Benha Journal of Applied Sciences (BJAS) **print: ISSN 2356–9751**
Vol. (9) Issue (*) (2024), (13-21) **online: ISSN 2356–976x**
http://bjas.bu.edu.eg

A comprehensive survey explores Drug-Drug interaction prediction using Machine-Learning techniques

Yasmin A.Radwan[1*], Karam A.Gouda[2], Ibrahim Z.Abdelbaky[3] and Mona M.Arafa[3]
[1]*Information System Department, Higher Institute of Computer Science and Information Technology, El-Shorouk Academy, Egypt*
[2]*Information Systems Department Faculty of Computers and Artificial Intelligence, Benha University, Egypt*
[3]*Information Systems Department Faculty of Computers and Artificial Intelligence, Benha University, Egypt*
E-mail: Yasmin.Atef.Radwan@gmail.com

Abstract

Drug-Drug Interactions is a critical health and safety concern that receives a lot of attention from both academia and business. Polypharmacy is often employed as a strategy to manage complex diseases such as cancer, diabetes, and age-related ailments. However, combining medications with other drugs can lead to unintended adverse reactions. Interactions between drugs may increase the chance of unanticipated negative effects and even unknown toxicity, putting patients at risk. Detecting and identifying Interactions not only helps clinicians avoid chronic but will also encourage the co-prescription of safe drugs for more effective therapies. It is expensive and time-consuming to identify drug-drug interactions and Adverse Reactions among several medication pairings, both in vivo and in vitro. Recent advancements in computer science, specifically in the field of Artificial Intelligence, have yielded techniques that enable researchers to identify drug-drug interactions. We present comprehensive approaches that enable in-depth analysis of potential interactions by taking into account various factors, including molecular structure, clinical data, network relationships, and existing literature. This paper offers an all-encompassing survey of research studies that utilize Machine Learning and Deep Learning algorithms for the prediction of Drug-Drug interactions.

Keywords: Drug-Drug interactions; Adverse Reactions; artificial intelligence.

1. Introduction

Failures during the final stages of clinical studies can be both costly and potentially harmful to the participants involved. Not only do they incur significant financial expenses, time, and labor, but they also pose risks to the individuals enrolled in the studies. While failures in clinical studies may hinder scientific progress, they can be mitigated through diligent research, curiosity, and adherence to rules and regulations. Learning from past mistakes is crucial in reducing the likelihood of future errors. This involves effectively acquiring, collaborating, and evaluating knowledge gained from both successful and unsuccessful drug examinations.

By analyzing the outcomes of previous studies, researchers can identify patterns and factors that contribute to failures, enabling them to make informed decisions and improve the development process. Furthermore, leveraging this knowledge to develop innovative approaches can help address the current shortcomings and improve the overall success rate of clinical studies. By incorporating lessons learned from previous errors, researchers can refine their methodologies, enhance study designs, and implement strategies to minimize the occurrence of mistakes (Scavetta and J, 2020). An ideal medication should possess certain characteristics such as non-toxicity, biodegradability, and biocompatibility.

Toxic reactions occur when drugs accumulate excessively in the bloodstream, leading to adverse effects on the human body. The ability of a drug and its related compounds to undergo metabolism within the body and the surrounding environment after fulfilling their therapeutic purpose is known as biodegradability and biocompatibility. A drug needs to break down into non-toxic byproducts that can be safely eliminated from the body's tissues without causing harm. Drug-drug interactions (DDIs) are more prevalent among older individuals who require ongoing treatment for one or more chronic conditions. As the number of approved medications increases, the likelihood of interactions between prescribed drugs also rises.

Recent advancements in machine learning have made a significant impact on the healthcare sector, particularly in the fields of disease diagnosis and prediction (Carleo et al.,2019). Over the past two decades, a wide range of machine learning algorithms, including feature selection algorithms, have been extensively employed in these areas. These techniques have shown promising results in tasks related to cancer evaluation and prediction, playing a crucial role in improving individual medical care. By leveraging machine learning techniques, healthcare professionals can enhance their ability to diagnose and predict diseases, leading to more effective and personalized treatment approaches.

This paper is structured as follows: Section 2 addresses the topic of DDIs and explores the associated challenges. In Section 3, we discuss the datasets used for DDIs prediction, including a detailed description of each dataset and the entities involved. Section 4 presents a comprehensive literature survey on the subject. Lastly, Section 5 offers a conclusion summarizing the key findings and insights discussed throughout the paper.

2. Drug-Drug Interactions

DDIs prediction holds immense significance in contemporary healthcare. It enhances patient safety, supports drug development, facilitates personalized medicine, and augments clinical decision-making

processes. By effectively managing drug interactions, healthcare professionals can optimize treatment outcomes and improve overall patient care. When a medication is combined with one or more drugs that either inhibit, enhance, or diminish its intended effects, it can, in severe cases, result in undesired adverse reactions, leading to substantial morbidity and increased mortality worldwide. It is important to recognize that a single medication alone may not always be the most effective treatment approach. The majority of diseases in patients are brought on by intricate biological mechanisms that cannot be cured by one particular medication and often require multiple medications (Bumgardner et al.,2021).

