Received: April 30, 2024. Revised: May 27, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Graph Representation Learning for Predicting Diverse Sources of Drug Interactions

Swathi Mirthika G. L[1] B. Sivakumar[1]*

[1]*Department of Computing Technologies, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur - 603 203, Chengalpattu, Tamil Nadu, India*
* Corresponding author's Email: sivakumb2@srmist.edu.in

Abstract: Drug treatment strategies to reduce dose-related hazards is a tried-and-true method for preventing drug resistance and enhancing the efficiency of the monotherapy. Except when certain drugs pile up. Most adverse medication effects are induced by antagonistic drug-drug interactions. New medications and monitoring patients' use of more effective medication combination therapies require precise Drug-Drug Interaction (DDI) prediction. Several machine learning-based DDI prediction methods exist. This wide range of strategies uses drug-related and substance-related traits covertly. Graph embeddings and deep learning are applied to benchmark datasets to overcome this. The Simplified Molecular Input Line Entry System (SMILE) method is introduced for preprocessing, and the GCNet is applied for DDI prediction. Moreover, the graph is also constructed based on that the similarity is identified using link prediction. The proposed method provides an accuracy range of 0.934, Mean Squared Error (MSE) of 0.082, and Root Mean Squared Error (RMSE) of 0.352, which assists in more effectively reducing adverse drug reactions.

Keywords: Adverse effect, Conv-LSTM, Drug - Drug interaction, DrugBank, Embedding, Heterogeneous graph.

1. Introduction

Patients are increasingly using various pharmacological medications to address their ailments. Drug interactions can diminish the efficacy of one or more medications, as well as cause negative impacts. The danger of side effects is increased in groups, such as elderly, who frequently yield many drugs at once. On the other hand, some drug interactions create synergistic action that allows the drug to treat diseases such as cancer [1]. Even though many technologies are available to predict cancer in humans, developing an anti-cancer treatment remains a difficult task [2]. Cancer patients require an effective combination of multiple medications known as drug-drug interaction (DDI) [3]. The DDI is the change in the effect of a drug on the body when taken together with other drugs [4]. It has the potential to raise, postpone, reduce, or produce undesirable effects with either medicine. However, predicting an efficient drug combination remains a difficult task [5]. The rapid accumulation of high-precision and large-scale biological data has resulted in an establishment of the research discipline known as computational pharmacology. Thus, this data enables systematic study of diverse datasets. Analysing this data can help enhance drug development and reduce risk.

Bonds within a chemical substance are commonly used in biological processes. As a result, biological data is typically represented using networks. The establishment of this biological network necessitates the development of new analytical computing techniques. New research has attempted to solve this problem. Detecting the presence of DDIs is an initial step toward avoiding potential negative consequences. DDIs is classified as pharmacokinetic (PK), pharmaceutical, or the pharmacodynamic (PD) [6]. In addition, DDIs predominantly produce the modification in PK results in the secondary change in their PD. Therefore, dividing the DDIs types is another task, usually through multiple tests in medical research [7], which

helps the manufacturers and scientific community to decreasing toxicity and maximizing the risk of these interactions [8]. However, there are also direct costs that occur during clinical studies. In addition, DDI is closely related to drugs. Due to the large number of different chemical compounds, it is difficult to explore the compound area by high-resolution analysis. These problems highlight the importance of developing new computational schemes for DDI prediction.

Several machine-learning DDI prediction methods are identified and classified into similarity, network, matrix synthesis, and learning-based. Similar schemes are a broad part of these techniques and refer to similar drugs that act on similar drugs. [9] identified new DDIs by structural features and fingertip interaction characteristics. The field has explored network approaches in which drug networks are built by known drug DDIs. Most of the graph-based techniques consider pairwise relationships among drugs. It works on the simple graphs in which every vertex is the vertex and every edge represents the connection among two vertices. Though, some of the techniques consider interactions among drugs and other biological entities to produce alternative designs. Network topology information is extracted to predict previously undiscovered drug interactions (i.e interactions). With recent improvements in graph neural networks (GNN), various GNN models have been introduced for DDI prediction problems, but some of them generate different graphs from multiple sources, others develop biological knowledge graphs by extracting duplicates from raw data, such as DrugBank [10].

The primary aim of this paper is to create a mechanism for DDI prediction. The DDI prediction is performed using GCNet, which has the benefits of the Squeeze-Excitation Networks (SENet) and Non-Local Networks (NLNet). GCNet gives the simple, rapid, and the effective approach to predictions. Furthermore, the Simple Molecule Intake Line Entry System (SMILE) algorithm is employed for pre-processing, which minimizes complexity and produces higher-quality pre-processed results. The graph is then created to determine the shortest path between two points. Graphs are extremely visual, making it simple to express complicated data and relationships in an understandable and concise manner. Following the building of graphs, similarity is determined using link prediction. Similarity-based approaches are efficient, because they use common neighbours among two nodes as the primary criterion to identify the structural similarity. Increase in the structural similarity may indicate the relationship among two nodes is likely to form. Finally, the DDI is predicted by GCNet.

The key contributions of this work include:

GCNet is the algorithm used to forecast DDI. The GCNet is trained on test data to predict drug interactions. The output represents a likelihood of interaction among input pairs. Two drugs produce the likelihood that is greater than threshold, termed as potential interacting drugs.The research article is ordered as follows: Section 2 discusses the literature on several DDI predictions. Section 3 provides a brief elucidation of suggested approach; Section 4 presents the architecture of GCNet; and Section 5 presents project findings and details. Section 6 summarizes paper.

2. Related works

Some of the recent studies relevant to the presented work are reviewed below.

Currently in the medical industry, multiple drugs are combined together to create an optimal medicine for eliminating pathologic processes. However, the collaboration of multiple drugs may create adverse impacts on patient health because of DDIs. The evolution of artificial intelligence (AI) in the pharma field provided a way to determine DDIs, and evaluate the side effects between the drugs. Thanh Hoa Vo et al. [11] presented an innovative algorithm named Ensemble Deep Neural Network (EDNN) for enhancing the DDI prediction performances. This methodology was trained and tested to predict 86 types of DDIs on a benchmark database, and it achieved an accuracy of 90.80%. Although this approach obtained moderate accuracy in predicting DDIs, it faced issues like limited generalization to new drug interactions.

