A biomedical knowledge graph-based method for drug–drug interactions prediction through combining local and global features with deep neural networks

Zhong-Hao Ren (ID), Zhu-Hong You (ID), Chang-Qing Yu, Li-Ping Li, Yong-Jian Guan, Lu-Xiang Guo and Jie Pan

Corresponding authors: Zhu-Hong You, School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China. Tel.: (+86)17392763836;
E-mail: zhuhongyou@nwpu.edu.cn; Chang-Qing Yu, School of Electronic Information, Xijing University, Xi'an 710100, China. Tel.: (+86)18991185758;
E-mail: xaycq@163.com; Li-Ping Li, College of Agriculture and Forestry, Longdong University, Qingyang, 745000, China. Tel.: (+86)13993147308;
E-mail: cs2bioinformatics@gmail.com

Abstract

Drug–drug interactions (DDIs) prediction is a challenging task in drug development and clinical application. Due to the extremely large complete set of all possible DDIs, computer-aided DDIs prediction methods are getting lots of attention in the pharmaceutical industry and academia. However, most existing computational methods only use single perspective information and few of them conduct the task based on the biomedical knowledge graph (BKG), which can provide more detailed and comprehensive drug lateral side information flow. To this end, a deep learning framework, namely DeepLGF, is proposed to fully exploit BKG fusing local–global information to improve the performance of DDIs prediction. More specifically, DeepLGF first obtains chemical local information on drug sequence semantics through a natural language processing algorithm. Then a model of BFGNN based on graph neural network is proposed to extract biological local information on drug through learning embedding vector from different biological functional spaces. The global feature information is extracted from the BKG by our knowledge graph embedding method. In DeepLGF, for fusing local–global features well, we designed four aggregating methods to explore the most suitable ones. Finally, the advanced fusing feature vectors are fed into deep neural network to train and predict. To evaluate the prediction performance of DeepLGF, we tested our method in three prediction tasks and compared it with state-of-the-art models. In addition, case studies of three cancer-related and COVID-19-related drugs further demonstrated DeepLGF's superior ability for potential DDIs prediction. The webserver of the DeepLGF predictor is freely available at http://120.77.11.78/DeepLGF/.

Keywords: drug–drug interactions, biomedical knowledge graph, graph neural network, deep learning, multi-feature aggregation

Introduction

The drug combination is a necessary approach to produce a synergistic and therapeutic effect. However, potential drug–drug interactions (DDIs) may cause drug withdrawal and patient death, which is a common and dangerous scenario [1–4]. With the constantly increasing need for drug therapy, the identification of unknown DDIs becomes more and more urgent. However, certain unrecognized interactions may not always be disclosed until they are approved for clinical use. Detecting DDIs based on traditional clinical trials and vitro-vivo experiments is a valid but time-consuming and laborious process [5, 6]. Consequently, a computer-aided DDIs prediction model is urgently needed to simplify the detection process and reduce the complexity of biological experiments. Recently, a large number of computational methods have been proposed that can prescreen the candidates for validation of the interaction mechanism. And most of them have achieved great success in reducing the time cost and labor cost of DDIs prediction [7–10, 14]. From the interaction network perspective, all the methods can be roughly classified into two groups:

the heterogeneous network-based method and the homogeneous network-based method.

Generally, homogeneous network-based methods predict potential DDIs by using the structure information of homogeneous network and node attribute information. For example, Yu *et al.* designed the DDINMF model which applies semi-nonnegative matrix factorization to conclude enhancive and degressive prediction of DDIs [12]. Shi *et al.* introduced a BRSNMF model, an optimization of the model DDINMF, which predicts DDIs by turning the DDIs adjacency matrix into several decomposed matrices and reestablishing adjacency [13]. Zhang *et al.* tried to enhance the prediction performance by offering a comprehensive ensemble model which consists of 29 models [15]. Similarly, Zhang *et al.* provided an ensemble model of SFLLN, based on sparse feature learning, to predict DDIs [16]. However, the abovementioned methods have common disadvantages that they lack the inductive ability and cannot handle new drugs not existing in the interaction network and cannot preserve the different types of associations between entities [25].

Zhong-Hao Ren is a graduate student of Xijing University. His research interests include drugs, proteins and non-coding RNAs, and machine learning.
Zhu-Hong You, PhD, is a Professor of School of Computer Science, Northwestern Polytechnical University, Xi'an, China. His research interests include neural networks, intelligent information processing, sparse representation and its applications in bioinformatics.
Chang-Qing Yu is currently a Senior Engineer with the School of Electronic information, Xijing University. His current research interests include neural networks and intelligent computing, and its applications in bioinformatics.
Li-Ping Li is a professor of College of Agriculture and Forestry, Longdong University. Her research interests include protein and non-coding RNAs, and data mining.
Yong-Jian Guan is a graduate student of Xijing University and is interested in computational biology, and machine learning.
Lu-Xiang Guo is a graduate student of Xijing University. His research interests include drug and RNAs.
Jie Pan is a graduate student of Xijing University and is interested in plant protein interactions prediction.

Compared with homogeneous network-based methods, heterogeneous network-based treatment is closer to reality, due to the different kinds of entities with various kinds of relations [22–24]. In recent years, knowledge graph (KG) has been utilized for data integration and federation [11, 17]. It allows the knowledge graph embedding (KGE) model to excel in the link prediction tasks [18, 19]. For example, Dai *et al.* provided a method using Wasserstein adversarial autoencoder-based KGE, which can solve the problem of vanishing gradient on the discrete representation and exploit autoencoder to generate high-quality negative samples [20]. The SumGNN model proposed by Yu *et al.* succeeds in integrating external information of KG by combining high-quality features and multi-channel knowledge of the sub-graph [21]. Lin *et al.* proposed KGNN to predict DDI only based on triple facts of KG [66]. Although these methods have used KG information, only focusing on the triple facts or simple data fusion can limit performance and inductive capability [69]. Su *et al.* successively proposed two DDIs prediction methods [55, 56]. The first one is an end-to-end model called KG2ECapsule based on the biomedical knowledge graph (BKG), which can generate high-quality negative samples and make predictions through feature recursively propagating. Another one learns both drug attributes and triple facts based on attention to extract global representation and obtains good performance. However, these methods also have limited ability or ignore the merging of information from multiple perspectives. Apart from the above, the single perspective has many limitations, such as the need to ensure the integrity of related descriptions, just as network-based methods cannot process new nodes [65]. So, the methods only based on network are not inductive, causing limited generalization [69]. However, it can be alleviated by fully using the intrinsic property of the drug seen as local information, such as chemical structure (CS) [40]. And a handful of existing frameworks can effectively integrate multi-information without losing induction [69]. Thus, there is a necessity for us to propose an effective model to fully learn and fuse the local and global information for improving performance of DDI identification through multiple information complementing.

In this paper, we propose a deep learning framework called DeepLGF, an inductive model, to predict DDIs through aggregating local–global multi-information based on the BKG. Global information (GI) is seen as integrating rich external information. Conversely, local information is regarded as single intrinsic property information. Local information contains drug CS and biological function (BF) features, which are considered as small but high-quality biochemical features. For fully learning BF features, we constructed heterogeneous networks of drug–receptor and proposed BFGNN to generate drug embedding vectors through aggregating receptor information from different biological functional spaces. And the global feature, extracted from BKG, contains complete drug lateral side information flow. After obtaining the feature under each view, for efficient fusion of the multi-feature, we present four feature integration methods to explore the best one and use it in the final model. Details of the process are introduced in the Materials and Methodology section. In addition, DeepLGF is tested in three different prediction tasks [16, 69] to confirm its robustness and stability. Compared with state-of-the-art models and machine learning methods demonstrates the high efficiency of our model. Focusing on cancer-related and COVID-19-related drugs, case studies of three drugs curing the disease can further indicate the prediction ability of the proposed model. And then, we further discuss the potential mechanisms of DDI to confirm the significance of prediction results and provide two predicted interactions with high confidence. Meaningfully, we built the online prescreening tool of DeepLGF for computational chemistry and drug discovery-related researchers to provide reliable drug candidates, which can simplify the screening process and is accessible at http://120.77.11.78/DeepLGF/. There is no doubt that our study plays a prominent contribution to predicting potential DDIs.

Materials and Methodology
Dataset and prediction tasks description

DrugBank is an online database [28–32], comprehensively integrating the information of about 14,528 detailed drugs and targets. DRKG [33], as a comprehensive BKG, contains 6 databases and possesses 97,238 entities with 13 entity types and 5,874,261 triples with 107 edge types, which has been used in DDI prediction, drug discovery and drug repurposing [40, 63, 64]. In this work, the datasets contain two parts, feature profiles and interaction labels. The feature profiles contain three parts: (i) the SMILES from DrugBank (version 5.1.8) are used as drug CS information. (ii) Several types of drug biological heterogeneous associations from DrugBank (version 5.1.8) are used to extract BF information. (iii) The BKG is used to extract GI. Due to the amount of noise existing in KG [21], we select 12 entities with 28 edge types belonging to the database of DrugBank and Hetionet [58], which virtually cover all the types of entity and entity-type pairs. In our KG, the triples are 62.6% among the whole DRKG and the direct interaction pairs among drugs are excluded to avoid label leakage. The details of our KG are shown in Figure 1. All the abbreviations in this paper and their descriptions can be found in the Supplementary Material (Table of S18).

For the drug interactions, inspired by Feng *et al.* [43], different scale datasets are fully used for validating DeepLGF. We processed DDIs information collected by Yue *et al.* [34], which is named DS1. To demonstrate the stability of DeepLGF, we also exploited 5-fold cross-validation (CV) on the DS2 dataset which is collected by Zhang *et al.* [15], and on the DS3 dataset, which is collected by Shi *et al.* [13], containing enhancive and depressive interactions. For convenience, two types of interaction datasets are respectively called E-DS3 and D-DS3. Table 1 illustrates more details. Remarkably, four datasets have different BF information; thus, we build four heterogeneous networks for four datasets to learn through the proposed model BFGNN.

To fully evaluate the predictive ability of our method, we tested DeepLGF on the three prediction tasks [16, 69], which are shown in Figure 2. And we deeply discussed the performance of DeepLGF on the first task. In task one called the T_{kk}, we randomly choose pairwise drugs without constraints, which can stand for predicting between known drugs. Prediction tasks two and three named the T_{kn} and the T_{nn}, are more representative of the medical therapy reality, containing prediction of interaction between known drugs and new drugs, and prediction of interaction among the new drugs. Specifically, we randomly select the 40 drugs and construct training positive and negative samples, which exclude all the interactions related to the 40 drugs. In type one, each test sample only needs to contain one drug from the 40 drugs. And in type two, the test set only contains the 40 drugs. The way of constructing sample sets can assure independence. Table 2 illustrates more details of the sample sets.

