Physicochemical Properties of Some Drugs by Using Machine Learning

MS Thesis

by

Amman Farzeen

CIIT/SP24-RMT-023/LHR

COMSATS University Islamabad
Pakistan

Fall 2025

Physicochemical Properties of Some Drugs by Using Machine Learning A thesis submitted to

COMSATS University Islamabad

In partial fulfillment
of the requirement for the degree of

Master of Science
in
Mathematics

by
Amman Farzeen

CIIT/SP24-RMT-023/LHR

Department of Mathematics
Faculty of Science

COMSATS University Islamabad
Pakistan

Fall 2025

Physicochemical Properties of Some Drugs by Using Machine Learning

This thesis is submitted to the department of Mathematics in partial fulfillment of the requirements for the award of the degree of Master of Science in Mathematics

Name	Registration Number
Amman Farzeen	CIIT/SP24-RMT-023/LHR

Supervisory Committee

Supervisor

Dr. Kashif Ali
Professor
Department of Mathematics
COMSATS University Islamabad (CUI)
Lahore Campus

Certificate of Approval

This thesis titled

Physicochemical Properties of Some Drugs by Using Machine Learning

By

Amman Farzeen

CIIT/SP24-RMT-023/LHR

Has been approved

For COMSATS University Islamabad, Lahore Campus.

External Examiner:_____

Prof. Dr. Muhammad Ozair Ahmad

University of Lahore(UOL)

Supervisor:_____

Prof. Dr. Kashif Ali

Department of Mathematics, (CUI) Lahore Campus

Head of Department:_____

Prof. Dr. Muhammad Hussain

Department of Mathematics, (CUI) Lahore Campus

Author's Declaration

I Amman Farzeen, CIIT/SP24-RMT-023/LHR, hereby declare that I have produced the work presented in this thesis, during the scheduled period of study. I also declare that I have not taken any material from any source except referred to wherever due to that amount of plagiarism is within an acceptable range. If a violation of HEC rules on research has occurred in this thesis, I shall be liable to punishable action under the plagiarism rules of HEC.

Date:_____ _____

Amman Farzeen
CIIT/SP24-RMT-023/LHR

Certificate

It is certified that Amman Farzeen, CIIT/SP24-RMT-023/LHR has carried out all the work related to this thesis under my supervision at the Department of Mathematics, COMSATS University Islamabad, Lahore Campus and the work fulfills the requirements for award of M.S degree.

Date:_____

Supervisor

_____

Dr. Kashif Ali
Professor,
Department of Mathematics,
CUI, Lahore Campus

Dedication

To my dearest parents,

With all my heart, I dedicate this thesis to you, my pillars of strength, my quiet dream weavers, and the unwavering light that guided every late-night study session and every moment of doubt. Your endless sacrifices, tender encouragement, and boundless belief in me turned an impossible journey into reality. You taught me that love is the most significant force, and hard work is its most beautiful expression. This work is but a small reflection of the values you instilled in me.

Forever grateful, forever yours.

This thesis belongs to you.

Acknowledgements

First and foremost, all praise and gratitude belong to Allah, the Most Gracious, the Most Merciful, who blessed me with strength, patience, and wisdom to complete this thesis on time, and to the Holy Prophet Muhammad (Peace Be Upon Him), the perfect role model and mercy for all mankind. I express my deepest gratitude to my supervisor, Dr. Kashif Ali, for his inspiring guidance, constant motivation, and invaluable suggestions throughout this journey. I am also sincerely thankful to the Head of the Department of Mathematics, Prof. Dr. Muhammad Hussain, for fostering an excellent and supportive research environment essential for the successful completion of this work. My heartfelt thanks go to my respected teacher, Mr. Wakeel Ahmed, whose continuous encouragement and belief in me never wavered. Above all, this achievement would not have been possible without the endless prayers, unwavering support, and boundless love of my parents, who stood by me every step of the way.

Amman Farzeen
CIIT/SP24-RMT-024/LHR

Abstract

Physicochemical Properties of Some Drugs by Using Machine Learning

By

Amman Farzeen

This study explores the integration of machine learning and multi-criteria decision-making techniques for the characterization and evaluation of potential therapeutics targeting Duchenne Muscular Dystrophy . A dataset comprising the physicochemical properties of twenty anti-Duchenne muscular dystrophy drug candidates is analyzed using topological indices derived from molecular graphs. These indices, computed via Python-based algorithms, served as input features for machine learning models. Artificial Neural Networks and Random Forests are employed to predict key physicochemical properties based on the structural descriptors. The results demonstrate that topological indices are effective predictors, enabling accurate and consistent performance across both models. To prioritize the drug candidates, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is applied as an multi-criteria decision-making tool, leveraging the model predictions to rank compounds based on multiple beneficial attributes. The proposed framework highlights the synergy between graph-theoretical descriptors, machine learning, and multi-criteria decision-making, offering a robust in-silico approach for early-stage drug evaluation. This methodology provides a systematic and objective strategy for ranking therapeutic candidates, particularly valuable in the context of rare diseases such as duchenne muscular dystrophy.

Key words: Machine learning; Molecular Graph; QSPR Analysis; Topological indices; Artificial Neural Networks; Random Forest; TOPSIS; Python Algorithm; Anti-DMD drugs.

Table of Contents

1 Introduction **1**
1.1 Historical View of Graph Theory 1
1.2 Background of the Study 2
1.3 Basic Definitions 2
1.3.1 Vertex 3
1.3.2 Edge 3
1.3.3 Order and Size of Graph 3
1.3.4 Simple Graph 4
1.4 An Overview of Topological Indices: 4
1.5 Literature Review 5

2 Predictive Modeling of Anti-DMD Drugs Properties Using Topological Descriptors and Machine Learning **7**
2.1 Results and Discussions 12
2.2 Supervised Machine Learning Techniques 19
2.2.1 Artificial Neural Network (ANNs) 19
2.2.2 Random Forest Algorithm (RF) 25
2.3 Computed vs. Actual: Physicochemical Properties Analysis 32
2.3.1 Assessing ANN and RF Predictions 32

3 Multi-Criteria Drug Ranking Using the TOPSIS Method. **37**
3.0.1 Create decision matrix 37
3.0.2 Normalize all criteria 38
3.0.3 Assign weights 39
3.0.4 Find Ideal Best and Ideal Worst solutions 40
3.0.5 Calculate Euclidean distance 41
3.0.6 Compute closeness score 41
3.0.7 Rank 41

4 Conclusion **45**

List of Figures

Figure1.1 Kneiphof bridge and their respective graph 1
Figure1.2 Graph 3
Figure1.3 A graph with $|V(G)| = 5$ and $|E(G)| = 7$. 3
Figure1.4 graph (a) has a loop (b) has multiple edges (c) has no loop and multiple edges so it is a simple graph 4

Figure2.1 Duchenne muscular dystrophy(source:https://www.mda.org/disease/duchenne-muscular-dystrophy/causes-inheritance) 8
Figure2.2 X-linked pattern of inheritance 8
Figure2.3 2D chemical structures of twenty Anti-DMD drugs 10
Figure2.4 Simple graph of Deflazacort, L-Citrulline and Epicatechin 11
Figure2.5 Graphical Representation Of Feature Importance For Density 22
Figure2.6 First Decision Trees For Surface Tension 26
Figure2.7 Second Decision Trees For Surface Tension 26
Figure2.8 Third Decision Trees For Surface Tension 27
Figure2.9 Fourth Decision Trees For Surface Tension 27
Figure2.10 First Decision Trees For Molar Volume 28
Figure2.11 Second Decision Trees For Molar Volume 28
Figure2.12 Third Decision Trees For Molar Volume 29
Figure2.13 Fourth Decision Trees For Molar Volume 29
Figure2.14 Comparison of actual values as predictions(ANNs and RF) for density and boiling point of anti-DMD drugs 33
Figure2.15 Comprehensive performance evaluation of ANNs and RF for density and boiling point predictions 33
Figure2.16 Comparison of actual values as predictions(ANNs and RF) for enthalphy and flash point of anti-DMD drugs 34
Figure2.17 Comprehensive performance evaluation of ANNs and RF for Enthalphy and Flash Point predictions 34
Figure2.18 Comparison of actual values as predictions(ANNs and RF) for index refraction and molar reactivity of anti-DMD drugs 34
Figure2.19 Comprehensive performance evaluation of ANNs and RF for Index Refraction and Molar Reactivity predictions 34
Figure2.20 Comparison of actual values as predictions(ANNs and RF) for polar surface and polarizabilty of anti-DMD drugs 35
Figure2.21 Comprehensive performance evaluation of ANNs and RF for Polar Surface and Polarizabilty predictions 35
Figure2.22 Comparison of actual values as predictions(ANNs and RF) for suface tension and molar volume of anti-DMD drugs 35
Figure2.23 Comprehensive performance evaluation of ANNs and RF for Surface Tension and Molar Volume predictions 35

Figure3.1 Graphical analysis of ranking anti-DMD drugs via TOPSIS 44

List of Abbreviations

DMD	Duchenne Muscular Dystrophy
QSPR	Quantitative Structure-Property Relationships
QSAR	Quantitative Structure-Activity Relationships
ANNs	Artifical Neural Networks
RF	Random Forest
D	Density
BP	Boiling Point
FP	Flash Point
IR	Index of Refrection
MR	Molar Refractivity
POL	Polarizability
ST	Surface Tension
MV	Molar Volume
PSA	Polar Surface Area
EV	Enthalpy of Vaporization
MSE	Mean Squared Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
R^2	Coefficient of determination
MCDM	Multi-Criteria Decision Making
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution

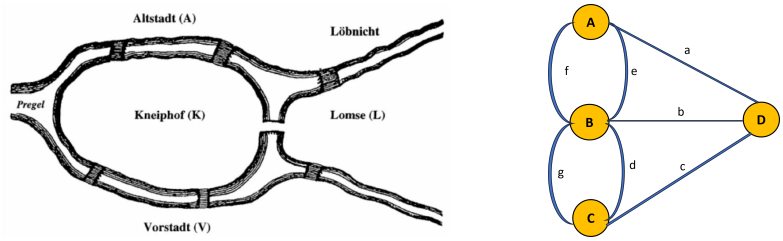
Chapter 1
Introduction

1.1 Historical View of Graph Theory

Leonhard Eulers

Graph theory was founded by the Swiss mathematician Leonhard Eulers solution of the Konigsberg Bridge Problem in 1736. He is recognised as the father of graph theory. The city of Konigsberg had seven bridges connecting four land masses as shown in Figure 1.1. The question was whether it was possible to walk across all seven bridges exactly once and return to the starting point. Euler proved it was impossible.

Graph theory is currently used in most fields of science, including Computer science, networks, chemical graphs, Electrical engineering, operations research, sewerage systems, traffic flow, and telephone lines. It is also potentially applied to an optimisation problem. Graph theory has several uses in that. Help a great deal for humanity. Graph theory is a part of mathematics, though its origins date back. Majoring in Economics, Biology, Statistics, Architecture, Communication, Management, Mechanical, Civil and Chemical engineering.



(a) **(b)**

Figure 1.1: Kneiphof bridge and their respective graph

Despite the graphical nature of chemical structures, the systematic use of graph theory in chemistry began with the introduction of the Wiener index, the first topological index, in 1947 by Harry Wiener, who showed that the index correlated strongly with the physicochemical properties of alkanes. In 1975, the Randic connectivity index, the Zagreb indices, and many others were introduced, and topological indices became the primary pinnacle de-

scriptors in QSAR/QSPR research. Hundreds of graph-theoretical indices have since been developed. With the advent of machine learning in the 21st century, these simple, interpretable, and computationally efficient descriptors remain an essential component, alongside modern deep graph neural networks, for predicting the physicochemical and biological properties of drug molecules.

1.2 Background of the Study

The discovery and development of a drug is a very resource-intensive process, and the average cost of developing a drug is over 2.6 billion USD The process takes 10-15 years before a drug is approved. One of the critical bottlenecks is the early identification of candidates with desirable physicochemical and ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profiles. Aqueous solubility, melting point, polar surface area and molecular volume are directly related to drug-likeness, bioavailability, and formulation feasibility. These properties are not only costly, time-consuming and often impractical to determine experimentally during the screening of large virtual libraries, comprising millions of compounds.

Chemoinformatics offers a strong alternative approach in the form of Quantitative Structure-Property Relationship (QSPR) modelling, which develops mathematical correlations between molecular structure and physicochemical properties. The key to QSPR research is molecular descriptors: numerical values that capture structural, electronic, topographic, or thermodynamic properties of molecules. Among thousands of possible descriptors, graph-based topological indices of molecular structures have become popular because of their conformational invariance, computational efficiency, and generally good predictive performance for many macroscopic properties.

1.3 Basic Definitions

A graph G is an ordered pair $(V(G), E(G))$ which contains the set of vertices $V(G)$ also called nodes and the set of edges $E(G)$ also called lines. In which each edge joins exactly two nodes as shown in Figure 2.1.

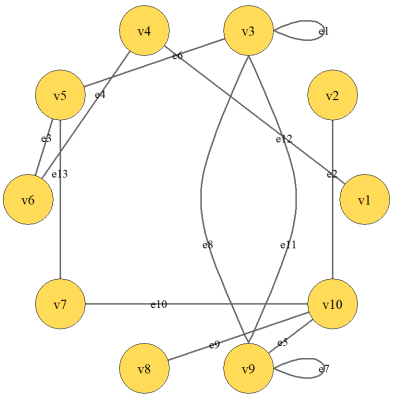


Figure 1.2: Graph

1.3.1 Vertex

A vertex is a common point which makes connection between two or more edges. It is represented by V. In the Figure 1.3 the set of vertices is $\{v_1, v_2, v_3, ..., v_{10}\}$.

1.3.2 Edge

Edge represents the connection between vertices of a graph, without vertices we cannot draw an edge. In the Figure 1.3 the set of edges is $\{e_1, e_2, e_3, ..., e_{13}\}$.

1.3.3 Order and Size of Graph

The total number of vertices in a graph G is called order of a given graph G. it is denoted as $|V(G)|$. The term size is associated with the cardinality (total number of edges present in a graph. The $|E(G)|$ represents the size of graph G. As in Figure 1.3 order and size of the graph is 5 and 7 respectively.

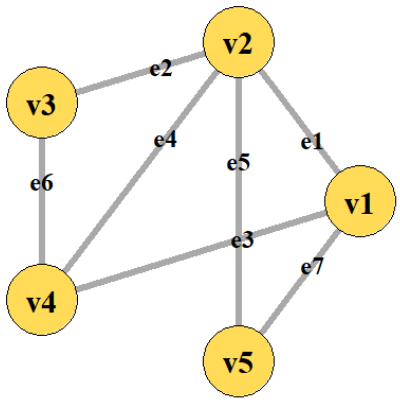
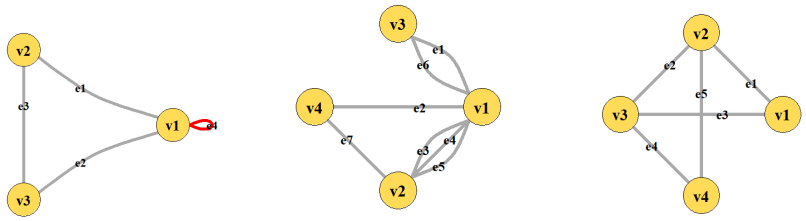


Figure 1.3: A graph with $|V(G)| = 5$ and $|E(G)| = 7$.

1.3.4 Simple Graph

In a graph, a edge joining a vertex to itself is loop while, two or more edges joining the same pair of vertices are multiple edge. Any graph is said to be a simple graph if all of its vertices do not contain any loop and there does not exist multiple edges between all the pairs of vertices. See Figure 1.4.



(a) **(b)** **(c)**

Figure 1.4: graph (a) has a loop (b) has multiple edges (c) has no loop and multiple edges so it is a simple graph

1.4 An Overview of Topological Indices:

Topological indices are numerical descriptors derived solely from the two-dimensional connectivity of a molecule, treating it as a simple hydrogen-suppressed graph in which atoms are vertices and bonds are edges. Topological indices are numerical descriptors that incorporate connectivity information within a molecular network [1, 2, 3, 4, 5], allowing for the quantitative investigation of molecular structures [7, 8, 9]. Degree-based topological indices are important in computational chemistry because they provide useful information about the molecular structures of drugs [10, 11, 12]. These indices are mathematical descriptors derived from atom connection patterns that provide a quantitative depiction of a molecules structure. Two popular degree-based indices are the Wiener index [13] and the Randic index [14]. In the present study, Table 1.1 represents a comprehensive set of such graph-theoretical descriptors extracted from the molecular graphs G.