On the other hand, Mei Ma, and Xiujuan Le [12] proposed a dual graph neural network DGNN for predicting DDIs. This study aimed at forecasting the DDIs and molecular representations in drugs. Firstly, a substructure attention module with directed message passing neural network (SA-DMPNN) was developed to extract substructures adaptively. Then, the DDIs are segregated into pairwise interactions, enabling it to predict molecular representations and DDIs more effectively. This methodology was validated using the real-world database, and it obtained accuracy of 89.65%. However, computational complexity is the biggest concern of this approach.

DDI prediction is significant for preventing the adverse effect of drugs in the human body. However, predicting DDIs is more complex particularly in case of new drugs where all the information is not

available. Hence, Khaled Mohammed Saifuddin et al. [13] proposed a distinct algorithm named Hypergraph Neural Network (HyGNN) based on SMILES for precise DDI forecasting. To extract the structure similarities of drug chemicals, a hypergraph was created from the SMILES strings. Further, an innovative attention-assisted hypergraph edge encoder was developed to illustrate the drug representation and to detect the interaction between the drug pairs. The implementation results of the study highlighted that this approach earned a f1-score of 92.61%. However, this strategy demands more computational resources and a high quality database for reliable DDI predictions. Shichao Liu et al. [14] presented a

DDI detection mechanism using Deep Attention Neural Network (DANN). Firstly, a multiple drug attribute network was created and then using a graph embedding approach the drug representations are learned from the constructed network. Further, drug-drug pairs are learned from the attention module, and DNN was employed for detecting the DDIs from the learned representations. This methodology incurred accuracy of 89.77%. However, training this approach is complex and consumes more time.

Shanchen Pang et al. [15] introduced an innovative DDI prediction framework using attention mechanism-assisted multidimensional feature encoder (AMDE). In this approach, the drug attributes are encoded from multiple dimensions, enabling the system to predict DDIs more effectively. This framework was validated using the publicly accessible DrugBank dataset, and it gained accuracy of 91.24%. However, this algorithm offers limited scalability and reliability across diverse drug cases. Consequently, the study conducted by Chengcheng Zhang et al. [16] used Convolutional Neural Networks for precisely learning and predicting DDIs. This research aimed at providing a reliable solution to the developing medical industry to reduce the adverse impacts of diverse drug combinations. However, this methodology is ineffective in handling drugs with unknown or limited information.

Hui Yu et al. [17] presented a DDI forecasting mechanism using a hybrid algorithm named Relation-Aware Neural Embedding (RANE) approach. This study aimed to resolve the challenges faced by the conventional embedding methods in DDI predictions. This approach not only analyzes multirelational data between drugs, but also hybrid relation-aware network architecture to determine drug embedding. The experimental results depict that this strategy earned better F1-score and accuracy. However, multidimensional data analysis introduces additional computational overhead. Shenggeng Lin et al. [18] presented a study to address the issues with the hybrid utilization of multiple drugs in the medical field for decelerating pathogenic diseases. This study employed a transformer self-attention mechanism (TSAM) for the accurate detection of DDIs. Firstly, the two drugs are combined in four distinct ways using the Siamese Network. Then, the drug pairs and their interactions are examined using TSAM. The implementation outcomes suggest that this strategy earned 0.889% f-measure in DDI prediction. However, this strategy lacks interpretability and it is highly data dependent. Xiaorui Su et al. [19] presented an Attention-assisted Knowledge Graph Representation learning for enhancing the performance of DDI prediction. Firstly, the drug representations are initialized with their embedding obtained from drug features. Then, the interactions between the drugs are learned recursively using the proposed technique. This strategy was experimented and validated using two distinct datasets with varying sizes, and it demonstrated this methodology offers improved scalability and accuracy in DDI predictions. However, understanding the reason behind DDI prediction is difficult, making it complex for medical experts.

The research survey highlights the challenges imposed by AI-based DDI models. Although the AI-based models including the utilization of deep learning, machine learning, etc., offered better DDI predictions than the conventional models, they face significant limitations such as moderate accuracy, lack of generalizability, high computational complexity, etc. These drawbacks hinder their applicability in real-world scenarios, demanding a technique which offers higher accuracy with less computational demand. Also, few techniques face difficulty in analyzing the new drugs whose information is limited or not available, making them ineffective in real-world scenarios where the new drugs are introduced frequently. Moreover, some techniques face issues like data dependency, lack of interpretability, reduced scalability, and often demands more computational resources. These challenges make those techniques reduce their applicability in DDI predictions. To resolve the issues faced by the existing techniques, a novel DDI prediction algorithm was proposed by leveraging the efficiency of GCNet. The motivation behind this research is to tackle the problem of unanticipated DDIs harming the patient's health. The GCNet model leverages its heterogeneous network architecture for addressing this issue. This unique architecture enables the system to integrate diverse sources of drug information, making it to capture complex and intricate interactions between drugs more effectively

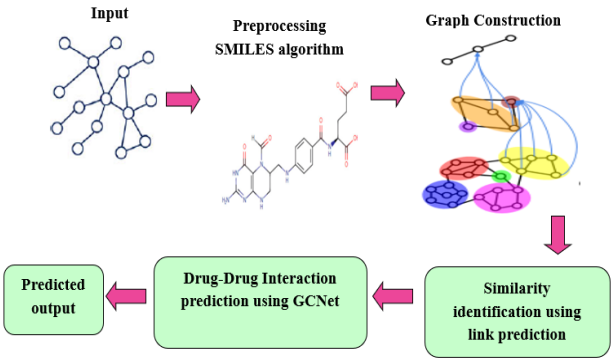


Figure. 1 Schematic view of GCNet for Drug-Drug interaction prediction

with less computational time. Also, examining the drug similarities obtained from undirected and unweighted networks pre-processed using the SMILES approach helps the proposed system to reduce the risk of unexpected adverse reactions. Furthermore, the reason behind the prediction of DDIs is transparent in the proposed approach; thereby enhancing its interpretability and applicability in real-world scenarios. This also helps in better decision-making and ensures patient safety.

3. Proposed Methodology

Combining numerous medications for complex disorders is becoming increasingly common due to the synergistic effects of DDIs. However, unanticipated DDI can cause side effects, unpleasant responses, and even severe toxicity, putting patients in the danger. As a demand for the multi-drug therapy grows, the identification of DDIs becomes increasingly critical. Nevertheless, detecting DDIs among the large amount of medication pairings, both in vivo and in vitro, is expensive and time-consuming. As a result, in this study, the GCNet is introduced to forecast DDI. The undirected and unweighted networks are first pre-processed using the SMILES approach, before the graph is formed. After the graph has been created, the similarity is determined using link prediction. Finally, the DDI is predicted using a heterogeneous network, specifically the GCNet. Fig. 1 displays GCNet's block structure for DDI prediction.

