

an acceptable range, and any out of range values and regarded outliers have been represented in a different form. By offering a graphical view, such a plot aids in comparative analysis of physicochemical property of drugs for treating tuberculosis and brings out variation, trends, and possibly relations between such traits. The presence of outliers in certain properties indicates significant deviations in specific drugs, which may influence their pharmacokinetic behavior and therapeutic effectiveness.

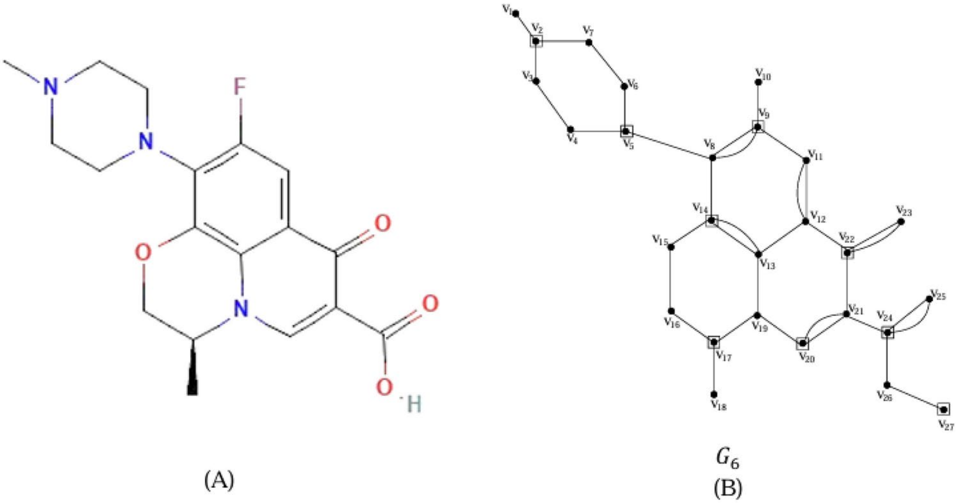


Fig. 6. (**A**) Chemical structure of levofloxacin (**B**) Chemical graph of levofloxacin.

Significance of physicochemical properties in tuberculosis drug analysis

In the following part, we examined a dataset consisting of various tuberculosis (TB) treatment drugs to investigate the relationships between their physicochemical properties. These properties include *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA*. The boiling point and melting point are measured in degrees Celsius, while the flash point is expressed in degrees Fahrenheit. Molecular refractivity and polarity are dimensionless, molecular volume is in cubic angstroms, molecular weight is in atomic mass units, and the log partition coefficient is also dimensionless. To analyze these properties, we applied three statistical models: linear, quadratic, and logarithmic regression. Linear regression predicts the value of a dependent variable based on an independent variable using a straight-line relationship. Quadratic regression builds on this by adding a squared term, which captures nonlinear trends in the data. Logarithmic regression models relationships where the rate of change of the dependent variable decreases as the independent variable increases, making it particularly useful for datasets with diminishing returns. These models were used to identify trends, correlations, and predictive relationships among the physicochemical properties of TB drugs, offering valuable insights into their pharmacokinetic behavior and potential therapeutic effectiveness. These models[27] are defined as:

$$\begin{aligned} Y &= a + bx, \\ Y &= a + bX + bX^2 \\ Y &= a + b\ln(X) \end{aligned}$$

In this study, X represents the independent variable, while Y denotes the dependent variable. We analyzed the physicochemical properties of TB treatment drugs, including *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA*, to develop predictive models. Using the least squares fitting procedure, we constructed regression models incorporating linear, quadratic, and logarithmic approaches to examine correlations and trends among these properties.

In our analysis, we employed R_v to measure the strength and direction of relationships between variables, while ζ_e was used as the standard error of estimation to assess the accuracy of predictions. The *F*-value determined the overall significance of the regression model, and ∇ represented the significance of *F*, indicating the reliability of the model in explaining variations in the data. For the physicochemical property values of drugs for TB, having a single predictive model with a basis in statistical regression analysis will make computation efficient and consistent and will capture inter-dependencies between such property values. In case performance discrepancies are high, or in case a property value shows high dependencies for a specific model, several such models can then be considered. For such scenarios, a statistical validation will have to be performed for increased predictive accuracy and confidence.

Linear regression models for physicochemical characteristics of TB treatment drugs using $\boldsymbol{EE_{M_2}}$

In this section, we identified the models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_{M_2}.

$$\begin{aligned} BP &= 47.113 + 1.831 \times EE_{M_2}, \\ R_v &= 0.849, \qquad \zeta_e = 104.091, \qquad F = 10.292, \qquad \nabla = 0.033, \\ MP &= 163.397 + 0.148 \times EE_{M_2}, \\ R_v &= 0.536, \qquad \zeta_e = 21.260, \qquad F = 1.612, \qquad \nabla = 0.273, \end{aligned}$$

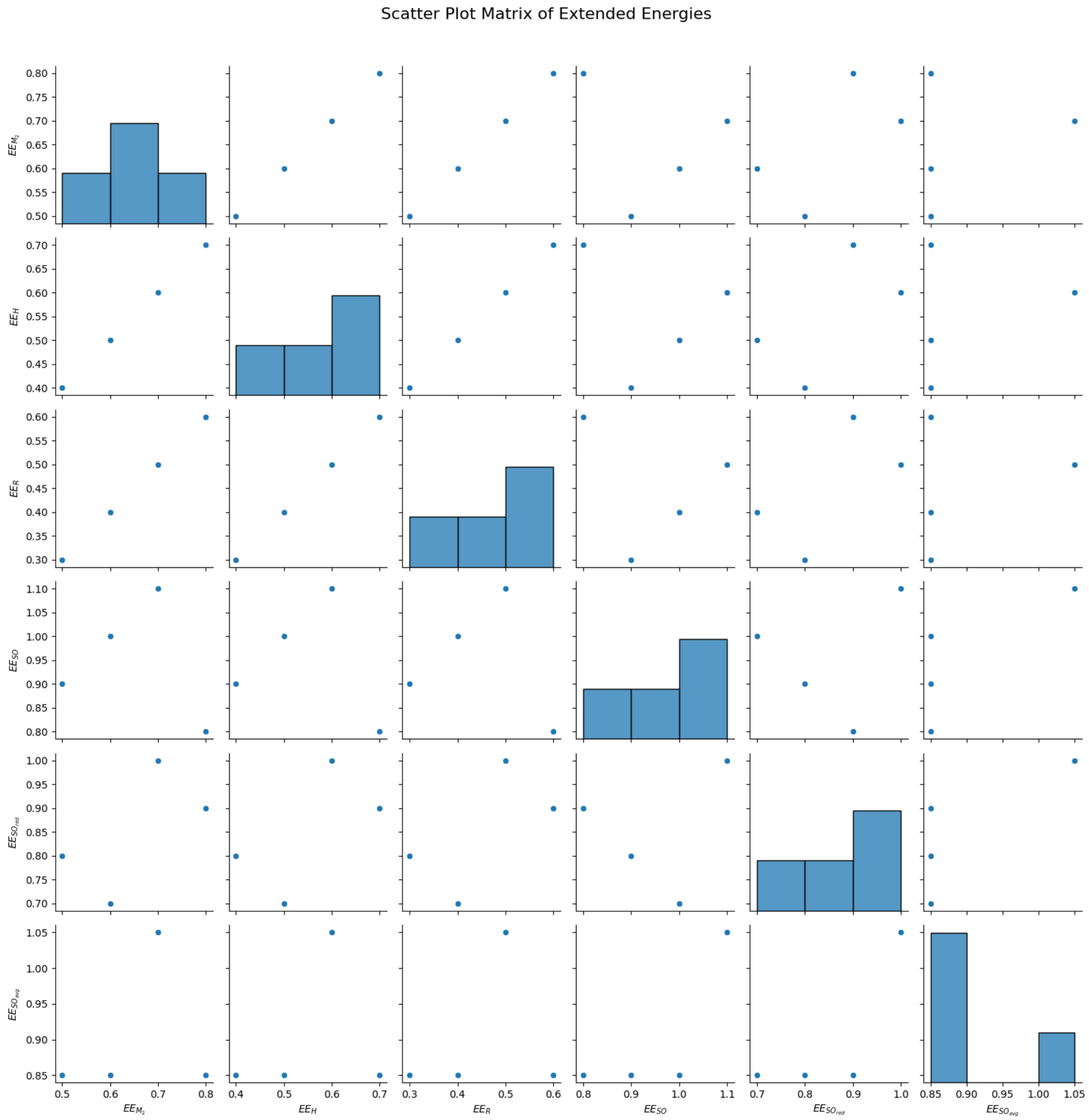


Fig. 7. Scatter plot matrix of extended energies of tuberculosis treatment drugs.

Drugs	*BP*	*MP*	*FP*	*MR*	*P*	*MV*	*MW*	log *P*	*SA*
Isoniazid	251.97	171.4	250	26.9	14.5	97.96	137.142	-0.64	68.01
Pyrazinamide	173.3	192	169.9	25.87	11.13	71.9	123.12	-1.884	68.87
Ethambutol	345.3	201	113.7	58.03	24.47	204.7	204.31	-0.14	64.52
Ethionamide	247.9	162	110	49.06	17.99	124.1	166.25	1.28	38.91
Linezolid	585.5	182	307.9	91.7	34.06	259	337.35	0.55	71.11
Levofloxacin	571.5	225	299.4	88.74	36.69	258.12	361.37	1.27	73.32

Table 2. Physicochemical characteristics of TB treatment drugs.

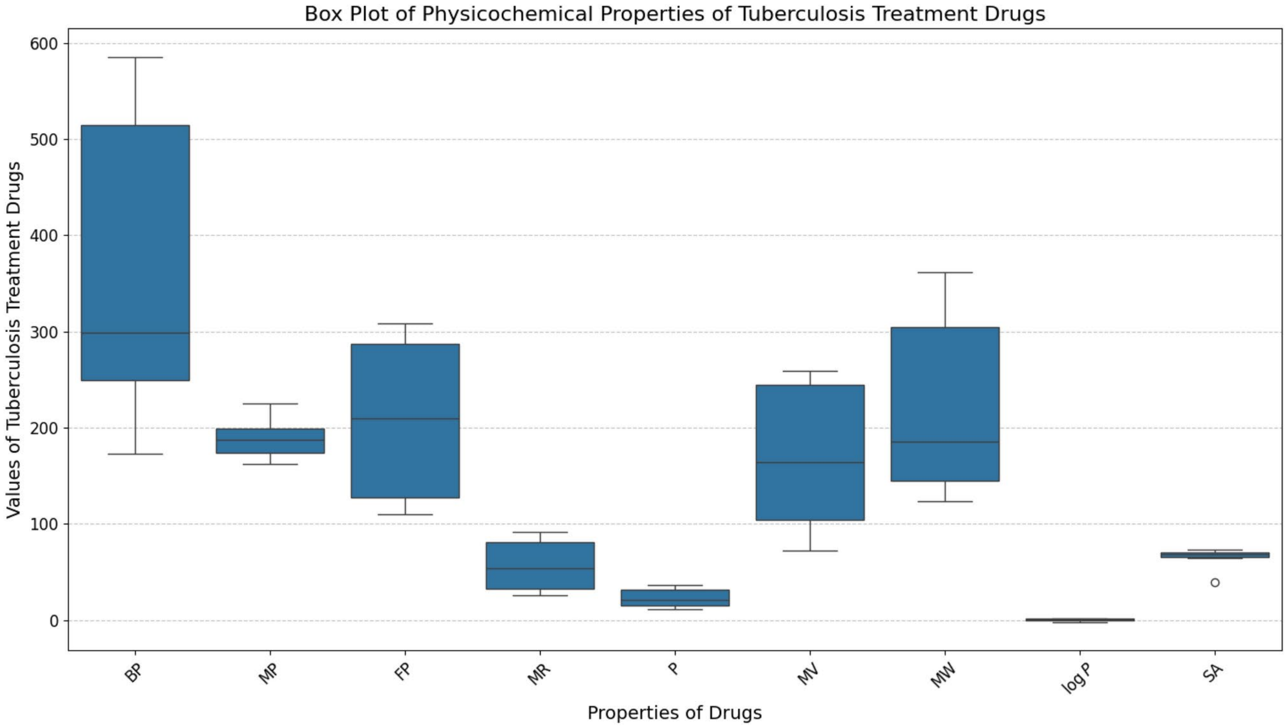


Fig. 8. Graphical analysis of tuberculosis treatment drugs.

$$
\begin{aligned}
FP =&54.933 + 0.891 \times EE_{M_2},\\
R_v =&0.812, \qquad \zeta_e = 58.363, \qquad F = 7.756, \qquad \nabla = 0.050,\\
MR =&8.542 + 0.280 \times EE_{M_2},\\
R_v =&0.792, \qquad \zeta_e = 19.683, \qquad F = 6.712, \qquad \nabla = 0.061,\\
P =&5.004 + 0.105 \times EE_{M_2},\\
R_v =&0.819, \qquad \zeta_e = 6.735, \qquad F = 8.125, \qquad \nabla = 0.046,\\
MV =&45.702 + 0.717 \times EE_{M_2},\\
R_v =&0.712, \qquad \zeta_e = 64.596, \qquad F = 4.102, \qquad \nabla = 0.113,\\
MW =&28.050 + 1.123 \times EE_{M_2},\\
R_v =&0.889, \qquad \zeta_e = 52.809, \qquad F = 15.050, \qquad \nabla = 0.018,\\
\log P =& - 1.433 + 0.009 \times EE_{M_2},\\
R_v =&0.582, \qquad \zeta_e = 1.114, \qquad F = 2.049, \qquad \nabla = 0.226,\\
SA =&53.079 - 0.064 \times EE_{M_2},\\
R_v =&0.412, \qquad \zeta_e = 12.944, \qquad F = 0.816, \qquad \nabla = 0.418.
\end{aligned}
$$

