

Article

# Exploring the Symmetry of Curvilinear Regression Models for Enhancing the Analysis of Fibrates Drug Activity through Molecular Descriptors

Suha Wazzan <sup>1,\*</sup>  and Nurten Urlu Ozalan <sup>2,†</sup><sup>1</sup> Department of Mathematics, Science Faculty, King Abdulaziz University, Jeddah 21589, Saudi Arabia<sup>2</sup> Faculty of Engineering and Natural Science, KTO Karatay University, Konya 42020, Turkey; nurten.ozalan@karatay.edu.tr

\* Correspondence: swazzan@kau.edu.sa

† These authors contributed equally to this work.

**Abstract:** Quantitative structure-property relationship (QSPR) modeling is crucial in cheminformatics and computational drug discovery for predicting the activity of compounds. Topological indices are a popular molecular descriptor in QSPR modeling due to their ability to concisely capture the structural and electronic properties of molecules. Here, we investigate the use of curvilinear regression models to analyze fibrates drug activity through topological indices, which modulate lipid metabolism and improve the lipid profile. Our QSPR approach predicts the physicochemical properties of fibrates based on degrees and distances from topological indices. Our results demonstrate that topological indices can enhance the accuracy of predicting physicochemical properties and biological activities of molecules, including drugs. We also conducted density functional theory (DFT) calculations on the investigated derivatives to gain insights into their optimized geometries and electronic properties, including symmetry. The use of topological indices in QSPR modeling, which considers the symmetry of molecules, shows significant potential in improving our understanding of the structural and electronic properties of compounds.

**Keywords:** topological indices; fibrates; curvilinear regression; QSPR analysis



**Citation:** Wazzan, S.; Urlu Ozalan, N. Exploring the Symmetry of Curvilinear Regression Models for Enhancing the Analysis of Fibrates Drug Activity through Molecular Descriptors. *Symmetry* **2023**, *15*, 1160. <https://doi.org/10.3390/sym15061160>

Academic Editors: Michel Planat and Ismail Naci Cangul

Received: 27 April 2023

Revised: 20 May 2023

Accepted: 25 May 2023

Published: 27 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Pharmacology has rapidly evolved, resulting in the introduction of numerous groundbreaking drugs each year. However, ensuring accurate testing performance requires the availability of appropriate equipment, a good rapport, and sufficient resources. Previous studies have shown that a drug's chemical properties are intricately linked to its molecular structure. Pharmacological and medical researchers often utilize topological indices to examine the molecules' properties and understand their impact on experimental outcomes. Hence, the topological index computation method is a useful tool for developing countries, allowing them to gather medical and biological data on upcoming drugs without the need for laboratory tests; see for example [1–4].

Fibrates are a type of medication that have been shown to lower high levels of bad cholesterol (also known as low-density lipoprotein or LDL), increase good cholesterol (also known as high-density lipoprotein or HDL), and decrease the amount of small dense LDL particles in the blood. They have been found to be effective in reducing the mortality and morbidity associated with cardiovascular disease (CVD) in individuals who are at risk for developing it. However, conducting laboratory studies to investigate the physicochemical properties of fibrates can be both expensive and time-consuming. To overcome this challenge, chemists can use topological indices to derive mathematical equations that provide valuable insights into the properties of fibrates. For more information on fibrates, please refer to sources [5,6].

Chemical graph theory is a field that integrates mathematical modeling of chemical phenomena with graph theory. It utilizes topological indices to establish a correlation between the properties of a chemical molecule and its structure [7]. These indices are also known as graph invariants or graph-based molecular descriptors, and they quantify the topological features of a molecule or molecules [8]. The application of quantitative structure-property/structure-activity relationship QSPR/QSAR models, which are commonly employed in this field, allows for the prediction of molecular properties using these topological indices. In 1947, Harold Wiener introduced the Wiener index, the first topological index; paraffin's physical properties were determined using it [9].

Topological indices, which are numerical values derived from the molecular graph of a chemical compound, have been extensively studied in the fields of QSPR/QSAR analysis. These indices encode the structural and topological information of molecules and have proven useful in predicting various physical, chemical, and biological properties [10–14]. The use of molecular graphs to represent unsaturated hydrocarbon structures provides a more intuitive and comprehensive understanding of the molecular characteristics and behavior of compounds [15–19]. In drug design, knowledge of molecular structure is essential in determining the potential therapeutic activity and overall effectiveness. In this study, we examine several vertex-degree-based topological indices, including the first and second Zagreb indices, hyper-Zagreb index, sigma index, inverse symmetric deviation index, max-min rodeg index, min-max rodeg index, inverse sum deviation index, atom-bond connectivity index, Randic index, and Albertson index [20–29]. Additionally, we investigate topological indices based on distance, such as the Wiener index, Schultz index, Harary index, and Gutman index [30–32]. These indices are used to classify the molecular descriptors and analyze the efficacy of curvilinear regression models in predicting the activity of fibrate drugs.

Molecular descriptors have been widely used to evaluate the physicochemical and bioactive properties of chemical structures, and their inclusion in curvilinear regression models can enhance the analysis of drug activity. Topological indices, such as the Zagreb indices, have shown promise in predicting the effectiveness of cancer treatments [33]. The max-min rodeg index has been found to give reliable predictions for octane isomers and polychlorobiphenyls in linear regression models [34]. A new index called the atom-bond connectivity index has been proposed to determine the complexity of alkanes [35]. The first hyper-Zagreb index has been found to be the preferred method for estimating the boiling points of benzenoid hydrocarbons [36]. Additionally, the indeg indices have been applied to predict topological polar areas [37]. The inverse sum deviation index has been used to calculate the vaporization and sublimation enthalpies of monocarboxylic acids [38,39]. Irregularity indices based on different degrees, in addition to Albertson and Sigma indices, have been found to predict the physicochemical properties of octane isomers [40]. The Wiener index was first introduced in QSPR studies, and has been shown to align well with the boiling points of alkanes [41]. The Wiener index has been further developed and used to explain different chemical and physical properties of molecules, as well as their biological activity [42]. The Schultz index has also been investigated to predict the boiling points of alkyl alcohols, and thus their suitability for various applications [43]. As indicated in Table 1, these indices are expressed mathematically and are shown with mathematical expressions.

**Notation 1.** A molecular graph is a simple graph  $\zeta = (V, E)$ . Its vertices and edges represent the atoms and the bonds, respectively. Note that hydrogen atoms are omitted. For any graph  $d(u)$  represent the degree of the vertex  $u \in V(\zeta)$  in the graph. While  $d(u, v)$  represent the distance of the shortest path from the vertex  $u$  to the vertex  $v$ .

**Table 1.** The mathematical expressions of topological indices.

<b>Vertex-Degree-Based Topological Indices</b>	<b>Mathematical Expression</b>
First Zagreb index	$M_1(\zeta) = \sum_{uv \in E(\zeta)} (d(u) + d(v))$
Second Zagreb index	$M_2(\zeta) = \sum_{uv \in E(\zeta)} (d(u) \cdot d(v))$
Hyper-Zagreb index	$HM(\zeta) = \sum_{uv \in E(\zeta)} (d(u) + d(v))^2$
Atom-bond connectivity index	$ABC(\zeta) = \sum_{uv \in E(\zeta)} \sqrt{\frac{d(u) + d(v) - 2}{d(u) \cdot d(v)}}$
Randić index	$R(\zeta) = \sum_{uv \in E(\zeta)} \frac{1}{\sqrt{d(u) \cdot d(v)}}$
Max-min rodeg index	$Mm_{s-de}(\zeta) = \sum_{uv \in E(\zeta)} \sqrt{\frac{\max(d(u), d(v))}{\min(d(u), d(v))}}$
Min-max rodeg index	$mM_{s-de}(\zeta) = \sum_{uv \in E(\zeta)} \sqrt{\frac{\min(d(u), d(v))}{\max(d(u), d(v))}}$
Albertson index	$irr(\zeta) = \sum_{uv \in E(\zeta)}  d(u) - d(v) $
Sigma index	$\sigma(\zeta) = \sum_{uv \in E(\zeta)} (d(u) - d(v))^2$
Inverse symmetric deg index	$ISDI(\zeta) = \sum_{uv \in E(\zeta)} \frac{d(u) \cdot d(v)}{d(u)^2 + d(v)^2}$
Inverse sum indeg index	$ISI(\zeta) = \sum_{uv \in E(\zeta)} \frac{d(u) \cdot d(v)}{d(u) + d(v)}$
<b>Distance-Based Topological Indices</b>	<b>Mathematical Expression</b>
Wiener index	$W(\zeta) = \sum_{\{u,v\} \subseteq V(\zeta)} d(u,v)$
Schultz index	$S(\zeta) = \sum_{\{u,v\} \subseteq V(\zeta)} (d(u) + d(v))d(u,v)$
Harary index	$H(\zeta) = \sum_{\{u,v\} \subseteq V(\zeta)} \frac{1}{d(u,v)}$
Gutman index	$Gut(\zeta) = \sum_{\{u,v\} \subseteq V(\zeta)} (d(u) \cdot d(v))d(u,v)$

Fenofibrate is an important component of a healthy diet and medication regimen, as it is used to reduce blood cholesterol and triglyceride levels. By decreasing triglyceride levels in the bloodstream, the risk of pancreatitis (inflammation of the pancreas) can be mitigated. To date, only one paper [44] has explored the use of topological indices in analyzing one of the drugs in the fibrate family. This study utilized *ve*-degree, *ev*-degree, and degree-based (*D*-based) approaches to compute the topological indices of fenofibrate’s chemical structure. With limited existing literature on fibrates that incorporate topological indices, this paper represents a pioneering effort in the investigation of novel physicochemical properties of fibrates using this technique. In this work, fenofibrate ( $C_{20}H_{21}ClO_4$ ), ciprofibrate ( $C_{13}H_{14}Cl_2O_3$ ), bezafibrate ( $C_{19}H_{20}ClNO_4$ ), clofibrate ( $C_{12}H_{15}ClO_3$ ) drugs used in the treatment of patients with high cholesterol are studied.

Fibrate drugs are a class of medications commonly used to treat dyslipidemia, a condition characterized by abnormal lipid levels in the blood. Despite their widespread use, the molecular mechanisms underlying the activity of fibrate drugs are not well understood.