3.1 Datasets

DrugBank: It is an encyclopedic Web library that contains all pharmacological and biochemical data about medications, which includes biological schemes and target data. The majority of content in the DrugBank is carefully chosen from scholarly publications. In current, the DrugBank contains 10376 medication entries and 577712 directed interactions between them. In this investigation, DrugBank version 5.0 is used that is downloaded from DrugBank Webpage (https://www.drugbank.ca) in April 2024. In addition, DDI information is analyzed from XML file to create edge list of drug identifier combinations.

KEGG: Kyoto Encyclopaedia of Genes and Genomes (KEGG) is the most comprehensive biomedical resources, containing metabolic pathways from numerous species. In addition, the KEGG DRUG is a thorough list of licensed pharmaceuticals in the Europe, the United States, and the Japan, unified based on the chemical structures. KEGG DRUG graphically represents chemical structural pattern groups, therapeutic categories, their interactions, and drug development history. The dataset is taken from, ftp://ftp.genome.jp/pub/kegg/medicus/drug, accessed on April 2024. The KEGG DRUG contains 10,340 drug entries and the 500,254 directed interactions. Moreover, the mapping to DrugBank identifiers yields 1194 distinct drugs and 52609 directed interactions.

PubMed: Provides authorised source that seeks to enhance global and individual health by facilitating the retrieval of information about published research in the fields of biomedicine and the biosciences. There are numerous references and descriptions of scientific publications that exceed 2 billion.

3.2 Technique

In this segment, data acquisition, preprocessing, graph construction, similarity identification using link prediction are discussed.

3.2.1. Data acquisition

The DDI requires the construction of a two-class classifier that, given the characteristics of two medications as input, generates the output indicating the medicines interact is present. In addition, the SMILES strings are accessible types of medication characteristics, despite the fact that numerous other types of characteristics can be used for medications.

Therefore, in the proposed methodology, SMILES characters are employed to establish a network of pharmaceuticals and generate more comprehensive drug characteristics. The information is extracted from the benchmark datasets. Fig.2 consists of the detailed information about the datasets used in the proposed architecture.

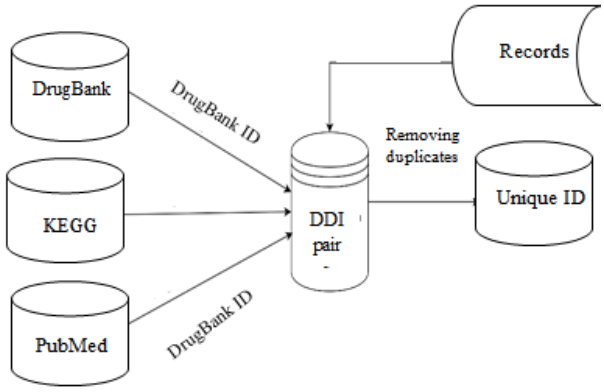


Figure. 2 Extraction of Drug-Drug interaction

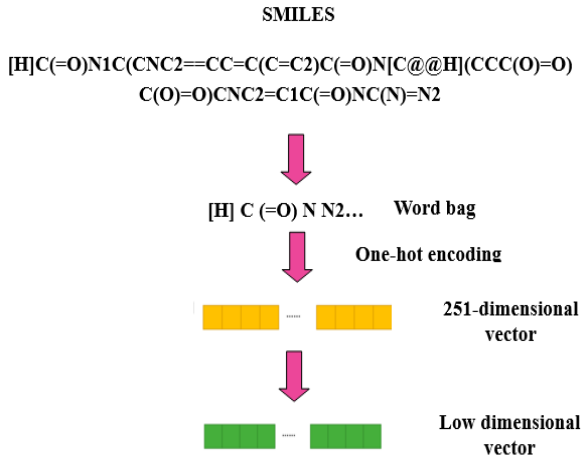


Figure. 3 SMILES pretreatment scheme

3.2.2. Preprocessing

The raw data obtained from the data sets must be cleansed to remove unwanted information. Preprocessing assists in enhancing accuracy and reducing redundant data. The data set for the drug bank comprises the name, weight, structure, description, and categories. For further classification, it is possible to acquire the necessary data of weight and structure.

The SMILES algorithm (B. Bumgardner et al., 2021) can be extracted from the PubMed dataset. Molecular editing software may import SMILES and convert them to two or three-dimensional molecules. The SMILES used Seq2seq technology on a SMILES string. During preprocessing, SMILES eliminates some of large (more than 250 letters) SMILES and performs one-hot coding on the remainder, transforming each SMILES into a vector of length 26. Fig. 3 depicts chemical structure of medication that is pretreated using this method. All SMILES saved in DrugBank are transformed into a word bag of 251 elements. After, one-hot encoding is employed to transform them into 251 dimensional vectors, which are then reduced to a specified dimension. As a result, a lower-dimensional vector is gained that can be utilized to represent a drug's structural feature.

3.2.3. Graph construction

The graph construction in this proposed work is founded on the substances. New varieties of components are used to produce the edges, which must be compared. Drugs are the vertices, and the edges are the connections between them. Assume an unweighted and undirected network that is represented by simple graph $H(U, C)$consisting set of nodes.

U refers to pharmaceuticals and set of edges C indicating drug interactions. Let {. indicate cardinality of set. The following equation is used for the comparison analysis.

$$H(U,C) = \begin{cases} 1, & if\ interaction\ exists \\ 0, & Otherwise \end{cases} \quad (1)$$

3.2.4. Similarity identification using link prediction

Each substance has a distinct molecular bond. Using drug similarity metrics, the link between a drug and another biological entity can be determined. Typically, drugs possess a number of characteristics that define their biological or chemical properties. Each element of the binary feature vector representing presence or absence of feature descriptor can be utilized to encode a single drug. The similarity between the drugs can be identified using the Jaccard similarity indexing techniques as follows:

$$SI = \frac{|y1 \cap y2|}{|y1 \cup y2|} \quad (2)$$

SI denotes the similarity index where $y1$ and $y2$are the drugs taken for calculating the similarity index. Heterogeneous graph will be created using the similarity index values.

Definition1:

A graph can be represented by the equation $X = (\mathrm{A,B,D},\varepsilon)$, where A is N nodes $(\mathrm{a1,a2,...,a}_n)$ and B is collection of edges connecting the various nodes. The letters D and ε denote the collection of different sorts of nodes and edges, in that order. When $A + B$'s sum is greater than 2, the graph X is said to be heterogeneous.