Linear regression models for physicochemical characteristics of TB treatment drugs using $\boldsymbol{EE_H}$

In this section, we identified the models of ΔH_f, *S*, *BP*, log In this section, we identified the models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, *logP*, and *SA* associated with EE_H.

$$\begin{aligned}
BP =& -23.943 + 40.833 \times EE_H, \\
R_v =& 0.870, \quad \zeta_e = 96.916, \quad F = 12.486, \quad \nabla = 0.024, \\
MP =& 153.171 + 3.774 \times EE_H, \\
R_v =& 0.629, \quad \zeta_e = 19.587, \quad F = 2.612, \quad \nabla = 0.181, \\
FP =& 105.800 + 10.848 \times EE_H, \\
R_v =& 0.455, \quad \zeta_e = 89.117, \quad F = 1.042, \quad \nabla = 0.365, \\
MR =& -4.517 + 6.469 \times EE_H, \\
R_v =& 0.842, \quad \zeta_e = 17.369, \quad F = 9.757, \quad \nabla = 0.035, \\
P =& -0.773 + 2.526 \times EE_H, \\
R_v =& 0.903, \quad \zeta_e = 5.028, \quad F = 17.757, \quad \nabla = 0.014, \\
MV =& -26.918 + 20.729 \times EE_H, \\
R_v =& 0.946, \quad \zeta_e = 29.933, \quad F = 33.733, \quad \nabla = 0.004, \\
MW =& 3.394 + 23.051 \times EE_H, \\
R_v =& 0.839, \quad \zeta_e = 62.756, \quad F = 9.490, \quad \nabla = 0.037, \\
\log P =& -1.502 + 0.166 \times EE_H, \\
R_v =& 0.509, \quad \zeta_e = 1.178, \quad F = 1.402, \quad \nabla = 0.302, \\
SA =& 49.048 + 1.593 \times EE_H, \\
R_v =& 0.470, \quad \zeta_e = 12.535, \quad F = 1.136, \quad \nabla = 0.347.
\end{aligned}$$

Linear regression models for physicochemical characteristics of TB treatment drugs using EE_R

In this section, we identified the models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_R.

$$\begin{aligned}
BP =& 431.382 - 4.302 \times EE_R, \\
R_v =& 0.296, \quad \zeta_e = 187.952, \quad F = 0.384, \quad \nabla = 0.569, \\
MP =& 180.728 + 0.511 \times EE_R, \\
R_v =& 0.274, \quad \zeta_e = 24.216, \quad F = 0.326, \quad \nabla = 0.599, \\
FP =& 216.547 - 0.504 \times EE_R, \\
R_v =& 0.068, \quad \zeta_e = 99.823, \quad F = 0.019, \quad \nabla = 0.898, \\
MR =& 68.243 - 0.721 \times EE_R, \\
R_v =& 0.303, \quad \zeta_e = 30.700, \quad F = 0.403, \quad \nabla = 0.560, \\
P =& 27.729 - 0.287 \times EE_R, \\
R_v =& 0.331, \quad \zeta_e = 11.065, \quad F = 0.492, \quad \nabla = 0.522, \\
MV =& 206.093 - 2.301 \times EE_R, \\
R_v =& 0.339, \quad \zeta_e = 86.505, \quad F = 0.518, \quad \nabla = 0.512, \\
MW =& 254.407 - 2.052 \times EE_R, \\
R_v =& 0.241, \quad \zeta_e = 111.854, \quad F = 0.246, \quad \nabla = 0.646, \\
\log P =& 1.265 - 0.075 \times EE_R, \\
R_v =& 0.736, \quad \zeta_e = 0.926, \quad F = 4.740, \quad \nabla = 0.095, \\
SA =& 57.730 + 0.400 \times EE_R, \\
R_v =& 0.381, \quad \zeta_e = 13.133, \quad F = 0.678, \quad \nabla = 0.456.
\end{aligned}$$

Linear regression models for physicochemical characteristics of TB treatment drugs using EE_{SO}

In this section, we identified the models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_{SO}.

$$\begin{aligned}
BP =& -28.127 + 4.399 \times EE_{SO}, \\
R_v =& 0.941, \quad \zeta_e = 66.530, \quad F = 30.985, \quad \nabla = 0.005, \\
MP =& 158.136 + 0.346 \times EE_{SO}, \\
R_v =& 0.579, \quad \zeta_e = 20.533, \quad F = 2.017, \quad \nabla = 0.229, \\
FP =& 39.910 + 1.898 \times EE_{SO}, \\
R_v =& 0.798, \quad \zeta_e = 60.237, \quad F = 7.036, \quad \nabla = 0.057, \\
MR =& -3.896 + 0.683 \times EE_{SO}, \\
R_v =& 0.892, \quad \zeta_e = 14.572, \quad F = 15.544, \quad \nabla = 0.017, \\
P =& 0.410 + 0.256 \times EE_{SO}, \\
R_v =& 0.919, \quad \zeta_e = 4.633, \quad F = 21.629, \quad \nabla = 0.010, \\
MV =& 5.794 + 1.841 \times EE_{SO}, \\
R_v =& 0.843, \quad \zeta_e = 49.471, \quad F = 9.814, \quad \nabla = 0.035,
\end{aligned}$$

$$\begin{aligned}
MW =&- 12.960 + 2.641 \times EE_{SO},\\
R_v =&0.965, \qquad \zeta_e = 30.418, \qquad F = 53.418, \qquad \nabla = 0.002,\\
\log P =&- 1.691 + 0.020 \times EE_{SO},\\
R_v =&0.610, \qquad \zeta_e = 1.085, \qquad F = 2.376, \qquad \nabla = 0.198,\\
SA =&50.416 + 0.154 \times EE_{SO},\\
R_v =&0.457, \qquad \zeta_e = 12.630, \qquad F = 1.058, \qquad \nabla = 0.362.
\end{aligned}$$

Linear regression models for physicochemical characteristics of TB treatment drugs using $\boldsymbol{EE_{SO_{red}}}$

In this section, we identified the models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with $EE_{SO_{red}}$.

$$\begin{aligned}
BP =&40.078 + 5.333 \times EE_{SO_{red}},\\
R_v =&0.880, \qquad \zeta_e = 93.441, \qquad F = 13.735, \qquad \nabla = 0.021,\\
MP =&161.371 + 0.455 \times EE_{SO_{red}},\\
R_v =&0.587, \qquad \zeta_e = 20.389, \qquad F = 2.102, \qquad \nabla = 0.221,\\
FP =&59.891 + 2.457 \times EE_{SO_{red}},\\
R_v =&0.797, \qquad \zeta_e = 60.384, \qquad F = 6.982, \qquad \nabla = 0.057,\\
MR =&7.168 + 0.819 \times EE_{SO_{red}},\\
R_v =&0.826, \qquad \zeta_e = 18.162, \qquad F = 8.583, \qquad \nabla = 0.043,\\
P =&4.366 + 0.310 \times EE_{SO_{red}},\\
R_v =&0.860, \qquad \zeta_e = 5.992, \qquad F = 11.321, \qquad \nabla = 0.028,\\
MV =&38.666 + 2.160 \times EE_{SO_{red}},\\
R_v =&0.763, \qquad \zeta_e = 59.436, \qquad F = 5.570, \qquad \nabla = 0.078,\\
MW =&24.556 + 3.258 \times EE_{SO_{red}},\\
R_v =&0.918, \qquad \zeta_e = 45.723, \qquad F = 21.412, \qquad \nabla = 0.010,\\
\log P =&- 1.451 + 0.025 \times EE_{SO_{red}},\\
R_v =&0.597, \qquad \zeta_e = 1.098, \qquad F = 2.220, \qquad \nabla = 0.210,\\
SA =&52.636 + 0.190 \times EE_{SO_{red}},\\
R_v =&0.434, \qquad \zeta_e = 12.794, \qquad F = 0.930, \qquad \nabla = 0.390.
\end{aligned}$$

Linear regression models for physicochemical characteristics of TB treatment drugs using $\boldsymbol{EE_{SO_{avg}}}$

In this section, we determined the models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with $EE_{SO_{avg}}$.

$$\begin{aligned}
BP =&139.650 + 6.540 \times EE_{SO_{avg}},\\
R_v =&0.914, \qquad \zeta_e = 79.943, \qquad F = 20.230, \qquad \nabla = 0.011,\\
MP =&182.207 + 0.196 \times EE_{SO_{avg}},\\
R_v =&0.214, \qquad \zeta_e = 24.598, \qquad F = 0.193, \qquad \nabla = 0.683,\\
FP =&116.329 + 2.703 \times EE_{SO_{avg}},\\
R_v =&0.743, \qquad \zeta_e = 66.991, \qquad F = 4.991, \qquad \nabla = 0.091,\\
MR =&21.469 + 1.034 \times EE_{SO_{avg}},\\
R_v =&0.882, \qquad \zeta_e = 15.150, \qquad F = 14.082, \qquad \nabla = 0.020,\\
P =&10.882 + 0.360 \times EE_{SO_{avg}},\\
R_v =&0.843, \qquad \zeta_e = 6.307, \qquad F = 9.827, \qquad \nabla = 0.035,\\
MV =&75.647 + 2.747 \times EE_{SO_{avg}},\\
R_v =&0.822, \qquad \zeta_e = 52.420, \qquad F = 8.303, \qquad \nabla = 0.045,\\
MW =&95.469 + 3.700 \times EE_{SO_{avg}},\\
R_v =&0.883, \qquad \zeta_e = 54.189, \qquad F = 14.092, \qquad \nabla = 0.020,\\
\log P =&- 0.792 + 0.025 \times EE_{SO_{avg}},\\
R_v =&0.509, \qquad \zeta_e = 1.179, \qquad F = 1.400, \qquad \nabla = 0.302,\\
SA =&57.601 + 0.191 \times EE_{SO_{avg}},\\
R_v =&0.370, \qquad \zeta_e = 13.193, \qquad F = 0.636, \qquad \nabla = 0.470.
\end{aligned}$$

The heatmap for the linear regression model as shown in Fig. 9 represents the correlation between the extended energy matrix and the physicochemical properties of TB treatment drugs. In this heatmap, the R_v values are shown in color, where darker shades indicate stronger correlations, suggesting a direct or inverse linear relationship between the energy matrix and the property. Lighter shades reflect weaker correlations, indicating

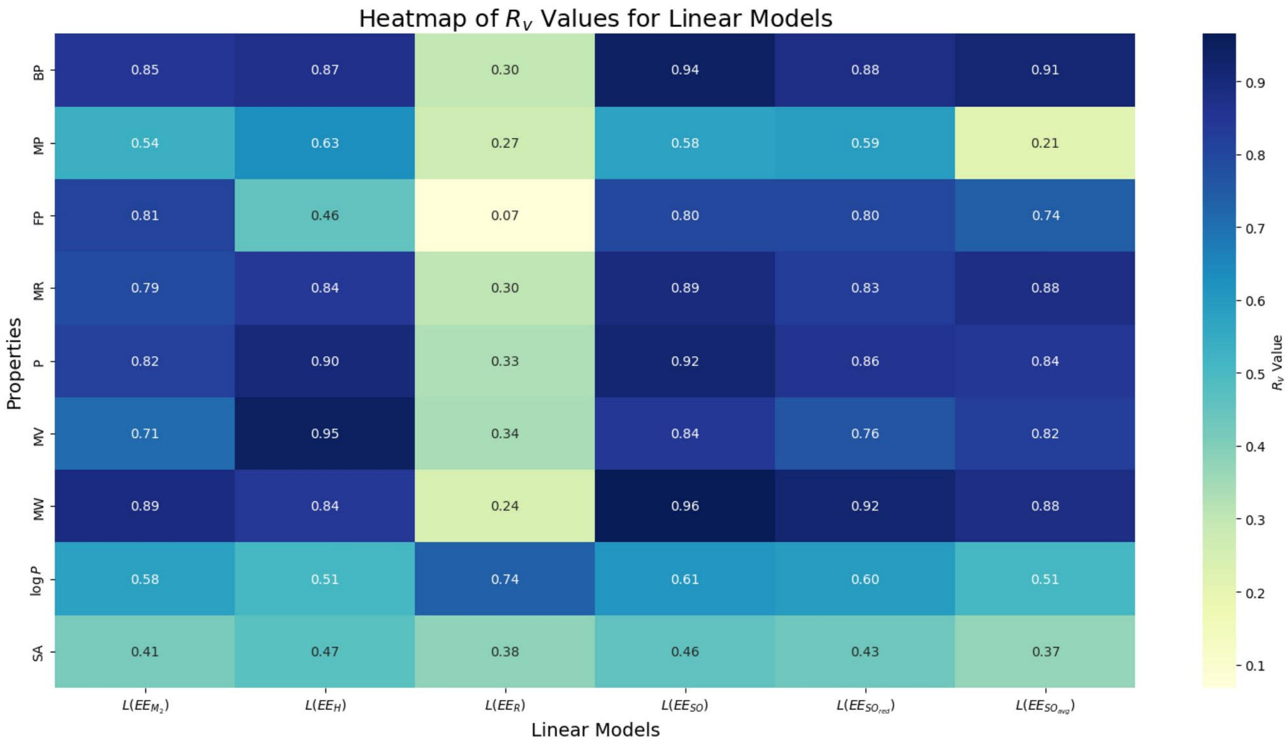


Fig. 9. Heatmap for linear models.

minimal linear dependence. This heatmap is useful for identifying properties that can be effectively predicted using a simple linear regression model, with higher R_v values suggesting a good fit. Properties with weak correlations in this heatmap indicate that a linear approach may not be the best model for those attributes.