Table 1.1: Topological indices

Indices Name	Notation	Mathematical formula
First Zagreb Index [15]	$M_1(G)$	$\sum_{pq\in E(G)}(d_p+d_q)$
Second Zagreb Index [15]	$M_2(G)$	$\sum_{pq\in E(G)}(d_p\times d_q)$
Harmonic Index [25]	$H(G)$	$\sum_{pq\in E(G)}\frac{2}{(d_p+d_q)}$
Forgotten Index[17]	$F(G)$	$\sum_{pq\in E(G)}[(d_p)^2+(d_q)^2]$
Sombor-Srednicki Index[18]	SS(G)	$\sum_{pq\in E(G)}\sqrt{\frac{d_p\times d_q}{d_p+d_q}}$
Atom Bond Connectivity Index [19, 20]	ABC(G)	$\sum_{pq\in E(G)}\frac{1}{\sqrt{d_p\times d_q}}$
Randic Index[21]	RI(G)	$\sum_{pq\in E(G)}\left(\frac{d_p\times d_q}{d_p+d_q-2}\right)^3$
Sum Connectivity Index[22]	$SCI(G)$	$\sum_{pq\in E(G)}\frac{1}{\sqrt{d_p+d_q}}$
Geometric Arithmetic Index [23]	$GA(G)$	$\sum_{pq\in E(G)}2\frac{\sqrt{d_p\times d_q}}{d_p+d_q}$
Hyper Zagreb Index[24]	$HZ_3(G)$	$\sum_{pq\in E(G)}(d_p+d_q)^2$
Redefined First Zagreb Index[25]	$RezG_1(G)$	$\sum_{pq\in E(G)}\frac{d_p+d_q}{d_p\times d_q}$
Redefined First Zagreb Index[25]	$RezG_2(G)$	$\sum_{pq\in E(G)}\frac{d_p\times d_q}{d_p+d_q}$
Neighborhood Degree Sum-Based Index[26]	$ND_1(G)$	$\sum_{pq\in E(G)}\sqrt{d_p\times d_q}$
Neighborhood Degree Sum-Based Index[26]	$ND_5(G)$	$\sum_{pq\in E(G)}(\frac{d_p}{d_q}+\frac{d_p}{d_q})$

1.5 Literature Review

Machine learning combined with graph theory has changed the way physicochemical properties of molecules are predicted, making drug discovery and development cycles much faster. Molecular graphs give an intuitive and mathematically beautiful visualization of chemical structures, with atoms and chemical bonds being the vertices and the edges respectively. This graph-theoretic model has facilitated the creation of robust computational modeling techniques to the quantitative structure-property/activity relationship (QSPR/QSAR) modeling [27]. Notably, Wakeel Ahmed et al. explored the role of topological descriptors, specifically topological indices, in predicting molecular properties using supervised machine learning algorithms [28]. Winner first introduced topological indices of molecules in 1947 [29], diverse topological indices employed to characterize graph structure. Some of them discussed, Topological indices, such as First and Second Zagreb index [30], Randic index [31], atom-bond connectivity index, the neighborhood degree-based ND(G) [32] and the Reverse and Reduce Reverse degree-based topological indices [33] and more indices,

quantitatively describe molecular structure and have been successfully applied to predict crucial properties like solubility, lipophilicity and toxicity. Their study extracted topological descriptors from molecular structures and applied machine learning algorithms like support vector machines, random forests, and gradient boosting, achieving impressive results. The advent of graph convolutional networks (GCNs) has been successfully applied to molecular property prediction tasks by Gilmer et al., who introduced GCNs as a type of neural network designed to work directly with graph-structured data [34]. This approach has shown remarkable performance in predicting molecular properties such as solubility, boiling point and toxicity. Furthermore, deep graph learning frameworks have achieved state-of-the-art performance in molecular property prediction, as demonstrated by Li et al., who proposed a framework combining graph convolutional layers and graph attention layers to learn molecular representations [35]. By developing a python algorithm, wakeel at al. [36] found the degree-based topological indices for sulfonamide drugs, Fraz et al. [37] found the distance base indices for some antiviral drugs. The current paper extends this fertile history by critically assessing an entire collection of classical and modern topological indexes alongside the latest machine learning algorithms to offer predictive power on important physicochemical characteristics of drug molecules, where model performance, interpretability of features, and applicability to real-life drug design are considered as the central aspects. The drug-related decision-making process shows practical uses in pharmaceutical supply chain optimization as well as drug selection and toxicity ranking to determine personalized medicine through multi-criteria analysis for data-driven optimal healthcare solutions [38, 39, 40].

Chapter 2
Predictive Modeling of Anti-DMD Drugs Properties Using Topological Descriptors and Machine Learning

Muscular dystrophy (MD) represents a collection of inherited diseases that cause muscles to weaken progressively along with their breakdown. The disease develops because of gene mutations which generate proteins vital for maintaining muscle health. When proteins essential for muscle fibers are missing the cells progressively break apart which causes weakness and functional decline. The MD mostly targets skeletal muscles yet it can harm heart and breathing muscles when the disease reaches severe stages. People with MD develop muscle weakness and problems with walking and experience recurrent falls together with delayed child development which eventually leads to complete immobility as the condition advances[41, 43]. There are almost 30 types of MD. The different types of muscular dystrophy include Duchenne Muscular Dystrophy (DMD) and Becker Muscular Dystrophy (BMD) together with myotonic dystrophy as well as limb-girdle muscular dystrophy (LGMD) and facioscapulohumeral muscular dystrophy (FSHD) and congenital muscular dystrophy (CMD). The distinct categories of MD differ concerning their genetic origins as well as when symptoms start and how they evolve. DMD represents the most significant and common form of muscular disorders affecting male patients through an X-linked inheritance pattern. Myotonic dystrophy presents mild BMD symptoms together with extensive systemic involvement and produces prolonged muscular active states. The shoulder and hip areas become weak because of LGMD whereas FSHD targets the facial and upper arm muscles yet CMD develops its symptoms during birth as a severe issue that results in muscle weakness and joint abnormalities [44, 45]. The French neurologist Guillaume Benjamin Amand Duchenne introduced the first description of DMD in the 1860s and the scientific community remained uninformed about MD origins until the 1980s. MDA-funded scientists isolated a certain X-chromosome gene in 1986 which becomes dysfunctional through mutation and triggers DMD. Scientists discovered the protein that associates with this gene in 1987 and gave it the name dystrophin. Muscle cells weaken when they lack dystrophin as shown in Figure 2.1 protein making them vulnerable to damage[46]. Each muscle consists

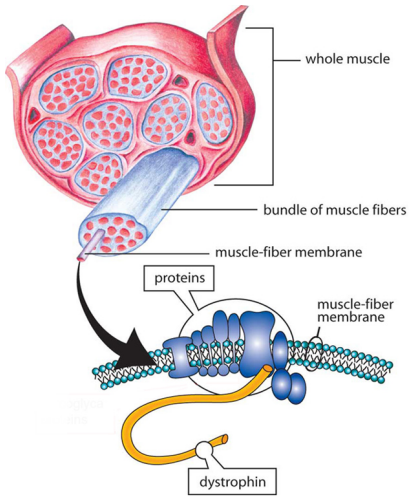


Figure 2.1: Duchenne muscular dystrophy(source:https://www.mda.org/disease/duchenne-muscular-dystrophy/causes-inheritance)

of various fibers which represent individual cells. Each muscle fiber functions properly through interdependent proteins that exist as a network within the membrane structure. People diagnosed with DMD have a complete absence of dystrophin protein.

DMD has an X-linked pattern of inheritance shown in Figure 2.2 because the gene containing mutations that cause DMD exists on X chromosomes [47]. A male inherits his X chromosome from his mother while getting his Y chromosome from his father which establishes his gender. Each female receives two X chromosomes which she gets from her mother and father [48]. DMD deases rate in male is grater then female because diseases transmitted through an X-linked recessive pattern primarily impact males and female carriers usually remain asymptomatic due to their two X chromosomes. So X-linked recessive patterns of inherited diseases primarily impact males because women have a safeguarding second X chromosome which blocks symptom expression. The prevalence of DMD remains sta-

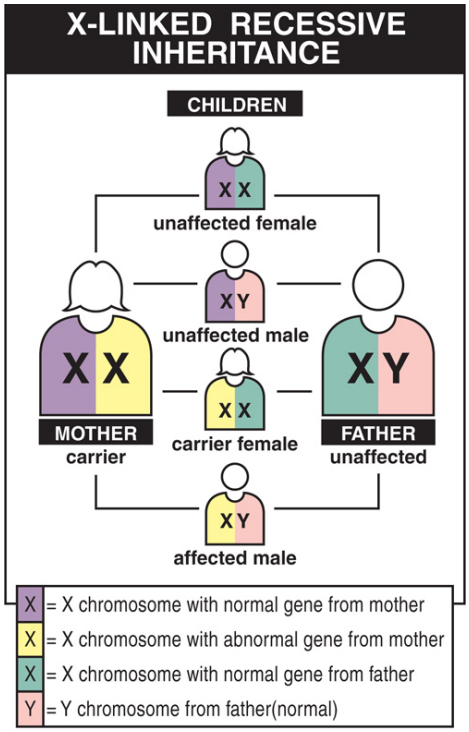


Figure 2.2: X-linked pattern of inheritance

ble between populations while the condition mostly impacts males because it follows an X-linked heredity pattern. The global statistics show that DMD affects 1 out of 3,500 to 5,000 boys who survive childbirth. DMD affects a percentage range of 0.02 percent to 0.03 percent in male patients in normal communities.

No treatment has been discovered to treat DMD but medical interventions focus on controlling symptoms and enhancing patient well-being. The medical community continues to search for treatment of muscular dystrophy because current approaches focus on symptom management and enhancing daily functions for patients. Physical therapy [49] combined with medications that include corticosteroids works together to keep joints active and protect muscles from deterioration. Assistive devices such as wheelchairs along with braces enable patients to maintain their ability to move and increase their self-dependence. Medical research has developed both gene therapy and exon- skipping which provide prospects for upcoming treatments to tackle the origin of the disease [50, 51]. The combined effort of performing early diagnosis and combining different medical experts leads to the best possible treatment results. In this work, twenty compounds that may have therapeutic potential against DMD ARE assessed. The chosen agents, Deflazacort, Citrulline, CoQ10, Epicatechin, Losartan, Halafuginone and Tadalafil, Ataluren, Givinostat, Clenbuterol, Prednisone, Oxandrolone, Rimeporide, Idebenone, Arginine, Ezutromid, Metformin, Casimersen, Creatine along with Agamree are a good mix of pharmacological approaches to dystrophin deficiency, inflammation, fibrosis, metabolism, and muscle functionality. Their respective chemical structures, as shown in the Figure 2.3, are obtained from authoritative databases such as DrugBank (https://go.drugbank.com/) and ChemSpider (https://www.chemspider.com/) to ascertain molecular identity. It should be noted that subset of these, namely deflazacort, ataluren, Prednisone, Tadalafil, Givinostat, Oxandrolone, and the exon-skipping drug casimersen, are currently FDA-approved for treating DMD. The others are under pre-clinical or clinical trials as investigational, repurposed, or nutritional supplements for this condition.

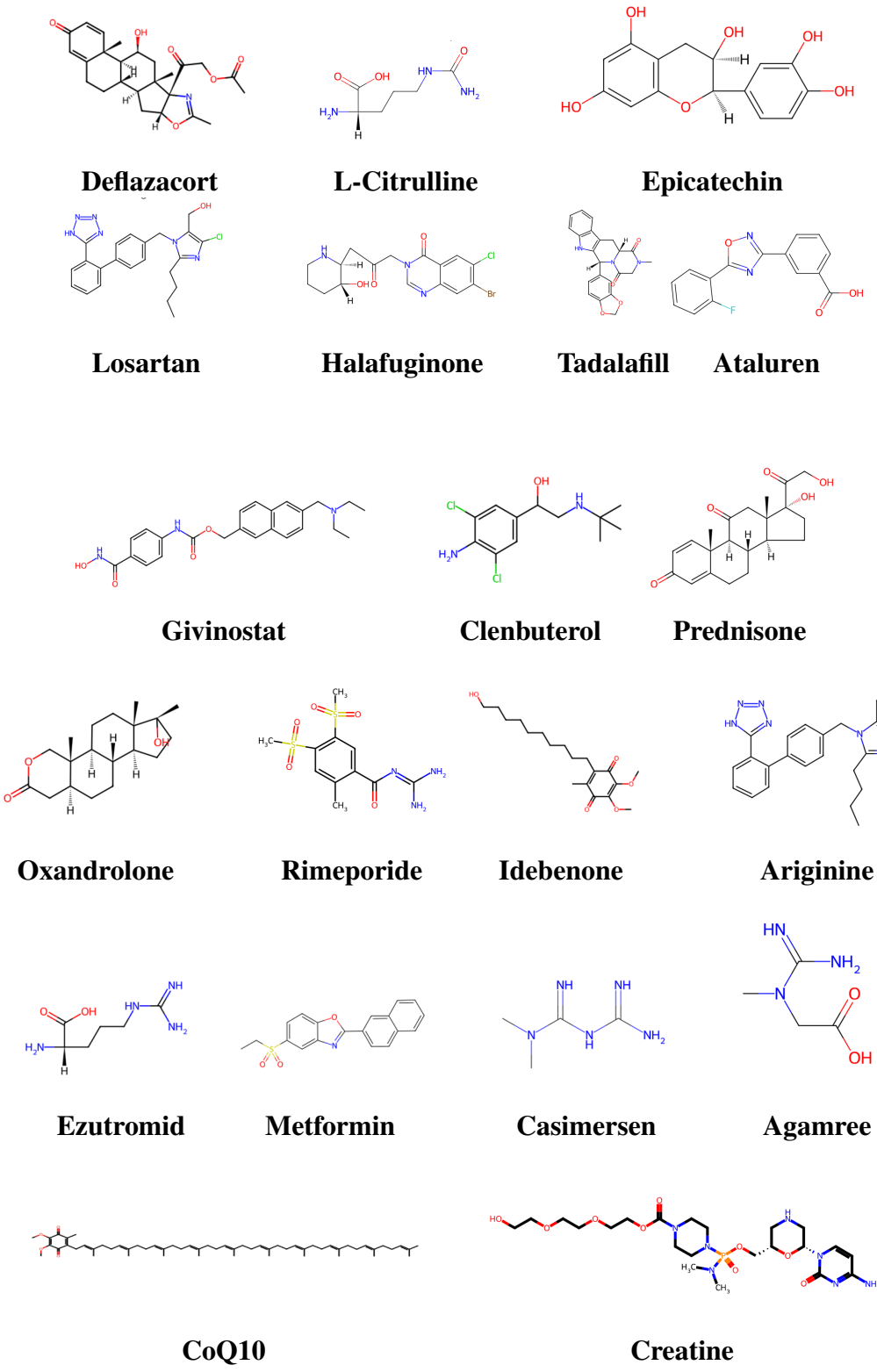
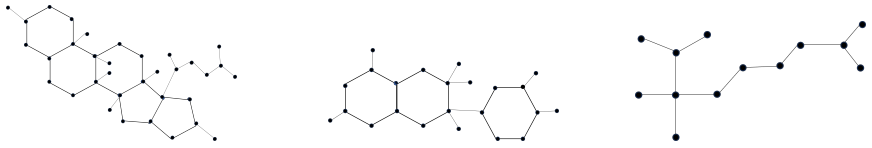


Figure 2.3: 2D chemical structures of twenty Anti-DMD drugs

Using graph models to represents molecular structures, chemical graph theory shows an interesting relationship between mathematics and chemistry. Under this method molecules

become visualized as graphs by assigning atoms to graph vertices and bonds to graph edges. The simple graph of Deflazacort, L-Citrulline and Epicatechin are shown in Figure 2.4 An illustration of chemical structures through graphs enables us to study molecular data using mathematical research methods. A graph's adjacency matrix displays square matrix properties with 1 in positions that mark bonded atoms while keeping all other positions at 0. The matrix provides information about molecular connectivity which forms the essential basis for additional computational analysis. Chemical graphs serve as the basis to generate topological indices which are numerical values that hold a correlation to molecular properties.



Deflazacort **L-Citrulline** **Epicatechin**

Figure 2.4: Simple graph of Deflazacort, L-Citrulline and Epicatechin

The important relationship exists between molecular graphs and topological indices because these indices serve as predictors of both physicochemical properties and biological activities such as Boiling point, enthalpy, flash point, index refraction, molar reactivity, polar surface, polarizabilty, surface tension and molar volume all influence a drugs pharmacokinetics and pharmacodynamicsfor QSAR and QSPR analysis fields. The fundamental structure of molecules determines their interactions with other molecules as well as their behavior because it is what drives their interactions and overall behavioral response. Research teams use molecular graph analysis and topological index evaluation to predict molecular structure properties and expected behavior without requiring extensive laboratory experiments so they consider topological indices essential for chemical discovery research in medicine. Beyond topological indices, the physical and chemical characteristics of anticervical cancer drugs are critical for understanding their activity in biological systems such as Density, Molar Refractivity (M.R), Polarizability, Molar Volume (MV), Complexity and Molecular Weight (M.W) all influence a drug's pharmacokinetics and pharmacodynamics [52, 53, 54]. Integrating these physicochemical attributes into the analysis by using machine learning algorithm, provides a comprehensive view of the medications' potential

efficacy and safety profiles. This multimodal investigation, which combines topological indices and physicochemical features, offers the framework for a thorough knowledge of the molecular characteristics of anti-cervical cancer therapies. Researchers are using machine learning techniques to better understand these drugs [55, 56]. Machine learning models such as Artificial Neural Networks and Random Forest are used to establish quantifiable correlations between computed molecular descriptors and analyze their biological activity[57].

