# KGCNA-CNN-BiLSTM：Knowledge graph and hybrid neural networks for drugs association prediction

**Shuang Liu**
Dalian Minzu University
**Xiaomin Zhu** ( 2224136339@qq.com )
Dalian Minzu University

**Research Article**

**Keywords:** drug association prediction, link forecasting, knowledge graph, graph convolutional network, Convolutional-BiLSTM network, attention mechanism

**Posted Date:** February 13th, 2023

**DOI:** https://doi.org/10.21203/rs.3.rs-2161967/v2



**Additional Declarations:** No competing interests reported.

# KGCNA-CNN-BiLSTM：Knowledge graph and hybrid neural networks for drugs association prediction

Shuang Liu, Xiaomin Zhu*

(*Department of Computer Science and Engineering, Dalian Minzu University , Dalian 116000, Liaoning,China*)

**Abstract** Drug association (DDIs) prediction is also called drug interaction prediction, which refers to the interactions between drug and drug that lead to unexpected side effects when two or more drugs are taken simultaneously or successively. Previous studies on DDIs prediction using methods such as molecular representation and network embedding were extremely complex, expensive and time-consuming, and were limited in acquiring rich neighborhood information about drug entities and their surroundings during the forecasting process. A drug linkage prediction method based on knowledge graphs and hybrid neural networks was proposed based on the deficiencies of the above methods. This method is mainly based on methods such as knowledge graphs, graph convolutional network, Convolutional-BiLSTM network and attention mechanisms to solve the limitations in acquiring rich neighborhood information about KG entities during the forecasting process. It transforms drug linkage prediction research into a link prediction problem and views drug relationships with known interactions as edges in the interaction graph. It can effectively discover interactions of unknown drugs; meanwhile, performance comparisons are performed with existing DDIs prediction methods. The results show that higher performance is achieved in terms of indicators such as ACC and F1 values, which validate the effectiveness of the

* Corresponding author：Xiaomin Zhu；Email：2224136339@qq.com

model. Finally, future directions in this field are proposed based on an analysis and summary of challenges faced by current DDIs predictions.

**Keywords** drug association prediction, link forecasting, knowledge graph, graph convolutional network, Convolutional-BiLSTM network, attention mechanism

# 1 Introduction

Drug association prediction is the application of link prediction that is important in the biomedical field, as shown in Figure 1: it refers to the simultaneous or sequential application of two or more drugs, under the participation of body factors, drugs because of interactions between each other and the occurrence of pharmacokinetic or pharmacodynamic changes[1], and clinically manifested as enhanced drug efficacy and aggravation of toxic side effects (reduced drug efficacy and reduced toxic side effects). As the probability of various complications gradually increases, the combination of medication has become a standard method for patients to treat diseases. However, after the use of drugs, those mentioned above, similar adverse reactions threatened the health and life of patients.

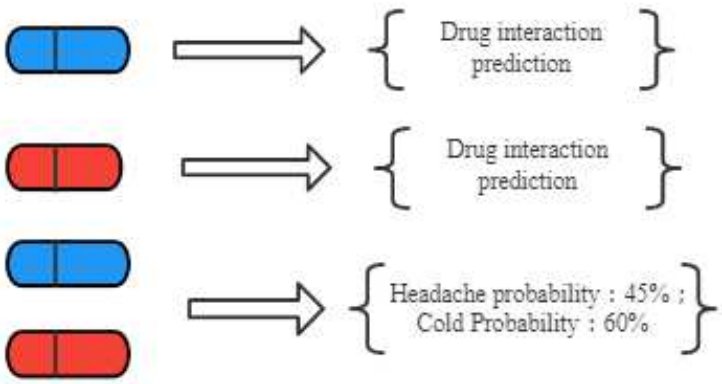


**Figure 1.** Drug association forecast introduction.

Data from the US Centers for Disease Control and prevention from 2016 to 2022 shows that 43.9% of Americans used more than five medications in the past 30 days, compared with only 10.3% using polypharmacy prior to this, illustrating that polypharmacy plays an important role in the medical treatment. In 2022, the adverse drug reactions in China showed the adverse effects of damaging body organs, as shown in Figure 2.

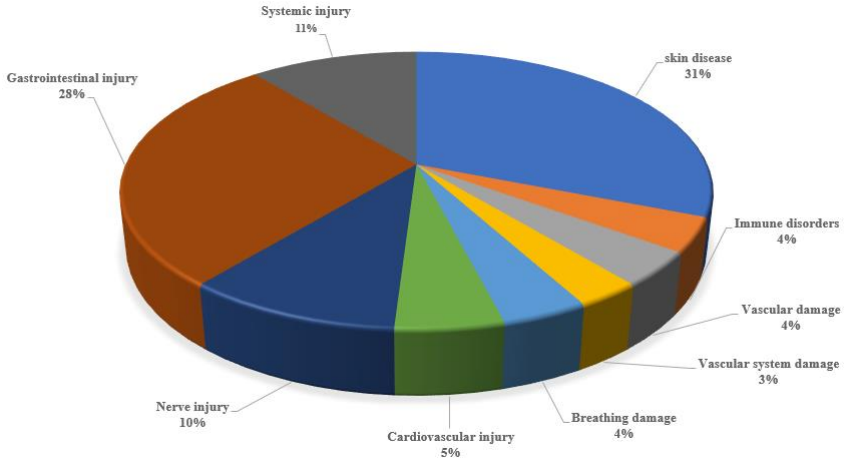


**Figure 2. 2022 Pharmaceutical adverse reactions and organ system conditions**

To prevent the potential threats and risks of DDIs on human health and improve the safety of drug combinations used[2], the prediction of DDIs is worthy of intensive research. Traditional prediction methods for DDIs rely on the results of data analysis after the drugs have been marketed, are time-consuming, and require statistical analysis of the data accumulated from drug use to obtain results. DDIs prediction significantly reduces the research scope of safety pharmacology and pharmaceuticals and reduces costs. The significance of this prediction is as follows:

（1）DDIs predictions can improve the chances of new adverse drug interaction discovery. By predicting a new DDIs role in advance, the drug development department or the drug safety department professionals can experiment with the expected drug pair to improve the efficiency of drug developers significantly.

（2）DDIs predictions can improve new drug development efficiency and safety. Through the fusion of the chemical structure information of the drug, DDIs predict that many medications have common chemical structure molecules, which will interact with some drug medicines. As a result, some high-risk chemical molecular structures intentionally avoiding these structures during drug development can reduce the chance of this to a certain extent.

（3）DDIs predictions can enhance the safety of personalized drug treatment plans. When recommending drugs for patients, adverse effects of the recommended and being taken can lead to severe consequences if the patient's medical history and medication status are not considered. DDIs prediction can recommend personalized treatments for patients based on their medical records, avoiding severe health threats to patients.

## 2 Related Work

DDIs prediction research will help the clinical research stage of drug research and development. The current DDIs method has the following:

（1）DDIs prediction based on similarity

This method mainly focuses on feature learning of drug molecules, representing those with similar embeddings that would exhibit similar DDIs performance, learning the similarity of drugs from multiple perspectives. It includes molecules' structural characteristics and other attribute characteristics to obtain a complete molecular representation[3].However, the disadvantages are limited to the expression learning of drug molecules themselves, which in most cases depends on the knowledge of relevant fields[4]. As shown in Figure 3: Assuming that drug A and drug B have similar molecular structures, it can be inferred that similar DDIs occur in drug B and drug C if drug

A and drug C interact.

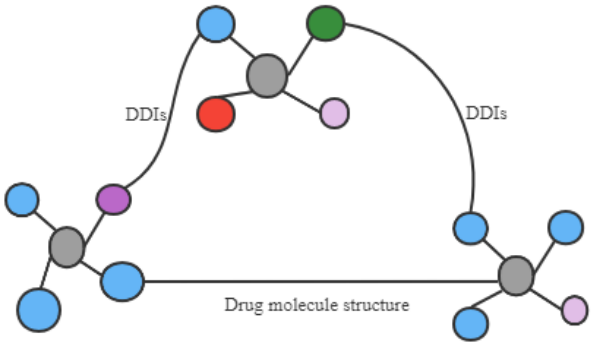


**Figure 3. Drug association prediction method based on molecular similarity**

（2）DDIs prediction based on network embedded

This method considers drugs as nodes to design the DDIs action relationship between drug nodes by building a biological network related to drugs[5]. The method is shown in Figure 4: there are also many methods to construct a mapping relationship network, such as matrix decomposition and random walk. Such methods aim to predict the label side between drugs, which only focuses on a single DDIs relationship and does not consider other drug-related relationships.