In the experiment, the known interactions are regarded as positive samples and randomly choosing negative samples from other pairwise drugs, equaling positive samples, to resolve the bias. We randomly shuffle the order of samples and regard 70% of them as training datasets. 20% of them were taken as test
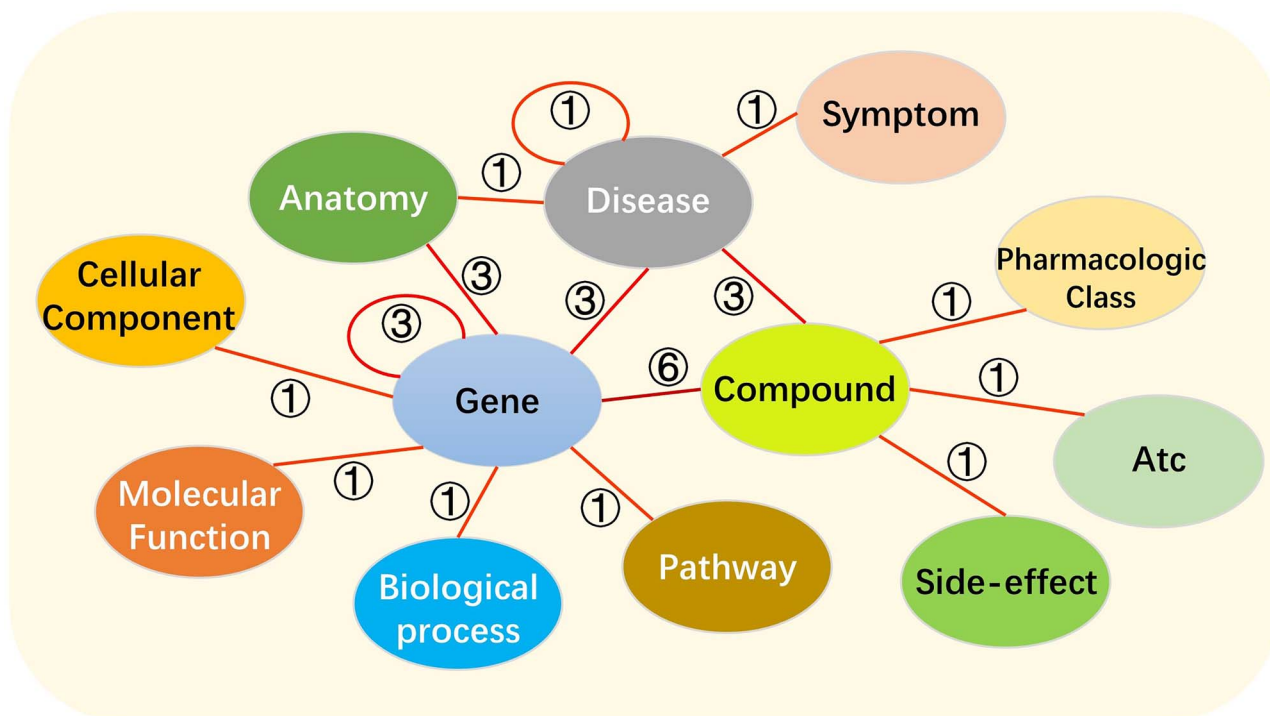
Figure 1. The structure of biomedical KG in this paper. Each number indicates the number of relationship types among the entities.

Table 1. The feature description and interaction details of the four datasets

Datasets Name	Reference	Number of Drugs	Number of Pairs	Number of Interaction	Number of Features	Features Description
DS1	Yue *et al.* [34]	1940	3 763 600	219 247	6	KG, Substructure, Target, Carrier, Enzyme, Transporter
DS2	Zhang *et al.* [15]	548	300 304	48 584	9	KG, Substructure, Target, Enzyme, Transporter, Pathway, Indication, Side effect, Offside effect
E-DS3	Shi *et al.* [13]	1562	2 439 844	125 298	6	KG, Substructure, Target, Carrier, Enzyme, Transporter
D-DS3	Shi *et al.* [13]	1562	2 439 844	55 278	6	KG, Structure, Target, Carrier, Enzyme, Transporter

datasets, and the rest 10% were treated as validation datasets to make a meaningful contrast under 5-fold CV. Due to DS2 being often used to evaluate the performance of the model, to save time and cost, we select DS2 as the primary dataset of the comparison experiment which is the same as the previous work. Meantime, DS1 and DS3 are utilized to evaluate the stability and robustness of the proposed model.

Overview of methods

In this paper, we propose a deep learning framework DeepLGF to fuse local–global information based on BKG, which can improve the performance of DDI identification through multiple information complementing. For analysis features in multi-view, local information of drug CS and drug BF, and GI of BKG are fully used. As figure 3 illustrates, DeepLGF possesses four modules, comprising a drug structure feature extracting module, drug GI learning module, learning module of drug BF feature and multi-perspective integrating prediction module. The first three modules are trained separately to learn multiple representations of drugs, and then the integration module fully fuses them and

makes the final prediction. Concretely, the drug with a unique characteristic of substructure is represented by learning corresponding SMILES with the method of natural language processing. After that, GI can be extracted from the BKG which includes the abundant relationship of drugs in kinds of bio-scale, through a KGE method of ComplEx [27]. ComplEx introduces a complex field for calculating deep-level interaction information about entities and relations, which is suitable for learning complex drug-related antisymmetric relations. Significantly, we proposed a new model of BFGNN and constructed heterogeneous networks of drug–receptor to learn drug information of different BFs, which generates the topology-preserving representation of drugs. Furthermore, we proposed four different integration strategies to fuse efficiently all the information and predict potential DDIs.

Representation of drug CS

Drug structure contains abundant potential chemical information. Disposing of abundant corpora, a sentence or word embedding method can be utilized to advance the performance of downstream tasks. In a biochemical scope, viewing one SMILES as a
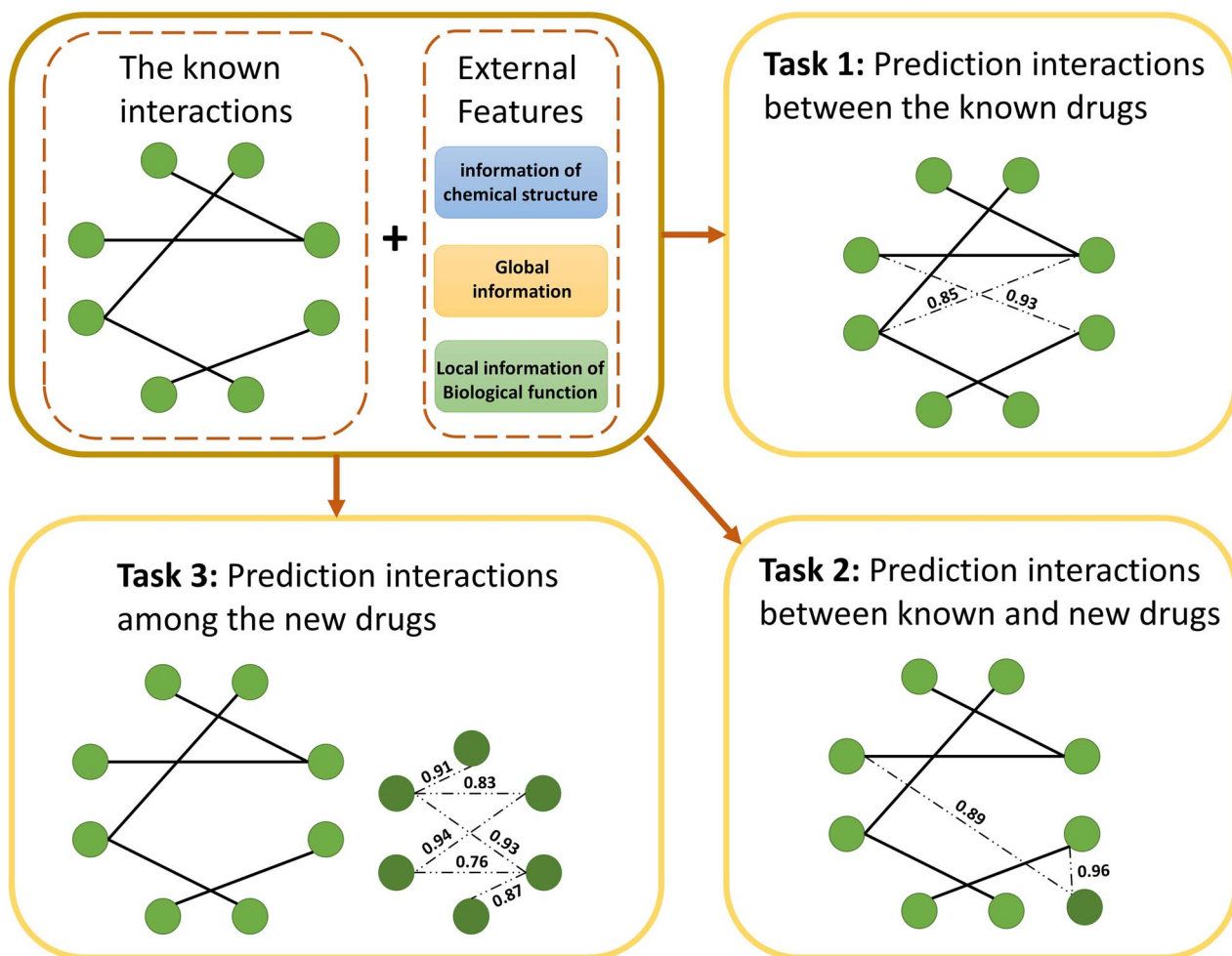
Figure 2. A detailed explanation of three prediction tasks. Dashed lines represent the prediction of our method.

Table 2. The details of the dataset in two prediction tasks

Prediction Task	Number of Drugs in Train Set	Number of Positive Samples in Train Set	Number of Negative Samples in Train Set	Number of Drugs in Test Set	Number of Samples in Test Set
T_{kn}	1900	209 909	209 909	>40	9338
T_{nn}	1900	209 909	209 909	40	106

sentence and regarding the chemical symbols as words in the sentence is a wonderful approach to the use of composition property for representing drug chemical information [35]. Therefore, based on the 'natural biochemical language', the embedding method of PV-DBOW [26] is used to learn the drug symbol embedding vectors. PV-DBOW can wonderfully pay attention to the order of drug symbols and the semantic information of drug symbols by extracting the distributed bag of words version of paragraph vector. In addition, it can be well applied to the feature extraction of sequences with variable lengths, which is suitable for learning the structure information of drugs.

PB-DBOW is utilized for extending the single word embedding to the vector representation of the whole sequence based on the method of skip-gram [36, 37]. Skip-gram can learn the representation through maximizing the probability of the context of a certain center word. And skip-gram contains an input layer, a hidden layer, and an output layer. The hidden layer calculates embedding vectors (h) of the center words through the weight matrix (ω), as follows:

$$h_c = \omega^{\mathrm{T}} x_c \tag{1}$$

where x_c represents one-hot vectors of C-th center words (v_c), and each center word has N context words. Through the context word embedding vectors (u_o) and weight matrix (ω') of the output layer, the probabilities between each context word and a certain center word are calculated. Moreover, the occurrence probability of the actual context word is maximized by the loss function:

$$E = \log p\left(v_o | v_c\right) = u_o^{\mathrm{T}} h_c - \log\left(\sum_{i \in N} \exp\left(u_i^{\mathrm{T}} h_c\right)\right) \tag{2}$$
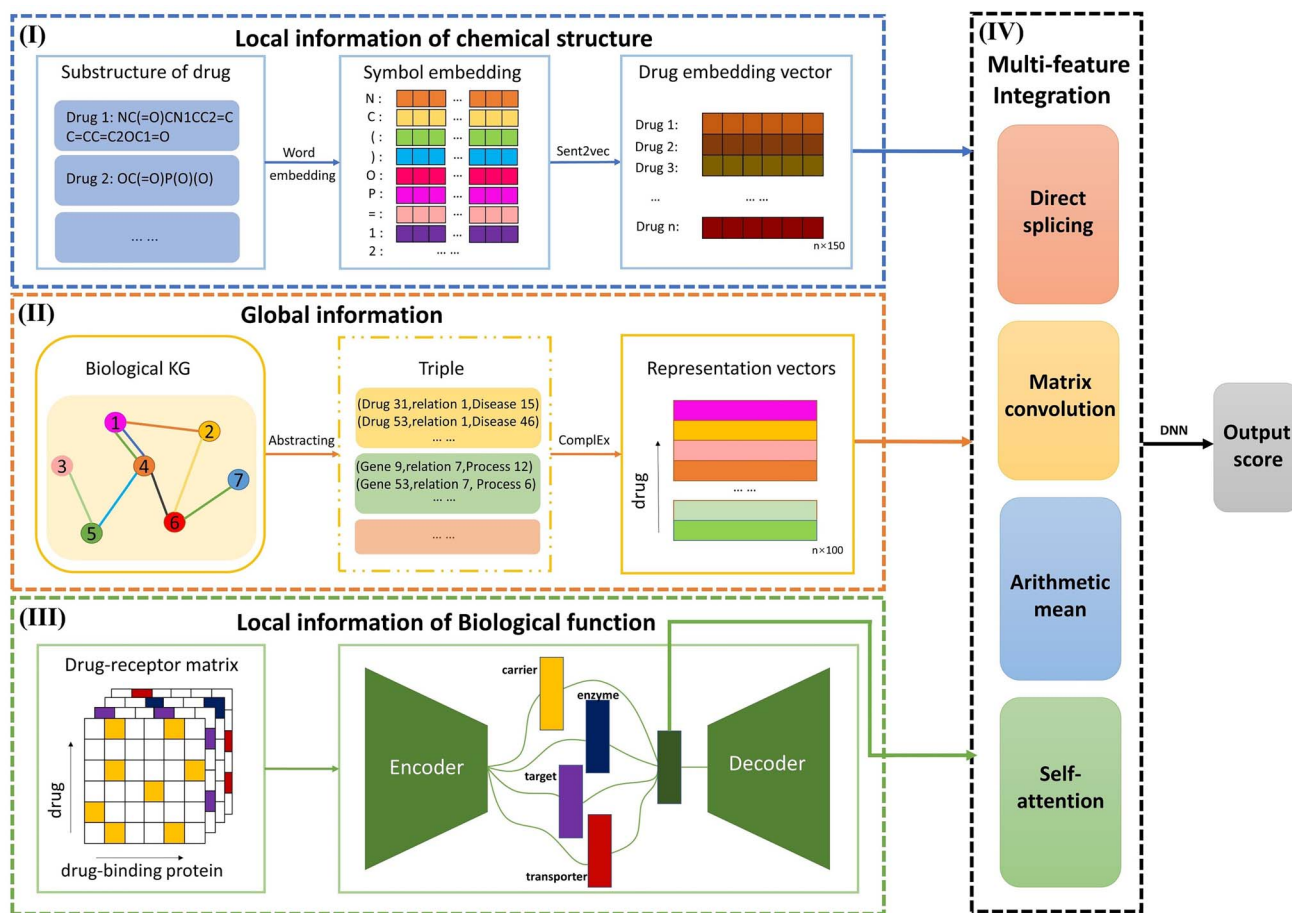
Figure 3. The pipeline of the model of DeepLGF. By giving the input data, sub-model (I) extracts embedding vectors of each symbol and uses the embedding vectors to obtain the complete vectors of the whole substructure of drugs. (II), Drug global features are extracted from the BKG, the same pair can be connected through different relationships, utilizing the method of ComplEx to the triples. And (III), several types of drug–receptor matrixes are put into BFGNN to generate drug representation. Finally, the (IV) module of feature aggregation, containing four types of fusing methods, fuses all features to predict potential DDIs through DNN.

where (u_i) means i-th column of weight matrix (ω'). PB-DBOW pro-motes this model to integrate the word feature of each sequence through a new weight matrix (ω''), and then generates the sen-tence(drug) embedding vector. The probability loss function of the context word on the condition of c-th sentence $(sent_c)$ is as follows:

$$E_{sent} = \log p\left(v_o | sent_c\right) = u_o^{\mathrm{T}} h_c' - \log\left(\sum_{i \in N} \exp\left(u_i^{\mathrm{T}} h_c'\right)\right) \quad (3)$$

where (h_c') means sentence (drug) embedding vector, c-th column of weight matrix (ω'). Finally, each SMILES can be represented as a feature vector.

Method of global feature extracting

The feature, extracted from the associations between the drug entities and other biological entities, is crucial to predicting poten-tial DDIs. In this section, the task of learning abundant lateral side information of drugs from BKG is focused on. Then, each node containing drugs can be represented by a unique embedding vector. However, the common network embedding algorithm only focusing on the connection between nodes is not suitable for KG learning, because abundant edge information and heterogeneous entities of KG will be ignored [55, 67], which can be overcome by KGE [68]. Inspired by Chen *et al.* [40], to exploit fully the

KG and to express globally the straight or potential informa-tion flow between nodes, the effective KGE method of ComplEx, proposed by Trouillon *et al.*, is used to gain the embedding representation [27].

Recently, many biological data and knowledge bases have been used to build a semantic network, which is published as Linked Open Data [38, 39]. The KGE can supply a strong model to set the kind of biological data and make full use of the underlying struc-ture of the graph to extract meaningful information. In addition, KGE can avoid the defect of network embedding that only focuses on the connection between nodes but ignores the types of edges and nodes [55, 67]. Each node in KG is embedded into a common space with comprehensive biomedical text semantic information.

The method KGE of ComplEx is chosen to extract heteroge-neous features with the superiority of capturing the antisym-metric relation. It introduces a complex field to calculate deep-level interaction information of entities and relations. First of all, BKG is defined. $G_{KG} = \{(h, r, t) | h, t \in E, r \in R\}$ is presented as KG, giving a set of various bio-entities E and the bio-relations R among the E. Each item (h_i, t_i, r_i) describes a bio-relation r_i between entity h_i and entity t_i, where $h_i, t_i \in E, r_i \in R, i \in \{1, 2, ..., N\}$, and N indicates the total number of items in our KG. For each entity and relation in G_{KG}, ComplEx is employed to obtain the KG representation. ComplEx conducts low-rank decomposition to the tensor of the knowledge base. Each slice of the tensor is one
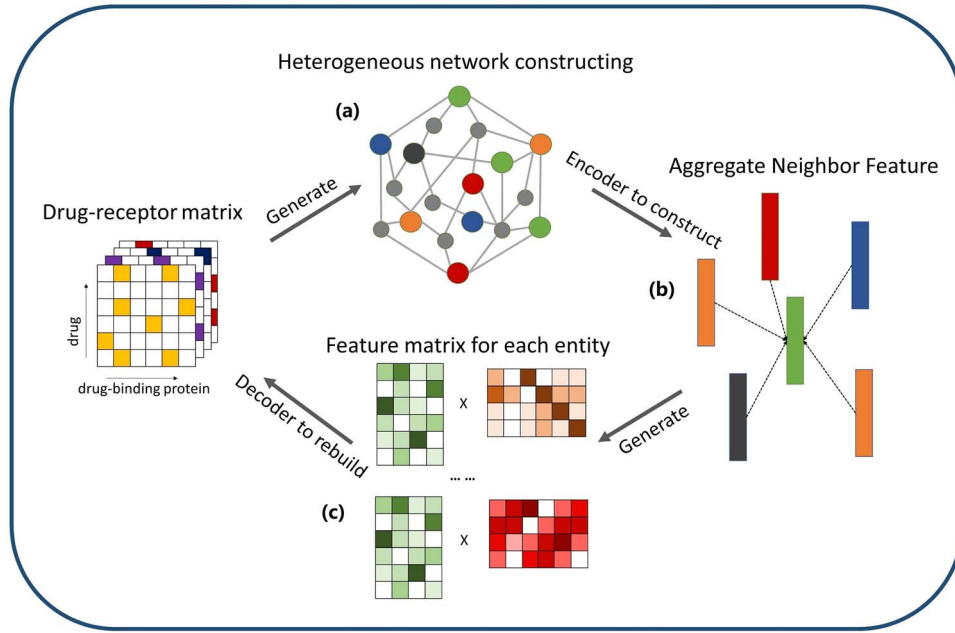
Figure 4. Process of the module of the GNN encoder–decoder machine. (a) The drug–receptor matrixes, as inputting data, generate heterogeneous network of the drug. (b) The feature of each entity comes from its fusion feature of the direct neighbors within itself. The fusion feature can catch complete local information. (c) Product between each pairwise feature vector rebuilds function matrixes under the different spaces through optimizing reconstruction loss. Finally, learning to obtain all aggregation representations of the drug, fusing different function spaces.

of the relation adjacent matrices in the knowledge base, as well as the decomposed matrix that is used to calculate the score. Specifically, the probability of existing interaction between entity h_i and entity t_i is:

$$P\left(Y_{h_i,r_i,t_i} = 1\right) = \sigma\left(\phi\left(r_i, h_i, t_i; \Theta\right)\right) \qquad (4)$$

where the right of the equation stands for using the function of sigmoid (σ) on the latent matrix of scores, then Θ denotes the parameters of the corresponding model, namely, embedding $w_{r_i}, e_{h_i}, \overline{e_{t_i}}$, and $Y_{h_i,r_i,t_i} = \{-1, 1\}$ indicates the true or false interaction fact of i-th triple (h_i, t_i, r_i) where -1 means no interaction otherwise exists interaction. In addition, $\phi(\bullet)$ is a scoring function and the definition is as follows:

$$\phi\left(r, h, t; \Theta\right) = \text{Re}\left(\sum_{k=1}^{K} w_{rk} e_{hk} \overline{e_{tk}}\right) \qquad (5)$$

where K means the dimension of the embedding vector. The ComplEx model is trained to minimize the loss function, described as follows:

$$Loss = \min_{\Theta} \sum_{i}^{N} \log\left(1 + \exp\left(-Y_{h_i,r_i,t_i}\phi\left(h_i, r_i, t_i; \Theta\right)\right)\right) + \lambda\|\Theta\|_2^2 \qquad (6)$$

where the λ is a hyperparameter.

BF feature learning module

The BF is another significant feature, which can be used to augment locally detailed information and boost the ability of feature learning expression. As shown in Table 1, the different dataset

has different scale information of BF. Thus, we respectively construct the different heterogeneous networks to extract BF features. Inspired by Wan *et al.* [41], we propose a model of BFGNN based on GNN. The feature vectors are learned from the structure of the heterogeneous network. Consisting of the following three steps, Figure 4 deploys the detail of this process. The mathematical formulations of the three steps are proposed as follows.

Firstly, the drug heterogeneous networks are constructed from different drug functional matrixes. In each functional matrix, values 1 or 0 indicate the existence or inexistence of the BF. Each element of matrixes is defined as $e(u, v, r) = \{0, 1\}$, where $v \in V$ represents a certain entity member in one entity type, $u \neq v$ stands for the heterogeneous neighbor of v, and $r \in R$ represents the function relation type between v and u. For example, there are 293 types of enzymes with 1940 types of drugs, and the function can be stood as a (1940×293)-dimensional 0–1 matrix. Then, an initial function $f : V \rightarrow R^d$ maps each entity type member to a d-dimension embedding vector, represented by $f(u)$. After that, due to the heterogeneity of the entity data, each embedding vector is mapped to different relation spaces. Thus, each node can integrate the embedding vector of neighbors under the different function relations. For effectively integrating neighbor vectors from different functional spaces, all of the embedding vectors are mapped to a common space and summed to obtain the aggregation feature a_v of the entity v. The process of information assembling for the member v is defined as:

$$a_v = \sum_{r \in R} \sum_{u \in N_r(v)} \sigma\left(W_c\left[\sigma\left(W_r f(u) + b_r\right) \cdot \lambda_{u,v,r}\right] + b_c\right) \qquad (7)$$

where $N_r(v) = \{u, u \in V, u \neq v, e(u, v, r) = 1\}$ indicates an adjacent set that all members of the entity have a relation with v through function type r, the $\sigma(\bullet)$ denotes a nonlinear activation function; the function of ReLU is used in this work. The W_r, b_r and W_c, b_c

stand for weights and bias under the circumstances of function relationship r and common mapping space c, respectively. And $\lambda_{u,v,r}$ means a normalization term, defined as follows:

$$\lambda_{u,v,r} = \frac{e(u,v,r)}{\sum\limits_{u \in N_r(v)} e(u,v,r)} \tag{8}$$

To obtain a feature representation with comprehensive information, we concatenate the vector of assembling neighbor information a_v and the vector of the entity member itself $f(v)$ which is shown as formula (9).

$$g(v) = \frac{\sigma\left(W' concat\left(f(v), a_v\right) + b'\right)}{\left\| \sigma\left(W' concat(f(v), a_v) + b'\right) \right\|_2} \tag{9}$$

where weights term meets $W' \in R^{d \times 2d}$, bias term meets $b' \in R^d$, and $\|\bullet\|_2$ is l_2 the norm used to do the operation of normalization.

Finally, the vectors, learned by BFGNN, are exploited to rebuild all function matrixes and optimized by a loss function which is given:

$$\min Loss = \min_{\substack{\{f(u), W_r, b_r, \\ W', b', W_c, b_c, \\ G_r, H_r \mid u \in V, r \in R\}}} \sum_{r \in R} \sum_{v,u \in V} \left[e(v,u,r) - g(u)^T G_r H_r^T g(v) \right]^2 \tag{10}$$

where G_r, H_r are edge-type specific projection matrixes. Eventually, the feature vector of assembling neighbor features is extracted by rebuilding the function matrix.