2.1 Results and Discussions

Theorem 2.1.1. *Let G be a graph and G_1 denotes the Deflazacort, then the following axioms holds for the graph G_1:*

a) $M_1(G_1) = 116$,

b) $M_2(G_1) = 280$,

c) $H(G_1) = 2.5816$,

d) $F(G_1) = 606$,

e) $SS(G_1) = 18.1776$,

f) $ABC(G_1) = 7.0686$,

g) $RI(G_1) = 2.6367$,

h) $SC(G_1) = 3.9168$,

i) $GA(G_1) = 11.7508$,

j) $HZI(G_1) = 1166$,

k) $ReZG_2(G_1) = 27.8255$,

l) $ReZG_3(G_1) = 2928$

m $ND_1(G_1) = 56.8012$,

n) $ND_5(G_1) = 26.1238$

Proof. As given Let us consider graph of Deflazacort drug is represented by G_1. Edges are represented by $E_{p,q}$ where E_{p} is the degree of first connecting vertex and q is the degree of second vertex connected to that edge the frequencies $|E_{p,q}|$ show the number of edges. The expression $|E_{4,7}| = 2$ denotes two edge between the vertices of degree 4 and 7 , while the expression $|E_{6,7}| = 2$ denotes two edges between the vertices of degree 6 and 7. Similarly,

$|E_{4,6}| = 1, |E_{3,6}| = 2, |E_{4,4}| = 1, |E_{4,5}| = 2, |\ E_{3,4}\ | = 2$. Then,

1) By using first Zagreb index , we get

$$\begin{aligned} M_1(G) &= \sum_{pq \in E(G)} (d_p + d_q) \\ &= 2(4+7) + 2(6+7) + 1(4+6) + 2(3+6) \\ &\quad +1(4+4) + 2(4+5) + 2(3+4) \\ &= 116 \end{aligned}$$

2) By using second Zagreb index , we get

$$\begin{aligned} M_2(G) &= \sum_{pq \in E(G)} (d_p \times d_q) \\ &= 2(4 \times 7) + 2(6 \times 7) + 1(4 \times 6) + 2(3 \times 6) \\ &\quad +1(4 \times 4) + 2(4 \times 5) + 2(3 \times 4) \\ &= 280 \end{aligned}$$

3) By using harmonic index , we get

$$\begin{aligned} H(G) &= \sum_{pq \in E(G)} \frac{2}{d_p + d_q} \\ &= 2 \times \frac{2}{4+7} + 2 \times \frac{2}{6+7} + 1 \times \frac{2}{4+6} + 2 \times \frac{2}{3+6} \\ &\quad +1 \times \frac{2}{4+4} + 2 \times \frac{2}{4+5} + 2 \times \frac{2}{3+4} \\ &= 2.5816 \end{aligned}$$

4) By using forgotten index , we get

$$\begin{aligned} F(G) &= \sum_{pq\in E(G)} \left[(d_p)^2+(d_q)^2\right] \\ &= 2\times(4^2+7^2)+2\times(6^2+7^2)+1\times(4^2+6^2)+2\times(3^2+6^2) \\ &\quad +1\times(4^2+4^2)+2\times(4^2+5^2)+2\times(3^2+4^2) \\ &= 606 \end{aligned}$$

5) By using sombor-srednicki Index , we get

$$\begin{aligned} SS(G) &= \sum_{pq\in E(G)} \sqrt{\frac{d_p\times d_q}{d_p+d_q}} \\ &= 2\times\sqrt{\frac{4\times 7}{4+7}}+2\times\sqrt{\frac{6\times 7}{6+7}}+1\times\sqrt{\frac{4\times 6}{4+6}}+2\times\sqrt{\frac{3\times 6}{3+6}} \\ &\quad +1\times\sqrt{\frac{4\times 4}{4+4}}+2\times\sqrt{\frac{4\times 5}{4+5}}+2\cdot\sqrt{\frac{4\times 5}{3+4}} \\ &= 18.1776 \end{aligned}$$

6) By using Atom Bond Connectivity Index , we get

$$\begin{aligned} ABC(G) &= \sum_{pq\in E(G)} \sqrt{\frac{d_p+d_q-2}{d_p\times d_q}} \\ &= 2\sqrt{\frac{4+7-2}{4\times 7}}+2\sqrt{\frac{6+7-2}{6\times 7}}+1\sqrt{\frac{4+6-2}{4\times 6}}+2\sqrt{\frac{3+6-2}{3\times 6}} \\ &\quad +1\sqrt{\frac{4+4-2}{4\times 4}}+2\sqrt{\frac{4+5-2}{4\times 5}}+2\sqrt{\frac{3+4-2}{3\times 4}} \\ &= 7.0686 \end{aligned}$$

7) By using Randic Index Index , we get

$$\begin{aligned} RI(G) &= \sum_{qp\in E(G)} \frac{1}{\sqrt{d_p\times d_q}} \\ &= 2\times\frac{1}{(4\times 7)}+2\cdot\frac{1}{(6\times 7)}+1\times\frac{1}{(4\times 6)}+2\times\frac{1}{(3\times 6)} \\ &\quad +1\times\frac{1}{(4\times 4)}+2\times\frac{1}{(4\times 5)}+2\times\frac{1}{(3\times 4)} \\ &= 2.6367 \end{aligned}$$

8) By using sum connectivity Index , we get

$$\begin{aligned} SCI(G) &= \sum_{pq\in E(G)} \frac{1}{\sqrt{d_p+d_q}} \\ &= 2\times\frac{1}{(4+7)}+2\times\frac{1}{(6+7)}+1\times\frac{1}{(4+6)}+2\times\frac{1}{(3+6)} \\ &\quad +1\times\frac{1}{(4+4)}+2\times\frac{1}{(4+5)}+2\times\frac{1}{(3+4)} \\ &= 3.9168 \end{aligned}$$

9) By using Geometric Arithmetic Index , we get

$$\begin{aligned} GA(G) &+ \sum_{qp\in E(G)} 2\frac{\sqrt{d_p\times d_q}}{d_p+d_q} \\ &= 2\times 2\frac{\sqrt{4\times 7}}{4+7}+2\times 2\frac{\sqrt{6\times 7}}{6+7}+1\times 2\frac{\sqrt{4\times 6}}{4+6}+2\times 2\frac{\sqrt{3\times 6}}{2+6} \\ &\quad +1\times 2\frac{\sqrt{4\times 4}}{4+4}+2\times 2\frac{\sqrt{4\times 5}}{4+5}+2\times 2\frac{\sqrt{3\times 4}}{3+4} \\ &= 11.7508 \end{aligned}$$

10) By using Hyper Zagreb Index , we get

$$\begin{aligned} HZ(G) &= \sum_{pq\in E(G)} (d_p+d_q)^2 \\ &= 2(4+7)^2+2(6+7)^2+1(4+6)^2+2(3+6)^2 \\ &\quad +1(4+4)^2+2(4+5)^2+2(3+4)^2 \\ &= 1166 \end{aligned}$$

11) By using Redefined second Zagreb Index , we get

$$\begin{aligned} ReZ_2(G) &= \sum_{pq\in E(G)} \frac{d_p+d_q}{d_p\times d_q} \\ &= 2\times\frac{4+7}{4\times 7}+2\times\frac{6+7}{6\times 7}+1\times\frac{4+6}{4\times 6}+2\times\frac{3+6}{3\times 6} \\ &\quad +1\times\frac{4+4}{4\times 4}+2\times\frac{4+5}{4\times 5}+2\times\frac{4+5}{3\times 4} \\ &= 27.8255 \end{aligned}$$

12) By using Redefined third Zagreb Index , we get

$$\begin{aligned} ReZ_3(G) &= \sum_{pq\in E(G)} \frac{d_p \times d_q}{d_p + d_q} \\ &= 2\times\frac{4\times 7}{4+7} + 2\times\frac{6\times 7}{6+7} + 1\times\frac{4\times 6}{4+6} + 2\times\frac{3\times 6}{3+6} \\ &\quad +1\times\frac{4\times 4}{4+4} + 2\times\frac{4\times 5}{4+5} + 2\times\frac{4\times 5}{3+4} \\ &= 2928 \end{aligned}$$

13) By using first Neighborhood Degree Sum-Based Index , we get

$$\begin{aligned} ND_1(G) &= \sum_{pq\in E(G)} \sqrt{d_p \times d_q} \\ &= 2\times\frac{4\times 7}{4+7} + 2\times\frac{6\times 7}{6+7} + 1\times\frac{4\times 6}{4+6} + 2\times\frac{3\times 6}{3+6} + 1\times\frac{4\times 4}{4+4} \\ &\quad +2\times\frac{4\times 5}{4+5} + 2\times\frac{4\times 5}{3+4} \\ &= 56.8012 \end{aligned}$$

14) By using fifth Neighborhood Degree Sum-Based Index , we get

$$\begin{aligned} ND_5(G) &= \sum_{pq\in E(G)} \left(\frac{d_p}{d_q} + \frac{d_p}{d_q}\right) \\ &= 2\times\left(\frac{4}{7}+\frac{4}{7}\right) + 2\times\left(\frac{6}{7}+\frac{6}{7}\right) + 1\times\left(\frac{4}{6}+\frac{4}{6}\right) \\ &\quad +2\times\left(\frac{3}{6}+\frac{3}{6}\right) + 1\times\left(\frac{4}{4}+\frac{4}{4}\right) + 2\times\left(\frac{4}{5}+\frac{4}{5}\right) + 2\times\left(\frac{4}{5}+\frac{3}{4}\right) \\ &= 26.1238 \end{aligned}$$

∎

By using similar method, One can compute topological indices of Citrulline, CoQ10, Epicatechin, Losartan, Halafuginone and Tadalafil, Ataluren, Givinostat, Clenbuterol, Prednisone, Oxandrolone, Rimeporide, Idebenone, Arginine, Ezutromid, Metformin, Casimersen, Creatine, and Agamree. Furthermore, an alternative, more efficient approach for computing these indices involves implementing algorithm in Python, which can process the data in a single pass and significantly reduce the overall computational time. The resulting numerical indices, which are given in Table 3.1. And Table 2.2 represent their physicochemical properties, collected from ChemSpider and PubChem.

Table 2.1: Topological Indices of anti-DMD drugs

Alternative	M_1	M_2	H	F	SS	ABC	RI	SCI	GA	HZ	$RezG_2$	$RezG_3$	ND_1	ND_5
Deflazacort	602	2390	6.0089	5320	74.0854	20.446	6.211	10.798	38.7421	10100	141.75084	44096	291.89	91.4343
Citrulline	116	280	2.5816	606	18.1776	7.0686	2.6367	3.9168	11.7508	1166	27.8255	2928	56.8012	26.1238
CoQ10	682	1858	13.1276	3918	102.9212	37.1168	13.3542	20.5155	64.0076	7634	165.7256	22332	336.0903	138.6695
Epicatechin	342	1150	4.1779	2446	46.02705	13.62832	4.2638	7.3187	25.5376	4746	82.7368	16796	168.176	55.9717
Losartan	370	1079	6.3466	2246	54.1625	18.3201	6.4347	10.12667	32.6047	4404	90.505	13486	182.9732	69.3321
Halafuginone	402	1408	4.3827	2996	52.6506	14.8974	4.4876	7.9238	28.4089	5812	96.9561	21220	197.3403	63.2571
Tadalafill	540	2032	5.3991	4244	69.269	18.5144	5.4762	9.9101	36.5127	8308	131.7665	32720	266.6879	78.2093
Ataluren	294	917	4.0953	1916	41.2515	12.8416	4.1558	6.9706	23.6853	3750	71.7568	12136	145.2238	50.6857
Givinostat	352	954	6.4675	1970	53.0454	18.5934	6.5776	10.2641	32.5892	3878	86.2976	10790	174.2677	69.4952
Clenbuterol	186	507	3.18135	1068	27.5695	9.5189	3.2487	5.1849	16.6991	2082	45.1023	5820	91.5623	36.6309
Prednisone	470	1712	4.7648	4010	57.8187	16.7324	5.0368	8.6251	30.4465	7434	107.3725	29096	224.1867	79.6527
Oxandrolone	446	1699	4.0106	3864	54.3424	14.588	4.1764	7.5656	27.904	7262	103.7476	29352	214.9111	68.027
Rimeporide	468	1551	5.9478	3414	62.398	19.3133	6.16304	10.2554	34.8621	6516	110.558	23140	227.3895	81.87301
Idebenone	246	670	5.2129	1444	37.0714	14.028	5.3463	7.8033	23.4482	2784	58.9468	8274	120.3618	52.8821
Ariginine	116	280	2.5816	606	18.1776	7.0686	2.6367	3.9168	11.7508	1166	27.8255	2928	56.8012	26.1238
Ezutromid	322	967	4.6793	2008	45.9234	14.5483	4.7504	7.9117	6.6712	3942	78.8181	12106	159.2873	56.8214
Metformin	72	162	1.8421	348	11.7041	4.8316	1.8828	2.7028	7.8344	672	17.3058	1572	35.293	17.4
Casimersen	486	1464	7.8652	3112	69.6091	23.1863	7.9993	12.7172	41.3129	6040	117.6669	19386	239.0902	89.8595
Creatine	102	295	1.6601	612	14.8813	4.9674	1.688	2.7175	8.8761	1202	24.9282	3604	50.4215	19.0285
Agamree	458	1678	4.9728	3752	57.8805	16.6199	5.1548	8.785	30.9288	7108	107.5169	27934	5.1548	73.6847

Table 2.2: Actual Physicochemical properties of anti-DMD drugs

Alternatives	D	BP	EV	FP	IR	MR	PS	Pol	ST	MV
Deflazacort	1.4	595.4	101.8	313.9	115.1	42.1	102	45.6	53.1	135.9
L-Citrulline	1.3	386.7	69.8	187.7	1.531	30.3	118	16.7	91.9	135.9
CoQ10	1	869	126.3	324.6	1.526	272.6	53	108.1	38.6	888.5
Epicatechin	1.6	630.4	98	335	1.742	73.6	110	29.2	88.1	182.2
Losartan	1.4	682	105.1	366.3	1.681	118.2	93	46.9	53.3	312.5
Halafuginone	1.7	595.8	93.4	314.1	1.712	93.4	82	37	62.5	238.6
Tadalafill	1.5	679.1	99.7	364.4	1.758	105.3	75	41.7	81.6	256.2
Ataluren	1.4	503.7	81.4	258.4	1.604	70.8	76	28.1	56.8	206
Givinostat	1.3	122.3	229	190.5	1.652	122.3	91	48.5	56.7	334.6
Clenbuterol	1.3	404.9	69.2	198.7	1.577	73.5	58	29.5	46.6	221.6
Prednisone	1.3	573.7	98.7	314.8	1.604	94.1	92	37.3	58.6	273.6
Oxandrolone	1.1	444.4	81	179.1	1.526	85.1	47	33.7	39.9	277.1
Rimeporide	1.6	705.9	103.3	380.7	1.64	77.4	167	30.7	62.4	214.8
Idebenone	1.1	497.3	88	170.1	1.502	92.1	73	36.5	41.1	312.1
L-Ariginine	1.5	367.6	167.4	176.1	1.601	40.7	125	16.1	66.1	118.7
Ezutromid	1.3	552.7	80.2	288.1	1.66	94.9	69	37.6	53.6	257.3
Metformin	1.3	172.5	40.9	58.1	1.576	33.4	89	13.2	50.8	100.8
Casimersen	1.5	729.4	121.6	394.9	1.62	137	191	54.3	62.8	390.2
Creatine	1.4	271.6	56.1	118.1	1.552	30.3	90	12	57.6	94.2
Agamree	1.2	548.3	95.2	299.4	1.603	98.3	75	39	54.1	286

2.2 Supervised Machine Learning Techniques

Artificial intelligence uses supervised machine learning to build algorithms that derive learning from datasets containing predefined output values known as ”labels.” The main objective entails creating a model that connects input features with their correct outputs and learns universal functions to forecast new data points. The success of supervised learning algorithms depends almost entirely on the quality along with quantity of training data that receives labels because these elements become essential for both learning and predicting new data. Our goal focuses on drug discovery speedup by realizing efficient identification of promising candidates by using two Supervised Machine Learning models Artificial Neural Network and Random Forest. Both models were specifically selected because they support different capabilities to interpret intricate data relations between experimental measurements with computed molecular descriptors. These models permit accurate prediction of density alongside Boiling point, enthalpy, flash point, index refraction, molar reactivity, polar surface, polarizabilty, surface tension and molar volume thus eliminating the requirement for expensive laboratory characterization in order to optimize in-silico library screening processes for selecting promising compounds.

2.2.1 Artificial Neural Network (ANNs)

Artificial Neural Network (ANN) constitute a machine learning model structure which mimics the biological nature of human brain operations. QSPR has seen an increased use of Artificial Neural Networks because ANN efficiently extract complicated non-linear molecular structures-property relationships. The essential structure of ANN-based QSPR models contains layers of interlinked artificial neurons which operate on molecular descriptor data as input data. The fundamental procedure uses multiple layers which process the structural data through connected weights. A neuron undertakes a weighted summation of preceding layer outputs that represent abstract molecule features before applying an activation function to create its output. The output from the previous layer moves through additional layers until it reaches the output layer to forecast a particular property of molecules (such as boiling point or toxicity or solubility). During training the network adjusts weights and

biases repeatedly until it reaches the minimal difference between predicted and experimental properties for each training molecule.

we use a feedforward neural network through the `TensorFlow/Keras` library as the ANN model implementation. The implementation of the ANN involves following steps:

1) Dataset Construction & Preprocessing

Input Features (X):14 Topological Indices for 20 drugs having shape (20, 14) matrix.

Target Variables (Y): 10 Physicochemical Properties for the same 20 drugs having Shape (01, 20) matrix each time.

The total unique data instances are 20 drugs. The 300 figure represents the total number of scalar values (20 drugs $\times$ 14 features + 20 drugs $\times$ 01 targets = 280 + 20 = 300). We work with the (20, 14) and (10, 20) matrices as input data at a time.

2) Normalization

By using Scikit-learn's or StandardScaler. Apply Min-Max scaling or Standardization Z-score normalization to both the input features and target variables because topological indices have different scales and magnitudes. Normalization ensures no single feature dominates the training and leads to faster convergence.

3) Model Architecture & Hyperparameter Optimization

A rigorous approach like K-Fold Cross-Validation is essential to avoid overfitting and get a robust performance estimate.

4) Network Architecture:

Input Layer: 14 neurons (one for each topological index).

Hidden Layers: Start with 1-2 hidden layers to prevent overfitting. A good starting point is a single hidden layer with 8-16 neurons.

Output Layer: 10 neurons (one for each physicochemical property, using a linear activation function for regression).

5) Hyperparameter Optimization :

we use GridSearchCV from Scikit-learn wrapped around a Keras model via

tf.keras.wrappers.scikit_learn.KerasRegressor.

hidden layer sizes: [(8,), (16,), (8, 4)]

activation: ['relu', 'tanh']

optimizer: ['adam', 'sgd']

learning rate: [0.001, 0.01, 0.1]

batch size: [2, 4]

epochs: [500, 1000]

6) Activation Functions and Minimization

Hidden Layers: ReLU

Output Layer: Linear

Loss Function: Mean Squared Error (MSE), Mean Absolute Error (MAE).