Definition 2:

Node wrapping in heterogeneous graphs. The goal of the node embedding algorithm, when applied to a heterogeneous network X, is to discover a function Ythat assigns a value to each node in X low-dimensional space with the coordinates

$$R^Z: Y: A \rightarrow R^Z \text{ where } Z << |A|. \quad (3)$$

4. Architecture

Heterogeneous Network:

A heterogeneous graph is the specialized type of the information network containing either distinct classes of entities or a large number of distinct classes of interconnections. Fig.4 provides the overall system architecture of the system and how the nodes embedded and the complete process has been depicted. It is a potent graph database application that provides precise results. The triples, which include the semantic relationship between nodes and edges, create a heterogeneous network. In the embedding space, network node and/or edge representations may take the form of compact yet informative vectors. Consequently, standard non-network machine learning techniques like Support Vector Machine (SVM), linear regression, and the decision forest, applied to vectors. All of these techniques have been demonstrated to be effective and cutting-edge. Resulting from their efficacy and potential in network analysis, network embedding techniques have spawned new research areas in biological data science. There are ongoing and forthcoming efforts to improve biomedical data analysis using network embedding. Numerous biological networks are composed of clinical text and other domain-specific data, and these networks are by nature scarce, chaotic, incomplete, and heterogeneous. It makes embedding duties more difficult compared to other application domains. To determine this objective, it is essential to analyze and compare existing network embedding models and investigate their applications to biomedical data. As a consequence, it can provide valuable information for future planning purposes.

GCNet:

GCNet is an advanced tool deployed for learning graph-structured data. This model functions by updating node attributes iteratively in accordance with the attributes of neighboring nodes using weighted average approach. The utilization of this approach enables this system to give importance to nodes with less connections, providing effective training and learning. Also, by deploying direct connections between the graphs and reconstructing embeddings as inputs for further layers, this model efficiently captures the interconnections and relationships within the constructed heterogeneous graph.

A procedure is used to predict the relationship between the two medications, which increases accuracy and reduces adverse effects. The primary benefit of GCNet is that it does not share weights between concealed layers. GCNet employs graph convolutional network to assist in the creation of the heterogeneous graph model.

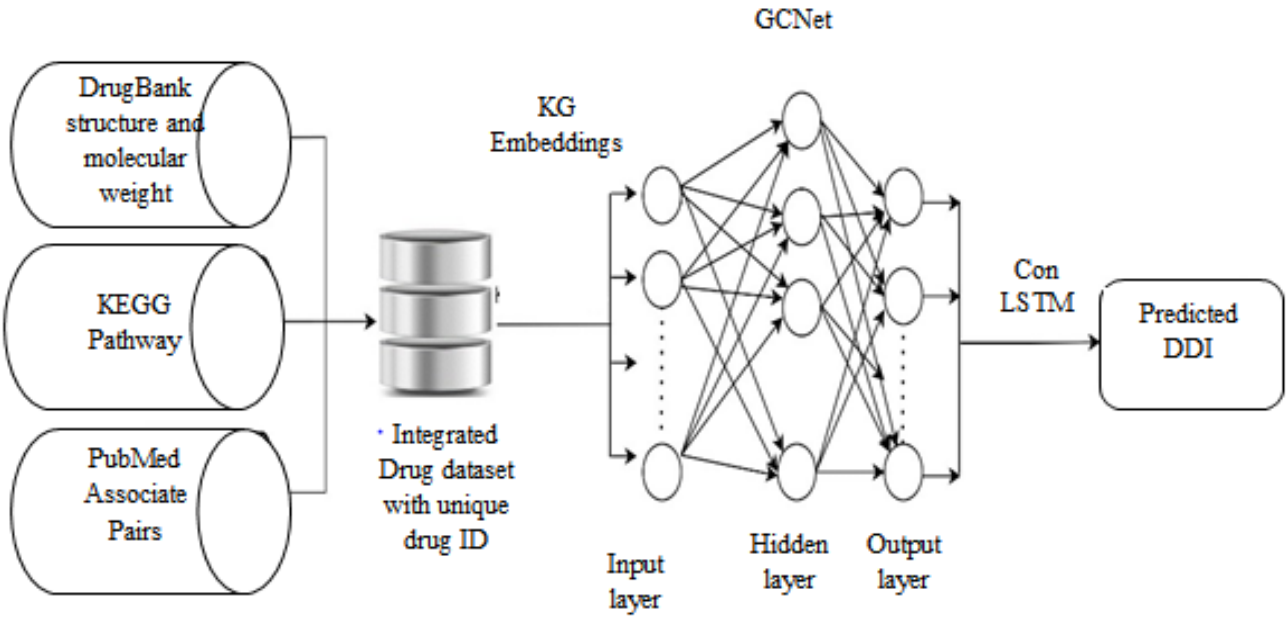


Figure. 4 GCNet architecture

Learning low dimensional representations of graph-structured nodes from architectures of multilayered, interconnected neural networks. In each layer of GCNet, direct connections between graphs collect neighbouring data to reconstruct embeddings as inputs for the subsequent layer. In semi-supervised graph learning, Convolutional Networks are utilized. The central principle of the GCN is to compute a weighted average of the characteristics of each node's neighbours, including itself, which is mathematically expressed in Eq. (3).

$$N_f(l+1) = \sigma\left(\tilde{D}^{1/2}\tilde{A}\tilde{D}^{1/2}N_f(l)W(l)\right) \quad (4)$$

Where $N_f(l+1)$ defines the node features at layer $(l+1)$, $\tilde{A}$ indicates the normalized adjacency matrix, $\tilde{D}$represents the degree matrix, σdenotes the activation function, and $W(l)$ indicates the weight matrix at layer l.

Nodes with a lesser degree has greater weights. Using the resulting feature vectors, a neural network is then trained.

KG (Knowledge Graph) Embedding:

Based on the information acquired, the knowledge graph is embedded, hypotheses is formed about the interactions between various medications. In contrast, Machine Learning classifiers frequently demand input vectors of fixed length. Encoding the graph's information into dense vectors, where the densely vector is generated by the KG embedding technique. KG embedding includes the steps of modelling objects and the relations, devising the scoring function, and the learning entity and relational representations. Knowledge graphs not only provide a precise framework for delineating the data, but also enable significant information extraction via the graph's underlying structure.

