A Neural Network on Biomedical Knowledge Graph for the Prediction of Drug-Drug Interactions

Author

IQRA NASEER KUNDI

MS-20 (EE)

00000328701

Supervisor

Dr. SHAHZAD AMIN SHEIKH

DEPARTMENT OF ELECTRICAL ENGINEERING COLLEGE OF ELECTRICAL AND MECHANICAL ENGINEERING NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY ISLAMABAD

AUGUST 2024

THESIS ACCEPTANCE CERTIFICATE

It is certified that final copy of MS/MPhil thesis written by Ms. Iqra Naseer Kundi (Registration No. 00000328701) Entry-2020, of (College of E&ME) has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistake and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC member of the scholar have also been incorporated in the said thesis.

Signature: Shahzad

Name of Supervisor Dr. Shahzad Amin Sheikh

Date: 05 - Sep - 2024

Signature (HoD): _____

Dr. Qasim Umar Khan

Date: 05 - Sep - 2024

Signature (Dean): _____

Brig Dr. Nasir Rashid

Date: 05 SEP 2024

DEDICATION

I would like to dedicate my thesis work to my beloved parents who have always supported me in achieving my goals.

ACKNOWLEDGEMENTS

All praises are for Allah, the most merciful and the most beneficent. First and foremost, thanks to Allah Almighty for His innumerable blessings and divine guidance throughout my life.

A particular thanks to my supervisor Dr. Shahzad Amin Sheikh, and GEC members, Dr. Fahad Mumtaz Malik and Asst. Prof. Kamran Aziz Bhatti, for their excellent guidance and very consistent support during this study. Without their help, I may not be here today. Moreover, I would also like to pay my gratitude to all our faculty members who advised or helped us in any form during the thesis. I also want to appreciate family members who have trusted in me and supporting me in continuing hard work in the most challenging times.

Grateful to all those who helped me, one way or the other in completing my research work and thesis. I consider myself very fortunate that I am surrounded by deft people of this field.

ABSTRACT

Effective drug combination prediction is crucial for the achievement of drug discovery, but it is a challenging task due to drug drug interactions and potential adverse drug reactions. This study presents an innovative technique named DDI-KGAT, which employs attention mechanisms to identify crucial characteristics and interrelationships between drugs and various entities, including targets and genes, using a knowledge graph-based strategy. By leveraging associated relations in the knowledge graph, our model adeptly captures drugs and their potential surroundings, thus extracting semantic relations and higher order structures of graph. The KEGG dataset is utilized in evaluating the model's effectiveness, and is compared to other state of the art techniques. The outcomes demonstrate that KGAT outperforms these methods. Additionally, our approach has several advantages, including simplicity, interpretability, and low-dimensional complexity, making it a favorable tool for accelerating the drug discovery and development. By identifying novel drug combinations with improved efficacy and safety profiles, our approach has the capability to improve the patient outcomes and support safer drug development. Our study highlights the potential of attention mechanisms in knowledge graph-based drug combination prediction, and we believe that KGAT framework has the potential to be a valuable foundation for future research in this field.

Key words: Artificial Intelligence, Attention Mechanisms, Drug-Drug Interactions, Graph Neural Networks, Knowledge Graphs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACC	Accuracy
ADR	Adverse Drug Reaction
AI	Artificial Intelligence
AUC-ROC	Area Under The Receiver Operating Characteristic Curve
AUPR	Area Under The Precision-Recall Curve
CNN	Convolutional Neural Network
DDI	Drug-Drug Interaction
FCNN	Fully Convolutional Neural Network
GAT	Graph Attention Networks
GNN	Graph Neural Network
KEGG	Kyoto Encyclopedia Of Genes And Genomes
KG	Knowledge Graph
KGAT	Knowledge Graph Attention Network
KGNN	Knowledge Graph Neural Network
ML	Machine Learning
NLP	Natural Language Processing
RDF	Resource Description Framework
RNN	Recurrent Neural Network
RSL	Recursive Least Squares
SPARQL	SPARQL Protocol And RDF Query Language
SVM	Support Vector Machine
SWA	Stochastic Weight Averaging
URLs	Universal Resource Locator
XML	Extensible Markup Language

CHAPTER 1: ANALYSIS AND EVALUATION OF THE SUGGESTED FRAMEWORK

This chapter offers a thorough description and analysis of the suggested system. A Neural Network on Biomedical Knowledge Graph for Drug-Drug Interaction Prediction called DDI-KGAT. The background section of the chapter covers the fundamentals of knowledge graphs, drug-drug interactions, hyper-parameters, evaluation parameters, and graph neural networks and graph attention networks.

The goal, scope, and introduction of the suggested system are outlined in this section. While the problem description part outlines the issue that the proposed system is intended to address, the block diagram section offers a visual depiction of the architecture of the proposed system. The particular aims of the research study are outlined in the objectives section.

The datasets and software tools used in the research project are listed in the resources used part, and the suggested system's importance in the biomedical area is highlighted in the applications and national need section.

In general, this chapter provides readers with a comprehensive overview of the origin, purpose, and relevance of the proposed system.

1.1. Background

1.1.1. Graph Neural Networks

In the discipline of deep learning, Graph Neural Networks (GNNs) have revolutionized the processing of graph-structured data. GNNs, in contrast to typical neural networks, are capable of handling non-Euclidean data, such as knowledge graphs, protein structures, and social networks, with different sizes, connectedness, and edge weights. GNNs are well-suited for tasks such as node classification, link prediction, and graph classification because they can learn a low-dimensional vector representation for each node in the graph, enabling them to capture intricate structural connections between nodes. GNNs take the place of older neural networks in these tasks because they produce predictions that are more comprehensible and accurate than those of classic neural networks.

Generally, GNNs are constructed with multiple layers that iteratively improve node representations. A message-passing method is used for this refinement process, in which each node sends and receives messages according to its own local features as well as the features of its neighbors. Activation functions and neural network layers are examples of non-linear changes that come after message aggregation in GNN computations. Each node's vector representation, which is the output of a GNN, can be utilized for a variety of downstream tasks, including graph-level prediction, link prediction, and node categorization. Overall, GNNs have demonstrated significant potential in numerous domains such as bioinformatics, social networks, and recommender systems, rendering them an effective and adaptable instrument for handling data with a graph structure. The ability of GNNs to handle non-Euclidean data and capture complex structural dependencies between nodes makes them a valuable addition to the deep learning toolbox.

1.1.2. Graph Attention Networks

During the message-passing process, Graph Attention Networks (GATs), a class of GNNs, use attention processes to determine the relevance of surrounding nodes. In traditional GNNs, the importance of each neighboring node is considered to be equal, regardless of its relevance to the target node. This can lead to suboptimal performance when dealing with large and complex graphs. GATs address this limitation by allowing each node to attend to its neighbors

selectively, based on the similarity of their features. An attention mechanism that learns a weight for each neighbor based on their feature vectors and the feature vector of the target node is used to do this. Then, using the attention weights, a weighted sum of the neighbor representations is calculated and added to the representation of the target node to create a new representation for the node.

The architecture of GATs is made up of several attention heads that are trained on distinct attention functions, giving each node's neighbors a variety of ways to receive attention. As a result, the model is able to learn more expressive representations and capture many facets of the graph structure. Each node's vector representation is the GAT's output, and it can be utilized for a number of downstream activities, including graph-level prediction, link prediction, and node classification.

In a number of applications, such as social networks, bioinformatics, and recommendation systems, GATs have demonstrated notable gains over conventional GNNs. Large-scale graph processing activities might benefit from their computational efficiency and capacity to capture intricate structural relationships between nodes. All things considered, GATs are a potent extension of GNNs that allow for more accurate and expressive modeling of graph-structured data.

1.1.3. Knowledge Graphs

One kind of structured graph data that depicts the relationships between things in a domain is called a knowledge graph (KG). They are built using a blend of expert curation and machine learning approaches, producing a carefully curated and superior knowledge base. Knowledge graphs (KGs) have become an important tool for reasoning and describing complicated knowledge areas like e-commerce and the biological sciences. Entities and relationships—represented as nodes and edges, respectively—are found in KGs. While edges show links between things, such as "is a subclass of" or "causes," nodes can have characteristics or attributes, such as a person's name or a disease's description. KGs can handle uncertainty, incompleteness, and inconsistencies, which is one of its many advantages over traditional databases. They are frequently utilized in many different applications, such as question-answering systems, recommender systems, and search engines. The need for KG-based solutions is growing across

multiple fields due to the exponential expansion of digital data, making this an essential area of study and development. KGs are especially crucial in the biomedical field for the identification of possible therapeutic targets, the study of disease mechanisms, and the discovery of new medications. They make it possible to combine information from many biomedical data sources, which opens up new avenues for research into intricate biological systems and the discovery of previously unknown linkages.

1.1.4. Drug-Drug Interactions

When two or more medications interact with one another in a way that modifies their safety or efficacy, this is known as a drug-drug interaction (DDI). Numerous processes, including as pharmacokinetic, pharmacodynamics, or pharmacogenetics interactions, may be involved in these interactions. When a medication influences the distribution, metabolism, excretion, or absorption of another medication, alterations in the drug's plasma concentration and therapeutic impact result in pharmacokinetic interactions. Pharmacogenetics interactions occur when genetic variations affect the metabolism or response to a drug, leading to different outcomes in different individuals. DDIs can have significant clinical consequences, such as increased toxicity, decreased efficacy, or unexpected adverse events. Thus, for safe and efficient medication therapy, DDI prediction and management are essential. Using these methods for DDI prediction and mitigation is becoming more popular as better machine learning models, like graph neural networks, become available and biomedical knowledge graphs become more widely available.

1.1.5. Hyper-parameter

Hyper parameters are aspects of a machine learning model that are predetermined and don't change during the training process. These parameters need to be carefully adjusted to get the best results because they have a big impact on the model's performance. You have established a number of hyper parameters for your model in your study, which comprise:

Embedding dimension:

The dimension of the vector that each node in the graph is represented by.

Neighborhood sample size:

The number of neighbors to take into account while aggregating data from a node's nearby neighborhood.

L2_weight:

The degree to which the model's weights are subjected to the L2 regularization penalty in order to avoid overfitting.

Learning rate:

The size of the step at which the model's parameters are updated using stochastic gradient descent.

Optimizer type:

An algorithm, like Adam or SGD that modifies the model's parameters while it is being trained.

Batch size:

The quantity of samples the model processes throughout a training cycle.

Aggregator type:

The technique—such as mean or max pooling—that is utilized to compile data from a node's immediate vicinity.

Regularization strength:

The degree to which the model's parameters are subjected to a regularization penalty in order to avoid overfitting.

Number of epochs:

The total number of times the model runs over the whole training set while being trained.

One can maximize the performance of your model and make sure it performs well when applied to new data by carefully adjusting these hyper parameters.

1.1.6. Evaluation parameter

Measures known as evaluation parameters are employed to evaluate a model's performance on a particular task. Evaluation metrics are used in the context of machine learning to assess a model's predicted performance. Four widely-used evaluation criteria were employed in this study: accuracy (ACC), F1 score, area under the precision-recall curve (AUPR), and area under the receiver operating characteristic curve (AUC-ROC).

Accuracy (ACC):

The ratio of accurately predicted samples to the total number of samples. It is a straightforward metric that expresses the proportion of accurate predictions the model makes.

F1 score:

The harmonic mean of recall and precision. When datasets are unbalanced means, this evaluation metric is frequently employed. This is a useful indicator for assessing models on unbalanced datasets since it takes precision and recall into account.

Area under the Receiver Operating Characteristic Curve (AUC-ROC):

A binary classifier performance is assessed using this. AUC-ROC plots the true positive rate (TPR) versus the false positive rate (FPR) at different categorization thresholds. The model's capacity to discriminate between positive and negative classes is also measured by AUC-ROC.

Area under the Precision-Recall Curve (AUPR):

A measure employed in the assessment of binary classifier performance. Plotting accuracy vs recall at different classification thresholds is what the precision-recall curve shows. AUPR is a useful metric for imbalanced datasets since it evaluates the trade-off between precision and recall.

1.2. Introduction

When used concurrently, certain medications have the potential to interact and alter one or both of the medications' effectiveness. We refer to this kind of interaction as a drug-drug interaction (DDI). Adverse drug responses (ADRs) [1], decreased drug efficacy, and even potentially fatal diseases can result with DDIs. ADRs can have unanticipated pharmacological consequences and possibly be fatal, which makes them a serious worry in the healthcare industry. ADRs are the fourth greatest cause of death in hospitalized patients, with a fatality rate of 0.32% [10], indicating that the overall prevalence of significant ADRs may be substantially higher than anticipated, according to recent studies [2].

The more medicines a person consumes, the higher their chance of developing DDIs. Over 10% of people take five or more prescriptions at the same time, according to the U.S. Centers for Disease Control and Prevention, and the problem gets worse for older adults—roughly 20% of them use ten or more medications at a time [3]. This greatly increases the likelihood of having ADRs. In order to reduce unexpected ADRs and maximize therapy synergy to lessen the impact of unanticipated pharmacological effects, it is imperative that potential DDIs be effectively identified [4].