Different types of drug interactions include pharmaceutical interactions, pharmacodynamics (PD) interactions, and pharmacokinetics (PK) interactions. Pharmaceutical interactions occur due to chemical reactions caused by improper dispensing before drugs are administered. For example, mixing tetracycline and calcium salt injection can lead to precipitation due to chelate formation in specific conditions. PD interactions happen when two drugs share the same receptor, altering each other's pharmacological effects. For instance, the combination of atropine and tubocurarine blocks the action of acetylcholine by binding to receptors. PK interactions occur when two or more drugs are taken together, affecting how they are absorbed, distributed, metabolized, and eliminated in the body. These interactions can either improve the effectiveness of the drugs or lead to adverse reactions. Unlike PD interactionsPK interactions primarily affect the blood concentration of the drugs involved. For instance, combining warfarin with nonsteroidal anti-inflammatory drugs can result in PK interactions (Zhao et al.,2024).

DDIs are typically examined in medicinal chemistry; however, a significant number of interactions go unnoticed, resulting in a wide range of medication combinations (Vilar et al.,2014). Given that many diseases require multiple medications due to complex biological mechanisms, it is impractical to identify every potential DDIs during the early stages of drug development (Bumgardner et al.,2021). In addition, a significant number of Adverse Drug Reactions often go unnoticed during pre-approval clinical trials. As a result, researchers analyze DDIs to predict potential interactions that may have been previously unidentified and to investigate their connections with the pharmacodynamic and pharmacokinetic properties of drugs (Kastrin et al.,2018).

Many DDIs are unexpectedly discovered after medicine has already entered the market or clinical use, highlighting the urgency for raising awareness about potential risks before a medication is made available to the public (Zhang et al.,2015). It is important to note that not all medications are equally effective in treating illnesses. However, for an ideal medication, the treatment would be 100% successful and safe for individuals, producing the intended outcomes. Furthermore, the medication would undergo simple metabolic processes and be easily eliminated from the human body, without causing adverse reactions in other organisms or body parts (Richelson and E,1994). Detecting and identifying DDIs not only helps healthcare providers avoid potential complications but also enables the co-prescription of safe drugs to enhance the effectiveness of treatments (Zhang et al.,2015). Gaining comprehensive information on new therapies, particularly regarding drug-drug interactions, can often be challenging when evaluating adverse medication events. In some cases, Drug A may affect the absorption of Drug B, resulting in decreased serum levels and potentially reduced efficacy. For instance, there is evidence of an interaction between Fluconazole, a triazole antifungal medication, and simvastatin, a cholesterol-lowering medication, which increases the risk of rhabdomyolysis and myopathy (Vilar et al.,2012).

3. Database Description

To predict DDIs, researchers leverage diverse datasets which offer valuable insights about drug characteristics, molecular structures, pharmacological profiles, and documented interactions. These datasets serve as a valuable resource for extracting relevant features that contribute to the accurate prediction of potential drug interactions.

***DrugBank*:** is a well-known bioinformatics and cheminformatics database that provides comprehensive details on medicines, drug targets, adverse effects, chemical structures, target proteins, drug interactions, and other associated data. The dataset contains over 4,100 drug entries and encompasses detailed information on drug interactions, including pharmacokinetic and pharmacodynamic interactions. It includes a collection of 365,984 direct interactions among these drugs (Wishart et al.,2018).

***MedLine*:** is an extensive biomedical literature database that encompasses scientific papers, publications, Drug Adverse Reactions, Targets and releases (Wang et al.,2014).

***KEGG* (Kyoto Encyclopedia of Genes and Genomes):** is a valuable database resource for understanding the high-level functions and utilities of biological systems. The latest version of KEGG DRUG comprises 12,316 drug entries, providing comprehensive information on drugs and their properties (Kanehisa and M,2019).

TWO-SIDES: is a dataset specifically designed for extracting drug-drug interaction details from medical literature using text-mining algorithms. It contains 645 drugs and 46,221 drug-drug pairs, each labeled with 200 distinct drug side effect types (Tatonetti et al.,2012).

***PharmGKB (Pharmacogenomics Knowledge Base)*:** contains detailed information on drug metabolism, drug targets, drug pathways, and genetic variations. It is a valuable resource for predicting drug-drug interactions. PharmGKB contains valuable information on drug reactions, side effects, and known interactions. The

PharmGKB dataset consists of 12,664 drug entries (Whirl-Carrillo et al.,2021; Karim et al.,2019).

SIDER (Side Effect Resource): contains information on drug adverse reactions reported in the literature, along with drug labels and targets. This dataset helps identify potential drug interactions based on shared adverse effects. SIDER includes 1,430 drugs and 5,880 adverse drug reactions (Kuhn et al.,2016; Han et al.,2022).

FDA Adverse Event Reporting System: serves as an online repository for data on adverse reactions and medication error reports reported to the FDA (Banda et al.,2016).

4. Literature Survey

When it comes to predicting DDIs, several primary approaches utilize different methodologies and techniques. These approaches can be categorized as follows: These categories include:

4.1. Literature-based approach

Text-mining tools employ natural language processing methods to extract logical connections between medications. These tools utilize text mining techniques to search and gather documented DDIs

from various databases, including procurement claims, the FDA Adverse Event Report, and electronic medical reports (Feng et al.,2022). Another architecture focuses on the identification and categorization of pharmacologic substances, such as Drug's Name, Drug's Brand Name, Drug's Group Name, and active substances not permitted for use by humans. This approach extracts DDIs from the DrugBank and MedLine abstracts corpus using non-linear kernel approaches (Segura-Bedmar et al.,,2013). In a different approach, a combination of drug embedding derived from biomedical texts from the 2013 corpus is used. These embeddings are generated through the Bio-BERT embedding-based method. Additionally, the drug molecular structure is represented as a multi-dimensional space using the Variational Autoencoder (VAE) algorithm (Mondal et al.,2020).

