3DGT-DDI: 3D graph and text based neural network for drug-drug interaction prediction

Article *in* Briefings in Bioinformatics · April 2022

DOI: 10.1093/bib/bbac134

3 authors, including:

Guanxing Chen
Sun Yat-Sen University
16 PUBLICATIONS **88** CITATIONS
SEE PROFILE

Chen Chen
Qufu Normal University
105 PUBLICATIONS **1,034** CITATIONS
SEE PROFILE

OXFORD

Briefings in Bioinformatics, 2022, **23(3)**, 1–15
https://doi.org/10.1093/bib/bbac134
Problem Solving Protocol

3DGT-DDI: 3D graph and text based neural network for drug–drug interaction prediction

Haohuai He†, Guanxing Chen† and Calvin Yu-Chian Chen
Corresponding authors: Calvin Yu-Chian Chen, Artificial Intelligence Medical Center, School of Intelligent Systems Engineering, Sun Yat-sen University,Shenzhen 510275, China; Department of Medical Research, China Medical University Hospital, Taichung, 40447, Taiwan; Department of Bioinformatics and Medical Engineering, Asia University, Taichung, 41354, Taiwan. TEL: 15626413023 E-mail: chenyuchian@mail.sysu.edu.cn
†These authors contributed equally to this work.

Abstract

Motivation: Drug–drug interactions (DDIs) occur during the combination of drugs. Identifying potential DDI helps us to study the mechanism behind the combination medication or adverse reactions so as to avoid the side effects. Although many artificial intelligence methods predict and mine potential DDI, they ignore the 3D structure information of drug molecules and do not fully consider the contribution of molecular substructure in DDI. **Results:** We proposed a new deep learning architecture, 3DGT-DDI, a model composed of a 3D graph neural network and pre-trained text attention mechanism. We used 3D molecular graph structure and position information to enhance the prediction ability of the model for DDI, which enabled us to deeply explore the effect of drug substructure on DDI relationship. The results showed that 3DGT-DDI outperforms other state-of-the-art baselines. It achieved an 84.48% macro F1 score in the DDIExtraction 2013 shared task dataset. Also, our 3D graph model proves its performance and explainability through weight visualization on the DrugBank dataset. 3DGT-DDI can help us better understand and identify potential DDI, thereby helping to avoid the side effects of drug mixing. **Availability:** The source code and data are available at https://github.com/hehh77/3DGT-DDI.

Keywords: Drug-Drug-interaction, 3D Graph Neural Network, Explainability, Attention-Mechanism

Introduction

Drug–drug interactions (DDIs) refer to the pharmacological action between drug components [4]. DDIs occur during pharmaceutical co-administration. They are a common source of adverse drug reactions and lead to rising healthcare costs [8, 36]. Many potential risks for patients were increased by DDI, such as the use of multiple drugs at the same time in cancer patients [6]. Many unknown DDIs in the clinical trial stage were reported after approval of clinical use. The incidence rate and mortality rate of DDIs is about 3–5% [19]. Therefore, it is critical to monitor DDI and explore potential DDI before medication, since it can effectively reduce the possible risk caused by combination medication.

DDIExtraction-2013 shared task (DDI2013) aimed to research DDI type judgment. Many researchers have conducted a series of research using a variety of methods. In the initial challenge and the following years, the models proposed by researchers are machine learning methods based on statistics, and the best model in the initial challenge is FBK IRST [5]. The model consisted of two independent steps: first, DDI was detected. Secondly, a binary support vector machine classifier was trained using context and shallow language features. Model representation based on machine learning depends on the effective extraction of data features. Moreover, the discrimination of long sentences is not satisfactory [30]. After that, the model based on deep learning has been widely considered because of its performance.

Compared with machine learning methods, deep learning methods do not rely on detailed pre-extraction of features. Quan et al. [28] used four word-embedding methods to generate multi-channel word embedding, and then input it into convolutional neural networks (CNNs) to output the prediction results. Zhang et al. [40] used multilayer CNN to fuse multiple features of drugs. Nevertheless, the CNN method uses a convolution kernel to extract the receptive field, which ignores the sequential characteristics of the text. Because of cyclical connections, recurrent neural networks (RNNs) are more suitable for the task of text recognition. RNN method can read the text sequentially to get context information. The ordinary RNN layer has the problems of gradient disappearance and gradient explosion. Long short-term memory (LSTM) and gate recurrent unit (GRU) models solve the problem of RNN. Wang et al. [35] integrated text

Haohuai He is a postgraduate candidate in the School of Intelligent Engineering, Sun Yat-Sen University. His research interests focus on natural language processing, graph neural network and drug discovery.
Guanxing Chen is a Ph.D. candidate in the School of Intelligent Engineering, Sun Yat-Sen University. His research interests focus on explainable artificial intelligence, drug discovery, deep learning, biosynthesis, and vaccine design.
Calvin Yu-Chian Chen is the Dean of Intelligent Medical Center and a professor of school of intelligent systems engineering at Sun Yat-sen University. He also had been served as an Advisor or guest Professor in China Medical University, Massachusetts Institute of Technology (MIT), Peking University, University of Pittsburgh, and adjunct professor in Zhejiang University. His research interests include the computer vision, natural language processing and deep learning.
Received: January 10, 2022. **Revised:** March 16, 2022. **Accepted:** March 21, 2022

information and syntax information, used embedded layer, bidirectional LSTM (BiLSTM) layer and maxpooling layer in series to output classification results. Fatehifar et al. [9] proposed text and a variety of embedded vectors, considered the part of speech relationship and input them into BiLSTM after concatenating. However, like the CNN method, the RNN method cannot effectively use the overall information of the text. Therefore, in recent years, many new methods have introduced the attention mechanism from the transformer [34] model attempting to solve the above problems. Mostafapour et al. [25] packaged BiLSTM using a two-layer attention mechanism. Hong et al. [13] applied bidirectional GRU (BiGRU) and self-attention mechanism to capture short-term and long-term dependencies, integrated the local context features of entities into sentence coding. In recent years, the bidirectional encoder representations from transformers (BERT) [7] model based on transformer has been proposed. And the pre-training technique can effectively improve the accuracy of the text model. Zhu et al. [41] proposed that BIOBERT be embedded in the BiGRU layer and connected with the multilayer entity perceived attention (MEA) layer. Huang et al. [15] processed the description with BIOBERT and packBiGRU, input it to the full connection layer and then fused the output.

