RESEARCH ARTICLE OPEN ACCESS

Artificial Intelligence Approach to Detect the Drug-to-Drug Interaction

[1] M.Arunkumar, [2] Dr.T.S.Baskaran,

Research Scholar, Research & PG Department of Computer Science,
A.V.V.M.Sri.Pushpam College,Poondi,Thanjavur

Associate professor, Research & PG Department of Computer Science,
A.V.V.M.Sri.Pushpam College,Poondi,Thanjavur
"Affiliated to Bharathidasan University", Tirchuirappalli

ABSTRACT

Drug–drug interactions play a vital rolnme in drug research. However, they may also cause adverse reactions in patients, with serious consequences. Manual detection of drug–drug interactions is time-consuming and expensive, so it is urgent to use computer methods to solve the problem. There are two ways for computers to identify drug interactions: one is to identify known drug interactions, and the other is to predict unknown drug interactions. In this paper, we review the research progress of machine learning in predicting unknown drug interactions. Among these methods, the literature-based method is special because it combines the extraction method of DDI and the prediction method of DDI. We first introduce the common databases, then briefly describe each method, and summarize the advantages and disadvantages of some prediction models. Finally, we discuss the challenges and prospects of machine learning methods in predicting drug interactions. This review aims to provide useful guidance for interested researchers to further promote bioinformatics algorithms to predict DDI.

Keywords: Drugs, DDI, Machine Learning Approach, Deep Learning Approach,Adverse Side effect, similarity, Predication

I. INTRODUCTION

Artificial intelligence (AI) has made important strides in recent years, impacting various industries, including healthcare. One of the most promising application of AI in healthcare is its potential to transfigure the field of pharmacology. Specifically, AI-powered druginteraction prediction is emerging as a new frontier in this domain, offering a more efficient and exact way to recognize potential drug interactions and optimize patient care.

Drug interactions are a considerable concern in contemporary medicine, as they can lead to adverse effects, reduced efficacy of medications, and even life-threatening complications. With the increasing number of medications available and the growing complication of treatment regimens, healthcare professionals face the daunting task of identifying and managing potential drug interactions. Conventional methods for predicting drug interactions rely on manual analysis of scientific literature, clinical trial data, and post-marketing surveillance reports. This process is time-consuming, labor-intensive, and prone to human error.

AI-powered drug interaction prediction has the potential to overcome these limitations by leveraging machine learning algorithms and vast amounts of data to identify potential drug interactions more accurately and efficiently. By analyzing large datasets, including electronic health records, genomic data, and chemical compound databases, AI can uncover patterns and relationships that may not be apparent to human researchers. This can lead to the identification of previously unknown drug interactions, as well as a better understanding of the underlying mechanisms behind known interactions.

II. SIGNIFICANCE

One of the key advantages of AI-powered drug interaction prediction is its ability to process and analyze vast amounts of data quickly. This allows for real-time monitoring of drug interactions, enabling healthcare professionals to make more informed decisions about patient care. For example, AI algorithms can analyze a patient's medical history, current medications, and genomic data to predict potential drug interactions and recommend alternative treatment options if necessary. This personalized approach to medicine can help optimize patient care and minimize the risk of adverse drug events.

The detection and classification of drug–drug interactions (DDI) from obtainable data are of high magnitude because recent reports show that DDIs are among the major causes of hospitalacquired conditions and readmissions and are also compulsory for elegant healthcare. Therefore, to avoid adverse drug interactions, it is necessary to have an up-to-date knowledge of DDIs. This knowledge could be extracted by applying text-processing techniques to the medical literature published in the form of 'Big Data' because, It is crucial to

automate the extraction of the interactions taking place between drugs because the medical literature is being published in immense volumes, and it is unfeasible for healthcare professionals to read and collect all of the investigate DDI reports from these Big Data. To avoid this time- consuming procedure, the Information Extraction (IE) and Relationship Extraction (RE) techniques that have been deliberate in depth in Natural Language Processing (NLP) could bevery shows potential.

In vivo and in vitro experiments can facilitate the identification of DDI, but cannot be performed in some cases due to laboratory limitations and/or high cost (Safdari et al., 2016). Thus, it is particularly important to develop computational methods to solve problems of identifying DDI.Current computational approaches to identify DDI can be divided into two categories:

1) extraction of DDI from literature, electronic medical records, and spontaneous reports;

2) use of known DDI to predict unknown DDI.

