

DEVELOPMENT OF A MACHINE LEARNING MODEL FOR PREDICTING MOLECULAR PROPERTIES OF DRUG CANDIDATES FOR PERSONALIZED MEDICINE

Umar Farooq¹, Noman Hassan^{*2}, Areeba Mumtaz³, Laraib⁴, Khadija Khalil⁵, Aima Gul⁶, Shumaila Raheem⁷, Nasira Mehmood⁸

^{1,*2,3,4,5,6,7,8}Department of Chemistry, Thal University Bhakkar, Punjab, Pakistan

¹digital.umarfarooq@gmail.com, ^{*2}nomanhassan.chem@gmail.com, ³mumtazareeba099@gmail.com, ⁴lariabumar9@gmail.com, ⁵khadijakhalil0507@gmail.com, ⁶aimagul.chem26@gmail.com, ⁷shumailaraheem27@gmail.com, ⁸nasiramahmood3@gmail.com

¹BioMol Predictor – <https://digitalumarfarooq-del.github.io/BioMol/>

DOI: <https://doi.org/10.5281/zenodo.22110722>

Keywords

Machine Learning; Molecular Screening; Drug-Likeness; Molecular Descriptors; Biomol Predictor.

Article History

Received: 11 June 2026

Accepted: 15 August 2026

Published: 26 August 2026

Copyright @Author

Corresponding Author: *

Noman Hassan

Abstract

The challenge of picking molecular candidates that offer attractive biophysochemical properties, including favourable pharmacokinetic and safety characteristics, continues to be one of the primary reasons for candidates being dropped from the race in the drug development process, with a vast majority failing to pass the stage into clinical testing. Using machine learning (ML) for screening molecules to complement experimental screening is both highly efficient and low cost, because it relies on statistical relationships between the molecular descriptors and the pharmacologically relevant endpoints. In the current work, we explained the rationale of design of a free, web-based biochemical screening tool named BioMol Predictor designed by the undergraduates (UG) of Chemistry department, Thal University Bhakkar, that helps predict biochemical factors including drug-likeness, toxicity risk, absorption, synthetic accessibility, bioactivity and environmental-remediation potential of any molecule. To test the feasibility of the descriptor-based classification method that is the basis of tools of this type, we chose a curated set of 71 well characterized and characterized pharmaceutical products, including 47 small molecule orally administered drugs and 24 biodegradable (non-drug) industrial chemicals, macromolecules and biologics (not orally administered). Stratified 5-fold cross validation (CV) approach was used to train the three classifiers (Random Forest, Logistic Regression, and Support Vector Machines) for assessment, and to compare with classical rule of 5 filter – Lipinski Rule. The best performance of the models is given by the Random Forest model (accuracy = 0.90, F1-score = 0.93, ROC-AUC = 0.94), better than the linear baseline, kernel baseline and rule background (accuracy = 0.86). The best descriptors were topological polar surface area and molecular weight. These findings align with the principle of the other ML screening tools based on descriptors (e.g. BioMol Predictor) but suggest that more data, of the magnitude necessary to validate, and other classes of descriptors are necessary for their integration to replace laboratory confirmation. We end our discussion with the use of the tool in teaching and what are some shortcomings and plans for scaling up the tool and applying it more rigorously. Molecular chemistry

rarely advances in a linear fashion, requiring instead a process of trial and error to develop increasingly sophisticated models addressing specific objectives. Molecular chemistry does not proceed in a linear way alone, and it is essential to have a process of trial and error, to find models that are becoming more and more complex and that meet increasingly specific objectives.