In addition, some recent studies have tried to use features contained in drugs that are beyond the scope of the text. Asada et al. [2] used molecular structure input to graph convolutional network (GCN) and then concatenated them to CNN. In 2021, they used SCIBERT input text and drug description concatenated graph neural networks (GNNs) of input drug smiles [3]. These methods make use of the structural characteristics or descriptive characteristics of drugs. Compared with the traditional 2D graph model, the 3D graph model inputs the coordinate features of nodes, which can obtain the exact position features of nodes. In other words, an aggregation function about position information is added. Schnet [29] used the embedding of distance as the location information into the aggregation information of the model. Physnet [33] applied the distance between atoms as an important feature. Its position aggregation function uses a radial basis function with a smooth cutoff. Dimenet [17] explicitly considered the distance between atoms and the angle between directed edges. The position aggregation function adopts the concatenation of distance and angle. Spherical net [24] further used three elements: distance, angle and torsion to accurately describe the position information between atoms, to stand for position aggregation.

Furthermore, when the molecular graph structure is used to build the model, we can see which substructures the model is focusing on by visualizing the attention weight of the molecular graph. Nyamabo et al. [26] plotted a heat map to show the contribution of substructures in drug molecules to the prediction of whether there is a DDI relationship. Li et al. [21] obtained the attention weight of the molecular graph model and the highest occupied molecular orbital and the lowest unoccupied molecular orbital. Through comparison, we can see the effectiveness of the model. They also used the relationship between sildenafil and nitrate drugs as a case to further analyze the explainability of the model.

As far as our knowledge is concerned, previous studies only used text information and 2D features of drugs, such as the molecular map constructed using a simplified molecular input line entry system (SMILES) of drug molecules. One limitation of these methods is that they did not consider the 3D features of drug molecular structure, which are richer than the 2D structure information. The traditional 2D smiles graph may have great limitations. In DDI, some molecules have high chemical possibilities. 3D conformations of these molecules play critical roles in DDI, while the 2D SMILES graph cannot capture that information. Moreover, the position information of entities can better help the model to combine the characteristics between entity structure and DDI description text. Previous studies did not use the structure information and position information of drug entities at the same time.

To solve the above problems, we proposed a new machine intelligence model called 3DGT-DDI. This model is composed of a 3D graph and SCIBERT. It can fuse text features, 3D structure and position information of drug molecular entities for the prediction and potential exploration of DDI relationships. First, to use the 3D characteristics of the drug molecule, we inputted the 3D conformation of SMILES into the 3D graph structure model to obtain the 3D structure characteristics of the drug entity. Then, to enrich the information of the drug entity, we carried out position embedding. In addition, to extract the information in the text more effectively, we adopted the pre-trained SCIBERT-based text feature extraction model. We also innovatively proposed a hidden layer feature fusion method. Finally, the 3DGT-DDI combined the entity features generated by the 3D graph and the semantic features generated by SCIBERT to learn the rich information of the context for DDI predictions. Our model can comprehensively consider text and entity information. For structural information, we used 3D structural information instead of traditional two-dimensional information. For text information, we considered the information in multiple hidden layers and use CNN for effective fusion.

Experiments have shown that 3DGT-DDI can effectively extract the DDI relationship. At the same time, our model has strong explainability and can be used to observe the substructures that play a prominent contribution to DDI drugs.

Materials and Methods

Datasets

DDI extraction 2013

In this study, we used the DDI2013 dataset. The datasets came from the DDIExtraction-2013 shared task (SemEval-2013 Task 9.2) [12].

Table 1. Statistics of DDI2013 dataset.

	Train		Test	
	DrugBank	MEDLINE	DrugBank	MEDLINE
Documents	572	142	158	33
Sentences	5675	1301	973	326
Drug pairs	26 005	1787	5265	451
Positive pairs	3789	232	884	95
Negative pairs	22 216	1555	4381	356
Advice	818	8	214	7
Effect	1535	152	298	62
Int.	179	10	94	2
Mechanism	1257	62	278	24

The dataset provides data from MEDLINE and DrugBank. MEDLINE part provides drug relationship descriptions extracted from abstracts of articles on MEDLINE and PubMed, while the DrugBank part obtained text description data and category labels on the DrugBank website [16, 37]. For each description text, the datasets provided the drug name and ID appearing in the text and the drug interaction type label between the two drug IDs.

The task defines the following four interaction labels and a negative label [12].

- Advice: this is assigned when a recommendation or advice regarding the concomitant use of two drugs is described.
- Effect: this type is assigned when the effect of the DDI is described.
- Int. (Interaction): this type is assigned when the sentence simply states that an interaction occurs and does not provide any detailed information about the interaction.
- Mechanism: this type is assigned when a pharmacokinetic mechanism is described in an input sentence.
- Negative: it means there is no relationship between the two drugs.

From Table 1, about 77% of the text descriptions were randomly selected as the training set, while the other samples were used as the test set. Whether in the training set or the test set, the labels of most samples are negative ones indicating that there is no interaction between drugs in the text.

To make the format of the DDI2013 dataset more consistent with our model, we preprocess the dataset as follows: Firstly, to input the data into the 3D Graph model, we need to eliminate the texts that cannot find smiles, remove the drugs which cannot carry out 3D conformation and eliminate the drugs with single atoms without relative positions of atoms. Detailed reasons and operation methods will be discussed in section 2.3. Then, for improving the reusability and robustness of the model, we replaced the name of the drug in the text and recorded the positions of the drugs. The specific reasons and operation methods are discussed in section 2.4.

Table 2. Statistics of DDIs 2013 dataset before and after processing.

	Train		Test	
	Before	After	Before	After
Drug pairs	27 792	19 947	5716	3957
Positive pairs	4021	2782	979	633
Negative pairs	23 771	17 165	4737	3324
Advice	826	534	221	143
Effect	1687	1065	360	193
Int.	189	144	96	47
Mechanism	1319	1039	302	250

Through the above methods, we filtered the dataset and eliminated a small part of the data. For each data, we added two attributes: the smiles of the drugs and the positions of drugs in the text.

From Table 2, although we screened the samples in datasets, the non-related drug type (negative pairs) still accounted for the majority in the training set and the test set. Therefore, we employed focal loss [22] in the training process, a method to solve the imbalance between samples.

The formula for focal loss is as follows:

$$Loss_{fl} = y * (-\alpha(1-y')^{\gamma}\log y') + (1-y)(-(1-\alpha)y'^{\gamma}\log(1-y')), \quad (1)$$

where y' is the prediction of the model, y is the groundtruth, γ and α are the parameters of focal loss [22]. The details of the use of focal loss in our study are described in section 3.

DrugBank

In the DDI2013 dataset, samples with the same drug pair might appear to have inconsistent DDI relationship due to different texts. Even the same drug pair can be classified as both as neg, which means that there is no DDI relationship or, other DDI relationship. Therefore, DDI2013 cannot evaluate the effect of the 3D graph model well.

Thus, we used the DrugBank [37] dataset for the DDI prediction task of the 3D Graph model. The dataset contains 1704 drugs and 191 400 DDIs classified into 86 general sentence structures, each of which describes a specific DDI type. Most of its samples only correspond to one DDI relationship, so it is almost equal to a multi-classification problem. Moreover, because all types in Drugbank have DDI relationships, there will be no contradiction between samples like DDI2013.