Optimizer / Minimization Algorithm: Adam (adaptive learning rate).

7) Convergence and Training Duration

Training stops when the validation loss stops improving for a specified number of epochs this prevents overfitting.

Monitoring: Plot the training loss vs. validation loss over epochs.

Good Convergence: Both curves decrease and stabilize.

Overfitting: Training loss decreases but validation loss starts to increase.

Training Duration: The approximate time of training occupies about 720 seconds (12 minutes) on a typical CPU.

Figure 2.5 illustrate the distribution of various topological indices used in ANNs for density prediction. Each plot reveals distinct data profiles. These varied distributions highlight the importance of feature selection, as the differing scales and shapes can significantly influence model performance and interpretability in ANN-based topological analyses. Predicted Physicochemical properties of anti-DMD via ANNs are shown in Table 2.3 and their respective error estimate are in Table 2.4.

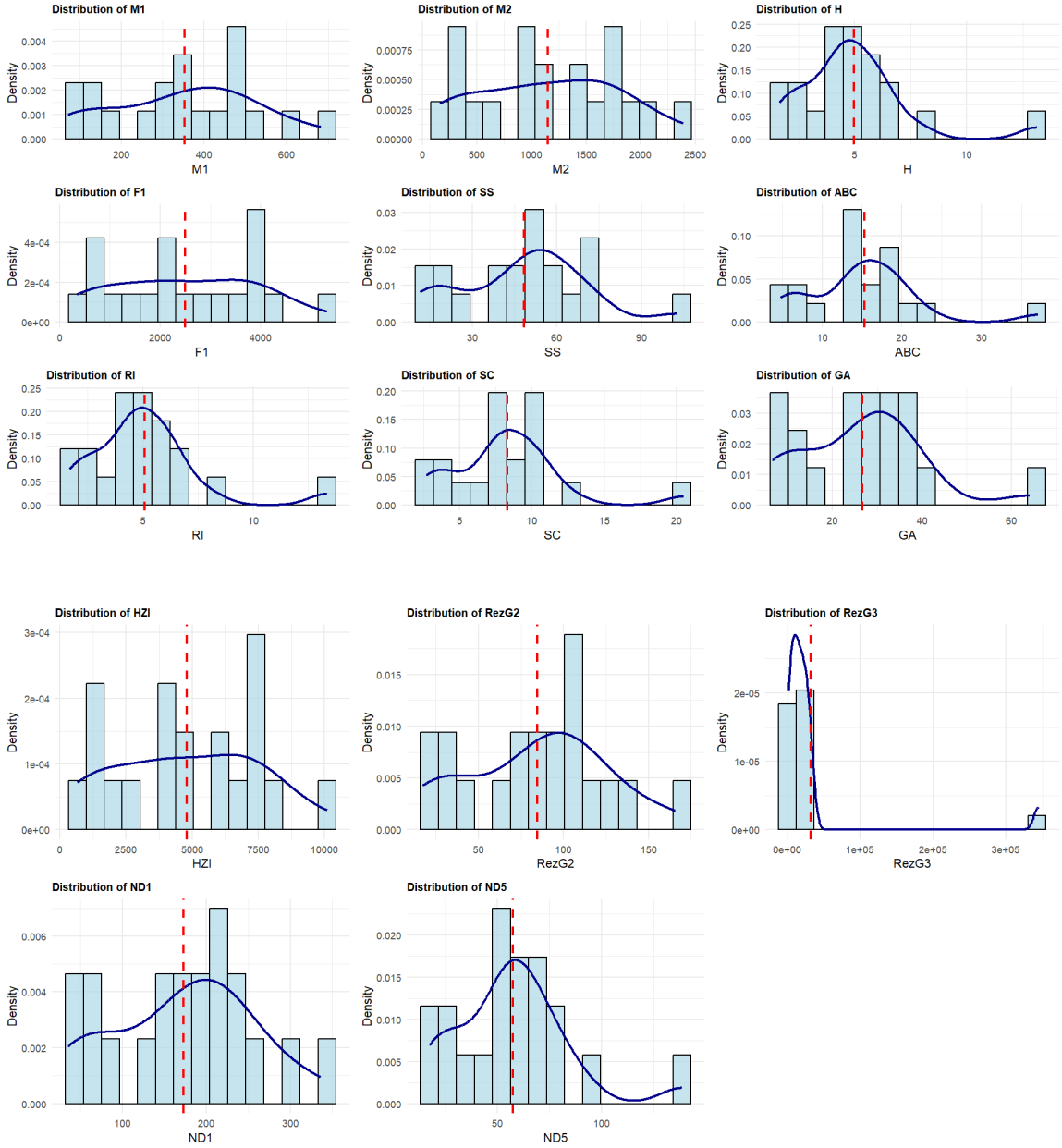


Figure 2.5: Graphical Representation Of Feature Importance For Density

Table 2.3: Predicted Physicochemical properties via Artificial Neural Network

Alternatives	D	BP	EV	FP	IR	MR	PSA	POL	ST	MV
Deflazacort	1.3889	601.36	102.62	313.429	112.92	54.156	91.131	44.384	54.554	131.529
L-Citrulline	1.3517	345.76	65.813	172.381	1.6353	39.006	111.86	18.582	74.774	139.533
CoQ10	1.0046	867.07	128.03	324.973	1.8578	272.417	57.488	107.48	41.668	883.216
Epicatechin	1.6019	623.86	90.783	334.091	2.0457	72.804	95.885	29.495	88.796	199.521
Losartan	1.4078	677.19	127.16	340.188	2.0058	118.239	103.92	43.779	56.979	305.054
Halafuginone	1.6916	620.09	92.477	314.681	1.4096	93.7109	88.129	35.337	62.387	241.530
Tadalafill	1.5148	675.90	99.087	368.969	10.893	93.504	90.035	39.223	81.478	256.210
Ataluren	1.4023	502.11	81.062	257.470	1.6782	72.067	78.898	29.412	56.840	227.609
Givinostat	1.2940	194.27	217.54	212.802	1.4759	122.396	81.818	47.096	53.911	332.210
Clenbuterol	1.3507	443.48	72.715	206.370	1.6582	71.876	78.897	25.007	48.499	175.187
Prednisone	1.2873	547.99	99.622	308.342	5.102	81.968	89.539	37.751	57.888	273.699
Oxandrolone	1.1113	443.81	84.841	182.414	1.5497	81.001	47.927	34.357	40.681	278.196
Rimeporide	1.6118	707.78	102.51	379.574	1.8322	91.975	178.529	37.461	63.014	213.192
Idebenone	1.1055	485.66	88.011	166.069	1.6154	91.703	69.7468	36.058	40.963	262.684
L-Ariginine	1.3517	345.76	65.813	172.381	1.6353	39.005	111.859	18.582	74.773	139.533
Ezutromid	1.3068	555.19	78.454	288.412	1.7561	95.098	67.0104	38.055	53.836	278.416
Metformin	1.3718	238.21	45.592	90.133	1.8903	31.606	100.553	11.864	56.534	104.406
Casimersen	1.4796	725.56	112.25	395.847	1.0644	141.93	174.156	56.507	62.644	408.541
Creatine	1.3970	247.73	54.269	93.6158	1.7616	31.744	83.7489	11.915	53.333	106.936
Agamree	1.2009	548.06	94.838	299.614	1.6142	98.482	75.6086	38.868	53.840	286.213

The models are assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) and R^2. The following formulas can be used to determine MAE, MSE, and RMSE:

$\text{MAE} = \frac{1}{K}\sum_{Q=1}^{K}|y_Q - \hat{y}_Q|$

$$\text{MSE} = \frac{1}{K}\sum_{Q=1}^{K}(y_Q - \hat{y}_Q)^2$$

$$\text{RMSE} = \sqrt{\frac{1}{K}\sum_{Q=1}^{K}(y_Q - \hat{y}_Q)^2}$$

ANNs exhibit strong predictive performance across a wide range of physicochemical properties. The large R, close to 0.94, indicates that the ANN captures the nonlinear relationships in the data. Interestingly, properties such as the index refraction and molar reactivity attain remarkable values. R scores of more than 0.985, which is excellent model accuracy. Yet some targets with larger measurement scales (such as boiling point and flash point) yield larger absolute errors (e.g., MAE and RMSE), which may be due to their numerical scale rather than inadequate model fit. The R values for surface tension and polar surface area are slightly lower, indicating that these properties can be improved through additional feature engineering or model tuning. On the whole, ANN is a valid and efficient method for forecasting the specified molecular characteristics, but better results can be achieved for particular endpoints through targeted optimisation.

Table 2.4: Artificial Neural Network Error Metrics.

	MAE	MSE	RMSE	R^2
Density	0.001694	0.022592	0.041152	0.944292
Boiling point	762.4838	17.87707	27.61311	0.977616
Enthalpy	41.89717	3.889702	6.472802	0.969029
Flash point	160.2881	7.788814	12.66049	0.981393
Index refraction	6.711408	3.916501	2.590639	0.989029
Molar reactivity	38.72575	5.390067	6.223002	0.985627
Polar surface	103.5572	8.461499	10.17630	0.914029
Polarizabilty	5.38033	1.674430	2.319554	0.986503
Surface tension	22.91229	2.627326	4.786679	0.884950
Molar volume	343.7940	11.99866	18.54168	0.957278

2.2.2 Random Forest Algorithm (RF)

The Random Forest machine learning method serves as a high-performing adaptable algorithm which benefits drug discovery operations and pharmaceutical development. During training Random Forest builds numerous decision trees which subsequently predict output through either class mode (classification) or mean prediction (regression) of their individual trees. RF operates according to several sequential steps during its principle application. Bootstrap aggregation (bagging) is deployed in Random Forest through a method that creates multiple training data subsets using random samples with replacement. A decision tree becomes part of the ensemble for every subset created. A decision tree examines only randomly selected features at each node for splitting rather than using all features available for splitting. The collection of these decision trees is the ”forest” and the final prediction is determined by calculating the average or weighted average of the predictions provided by each individual tree. The prediction of targeted variable y of a Random Forest Regression model can be expressed as:

$$y = \frac{1}{N} \sum_{i=1}^{N} f_i(x)$$

where N is the number of trees in the forest and $f_i(x)$ is the prediction of the $i-th$ decision tree as its output displayed.

The decision trees as shown in Figure 2.6 , Figure 2.7, Figure 2.8, and Figure 2.9 is for surface tension and the decision trees as shown in Figure 2.10 , Figure 2.11, Figure 2.12, and Figure 2.13 is for molar volume is highly interpretable, using molecular descriptors to create clear splitting rules. It shows strong performance on the training data, achieving squared errors as low as zero in some leaves, indicating excellent fit for specific subgroups.

