ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 15, Issue, 05, pp.4717-4720, May, 2024

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

DRUG DISCOVERY AND TOXICITY PREDICTION

Sahana R., Sampada Purushotham., Shreya M and Dr. Mahanthesha U

Department of Artificial Intelligence & Machine Learning
B N M Institute of Technology Bangalore, India

DOI: http://dx.doi.org/10.24327/ijrsr.20241505.0880

ARTICLE INFO

Article History:

Received 16th April, 2024
Received in revised form 30th April, 2024
Accepted 15th May, 2024
Published online 28th May, 2024

Keywords:

Drug Discovery, ANN, Deep Learning, Toxicity Prediction, Regression

ABSTRACT

Drug discovery, a pivotal aspect of pharmaceutical research, involves the identification and development of new therapeutic compounds. However, the process is hampered by challenges such as the labour- intensive nature of screening vast chemical libraries and the need to predict potential toxicity accurately. Common issues include limited labelled toxicity data, the complexity of molecular structures, and the time-consuming nature of experimental validation. This project aims to leverage deep learning for drug discovery and toxicity prediction, addressing these challenges by designing robust models capable of simultaneously predicting drug efficacy and toxicity. Through the analysis of diverse datasets, the project seeks to expedite the identification of promising drug candidates while ensuring safety. The approach involves training deep neural networks on comprehensive datasets and implementing multi-task learning strategies, to enhance the model's performance. In drug discovery and toxicity prediction, deep learning methods have proven to be powerful tools for extracting meaningful patterns and insights from complex biological data. Random Forest Regressor, Lazy Regressor and Decision Trees are commonly employed in these applications. By integrating computational intelligence with biological insights, this project strives to revolutionize pharmaceutical research and contribute to the development of safer and more effective medications.

INTRODUCTION

Deep learning is opening up new avenues for drug development and toxicity prediction in pharmaceutical research by utilizing artificial intelligence to quickly examine large chemical structure databases. The typically drawn-out process of developing new drugs is sped up by these sophisticated models, which are made to forecast both the safety profiles and the efficacy of possible therapeutic options. Deep learning provides a detailed comprehension of chemical interactions by utilizing a variety of representations, including molecular fingerprints, and multi-task learning guarantees thorough evaluations of drug toxicity and efficacy.

In addition to speeding up the process of finding new drugs, the combination of computational intelligence and biological insights holds great promise for producing safer and more effective medications. This highlights the profound influence of deep learning on the way that pharmaceutical research is conducted.

EXISTING MODELS

Many deep-learning-based approaches are emerging at every stage of the drug development process due to the evolution of deep learning (DL) technology and the expansion of drug-related data. Predicting how medications will interact with druggable targets and creating innovative molecular structures appropriate for a target of interest are the two main hurdles. As a result, we examined current deep learning uses in de novo drug design and drug–target interaction (DTI) prediction.

Fig. 1 Application of AI in Drug Discovery

As shown in Fig. 1, the use of artificial intelligence (AI) in drug development has completely changed several aspects of the pharmaceutical research environment. By evaluating biological data and forecasting druggability, AI systems are essential to the identification and validation of possible drug targets. AI-poweredvirtual screening finds potential medication candidates

Corresponding author:* **Sahana R
Department of Artificial Intelligence & Machine Learning, B N M Institute of Technology Bangalore, India

more quickly and optimises their chemical structures for both safety and efficacy.