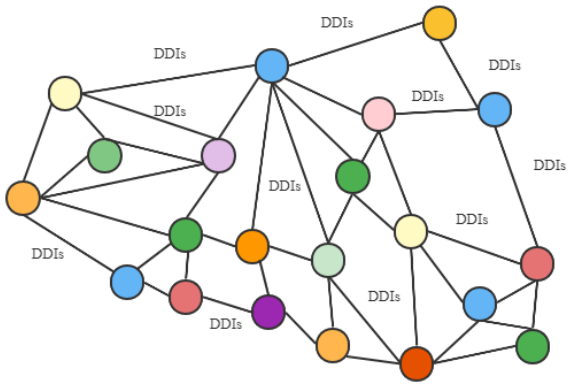


**Figure 4. Predictive prediction method based on network embedded drugs**

# 3 Experiment

## 3.1 Datasets and Experimental Settings

The DrugBank dataset contains 13791 drug entries, including 2653 approved small molecule drugs, 1417 approved biotechnology drugs, 131 nutrition products and 6451 experimental drugs. The KEGG dataset is a comprehensive database of drug-related information that distinguishes drugs according to chemical results and different chemical components. The basic statistics of the used datasets are summarized in Table 1.

Table 1 data set introduction

| | **Drugbank** | **KEGG** |
|---|---|---|
| **Drugs** | 2,578 | 1,925 |
| **Interactions** | 12,388 | 56,983 |
| **Entities** | 2,129,712 | 129,910 |
| **Relation Types** | 72 | 167 |
| **KG Triples** | 7,852,852 | 362,870 |

We compare KGCNA-CNN-BiLSTM against a variety of baselines which can be categorized as follow. The hyper-parameter settings of the two datasets are summarized in Table 2.

Table 2 The hyper-parameter settings for the two datasets

| **Parameter** | **Setting** | **Parameter** | **Setting** |
|---|---|---|---|
| **neighbor sampling size** | 4 | **learning rate** | 2e-1 |
| **dimension ofembedding** | 32 | **self.batch_size** | 70000 |
| **regularizer weight** | 2 | **self.n_epoch** | 50 |

| | | | |
|---|---|---|---|
| **learning rate** | 1e-7 | **self.checkpoint_verbose** | 1 |

### 3.2 Main Results

The KGCNA-CNN-BiLSTM-DDIs performance is evaluated through the two datasets of DRUGBANK and KEGG: the DDI approved by 8: 1: 1 is randomly divided into training, verification, and test sets. Sample replacement is used as a negative sample for model training. In terms of evaluation indicators such as ACC[22] and F1 value[23], compared with the current baseline method, it proves the model's effectiveness. MLP[24]， Deepwalk[25] and other methods are selected as experimental benchmarks on the Drugbank data set.

In this section, we compare the performance of the proposed method with the baselines. Tables 3 and 4 reports the average ACC and F1 scores across 67 runs on DrugBank and KEGG datasets, respectively. From the tables, we found that KGCNA-CNN-BiLSTM-DDIs significantly outperform the baselines across the two datasets. Specifically, it achieved at least 25% on ACC and 21% on F1 in DrugBank datasets. In addition, it reached at least 10% on ACC and 11% on F1, higher performance than other methods in KEGG datasets. The experimental results are shown in Table 3:

Table 3 Table of Drugbank datasets on each model experimental results

| Methods | ACC | F1 |
|---|---|---|
| **MLP (Rogers and Hahn, 2010)** | 0.61 | 0.81 |
| **DeepWalk (Perozzi et al, 2014)** | 0.64 | 0.76 |
| **Line (Tang et al, 2015)** | 0.60 | 0.68 |
| **Node2vce (Grover and Leskovec, 2016)** | 0.75 | 0.71 |
| **PRD (Wang, 2017)** | 0.70 | 0.80 |
| **Decagon (Zitnik et al, 2018)** | 0.67 | 0.87 |

| | | |
|---|---|---|
| **GAT (Velickovic et al, 2018)** | 0.73 | 0.77 |
| **SkipGNN (Huang et al, 2020b)** | 0.59 | 0.80 |
| **KG-DDIs（Karim *et al*,2021）** | 0.71 | 0.70 |
| **KGCNA-CNN-BiLSTM -DDIs（2022 Ours）** | **0.86** | **0.82** |

In the Drugbank dataset, the KGCNA-CNN-BiLSTM-DDIs method improved ACC metrics by 25%, 22%, 26%, 11%, 16%, 19% 13%, 27%, and 15%, respectively, compared with other benchmark methods. In addition, in F1 indexes compared with other benchmark methods, the improvement was 1%, 6%, 11%, 1%, 5% ,2%and 12%, respectively. Then select Struc2vec[26], DeepDDIS[27],KG-DDIs[28]and other methods on the KEGG dataset to compare the results of the experimental benchmark.The experimental results are shown in Table4:

Table 4 Experimental results of each model on the KEGG dataset

| **Methods** | **ACC** | **F1** |
|---|---|---|
| **Laplacian（Belkin and Niyogi,2013）** | 0.80 | 0.80 |
| **DeepWalk (Perozzi *et al,* 2014)** | 0.85 | 0.85 |
| **GreRepGraRep（Cao *etal*, 2015）** | 0.87 | 0.87 |
| **LINE (Tang *et al*, 2016)** | 0.86 | 0.86 |
| **SDNE(Wang *et al*, 2017)** | 0.86 | 0.85 |
| **GAE(Kipf and Welling, 2018)** | 0.75 | 0.78 |
| **Struc2Vec (Ribeiro *et al*, 2019)** | 0.84 | 0.84 |
| **DeepDDIs (Ryu *et al*, 2020)** | 0.82 | 0.79 |
| **KG-DDIs（Karim *et al*,2021）** | 0.81 | 0.81 |
| **KGCNA-CNN-BiLSTM -DDIs（2022.Ours）** | **0.90** | **0.90** |

In the KEGG dataset, compared with other benchmark methods, the method has improved ACC indicators by 10%, 5%, 3%, 4%, 4%, 15%, 6%, 28% and 9% respectively. In addition, the F1 index is increased by 1%, 6%, 14%, 11%, 2% and 5% respectively compared with other benchmark methods.

This is because KGCNA-CNN-BiLSTM-DDIs explore drug features and related entities in the knowledge graph, while the others like Laplacian, DeepWalk, and struc2vec only learn from similar drug features. Compared with LINE-based methods, which leverages graph features with similar connections for DDI prediction, it also achieved superior results. This is due to the fact that the drug embeddings in KGCNA-CNN-BiLSTM-DDIs can better capture the semantic similarity of relations than the graph embeddings used in NN-based methods. Moreover, compared to DeepDDI and KG-did, it achieved stable performances across datasets, and our three variants all achieved similar results. This is a very encouraging result. The reason could be that compared to DeepDDI, our method jointly considers topological neighborhood structures and related entities in the knowledge graph, which benefits the performance. It compared to KG-did, our model incorporates the GCN model to obtain the topological neighborhood representations of drugs and related entities, which can obtain more high-order structures and semantic relations than embedding-based methods for modeling drugs between drug pairs.

### 3.3 **Analysis**

The model's training process in this paper on the DrugBank dataset is shown in Figure 5 and Figure 6, which is the model's curve of ACC and F1 values. It can be seen from the figure that the ACC and F1 values of the front part of the model training increase rapidly, and the subsequent constant fluctuations seek the optimal local value and gradually become stable.