A Convolutional Neural Network (CNN) was utilized to predict types of DDIs in a multi-class setting. This approach was applied to biomedical texts from the DrugBank corpus and MEDLINE abstracts. The preprocessing steps involved blinding drugs and tokenizing sentences using the Natural Language Toolkit. Words and their positions within sentences were embedded and combined, then fed into the CNN model for prediction (Liu et al.,2016).

For multiple-label DDIs prediction, three key processes were employed. Firstly, a knowledge graph was created using four knowledge graphs from Bio2RDF (DrugBank, KEGG, PharmaGKB, and Comparative Toxicogenomics Database). Secondly, in addition to the drug knowledge graph, biological DDIs text, consisting of DDIs documents from DrugBank and MEDLINE Abstracts, was embedded into a low-dimensional vector. Finally, DDIs prediction was performed using the learned embeddings as a link prediction methodology (Wang et al.,2021).

To identify relevant abstracts and PubMed articles confirming chemical DDIs, the efficiency of various classifiers such as Logistic Regression (LR), Support Vector Machines (SVM), and discriminatory analysis has been evaluated (Kolchinsky et al.,2013). Moreover, their methodology enables the connection of causal processes to potential DDIs. In this approach, a parsing tree is utilized to obtain different interactions, and logical rules are employed to predict interactions based on extracted interactions between the novel and existing drugs (Tari et al.,2010).

One advantage of the Text-Based Approach over other methods, such as the Classification-Based Approach and Graph-Based Approaches, is its ability to employ natural language processing techniques to extract and analyze textual data. By leveraging these techniques, text-based methods can gain a semantic understanding of the text, allowing them to capture nuanced information regarding drug-drug relationships, drug targets, and biological pathways. This semantic understanding offers an advantage in terms of enhancing the accuracy and depth of the predictions, surpassing the capabilities of other methods that may not have access to such detailed textual information.

4.2. Similarity-based approach

The similarity-based approach is a widely used framework in machine learning and data analysis that aims to measure the degree of similarity or distance between data points to make predictions. This approach is based on the assumption that data points with similar characteristics or behaviors are likely to have similar outcomes. In the context of DDIs, the similarity-based approach operates on the premise that medications with similar properties may interact with each other. By considering the chemical structures of drugs, similarity-based methods assume that medications with chemically identical or similar structures are likely to have similar interactions with other drugs. This approach leverages the concept of similarity to identify potential drug interactions and can be valuable in predicting and understanding the effects of combinations of medications (Kastrin et al.,2018).

To measure the similarity between medications, common substructures are potentially used instead of entire chemical structures (Bumgardner et al.,2021). If medications both A and B interact to generate a particular response, then medications similar to drug A (or drug B) are probable to generate an identical impact as drug B (or drug A). In terms of medication similarity, interactions among novel drugs can be predicted by combining similar properties among various drugs (Shaker et al.,2021; Mo et al.,2020).

A significant technique for identifying and predicting DDIs is proposed using molecular structural similarity data obtained from fingerprint-based modeling. The Tanimoto Coefficient (TC) method is employed to calculate five DDIs similarity scores, which are then combined to form a single DDI score (Vilar et al.,2014)

By utilizing the Tanimoto coefficient method, a total of 58,403 novel predicted interactions were generated. Among these, 430,128 new DDIs were identified, including 9,454 DDIs from the DrugBank database, which were used to construct the model. This approach leverages molecular structural similarity data and computational methods to predict and analyze potential interactions between drugs, thereby aiding in the identification of new DDIs and expanding our understanding of drug interactions (Vilar et al.,2013)

The Rus-Rao approach has been employed for calculating the similarity among drug pairs according to similarity measurements of binary vectors as depicted in **Figure 1**. The higher the similarity, the more likely the drug interactions (Ferdousi et al.,2017)

A novel method for predicting DDIs is proposed, which incorporates several drug-related similarities. These similarities encompass various aspects such as Target, Pathway, Substructure of drugs, Transporter, Gene Ontology (GO), Anatomical Therapeutic Chemical classification (ATC), Gaussian Interaction Profile, Indication, Side Effects, Off-label Side Effects, and Interaction datasets. To ensure informative and non-redundant similarities, a heuristic strategy is employed for selecting a collection of similarities. The selected similarities are then integrated using the Similarity Network Fusion (SNF) method. Additionally, a Deep Neural Network (DNN) model is applied to utilize the interaction data and the combined similarity matrix (Song et al.,2019).