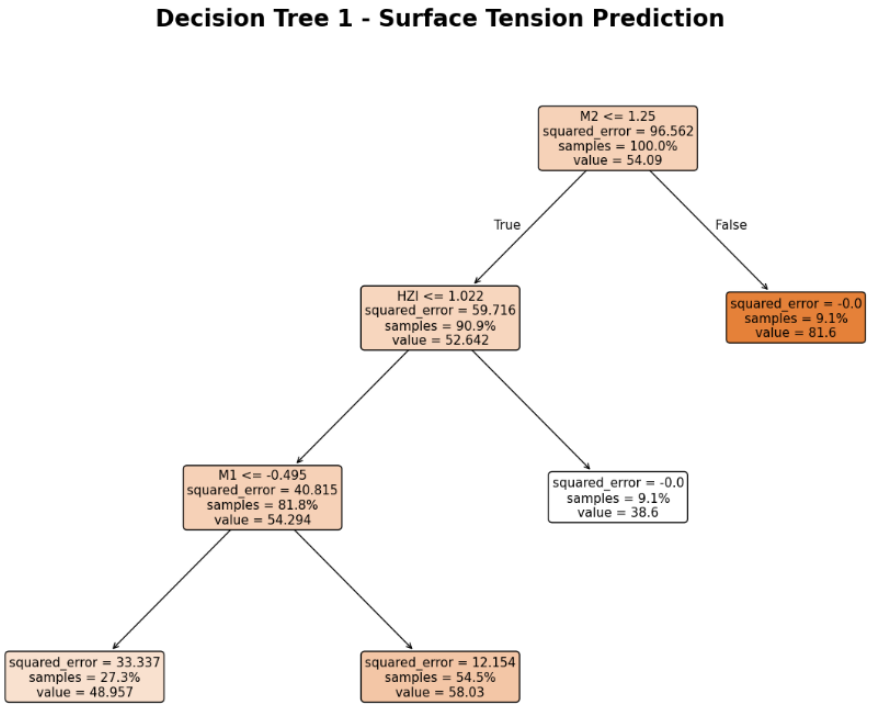


Figure 2.6: First Decision Trees For Surface Tension

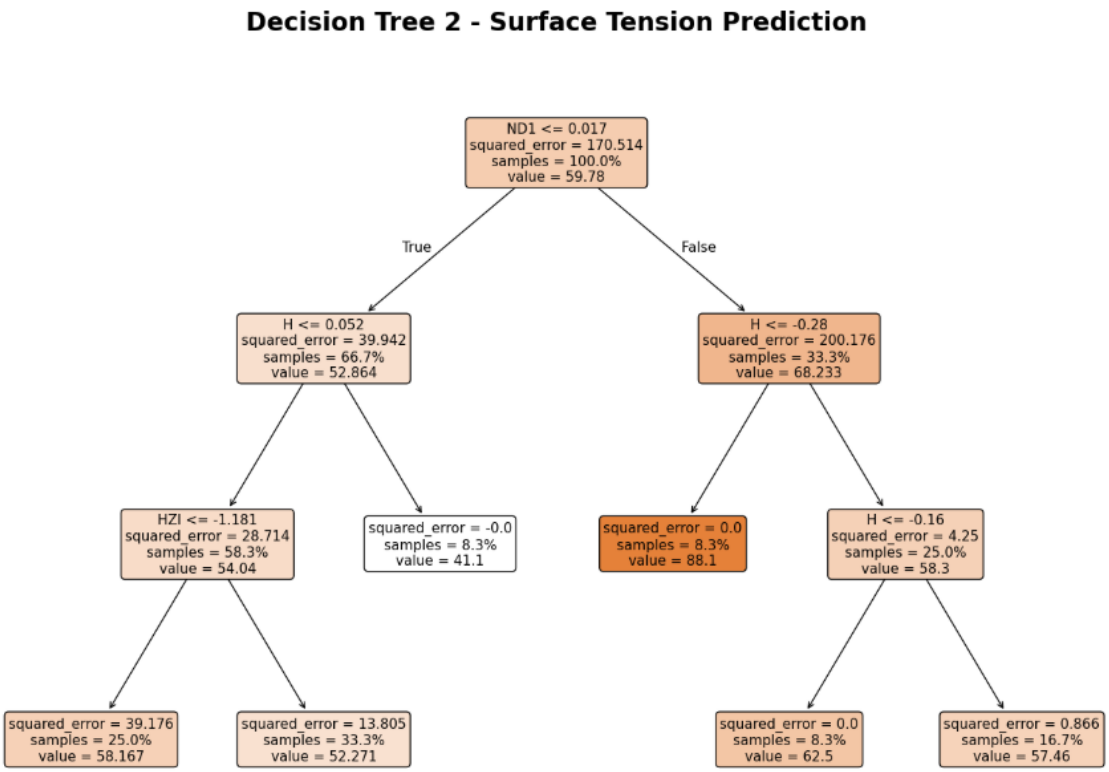


Figure 2.7: Second Decision Trees For Surface Tension

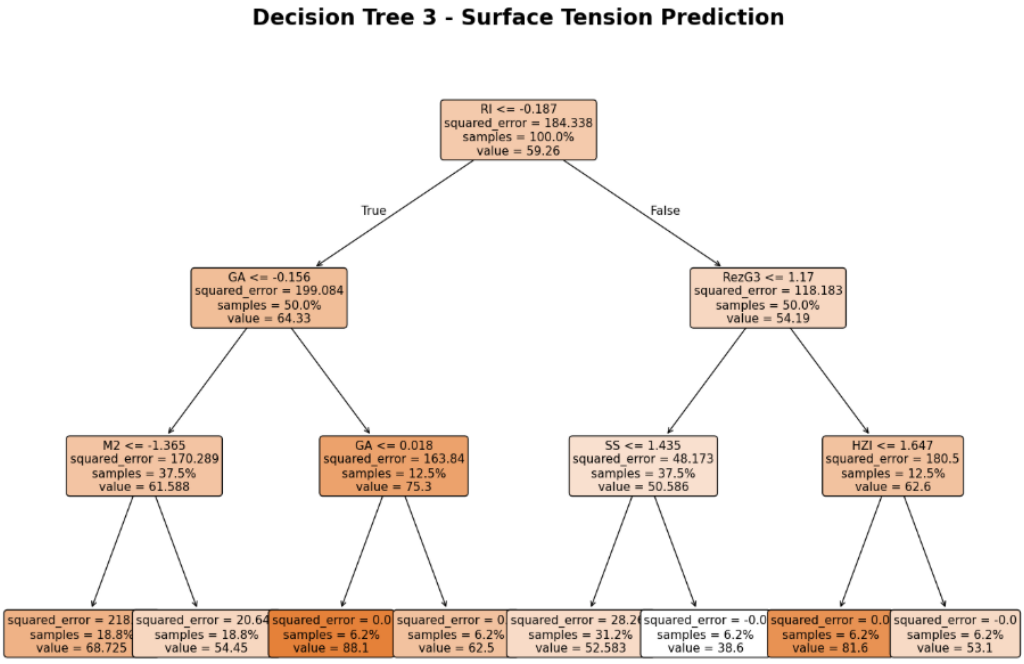


Figure 2.8: Third Decision Trees For Surface Tension

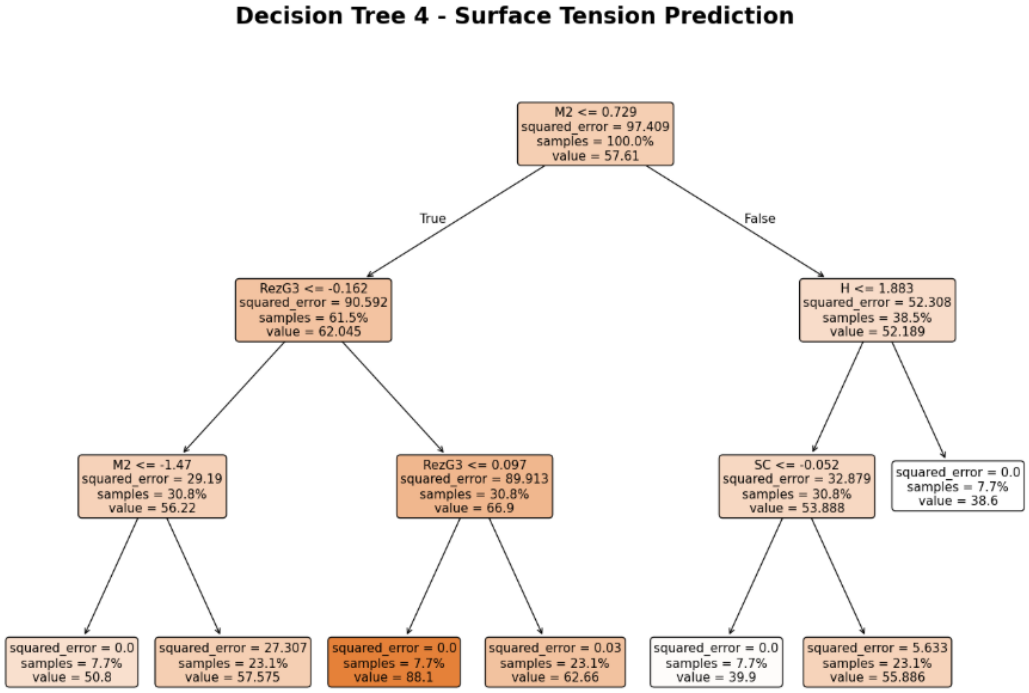


Figure 2.9: Fourth Decision Trees For Surface Tension

The structure of this decision tree is refined, and it can isolate specific groups in which surface tension can be predicted with considerable accuracy, as some branches have near-zero errors. Nevertheless, there is substantial uncertainty in the predictions, especially for intermediate values. The model is quite effective at exploiting a variety of molecular prop-

erties to identify apparent thresholds in extreme cases. Still, it fails to provide consistent explanations of variations across the entire chemistry space. This emphasises the subtle, situational character of the property.

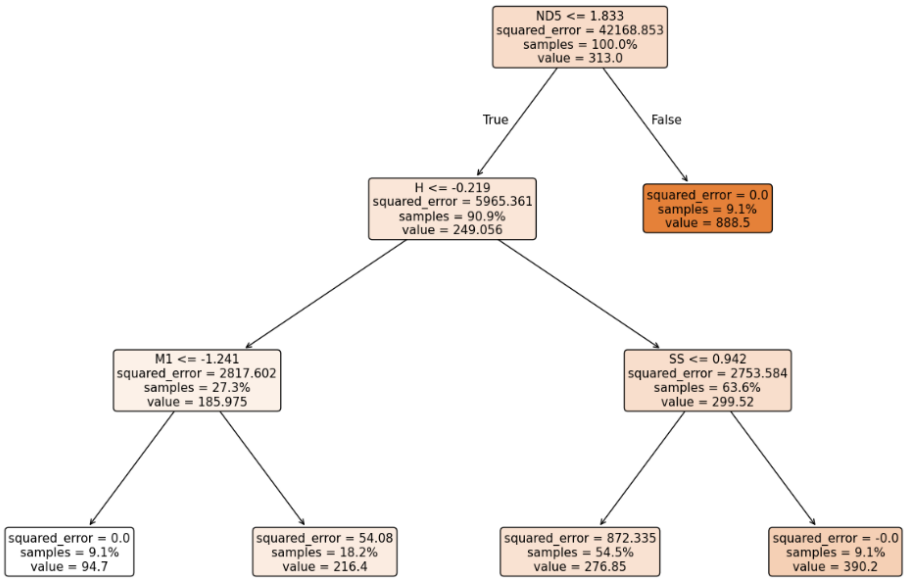


Figure 2.10: First Decision Trees For Molar Volume

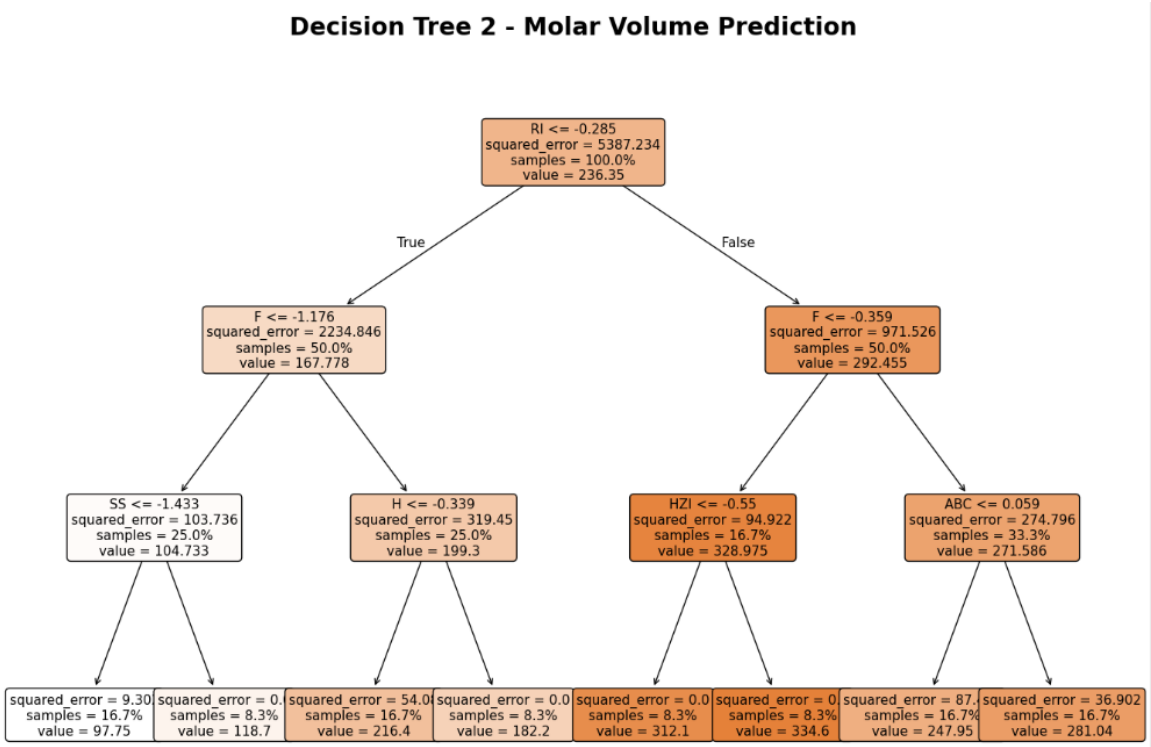


Figure 2.11: Second Decision Trees For Molar Volume

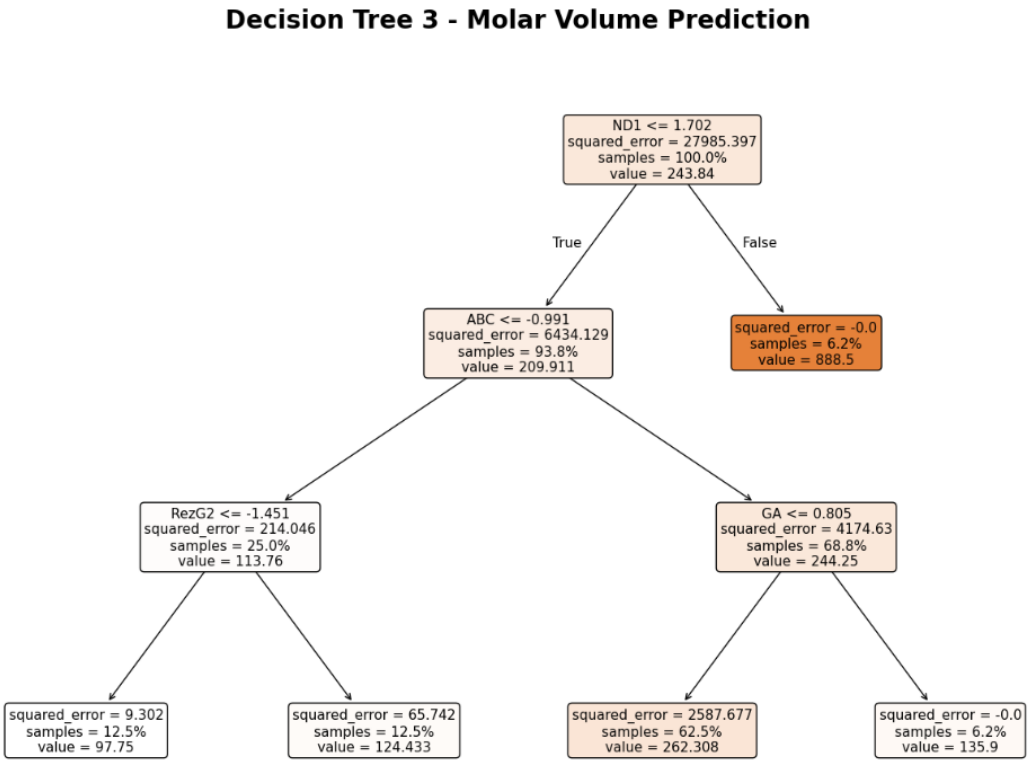


Figure 2.12: Third Decision Trees For Molar Volume

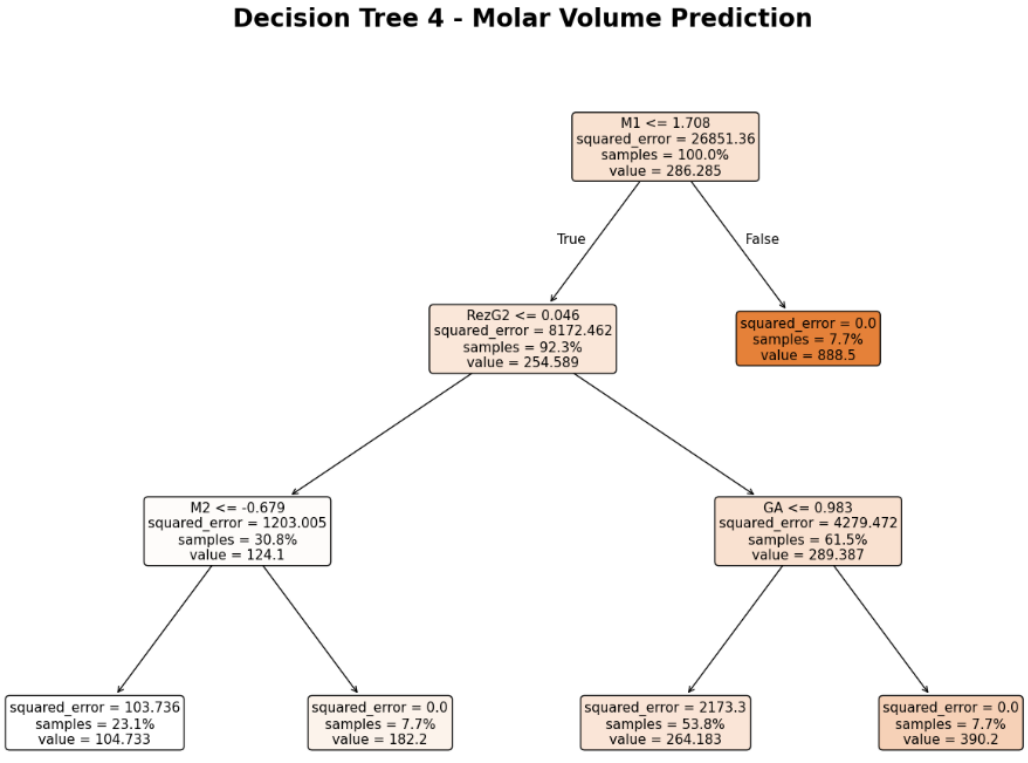


Figure 2.13: Fourth Decision Trees For Molar Volume

This fourth decision tree predicting molar volume shows a model that can isolate ex-

treme outliers with complete accuracy, as evidenced by the branch that predicts a large molar volume with no error. The feature splits, such as M1, RezG2, M2, and GA, enable effective splits that yield substantial reductions in prediction error for some subgroups, especially those with lower volume values. Yet, the enormous squared initial mistake and the fact that a significant portion of the data variability persists across many of the larger branches suggest that a substantial portion of the data variability remains hard to model. The tree can find clear, rule-based boundaries for specific chemical profiles. Still, it implies that the molar volume of a wider group of compounds results from more complicated, nonlinear interactions that are not entirely resolved by the structure of this particular tree.

The predicted properties and error metrics of anti-DMD drugs via RF are shown in Table 2.6 and Table 2.5 respectively. The Random Forest model shows both sufficient and heterogeneous predictions of the physicochemical properties. It has good predictive power for some of its targets, including boiling point, flash point, and polarizability, with each having an R2 value of over 0.91, but not for the rest, especially surface tension $R^2 = 0.723$ and index of refraction $R^2 = 0.807$. The large numerical values of boiling point and molar volume are likely the cause of high absolute errors. The model, though, is effective in many applications and may need additional improvements, such as parameter tuning or boosted ensemble techniques, to achieve consistency across all predicted variables.

Table 2.5: Random Forest Error Metrics.

	MAE	MSE	RMSE	R^2
Density	0.006175	0.061574	0.078580	0.796847
Boiling point	3032.432	35.81249	55.06752	0.910977
Enthalpy	162.9599	7.234199	12.79035	0.879070
Flash point	639.7005	18.74503	25.29230	0.925741
Index refraction	118.3514	3.419822	10.87894	0.806539
Molar reactivity	231.6836	8.332000	15.22116	0.9140139
Polar surface	179.2310	9.844500	13.38772	0.8512062
Polarizabilty	22.92613	2.417199	4.788124	0.9424875
Surface tension	55.09231	5.400800	7.422420	0.7233639
Molar volume	2693.381	28.38712	51.89780	O.9003348

Table 2.6: Predicted Physicochemical properties via Random Forest

Alternatives	D	BP	EV	FP	IR	MR	PSA	POL	ST	MV
Deflazacort	1.405	639.03	110.71	334.17	69.718	76.608	103.91	48.25	58.947	235.561
L-Citrulline	1.401	361.38	66.43	172.36	1.569	36.584	115.64	16.173	73.439	125.768
CoQ10	1.148	795.6	133.41	337.01	18.58	219.01	86.040	88.713	46.008	694.78
Epicatechin	1.486	576.73	91.46	298.32	1.686	77.984	96.65	30.772	73.603	204.335
Losartan	1.405	648.30	124.89	356.73	1.673	115.12	95.15	45.264	57.496	306.205
Halafuginone	1.585	596.55	93.76	315.21	1.703	91.108	87.84	36.089	62.531	243.219
Tadalafill	1.443	663.65	104.23	350.97	3.970	106.92	91.26	42.36	70.699	274.186
Ataluren	1.379	497.45	82.896	248.18	1.608	72.954	77.99	28.874	57.62	207.812
Givinostat	1.345	316.99	181.51	253.93	1.653	118.55	94.92	47.048	56.108	319.063
Clenbuterol	1.320	404.33	69.82	191.52	1.570	64.587	69.72	26.111	51.903	195.063
Prednisone	1.347	586.28	98.28	319.93	1.621	94.209	88.08	37.126	58.91	267.652
Oxandrolone	1.268	505.79	86.78	236.77	1.597	85.704	65.63	33.886	50.587	261.783
Rimeporide	1.559	685.69	108.9	369.95	1.635	90.032	139.13	37.11	61.405	242.7
Idebenone	1.208	464.08	87.22	204.11	1.551	89.062	73.9	35.099	47.598	275.768
L-Ariginine	1.402	361.38	66.43	172.37	1.569	36.584	115.63	16.173	54.603	125.768
Ezutromid	1.313	522.88	84.13	276.92	1.656	89.795	76.56	36.319	54.603	249.787
Metformin	1.3395	229.87	49.27	91.66	1.569	34.252	97.93	13.946	59.974	108.275
Casimersen	1.508	706.27	138.07	382.45	5.032	124.53	169.44	51.892	60.989	348.53
Creatine	1.381	280.83	57.08	128.62	1.558	32.549	93.345	13.443	60.066	106.354
Agamree	1.257	564.02	96.34	305.58	1.621	93.395	77.27	37.727	56.497	277.744

2.3 Computed vs. Actual: Physicochemical Properties Analysis

A critical part of this research conducted an extensive analysis to validate the accuracy and reliability of predicted physicochemical properties generated from ANN and RF models. The analysis became the essential step to prove that models properly converted the intricate linkages in topological indices from the anti-DMD drug dataset into significant molecular characteristic estimates. The core evaluation process consisted of exact comparison work between the ANN and RF model-generated physicochemical results against experimental or database-sourced actual values. Through an exact confrontation of predictions with actual measurements we were able to determine the level of accuracy between model outputs and drug candidate properties.

