International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

PREDICTING DRUG-DRUG INTERACTION USING MACHINE LEARNING

Prof. M. V. Raut[1], Shreya Deshmukh[2], Sachin Deshpande[3], Kalash Changal[4], Renuka Modak[5]

Information Technology/ Smt. Kashibai Navale College of Engineering/ Savitribai Phule Pune University/ Pune

---

Abstract -The ever-expanding landscape of pharmaceutical compounds and medical treatments has underscored the critical need for an advanced system to predict drug-drug interactions (DDIs) using machine learning. This paper delves into the development of a novel predictive model that leverages diverse data sources, including chemical structures, pharmacological properties, and scientific literature, to comprehensively analyse and foresee potential DDIs. The current state of DDIs is rife with limitations in scope and accuracy, making the development of an advanced system not only a necessity for enhancing patient safety but also a boon to pharmaceutical research and regulatory oversight. The relationship between two medications in which one drug's pharmacological actions are altered by another is known as a drug-drug interaction (DDI). Favourable drug-drug interactions (DDIs) generally contribute to improved therapeutic results for patients. On the contrary, adverse DDIs pose a significant risk, often resulting in undesirable drug reactions that can potentially lead to the withdrawal of a drug from the market and even patient fatalities. Consequently, the identification of DDIs is essential for advancing new drug development and effectively managing diseases.

Keywords: Drug, Drug-drug interaction, Machine learning, Bioinformatics, Drug development.

1.INTRODUCTION

A drug interaction occurs when two or more drugs, or a drug and a substance like food, drink, or a supplement, interact with each other. In the case of drug-drug interactions (DDIs), the combined use of multiple drugs can alter their individual effects or potentially result in undesirable side effects. It's important to be aware that taking certain drugs together may lead to unexpected adverse events, emphasizing the need for caution and awareness of potential interactions when using multiple medications simultaneously. In this project we aim to develop and implement an effective and reliable model to predict the drug-drug interactions aiding in reduction of computational time and cost.

The increasing presence of chronic illness in our generation has led to a rise in polypharmacy, the simultaneous use of multiple medications. While this approach offers tailored treatment options for complex conditions and rare genetic conditions, it also introduces the increased risk of inter-drug interactions (DDIs). These interactions occurs when medications influence each other's absorption, metabolism, excretion, distribution, potentially leading to adverse side effect.

Current methods for DDI detection primarily work with manual review of medication lists and reference materials by healthcare professionals. This approach can be time-consuming, prone to human error, and may not capture all potential interactions, particularly for complex medication regimens. To address these limitations, there is a growing need for robust and efficient tools to predict DDIs. This survey paper explores the potential of machine learning (ML) algorithms for predicting DDIs. To expand further we will examine how ML techniques can analyze vast datasets of drug information, including their mechanisms of action, metabolism pathways, and known interaction types. By identifying patterns within this data, ML algorithms can learn to predict potential interactions between provided drugs.

Furthermore, this paper will give views on the concept of symptom-aware DDI prediction. By providing information on the potential consequences of specific interactions, the algorithm can not only predict an interaction but also outline the associated symptoms. This information is very useful for healthcare professionals as well as patients. Clinical professionals can work with this algorithm as a decision-support tool, enhancing medication safety while suggesting medications to patients. Patients can gain valuable knowledge about potential interactions, enabling them to report concerning symptoms and participate actively in their healthcare.

This survey aims to provide a comprehensive overview of the current state-of-the-art in DDI prediction using ML. We will discuss the machine learning approaches used for this project. Ultimately, the goal is to promote the development of safe, reliable and user-friendly DDI prediction tools that can significantly improve medication safety and patient outcomes.

2. Body of Paper

2.1 RELATED WORKS

A thorough summary of the several machine learning approaches used to predict unknown drug-drug interactions (DDIs) can be found in the publication [1]. Because DDIs can result in serious adverse effects, manual detection is an expensive and time-consuming operation. This is why it is important to predict DDIs. The machine learning methodologies are divided into multiple categories by the authors: ensemble-based, similarity-based, classification-based, and network propagation techniques. Every category approaches the prediction problem differently, making use of a variety of data sources and computational methods to increase the precision and effectiveness of DDI predictions. While classification-based methods frame the prediction

challenge as a binary classification problem using known interaction pairs to train models, similarity-based approaches use the intrinsic properties of medications to infer possible interactions. Drug interaction networks have topological characteristics that can be used by network propagation techniques like graph embedding and link prediction to uncover previously unidentified interactions. Ensemble-based techniques, which frequently include genetic algorithms and multi-label learning, combine several strategies to improve predictive accuracy. Furthermore, literature-based methods gather interaction data from unstructured text sources by utilising natural language processing. The study outlines the benefits and drawbacks of each approach and stresses the need for greater development in order to successfully integrate various data sources and forecast DDIs.

