# Prediction of Molecular and Physical Properties of Non-small Cell Lung Cancer (NSCLC) Drugs using Mathematical Modelling and M-Polynomial Indices

Faiqa Suleman [1], Aamir Shahzad [2]*, Tasadduq Niaz [1], Muhammad Ali [3], Sidra Ashraf [4]
[1]Faculty of Sciences, Superior University Lahore, Lahore 54000, Pakistan
[2]Department of Mathematics, Faculty of Natural Science and Technology, Baba Guru Nanak University, Nankana Sahib 39100, Pakistan
[3]Department of Pharmacology, College of Pharmacy, University of Sargodha, Sargodha 40100, Pakistan
[4]Department of Mathematics, Faculty of Sciences, University of Sargodha, Sargodha 40100, Pakistan
***Correspondence:** aamir.shahzad@bgnu.edu.pk

**Citation |** Suleman. F, Shahzad. A, Niaz. T, Ali. M, Ashraf. S, "Prediction of Molecular and Physical Properties of Non-small Cell Lung Cancer (NSCLC) Drugs using Mathematical Modeling and M-Polynomial Indices", IJIST, Vol. 07 Issue 04 pp 2900-2912, November 2025
**Received |** October 15, 2025 **Revised |** November 12, 2025 **Accepted |** November 19, 2025 **Published |** November 26, 2025.

The computation of M-Polynomial indices for Erlotinib, a tyrosine kinase receptor inhibitor and the most widely recognized anti-cancer drug for the treatment of patients with NSCLC and advanced pancreatic cancer, is the main focus of this study. In order to efficiently calculate these M-polynomial indices, we used a graph-based method that renders use of the edge partitioning technique based on adjacent matrices and vertex degrees. Using Python software, we applied numerous regression models, such as Linear Regression (LR), Elastic Net Regression (ENR), Lasso Regression (LR), Ridge Regression (RR), and Support Vector Regression (SVR), to develop Quantitative Structure-Property Relationships (QSPR). Based on the M polynomial indices, these models were utilized to forecast the physical properties such as melting point, enthalpy of vaporization, molar refractivity, molar volume, and polarizability, molecular weight, molecular mass, surface area, and chemical hardness of NSCLC medications. According to our research, the M-polynomial indices predict these physical attributes with remarkable accuracy, providing crucial information on structural traits that maximize anticancer effectiveness. Additionally, we suggested predictive models for every physical attribute examined, proving the value of the M-polynomial index in comprehending molecular behaviour and directing the creation of innovative therapeutic medicines. This study not only facilitates the accurate prediction of physical properties for known NSCLC drugs but also holds the potential to accelerate the novel drug discovery and development, uncharacterized anti-cancer compounds, thus contributing to the advancement of cancer therapeutics.

**Keywords:** M Polynomial Indices; Statistical Analysis; NSCLC; QSPR; Physical Properties.

**Introduction:**

Lung cancer continues to pose a major global health challenge due to its high mortality rate and complex biological characteristics. Among its primary subtypes, non-small cell lung cancer (NSCLC) accounts for approximately 85% of all diagnosed cases and frequently necessitates carefully structured and targeted therapeutic interventions. In recent years, mathematical chemistry has emerged as a powerful tool for studying the structural characteristics of chemical compounds. One of the central approaches in this field involves the use of topological indices, which translate molecular structure into numerical values that reflect various physical and chemical attributes. Among these indices, the M- polynomial framework has attracted considerable attention because it provides a generalized representation from which several degree-based topological indices can be derived. These indices often show strong correlations with key physicochemical properties such as boiling point, stability, entropy, molar volume, and biological activity.

Given the increasing complexity of modern anticancer agents, particularly those used in NSCLC treatment, a mathematical modeling approach can offer deeper insights into structure-property relationships. Modeling drug molecules as graphs and deriving their M-polynomials enables the prediction of various physical properties, reducing dependence on purely experimental methods. These models help lower overall research costs while also facilitating early-stage processes such as drug design, candidate screening, and molecular optimization. Compared with traditional experimental or computational approaches, the M-polynomial framework provides a faster and more efficient way to characterize drug molecules. Experimental measurements can be time-consuming and costly, while many computational methods require intensive resources. In contrast, M-polynomial indices capture essential structural information through simple graph-based calculations, enabling rapid property estimation and streamlined analysis [1].

Erlotinib is an oral, low-molecular-weight quinazoline derivative that competes with adenosine triphosphate (APT) for binding in the tyrosine kinase domain of the receptor, thereby selectively and reversibly reducing the tyrosine kinase activity of EGFR [2]. Erlotinib exhibits approximately 93% protein binding following absorption. Its primary metabolic pathway involves CYP3A4-mediated biotransformation via the cytochrome P450 system. The drug has an elimination half-life of 36 hours and is predominantly excreted in faeces. In long-term daily administration, the established maximum tolerated dose of erlotinib is 150 mg per day. Diarrhea and skin rash are the dose-limiting adverse effects. As a single-agent treatment, erlotinib has been shown to have antitumor efficacy in patients with advanced ovarian cancer, head and neck cancer, and non-small cell lung cancer (NSCLC) who have received extensive pretreatment [3]

Erlotinib therapy as a single agent was evaluated in advanced (NSCLC)non-small cell lung cancer patients who had not responded to the best supportive care after one or two standard chemotherapy regimens in a large (731 patient) multicentre randomized phase III clinical trial (BR21 research). These patients were treated for metastatic non-small cell lung cancer (NSCLC) with either one standard chemotherapy regimen (for 50% of patients) or two chemotherapy regimens (50% of patients) [4]. Almost all patients received treatment with a platinum-based medication. The response rate (RR) for the erlotinib group was 8.9%, compared to less than 1% in the placebo group (P ¡0.001). The median response times were 3.7 months and 7.9 months, respectively. The overall survival (OS) for the erlotinib regimen was 6.7 months, compared to 4.7 months for the placebo arm [P ¡ 0.001, hazard ratio (HR) = 0.7]. The progression-free survival (PFS) was 1.8 months for the placebo group and 2.2 months for the erlotinib group (P ¡ 0.001, HR = 0.70). Five percent of patients discontinued erlotinib due to toxicity [5].