One approach to addressing this challenge is to develop quantitative structure-activity relationship (QSAR) models that can predict the activity of fibrates drugs based on their molecular descriptors. In this study, we investigate the efficacy of curvilinear regression models in enhancing the analysis of fibrate drug activity through molecular descriptors. Curvilinear regression models are a type of non-linear regression model that can capture non-linear relationships between variables, making them useful for analyzing complex systems such as the interactions between drugs and their molecular targets. Our study builds upon previous research that has investigated the use of QSAR models to predict the activity of drugs. Several articles published in *Symmetry* have explored the use of topological indices and other mathematical methods to predict various properties of organic compounds, including their biological activity. For example, a study by Berinde et al. [45] used the QSPR technique to investigate the molar refraction, polarizability, and refractive index of monocarboxylic, dicarboxylic, and unsaturated monocarboxylic acids. Their approach relied on a single molecular descriptor, the ZEP topological index, calculated using weighted electronic distances and chemical structure. The QSPR models obtained had high-quality performance and predictive ability, with  $R^2 > 0.99$ . Their approach provides an alternative to existing additive methods for predicting these properties. Zuo and Hu [46] developed QSPR models for predicting the melting points of organic compounds using molecular topology and quantum chemical descriptors. The authors used a dataset of 893 organic compounds and developed multiple linear regression models using the partial least squares (PLS) method. They compared their models with other models reported in the literature and found that their models were more accurate in predicting the melting points of organic compounds. Zhang et al. [47] developed QSPR models for predicting the melting points of organic compounds based on molecular topology. The authors used a dataset of 1427 organic compounds and developed models using the neural network algorithm. They compared their models with other models reported in the literature and found that their models were more accurate in predicting the melting points of organic compounds. Naghipour and Kiasat [48] developed a QSPR model for predicting the fullerene-like behavior of C60 derivatives using topological indices. The authors used a dataset of 46 C60 derivatives with known fullerene-like behavior and developed a model using multiple linear regression. They compared their model with other models reported in the literature and found that their model had higher accuracy in predicting the fullerene-like behavior of C60 derivatives. Wang and Xu [49] developed QSPR models for predicting the boiling points of alkyl alkanes based on the novel vertex degree valence topological index. The authors used a dataset of 388 alkyl alkanes and developed models using multiple linear regression and artificial neural network methods. They compared their models with other models reported in the literature and found that their models were more accurate in predicting the boiling points of alkyl alkanes.

In our study, we apply curvilinear regression models to analyze the activity of fibrate drugs based on their molecular descriptors. By incorporating non-linear relationships between variables, we aim to enhance the accuracy and predictive power of QSAR models for analyzing the activity of fibrate drugs. Ultimately, our research may contribute to a better understanding of the molecular mechanisms underlying the activity of fibrate drugs, and to the development of more effective treatments for dyslipidemia. These studies demonstrate the usefulness of QSAR modeling and related techniques for predicting the activity of various compounds based on their molecular descriptors. By building on this previous work, we hope to further advance our understanding of the molecular mechanisms underlying the activity of fibrate drugs.

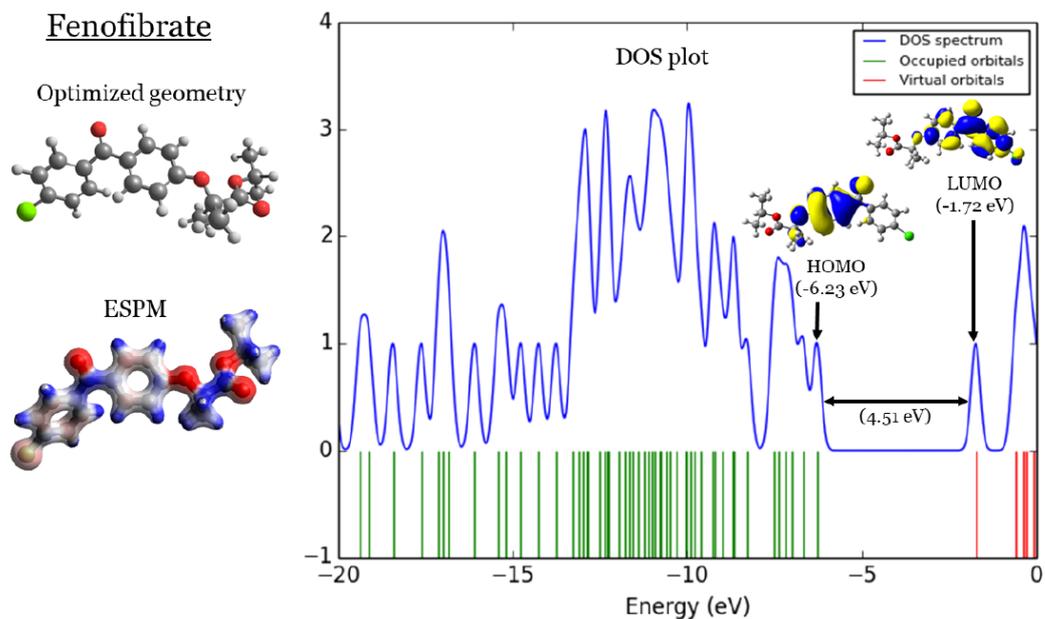
The QSPR model is a highly effective tool for predicting a wide range of physicochemical properties of drugs. To make these predictions, the model employs degree-based indices and distance-based topological indices (as detailed in Table 1). The properties considered include polarizability, Sum of electronic and zero-point Energies, the sum of electronic and thermal energies, the sum of electronic and thermal enthalpies, the sum of electronic and thermal free energies, zero-point vibrational energy, complexity, topological polar area,

dipole moment, heat capacity, molar entropy, and octanol-water partition coefficients. To analyze the relationships between these properties and the topological indices, curvilinear regression (linear, quadratic, and cubic) is utilized. The model generates statistical parameters using SPSS and MATLAB statistical functions. In addition, DFT calculations are conducted at the B3LYP/6-31G(d,p) to gain insight into the optimized geometries, DOS plots, HOMO and LUMO orbitals energies, and distribution of the four derivatives studied in the next section. Section 3 examines the contributions of different topological indices as molecular structural descriptors. Finally, Section 4 concludes the paper.

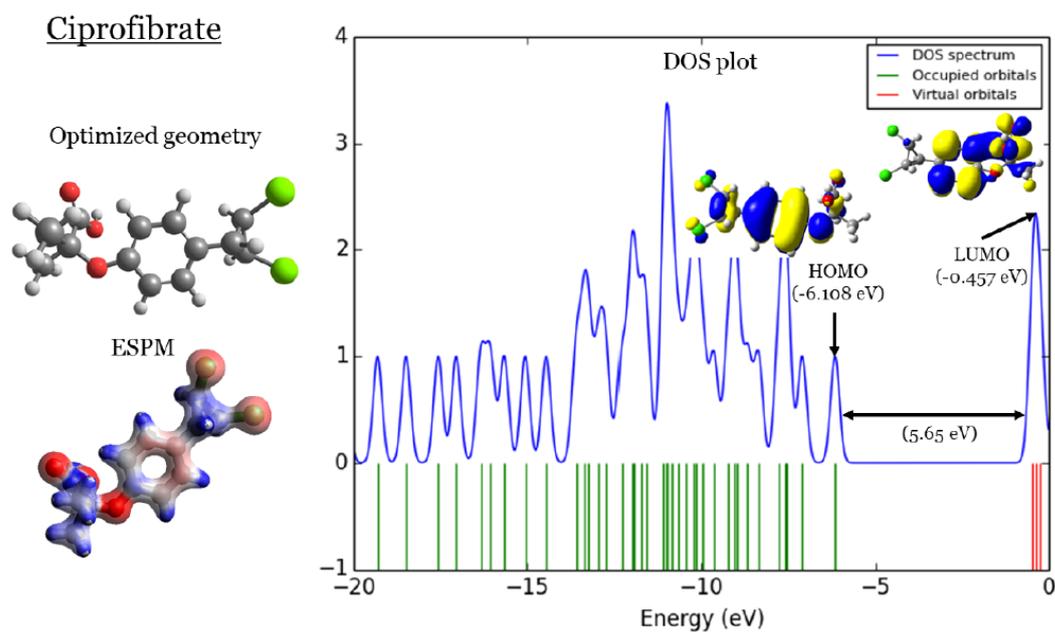
## 2. DFT Part

In Figures 1–4, four important characteristics of the four investigated fibrate derivatives were indicated, including (1) optimized geometries, (2) electron density mapped with electrostatic potential (ESPM), (3) total density of states (DOS) plots, and (4) the special distributions of the highest occupied molecular orbitals (HOMOs) and the lowest unoccupied molecular orbitals (LUMOs). Density functional theory (DFT) calculations of the investigated fibrate derivatives utilized one of the well-known hybrid functionals, Becke, 3-parameter, Lee–Yang–Parr (B3LYP) [50]. In DFT, hybrid functionals incorporate a portion of the Hartree–Fock exchange, as well as extra exchange from other sources (empirical/ab initio) to approximate the exchange–correlation energy. The B3LYP as a representation of a Hamiltonian term in the Schrödinger equation was combined with a 6–31 G(*d*, *p*) basis set as a representation of the eigenvalue wavefunction. It is a moderate double zeta ( $\zeta$ ) basis set enlarged with two polarization basis functions, a *d*–function for heavy atoms (carbon, oxygen, and chlorine), and a *p*–function for all hydrogen atoms. Most of the physicochemical properties of the investigated fibrate derivatives discussed in next section were obtained from the frequency calculations carried out at the same level of theory of optimization. Calculations were carried out using Gaussian 09 software suite [51]. Visualizations of molecular structures were performed by using GaussView (version 5.0.8) [52], ESPMs were drawn used the Avogadro package [53], and the GaussSum program [54] was used to DOS plots. ESPMs show how electron density is distributed in the four non-planar molecules considering the electrostatic potentials, and this gives information about the region in the molecule that has the highest or lowest electron density, and thus is most likely to be attacked by electrophilic or nucleophilic agents. Keep in mind that the nucleophilic and electrophilic attack regions are represented by blue (positively charged) and red colors (negatively charged). The red color is concentrated on the more electronegative atoms such as the oxygen (deep red) and chlorine atoms (light red), the blue color covers the hydrogen atoms (the least electronegative atoms), while the carbon atoms are covered by white, indicating the intermediate electronegativity of the carbon atom. Thus, it is possible to determine the position and region in a molecule attacked by an electrophile or nucleophile using ESPMs. The molecule DOS plot indicates how many energy states electrons are allowed to occupy in the system. The HOMO energies of the four investigated fibrate derivatives are  $-6.230$ ,  $-6.108$ ,  $-6.166$ , and  $-6.422$  eV for fenofibrate, ciprofibrate, bezafibrate, and clofibrate, respectively. Since the HOMO energy is used as a measure of the electron-donating power of a molecule, destabilized HOMO (less negative) leads to a greater ability to donate electrons. The capacity for electron donation of the five derivatives can be arranged as follows: ciprofibrate > bezafibrate > fenofibrate > clofibrate. On the other hand, the LUMO energy measures the electron-accepting ability of a molecule; more combined ability stabilizes LUMO (more negative). Therefore, the derivatives' ability to accept electrons is as follows: fenofibrate ( $-1.720$  eV) > bezafibrate ( $-1.220$  eV) > clofibrate ( $-0.461$  eV) > ciprofibrate ( $-0.457$  eV). The energy gap (HOMO energy subtracted from LUMO energy) measures the chemical reactivity. A smaller gap is a more reactive molecule, and the reactivity of the four derivatives is: fenofibrate ( $4.51$  eV) < bezafibrate ( $4.95$  eV) < ciprofibrate ( $5.65$  eV) < clofibrate ( $5.96$  eV). Finally, the 2D special distribution of HOMO and LUMO orbitals is another indicator of the position/region subjected to electrophilic and nucleophilic attack. The HOMO and LUMO orbitals in

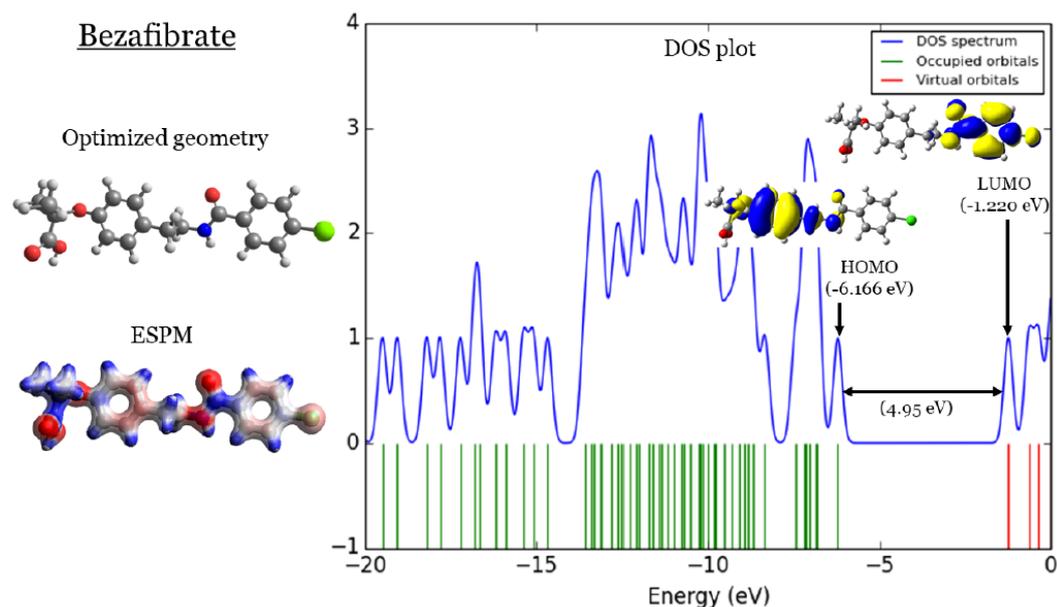
ciprofibrate are distributed on similar parts of a molecule, except that the two chlorine atoms have more HOMO character. Other molecules have HOMO orbitals delocalized over different regions compared to the LUMO orbitals distribution.