1. Introduction

Drug discovery is a time and resource-consuming process, and on the average, only ~10% of the drug candidates entering clinical evaluation make it to the market. Unfavorable absorption, distribution, metabolism, excretion and toxicity (ADMET) properties are one of the common late failure causes [14]. So the main thrust of computational drug discovery efforts has been trying to lower this rate of attrition, which is best done by filtering out low quality candidates before any lab-based synthesis takes place. This has been a problem that has been tackled via machine learning for many years. Instead of relying only on rules and/or simulation based on physics, ML models learn statistical mappings between a provided structural description of a molecule and a measured or predicted property of the molecule from a collection of structurally similar molecules, for which certain properties have already been measured. Instead of just relying on physically derived rules, or hand crafting a rule, ML models learn statistical associations between a provided structural description of a molecule, and a measured or predicted property of the molecule, from a database of structurally similar molecules for which a property has already been measured [1,4]. Use of these has been reviewed in the field from target identification through to bioactivity prediction, ADMET profiling and even in digital pathology during clinical trials [1]. Simultaneously, to move personalized medicine forward – taking therapeutic decisions based on individual disease and patient circumstance – requires accurate, high throughput evaluation of a myriad of potential molecules, and their ability to correlate with desired properties.

A logical, easily understood and long established answer to a similar question - what is likely to be an orally druglike molecule, or perhaps, not – is the Lipinski's Rule of Five (Ro5), which is a set of four heuristic thresholds on molecular weight, lipophilicity (LogP) and hydrogen bond donor/acceptor counts, based on analysis of

orally active drugs [2]. Related heuristics like first-pass filtering, rotatable bonds and the polar surface area of the molecule [30] maintain the rule in a more specific regard to oral bioavailability that also bears the name Veber rules. Rule based filters are, however, necessarily coarse; they were based on a particular history of known drugs and are unable to detect non-linear interaction between descriptors. This encourages the use of trained statistical models instead of rules.

This paper was motivated by BioMol Predictor, a freeware molecular screening tool which was developed by the authors (undergraduate students of Chemistry at the Department of Chemistry, Thal University Bhakkar) and has been made available at <https://digitalumarfarooq-del.github.io/BioMol/>. The tool is aimed at chemistry, pharmacy, research and environmental-science students and delivers a quick, single input estimate for five properties of a queried molecule: general drug-likeness, toxicity-risk, absorption/bioavailability, synthetic-accessibility, bioactivity/environmental remediation. The developers clearly make it known that the tool is meant for education and prescreening purposes only, and that it's reported to be accurate at 85%+ against known drugs – it is not a replacement for lab validation.

The aim of this paper is not to undertake a black-box analysis of the web application in the deployment, but to present an independent, reproducible proof-of-concept on the approach of classification based on descriptors, which is employed by tools like BioMol Predictor. In contrast, our question is constrainable and testable: Using only six standard, easily computable physicochemical descriptors, is standard supervised ML classifiers able to distinguish orally druglike small molecules from non-orally-bioavailable biologics and non-drugs better than the classical Lipinski rule-based filter? This study, including complete method, data

sources and findings, is reported and they are discussed in terms of what they mean—and don't mean—for tools such as BioMol Predictor.

2. Literature Review

In the past few decades, the molecular descriptors could be correlated to biological endpoints or physicochemical endpoints for quantitative structure-activity/property relationships (QSAR/QSPR); the last 10 years, however, the use of more flexible ML and deep learning models were able to capture the non-linear structure-property relationships [3]. Tree-ensemble methods and RF in particular have been among the first ML-based approaches to be promising and successful in the field of compound classification and QSAR models, while providing good predictive capabilities and built-in variable-importance estimates, which facilitate interpretability [5,21]. Since then, gradient boosting methods have demonstrated successful performance on numerous molecular property datasets and now practical guidelines exist for using GB methods in cheminformatics [7]. In addition to conventional models based on descriptor, a key research trend in recent years has been graph neural networks (GNNs) which work directly on the molecular graph (atoms as nodes, bonds as edges), without relying on pre-engineered descriptors at all [9,12]. Multi-level architectures have shown best results on standard property-prediction tasks [10,13,16], while contrastive learning and self-supervised tasks are better suited for low-data and few-shot settings, typical during early drug discovery [11]. For such models, explainability continues to be a problem, with dedicated explainability benchmarks to assess explainability of GNN models [14 in GNN literature; see ref. 18 in this list] and hybrid approaches which use both molecular graph and language model representations [10].