2.3.1 Assessing ANN and RF Predictions

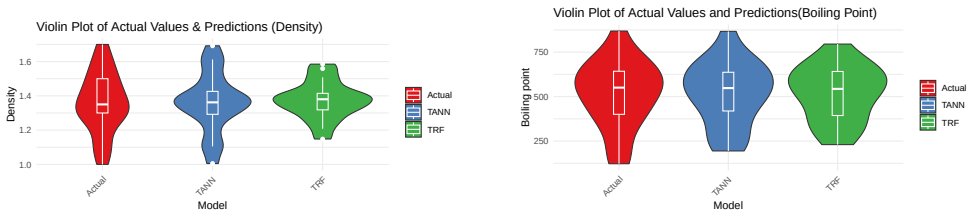
The analysis of predictive abilities for ANN and RF models used a computed-against-actual evaluation charted out in a detailed violin box plot. The plot showed a meticulous analysis between predictions from both models versus confirmed actual values for vital physicochemical properties of anti-DMD drug candidates. Research involved evaluation of Density, Boiling point, Enthalpy, Flash point, Index of refraction, Molar reactivity, Polar surface area, Polarizability, Surface tension, and Molar volume parameters. The violin box plot functioned as an advanced visualization method which displayed both standard interquartile ranges together with density distributions of calculated and actual data points for every parameter. Such representation uncovered the distribution spread and non-untypical prediction patterns between actual and estimated values of the models.

Note 1: Studying a violin plot requires examining its dual representational aspect. The central box plot uses the same method to display median and interquartile range just like a traditional box plot. The distribution density of the data appears as a violin-shaped figure that surrounds the box visualization. Areas that extend further from center represent denser levels of data points. The median appears within the central box for central tendency while box and whiskers show spread and outliers and the violin shape reveals distribution characteristics such as unimodality and skewness. A violin plot study can provide more nuanced findings about distribution differences between groups than traditional box plots do.

After training both the ANN and RF models on the dataset of Anti-DMD drugs, we pro-

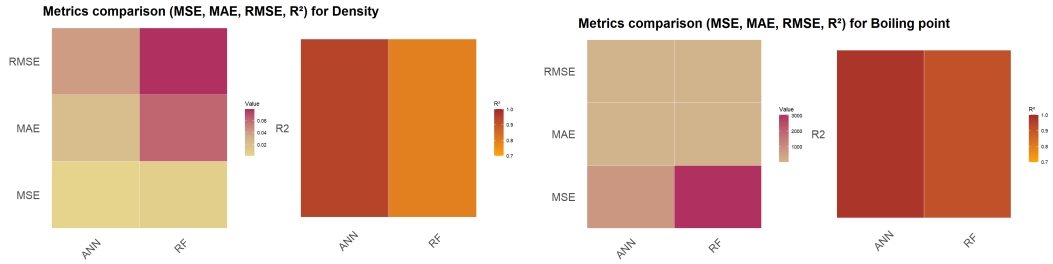
ceeded to evaluate their performance on a separate, unseen test set using the chosen error metrics: MSE, MAE, RMSE, and R^2. For each physicochemical property predicted, we calculated these four metrics for both the ANN and RF models. To provide a clear and intuitive overview of this performance comparison, we visualized the calculated error metrics using a heatmap.

Note 2: The heatmap features physicochemical properties as rows and the different error metrics MSE, MAE, RMSE, and R^2 for both models (ANN and RF) as columns. The color intensity within each cell directly reflects the magnitude of the error metric, allowing for a rapid and comprehensive visual assessment. Lower values for MSE, MAE, and RMSE, represented by lighter shades, indicate that the model's predictions are closer to the actual values on average, signifying better accuracy. Conversely, a higher R-Squared value, closer to 1, represented by darker shades, suggests that the model explains a larger proportion of the variance in the physicochemical property, indicating a better fit to the data. By examining this heatmap, we can quickly identify which model, the ANN or the RF, consistently exhibits superior performance across the various physicochemical properties of Anti-DMD drugs on previously unseen data.



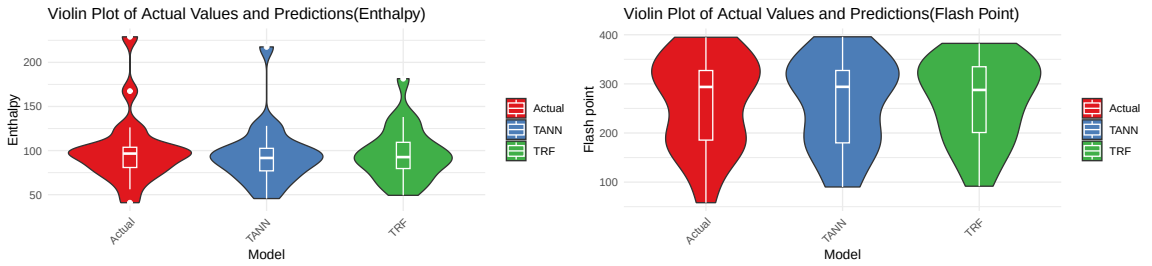
(a) Comparison For Density (b) Comparison For Boiling Point

Figure 2.14: Comparison of actual values as predictions(ANNs and RF) for density and boiling point of anti-DMD drugs



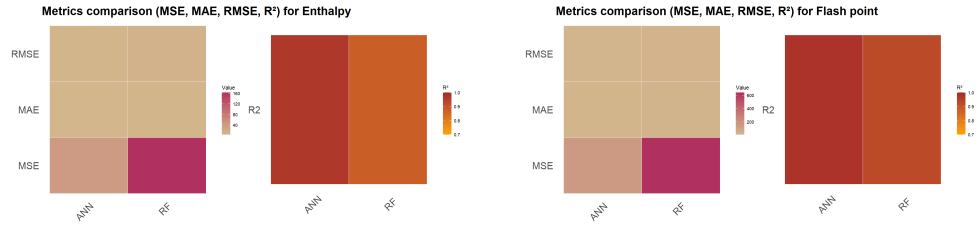
(a) Metrics Comparison For Density (b) Metrics Comparison For Boiling Point

Figure 2.15: Comprehensive performance evaluation of ANNs and RF for density and boiling point predictions



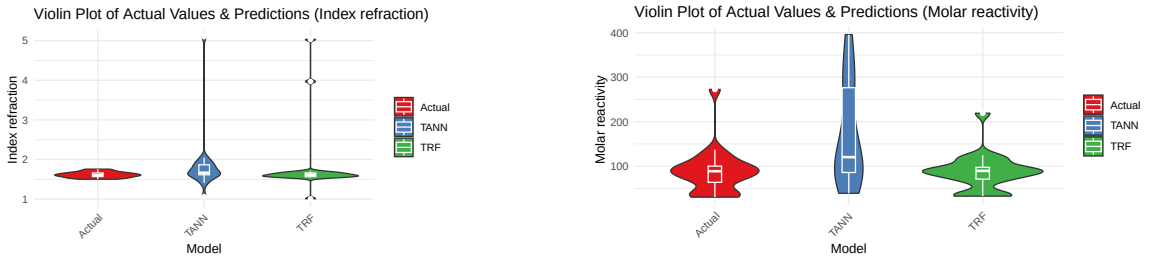
(a) Comparison For Enthalpy (b) Comparison For Flash Point

Figure 2.16: Comparison of actual values as predictions(ANNs and RF) for enthalphy and flash point of anti-DMD drugs



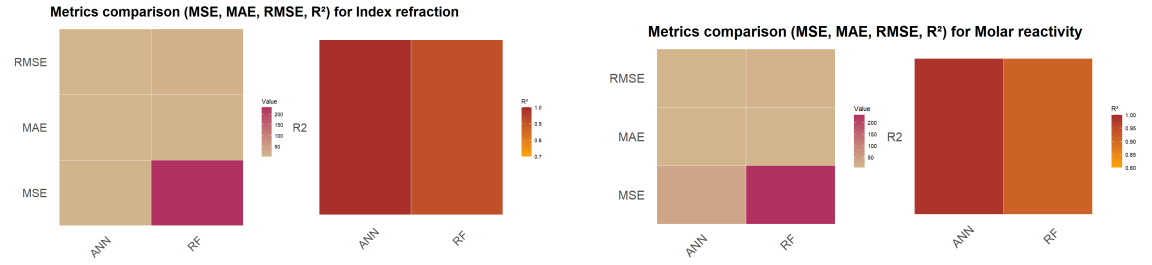
(a) Metrics Comparison For Enthalphy (b) Metrics Comparison For Flash Point

Figure 2.17: Comprehensive performance evaluation of ANNs and RF for Enthalphy and Flash Point predictions



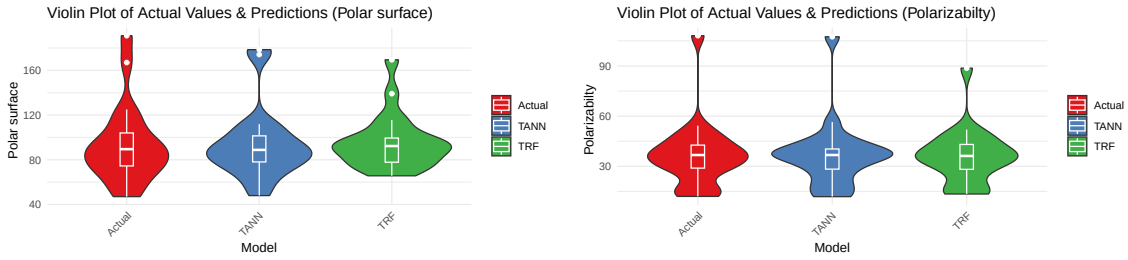
(a) Comparison for index refraction (b) Comparison for molar reactivity

Figure 2.18: Comparison of actual values as predictions(ANNs and RF) for index refraction and molar reactivity of anti-DMD drugs



(a) Metrics Comparison For Index Refraction (b) Metrics Comparison For Molar Reactivity

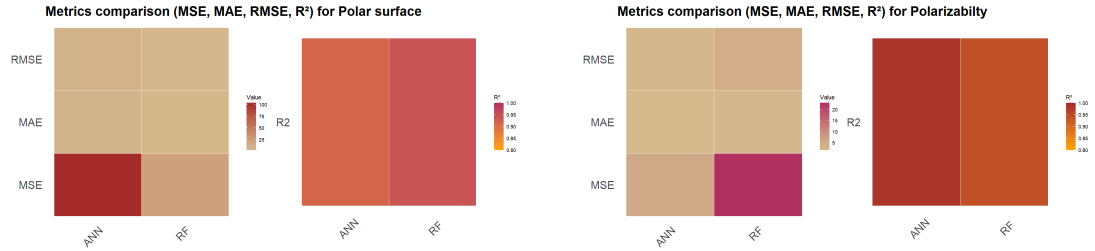
Figure 2.19: Comprehensive performance evaluation of ANNs and RF for Index Refraction and Molar Reactivity predictions



(a) Comparison for polar surface

(b) Comparison for polarizabilty

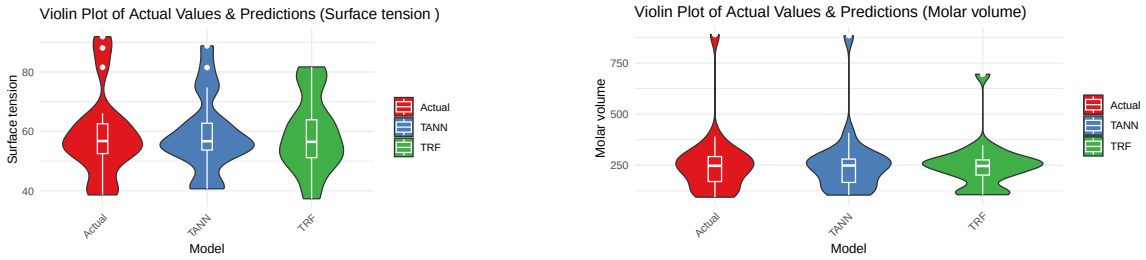
Figure 2.20: Comparison of actual values as predictions(ANNs and RF) for polar surface and polarizabilty of anti-DMD drugs



(a) Metrics Comparison For Polar Surface

(b) Metrics Comparison For Polarizabilty

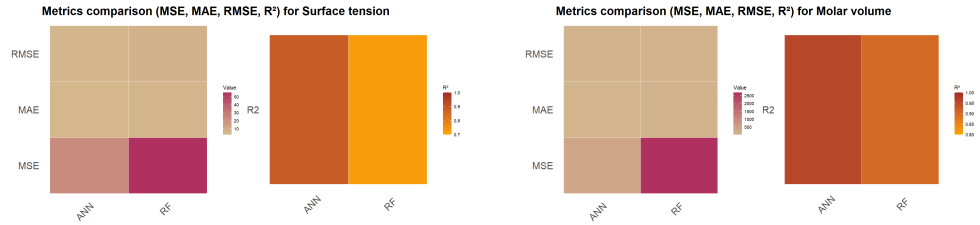
Figure 2.21: Comprehensive performance evaluation of ANNs and RF for Polar Surface and Polarizabilty predictions



(a) Comparison for surface tension

(b) Comparison for molar volume

Figure 2.22: Comparison of actual values as predictions(ANNs and RF) for suface tension and molar volume of anti-DMD drugs



(a) Metrics Comparison For Surface Tension

(b) Metrics Comparison For Molar Volume

Figure 2.23: Comprehensive performance evaluation of ANNs and RF for Surface Tension and Molar Volume predictions

Figure 2.14, Figure 2.16, Figure 2.18, Figure 2.20, and Figure 2.22 shows the graphical comparison of the actual and predicted values of 10 selected physicochemical properties by the RF and ANNs models, which indicates the extent of the accuracy and distribution overlap of the various predicted and actual values of each property on the test set. As Figure 2.14, Figure 2.16, Figure 2.18, Figure 2.20, and Figure 2.22 illustrates, both models are largely consistent with the actual values of most properties, although ANNs tend to cluster closer to the line of perfect prediction (y = x), especially in the density, boiling point, enthalpy, flash point, index of refraction and molar volume, but the Random Forest model is somewhat more scattered at the higher magnitude regions, the boiling point, flash point, and molar reactivity. A quantitative assessment of the model performance is presented in Figure 2.15, Figure 2.17, Figure 2.19, Figure 2.21, Figure 2.23 and the error measures in Tables 2.4 and 2.6. In terms of density, both models perform well, but ANN is slightly better than RF (R^2 = 0.944 vs. 0.797; RMSE= 0.041 vs. 0.078). The ANN model would considerably perform better than the Random Forest in boiling point (R^2 = 0.978 vs. 0.910), enthalpy (R^2 = 0.969 vs. 0.880), flash point (R^2 = 0.981 vs. 0.926), and molar volume (R^2 = 0.957 vs. 0.900), which is its ability to represent the non-linear associations, which are usually rather complex. ANNs, on the same note, is more predictive accurate in polar surface area, polarizability and surface tension. Random Forest competes well with molar reactivity and index of refraction (R^2 = 0.914 and R^2 = 0.807), although the ANN invariably reports lower values of MAE, MSE, and RMSE values in almost all properties, which justifies its overall performance in predictive power and generalization in this dataset.

Chapter 3
Multi-Criteria Drug Ranking Using the TOPSIS Method.

Technique for order of preference by similarity to ideal solution(TOPSIS) play a curical role in the MSDM due to its intuitive framework that simplifies complex decision-making processes, making it particularly valuable in scenarios where both qualitative and quantitative factors need to be considered, It offer a systematic and efficient approach to evaluate and rank a set of alternatives based on multiple conflicting criteria. Pharmaceutical sciences leverage the TOPSIS method as a vital decision-making tool to establish systematic drug candidate rankings through multiple competing criteria assessment. The process of drug discovery and development requires us to choose between various compounds which show different characteristics regarding potency together with safety profiles and pharmacokinetics and manufacturability metrics. The TOPSIS method solves these types of evaluation challenges through its quantitative approach which assesses drugs according to best and worst theoretical outcomes. The technique proves suitable for this purpose because it combines positive attributes with negative ones within a unified analytical framework to maintain therapeutic safety alongside beneficial effects. The TOPSIS approach enables the pharmaceutical field to utilize lead compound assessment in discovery and formulation development until it reaches commercial surveillance analysis phases. TOPSIS transforms drug profiles with multiple dimensions into structured rankings thus minimizing personal opinions in decision processes while keeping decisions transparent. [58, 59, 60, 61]

3.0.1 Create decision matrix

Create a matrix with the alternatives in the rows and the criteria in the columns. Let p be the number of alternatives and q be the number of criteria then r_{xy} represents the performance rating of alternative x (where x=1,2,3,...,p) with respect to criterion y (where y=1,2,3,...,q).