The overview of 3DGT-DDI

We proposed 3DGT-DDI, a novel DDI prediction architecture. We used the 3D graph model to obtain the three-dimensional structure of drug molecules in drug pairs. Besides, we used the feature extraction model of text

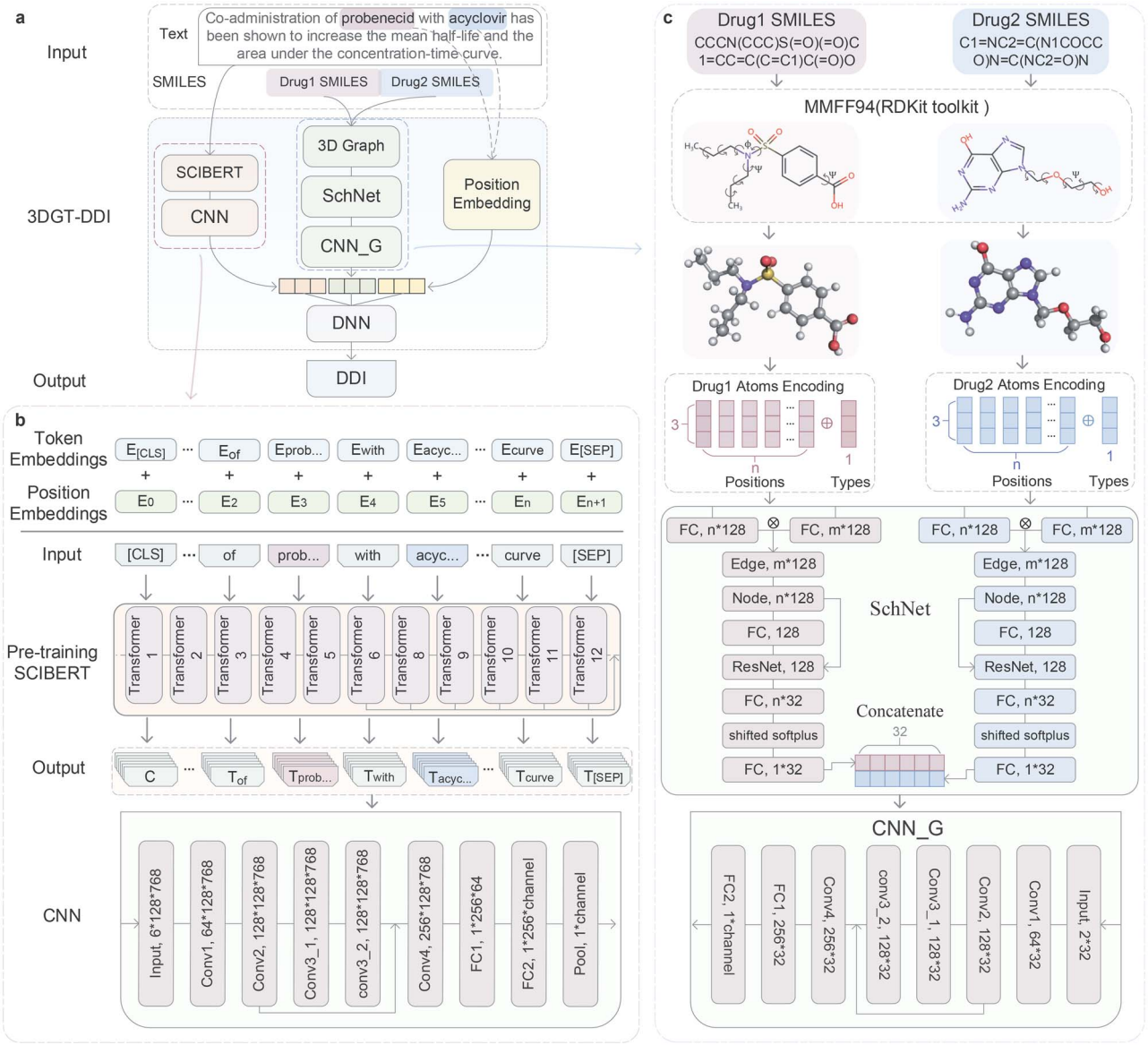

Figure 1. Overview of the 3DGT-DDI framework. **(A)** illustrates the work flow of the 3DGT-DDI framework. The 3DGT-DDI framework includes two key components—3d graph and text. **(B)** illustrates the details of text part. **(C)** illustrates the details of 3d graph part.

description based on BERT [7] to obtain the text information in DDI. Moreover, we extracted the position embedding vector of the drug name, to enhance the ability of context information feature extraction of the model.

Figure 1 is an overview of the 3DGT-DDI model. As can be seen from the figure, we entered the token of the text into the Bert model. In addition, we performed 3D conformation of drug molecule smiles through MMFF and then inputted it into the 3D graph network. Then, the position information of Drug1 and Drug2 was inputted into the embedding layer. Finally, we made feature concatenation and output the prediction of DDI.

3D Graph

The atomic coordinate information of molecular 3D structure comes from the force field for the optimal conformation of molecular energy. To obtain molecular 3D conformation, we need to obtain the corresponding smiles of the drug in the DDI2013 dataset. Through PubChem datasets [16], we found the smiles corresponding to most drugs in DDI2013. Then we used the RDKit [18] Library in Python to optimize the 3D conformation. Previous studies have shown MMFF is a common force field optimization method [32]. It provides a possible energy optimal conformation of compounds in space. Moreover, we proposed the MMFF force field of the molecule to obtain the energy optimal conformation of the drug molecule in a three-dimensional space.

Compared with a general GNN, a 3D GNN can obtain the coordinate position information of nodes. Using the coordinate information can effectively obtain more molecular structure information. We use the Schnet as the infrastructure to build a neural network model for DDI prediction. By generating three-dimensional

conformations of drug molecules, we input them into the 3D GNN. After feature fusion, we can get the prediction of the DDI relationship.

The general model of 3D graph structure can be represented by a 4-tuple Graph=(U,V,E,P), where U is the global eigenvector, V is the node eigenvector, E represents the set of edges and finally, P represents the coordinate information of each node. The $E_i = (e, s, r) \in E$, where e means the features of edge, s and r mean the index of the sender and receiver node. The general 3D graph model message passing update equation is like eq 2:

$$\begin{aligned} e_k^{'} &= \phi^e(e_k, v_{r_k}, v_{s_k}, E_{s_k}, u, \rho^{p\to e}(\{r_h\}_{h=r_k\cup s_k\cup \aleph s k})), \\ v_i^{'} &= \phi^v(v_i, \rho^{e\to v}(E_i), \rho^{p\to v}(\{r_h\}_{h=i\cup \aleph_i})), \\ u^{'} &= \phi^u(\rho^{e\to u}(E^{'}), \rho^{v\to u}(V^{'}), \rho^{p\to u}(\{r_h\}_{h=1:n}), \end{aligned} \tag{2}$$

where φ represents the update function for the same class of features and ρ represents the transfer function from one class of features to another. Customized for a specific 3D graph structure, the edge feature may receive the old edge feature e_k, node features $v_r k$, $v_s k$ at both ends, the feature of the set of edges pointing to s_k, the global feature u and the feature from the union of $r_k \cup s_k \cup \aleph_s k$ using $\rho^{p\to e}$; the node may receive the old node feature v_i. And the features of edges aggregated by $\rho^{p\to v}$ and the structural information of adjacent nodes$\{r_h\}_{h=i\cup \aleph_i}$; global features u may receive the aggregation of features of all edges, nodes and positions in the graph.