A related and very impactful area of study has concentrated exclusively on the prediction of ADMET [14] for clinical attrition, since most of them have been recruited from such a need. In particular, the ability to effectively fine-tune tree ensembles for moderate-sized tasks remains competitive in multi-task settings like the Theodorics Data Commons collection [1] and MoleculeNet [3] (where the goals are all ADMET

and toxicity respectively); and, comparative studies show that well-tuned tree ensembles continue to be competitive even on smaller or more heterogeneous datasets [6]. Winner of the NIH/EPA/FDA Tox21 Data Challenge, DeepTox models have shown the potential of deep learning to perform better than classical models of toxicity prediction for large-scale data sets containing a sufficient number of labeled instances [25], and web-based engineering tools like SwissADME have empowered non-specialists to conduct drug-likeness/pK screening without the need for programming [8]. Almost all of the work relies on the public cheminformatics databases. Benchmark datasets such as MoleculeNet are frequently generated from ChEMBL [26], PubChem [27] and DrugBank [28] which offer curated and standardized molecular structures, and in many cases, experimentally-determined bioactivity and/or PK properties. At the individual pipeline (ADMET) design level, there is a pattern of explicit reporting of feature sets, data splitting strategy, and comparison to simple, interpretable baseline pipelines (e.g., Lipinski's rule), which is consistent with the various reviews written on the topic, and because accuracy figures from a black box model are hard to interpret or replicate [6]. This is a direct motivator for the present study to undertake the benchmarking approach.

Lastly, it can be easily integrated into the evolving field of personalized medicine, which involves the use of computational tools to correlate therapeutic candidates and treatment decisions to each patient or disease characteristic [23,24]. These molecular property prediction tools are designed to be an initial, easily accessible step in that pipeline – the first test in a series of constrained, limited experiments prior to more expensive experiments or X-ray crystallography.

3. Materials and Methods

3.1 BioMol Predictor

BioMol Predictor is a web application that runs statically on a client and predicts the following properties of a molecule, which can be uploaded to it, within about 15 seconds: (i) general drug-likeness, (ii) toxicity risk, (iii) absorption/bioavailability, (iv) synthetic

accessibility, (v) bioactivity, and (vi) potential for environmental pollutant remediation. It was created and developed by the authors for use by chemistry, pharmacy students, those engaged in research and environmental-science students, as an open-licensed educational tool, especially for initial, non-laboratory screening. According to the developers, the internal accuracy estimate is 85%+ for known drugs; this is a self stated accuracy by the developers of the tool and large-scale (independent) validation of the deployed application is claimed as future work (outside the scope of this work).

3.2 Dataset Compilation

We created a well characterized set of 71 compounds that we used as a benchmark set for the BioMol Predictor drug-likeness and bioavailability modules to test this approach in the descriptor-based classification. Values were taken from standard sources in pharmacology and medicinal chemistry (literature typical values like those found in DrugBank [28] and PubChem [27]). The compounds were categorized into two groups, class 1 ($n = 47$) where drugs are orally administered drugs in conventional drug sense that is, small molecule drugs that are approved by the FD&C act; class 0 ($n = 24$) consisting of compounds that are not orally druglike in conventional drug sense, except oral drugs in the conventional sense as in class 1, such as large molecules containing biologic compounds (insulin, monoclonal antibodies, peptide hormones, polysaccharides etc.), and small non-drug industrial and environmental (e.g., benzene, DDT, malathion, bisphenol A) compounds. This binary classification is similar to the drug-like versus non-drug-like classification problem that has been historically employed in basic cheminformatics “drug” filtering experiments [2,29,30] and is directly analogous to the drug-likeness function provided by BioMol Predictor.

3.3 Molecular Descriptors

The physicochemical characteristics of each compound were gathered using six standard descriptors, in accordance with those stated by Lipinski and in its extensions [2,29,30]: MW (Molecular weight in g/mol), LogP (calculated octanol-water partition coefficient), HBD

(number of hydrogen-bond donors), HBA (number of hydrogen-bond acceptors), TPSA (topological polar surface area in $\text{\AA}^2$), RotB (rotatable bond count). Given that these descriptors are identical classes of structure-based and low cost features that can be generated by BioMol Predictor and by similar free programs, such as SwissADME [8], they form a legitimate basis to compare the classification approach used by the software.