The mainstay of current DDI prediction techniques is the integration of several data sources to extract drug properties, including multi-task learning [7], adverse or side effect [6], and similarity features [5]. These techniques operate under the premise that medications with comparable representations will also have comparable DDIs. Machine learning approaches have emerged as a viable method for predicting drug-disease interactions (DDIs) from a variety of sources, including drug databases, scientific literature, and electronic health records. However, the majority of computer models that are now in use that make use of AI techniques tend to concentrate more on merging various data sources and including widely used embedding techniques [8], with less emphasis on possible connections between medications and other entities like targets and genes. Although some recent studies have utilized KGs for DDI prediction, these methods only learn node hidden embedding directly, limiting the ability to obtain comprehensive neighborhood details for every entity in the KG.

Therefore, this research aims to address these research gaps by leveraging comprehensive neighborhood details for every entity in KG and employing the knowledge graph attention network (KGAT), a technique that has not been previously used for DDI prediction despite its popularity in various applications. The objectives of this research are to demonstrate the effectiveness of KGs and GNN in solving complex prediction tasks, to provide a simple and innovative approach that can potentially increase the prediction accuracy of DDIs with low-dimensional complexity, and to improve reliability and the accuracy of DDI prediction, potentially leading to better patient outcomes and safer drug development.

To achieve these objectives, this thesis proposes a novel KGAT model that utilizes both attention mechanisms and sum aggregation layers to capture important features and enhance performance of DDI prediction. The Kyoto Encyclopedia of Genes and Genomes (KEGG) dataset is used to evaluate the suggested model, and the outcomes are compared to state-of-the-art methods. This research advances the field of DDI prediction by proving the efficacy of this method and offering a novel, straightforward technique that may improve the prediction accuracy of DDIs with low dimensional complexity. Additionally, this research highlights the importance of leveraging comprehensive neighborhood details for every entity in the KG and employing KGAT for DDI prediction.

Overall, Because ADRs are a major cause of morbidity and mortality in healthcare, accounting for an estimated 7k deaths annually [9], and because they can be accurately predicted, this research has the potential to both improve patient outcomes and aid in the development of safer drugs.

1.3. Block Diagram

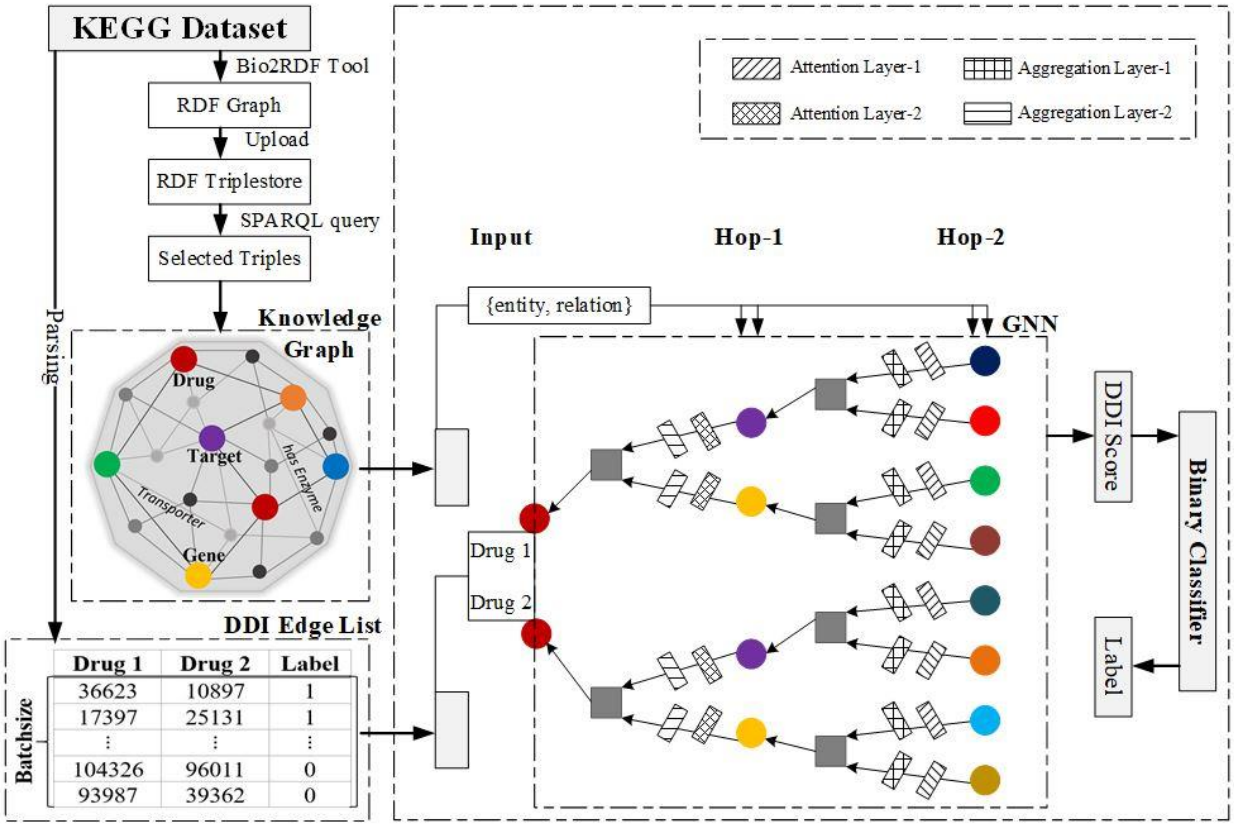

FIGURE 1: BLOCK DIAGRAM

1.4. Problem Statement

DDIs are a major concern in healthcare as they can lead to ADRs and potentially life-threatening outcomes. While traditional methods for identifying DDIs rely on experimental studies, these methods are time-consuming, expensive, and often lack comprehensive coverage of all possible drug combinations. More and more biomedical literature and electronic health data are becoming available, which has increased the demand for precise and effective computational techniques to identify possible DDIs. Nevertheless, current computational models suffer from a number of drawbacks, including a failure to consider plausible associations between medicines and other things and an incapacity to gather thorough neighborhood information for each object in the knowledge network. Therefore, for better DDI prediction, a knowledge graph-based model that can accurately reflect the intricate interactions between medications and other things needs to be developed.

1.5. Objectives

- To show how well knowledge graphs and graph neural networks perform at handling challenging prediction problems..
- To offer a novel and straightforward method that may improve the accuracy of drug-drug interactions predictions with low dimensional complexity.
- To take advantage of the extensive neighborhood data for every entity in the KG, since most approaches just learn the latent embedding of nodes directly, which yields poor accuracy results.
- To employ the knowledge graph attention network, a technique that has not been previously used for DDI prediction despite its popularity in various applications.

1.6. Resources Used

The present study utilized various resources to conduct the research. The first resource used was the virtual access to the supercomputer of the university. Through remote desktop connection, we were able to log in to their account and access the computing resources provided by the supercomputer. Following a successful login, the open-source Python distribution Anaconda was installed. It combines a number of packages and libraries to offer a platform for data science and machine learning applications. A virtual environment with the project-specific dependencies was established in the Anaconda environment. Versions of Python 3.6.6, Keras 2.3.0, TensorFlow 1.13.1, and scikit-learn 0.22 were among these dependencies. Popular deep learning frameworks TensorFlow and Keras are used to create and train neural networks, and scikit-learn is a machine learning library used for a variety of tasks like clustering, regression, and classification. Once the virtual environment was set up, we installed Spyder, an open-source integrated development environment (IDE), which provided a convenient interface for editing, running, and debugging the project files. Together, these resources provided the necessary tools and infrastructure for conducting the research and developing the KG based model for DDI prediction.

1.7. Applications and National Need

The applications and national need for this study are significant and far-reaching. The current ranking of Pakistan at 154/195 in the year 2021 Global health security Index indicates a dire need for advancements in healthcare technologies. This study aligns with national needs by providing a solution for improving human health through the development of a knowledge graph centered model for drug-drug interaction prediction. This model offers several advantages over traditional biomedical methods as it is less costly, less time-consuming, and has higher scalability.

The area of application for this study is vast and includes various fields such as biomedicine, bioinformatics, computational biology, and health informatics. The results of this study will have applications in personalized medicine, pharmacovigilance, adverse drug reaction monitoring, and drug development and discovery. The application of knowledge graphs in drug discovery has the potential to transform the sector and enhance patient outcomes.

CHAPTER 2: LITERATURE REVIEW

An extensive evaluation of the relevant research in the area of drug-drug interaction prediction is given in the literature review chapter. An overview of the different methods used to predict drug-drug interactions is given at the beginning of the chapter. These methods include neural network, ensemble-based, matrix factorization, graph embedding, literature-based, similarity-based, and classification-based, link prediction/random walk, neural network, and knowledge graph-DDI methods. In addition, the chapter examines earlier studies that were especially connected to the KGAT model.

Moreover, the chapter also emphasizes the necessity for creating a more precise and effective model for forecasting drug-drug interactions and points out research gaps in the body of current knowledge. The literature evaluation offers insight into the state of the art in the field of drug-drug interaction prediction and forms the basis for the proposed research project.

2.1. Related Work

DDIs can have serious consequences for patients, including reduced therapeutic efficacy and increased risk of ADRs. Therefore, predicting potential DDIs is crucial for ensuring patient safety. Various approaches have shown promise in predicting DDIs by leveraging various data sources. In this literature review, we have discussed 9 different approaches for DDI prediction and a general KGAT approach on different applications as this approach has not been implemented for DDI prediction.

2.1.1. Similarity-Based

Pairwise similarities between medications are used by similarity-based approaches to predict DDIs. Based on a variety of characteristics, including pharmacological characteristics and chemical structure, these similarities can be computed. Similarity-based techniques are simple to use and comprehend. An overview of medications and their targets is given by Imming et al. (2006) [11], who stress the significance of comprehending drug-target interactions in drug discovery. The research emphasizes how difficult it is to forecast DDIs because of the variety of molecular targets and the intricacy of biological systems. The FDA's viewpoint on DDI prediction is covered by Zhang et al. (2009) [12], along with the significance of spotting possible interactions at an early stage of the drug development process. A summary of the in vitro and in vivo techniques for DDI evaluation is also included in the study. A computational system named INDI is introduced by Gottlieb et al. (2012) [13] in order to predict medication interactions and recommendations related to them. In order to forecast DIs, INDI blends drug similarity measurements based on drug-target interactions and side effects. A strategy for DDI prediction based on molecular structure similarity analysis is put out by Vilar et al. (2012) [14]. The method described in this research represents the structure of pharmaceuticals using molecular fingerprints, and then evaluates the similarity of drug pairs based on these fingerprints. A computational approach for DDI prediction based on functional similarity across medications is developed by Ferdousi et al. (2017) [15]. The model predicts DDIs by estimating the functional similarity of medications based on gene ontology annotations and drug target information.

The similarity-based method may not take into consideration context-specific elements that affect DIs and has limited capacity to capture complicated interactions and dependencies between medicines and DDIs.

2.1.2. Classification-Based

Classification-based approaches learn a classification model that can predict DDIs for novel drug combinations using a training set of known DDIs. These techniques can distinguish between different kinds of DDIs and manage more intricate medication interactions. Using a classification-based methodology, Li et al. (2015) [16] carried out a thorough investigation and analysis of medication combinations. They discovered new DIs by using ML algorithms to forecast DDIs based on the chemical and biological characteristics of medications. A unique method, called LCM-DS, for DDI prediction utilizing the Dempster-Shafer theory of evidence was proposed by Shi et al. (2016) [17]. They obtained encouraging findings by predicting DDIs for novel medications using a classification-based method. A statistical learning-based method was created by Kastrin et al. (2018) [18] to forecast possible DDIs. They used topological plus semantic parallel features and applied ML algorithms to classify DIs as either interacting or non-interacting. A compact integration of heterogeneous networks for the prediction of pharmacological side effects was proposed by Zhao et al. (2019) [19]. They created a classification-based strategy based on machine learning algorithms to forecast the probability of a medicine having a specific side effect based on gene ontology annotations, drug-target interaction networks, and protein-protein interaction networks.

In order to attain high accuracy, the Classification-Based approach needs a large and diversified training set of known DDIs, which is restricted by the quality and quantity of available data. There's a chance that it will over fit to training data, which would lead to inadequate generalization to fresh data.

2.1.3. Literature-Based

Drug pairings and the interactions they are linked with are identified by literature-based approaches that extract DDI data from scientific publications and apply natural language processing (NLP) techniques. These techniques are able to make use of large volumes of unstructured text data. In order to find DDIs, Tari et al. (2010) [20] suggested using text mining

and reasoning to examine drug metabolism characteristics. The approach was based on searching scientific literature for words that provided evidence of medication interactions, after which logical criteria were applied to deduce possible interactions. Tatonetti et al. (2012b) [21] created a data-driven method that makes use of electronic health information to forecast medication interactions and effects. To forecast novel interactions and adverse events, their approach combined DDIs, drug-gene interactions, and phenotypic side effects.