Figure 1c shows details of the 3d graph part of 3DGT-DDI. The Schnet framework constructs continuous distance features. The continuous filter operation is used to input the distance between atoms into a continuous mapping. For a given n atom features:

$$X^l = (\boldsymbol{x}_1^l, \ldots, \boldsymbol{x}_n^l), \boldsymbol{x}_i^l \in R^F. \tag{3}$$

Their location information is $\mathbf{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)$ where $\boldsymbol{r}_i \in R^D$. The continuous filter revolution operation of the first layer requires one Filter generating function:

$$\boldsymbol{x}_i^{l+1} = (X^l * \mathrm{W}^l)_i = \sum_j \boldsymbol{x}_j^l \circ \mathrm{W}^l(\boldsymbol{r}_i - \boldsymbol{r}_j), \mathrm{W}^l : R^D \to R^F. \tag{4}$$

Then, the above information is aggregated. The Schnet message update function is shown in eq 5:

$$\begin{aligned} e_k^{'} &= \phi^e(v_{r_k}, \rho^{p\to e}(\{r_h\}_{h=r_k\cup s_k})), \\ v_i^{'} &= \phi^v(v_i, \rho^{e\to v}(E_i)), \\ u^{'} &= \phi^u(u, \rho^{v\to u}(V^{'})). \end{aligned} \tag{5}$$

The 3D information in P is converted and used to update each edge feature e_k. Hence, for the three positional aggregation functions, the Schnet only uses $\rho^{p\to e}$ for edge updates.

When applying Schnet to the 3D graph structure of drug molecules, we encountered some problems. Schnet is suitable for the regression prediction tasks of some molecular attributes, rather than the prediction problems of DDI types. Moreover, because Schnet needs to calculate the distance of nodes, we met errors when inputting molecules with only one atom. In order to solve the above problem, we modify the number of node characteristic output channels of Schnet and use a multilayer one-dimensional convolutional neural network CNN_g. The structural features of the two drugs were extracted. And we filtered out molecules with only one atom in the datasets. We used the PyG [10] and DIG [23] repositories in Python to build 3d graph part of 3DGT-DDI. After 3D reconstruction of drug pairs molecule D_1 and D_2, we obtained the corresponding $pos1 \in R^{N1,3}$, $pos2 \in R^{N2,3}$ and atomic category numbers Z_1 and Z_2. Those features were inputted into Schnet, then calculated the RBF coordinates of POS, and embedded the atomic categories,

$$\begin{aligned} F_1(pos, z) &= SchN(D1), \\ F_2(pos, z) &= SchN(D2), \end{aligned} \tag{6}$$

then we concatenated the output features from Schnet, got the output features F, as follows:

$$F = CNN_g(Cat(F_1, F_2)). \tag{7}$$

Text and position embedding

In the DDI2013 data set, for each drug pair, there is a corresponding text description. We considered DDI extraction as a task to identify drug pairs in an input sentence in which the interaction of the pairs is described and to assign the right types of interactions to the pairs [3]. The human attention mechanism is derived from intuition. It is a strategy by which humans use limited attention resources to quickly filter out high-value information from high-volume data. The attention mechanism in deep learning draws on human attention thinking and is widely used in natural language processing [14]. So, we introduced the multi-head attention mechanism in our research, hoping to extract the key information of the DDI relationship in the text description [34].

For a text X, it is divided into

$$X = (w_1, w_2, \ldots, w_n), \tag{8}$$

where each w_i represents a token. For each vector w, we pass in three full connect layers to get three feature vectors: query Q, key K and value V.

$$\begin{aligned} Q &= W_q X + B_q, \\ K &= W_k X + B_k, \\ V &= W_v X + B_v, \end{aligned} \tag{9}$$

where the dimensions of Q, K and V are d_q,$d_k$and d_v, respectively.

Using Q to query the K of each word, we can get the total attention parameters, where the attention score is calculated by

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}}V). \quad (10)$$

Figure 1b shows the details of the text part of 3DGT-DDI. In our model, we use BERT as the infrastructure [7]. BERT uses WordPiece [39] for unsupervised tokenization of the input text. The vocabulary is built such that it contains the most frequently used words or subword units. BERT can better extract the feature information from the text. BERT is trained on two tasks: predicting randomly masked tokens and predicting whether two sentences follow each other. We know that BERT has many parameters and a complex model structure. For better training, we usually load a pre-trained model parameter before training a specific task and apply the idea of pre-training to a new training task. Therefore, we introduce the pre-training model, SCIBERT, into our DDI model, to improve the efficiency of training and better extract text-related features.

SCIBERT was trained on a random sample of 1.14M papers from Semantic Scholar [1]. This corpus consists of 18% papers from the computer science domain and 82% from the broad biomedical domain. SCIBERT follows the same architecture as BERT but is instead pre-trained on the scientific text and refers to the original vocabulary released with BERT as BASEVOCAB. Considering that most papers trained by SCIBERT come from medicine, we believe that the migration of SCIBERT's pre-training model can better extract the relevant features of DDI categories when applied to downstream tasks.

In the model of fusing multiple hidden layers, similar to the study [40], we use a multilayer convolution neural network to convolute the fused multi hidden layer features to obtain high-dimensional features and finally output them through the full connection layer. We consider that the 'CLS' token at the beginning of the text will be extracted in the classification task. In order to improve the feature extraction range, we use an adaptive multiple fusion layer for hidden layer features to improve the accuracy of feature extraction. Considering the existence of multiple hidden layers h_i in BERT model, we used a learnable parameter w to fuse the output of multilayer hidden layers.

$$Out = W[CNN(ConCat(H_n, H_{n-1}, H_{n-2}, \ldots, H_{n-r+1}))], \quad (11)$$

where r is a super parameter which represents the range of hidden layers fused in an adaptive multiple fusion layer. In the experiment, we set it to 6.

We used the huggingface [38] repository in Python to build the text part of 3DGT-DDI. To improve the ability of our deep learning model to extract the context relationship in the sentence, like the previous research, we also introduced position embedding for the position of the drug pairs in the text [3, 9, 11, 31].

In addition, since the text model will train a corpus of words, if the drug name is not processed, it will not be able to predict DDI once a new drug beyond the scope of the training set is encountered.