4.3. Graph-Embedding Approach

The primary goal of graph-embedding methods is to represent a graph in a low-dimensional vector while preserving its structural details. In the field of biological sciences, graph embedding methods have shown superior performance compared to conventional techniques and employed a Graph Neural Network (GNN) model with a hyper-graph edge encoder and decoder to generate drug embeddings. By concatenating the representations of drugs, they successfully predicted DDIs scores for drug pairs (Saifuddin et al.,2023). Deep learning techniques have proven effective in extracting drug features from datasets and conducting self-training through multiple layers of the neural network to predict previously unidentified DDIs. It proposed a DNN-based approach that constructs an architecture utilizing various types of drug data. By encoding SMILES as low-dimensional vectors using one-hot encoding and incorporating topological features acquired from a knowledge graph (using GNN, they achieved an accurate prediction of unidentified DDIs (Dang et al.,2021).

another study, employed GNN or Multi-Layer Perceptron (MLP) to generate Morgan fingerprints for drug pairs. These fingerprints, representing SMILES as binary vectors, were then fed into a MLP to predict output labels (Long et al.,2022; Herrero-Zazo et al.,2016). Knowledge graphs and machine learning models, combined with embedding techniques, are utilized to detect DDIs (Celebi et al.,2019). The Explainable Substructure Partition Fingerprint (ESPF) algorithm is employed to partition fingerprints and extract common and frequent substructures of drugs. GNN techniques are then used to embed each drug node and the embeddings of drug pairs are concatenated and utilized as features in Machine-Learning models (Bumgardner et al.,2021).

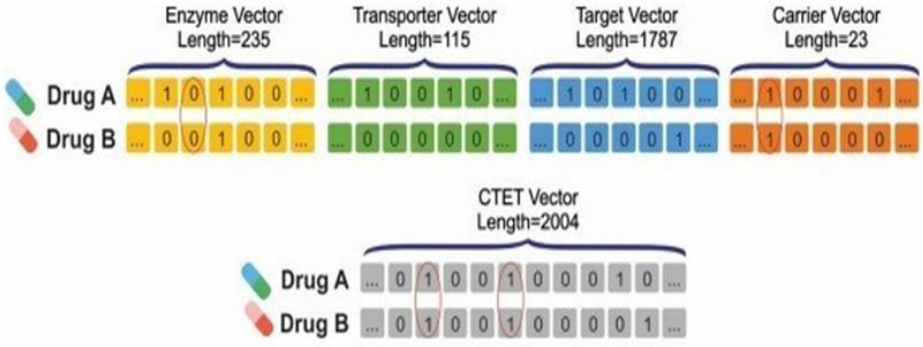

Fig. (1) Similarity-based approach to predicting drug-drug interactions (DDIs)

The SmileGNN model focuses on the structural features of drugs represented as strings of characters (SMILES). One-Hot Encoding is applied to smiles, followed by PCA dimensional reduction to represent smiles as Low-Dimensional Vectors. These vectors are then aggregated with the topological features of drugs, which capture the connections between the drug and other compounds in the knowledge graph obtained using GNN. The sigmoid function is used to calculate the drug pair score, which is scaled to the range of (0, 1). Scores exceeding 0.5 indicate the presence of a DDIs between drug pairs (Kastrin et al.,2018). The PyBioMed package is a Python library that includes the PyInteraction module, which was utilized for calculating features of drug pairs and multiplying them (Dang et al.,2021). In the context of predicting synergy scores, a combination of Recurrent Neural Networks (RNNs), CNN, and Mixture Density Networks (MDNs) is employed. These models leverage high-dimensional input features obtained using CNN from cell lines. To illustrate drug

similarities, substructures extracted from drug smiles are utilized to create a hyper-graph, where each node represents a drug substructure. Drug embeddings are generated using a GNN model with a hyper-graph edge encoder and decoder. The representations of drugs are concatenated to predict DDIs scores for each drug pair (Saifuddin et al.,2023).

Morgan fingerprints of drug pairs are generated using GNN and MLP. These fingerprints encode drug smiles into binary vectors, and the separated representations of drug pairs are either concatenated or added to form a single vector. This vector is then fed into a MLP to predict output labels (Long et al.,2022).

A GNN-DDI model is constructed to predict potential DDIs. It utilizes a GNN architecture with five layers to find k-hops for each drug, generating low-dimensional representations from molecular graphs. All features are concatenated for each drug pair and fed into an MLP algorithm to predict the final result (Feng et al.,2022).

A novel method for predicting DDIs incorporates various drug-related similarities, such as Target,

Pathway, Substructure of Drugs, Transporter, Gene Ontology, Anatomical Therapeutic Chemical, Gaussian Interaction Profile, Indication, Side-effect, and Off-label Side Effects, as well as interactions from datasets. A heuristic strategy is employed to select informative and nonrepetitive similarities, which are then integrated using the SNF method. Finally, a Deep Neural Network DNN model is applied using the interaction data and the combined matrix of similarities (Rohani et al.,2019). In another study, a DNN model is utilized for DDIs prediction, where drugs are represented as SMILE code features and fed into the DNN (Hou et al.,2019). A deep-learning framework is employed for predicting DDIs by integrating chemical-based substructures, targets, enzymes, and pathways within a deep-learning architecture. Four drug description vectors are computed and trained in the DNN network (Deng et al.,2020).