A thorough overview of machine learning techniques for drug-drug interaction (DDI) prediction can be found in the publication [2]. Adverse reactions to DDIs can pose serious health hazards to patients. The accuracy and data sufficiency of traditional computational techniques for DDI prediction are limited, which makes machine learning approaches more and more useful. Support vector machines (SVMs), deep neural networks, and ensemble classifiers are just a few of the machine learning and deep learning methods that are highlighted in the study. These algorithms have demonstrated promise in improving prediction accuracy by utilising big datasets and intricate data representations. A number of case studies are examined to demonstrate how successful these methods are. SVMs have been used, for example, to categorise relevant data from big datasets, increasing drug development efficiency. By combining structural and genetic data, deep learning models—such as those that use auto-encoders and feed-forward neural networks—have achieved great accuracy in DDI prediction. These techniques have drawbacks despite their benefits, such as the need for huge datasets and high processing demands. The research highlights the necessity for ongoing studies to successfully manage the dynamic and complicated nature of drug interactions. The paper adds that combining several machine learning algorithms could further improve DDI prediction.

The paper [3] by Brigitte Grau, Lamia Makour, Anne-Laure Ligozat, Anne-Lyse Minard, and others examines machine learning-based approaches for drug-drug interactions (DDIs). The writers talk about how they took part in the DDI Extraction challenge, which looks for drug interactions when they are mentioned in the same sentence. They developed a technique to categorise drug interactions using support vector machines (SVM), more especially LIBSVM and SVMPerf tools. The F-score served as the basis for the feature selection process, which made use of both classical and corpus-specific features. With an F-measure of 0.5965, their top model showed how effective their feature selection and machine learning strategy were. Two corpora from the DrugBank database that were annotated with drugs and drug-drug interactions were used in the study. While the assessment corpus had 144 files and 7,026 candidate pairings with 755 interactions, the development corpus included 435 files with 23,827 potential drug pairs, including 2,402 interactions. To build the model, the development corpus was divided into training and test sub-corpora. The paper compares the efficacy of LIBSVM and SVMPerf, especially for handling the imbalanced class distribution in DDI detection, and emphasises the significance of feature selection in enhancing the classification performance of SVMs.

The paper [4] by Idris Demirsoy and Adnan Karaibrahimoglu addresses the critical issue of adverse drug reactions (ADRs), which result in nearly 100,000 deaths annually in the USA. In order to reduce these adverse drug reactions, the study uses machine learning to forecast drug-drug interactions (DDIs). The scientists created eight similarity matrices including information from sources like DrugBank, BioGRID, and the Comparative Toxicogenomics Database in order to evaluate their impact on DDIs. The study examined the relationships between 22 well-known medications and 841 others using three machine learning algorithms: neural networks, eXtreme Gradient Boosting (XGBoost), and logistic regression. The models' accuracy rates ranged from 68% to 78%, while their F1 scores were between 78% and 83%. The study demonstrated the potential of machine learning to provide precise, fast, and economical predictions by determining that enzyme and target similarity were the most important criteria in detecting DDIs. The process comprised merging data from several sources and applying a range of similarity criteria to produce extensive feature matrices. The study found that sophisticated machine learning approaches could improve the predicted accuracy by examining the significance of these parameters in DDI prediction. Similarities between the enzyme and the target were highlighted by the researchers, who recommended that they be important factors for DDI study in the future. The findings highlight the potential for machine learning techniques to enhance DDI forecasts, which would eventually lead to increased patient safety and more successful medication delivery plans. The results show a promising path for the use of machine learning in pharmacology and provide insightful information for researchers and healthcare providers.