For improving the generalization ability and robustness of the model for DDI relationship prediction, like [3], we processed the text as follows: replaced the names of the two drugs of the drug pair with 'Drug1' and 'Drug2' according to the front and rear positions, and modified the names of other drugs in the statement to 'Drugother', which may enhance the new drug description prediction capacity of the model.

Experiment

Metrics

When measuring the effectiveness of a model in a multiclassification problem, the common measures in a binary classification problem like precision, recall and pr curve, auc, etc., are not applicable. In this case, we consider a modified version of the F1 score of the dichotomous model to make it applicable to the multi-classification problem. To evaluate the performance of our 3D graph model on the DrugBank dataset, we used three metrics: the 'accuracy' (ACC), the 'area under the precision-recall curve' (AUPRC) and the 'area under the receiver operating characteristic' (AUROC).

There are two common F1 score revisions: micro-f1($F1_{micro}$) and macro-f1($F1_{macro}$) score. These two F1 score metrics have corresponding formulas for calculating precision(P) and recall(R). Micro precision, recall and F1 score are calculated as follows:

$$P_{micro} = \frac{\Sigma_{i=1}^{n} TP_i}{\Sigma_{i=1}^{n} TP_i + \Sigma_{i=1}^{n} FP_i},$$
$$R_{micro} = \frac{\Sigma_{i=1}^{n} TP_i}{\Sigma_{i=1}^{n} TP_i + \Sigma_{i=1}^{n} FN_i},$$
$$F1_{micro} = \frac{2 * P_{micro} * R_{micro}}{P_{micro} + R_{micro}}, \quad (12)$$

where TP represents the true positive case, FP represents the false positive, and FN represents the false negative. i represents the i-th type.

According to [20, 27], the authors pointed out that there are two formulas for calculating $F1_{macro}$ scores. One is to calculate the average precision and recall of each type, and $F1_{macro}$ is the harmonic average of P and R:

$$F1 = \frac{2 * P * R}{P + R}; \quad (13)$$

and the other formula is calculated $F1_{macro}$ as the average value of $F1_{macro}$ in each type i:

$$\frac{\sum_{i=1}^{n} F1_{macro,i}}{n}. \quad (14)$$

Therefore, researchers must clarify which formula of $F1_{macro}$ was used. In this paper, the harmonic mean of P and R was used to calculate $F1_{macro}$ score.

So, Macro precision, recall and F1 score are calculated as follows:

$$P_{macro} = \frac{\sum_{i=1}^{n} P_i}{n},$$
$$R_{macro} = \frac{\sum_{i=1}^{n} R_i}{n},$$
$$F1_{macro} = \frac{2 * P_{macro} * R_{macro}}{P_{macro} + R_{macro}}. \quad (15)$$

We know by the formula that $F1_{micro}$ will give more weight to the categories with more samples. And $F1_{macro}$ looks more equally at each type, regardless of its sample size. According to Opitz [27], it is more appropriate to use $F1_{macro}$ when the label of datasets is unbalanced. We also prove that $F1_{micro}$ score, P_{micro} and $F1_{micro}$ are equivalent. They are equivalent to accuracy, the details of proving was shown in Supplementary materials

Training setting

According to the dataset section, we know that even after resampling the dataset, 'negative' types were still in the majority, and more than 80% of the data are still in the 'negative' label in both the test and training sets. There is a serious imbalance in the dataset. To solve this problem, instead of using the cross-entropy loss function, which is a common loss function for multi-category problems, we used focal loss [22] to solve the sample imbalance problem during training. For FocalLoss's weights, we used the inverse normalization of the number of categories of the dataset as weights for each type:

$$Weight_i = \frac{\frac{TotalNum}{Num_i}}{\sum_{i=1}^{n} \frac{TotalNum}{Num_i}}. \quad (16)$$

In particular, for solving the problem of dataset imbalance, we set the weight of 'negative' to a small value, let the model first focus on the identification training of samples with DDI relationship. And then gradually increase the weight of 'negative', so as to reduce the impact of sample imbalance. In the DDIExtraction 2013 shared task, the official development dataset is not provided; thus, we prepared a development dataset from the official training dataset. We used focal loss as the loss function and employed Adam optimizer.

We used the 3D graph model alone to train on the DrugBank dataset. By changing the head or tail of the drug pairs, we constructed the drug pair that did not exist in DrugBank as a negative class and used the drug pair that exists in DrugBank as a positive class. Then, the training set, verification set and test set are randomly divided on the DrugBank dataset. Because there is no serious problem of sample imbalance, the cross-entropy loss function is used. Table 3 shows the superparameter settings in the experiment.

Table 3. Training setting

	Parameter	Value
DDI2013	Learning rate	2e-6
	L2 weight decay	1e-2
	Number of epochs	200
	Mini-batch size	6
	Hidden_size	768
	Word embedding size	64
	Max length	128
	Number of Multiple hidden layers	6
	Cutoff	10
	Number of Layers	6
	Number of Hidden	128
	Number of Filters	128
	Word position embedding size	128
drugbank	Learning rate	2e-5
	L2 weight decay	1e-2
	Number of epochs	200
	Mini-batch size	16

Table 4. Evaluation on DDI2013 on the test set

Methods	P_{macro}	R_{macro}	$F1_{macro}$ (%)
FBK-irst [5]	64.6	65.6	65.1
MCCNN [28]	76.0	65.3	70.2
DLSTM [35]	72.5	71.5	72.0
POS with Bi-LSTM [9]	80.2	77.1	78.6
Attention Bi-LSTM [25]	80.0	77.0	78.5
BERE [13]	76.8	71.3	73.9
MAE [41]	81.0	80.9	80.9
EGFI [15]	83.4	85.2	84.2
GCN with CNN [2]	71.97	68.44	70.16
SCIBERT with GNN [3]	**85.36**	82.83	84.08
3DGT-DDI (ours)	81.17	**88.07**	**84.48**

Results and Discussion

Model performance

We benchmarked the 3DGT-DDI model on the test dataset of the DDI 2013 dataset to compare the performance of 3DGT-DDI with other state-of-the-art models. Table 4 and Figure 2 show the performance of the 3DGT-DDI model and other models on the DDI2013 test set. We use $F1_{macro}$, P_{macro} and R_{macro} to evaluate the performance of the model, and the most important model performance metric is $F1_{macro}$ score.