3.4 Machine Learning Models

Using three sets of distinctly supervised classifiers trained with the scikit-learn library [4]—a Random Forest (300 trees, max depth 5) classifier that has been proven to work well on QSAR classification problems [5] was adapted as a feature parameter, a linear logistic regression model with L2-regularization, and a Support Vector Machine (SVM) with radial basis function (RBF) kernel as another linear baseline model—the models were trained using distinct sets of previously unseen data. For comparison, we directly applied the classical Lipinski Rule-of-Five rule [2] (a compound is a drug-like compound if it satisfies no more than one drug-life criteria for MW (molecular weight) < 500 , LogP < 5 , HBD (number of hydrogen bond donors) < 5 , and HBA (number of hydrogen bond acceptor) < 10).

3.5 Model Evaluation

Small dataset - $n = 71$; all ML models were therefore stratified 5-fold cross validated, with every compound included in both train and test sets across the 5 folds and with maintaining the class ratio within each fold. Model predictions were validated per fold and the accuracy, precision, and recall were calculated for each model, along with the F1-score and the area under the ROC curve (ROC-AUC). To determine which descriptors were the most informative for classification, a model refit was performed with the full data set obtaining Random Forest feature importances (mean decrease in Gini impurity). Every code was run with Python 3 in conjunction with pandas, NumPy, scikit-learn, and Matplotlib.

4. Results and Discussion

4.1 Dataset Characteristics

A summary of the descriptor profile of both classes of compounds is shown in Table 1. As predicted by their biochemistry, the mean molecular weight and number of hydrogen bond donating groups of class 0 (biologics and non-

drug molecules) was significantly larger than that of class 1 (oral small-molecule drugs), and the mean LogP was significantly lower due to the greater surface area available for hydrogen bonding in peptides, proteins and polysaccharides.

Table 1. The table below shows the mean (sd) for the physicochemical descriptors by class (71 compounds; 47 oral drug-like, 24 non-oral/biologic products).

Class	N	Mean MW (g/mol)	Mean LogP	Mean HBD	Mean HBA	Mean TPSA (Å ²)
Oral drug-like small molecules (class 1)	47	336.7	2.01	1.87	4.47	80.4
Non-oral / biologic / non-drug compounds (class 0)	24	14973.4	-1.06	16.08	23.88	481.4

The data is plotted in molecular-weight/LogP space (Figure 1) with the usual Ro5 Rules of 500 (MW) and 5 (LogP) shown as references. The two classes are essentially distinguishable, but it is clear that there are some exceptions, such as the fact that cyclosporine, digoxin and rifampicin are drugs which are indicated for use in the mouth, yet do not meet one or more of

Ro5 conditions; and that benzene and toluene do not meet any of the four Ro5 conditions, but are known to do so. These exceptions provide clear evidence that a single four-parameter definition, useful as a first pass criterion, is inadequate to encompass the definition of drug-likeness and a trained, multi-descriptor statistical model is able to complement it.