Graph-based methods offer interpretability by visualizing the molecular graphs and highlighting key interactions. This visual representation facilitates the interpretation of results and aids in identifying important molecular features or regions that contribute to the predicted interactions. The Deep Predictor Drug-Drug Interaction (DPDDI) method incorporates the structural features of drugs, such as ATC, Drug Binding Proteins (DBP), and Chemical Structures (CS), into a DNN model.The DNN model concatenates the feature vectors of two drugs into one feature vector for drug pairs and is trained to predict DDIs (Feng et al.,2020) as depicted in **Figure 2**. A deep learning-based model named SSF-DDI has been developed for DDI prediction. This model aims to tackle the limitation of existing DDI prediction methods, which primarily rely on topological information between drug molecules and overlook features of drug molecule sequences. SSF-DDI integrates both drug sequence features and structural features derived from the drug molecule graph to enhance DDI prediction. By combining these features, the model offers a more comprehensive and accurate representation of drug molecules, thereby improving the prediction accuracy of DDIs (Zhu et al.,2024).

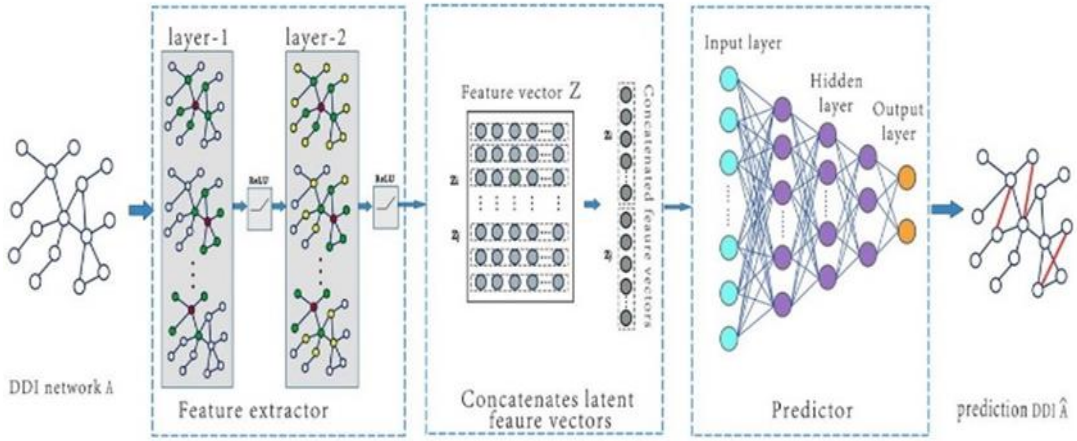

Fig (2) Architecture of the DPDDI Method

4.4. Classification-Based Approach

In the conventional classification-based approach, the prediction of DDIs is treated as a binary classification problem. This means that the task involves assigning a label to each drug pair, indicating whether they interact or not. However, it is important to note that not all drug pairs have known interactions, and there may be undiscovered or unnoticed interactions that need to be predicted. To address this challenge in DDIs prediction, algorithms are developed using various classifiers, including LR, Bayesian methods, K-Nearest Neighbor, Random Forest, and SVM. These classifiers are commonly used to build predictive models that can classify drug pairs as interacting or non-interacting based on input features (Han et al.,2022).

To predict drug pairs using both molecular and pharmacological features, scientists have developed a probability ensemble approach that incorporates a to improve the accuracy and reliability of drug pair prediction (Shi et al.,2016). According to structural

along-side-effect similarities, novel DDIs can be predicted (Zhao et al.,2019). Bayesian network model and a similarity technique. This approach leverages the strengths of both methods The conventional approach of using similarity and classification techniques can be employed to predict unknown drug-drug interactions. However, when solely relying on these approaches, the features of drugs and their interactions may not effectively collaborate with the known interactions, leading to inaccurate predictions. Therefore, more

advanced computational techniques are needed to improve the prediction of unknown drug interactions (Han et al.,2022). Classification-based models have the advantage of easily integrating multiple types of data, including structural and physicochemical properties, into a unified predictive framework. This allows for a comprehensive analysis of diverse data sources, which can lead to more robust predictions compared to text-based and graph-based methods that may focus on a single data modality.

4.5. Ensemble-based approach

An ensemble-based methodology for DDIs prediction involves integrating predictions from multiple independent models to improve overall accuracy and prediction robustness. After training various algorithms, these models are utilized to predict DDIs for a test set or previously unknown drug pairs. Each model provides a distinct prediction based on its specific methodology and characteristics. For multiple-label DDIs prediction, three key processes are involved, as illustrated in Figure 3. Firstly, a knowledge graph is created using four knowledge graphs established in Bio2RDF, namely DrugBank, KEGG, PharmaGKB, and the Comparative Toxicogenomics Database. Secondly, in addition to the drug knowledge graph, biological DDIs text, consisting of DDIs documents from DrugBank and MEDLINE Abstract is embedded into a low-dimensional vector. Finally, DDIs prediction is performed by effectively using the learned embeddings as a link prediction methodology (Wang et al.,2021).

A newly developed method called the" feature selection-based multi-label k-Nearest Neighbor method" (FS-MLKNN) can identify important feature dimensions and construct high-accuracy multi-label prediction models (Zhang et al.,2015). The ensemble procedure incorporates the Gradient Boosting classifier, Adaboost, and Gaussian Naive Bayes to predict the indications of drug-drug interactions, benefiting from the comprehensive and accurate medicinal data provided by DrugBank (Abbas et al.,2023). Another ensemble model, based on a DNN, was developed to enhance the predictive accuracy of DDIs and successfully identified 86 types of interactions. This ensemble model combined six Machine Learning models, including DNN, Random Forest, and XGBoost, as base models for the classification process. The outcome of the ensemble model was obtained by stacking the results of the individual models using LR classifier (Vo et al.,2023). To develop prediction models that incorporate various types of drug data influencing drug-drug interactions, three representative approaches have been utilized. These approaches consider drug substructure data, drug target data, drug enzyme data, drug transporter data, drug pathway data, drug indication data, drug side effect data, drug offside effect data, and identified drug-drug interactions (Zhang et al.,2017).