III. LITERATURE SURVEY

Hunta et al. presented an enzyme and transporter protein Integrated Action Crossing (IAC) technique to predict non-communicable disease's drug-drug correlation depending on pharmacokinetic strategy. In this technique, the drug-drug correlation data were collected from the web. Then, the novel characteristics were created and applied to the different machine learning algorithms to generate the prediction model. But this technique was not suitable for large-scale databases.

Peng et al. developed a learning-based technique depending on feature interpretation training and deep learning called DTI-CNN to predict drug-target correlations. Initially, the Jaccard similarity coefficient extracted the related attributes of drugs and proteins from heterogeneous networks and restarted the random walk scheme. After that, a denoising AE was used to minimize the dimension and detect the important attributes. According to these attributes, the CNN was created to predict the correlation between drugs and proteins. But, it needs more related data and network design to contain more sophisticated input networks.

Ding et al. designed a Multi-view Graph Regularized Link Propagation (MvGRLP) framework to predict novel drug-target correlations. This framework merged complementary data among various views in drug and target space. Also, an iterative scheme was applied to resolve the objective function. But, it needs to define structural correlations among entities to increase the prediction efficiency.

Khaled Mohamad Alumstafa K-nearest neighbor,Decision tree and SVM classifiers show the performance of the selected classification to classify the best or predict the heart disease cases.
Rayan AlanaziConvolutional Neural Network(CNN) is for the prediction of the disease, and K-nearest neighbor (KNN) is used for calculating the distance to find the match that is generated in the data set for the prediction of diseases.

C K GomathyNaïve Bayes Supervised Machine Learning algorithm predicts the disease. The probability of the disease is calculated using the Naïve Bayes algorithm.

IV. MATERIALS AND METHODSDATA COLLECTION

The DDI dataset was collected from the DrugBank database (version 5.1.7, released 2020-07- 02) within the histamine-antagonist category. The drug interaction pairs heading "approved" was retained, and the drugs with labels "has been revoked", "experimental", and "investigational" headers were removed from the dataset. We used specific descriptions of the interaction (also collected from the DrugBank database) between two drugs to classify the DDI type, for example: "DRUG_A can cause a decrease in the absorption of DRUG_B, resulting in a reduced serum concentration and potentially a decrease in efficacy"; meanwhile, “DRUG_A belongs to the histamine antagonist group, and DRUB_B is a drug that, when combined with DRUG_A, can induce ‘a reduced serum concentration and potentially a decrease in efficacy’". There were 92 types of interactions based on this specific description, which were categorized from 0 to 91. In total, we collected 67,317 DDI pairs together with their detailed interactions using SMILE. However, after considering the statistics of the data, we found that some input interaction pairs were present at very small amounts (less than 50 pairs), which will lead to a much lower prediction rate for these types of interactions due to insufficient training data. Therefore, we aggregated the interaction types that were present at small amounts in the input data based on the following condition: if the input data of each interaction type include less than 50 drug pairs, the type will be categorized into an interaction group numbered 51.This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3861479Preprint not

peer reviewed To further demonstrate the ability of HAINI to predict unknown DDIs, we investigated the performance of the model on the validation dataset prepared from 19,971 DDI pairs from the DrugBank database with the statuses "experimental" and "investigational".Interaction features extractionPyBioMed is a package written in the Python programming language that can be used to create numerous future vectors from molecular structure, protein sequences, and DNA sequences. PyBioMed is a remarkable tool and can be applied to a wide range of tasks in areas related to cheminformatics, bioinformatics, and systems biology. The PyBioMed package includes six main modules, PyInteraction, PyDNA, PyMolecule, PyProtein, PyGetMol, and PyPretreat, to compute various molecular descriptors as well as assist in processing the input data. In this study, we mainly focused on identifying drug interactions based on chemical structure; hence, the Interaction module was used to calculate the features between drug pairs with interactions based on the SMILE structures of 67,317 DDI pairs.