Constructing Network using GCNet:

After embedding the drug information, the convolution technique is used to construct the network. Create a graph convolution network. The path is determined using the conv LSTM technique. The input layer, the concealed layer, and the output layer are the layers present in a network. As input, embedded data will be provided. In previous experiments, negative samples from unknown interactions were chosen at random because all classifiers require them for the link estimation problem [20], [21].

When all unknown interactions are assigned to negative samples, there is the data imbalance. However, metrics, like AUPR and the F1-score are affected. To circumvent this problem, unsupervised clustering analysis or sampling from the unidentified interaction at a ratio equal to the optimistic set has been utilised.

Considering (1), (2), (3) and (4) are the four different localized computation graphs. The primary node in (1) is the drug d1, and its neighbors are all other drugs in (2), the central node is the drug d3, and its neighbors are all either other drugs or targets.

A target node can be found in the middle of (3) and (4). During the graph convolution, every drug interaction with its neighboring drugs by comparing the molecules, and then combining the interaction it received. The GCN is then trained with an activation function that improves its distribution-fitting skills. Each row of the adjacency matrices has been preprocessed with a softmax function to standardize the input data. More information will be incorporated into the forecast as the receptive field expands along with the GCN's depth.

To get an embedded representation of a drug node, the characteristics determined in steps 1 and 2 are added together. To a similar degree, the feature set of the targeted nodes can be derived in accordance with Eq. (5) and (6). The embedding of node is computed by:

$$h_d = h_d^1 + h_d^2 \quad (5)$$

$$h_t = h_t^3 + h_t^4 \quad (6)$$

where, $h_d$signifies embedding illustration of the drug node d ; the terms h_d^1 and h_d^2 signify the concealed positions of the node d in local design graphs (1) and (2); h_t denotes embedding depiction of target node $t$and the h_t^3, h_t^4 signify hidden states of node t in local calculation graphs (3) and (4), correspondingly. The classifiers effort better with the generated features. The primary layer is called the embedding layer, and its job is to convert a ‘sequence' representing a drug sample into an actual vector domain. The embed model, which has a form of 100x200, is then sent into a one-dimensional convolutional layer that has 100 filters and the kernel size of 3.

5. Results and discussion

The findings and analyses of a GCNet for DDI interaction prediction are discussed in this section.

5.1 Experimental setup

GCNet is executed in Python on Windows 10. Here, ideal layers and neurons in each layer are simulated. The GCNet performance is measured by accuracy and loss, and the expression is as follows.

Accuracy: It measures how closely expected and predicted values match. Furthermore, the equation used for the accuracy test is given below:

$$Accuracy = \frac{Number\ of\ predicted\ drugs}{Total\ number\ of\ given\ drugs} \quad (7)$$

MSE: The fitness function computes the optimal solution using an error function, as seen in the equation below:

$$MSE = \frac{1}{J}\sum_{k=1}^{J}[H - \gamma_u]^2 \quad (8)$$

where, $D$is GCNet output, $\gamma$is target output, and $J$is training samples.

RMSE: The square root of MSE is designated as the RMSE, and is shown as,

$$RMSE = \sqrt{\frac{1}{J}\sum_{k=1}^{J}[H - \gamma_u]} \quad (9)$$

5.2 Sample outcome

The accuracy and loss functions for training and validation for data sets are given in the Fig. 5 a) and Fig. 5 b). When comparing GCNet and its improvements to confGCN, GCNet comes out on top every time. The use of cross-entropy SoftMax V2 function is identified for best performance, rather than just plain cross-entropy, which resulted in a significant reduction in computing cost while performing hyper-parameter optimization. Therefore, loss function is employed in all following tests. GCNet is the fastest running discovered interactions.

5.3 Similarity measures evaluation

Topological and the semantic similarity scores are averaged across the drug-drug combinations. Prior to averaging, all similarities were scaled. Moreover, using the similarity method, the average similarity of the positive DDI pairings is considerably higher than that of the negative drug pairs and the random drug pairs for entire networks. The rank sum test is conducted to identify statistical significance, and all values were $p < 0.001$.

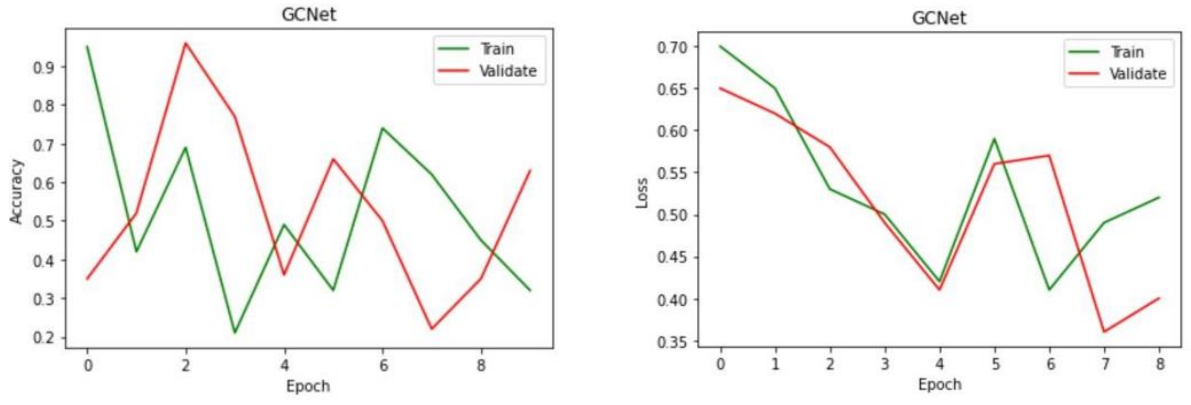
Fig. 6.i, ii, iii) visually depicts the differences between positive, negative, and random drug combinations for DrugBank, KEGG, and PubMed datasets. This finding verifies our main premise that related chemicals have a high potential for DDIs.

5.4 Comparative methods

In this module, we compare the performances obtained by the proposed framework with the state-of-the-art techniques. The metrics used for performance evaluation include accuracy, RMSE, and MSE. The techniques used in comparative evaluation include EDNN [11], DGNN [12], SA-DMPNN [13], HyGNN [14], DANN [15], and AMDE [16].

The Accuracy metric measures the overall correctness of the models in predicting DDIs. The accuracy of the model was determined by increasing the number of drugs.