The present work focuses on the mathematical modelling of selected NSCLC drug molecules using the M-polynomial technique. This study formulates molecular graph representations, computes the respective M-polynomials, and evaluates their associated topological indices. Furthermore, it investigates how these indices correlate with experimentally reported physical properties. The findings aim to highlight the effectiveness of M-polynomial-based modelling as a predictive tool for anticancer drug analysis and to provide a foundation for future computational drug-design strategies.

**Objective of the Study:**

Accurate prediction of the physical properties of NSCLC drugs—such as Erlotinib—remains a major challenge in the drug development process. Traditional methods for determining topological indices are often slow, error-prone, and computationally demanding. Such constraints are an impediment to the effective design and optimization of anticancer drugs. Hence, an automated and efficient method to calculate M-polynomial indices is necessary so that more reliable and quicker prediction of the physical properties required to predict drug efficacy and development can be made.

**Novelty of the Statement:**

This study presents the application of M-polynomial indices to forecast the physical characteristics of anticancer drugs, namely Erlotinib, which presents a new method of computation.

**Literature Review:**

In 1878, James J. Sylvester [1] coined the term ”graph”. Graph theory, a subfield of math- ematics is fast growing field of the modern era. Furthermore, graph theory has applications in a wide range of fields, including chemistry, statistical mechanics, biology, physics, computer science, optimization theory, and operations research [6]. One of the most important sub- fields in Mathematical Chemistry is Chemical Graph Theory, which was first established by Alexander Balaban [6], Haruo Hosoya [7], Milan Randi [5], and Ante Graovac [8]. Undirected linked molecular graphs' topological indices offer insight into chemical compounds' physicochemical properties and biological activities. QSPR and QSAR are two essential techniques used in cheminformatics to forecast the physicochemical characteristics of molecules. These approaches make a substantial contribution to the study of topological indices [9]. The vertices (atoms) and edges (covalent bonds) of a molecular graph, a topological representation of a molecule, provide a mathematical framework for the analysis of molecular structures. The analysis of molecular characteristics and activities is made possible by this graph-theoretic method [10]. The M-polynomial technique has been used in recent developments in chemical graph theory to investigate a variety of chemical structures. M-polynomials for a variety of indices have been derived by numerous researchers. Zagreb values for infinite dendrimer nanostars and M-polynomials for benzene rings contained in P-type surfaces and polyhex nanotubes are among the initial uses. Additionally, M-polynomial (MPI) forms for certain nanostructures have been widely accepted [11].

The M-polynomial, a recent advancement in polynomial theory, has the potential to transform the fields of degree-based topological indices. This flexible mathematical framework allows precise evaluation of more than ten degree-related indices, thereby creating fresh opportunities for scholarly exploration. Research on the M-polynomial has progressed rapidly in recent years. Notably, the work of Kwun et al. [12] has been instrumental, as they formulated M-polynomial expressions for a range of nanotube structures, demonstrating their significant value in modern scientific investigations.

**Methodology*:***

The methodology section details the procedures used to calculate M-Polynomial Indices (MPI) and to assess their relationship with the molecular and physical characteristics of chemical compounds. Let G = (V, E) be a simple and connected graph, where V represents

the number of vertices and E represents the set of edges. In this graph theory, a vertex corresponds to an individual entity, and an edge signifies a link between a pair of vertices. For any vertex u ∈ V, the degree of u, represented as du, refers to the number of edges incident to u; equivalently, it is the number of immediate neighbors of that vertex. The degree of a vertex is a fundamental parameter used to study the structural properties of a graph.
Following Kwun [12], the M-polynomial of the graph G is illustrated as:

$$M(G; x, y) = \sum_{p \le q} |N(p, q)| x^p y^q$$

where |N (p, q)| denotes the count of all edges uv ∈ E for which the endpoint degrees meet the condition (du, dv) = (p, q) with p ≤ q. In other words, every edges are recorded according to the degrees of its adjacent vertices, and the expression sums over all such degree-classified edges in the graph. The symbols x and y serve as formal variables that represent this degree-dependent edge distribution. Wiener [13] introduced the number of paths as the earliest topological index in 1947. The Wiener index has numerous uses within the field of chemistry. Subsequently, Milan Randic put forward the idea of what is now known as the Randic index [5]:

$$R_{-1/2}(G) = \sum_{uv \in E} \sqrt{\frac{1}{d_u d_v}}$$

Bollobs and Amic et al [14] developed the inverse and general Randi index, defined as:

$$GR_\alpha(G) = \sum_{uv \in E} (d_u d_v)^\alpha, \quad R_\alpha(G) = \sum_{uv \in E} \frac{1}{(d_u d_v)^\alpha}$$