Kolchinsky et al. (2013) [22] assessed how well linear classifiers performed on publications that included DDI pharmacokinetic findings. They contrasted several feature extraction techniques and classifiers and showed that employing literature-based methods to predict DDIs is feasible.

Literature-based methodologies depend on the reliability and accessibility of literature sources, which can be unreliable, prejudiced, or omit crucial information that isn't made clear in the literature.

2.1.4. Link Prediction / Random Walk

Graph theory is used by link prediction methods to forecast new drug connections based on a drug's connectedness within a drug distribution network. Indirect medication interactions can also be found using random walk-based techniques. A pharmacointeraction network model was presented by Cami et al. (2013) [23] to predict unknown drug-target interactions (DDIs) by combining several data sources, such as drug indications, side effects, and drug-target interactions, into a single network representation. DeepWalk is an online learning method that was presented by Perozzi et al. (2014) [24]. It learns node embeddings by using the Skip-gram model from natural language processing and using random walks as sentences. They learned node embeddings in social networks using DeepWalk, and they attained state-of-the-art results on multiple common benchmarks.A drug recommendation approach based on a combination of ensemble learning and matrix factorization was presented by Zhang et al. (2016) [25]. By taking use of the similarities between medications and past preferences, their approach forecasts possible adverse effects of pharmaceuticals. A technique for forecasting pharmacodynamic drug-drug interactions was put out by Park et al. (2015) [26] through an examination of signaling interference propagation on protein-protein interaction networks. Their strategy is predicated on the idea that medications that

disrupt the same protein are more likely to interact with one another. A probabilistic method for forecasting DDIs was presented by Sridhar et al. (2016) [27] by simulating the overall similarity of medications. To capture the joint distribution of drug pairs, side effects, and chemical properties, they employed a generative model and inference to identify potential drug interactions. Struc2vec is a technique for learning node embeddings that extracts structural information from graphs, as described by Ribeiro et al. (2017) [28]. By taking into account a node's structural identity—which is defined as the collection of neighborhoods around it with progressively larger radii—their method generalizes DeepWalk. An integrated drug similarity network technique (IDNDDI) was introduced by Yan et al. (2019) [29] to predict drug-drug interactions. By combining information from several sources, like as chemical structures, targets, and side effects, they created a drug similarity network. Then, they used a random walk-based method to forecast possible DIs.

Large networks may provide scalability problems for these investigations, and their scope is restricted by the accuracy and completeness of the network data. Moreover, they might miss intricate relationships and interactions between medications and DDIs.

2.1.5. Neural Network

Deep learning techniques are used by neural network-based approaches to learn intricate drug representations and forecast DDIs. These techniques are capable of processing sizable and intricate data sets and capturing intricate drug interactions. LINE (Large-scale Information Network Embedding) is a scalable method for network embedding that maintains the first- and second-order proximity of nodes. It was first presented in a publication by Tang et al. (2015) [30]. LINE is appropriate for large-scale network applications because it uses both the local and global network topologies to learn node embeddings. Wang et al. (2016) [31] describe a unique method for learning the network representation termed Structural Deep Network Embedding (SDNE) in their paper. SDNE is a deep autoencoder that uses deep neural networks to learn non-linear representations while maintaining both first- and second-order closeness of nodes. The study shows that SDNE performs better on a variety of tasks than other cutting-edge network embedding techniques. A Variational Graph Autoencoder (VGAE) model for node clustering is proposed by Mrabah et al. (2016) [32]; this model circumvents the problem of feature twisting in the latent space of conventional graph autoencoders. VGAE employs the reparameterization method for effective training, modeling the probability distribution of latent variables. The study demonstrates

that VGAE can successfully cluster nodes in different real-world datasets by evaluating the suggested model.

Neural network techniques are computationally costly, needing enormous quantities of processing power and time to train, as they may suffer from overfitting and require a large amount of data to train. It could be challenging to decipher and comprehend the model's prediction-making process.

2.1.6. Ensemble-Based

Several ML models are combined in ensemble-based strategies to increase prediction accuracy. These techniques are capable of handling various data formats and capturing various facets of the DDI issue. The first study, by Zhang et al. (2017) [33], integrates phenotypic, chemical, biological, and network data to provide an ensemble-based method for DDIs. The suggested approach uses an ensemble approach to aggregate the output of multiple machine learning methods, such as logistic regression, decision trees, and random forests. The second paper, A meta-learning approach for representation learning-based drug-drug interaction prediction is presented by Deepika and Geetha (2018) [34]. Learning a meta-learner that can forecast the best classifier for a given dataset is the suggested approach. The technique leverages a number of well-known representation learning methods, such as autoencoders and deep belief networks, and trains a meta-learner to select the best classifier for a particular dataset based on the features that the algorithms have learnt.

These techniques run the danger of overfitting if individual models have a high degree of correlation, can be challenging to integrate numerous models, and may demand a large amount of processing power.

2.1.7. Matrix Factorization

Drugs and interactions are represented in lesser dimensions when a DI matrix is broken down using matrix factorization techniques. Next, new interactions can be predicted using these representations. These techniques can capture latent correlations between medications and handle missing data. A technique for dimensionality reduction and data representation known as Laplacian eigenmaps is presented in Belkin and Niyogi's 2003 [35] study. It preserves the local

structure of high-dimensional data points by embedding them into a low-dimensional space using graph theory. This method has been extensively used in many domains, including medication development. A technique for identifying possible DDIs through interaction profile fingerprint modeling is presented by Vilar et al. (2013) [36]. To find possible interactions, the method entails creating binary interaction profiles for every medication and comparing them. This approach works well in identifying potential interactions between drugs that have not been previously reported. Grarep is a matrix factorization technique that Cao et al. (2015) [37] propose for learning network representations by utilizing global structural information. This method has been applied to drug discovery to forecast putative targets for drugs since it can capture intricate interactions between nodes in a graph. For DDI prediction, Zhang et al. (2018) [38] suggest a manifold regularized matrix factorization method. This approach has demonstrated promising results in predicting hitherto unidentified interactions and uses manifold regularization to better capture the underlying structure of DDIs. A semi-nonnegative matrix factorization method is suggested by Yu et al. (2018) to forecast and comprehend comprehensive DDIs. This method makes use of similarity matrices for both medicines and targets in order to capture the intricate relationships between them. TMFUF, a triple matrix factorization-based unified framework for predicting complete DDIs of novel medications, is proposed by Shi et al. (2018) [40]. To more correctly forecast DDIs, this approach incorporates many data sources, such as target similarity, side effect similarity, and drug chemical structure. A technique for DDI detection via artificial neural networks and traditional graph similarity metrics is put forth by Shtar et al. (2019) [41]. Compared to conventional methods, our methodology is more accurate in capturing the complicated relationships between pharmaceuticals and predicting probable interactions. For DDI prediction, Rohani et al. (2020) [42] suggest an integrated similarity-constrained matrix factorization method. This method has demonstrated better accuracy over existing matrix factorization algorithms by utilizing both target and drug chemical structure similarities to forecast possible interactions.

Matrix Factorization based techniques are dependent on the completeness and quality of the input data and do not work well with sparse data sets since they may not capture complicated interactions and dependencies between medications and DDIs.

2.1.8. Graph Embedding

Drugs and interactions within a DDI network are represented in low dimensions by graph embedding techniques. Next, new interactions can be predicted using these representations. These techniques can handle incomplete data and capture intricate drug connections. A deep learning method for forecasting drug-drug and drug-food interactions was presented by Ryu et al. (2018) [43]. This method makes use of a variety of data sources, such as chemical and genomic data. In order to forecast DDIs, Ma et al. (2018) [44] suggested an attentive multi-view graph autoencoder model that combines drug similarity data from several sources. A multi-modal deep autoencoder technique was presented by Liu et al. (2019) [45] to embed structural networks and predict DDIs. In order to increase the precision of DDI prediction, Lee et al. (2019) [46] developed a novel deep learning (DL) model that incorporates a variety of data sources, such as side effect profiles, genetic data, and chemical structures. A deep neural network (DNN) model for predicting drug-dose interactions (DDIs) was presented by Hou et al. (2019) [47]. This model makes use of several drug properties, such as chemical structures, side effects, and drug targets. In order to predict drug-drug interactions (DDIs), Deng et al. (2020) [48] developed a multimodal DL framework that incorporates several sorts of data, such as chemical structures, drug side effects, and drug-target interactions. A DL model called DPDDI was presented by Feng et al. (2020) [49] to predict DDIs by incorporating several pharmacological properties, such as chemical structures, drug targets, and side effects. A thorough analysis of GNNs, a class of DL models made for graph data, including graphs utilized in DDI prediction, was provided by Wu et al. (2021) [50].

Studies that rely on graph embedding could miss context-specific data regarding medications and DDIs. They cannot function successfully in vast and complicated networks because they are constrained by the completeness and quality of the input data and graph structure.

2.1.9. Knowledge Graph-DDI

DDIs are represented as a KG in knowledge graph-based approaches, which employ graph-based algorithms to forecast future interactions. These techniques can manage missing data and capture intricate drug connections. Using linked open data, Celebi et al. (2019) [51] assessed several KG embedding techniques for DDI prediction. They discovered that TransE performed better than alternative strategies and that performance was enhanced by using connected open data.

A DDI prediction technique based on KG embeddings and a convolutional-LSTM network was presented by Karim et al. (2019). Their method outperformed baseline models, and they demonstrated that the prediction accuracy increased with the inclusion of contextual information. SumGNN, a technique for multi-typed DDI prediction via effective KG summarization, was presented by Yu et al. (2021) [52]. On two benchmark datasets, they demonstrated that their method reached state-of-the-art performance and outperformed existing approaches. BioDKG-DDI, a technique for forecasting DDIs based on drug KG including biochemical data, was introduced by Ren et al. (2022) [53]. In comparison to cutting edge techniques, their strategy produced competitive results by combining several forms of data. They also showed how well their approach worked for locating possible drug-drug interactions (DDIs) between medications used to treat COVID-19.

In addition to needing high-quality data and possibly having scalability problems, KG techniques for DDI may miss new or developing DIs that aren't clearly reflected in the KG. The accuracy and comprehensiveness of the KG and related data sources also place restrictions on these strategies. The ability of KGs to integrate and depict various biological data types has been demonstrated by their recent breakthroughs in drug discovery. A Knowledge Base (KG) is a type of structured, semantic database used to store information about items and their relationships. Through the use of KGs, scientists may combine information from multiple sources, like as proteomics, metabolomics, and genomes, to produce a holistic picture of the biological system. In drug discovery, KGs can be used to forecast DTIs, DDIs, and drug.

A more recent advancement in the field of KGs that can be applied to drug discovery prediction is Knowledge Graph Neural Networks, or KGNNs. KGNNs are a kind of neural network that uses node characteristics and graph topology together to generate predictions. Among other things, they can be used to forecast DTIs, DDIs, and medication efficacy. Nevertheless, KGNNs still have a number of drawbacks, including the challenge of managing large-scale KGs and the requirement for a substantial quantity of labeled data.

In summary, a variety of strategies exist for predicting DTIs and DDIs, such as methods based on similarity, classification, literature, and link prediction/random walk. Recent advancements in KGs have demonstrated their potential in drug discovery, and KGNNs have become a viable method for prediction in this domain.

2.1.10. KGAT (GENERAL)

In order to forecast DDIs, a deep learning technique called knowledge graph attention network (KGAT) combines neural network- and knowledge graph-based techniques. Additionally, KGAT is capable of handling partial data and capturing intricate drug connections. Nevertheless, no study has employed the KGAT for DDI prediction, despite its widespread use in other areas. This general review of studies from various fields of applications is provided here in order to get insight into the process of applying attention mechanisms to pharmacological knowledge graphs. KGAT, a knowledge graph attention network, was proposed by Wang et al. (2019) [54]. It learns to recommend things by capturing both entity and relation information in a KG. They used real-world recommendation datasets to show how effective their approach is. DETERRENT, a knowledge-guided graph attention network for identifying healthcare disinformation, was introduced by Cui et al. (2020) [55]. Incorporating domain-specific knowledge into the model with a KG, they demonstrated how their approach works better than current techniques. Message-aware GAT for large-scale multi-robot path planning was introduced by Li et al. (2021) [56]. They presented a novel attention mechanism that takes into account the messages delivered from each edge to its nearby nodes, and they used a difficult robotic path planning challenge to show how effective their approach is. DisenKGAT, a KG embedding method with a disentangled graph attention network, was introduced by Wu et al. (2021) [57]. In order to better capture the diverse information in the KG, their method disentangles the embeddings of various relation types and employs graph attention to capture the relations between entities. KGANCDA, a strategy for forecasting circRNA-disease correlations based on a KGAT, was proposed by Lan et al. (2022) [58]. They demonstrated how their method works better than current approaches by representing the relationships between circRNAs and diseases using a heterogeneous knowledge network. For knowledge graph completion, Zhang et al. (2023) [59] created a GAT with dynamic relation representation. Their method changes the relation embeddings dynamically based on the learnt importance by using attention processes to learn the relative importance of various relation kinds. They used a number of benchmark datasets to show how effective their strategy was.