Quadratic models related to EE_{M_2}

In this portion, we determined the quadratic models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_{M_2}.

$$\begin{aligned}
BP =&83.326 + 1.396 \times EE_{M_2} + 0.001 \times (EE_{M_2})^2,\\
R_v =&0.849, \qquad \zeta_e = 120.055, \qquad F = 3.872, \qquad \nabla = 0.148,\\
MP =&306.295 - 1.521 \times EE_{M_2} + 0.004 \times (EE_{M_2})^2,\\
R_v =&0.932, \qquad \zeta_e = 10.536, \qquad F = 9.926, \qquad \nabla = 0.048,\\
FP =&-144.091 + 3.216 \times EE_{M_2} - 0.006 \times (EE_{M_2})^2,\\
R_v =&0.855, \qquad \zeta_e = 59.900, \qquad F = 4.080, \qquad \nabla = 0.139,\\
MR =&21.847 + 0.124 \times EE_{M_2} + 0.000 \times (EE_{M_2})^2,\\
R_v =&0.794, \qquad \zeta_e = 22.634, \qquad F = 2.551, \qquad \nabla = 0.225,\\
P =&19.405 - 0.063 \times EE_{M_2} + 0.000 \times (EE_{M_2})^2,\\
R_v =&0.835, \qquad \zeta_e = 7.449, \qquad F = 3.456, \qquad \nabla = 0.167,\\
MV =&178.557 - 0.835 \times EE_{M_2} + 0.004 \times (EE_{M_2})^2,\\
R_v =&0.738, \qquad \zeta_e = 71.684, \qquad F = 1.790, \qquad \nabla = 0.308,\\
MW =&102.740 + 0.251 \times EE_{M_2} + 0.002 \times (EE_{M_2})^2,\\
R_v =&0.893, \qquad \zeta_e = 59.868, \qquad F = 5.911, \qquad \nabla = 0.091,\\
\log P =&-1.336 + 0.008 \times EE_{M_2} + 2.802E - 6 \times (EE_{M_2})^2,\\
R_v =&0.582, \qquad \zeta_e = 1.286, \qquad F = 0.769, \qquad \nabla = 0.538,\\
SA =&74.808 - 0.190 \times EE_{M_2} + 0.001 \times (EE_{M_2})^2,\\
R_v =&0.460, \qquad \zeta_e = 14.562, \qquad F = 0.403, \qquad \nabla = 0.700.
\end{aligned}$$

Quadratic models related to EE_H

In this part, we determined the quadratic models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_H.

$$\begin{aligned}
BP =&96.359 + 10.546 \times EE_H + 1.642 \times (EE_H)^2,\\
R_v =&0.872, \qquad \zeta_e = 111.215, \qquad F = 4.760, \qquad \nabla = 0.117,\\
MP =&330.285 - 40.815 \times EE_H + 2.418 \times (EE_H)^2,\\
R_v =&0.890, \qquad \zeta_e = 13.252, \qquad F = 5.722, \qquad \nabla = 0.095,\\
FP =&- 11.326 + 40.335 \times EE_H - 1.599 \times (EE_H)^2,\\
R_v =&0.467, \qquad \zeta_e = 102.187, \qquad F = 0.417, \qquad \nabla = 0.692,\\
MR =&48.668 - 6.921 \times EE_H + 0.726 \times (EE_H)^2,\\
R_v =&0.855, \qquad \zeta_e = 19.286, \qquad F = 4.079, \qquad \nabla = 0.139,\\
P =&14.589 - 1.341 \times EE_H + 0.210 \times (EE_H)^2,\\
R_v =&0.911, \qquad \zeta_e = 5.584, \qquad F = 7.320, \qquad \nabla = 0.070,\\
MV =&67.264 - 2.983 \times EE_H + 1.286 \times (EE_H)^2,\\
R_v =&0.950, \qquad \zeta_e = 33.161, \qquad F = 13.872, \qquad \nabla = 0.030,\\
MW =&224.793 - 32.688 \times EE_H + 3.022 \times (EE_H)^2,\\
R_v =&0.856, \qquad \zeta_e = 68.747, \qquad F = 4.121, \qquad \nabla = 0.138,
\end{aligned}$$

$$\begin{aligned}
\log P =&- 4.448 + 0.908 \times EE_H - 0.040 \times (EE_H)^2,\\
R_v =&0.545, \qquad \zeta_e = 1.326, \qquad F = 0.633, \qquad \nabla = 0.590,\\
SA =&84.142 - 7.243 \times EE_H + 0.479 \times (EE_H)^2,\\
R_v =&0.520, \qquad \zeta_e = 14.011, \qquad F = 0.555, \qquad \nabla = 0.624.
\end{aligned}$$

Quadratic models related to EE_R

In this portion, we determined the quadratic models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_R.

$$\begin{aligned}
BP =&- 176.047 + 70.204 \times EE_R - 1.546 \times (EE_R)^2,\\
R_v =&0.898, \qquad \zeta_e = 100.137, \qquad F = 6.221, \qquad \nabla = 0.086,\\
MP =&115.695 + 8.488 \times EE_R - 0.166 \times (EE_R)^2,\\
R_v =&0.760, \qquad \zeta_e = 18.892, \qquad F = 2.054, \qquad \nabla = 0.274,\\
FP =&30.498 + 22.316 \times EE_R - 0.474 \times (EE_R)^2,\\
R_v =&0.515, \qquad \zeta_e = 99.036, \qquad F = 0.541, \qquad \nabla = 0.630,\\
MR =&- 26.761 + 10.932 \times EE_R - 0.242 \times (EE_R)^2,\\
R_v =&0.864, \qquad \zeta_e = 18.701, \qquad F = 4.433, \qquad \nabla = 0.127,\\
P =&- 8.572 + 4.166 \times EE_R - 0.092 \times (EE_R)^2,\\
R_v =&0.912, \qquad \zeta_e = 5.554, \qquad F = 7.417, \qquad \nabla = 0.069,\\
MV =&- 93.529 + 34.450 \times EE_R - 0.763 \times (EE_R)^2,\\
R_v =&0.957, \qquad \zeta_e = 30.937, \qquad F = 16.161, \qquad \nabla = 0.025,\\
MW =&- 31.418 + 40.366 \times EE_R - 0.880 \times (EE_R)^2,\\
R_v =&0.858, \qquad \zeta_e = 68.301, \qquad F = 4.194, \qquad \nabla = 0.135,\\
\log P =&- 0.019 + 0.078 \times EE_R - 0.003 \times (EE_R)^2,\\
R_v =&0.778, \qquad \zeta_e = 0.994, \qquad F = 2.295, \qquad \nabla = 0.249,\\
SA =&23.433 + 4.607 \times EE_R - 0.087 \times (EE_R)^2,\\
R_v =&0.764, \qquad \zeta_e = 10.572, \qquad F = 2.109, \qquad \nabla = 0.268.
\end{aligned}$$

Quadratic models related to EE_{SO}

In this part, we determined the quadratic models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_{SO}.

$$\begin{aligned}BP =& -728.906 + 20.114 \times EE_{SO} - 0.077 \times (EE_{SO})^2,\\ R_v =&0.972, \qquad \zeta_e = 53.523, \qquad F = 25.528, \qquad \nabla = 0.013,\\ MP =&384.402 - 4.727 \times EE_{SO} - 0.025 \times (EE_{SO})^2,\\ R_v =&0.842, \qquad \zeta_e = 15.670, \qquad F = 3.666, \qquad \nabla = 0.156,\\ FP =& -152.875 + 6.221 \times EE_{SO} - 0.021 \times (EE_{SO})^2,\\ R_v =&0.809, \qquad \zeta_e = 67.883, \qquad F = 2.845, \qquad \nabla = 0.203,\\ MR =& -120.218 + 3.291 \times EE_{SO} - 0.013 \times (EE_{SO})^2,\\ R_v =&0.925, \qquad \zeta_e = 14.123, \qquad F = 8.904, \qquad \nabla = 0.055,\\ P =& -27.377 + 0.879 \times EE_{SO} - 0.003 \times (EE_{SO})^2,\\ R_v =&0.933, \qquad \zeta_e = 4.883, \qquad F = 10.036, \qquad \nabla = 0.047.\end{aligned}$$

$$\begin{aligned}MV =& -311.194 + 8.949 \times EE_{SO} - 0.035 \times (EE_{SO})^2,\\ R_v =&0.875, \qquad \zeta_e = 51.398, \qquad F = 4.899, \qquad \nabla = 0.114,\\ MW =& -231.152 + 7.534 \times EE_{SO} - 0.024 \times (EE_{SO})^2,\\ R_v =&0.973, \qquad \zeta_e = 30.648, \qquad F = 26.781, \qquad \nabla = 0.012,\\ \log P =& -5.908 + 0.114 \times EE_{SO} - 0.000 \times (EE_{SO})^2,\\ R_v =&0.446, \qquad \zeta_e = 1.208, \qquad F = 1.071, \qquad \nabla = 0.249,\\ SA =&85.568 - 0.634 \times EE_{SO} + 0.004 \times (EE_{SO})^2,\\ R_v =&0.487, \qquad \zeta_e = 14.320, \qquad F = 0.467, \qquad \nabla = 0.666.\end{aligned}$$

Quadratic Models related to $EE_{SO_{red}}$

In this portion, we determined the quadratic models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with $EE_{SO_{red}}$.

$$\begin{aligned}BP =& -265.885 + 15.347 \times EE_{SO_{red}} - 0.069 \times (EE_{SO_{red}})^2,\\ R_v =&0.898, \qquad \zeta_e = 99.795, \qquad F = 6.275, \qquad \nabla = 0.085,\\ MP =&308.808 - 4.370 \times EE_{SO_{red}} + 0.033 \times (EE_{SO_{red}})^2,\\ R_v =&0.898, \qquad \zeta_e = 12.788, \qquad F = 6.256, \qquad \nabla = 0.085,\\ FP =& -208.799 + 11.251 \times EE_{SO_{red}} - 0.060 \times (EE_{SO_{red}})^2,\\ R_v =&0.856, \qquad \zeta_e = 59.699, \qquad F = 4.118, \qquad \nabla = 0.138,\\ MR =& -36.158 + 2.237 \times EE_{SO_{red}} - 0.010 \times (EE_{SO_{red}})^2,\\ R_v =&0.841, \qquad \zeta_e = 20.151, \qquad F = 3.611, \qquad \nabla = 0.159,\\ P =& -0.657 + 0.475 \times EE_{SO_{red}} - 0.001 \times (EE_{SO_{red}})^2,\\ R_v =&0.861, \qquad \zeta_e = 6.886, \qquad F = 4.300, \qquad \nabla = 0.132,\\ MV =& -7.709 + 3.678 \times EE_{SO_{red}} - 0.010 \times (EE_{SO_{red}})^2,\\ R_v =&0.765, \qquad \zeta_e = 68.349, \qquad F = 2.118, \qquad \nabla = 0.267,\\ MW =& -68.594 + 6.307 \times EE_{SO_{red}} - 0.021 \times (EE_{SO_{red}})^2,\\ R_v =&0.923, \qquad \zeta_e = 51.299, \qquad F = 8.594, \qquad \nabla = 0.057,\\ \log P =& -2.654 + 0.065 \times EE_{SO_{red}} + 0.000 \times (EE_{SO_{red}})^2,\\ R_v =&0.606, \qquad \zeta_e = 1.258, \qquad F = 0.871, \qquad \nabla = 0.503,\\ SA =&67.879 - 0.309 \times EE_{SO_{red}} + 0.003 \times (EE_{SO_{red}})^2,\\ R_v =&0.452, \qquad \zeta_e = 14.631, \qquad F = 0.385, \qquad \nabla = 0.710.\end{aligned}$$

Quadratic models related to $EE_{SO_{avg}}$

In this portion, we determined the quadratic models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with $EE_{SO_{avg}}$.