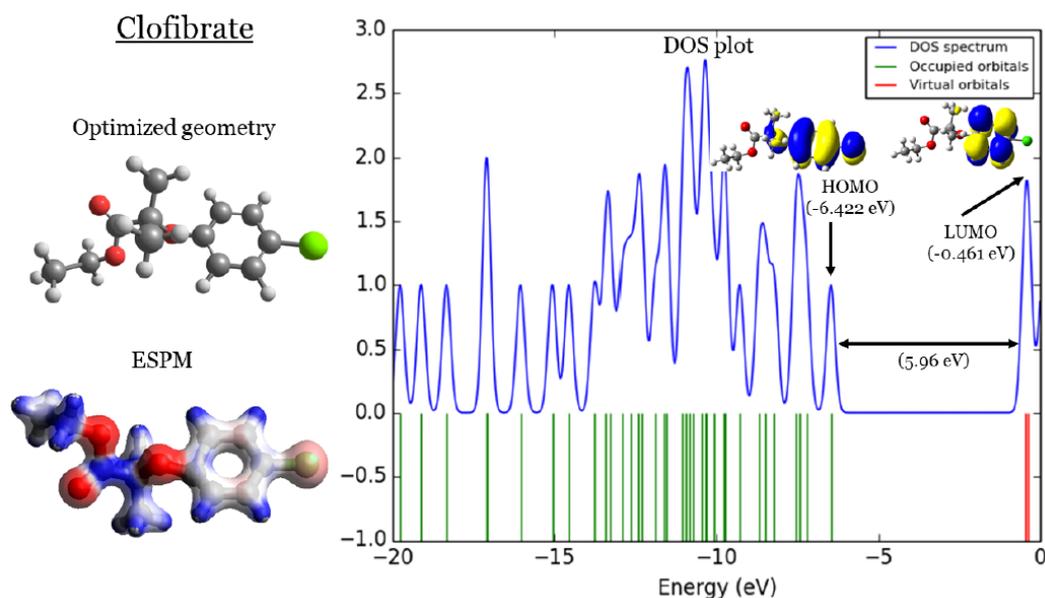
**Figure 1.** Fenofibrate: (1) optimized geometries, (2) ESPM, (3) DOS plots, and (4) HOMOs and LUMOs.



**Figure 2.** Ciprofibrate: (1) optimized geometries, (2) ESPM, (3) DOS plots, and (4) HOMOs and LUMOs.



**Figure 3.** Bezafibrate: (1) optimized geometries, (2) ESPM, (3) DOS plots, and (4) HOMOs and LUMOs.



**Figure 4.** Clofibrate: (1) optimized geometries, (2) ESPM, (3) DOS plots, and (4) HOMOs and LUMOs.

### 3. Materials and Method

In this section, the overall objective is to establish a QSPR relationship between the various topological indices and some physicochemical properties/activities of the fibrate drugs under study in order to assess the effectiveness of these drugs. Eleven degree-based and four distance topological indices were used for modeling antiviral activity, based on DMol3-optimized geometries for the fibrate drugs investigated. The Gaussian 09 software package was used to perform DFT calculations, which are as follows: polarizability ( $P$ ), sum of electronic and zero-point energies ( $SEZPE$ ), sum of electronic and thermal energies ( $SETEnergy$ ), sum of electronic and thermal enthalpies ( $SETEnthalpy$ ), sum of electronic and thermal free energies ( $SETFEnergy$ ), zero-point vibrational energy ( $ZPVE$ ), complexity ( $C$ ), topological polar area ( $TPA$ ), dipole moment ( $DM$ ), heat capacity ( $CV$ ), molar entropy ( $S$ ), and octanol-water partition coefficients ( $XlogP3$ ) of several drugs currently being investigated for the treatment of high cholesterol, which include fenofibrate,

ciprofibrate, bezafibrate, and clofibrate drugs. It is possible to use curvilinear regression analysis to fit curves instead of straight lines, and SPSS statistical software is used to analyze curvilinear regressions. As described below, the independent variables in the curvilinear regression models are topological indices. Indicators are derived from cholesterol-lowering drugs. Based on the equations below, tests were conducted.

$$y = a + bx; \quad n, R^2, F, Se, SF \quad (\text{Linear equation})$$

$$y = a + b_1x + b_2x^2; \quad n, R^2, F, Se, SF \quad (\text{Quadratic equation})$$

$$y = a + b_1x + b_2x^2 + b_3x^3; \quad n, R^2, F, Se, SF \quad (\text{Cubic equation})$$

In this context,  $y$  represents the response or dependent variable, while  $a$  denotes the regression model constant, and  $b_i (i = 1, 2, 3)$  refers to the coefficients for each individual descriptor. The independent variable is represented by  $x$ , and  $n$  signifies the number of samples used in building the regression equation.  $R^2$  denotes the coefficient of determination,  $R$  signifies the correlation coefficient,  $F$  represents the calculated value of the Fischer  $F$ -values test,  $Se$  denotes the standard error of estimate, and  $SF$  stands for  $F$ -significance. It should be noted that when the experimental and theoretical results are in close proximity to each other, the correlation coefficient approaches 1. To gauge the predictability of a model, it is necessary to compare the observed values and the model predictions, for which the root mean square error ( $RMSE$ ) metric is used. The predictive quality of a model is higher when the error or  $RMSE$  is lower, which is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

where  $y_i$  is the observed value of the independent variable in the test set,  $\hat{y}_i$  is the predicted value of the independent variables in the test set,  $n$  is the number of samples in the test, and topological indices serve as independent variables. To evaluate our initial model, we used the  $RMSE$  metric and then normalized the data to enhance our predictions' accuracy. We measured the difference between predicted and actual values using the  $RMSE$  score, which revealed that our model needed improvement. To address issues such as outliers and varying scales of measurement that could negatively affect model performance, we applied normalization techniques to our data. The normalization step was essential in improving the model's accuracy, as it scaled variables to a common range, reduced the impact of outliers, and ensured that all variables were weighted equally. After normalization, we re-evaluated the model using the  $RMSE$  metric, and the updated score showed a significant improvement in our predictions' accuracy. Computed topological index values are shown in Table 2. We computed the values using combinatorial computations and edge partitioning as follows: the molecular graph of fenofibrate has 25 vertices and 26 edges. Its edges can be partitioned as  $|E_{1,4}| = 2$ ,  $|E_{1,3}| = 5$ ,  $|E_{2,3}| = 11$ ,  $|E_{2,2}| = 4$ ,  $|E_{3,4}| = 1$ ,  $|E_{3,3}| = 2$ , and  $|E_{2,4}| = 1$ . The molecular graph of ciprofibrate has 18 vertices and 19 edges. Its edges can be partitioned as  $|E_{2,2}| = 2$ ,  $|E_{1,4}| = 4$ ,  $|E_{2,4}| = 2$ ,  $|E_{1,3}| = 2$ ,  $|E_{2,3}| = 6$ ,  $|E_{1,4}| = 2$ , and  $|E_{1,3}| = 1$ . The molecular graph of bezafibrate has 25 vertices and 26 edges. Its edges can be partitioned as  $|E_{1,3}| = 4$ ,  $|E_{2,3}| = 11$ ,  $|E_{2,2}| = 6$ ,  $|E_{3,3}| = 1$ ,  $|E_{2,4}| = 1$ ,  $|E_{3,4}| = 1$ , and  $|E_{1,4}| = 2$ . The molecular graph of fenofibrate has 16 vertices and 16 edges. Its edges can be partitioned as  $|E_{1,3}| = 2$ ,  $|E_{2,3}| = 6$ ,  $|E_{1,2}| = 1$ ,  $|E_{2,2}| = 3$ ,  $|E_{3,4}| = 1$ ,  $|E_{1,4}| = 2$ , and  $|E_{2,4}| = 1$ . Using MATLAB, it is possible to efficiently compute degree-based and distance-based topological indices, as explained in Algorithms 1 and 2. To calculate the topological indices of molecules based on distance and degree, MATLAB utilizes various mathematical expressions. The fibrate family and the drugs under consideration, namely fenofibrate, ciprofibrate, bezafibrate, and clofibrate, have been studied and are presented in Table 3, including their experimental data [51] and optimized geometries obtained through DFT calculations using the DMol3 module of Version 8.0 of Material Studio from BIOVIA. Table 4 shows the correlation coefficient ( $R$ ) between degree-based topological indices and

some physicochemical properties, computed using a linear regression model. A quadratic regression model is used in Table 5 to calculate the correlation coefficient ( $R$ ) between these indices and some physicochemical properties. The cubic model is employed for this purpose in Table 6. Similarly, for the distance-based topological indices, linear, quadratic, and cubic regression models are utilized, and the results are presented in Table 7. Once the correlation coefficient for a physicochemical property is obtained, the model with the maximum  $R$  becomes the most accurate predictor of the regression model. This is indicated in Tables 8–11. By leveraging the power of MATLAB, it is possible to efficiently and accurately compute topological indices and use them to predict the physicochemical properties of molecules, which can be incredibly useful in various fields, including drug discovery and materials science.