The salient features of the research examined in this study are enumerated in the following table:

Table 1: Literature Review of DDIs

Category	Papers	Description	Limitations
Similarity-Based	Imming et al. (2006) [11], Zhang et al. (2009)[12], Gottlieb et al. (2012)[13], Vilar et al. (2012)[14], Ferdousi et al. (2017) [15].	ML algorithms, which rely on similarity measures, are used in these approaches to predict DDIs. The algorithms employ similarity measures, such as Jaccard index, cosine similarity, and Tanimoto coefficient, to compare the drug features and predict DDIs.	• Limited ability to capture complex relationships and dependencies between drugs and DDIs. • May not account for context-specific factors that impact drug interactions.
Classification-Based	Li et al. (2015)[16], Jian-Yu et al. (2016)[17], Kastrin et al. (2018)[18], Zhao et al. (2019)[19].	The enlisted methodologies involve the classification of drugs into different categories based on their pharmacological properties and therapeutic effects using ML algorithms such as SVMs.	• Limited by the quality and quantity of available data. • Require a large and diverse training set of known DDIs to achieve high accuracy. • Risk of overfitting to training data, which can result in poor generalization to new data.
Literature-Based	Tari et al. (2010)[20], Tatonetti et al. (2012b)[21], Kolchinsky et al. (2013)[22].	These studies involve NLP techniques to extract drug interaction information from text sources, followed by text or statistical analysis techniques to establish drug relationships and ML techniques to predict the unknown drug interactions.	• Relies on availability and accuracy of literature sources, which may be incomplete or biased. • May miss important information that is not explicitly mentioned in the literature. • Suffer from low accuracy due to noise and variability in language use.
Link Prediction / Random Walk	Cami et al. (2013)[23], Perozzi et al. (2014)[24], Zhang et al. (2015)[25], Park et al. (2015)[26], Sridhar et al. (2016)[27], Ribeiro et al. (2017)[28], Yan et al. (2019)[29].	Drugs are represented as nodes, and their connections and interactions as edges to predict unknown interactions. The label propagation, recursive least squares (RSL), graph traversal and random walks strategies are employed for link prediction in these techniques.	• Limited by the quality and completeness of network data. • May suffer from scalability issues with large networks. • May not capture complex interactions and dependencies between drugs and DDIs.
Neural Network	Tang et al. (2015)[30], Wang et al. (2016)[31], Kipf et al. (2016)[32].	Such approaches involves the use of deep learning algorithms, such as CNNs and RNNs, to predict DDIs. These algorithms use the features of drugs and their interactions to predict DDIs.	• Training requires large no of data. • Computationally intensive, requiring large amounts of computing power and time to train. • Can be difficult to interpret and understand how the model is making predictions.

Ensemble-Based	Zhang et al. (2017)[33], Deepika et al. (2018)[34].	The techniques were used for DDI prediction by combining multiple types of features and models, such as similarity-based features and deep neural networks.	• Can be difficult to integrate multiple models and may require significant computational resources. • Risk of overfitting if individual models are highly correlated
Matrix Factorization	Belkin et al. (2003)[35], Vilar et al. (2013)[36], Cao et al. (2015)[37], Zhang et al. (2018)[38], Yu et al. (2018)[39], Shi et al. (2018)[40], Shtar et al. (2019)[41], Rohani et al. (2020)[42].	The proposed algorithms employ a decomposition process on the known DDI matrix into several potential matrices, which are constrained by collective similarity. Then, the potential matrix is reconstructed to obtain a new low-dimensional interaction matrix.	• May not capture complex interactions and dependencies between drugs and DDIs so do not perform well with sparse datasets. • Limited by the quality and completeness of the input data.
Graph Embedding	Ryu et al. (2018)[43], Ma et al. (2018)[44], Liu et al. (2019)[45], Lee et al. (2019)[46], Hou et al. (2019)[47], Deng et al. (2020)[48], Feng et al. (2020)[49], Wu et al. (2021)[50].	Such methods involve the use of graph embedding algorithms that automagically learn node representation in low-dimensional space for predicting DDI.	• They are limited by the quality and completeness of the input data and graph structure so they cannot perform well with large and complex networks. • Cannot perform well with large and complex networks.
Knowledge Graph-DDI	Celebi et al. (2019)[51], Karim et al. (2019), Yue et al. (2021)[52], Ren et al. (2022)[53].	These studies extracts features from different biological entities such as drugs, targets, and enzymes, and integrates them into a multi-modal KG to perform graph embedding.	• Limited by the quality and completeness of the KG and associated data sources. • Require high-quality data and may suffer from scalability issues. • May not capture novel or emerging DI that are not explicitly represented in the KG.
KGAT (General)	Wang et al. (2019)[54], Cui et al. (2020)[55], Li et al. (2021)[56], Wu et al. (2021)[57], Lan et al. (2022)[58], Zhang et al. (2023)[59].	GNN techniques that utilizes attention mechanisms to model complex relationships in KGs. KGAT has become a prevalent tool and has been applied in various applications, including recommendation systems, collaborative filtering, KG completion, entity alignment, text-to-speech synthesis, healthcare and drug discovery.	• May have same limitations as mentioned above in the knowledge graph category as for this technique only an additional mechanism of attention is introduced.

2.2. Research Gaps

The fundamental focus of many AI-based computational models for DDI prediction is merging popular embedding techniques and different data sources. Nevertheless, scientists frequently ignore the possible relationships that medications may have with other objects, such targets and genes. Additionally, some recent studies have utilized KGs for DDI prediction, but these methods typically learn node hidden embeddings directly, limiting the ability to obtain detailed information on the neighborhoods of every entity in the KG [60]. Furthermore, despite its popularity in various applications, no research has used the KGAT for DDI prediction.

CHAPTER 3: EXPERIMENTAL PIPELINE

In this chapter offers a thorough explanation of the experimental procedure we used to assess the suggested DDI-KGAT model for drug-drug interaction prediction. The chapter is organized into multiple sections, with Pseudocode serving as the introduction, outlining the proposed model's algorithmic flow. We then go over the steps needed in gathering data for drug-drug interactions and creating the biological knowledge network, which we call data collection and preprocessing. We go over feature extraction in section 3.3. Feature extraction is the process of taking pertinent features out of the knowledge network and using them to represent medications and their interactions. We describe the specifics of the Model Development process, which included putting the KGAT model into practice, in section 3.4. Finally, in section 3.5, we conclude with a discussion of model training and assessment. This process entailed using the extracted features to train the model and a variety of assessment measures to assess its performance. An extensive description of the experimental procedure we used to assess the suggested KGAT model for drug-drug interaction prediction is given in this chapter.

3.1. Experimental Work

3.1.1. Pseudocode

1. Define the main function
2. Load the necessary modules and libraries.
3. Define the get_optimizer function.
    a. Check the type of optimizer.
    b. Return the corresponding Keras optimizer.
4. Define the train function that takes several inputs.
5. Initialize the config object with the default values.
6. Adjust the model's hyper parameters.
7. Get the information from the dataset.
8. Create an instance of the KGAT class.
    a. Initialize the class with the hyper parameters.
    b. Define architecture of generator plus discriminator.
    c. Define the generator and discriminator's loss functions.
    d. Define the training procedure for the generator and discriminator.
    e. Define the method to score the model.
    f. Define the method to load the best model.
    g. Define the method to load the SWA model.
9. Use the training data to fit the model.
10. Use the validation data to assess the model.
11. Print the results.

3.1.2. Block diagram of experimental work

The block diagram illustrates the main experimental workflow of this research, from data collection and DDI extraction through to model development, training, and evaluation. Predicting DDIs is made rigorous and methodical with this organized technique.

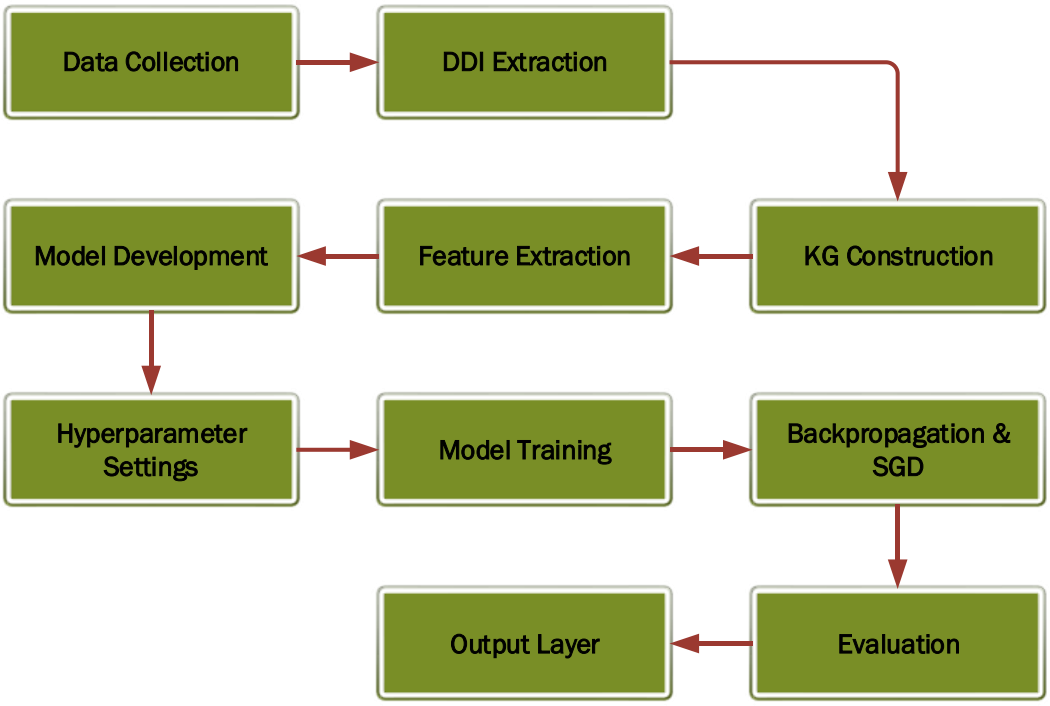

FIGURE 2: BLOCK DIAGRAM OF THE EXPERIMENTAL WORKFLOW

3.2. Data Collection and Preprocessing

3.2.1. DDI Extraction

1. Download data from public drug database (KEGG): The first step in DDI extraction is to obtain the necessary data from the KEGG drug database. This can be done by downloading the XML file containing information about drugs and their interactions.

2. Parse DDI information from XML file: The next step is to extract the DDI information from the XML file. This involves parsing the file and identifying the relevant information, such as the names and identifiers of the drugs involved in each interaction.

3. Compile an edge list of drug identifier combinations: Once the DDI information has been extracted, the next step is to create an edge list of drug identifier combinations. This involves creating a list of all the possible drug pairs involved in DDIs, along with a binary label indicating whether or not a DDI exists between each pair.

3.2.2. KG Construction

1. Retrieve raw data from KEGG dataset: The first step in constructing the KG is to obtain the necessary data from the KEGG dataset. This can be done by downloading the relevant files from the KEGG website.

2. Convert data into RDF graph using Bio2RDF tool: The next step is to convert the raw data into an RDF graph using the Bio2RDF tool. This involves mapping the data to appropriate ontologies and creating RDF triples that represent the entities and relationships in the KG.

3. Upload RDF graph to RDF triplestore: Once the RDF graph has been created, the next step is to upload it to an RDF triplestore. This allows for efficient querying and retrieval of data from the KG.

4. Execute federated SPARQL queries based on billion triples benchmark to extract selected triples: In order to extract the relevant triples from the KG, federated SPARQL queries based on the billion triples benchmark are executed. These queries retrieve triples that represent the entities and relationships involved in DDIs.

5. Construct KG from extracted triples (entity, relation, entity): The final step in KG construction is to use the extracted triples to construct the KG. This involves creating nodes for each entity in the KG (drugs, genes, and targets) and edges for each relationship between entities. The resulting KG can be used as input to train the KGAT model for predicting DDIs.

Create two data sources:

1. Parsed DDI matrix containing drug-drug pairs.
2. Constructed knowledge graph.

3.3. Feature Extraction

The next step is to extract features from the preprocessed data. We used various techniques to convert categorical data into numerical data. We also used graph-based methods to represent the drug interactions and associations in a more structured format. This step helped us to obtain the necessary features required for the drug combination prediction.

3.4. Model Development

We developed a knowledge graph-based approach called KGAT for drug combination prediction using attention mechanisms. KGAT uses GNN with attention mechanisms to learn embeddings for the drugs and their interactions in the KG. The attention mechanisms allow KGAT to capture important features and enhance the performance of drug combination prediction.

