International Journal of Chemical and Biochemical Sciences
(ISSN 2226-9614)

Journal Home page: www.iscientific.org/Journal.html

The Application of Machine Learning in Small Molecule Drug Discovery across Academic and Industrial Settings

Arunkumar DT[1], Kalyan Acharjya[2], KG Patel[3], Simran Kalra[4], Mohan Vishal Gupta[5]

[1]Assistant Professor, Department of Mechanical Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bangalore, India
[2]Assistant Professor, Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India
[3]Dean and Principal, College of Agriculture, College of Agriculture, Parul University, PO Limda, Vadodara, Gujarat, India
[4]Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India.
[5]Assistant Professor, College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India

Abstract

Small molecule drug development is an important endeavor that is explored in academia and industrial environments, which provides a particular benefits and issues. Academic and Industrial settings prioritize medication advancement through specialized skills, infrastructure, efficient processes and long-term objectives using the development in field of molecular machine learning (ML) that depend on both industry and academic research. High attrition rates in drug discovery pose challenges to academia and business due to unfavorable pharmacokinetics and off-target effects. To overcome this issue, we introduced a method termed as Stochastic chimp-optimized dynamic decision tree (SCO-DDT) method to optimize small molecule pharmacokinetics, reducing off-target effects and attrition in drug discovery, improving candidate selection, reducing trial failures in academic and industrial settings. We gather a dataset of drugs, which includes Drug Bank, PubChem and Therapeutics Data Commons (TDC). Z-score normalization is a technique which is used to preprocess the data that removes the distortions caused by different scales of features. Linear Discriminant Analysis (LDA) method is employed to extract the features of data. The findings of SCO-DDT model with other traditional methods revealed considerable gains in performance parameters such as accuracy (89.98%), prediction (78.62%), specificity (99.90%) and recall (85.32%). Finally, we provide several ideas that will further progress the area and can enhance collaboration between academic and industry organizations.

Keywords: Drug Discovery, Stochastic Chimp Optimized Dynamic Decision Tree, Academic and Industrial, Small molecule

Full length article **Corresponding Author*, e-mail: dt.arunkumar@jainuniversity.ac.in

1. Introduction

Molecular pharmacology is a small molecule of drug development, which closes the gap between scientific advancement and clinical use. Discovering new small molecule treatments was a dynamic convergence of fundamental science, clinical medical treatment and drug discovery that occurs in academic and industry contexts [1]. To improve patient care and solve unmet medical needs, coordinated activities with these different ecosystems use knowledge, resources, and technologies. Developments in small molecules represent the pinnacle of the attempt for fundamental knowledge and the real-world application of scientific discoveries to clinical issues in academic settings. [2]. Small molecules, on the foundation of innovation in pharmaceuticals and product development in the field of industrial drug discovery. Pharmaceutical industrials use high-throughput screening technology, computer modeling, medicinal chemistry knowledge, and large resources to accelerate the drug development procedure [3]. The meticulous experimental based on assumptions experiments, academic researchers investigate complex biological processes, reveal illness causes, and discover possible treatment targets. To bring new medicines to market, industrial investigators optimize chemical libraries, conduct preclinical research and navigate the challenging drug