---

**Algorithm 1:** Computational procedure of calculation of degree-based indices.

---

Input: Edges and nodes of molecule

Output:  $e \leftarrow$  Topological indices vector

Step 1. Start

Step 2.  $G \leftarrow$  Graph of undirected edges

Step 3.  $A \leftarrow$  Adjacency matrix of  $G$

Step 4.  $d \leftarrow$  Distances of  $G$

Step 5.  $d_1 \leftarrow$  Vertex degree of  $G$

Step 6. Calculate size of matrix  $d$

Step 4. Construct  $A_N$  :

**for**  $i = 1$  to number of columns **do**

**for**  $j = 1$  to number of rows **do**

**if**  $i = j$  **then**

$A_N(i, j) = 0$

**elseif**  $A(i, j) = 1$  **then**

$A_N(i, j) = d_1(i) + d_1(j)$  First Zagerb index

$A_N(i, j) = d_1(i) * d_1(j)$  Second Zagerb index

$A_N(i, j) = (d_1(i) + d_1(j))^2$  Hyper-Zagerb index

$A_N(i, j) = \sqrt{\frac{d_1(i) + d_1(j) - 2}{(d_1(i) * d_1(j))}}$  Atom-Bond Connectivity index

$A_N(i, j) = \frac{1}{\sqrt{d_1(i) * d_1(j)}}$  Randic index

$A_N(i, j) = \sqrt{\frac{\min(d_1(i), d_1(j))}{\max(d_1(i), d_1(j))}}$  min-max rodeg index

$A_N(i, j) = \sqrt{\frac{\max(d_1(i), d_1(j))}{\min(d_1(i), d_1(j))}}$  max-min rodeg index

$A_N(i, j) = |d_1(i) - d_1(j)|$  Alberston index

$A_N(i, j) = (d_1(i) - d_1(j))^2$  Sigma index

$A_N(i, j) = \frac{d_1(i).d_1(j)}{d_1(i)^2 + d_1(j)^2}$  Inverse symmetric deg index

$A_N(i, j) = \frac{d_1(i).d_1(j)}{d_1(i) + d_1(j)}$  Inverse sum deg index

**end if**

**end for**

**end for**

Step 5.  $e = (\text{summation of } A_N) / 2.$

---

**Table 2.** Values of topological indices in fibrates' molecular structures.

Topological Index	Fenofibrate	Ciprofibrate	Bezafibrate	Clofibrate
$M_{1(\zeta)}$	126	98	124	76
$M_{2(\zeta)}$	143	115	139	84
$H(\zeta)$	626	520	606	374
$ABC(\zeta)$	19.1	14.12	19.03	11.78
$R(\zeta)$	11.68	8.22	11.77	7.45
$mM_{s-de}(\zeta)$	20.44	14.19	20.864	12.33
$Mm_{s-de}(\zeta)$	34.7	26.95	33.9693	21.79
$irr(\zeta)$	30	28	28	20
$\sigma(\zeta)$	54	60	50	38
$ISDI(\zeta)$	10.92	7.5704	11.12	6.61
$ISI(\zeta)$	28.59	21.4952	28.34	17.01
$W(\zeta)$	1716	660	1882	468
$S(\zeta)$	6872	2652	7600	1776
$H(\zeta)$	87.5476	55.1468	84.5541	45.5162
$Gut(\zeta)$	6846	2638	7650	1670

**Algorithm 2:** Computational procedure of calculation of distance-based indices.

```

Input: Edges and nodes of molecule
Output:  $e \leftarrow$  Topological indices vector
Step 1. Start
Step 2.  $G \leftarrow$  Graph of undirected edges
Step 3.  $A \leftarrow$  Adjacency matrix of  $G$ 
Step 4.  $d \leftarrow$  Distances of  $G$ 
Step 5.  $d_1 \leftarrow$  Vertex degree of  $G$ 
Step 6. Calculate size of matrix  $d$ 
Step 4. Construct  $A_N$  :
for  $i = 1$  to number of rows-1 do
   $aa = 0$ ;
  for  $j = i + 1$  to number of columns do
     $aa = \sum d(i, j)$  Wiener index
     $aa = \sum d(i, j) * (d_1(i) + d_1(j))$  Schultz index
     $aa = \sum \frac{1}{d(i, j)}$  Harary index
     $aa = \sum d(i, j)(d_1(i) * d_1(j))$  Gutman index
  end for  $A_N(i) = aa$ 
end for
Step 5.  $e =$  summation of  $A_N$ .

```

*Results and Discussion*

Fibrate drugs are predicted by numerous topological indices. In QSPR, linear, quadratic, and cubic regression models are examined. Several topological indices are calculated for fibrate drugs, including the vertex degree, and distance between vertices. The models are analyzed using twelve descriptors and thirteen topological indices. Using a linear regression model, a correlation coefficient ( $R$ ) between these indices and some physicochemical properties can be seen in Table 4. In Table 5, using a quadratic regression model, a correlation coefficient ( $R$ ) between these indices and some physicochemical properties is computed. When a correlation coefficient is obtained for a physicochemical property,

the model that has maximum  $R$  is the most accurate predictor of the regression model. In Table 4, we display  $maximum(R)$  for each physicochemical property, based upon the analysis of the data (linear and quadratic). We have excluded values less than 0.64 from the Tables 4 and 5, for convenience.

With linear regression models, the following Table 8 illustrates the most appropriate topological index for estimating physicochemical properties. A diagram depicting this is shown in Figure 5.

**Table 3.** The physicochemical properties of potential drugs of fibrates.

Physicochemical Properties	Fenofibrate	Ciprofibrate	Bezafibrate	Clofibrate
(DM)	3.98025	3.94641	3.01127	2.19815
(P)	164.27567	244.49533	232.43367	144.46
(SEZ <sub>pE</sub> )	−1649.62662	−1535.54779	−1551.61567	−1151.954
(SET <sub>Energy</sub> )	−1649.60875	−1535.52308	−1551.59139	−1151.93749
(SET <sub>Enthalpy</sub> )	−1649.6078	−1535.52214	−1551.59044	−1151.93749
(SET <sub>FEnergy</sub> )	−1649.67518	−1535.60609	−1551.6753	−1152.00027
(Z <sub>pVE</sub> )	155.46481	231.67184	225.46799	157.25221
(CV)	66.502	92.538	91.009	61.172
(S)	141.803	176.701	178.604	134.118
(XLogP3)	5.2	3.4	3.8	3.3
(C)	458	333	452	232
(TPA)	52.6	46.5	75.6	35.5

**Table 4.** The correlation coefficient (R) obtained by linear regression model between topological indices and physicochemical properties of various drugs of fibrates.

T.I.	SEZ <sub>pE</sub> SET <sub>Energy</sub> SET <sub>Enthalpy</sub> SET <sub>FEnergy</sub>	XLogP3	C	TPA
$M_1(\zeta)$	−0.902	0.74	1	0.811
$M_2(\zeta)$	−0.941	0.729	0.995	0.786
$H(\zeta)$	−0.89	0.771	0.998	0.791
$ABC(\zeta)$	−0.84	0.748	0.992	<b>0.826</b>
$R(\zeta)$	−0.765	0.746	0.967	<b>0.826</b>
$mM_{s-de}(\zeta)$	−0.914	<b>0.833</b>	0.985	0.705
$Mm_{s-de}(\zeta)$	−0.796	0.758	0.978	0.819
$irr(\zeta)$	<b>−0.999</b>	0.647	0.887	—
$\sigma(\zeta)$	−0.848	—	—	—
$ISDI(\zeta)$	−0.669	0.736	0.922	0.805
$ISI(\zeta)$	−0.855	0.75	0.995	0.821

**Table 5.** The correlation coefficient (R) obtained by quadratic regression model between topological indices and physicochemical properties of various drugs of fibrates.

T.I.	DM	P	Z <sub>p</sub> VE	CV	S	XLogP3	C	TPA	SEZ <sub>p</sub> E SET <sub>Energy</sub> SET <sub>Enthalpy</sub> SET <sub>F</sub> Energy
M <sub>1</sub>	0.850	0.881	0.803	0.848	0.820	0.807	<b>1.000</b>	0.811	0.979
M <sub>2</sub>	0.837	0.908	0.843	0.878	0.850	0.850	0.998	0.786	0.981
H	0.808	0.930	0.872	0.904	0.879	0.852	0.999	0.804	0.973
ABC	0.874	0.884	0.752	0.808	0.779	0.768	<b>1.000</b>	0.829	0.981
R	0.929	0.756	–	0.714	0.684	0.746	<b>1.000</b>	0.831	0.993
mM <sub>s-de</sub>	0.868	0.851	0.760	0.815	0.787	0.768	<b>1.000</b>	0.832	0.979
Mm <sub>s-de</sub>	0.746	<b>0.997</b>	0.984	0.990	0.979	0.964	0.990	0.796	0.971
irr	0.894	0.995	<b>0.998</b>	<b>0.999</b>	<b>0.999</b>	<b>0.983</b>	0.893	0.712	<b>1.000</b>
σ	<b>0.947</b>	0.708	–	0.667	–	0.677	0.996	<b>0.882</b>	0.991
ISDI	–	–	–	–	–	0.736	0.922	0.810	0.669
ISI	0.861	0.863	0.777	0.828	0.800	0.782	<b>1.000</b>	0.825	0.979

**Table 6.** The correlation coefficient (R) obtained by cubic regression model between topological indices and physicochemical properties of various drugs of fibrates.

T.I.	SEZ <sub>p</sub> E SET <sub>Energy</sub> SET <sub>Enthalpy</sub> SET <sub>F</sub> Energy	P	C	TPA	XLogP3	S	DM	CV
M <sub>1</sub>	0.979	0.886	<b>1.000</b>	0.811	0.813	0.826	0.850	0.854
M <sub>2</sub>	0.981	0.915	0.998	0.786	0.859	0.858	0.837	0.885
H	0.973	0.939	0.999	0.806	0.863	0.890	0.808	0.914
ABC	0.981	0.846	<b>1.000</b>	0.829	0.769	0.782	0.874	0.810
R	0.994	0.756	<b>1.000</b>	0.831	0.746	0.684	0.934	0.714
mM <sub>s-de</sub>	0.979	0.854	<b>1.000</b>	0.832	0.769	0.791	0.868	0.819
Mm <sub>s-de</sub>	0.971	<b>0.999</b>	0.991	0.806	0.973	0.985	0.746	0.994
irr	<b>1.000</b>	0.995	0.893	0.712	0.983	<b>0.999</b>	0.894	<b>0.999</b>
σ	0.992	0.708	0.998	0.882	0.689	0.645	<b>0.948</b>	0.667
ISDI	0.691	/	0.923	<b>0.970</b>	<b>0.997</b>	0.690	/	0.650
ISI	0.979	0.867	<b>1.000</b>	0.825	0.785	0.805	0.861	0.833

Table 9 illustrates the best topological index, which gives the best estimate for physicochemical properties using quadratic regression models; we only consider topological index with  $R^2 \geq 0.8$ . A diagram depicting this is shown in Figure 6.

**Table 7.** The curvilinear models, along with the linear, quadratic, and cubic regression models, were used to determine the correlation coefficient (R) between the physicochemical properties of various fibrate drugs and their distance topological indices.