A thorough assessment of the present state of machine learning methodologies and databases for drug-target interaction (DTI) prediction is given in the publication [5]. The authors stress the importance of DTI prediction in the drug discovery process, with the goal of increasing effectiveness and lowering the expense of experimental techniques. They list the kinds of data required for DTI prediction, including as information about drugs and targets, and they explain the many machine learning techniques used in this field. These strategies cover everything from

conventional techniques like random forests and support vector machines to sophisticated methods like deep learning and ensemble approaches. The paper also discusses the benefits and limits of a comprehensive catalogue of datasets that are currently accessible to help DTI prediction. The study also discusses the difficulties in predicting DTIs by machine learning, including the lack of data, unbalanced datasets, and the requirement for better model interpretability. It demonstrates the benefits and drawbacks of various machine learning techniques and provides information on how well they work in diverse contexts. In their concluding remarks, the authors provide several recommendations for future study, such as integrating multi-omics data, creating more comprehensible models, and standardising benchmarks for technique evaluation. Researchers and practitioners in the field will benefit greatly from this survey, which offers a comprehensive overview of the tools and approaches available for DTI prediction and their potential to expedite drug discovery procedures.

The HAINI framework, which predicts drug-drug interaction (DDI) types for histamine antagonist drugs using simplified molecular-input line-entry systems (SMILES) combined with interaction features based on the CYP450 group, is presented in the paper [6]. The DrugBank database's 26,344 DDI pairings of authorised histamine antagonist medications were used in the investigation. Five-fold cross-validation was used with a variety of machine learning algorithms, such as XGBoost, Naive Bayes, Decision Tree, Random Forest, and Logistic Regression. With a precision of 0.788, recall of 0.921, and F1-score of 0.838 among 19 DDI types, the HAINI model outperformed the other studies in a way that was clearly noticeable. This model's exclusive dependence on SMILES and CYP450 inputs, which permits early-stage application in drug development, is a crucial feature. SMILES and drug metabolism profiles were gathered from the DrugBank database as part of the technique, and the data was split into authorised and non-approved medications for training and validation, respectively. The study created new features based on CYP450 interactions and presented a novel hybrid feature extraction method utilising the PyBioMed package. Because of the models' excellent prediction accuracy, machine learning may be able to significantly cut down on the time and expense of DDI examinations in clinical settings. The study emphasises how important it is to integrate different kinds of data, such as genetic and chemical information, in order to improve prediction accuracy and lessen the financial strain on healthcare systems.

The advancements in machine learning (ML) models for predicting drug-drug interactions (DDIs) are discussed in the paper [7]. DDIs are a critical issue due to polypharmacy, especially among the elderly and patients with chronic conditions like those observed during the COVID-19 pandemic. In reviewing models created since 2018, the paper emphasises how crucial it is to incorporate a variety of datasets and advanced feature extraction techniques. These methods greatly improve prediction accuracy and offer more in-depth insights into possible DDIs. They include molecular structure analysis and the incorporation of biological information such as drug target proteins, DDI networks, and knowledge graphs. The review classifies DDI prediction models according to the sources of their input data, the machine learning methods used to extract features, and the kinds of outcomes they anticipate (e.g., multilabel classification for different kinds of interactions and binary classification for the presence of DDIs). DDIMDL, which integrates numerous biological aspects, and DeepDDI, which leverages structural similarity profiles for multilabel classification, are two of the key models presented. The study highlights how knowledge graph embeddings and graph neural networks (GNNs) can better predict DDIs by capturing intricate linkages in biological data. The authors also suggest future directions for study, urging the continuous creation of models that make use of large datasets and cutting-edge machine learning methods to overcome the difficulties associated with DDI prediction.

The integration of computational techniques to anticipate and control drug-drug interactions (DDIs) is explored in the study [8] by Safdari et al. The authors stress that adverse drug reactions (ADRs), which pose serious concerns to patient safety and raise healthcare expenditures, are largely caused by DDIs. In order to forecast possible DDIs, the study examines a variety of computational techniques, such as artificial neural networks (ANN), machine learning algorithms, and pharmacokinetic and pharmacodynamic modelling. These techniques make use of biological information, such as signalling pathways and drug target proteins, to find hidden interactions and advance our knowledge of DDI mechanisms. The study emphasises how computerised methods have the potential to revolutionise personalised medicine and medication discovery. Researchers can predict and prevent dangerous drug-drug interactions (DDIs), improving therapeutic efficacy and patient safety, by utilising data mining, biological simulations, and machine learning. In order to create thorough DDI prediction models, the authors also stress the significance of combining data from other sources, such as clinical, molecular, and pharmacogenomic information. With the ultimate goal of optimising therapeutic outcomes and lessening the strain on healthcare systems, they promote the implementation of computerised physician order entry (CPOE) systems and public education platforms to decrease the hazards associated with numerous pharmacological regimens.