Nikolic proposed a modified version of the M2 index as mM2(G) and defined it as:

$$mM_2(G) = \sum_{uv \in E} \frac{1}{d_u d_v}$$

In 2011, Fath-Tabar introduced the concept of the M3 index:

$$M_3(G) = \sum_{uv \in E} |d_v - d_u|$$

The Symmetric Division (SDD) index and Augmented Zagreb (AZI) index are defined as:

$$SDD(G) = \sum_{uv \in E} \frac{\max(d_v, d_u)}{\min(d_v, d_u)} + \frac{\min(d_v, d_u)}{\max(d_v, d_u)}$$

$$AZI(G) = \sum_{uv \in E} \left( \frac{d_v d_u}{d_v + d_u - 2} \right)^3$$

The inverse sum I index was analyzed as a fundamental characteristic of octane and is precisely described as:

$$I(G) = \sum_{uv \in E} \frac{d_v d_u}{d_v + d_u}$$

The Harmonic index *H* is defined as:

$$H(G) = \sum_{uv \in E} \frac{2}{d_v + d_u}$$

Many polynomials have been proposed, including Tutte, matching, Schultz, Hosoya [7], and Zhang-Zhang polynomials. This study focuses on the M-polynomial and shows how it may be used to calculate degree-based indices. The Hosoya polynomial for distance-based indices is similar to the *function*. The M-polynomial is derived as follows, based on the works of Munir et al. [15]:

$$M(G; x, y) = p(x, y), \quad D_x = x\frac{\partial p(x, y)}{\partial x}, \quad D_y = y\frac{\partial p(x, y)}{\partial y}$$

$$\int_x \qquad \int_y$$

$$I_x = \int_0 p(t, y)\, t\, dt, \quad I_y = \int_0 p(x, t)\, t\, dt$$

$$J(p(x, y)) = p(x, x), \quad Q_\alpha(p(x, y)) = x^\alpha p(x, y)$$

This shows the mathematical form of the M-polynomial indices. Moreover, it highlights the utility of the M-polynomial in deriving degree-based indices and provides a comprehensive framework for using this method in chemical graph theory.

**Computation of M-Polynomial:**

To begin, we evaluated the M-polynomial indices for the anticancer medication erlotinib in order to examine their usefulness in forecasting its physical characteristics. The process is described in the steps below.

The molecular configuration of erlotinib is transformed into a graph model, in which atoms are nodes and chemical bonds are represented as edges.

The edges and nodes of this graph are categorized according to their respective vertex degrees.

By utilizing the edge distribution derived from vertex degrees, the corresponding M-polynomial(MP) is formulated.

The computed M-polynomial indices are illustrated through graphical plots generated using the software MATLAB.

**Algorithm for M-Polynomial Indices Computation:**

We used a Python-written algorithm to compute M-polynomial indices of a given molecular graph in an automated manner. The algorithm used the graph's adjacency matrix, which was made with Python. The degree of every vertex was determined by an algorithm, and the M-polygon was built by counting how many edges there were between the vertices with different degrees. M-polynomial indices were calculated for 13 NSCLC drugs. The molecular SMILES representations and physicochemical properties were retrieved from PubChem and ChemSpider [5]. Any entries with missing or incomplete values were removed prior to analysis to ensure data consistency. Statistical analysis of M-polynomial indices. To evaluate the utility of M-polynomial indices as descriptors of molecular structure, a statistical analysis was performed on a selected subset of NSCLC drugs. Specifically, only a defined set of NSCLC medications was included, and for each compound, the molecular graph of its chemical structure was generated using the same methodology previously applied to Erlotinib. The adjacency matrix for each molecular graph was computed using a Python-based approach, utilizing the RDKit library for molecular structure processing and Network X for graph representation. The indices of the M-polymorphic of every drug were determined by the Python-based algorithm that was specifically created to perform this calculation. Such physical properties of these drugs as molecular weight (MW), boiling point (BP), melting point (MP), and solubility were obtained in favourable open public databases, such as ChemSpider and PubChem. All computational steps, including adjacency matrix construction, symbolic M-polynomial formulation, derivative-based index extraction, and statistical correlation analysis, were executed using Python. A fully documented version of the Python code used in this study is provided in Supplementary File S1, enabling complete reproducibility of the methodology.

The objective of our statistical modelling is to identify the correlation between M-polynomial indices and physical properties through the application of various machine

learning regression models, including Linear regression (LR), Ridge regression (RR), Lasso regression (LR), Elastic Net regression (ER), and Support Vector regression (SVR). The models are measured by applying common evaluation measures, including the Mean Squared Error (MSE) and the coefficient of determination (R2). It is a unique methodological strategy that creates new molecular descriptors and empirical validation of these descriptors using statistical modeling. Linear Regression, Ridge Regression, Lasso Regression, ElasticNet Regression, and Support Vector Regression (SVR) were chosen for a better outcome due to their complementary strengths and capacity to capture the relationship between M-polynomial indices and physicochemical properties. The Linear Regression is the simple interpretive baseline model, whereas Ridge and Lasso help manage multi-collinearity. The ElasticNet method is useful when the predictors are correlated, and we want variable selection. SVR was added in order to capture non-linear patterns that cannot usually be captured by linear models [8].

**Results:**

This study computes the degree-based M-polynomial indices of Erlotinib. We utilize the edge partitioning method to compute the indices. A molecular graph is being constructed from the chemical structure of Erlotinib for the purpose of edge partitioning.