The field of drug discovery and toxicity prediction, propelled by advancements in deep learning, showcases a dynamic landscape of innovation and research. Various studies delve into the intersection of artificial intelligence (AI) and biomedical sciences, offering insights into methodologies, applications, and challenges. Wodzisaw Duch et al. discuss the application of AI methods such as docking, screening, and QSAR studies in drug discovery, highlighting the broader role of machine learning in biomedical research.[1] Janaina Cruz Pereira et al. address the challenge of exploring chemical space in small molecule drug discovery and introduce SynDiR for optimizing CDK2 inhibitors.[2] Lu Zhang et al. provide a comprehensive overview of ANNs and their architectures, emphasizing their role in processing information for desired outcomes.[3] H.C Stephen Chan et al. underscore the multidisciplinary nature of drug design and the pivotal role of AI in accelerating knowledge acquisition in fields like genetics and pharmacology.[4] Mallikarjuna Rao et al. focus on deep learning- based approaches for drug-target interactions (DTIs) prediction, exploring various architectures and challenges in DTI prediction.[5] Hao Zhu et al. propose a deep learning-based method (deepACTION) for predicting potential DTIs, addressing challenges like high dimensionality and class imbalance.[6] Han Li et al. highlight the importance of accurate toxicity evaluation in drug development and discuss computational methods for drug toxicity prediction.[7] Ola Engkvist et al. explore recent advances in AI-based drug toxicity prediction, emphasizing the role of AI in reducing late-stage failures and improving efficiency.[8] Jintae et al. underscore the role of deep learning in accelerating drug discovery processes, particularly in drug-target interaction prediction and de novo drug design.[9] Tripti Shukla et al. discuss the integration of big data in drug discovery to enhance efficiency and reliability.[10] Lucas M Glass et al. introduce DeepPurpose for accurate drug-target interaction prediction and provide insights into deep learning approaches for DTI prediction.[11] Abbasi Karim et al. explore various architectures for feature extraction in drug and protein sequences.[12] Hosney Jahan et al. present a deep learning-based method for DTI prediction, addressing challenges like class imbalance and high dimensionality.[13] Zhang Li et al. introduce deepACTION for DTI prediction, achieving high performance in cross-validation tests on Drug Bank data.[14] Kexin Huang et al. highlight recent advances in AI-driven toxicity prediction in drug discovery.[15] Zhang Hui et al. discuss machine learning-based drug toxicity prediction models and their performance metrics.[16] Nathan Brown et al. explore computational methods like de novo design for optimizing drug compounds.[17] I.N. da Silva et al. provide an overview of artificial neural networks and their applications in solving complex problems.[18] Oscar Alvarez-Machancoses et al. discuss the multidisciplinary nature of drug and the effectiveness of Random Forest algorithm.[19] Ulrike Gromping et al. introduce Random Forest Regression as a powerful technique for solving regression problems in drug discovery.[20] Collectively, these studies highlight the diverse applications of deep learning in drug discovery and toxicity prediction, offering valuable insights into methodologies, challenges, and future directions in the field.

METHODOLOGY

The project involves working with datasets in CSV and ChEMBL formats using Python and the Pandas, Seaborn, NumPy, RDKit, and Scikit-Learn libraries. The pipeline consists of transformation, feature engineering, and data cleansing. During the training, testing, and validation phases, ML techniques such as random forest, lazy regressor, and decision tree classifier are used to help with drug discovery, toxicity assessment, and replacement finding tasks. An online application that meets the needs of pharmaceutical research is the end result.

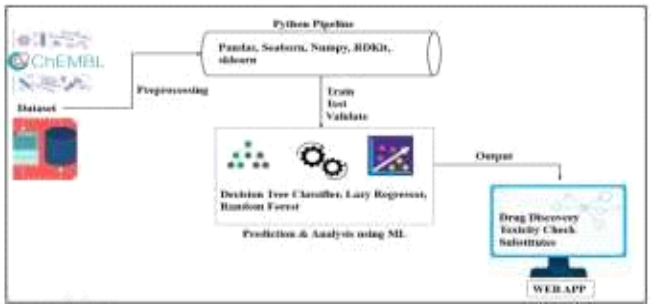

Fig. 2 Architectural design in Drug Discovery & ToxicityPrediction

The dataflow for drug discovery and toxicity prediction using deep learning typically follows a structured process encompassing data preparation, model development, and validation.