development process, led by considerations of clinical viability, regulatory issues, and consumer demand [4]. Academic provides the basis for commercial drug development and industry collaborations that give academics access to vital infrastructure, resources and clinical translation routes. Together, they promote creativity, decrease the time it takes to find new drugs, and increase the probability that medical advancements affect patient care. The comprehensive investigation of small molecular drug development delves into the complex terrain that spans both academic and industry settings [5]. The perspective provided by the industry addresses large pharmaceutical businesses and it could not always reflect standard procedures in smaller biotechnology companies or startups. The convergence of academics and business produces a beneficial platform that stimulates creativity and advances the creation of novel treatments [6]. The introduction lays the groundwork for examining the complex interactions among small molecule drug development in the scientific and commercial domains, emphasizing the cooperative endeavors that propel advancement in the pursuit of revolutionary medicines. A few disadvantages of small molecule drug discovery in both academic and industrial contexts are possible for academic resource limitations that restrict access to specialized tools and knowledge, as well as difficulties in coordinating research goals and schedules between academic and industrial collaborators, which could impede progress [7]. The objective of the study is a small Molecule of Drug Development across Academic and Industrial Settings which aims to identify important factors that impact drug discovery results and promote cross-sector collaboration by analyzing and comparing the procedures, difficulties, and outcomes of medication discovery in both educational and industrial environments. The paper [8] described the current developments and ML applications in experimental drug discovery. They give an explanation of the improvements and strategies that are used at the stage to forecast the absorption, distribution, metabolism, and excretion (ADME) characteristics of small molecules based on their structures, as well as the structures that are predicted based on the qualities that are sought for molecular screening and optimization. The study [9] evaluated the related methodologies including ML and deep learning algorithms used in drug development. They explored the applications that yield approaches and outcomes that seem promising. The development of lead synthesis routes has emphasized the use of these simulations and comprehensive online information. The study [10] determined biotechnology and pharmaceutical sectors are becoming more and more interested in graph machine learning (GML) because of its capacity to combine information, model biomolecular structures, and the functional interactions between them. In this article, they provided a comprehensive academic-industrial evaluation of the subject of medication development and discovery. The paper [11] described outcomes that validated the model's strong performance and capacity for generalization. Furthermore, the results showed that a significant percentage of small compounds that were categorized as non-drugs determine the circumstances of bioavailability and it can be studied. Additionally, model tried to make use of such openings as a drug filter throughout the drug development process. The paper [12] determined molecular docking remains perfect considering its tremendous value to the drug development process. The study determined to present an overview of molecular docking and its techniques, emphasizing the importance of certain protocols and factors that can enhance the docking outcomes. These include consensus, active site waters and protonation states. The article [13] developed a drug discovery that was crucial for pharmaceutical firms as it involves discovering new candidate medications. Currently, drug discovery was costly and time-consuming. ML techniques, particularly deep learning, have demonstrated excellent performance in a variety of domains, and AI techniques are important for drug discovery. The paper [14] focused on associated uses after describing drug discovery. These uses can be condensed into two primary tasks: molecule creation and molecular characteristic predictions. Benchmark platforms, molecular representations and shared data sources are shown. Model architecture and learning frameworks are used to analyze AI approaches. The study [15] determined to drug investigation appears to benefiting from ML methods such as support vector machines (SVM), naïve Bayesian (NB), and deep neural networks (DNN). These make use of the larger datasets that are produced from large amounts of screening data and enable more accurate prediction of targets' bioactivities and molecular features. The study [16] evaluated the degradation of target proteins. Because it redefines the fundamentals of traditional drug discovery and it was driven by target activity that is event-based rather than occupancy-driven, Target Protein Degradation (TPD) offered a novel and innovative approach to treatments, with applications in chemical biology and drug discovery. The paper [17] provided brief findings on the key technological innovations that raised the status of alchemical free energy methods (AFEMs) from theoretical concepts to technology with widespread applications in the pharmaceutical and biotechnology sectors. The demanding absolute binding free energy (ABFE) calculations, which are used in computer-aided drug design (CADD) campaigns, should be avoided in favor of relative binding free energy (RBFE) computations. The study recognizes the unique advantages and difficulties comes with working in academia and industry settings while developing small-molecule drugs. The study emphasizes the value of cooperation between business and academia to drive advancements in the field of molecular ML method. Numerous large-scale databases, such as PubChem, Drug Bank, and TDC, offer invaluable insights for the discovery of small molecules and support research endeavors in both academic and industrial settings. The paper discusses the use of Z-score normalization in preprocessing, linear discriminant analysis (LDA) use of feature extraction, and the proposed method Stochastic Chimp Optimized Dynamic Decision Tree (SCO-DDT) model is employed for efficient chemical selection procedures. The rest of the paper is divided into several Sections. The data collection and methodology are covered in Section 2. The result analysis was the main focus of Section 3, while the discussion and conclusion were covered in Sections 4 and 5, respectively.

2. Materials and Methods

The following activities can be completed using the suggested approach. The dataset was collected and then preprocessed with the z-score normalization approach. In the section, a stochastic chimp optimized dynamic decision tree (SCO-DDT) is proposed to achieve the greatest performances in terms of small molecule drug discovery across academic and industrial settings with feature extracting capability of linear discriminant analysis (LDA) method.

2.1. Samples

Several extensive databases, including DrugBank, Therapeutics Data Commons (TDC), and PubChem, include information on various molecular characteristics crucial to the drug development process. A collection called DrugBank focuses exclusively on registered medications that are sold commercially and their intended uses. There are 3398 biologics and 15799 pharmaceuticals in the present edition, the bulk of which are small molecules. Another significant resource, PubChem, contains 307 million reported bioactivity and toxicity data points for around 117 million chemicals. A platform and program called TDC was created to make it easier to create new ML tools across a range of therapeutic domains. The TDC contains a total of 15919337 data points from 68 different datasets that have been carefully selected and prepared for the creation of ML models for 27 distinct prediction tasks.

2.2. Data pre-processing using z-score normalization

The small molecule drug development in both academic and corporate contexts that employs the statistical technique known as Z-score normalization. Drug development methods are made more accurate and efficient by standardizing data through the process of removing the mean and dividing by the standard deviation using equation (1).

$$d' = \frac{d - mean(P)}{std(p)} \quad (1)$$

When $mean(P)$= sum of each feature value for P, $std(p)$ =The standard deviations for each value of p.