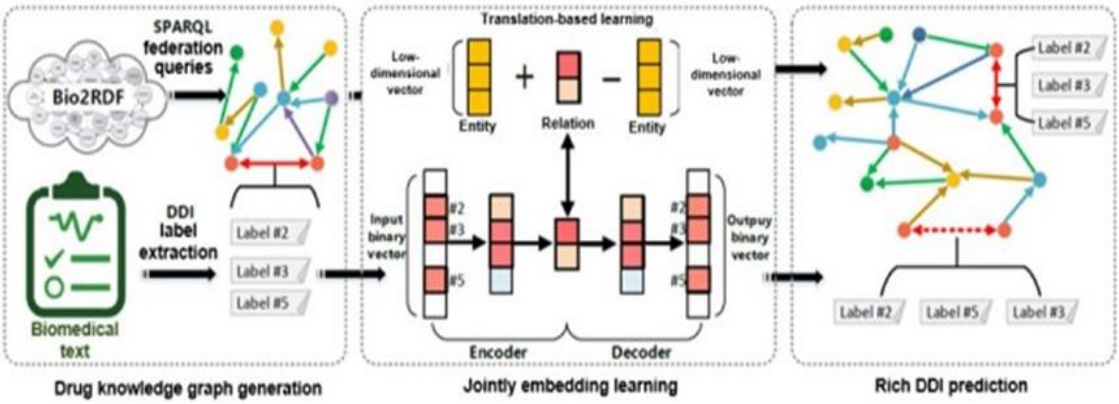

Fig (3) Architecture of the 3 phases: drug KG generation, joint embedding learning, DDIs prediction

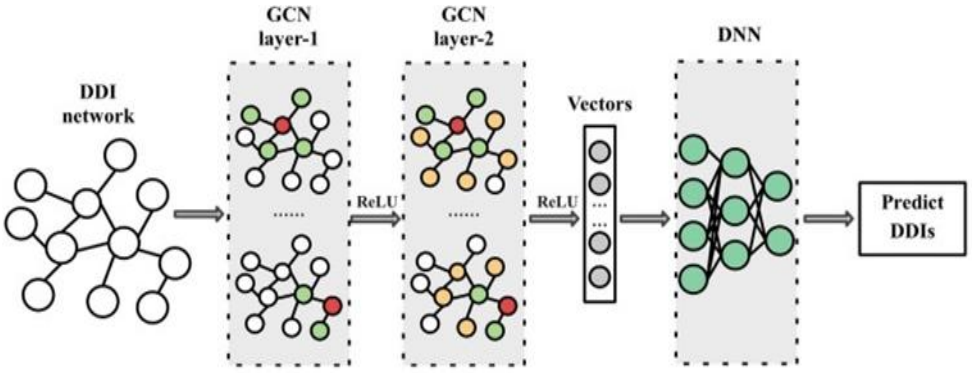

Fig. (4) Ensemble model combining GCN and DNN for DDI prediction

For predicting drug interactions, effective techniques such as Node2vec (a network representation technique for learning), bagging SVM, and positive-unlabeled (PU) learning have been employed. Since drug-drug interactions only contain positively identifiable data, a PU learning-based algorithm is used to generate meta-knowledge from feature networks. Finally, a meta-classifier is developed to integrate the expected interactions based on the learned meta-knowledge (Deepika et al.,2018). Researchers have initiated the integration of various neural networks to forecast potential Drug-Drug Interactions (DDIs), leading to the creation of numerous advanced computing techniques that exhibit superior performance and robustness. The combined utilization of GCN and DNN involves a multi-step process. Initially, GCN is employed to meticulously extract intricate features from the DDIs network.

These extracted features are then seamlessly integrated and forwarded into the DNN network, which subsequently conducts predictive analysis as depicted in **Figure 4** (Luo et al.,2024). A hierarchical classification approach has been developed to extract DDIs from clinical text, making use of the shared DDIExtraction 2013 dataset. Initially, the input document is cleaned using established pre-processing techniques. Subsequently, the most effective features are meticulously selected for training. In the third phase, hierarchical classification is applied using deep learning methods. these models are trained using a variety of algorithms, such as DeepFFNN, ConvFFNN, ConvBLSTM, and ConvStackedLSTM. Remarkably, ConvBLSTM emerges as the most promising model among them (Malik et al.,2024).

5. Conclusion

Machine learning technology indeed has tremendous potential in addressing complex problems, including predicting DDIs, at a lower cost and with the potential to reduce morbidity and toxicity. By leveraging machine learning algorithms and techniques, researchers and healthcare professionals can analyze large datasets, identify patterns, and make accurate predictions regarding potential drug interactions. Different approaches in predicting DDIs provide complementary perspectives that, when combined, can improve the overall accuracy and reliability of DDIs prediction. By integrating multiple approaches, researchers can leverage the strengths of each method and overcome their limitations. This can lead to more effective and personalized treatment plans, improved patient safety, and reduced healthcare costs.