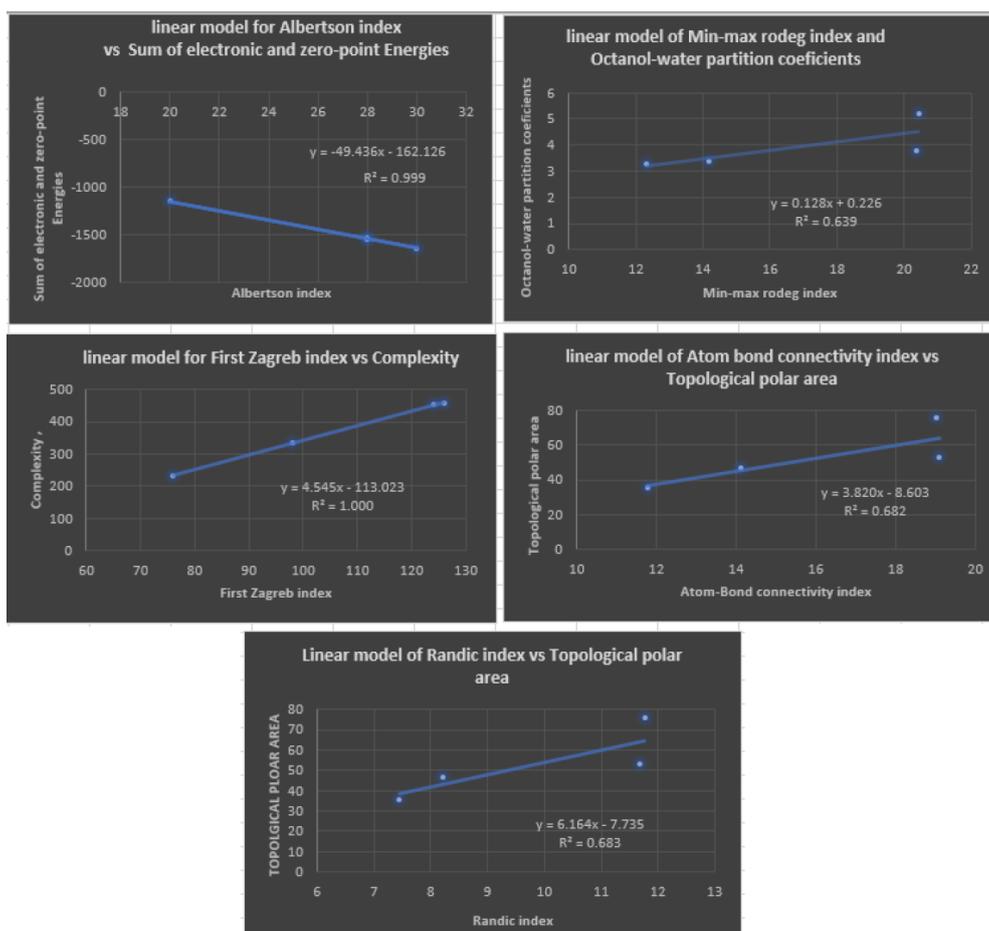
P.P.	$\underbrace{W}$	$\underbrace{S}$	$\underbrace{H}$	$\underbrace{Gut}$
	Linear, Quadratic, cubic	Linear, Quadratic, cubic	Linear, Quadratic, cubic	Linear, Quadratic, cubic
DM	0.334, 0.991, <b>1</b>	0.335, 0.989, <b>1</b>	0.465, 0.750, <b>0.750</b>	0.349, 0.997, <b>1</b>
P	0.185, 0.332, <b>1</b>	0.198, 0.321, <b>1</b>	0.166, 0.958, <b>0.971</b>	0.209, 0.383, <b>1</b>
Z <sub>p</sub> VE	0.086, 0.152, <b>1</b>	0.1, 0.144, <b>1</b>	0.042, 0.908, <b>0.950</b>	0.108, 0.205, <b>1</b>
CV	0.207, 0.297, <b>1</b>	0.221, 0.292, <b>1</b>	0.177, 0.937, <b>0.954</b>	0.230, 0.345, <b>1</b>
S	0.258, 0.305, <b>1</b>	0.272, 0.306, <b>1</b>	0.220, 0.917, <b>0.936</b>	0.280, 0.348, <b>1</b>
SEZ <sub>p</sub> E	0.731, 0.977, <b>1</b>	0.734, 0.976, <b>1</b>	0.807, 0.950, <b>0.950</b>	0.744, 0.986, <b>1</b>
XLogP3	0.696, 0.819, <b>1</b>	0.688, 0.819, <b>1</b>	0.789, 0.839, <b>0.859</b>	0.690, 0.793, <b>1</b>
C	0.954, 0.995, <b>1</b>	0.955, 0.996, <b>1</b>	0.979, 0.999, <b>0.999</b>	0.960, 0.998, <b>1</b>
TPA	0.859, 0.876, <b>1</b>	0.866, 0.885, <b>1</b>	0.790, 0.854, <b>0.856</b>	0.867, 0.881, <b>1</b>

**Table 8.** Linear regression models that give the best estimate for physicochemical properties.

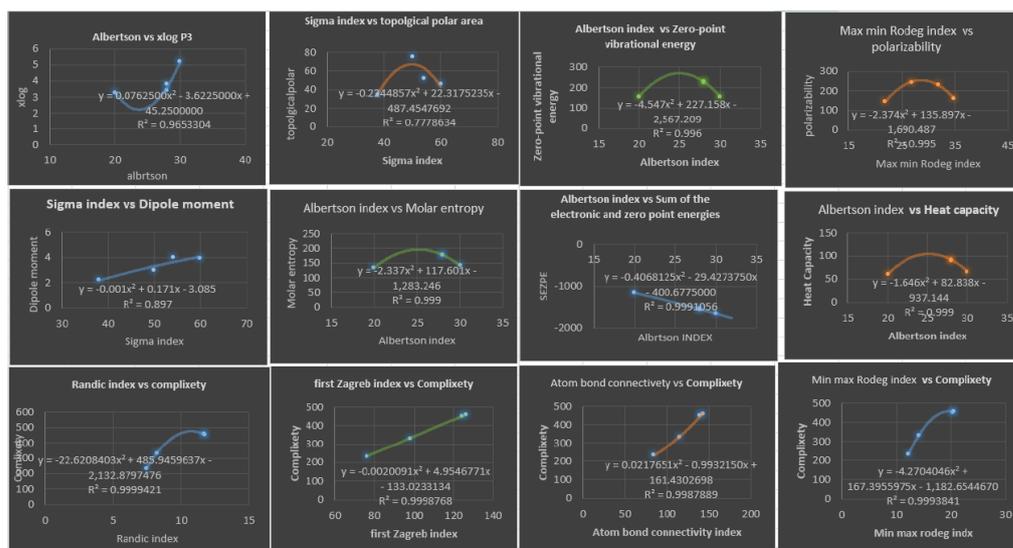
Linear Regression Model	R <sup>2</sup>	F	Se	SF	RMSE
SEZ <sub>p</sub> E = −162.126 − (49.436)irr(ζ)	0.999	1747.706	9.083	0.0005	6.4227673
XLogP3 = 0.226 + (0.128)mM <sub>s-de</sub> (ζ)	0.639	4.522	0.595	0.167	0.4953885
C = −113.023 + (4.545)M <sub>1</sub> (ζ)	1	13088.633	1.632	0.000076	1.1555164
TPA = −8.603 + (3.820)ABC(ζ)	0.682	4.293	16.390	0.174	8.2595015
TPA = −7.735 + (6.164)R(ζ)	0.683	4.308	11.667	0.174	8.2495412

**Table 9.** Quadratic regression model that give the best estimate for physicochemical properties.

Quadratic Regression Model	R <sup>2</sup>	F	Se	SF	RMSE
DM = −3.085 + (0.171)σ − (0.001)σ <sup>2</sup>	0.897	4.360	0.473	0.321	0.4332951
P = −1690.487 + (135.897)Mm <sub>s-de</sub> − (2.374)Mm <sub>s-de</sub> <sup>2</sup>	0.995	97.646	6.116	0.071	3.057903
Z <sub>p</sub> VE = −2567.209 + (227.158)irr − (4.547)irr <sup>2</sup>	0.996	135.526	4.387	0.061	2.193392
CV = −937.144 + (82.838)irr − (1.646)irr <sup>2</sup>	0.999	339.909	1.081	0.038	0.540583
S = −1283.246 + (117.601)irr − (2.337)irr <sup>2</sup>	0.999	443.193	1.346	0.034	0.732681
XLogP3 = 0.0763irr <sup>2</sup> − 3.6225irr + 45.25	0.965	6.644	0.283	0.265	0.141424
C = −0.0020091M <sub>1</sub> <sup>2</sup> + 4.9546771M <sub>1</sub> − 133.0233134	1.000	4059.09	2.07	0.01	1.036262
C = −22.6208403R <sup>2</sup> + 485.9459637R − 2132.8797476	1.000	8639.90	1.42	0.01	0.710303
C = −4.2704046mM <sub>s-de</sub> <sup>2</sup> + 167.3955975mM <sub>s-de</sub> − 1182.6544670	0.999	811.32	4.63	0.02	2.317288
C = −0.7828926ISI <sup>2</sup> + 55.0686241ISI − 478.1259314	1.000	3041.59	2.39	0.01	1.297084
C = 0.0217651ABC <sup>2</sup> − 0.9932150ABC + 161.4302698	0.999	155.21	10.58	0.06	5.291185
TPA = −0.2245σ <sup>2</sup> + 22.318σ − 487.45	0.778	1.751	13.810	0.471	6.90523
SEZ <sub>p</sub> E = −0.4068irr <sup>2</sup> − 29.427irr − 400.68	0.999	558.642	11.362	0.030	5.68161



**Figure 5.** Plots of Linear Regression Equations for the Best Physicochemical Properties Predicted by Degree-Based Topological Indices.

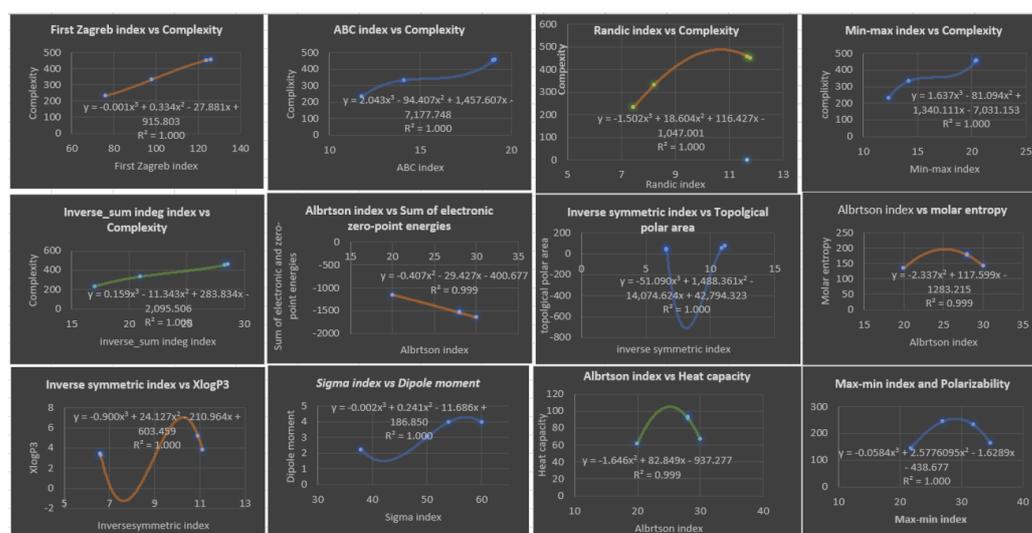


**Figure 6.** Plots of QuadraticRegression Equations for the Best Physicochemical Properties Predicted by Degree-Based Topological Indices.