2.3. Feature extraction using Linear Discriminant Analysis (LDA)

The field of academic and industry in LDA is a powerful method for small molecule drug discovery. It is used at different phases of the drug development process that helps to distinguish and categorize substances according to their chemical characteristics and biological activity. Using LDA makes more feasible to find possible drug candidates, which advances pharmacological research and increases the search for innovative therapeutic approaches in a variety of contexts. Consider the following: $Yi, \mu i, \Sigma i$ denotes the samples of dataset, average, and correlation matrix. The matrix of the two samples covariance is represented by $x^t\Sigma_{1x}$and $x^t x^t \Sigma_{2t}$, whereas the midpoint of samples $J\epsilon\{1,2\}$ is projected on line ω in $x^t(\mu_1 - \mu_2)$ and $\omega x^t(\mu_1 - \mu_2)^t x$. The long-term of LDA is to decrease the distance between homogeneous sample points,$x^t\Sigma_{1x} + x^t\Sigma_{2t}$ while minimizing the distance between varied sample points, i.e. $x^t(\mu_1 - \mu_2)(\mu_1 - \mu_2)^t$. Give the purpose of function J as follows using equation (2):

$$J = \frac{\|x^t\mu_1 - x^t\mu_2\|_2^2}{x^t\Sigma_{1x} + x^t\Sigma_{2t}} = \frac{x^t(\mu_1 - \mu_2)(\mu_1 - \mu_2)^t x}{x^t(\Sigma_1 + \Sigma_2)x} \quad (2)$$

The inner-class diverging matrix is defined as follows using equation (3):

$$S_W = \Sigma_1 + \Sigma_2 = \sum_{y\in Y_1}(y - \mu_1)(y - \mu_2)^t + \sum_{y\in Y_2}(y - \mu_1)(y - \mu_2)^t \quad (3)$$

Consequently, the divergence matrix among classes can be expressed as equation (4):

$$S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^t \quad (4)$$

$$J = \frac{x^t S_b x}{x^t S_w x} \quad (5)$$

The lagrangiis a multiplier approach and the singular value decomposition (SVD) method are used to solve this x matrix, which is $J = x^t S_b x$, or the best projection direction. Using the training data ω as $J = x^t S_w x$, the next step is to choose the best course of action using equation (5).

$$\begin{cases} Z > z_0 \Rightarrow y\epsilon Class_1 \\ Z < z_0 \Rightarrow y\epsilon Class_2 \end{cases} \quad (6)$$

$$z_0 = \frac{N_1\mu_1 + N_2\mu_2}{N_1 + N_2} \quad (7)$$

The discriminating outcome of N_1 and N_2 is found using equations (6) and (7), which display the quantities of real and simulated collisions.

2.4. Stochastic chimp optimized dynamic decision tree (SCO-DDT)

A potential strategy that can be used in both academic and corporate settings is the integration of the SCO-DDT technique with small-molecule drug development methods. Through the use of SCO-DDT properties and dynamic decision-making skills, scientists and professionals can improve the efficacy as well as effectiveness of drug development initiatives. With the potential to foster industry-academia collaboration on innovative treatment solutions, this combination might significantly advance drug research and development.

2.4.1. Dynamic decision tree (DDT)

An important development in the fields of academic and commercial research is the incorporation of the DDT approach in small molecule drug discovery. Its use makes a real-time adaptation quicker that enable investigators to negotiate the challenging landscape of drug development. The use of SCO-DDT speeds up the conversion of scientific discoveries into novel pharmacological treatments and improves decision-making

in a variety of contexts, including educational institutions labs, and industrial facilities. The classification procedure incorporates certain traits and recursively determines classes that differentiate the target application on all fronts. Let W represent the characteristics of a data point and Z represent the class in this example. Choosing the right category for the data point entails computing the ratio between Wand Z for equation (8).

$$\text{RATIO (W|Z)} = G(W) - G(\text{W|Z})G(W) \quad (8)$$

When variable W is given variable Z, the conditional entropy, represented by the notation $G(\text{W|Z})$, calculates the uncertainty of variable W. In contrast, the marginal entropy does not consider any other variables; rather, it just evaluates the uncertainty of variable W. Then $G(W)$ is the term used for equation (9).

$$C = \{(w1, z1), (w1, w2), \dots, (wM, zM)\} \quad (9)$$

Let $D_j = w1, w2, \dots, wm$ represent the feature vector of sample $j$in the regression tree, where W_{ji}represents the feature i of sample$j$. The input space is divided into L regions $(Q1, Q2, \dots, QL)$by the regression tree, which is associated with a distinct set of outcomes$(d1, c2, \dots., dl)$. As a consequence, we may express the regression model as follows equation (10):

$$z = e(w) = \sum dl * J(w \ \epsilon \ Ql) \quad (10)$$

When the specific result associated with area QL is denoted by dl, the regression function is represented by $e(w)$, the expected output variable is represented by z, and an indicator function, denoted by $J(w\epsilon Ql)$, evaluations to 1 ,when the input variable w falls inside QL and 0 otherwise using equation (11).