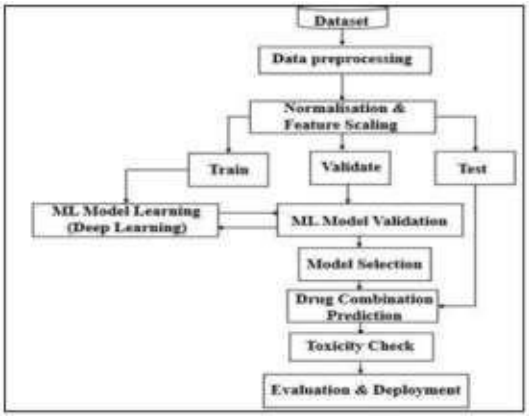

Fig. 3 Dataflow Diagram in Drug Discovery & ToxicityPrediction

As illustrated in Fig.3, this is a concise overview of the methodology of implementation: Data Collection and Preprocessing, Model Architecture Selection, Data Splitting, Model Training, Hyperparameter Tuning, Model Evaluation, Interpretability and Visualization, Validation on Test Set, Integration with Experimental Validation.

The algorithmic approach provides a methodical framework for developing machine learning-based predictive models that are specific to acetylcholinesterase inhibitors. The goal variable (pIC50 values) and molecular descriptors are identified as characteristics for model construction in the first definition of the issue statement and objectives. In order to handle missing values, encode categorical variables, and normalise numerical features, data is gathered from the ChEMBL database and preprocessed. RDKit and PaDEL-Descriptor are used to compute molecular descriptors. To comprehend data distribution and patterns, exploratory data analysis, or EDA, is carried out. Building a model entails dividing the data, creating

a Random Forest Regression model, and utilising Lazy Regressor to compare the results. Metrics and statistical studies like the Mann-Whitney U Test are evaluated as part of the model evaluation process. Lastly, a production environment with maintenance monitoring systems is used to implement the trained model, while documentation ensures reproducibility and future reference.

The dataset has been taken from Kaggle and ChemBL websites.Each dataset serves a specific purpose in different parts of the project, such as data preprocessing, feature engineering, model building, and evaluation.

Module Description: Pandas, numpy, train_test_split: This Pythoncode uses the scikit-learn and pandas libraries to prepare data for machine learning. It extracts features (X) that affect the result andtarget labels (Y) for prediction from data it reads from a CSV file. To make sure the model generalises well beyond training data, train_test_split divides the data into training and testing sets. RDKit: This Python code calculates molecular weight using the rdkit. Chem library. It defines a molecule's structure from a SMILES string, converts it into a molecule object, and computes molecular weight using rdkit. Chem.

sklearn (Random Forest Regressor): The Random Forest Regressor model for prediction tasks is constructed using this code. The number of trees in the model is specified at the time of creation. Itis then trained on known features and outcomes to predict outcomes for unseen data. By integrating numerous decision trees,the model's predictions are improved.

Seaborn, Matplotlib: This code uses Seaborn to visualise model performance, which improves upon matplotlib for statistical visualisations. To help evaluate prediction accuracy, it generates a scatter plot that contrasts actual values (Y_test) with model predictions (Y_pred).

Lazy Predict, Lazy Regressor: This method uses Lazy Regressor to assess model performance on training data, summarising measures like accuracy and precision. However, in order to prevent over fitting, genuine performance validation needs to be done on unseen data. wget: This code imports wget, defines a file's URL, and uses wget. Download to download it locally. It is helpful for getting online resources such as datasets. It uses the wget module to download files from the web.chembl_ webr esource_client: This code uses chembl _webresource_client to acquire bioactivity information for a target molecule, most likely a protein like "Acetyl cholinesterase". In order to facilitate effective analysis, it searches, extracts data (especially records of the standard type "IC50"), and transforms them into a pandas Data Frame.