**Theorem 1.** Let G be the molecular graph of erlotinib. Then the M-polynomial is given by:

$$M(G_t; x, y) = t(3xy^2 + 10x^2y^2 + 15x^2y^3 + 3x^3y^3) + (x^2y + x^2y^3).$$

The chemical structure and molecular graph of Erlotinib are shown in the figure-1,2

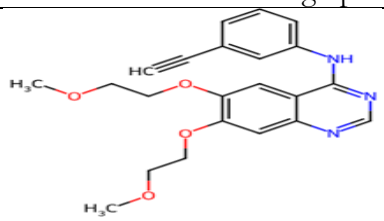


**Figure 1.** Chemical structure.

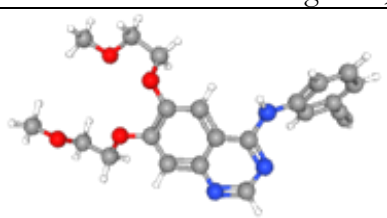

**Figure 2.** Molecular graph.

**Proof.** A molecular graph (Gt) is a graphical representation of this molecule in which chemical bonds are shown by edges and atoms by vertices. The number of chemical bonds that are attached to an atom is known as the degree (du) of a vertex (u) in Gt. Using common valency standards, this information is obtained directly from the molecule structure. Let Gt be the molecular graph corresponding to the M-polynomial.

$$M(G_t; x, y) = t(3xy^2 + 10x^2y^2 + 15x^2y^3 + 3x^3y^3) + (x^2y + x^2y^3).$$

We analyzed the vertex degrees and edge distribution as follows:

Based on the molecular connectivity implied by M (Gt; x, y), the vertex degrees were distributed as:

Degree 1 vertices: correspond to edges with $x_2$ terms in $M(G_t)$,

Degree 2 vertices: appear in edges with $x$

Degree 3 vertices: appear in edges with $x$

Degree 4 vertices: appear in edges with $x$

We define the edge sets:

terms,

$$N(p, q) = \{uv \in E(G_t) \mid d_u = p, d_v = q, p \leq q\}.$$

Then the edge counts are:

$|N(1, 2)| = 3t,$

$|N(2, 2)| = 10t,$

$|N(2, 3)| = 15t,$

$|N(3,3)| = 3t,$
$|N(2,1)|_{\text{single}} = 1,$
$|N(2,3)|_{\text{single}} = 1.$
Total Vertices $= 13t + 63$ Total Edges $= 31t + 2$
For $t = 1$:
Number of Vertices: $13(1) + 63 = 76$
Number of Edges: $31(1) + 2 = 33$ By definition, the M-polynomial is

$$M(G_t; x, y) = \sum_{p \le q} |N(p,q)| x^p y^q.$$

Substituting the edge counts gives

$$M(G_t; x, y) = 3txy^2 + 10tx^2y^2 + 15tx^2y^3 + 3tx^3y^3 + x^2y + x^2y^3 = t(3xy^2 + 10x^2y^2 + 15x^2y^3 + 3x^3y^3) + (x^2y + x^2y^3),$$

which exactly matches the given polynomial.

A molecular graph (G) is a molecular representation where atoms are shown by vertices and chemical bonds by edges. The number of chemical bonds (edges) that are attached to an atom (vertex) in H is its degree ($d_e$).

**Theorem 2.** Let $G_t$ be the graph of Erlotinib and the M-polynomial for the Erlotinib with a repeating factor t is:

$$M(G_t; x, y) = t(3xy^2 + 10x^2y^2 + 15x^2y^3 + 3x^3y^3) + (x^2y + x^2y^3)$$

Then

**Table 1.** Analytical expressions of selected degree-based topological indices as functions of the parameter t.

| Index | Value |
|---|---|
| First Zagreb index ($M_1$) | 142 t + 8 |
| Second Zagreb index ($M_2$) | 163 t + 8 |
| Modified second Zagreb index ($MM_2$) | 955 t + 40 |
| Forgotten index (F) | 874t + 44 |
| Redefine the third Zagreb index (ReZG3) | 3244t + 152 |
| Harmonic index (HM) | 69t + 2.33 |
| Symmetric division index SDD | 66t + 4.6667 |
| Inverse sum index (I) | 7t + 0.533 |

*Proof.* For a molecular graph $G_t$ with M-polynomial $M(G_t; x, y)$:

$$M(G) = \left(x\frac{\partial M}{\partial x} + y\frac{\partial M}{\partial y}\right)_{x=y}, \quad M(G) = \left(x\frac{\partial M}{\partial x} \cdot y\frac{\partial M}{\partial y}\right)_{x=y}$$

**Step-by-Step Calculation:**

**Table 2.** Coefficients and corresponding exponent values (p,q)(p, q)(p,q) of the polynomial terms in variables xxx and yyy

| Term | c_pq | p | q |
|---|---|---|---|
| 3txy^2 | 3t | 1 | 2 |
| 10tx^2 y^2 | 10t | 2 | 2 |
| 15tx^2 y^3 | 15t | 2 | 3 |
| 3tx^3 y^3 | 3t | 3 | 3 |
| x^2 y | 1 | 2 | 1 |
| x^2 y^3 | 1 | 2 | 3 |

**First Zagreb Index ($M_1$):**

(M_1 (G_t)&=3t(1+2)+10t(2+2)+15t(2+3)+3t(3+3)+1(2+1)+1(2+3)@&=142t+8)