2.2 PROPOSED SYSTEM

Our project holds immense importance in the field of drug-drug interaction (DDI) prediction, addressing a critical need for more accurate and efficient methods. With the rapid expansion of pharmaceutical compounds and increasingly complex treatment regimens, the risk of unforeseen interactions between drugs has become a pressing concern in healthcare. Current DDI prediction methods often fall short in providing comprehensive and precise solutions. By conducting a thorough study of various methodologies and their respective accuracies, our project aims to identify and implement the most efficient algorithm. This initiative plays a

crucial role in improving patient safety by providing healthcare professionals with a dependable resource to make well-informed decisions regarding prescriptions, thereby reducing the risk of negative consequences. Additionally, it has the capacity to streamline the drug development timeline, allowing the pharmaceutical sector to detect and manage potential drug-drug interactions at an earlier stage in the research process. Our project, in essence, is a pivotal step towards revolutionizing healthcare and pharmaceutical research, addressing a multifaceted challenge that continues to grow in significance, and ultimately advancing patient well-being and drug safety.

In the initial phase of project we are going to use some rudimentary machine learning techniques like supervised learning which would further use the implementation of various algorithms which include Random forest, logistic regression, Naive bayes and KNN etc.

We intend to integrate proven methods into our approach, incorporating well-established techniques alongside innovative semi-supervised learning methods. This includes the utilization of semi-supervised ensemble learning and semi-supervised support vector machines. The reason for using this in the initial phase is to determine their individual accuracy and compatibility with the dataset we are going to use.

DrugBank is the main dataset we are going to be using in this project. It is available on Kaggle at https://www.kaggle.com/datasets/sergeguillemart/drugbank/data. This dataset contains information about various drugs, including their structures, reactivity to certain compounds, various chemical attributes including pH value etc. It also contains some information about some commonly known Drug-Drug interaction examples including Fluconazole and simvastatin etc. This dataset has been selected due to its robustness, availability and is relatively concise.

One of the major focuses of our project is to test multiple algorithms and find the best suited, fastest and most efficient one from the lot. We are eyeing for better optimization of currently used techniques. This may include the usage of different algorithms for different tasks which are suitable for that 1 task and combining them which would increase the overall accuracy of our system.

After incorporating these fundamental algorithms, we plan to integrate more advanced approaches, such as deep learning, to assess its precision when combined with the previously mentioned techniques. This will allow us to harness its complete potential for achieving exceptionally accurate results.

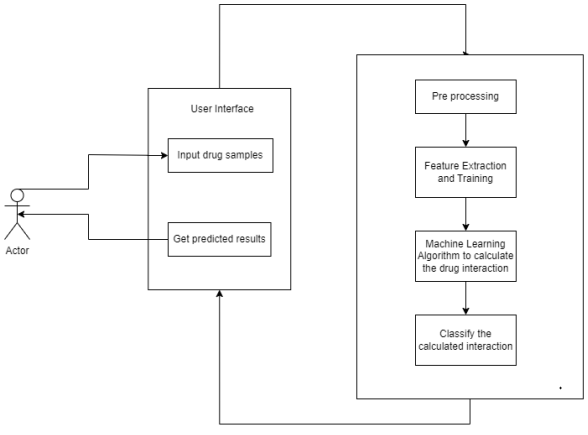

Fig -2.1: System Architecture

The drug-drug interaction prediction system consists of an user-friendly interface and a trained ML model for the prediction of DIs. The user interface can be used by user to login to the system and give inputs to the model. The input can be in the form of drug samples (names, structures). The output of the system will be presented to the user using the interface. The model will be able to do tasks like taking inputs from user, preprocessing the input data, feature extraction and training and applying algorithms to calculate the drug interaction between the input drugs. The output of the model will be passed on to the user interface to be available to the user.

3. CONCLUSIONS

In order to increase the effectiveness of treatment and lessen the burden on patients, multidrug therapies have been frequently employed to treat illnesses, particularly complex disorders like cancer. Multi-drug therapies do, however, sometimes have unfavorable side effects that might result in serious problems or even patient death. Consequently, finding drug-drug interactions aids in bettering the treatment of diseases and lessens the complexity of developing new drugs. In particular, it is imperative to create new computational techniques for DDI identification. These algorithms evince promising results in identifying potential interactions between medications. Furthermore, the concept of symptom-aware DDI prediction holds significant value. By providing information on potential consequences and associated symptoms, such algorithms can enhance medication safety and patient awareness. However, challenges remain in the development and implementation of ML-based DDI prediction tools. The accuracy and generalizability of these algorithms depend heavily on the quality and comprehensiveness of training data. Additionally, continuous validation and monitoring are crucial to ensure the tool's effectiveness in real-world clinical settings. Despite these challenges, the potential benefits of ML-based DDI prediction are clear. As research progresses and these algorithms are further refined, they have the potential to revolutionize medication safety practices. By empowering clinicians with accurate and timely DDI information and fostering patient awareness of potential interactions, these tools can significantly improve patient outcomes and contribute to a safer and more informed healthcare landscape.