RESULTS AND DISCUSSION

The research intends to thoroughly assess machine learning models' efficacy in forecasting medication toxicity by utilising cutting-edge testing techniques, enabling well-informed decision- making. For dataset partitioning, it makes use of train_test_split, guaranteeing a reliable evaluation of model generalisation. Incorporating k-fold cross-validation would improve model robustness, even though it isn't demonstrated explicitly. Techniques for hyper parameter adjustment, including random or grid search, could improve model performance even further. Beyond Random Forest Regression, ensemble techniques like Gradient Boosting might be investigated to increase prediction accuracy. Furthermore, combining R-squared with assessment metrics like RMSE, MAE, or MAPE will yield a thorough evaluation of model accuracy and error. These methods improve the project's usefulness and efficacy in drug toxicity prediction by being in line with its goals.

The project's model performance demonstrates efficiency and resilience in both training and test sets. Several machine learning methods were used on the training set, with encouraging outcomes. The models' capacity to explain the variation in the data is demonstrated by the R-squared values, which show a strong correlation between anticipated and actual values. Low RMSE values also indicate precise forecasts with little errors. High R-squared scores, low RMSE values, and reasonable calculation times are displayed in the bar plots, which graphically represent these metrics and highlight how well the machine learning models predict drug activity. All things considered, the models show outstanding predictive performance, which makes them useful resources for toxicity evaluation and medication discovery.

CONCLUSION

Pharmaceutical research is revolutionised by the incorporation ofmachine learning into drug discovery and toxicity prediction, which provides sophisticated tools for molecular dataset analysis.Drug names are routinely predicted using decision tree classifiers and other algorithms based on medical conditions and toxicity levels. This improves identification speed and precision in matching pharmaceuticals to specific ailments, improving patient outcomes. The code built ensures that inputs are easy to use and that errors are handled robustly, which encourages broader adoption and effective decision making in healthcare contexts. Multi-dimensional studies that provide nuanced insights into drugefficacy and safety for well-informed treatment decisions are made possible by utilising varied datasets. Personalised medicine is made possible by machine learning, which speeds upmedication discovery by customising therapies based on the unique characteristics of each patient. This confluence points to a new and exciting direction in healthcare, providing physicians and researchers with the tools they need to navigate intricate datasets and improve patient outcomes through safer, more individualised,and more effective therapeutic approaches.

Future work in the context of drug discovery and toxicity prediction using machine learning models involves several key areas of focus and improvement: Enhanced Model Performance, Integration of Additional Data Sources, Explainable AI (XAI), Scalability and Deployment, Collaborative Research. The field of drug discovery can continue to advance by addressing these areas in subsequent research, providing revolutionary solutions for personalised medicine, better patient outcomes, and breakthroughs in pharmaceutical research and healthcare delivery.

ACKNOWLEDGMENT

We are grateful to the faculty members of Department of Artificial Intelligence and Machine, BNMIT for their insightful feedback and constant encouragement.

References

1. Duch, Wlodzislaw, Karthikeyan Swaminathan, and Jaroslaw Meller, "Artificial intelligence approaches for rational drug design and discovery", *Current Pharmaceutical Design*, 13(14), 1497-1508 (2007). DOI: 10.2174/138161207780765954