Strategies of information integration

To find the best way to merge efficiently multi-perspective feature vectors, we put forward four kinds of fusing methods to conclude the best fusion method. In this section, before describing the detail of the four proposed fusion methods, we denote three types of feature representations as d-dimension vectors $u_i = \{u_{i1}, u_{i2}, ..., u_{id}\}$, $v_i = \{v_{i1}, v_{i2}, ..., v_{id}\}$ and $w_i = \{w_{i1}, w_{i2}, ..., w_{id}\}$.

(a) Directly connecting-based. It is easy to describe the concatenating of all the feature vectors. The fusion vector can be formulated as follows:

$$f_i = concatenate(u_i, v_i, w_i) \tag{11}$$

where f_i represents i-th fusion feature. Then, the prediction result can be drawn by hidden layers which contain three layers and a softmax function used to calculate the score.

(b) Crossing matrix-based. We firstly construct a cross matrix C_i' by cross-product operation $\otimes$, which indicates the interaction between u_i and v_i as follows:

$$C_i' = u_i \otimes v_i = \begin{bmatrix} u_{i1}v_{i1} & u_{i1}v_{i2} & ... & u_{i1}v_{id} \\ ... & ... & ... & ... \\ u_{id}v_{i1} & u_{id}v_{i2} & ... & u_{id}v_{id} \end{bmatrix} \tag{12}$$

Then, we use the convolution neural network model with a batch normalization layer, a flatten layer, and a dense layer to obtain the d-dimension fusion vector c_i' between u_i and v_i. Therewith, similarly, we aggregate feature vector w_i through

cross-product operation to the final fusion matrix C_i as follows:

$$C_i = c_i' \otimes w_i = \begin{bmatrix} c_{i1}'w_{i1} & c_{i1}'w_{i2} & ... & c_{i1}'w_{id} \\ ... & ... & ... & ... \\ c_{id}'w_{i1} & c_{id}'w_{i2} & ... & c_{id}'w_{id} \end{bmatrix} \tag{13}$$

where c_{id}' means d-th element of i-th vector c'. In the end, through a fusion vector f, generated by another convolution layer, and multiple hidden layers, the result can be predicted. This method has some defects, which are discussed in detail later.

(c) Average-based. We utilize the element-wise average to encode the global interaction feature with the formula:

$$f_i = average(u_i, v_i, w_i) = \frac{u_i + v_i + w_i}{3} \tag{14}$$

where f_i means i-th fusion vector, calculated by u_i, v_i and w_i. After that, multiple hidden layers make the final projection.

(d) Self-attention-based. We propose an attentional multi-feature integration and prediction module for integrating features. The attention [42] module utilizes attention scores to reflect adaptive weights before fusing multi-type features. The attention-enhancing vectors θ' can be calculated by the function:

$$\theta' = Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{15}$$

where Q, K and V, respectively, are on behalf of the product results of the input vector and three learnable matrices W_Q, W_K and W_V. The d_K stands for a normalization coefficient. Finally, the prediction results can be captured through multiple hidden layers.

Experimental Results and Discussion
Evaluation criteria

In this study, the metrics frequently utilized in classification are used to evaluate the effectiveness and robustness of our framework from distinct perspectives containing five indicators: Accuracy (Acc.), Sensitivity (Sen.), Precision (Prec.), F1 and Matthews's Correlation Coefficient (MCC), respectively. These evaluation indicators are defined as follows:

$$Acc. = \frac{TP + TN}{TN + TP + FN + FP} \tag{16}$$

$$Prec. = \frac{TP}{TP + FP} \tag{17}$$

$$Sen. = \frac{TP}{TP + FN} \tag{18}$$

$$F1 = \frac{2 \times Prec. \times Sen.}{Prec. + Sen.} \tag{19}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TN + FN) \times (TN + FP) \times (TP + FN)}} \tag{20}$$

where TP, FN, TN and FP, respectively, stand for true positive, false negative, true negative and false positive. AUC, the area under the receiver operating characteristic (ROC) curves, and AUPR, the area under the precision-recall (PR) curve, are also exploited to evaluate the performance of the model. Meanwhile, the mean results of 5-fold CV can ensure low-variance and unbiased evaluations of the results. In addition, we utilize some range-based metrics to demonstrate the performance of our framework, which are reported in Supplementary Material (Table of S21), containing

Table 3. The 5-fold CV average results through DeepLGF with four aggregation methods on T_{kk} task

Dataset	Method	Acc.	Prec.	Sen.	F1	MCC	AUC	AUPR
DS1	Ds	**0.9432 ± 0.0056**	0.9213 ± 0.0063	**0.9691 ± 0.0073**	0.9446 ± 0.0055	**0.8875 ± 0.0112**	**0.9859 ± 0.0028**	**0.9845 ± 0.0032**
	Mc	0.9194 ± 0.0014	0.8879 ± 0.0038	0.9599 ± 0.0045	0.9225 ± 0.0013	0.8415 ± 0.0027	0.9765 ± 0.0007	0.9741 ± 0.0008
	Am	0.9421 ± 0.0040	**0.9215 ± 0.0046**	0.9665 ± 0.0082	0.9434 ± 0.0040	0.8852 ± 0.0082	0.9852 ± 0.0021	0.9836 ± 0.0025
	S-a	0.9215 ± 0.0078	0.8973 ± 0.0087	0.9521 ± 0.0151	0.9238 ± 0.0079	0.8447 ± 0.0160	0.9758 ± 0.0041	0.9732 ± 0.0046
DS2	Ds	0.9082 ± 0.0015	0.8945 ± 0.0021	**0.9256 ± 0.0052**	0.9098 ± 0.0017	0.8169 ± 0.0031	0.9718 ± 0.0009	0.9705 ± 0.0012
	Mc	0.9059 ± 0.0031	**0.8963 ± 0.0052**	0.9180 ± 0.0081	0.9070 ± 0.0032	0.8121 ± 0.0063	0.9690 ± 0.0019	0.9665 ± 0.0018
	Am	**0.9087 ± 0.0009**	0.8954 ± 0.0040	0.9255 ± 0.0055	**0.9102 ± 0.0011**	**0.8178 ± 0.0019**	**0.9721 ± 0.0008**	**0.9712 ± 0.0012**
	S-a	0.9057 ± 0.0007	0.8931 ± 0.0036	0.9217 ± 0.0041	0.9072 ± 0.0006	0.8118 ± 0.0013	0.9676 ± 0.0011	0.9647 ± 0.0017
D-DS3	Ds	0.9297 ± 0.0041	0.9097 ± 0.0066	0.9540 ± 0.0033	0.9313 ± 0.0037	0.8604 ± 0.0078	0.9742 ± 0.0021	0.9672 ± 0.0028
	Mc	**0.9375 ± 0.0022**	**0.9214 ± 0.0029**	**0.9566 ± 0.0037**	**0.9387 ± 0.0022**	**0.8756 ± 0.0045**	**0.9753 ± 0.0013**	**0.9674 ± 0.0011**
	Am	0.9282 ± 0.0031	0.9054 ± 0.0075	0.9564 ± 0.0055	0.9302 ± 0.0028	0.8578 ± 0.0059	0.9732 ± 0.0016	0.9659 ± 0.0032
	S-a	0.9313 ± 0.0019	0.9154 ± 0.0023	0.9505 ± 0.0062	0.9326 ± 0.0021	0.8633 ± 0.0041	0.9723 ± 0.0013	0.9631 ± 0.0027
E-DS3	Ds	0.9393 ± 0.0059	0.9154 ± 0.0089	**0.9682 ± 0.0032**	0.9410 ± 0.0055	0.8801 ± 0.0114	0.9810 ± 0.0022	0.9778 ± 0.0025
	Mc	0.9379 ± 0.0092	0.9201 ± 0.0115	0.9591 ± 0.0068	0.9392 ± 0.0088	0.8766 ± 0.0180	0.9799 ± 0.0038	0.9765 ± 0.0046
	Am	0.9338 ± 0.0021	0.9084 ± 0.0039	0.9650 ± 0.0025	0.9359 ± 0.0019	0.8694 ± 0.0040	0.9788 ± 0.0007	0.9753 ± 0.0001
	S-a	**0.9440 ± 0.0043**	**0.9289 ± 0.0062**	0.9616 ± 0.0026	**0.9450 ± 0.0041**	**0.8885 ± 0.0084**	**0.9820 ± 0.0015**	**0.9786 ± 0.0017**

The bold values represent the higher values of each column on the different datasets.

HIT@15, Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR).

Assessment of prediction ability

To validate the prediction performance and robustness of DeepLGF in T_{kk}, both the jackknife test and q-fold CV test [44] can examine whether the predictor is effective or not. In this work, we adopt the 5-fold CV to enhance results persuasion and use four information fusion methods, direct splicing (Ds), matrix convolution (Mc), arithmetic mean (Am) and self-attention (S-a), as feature aggregation to evaluate. All the 5-fold CV average results are displayed in Table 3. The high scores claim that utilizing local and global information is a promising method. The detailed results of each fold in all datasets are reported in Supplementary Material (Figures of S1–S16 and Tables of S1–S16).

As shown in Table 3, AUC and AUPR on DS1 reach the highest score of 0.9859 and 0.9845 with the method of Ds. In addition, with the best aggregation method, the mean scores of the other evaluation criteria are also high. It can be drawn that the performance of the model is impacted by the scale of data. The large network with many nodes and edges is more suitable for Ds, due to each entity integrating complex information being not appropriate for complex fusion methods. Moreover, the most anticipated S-a achieves a worse performance than Am, which is probably attributed to overfitting led by intricate strategy. And Mc is most affected by features. Each dataset can get the best performance with Ds, Am, Mc and S-a, respectively. Obviously, DeepLGF has an excellent capability to identify positive and negative samples and predict novel interactions. To further reflect the effect of application in different datasets with different aggregation methods, the mean results of the 5-fold CV of ROC curves and PR curves are shown in Figure 5.

Ablation experiments

A set of ablation experiments on all datasets are conducted to validate the contribution degree of the GI, the local information of drug CS, and the information on BF. The most suitable fusion strategies for each dataset are chosen, and results of ablation experiment are presented in Table 4. After the analysis, it can be seen that the GI is more effective and crucial to the model's performance. Meanwhile, the other local information seems to have little impact on the performance, but fully using biochemical

characteristics can ensure inductive and scalable capability to the model through information integrity. And only when using them all together, the performance can be the best. Moreover, we test performance of every single feature to obviously assess their importance. After the analysis, we can find the results with adding other information are always better than the results with the single information, which demonstrates fuse local and global information can improve the performance of DDI identification through multiple information complementing.

Comparison of different methods for extracting GI

Since GI is a crucial feature, to demonstrate the predictive ability of our GI learning module, various methods of learning GI are compared.

- TransE [59]: Translating Embedding is a KGE, which generates the distributed vectors of entities and relations based on optimizing tail vector equal to head vector adding to relation vector.
- GraRep [60]: GraRep is a graph embedding method of learning graph representations through global structural information, based on matrix factorization.
- LINE [61]: LINE utilizes and optimizes simultaneously first-order proximity and second-order proximity to obtain embedding vectors, based on the neural network.
- SDNE [62]: SDNE is a deep learning method to represent the network, which can capture non-linear structures through multiple nonlinear functions.