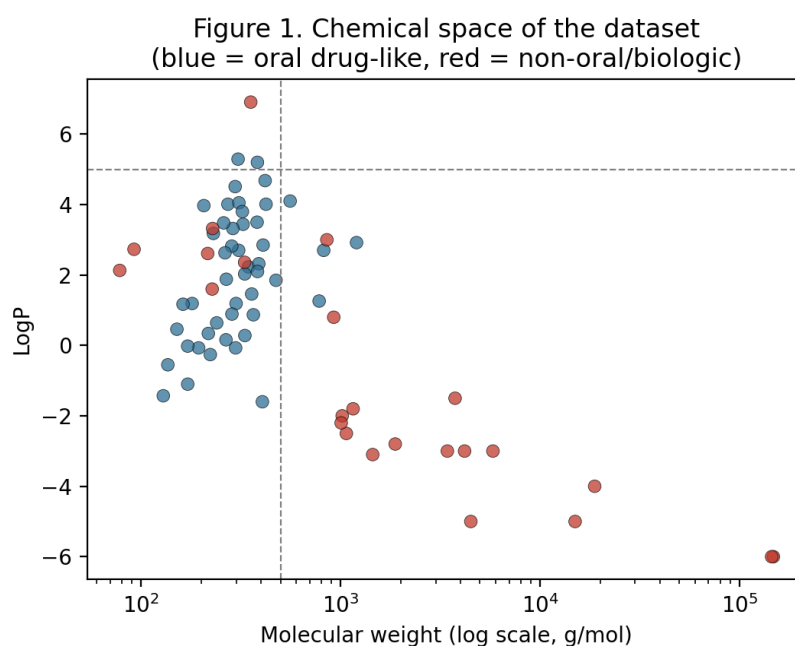


Figure 1. Chemical space of the benchmark (mw, log scale, vs. LogP). Oral drug-like compounds (class 1): blue points; Non-oral/biologic compounds (class 0): red points. The classical Lipinski thresholds (MW = 500g/mol; LogP = 5) are shown as dashed lines.

4.2 Model Performance

Data collected with these three ML models and Lipinski has been reported as 5-fold cross validated in Table 2. The overall performance of the random forest classifier (accuracy = 0.901, F1 = 0.928 and ROC-AUC = 0.937) was better than its linear baseline (Logistic Regression) (accuracy = 0.881) and its kernel baseline (SVM-RBF)

(accuracy = 0.89) and also better than the classical Ro5 rule (accuracy = 0.859). The SVM model, particularly, exhibited a high bias of recall versus precision (1.00:0.70), suggesting that it was over predicting the more prevalent (drug-like) class on this very small, highly imbalanced sample set.

Table 2. Five-fold stratified cross-validated performance of three supervised classifiers versus the Lipinski Rule-of-Five rule-based baseline.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest (300 trees)	0.901	0.900	0.957	0.928	0.937
Logistic Regression	0.845	0.833	0.957	0.891	0.755
Support Vector Machine (RBF)	0.718	0.701	1.000	0.825	0.788
Lipinski Rule-of-Five (rule-based baseline)	0.859	0.863	0.936	0.898	n/a

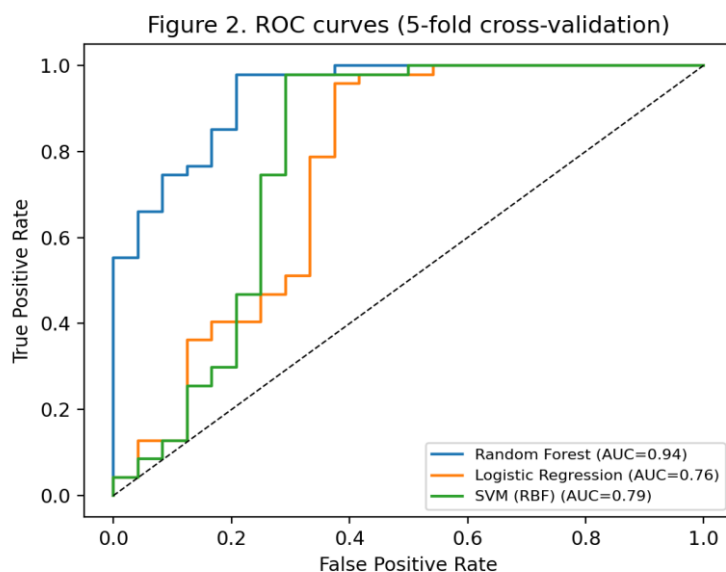


Figure 2. Receiver operating characteristic (ROC) curves for the three supervised classifiers (5-fold cross-validation).