**Remark 1.** Initially, linear regression was attempted on all physicochemical properties using degree-based topological indices. Correlation coefficients were calculated for 7 out of 12 properties that showed satisfactory results, as presented in Table 4. For the remaining properties with correlation

coefficients less than 0.64, Table 5 explored alternative models. An additional 5 properties were tested, and if their correlation coefficients exceeded 0.64, the quadratic regression model was used. Note that some properties, such as the sum of the electronic and zero-point energies ( $SEZ_{PE}$ ), sum of the electronic and thermal energies ( $SET_{Energy}$ ), sum of the electronic and thermal enthalpies ( $SET_{Enthalpy}$ ), and sum of the electronic and thermal free energies ( $SET_{FreeEnergy}$ ) have identical correlation coefficients, and only ( $SEZ_{PE}$ ) is listed in Tables 8 and 9.

The cubic model is used for all the physicochemical properties and degree-based topological indices in order to provide a comprehensive analysis. Table 6 presents the correlation coefficients, which are high, as anticipated. Table 10 and Figure 7 display the best predictions of the properties.



**Figure 7.** Plots of Cubic Regression Equations for the Best Physicochemical Properties Predicted by Degree-Based Topological Indices.

Based on three curvilinear models, linear, quadratic, and cubic, the following Table 7, illustrates the correlation coefficient  $R$  for the four distance topological indices. The next table shows the most accurate prediction of the physicochemical properties based on linear or quadratic models. It should be noted that the physicochemical properties: sum of the electronic and zero-point energies ( $SEZ_{PE}$ ), sum of the electronic and thermal energies ( $SET_{Energy}$ ), sum of the electronic and thermal enthalpies ( $SET_{Enthalpy}$ ), and Sum of the electronic and thermal free energies ( $SET_{FreeEnergy}$ ) have the same correlation coefficients, which is why the ( $SEZ_{PE}$ ) is the only one listed in Table 7. It is evident that the cubic model is the optimal model to predict all physicochemical properties of fibrates. Notice that we displayed the correlation coefficient in bold for the cubic model. Table 11 and Figure 8 illustrate the best linear and quadratic model of distance-based topological indices with the properties.

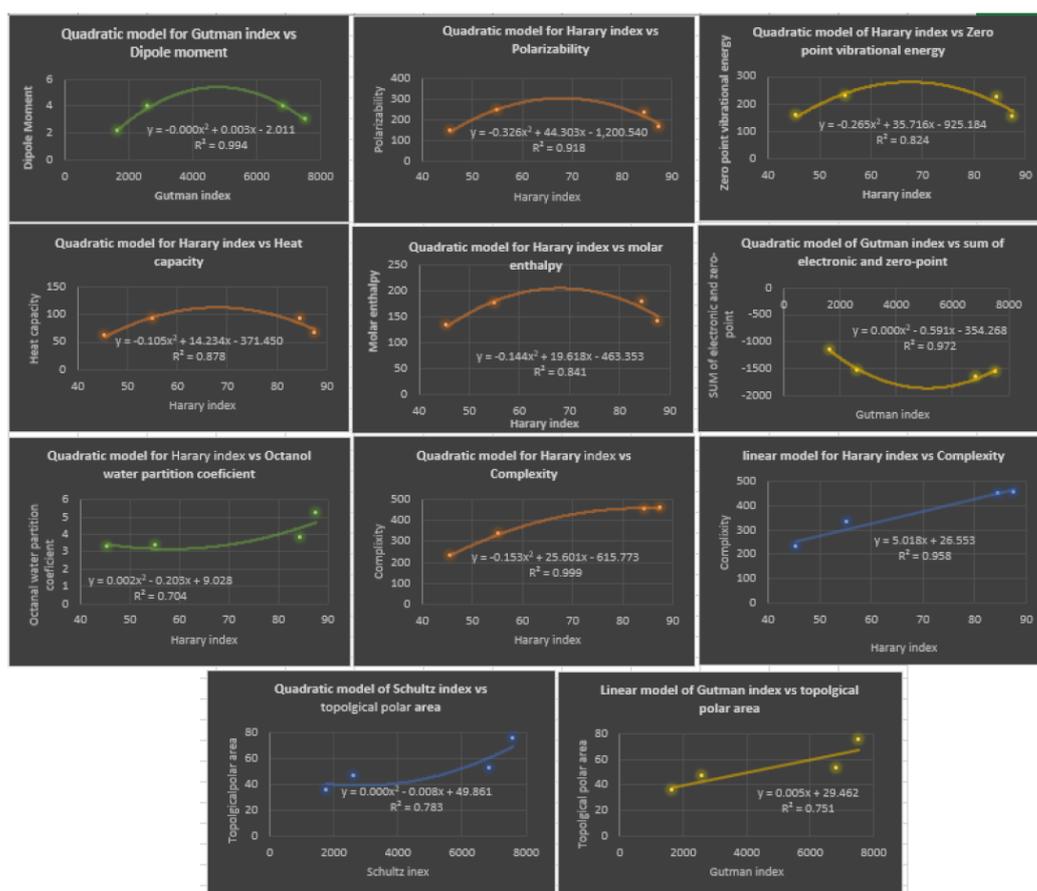
The physicochemical properties of fibrate drugs and their corresponding degree-based and distance-based topological indices were analyzed using three curvilinear models: linear, quadratic, and cubic. The aim was to determine the most accurate correlation coefficient for the properties studied.

**Table 10.** Cubic regression models that give the best estimate for physicochemical

Cubic Regression Model	R <sup>2</sup>	F	Se	SF	RMSE
$SEZ_pE = -0.407irr^2 - 29.427irr - 400.677$	0.999	558.642	11.362	0.030	5.68160
$P = \begin{matrix} -0.0584988Mm_{s-de}^3 + \\ 2.5776095Mm_{s-de}^2 - \\ 1.62895Mm_{s-de} - 438.67733 \end{matrix}$	1.000	361.397	3.185	0.037	0.00032
$C = \begin{matrix} -0.001M_1^3 + 0.334M_1^2 - \\ 27.881M_1 + 915.803 \end{matrix}$	1.000	4108.744	2.060	0.011	0.06846
$C = \begin{matrix} 2.043ABC^3 - 94.407ABC^2 \\ +1457.607ABC - 7177.748 \end{matrix}$	1.000	1315.359	3.640	0.019	0.00024
$C = \begin{matrix} -1.502R^3 + 18.604R^2 \\ +116.427R - 1.047.001 \end{matrix}$	1.000	4237.196	0.641	0.003	0.00006
$C = \begin{matrix} 1.637mM_{s-de}^3 - 81.094mM_{s-de}^2 \\ +1340.111mM_{s-de} - 7,031.15 \end{matrix}$	1.000	811.323	4.635	0.025	0.00005
$C = \begin{matrix} 0.159ISI^3 - 11.343ISI^2 \\ +283.834ISI - 2095.506 \end{matrix}$	1.000	3041.588	2.394	0.013	0.00052
$TPA = \begin{matrix} -51.090ISDI^3 + 1488.36ISDI^2 \\ -14,074.62ISDI + 42,794.32 \end{matrix}$	1.000	7.894	7.151	0.244	0.00003
$S = -2.337irr^2 + 117.599irr - 1283.215$	0.999	443.193	1.346	0.034	0.67175
$XLogP3 = \begin{matrix} -0.900ISDI^3 + 24.127ISDI^2 \\ -210.964ISDI + 603.459 \end{matrix}$	0.900	85.350	0.116	0.076	0.00001
$DM = \begin{matrix} -0.002\sigma^3 + 0.241\sigma^2 \\ -11.686\sigma + 186.850 \end{matrix}$	1.000	4.396	0.471	0.320	0.00612
$CV = -1.646irr^2 + 82.849irr - 937.277$	0.999	339.909	1.081	0.038	0.54094

**Table 11.** The linear and quadratic regression models provide the most accurate predictions for the physicochemical properties.

Linear and Quadratic Best Regression Model	R <sup>2</sup>	F	Se	SF	RMSE
$DM = \begin{matrix} -2.010 + (0.003)Gut \\ - (3.239E^{-7})Gut^2 \end{matrix}$	0.994	84.508	0.113	0.077	0.9007267
$P = -1200.200 + (44.294)H - (0.326)H^2$	0.918	5.597	24.537	0.286	12.239175
$Z_pVE = \begin{matrix} -925.184 + (35.716)H \\ - (0.265)H^2 \end{matrix}$	0.824	1.252	30.377	0.534	15.16917
$CV = -371.268 + (14.228)H - (0.105)H^2$	0.878	3.585	9.869	0.350	4.9268046
$S = -463.186 + (19.618)H - (0.144)H^2$	0.840	2.634	16.012	0.399	7.9943483
$SEZ_pE = \begin{matrix} -354.281 - (0.591)Gut \\ + (5.778E^{-5})Gut^2 \end{matrix}$	0.972	17.347	63.547	0.167	36.510223
$XLogP3 = 9.019 - (0.202)H + (0.002)H^2$	0.704	1.188	0.827	0.544	0.413171
$C = -615.681 + (25.600)H - (0.153)H^2$	0.999	475.806	6.051	0.032	3.0262032
$C = 26.598 + (5.018)H$	0.958	45.340	27.142	0.021	19.193309
$TPA = 49.861 - (0.008)S + (1.334E^{-6})S^2$	0.783	1.799	13.664	0.466	6.9466441
$TPA = 29.462 + (0.005)Gut$	0.751	6.031	10.340	0.133	7.3111917



**Figure 8.** Plots of Linear and Quadratic Regression Equations for the Best Physicochemical Properties Predicted by Distance-Based Topological Indices.

Table 4 shows the correlation coefficients ( $R$ ) obtained by a linear regression model between various topological indices and physicochemical properties of fibrate drugs. The topological indices include degree-based topological indices. The results show that the correlation coefficients vary across the different topological indices and physicochemical properties. A positive correlation indicates two variables that tend to move strongly in opposite directions, while a negative correlation indicates two variables that move strongly in opposite directions. In particular, for the first Zagreb index  $M_1(\zeta)$ , the correlation coefficient lies between 0.740 and 1, with the best prediction for complexity ( $C$ ) being 1. For the second Zagreb index  $M_2(\zeta)$ , the range of the correlation coefficient is  $0.729 \leq R \leq -0.941$ , which indicates a high prediction of all physicochemical properties under study. The highest correlation coefficient values were observed for the ( $SEZ_pE$ ) property with values ranging from 0.887 to 0.998, followed by the ( $TPA$ ) index, with values ranging from 0.786 to 0.826. The other topological indices showed weaker correlations with the physicochemical properties, with correlation coefficients ranging from 0.647 to 0.967 for the remaining indices. Table 8 provides a list of five linear regression models and their corresponding  $R^2$  and  $RMSE$  values.  $R^2$ , or the coefficient of determination, is a measure of how well the independent variables in a linear regression model explain the variation in the dependent variable. It ranges from 0 to 1, with 1 indicating a perfect fit.  $RMSE$ , or root mean squared error, is a measure of how well the regression model's predictions match the actual values. It represents the average distance between the predicted and actual values, and lower values indicate better accuracy. All five models have relatively high  $R^2$  values, indicating that they explain a significant amount of the variation in the dependent variable. The lowest  $R^2$  value is 0.639, which is still considered a relatively good fit. However, the models have different levels of prediction accuracy, as measured by

*RMSE*. The *XLogP3* with min-max rodeg index  $mM_{s-de}(\zeta)$  model has the lowest *RMSE* value of 0.495, which suggests that it has the most accurate predictions among the five models. The *C* model with the first Zagreb index  $M_1$  has the second lowest *RMSE* value of 1.156, followed by the *SEZ<sub>p</sub>E* model with an *RMSE* of 6.423. The *TPA*(*ABC* index) and *TPA*(*R* index) models have the highest *RMSE* values of 8.260 and 8.250, respectively, indicating that their predictions are the least accurate among the five models. In summary, while all five models have relatively high  $R^2$  values, indicating good fit to the data, the *XLogP3* model is the most accurate based on its low *RMSE* value, followed by the *C* and *SEZ<sub>p</sub>E* models, and then the *TPA* (*ABC* index) and *TPA* (*R* index) models, which have the highest *RMSE* values.