For a fair comparison, all methods are tested with the same data division and the same local information. And the fusion strategy corresponds to the dataset, except Mc. Because of the sparsity of the graph, too much noise learned from poor graph embedding methods is diffused by Mc. Thus, Mc is replaced by Ds. As shown in Figure 6, our method has appreciable advantages on datasets of DS1, D-DS3 and E-DS3 in all metrics, especially MCC. In the KG, different types of knowledge entities contain various types of relations that are neglected by the ordinary graph embedding methods, which cause poor performance. Although the TransE is designed to solve the problem of KGE, the oversimplified way of embedding leads to no ability to capture complex interactions. DeepLGF on DS2 also reaches the highest scores in five evaluation
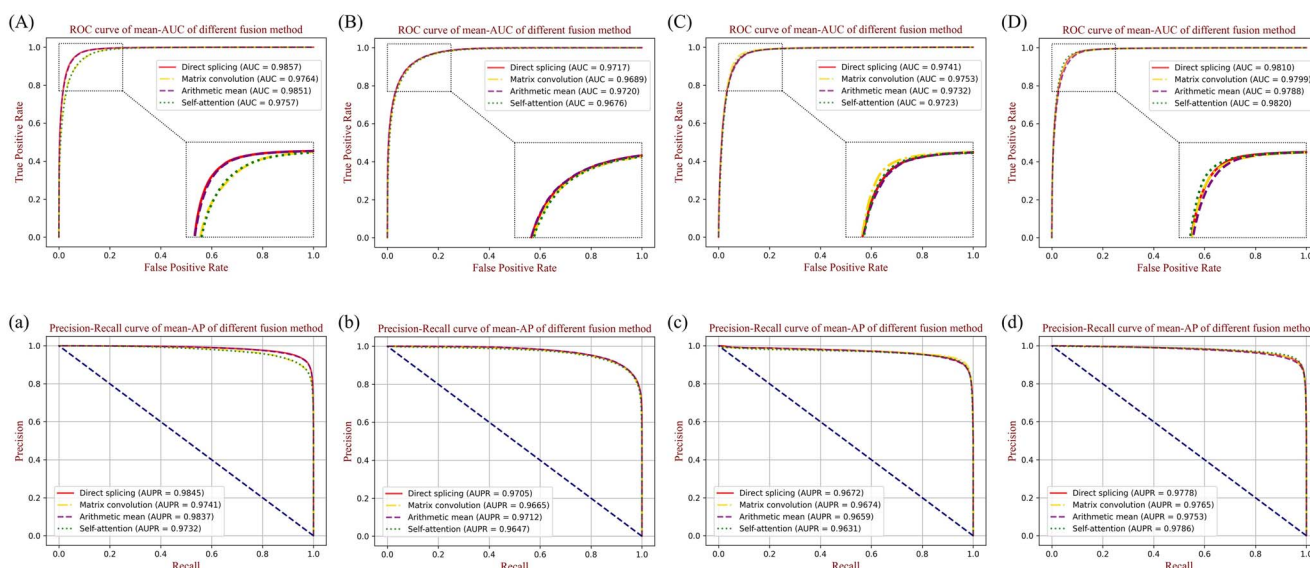
Figure 5. (A), (B), (C) and (D), respectively, indicate the 5-fold CV average result of ROC curves of DS1, DS2, D-DS3 and E-DS3. (a), (b), (c) and (d), respectively, indicate the 5-fold CV average result of PR curves of DS1, DS2, D-DS3 and E-DS3.

Table 4. Results of ablation test on DeepLGF

Dataset	CS	GI	BF	AUC	AUPR
DS1 (with Ds)	T	F	F	0.9095	0.8997
	F	T	F	0.9762	0.9743
	F	F	T	0.8589	0.8594
	F	T	T	0.9791	0.9774
	T	F	T	0.9502	0.9452
	T	T	F	0.9823	0.9803
	T	T	T	**0.9857**	**0.9845**
DS2 (with Am)	T	F	F	0.9254	0.9204
	F	T	F	0.9628	0.9606
	F	F	T	0.9567	0.9545
	F	T	T	0.9667	0.9657
	T	F	T	0.9634	0.9626
	T	T	F	0.9667	0.9659
	T	T	T	**0.9720**	**0.9712**
D-DS3 (with Mc)	T	F	F	0.9147	0.9021
	F	T	F	0.9669	0.9583
	F	F	T	0.9003	0.8987
	F	T	T	0.9725	0.9642
	T	F	T	0.9445	0.9356
	T	T	F	0.9707	0.9631
	T	T	T	**0.9753**	**0.9674**
E-DS3 (with S-a)	T	F	F	0.9100	0.8975
	F	T	F	0.9623	0.9580
	F	F	T	0.8803	0.8732
	F	T	T	0.9715	0.9675
	T	F	T	0.9345	0.9275
	T	T	F	0.9637	0.9625
	T	T	T	**0.9820**	**0.9786**

The F indicates False; the T indicates True. The bold values represent the higher AUC and AUPR in the current dataset.

criteria, but without significant advantages. This may be due to the sample size of DS2 being too small. Totally, our method has superior ability in GI extracting. Moreover, to prove the effectiveness of our local information learning module, we compare our method with three fingerprint methods, containing Morgan, MACCSkeys and RDK, whose results and discussion are reported in Supplementary Material (Table of S19).

Comparison of machine learning classifiers

To demonstrate our framework's superior capability of integrating multi-information, we select five different classifiers commonly used to perform classification tasks, containing Gaussian NB(GNB), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM) and Random Forest. The same local–global feature vectors can ensure a fair comparison. As shown in Figure 7, the proposed framework achieves superior performance in each dataset, regardless of dataset size, which indicates the proposed feature integration strategies are efficient. The machine learning classifiers simply connect the information to calculate the prediction results, which causes the poor performance.

Comparison of other methods

To comprehensively evaluate the performance of the proposed method, several representative methods are chosen as baselines to compare with our method. As previously mentioned, we utilize DS2, usually used to test model performance, as the comparison dataset. First the prediction methods based on graph embedding algorithms (Laplacian Eigenmap and DeepWalk) and three classifiers (GNB, LR and DT) are used to test the robustness and stability of the proposed framework on all datasets. For a fair comparison, the local information vectors are also used by each method. The comparison results are shown in Figure 8. Our method is significantly better than the other methods because our deep learning framework has a more powerful capability to exploit local and global information and suitable fusing methods further improving the performance. The details are reported in Supplementary Material (Table of S17).

Then, for comprehensive assessment, we compared DeepLGF with other state-of-the-art prediction models on DS2 based on three model views [69], including NDD [45], the ensemble model [15], BioDKG-DDI [54], DPDDI [43], CSEP [48], ISCMF [49], AttentionDDI [50], GCN-BMP [51], BioChemDDI [57], Vilar1 and Vilar2 [46, 47], whose results are shown in Table 5. Those results are average values based on 5-fold CV that can ensure fairness, and all the comparison results are taken directly from the respective papers. For further discussion, we also compared baselines on
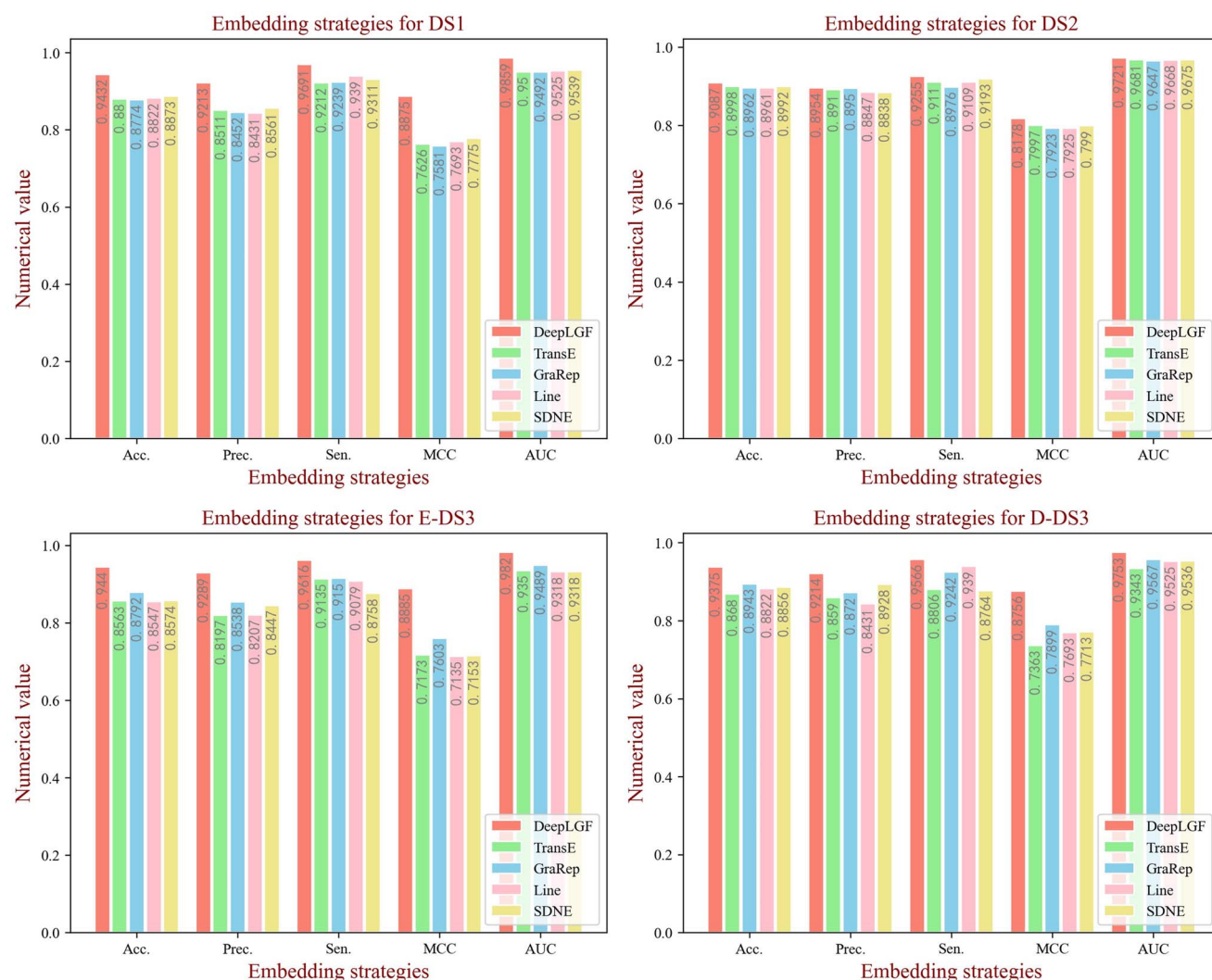
Figure 6. The performance of different GI learning methods on each dataset.

the other datasets to demonstrate the superior performance of DeepLGF, which is reported in Supplementary Material (Table of S20). The comparison results indicate that DeepLGF (using Am) performs best with outstanding improvements of 0.0425–0.2555, 0.0614–0.1872, 0.1471–0.2894, 0.0241–0.1808 and 0.0326–0.2304 against others in terms of Sen., F1, MCC, AUC and AUPR, respectively. It is lower 0.0463 than Ensemble Model in terms of Acc., whereas it still achieves a comparable result of 0.9087. In conclusion, the prediction results are satisfactory, and the results fully demonstrate that DeepLGF has a powerful capability to predict potential DDIs, which is mainly attributed to the advanced representation of local–global information and suitable feature fusing methods.

Evaluation of prediction on T_{kn} and T_{nn} tasks

The performance of DeepLGF is evaluated through the other two prediction tasks, which are closer to real life. Although lower performance scores are produced, as expected, the results are compelling. Due to the common graph embedding method being not suitable for these tasks as mentioned previously, we compare two methods of KG link prediction, KG-TransE and KG-ComplEx, and three sequence-similarity-based methods of S-GNB, S-DT and S-LR to demonstrate the advancement of our method. Tables 6

and 7 show the comparison results. In the task of predicting interaction between known and new drugs, the proposed model gets the obviously highest scores of Acc. and AUC. In the task of prediction interaction among new drugs, because of the completely independent test set, Acc. and AUC just get scores of 0.6557 and 0.7361, but the scores are much higher than the other comparison methods. In terms of Sen. and F1, our method is lower than the method of S-LR. Because the S-LR is more predisposed to group the test samples into positive samples, which can ensure the highest Sen. score but the lowest Prec. value. That is clearly unjustified. Furthermore, HIT@15, MAP and MRR are reported in Supplementary Material (Table of S21). In summary, it can demonstrate that the proposed method is effective and meaningful in real life.