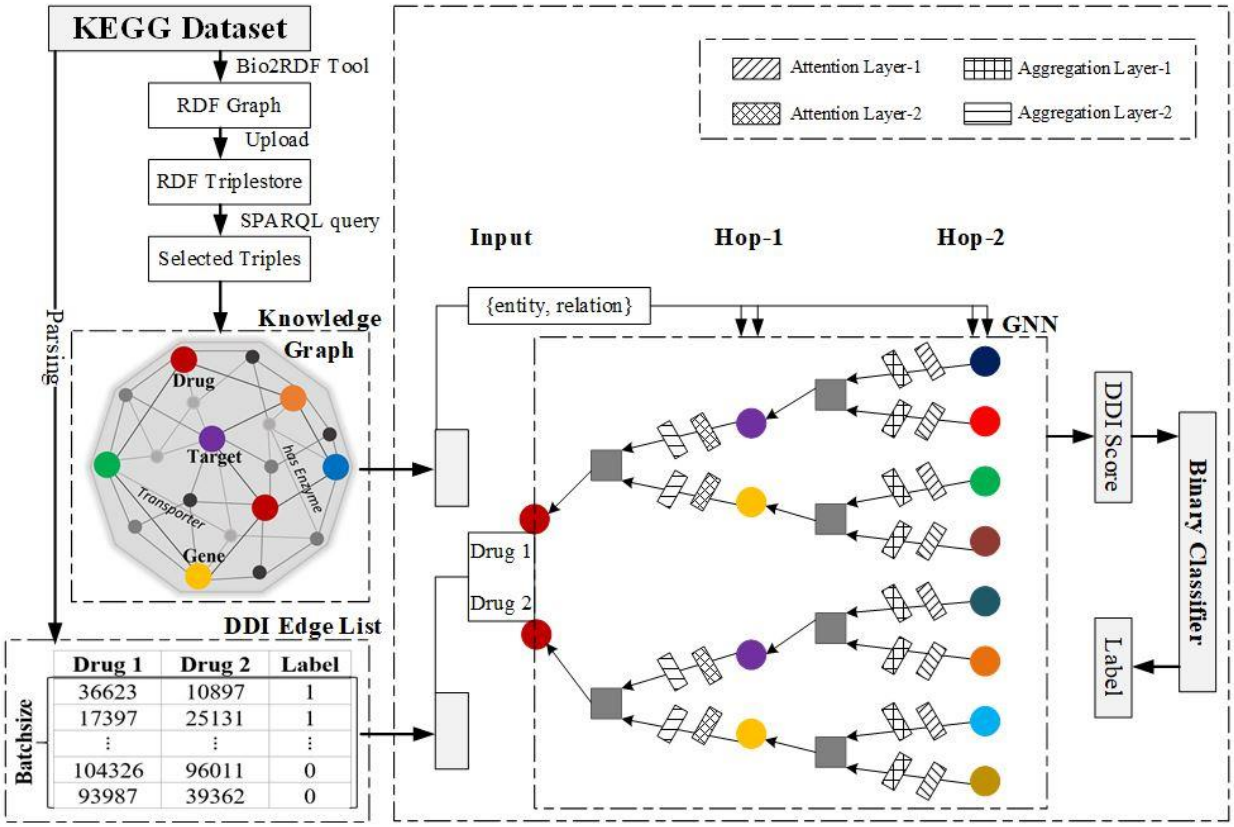

FIGURE 3: OVERVIEW OF OUR TECHNIQUE

3.5. Model Training and Evaluation

We Our KGAT model was trained using a sizable dataset of medication combinations, and its effectiveness was assessed using the KEGG dataset. In order to prevent overfitting and guarantee its robustness, we carried out cross-validation. After that, we split the dataset into training and testing sets by preparing the data and extracting its features. Before training the model using backpropagation and stochastic gradient descent to minimize the loss function, we defined the hyper-parameters, which included the embedding dimension, neighborhood sample size, l2_weight, learning rate, optimizer type, batch size, aggregator type, regularization strength, and number of epochs. Using widely used evaluation metrics including accuracy (ACC) and F1 score, as well as Area under the Precision-Recall Curve (AUPR) and Receiver Operating Characteristic Curve (AUC-ROC), we assessed the model's performance on the testing set.

Furthermore, we visualized the model's predictions and the relationships between the drugs in the KG and compared the performance with other state-of-the-art approaches for predicting DDIs.

CHAPTER 4: METHODOLOGICAL FRAMEWORK

In this chapter, experimental setup and technique for the proposed KGAT model are provided. The methodology section details the architecture of DDI-KGAT, which consists of input, graph embedding, graph attention, graph convolution, aggregation, and output layers. Each layer is described in detail, including the purpose and functionality of the layer in the model. Additionally, the training process's loss function is also covered.

The experimental setup, which includes the data sources and pre-processing methods used to build the biomedical knowledge graph (KG) and extract drug-drug interactions (DDIs) for the model training and evaluation, is described after the methodology section. The main data source for building the KG is the KEGG database, which includes details on metabolic pathways and biological reactions. This section also includes a description of the hyper-parameter values that were used in the tests. All things considered, this chapter offers a thorough rundown of the experimental design and technique for the DDI-KGAT model.

4.1. Methodology

Complete architecture of Knowledge Graph Attention Network (KGAT) for predicting drug-drug interactions (DDI):

4.1.1. Input Layer

-Knowledge graph (KG) with entities (drugs, targets, genes) and relations between them

-DDI labels (binary classification)

4.1.2. Graph Embedding Layer

In the DDI prediction task, the Embedding Layer maps every drug and feature in the KG to a low-dimensional vector space using embedding matrices. The embedding matrices are trained through backpropagation while training the model, during the training phase.

In this project, we use an entity embedding approach to encode the drug nodes and their features. We create a dictionary of all drugs and features in the KG and assign a unique index to every of them. We then use these indices to create a dense matrix representation of the KG.

To obtain the entity embeddings, we use an embedding matrix that maps the indices to low-dimensional vectors. Every row of the embedding matrix represents a unique entity in the KG. During training, the embedding matrix is learned through backpropagation to minimize the loss function.

The Embedding Layer is an essential component of the KGAT model because it enables the model to represent every entity in the KG as a low-dimensional vector. This vector representation captures the characteristics of the entity and its relationships with other entities in the KG. These embeddings are then fed into the Attention Layer for further processing.

4.1.3. Graph Attention Layer

The Attention Layer in our project is a crucial component in capturing the relationships between entities in the KG and incorporating this information into the drug representation. The layer operates by computing attention weights for every neighbor of a drug node in the KG, which are a measure of the importance of every neighbor in the context of the drug node.

In the attention computation process, a multi-head attention mechanism is employed, enabling the model to concurrently focus on diverse subsets of neighbors. The attention weights (AW) are determined through a dot product between the drug node embedding and the embedding of each neighbor (ne), followed by the addition of the attention bias term (ab) and an attention weight tensor (aw). The resulting values undergo a softmax operation (σ) along the specified axis to generate a probability distribution over the neighbors. The mathematical representation is expressed as:

$$AW = \sigma((ne\ .\ aw) + ab)_{axis=1} \quad \textbf{(1)}$$

The attention weights are then used to obtain a context vector for the drug node by summing the weighted neighbor embeddings. The context vector captures the most important information from the neighbors in the KG and is used as a supplement to the original drug representation obtained from the Embedding Layer.

Overall, the Attention Layer is an essential component of current work as it allows the model to effectively capture the intricate connections among entities in the KG, which is crucial for accurately predicting DDIs.

4.1.4. Graph Convolution Layer

For the purpose of learning node representations in a graph, the graph convolution layer is a crucial part of GNNs, which include graph attention networks (GATs). This layer's objective is to compile data from a node's immediate neighborhood and create a new representation for the node that includes its neighbors' and the node's own features.

While there are other methods for creating graph convolution layers, a popular strategy is to leverage a message passing scheme variation. Messages are passed iteratively between nodes in a graph using the message passing technique. Each message is a function of the node's own properties as well as those of its neighbors. After that, a mean or summing operation is applied to all of the neighbors' messages, and the resulting aggregate message is used to update the node's representation.

Our study explores the effectiveness of three distinct types of aggregators, including the sum, mean, and max aggregators, each influencing the aggregation process differently. In the subsequent formulas (2), (3), and (4) we utilize the following terms:

Drug Entity (e), represents the intrinsic properties of the drug node in the KG; Neighbor (n), signifies the features of neighboring nodes that interact with the drug node; Output (y), denotes the resultant representation of the drug node after the aggregation process; Activation Function (σ) as we are using sigmoid, it introduces non-linearity to the output, facilitating the model to learn complex relationships and patterns; Dot Product Operation (.) , a mathematical operation used extensively in the aggregation process to determine feature significance; Weight Tensor (w), refers to the tensor responsible for weighting the importance of different features during aggregation; Bias Term (b), represents the bias term, which is introduced to the weighted sum during aggregation.

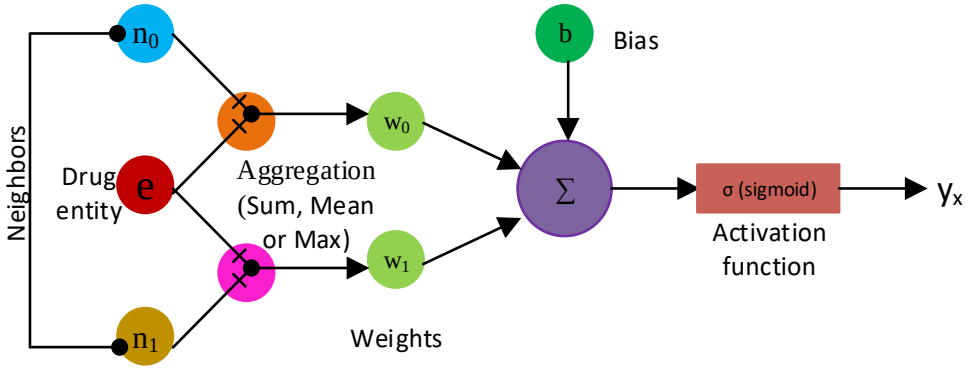

FIGURE 4: STRUCTURE OF GRAPH CONVOLUTION LAYER

4.1.4.1. Sum Aggregation Layer

In the Sum Aggregation Layer, the drug features and the context vector obtained from the Attention Layer are combined ($e + n$) to create a more inclusive illustration of the drug node in the KG. The drug features represent the intrinsic properties of the drug, while the context vector captures the influence of the drug's neighbors in the KG. By weighting and combining these two sources of information, the model is capable to capture both the intrinsic properties of the drug and its interactions with other entities in the KG, resulting in a more accurate prediction of DDIs.

$$y_{sum} = \sigma((e + n) \cdot w + b) \qquad \textbf{(2)}$$

4.1.4.2. Mean Aggregation Layer

In the Mean Aggregation Layer, the drug features and the context vector obtained from the Attention Layer are combined using the element-wise mean operation $((e + n)/2)$ to create a more balanced representation of the drug node in the KG. This operation allows the model to capture both the individual properties of the drug and the overall characteristics of its neighbors in the KG, leading to a more complete understanding of the drug and its potential interactions.

$$y_{mean} = \sigma(((e + n)/2) \cdot w + b) \quad \textbf{(3)}$$

4.1.4.3. Max Aggregation Layer

In the Max Aggregation Layer, the drug features and the context vector obtained from the Attention Layer are combined using the element-wise maximum operation $(max(e, n))$ to emphasize the most salient features of the drug node in the KG. This operation permits the model to focus on most important properties of the drug and its neighbors, potentially leading to more accurate predictions of DDIs. However, the approach may also be more sensitive to outliers and noise in the data, and may require more careful tuning.

$$y_{max} = \sigma(max(e, n) \cdot w + b) \quad \textbf{(4)}$$

4.1.5. Output Layer

The composite representation of the drug node is then sent into the FCNN for binary classification. A particular kind of neural network called an FCNN is made up of fully connected layers that are connected one after the other so that every neuron in one layer is connected to every other neuron in the layer above it. A single value that represents the anticipated chance of a DDI between the two medications is the output of the FCNN's last layer. The model is trained using the binary cross-entropy loss function, which calculates the variance between each training example's predicted and actual label and uses the Adam optimizer to apply gradient descent to modify the model's weights.

4.1.6. Loss Function

The binary cross entropy loss, or loss function (L), is a frequently used function for binary classification issues such as DDI prediction. The difference between the true labels, which is either $y = 0$ (no DDI) or $y = 1$ (DDI present) and the predicted probability (p) of a DDI is measured by the binary cross entropy loss.

$$L = -[y * log(p) + (1 - y) * log(1 - p)] \quad \textbf{(5)}$$

During training, model employs the Adam optimizer, a commonly used algorithm that dynamically adjusts learning rates for each parameter based on gradients. The optimizer aims to find the best parameter configuration, minimizing the loss function and maximizing predictive performance on test data. By integrating a loss function and optimizer, the model learns from errors, fine-tuning parameters for improved predictions. Evaluating the resulting predictions on a separate test set allows performance measurement and facilitates comparison with other models or baselines.