V. DEVELOPMENT OF THE PREDICTIVE HAINI MODEL

In this study, we applied different algorithms: Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) for Machine Learning, and Multilayer Perceptron (MLP) This preprint research paper has not been peer reviewed. for Deep Learning models. The traditional way of performing parameter optimization is a grid search, which is simply an exhaustive search of a manually specified subset of the hyperparameter space of each machine learning algorithm

Naive Bayes (NB)

In this study, the Naive Bayes method was applied as a supervised learning algorithm for our data set Since the 1960s, Naive Bayes has been applied in machine learning and is highly accurate, fast, and simple, and it also performs well in various class predictions. The posterior probability of our data set was determined using the equation below:

$P(c \mid x1,\ldots, xn) = P(c) * P(x1,\ldots, xn|c)$
$P(x1,\ldots, xn)$
$P(c \mid x$ is the posterior probability of a class (c, target) of a given predictor(x, 1,…., xn)attributes).

1. P(c) is the prior probability of a class.
2. P(x is the likelihood, which is the probability of a predictor of a given class. 1,…,xn|c)
3. P(x is the prior probability of the predictor. 1,…, xn)
4. Vector represents some n features. (x1,…, xn)

Decision Tree (DT)

In the field of machine learning, Decision Tree (DT) is a predictive model, that is, it generates a map from observations of an object/phenomenon to allow conclusions to be drawn about the target value of the object/phenomena. Each internal node in a tree corresponds to a variable, and each subset represents a value specific to that variable. Each leaf node represents the predicted value of the target variable given the values of the variables represented by the path from the root node to that leaf node. In this study, the data were given as records of the form:

(X, Y) = (X1, X2, X3 ..., Xi, Y)

In which the interaction types or Y are denoted (51 classes - dependent variables) as the variables for prediction. X1, X2, X3, and so on are variables equivalent to descriptors that act as input data to contribute to the decision of the type of interaction between the given drug pair.

Random Forest (RF)

Random forest (RF) or random decision forest describes an ensemble learning algorithm that was first introduced by Tin Kam Ho in 1995 .This statistical learning methodology is used for classification, regression, and other tasks that operate by generating multiple decision trees at training time and outputting the class that is the mode of the classes (classification) or the mean prediction (regression) of the individual trees in the forest implemented on three similarities of the DDIs.

Logistic Regression (LR)

The Logistic Regression (LR) method is a regression model that predicts a discrete target variable y corresponding to an input vector x. This is equivalent to whether the feature (x) extracted from the SMILE constructs of the drug pair belongs to any of the 51 classes.

Multilayer Perceptron (MLP)

We also tested predictability through a deep learning model with a supervised learning algorithm, Multilayer Perceptron (MLP) .MLP is a feedforward artificial neural network (ANN) that refers to networks composed of multiple layers of perceptron (one of the classification algorithms proposed by Frank Rosenblatt) stacked in several layers to solve complex problems. We use the function f (·): Rm→Rn by training on a dataset, where m is the number of dimensions of the input and n is the number of dimensions of the output, given a set of 3600 features X= x1, x2, x3, ..., x3600 and a y target (a type of interaction). In this study, we applied 3 layers while using the MLP algorithm with loss = "binary crossentropy" and an “adam” optimizer.

RESULTS

Evaluation of HAINI performance

In this study, we trained and tested the HAINI model on the compiled drug-drug interaction (DDI) dataset from the DrugBank database (version 5.1.7, released 2020-07-02). Here, we focused on adverse drug interactions where the chosen drugs belong to the histamine antagonist group or interact with them. The framework of HAINI, which is clearly described in Figure 1, consists of the following three main steps:

(i) filtering drugs (including drug ID, SMILE, and detailed interaction) of the histamineantagonist group and drugs that interact with them from the DrugBank database

(ii) labeling the interaction types and extracting features of the interactive drug pairs based on SMILE and the PyInteraction module and

(iii) applying the machine learning and deep learning models via common classifier algorithms, such as Naive Bayes, Decision Tree, Random Forest, Logistic Regression, and Multilayer Perceptron. All machine learning models have been implemented and visualized using Python scikit-learn and matplotlib packages.

We used 5-fold cross-validation (k-CV) to evaluate the efficiency of the HAINI model.