**Second Zagreb Index ($M_2$):**

M_2 (G_t)=3t(1·2)+10t(2·2)+15t(2·3)+3t(3·3)+1(2·1)+1(2·3)=163t+8

**Modified Second Zagreb Index (MM2):**

MM2(G_t)=3t(1·2)^2+10t(2·2)^2+15t(2·3)^2+3t(3·3)^2+1(2·1)^2+1(2·3)^2=955t+40

**Forgotten Index (F):**

F(G_t)=∑c_pq (p^3+q^3)=874t+44

**Redefined Third Zagreb Index (ReZG3):**

ReZG3(G_t)=∑c_pq (p+q)^3=3244t+152

**Harmonic Index (HM):**

HM(G_t )=∑c_pq 2pq/(p+q)=69t+2.33

**Symmetric Division Degree Index (SDD):**

SDD(G_t)=∑c_pq (p^2+q^2)/(p·q)=66t+4.6667

**Inverse Sum Index (I):**

I(G_t)=∑c_pq 1/(p+q)=7t+0.533

**Statistical Analysis:**

In order to assess the predictive contribution of the M-polynomial indices, we evaluated each model using the coefficient of determination (R), Mean Squared Error (MSE), and Pearson correlation (r), where higher R and r values, together with lower MSE, indicated indices with stronger predictive relevance for the physicochemical properties.

**Table 3.** Statistical analysis for Boiling Point (BP)

| Model | R | Mean Squared Error | Pearson R | Property |
|---|---|---|---|---|
| Linear Regression | -5E+15 | 8.93E+16 | 0.97736 | BP |
| Ridge Regression | -512737 | 9237256 | 0.96858 | BP |
| Lasso Regression | -1125.32 | 20291.24 | 0.79718 | BP |
| ElasticNet Regression | -1806.66 | 32566 | 0.9742 | BP |
| SVR | -125.192 | 2273.418 | 0.743 | BP |

As shown in Table 1, the regression models exhibited varying performance in predicting the target variable. Linear Regression demonstrated the poorest fit, with a very low $R^2$ and a high mean squared error (MSE). In contrast, Ridge Regression performed most effectively, showing a comparatively strong correlation. Lasso Regression, however, yielded suboptimal results, characterized by low $R^2$ and high MSE values. ElasticNet did not work well with a negative correlation. Although it had a negative correlation, SVR reduces error to a minimal level, which means that it gave the highest accuracy of the prediction in this instance.

**Table 4.** Statistical analysis for Enthalpy of Vaporization (EoV)

| Model | R | Mean Squared Error | Pearson R | Property |
|---|---|---|---|---|
| Linear Regression | -6.8E+15 | 1.93E+15 | 0.887356 | EoV |
| Ridge Regression | -199351 | 56261.49 | 0.853831 | EoV |
| Lasso Regression | -11418 | 3222.689 | -0.96068 | EoV |
| ElasticNet Regression | -13794.3 | 3893.336 | -0.9556 | EoV |
| SVR | -78.177 | 22.3455 | -0.89279 | EoV |

In the table 2, the Enthalpy of Vaporization (EoV) regression models demonstrated different performance. The R2, MSE, and Pearson R of Linear Regression were -6.8E+15, 1.93E +15, 0.887356, respectively, meaning that it was fitting with an average correlation. Ridge Regression had a weaker correlation with an R2 of -199351, a MSE of 56261.49, and a Pearson R of 0.853831. The R2 of Lasso Regression was -11418, MSE = 3222.689, and the correlation was strong (Pearson = -0.96068). Similar results were also observed with ElasticNet Regression with R2 of -13794.3, MSE of 3893.336, and Pearson R of -0.9556. The MSE of SVR was lowest (22.3455) with a Pearson R of -

0.89279, giving the best predictions, but with a negative correlation.

**Table 5.** Statistical analysis for Flash Point (FP)

| Model | R | Mean Squared Error | Pearson R | Property |
|---|---|---|---|---|
| Linear Regression | -1.4E+16 | 2.51E+17 | 0.085324 | FP |
| Ridge Regression | -235452 | 4316773 | 0.062658 | FP |
| Lasso Regression | -1.8E+07 | 3.25E+08 | 0.082239 | FP |
| ElasticNet Regression | -59914.5 | 1098484 | 0.024401 | FP |
| SVR | -400.02 | 7352.253 | -0.09192 | FP |

The regression models for predicting Flash Point (FP) exhibited notably poor performance (Table 3). The $R^2$ value was $-1.4 \times 10^{16}$, the mean squared error (MSE) was $2.51 \times 10^{17}$, and the Pearson correlation coefficient (R) was 0.0853, indicating a very poor fit and only a weak positive relationship between the predicted and observed values. Ridge Regression performed slightly better, with an $R^2$ of −235,452, an MSE of 4,316,773, and a Pearson correlation coefficient (R) of 0.0627; however, it still demonstrated very limited predictive capability Lasso Regression exhibited an $R^2$ of $-1.8 \times 10^7$, an MSE of $3.25 \times 10^8$, and a Pearson correlation coefficient (R) of 0.0822, indicating low predictive power. Similarly, ElasticNet Regression performed poorly, with an $R^2$ of −59,914.5, an MSE of 1,098,484, and a very low Pearson R of 0.0244, reflecting minimal predictive capability. The lowest MSE of 7352.253 was obtained using SVR, although the Pearson R of -0.09192 was negative, which proves the least accurate prediction with a weak negative relationship.