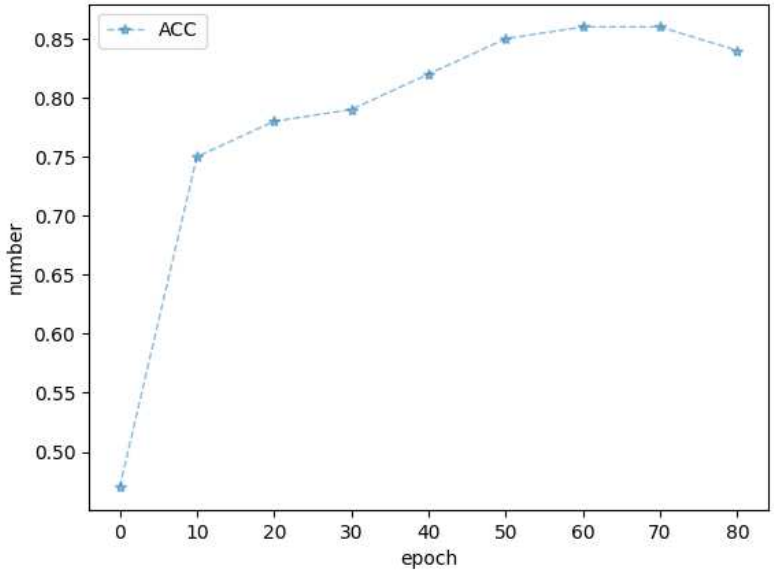


**Fig.5 ACC value graph on the DrugBank dataset**

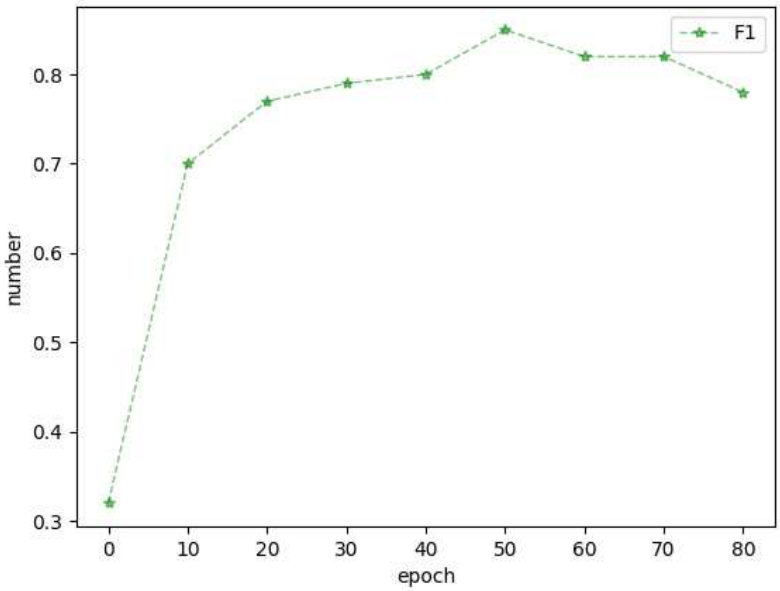


**Fig.6 F1 value graph on the Drugbank dataset**

The model's training process in this paper on the KEGG dataset is shown in Figure 7 and Figure 8, which is the model's curve of acc and F1 values. It can be seen from the figure that the acc and F1 values of the front part of the model training increase rapidly, and the subsequent constant fluctuations seek the optimal local value and gradually become stable.

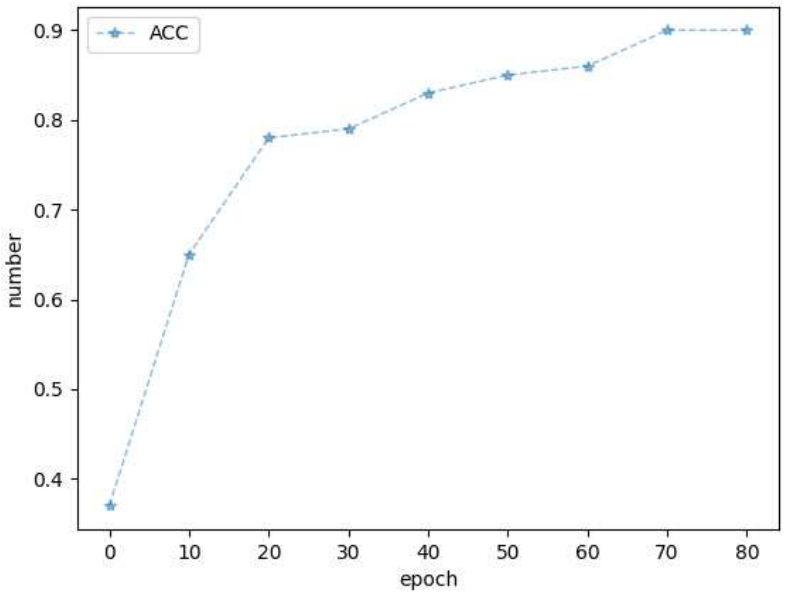


**Fig.7 ACC value graph on the KEGG dataset**

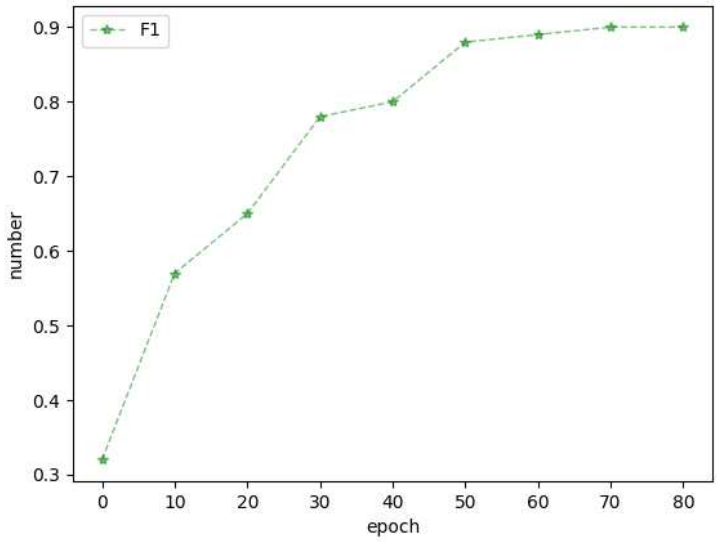


**Fig.8 F1 value graph on the KEGG dataset**

### 3.4 Case Study

The experimental results found that KGCNA-CNN-BiLSTM-DDIs achieved the best effect on both datasets. In addition to the above experiments, this article also sets up several key parameters: the impact of critical parameters such as K, H, D and other vital parameters on the proposed model performance. When studying one of the parameters, other parameters are fixed. First, change the

size of the neighborhood. The result is shown in Figure 9:K=16 is the best performance. Explain that if the number of neighbor nodes is too small, the information the neighborhood can contain is insufficient.

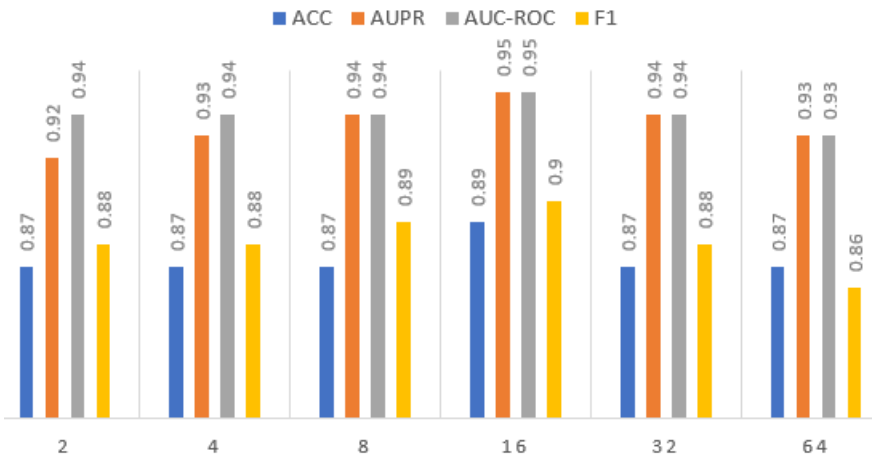


Figure 9. The impact of the size K on model on the model

Secondly, by setting the depth parameter H, the H value is set to 1 to 6, and the effect of research H on model performance. The experimental results are shown in Figure 10: It can be seen from the figure that when h = 2 can learn more features, the model performance is the best when the H = 3 model performance begins to decrease.