Table 4 shows the performance of 3DGT-DDI models and the state-of-the-art models. Our model reached 88.07% on R_{macro} and 84.48% on $F1_{macro}$, outperforming other state-of-the-art models. In general, the models of

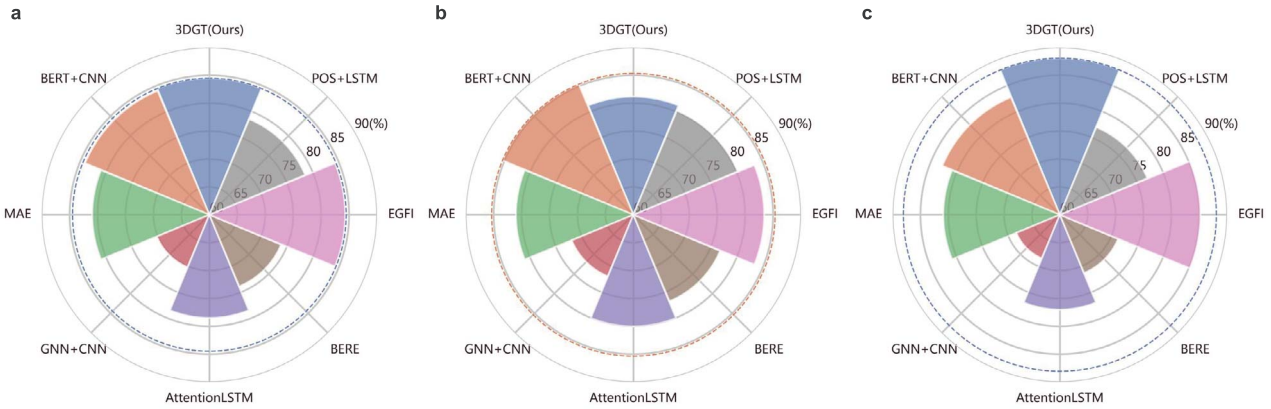

Figure 2. Test result compare on DDI2013, **(A)**:$F1_{macro}$, **(B)**:P_{macro}, **(C)**:R_{macro}.

Table 5. Performance on individual DDI types in $F1_{macro}$

	DDI Types				
Method	**Negative**	**Advice**	**Effect**	**Int.**	**Mech. (%)**
POS with Bi-LSTM [9]	95.23	82.92	75.86	50.09	84.51
Attention Bi-LSTM [25]	**96.8**	81.9	77.4	58.4	78.0
MAE [41]	–	86.0	80.1	56.6	84.6
EGFI [15]	95.05	87.7	**84.8**	57.9	**87.2**
SCIBERT with GNN [3]	–	**92.07**	80.48	49.25	86.33
3DGT-DDI(ours)	**96.89**	84.34	82.60	74.47	82.57
–Without Pos	96.58	82.10	81.08	**77.42**	80.88
–Without 3D Graph	96.33	81.37	78.80	76.09	80.71
–Text Only	96.25	81.23	79.65	72.16	80.98

using attention mechanism and pre-training method are better than using machine learning algorithms, CNN and RNN. In addition, we can see that the previous works [2, 3] use text model, 2D GNN and position embedding to form DDI prediction model. According to table 4, 3DGT-DDI has better performance than these models. This can further prove that the molecular structure feature extraction ability of 3D graph model is better than that of 2D graph model.

Moreover, our 3DGT-DDI model comprehensively considers the text information and the 3D graph structure information of drug molecules, which is superior to other attention-based models.

The performance of the 3DGT-DDI model under different types is shown in Table 5. Except for negative classes, 3DGT-DDI model performs best on advice. The performance of the 'int.' type is the worst. We think it may be the imbalance of label affected model on the ability to predict DDI type.

In Table 5, we also selected five other state-of-the-art models with the highest score of $F1_{macro}$ for comparison. We can see that our model is the best in 'negative' and int. types. In particular, the type of 'int.' achieved noticeable performance improvement (58.4% versus 74.47%).

To further verify the prediction ability and the robustness of the model, we also conducted a 5-fold cross validation for 3DGT-DDI on the DDI2013 training dataset. Table 6 shows the validation results. It can be seen that the difference of model performance between 5-fold on $F1_{macro}$, $F1_{micro}$, and other metrics is small (< 1%), which further proves that the robustness and prediction ability of the model.

On the DrugBank dataset, our test results of 3D graph model are as follows: Our model achieved 0.915 accuracy, 0.970 auroc and 0.968 auprc. Figure 3 shows the ROC curve and PR curve, which proves that our 3D graph model has a strong DDI relationship recognition ability.

We used the 3D graph model trained in DrugBank to conduct a cold start test in DDI2013 dataset. Since the labels of DDI2013 and DrugBank cannot be mapped, we modified DDI2013 to a binary label dataset with or without DDI. The results are shown in table 7: Compared with two 2D GNNs [2], CNNs for finger prints (NFP) and Gated Graph Neural Networks (GGNN), our 3D graph model has great advantages in accuracy. Precision and F1 score are also better than 2D GNN model. Therefore, the experiment has further shown that our 3D graph model has stronger prediction ability and versatility.

Ablation Experiment

For the sake of exploring the role of each component in the DDI model, we take the model using only the

Table 6. Results of 5-fold cross validation

	$F1_{micro}$	P_{macro}	R_{macro}	$F1_{macro}$(%)
FOLD1	94.21	81.72	87.22	84.38
FOLD2	94.19	81.61	88.13	84.74
FOLD3	93.98	81.21	88.18	84.55
FOLD4	94.16	81.57	87.28	84.33
FOLD5	93.88	81.00	88.30	84.49
Average	94.08 ± 0.13	81.42 ± 0.27	87.82 ± 0.47	84.49 ± 0.14

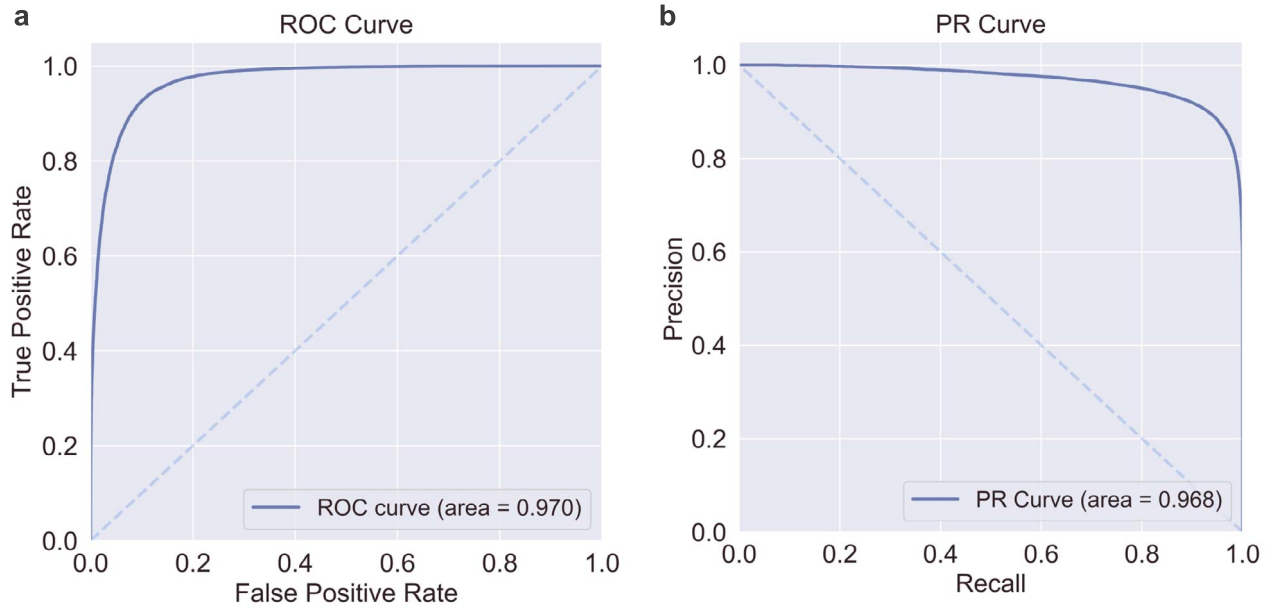

Figure 3. ROC curve and PR curve on DrugBank dataset.