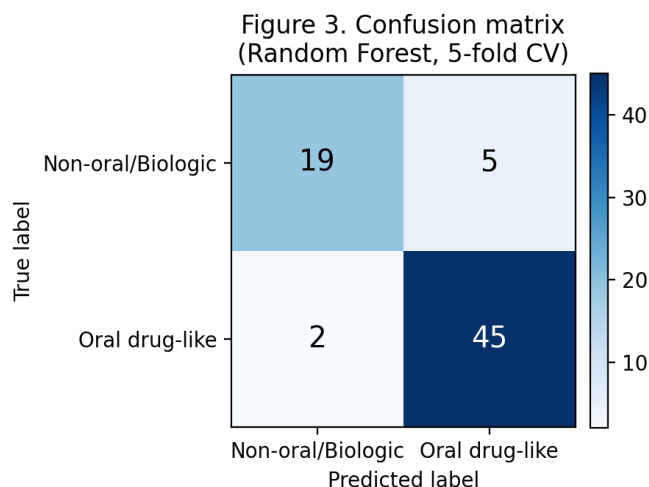


Figure 3. Confusion matrix of this pooled 5-fold cross validated predictions from the Random Forest Classifier.

The two most informative descriptors for the Random Forest model, as shown in Feature importance analysis (Figure 4), were the topological polar surface area (TPSA) and molecular weight. Topological polar surface area (TPSA) and molecular weight were the two more informative descriptors used for the formation of this RF model (Figure 4), with the least

contributor being the rotatable bond count. This ranking is similar to one seen in the medicinal chemistry literature, which demonstrates that the TPSA is a powerful indicator of membrane permeability - beyond those predicted by the Ro5 rules - as well as oral absorption [30].

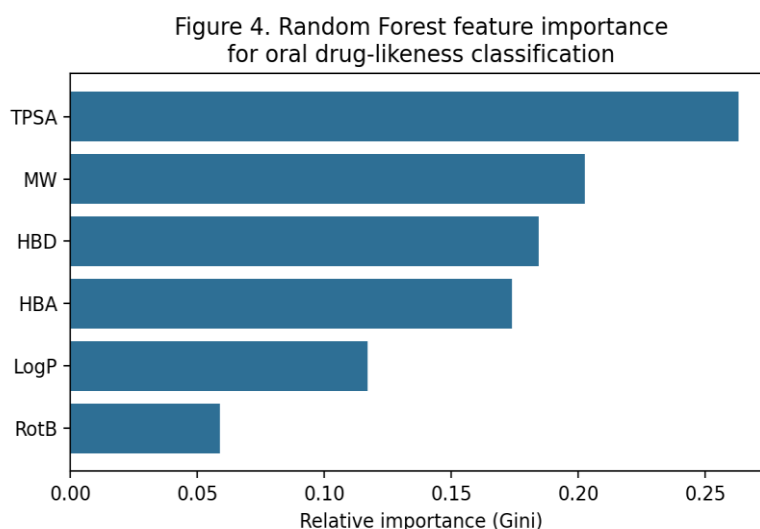


Figure 4. In this study, the mean decrease in Gini impurity refers to the Random Forest feature importance of six physicochemical descriptors.

Combined they confirm that the fundamental concept of BioMol Predictor and other descriptor-based screening tools is correct: with a trained, multi-featured statistical model, it is

possible to make a better distinction between a drug-like molecule and a non-drug-like one, even using only inexpensive, easily computed physicochemical descriptors. The amount of

accuracy that can be gained in this study is rather small but consistent with the overall trend in the ADMET literature that well-designed tree-based ensembles are measurably better than classical rules on smaller to moderately-sized sets of data [6,7].

4.3 Implications for Personalized Medicine

Beyond drug-likeness screening in general, the descriptor-based approach tested here speaks directly to the personalized medicine motivation stated at the outset of this paper. Personalized medicine depends on being able to quickly narrow down large numbers of candidate molecules to those most likely to suit a particular patient's metabolic profile, drug-target interaction, or tolerance for side effects, rather than relying on one-size-fits-all screening [23,24]. A lightweight tool such as BioMol Predictor, built on the same class of physicochemical descriptors evaluated in this study, could in principle be extended to flag candidates whose predicted absorption or toxicity profile is more or less suitable for a specific patient sub-group, before any resource-intensive pharmacogenomic or clinical work begins. At present this remains a direction for future development rather than a demonstrated capability of the tool: doing so credibly would require training on patient-stratified outcome data, not only the aggregate drug/non-drug labels used here, which is one reason we regard this study as an early step toward, rather than a finished application for personalized medicine.