The DDI-KGAT method is composed of multiple graph convolutional layers that are succeeded by attention layers that determine the significance of certain graph segments for the DDI prediction job. This enables the model to filter out noise and unimportant data and concentrate only on pertinent regions of the graph. To forecast DDI, the final output layer combines drug characteristics and learned node embeddings.

Overall, the architecture of KGAT for DDI prediction involves multiple layers that are specifically designed to influence the graph structure of the data and incorporate information from the graph in a principled way. The attention mechanism in the DDI-KGAT provides a flexible way to weight the contributions of different parts of the graph to the prediction task, making it a powerful approach for modeling complex relationships in the data.

4.2. Experimental Setup

4.2.1. KEGG Database

In order to make it easier to analyze complicated biological systems, KEGG is a comprehensive database resource that combines different biological information, including genetic, chemical, and network information. KEGG is an invaluable resource for biomedical research since it offers a significant amount of manually selected data on numerous biological processes, illnesses, and medications. In the realm of bioinformatics, the KEGG dataset is extensively utilized and has been useful in deepening our understanding of biological systems.

In this study, we utilized a combination of two data sources: a parsed DDI matrix and a constructed biomedical knowledge graph. The DDI matrix was extracted from publicly available drug database, KEGG, and parsed to assemble an edge list of the drug identifier combinations. The KG was constructed by retrieving raw data from the KEGG dataset and converting it into an RDF graph using the Bio2RDF tool. The RDF graph was then uploaded to an RDF triplestore, and to extract selected triples federated SPARQL queries built on the billion triples benchmark were executed. The constructed KG was composed of entities, relations, and entities. The statistics of both data sources were analyzed to ensure data quality and accuracy.

Table 2: Basic statistics of the used KEGG dataset.

Drugs	1925
DDIs	113965
Entities	129910
Relation	167
KG triplets	362870

To preserve the original integrity of the KG, we collected and constructed it using a careful approach. However, as the KG should not include any explicit information on DDIs, we excluded the information in the form of URLs, specifically "Drug-Drug-Interaction" from KEGG-drug

datasets. This process ensured that the KG did not include any data that could bias the prediction of DDIs.

4.2.2. Hyper-Parameter Settings

After conducting a thorough case study and examining the effects of hyper parameters at various settings, we have carefully tuned our model to the optimal configuration that yields the best results for our task of DDI prediction.

All authorized DDIs were split into positive samples at random and then distributed using an 18/1/1 ratio among training, validation, and testing sets. We randomly selected an equal number of negative samples from the complement set of positive samples in each phase to maintain equilibrium. We used eight-fold cross-validation tests and a random search to optimize all trainable parameters with Adam. We optimized additional hyper-parameter values for the validation set, which are shown in a table for reference, and the training was set for 30 epochs.

Table 3: Hyper-parameter settings.

Hyper-parameter	Value
Neighbor sampling size	16
Dimension of embedding	32
Depth of receptive field	2
Batch size	2048
L2 regularizer weight	2e-8
Learning rate	5e-3

CHAPTER 5: RESULTS AND DISCUSSIONS

The findings and explanations of the KGAT model for forecasting drug-drug interactions are presented in this chapter. The findings of the experiments done on the KGAT model are presented at the start of the chapter. Several performance indicators, including accuracy, F1 score, AUC-ROC, and AUPR, are included in the results section. To assess KGAT's efficacy, its performance is also contrasted with that of other cutting-edge techniques.

The results and their consequences are thoroughly analyzed in the discussion section. The study's contributions are highlighted and the results are compared with earlier research. A overview of the findings and their importance in the context of KGAT-based drug-drug interaction prediction round off the chapter.

5.1. Results

The outcomes of our tests assessing the effectiveness of our suggested KGAT model for DDI prediction on the KEGG dataset are shown in this section. We evaluate the efficacy of three distinct aggregator types—sum, mean, and max—that are frequently employed in GNNs.

First, we use the sum aggregator to analyze our methods' performance. The findings demonstrate that, in comparison to state-of-the-art methods, our KGAT model obtains the greatest average AUC-ROC of 0.9447, ACC of 0.8869, F1 score of 0.8917, and AUPR of 0.9255.

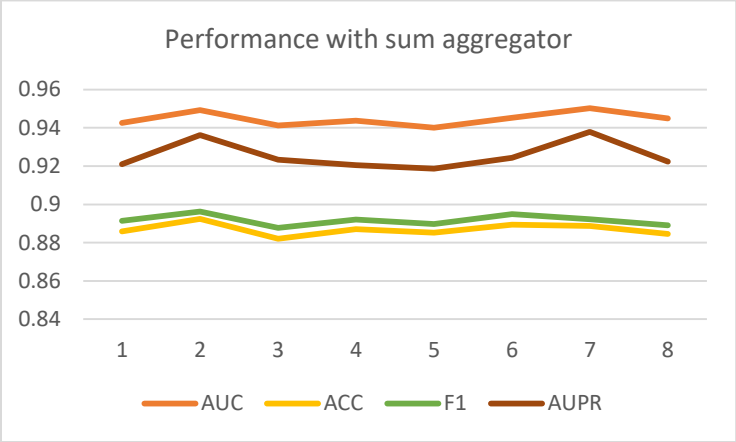

FIGURE 5: RESULTS OF OUR MODEL USING SUM AGGREGATOR

Next, we use the mean aggregator to assess how well our plan is working. The findings show that the KGAT model performs comparably to the sum aggregator, achieving an average AUC-ROC of 0.9419, ACC of 0.8823, F1-score of 0.8867, and AUPR of 0.9226.

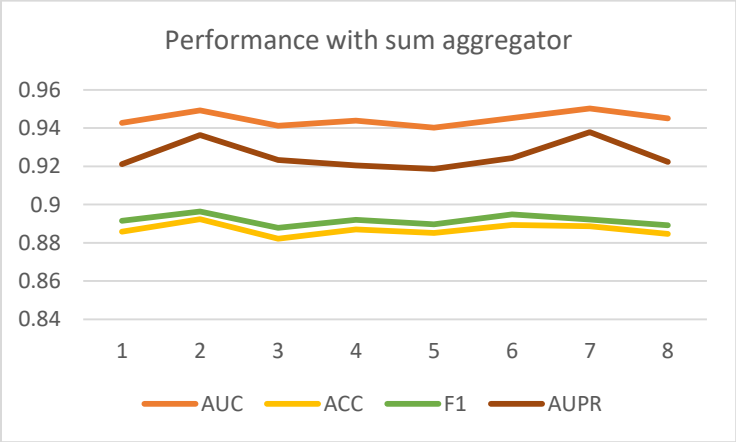

FIGURE 6: RESULTS OF OUR MODEL USING MEAN AGGREGATOR

We then examine the performance of our model using the max aggregator. The results indicate that the KGAT model achieves the average AUC-ROC of 0.9397, ACC of 0.8772, F1-score of 0.8816 and AUPR of 0.9207, which is slightly lower than the performance achieved with the sum aggregator.

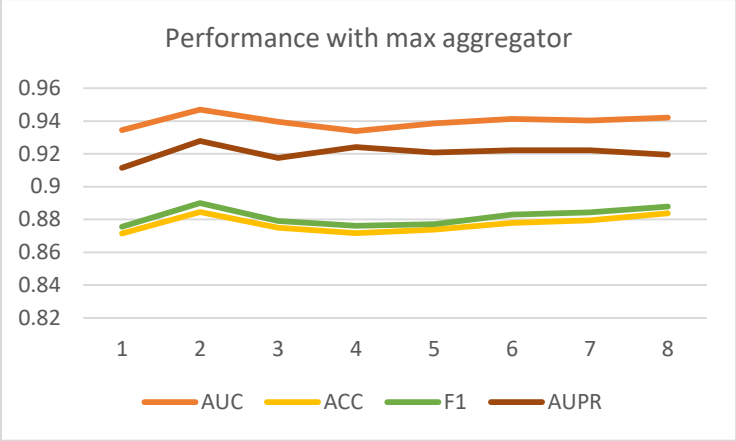

FIGURE 7: RESULTS OF OUR MODEL USING MAX AGGREGATOR

Finally, we present a comparison of the performance of our KGAT model using the three different types of aggregators. The results show that the sum aggregator outperforms both the mean and max aggregators in terms of AUC-ROC, ACC, F1-score and AUPR. Our findings suggest that the sum aggregator is the most effective aggregation function for our proposed KGAT model in the task of predicting drug-drug interactions.

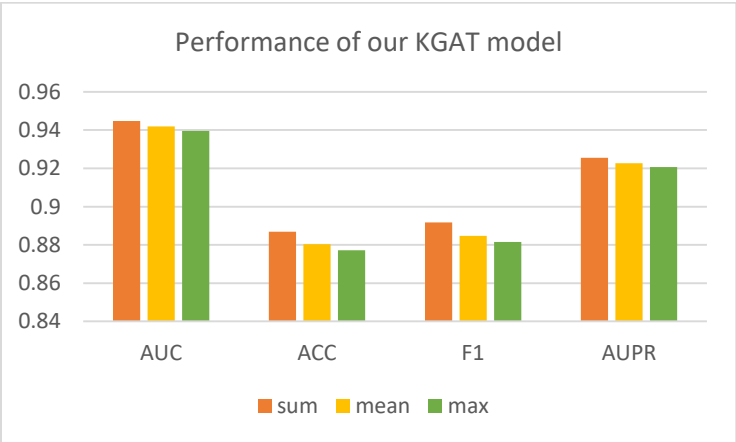

FIGURE 8: OVERALL COMPARISON OF OUR MODEL RESULTS

5.2. Discussion

GNNs have emerged as powerful tools for DDI prediction using KGs. Aggregators play a crucial role in these models as they determine how neighboring nodes are combined to update the embedding of a central node. The selection of aggregator depends on the precise task and the characteristics of the graph being used.

There are many main types of aggregators commonly used in GNNs. We have used: sum, mean, and max aggregators. The sum aggregator is a simple function that computes the sum of the embeddings of neighboring nodes, making it widely used to update the embedding of a central node by combining info from all neighboring nodes. However, it can be sensitive to outliers and may not always capture the most relevant information from neighboring nodes. On the other hand, the mean aggregator computes the mean of the embeddings of neighboring nodes, making it more robust to outliers and commonly used in graph neural networks. Meanwhile, the max aggregator computes the maximum of the embeddings of neighboring nodes and is useful when the most relevant information is likely to come from a single neighbor.

The choice of aggregator in a graph neural network can have significant implications for memory usage and computational power. The sum aggregator is a memory-efficient function that requires less computational resources than the mean and max aggregators. Moreover, our research study found that the sum aggregator outperformed the mean and max aggregators in predicting DDIs, mainly because it can effectively capture the most relevant information from neighboring nodes. Additionally, the sum aggregator is a simple and intuitive function that is easy to implement and interpret, making it a prevalent choice for many applications. Overall, the sum aggregator is a reliable and efficient function that can be a useful tool for updating the embeddings of central nodes in GNNs.

5.2.1. Comparison with similar work on predicting DDIs using knowledge graph neural network (KGNN) on KEGG dataset.

Our research study has shown that incorporating an attention mechanism with a simpler aggregator has yielded even better results than the concatenate aggregator, which was claimed to be the best so far by Lin et al. [60].

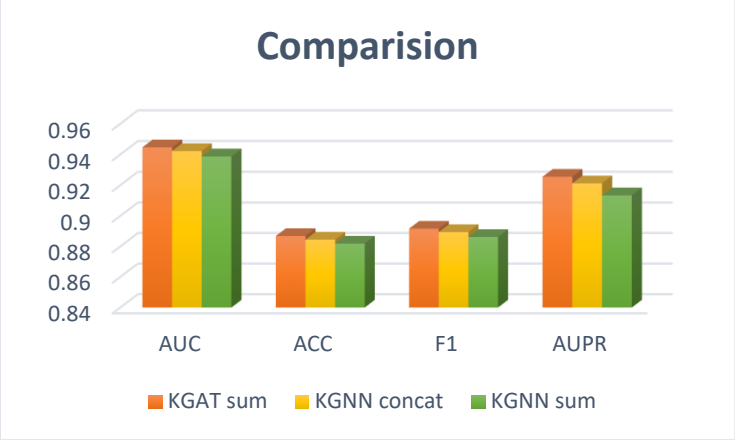

FIGURE 9: OUR TECHNIQUE CONTRAST TO THE 2ND BEST TECHNIQUE UNTIL NOW

This finding suggests that our technique has surpassed the state of the art approaches and achieved the highest ACC, F1-score, AUC-ROC, and AUPR values for the DDI prediction task.

Table 4: Comparison of DDI-KGATX vs KGNNX technique.

Methods / Metrics	KGNNx			DDI-KGATx		
	neighbor	*sum*	*concat*	*max*	*mean*	*sum*
AUC-ROC	0.9295	0.9387	0.9422	0.9397	0.9419	0.9447
ACC	0.8681	0.8818	0.8844	0.8799	0.8853	0.8901
F1 score	0.8726	0.8862	0.8894	0.8816	0.8867	0.8917
AUPR	0.9052	0.9134	0.9212	0.9207	0.9226	0.9255

The results of our study have significant inferences for the field of KG embedding, especially in domains where high accuracy and low-dimensional complexity are essential, such as drug discovery. By combining the attention mechanism with a simpler aggregator, our technique has not only improved the performance of the KGNN model [61] but has also reduced its complexity, making it more computationally efficient.