Table 3.1: Decision Matrix

Alternative	M_1	M_2	H	F	SS	ABC	RI	SCI	GA	HZ	$RezG_2$	$RezG_3$	ND_1	ND_5
Deflazacort	602	2390	6.0089	5320	74.0854	20.446	6.211	10.798	38.7421	10100	141.75084	44096	291.89	91.4343
Citrulline	116	280	2.5816	606	18.1776	7.0686	2.6367	3.9168	11.7508	1166	27.8255	2928	56.8012	26.1238
CoQ10	682	1858	13.1276	3918	102.9212	37.1168	13.3542	20.5155	64.0076	7634	165.7256	22332	336.0903	138.6695
Epicatechin	342	1150	4.1779	2446	46.02705	13.62832	4.2638	7.3187	25.5376	4746	82.7368	16796	168.176	55.9717
Losartan	370	1079	6.3466	2246	54.1625	18.3201	6.4347	10.12667	32.6047	4404	90.505	13486	182.9732	69.3321
Halafuginone	402	1408	4.3827	2996	52.6506	14.8974	4.4876	7.9238	28.4089	5812	96.9561	21220	197.3403	63.2571
Tadalafill	540	2032	5.3991	4244	69.269	18.5144	5.4762	9.9101	36.5127	8308	131.7665	32720	266.6879	78.2093
Ataluren	294	917	4.0953	1916	41.2515	12.8416	4.1558	6.9706	23.6853	3750	71.7568	12136	145.2238	50.6857
Givinostat	352	954	6.4675	1970	53.0454	18.5934	6.5776	10.2641	32.5892	3878	86.2976	10790	174.2677	69.4952
Clenbuterol	186	507	3.18135	1068	27.5695	9.5189	3.2487	5.1849	16.6991	2082	45.1023	5820	91.5623	36.6309
Prednisone	470	1712	4.7648	4010	57.8187	16.7324	5.0368	8.6251	30.4465	7434	107.3725	29096	224.1867	79.6527
Oxandrolone	446	1699	4.0106	3864	54.3424	14.588	4.1764	7.5656	27.904	7262	103.7476	29352	214.9111	68.027
Rimeporide	468	1551	5.9478	3414	62.398	19.3133	6.16304	10.2554	34.8621	6516	110.558	23140	227.3895	81.87301
Idebenone	246	670	5.2129	1444	37.0714	14.028	5.3463	7.8033	23.4482	2784	58.9468	8274	120.3618	52.8821
Ariginine	116	280	2.5816	606	18.1776	7.0686	2.6367	3.9168	11.7508	1166	27.8255	2928	56.8012	26.1238
Ezutromid	322	967	4.6793	2008	45.9234	14.5483	4.7504	7.9117	6.6712	3942	78.8181	12106	159.2873	56.8214
Metformin	72	162	1.8421	348	11.7041	4.8316	1.8828	2.7028	7.8344	672	17.3058	1572	35.293	17.4
Casimersen	486	1464	7.8652	3112	69.6091	23.1863	7.9993	12.7172	41.3129	6040	117.6669	19386	239.0902	89.8595
Creatine	102	295	1.6601	612	14.8813	4.9674	1.688	2.7175	8.8761	1202	24.9282	3604	50.4215	19.0285
Agamree	458	1678	4.9728	3752	57.8805	16.6199	5.1548	8.785	30.9288	7108	107.5169	27934	5.1548	73.6847

3.0.2 Normalize all criteria

Normalized the r_{xy} matrix as shown in Table 3.2 by using vector normalization:

$$r_{xy} = \frac{m_{xy}}{\sqrt{\sum_{z=1}^{p} (m_{zy})^2}} \tag{3.0.1}$$

for all x=1,2,3,...,p and y=1,2,3,...,q

Table 3.2: Normalized Decision matrix

alternative	M_1	M_2	H_1	F_1	SSG_1	ABC_1	RI_1	SC_1	GA_1	HZI_1	$RezG_2$	$RezG_3$	ND_1	ND_5
Deflazacort	0.3435	0.4063	0.2428	0.4153	0.3107	0.2704	0.2454	0.2634	0.2892	0.4110	0.3378	0.4819	0.3526	0.2988
L-Citrulline	0.0662	0.0476	0.1043	0.0473	0.0762	0.0935	0.1042	0.0956	0.0877	0.0475	0.0663	0.0320	0.0686	0.0854
CoQ10	0.3892	0.3158	0.5304	0.3058	0.4316	0.4908	0.5277	0.5005	0.4777	0.3107	0.3950	0.2441	0.4060	0.4532
Epicatechin	0.1952	0.1955	0.1688	0.1909	0.1930	0.1802	0.1685	0.1786	0.1906	0.1931	0.1972	0.1836	0.2032	0.1829
Losartan	0.2111	0.1834	0.2564	0.1753	0.2271	0.2423	0.2543	0.2471	0.2434	0.1792	0.2157	0.1474	0.2210	0.2266
Halafuginone	0.2294	0.2393	0.1771	0.2339	0.2208	0.1970	0.1773	0.1933	0.2120	0.2365	0.2311	0.2319	0.2384	0.2067
Tadalafill	0.3081	0.3454	0.2181	0.3313	0.2905	0.2448	0.2164	0.2418	0.2725	0.3381	0.3140	0.3576	0.3222	0.2556
Ataluren	0.1678	0.1559	0.1655	0.1496	0.1730	0.1698	0.1642	0.1701	0.1768	0.1526	0.1710	0.1326	0.1754	0.1657
Givinostat	0.2009	0.1622	0.2613	0.1538	0.2224	0.2459	0.2599	0.2504	0.2432	0.1578	0.2057	0.1179	0.2105	0.2271
Clenbuterol	0.1061	0.0862	0.1285	0.0834	0.1156	0.1259	0.1284	0.1265	0.1246	0.0847	0.1075	0.0636	0.1106	0.1197
Prednisone	0.2682	0.2910	0.1925	0.3130	0.2425	0.2213	0.1990	0.2104	0.2272	0.3025	0.2559	0.3180	0.2708	0.2603
Oxandrolone	0.2545	0.2888	0.1620	0.3016	0.2279	0.1929	0.1650	0.1846	0.2083	0.2955	0.2473	0.3208	0.2596	0.2223
Rimeporide	0.2671	0.2637	0.2403	0.2665	0.2617	0.2554	0.2435	0.2502	0.2602	0.2652	0.2635	0.2529	0.2747	0.2676
Idebenone	0.1404	0.1139	0.2106	0.1127	0.1555	0.1855	0.2112	0.1904	0.1750	0.1133	0.1405	0.0904	0.1454	0.1728
L-Ariginine	0.0662	0.0476	0.1043	0.0473	0.0762	0.0935	0.1042	0.0956	0.0877	0.0475	0.0663	0.0320	0.0686	0.0854
Ezutromid	0.1837	0.1644	0.1891	0.1567	0.1926	0.1924	0.1877	0.1930	0.0498	0.1604	0.1878	0.1323	0.1924	0.1857
Metformin	0.0411	0.0275	0.0744	0.0272	0.0491	0.0639	0.0744	0.0659	0.0585	0.0273	0.0412	0.0172	0.0426	0.0569
Casimersen	0.2773	0.2489	0.3178	0.2429	0.2919	0.3066	0.3161	0.3103	0.3084	0.2458	0.2804	0.2119	0.2888	0.2937
Creatine	0.0582	0.0501	0.0671	0.0478	0.0624	0.0657	0.0667	0.0663	0.0663	0.0489	0.0594	0.0394	0.0609	0.0622
Agamree	0.2614	0.2852	0.2009	0.2929	0.2427	0.2198	0.2037	0.2143	0.2308	0.2893	0.2562	0.3053	0.0062	0.2408

3.0.3 Assign weights

Define the weight for each criteria such that the sum of all weight should be 1 as shown in Table 3.3. As in the case with many other techniques, the TOPSIS method also has its own way of dealing with problems. Here too, the order of execution is very important and assigning criteria weights is one of the most critical factors that determines the overall success of the method. The Entropy method is a technique employed to assign weights to criteria in Multi-Criteria Decision Making (MCDM) procedures such as TOPSIS and it is quite objective. It employs Shannon's entropy to measure the level of information each criterion carries. Criteria with higher entropy values, meaning more variety across alternatives, are given lower weights and those with lower values, meaning less similarity across alternatives are weighted higher. With higher weights, these criteria are deemed to be more important and able to distinguish between the alternatives.

Table 3.3: weight allocation to topological indices

alternative	M_1	M_2	H_1	F_1	SSG_1	ABC_1	RI_1	SC_1	GA_1	HZI_1	$RezG_2$	$RezG_3$	ND_1	ND_5
Weight	0.0087341	0.073116	0.007411	0.003208	0.0054899	0.00764510	0.204769	0.06821	0.003416	0.000844	0.429403	0.064317	0.003381	0.120056

Calculate the weighted normalized decision matrixas shown in Table 3.4 with the help of Table 3.2 and Table 3.3.

$$W_{xy} = w_y \times m_{xy} \tag{3.0.2}$$

for all y=1,2,3...,q.

Table 3.4: Weighted Normalized Decision matrix

Alternative	M_1	M_2	H_1	F_1	SSG_1	ABC_1	RI_1	SC_1	GA_1	HZI_1	$RezG_2$	$RezG_3$	ND_1	ND_5
Deflazacort	0.00300	0.02971	0.00180	0.00133	0.00171	0.00207	0.05025	0.01797	0.00099	0.00035	0.14506	0.03099	0.00119	0.03588
L-Citrulline	0.00058	0.00348	0.00077	0.00015	0.00042	0.00071	0.02133	0.00652	0.00030	0.00004	0.02848	0.00206	0.00023	0.01025
CoQ10	0.00340	0.02309	0.00393	0.00098	0.00237	0.00375	0.10805	0.03414	0.00163	0.00026	0.16960	0.01570	0.00137	0.05441
Epicatechin	0.00170	0.01429	0.00125	0.00061	0.00106	0.00138	0.03450	0.01218	0.00065	0.00016	0.08467	0.01181	0.00069	0.02196
Losartan	0.00184	0.01341	0.00190	0.00056	0.00125	0.00185	0.05206	0.01685	0.00083	0.00015	0.09262	0.00948	0.00075	0.02720
Halafuginone	0.00200	0.01750	0.00131	0.00075	0.00121	0.00151	0.03631	0.01319	0.00072	0.00020	0.09922	0.01492	0.00081	0.02482
Tadalafill	0.00269	0.02526	0.00162	0.00106	0.00159	0.00187	0.04431	0.01649	0.00093	0.00029	0.13484	0.02300	0.00109	0.03069
Ataluren	0.00147	0.01140	0.00123	0.00048	0.00095	0.00130	0.03362	0.01160	0.00060	0.00013	0.07343	0.00853	0.00059	0.01989
Givinostat	0.00175	0.01186	0.00194	0.00049	0.00122	0.00188	0.05322	0.01708	0.00083	0.00013	0.08831	0.00758	0.00071	0.02727
Clenbuterol	0.00093	0.00630	0.00095	0.00027	0.00063	0.00096	0.02629	0.00863	0.00043	0.00007	0.04616	0.00409	0.00037	0.01437
Prednisone	0.00234	0.02128	0.00143	0.00100	0.00133	0.00169	0.04075	0.01435	0.00078	0.00026	0.10988	0.02045	0.00092	0.03125
Oxandrolone	0.00222	0.02112	0.00120	0.00097	0.00125	0.00147	0.03379	0.01259	0.00071	0.00025	0.10617	0.02063	0.00088	0.02669
Rimeporide	0.00233	0.01928	0.00178	0.00085	0.00144	0.00195	0.04986	0.01707	0.00089	0.00022	0.11314	0.01626	0.00093	0.03212
Idebenone	0.00123	0.00833	0.00156	0.00036	0.00085	0.00142	0.04326	0.01299	0.00060	0.00010	0.06032	0.00582	0.00049	0.02075
L-Ariginine	0.00058	0.00348	0.00077	0.00015	0.00042	0.00071	0.02133	0.00652	0.00030	0.00004	0.02848	0.00206	0.00023	0.01025
Ezutromid	0.00160	0.01202	0.00140	0.00050	0.00106	0.00147	0.03844	0.01317	0.00017	0.00014	0.08066	0.00851	0.00065	0.02229
Metformin	0.00036	0.00201	0.00055	0.00009	0.00027	0.00049	0.01523	0.00450	0.00020	0.00002	0.01771	0.00110	0.00014	0.00683
Casimersen	0.00242	0.01820	0.00236	0.00078	0.00160	0.00234	0.06472	0.02116	0.00105	0.00021	0.12042	0.01363	0.00098	0.03526
Creatine	0.00051	0.00367	0.00050	0.00015	0.00034	0.00050	0.01366	0.00452	0.00023	0.00004	0.02551	0.00253	0.00021	0.00747
Agamree	0.00228	0.02086	0.00149	0.00094	0.00133	0.00168	0.04171	0.01462	0.00079	0.00024	0.11003	0.01963	0.00002	0.02891

3.0.4 Find Ideal Best and Ideal Worst solutions

Determine the positive ideal solution and negative ideal solution Table 3.5 defined as:

$$I^+ = \{W_x^+, ..., W_y^+\} = (max(ormin)W_{xy} \| y \varepsilon Y) \tag{3.0.3}$$

$$I^- = \{W_x^-, ..., W_y^-\} = (max(ormin)W_{xy} \| y \varepsilon Y) \tag{3.0.4}$$

The ideal positive solution represents a hypothetical drug with maximizes all benefit criteria like bio availability and minimizes all cost criteria. Conversely, the ideal negative solution represents a drug that minimizes all benefit criteria and maximizes all cost criteria, essentially the worst-case scenario.

Table 3.5: Positive ideal solution and negative ideal solution

alternative	M_1	M_2	H_1	F_1	SSG_1	ABC_1	RI_1	SC_1	GA_1	HZI_1	$RezG_2$	$RezG_3$	ND_1	ND_5
Ideal Best(A^+)	0.0003588	0.0020135	0.0004971	0.0000871	0.0002694	0.0004884	0.0136575	0.0341411	0.0016320	0.0000231	0.1695967	0.0309946	0.0000211	0.0068272
Ideal Best (A^-)	0.0033991	0.0297050	0.0039309	0.0013321	0.0023694	0.0037523	0.1080480	0.0044979	0.0001701	0.0003469	0.0177100	0.0011049	0.0013727	0.0544095

3.0.5 Calculate Euclidean distance

Determine the n-dimensional euclidean distance. A longer euclidean distance represents a greater distance to undesirable attributes while a shorter euclidean distance represents the higher similarities to the desired character.

$$D_x^+ = \sqrt{\sum_{y=1}^{q} (W_{xy} - I_y^+)} \tag{3.0.5}$$

$$D_x^- = \sqrt{\sum_{y=1}^{q} (W_{xy} - I_y^-)} \tag{3.0.6}$$

3.0.6 Compute closeness score

Determine the scores by using the best and worst distances which enable us to know how near the answer you are. The defination is

$$S_x^{'} = \frac{D_x^-}{D_x^+ + D_x^-}, where 0 < S_x^{'} < 1, x = 1, 2, 3, ..., p. \tag{3.0.7}$$

it is clear that $S_x^{'} = 1 if D_x = D_x^+ and S_x^{'} = 0 if D_x = D_x^-$.

3.0.7 Rank

Ranks the drugs in decreasing order using $S_x^{'}$ as shown in Table 3.6 and and its graphical analysis is shown in Figure 3.1.

Table 3.6: Ranking anti-DMD drugs

Serial no.	Alternative	D_x^+	D_x^-	score	Rank
A_1	Deflazacort	0.0618871	0.1448649	0.7006701	1.0
A_2	Citrulline	0.1469366	0.1015474	0.4086678	18.5
A_3	CoQ10	0.1090449	0.1555866	0.5879369	7.0
A_4	Epicatechin	0.0942488	0.1066570	0.5308806	10.0
A_5	Losartan	0.0933587	0.0999350	0.5170113	11.0
A_6	Halafuginone	0.0820626	0.1144439	0.5823926	9.0
A_7	Tadalafill	0.0603744	0.1378397	0.6954080	2.0
A_8	Ataluren	0.1045081	0.1014747	0.4926368	14.0
A_9	Givinostat	0.0976425	0.0962275	0.4963508	13.0
A_{10}	Clenbuterol	0.1298062	0.0984837	0.4313976	15.0
A_{11}	Prednisone	0.0760291	0.1188114	0.6097880	3.0
A_{12}	Oxandrolone	0.0759464	0.1210133	0.6144063	5.0
A_{13}	Rimeporide	0.0771739	0.1161734	0.6008533	6.0
A_{14}	Idebenone	0.1188964	0.0878531	0.4249252	16.0
A_{15}	Ariginine	0.1469366	0.1015474	0.4086678	18.5
A_{16}	Ezutromid	0.0990726	0.1014912	0.5060294	12.0
A_{17}	Metformin	0.1576269	0.1080940	0.4067952	20.0
A_{18}	Casimersen	0.0811253	0.1156140	0.5876506	8.0
A_{19}	Creatine	0.1498444	0.1090496	0.4212132	17.0
A_{20}	Agamree	0.0754770	0.1187981	0.6114941	4.0

The corticosteroid medication Deflazacort functions as primary treatment for Duchenne Muscular Dystrophy (DMD) which causes severe progressive muscle weakness and degeneration. The treatment received FDA and EMA authorization as first-line therapy for

patients who reached the age of five years and above with substantial advantages in delaying disease progression. Deflazacort offers different steroid effects as it combats inflammatory processes and defends muscles against immune damage which allows patients to maintain their strength and functioning capabilities longer. The administration of deflazacort treatment helps patients maintain walking ability longer by 1.4 to 2.5 years than untreated patients while enhancing respiratory and cardiac measures. The preferred side-effect profile of deflazacort over prednisone includes minimal weight increases alongside lower incidence of behavioral side effects. Additional monitoring is necessary to mitigate risks including cataracts and both bone thinning and growth suppression and metabolic effects because of this medication. The medical administration of deflazacort should begin at 0.9 mg/kg/day but does adjustments may occur according to patient responses and treatment tolerability. The effectiveness of deflazacort as a preservation agent for mobility and muscular function positions the drug as a leading therapeutic approach in DMD treatment standards advocated by international guidelines. Scientific teams actively study how deflazacort stands against new treatments including gene therapy and exon-skipping drugs for long-term effectiveness. Research has assessed citrulline and metformin and arginine as therapeutic options for DMD but these strategies maintain low status because appropriate clinical benefits have not been verified. Human tests of citrulline as a nitric oxide precursor failed to establish any meaningful medical advantages in muscle function or disease progression for DMD patients despite theoretical benefits for muscle perfusion and fatigue reduction. Results from small and nonconclusive tests rule out the meaningful therapeutic value of metformin for DMD patients. Arginine stands as a nitric oxide booster but medical authorities do not recommend its use since clinical evidence shows it does not prevent muscle loss or improve patient mobility.Epicatechin and losartan and ataluren and halofuginone operate as average therapeutic choices for treating Duchenne Muscular Dystrophy (DMD) since clinical trials indicate promising but inconclusive results regarding their effectiveness. Research indicates that natural flavonoid Epicatechin shows promise for muscle strength enhancement together with fibrosis reduction yet additional extensive research is required for confirmation. The anti-fibrotic properties of losartan (an angiotensin receptor blocker) are being investigated while researchers observe some signs indicating it might minimize muscle deterioration when used with steroids although research data is inconsistent. Ataluren serves as a targeted therapeutic approach for nonsense mutation

read-through in certain DMD patients however benefits remain clinical and the general disease progression remains under evaluation. Data on the use of anti-fibrotic compound halofuginone for reducing muscle scarring and improving regeneration remains limited to preclinical research only. The drugs show promise as additional treatments for corticosteroids yet they cannot replace steroids because their effectiveness varies and they have restricted availability or particular usage requirements.