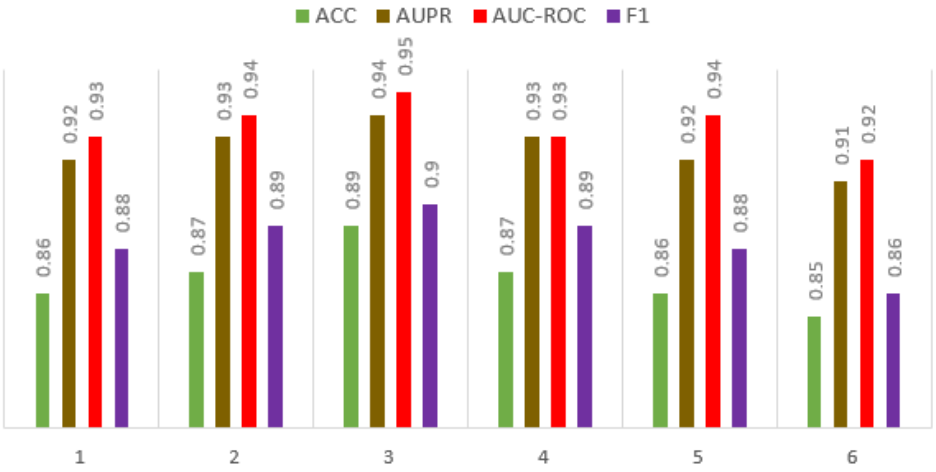


Figure 10. The effect of perception domain h on the model

Finally, the impact of the inspection embedded dimension D changes from 8 to 512. As shown

in Figure 11, D = 32 and 64 influence the model the most. It improves its learning ability by setting the appropriate dimension size. If the value is too large, it will bring overfitting.

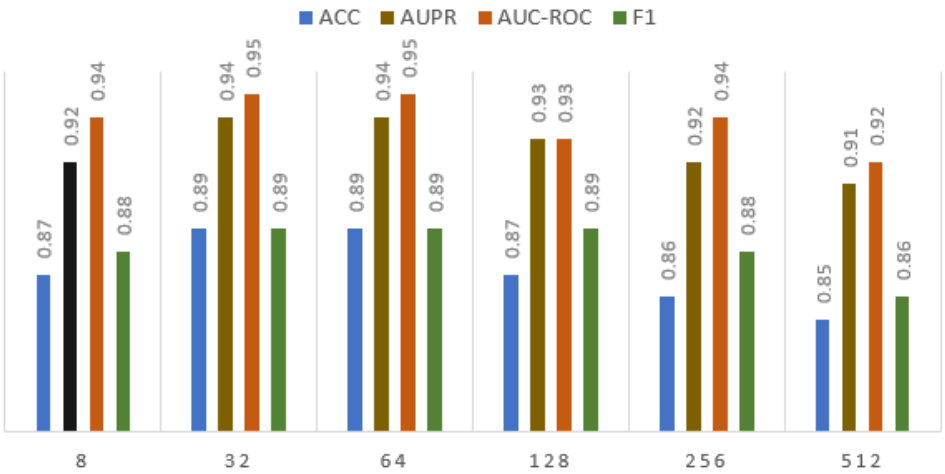


Figure 11. A sensitive analysis of the sample size K in the neighbor

### 3.5 Ablation experiments

In this paper, the ablation experiment is designed to analyze differently better. The influence of the module on this prediction is verified, as shown in Table 5. The influence of Graph convolution neural network, Convolutional BiLSTM,and Attention mechanism on the experimental results is discussed.

Table 5. The influence

| | **ACC （Drugbank）** | **F1 （Drugbank）** | **ACC （KEGG）** | **F1 （KEGG）** |
|---|---|---|---|---|
| **KGCNA-CNN-BiLSTM** | **0.86** | **0.82** | **0.90** | **0.90** |
| KGA- ConBiLSTM | 0.82 | 0.79 | 0.872 | 0.881 |
| KGCNA | 0.855 | 0.81 | 0.868 | 0.892 |
| ConBiLSTM-A | 0.84 | 0.815 | 0.883 | 0.895 |

The results of this ablation experiment illustrate that the model proposed in this paper can

thoroughly combine the advantages of three kinds of graph convolutional neural networks, convolutional-BiLSTM and attention mechanism to improve the prediction effect of DDIs of the whole model. Compared with other benchmark models, our model uses KGCNA-CNN-BiLSTM -DDIs to enable Convolutional-BiLSTM to first process the interacting drugs into sequence form, which can capture the long-distance dependence and location information of the original text and learn the protein representation of the amino acid sequence composition of drugs through Convolutional-BiLSTM recurrent neural network, thus enriching the feature extraction process. Additionally, the KGCNA-CNN-BiLSTM-DDIs DDIs attention mechanism-based feature method can simultaneously capture entity and relationship features in the neighborhood of any given entity. Both relationship clustering and multi-hop relationships were encapsulated in the model, thus capturing the information and association features in the multi-hop neighborhood of a given drug.

# 4 Conclusion

DDIs drug association forecast is one of the important topics of public health and drug safety monitoring. This article focuses on the background of the prediction of drug associations and the problems and significance of the existing DDIs prediction. The model of this article first applies the diagram curl network to the knowledge map. At the same time, the topology information of the drug entity in the knowledge map and the associated semantic information attached to it will help the clinical research stage of drug research and development. It can greatly help doctors to avoid side effects or adverse reactions when combined with medication, which has unusual significance for the clinical treatment of drugs. It can be seen from the research results in recent years at home and abroad that more and more researchers have conducted profound DDIs research through multi-channel, multi-method and multi-angle. In summary, this article blends the GCN model and

knowledge map and introduces the Convolutional-BiLSTM and attention mechanism models to DDIs prediction research value.

# 5 Methods

This paper begins by incorporating graph convolutional neural networks into knowledge maps to effectively solve the problem that only two drugs traditionally can be derived if they interact. Then, the Convolutional-BiLSTM network is introduced to enable the entity feature vector to fuse all neighborhood entity and relationship features, enriching the feature extraction process. Finally, introducing attention mechanisms can better capture the information and association features in a given multi-hop neighborhood, improving the drug association prediction quality.

## 5.1 Knowledge Graph

The KGCNA-CNN-BiLSTM method combines the knowledge map with the graph convolution network for drug association prediction. It better solves complex problems in the medical field by combining knowledge and data. The knowledge atlas is developed from a semantic network. A directed graph is constructed by representing entities as nodes and semantic relations between entities as edges connecting nodes[6]. That is, knowledge is represented as a more intuitive network structure. This way of representing data through graph structure enables computers to organize and manage a large amount of information data more efficiently and further realize vector embedding, retrieval, prediction and reasoning of knowledge[7]. An example of knowledge map is shown in Figure 12：

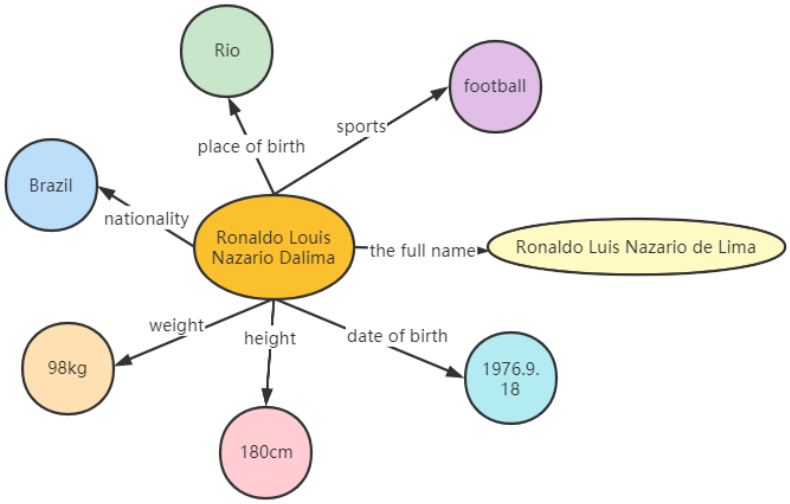


**Figure 12. an example of knowledge graph**

### 5.2 Graph Convolutional Network

The graph convolutional network is introduced into the knowledge map to sample the neighborhood of each node and aggregate the neighborhood information to obtain the entity-embedded representation. The graph neural network is introduced into the knowledge map to sample the neighborhood of each node and aggregate the neighborhood information to obtain the entity-embedded representation[8].The model is shown in Figure 13:

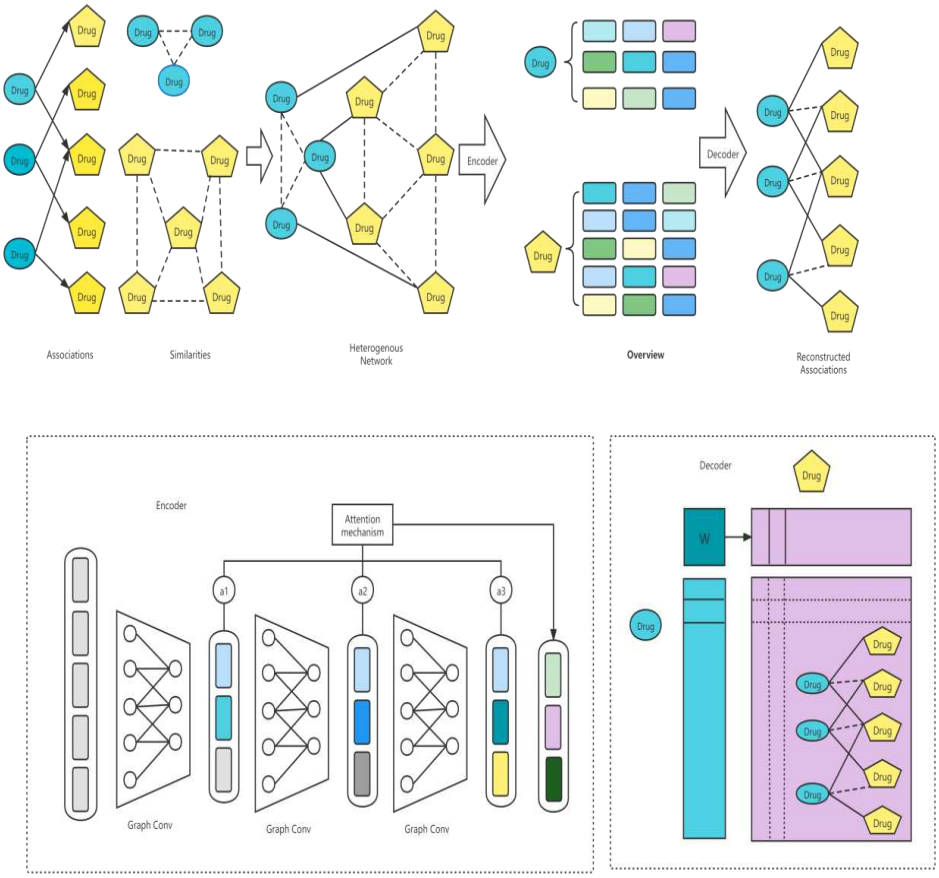


**Figure 13. graph convolutional network**

### 5.3 Convolutional-BiLSTM Network

Convolutional-BiLSTM networks can be used as a module by more complex structures, such as a coding network and a prediction network, where the initial state and cell output of the prediction network are copied directly from the last state of the encoding network. The network is mainly composed of several Convolutional-BiLSTM layers. Because the predicted target is the same as the input dimension[9], all states are spliced in the prediction network, and then the final prediction is generated through the convolution layer. Encoding's BiLSTM compresses the entire input sequence into a hidden layer state tensor, then predicts BiLSTM to expand the hidden layer state, giving the final prediction. Its input is an M * N grid, and the channel size of each grid is P. That is, the input is a 3D tensor. Padding will be used during convolution to ensure that these states have duplicate

rows and columns as the input. The padding of hidden layer boundary points is calculated using the external world's state[10]. Before the first input, all States of the BiLSTM to be initialized are 0. If we use 0-padding in the hidden layer, we set the external world state to zero and pad different boundary points[11]. The model is shown in Figure 14:

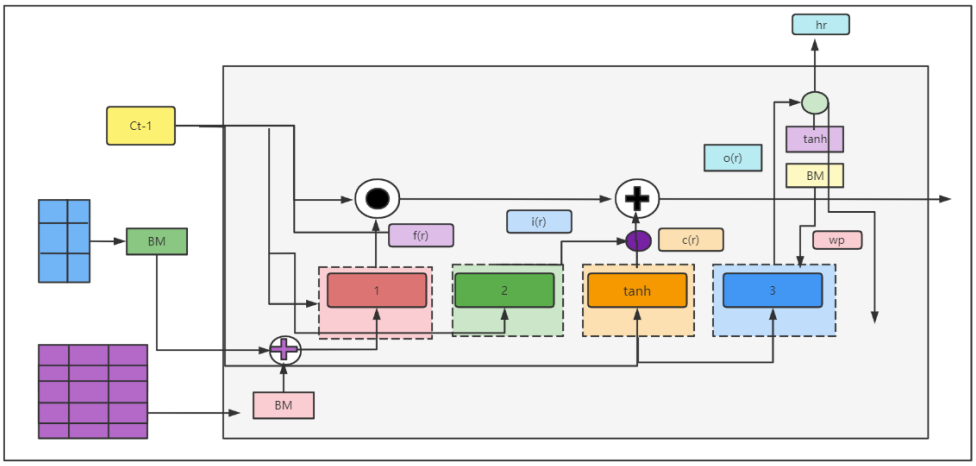


**Figure 14. Convolutional-BiLSTM Network**

### 5.4 Attention mechanism

The framework introduces the attention mechanism model to capture the entity and relationship characteristics of the given entity multi-neighborhood so that the model can specify different rights on different neighbor nodes. It can avoid the excessive amount of information in collecting influential neighbor nodes that cause noise to affect the prediction results and achieve explanations[12]. In addition, it also solves the critical data of semantic vectors that cannot pay attention to the sequence. When the acquisition vector is sent to the network model one by one, a series of the coding-end will hide to participate in calculating the attention coefficient, and the atomicity weight integrates the atom into the molecular representation. Based on the characteristics and methods of attention mechanism, the entity and relationship characteristics can capture simultaneously in any given entity. It can be encapsulated in the model association characteristics[13].

Provide compelling insights for DDIs prediction. The model is shown in Figure 15:

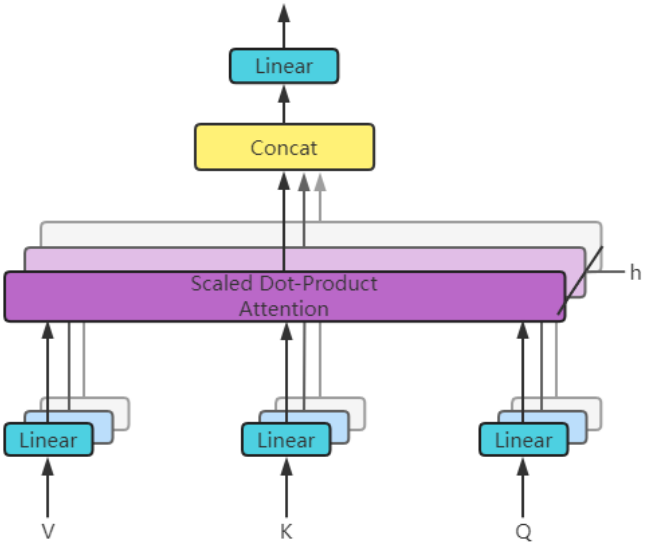


**Figure 15. Attention Mechanism Model**

# 6 The Proposed KGCNA-CNN-BiLSTM

This article proposes the following improvement ideas based on the traditional method: combining the knowledge map technology with the convolutional neural network. At the same time, the Convolutional-BiLSTM network and the attention mechanism are introduced to realize the drug interaction prediction[14].The overall model framework is shown in Figure 16:

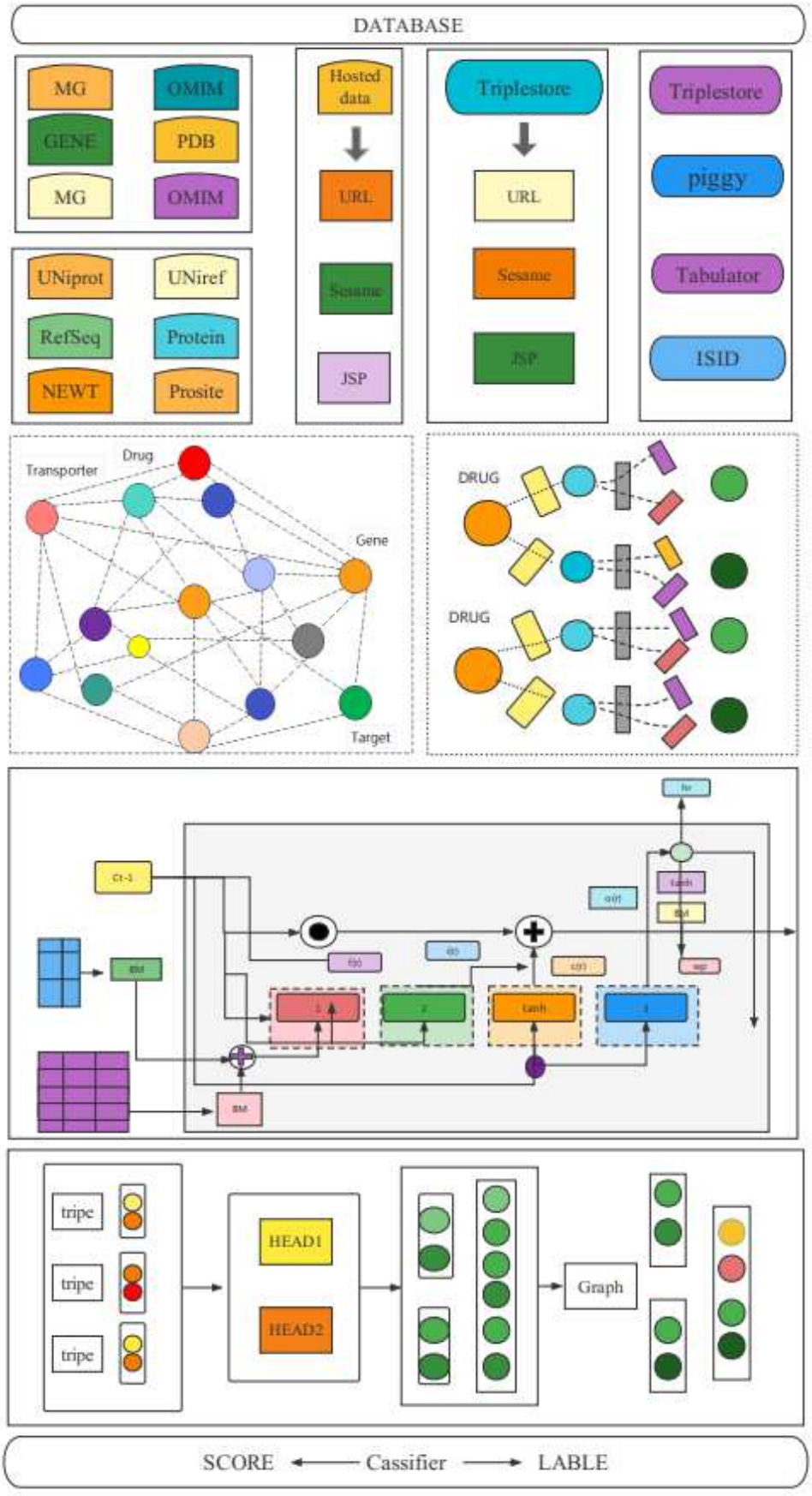


**Figure 16. Overall model framework**

Drug-related prediction technology routes based on this model have been divided into four parts: (1) extraction of DDIs drugs. (2) Construction of the knowledge graph. (3) Model Framework Part: Combining diagram curly neural networks, Convolutional-BiLSTM networks with the Attention model, and then combined with drug-associated predictions. (4) Drug-Drug similarity Calculation: Calculate the similarity of the two drugs through the predictive module of drugs and drugs and feedback on the interactive output value.

### 6.1 Problem Formulation

(1) Based on BIO2RDF Knowledge Graph Construction: Use the BIO2RDF[15] tool to build a link data network. Based on transmission definitions, obtain data from different format data sources, and create link data compatible with the RDF data format. Convert the data to the three yuan group format to build the knowledge map.

（2）Methods based on knowledge maps and diagrams: Some of the previous DDIs calculation methods tend to combine advanced embedded techniques, try to use automatic learning drugs, and then model DDIs through operations such as matrix decomposition and random walking. Because DDIs modeling does not consider other correlations, in response to this problem, the prediction problem is defined as a link prediction problem on the graph. It is extended to the neural network method on the graph structure to be embedded in non-linearly learning nodes. In addition, a potential DDIs relationship is obtained by reconstructing a new adjacent matrix or reconstruction of the link on the new graph. Based on the knowledge diagram, the model introduces the diagram convolutional neural network for drug association prediction[16]. Borrowing the diagram convolutional neural network is sampled from each node, and the entity embedded by the neighborhood information is obtained by aggregating neighborhood information. The intellectual sheet generated by the three

yuan group in the DDIs prediction research combination achieves better prediction results. Effective capture of KG Chinese drug pairs of high-level neighborhood topology verification based on knowledge maps and diagram convolutional neural networks have advantages and effectiveness in DDIs prediction tasks.

（3）Methods based on the Convolutional-BiLSTM network model: Because the existing method has the problem of insufficient path structure and noise in the path, it will adopt a long and short-term memory network Convolutional-BiLSTM to capture the way characteristics comprehensively. The design is based on the Convolutional-BiLSTM model and attention mechanism Multi-path fusion algorithm. Introduce the Convolutional-BiLSTM network to learn the drug atom's protein sequence of amino acids, which is expressed by enriching the feature extraction. At the same time, combined with the local sub-structure structure and the global path structure, joint reasoning, fully tap and use the diagram structure information of the knowledge map, and verify the effectiveness of prediction through experimental verification of the Convolutional-BiLSTM network.

（4）Based on the method of attention mechanism attention: Due to the lack of power building capabilities, noise, noise, and the traditional way of knowledge map link forecasting, it is not adequate to learn the entity during the prediction of the knowledge map link forecasting. The rich semantics and potential relationship information exist in the neighboring domain. Based on the knowledge map convolution network and Convolutional-BiLSTM,the attention mechanism is added to obtain the ternary group information associated with each entity. With attention to the impact of different neighborhood triad groups on the target entity, the characteristic vector of the physical features with attention weights is used to make the biological feature vector integrate all neighboring

physical characteristics and their corresponding relationship features. Finally, calculate the similarity of the two drugs and feedback on the interactive output value, and perform DDIs prediction.

### 6.2 DDIs Extraction

Filter the data required from the data set for drugs for extraction; DDI extraction mainly uses Drugbank[17] and KEGG[18] datasets, first screening out the database. In the analysis of the DDI matrix format, each line has three columns. The first two columns represent the code number of the two drugs in the dataset.” DB” represents the data set from Drugbank, and the third column represents the label. When the Lable is equal to "1", it indicates an interaction relationship between the two drugs; When the Lable is equal to "0", it does not exist. The model is shown in Figure 17:

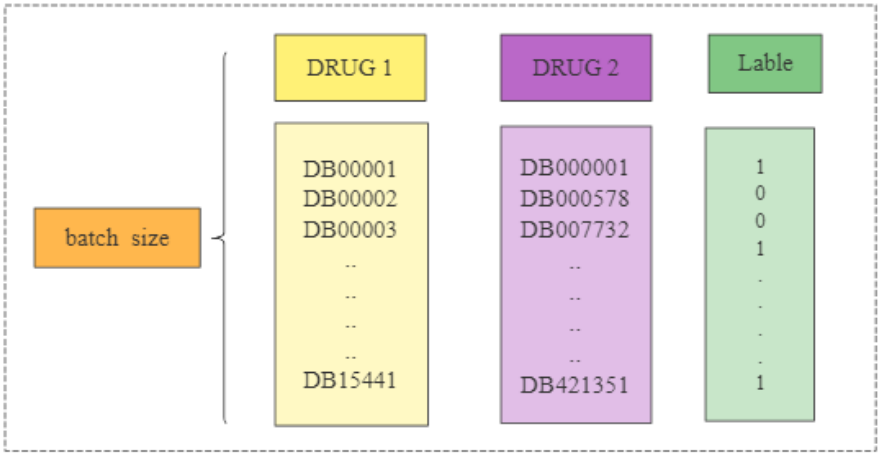


**Figure 17. DDI matrix**

### 6.3 KG Construction

KG Construction: Data uses the BIO2RDF tool[19] to build a link data network. After obtaining the data from different format data sources, the data format data compatible with the data format of the RDF[20] data is received. Treat the dataset as a ternary group and create about 5 billion RDF

ternary groups[21] for constructing the knowledge graph. The knowledge is shown in Figure 18:

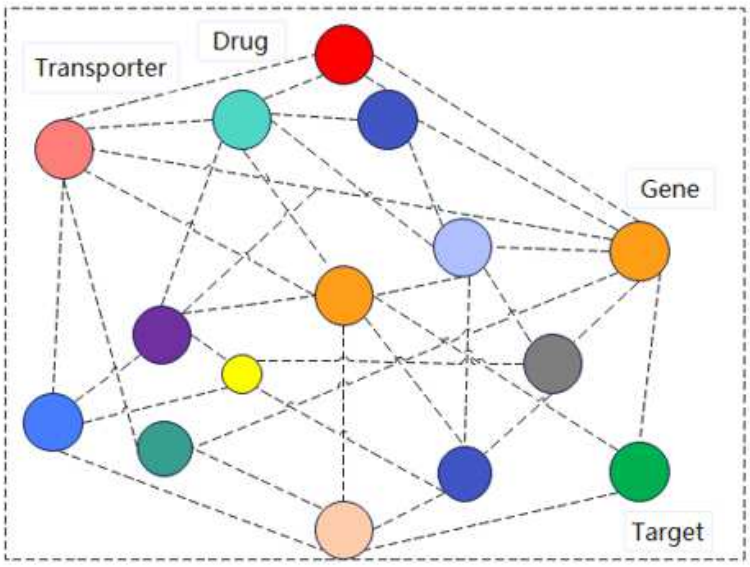


**Figure 18. Knowledge graph**

Select the constructive knowledge map for the knowledge diagram. It is known for constructing and generating the pharmaceutical diagram by crawling the Drugbank database pharmaceutical relationship data set. The expected effect is shown in Figure 19:

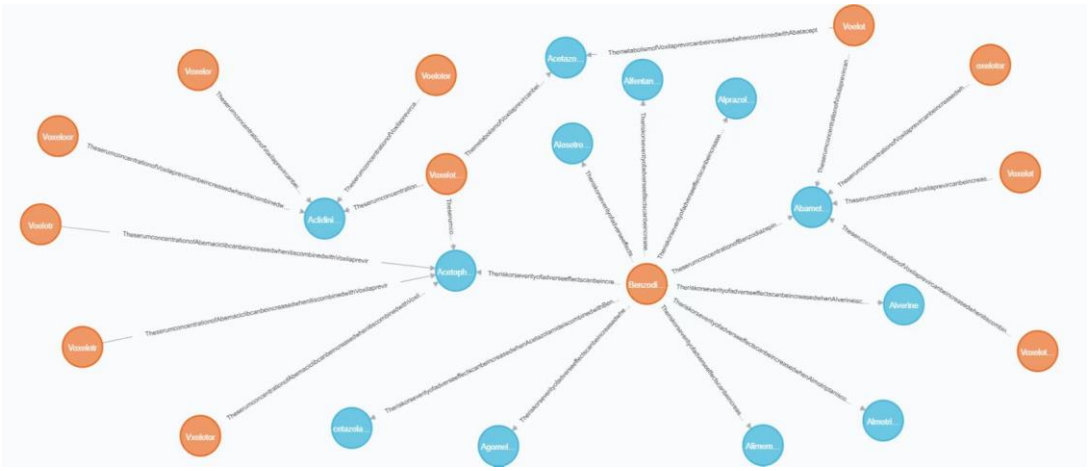

**Figure 19. Given knowledge graph**

## 6.4 KGCNA-CNN-BiLSTM Layer

First, the information input model is used to obtain the characteristics of drugs and their related

entity neighborhoods. The models transported first GCN models:

（1）Use the constructed knowledge map and DDIs matrix as input. After the input is obtained, the neighbor sample must be sampled for the drug entity. Then converge the information as the input of the following network. When h =1, consider only neighbor nodes directly connected to the current node. When h =2, consider the second-order node situation, and you can learn more about the physical information of the neighborhood. After the sampling is completed, the embedded representation of the entity and the neighborhood information is aggregated by the aggregation method. The embedded representation of the current entity is finally obtained.The drug characteristics will go through a linear transformation and a layer of GCN, and then there will be an interactive process between the two pictures. The formula is as follows:

$$h_{I\,I}^{1}=F_1(h_I^0 W_I^0) \quad (1)$$

$$h_{I\,D}^{1}=F_1(h_I^0 W_I'^0) \quad (2)$$

$$h_{D\,D}^{1}=F_D(H_D^0 W_D^0) \quad (3)$$

$$h_{D\,D}^{1}=F_D(H_D^0 W'^0_D) \quad (4)$$

(2)It is the diagonal matrix, D is the degree matrix, and W is the training weight matrix. Before activating the function, the modules on the left and right sides should be interacting and fusion of information. The formula is as shown below:

$$h_I^1=\sigma h_{II}^0+\sigma W_I^0 \quad (5)$$

$$h_D^1=\sigma h_{DD}^1+\sigma W_I^0 \quad (6)$$

(3)After aggregating the upper GCN model as input, the neighbor information is used as input, and the Convolutional-BiLSTM model is used to mine the drug serialization. Before using the Convolutional-BiLSTM model, the long-distance dependency relationship and location information

is represented by the protein composed of a pharmaceutical amino acid sequence through the Convolutional-BiLSTM cycle neural network and the process of rich feature extraction.

$$H_t = \sigma\left(W_{xh} * X_t + W_{yh} * H_t + W_{ch} * C_t\right) + b_h \quad (7)$$

$$F_t = \sigma\left(W_{xf} * X_t + W_{hf} * F_t + W_{cf} * C_t\right) + b_f \quad (8)$$

$$C_t = \sigma(W_{xc} * X_t + W_{hc} * F_t + W_{cc} * C_t) + b_c \quad (9)$$

$$O_t = \sigma(W_{xo} * X_t + W_{ho} * F_t + W_{co} * C_t) + b_o \quad (10)$$

$$\mathrm{Z}=O_t * \tanh(C_t) \quad (11)$$

（4）Attention mechanism layer:Input the upper-layer output as an attention mechanism model, and design an attention mechanism for each feature diagram to effectively capture the importance of local and global neighbors and the local and global representations of learning nodes. Use formula calculation to score:

①For a node, the node directly connected to the node in the figure is defined as a local neighbor. Considering that the importance of different neighbors is different, the nodes are learned through the design attention mechanism:

$$a_{mn}^{l}(v_m, v_n)=\mathrm{f}(W_{a,a'}^{m}, W_{a,a'}^{n}) \quad (12)$$

$a_{mn}^{l}$indicates that the importance of neighbor $v_m$ to $v_n$ is the attention score，$v_m$ represents the special n-dimensional vector of gene m, f (·) represents a single layer feed neural network and W represents the weight matrix of the learning right.

②The formula is performed as follows:

$$\propto_{mn}^{l}=\frac{\exp\left(a_{mn}^{l}\right)}{\sum_{x\epsilon N_m^n}\exp\left(a_{mx}^{l}\right)} \quad (13)$$

$\propto_{mn}^{l}$ represents the attention scoring after the normalization, x represents the selected node, $N_m^n$ represents the global neighbor set of the node and $a_{mx}^{l}$ represents the calculated attention

score.