$$\begin{aligned}BP =& -108.629 + 22.089 \times EE_{SO_{avg}} - 0.169 \times \left(EE_{SO_{avg}}\right)^2,\\ R_v =&0.981, \qquad \zeta_e = 44.621, \qquad F = 37.387, \qquad \nabla = 0.008,\\ MP =&131.907 + 3.347 \times EE_{SO_{avg}} + 0.034 \times \left(EE_{SO_{avg}}\right)^2,\\ R_v =&0.602, \qquad \zeta_e = 23.210, \qquad F = 0.854, \qquad \nabla = 0.509,\\ FP =& -75.496 + 5.261 \times EE_{SO_{avg}} - 0.028 \times \left(EE_{SO_{avg}}\right)^2,\\ R_v =&0.752, \qquad \zeta_e = 76.205, \qquad F = 1.948, \qquad \nabla = 0.287,\end{aligned}$$

$$MR = -21.397 + 3.719 \times EE_{SO_{avg}} - 0.029 \times \left(EE_{SO_{avg}}\right)^2,$$
$$R_v = 0.959, \quad \zeta_e = 10.553, \quad F = 17.132, \quad \nabla = 0.023,$$
$$P = -8.690 + 1.585 \times EE_{SO_{avg}} - 0.013 \times \left(EE_{SO_{avg}}\right)^2,$$
$$R_v = 0.965, \quad \zeta_e = 3.530, \quad F = 20.573, \quad \nabla = 0.018,$$
$$MV = -55.909 + 10.987 \times EE_{SO_{avg}} - 0.090 \times \left(EE_{SO_{avg}}\right)^2,$$
$$R_v = 0.915, \quad \zeta_e = 42.783, \quad F = 7.735, \quad \nabla = 0.065,$$
$$MW = -79.232 + 14.647 \times EE_{SO_{avg}} - 0.119 \times \left(EE_{SO_{avg}}\right)^2,$$
$$R_v = 0.981, \quad \zeta_e = 26.053, \quad F = 37.637, \quad \nabla = 0.008,$$
$$\log P = -3.890 + 0.219 \times EE_{SO_{avg}} - 0.002 \times \left(EE_{SO_{avg}}\right)^2,$$
$$R_v = 0.816, \quad \zeta_e = 0.914, \quad F = 2.991, \quad \nabla = 0.193,$$
$$SA = 59.686 + 0.061 \times EE_{SO_{avg}} + 0.001 \times \left(EE_{SO_{avg}}\right)^2,$$
$$R_v = 0.373, \quad \zeta_e = 15.219, \quad F = 0.242, \quad \nabla = 0.799.$$

The quadratic regression model heatmap as shown in Fig. 10 shows how the relationships between the physicochemical properties and extended energies change when a squared term is introduced. A higher R_v value in the quadratic model heatmap, compared to the linear model, indicates that the property follows a nonlinear trend and benefits from the inclusion of the squared term. This heatmap helps in identifying properties with a parabolic relationship, where the impact of extended energies on the property either increases or decreases at an accelerating rate, highlighting properties that require a more complex regression model for accurate prediction

Logarithm models related to EE_{M_2}

In this portion, we determined the logarithm models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_{M_2}.

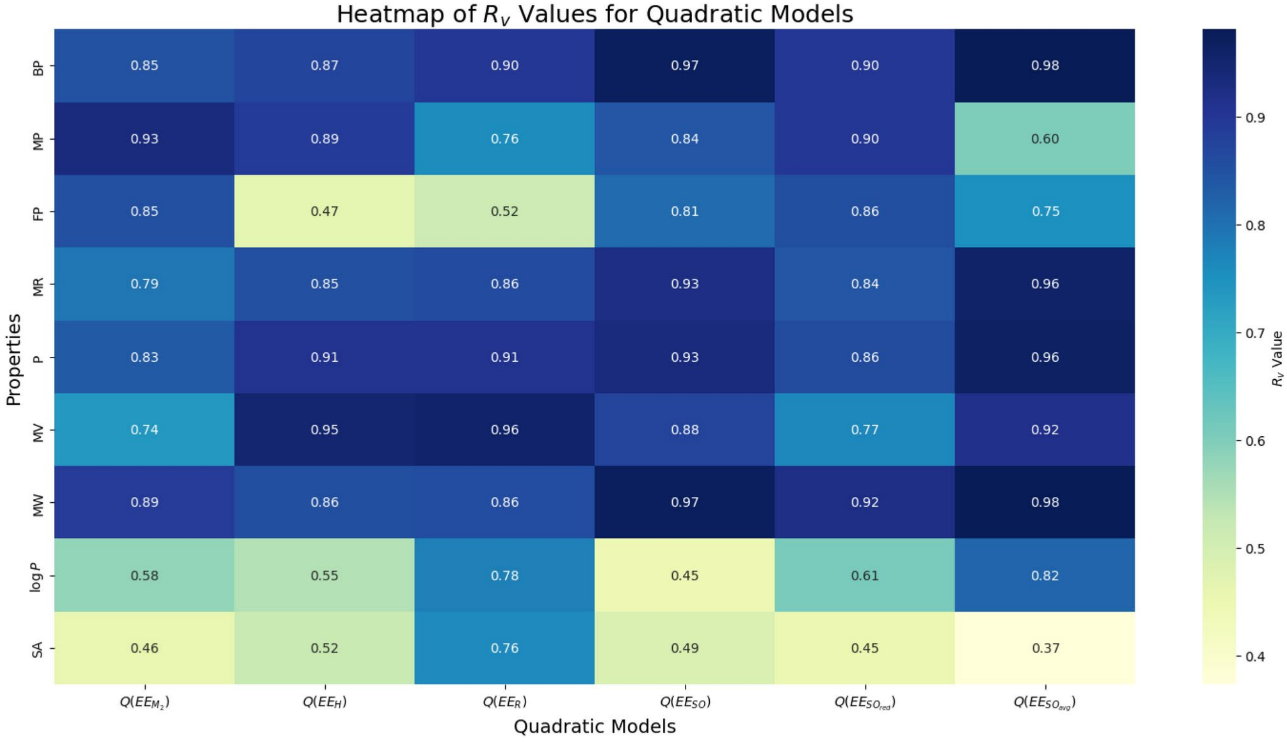


Fig. 10. Heatmap for quadratic models.

$$\begin{aligned}
BP =& -1289.992 + 326.308 \times \ln(EE_{M_2}),\\
R_v =& 0.817, \quad \zeta_e = 113.584, \quad F = 8.003, \quad \nabla = 0.047,\\
MP =& 81.337 + 21.239 \times \ln(EE_{M_2}),\\
R_v =& 0.415, \quad \zeta_e = 22.909, \quad F = 0.833, \quad \nabla = 0.413,\\
FP =& 654.611 + 170.422 \times \ln(EE_{M_2}),\\
R_v =& 0.839, \quad \zeta_e = 54.502, \quad F = 9.481, \quad \nabla = 0.037,\\
MR =& -193.321 + 49.371 \times \ln(EE_{M_2}),\\
R_v =& 0.755, \quad \zeta_e = 21.134, \quad F = 5.292, \quad \nabla = 0.083,\\
P =& -69.377 + 18.268 \times \ln(EE_{M_2}),\\
R_v =& 0.767, \quad \zeta_e = 7.524, \quad F = 5.716, \quad \nabla = 0.075,\\
MV =& -447.376 + 121.765 \times \ln(EE_{M_2}),\\
R_v =& 0.652, \quad \zeta_e = 69.697, \quad F = 2.960, \quad \nabla = 0.160,\\
MW =& -786.483 + 199.049 \times \ln(EE_{M_2}),\\
R_v =& 0.850, \quad \zeta_e = 60.641, \quad F = 10.448, \quad \nabla = 0.032,\\
\log P =& -7.967 + 1.588 \times \ln(EE_{M_2}),\\
R_v =& 0.571, \quad \zeta_e = 1.124, \quad F = 1.933, \quad \nabla = 0.237,\\
SA =& 10.495 + 10.589 \times \ln(EE_{M_2}),\\
R_v =& 0.367, \quad \zeta_e = 13.211, \quad F = 0.623, \quad \nabla = 0.474.
\end{aligned}$$

Logarithm models related to EE_H

In this portion, we determined the logarithm models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_H.

$$\begin{aligned}
BP =& -392.395 + 347.396 \times \ln(EE_H),\\
R_v =& 0.860, \quad \zeta_e = 100.318, \quad F = 11.387, \quad \nabla = 0.028,\\
MP =& 125.898 + 28.990\times\\
R_v =& 0.561, \quad \zeta_e = 20.849, \quad F = 1.836, \quad \nabla = 0.247,\\
FP =& 2.593 + 94.739 \times \ln(EE_H),\\
R_v =& 0.461, \quad \zeta_e = 88.772, \quad F = 1.082, \quad \nabla = 0.357,\\
MR =& -61.634 + 54.458 \times \ln(EE_H),\\
R_v =& 0.824, \quad \zeta_e = 18.263, \quad F = 8.443, \quad \nabla = 0.044,\\
P =& -23.326 + 21.381 \times \ln(EE_H),\\
R_v =& 0.888, \quad \zeta_e = 5.384, \quad F = 14.976, \quad \nabla = 0.018,\\
MV =& -212.645 + 175.748 \times \ln(EE_H),\\
R_v =& 0.931, \quad \zeta_e = 33.458, \quad F = 26.200, \quad \nabla = 0.007,\\
MW =& -198.909 + 193.490 \times \ln(EE_H),\\
R_v =& 0.818, \quad \zeta_e = 66.290, \quad F = 8.090, \quad \nabla = 0.047,
\end{aligned}$$

$$\begin{aligned}
\log P =& -3.208 + 1.510 \times \ln(EE_H),\\
R_v =& 0.537, \quad \zeta_e = 1.155, \quad F = 1.622, \quad \nabla = 0.272,\\
SA =& 36.451 + 12.733 \times \ln(EE_H),\\
R_v =& 0.437, \quad \zeta_e = 12.776, \quad F = 0.943, \quad \nabla = 0.386.
\end{aligned}$$

Logarithm models related to EE_R

In this portion, we determined the logarithm models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_R.

$$\begin{aligned}
BP =&404.569 - 16.237 \times \ln(EE_R),\\
R_v =&0.059, \qquad \zeta_e = 196.419, \qquad F = 0.014, \qquad \nabla = 0.912,\\
MP =&146.422 + 16.425 \times \ln(EE_R),\\
R_v =&0.463, \qquad \zeta_e = 22.327, \qquad F = 1.089, \qquad \nabla = 0.356,\\
FP =&180.707 + 10.740 \times \ln(EE_R),\\
R_v =&0.076, \qquad \zeta_e = 99.765, \qquad F = 0.023, \qquad \nabla = 0.886,\\
MR =&66.105 - 3.630 \times \ln(EE_R),\\
R_v =&0.080, \qquad \zeta_e = 32.108, \qquad F = 0.026, \qquad \nabla = 0.880,\\
P =&27.094 - 1.529 \times \ln(EE_R),\\
R_v =&0.092, \qquad \zeta_e = 11.676, \qquad F = 0.034, \qquad \nabla = 0.862,\\
MV =&198.371 - 11.242 \times \ln(EE_R),\\
R_v =&0.087, \qquad \zeta_e = 91.588, \qquad F = 0.030, \qquad \nabla = 0.870,\\
MW =&227.376 - 2.237 \times \ln(EE_R),\\
R_v =&0.014, \qquad \zeta_e = 115.236, \qquad F = 0.001, \qquad \nabla = 0.979,\\
\log P =&3.327 - 1.258 \times \ln(EE_R),\\
R_v =&0.652, \qquad \zeta_e = 1.039, \qquad F = 2.954, \qquad \nabla = 0.161,\\
SA =&35.161 + 11.199 \times \ln(EE_R),\\
R_v =&0.559, \qquad \zeta_e = 11.775, \qquad F = 1.820, \qquad \nabla = 0.249.
\end{aligned}$$

Logarithm models related to EE_{SO}

In this portion, we determined the logarithm models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with EE_{SO}.

$$\begin{aligned}
BP =& - 1553.783 + 433.606 \times \ln(EE_{SO}),\\
R_v =&0.956, \qquad \zeta_e = 57.693, \qquad F = 42.523, \qquad \nabla = 0.003,\\
MP =&30.945 + 52.134 \times \ln(EE_{SO}),\\
R_v =&0.533, \qquad \zeta_e = 21.306, \qquad F = 1.588, \qquad \nabla = 0.276,\\
FP =& - 609.952 + 185.183 \times \ln(EE_{SO}),\\
R_v =&0.803, \qquad \zeta_e = 59.642, \qquad F = 7.257, \qquad \nabla = 0.054,\\
MR =& - 241.029 + 67.369 \times \ln(EE_{SO}),\\
R_v =&0.907, \qquad \zeta_e = 13.542, \qquad F = 18.632, \qquad \nabla = 0.012,\\
P =& - 87.890 + 25.122 \times \ln(EE_{SO}),\\
R_v =&0.929, \qquad \zeta_e = 4.328, \qquad F = 25.367, \qquad \nabla = 0.007,\\
MV =& - 635.151 + 182.019 \times \ln(EE_{SO}),\\
R_v =&0.859, \qquad \zeta_e = 47.081, \qquad F = 11.252, \qquad \nabla = 0.028,\\
MW =& - 918.717 + 258.012 \times \ln(EE_{SO}),\\
R_v =&0.971, \qquad \zeta_e = 27.447, \qquad F = 66.525, \qquad \nabla = 0.001,\\
\log P =& - 8.757 + 1.998 \times \ln(EE_{SO}),\\
R_v =&0.633, \qquad \zeta_e = 1.060, \qquad F = 2.673, \qquad \nabla = 0.177,\\
SA =&0.781 + 14.332 \times \ln(EE_{SO}),\\
R_v =&0.438, \qquad \zeta_e = 12.770, \qquad F = 0.948, \qquad \nabla = 0.385.
\end{aligned}$$

Logarithm models related to $EE_{SO_{red}}$

In this portion, we determined the logarithm models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with $EE_{SO_{red}}$.