$$e(w) = \sum_{l=1}^{l} D_l J(w\epsilon Ql) \quad (11)$$

To get the values of i and t, it is essential to solve the following optimization problems.

$$\min i, t \min d1 \sum wi\epsilon Q1(i, r)(zi - d1)2 + \min d2 \sum wi\epsilon Q2(i, r)(zi - d2)2 \quad (12)$$

$$D1 = ave((zj|wj\epsilon Qj(i, t),), D2 = ave((zj|wj\epsilon Qj(i, t),) \quad (13)$$

The process comprises in selecting the optimal split variable$i$ after calculating the output values for each of the input variables. Every variable serves as a dividing line, splitting the input space into two distinct sections (i, t)using equation (12) and (13). After splitting up each region, the process is carried out again until a stop requirement is met.

2.4.2. Stochastic chimp optimization (SCO)

There is promise for both academic and industry contexts when integrating SCO into small molecule drug development methods. SCO-DDT provides an adaptable method for effectively exploring complicated search areas, modeled after the foraging procedures of primates. Its use improves chemical compound research, stimulating creativity and quickening the pace of medicine development. Academic-industry collaborations benefit from SCO-DDT flexible approaches, which push the boundaries of medicinal study and development. The capacity to think for themselves allows all chimp species to locate prey and employ their unique search approach to find it. Even while they carry out their responsibilities, individuals are also socially driven to get advantages and sex during the last phases of the search. The disorderly, solitary gathering conduct takes place throughout this phase. To say that there are Mchimpanzees and that Wj is the i^{th} chimp's location. The following describes the behavior of the chimps as they approach and surround it, as well as their position update with equations (14) and (15):

$$C = |D.W_{prey}(s) - n.W_{chimp}(s)| \quad (14)$$

$$W_{chimp}(t+1) = W_{chimp}(s) - B - C \quad (15)$$

$$B = e.(2.q_1 - 1), D = 2.q_2 \quad (16)$$

$$n - chaotic - value \quad (17)$$

Where the values of the random vectors $q1$ and $q2$ are between 0 and 1. The value of the non-linear decay factor, or e, drops linearly from 2.7 to 0 as the quantity of iterations rises. The total of iterations in use is indicated by s. The value of random vector B is a random number in the interval $[-e, e]$. When d is the random variable using equation (16) and (17).

$$\begin{matrix} C_{attacker} = |D_1 * W_{attacker} - w_1 W| \\ C_{Barrier} = |D_2 * W_{Barrier} - w_2 W| \\ C_{chaser} = |D_1 * W_{chaser} - w_3 W| \\ C_{Driver} = |D_1 * W_{Driver} - w_4 W| \end{matrix} \quad (18)$$

$$\begin{cases} W_1 = W_{Attacker} - B_1 * C_{Attacker} \\ W_2 = W_{Barrier} - B_2 * C_{Barrier} \\ W_3 = W_{Chaser} - B_3 * C_{Chaser} \\ W_4 = W_{Driver} - B_4 * C_{Driver} \end{cases} \quad (19)$$

$$W(s+1) = (W_1 + W_2 + W_3 + W_4)/4 \quad (20)$$

As $A1, A2, A3, and\ A4$ are similar to A, then $C1, C2, C3, and\ C4$ are similarly to C with both equation (18), (19) and (20). Also comparable to m are $m1, m2, m3, and\ m4$. Algorithm 1 contains a list of the SCO-DDT pseudo-code.

Algorithm 1- Pseudocode of SCO-DDT

def_Stochastic_ chimp_ optimized_ dynamic_ decision _tree ($W\_chimp\_initial, iterations$):
for s in range(iterations):
$q1, q2 = random\_vector(), random\_vector()$
$e = linear\_decay\_factor(s)$
$B = e * (2 * q1 - 1)$
$D = 2 * q2$
$C = calculate\_c(W\_prey, W\_chimp, D)$
$W\_chimp = update\_position(W\_chimp, B, C)$
defdynamic_decision_tree(data, labels, max_depth):
if stopping_criteria(data, labels, max_depth):

returncreate_leaf_node(labels)
$i, t = find\_optimal\_split(data, labels)$
data_left, labels_left, data_right, labels_right = split_data
left_subtree = dynamic_decision_tree
right_subtree = dynamic_decision_tree
returncreate_decision_node
defintegrate_sco_ddt(data,labels,max_depth, sco_iterations):
stochastic_chimp_optimization
decision_tree = dynamic_decision_treereturndecision_tree
data, labels = load_data()
max_depth = 5
sco_iterations = 100
final_decision_tree = integrate_sco_ddt

3. Results and discussion

In this research, Python 3.1 was used extensively during the inquiry. It gives Intel Core i5 laptops with 32GB SSDs and Windows 8. Here, the recommended system's efficacy is evaluated. The assessment factors include recall, specificity, accuracy, and precision. A comparative analysis has been conducted using the proposed techniques SCO-DDT, Adaboost decision tree (ABDT) [18], decision trees (DT) [18], and random forest (RF) [18]. (Table 1) shows the outcomes of suggested and existing methods.