6. Future Work

In our manuscript, we discuss different methods for predicting drug interactions and find that the classification-based approach outperforms other methods. This is because it can effectively combine different types of data, allowing for a more comprehensive analysis. As a result, we plan to prioritize further research and development of the classification-based method in our future work.

References

[1] Scavetta, J.: Prediction of Drug-drug Interaction Potential Using Machine Learning Approaches. Rowan University, (2020)

[2] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., VogtMaranto, L., Zdeborov´a, L.: Machine learning and the physical sciences. Rev. Mod. Phys. **91**, 045002 (2019) https://doi.org/10.1103/RevModPhys.91. 045002

[3] Bumgardner, B., Tanvir, F., Saifuddin, K.M., Akbas, E.: Drug-drug interaction prediction: a purely smiles based approach. In: 2021 IEEE International Conference on Big Data (Big Data), pp. 5571–5579 (2021). IEEE

[4] Vilar, S., Uriarte, E., Santana, L., Lorberbaum, T., Hripcsak, G., Friedman, C., Tatonetti, N.P.: Similarity-based modeling in large-scale prediction of drug-drug interactions. Nature protocols **9**(9), 2147–2163 (2014)

[5] Kastrin, A., Ferk, P., Lesko˘sek, B.: Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. PloS one **13**(5), 0196865 (2018)

[6] Zhang, P., Wang, F., Hu, J., Sorrentino, R.: Label propagation prediction of drug-drug interactions based on clinical side effects. Scientific reports **5**(1), 1–10 (2015)

[7] Richelson, E.: Pharmacology of antidepressants—characteristics of the ideal drug. In: Mayo Clinic

Proceedings, vol. 69, pp. 1069–1081 (1994). Elsevier
[8] Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., Friedman, C.: Drug-drug interaction through molecular structure similarity analysis. Journal of the American Medical Informatics Association **19**(6), 1066–1074 (2012)
[9] Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., *et al.*: Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic acids research **46**(D1), 1074–1082 (2018)
[10] Kanehisa, M.: Toward understanding the origin and evolution of cellular organisms. Protein Science **28**(11), 1947–1951 (2019)
[11] Banda, J.M., Evans, L., Vanguri, R.S., Tatonetti, N.P., Ryan, P.B., Shah, N.H.: A curated and standardized adverse drug event resource to accelerate drug safety research. Scientific data **3**(1), 1–11 (2016)
[12] Wang, Y., Suzek, T., Zhang, J., Wang, J., He, S., Cheng, T., Shoemaker, B.A., Gindulyte, A., Bryant, S.H.: Pubchem bioassay: 2014 update. Nucleic acids research **42**(D1), 1075–1082 (2014)
[13] Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C.F., Whaley, R., Klein, T.E.: An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. Clinical Pharmacology & Therapeutics **110**(3), 563–572 (2021)
[14] Karim, M.R., Cochez, M., Jares, J.B., Uddin, M., Beyan, O., Decker, S.: Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-lstm network. In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, pp. 113–123 (2019)
[15] Kuhn, M., Letunic, I., Jensen, L.J., Bork, P.: The sider database of drugs and side effects. Nucleic acids research **44**(D1), 1075–1079 (2016)
[16] Han, K., Cao, P., Wang, Y., Xie, F., Ma, J., Yu, M., Wang, J., Xu, Y., Zhang, Y., Wan, J.: A review of approaches for predicting drug–drug interactions based on machine learning. Frontiers in pharmacology **12**, 814858 (2022)
[17] Tatonetti, N.P., Ye, P.P., Daneshjou, R., Altman, R.B.: Data-driven prediction of drug effects and interactions. Science translational medicine **4**(125), 125– 3112531 (2012)
[18] Shaker, B., Tran, K.M., Jung, C., Na, D.: Introduction of advanced methods for structure-based drug discovery. Current Bioinformatics **16**(3), 351–363 (2021)
[19] Mo, F., Luo, Y., Fan, D., Zeng, H., Zhao, Y., Luo, M., Liu, X., Ma, X.: Integrated analysis of mrna-seq and mirna-seq to identify c-myc, yap1 and mir-3960 as major players in the anticancer effects of caffeic acid phenethyl ester in human small cell lung cancer cell line. Current Gene Therapy **20**(1), 15–24 (2020)
[20] Vilar, S., Uriarte, E., Santana, L., Tatonetti, N.P., Friedman, C.: Detection of drug-drug interactions by modeling interaction profile fingerprints. PloS one **8**(3), 58321 (2013)
[21] Ferdousi, R., Safdari, R., Omidi, Y.: Computational prediction of drug-drug interactions based on drugs functional similarities. Journal of biomedical informatics **70**, 54–64 (2017)
[22] Song, D., Chen, Y., Min, Q., Sun, Q., Ye, K., Zhou, C., Yuan, S., Sun, Z., Liao, J.: Similarity-based machine learning support vector machine predictor of drug-drug interactions with improved accuracies. Journal of clinical pharmacy and therapeutics **44**(2), 268–275 (2019)
[23] Shi, J.-Y., Gao, K., Shang, X.-Q., Yiu, S.-M.: Lcm-ds: a novel approach of predicting drug-drug interactions for new drugs via dempster-shafer theory of evidence. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 512–515 (2016). IEEE
[24] Zhao, X., Chen, L., Guo, Z.-H., Liu, T.: Predicting drug side effects with compact integration of heterogeneous networks. Current Bioinformatics **14**(8), 709–720 (2019)
[25] Saifuddin, K.M., Bumgardner, B., Tanvir, F., Akbas, E.: Hygnn: Drug-drug interaction prediction via hypergraph neural network. In: 2023 IEEE 39th International Conference on Data Engineering (ICDE), pp. 1503–1516 (2023). IEEE
[26] 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 **10**(11), 3092 (2021)
[27] Long, Y., Pan, H., Zhang, C., Song, H.T., Kondor, R., Rzhetsky, A.: Molecular fingerprints are a simple yet effective solution to the drug–drug interaction problem. drugs **500**, 1 (2022)
[28] Herrero-Zazo, M., Lille, M., Barlow, D.J.: Application of machine learning in knowledge discovery for pharmaceutical drug-drug interactions. In: KDWeb (2016)
[29] Celebi, R., Uyar, H., Yasar, E., Gumus, O., Dikenelli, O., Dumontier, M.: Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. BMC bioinformatics **20**(1), 1–14 (2019)
[30] Feng, Y.-H., Zhang, S.-W., Shi, J.-Y.: Dpddi: a deep predictor for drug-drug interactions. BMC bioinformatics **21**(1), 1–15 (2020)
[31] Feng, Y.-H., Zhang, S.-W.: Prediction of drug-drug interaction using an attention-based graph neural network on drug molecular graphs. Molecules **27**(9), 3004 (2022)
[32] Rohani, N., Eslahchi, C.: Drug-drug interaction predicting by neural network using integrated similarity. Scientific reports **9**(1), 13645 (2019)
[33] Hou, X., You, J., Hu, P.: Predicting drug-drug interactions using deep neural network. In:

Proceedings of the 2019 11th International Conference on Machine Learning and Computing, pp. 168–172 (2019)

[34] Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., Liu, S.: A multimodal deep learning framework for predicting drug–drug interaction events. Bioinformatics **36**(15), 4316–4322 (2020)

[35] Segura-Bedmar, I., Mart´ınez Fern´andez, P., Herrero Zazo, M.: Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). (2013). Association for Computational Linguistics

[36] Mondal, I.: Bertchem-ddi: improved drug-drug interaction prediction from text using chemical structure information. arXiv preprint arXiv:2012.11599 (2020)

[37] Liu, S., Tang, B., Chen, Q., Wang, X., et al.: Drug-drug interaction extraction via convolutional neural networks. Computational and mathematical methods in medicine **2016** (2016)

[38] Wang, M., Wang, H., Liu, X., Ma, X., Wang, B.: Drug-drug interaction predictions via knowledge graph and text embedding: instrument validation study. JMIR Medical Informatics **9**(6), 28277 (2021)

[39] Kolchinsky, A., Lourenc¸o, A., Li, L., Rocha, L.M.: Evaluation of linear classifiers on articles containing pharmacokinetic evidence of drug-drug interactions. In: Biocomputing 2013, pp. 409–420. World Scientific, (2013)

[40] Tari, L., Anwar, S., Liang, S., Cai, J., Baral, C.: Discovering drug–drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. Bioinformatics **26**(18), 547–553 (2010)

[41] Zhang, W., Liu, F., Luo, L., Zhang, J.: Predicting drug side effects by multilabel learning and ensemble learning. BMC bioinformatics 16(1), 1–11 (2015)

[42] Abbas, S., Sampedro, G.A., Abisado, M., Almadhor, A., Kim, T.-H., Zaidi, M.M.: A novel drug-drug indicator dataset and ensemble stacking model for detection and classification of drug-drug interaction indicators. IEEE Access (2023)

[43] Vo, T.H., Nguyen, N.T.K., Le, N.Q.K.: Improved prediction of drug-drug interactions using ensemble deep neural networks. Medicine in Drug Discovery **17**, 100149 (2023)

[44] Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G., Li, X.: Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC bioinformatics **18**, 1–12 (2017)

[45] Deepika, S., Geetha, T.: A meta-learning framework using representation learning to predict drug-drug interaction. Journal of biomedical informatics **84**, 136–147 (2018)

[46] Zhao, Yan, Jun Yin, Li Zhang, Yong Zhang, and Xing Chen. "Drug–drug interaction prediction: databases, web servers and computational models." *Briefings in Bioinformatics* 25, no. 1 (2024)

[47] Luo, Huimin, Weijie Yin, Jianlin Wang, Ge Zhang, Wenjuan Liang, Junwei Luo, and Chaokun Yan. "Drug-drug interactions prediction based on deep learning and knowledge graph: A review." *Iscience* (2024)

[48] Malik, M. Shoaib, Sara Jawad, Syed Atif Moqurrab, and Gautam Srivastava. "DeepMedFeature: An Accurate Feature Extraction and Drug-Drug Interaction Model for Clinical Text in Medical Informatics." *ACM Transactions on Asian and Low-Resource Language Information Processing* (2024)

[49] Zhu, Jing, Chao Che, Hao Jiang, Jian Xu, Jiajun Yin, and Zhaoqian Zhong. "SSF-DDI: a deep learning method utilizing drug sequence and substructure features for drug–drug interaction prediction." *BMC bioinformatics* 25, no. 1 (2024)