$$\begin{aligned}
BP =& -1087.979 + 360.976 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.886, \quad \zeta_e = 91.268, \quad F = 14.590, \quad \nabla = 0.019,\\
MP =& 84.312 + 26.027 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.499, \quad \zeta_e = 21.823, \quad F = 1.327, \quad \nabla = 0.314,\\
FP =& -480.133 + 171.364 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.827, \quad \zeta_e = 56.248, \quad F = 8.657, \quad \nabla = 0.042,\\
MR =& -165.394 + 55.273 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.829, \quad \zeta_e = 18.033, \quad F = 8.762, \quad \nabla = 0.042,\\
P =& -59.894 + 20.663 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.851, \quad \zeta_e = 6.160, \quad F = 10.494, \quad \nabla = 0.032,\\
MV =& -406.699 + 143.338 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.753, \quad \zeta_e = 60.507, \quad F = 5.234, \quad \nabla = 0.084,\\
MW =& -655.840 + 218.351 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.915, \quad \zeta_e = 46.528, \quad F = 20.541, \quad \nabla = 0.011,\\
\log P =& -6.871 + 1.728 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.609, \quad \zeta_e = 1.086, \quad F = 2.362, \quad \nabla = 0.199,\\
SA =& 15.953 + 11.987 \times \ln(EE_{SO_{red}}),\\
R_v =& 0.408, \quad \zeta_e = 12.970, \quad F = 0.797, \quad \nabla = 0.423.
\end{aligned}$$

Logarithm models related to $EE_{SO_{avg}}$

In this portion, we determined the logarithm models for *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log P$, and *SA* associated with $EE_{SO_{avg}}$.

$$\begin{aligned}
BP =& -517.250 + 263.223 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.961, \quad \zeta_e = 54.519, \quad F = 48.099, \quad \nabla = 0.002,\\
MP =& 156.153 + 9.797 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.279, \quad \zeta_e = 24.180, \quad F = 0.339, \quad \nabla = 0.592,\\
FP =& -130.757 + 101.493 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.729, \quad \zeta_e = 68.539, \quad F = 4.524, \quad \nabla = 0.101,
\end{aligned}$$

$$\begin{aligned}
MR =& -83.895 + 42.068 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.938, \quad \zeta_e = 11.166, \quad F = 29.288, \quad \nabla = 0.006,\\
P =& -26.813 + 14.945 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.915, \quad \zeta_e = 4.722, \quad F = 20.670, \quad \nabla = 0.010,\\
MV =& -210.405 + 113.597 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.887, \quad \zeta_e = 42.370, \quad F = 14.832, \quad \nabla = 0.018,\\
MW =& -281.645 + 150.556 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.938, \quad \zeta_e = 39.866, \quad F = 29.428, \quad \nabla = 0.006,\\
\log P =& -4.038 + 1.230 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.645, \quad \zeta_e = 1.046, \quad F = 2.851, \quad \nabla = 0.167,\\
SA =& 42.943 + 6.337 \times \ln(EE_{SO_{avg}}),\\
R_v =& 0.320, \quad \zeta_e = 13.454, \quad F = 0.458, \quad \nabla = 0.536.
\end{aligned}$$

The logarithmic regression model heatmap as shown in Fig. 11 examines the relationships where the rate of change of the dependent variable decreases as the independent variable increases. This heatmap reveals whether the logarithmic transformation improves the fit compared to the linear and quadratic models. Strong correlations in the logarithmic model suggest that the relationship between extended engeries and certain properties follows a diminishing return pattern, where the effect of extended engeries on the property diminishes as its value increases. By comparing the heatmaps of the three models, it is possible to determine which transformation best captures the behavior of each physicochemical property.

Statistical validation of predictive model consistency

In statistical analysis, correlation is a fundamental measure used to assess the strength and direction of the relationship between two variables. It provides insights into how variations in one variable correspond to changes in another, making it a crucial tool in predictive modeling. The correlation coefficient (*r*) between 1 and 1 varies, with positive values for a direct relation and a negative value for an inverse relation, and values near zero for no relation and a weak relation[28]. As a larger value in terms of its absolute value, a larger association between two variables is denoted. In chemical graph theory, a function of significant role in predictive capability checking of topological indices in molecular property characterization is played through correlation analysis.

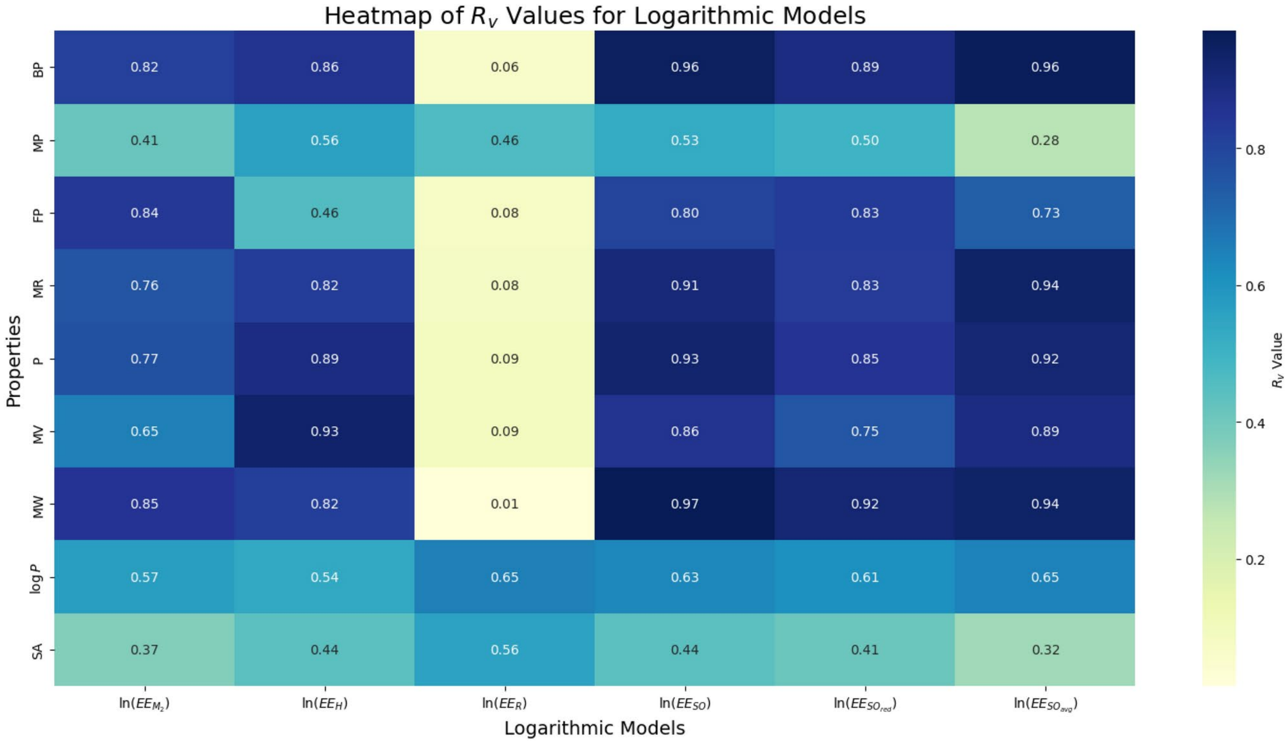


Fig. 11. Heatmap for logarithm models.

For further evidence of the quadratic model effectiveness, the actual values of three critical physicochemical characteristics, boiling point, melting point, and flash point, are graphically exhibited alongside their respective predicted values. These are included as Figs. 12, 13, and 14, respectively, within the manuscript.

The quadratic model for regression was determined to provide the highest prediction accuracy for such properties, with actual vs. predicted values plotting very near the regression line, reflecting the presence of a strong correlation as well as a good prediction. R^2 values as well as the RMSE values for each of the properties further support the accuracy of the model. Similar graphs for molar refractivity, polarizability, molar volume, molecular weight, log of partition coefficient, and surface area can be drawn by following the same procedure. These plots give an overall insight into the model's stability with respect to varied drug properties and enhance the usefulness of the quadratic model for QSPR analysis. Table 3 shows the extended energies of a variety of topological descriptors and physicochemical descriptors of drugs for antitubercular activity. The indices analyzed include EE_{M_2}, EE_H, EE_R, EE_{SO}, $EE_{SO_{red}}$, $EE_{SO_{avg}}$, while the molecular properties considered are *BP*, *MP*, *FP*, *MR*, *P*, *MV*, *MW*, $\log(P)$, and *SA*. The correlation values denote the intensity of association between each molecular property and topological index, with high values signifying strong relations. In a striking observation, Sombor index (EE_{SO}) reflects strong relations with a range of significant properties, such as MW, $(r = 0.965)$ and BP, $(r = 0.941)$, indicative of its use in explaining molecular behavior. In contrast, the Randic index EE_R reflects a negative relationship with properties such as MR, $(r = -0.303)$ and $\log(P)$, $(r = -0.736)$, indicative of an inverse relationship. The variation in correlation values across different indices emphasizes the importance of selecting appropriate descriptors for predictive modeling, as certain indices consistently exhibit stronger associations with molecular properties. This analysis reinforces the reliability of predictive models and provides valuable insights into the most influential topological indices for understanding the physicochemical behavior of TB drugs, contributing to the development of more accurate pharmaceutical property predictions.

In addition to the analysis presented in Tables 4, 5, and 6, where $EE_{SO_{avg}}$ consistently yielded the lowest RMSE values across linear, quadratic, and logarithmic models, a comparison using Python and R revealed significant differences in the predictive accuracy of each model. This was further illustrated by a bar plot chart of RMSE values as shown in Figs. 15, 16, and 17 for each model, which visually demonstrated the performance of the different modeling approaches in predicting drug properties. The findings underscore the importance of model selection and optimal descriptor choice for accurate molecular property predictions

The code, as shown in Fig. 18, provides an algorithm to compare the RMSE values and R_v values for different models (Linear, Quadratic, and Logarithmic) in predicting drug properties. It first loads the RMSE and R_v value data from separate Excel files and ensures that the drug properties match across both datasets. The script then extracts the relevant data for each model and property and determines the best model for each drug property based on the minimum RMSE and maximum R_v value. The result is summarized in a new Excel file, listing the best model for each property along with its corresponding R_v value and RMSE. According to the comparison, the quadratic model emerges as the best for predicting the drug properties. The corresponding chart and summary are captioned as the "Algorithm" for visualization.

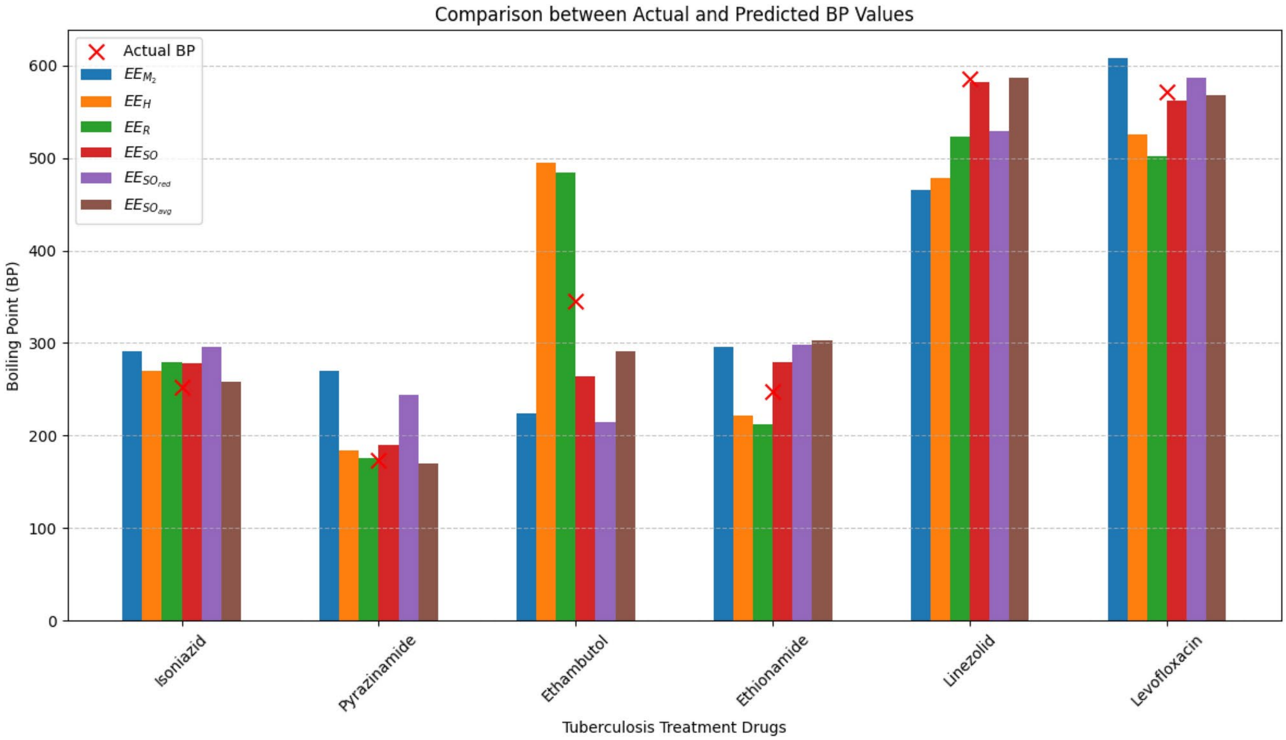


Fig. 12. Boiling point - actual vs. predicted values for the quadratic regression model.