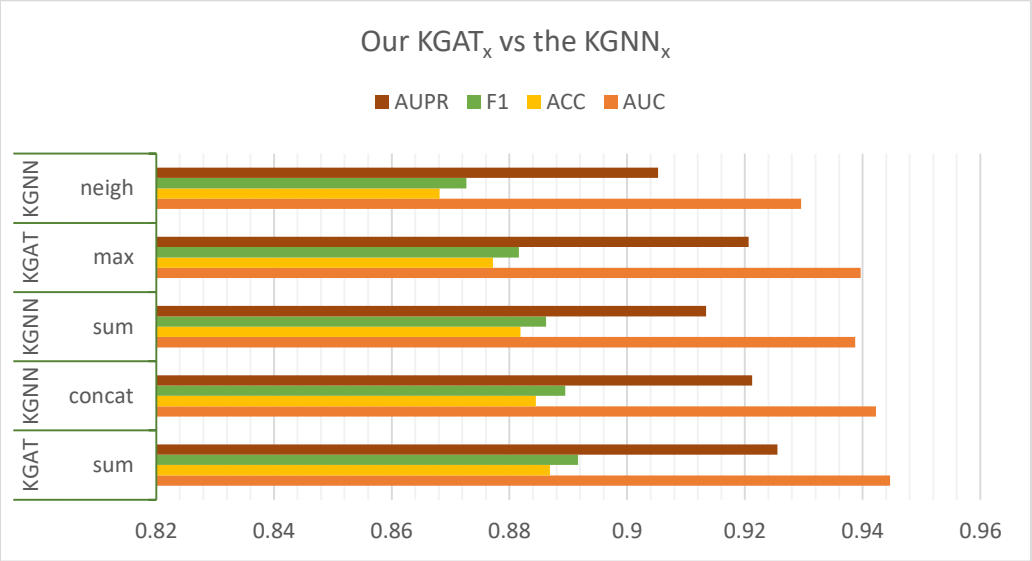

FIGURE 10: COMPARISON OF OUR KGATx TECHNIQUE VS KGNNx

All things considered, our research study has advanced the state of the art in KG embeddings by putting forth a unique method that works better than current models. It emphasizes how crucial it is to investigate various aggregator kinds and include attention techniques in order to improve GNN performance.

5.2.2. Comparison with state-of-the-art techniques using knowledge graph for predicting DDIs.

In the study by Feng et al. on DPDDI [49], deep learning frameworks are employed; however, a limitation arises in terms of interpretability, making it challenging to comprehend the underlying factors influencing predictions. Yu et al. propose SumGNN [52], which focuses on efficient KG summarization for multi-typed drug interaction prediction. Nonetheless, the method encounter difficulties in capturing intricate relationships within the KG, potentially impacting prediction accuracy. Ren et al. introduce BioDKG-DDI [53], which combines biochemical information with a drug KG. Nevertheless, the integration of multiple data sources presents complexities and potential challenges in handling heterogeneous data. SA-DDI technique by Yang et al. [62] include challenges in interpreting the identified substructures, requiring further investigation for meaningful interpretation. Lastly, Hao et al. [63] utilize a three-way decision and stack multiple complex techniques, which can lead to increased computational complexity and hinder scalability. Although their performance is comparable to DDI-KGAT, but it is not offering the same level of simplicity and computational efficiency.

FIGURE 11: COMPARISON OF DDI-KGAT VS STATE-OF-THE-ART TECHNIQUES

The effectiveness of present strategies is validated by the outstanding F1 score, accurately capturing true positives and reducing false positives in DDI prediction. Prioritizing this metric allows for meaningful comparisons with existing literature, offering a comprehensive assessment of precision and recall. While AUPR, ACC, and AUC provide insights into classifier performance, the emphasis on the F1 score highlights the ability of the current technique to strike a balance

between precision and recall. This focus underscores our commitment to addressing DDI prediction challenges and establishes a new benchmark in the field. In conclusion, the F1 score serves as a reliable metric, demonstrating the effectiveness of DDI-KGAT in predicting DDIs.

Table 5: Comparison of DDI-KGAT vs state-of-the-art techniques

Methods	F1-score (%)
DPDDI (Feng et al., 2020) [49]	85.5
SumGNN (Yu et al., 2021) [52]	86.9
BioDKG-DDI (Ren et al., 2022) [53]	88
SA-DDI (Yang et al., 2022) [62]	88
3WDDI (Hao et al., 2023) [63]	89
DDI-KGAT (Ours)	89

Our analysis reveals that DDI-KGAT stands out from the aforementioned publications due to its exceptional performance, high accuracy, low-dimensional complexity, and computational efficiency. The approach offers a simpler and more interpretable solution, while still achieving comparable or even superior results, demonstrating its effectiveness and practicality in the context of DDI prediction. Additionally, DDI-KGAT demonstrates superior computational efficiency, scalability, and integration ease compared to current methods. It requires less computational resources and proposes faster inference times due to its optimized architecture, maintaining robust performance with increasing data volumes and knowledge graph complexity. Its interpretable predictions and user-friendly interface facilitate seamless integration into clinical workflows, making it the preferred choice for DDI prediction in real-world settings. This positions DDI-KGAT as a benchmark in the field, setting new standards for efficiency and usability in DDI prediction. See the table below for a summary of the performance of DDI-KGAT compared to state-of-the-art techniques.

Table 6: Performance overview of DDI-KGAT vs state-of-the-art techniques

Methods	Computational Efficiency	Scalability	Integration Ease
DPDDI	Moderate to High	Limited	Moderate
SumGNN	Moderate	Moderate	Limited
BioDKG-DDI	Moderate	Moderate	Limited
SA-DDI	Moderate to High	Limited	Moderate
3WDDI	Low to Moderate	Limited	Limited
DDI-KGAT (Ours)	High	High	Easy

5.2.3. Case Study

EVALUATING DDI PREDICTIONS FOR METFORMIN AND DAPAGLIFLOZIN

To demonstrate the clinical relevance and interpretability of our DDI-KGAT model, we conducted a detailed case study focused on the predicted interaction between Metformin, a widely used antidiabetic drug, and Dapagliflozin, an SGLT2 inhibitor. Both drugs are commonly prescribed to manage type 2 diabetes but their interaction potential is not extensively documented in clinical practice.

Our DDI-KGAT model predicted a significant interaction between Metformin and Dapagliflozin. This interaction could potentially enhance the hypoglycemic effects when used concomitantly, thereby increasing the hazard of lactic acidosis [64] , a rare but thoughtful complication that involves the enhanced glucose regulatory effect. The binary output of the model indicated a "positive" interaction, suggesting heightened attention when these drugs are prescribed together.

SUMMARY

The goal of the proposed study is to predict drug-drug interactions (DDI) by employing a unique knowledge graph-based method known as KGAT. This method efficiently extracts the drug and its possible neighbors by capturing high-order structures and semantic relations of the graph. To capture significant aspects and connections between medications and other things, like targets and genes, the model makes use of attention mechanisms. This smaller dimensionality, computationally efficient method makes use of detailed neighborhood information for each entity in the knowledge graph (KG). On the Kyoto Encyclopedia of Genes and Genomes (KEGG) dataset, the study demonstrates that the KGAT model performs better than the state-of-the-art methods currently in use. The model emphasizes the significance of accurate drug combination prediction and has the potential to enhance patient outcomes and safer drug development. The usage of the knowledge graph attention network (KGAT), a method not previously investigated in DDI prediction, is also investigated in this study, as is its potential for further investigation in KG embedding.

The study draws attention to the shortcomings in current computational models that use AI techniques for DDI prediction. These models typically concentrate on merging popular embedding methods and integrating multiple data sources, but they pay less attention to possible correlations between drugs and other entities like targets and genes. Furthermore, KGs have been used for DDI prediction in recent research; however, these approaches directly train node hidden embedding, which restricts the capacity to acquire detailed neighborhood information for each entity in the KG. The study's findings show how well the KGAT model performs when using the sum, mean, and max aggregators—three aggregator types that are frequently employed in graph neural networks (GNNs). The findings demonstrate that in terms of AUC-ROC, ACC, F1-score, and AUPR, the sum aggregator performs better than the mean and max aggregators.

The talk focuses light on the aggregator selection that is frequently employed in GNNs. The particular task at hand and the properties of the graph being used determine which aggregator is best. The sum aggregator is a frequently used function that combines information from all nearby nodes to update the embedding of a central node. It is a simple function that computes the sum of the embeddings of neighboring nodes. Often utilized in graph neural networks, the mean

aggregator calculates the mean of the embeddings of nearby nodes, increasing its resilience to outliers. In the meanwhile, the max aggregator is helpful when the most pertinent data is probably going to originate from a single neighbor since it calculates the maximum of the embeddings of nearby nodes. The choice of aggregator in a GNN can have significant implications for memory usage and computational power. The sum aggregator is a memory-efficient function that requires less computational resources than the mean and max aggregators.

Overall, the research provides a valuable framework for future research in knowledge graph-based drug combination prediction, highlighting the potential of attention mechanisms in this field. The suggested method has the ability to enhance patient outcomes and promote safer medication development by discovering innovative drug combinations with enhanced efficacy and safety characteristics.

CONCLUSIONS

In conclusion, our research has effectively created a novel method for leveraging graph neural networks and knowledge graphs to predict drug-drug interactions. Our results show that this approach has the ability to solve challenging prediction tasks because it makes use of the structured information found in knowledge graphs to learn the complex relationships between medications and their interactions. Through the addition of a sum aggregator as an attention mechanism to the KGAT model, we were able to show a notable increase in accuracy while maintaining the model's computational efficiency and reduced dimensionality. Our utilization of the knowledge graph attention network added further novelty to our study and holds great promise for future research in knowledge graph embedding, especially in domains where high accuracy and low-dimensional complexity are of utmost importance.

This method has many advantages, such as improving drug-drug interaction prediction accuracy, avoiding costly and time-consuming experimental studies, and offering transparency and interpretability to aid in understanding the underlying mechanisms of drug interactions. The method's simplicity makes it accessible to a larger audience, and it may also be expanded to other fields where graph-based data is common, like social networks, recommendation systems, and fraud detection.

Overall, our research shows how well knowledge graphs and graph neural networks perform complex prediction tasks, and our methodology offers a straightforward and novel means of potentially revolutionizing the fields of personalized medicine and drug discovery, resulting in safer and more effective treatments for a range of illnesses.

REFERENCES

[1]. "Drug-Drug Interaction: NIH," Drug-Drug Interaction | NIH, https://clinicalinfo.hiv.gov/en/glossary/drug-drug-interaction#:~:text=A%20change%20in%20a%20drug's,drugs%20or%20cause%20adverse%20effects. (accessed Mar. 6, 2023).

[2]. L. L. Leape, "Systems analysis of adverse drug events. Ade Prevention Study Group," *JAMA: The Journal of the American Medical Association*, vol. 274, no. 1, pp. 35–43, 1995. doi:10.1001/jama.274.1.35.

[3]. M. Hohl, J. Dankoff, A. Colacone, and M. Afilalo, "Polypharmacy, adverse drug-related events, and potential adverse drug interactions in elderly patients presenting to an emergency department," *Annals of Emergency Medicine*, vol. 38, no. 6, pp. 666–671, 2001. doi:10.1067/mem.2001.119456.

[4]. J. Lehár *et al.*, "Synergistic drug combinations tend to improve therapeutically relevant selectivity," *Nature Biotechnology*, vol. 27, no. 7, pp. 659–666, 2009. doi:10.1038/nbt.1549.

[5]. S. Vilar, E. Uriarte, L. Santana, T. Lorberbaum, G. Hripcsak, C. Friedman, and N. P. Tatonetti, "Similarity-based modeling in large-scale prediction of drug-drug interactions," *Nature Protocols*, vol. 9, no. 9, pp. 2147–2163, 2014. doi:10.1038/nprot.2014.151.

[6]. I. Abdelaziz, A. Fokoue, O. Hassanzadeh, P. Zhang, and M. Sadoghi, "Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions," *Journal of Web Semantics*, vol. 44, pp. 104–117, 2017. doi:10.1016/j.websem.2017.06.002.

[7]. A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in Neural Information Processing Systems*, vol. 26, 2013, pp. 2787-2795.

[8]. Md. R. Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, and S. Decker, "Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network," *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2019. doi:10.1145/3307339.3342161.

[9]. K. L. J. MS;, “To err is human: Building a safer health system,” National Center for Biotechnology Information, https://pubmed.ncbi.nlm.nih.gov/25077248/ (accessed Mar. 6, 2023).

[10]. Incidence of adverse drug reactions in hospitalized patients: A Meta ..., https://www.ncbi.nlm.nih.gov/books/NBK67323/ (accessed Mar. 6, 2023).