3.1. Accuracy

The accuracy is essential for small molecule drug discovery in both academic and corporate contexts since it helps to identify and produce therapeutic molecules that work. Drug candidates with a high chance of clinical success and patient benefit must be advanced through the integration of exact techniques and stringent validation processes. ABDT scored 72.62%, DT scored 79.22%, and RF scored of 76.76%. With an excellent accuracy of 89.98%, the suggested algorithm, SCO-DDT, greatly surpassed the others. The suggested SCO-DDT algorithm demonstrates great promise in its classification tasks, with these percentages representing the effectiveness and performance of each approach. The comparison of Accuracy is shown in (Fig.1).

3.2. Precision

The precision in small molecule drug discovery have the potential to improve target specificity, optimize drug development pipelines, and promote teamwork in academic and industrial contexts. This increases the efficacy and achievement of drug-discovering processes in both domains. ABDT scored 32.47%, while DT scored 60.10%, and RF scored 39.22%, respectively. The suggested SCO-DDT algorithm has the greatest accuracy of 78.62%.The comparison of precision is shown in (Fig.2).

3.3. Specificity

The specificity considerations are essential for improving target interactions, reducing off-target effects, and increasing therapeutic efficacy in academic and industrial settings where small molecule drug discovery is being conducted. This ultimately lead to the development of more accurate and potent drugs with a variety of study and development systems. ABDT scored 88.17%, while DT scored 98.92%, and RF scored 94.17%, respectively. With an excellent accuracy of 99.90%, the suggested algorithm, SCO-DDT, greatly surpassed the others. These percentages show the performance and efficacy of each method, with the proposed SCO-DDT algorithm showing tremendous potential in its classification tasks. The comparison of Specificity is shown in (Fig.3).

3.4. Recall

The recall rates for small molecules must be high in both academic and industrial settings to identify candidates for potential drugs, reduce the possibility of the absence of important compounds and improve the general efficacy and dependability of the process of drug discovery in a variety of research settings. ABDT (22.5%), DT (70.80%), and RF (39.22%), the suggested approach SCO-DDT has the maximum recall level of 85.32%. The comparison of Recall is shown in (Fig.4). The ABDT is susceptible to adjust because of its too-noisy input and errors. Because boosting is iterative, it can be computationally costly and since the quality, level of initial learner has a significant impact on performance, it may not be as successful in some situations [18]. The RF has limitations despite its strength and adaptability. Particularly for big datasets and intricate forests, RF can be operationally costly. Compared to simpler models, the model's intricacy makes interpretation difficult. Furthermore, RF does not have clear data, resulting in a decrease in generalization performance. Finally, unbalanced datasets may cause RF to perform poorly and necessitate the use of extra handling methods. Firstly, they are prone to adapting the training set, which leads to inadequate extrapolation to new data. Second, they produce distinct tree architectures that respond even minute differences in the training set. Finally, DT tends to favor majority classes over minority ones, they may not perform well on unbalanced datasets and produce incorrect predictions. The SCO-DDT approach lessens the drawbacks of conventional techniques. It increases flexibility and lessens sensitivity to noisy input by adding stochasticity. Its dynamic character corrects overcomes and balances disparities in categories. By optimizing, SCO-DDT enhances generalization and allows for accurate findings on a variety of datasets.

Table 1. Numerical outcomes of parameters

Method	Accuracy (%)	Precision (%)	Specificity (%)	Recall (%)
ABDT [18]	72.62	32.47	88.17	22.5
DT [18]	79.22	60.10	98.92	70.80
RF [18]	76.76	39.22	94.17	39.22
SCO-DDT [Proposed]	89.98	78.62	99.90	85.32