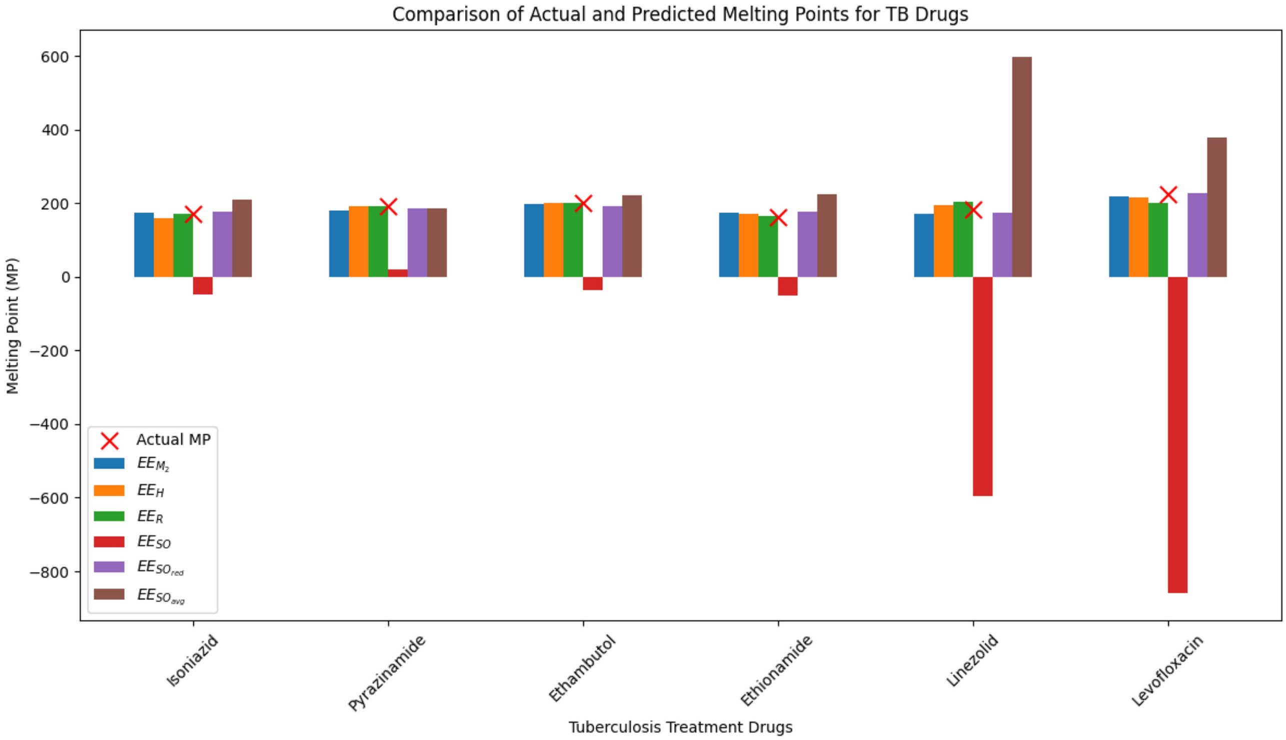


Fig. 13. Melting Point - Actual vs. Predicted Values for the Quadratic Regression Model.

Algorithm

Following the comparison of RMSE and R_v values across different models, a “Quadratic Scatter Plot between Extended Energy of M_2 and Drug Properties” was generated as shown in Fig. 19. This plot illustrates the relationship between the extended energy of M_2 and the drug properties, with the quadratic model effectively

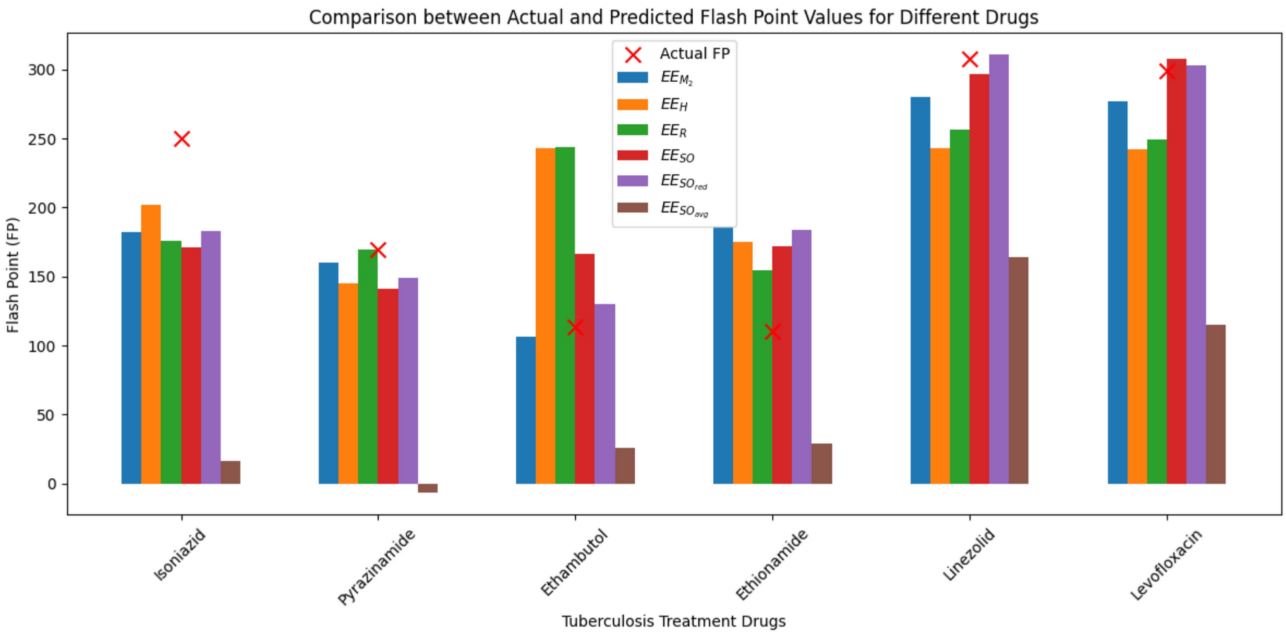


Fig. 14. Flash point - actual vs. predicted values for the quadratic regression model.

Properties of drugs	EE_{M_2}	EE_H	EE_R	EE_{SO}	$EE_{SO_{red}}$	$EE_{SO_{avg}}$
BP	0.849	0.87	-0.296	0.941	0.88	0.914
MP	0.536	0.629	0.274	0.579	0.587	0.214
FP	0.812	0.455	-0.068	0.798	0.797	0.743
MR	0.792	0.842	-0.303	0.892	0.826	0.882
P	0.819	0.903	-0.331	0.919	0.860	0.843
MV	0.712	0.946	-0.339	0, 843	0.763	0.822
MW	0.889	0.839	-0.241	0.965	0.918	0.883
log P	0.582	0.509	-0.736	0.61	0.597	0.509
SA	0.412	0.47	0.381	0.457	0.434	0.370

Table 3. Correlation among extended energies of topological indices and molecular properties of pharmaceutical compounds.

Properties of Drugs	$L(EE_{M_2})$	$L(EE_H)$	$L(EE_R)$	$L(EE_{SO})$	$L(EE_{SO_{red}})$	$L(EE_{SO_{avg}})$
BP	258.7039	1188.673	192.3846	259.679	168.0236	92.1292
MP	82.6495	112.2976	18.9123	26.2853	21.6081	19.2096
FP	122.9248	325.3311	83.4822	126.8023	98.2937	76.485
MR	55.8469	186.5904	32.0912	40.1079	25.3788	12.1117
P	91.0768	73.3699	11.8027	15.0491	9.6422	5.0629
MV	77.9561	605.6785	92.141	111.8278	73.5551	44.4363
MW	129.4312	666.3083	110.1013	153.2627	97.3105	47.0819
log P	115.2049	4.7085	1.6783	1.3745	1.0449	0.8193
SA	57.4888	50.0912	11.9985	14.4434	12.8836	11.9462

Table 4. Root mean square error analysis of linear models for various drug properties.

capturing the correlation. The scatter plot demonstrates the predictive accuracy of the quadratic model in depicting the influence of M_2 on the drug properties. A similar approach can be applied to other extended energies, allowing for comparative analysis of their effects on drug properties. These plots provide a clear visualization of the performance and predictive potential of each extended energy descriptor when used within a quadratic framework.

Properties of Drugs	$Q(EE_{M_2})$	$Q(EE_H)$	$Q(EE_R)$	$Q(EE_{SO})$	$Q(EE_{SO_{red}})$	$Q(EE_{SO_{avg}})$
BP	268.5345	3301.967	992.4777	527.4754	224.2947	131.4076
MP	60.0561	3284.867	85.9689	122.4507	50.8526	143.2644
FP	106.0972	1819.049	308.0136	198.7778	169.1644	165.5066
MR	77.6443	1132.483	156.9403	83.8254	31.6855	16.1519
P	111.8858	346.432	59.2775	24.958	9.9512	8.2106
MV	66.2894	2273.725	485.1814	229.8079	76.9526	57.667
MW	113.9163	4605.032	539.4	230.5861	110.9487	75.1085
log P	135.7957	48.1505	3.8159	2.352	0.8408	1.1652
SA	74.7031	676.0449	41.1265	13.83	11.6304	11.6231

Table 5. Root mean square error analysis of quadratic models for various drug properties.

Properties of drugs	EE_{M_2}	EE_H	EE_R	EE_{SO}	$EE_{SO_{red}}$	$EE_{SO_{avg}}$
BP	560.3113	449.711	168.4461	478.6615	282.9367	119.1011
MP	415.4914	41.0695	22.9365	35.1669	129.7912	18.8308
FP	437.2218	147.2886	81.2024	226.9528	159.7251	92.7285
MR	295.0245	68.736	28.3869	72.1957	93.6558	13.3887
P	278.6169	27.3397	10.4052	27.3249	124.5838	5.8477
MV	379.6025	225.6554	80.992	198.4435	93.472	48.5876
MW	424.3646	247.4303	95.2007	281.201	135.0338	59.4204
log P	267.1371	2.0014	2.0323	2.2066	145.0413	0.7614
SA	315.4029	21.2677	16.6237	20.9835	111.9893	12.5184

Table 6. Root mean square error analysis of logarithm models for various drug properties.

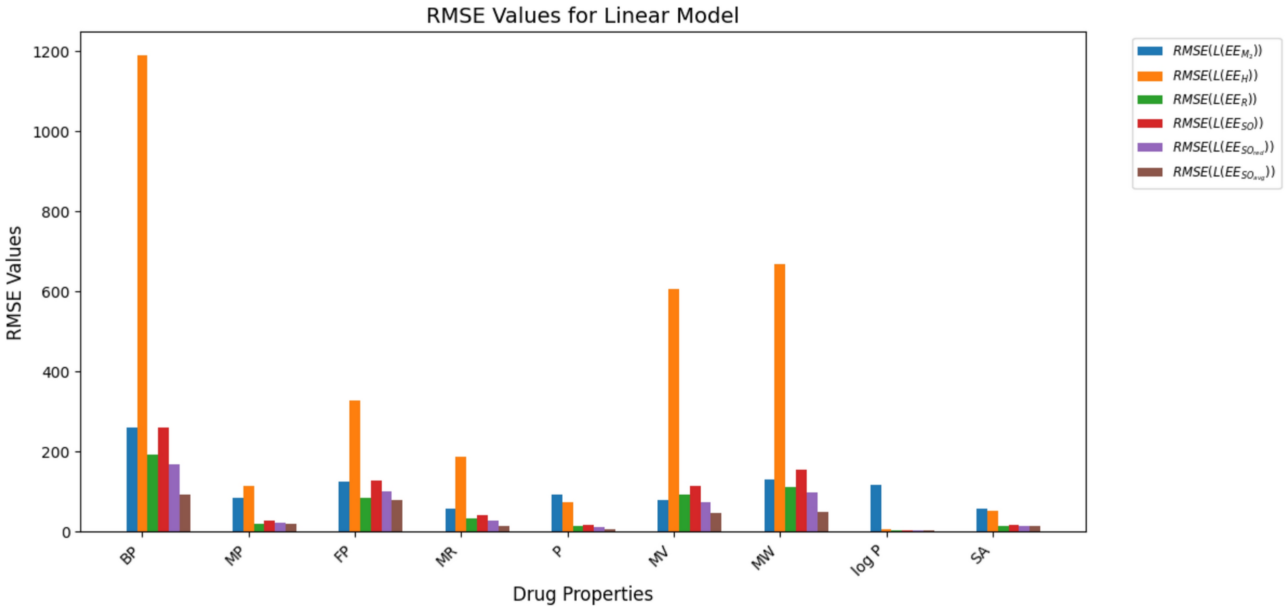


Fig. 15. Barplot for linear models.

Model significance and validation criteria

In the current study, the topological indices based on extended energy were assessed by application of several regression models-linear, quadratic, and logarithmic-to model the following nine essential physicochemical characteristics: boiling point, melting point, flash point, molar refractivity, polarizability (P), molar volume, molecular weight, logarithm of the partition coefficient, and surface area. Standard statistical measures such as the coefficient of determination (R^2), root mean square error (RMSE), and the adjusted R^2 were employed to evaluate the performance of the model. A model is statistically significant when R^2 is high, and RMSE is low.