Table 7. Classification of DDIs in DDI2013 by 2D and 3D graph DDI model

Model	P_{macro}	R_{macro}	$F1_{macro}$	Accuracy(%)
NFP	15.56	48.93	23.61	45.78
CGNN	15.11	**57.10**	23.90	37.72
3D Graph(Ours)	**17.26**	39.81	**24.08**	**59.84**

text part as the baselines and observe the micro-F1 and macro scores of different models by adding or removing other parts. From Table 8, we found that the performance of the model is the worst when only text is used, but the effect is not significantly improved when position embedding is added. This may be because there is not too much relationship between text model and position embedding. And the effect is better when text plus 3D graph is used than when position embedding is added, which may be owed to the 3D graph model carrying the three-dimensional structure information of the drug, and the 3D-structure information is more abundant than the position information. After using position embedding and 3D graph models, the complete 3DGT-DDI model has a larger improvement, which is greater than the sum of one of position embedding and 3D graph. It indicated that the use of position embedding may play an auxiliary role in the prediction of the 3D graph and help the recognition of the 3D graph. Therefore, although the promotion of position embedding is smaller than 3D graph, it is still an important part of the model. So, we added position embedding to our model. Finally, the DDI model that uses all the components performed best.

Error Analysis

We observed that the precision of our model is always lower than the recall. Similar to previous work [15, 41], Figure 4 shows the confusion matrix of the test set predicted by the model before and after normalization for error analysis.

First of all, we noted that many 'int.' samples were misjudged as 'effect', while the samples of the 'effect' label were almost not misclassified into 'int.', which is the main reason for the low macrof1 score of 'int.' type. By observing the previous work, we found that their model predictions were the same. Many of the 'int.' type samples were misjudged as 'effect'. This may indicate that the samples of the 'int.' label on the testset of DDI2013 have characteristics similar to the 'effect'.

We can review the definition of tags in the DDI2013 dataset, where the 'int.' type is assigned when the

Table 8. Comparisons of $F1_{macro}$ and other metrics on different parts

Features	$F1_{micro}$	P_{macro}	R_{macro}	$F1_{macro}$	$\Delta(\%)$
3DGT-DDI(ours)	94.41	81.17	88.07	84.48	–
–Without Pos	93.86	80.19	88.30	84.05	–0.43
–Without 3D Graph	93.45	79.38	87.26	83.13	–1.35
–Text Only	93.33	77.23	88.53	82.49	–1.99

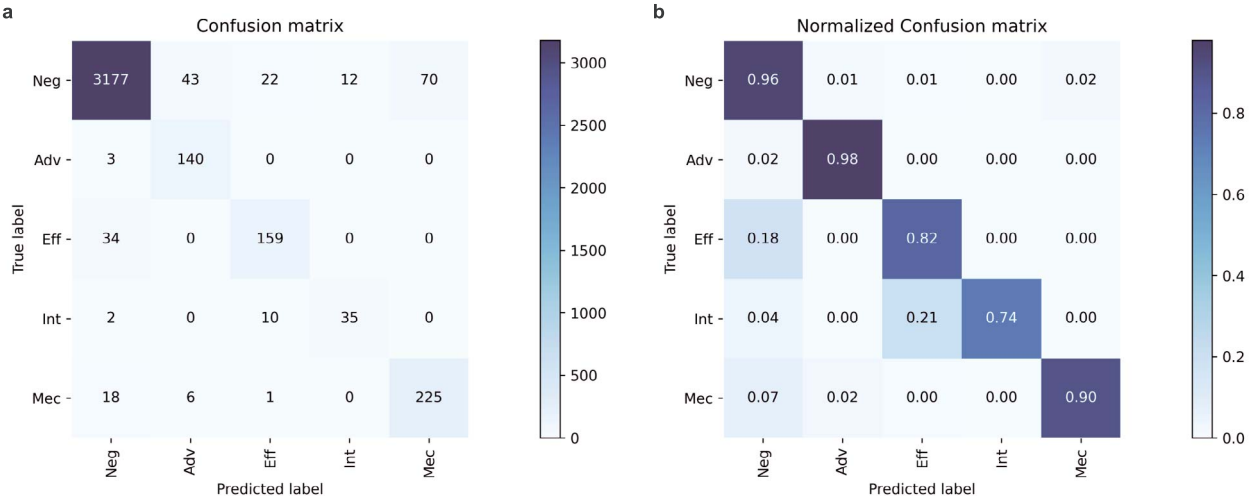

Figure 4. Confusion matrix without and with Normalization.

sentence simply states that interaction occurs and does not provide any detailed information about the interaction. Therefore, 'int.' represents a vague category. If there is a semantic that clearly points to a specific DDI category ('advice','effect' and 'mechanism') in the statement, misjudgment will occur.

For sample, here are three examples of the DDI2013 test set. An '…' indicates that part of the statement is omitted. The first two samples in the DDI2013 test set that our model misclassifies 'int.' into 'effect':

- 'Other drugs which may enhance the neuromuscular blocking action of DRUGOTHER such as DRUG1 include certain DRUG2 …'
- 'DRUG1 can interact with the drugs of the following categories: …-DRUG2: can increase seizure activity'
- 'DRUG1 can interact with the drugs of the following categories: …DRUGOTHER: can increase seizure activity …-DRUG2: may cause shortness of breath and bronchospasm'

The third sentence is similar to the second sentence, but drug2's choice is different. And it is correctly classified as 'int.' by our model

Firstly, it can be seen that there are two words 'enhance' and 'increase' in the semantics of the two sentences. The model may be confused because Drug2 is close to a word with a positive effect, resulting in misclassification into 'effect'. Moreover, for the second sentence, if drug2 chooses other drugs instead of the drug closed to 'increase', the model can correctly judge the int category, which further verifies our idea. Although the score of 'int.' is still not high, our model has been greatly improved in the 'int.' type compared with other SOTA models. It may be due to the use of focal loss to balance the extremely unbalanced type 'int.', and the use of a 3D graph model for better extracting and utilizing the drug features.