2. Janaina Cruz Pereira, Ernesto Raúl Caffarena, and Cicero Nogueira dos SantosBoosting, "Docking-Based Virtual Screening with Deep Learning", *Journal of Chemical Information and Modeling*, 56(12), 2495-2506 (2016). DOI: 10.1021/acs.jcim.6b00355
3. Lu Zhang, Jianjun Tan, Dan Han, Hao Zhu, "From machine learning to deep learning: progress in machine intelligence for rational drug discovery", *Drug Discovery Today*, 22(11), (2017). DOI: 10.1016/j.drudis.2017.08.010
4. Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S, "Advancing Drug Discovery via Artificial Intelligence", *Trends Pharmacol Sci*, 40(8), 592-604 (2019). DOI: 10.1016/j.tips.2019.06.004
5. Kit-Kay Mak, Mallikarjuna Rao Pichika, "Artificial intelligence in drug development: present status and future prospects", *Drug Discovery Today*, 24(3), 773-780 (2019). DOI: 10.1016/j.drudis.2018.11.014
6. Zhu H, "Big Data and Artificial Intelligence Modeling for Drug Discovery", *Annu Rev Pharmacol Toxicol*, 60, 573-589 (2020). DOI: 10.1146/annurev-pharmtox-010919-023324
7. Li H, Zhang R, Min Y, Ma D, Zhao D, Zeng J, "A knowledge-guided pre-training framework for improving molecular representation learning", *Nat Commun*, 14(1), 7568 (2023). DOI: 10.1038/s41467-023 43214-1
8. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T, "The rise of deep learning in drug discovery", *Drug Discov Today*, 23(6), 1241-1250 (2018). DOI: 10.1016/j.drudis.2018.01.039
9. Kim J, Park S, Min D, Kim W. Comprehensive Survey of Recent Drug Discovery Using Deep Learning. *Int J Mol Sci*. 22(18), 9983 (2021). DOI: 10.3390/ijms22189983
10. Patel L, Shukla T, Huang X, Ussery DW, Wang S, "Machine Learning Methods in Drug Discovery", *Molecules*, 25(22), 5277 (2020). DOI: 10.3390/molecules25225277
11. Huang K,Fu T,Glass LM,Zitnik M,Xiao C, Sun J, "DeepPurpose: a deep learning library for drug-target interaction prediction", *Bioinformatics*, 36(22-23), 5545-5547 (2021). DOI: 10.1093/bioinformatics/btaa1005
12. Abbasi K, Razzaghi P, Masoudi-Nejad A, "Deep Learning in Drug Target Interaction Prediction: Current and Future Perspectives", *Curr Med Chem*, 28(11), 2100-2113 (2021). DOI: 10.2174/0929867327666200907141016
13. Hasan Mahmud, Dai B, Din SU, Dzisoo AM, "DeepACTION: A deep learning-based method for predicting novel drug-target interactions", *Anal Biochem*, 610, 113978 (2020). DOI: 10.1016/j.ab.2020.113978
14. Zhang L, Zhang H, Ai H, Hu H, Li S, Zhao J, Liu H. Applications of Machine Learning Methods in Drug Toxicity Prediction. *Curr Top Med Chem*. 18(12), 987-997 (2018). DOI: 10.2174/1568026618666180727152557
15. Cavasotto CN, Scardino V. Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point. *ACS Omega*. 7(51), 47536-47546 (2022). DOI: 10.1021/acsomega.2c05693
16. King Scarbrown, RossHirst G, Jonathan Michael, Razzaghi P, "Comparison of artificial intelligence methods for modelling pharmaceutical QSARS", *Applied Artificial Intelligence: An International Journal*, 233, STERNBERG, MICHAEL (2007/04/27), 213, VL- 9. DOI: 10.1080/08839519508945474
17. Firth NC, Atrash B, Brown N, Blagg J, "MOARF, an Integrated Workflow for Multiobjective Optimization: Implementation, Synthesis, and Biological Evaluation", *J Chem Inf Model*, 55(6), 1169-1180 (2015). DOI: 10.1021/acs.jcim.5b00073
18. Silva, Ivan, Spatti, Danilo, Flauzino, R.A., Bartocci Liboni, Luisa, Reis Alves, Silas, "Artificial Neural Network Architectures and Training Processes", (2017/01/01), 21, 28. DOI: 10.1007/978-3-319-43162-8_2
19. Álvarez-Machancoses Ó, Fernández-Martínez JL. Using artificial intelligence methods to speed up drug discovery. *Expert Opin Drug Discov*. 14(8), 769-777 (2019). DOI: 10.1080/17460441.2019
20. Grömping, Ulrike , "Variable Importance Assessment in Regression: Linear Regression versus Random Forest", *The American Statistician*, 308,319, (2009/11/01), 2009. DOI: 10.1198/tast.2009.0819

How to cite this article:
Sahana R., Sampada Purushotham., Shreya M and Mahanthesha U. (2024). Drug discovery and toxicity prediction. *Int J Recent Sci Res.* 15(05), pp.4717-4720.

****