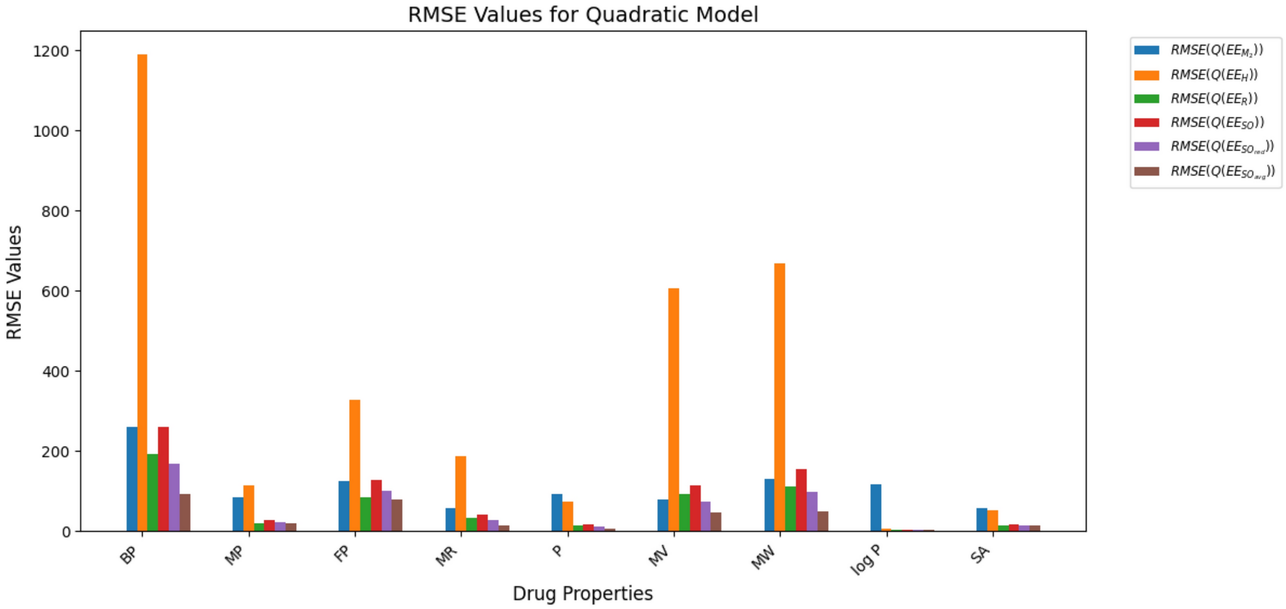


Fig. 16. Barplot for quadratic models.

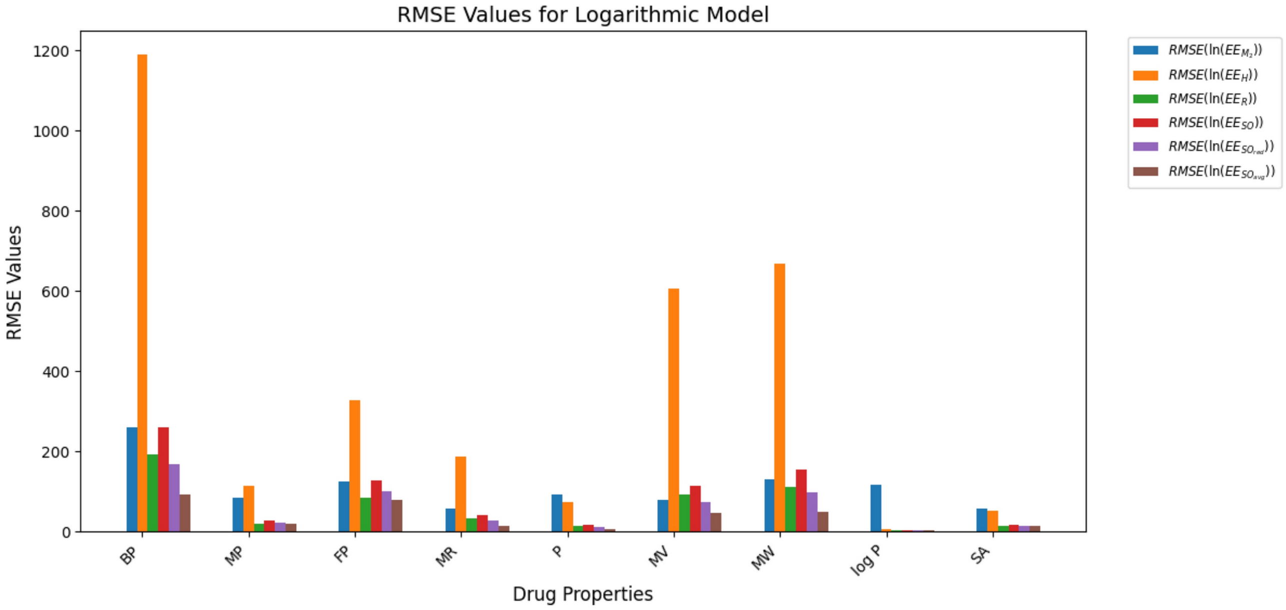
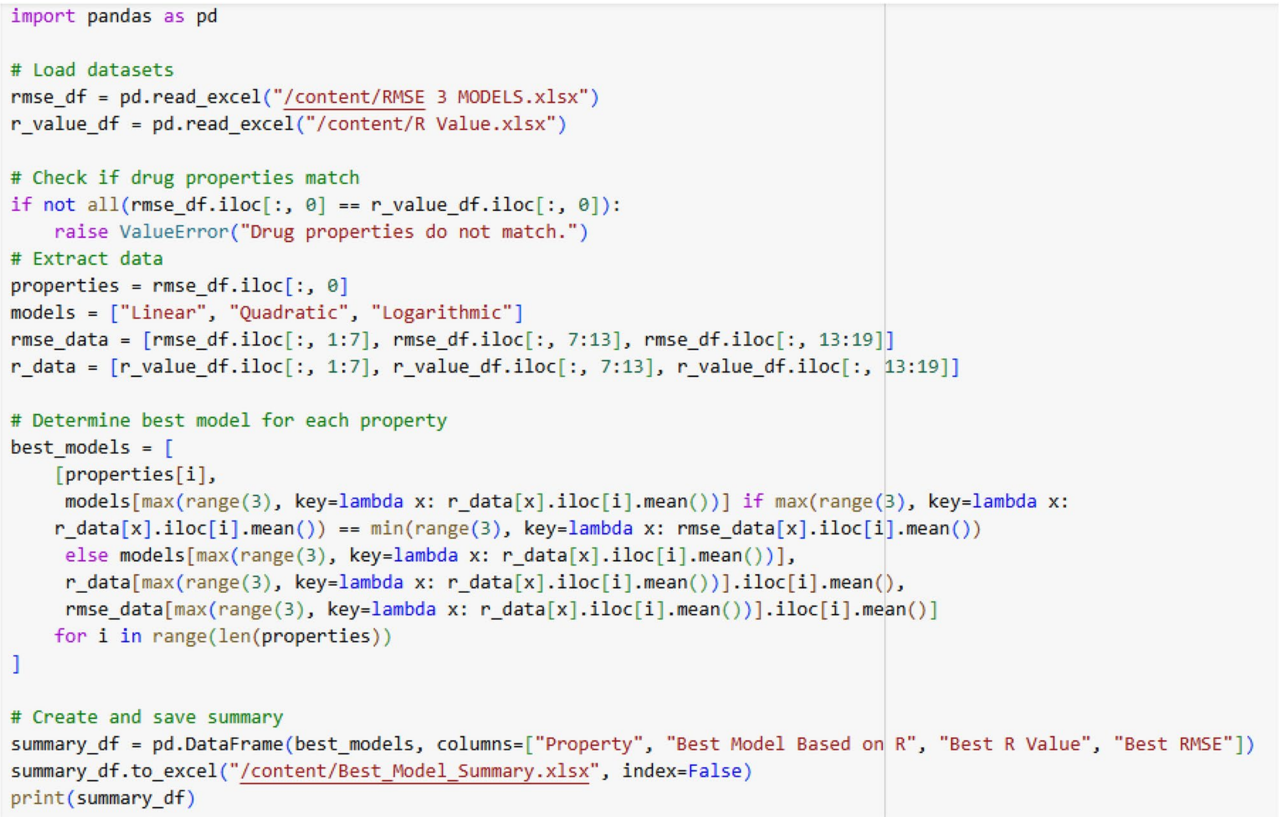


Fig. 17. Barplot for logarithm models.

Among the tested models, the quadratic model outperformed the remainder in the prediction of the majority of drug properties, such as BP, FP, MR, MV, and $\log P$, as demonstrated by greater R^2 and smaller RMSE values. The model, however, had low predictivity for MP, suggesting that the topological descriptors employed are perhaps insufficient to account for the underlying structural or energetics that impact the melting point. This points toward the possibility of investigating more complex models or the inclusion of extra descriptors that are more specialized in the case of MP in subsequent work.

Limitations and future work

The research has some limitations. The dataset contains only a few tuberculosis drugs, so the generalizability of the findings might be impacted by it. The employed QSPR models, including the linear, quadratic, and logarithmic ones, are simplifications and do not necessarily represent complex molecular interactions. No external validation using independent datasets, a factor that can increase the robustness of the model, was practiced. Further research in the future will extend the dataset, employ more sophisticated machine learning methods



```python
import pandas as pd

# Load datasets
rmse_df = pd.read_excel("/content/RMSE 3 MODELS.xlsx")
r_value_df = pd.read_excel("/content/R Value.xlsx")

# Check if drug properties match
if not all(rmse_df.iloc[:, 0] == r_value_df.iloc[:, 0]):
    raise ValueError("Drug properties do not match.")
# Extract data
properties = rmse_df.iloc[:, 0]
models = ["Linear", "Quadratic", "Logarithmic"]
rmse_data = [rmse_df.iloc[:, 1:7], rmse_df.iloc[:, 7:13], rmse_df.iloc[:, 13:19]]
r_data = [r_value_df.iloc[:, 1:7], r_value_df.iloc[:, 7:13], r_value_df.iloc[:, 13:19]]

# Determine best model for each property
best_models = [
    [properties[i],
     models[max(range(3), key=lambda x: r_data[x].iloc[i].mean())] if max(range(3), key=lambda x:
    r_data[x].iloc[i].mean()) == min(range(3), key=lambda x: rmse_data[x].iloc[i].mean())
     else models[max(range(3), key=lambda x: r_data[x].iloc[i].mean())],
     r_data[max(range(3), key=lambda x: r_data[x].iloc[i].mean())].iloc[i].mean(),
     rmse_data[max(range(3), key=lambda x: r_data[x].iloc[i].mean())].iloc[i].mean()]
    for i in range(len(properties))
]

# Create and save summary
summary_df = pd.DataFrame(best_models, columns=["Property", "Best Model Based on R", "Best R Value", "Best RMSE"])
summary_df.to_excel("/content/Best_Model_Summary.xlsx", index=False)
print(summary_df)
```

Fig. 18. Comparison algorithm for RMSE and R_v values across Linear, Quadratic, and Logarithmic models, highlighting the quadratic model as the best.

such as graph neural networks and random forests, and incorporate hybrid indices to enhance the accuracy of the predictions. More detailed and open-source code implementations can also facilitate reproducibility and stimulate further studies in the topic.

Conclusion

In this study, we analyzed the physicochemical properties of six Tuberculosis (TB) drugs using extended energies of topological indices, including the Zagreb second index, Harmonic index, Randic index, Sombor index, and others. Linear, quadratic, and logarithmic regression models were applied to explore the relationships between the indices and drug properties. The quadratic regression model provided the best fit, showing the highest R_v values and lowest RMSE, outperforming the other models. A comparison algorithm was added to validate the results, further supporting the superiority of the quadratic model. Various visualizations, including heatmaps, scatter plots, and a bar plot matrix, were created to better understand the correlations.

The results of this study offer valuable insights for drug design and optimization, particularly for Tuberculosis treatments. By identifying the most accurate models for predicting physicochemical properties, this work can guide the development of more effective TB drugs with better therapeutic outcomes. Additionally, leveraging topological indices and advanced regression modeling allows for a deeper understanding of drug properties at a molecular level, enhancing the potential for novel drug discovery and optimization in the fight against TB.