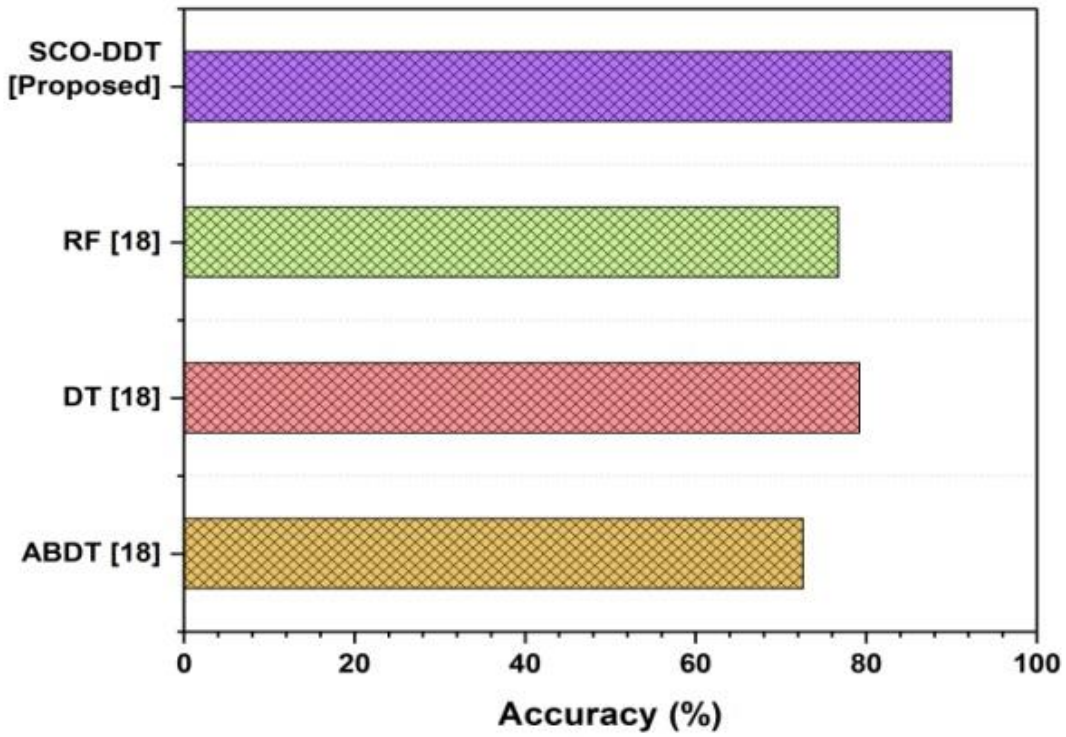

Figure 1. Comparison of Accuracy

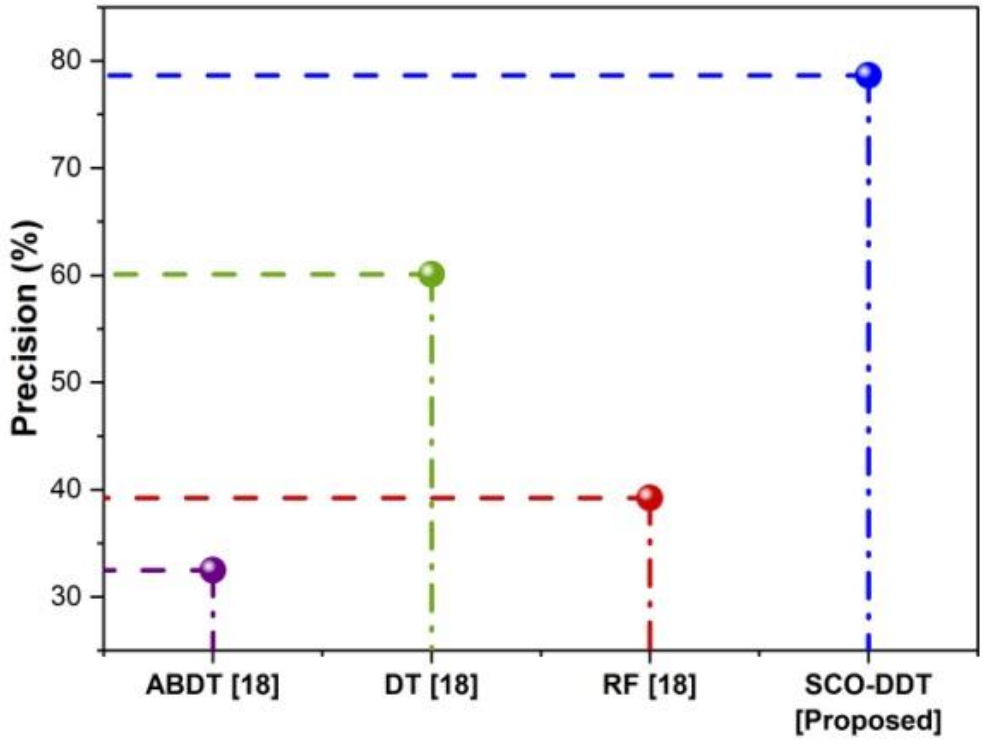

Figure 2. Comparison of Precision

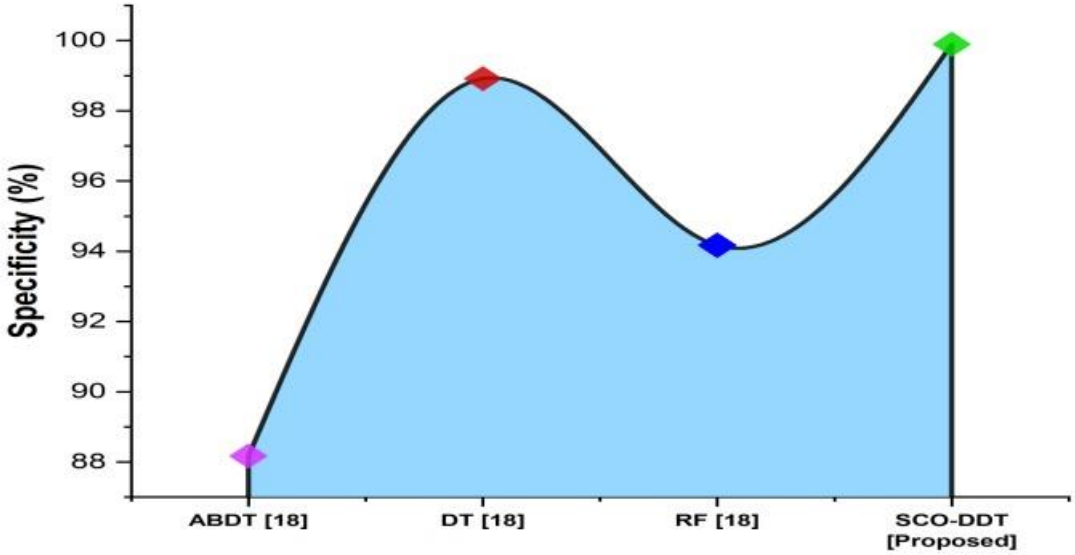

Figure 3. Comparison of Specificity

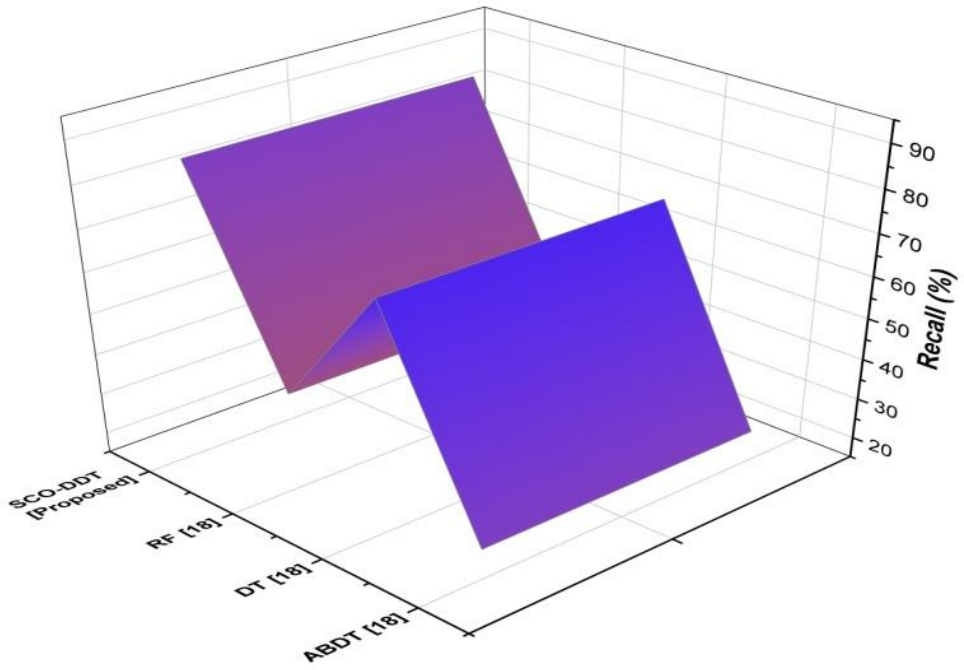

Figure 4. Comparison of Recall

4. Conclusions

The creation of small molecule medications is an important endeavor carried out by both business and academics, each with its benefits and difficulties. Academic settings are driven by scientific curiosity to perform basic research and start drug development at an early level, whereas industrial contexts use specialized expertise, a strong infrastructure, and simplified procedures to progress drug candidates rapidly. Both academic and corporate research share common study areas and long-term aims in advancing molecular ML, despite differing techniques and ranges. Furthermore, methods such as Z-score normalization and LDA are essential for data preparation and enhancing the performance of ML. The study comparison of the SCO-DDT model with other traditional methods revealed considerable gains in performance parameters such as accuracy (89.98%), prediction (78.62%), specificity (99.90%), and recall (85.32%).The potential for improving and expediting chemical selection procedures is highlighted in this study, specifically about SCO-DDT models. Then the SCO-DDT have a few limitation of this process might result in unexpected behavior and in dynamic contexts, inferior or inconsistent decision-making outcomes. Academic and industry environments for small molecule drug development in the future. Drug development is expected to speed when novel chemical synthesis methodologies are combined with omics data, such as proteomics and genomes, to address a variety of therapeutic targets.