The state-of-the-art techniques including EDNN, DGNN, SA-DMPNN, HyGNN, DANN and AMDE incurred accuracy rates of 0.89%, 0.84%, 0.86%, 0.924%, 0.89%, and 0.91%, respectively for four drugs. On the other hand, the developed strategy achieved comparatively higher accuracy of 0.934%, demonstrating its effectiveness in predicting DDIs. Moreover, it is observed that the proposed algorithm maintained consistent accuracy rates over increasing numbers of drugs, which highlights its scalability and reliability in processing numerous drugs with greater accuracy. Fig. 7 (i) presents the comparative



(a) (b)

Figure. 5 Estimated output: (a) Accuracy and (b) Loss

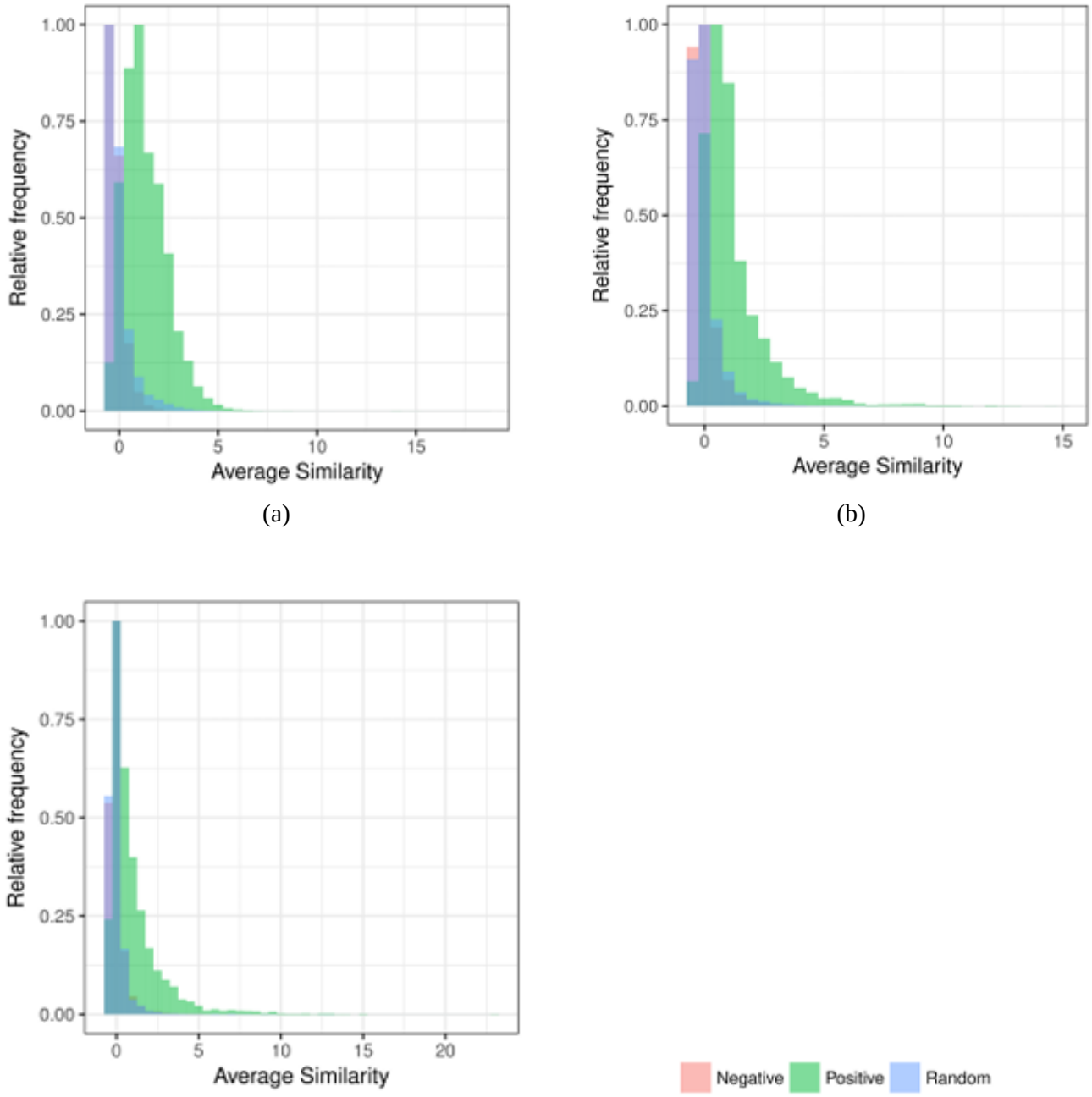


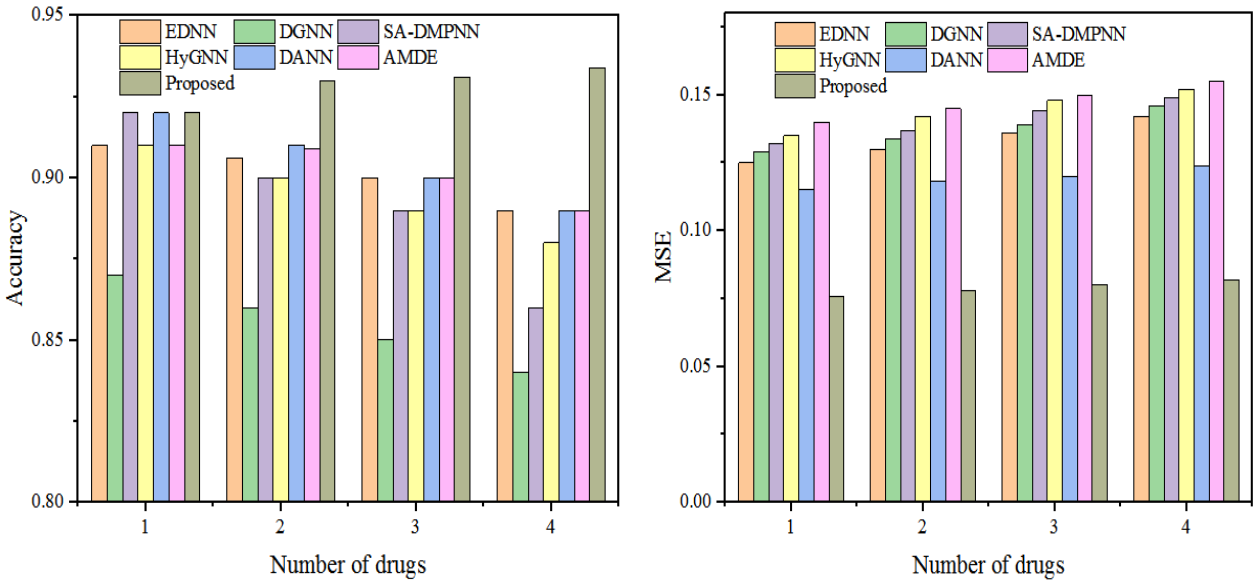
Figure. 6 Distribution of the average similarity among positive, negative, and the random drug-drug pairs: (a) DrugBank, (b) KEGG, and (c) PubMed

evaluation of prediction accuracy with existing techniques. Consequently, the MSE metric is determined and compared with the above-stated state-of-the-art techniques. The MSE metric measures the deviation between the actual and predicted DDIs in the input dataset. The lower MSE manifests greater precision in identifying DDIs. The above-mentioned approaches incurred MSE of 0.142%, 0.146%, 0.149%, 0.152%, 0.124%, and 0.155%, respectively for four drugs.