Case studies: cannabidiol, cisplatin and dexamethasone

To validate the performance of DeepLGF in real-life predictive tasks, we conduct case studies of three drugs that are respectively given the ability to alleviate three hard-to-cure cancers, containing advanced cancer, breast cancer, and malignancies. Specifically, we utilize DS1 to train a predicting model for identifying the potential DDIs of *Cannabidiol*, *Cisplatin* and *Dexamethasone*. Notably,
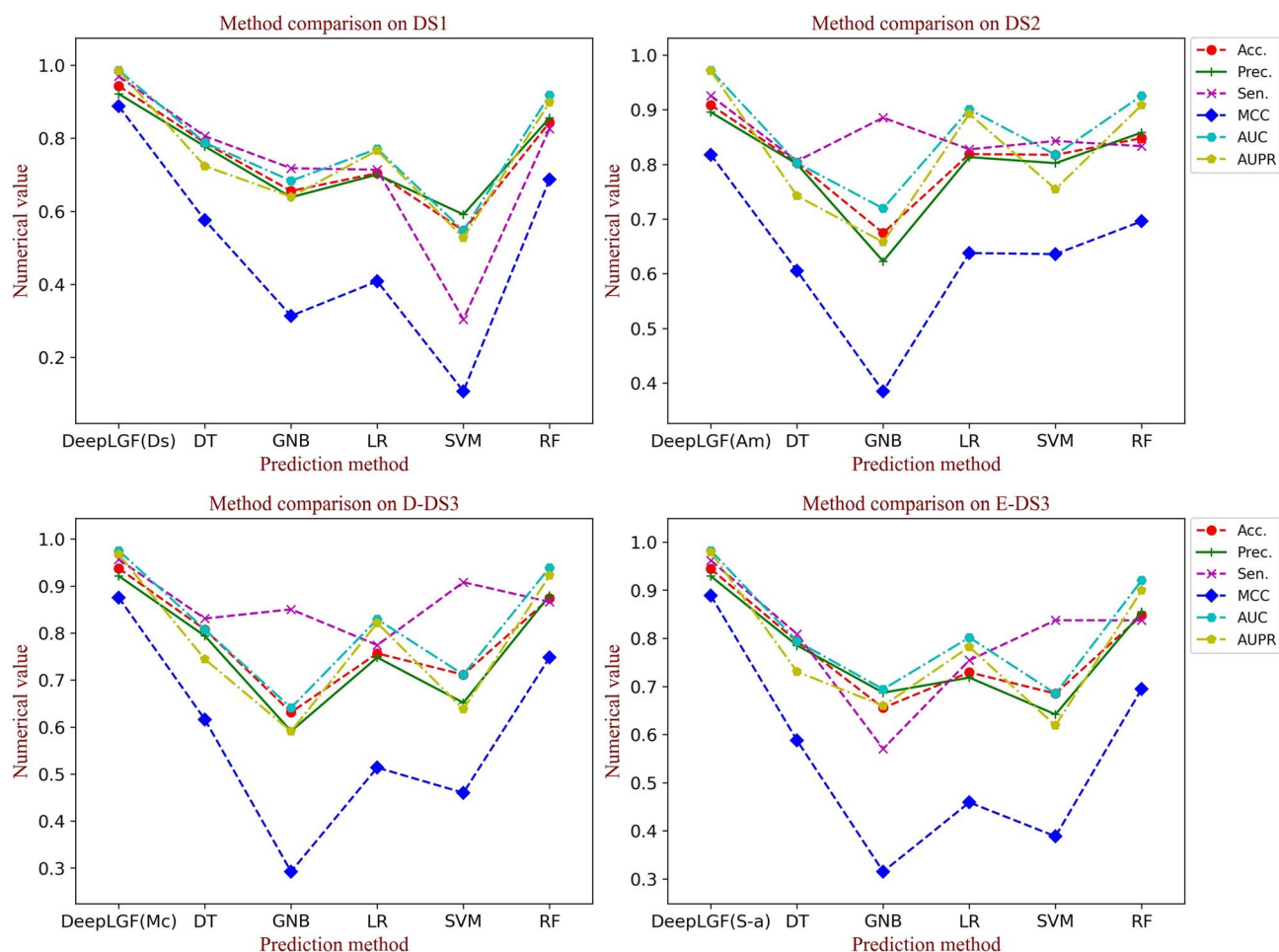
Figure 7. Comparison with different machine learning classifiers on each dataset.

since *Cannabidiol* [52] and *Dexamethasone* are related to COVID-19 [53], exploring their drug interaction structure is very important. The probability score results of the top 15 ranked predictions are shown in Tables 8–10. As seen in the tables, the top five predictions are basically confirmed, and 14, 9 and 14 interactions can be confirmed by DrugBank, which can demonstrate the reliable capability of DeepLGF. In addition, the unconfirmed interactions also have vital significance to further study.

To verify and explain the potential interaction mechanism of unconfirmed DDI from the biomedical perspective, we consult the relevant pharmacokinetic knowledge of the drugs of *Cannabidiol*, *Esomeprazole*, *Dexamethasone* and *Caffeine* from DrugBank. On the one hand, we find that *Cannabidiol* is the inhibitor of the Cytochrome P450 3A4 enzyme, and meanwhile, *Esomeprazole* is the substrate of the same enzyme, which means *Cannabidiol* may impact the metabolism process of *Esomeprazole*. On the other hand, *Dexamethasone* is an inducer of the Cytochrome P450 2C8, Cytochrome P450 2E1 and Cytochrome P450 3A4, and *Caffeine* is the substrate of the three enzymes, which increases the probability of changing the safety and efficacy profile of the drug. Therefore, there is a high possibility of potential interaction existing in these two drug pairs, deserving further investigation. The newly discovered and known interactions of the top 30 in each case study are visualized in Figure 9.

As shown in Figure 9, 75 pairs of 90 pairs of new interactions are predicted successfully by DeepLGF, which means our framework can be applied to predict potential DDIs in reality. There are still 15 pairs of interactions having no support, which deserve further study and validation. Especially, the interactions of *Cannabidiol* with *Esomeprazole* and *Cannabidiol* with *Rolapitant* are most likely to have potential interaction with high confidence. In addition, the complete prediction scores from each case study are reported in Supplementary Material.

Conclusion

Combining drugs has been playing a critical role in complex disease treatment, which may generate severe drug–drug interactions. To avoid the negative impacts on patients, potential DDIs are urgently needed to be identified. In this paper, for solving the drawbacks of existing methods, we proposed a deep learning framework of DeepLGF to predict potential DDIs through fully exploiting local–global information, which has the inductive and scalable capability. The novel deep learning method not only fully considers drug biochemical information and KG to extract the local–global information, but also employs the suitable feature fusing method to enhance the features to obtain the advanced embeddings of drugs. Meanwhile, comparison experiments on multiple prediction tasks and multiple datasets can demonstrate our model has remarkable performance and is adapted to various prediction tasks. Furthermore, case studies of cancer-associated
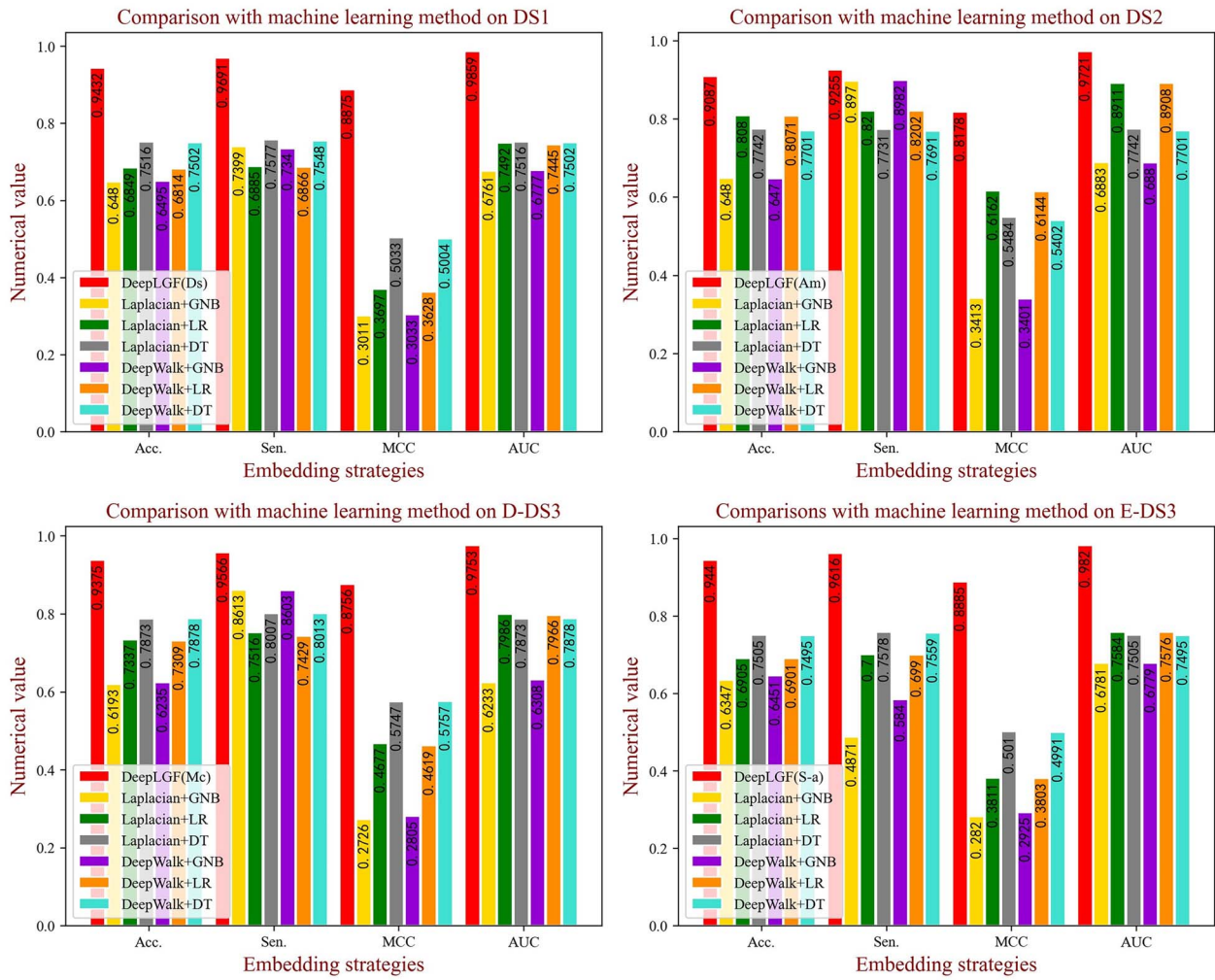
Figure 8. Comparison results on each dataset in T_{kk} tasks.

Table 5. Results of comparison with state-of-the-art prediction models on DS2

Model view	Method	Acc.	Prec.	Sen.	F1	MCC	AUC	AUPR
Lower	Vilar1 [46]	0.7190	0.2530	0.4950	0.3340	(−)	0.7070	0.2620
	Vilar2 [47]	0.8620	0.5150	0.5690	0.5400	(−)	0.8260	0.5330
	DPDDI [43]	0.9400	0.7540	0.8100	0.8400	(−)	0.9560	0.9070
	GCN-BMP [51]	(−)	(−)	(−)	0.8500	(−)	0.9666	0.9402
Higher	CSEP [48]	0.8090	0.7290	0.6850	0.7060	(−)	0.8510	0.7990
	ISCMF [49]	0.8510	**0.9880**	0.8510	0.8850	(−)	0.8990	0.8640
	AttentionDDI [50]	(−)	(−)	(−)	(−)	(−)	0.9540	0.9240
Hierarchi-cal	BioChemDDI [57]	0.9072	0.8954	0.9221	0.9086	0.8148	0.9711	0.9701
	BioDKG-DDI [54]	0.8984	0.8835	0.9178	0.9003	0.7974	0.9668	(−)
	NDD [45]	(−)	0.8330	0.8360	0.8350	(−)	0.9540	0.9220
	Ensemble Model [15]	**0.9550**	0.7850	0.6700	0.7230	(−)	0.9570	0.8070
	DeepLGF (Ds)	0.9082	0.8945	**0.9256**	0.9098	0.8169	0.9718	0.9705
	DeepLGF (Mc)	0.9059	0.8963	0.9180	0.9070	0.8121	0.9690	0.9665
	DeepLGF (Am)	0.9087	0.8954	0.9255	**0.9102**	**0.8178**	**0.9721**	**0.9712**
	DeepLGF (S-a)	0.9057	0.8931	0.9217	0.9072	0.8118	0.9676	0.9647

The bold values represent the higher values of each column. The symbol of (−) indicates the evaluation criteria are not reported in the original articles.