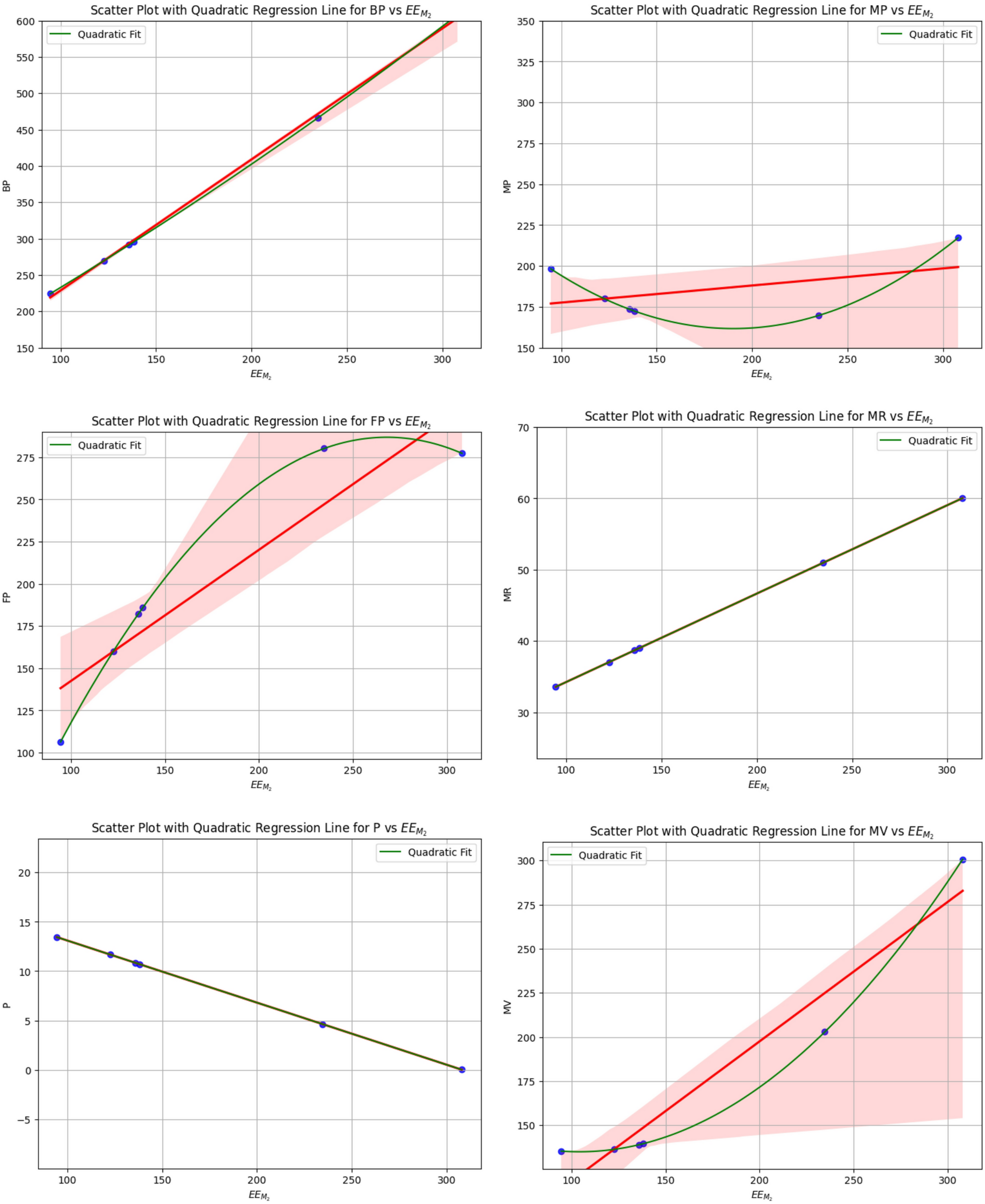


Fig. 19. Quadratic scatter plot between extended energy of M_2 and drug properties.

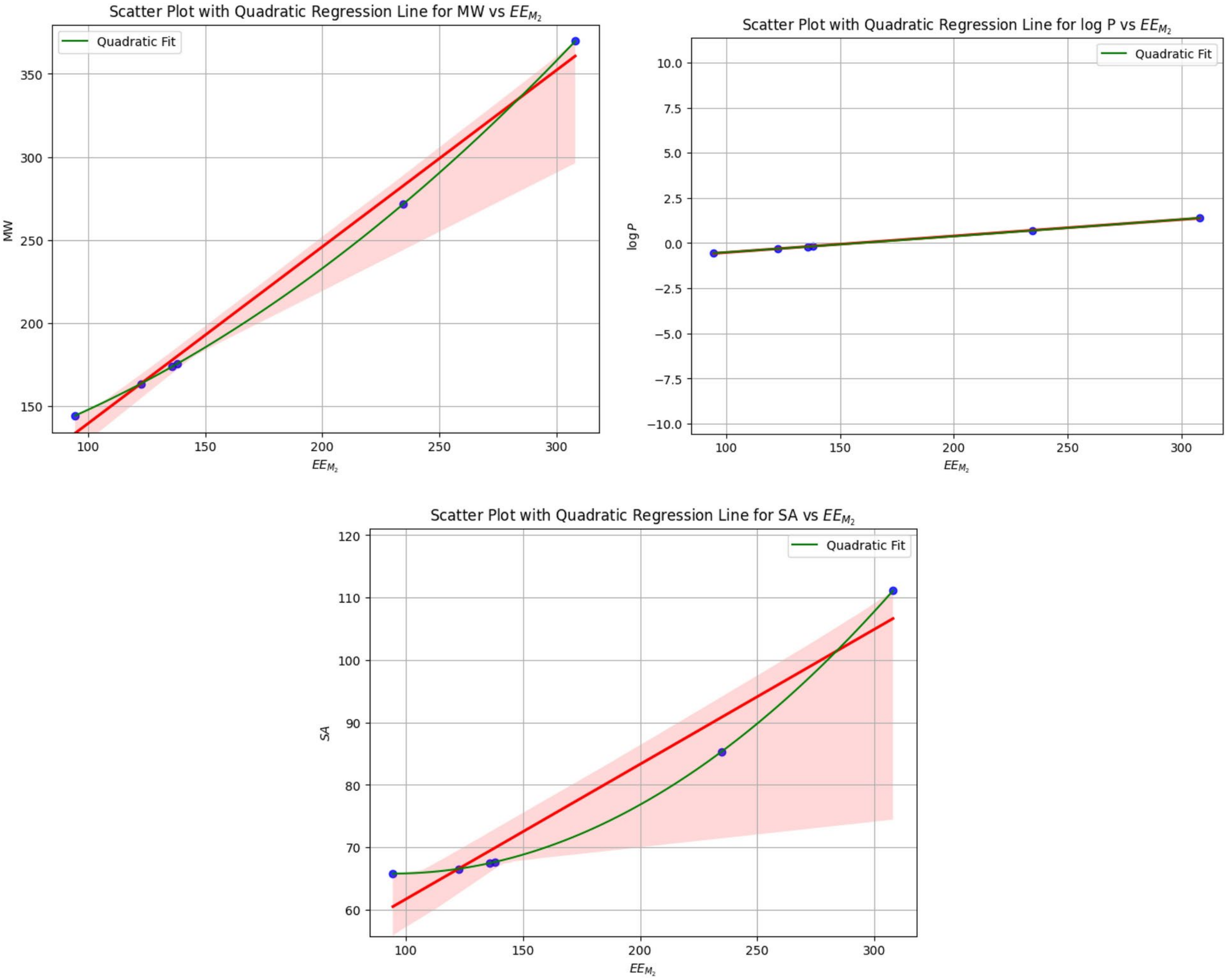


Figure 19. (continued)

Data availability

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

Code availability

The custom Python code used in this study for data analysis and modeling is publicly available in a GitHub repository: github.com/kirannaz145/Linear-Quadratic-Logarithmic. To ensure long-term accessibility and reproducibility, the version of the code referenced in this publication has been archived on Zenodo and can be accessed via the following DOI: https://doi.org/10.5281/zenodo.15240618. This ensures that the code remains accessible even if modifications are made to the GitHub repository in the future. The archived version can be cited and used by other researchers to replicate and extend our findings. No restrictions apply to access or use of the provided code.

Received: 6 February 2025; Accepted: 29 April 2025
Published online: 05 May 2025

References

1. Leite, L. S., Banerjee, S., Wei, Y., Elowitt, J. & Clark, A. E. Modern chemical graph theory. *Wiley Interdiscipl. Rev.* **14**(5), e1729 (2024).
2. Bommahalli Jayaraman, B. & Siddiqui, M. K. Exploring the properties of antituberculosis drugs through QSPR graph models and domination-based topological descriptors. *Sci. Rep.* **14**(1), 24387 (2024).
3. Fernandes, G. F. D. S., Salgado, H. R. N. & Santos, J. L. D. Isoniazid: A review of characteristics, properties and analytical methods. *Crit. Rev. Anal. Chem.* **47**(4), 298–308 (2017).
4. Njire, M. et al. Pyrazinamide resistance in *Mycobacterium tuberculosis*: Review and update. *Adv. Med. Sci.* **61**(1), 63–71 (2016).
5. Feng, X., Ma, Z., Yu, C. & Xin, R. MRNDR: Multihead attention-based recommendation network for drug repurposing. *J. Chem. Inf. Model.* **64**(7), 2654–2669 (2024).
6. Zhou, Y. et al. Dermatophagoides pteronyssinus allergen Der p 22: Cloning, expression, IgE-binding in asthmatic children, and immunogenicity. *Pediatr. Allergy Immunol.* **33**(8), e13835 (2022).

7. Hu, S. et al. Races of small molecule clinical trials for the treatment of COVID-19: An up-to-date comprehensive review. *Drug Dev. Res.* **83**(1), 16–54 (2022).
8. Pu, X., Sheng, S., Fu, Y., Yang, Y. & Xu, G. Construction of circRNA-miRNA-mRNA ceRNA regulatory network and screening of diagnostic targets for tuberculosis. *Ann. Med.* **56**(1), 2416604 (2024).
9. Naz, K., Ahmad, S., Bilal, H. M. & Siddiqui, M. K. Computing degree based topological indices for bulky and normal polymers. *Int. J. Quant. Chem.* **124**(12), e27435 (2024).
10. Ismail, R. et al. Investigating Seidel energies and thermodynamic properties of benzenoid hydrocarbons through regression models. *Sci. Rep.* **15**(1), 867 (2025).
11. Wu, Z., Shangguan, D., Huang, Q. & Wang, Y. Drug metabolism and transport mediated the hepatotoxicity of *Pleuropterus multiflorus* root: A review. *Drug Metab. Rev.* **56**(4), 349–358 (2024).
12. Wang, H. et al. NIR-II AIE luminogen-based erythrocyte-like nanoparticles with granuloma-targeting and self-oxygenation characteristics for combined phototherapy of tuberculosis. *Adv. Mater.* **36**(38), 2406143 (2024).
13. Liu, H., You, L., Tang, Z. & Liu, J. B. On the reduced Sombor index and its applications. *MATCH Commun. Math. Comput. Chem* **86**, 729–753 (2021).
14. Liu, J. B. & Pan, X. F. Asymptotic incidence energy of lattices. *Physica A* **422**, 193–202 (2015).
15. Sarkar, P., Dey, A., Kumar, S. & Pal, A. On some extended energy of graphs and their applications. *Yugoslav J. Oper. Res.* **00**, 40–50 (2024).
16. Milovanovic, I. Z., Milovanovic, E. I. & Zakic, A. A short note on graph energy. *MATCH Commun. Math. Comput. Chem* **72**(1), 179–182 (2014).
17. Li, W. et al. Puerarin-loaded PEG-PE micelles with enhanced anti-apoptotic effect and better pharmacokinetic profile. *Drug Deliv.* **25**(1), 827–837 (2018).
18. Li, H. et al. The effects of ferulic acid on the pharmacokinetics of warfarin in rats after biliary drainage. *Drug Des. Dev. Ther.* **10**, 2173–2180 (2016).
19. Zeng, M. et al. The integration of nanomedicine with traditional Chinese medicine: Drug delivery of natural products and other opportunities. *Mol. Pharm.* **20**(2), 886–904 (2023).
20. Li, H. et al. The effects of warfarin on the pharmacokinetics of Senkyunolide I in a rat model of biliary drainage after administration of Chuanxiong. *Front. Pharmacol.* **9**(1461), d25-35 (2018).
21. Gutman, I. The energy of a graph. *Ber. Math.-Stat. Sekt. Forschungszent. Graz* **103**, 1–22 (1978).
22. Ilić, A. & Stevanović, D. The energy of graphs and matrices. *Linear Algebra Appl.* **431**, 2195–2203 (2010).
23. Das, K. C. & Gutman, I. Some properties of the Laplacian energy of a graph. *MATCH Commun. Math. Comput. Chem.* **52**, 103–112 (2004).
24. Cavers, M., Fallat, S. M. & Kirkland, S. J. On the normalized Laplacian energy and general Randi? index. *Linear Algebra Appl.* **433**, 172–190 (2010).
25. Chellali, M., Kiani, D. & Gutman, I. Recent developments in energy-like graph invariants. *MATCH Commun. Math. Comput. Chem.* **82**, 5–28 (2019).
26. Dehmer, M., Emmert-Streib, F. & Mehler, A. Graph entropy and information functionals for the analysis of complex networks. *Appl. Math. Comput.* **201**, 82–94 (2009).
27. Huang, J. C., Ko, K. M., Shu, M. H. & Hsu, B. M. Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Comput. Appl.* **32**(10), 5461–5469 (2020).
28. Asuero, A. G., Sayago, A. & González, A. G. The correlation coefficient: An overview. *Crit. Rev. Anal. Chem.* **36**(1), 41–59 (2006).
29. Siddiqui, M. K. Exploring the properties of antituberculosis drugs through QSPR graph models and domination-based topological. *Sci. Rep.* **14**, 24387 (2024).

Author contributions

Kiran Naz was responsible for data analysis, computation, and verification of calculations. Hafiz Muhammad Bilal contributed to enhancing the graphical representations using Python and MATLAB. Muhammad Kamran Siddiqui supervised the project, conceptualized and structured the methodology, coordinated the research, secured resources, and drafted the initial version of the paper. Sarfraz Ahmad assisted with computation, data analysis, and reviewing the final draft of the paper. Mustafa Ahmed Ali contributed for Validation, formal analysis of experiments, funding acquisition, and software development. Each author reviewed and approved the final version of the work.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

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

Reprints and permissions information is available at www.nature.com/reprints.

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