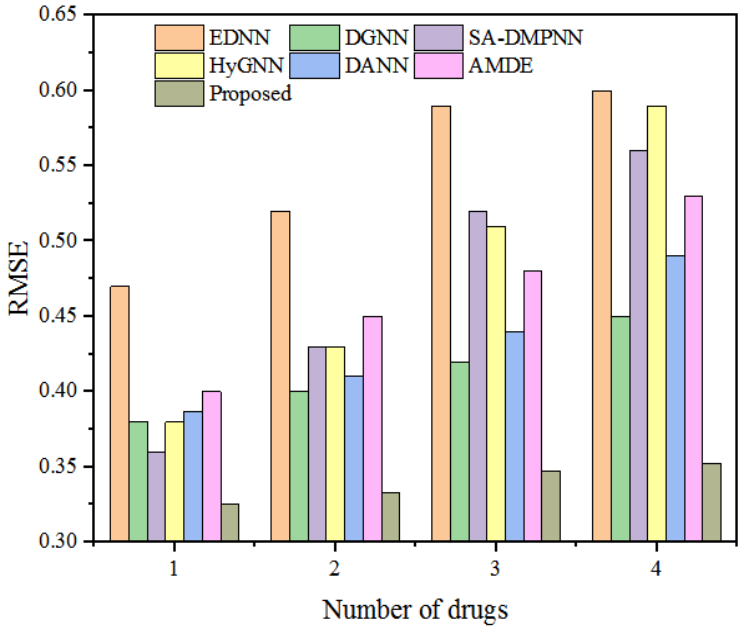
But the developed approach earned a minimum MSE of 0.082%, which illustrates that the error in DDI prediction is minimum and negligible. This significant reduction of MSE manifests the proposed model's effectiveness in accurately identifying DDIs. Also, the developed algorithm maintained a consistent MSE rate over increasing drug count, demonstrating its scalability and effectiveness in processing the drug features. Fig. 7 (ii) provides the comparative assessment of MSE.

Furthermore, the RMSE rate was compared and evaluated with the conventional models. Fig. 7 (iii) provides the comparison of RMSE. The RMSE metric determines the squared difference of the deviation between the real and the predicted results.

The existing techniques including EDNN, DGNN, SA-DMPNN, HyGNN, DANN and AMDE obtained RMSE of 0.60%, 0.450%, 0.560%, 0.590%, 0.490%, and 0.530%, respectively, while the developed algorithm attained minimum RMSE of 0.352%. The reduction of RMSE in the developed strategy highlights its effectiveness and reliability in predicting DDIs. From this analysis, it is clear that the proposed algorithm achieved higher accuracy than the conventional models. Also, the developed algorithm obtained minimum MSE and RMSE than the state-of-the-art techniques. This illustrates that the designed methodology is effective and reliable in identifying DDIs.



(a) (b)



(c)

Figure. 7 Performance comparison: (a) Accuracy, (b) MSE, and (c) RMSE

Table 1. Comparative analysis with state-of-the-art techniques

Algorithms	Accuracy	MSE	RMSE
EDNN	0.89	0.142	0.60
DGNN	0.84	0.146	0.450
SA-DMPNN	0.86	0.149	0.560
HyGNN	0.92	0.152	0.590
DANN	0.89	0.124	0.490
AMDE	0.91	0.155	0.530
Proposed	0.934	0.082	0.352

5.5 Discussion

This study presented a unique approach for predicting DDIs using the GCNet algorithm. The major objective of this study is to address the issues faced by the existing techniques such as higher computational time, limited scalability, reduced generalization, etc., in DDI prediction. The proposed framework uses three different databases namely DrugBank, KEGG, and PubMed, which provides important medication information needed for analysis. Also, we use SMILES strings for preprocessing, ensuring data standardization and consistency for subsequent analysis. Furthermore, a heterogeneous network was used in the proposed work to establish connections between drugs, enabling the system to capture and examine intricate relations and patterns within the drug features. The GCNet model leverages its efficiency in analyzing underlying mechanisms of drug interactions, offering improved accuracy in DDIs.

By leveraging the efficiency of embedding algorithms and convolutional networks, the proposed algorithm offers a reliable and promising solution for DDI prediction. The proposed strategy was modeled and implemented in a Python tool, and the results are determined using metrics such as accuracy, MSE, and RMSE. The proposed algorithm achieved an enhanced accuracy of 0.934, minimum MSE of 0.082, and lower RMSE of 0.352. Furthermore, a comprehensive comparative study was conducted with the state-of-the-art techniques including EDNN, DGNN, SA-DMPNN, HyGNN, DANN and AMDE to validate the effectiveness of the developed mechanism in predicting DDIs. The comparative evaluation manifests that the proposed strategy outperformed these techniques in terms of accuracy by 0.024%, and the metrics like MSE and RMSE are reduced by 0.042%, and 0.092%. Table 1 presents the comparative evaluation of proposed model performance with the conventional techniques. These performances make the developed strategy more effective and reliable for DDI prediction. Also, the designed technique maintained a consistent performance over increasing numbers of drugs, highlighting its scalability in processing large drug features.

6. Conclusion

This study proposed a novel DDI prediction strategy leveraging the efficiency of graph embeddings and deep learning approaches. The objective of this research is to analyze the interactions between the drugs to prevent its adverse impact on the human body. The proposed approach utilizes three different databases namely DrugBank, KEGG, and PubMed, which act as the foundation providing extensive medical information for drug feature analysis. Furthermore, the proposed framework uses the SMILE method for preprocessing, converting the dataset into an appropriate format for subsequent analysis. In addition, it uses heterogeneous graphs for capturing the intricate attributes in drug interaction analysis. The GCNet approach learns the interactions between the drugs through link prediction, offering optimal DDI prediction. The experimental results achieved notable accuracy of 0.934, MSE of 0.082, and RMSE of 0.370, offering a valuable tool for reducing adverse drug reactions more effectively. Furthermore, an intensive comparative analysis with the state-of-the-art techniques highlighted the proposed model's effectiveness and scalability across real-world scenarios. The comparative assessment depicts that compared to the existing models, the accuracy metric was improved by 0.024%, MSE was reduced by 0.042%, and RMSE was minimized by 0.092% in the developed approach.