③At the same time, $v_i$ based on the representation of local neighbor information aggregation nodes, the formula is shown below:

$$g_m^l=\emptyset(\sum_{n\epsilon N_m^n} \alpha_{mn}^l . \omega_1^y g_n) \quad (14)$$

$g_m^l$ represents the overall situation of the node, $N_m^n$ represents the global neighbor's set of nodes, $\omega_1^y$ represents the weight shared by local global neighbors and $g_n$ represents the representation of local neighbors.

④Due to the unstable attention coefficient, a single node attention mechanism may introduce noise. The formula is as follows:

$$g_m^l=\coprod_{y=1}^{y} = \emptyset(\sum_{n\epsilon N_m^n} . W_1^y g_n) \quad (15)$$

$N_m^n$ indicates the global neighbor set of nodes.

**6.5 DDIs Prediction**

To predict DDIs, potential representations of output drugs and neighboring topology information between the current drug pair. Translate the medication into multiple types of feature vectors, and use different similarities to calculate the similarity of various drugs and drugs based on the characteristics to obtain DDIS similarity measurement and compare the two similarities to other features. Calculate the score by the score function, the output interaction prediction value.

⑤ Use three full connection layers to perform DDIs predictions and use cross-entropy loss：

$$L_1=-\frac{1}{N}\sum_{ij}[y_{ij}\ log_{p_{ij}}+(1-y_{ij})\ \log(1-_{P_{ij}})] \quad (16)$$

⑥ In order to avoid overfitting conditions, $L_2$ regular regularity：

$$L_2=\frac{\lambda}{2N}\sum_w w^2 \quad (17)$$

λ is a super-recession, W is the element in each training parameter matrix, and the final total Loss is this:

In summary, the limitations of the existing DDIs prediction methods inspire us to find new improvements or models for this research.

$$\text{Loss}=L_1+L_2 \quad (18)$$

## Availability of data and materials

The datasets used and analysed during the current study available from the corresponding author on reasonable request.


## Funding

This research was funded by Liaoning Province Economic and Social Development Research Project 2023 of Provincial Social Science Association (grant no.2023lslybkt-039), 2019 National Natural Science Foundation of China (grant no.61876031).


## Authors' contributions

This article is jointly completed by Shuang Liu and Xiaomin Zhu . Shuang Liu is mainly responsible for charts, Xiaomin Zhu is mainly responsible for experimental part and article writing.


## Acknowledgements

First of all, I would like to give my heartfelt thanks to all the people who have ever helped me in this paper.

My sincere and hearty thanks and appreciations go firstly to my supervisor, whose suggestions and encouragement have given me much insight into these translation studies. It has been a

great privilege and joy to study under his guidance and supervision. Furthermore, it is my honor to benefit from his personality and diligence, which I will treasure my whole life. My gratitude to him knows no bounds.I am also extremely grateful to all my friends and classmates who have kindly provided me assistance and companionship in the course of preparing this paper.

In addition, many thanks go to my family for their unfailing love and unwavering support.

Finally, I am really grateful to all those who devote much time to reading this thesis and give me much advice, which will benefit me in my later study.

## Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

## References

1. Tengfei Ma, Cao Xiao, Jiayu Zhou, and Fei Wang. Drug similarity integration through attentive multiview graph auto-encoders. In IJCAI, 2018.
2. Xu Chu, Yang Lin, Yasha Wang, LeyeWang, Jiangtao Wang, and Jingyue Gao. Mlrda: a multitask semi-supervised learning framework for drug-drug interaction prediction. In IJCAI, pages 4518–4524, 2019.
3. Huang, and Xiangxiang Zeng. A novel molecular representation with bigru neural networks for learning atom.Briefings in bioinformatics, 11 2019.

4. Ryu, J.Y. et al. (2018) Deep learning improves prediction of drug-drug and drug-food interactions. Proc. Natl. Acad. Sci. USA, 115, E4304–4311.
5. Su,C. et al. (2020) Network embedding in biomedical data science. Brief.Bioinf., 21, 182–197.
6. Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics,34(13):i457–i466, 2018.
7. Hongwei Wang, Miao Zhao, Xing Xie,Wenjie Li, and Minyi Guo. Knowledge graph convolutional networks for recommender systems. In WWW, pages 3307–3313, 2019.
8. Shi,H. et al. (2020). Predicting origin-destination flow via multi-perspective graph convolutional network. In: ICDE, Virtual.
9. Wang, M. (2017) Predicting rich drug-drug interactions via biomedical knowledge graphs and text jointly embedding. arXiv Preprint arXiv, 1712,08875.
10. Huang, and Xiangxiang Zeng. A novel molecular representation with bigru neural networks for learning atom.Briefings in bioinformatics, 11 2019.
11. Bo Jin, Haoyu Yang, Cao Xiao, Ping Zhang, Xiaopeng Wei, and Fei Wang. Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. In AAAI, 2017.
12. Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deep walk: Online learning of social representations. In SIGKDD, pages 701–710, 2014.
13. Zhe Quan, Yan Guo, Xuan Lin, Zhi-Jie Wang, and Xiangxiang Zeng. Graphcpi: Graph neural representation learning for compound-protein interaction. In BIBM, pages 717–722, 2019.
14. T. Dettmers, P. Minervini, P. Stenetorp, et al. Convolutional 2D knowledge graph embeddings[J]. ar Xiv preprint ar Xiv: 1707.01476, 2018
15. Q. N. Dai, D. N. Tu, D. Q. Nguyen, et al. A novel embedding model for knowledge base completion based on the convolutional neural network[J]. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), 2018
16. D. Q. Nguyen, T. Vu, T. D. Nguyen, et al. A capsule network-base embedding model for knowledge graph completion and search personalization[C]. Proceedings of the 2019 Conference of the North. 2019

17. Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. Scientific data 3 (2016), 160035.

18. ZHANG Wen, CHEN Yanlin, LIU Feng, et al. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data[J]. BMC bioinformatics, 2017, 18 (1): 18

19. J. Urbani, C. Jacobs. RDF-SQ: Mixing parallel and sequential computation for Top-Down OWL RL inference[M]. Graph structures for Knowledge representation and reasoning. springer international Publishing. 2015, 125-138

20. M. Nickel , K. Murphy, V. Tresp et al. A review of relational machine learning for knowledge graphs[J]. Proceedings of the IEEE, 2016,104(1):11-33

21. B. Motik, Y. Nenov, R. Piro, et al. Parallel materialisation of datalog programs in centralised, main-memory RDF systems[C]. aaai conference on Artifcial Intelligence, 2014

22. VILAR Santiago, HARPAZ Rave, URIARTE Eugenio, et al. Drug-drug interaction through molecular structure similarity analysis[J]. Journal of the American Medical Informatics Association, 2012, 19 (6): 1066-1074

23. Sebastian Köhler, Michael Gargano, Nicolas Matentzoglu, Leigh C Carmody, David Lewis-Smith, Nicole A Vasilevsky, Daniel Danis, Ganna Balagura, Gareth Baynam, Amy M Brower, et al. 2021. The Human Phenotype Ontology in 2021. Nucleic acids research 49, D1 (2021), D1207–D1217.

24. Soo Yeon Kim, Saehoon Kim, Joongbum Cho, Young Suh Kim, In Suk Sol, Youngchul Sung, Inhyeok Cho, Minseop Park, Haerin Jang, Yoon Hee Kim, et al. 2019. A deep learning model for real-time mortality prediction in critically ill children. Critical care 23, 1 (2019), 1–10.

25. SINGH Harpreet, RANA Prashant Singh, SINGH Urvinder.Prediction of drug synergy in cancer using ensemble-based machine learning techniques[J]. Modern Physics Letters B, 2018, 32 (11): 1850132

26. WANG Chi-Shiang, LIN Pei-Ju, CHENG Ching-Lan, et al. Detecting potential adverse drug reactions using a deep neural network model[J]. Journal of medical Internet research, 2019, 21 (2): e11016

27. Jay L Koyner, Kyle A Carey, Dana P Edelson, and Matthew M Churpek. 2018.The development of a machine learning inpatient acute kidney injury prediction model. Critical care medicine 46, 7 (2018), 1070–1077.

28. Alistair EW Johnson and Roger G Mark. 2017. Real-time mortality prediction in the Intensive Care Unit. In AMIA Annual Symposium Proceedings, Vol. 2017. American Medical Informatics Association, 994.