[11]. P. Imming, C. Sinning, and A. Meyer, “Drugs, their targets and the nature and number of drug targets,” *Nature Reviews Drug Discovery*, vol. 5, no. 10, pp. 821–834, 2006. doi:10.1038/nrd2132.

[12]. L. Zhang, Y. Zhang, P. Zhao, and S.-M. Huang, “Predicting drug–drug interactions: An FDA perspective,” *The AAPS Journal*, vol. 11, no. 2, pp. 300–306, 2009. doi:10.1208/s12248-009-9106-3.

[13]. A. Gottlieb, G. Y. Stein, Y. Oron, E. Ruppin, and R. Sharan, “Indi: A computational framework for inferring drug interactions and their associated recommendations,” *Molecular Systems Biology*, vol. 8, no. 1, 2012. doi:10.1038/msb.2012.26.

[14]. S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, and C. Friedman, “Drug—drug interaction through molecular structure similarity analysis,” *Journal of the American Medical Informatics Association*, vol. 19, no. 6, pp. 1066–1074, 2012. doi:10.1136/amiajnl-2012-000935.

[15]. R. Ferdousi, R. Safdari, and Y. Omidi, “Computational prediction of drug-drug interactions based on drugs functional similarities,” *Journal of Biomedical Informatics*, vol. 70, pp. 54–64, 2017. doi:10.1016/j.jbi.2017.04.021.

[16]. P. Li, C. Huang, Y. Fu, J. Wang, Z. Wu, J. Ru, Y. Wang, *et al.*, “Large-scale exploration and analysis of drug combinations,” *Bioinformatics*, vol. 31, no. 12, pp. 2007–2016, 2015. doi:10.1093/bioinformatics/btv080.

[17]. J.-Y. Shi, K. Gao, X.-Q. Shang, and S.-M. Yiu, “LCM-DS: A novel approach of predicting drug-drug interactions for new drugs via Dempster-Shafer Theory of Evidence,” in *Proc. 2016 IEEE Int. Conf. Bioinformatics Biomedicine (BIBM)*, 2016. doi: 10.1109/bibm.2016.7822571.

[18]. A. Kastrin, P. Ferk, and B. Leskošek, “Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning,” *PLOS ONE*, vol. 13, no. 5, 2018. doi:10.1371/journal.pone.0196865.

[19]. X. Zhao, L. Chen, Z.-H. Guo, and T. Liu, “Predicting drug side effects with compact integration of heterogeneous networks,” *Current Bioinformatics*, vol. 14, no. 8, pp. 709–720, 2019. doi:10.2174/1574893614666190220114644.

[20]. L. Tari, S. Anwar, S. Liang, J. Cai, and C. Baral, “Discovering drug–drug interactions: A text-mining and reasoning approach based on properties of Drug Metabolism,” *Bioinformatics*, vol. 26, no. 18, pp. i547–i553, 2010. doi:10.1093/bioinformatics/btq382.

[21]. N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, “Data-driven prediction of drug effects and interactions,” *Science Translational Medicine*, vol. 4, no. 125, 2012. doi:10.1126/scitranslmed.3003377.

[22]. A. Kolchinsky, A. Lourenço, L. LI, and L. M. Rocha, “Evaluation of linear classifiers on articles containing pharmacokinetic evidence of drug-drug interactions,” *Biocomputing 2013*, 2012. doi:10.1142/9789814447973_0040.

[23]. A. Cami, S. Manzi, A. Arnold, and B. Y. Reis, “Pharmacointeraction network models predict unknown drug-drug interactions,” *PLoS ONE*, vol. 8, no. 4, 2013. doi:10.1371/journal.pone.0061468.

[24]. B. Perozzi, R. Al-Rfou, and S. Skiena, “Deepwalk,” *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014. doi:10.1145/2623330.2623732.

[25]. W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, “Predicting potential side effects of drugs by recommender methods and ensemble learning,” Neurocomputing, vol. 173, pp. 979–987, 2016. doi:10.1016/j.neucom.2015.08.054.

[26]. K. Park, D. Kim, S. Ha, and D. Lee, “Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks,” PLOS ONE, vol. 10, no. 10, p. e0140816, 2015. doi:10.1371/journal.pone.0140816.

[27]. D. Sridhar, S. Fakhraei, and L. Getoor, “A probabilistic approach for collective similarity-based drug–drug interaction prediction,” *Bioinformatics*, vol. 32, no. 20, pp. 3175–3182, 2016. doi:10.1093/bioinformatics/btw342.

[28]. L. F. Ribeiro, P. H. Saverese, and D. R. Figueiredo, “struc2vec: Learning node representations from structural identity,” in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 385–394. doi:10.1145/3097983.3098061.

[29]. C. Yan, G. Duan, Y. Zhang, F.-X. Wu, Y. Pan, and J. Wang, “IDNDDI: An integrated drug similarity network method for predicting drug-drug interactions,” *Bioinformatics Research and Applications*, pp. 89–99, 2019. doi:10.1007/978-3-030-20242-2_8.

[30]. J. Tang *et al.*, “Line,” *Proceedings of the 24th International Conference on World Wide Web*, 2015. doi:10.1145/2736277.2741093.

[31]. D. Wang, P. Cui, and W. Zhu, “Structural Deep Network embedding,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. doi:10.1145/2939672.2939753.

[32]. N. Mrabah, M. Bouguessa, and R. Ksantini, “Escaping feature twist: A variational graph auto-encoder for node clustering,” *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 2022. doi:10.24963/ijcai.2022/465.

[33]. W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, and X. Li, “Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data,” *BMC Bioinformatics*, vol. 18, no. 1, 2017. doi:10.1186/s12859-016-1415-9.

[34]. S. S. Deepika and T. V. Geetha, “A meta-learning framework using representation learning to predict drug-drug interaction,” *Journal of Biomedical Informatics*, vol. 84, pp. 136–147, 2018. doi:10.1016/j.jbi.2018.06.015.

[35]. M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003. doi:10.1162/089976603321780317.

[36]. S. Vilar, E. Uriarte, L. Santana, N. P. Tatonetti, and C. Friedman, “Detection of drug-drug interactions by modeling interaction profile fingerprints,” *PLoS ONE*, vol. 8, no. 3, 2013. doi:10.1371/journal.pone.0058321.

[37]. S. Cao, W. Lu, and Q. Xu, “Grarep,” *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015. doi:10.1145/2806416.2806512.

[38]. W. Zhang, H. Zou, L. Luo, Q. Liu, W. Wu, and W. Xiao, “Predicting potential side effects of drugs by recommender methods and ensemble learning,” Neurocomputing, vol. 173, pp. 979–987, 2016. doi:10.1016/j.neucom.2015.08.054.

[39]. K. Park, D. Kim, S. Ha, and D. Lee, "Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks," PLOS ONE, vol. 10, no. 10, p. e0140816, 2015. doi:10.1371/journal.pone.0140816.

[40]. J.-Y. Shi, H. Huang, J.-X. Li, P. Lei, Y.-N. Zhang, K. Dong, and S.-M. Yiu, "TMFUF: A triple matrix factorization-based unified framework for predicting comprehensive drug-drug interactions of new drugs," *BMC Bioinformatics*, vol. 19, no. S14, 2018. doi:10.1186/s12859-018-2379-8.

[41]. G. Shtar, L. Rokach, and B. Shapira, "Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures," PLOS ONE, vol. 14, no. 8, p. e0219796, 2019. doi:10.1371/journal.pone.0219796.

[42]. N. Rohani, C. Eslahchi, and A. Katanforoush, "Iscmf: integrated similarity-constrained matrix factorization for drug–drug interaction prediction," Network Modeling Analysis in Health Informatics and Bioinformatics, vol. 9, pp. 1–8, 2020. doi:10.1007/s13721-019-0215-3.

[43]. J. Y. Ryu, H. U. Kim, and S. Y. Lee, "Deep learning improves prediction of drug–drug and drug–food interactions," *Proceedings of the National Academy of Sciences*, vol. 115, no. 18, 2018. doi:10.1073/pnas.1803294115.

[44]. T. Ma, C. Xiao, J. Zhou, and F. Wang, "Drug similarity integration through attentive Multi-view graph auto-encoders," *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018. doi:10.24963/ijcai.2018/483

[45]. S. Liu, Z. Huang, Y. Qiu, Y.-P. P. Chen, and W. Zhang, "Structural Network embedding using Multi-modal Deep auto-encoders for predicting drug-drug interactions," *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019. doi:10.1109/bibm47256.2019.8983337.

[46]. G. Lee, C. Park, and J. Ahn, "Novel deep learning model for more accurate prediction of drug-drug interaction effects," BMC Bioinformatics, vol. 20, no. 1, p. 1-8, 2019. doi:10.1186/s12859-019-3013-0.

[47]. X. Hou, J. You, and P. Hu, "Predicting drug-drug interactions using deep neural network," in Proceedings of the 2019 11th International Conference on Machine Learning and Computing, 2019, pp. 168-172. doi:10.1145/3318299.3318323.

[48]. Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, and S. Liu, “A multimodal deep learning framework for predicting drug–drug interaction events,” Bioinformatics, vol. 36, no. 15, pp. 4316–4322, 2020. doi:10.1093/bioinformatics/btaa501.

[49]. Y.-H. Feng, S.-W. Zhang, and J.-Y. Shi, “DPDDI: A deep predictor for drug-drug interactions,” *BMC Bioinformatics*, vol. 21, no. 1, 2020. doi:10.1186/s12859-020-03724-x.

[50]. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on Graph Neural Networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021. doi:10.1109/tnnls.2020.2978386.

[51]. R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, and M. Dumontier, “Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings,” *BMC Bioinformatics*, vol. 20, no. 1, 2019. doi:10.1186/s12859-019-3284-5.

[52]. Y. Yu, K. Huang, C. Zhang, L. M. Glass, J. Sun, and C. Xiao, “SumGNN: Multi-typed drug interaction prediction via efficient knowledge graph summarization,” *Bioinformatics*, vol. 37, no. 18, pp. 2988–2995, 2021. doi:10.1093/bioinformatics/btab207.

[53]. Z.-H. Ren, C.-Q. Yu, L.-P. Li, Z.-H. You, Y.-J. Guan, X.-F. Wang, and J. Pan, “BioDKG–DDI: Predicting drug–drug interactions based on drug knowledge graph fusing biochemical information,” *Briefings in Functional Genomics*, vol. 21, no. 3, pp. 216–229, 2022. doi:10.1093/bfgp/elac004.

[54]. X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, “KGAT,” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019. doi:10.1145/3292500.3330989.

[55]. L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, and D. Lee, “DETERRENT: Knowledge guided graph attention network for detecting healthcare misinformation,” *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020. doi:10.1145/3394486.3403092.

[56]. Q. Li, W. Lin, Z. Liu, and A. Prorok, “Message-aware graph attention networks for large-scale multi-robot path planning,” *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5533–5540, 2021. doi:10.1109/lra.2021.3077863.

[57]. J. Wu, W. Shi, X. Cao, J. Chen, W. Lei, F. Zhang, and X. He, “DisenKGAT,” *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021. doi:10.1145/3459637.3482424.

[58]. W. Lan, Y. Dong, Q. Chen, R. Zheng, J. Liu, Y.-P. P. Pan, and Y. Chen, "KGANCDA: Predicting circrna-disease associations based on knowledge graph attention network," *Briefings in Bioinformatics*, vol. 23, no. 1, 2021. doi:10.1093/bib/bbab494.

[59]. X. Zhang, C. Zhang, J. Guo, C. Peng, Z. Niu, and X. Wu, "Graph attention network with dynamic representation of relations for knowledge graph completion," *Expert Systems with Applications*, vol. 219, p. 119616, 2023. doi:10.1016/j.eswa.2023.119616.

[60]. X. Lin, Z. Quan, Z.-J. Wang, T. Ma, and X. Zeng, "KGNN: Knowledge Graph Neural Network for drug-drug interaction prediction," *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020. doi:10.24963/ijcai.2020/380.

[61]. xzenglab, "KGNN," GitHub, https://github.com/xzenglab/KGNN/blob/master/log/kegg_result.txt (accessed Mar. 6, 2023).

[62]. Z. Yang, W. Zhong, Q. Lv, and C. Yu-Chian Chen, "Learning size-adaptive molecular substructures for explainable drug–drug interaction prediction by substructure-aware graph neural network," *Chemical Science*, vol. 13, no. 29, pp. 8693–8703, 2022. doi:10.1039/d2sc02023h.

[63]. X. Hao, Q. Chen, H. Pan, J. Qiu, Y. Zhang, Q. Yu, and X. Du, "Enhancing drug–drug interaction prediction by three-way decision and knowledge graph embedding," *Granular Computing*, vol. 8, no. 1, pp. 67–76, 2022. doi:10.1007/s41066-022-00315-4.

[64]. "Dapagliflozin and metformin uses, side effects & warnings," Drugs.com, https://www.drugs.com/mtm/dapagliflozin-and-metformin.html (accessed Mar. 6, 2023).