**Table 6.** Statistical analysis for Molar Refractivity (MR)

| Model | R | Mean Squared Error | Pearson R | Property |
|---|---|---|---|---|
| Linear Regression | -2.1E+13 | 3.87E+15 | -0.244 | MR |
| Ridge Regression | -6886.94 | 1242140 | -0.2288 | MR |
| Lasso Regression | -1281.65 | 231306.6 | -0.2143 | MR |
| ElasticNet Regression | -92.9515 | 16942.8 | -0.1378 | MR |
| SVR | -0.26862 | 228.777 | 0.96557 | MR |

Regression equations of the MR demonstrated different results (Table-4). The $R^2$ of Linear Regression was -2.1E +13, MSE was 3.87E +15, and the Pearson R was -0.244, which showed that it was poorly fitted and was negatively correlated. Ridge Regression was a little higher with $R^2$ = -6886.94, MSE = 1242140, and Pearson = -0.2288, which indicates a slightly weaker negative correlation. The $R^2$ of Lasso Regression was -1281.65, MSE was 231306.6, and Pearson R was -0.2143, and the predictive ability was equally weak. ElasticNet Regression performed poorly with an $R^2$ = -92.9515, MSE = 16942.8, and Pearson R = -0.1378. The highest predictive accuracy and high positive Pearson R of 0.96557 showed that SVR had the best performance with an R 2 of -0.26862, MSE of 228.777, and positive Pearson R. □

**Table 7.** Statistical analysis for Molar Volume (MV)

| Model | R | Mean Squared Error | Pearson R | Property |
|---|---|---|---|---|
| Linear Regression | -4.6E+13 | 5.38E+16 | -0.68549 | MV |
| Ridge Regression | -1251.32 | 1474696 | 0.730584 | MV |
| Lasso Regression | -3.55675 | 5365.921 | 0.997262 | MV |
| ElasticNet Regression | -26.4794 | 32359.11 | 0.926087 | MV |
| SVR | -0.32559 | 1560.986 | 0.472211 | MV |

**Discussion:**

This section highlights the significant aspects and defining characteristics of the proposed model.

**Regression models:**

**Table 8.** Regression Models

| Model | R | Mean Squared Error | Pearson R | Property |
|---|---|---|---|---|
| Linear Regression | -2.2E+12 | 8.95E+15 | 0.518305 | PSA |
| Ridge Regression | -1267.1 | 5088759 | -0.50099 | PSA |
| Lasso Regression | 430.822 | 1732861 | -0.50188 | PSA |
| ElasticNet Regression | -62.0386 | 252968 | -0.47456 | PSA |
| SVR | -0.51494 | 6079.316 | 0.995896 | PSA |
| Linear Regression | -1.4E+13 | 2.81E+14 | -0.15554 | HAC |
| Ridge Regression | -3462.83 | 71585.87 | -0.13522 | HAC |
| Lasso Regression | -5.95335 | 143.7027 | 0.553452 | HAC |
| ElasticNet Regression | -9.89884 | 225.2428 | 0.456641 | HAC |
| SVR | 0.052498 | 19.58171 | 0.938189 | HAC |
| Linear Regression | -9.5E+12 | 1.31E+17 | 0.148763 | C |
| Ridge Regression | -218.687 | 3020010 | 0.233267 | C |
| Lasso Regression | -67.373 | 939916.4 | 0.231159 | C |
| ElasticNet Regression | -6.3501 | 101041.1 | 0.278722 | C |
| SVR | -1.53429 | 34838.66 | 0.984829 | C |

The regression equations of the MV property demonstrated different performance (Table-6). Linear Regression did not fit well with a low R2 of -4.6E +13 and a large MSE of 5.38E +16, and a low negative correlation (Pearson R = -0.68549). Ridge Regression was more successful, as it has an R2 of -1251.32, MSE of 1474696, and a positive correlation is moderate (Pearson R = 0.730584). The overall results of Lasso Regression were the best with an R-Sq = -3.55675, a low MSE = 5365.921, and a very good positive correlation (Pearson R = 0.997262), showing that it is very accurate. Elastic Net Regression was also a good model with an R 2 of -26.4794, MSE of 32359.11, and Pearson R of 0.926087. SVR has the lowest MSE (1560.986) but a lower Pearson R of 0.472211, which means that it has a lower predictive power. Overall, Lasso Regression demonstrated the most effective performance with this dataset. The regression models for the three properties—Polar Surface Area (PSA), Heavy Atom Count (HAC), and Complexity (C)—exhibited varying predictive capabilities (Table 7). For PSA, Linear Regression showed a moderate positive correlation (Pearson R = 0.5183); however, its predictive power was limited, as reflected by an $R^2$ of $-2.2 \times 10^{12}$ and an MSE of $8.95 \times 10^{15}$. Ridge Regression and Lasso Regression were weakly negatively correlated, with the latter being a little worse. ElasticNet Regression had the least R 2 of -62.0386 and a negative correlation. The most successful model was the SVR, which has the highest predictive power with the minimum possible error (MSE = 6079.316) and a high positive correlation (Pearson R = 0.995896). In the case of HAC, Linear Regression and Ridge Regression failed to perform well, and their negative correlation was poor. Lasso Regression had a stronger positive correlation (0.553452) but had a weak predictive power. The highest Pearson R of 0.938189 and a low MSE of 19.58171 prove that SVR gave the best results, and it implies that it has good predictive power. In the instance of Complexity (C), the performance of all the models except SVR was weak. The highest performance of SVR is reported as a high Pearson R of 0.984829 and low MSE (34838.66), which shows that it is effective in the prediction of this property. Generally, SVR has always demonstrated high predictive accuracy on all properties, especially PSA and HAC, with high correlations and minimum MSE.