The extracted features from known drug-drug interactions were randomly split into five subsets with equal sizes. In each fold, one subset was used as a testing set with the fraction of 0.2, and the other four were used as the training set. Several evaluation metrics were applied to investigate the performances of the prediction models for various algorithms, i.e., Precision, Recall, F-measure (F1) as follows:

Precision = TP/(TP+FP)
Recall = TP/(TP+FN)
F–1 score = (2*Precision*Recall)/(Precision + Recall)

where TP, FP, TN, and FN stand for True Positive, False Positive, True Negative, and False Negative, respectively. Precision is the ratio between the numbers of correctly predicted DDIs and all DDIs; recall is the ratio between number of correctly predicted DDIs and all true DDIs. Since recall and precision affect each other, improving one of them may lead to a reduction of the other. Therefore, we used the F-1 score, which is the geometric mean of precision and recall, to make the predictive results more reasonable.

VI. DISCUSSIONS

In this paper, our main goal was to use the properties of the chemical structures of interactive drug pairs to generate accurate predictive models rather than having to combine too many additional similarities, including ATC, target, transporter, side effect, etc. Using a single similarity will save time in having to use encoders to simplify input data, avoid overfitting in the machine learning model, and also help to smooth the progress of the accurate prediction of drug interactions as well as adverse drug interactions. In addition, while our model has a simple input type, a large number of features (approximately 3600) can still be produced with each pair of drugs when using the PyInteraction module, and the predicted value of the HAINI model in each class (equivalent to a detailed interaction type) is also very high (over 0.82). With a large quantity of concentration drawn to drug-drug interactions in the field of computer science research, computational predictions using machine learning

are highly encouraged. Compared to the results of some previous studies using only chemical similarity in machine learning or deep learning models the direct use of interactive features generated from interactions of drug pairs has a stronger effect than when separating the features of each structure and then building the matrix network or assessing the similarity of each drug with each other. For example, compared with the Deep Learning - Neural Network model of Narjes Rohani and colleagues we can notice that the creation of a matrix comparing the similarity between SMILE structures of drug pairs serves as a diagnostic database for precision, which is extensively lower than in the HAINI model while HAINI is also amulticlass model and is clinically more significant than a binary model (allowing only cases for which an interaction is present). It is important to have a model to accurately predict DDIs for the reasons outlined above, such as saving on testing costs and quickly identifying drug interactions that cause many side effects. Such models are also effective tools to support doctors in prescribing drugs to people who are taking different medications at the same time. When prescribe a drug to a patient, it is more helpful to specifically identify the details of the interaction between drugs than only prominent whether the drug has side effects. Some drugs from time to time have only mild side effects for patients that are within good enough levels; therefore, only defining a binary class problem with or without information about an interaction does not make sense in contributing to clinical research or drug development.

Moreover, in our study, we found that the high predictability rates of some types of interactions are not entirely reliant on the amount of data input (i.e., number of interacting drug pairs). We can clearly observe this based on supplemental data in interaction types 27, 1,12, 27, and 39; the number of input data pairs of these interactions is much lower than the topclasses

Currently, although the quality of health management in the world is increasing, the amount of histamine antagonists prescribed, as well as the cost of treating related diseases, have not shown signs of decline. The management and evaluation of factors related to using antihistamine or histamine antagonists in clinical treatments are necessary to improve the quality of medical treatments and to reduce the economic burden on the health care system. This study shows the potential application of Artificial Intelligence technology to reduce unnecessary costs of evaluating clinical drug interactions, especially with the incorporation of a machine learning method. The performance of computational-assisted prediction also demonstrates a simple method to analyze the meaning of drug-interaction networks. Using machine learning analysis could lead to the identification of many potential undiscovered- interactions and will also reduce the cost of clinical trials of new drugs. Moreover, there will be more potential in the analysis of drug interactions when various genomic, chemical, or other datasets related to human metabolism processes are combined. The limitation in our analysis is that the findings are only based on large-scale databases and do not include more relevant clinical databases (e.g., Drugs.com, Medscape Multi-Drug Interaction Checker, RxList). on the other hand, through this study, we have shown the potential use of these downloadable databases, which would greatly expand the possibilities of massive data miningin the medical field.