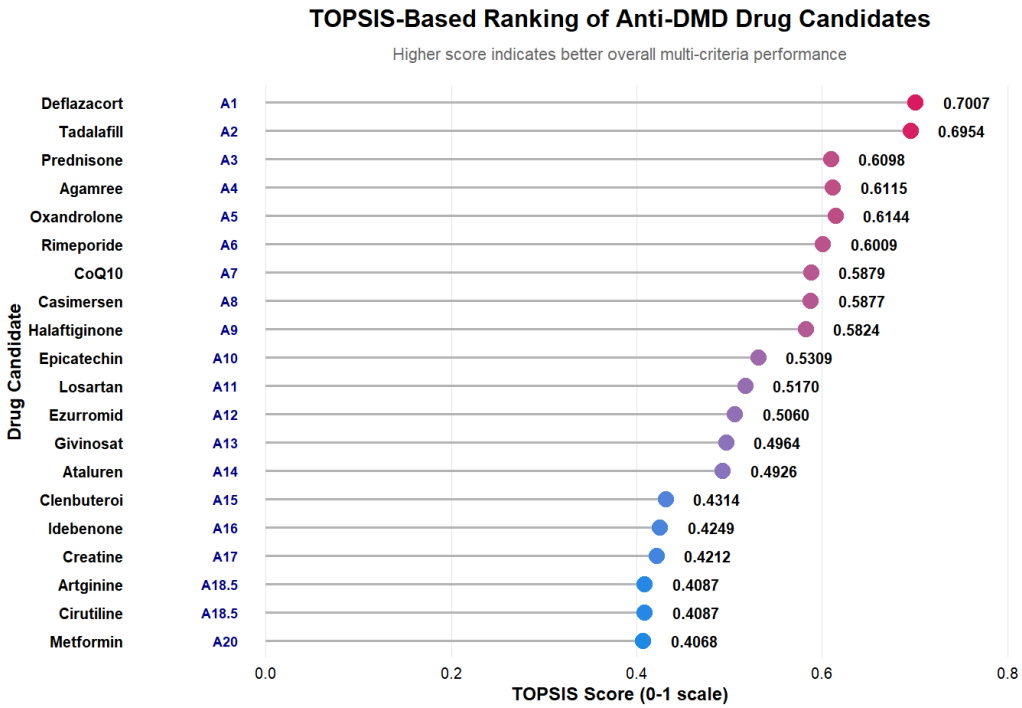


Figure 3.1: Graphical analysis of ranking anti-DMD drugs via TOPSIS

Chapter 4
Conclusion

This thesis demonstrated the strength of machine learning in predicting the physicochemical characteristics of (anti-DMD) drugs. The models, which were trained on a combination of experimental properties and molecular topological indices, were accurate predictors of nine critical properties, with ANNs providing a line of R^2 between 0.943 and 0.986 and RF confirming the fact that the index was correct and had an R^2 of at least 0.91 (and a high of 0.942). These findings validate the appropriateness of topological descriptors as cost-effective, robust features for accurately predicting properties. Detailed graphical displays, such as predicted-versus-actual plots, residual distributions, and feature correlation plots, confirmed the models reliability and provided interpretable insights into the relationships between molecular topology and key drug-like properties. The primary contribution of this work is the use of the TOPSIS multicriteria decision-making technique, which objectively ranked 20 anti-DMD drug candidates. Deflazacort was the highest-ranked substance, as suggested by Tadalafil, Prednisone, Agamore (vamorolone), and Oxandrolone, which provide a good direction for lead prioritisation when developing a therapy against DMD. However, it has constraints such as size and diversity of data sets, which might be associated with constrained generalizability to new scaffolds, and the omission of dynamic effects such as 3D conformations or solvent interactions. Future opportunities may include further increases in datasets, the use of more sophisticated architectures, such as Graph Neural Networks, to process molecular graphs, or complementary use with quantum simulations to make even more refined predictions. Computational knowledge would be further corroborated by clinical practicability through empirical validation of the TOPSIS-ranked leads using in vivo/in vitro studies. Overall, this paper has developed a predictive model that is both predictive, practical, and understandable, combining machine learning, cheminformatics, and decision-making tools to enhance the potential of drug discovery in the context of anti-DMD. The suggested methodology has significant potential for further application to other rare diseases and has a considerable impact on more sensitive, evidence-based pharmaceutical research.

References

[1] Yousaf, S., Bhatti, A. A., and Aslam, A. (2022). Study of carbon nanbotubes and boron nanotubes using degree based topological indices. Polycyclic Aromatic Compounds, 42(10), 7724-7737.

[2] Ahmed, W., Ali, K., Zaman, S., and Raza, A. (2024). Molecular insights into anti-Alzheimers drugs through predictive modeling using linear regression and QSPR analysis. Modern Physics Letters B, 2450260.

[3] Zaman, S., Majeed, M. U., Ahmed, W., and Saleem, M. T. (2024). On Neighborhood Eccentricity-Based Topological Indices with QSPR Analysis of PAHs Drugs. Measurement: Interdisciplinary Research and Perspectives, 1-14.

[4] Tang, J. H., Yousaf, S., Ashraf, M. J., Tawfiq, F. M., & Aslam, A. (2024). Expected values of sum-based topological indices of random cyclodecane chains. Physica Scripta, 99(3), 035020.

[5] Ahmed, W., Zaman, S., & Ashraf, S. (2024). Topological characterisation of three classes of complex networks and their graphical representation and analysis. Journal of Micromechanics and Molecular Physics, 9(02n03), 77-89.

[6] Zhou, H., Mahboob, A., Rasheed, M. W., Ovais, A., Siddiqui, M. K., and Cheema, I. Z. (2023). On QSPR Analysis of Molecular Descriptor and Thermodynamic Features of Narcotic Drugs. Polycyclic Aromatic Compounds, 1-21.

[7] Ahmed, W., Zaman, S., & Khalid, F. (2024). Analyzing polycyclic aromatic hydrocarbons using topological indices and QSPR analysis to reveal molecular complexity. Journal of Micromechanics and Molecular Physics, 2450006.

[8] Balasubramaniyan, D., and Chidambaram, N. (2023). On some neighbourhood degree-based topological indices with QSPR analysis of asthma drugs. The European Physical Journal Plus, 138(9), 823.

[9] Meharban, S., Ullah, A., Zaman, S., Hamraz, A., and Razaq, A. (2024). Molecular structural modeling and physical characteristics of anti-breast cancer drugs via some novel topological descriptors and regression models. Current Research in Structural Biology, 100134.

[10] Hayat, S., Khan, M. A., Khan, A., Jamil, H., and Malik, M. Y. H. (2023). Extremal hyper-Zagreb index of trees of given segments with applications to regression modeling in QSPR studies. Alexandria Engineering Journal, 80, 259-268.

[11] Hayat, S., Khan, A., Ali, K., and Liu, J. B. (2024). Structure-property modeling for thermodynamic properties of benzenoid hydrocarbons by temperature-based topological indices. Ain Shams Engineering Journal, 15(3), 102586.

[12] Hayat, S. (2023). Distance-based graphical indices for predicting thermodynamic properties of benzenoid hydrocarbons with applications. Computational Materials Science, 230, 112492.

[13] Wiener, Harry. (1947). Structural determination of paraffin boiling points. Journal of the American chemical society 69, no. 1 : 17-20.

[14] Delorme, Charles, Odile Favaron, and Dieter Rautenbach. (2002). On the Randic index. Discrete Mathematics 257, no. 1 : 29-38.

[15] Gutman, I., and Polansky, O. E. (2012). Mathematical concepts in organic chemistry. Springer Science and Business Media.

[16] Martınez-Martınez, C. T., Mendez-Bermudez, J. A., Rodrıguez, J. M., and Sigarreta, J. M. (2021). Computational and analytical studies of the harmonic index in Erdös–Rényi models. MATCH Commun. Math. Comput. Chem, 85, 395-426.

[17] Furtula, B., and Gutman, I. (2015). A forgotten topological index. Journal of mathematical chemistry, 53(4), 1184-1190.

[18] Zhao, W., Shanmukha, M. C., Usha, A., Farahani, M. R., and Shilpa, K. C. (2021). Computing SS index of certain dendrimers. Journal of Mathematics, 2021, 1-14.

[19] Estrada, E., Torres, L., Rodriguez, L., and Gutman, I. (1998). An atom-bond connectivity index: modeling the enthalpy of formation of alkanes. Indian J. Chem, 37A, 849–855.

[20] Gao, W., Wang, W. F., Jamil, M. K., Farooq, R., and Farahani, M. R. (2016). Generalized atom-bond connectivity analysis of several chemical molecular graphs. Bulgarian Chemical Communications, 48(3), 543-549.

[21] Randic, M. (1975). Characterization of molecular branching. Journal of the American Chemical Society, 97(23), 6609-6615.

[22] Ali, A., Furtula, B., Redžepović, I., Gutman, I. (2022). Atom-bond sum-connectivity index. Journal of Mathematical Chemistry, 60(10), 2081-2093.

[23] Vujošević, S., Popivoda, G., Vukićević, Ž. K., Furtula, B., and Škrekovski, R. (2021). Arithmetic–geometric index and its relations with geometric–arithmetic index. Applied Mathematics and Computation, 391, 125706.

[24] Gutman, I. (2017). On hyper–Zagreb index and coindex. Bulletin (Académie serbe des sciences et des arts. Classe des sciences mathématiques et naturelles. Sciences mathématiques), (42), 1-8.

[25] Ranjini, P. S., Lokesha, V., and Usha, A. (2013). Relation between phenylene and hexagonal squeeze using harmonic index. Int. J. Graph Theory, 1(4), 116-121.

[26] Mondal, S., Dey, A., De, N., and Pal, A. (2021). QSPR analysis of some novel neighbourhood degree-based topological descriptors. Complex and intelligent systems, 7, 977-996.

[27] Furtula, B., Gutman, I., & Dehmer, M. (2013). On structure-sensitivity of degree-based topological indices. Applied Mathematics and Computation, 219(17), 8973-8978.

[28] Ahmed, W., Ashraf, T., AlMutairi, D., Zaman, S., Ahmed, S., & Ehsan, H. (2025). A Deep Dive into Machine Learning: The Roles of Neural Networks and Random Forests in QSPR Analysis. BioNanoScience, 15(1), 89.

[29] Hosoya, H. (2002). The topological index Z before and after 1971. Int. Elect. J. Mol. Des, 1(9), 428-442.

[30] Zhou, B. (2004). Zagreb indices. MATCH Commun. Math. Comput. Chem, 52(1), 113-118.

[31] Gutman, I., Furtula, B., & Katanic, V. (2018). Randic index and information. AKCE International Journal of Graphs and Combinatorics, 15(3), 307-312.

[32] Mondal, S., De, N., and Pal, A. (2019). On some new neighbourhood degree based indices. arXiv preprint arXiv:1906.11215.

[33] Zaman, S., Hakami, K. H., Rasheed, S., and Agama, F. T. (2024). Reduced reverse degree-based topological indices of graphyne and graphdiyne nanoribbons with applications in chemical analysis. Scientific Reports, 14(1), 547.

[34] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., and Sun, M. (2020). Graph neural networks: A review of methods and applications. AI open, 1, 57-81.

[35] Li, Y., Hu, J., Wang, Y., Zhou, J., Zhang, L., and Liu, Z. (2019). DeepScaffold: a comprehensive tool for scaffold-based de novo drug discovery using deep learning. J. chem. information and modeling, 60(1), 77-91.

[36] Wakeel A., Ali K., Zaman S., Agama F.T., A python based algorithmic approach to optimize sulfonamide drugs via mathematical modeling, Scientific Reports (2024) 14:12264.

[37] Fraz M., Ali K., Nadeem M. F., Computational analysis of antiviral drugs using topological descriptors, Computer in biology and medicine, Volume 175, (2024), 108529.

[38] Farooq, F. B., Sultana, S., Alqahtani, N. A., & Imran, M. (2025). Optimization of HIV drugs through MCDM technique Analytic Hierarchy Process (AHP). PloS one, 20(1), e0316617. Journal of Pharmaceutical Innovation, 16(3), 456-467

[39] Sultanoglu, N., Uzun, B., Sayan, M., & Sanlidag, T. (2021). Evaluation of the effectiveness of adult HIV antiretroviral treatment regimens using TOPSIS. In Applications of Multi-Criteria

Decision-Making Theories in Healthcare and Biomedical Engineering (pp. 217-228). Academic Press.

[40] MC, S., B, K., A, U., & KC, S. (2025). Computational insights and study of drugs for dry eye disease through QSPR and MCDM approaches using topological indices. Scientific Reports, 15(1), 22245.

[41] Emery, A. E. The muscular dystrophies. Lancet 2002, 359 (9307), 687-695.

[42] Mercuri, E.; Bonnemann, C. G.; Muntoni, F. Muscular dystrophies. Lancet 2019, 394 (10213), 2025-2038.

[43] LaPelusa, A., Asuncion, R. M., & Kentris, M. (2024). Muscular dystrophy. StatPearls.

[44] Mercuri, E.; Muntoni, F. Muscular dystrophies. Lancet 2013, 381 (9869), 845-860.

[45] Mercuri, E.; Bonnemann, C. G.; Muntoni, F. Muscular dystrophies. Lancet 2019, 394 (10213), 2025-2038.

[46] Emery, A. E. H. The muscular dystrophies. Bmj 1998, 317 (7164), 991-995.

[47] Riguzzi, P., Sabbatini, D., Fusto, A., Vianello, S., Merlo, B., Zangaro, V., ... & Pegoraro, E. (2025). Deep characterization of females with heterozygous Duchenne muscular dystrophy mutations. Journal of Neurology, 272(3), 244.

[48] Sun, M. X., Jing, M., Hua, Y., Wang, J. B., Wang, S. Q., Chen, L. L., ... & Liu, Y. S. (2024). A female patient carrying a novel DMD mutation with non-random X-chromosome inactivation from a DMD family. BMC Medical Genomics, 17(1), 46.

[49] Goyenvalle, A.; Seto, J. T.; Davies, K. E.; Chamberlain, J. Therapeutic approaches to muscular dystrophy. Hum. Mol. Genet. 2011, 20 (R1), 69-78.

[50] Bez Batti Angulski, A.; Hosny, N.; Cohen, H.; Martin, A. A.; Hahn, D.; Bauer, J.; Metzger, J. M. Duchenne muscular dystrophy: Disease mechanism and therapeutic strategies. Front. Physiol. 2023, 14, 1183101.

[51] Vincik, L. Y., Dautel, A. D., Staples, A. A., Lauck, L. V., Armstrong, C. J., Howard, J. T., ... & Kaye, A. D. (2024). Evolving role of viltolarsen for treatment of duchenne muscular dystrophy. Advances in Therapy, 41(4), 1338-1350.

[52] Imran, M., Azeem, M., Jamil, M. K., and Deveci, M. (2024). Exploring innovative single-value neutrosophic fuzzy topological graph parameters. Granular Computing, 9(2), 37.

[53] Al Khabyah, A., Ahmad, H., Ahmad, A., & Koam, A. N. (2024). A uniform interval-valued intuitionistic fuzzy environment: topological descriptors and their application in neural networks. AIMS Mathematics, 9(10), 28792-28812.

[54] Khabyah, A., Ahmad, A., Azeem, M., Ahmad, Y., and Koam, A. N. (2023). Reverse-degree-based topological indices of two-dimensional coronene fractal structures. Physica Scripta, 99(1), 015216.

[55] Huang, X. Y., Zhang, B., Tian, Q., Wu, H. H., Gan, B., Bi, Z. N., ... and Wang, H. (2023). Machine learning study on time-temperature-transformation diagram of carbon and low-alloy steel. Journal of Iron and Steel Research International, 30(5), 1032-1041.

[56] Tawhari, Q. M., Rehman, M., Ahmed, W., Ahmad, A., & Koam, A. N. (2025). Exploring the potential of artificial neural networks in predicting physicochemical characteristics of anti-biofilm compounds from 2D and 3D structural information. Modern Physics Letters B, 2550157.

[57] Zaman, S., Yaqoob, H. S. A., Ullah, A., and Sheikh, M. (2023). QSPR Analysis of Some Novel Drugs Used in Blood Cancer Treatment Via Degree Based Topological Indices and Regression Models. Polycyclic Aromatic Compounds, 1-17.

[58] Chen, C.-T. (2000). "Extensions of TOPSIS for fuzzy multiple attribute decision-making." Fuzzy Sets and Systems, 114(1), 1-9.

[59] Yoon, K., and Hwang, C.-L. (1981). "Multiple Attribute Decision Making." Springer-Verlag.

[60] Mardani, A., Zavadskas, E. K., and Amiri, M. (2021). "A comprehensive review of multi-criteria decision-making methods in sustainable supply chain management." Sustainable Production and Consumption, 27, 674-691.

[61] Opricovic, S., and Tzeng, G.-H. (2020). "Extended TOPSIS for multi-criteria decision making with fuzzy and interval data." Expert Systems with Applications, 157, 113439.