**Table 9.** Regression Equations for Boiling Point (BP)

| Model | Equation |
|---|---|
| Linear Regression | $BP = 697.37 + (18465.7168)\mathrm{AZI} + (429962.1793)M_1 + (274244.4798)M_2 + (-1448551.9618)mM_2 + (1497491.6541)H + (-11504.1322)\mathrm{ReZG_3} + (-82989.8275)\mathrm{SDD} + (-2590690.9016)I + (-51297.7970)F$ |
| Ridge Regression | $BP = 324.56 + (-8.0766)\mathrm{AZI} + (19.3593)M_1 + (5.4538)M_2 + (-13.1955)mM_2 + (-4.3436)H + (2.1345)\mathrm{ReZG_3} + (20.6202)\mathrm{SDD} + (8.1577)I + (-13.1321)F$ |
| Lasso Regression | $BP = 312.05 + (-1.0387)\mathrm{AZI} + (7.9282)M_1 + (0.6396)M_2 + (0.0000)mM_2 + (-34.5877)H + (-0.1980)\mathrm{ReZG_3} + (4.1902)\mathrm{SDD} + (10.1673)I + (-1.9813)F$ |
| ElasticNet Regression | $BP = 307.83 + (-2.0348)\mathrm{AZI} + (7.9989)M_1 + (1.4271)M_2 + (-1.1194)mM_2 + (-6.3037)H + (-0.2996)\mathrm{ReZG_3} + (-0.8104)\mathrm{SDD} + (6.0451)I + (-1.1382)F$ |

**Heat Map:**

The correlation heatmap (Figure 3) shows linear relationships between the topological indices and the studied physicochemical properties. Most of the indices showed a positive correlation with physicochemical variable indicating that the descriptors of M-polynomial represent structural properties that affect molecular performance. AZI, $M_1$, $M_2$, $ReZG_3$, I, F, SDD, and HAC were mutually correlated with $r \geq 0.90$. It indicates that these descriptors were coded for similar structural information. Many of the physical properties also correlated well with these indices ($r > 0.85$). On the other hand, $mM_2$ was the only descriptor that consistently had either weak or negative correlations with the other indices and with the physicochemical properties: correlation coefficients between approximately −0.45 and 0.10. The pattern suggests that $mM_2$ captured a structural feature that has not been tightly related to the measured traits.

The boiling point (BP), enthalpy of vaporization (EoV), molar refractivity (MR), molar volume (MV), and polarizability (P) of the compounds are, however, almost all correlated ($r \geq 0.90$), being dependent on size and electron delocalization. These properties also exhibited some connection with the main group of topological indices, with values generally varying from $r = 0.85$ to 0.98. Chemical hardness (C) and HAC correlate well with most indices ($r \geq 0.90$), indicating that the chemical hardness (C) and the HAC were influenced by structural features represented by the descriptors.

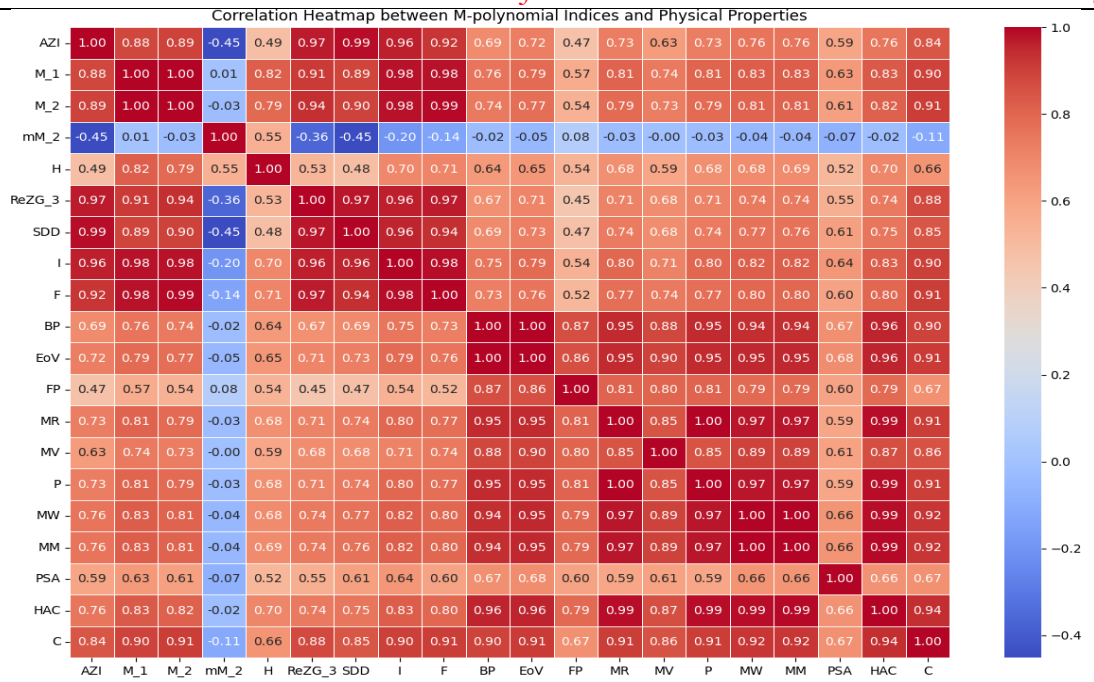


**Figure 3.** Heatmap of all variables in the dataset.

## Conclusion:

This study used edge-partitioning based on vertex degrees and adjacency matrices to successfully calculate the M-polynomial indices of erlotinib. Computational efficiency was greatly increased by a specially created Python script, which minimized human mistakes and cut processing times from days to minutes. The combination of machine learning models and graph- based indexes shows an effective approach for expediting the drug development process. The results set the stage for further research in computational drug design, namely in the creation of novel therapeutic molecules to combat NSCLC.