Table 5 presents the correlation coefficients ( $R$ ) obtained by a quadratic regression model between the topological indices and physicochemical properties of various drugs of fibrate. Upon analyzing the data in Table 5, several noteworthy findings can be observed. Firstly, many of the correlation coefficients ( $R$ ) are relatively high, indicating a strong linear relationship between the topological indices and physicochemical properties of the fibrate drugs. For instance,  $\sigma(\zeta)$  has a high correlation coefficient of 0.947 with (*DM*), indicating a strong positive linear relationship between these two variables. Similarly,  $Mm_{s-de}(\zeta)$  has a high correlation coefficient of 0.997 with (*P*), suggesting a strong positive linear relationship between these variables as well. Furthermore, some of the correlation coefficients are close to 1, indicating a perfect positive linear relationship between the variables. For example,  $M_1$ , *ABC*, *R*,  $mM_{s-de}$ , and *ISI* indices have a correlation coefficient of 1.000 with (*C*), suggesting a perfect positive linear relationship between these two variables. Similarly, *irr* index has a correlation coefficient of 1.000 with *SEZ<sub>p</sub>E*, *SET<sub>Energy</sub>*, *SET<sub>Enthalpy</sub>*, and *SET<sub>F<sub>Energy</sub></sub>*, indicating a perfect positive linear relationship between these variables. On the other hand, some correlation coefficients are relatively low, indicating a weak linear relationship between the variables. For instance, the *ISDI* index has a correlation coefficient less than 0.64 for most of the properties except for (*C*) ( $R = 0.922$ ) and (*TPA*) ( $R = 0.882$ ), suggesting a weak positive linear relationship between these two variables. It is also interesting to note that we do not have any negative values which would indicate an inverse relationship between the variables. In addition, some of the correlation coefficients are moderate, suggesting a moderate linear relationship between the variables. For instance, (*TPA*) has a correlation coefficient of ( $0.712 \leq R \leq 0.882$ ), indicating a moderate positive linear relationship between these variables. Overall, the findings from Table 5 suggest that there are varying degrees of linear relationships between the topological indices and physicochemical properties of fibrate drugs. Some of the relationships are strong, while others are weak or moderate. Looking at Table 9, we see that all five models for the complexity property (*C*) have high  $R^2$  values, with the lowest being 0.999 and the highest being 1.000. This suggests that all five models are good at explaining the variation in the physicochemical property they are modeling. The second thing to consider is the *RMSE* value; a lower *RMSE* value indicates that the model has a better fit. In this table, we can see that the *RMSE* values range from 0.710303 to 5.291185. The model with the lowest *RMSE* value is the second model:  $C = -22.6208403R^2 + 485.9459637R - 132.8797476$  for the Randic index. This indicates that this model has the best fit for estimating the physicochemical property. However, it is important to note that all five models have high  $R^2$  values, suggesting that they all provide good estimates for the physicochemical property. After analyzing the table, we found that there are five quadratic regression models with both high  $R^2$  values and low *RMSE* values. The quadratic regression model for *S* has a high  $R^2$  value of 0.999 and a low *RMSE* value of 0.732681, making it one of the best models in terms of accurately predicting the target variable. The other models are for (*Z<sub>p</sub>VE*), (*CV*), (*SEZ<sub>p</sub>E*), and (*P*). The model for (*Z<sub>p</sub>VE*) has an  $R^2$  value of 0.996 and an *RMSE* of 2.193392, the model for (*CV*) has an  $R^2$  value of 0.999 and an *RMSE* of 0.540583, the model for (*SEZ<sub>p</sub>E*) has an  $R^2$  value of 0.999 and an *RMSE* of 5.68161, and the model for (*P*) has an  $R^2$  value of 0.995 and an *RMSE* of 3.057903. These models can be considered the best in terms of their ability to fit the data and accurately predict the target variable.

Table 6 presents the correlation coefficient ( $R$ ) obtained by cubic regression models between topological indices and physicochemical properties of various drugs of fibrates. Looking at the table, we can see that the range of correlation coefficients varies for each row. For instance, the correlation coefficient for the row of the first Zagreb index ( $M_1$ ) ranges from 0.811 to 1.0, while for the row inverse symmetric deg index ( $ISDI$ ), the correlation coefficient ranges from 0.650 to 0.970. Overall, most of the correlation coefficients are relatively high, with many of them being close to 1.0. This suggests a strong correlation between the topological indices and the physicochemical properties of the drugs of fibrates. The high correlation coefficients could indicate that the topological indices could be used to predict the physicochemical properties of the drugs with high accuracy. Based on Table 10, it appears that the cubic regression model provides the highest correlation coefficients for most of the topological indices and physicochemical properties of fibrate drugs. The range of correlation coefficients for each row varies, but in general, they are relatively high, indicating a strong relationship between the topological indices and physicochemical properties. Furthermore, the high correlation coefficients suggest that the cubic regression model is an effective tool for predicting physicochemical properties based on the topological indices of fibrate drugs. Overall, the results of the table suggest that the cubic regression model is the best choice for analyzing the relationship between topological indices and physicochemical properties in fibrate drugs. Based on Table 9, we can analyze the four topological indices with respect to high  $R^2$  and minimum  $RMSE$ . ( $XLogP3$ ) ( $RMSE = 0.00001$ ,  $R^2 = 0.900$ ) indicating a strong correlation between the physicochemical properties and this index. Additionally, its  $RMSE$  value of 0.00001 is also very low, suggesting that the predicted values using this index are very close to the actual values. ( $TPA$ ) ( $RMSE = 0.00003$ ,  $R^2 = 1.000$ ) indicating a perfect correlation with the physicochemical properties.

By looking closely at Table 7, considering only the distance-based topological indices, we can notice that the model which gives the highest correlations with all the investigated physicochemical properties of fibrate drugs is the cubic model, since the correlation coefficients range from 0.750 to 1.000. In second place is the quadratic model, since it gives good correlations with most of these properties, and the correlation coefficients range from 0.750 to 0.999. While the linear model comes in the third place, it shows good correlation but with the least number of properties, and the correlation coefficients range 0.688 to 0.979. An important note is that in most cases, the linear and quadratic models give comparable correlation coefficients, while there is a significant improvement in the correlation coefficients when the cubic model is used for most of properties. For instance, for the polarizability ( $P$ ) property estimated using the Wiener index, correlations are comparable,  $R = 0.185$  and  $R = 0.322$  for the linear and quadratic models, respectively, and it improves to 1 with the cubic model. As a result, we should consider our model type when dealing with such properties. Generally speaking, the four properties at the end of Table 7 are estimated very well with the three models compared to the first five properties in the table. The complexity ( $C$ ) property can be best estimated using the various models, since the correlations with each model reach  $\sim 1$ . The topological polar area ( $TPA$ ) can be nominated as the second-best estimated property by the three models, followed by sum of electronic and zero-point energies ( $SEZ_pE$ ) property. Conversely, the zero-point vibrational energy ( $Z_pVE$ ) and heat capacity ( $CV$ ) properties seems to be the properties which can be estimated the least accurately using the two models (linear and quadratic), the correlations not exceeding 0.345. The exception is the quadratic model of the hyper-Zagreb index  $H(\zeta)$ ,  $R = 0.824$  and 0.937, respectively. Based on the  $RMSE$  values given in Table 11, the three best predictors with the lowest  $RMSE$  values are: linear regression ( $DM = -2.010 + (0.003)Gut$ ) with  $RMSE = 0.9007$ , quadratic regression ( $P = -1200.200 + (44.294)H - (0.326)H^2$ ) with  $RMSE = 12.2392$ , and curvilinear regression ( $XLogP3 = 9.019 - (0.202)H + (0.002)H^2$ ) with  $RMSE = 0.4131$ . These three regression models exhibit the lowest  $RMSE$  values, indicating higher accuracy and better predictive performance compared to the other regression models. Therefore, these three regression models, namely, linear, quadratic, and curvilinear, can be considered as the best predictors for enhancing the analysis of fibrate drug activity

through molecular descriptors in this study. Therefore, based on the results obtained, it can be concluded that the cubic and quadratic regression models are the top predictors for the physicochemical properties analyzed in this investigation, as they exhibit both high  $R^2$  values and minimum *RMSE* values simultaneously. These findings highlight the effectiveness of these regression models in enhancing the analysis of fibrate drug activity through molecular descriptors and provide valuable insights for future research in this area.

#### 4. Conclusions

Based on our comprehensive analysis, we have demonstrated that the use of curvilinear regression models can significantly enhance the analysis of fibrate drug activity through molecular descriptors. Our results have revealed that these models have superior predictive power compared to linear regression models, especially when the underlying data exhibit nonlinear relationships. Furthermore, the incorporation of molecular descriptors as independent variables has substantially improved the accuracy and robustness of the models. Our findings have several important implications for the field of drug discovery and development. Firstly, the use of curvilinear regression models, in conjunction with molecular descriptors, can facilitate the identification and optimization of more potent and selective drugs, thus reducing the time and cost associated with drug development. Secondly, our study underscores the importance of considering nonlinear relationships between molecular descriptors and drug activity, which has traditionally been overlooked in conventional linear regression analyses. Lastly, the efficacy of curvilinear regression models and molecular descriptors in predicting drug activity may be extended to other drug classes and further elucidated through future studies. In summary, our investigation demonstrates that curvilinear regression models represent a powerful approach for analyzing drug activity, particularly when coupled with molecular descriptors. Our results provide a basis for the development of improved drug discovery pipelines and offer insights into the molecular mechanisms governing drug activity. In summary:

- Despite the limited number of input molecules used in our study, we have taken great care to ensure the reliability and validity of our findings. We have rigorously tested our models using appropriate statistical methods and validated their predictive performance through external testing. Furthermore, we have provided a clear and transparent description of our methodology, including the selection and preparation of our data, the choice of input features, and the modeling approach. We believe that our manuscript reflects a well-designed and carefully executed study that contributes to the field of predictive modeling.
- While we acknowledge the limitation of our small dataset, we would like to emphasize that our study is not meant to provide a definitive model for predicting molecular properties. Rather, it aims to demonstrate the feasibility and potential of using DFT calculations and topological indices as input features for predictive modeling. Our results show promising predictive performance and highlight the importance of selecting appropriate input features and modeling approaches. Our study will inspire further investigations on larger datasets and lead to the development of more robust and accurate models.
- By evaluating three distinct models, we have provided a comprehensive and nuanced analysis of the relationship between molecular structure and properties. Our models include both linear and non-linear approaches, which allowed us to capture both linear and non-linear relationships between input features and output properties. This approach is particularly important in the field of predictive modeling, where complex relationships are often present. Moreover, by comparing and contrasting the performance of different models, we were able to identify the most effective approach for our specific research question. Our study demonstrates the importance of model selection and the need for careful evaluation of different modeling approaches.