Table 6. Evaluation results of prediction interaction between the known and new drugs

Method	Acc.	Prec.	Sen.	F1	MCC	AUC	AUPR
S-GNB	0.6491	0.6520	0.6395	0.6457	0.2982	0.7005	0.6610
S-DT	0.6472	0.6742	0.5696	0.6175	0.2980	0.6472	0.5992
S-LR	0.6590	0.6400	0.7268	0.6807	0.3210	0.7079	0.6747
KG-TransE	0.6437	0.6787	0.5456	0.6049	0.2930	0.6878	0.7078
KG-ComplEx	0.7306	0.7708	0.6565	0.7091	0.4664	0.8032	0.8245
Our method	0.7947	0.8358	0.7333	0.7812	0.5938	0.8680	0.8802
Range of difference	**0.0641 to 0.1510**	**0.0650 to 0.1958**	**0.0065 to 0.1887**	**0.0721 to 0.1763**	**0.1274 to 0.3008**	**0.0648 to 0.2208**	**0.0557 to 0.2810**

The bold values represent the difference in each column.

Table 7. Evaluation results of prediction interaction among the new drugs

Method	Acc.	Prec.	Sen.	F1	MCC	AUC	AUPR
S-GNB	0.6038	0.5714	0.8302	0.6769	0.2328	0.6275	0.5863
S-DT	0.5755	0.6081	0.4245	0.5001	0.1583	0.5755	0.5459
S-LR	0.5849	0.5577	0.8208	0.6641	0.1926	0.6165	0.5995
KG-TransE	0.5566	0.5998	0.3491	0.4405	0.1244	0.5801	0.6009
KG-ComplEx	0.6274	0.6311	0.6132	0.6220	0.2548	0.6555	0.6361
Our method	0.6557	0.6813	0.5849	0.6294	0.3145	0.7361	0.7467
Range of difference	**0.0283 to 0.0991**	**0.0502 to 0.1236**	**−0.2359 to 0.2358**	**−0.0347 to 0.1889**	**0.0597 to 0.3021**	**0.0806 to 0.1606**	**0.1106 to 0.2008**

The bold values represent the difference in each column.

Table 8. The top 15 interactions of *Cannabidiol* drug-related

Rank	Drug name	DrugBank ID	Score	Evidence	Rank	Drug name	DrugBank ID	Score	Evidence
1	*Caffeine*	DB00201	0.9938	Confirmed	9	*Temazepam*	DB00231	0.9784	Confirmed
2	*Doxorubicin*	DB00997	0.9864	Confirmed	10	*Warfarin*	DB00682	0.9763	Confirmed
3	*Cyclophospha-mide*	DB00531	0.9852	Confirmed	11	*Cimetidine*	DB00501	0.9756	Confirmed
4	*Clomipramine*	DB01242	0.9851	Confirmed	12	*Terbinafine*	DB00857	0.9651	Confirmed
5	*Mirabegron*	DB08893	0.9844	Confirmed	13	*Esomeprazole*	DB00736	0.9600	N.A.
6	*Theophylline*	DB00277	0.9831	Confirmed	14	*Cilostazol*	DB01166	0.9584	Confirmed
7	*Azelastine*	DB00972	0.9812	Confirmed	15	*Aminophylline*	DB01223	0.9473	Confirmed
8	*Buprenorphine*	DB00921	0.9810	Confirmed					

Table 9. The top 15 interactions of *Cisplatin* drug-related

Rank	Drug name	DrugBank ID	Score	Evidence	Rank	Drug name	DrugBank ID	Score	Evidence
1	*Daunorubicin*	DB00694	0.9514	Confirmed	9	*Cabazitaxel*	DB06772	0.9045	Confirmed
2	*Doxorubicin*	DB00997	0.9441	Confirmed	10	*Methotrexate*	DB00563	0.8976	Confirmed
3	*Docetaxel*	DB01248	0.9381	Confirmed	11	*Ibuprofen*	DB01050	0.8966	Confirmed
4	*Indomethacin*	DB00328	0.9245	Confirmed	12	*Ulipristal*	DB08867	0.8944	N.A.
5	*Paclitaxel*	DB01229	0.9156	Confirmed	13	*Vismodegib*	DB08828	0.8942	N.A.
6	*Regorafenib*	DB08896	0.9147	N.A.	14	*Ponatinib*	DB08901	0.8926	Confirmed
7	*Taurocholic acid*	DB04348	0.9141	N.A.	15	*Sparfloxacin*	DB01208	0.8891	N.A.
8	*Pitavastatin*	DB08860	0.9124	N.A.					

three drugs and potential interaction mechanisms analysis indicate that our model can be applied to real life. More importantly, we give two predicted interactions with high confidence.

Although DeepLGF shows good prediction performance, this work is still faced with many challenges. First, the negative samples are randomly selected, which can bring certain noise to limit predicting ability. Proposing a suitable negative samples selected strategy is needed in the future. Another challenge is that the simple sequence of the drug is used as CS information, which may only provide limited one-dimensional information.

Thus, if the three-dimensional structure of drugs can be utilized to model, the performance can be further optimized.

In summary, DeepLGF is an effective model for discovering potential DDIs by narrowing the search space. It can be used for preliminary screening of potential DDIs, before the wet laboratory experiments. The accurate identification of DDIs not only can reduce medical accidents but also can promote the drug discovery process and supply important hints for understanding the drug action mechanisms. Overall, our study can provide new insights for rapidly predicting potential DDIs.

Table 10. The top 15 interactions of *Dexamethasone* drug-related

| Rank	Drug name	DrugBank ID	Score	Evidence	Rank	Drug name	DrugBank ID	Score	Evidence
1	*Estrone*	DB00655	0.9997	Confirmed	9	*Cyclophospha-mide*	DB00531	0.9988	Confirmed
2	*Docetaxel*	DB01248	0.9997	Confirmed	10	*Cerivastatin*	DB00439	0.9984	Confirmed
3	*Estradiol*	DB00783	0.9997	Confirmed	11	*Ethinylestradiol*	DB00977	0.9984	Confirmed
4	*Doxorubicin*	DB00997	0.9995	Confirmed	12	*Vincristine*	DB00541	0.9984	Confirmed
5	*Caffeine*	DB00201	0.9994	N.A.	13	*Loperamide*	DB00836	0.9984	Confirmed
6	*Mycophenolate mofetil*	DB00688	0.9993	Confirmed	14	*Everolimus*	DB01590	0.9984	Confirmed
7	*Pravastatin*	DB00175	0.9992	Confirmed	15	*Paclitaxel*	DB01229	0.9984	Confirmed
8	*Testosterone*	DB00624	0.9989	Confirmed					
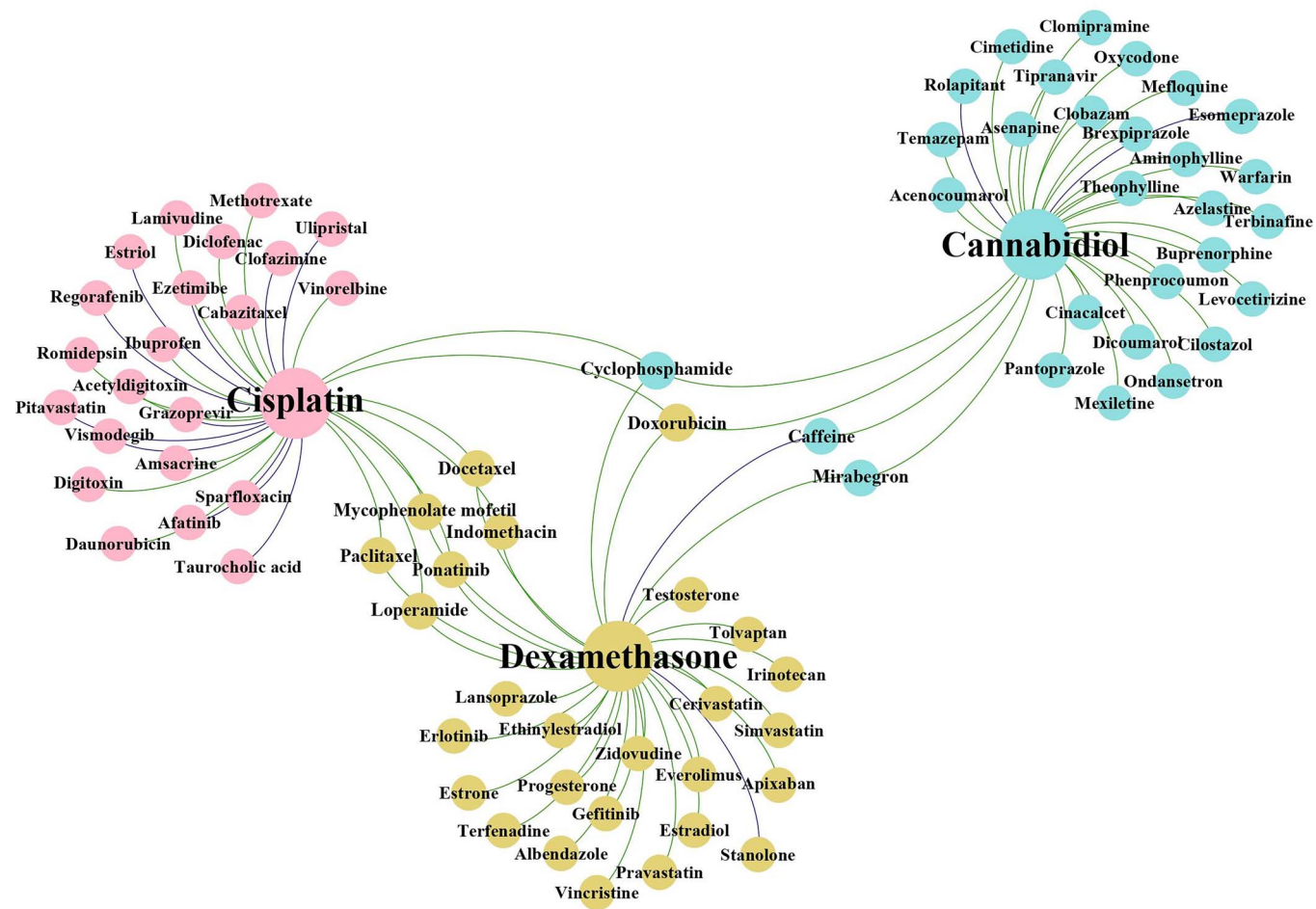
Figure 9. Network visualization of newly discovered and known DDIs of the top 30 in each case study. Three colors of the node represent three case studies respectively. Purple lines mean the interactions with no support and green lines mean confirmed interactions.

Key points

- We propose a novel framework based on biochemical KG to predict potential DDIs and fully utilize local–global information to improve the performance.
- Technically, we propose a model of BFGNN to represent the BF information and four integration methods to obtain the most suitable, which are crucial for model's prediction ability.
- We have conducted extensive experiments on three prediction tasks and four datasets. Compared with baseline methods demonstrates the promising performance and robustness of our method.
- Case studies of three cancer-related and COVID-19-related drugs further indicate the prediction ability of the proposed model when facing practical concerns, which can also help to study cancer disease. And, we give two predicted interactions with high confidence.
- The online prescreening platform is built for computational chemistry and drug discovery-related researchers to simplify the screening process. Available at http://120.77.11.78/DeepLGF/.