## References:

[1] J. J. Sylvester, "On an Application of the New Atomic Theory to the Graphical Representation of the Invariants and Covariants of Binary Quantics, with Three Appendices," *Am. J. Math.*, vol. 1, no. 1, p. 64, 1878, doi: 10.2307/2369436.

[2] A. A. Abdelgalil, H. M. Al-Kahtani, and F. I. Al-Jenoobi, "Erlotinib," *Profiles Drug Subst. Excipients Relat. Methodol.*, vol. 45, pp. 93–117, Jan. 2020, doi: 10.1016/BS.PODRM.2019.10.004.

[3] M. A. Bareschino, C. Schettino, T. Troiani, E. Martinelli, F. Morgillo, and F. Ciardiello, "Erlotinib in cancer treatment," *New Trends Clin. Oncol. - 9th Natl. GOIM Congr. 17-19 June 2007, Potenza, Italy*, vol. 18, pp. vi35–vi41, 2007, doi: 10.1093/annonc/mdm222.

[4] U. Smrdel and V. Kovač, "Erlotinib in previously treated non-small-cell lung cancer," *Radiol. Oncol.*, vol. 40, no. 1, Jul. 2006, doi: 10.1056/NEJMOA050753;PAGEGROUP:STRING:PUBLICATION.

[5] "On history of the Randić index and emerging hostility toward chemical graph theory | Request PDF." Accessed: Dec. 09, 2025. [Online]. Available: https://www.researchgate.net/publication/267081398_On_history_of_the_Randic_index_and_emerging_hostility_toward_chemical_graph_theory

[6] E. Mohyedinbonab, M. Jamshidi, and Y. F. Jin, "A Review on Applications of Graph Theory in Network Analysis of Biological Processes," *Int. J. Intell. Comput. Med. Sci. Image Process.*, vol. 6, no. 1, pp. 27–43, 2014, doi: 10.1080/1931308X.2014.938492;WGROUP:STRING:PUBLICATION.

[7] H. Hosoya, "Topological Index. A Newly Proposed Quantity Characterizing the

Topological Nature of Structural Isomers of Saturated Hydrocarbons," *Bull. Chem. Soc. Jpn.*, vol. 44, no. 9, pp. 2332–2339, Sep. 1971, doi: 10.1246/BCSJ.44.2332.

[8] A. Graovac, I. Gutman, and N. Trinajstić, "Topological Approach to the Chemistry of Conjugated Molecules," 1977, Accessed: Dec. 09, 2025. [Online]. Available: https://books.google.com/books/about/Topological_Approach_to_the_Chemistry_of.html?id=Hq7rCAAAQBAJ

[9] H. Parastar and R. Tauler, "Big (Bio)Chemical Data Mining Using Chemometric Methods: A Need for Chemists," *Angew. Chemie*, vol. 134, no. 44, p. e201801134, Nov. 2022, doi: 10.1002/ANGE.201801134.

[10] S. Sorgun and K. Birgin, "Vertex-Edge-Weighted Molecular Graphs: A Study on Topological Indices and Their Relevance to Physicochemical Properties of Drugs Used in Cancer Treatment," *J. Chem. Inf. Model.*, vol. 65, no. 4, pp. 2093–2106, Feb. 2025, doi: 10.1021/ACS.JCIM.4C02013.

[11] H. Saeidi, H. Hassani, M. S. Dahaghin, and S. Mehrabi, "An optimal solution of lung cancer mathematical model using generalized Bessel polynomials," *Phys. Scr.*, vol. 99, no. 12, p. 125269, Nov. 2024, doi: 10.1088/1402-4896/AD9095.

[12] Y. C. Kwun, M. Munir, W. Nazeer, S. Rafique, and S. M. Kang, "M-Polynomials and topological indices of V-Phenylenic Nanotubes and Nanotori," *Sci. Reports 2017 71*, vol. 7, no. 1, pp. 8756-, Aug. 2017, doi: 10.1038/s41598-017-08309-y.

[13] H. Wiener, "Structural Determination of Paraffin Boiling Points," *J. Am. Chem. Soc.*, vol. 69, no. 1, pp. 17–20, 2002, doi: 10.1021/JA01193A005.

[14] D. Amić, D. Bešlo, B. Lučić, S. Nikolić, and N. Trinajstić, "The Vertex-Connectivity Index Revisited," *J. Chem. Inf. Comput. Sci.*, vol. 38, no. 5, pp. 819–822, 1998, doi: 10.1021/CI980039B.

[15] M. Munir, W. Nazeer, S. Rafique, and S. M. Kang, "M-Polynomial and Degree-Based Topological Indices of Polyhex Nanotubes," *Symmetry 2016, Vol. 8, Page 149*, vol. 8, no. 12, p. 149, Dec. 2016, doi: 10.3390/SYM8120149.