ACKNOWLEDGEMENT

We are very thankful to all the teachers who have provided us valuable guidance towards the completion of this project work on Predicting Drug-Drug Interaction during Machine Learning. We express our sincere gratitude towards cooperative department who has provided us with valuable assistance and requirements for the project work.

We are very grateful and want to express our thanks to Prof. M. V. Raut for guiding us in right manner, correcting our doubts by giving her time whenever we required, and providing her knowledge and experience in making this

project work. We are also thankful to the HOD of our Information Technology department Dr. M. L. Bangare for his moral support and motivation which has encouraged us in making this project work. We are also thankful to our Vice Principal Dr. K. R. Borole and our Principal Dr. A.V. Deshpande, who provided his constant support and motivation that made a significant contribution to the success of this project.

REFERENCES

1. Han K, Cao P, Wang Y, Xie F, Ma J, Yu M, Wang J, Xu Y, Zhang Y and Wan J (2022) A Review of Approaches for Predicting Drug–Drug Interactions Based on Machine Learning. Front. Pharmacol. 12:814858. doi: 10.3389/fphar.2021.814858

2. A.Saran Kumar, Dr. R.Rekha. A Critical Review On Predicting Drug-Drug Reactions Using Machine Learning Techniques. Int J Comput Neural Eng. 2021;7(1):119-121.

3. Anne-Lyse Minard, Lamia Makour, Anne-Laure Ligozat, Brigitte Grau. Feature selection for drug drug interaction detection using machine-learning based approaches. Challenge Task on Drug-Drug Interaction Extraction (DDI), SEPLN, Sep 2011, Huelva, Spain. pp.43–50. hal-02289969

4. Demirsoy I, Karaibrahimoglu A. Identifying drug interactions using machine learning. Adv Clin Exp Med. 2023;32(8):829–838. doi:10.17219/acem/169852

5. Maryam Bagherian, Elyas Sabeti, Kai Wang, Maureen A. Sartor,Zaneta Nikolovska-Coleska and Kayvan Najarian. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. Briefings in Bioinformatics, 22(1), 2021, 247–269

6. Dang, L.H.; Dung, N.T.; Quang, L.X.; Hung, L.Q.; Le, N.H.; Le, N.T.N.; Diem, N.T.; Nga, N.T.T.; Hung, S.-H.; Le, N.Q.K. Machine Learning-Based Prediction of Drug-Drug Interactions for Histamine Antagonist Using Hybrid Chemical Features. Cells 2021, 10, 3092. https://doi.org/10.3390/ cells10113092

7. Eujin Hong, Junhyeok Jeon, and Hyun Uk Kim. Recent development of machine learning models for the prediction of drug-drug interactions. Korean J. Chem. Eng., 40(2), 276-285 (2023) DOI: 10.1007/s11814-023-1377-3

8. Reza Safdari, Reza Ferdousi, Kamal Aziziheris, Sharareh R. Niakan-Kalhori, Yadollah Omidi. Computerized techniques pave the way for drug-drug interaction prediction and interpretation. Safdari R., et al., BioImpacts**,** 2016, 6(2), 71-78 doi: 10.15171/bi.2016.10

9. Cheng F, Zhao Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties J Am Med Inform Assoc 2014;21:e278–e286.

10. Sun Kim, Haibin Liu, Lana Yeganova , W. John Wilbur. Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. National Center for Biotechnology Information (NCBI), Bethesda, MD, USA.

11. Cheng Yan, Guihua Duan, Yayan Zhang, Fang-Xiang Wu, Yi Pan and Jianxin Wang. Predicting drug-drug interactions based on integrated similarity and semi supervised learning. DOI 10.1109/TCBB.2020.2988018, IEEE/ACM Transactions on Computational Biology and Bioinformatics.

12. Wen Zhang, Yanlin Chen, Dingfang Li, Xiang Yue. Manifold regularized matrix factorization for drug-drug interaction prediction. Journal of Biomedical Informatics 88 (2018) 90–97. https://doi.org/10.1016/j.jbi.2018.11.005