Supplementary Data

Acknowledgments

Funding

Data Availability

DeepLGF is also publicly available as an online predictor at: http://120.77.11.78/DeepLGF/. All the codes are available at https://github.com/MrPhil/DeepLGF and original data of drugs are available at https://go.drugbank.com/.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Prueksaritanont T, Chu X, Gibson C, *et al.* Drug–drug interaction studies: regulatory guidance and an industry perspective. *AAPS J* 2013;**15**:629–45.

2. Percha B, Altman RB. Informatics confronts drug–drug interactions. *Trends Pharmacol Sci* 2013;**34**:178–84.

3. Kusuhara K. How far should we go? Perspective of drug-drug interaction studies in drug development. *Drug Metab Pharmacokinet* 2014;**29**:227–8.

4. Qiu Y, Zhang Y, Deng Y, *et al.* A Comprehensive Review of Computational Methods for Drug-drug Interaction Detection. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**:1.

5. Whitebread S, Hamon J, Bojanic D, *et al.* Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. *Drug Discov Today* 2005;**10**:1421–33.

6. Yu H, Zhao S, Shi J. STNN-DDI: a substructure-aware tensor neural network to predict drug-drug interactions. *arXiv preprint arXiv:2111.05708* 2021.

7. Wang Y; Min Y, Chen X, Wu J. *Multi-view graph contrastive representation learning for drug-drug interaction prediction*. In: *Proceedings of the Proceedings of the Web Conference*. Association for Computing Machinery, New York, NY, United States, 2021, **2021**; 2921–33.

8. Nyamabo AK, Yu H, Liu Z, *et al.* Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Brief Bioinform* 2021;**22**:1–12.

9. Wang F, Lei X, Liao B, *et al.* Predicting drug–drug interactions by graph convolutional network with multi-kernel. *Brief Bioinform* 2021;**23**(1):bbab511.

10. Deng Y, Qiu Y, Xu X, *et al.* META-DDIE: predicting drug–drug interaction events with few-shot learning. *Brief Bioinform* 2021;**23**(1):bbab514.

11. Mondal I. BERTKG-DDI: towards incorporating entity-specific knowledge graph information in predicting drug-drug interactions. In: *Proceedings of the Workshop on Scientific Document Understanding co-located with 35th AAAI Conference on Artificial Intelligence (AAAI)*. 2021; 35–40.

12. Yu H, Mao K-T, Shi J-Y, *et al.* Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol* 2018;**12**:101–10.

13. Shi J-Y, Mao K-T, Yu H, *et al.* Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. *J Cheminform* 2019;**11**:1–16.

14. Zhu Y, Li L, Lu H, *et al.* Extracting drug-drug interactions from texts with BioBERT and multiple entity-aware attentions. *J Biomed Inform* 2020;**106**:103451.

15. Zhang W, Chen Y, Liu F, *et al.* Predicting potential drug-drug interactions by integrating chemical, biological phenotypic and network data. *BMC Bioinformatics* 2017;**18**:1–12.

16. Zhang W, Jing K, Huang F, *et al.* SFLLN: a sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. *Inf Sci* 2019;**497**:189–201.

17. Katayama T, Wilkinson MD, Aoki-Kinoshita KF, *et al.* Bio-Hackathon series in 2011 and 2012: penetration of ontology and linked data in life science domains. *J Biomed Semantics* 2014;**5**: 1–13.

18. Lacroix, T.; Usunier, N.; Obozinski, G. *Canonical tensor decomposition for knowledge base completion*. In: *Proceedings of the International Conference on Machine Learning*, Stockholmsmässan, Stockholm, Sweden, 2018; 2863–72.

19. Dettmers T, Minervini P, Stenetorp P, Riedel, S. Convolutional 2d knowledge graph embeddings. In: *Proceedings of the Thirty-second AAAI Conference on Artificial intelligence*, New Orleans, Louisiana, USA, The Association for the Advancement of Artificial Intelligence, 2018.

20. Dai Y, Guo C, Guo W. Drug–drug interaction prediction with Wasserstein Adversarial Autoencoder-based knowledge graph embeddings. *Brief Bioinform* 2021;**22**:bbaa256.

21. Yu Y, Huang K, Zhang C, *et al.* SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics* 2020;**37**:2988–95.

22. Liu S, Zhang Y, Cui Y, *et al.* Enhancing drug-drug interaction prediction using deep attention neural networks. *IEEE/ACM Trans Comput Biol Bioinform* 2022;1. Early Access Articles.

23. Gottlieb A, Stein GY, Oron Y, *et al.* INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol* 2012;**8**:592.

24. Park K, Kim D, Ha S. Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks. *PloS one* 2015;**10**: e0140816.

25. Mohamed SK, Nounu A. Biological applications of knowledge graph embedding models. *Brief Bioinform* 2021;**22**:1679–93.

26. Le Q, Mikolov T *Distributed representations of sentences and documents*. *In Proceedings of the International conference on machine learning*, Bejing, China, 2014; 1188–96.

27. Trouillon T, Welbl J, Riedel S, *et al. Complex embeddings for simple link prediction*. In: *Proceedings of the International Conference on Machine Learning*, New York, New York, USA, 2016; 2071–80.

28. Knox C, Law V, Jewison T, *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* 2010; **39**:D1035–41.

29. Law V, Knox C, Djoumbou Y, *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014;**42**:D1091–7.

30. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;(46):D1074–82.

31. Wishart DS, Knox C, Guo AC, *et al.* DrugBank: a knowledge-base for drugs, drug actions and drug targets. *Nucleic Acids Res* 2008;**36**:D901–6.

32. Wishart DS, Knox C, Guo AC, *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;**34**:D668–72.

33. Ioannidis VN, Song X, Manchanda S, *et al.* Drkg-drug repurposing knowledge graph for covid-19. *arXiv preprint arXiv: 2010.09600.* 2020.

34. Yue X, Wang Z, Huang J, *et al.* Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics* 2020;**36**:1241–51.

35. Arora S, Liang Y, Ma T. A simple but tough-to-beat baseline for sentence embeddings. *ICLR Princeton University Press,* 2016.

36. Mikolov T, Chen K, Corrado G, *et al.* Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.* 2013; 1–12.

37. Mikolov T, Sutskever I, Chen K, *et al. Distributed representations of words and phrases and their compositionality.* In: *Proceedings of the Advances in Neural Information Processing Systems,* Harrahs and Harveys, Lake Tahoe, 2013; 3111–9.

38. Jupp S, Malone J, Bolleman J, *et al.* The EBI RDF platform: linked open data for the life sciences. *Bioinformatics* 2014;**30**:1338–9.

39. Smith B, Ceusters W, Klagges B, *et al.* Relations in biomedical ontologies. *Genome Biol* 2005;**6**:1–15.

40. Chen Y, Ma T, Yang X, *et al.* MUFFIN: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics* 2021;**37**:2651–8.

41. Wan F, Hong L, Xiao A, *et al.* NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics* 2019;**35**:104–11.

42. Vaswani A, Shazeer N, Parmar N, *et al. Attention is all you need.* In: *Proceedings of the Advances in Neural Information Processing Systems,* Long Beach Convention Center, Long Beach, 2017; 5998–6008.

43. Feng Y-H, Zhang S-W. DPDDI: a deep predictor for drug-drug interactions. *BMC Bioinformatics* 2020;**21**:1–15.

44. Chou K-C. Prediction of protein structural classes. *Crit Rev Biochem Mol Biol* 1995;**30**:275–349.

45. Rohani N. Drug-drug interaction predicting by neural network using integrated similarity. *Sci Rep* 2019;**9**:1–11.

46. Vilar S, Harpaz R, Uriarte E, *et al.* Drug—drug interaction through molecular structure similarity analysis. *Am Med Inform Assoc* 2012;**19**:1066–74.

47. Vilar S, Uriarte E, Santana L, *et al.* Detection of drug-drug interactions by modeling interaction profile fingerprints. *PloS one* 2013;**8**:e58321.

48. Zhang P, Wang F, Hu J. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci Rep* 2015;**5**:1–10.

49. Rohani N, Eslahchi C, Katanforoush A. Iscmf: integrated similarity-constrained matrix factorization for drug–drug interaction prediction. *NetwModel Anal Health Informatics Bioinformatics* 2020;**9**:1–8.

50. Schwarz K, Allam A, Perez Gonzalez NA, *et al.* AttentionDDI: siamese attention-based deep learning method for drug–drug interaction predictions. *BMC bioinformatics* 2021;**22**:1–19.

51. Chen X, Liu X, Wu J. GCN-BMP: investigating graph representation learning for DDI prediction task. *Methods* 2020;**179**:47–54.

52. Crippa JAS, Zuardi AW, Guimarães FS, *et al.* Efficacy and safety of cannabidiol plus standard care vs standard care alone for the treatment of emotional exhaustion and burnout among frontline health care workers during the COVID-19 pandemic: a randomized clinical trial. *JAMA Netw Open* 2021;**4**:e2120603.

53. Villar J, Añón JM, Ferrando C, *et al.* Efficacy of dexamethasone treatment for patients with the acute respiratory distress syndrome caused by COVID-19: study protocol for a randomized controlled superiority trial. *Trials* 2020;**21**:1–10.

54. Ren Z-H, Yu C-Q, Li L-P, *et al.* BioDKG–DDI: predicting drug–drug interactions based on drug knowledge graph fusing biochemical information. *Brief Funct Genomics* 2022.

55. Su X, You Z-H, Huang DS, *et al.* Biomedical Knowledge Graph Embedding with Capsule Network for Multi-label Drug-Drug Interaction Prediction. *IEEE Trans Knowl Data Eng* 2022;1. Early Access Articles.

56. Su X, Hu L, You Z, *et al.* Attention-based Knowledge Graph Representation Learning for Predicting Drug-drug Interactions. *Brief Bioinform* 2022;**23**:bbac140.

57. Ren Z-H, Yu C-Q, Li L-P, *et al.* BioChemDDI: predicting drug-drug interactions by fusing biochemical and structural information through a self-attention mechanism. *Biology* 2022;**11**:758.

58. Himmelstein DS, Lizee A, Hessler C, *et al.* Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife* 2017;**6**:e26726.

59. Bordes A, Usunier N, Garcia-Duran A, *et al.* Translating embeddings for modeling multi-relational data. *Adv Neural Inf Process Syste* 2013;**26**.

60. Cao S, Lu W, Xu Q. Grarep: learning graph representations with global structural information. In: *Proceedings of the Proceedings of the 24th ACM International on Conference on Information and Knowledge Management.* Association for Computing Machinery, New York, NY, United States, 2015; 891–900.

61. Tang J, Qu M, Wang M , *et al.* Line: large-scale information network embedding. In: *Proceedings of the 24th International Conference on World Wide Web.* International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2015; 1067–77.

62. Wang D, Cui P, Zhu W Structural deep network embedding. In: *Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 2016; 1225–34.

63. Kanatsoulis CI, Sidiropoulos ND. *TeX-Graph: coupled tensor-matrix knowledge-graph embedding for COVID-19 drug repurposing.* In: *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM).* 2021, p. 603–11. Society for Industrial and Applied Mathematics (SIAM). https://doi.org/10.1137/1.9781611976700.68

64. Bonner S, Barrett IP, Ye C, *et al.* Understanding the performance of knowledge graph embeddings in drug discovery. *Artif Intell Life Sci* 2022;**2**:100036.

65. Natarajan N, Dhillon IS. Inductive matrix completion for predicting gene–disease associations. *Bioinformatics* 2014;**30**:i60–8.

66. Lin X, Quan Z, Wang Z-J, *et al.* KGNN: knowledge graph neural network for drug-drug interaction prediction. *IJCAI* 2020;**380**:2739–45.

67. Nickel M, Murphy K, Tresp V, *et al.* A review of relational machine learning for knowledge graphs. *Proc IEEE* 2015;**104**:11–33.

68. Guan N, Song D, Liao L. Knowledge graph embedding with concepts. *Knowledge-Based Systems* 2019;**164**:38–44.

69. Rozemberczki B, Bonner S, Nikolov A, *et al. A unified view of relational deep learning for drug pair scoring* arXiv preprint arXiv:2111.02916 2021.