**Author Contributions:** Conceptualization, S.W. and N.U.O.; methodology, S.W.; validation, S.W. and N.U.O.; formal analysis, S.W.; investigation, S.W.; resources, N.U.O.; data curation, N.U.O.; writing—original draft S.W.; preparation, S.W.; writing—review and editing, S.W. and N.U.O.; supervision, S.W.; project administration, S.W.; funding acquisition, S.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 214-247-1443). The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

**Data Availability Statement:** The article contains the data that supported the study's findings.

**Acknowledgments:** This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 214-247-1443). The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia. The authors acknowledge Nuha Wazzan from the Chemistry department at King Abdulaziz University for her contribution with the DFT calculations, and King Abdulaziz University's High-Performance Computing Centre (Aziz Supercomputer) (<http://hpc.kau.edu.sa>) for supporting the computation for the work described in this paper.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## References

1. Gonzalez-Diaz, H.; Vilar, S.; Santana, L.; Uriarte, E. Medicinal chemistry and bioinformatics-current trends in drugs discovery with networks topological indices. *Curr. Top. Med. Chem.* **2007**, *7*, 1015–1029. [[CrossRef](#)] [[PubMed](#)]
2. Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2001**, *8*, 1573–1588. [[CrossRef](#)] [[PubMed](#)]
3. Gao, W.; Wang, W.; Farahani, M.R. Topological indices study of molecular structure in anticancer drugs. *J. Chem.* **2016**, *2016*, 3216327. [[CrossRef](#)]
4. Gao, W.; Farahani, M.R.; Shi, L. Forgotten topological index of some drug structures. *Acta Medica Mediterr.* **2016**, *32*, 579–585.
5. McCullough, P.A.; Loreto, M.J.D. Fibrates and cardiorenal outcomes. *J. Am. Coll. Cardiol.* **2012**, *60*, 2072–2073. [[CrossRef](#)] [[PubMed](#)]
6. Brea, A.; Millán, J.; Ascaso, J.F.; Blasco, M.; Díaz, A.; Hernxandez-Mijares, A.; Mantilla, T.; Pedro-Botet, J.C.; Pintó, X. Fibrates in primary prevention of cardiovascular disease. Comments on the results of a systematic review of the Cochrane Collaboration. *Clínica Investigación Arterioscler. (Engl. Ed.)* **2018**, *30*, 188–192. [[CrossRef](#)]
7. Devillers, J.; Balaban, A.T. *Topological Indices and Related Descriptors in QSAR and QSPAR*; CRC Press: Boca Raton, FL, USA, 2000.
8. Gutman, I. A Property of the Simple Topological Index. *Match Commun. Math. Comput. Chem.* **1990**, *25*, 131–140.
9. Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem.* **1947**, *69*, 17–20. [[CrossRef](#)]
10. Gao, W.; Wang, Y.; Basavanagoud, B.; Jamil, M.K. Characteristics Studies of Molecular Structures in Drugs. *Saudi Pharm. J.* **2017**, *25*, 580–586. [[CrossRef](#)]
11. Doslic, T.; Reti, T.; Ali, A. On the Structure of Graphs with Integer Sombor Indices. *Discret. Lett.* **2021**, *7*, 1–4.
12. Gutman, I. Geometric Approach to Degree-Based Topological Indices: Sombor Indices. *Match Communications Math. Comput. Chem.* **2021**, *86*, 11–16.
13. Ediz, S.; Çiftçi, İ.; Cancan, M.; Farahani, M.R. On k-total distance degrees and k-total Wiener polarity index. *J. Inf. Optim. Sci.* **2021**, *42*, 1469–1477. [[CrossRef](#)]
14. Matejić, M.; Zogić, E.; Milovanović, E.; Milovanović, I. A Note on the Laplacian Resolvent Energy of Graphs. *Asian-Eur. J. Math.* **2020**, *13*, 2050119. [[CrossRef](#)]
15. Gutman, I.; Trinajstić, N. Graph Theory and Molecular Orbitals. Total p-Electron Energy of Alternant Hydrocarbons. *Chem. Phys. Lett.* **1972**, *17*, 535–538. [[CrossRef](#)]
16. Randić, M. On Characterization of Molecular Branching. *J. Am. Chem.* **1975**, *97*, 6609–6615. [[CrossRef](#)]
17. Estrada, E. Characterization of 3D Molecular Structure. *Chem. Phys. Lett.* **2000**, *319*, 713–718. [[CrossRef](#)]
18. Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339. [[CrossRef](#)]
19. Estrada, E.; Bonchev, D. *Chemical Graph Theory*; Chapman and Hall/CRC: New York, NY, USA, 2013.
20. Gutman, I.; Ruscic, B.; Trinajstić, N.; Wilson, C.F., Jr. Graph theory and molecular orbitals. XII. Acyclic polyenes. *J. Chem. Phys.* **1975**, *62*, 3399–3405. [[CrossRef](#)]
21. Shirdel, G.H.; Rezapour, H.; Sayadi, A.M. The hyper Zagreb index of graph operations. *Iran. J. Math. Chem.* **2013**, *4*, 213–220.
22. Togan, M.; Yurttas, A.; Cevik, A.S.; Cangul, I.N. Effect of edge deletion and addition on Zagreb indices of graphs. In *Mathematical Methods in Engineering*; Springer: Cham, Switzerland, 2019; pp. 191–201.
23. Togan, M.; Yurttas, A.; Çevik, A.S.; Cangul, I.N. Zagreb indices and multiplicative Zagreb indices of double graphs of subdivision graphs. *Twms J. Appl. Eng. Math.* **2019**, *9*, 404–412.

24. Gutman, I.; Togan, M.; Yurttas, A.; Cevik, A.S.; Cangul, I.N. Inverpe problem for sigma index. *Match Commun. Math. Comput. Chem.* **2018**, *79*, 491–508.
25. Ghorbani, M.; Zangi, S.; Amraei, N. New results on symmetric division deg index. *J. Appl. Math. Comput.* **2021**, *65*, 161–176. [[CrossRef](#)]
26. Richardson, C.W.; Foster, G.R.; Wright, D.A. Estimation of erosion index from daily rainfall amount. *Trans. Asae* **1983**, *26*, 153–156. [[CrossRef](#)]
27. Das, K.C.; Gutman, I.; Furtula, B. On atom-bond connectivity index. *Chem. Phys. Lett.* **2011**, *511*, 452–454. [[CrossRef](#)]
28. Dalfó, C. On the Randić index of graphs. *Discret. Math.* **2019**, *342*, 2792–2796. [[CrossRef](#)]
29. Jahanbani, A. Albertson energy and Albertson Estrada index of graphs. *J. Linear Topol. Algebra* **2019**, *8*, 11–24.
30. Klavžar, S.; Rajapakse, A.; Gutman, I. The Szeged and the Wiener index of graphs. *Appl. Math. Lett.* **1996**, *9*, 45–49. [[CrossRef](#)]
31. Xu, K.; Das, K.C. On Harary index of graphs. *Discret. Appl. Math.* **2011**, *159*, 1631–1640. [[CrossRef](#)]
32. Mukwembi, S. On the upper bound of Gutman index of graphs. *Match-Commun. Math. Comput. Chem.* **2012**, *68*, 343.
33. Havare, O.Ç. Topological indices and QSPR modeling of some novel drugs used in the cancer treatment. *Int. J. Quantum Chem.* **2021**, *121*, e26813. [[CrossRef](#)]
34. Vukicevic, D. Bond additive modeling 2. Mathematical properties of max-min rodig index. *Crica Chem. Actata* **2010**, *83*, 261–273.
35. Estrada, E.; Torres, L.; Rodríguez, L.; Gutman, I. An atom-bond connectivity index: Modelling the enthalpy of formation of alkanes. *Indian J. Chem.* **1998**, *37*, 849–855.
36. Rajasekharaiyah, G.V.; Murthy, U.P. Hyper-Zagreb indices of graphs and its applications. *J. Algebra Comb. Discret. Struct. Appl.* **2020**, *8*, 9–22. [[CrossRef](#)]
37. Vukiccević, D.; Gasparov, M. Bond additive modeling 1. Adriatic indices. *Crica Chem. Actata* **2010**, *83*, 243–260.
38. Çolaköglü Havare, Ö. Determination of some thermodynamic properties of monocarboxylic acids using multiple linear regression. *Beu J. Sci.* **2019**, *8*, 466–471. [[CrossRef](#)]
39. Loksha, V.; Shruti, R.; Sinan, Cevik, A. On certain topological indices of Nanostructures using  $Q(G)$  and  $R(G)$  operators. *Commun. Fac. Sci. Univ. Ank. Ser. Math. Stat.* **2018**, *67*, 178–187.
40. Reti, T.; Sharafdini, R.; Drégelyi-Kiss, A.; Haghbin, H. Graph irregularity indices used as molecular descriptors in QSPR studies. *Match Commun. Math. Comput. Chem.* **2018**, *79*, 509–524.
41. Wiener, H. Relation of the physical properties of the isomeric alkanes to molecular structure Surface tension, specific dispersion, and critical solution temperature in aniline. *J. Phys.* **2018**, *52*, 1082–1089. [[CrossRef](#)]
42. Dobrynin, A.A.; Entringer, R.; Gutman, I. Wiener index of trees: Theory and applications. *Acta Appl. Math.* **2001**, *66*, 211–249. [[CrossRef](#)]
43. Castro, E.A.; Tueros, M. QSPR Study of boiling points of alkyl alcohols via improved polynomial relationships. *Philipp. J. Sci.* **2001**, *130*, 111–118.
44. Delen, S.; Khan, R.H.; Kamran, M.; Salamat, N.; Baig, A.Q.; Naci Cangul, I.; Pandit, M.K. Ve-Degree, Ev-Degree, and Degree-Based Topological Indices of Fenofibrate. *J. Math.* **2022**, *2022*, 4477808. [[CrossRef](#)]
45. Berinde, Z.M. QSPR Models for the Molar Refraction, Polarizability and Refractive Index of Aliphatic Carboxylic Acids Using the ZEP Topological Index. *Symmetry* **2021**, *13*, 2359. [[CrossRef](#)]
46. Zuo, J.; Hu, L. QSPR modeling of the melting points of organic compounds using molecular topology and quantum chemical descriptors. *Symmetry* **2020**, *12*, 1104.
47. Zhang, Y.; Li, H.; Liu, Y.; Zhou, P. QSPR models for predicting melting points of organic compounds based on molecular topology. *Symmetry* **2019**, *11*, 25. [[CrossRef](#)]
48. Naghipour, S.; Kiasat, A.R. Application of topological indices in QSPR modeling of C60 derivatives' fullerene-like behavior. *Symmetry* **2019**, *11*, 368.
49. Wang, J.; Xu, L. QSPR models for predicting the boiling points of alkyl alkanes based on the novel vertex degree valence topological index. *Symmetry* **2018**, *10*, 282.
50. Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785. [[CrossRef](#)] [[PubMed](#)]
51. Frisch, M.J. *Gaussian 09 Programmer's Reference*; Gaussian: Wallingford, CT, USA, 2009; 25p.
52. Dennington, R.; Keith, T.; Millam, J. *GaussView Version 5*; Semichem Inc.: Shawnee Mission, KS, USA, 2009.
53. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **2008**, *4*, 17. [[CrossRef](#)]
54. O'boyle, N.M.; Tenderholt, A.L.; Langner, K.M. Cclib: A library for package-independent computational chemistry algorithms. *J. Comput. Chem.* **2008**, *29*, 839–845. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.