5. Limitations

These findings must be interpreted in the light of the fact that they have some limitations. First, the size of dataset ($n = 71$) is small by currently accepted standards in cheminformatics (molecules commonly have thousand to hundreds of thousands of compounds in this type of benchmark collections [3]). Second, small sets of "small performance" cannot be relied upon as definitive, so reported performance should be interpreted as indicative only: if the dataset is expanded, the results might be different. Secondly, the values of descriptors used here are literature typical values, rather than new values computed from three dimensional structures using cheminformatics

computer software (such as RDKit), so are only approximate. Third, as noted, the task addressed here (oral drug-like vs. non-oral/biologic) is separate from, and more complex than, the multi-property predictions that BioMol Predictor makes concerning toxicity, bioactivity, synthetic accessibility and environmental remediation for particular drugs, which were not independently assessed in this study but would each warrant its own labeled dataset and validation. Fourth, this investigation was not network-accessible to query the web-based version of BioMol predictor deployed in the investigation directly, but rather the result of the predictor's accuracy was based on the accuracy of the predictions that BioMol reported itself produced in the study, which was a developer-stated figure of 85%+. The next critical step is closing this gap, the operation of the deployed tool against an appropriate, large benchmark labeled by other means, like MoleculeNet or Therapeutics Data Commons, and is discussed further below.

6. Conclusion

This study was undertaken by keeping in mind about BioMol Predictor, an easy, free and educational molecular drug-likeness classification tool developed by undergraduate chemistry students of Department of chemistry, Thal University Bhakkar with the intention of evaluation of feasibility of a descriptor-based machine learning approach for molecular drug-likeness classification. A Random Forest classifier applied this curated set of 71 compounds and six standard physicochemical descriptors yielded an accuracy of 90.1% (cross-validated) with an ROC-AUC of 0.937, significantly higher than that of a linear baseline, kernel baselines and the classical Lipinski Rule-of-Five filter. Two descriptors, namely the topological polar surface area (TPSA) and the molecular weight were selected as the most suitable descriptors. The results suggest empirical validation of the generic methodology implemented with the lightweight drug candidate screening tools based on descriptors and provide additional justification for the use of such tools as educational and pre-screening tools and not standalone experimental validation methods. Future work should focus

on: (i) direct, large-scale comparisons (benchmarking) of the deployed BioMol Predictor application with curated public datasets like MoleculeNet or the Therapeutics Data Commons; (ii) introduction of descriptor representations for 3D (or 3DMole) and graph descriptors; (iii) individual validation of each of the BioMol Predictor properties (toxicity, absorptivity, synthetic accessibility, bioactivity, environmental remediation) to their known experimental endpoint; (iv) reporting of per-module accuracy statistics as opposed to just an overall accuracy statistic, so that users of the application have an idea of its reliability when applied to specific tasks.

Author Contributions

Umar Farooq and Noman Hassan conceived the BioMol Predictor tool, designed the benchmark dataset and evaluation methodology, led model development, and wrote the manuscript. Areeba Mumtaz and Laraib contributed to compiling and verifying the physicochemical descriptor dataset. Khadija Khalil and Aima Gul contributed to the literature review and reference compilation. Shumaila Raheem and Nasira Mehmood contributed to reviewing and proofreading the manuscript. All authors read and approved the final manuscript.

Declarations

Funding: This research received no external funding.

Conflicts of interest: The authors declare no conflicts of interest.

REFERENCES

Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*. 2019;18(6):463-477. doi:10.1038/s41573-019-0024-5

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*. 1997;23(1-3):3-25.

Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V. MoleculeNet: a benchmark for molecular machine learning. *Chemical Science*. 2018;9(2):513-530. doi:10.1039/C7SC02664A

Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.

Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*. 2003;43(6):1947-1958.

An T, Chen Y, Chen Y, Ma L, Wang J, Zhao J. A machine learning-based approach to ER α bioactivity and drug ADMET prediction. *Frontiers in Genetics*. 2023;13:1087273. doi:10.3389/fgene.2022.1087273

Boldini D, Grisoni F, Kuhn D, Friedrich L, Sieber SA. Practical guidelines for the use of gradient boosting for molecular property prediction. *Journal of Cheminformatics*. 2023;15(1):73. doi:10.1186/s13321-023-00743-7

Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*. 2017;7:42717. doi:10.1038/srep42717

Rittig JG, Ben Hicham M, Schweidtmann AM, Dahmen M, Mitsos A. Graph neural networks for the prediction of molecular structure-property relationships. *arXiv:2208.04852*. 2022.

Wang Z, Liu M, Luo Y, Xu Z, Xie Y, Wang L, Cai L, Qi Q, Yuan Z, Yang T, Ji S. Advanced graph and sequence neural networks for molecular property prediction and drug discovery. *Bioinformatics*. 2022;38(9):2579-2586. doi:10.1093/bioinformatics/btac112

Berry K, Cheng L. A survey of graph neural networks for drug discovery: recent developments and challenges. *arXiv:2509.07887*. 2025.

- Deng J, Yang Z, Wang H, Ojima I, Samaras D, Wang F. A systematic study of key elements underlying molecular property prediction. *Nature Communications*. 2023;14:6395. doi:10.1038/s41467-023-41948-6
- Chandrasekaran B, Abed SN, Al-Attraqchi O, Kuche K, Tekade RK. Computer-aided prediction of pharmacokinetic (ADMET) properties. In: *Dosage Form Design Parameters*. Elsevier; 2018:731-755. doi:10.1016/B978-0-12-814421-3.00021-X
- Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*. 2022;12(7):3049-3062. doi:10.1016/j.apsb.2022.02.002
- Liu K, Sun X, Jia L, et al. Chemi-Net: a molecular graph convolutional network for accurate drug property prediction. *International Journal of Molecular Sciences*. 2019;20(14):3390.
- Jo J, Kwak B, Choi HS, et al. The message passing neural networks for chemical property prediction on SMILES. *Methods*. 2020;179:65-72.
- Durrant JD, McCammon JA. NNScore: a neural-network-based scoring function for the characterization of protein-ligand complexes. *Journal of Chemical Information and Modeling*. 2010;50(10):1865-1871.
- Rao J, Zheng S, Yang Y. Quantitative evaluation of explainable graph neural networks for molecular property prediction. *arXiv:2107.04119*. 2021.
- Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*. 2018;4(2):268-276.
- Lipinski CF, Maltarollo VG, Oliveira PR, da Silva ABF, Honorio KM. Advances and perspectives in applying deep learning for drug design and discovery. *Frontiers in Robotics and AI*. 2019;6:108.
- Sheridan RP. Using random forest to model the domain applicability of another random forest model. *Journal of Chemical Information and Modeling*. 2013;53:2837-2850.
- Duvenaud DK, Maclaurin D, Iparraguirre J, et al. Convolutional networks on graphs for learning molecular fingerprints. *Advances in Neural Information Processing Systems*. 2015:2224-2232.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 2019;25(1):44-56.
- Hamburg MA, Collins FS. The path to personalized medicine. *New England Journal of Medicine*. 2010;363(4):301-304.
- Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: toxicity prediction using deep learning. *Frontiers in Environmental Science*. 2016;3:80.
- Gaulton A, Hersey A, Nowotka M, et al. The ChEMBL database in 2017. *Nucleic Acids Research*. 2017;45(D1):D945-D954.
- Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research*. 2019;47(D1):D1102-D1109.
- Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*. 2018;46(D1):D1074-D1082.
- Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. *Journal of Combinatorial Chemistry*. 1999;1(1):55-68.
- Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*. 2002;45(12):2615-2623.