VII. CONCLUDING REMARKS

We propose a new multilayer prediction model, HAINI, using machine learning in conjunction with the PyBioMed package to extract DDIs. In addition, the performance of the HAINI model is robust based on a single similarity and is ready to integrate more drug- and targetrelated information. The model takes advantage of multitasking learning by predicting whether two drugs interact with each other and more clearly distinguishing types of interactions only using the SMILE structure of the drug. The results of the tests on both the training and validation datasets were high for most specific types of interactions. In the future, we will continue to explore other structures of the neural network for multitasking learning to improve the performance of the model to detect DDIs of various drugs at the same time, not only for drugs of the histamine antagonist group. Moreover, our prediction is based solely on the SMILE and thus can be applied at early stage of drug development.

VIII. REFERENCES

1. Ogu, C.C. and J.L. Maxa. Drug interactions due to cytochrome P450. in Baylor UniversityMedical Center Proceedings. 2000. Taylor & Francis.

2. Lynch, T. and A.L. Price, The effect of cytochrome P450 metabolism on drug response,interactions, and adverse effects. American family physician, 2007. 76(3): p. 391-396.

3. Dipiro, J.T., et al., Pharmacotherapy: A Pathophysiologic Approach, ed. 2014: McGraw- Hill Medical, New York.

4. Sproule, B.A., et al., Selective serotonin reuptake inhibitors and CNS drug interactions. Clinical

pharmacokinetics, 1997. 33(6): p. 454-471.

5. Bui, Q.-C.; Sloot, P.M.A.; van Mulligen, E.M.; Kors, J.A. A novel feature-based approach to extract drug-drug interactions from biomedical text. Bioinformatics 2014, 30, 3365–3371. [CrossRef]

6. Raihani, A.; Laachfoubi, N. Extracting drug-drug interactions from biomedical text using a feature-based kernel approach. J.Theor. Appl. Inf. Technol. 2016, 92, 109–120.

7. Zhang, J.; Li, H.; Mo, D.; Chang, L. Mining Multispectral Aerial Images for Automatic Detection of Strategic Bridge Locations for Disaster Relief Missions. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer Nature Switzerland AG: Cham, Switzerland, 2019; Volume 11607, pp. 189–200. [CrossRef]

8. Sun, X.; Dong, K.; Ma, L.; Sutcliffe, R.; He, F.; Chen, S.; Feng, J. Drug-Drug Interaction Extraction via Recurrent Hybrid Convolutional Neural Networks with an Improved Focal Loss. Entropy 2019, 21, 37. [CrossRef]

9. Ali Awan, A.; Khalid, U.; Maqsood, A. Revolutionizing Telemedicine by Instilling H.265. Int. J. Image Graph. Signal Processing2017, 9, 20–27. [CrossRef]

10. Wu, H.; Xing, Y.; Ge, W.; Liu, X.; Zou, J.; Zhou, C.; Liao, J. Drug-drug interaction extraction via hybrid neural networks on biomedical literature. J. Biomed. Inform. 2020, 106, 103432. [CrossRef] [PubMed]

11. Park, C.; Park, J.; Park, S. AGCN: Attention-based graph convolutional networks for drug-drug interaction extraction. Expert Syst.Appl. 2020, 159, 113538. [CrossRef]

12. Agatonovic- Kustrin S, & Morton D, "Data Mining in Drug Discovery and Design," In Artificial Neural Network for Drug Design, Delivery and Disposition, Academic Press, pp. 181-193, 2016.

13. Chittora, A., &Mekala A. M, "Discovery of Drug and Medicine Using Data Mining Techniques," Research Journal of Pharmacy and Technology, vol. 10, no. 12, pp. 4147-4151, 2017.

14.Yosipof, A., Guedes, R. C., & García-Sosa A. T, "Data Mining and Machine Learning Models for Predicting Drug Likeness and Their Disease or Organ Category," Frontiers in Chemistry, vol. 6, pp. 1-11, 2018.

15.Dunne, S., Shannon, B., Dunne, C., & Cullen W, "A Review of the Differences and Similarities Between Generic Drugs and their Originator Counterparts, Including Economic Benefits Associated with Usage of Generic Medicines, using Ireland as a Case Study," BMC Pharmacology and Toxicology, vol. 14, no. 1, pp. 1-19, 2013.

16. Tamargo, J., Le Heuzey, J. Y., & Mabo P, "Narrow Therapeutic Index Drugs: A Clinical Pharmacological Consideration to Flecainide," European Journal of Clinical Pharmacology, vol. 71, no. 5, pp. 549-567, 2015