Notations:

U	Pharmaceuticals
C	Set of edges
SI	Similarity Index
{	Cardinality of set
y1, y2	Drugs taken to calculate SI
ε	Edges
X	Heterogeneous graph

A	Number of nodes
B	Connectivity edges
R	Wrapping node
h_d	Embedding illustration of the drug node d
t	Target node
h_t	Hidden states of node
γ	Target output
J	Training Samples
σ	Activation Function

Conflicts of Interest

There is no Conflict of Interest.

Author Contributions

"Conceptualization, Swathi Mirthika G. L and B. Sivakumar; methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation Swathi Mirthika G. L; writing—review and editing, Swathi Mirthika G. L and B. Sivakumar; visualization, Swathi Mirthika G. L and B. Sivakumar; supervision, B. Sivakumar; project administration, B. Sivakumar".

Acknowledgments

This work was not supported by any organization and funding agencies.

References

[1] Z. Yu, F. Huang, X. Zhao, W. Xiao, and W. Zhang, "Predicting drug-disease associations through layer attention graph convolutional network", *Briefings in Bioinformatics*, Vol. 22, No. 4, 2020.

[2] E. A. Souri, R. Laddach, S. N. Karagiannis, L. G. Papageorgiou, and S. Tsoka, "Novel drug-target interactions via link prediction and network embedding", *BMC Bioinformatics*, Vol. 23, No. 1, 2022.

[3] X. Y. Yan, P. W. Yin, X. M. Wu, and J. X. Han, "Prediction of the Drug-Drug Interaction Types with the Unified Embedding Features from Drug Similarity Networks", *Frontiers in Pharmacology*, Vol. 12, No. 1, 2021.

[4] 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.

[5] B. Bumgardner, F. Tanvir, K. M. Saifuddin, and E. Akbas, "Drug-Drug Interaction Prediction: a Purely SMILES Based Approach", In: *Proc. of 2021 IEEE International Conf. on Big Data (Big Data)*, 2021.

[6] N. N. Wang, et al, "Machine learning to predict metabolic drug interactions related to cytochrome P450 isozymes", *Journal of Cheminformatics*, Vol. 14, No. 1, 2022.

[7] T. H. Vo, N. T. K. Nguyen, Q. H. Kha, and N. Q. K. Le, "On the road to explainable AI in drug-drug interactions prediction: A systematic review", *Computational and Structural Biotechnology Journal*, Vol. 20, pp. 2112-2123, 2022.

[8] K. Shao, Y. Zhang, Y. Wen, Z. Zhang, S. He, and X. Bo, "DTI-HETA: prediction of drug-target interactions based on GCN and GAT on heterogeneous graph", *Briefings in Bioinformatics*, Vol. 23, No. 3, 2022.

[9] D. Cao, N. Xiao, Y. J. Li, W.B. Zeng, Y. Z. Liang, A. P. Lu, et al, "Integrating Multiple Evidence Sources to Predict Adverse Drug Reactions Based on a Systems Pharmacology Model", *CPT: Pharmacometrics & Systems Pharmacology*, Vol. 4, No. 9, pp. 498-506, 2015.

[10] N. Rohani, and C. Eslahchi, "Drug-Drug Interaction Predicting by Neural Network Using Integrated Similarity", *Scientific Reports*, Vol. 9, No. 1, 2019.

[11] T. H. Vo, N. Thi, & Le. "Improved prediction of drug-drug interactions using ensemble deep neural networks", *Medicine in Drug Discovery*, Vol. 17, pp. 100149-100149, 2023.

[12] M. Mei, and X. Lei. "A dual graph neural network for drug-drug interactions prediction based on molecular structure and interactions", *PLOS Computational Biology*, Vol. 19, No. 1, 2023.

[13] Saifuddin, K. Mohammed, B. Bumgardner, F. Tanvir, and E. Akbas, "Hygnn: Drug-drug interaction prediction via hypergraph neural network", In: *Proc. of IEEE 39th International Conf. on Data Engineering (ICDE)*, pp. 1503-1516, 2023.

[14] L. Shichao, Y. Zhang, Y. Cui, Y. Qiu, Y. Deng, Z. Zhang, and W. Zhang. "Enhancing drug-drug interaction prediction using deep attention neural networks", *IEEE/ACM Transactions on*

Computational Biology and Bioinformatics, Vol. 20, No. 2, pp 976-985, 2022.

[15] S. Pang, Y. Zhang, T. Song, X. Zhang, X. Wang, & A. R. Patón, "AMDE: a novel attention-mechanism-based multidimensional feature encoder for drug-drug interaction prediction", *Briefings in Bioinformatics*, Vol. 23, No. 1, 2021.

[16] Zhang, Chengcheng, Y. Lu, and T. Zang, "CNN-DDI: a learning-based method for predicting drug-drug interactions using convolution neural networks", *BMC Bioinformatics*, Vol. 23, No. 1, pp. 88, 2022.

[17] Yu, Hui, W. Dong, and J. Shi. "Raneddi: relation-aware network embedding for drug-drug interaction prediction", *Information Sciences*, Vol. 58, No. 2, pp. 167-180, 2022.

[18] Lin, Shenggeng, Y. Wang, L. Zhang, Y. Chu, Y. Liu, Y. Fang, M. Jiang, et al. "MDF-SA-DDI: predicting drug-drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism", *Briefings in Bioinformatics*, Vol. 23, No. 1, 2022.

[19] Su, Xiaorui, L. Hu, Z. You, P. Hu, and B. Zhao. "Attention-based knowledge graph representation learning for predicting drug-drug interactions", *Briefings in Bioinformatics*, Vol. 23, No. 3, 2022.

[20] G. G. Graham, et al, "Clinical Pharmacokinetics of Metformin", *Clinical Pharmacokinetics*, Vol. 50, No. 2, pp. 81-98, 2011

[21] P. Zhang, F. Wang, J. Hu, and R. Sorrentino, "Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects", *Scientific Reports*, Vol. 5, No. 1, 2015.