References

[1] S. Ekins, A.C. Puhl, K.M. Zorn, T.R. Lane, D.P. Russo, J.J. Klein, A.M. Clark. (2019). Exploiting machine learning for end-to-end drug discovery and development. Nature materials. 18 (5) 435-441.

[2] R.J. Young, S.L. Flitsch, M. Grigalunas, P.D. Leeson, R.J. Quinn, N.J. Turner, H. Waldmann. (2022). The time and place for nature in drug discovery. JACS Au. 2 (11) 2400-2416.

[3] T.S. Lee, B.K. Allen, T.J. Giese, Z. Guo, P. Li, C. Lin, D.M. York. (2020). Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. Journal of Chemical Information and Modeling. 60 (11) 5595-5623.

[4] G. Nishiguchi, S. Das, J. Ochoada, H. Long, R.E. Lee, Z. Rankovic, A.A. Shelat. (2021). Evaluating and evolving a screening library in academia: the St Jude approach. Drug discovery today. 26 (4) 1060-1069.

[5] P. Agarwal, J. Huckle, J. Newman, D.L. Reid. (2022). Trends in small molecule drug properties: A developability molecule assessment perspective. Drug Discovery Today. 103366.

[6] S. Ramakrishnan, J.S. Weerakkody. (2022). Suspended Lipid Bilayer: A Versatile Platform for Nextgen Drug Discovery and Biomedical Applications. Accounts of Materials Research. 3 (10) 996-998.

[7] J.V. Chari, R.R. Knapp, T.B. Boit, N.K. Garg. (2022). Catalysis in Modern Drug Discovery: Insights from a Graduate Student-Taught Undergraduate Course. Journal of Chemical Education. 99 (3) 1296-1303.

[8] N. Pillai, A. Dasgupta, S. Sudsakorn, J. Fretland, P.D. Mavroudis. (2022). Machine learning guided early drug discovery of small molecules. Drug Discovery Today. 27 (8) 2209-2215.

[9] L. Patel, T. Shukla, X. Huang, D.W. Ussery, S. Wang. (2020). Machine learning methods in drug discovery. Molecules. 25 (22) 5277.

[10] T. Gaudelet, B. Day, A.R. Jamasb, J. Soman, C. Regep, G. Liu, J.P. Taylor-King. (2021). Utilizing graph machine learning within drug discovery and development. Briefings in bioinformatics. 22 (6) bbab159.

[11] S.A. Hooshmand, S.A. Jamalkandi, S.M. Alavi, A. asoudi-Nejad. (2021). Distinguishing drug/non-drug-like small molecules in drug discovery using deep belief network. Molecular Diversity. 25 827-838.

[12] F. Stanzione, I. Giangreco, J.C. Cole. (2021). Use of molecular docking computational tools in drug discovery. Progress in Medicinal Chemistry. 60 273-343.

[13] N. Stephenson, E. Shane, J. Chase, J. Rowland, D. Ries, N. Justice, R. Cao. (2019). Survey of machine learning techniques in drug discovery. Current drug metabolism. 20 (3) 185-193.

[14] J. Deng, Z. Yang, I. Ojima, D. Samaras, F. Wang. (2022). Artificial intelligence in drug discovery: applications and techniques. Briefings in Bioinformatics. 23 (1) bbab430.

[15] F. Boniolo, E. Dorigatti, A.J. Ohnmacht, D. Saur, B. Schubert, M.P. Menden. (2021). Artificial intelligence in early drug discovery enabling precision medicine. Expert Opinion on Drug Discovery. 16 (9) 991-1007.

[16] N. Guedeney, M. Cornu, F. Schwalen, C. Kieffer, A.S. Voisin-Chiret. (2023). PROTAC technology: A new drug design for chemical biology with many challenges in drug discovery. Drug Discovery Today. 28 (1) 103395.

[17] L.F. Song, K.M. Merz Jr. (2020). Evolution of alchemical free energy methods in drug discovery. Journal of Chemical Information and Modeling. 60 (11) 5308-5318.

[18] M.A. Wani, K.K. Roy. (2022). Development and validation of consensus machine learning-based models for the prediction of novel small molecules as potential anti-tubercular agents. Molecular Diversity. 26 (3) 1345-1356.