It can be seen that there are many 'negative' samples misjudged as 'advice', 'effect' and 'mechanism'. Moreover, due to a large number of samples in the negative, only a small part of the misclassification also leads to a significant decline in the precision of the four relationships in the positive label. It may be the main reason why the precision of our model is lower than the recall. However, there are relatively few samples whose positive types were misclassified into negative, which may be because the loss function was given sample weight, so the imbalance of samples was eliminated. Even without negative category filtering, it still achieved a high $F1_{macro}$ score.

Visualization with attention weights

According to the last layer output of the pre-training 3d graph network, we get the attention weight and color of each atom on the input molecule. The attention weight represents the contribution of atoms to the global feature U of the whole molecule, indicating its significance to the DDI. We take the first DDI type in the DrugBank dataset as an example. The description of that category in the dataset is that 'drug 1 may increase the light

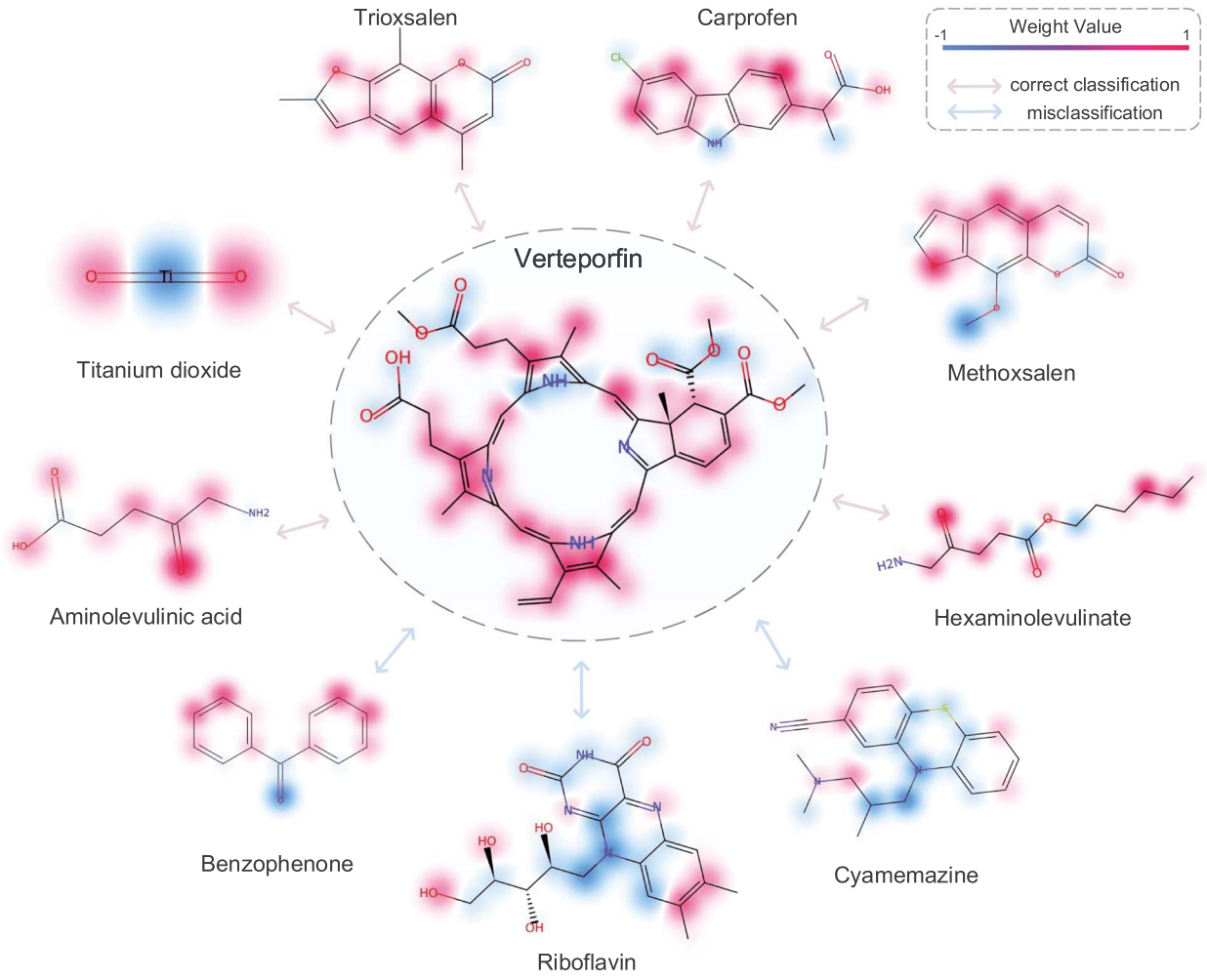

Figure 5. The test of 3DGT-DDI at DrugBank. Verteporfin is in the center of the figure, surrounded by drugs that react with Verteporfin to enhance photosensitivity. The weight value was used to indicate the importance of the model to the atoms in the drug molecule.

sensitivity of drug 2'. Moreover, we note that among all sample drug pairs belonging to label 1, drug 2 is verteporfin (DBid: db00460). Verteporfin is a benzoporphyrin derivative used to treat pathological myopia, ocular histoplasmosis and choroidal neovascularization in macular degeneration.

Figure 5 shows the reaction example of type 1 in the DrugBank dataset, that is, the drug relationship of other drugs with enhanced light sensitivity to verteporfin. And we show the weight of each atom in different drug molecules through the output of the 3D graph model. Red and blue represent more attention, while purple represents less attention. The shade of the color represents the degree of more or less attention. Among them, red indicates a positive correlation with the final predicted category and vice versa. For the example that the model correctly judges the DDI category, we can see that the model usually pays more attention to oxygen, OH, benzene ring and Ti in molecules interacting with verteporfin, and less attention to S and Cl. Carbon atoms at different positions on the benzene ring have received varying degrees of attention. In addition, the model captures both the functional groups of the molecule and the benzene ring correctly. These weights indicate that our model can correctly learn some information about the 3D structure of drug molecules. Through DrugBank, we can know that Verteporfin is transported in the plasma primarily by lipoproteins. Once verteporfin is activated by light in the presence of oxygen, highly reactive, short-lived singlet oxygen and reactive oxygen radicals are generated. According to the action mechanism of verteporfin, we can find that it will react through the excitation of light in the presence of oxygen. Therefore, its light-sensitive characteristics may depend on oxygen. And it is reasonable for the model to pay more attention to oxygen of drugs that improve the light sensitivity of verteporfin.

PDE5 inhibitors are effective drugs for the treatment of pulmonary hypertension and erectile dysfunction. However, when nitrate drugs are combined with PDE5 inhibitors, cGMP will be increased, which may lead to a sudden drop in blood pressure and even heart attack [21, 26]. We selected three PDE5 inhibitors (sildenafil, vardenafil and tadalafil) and nitrate drugs for testing to

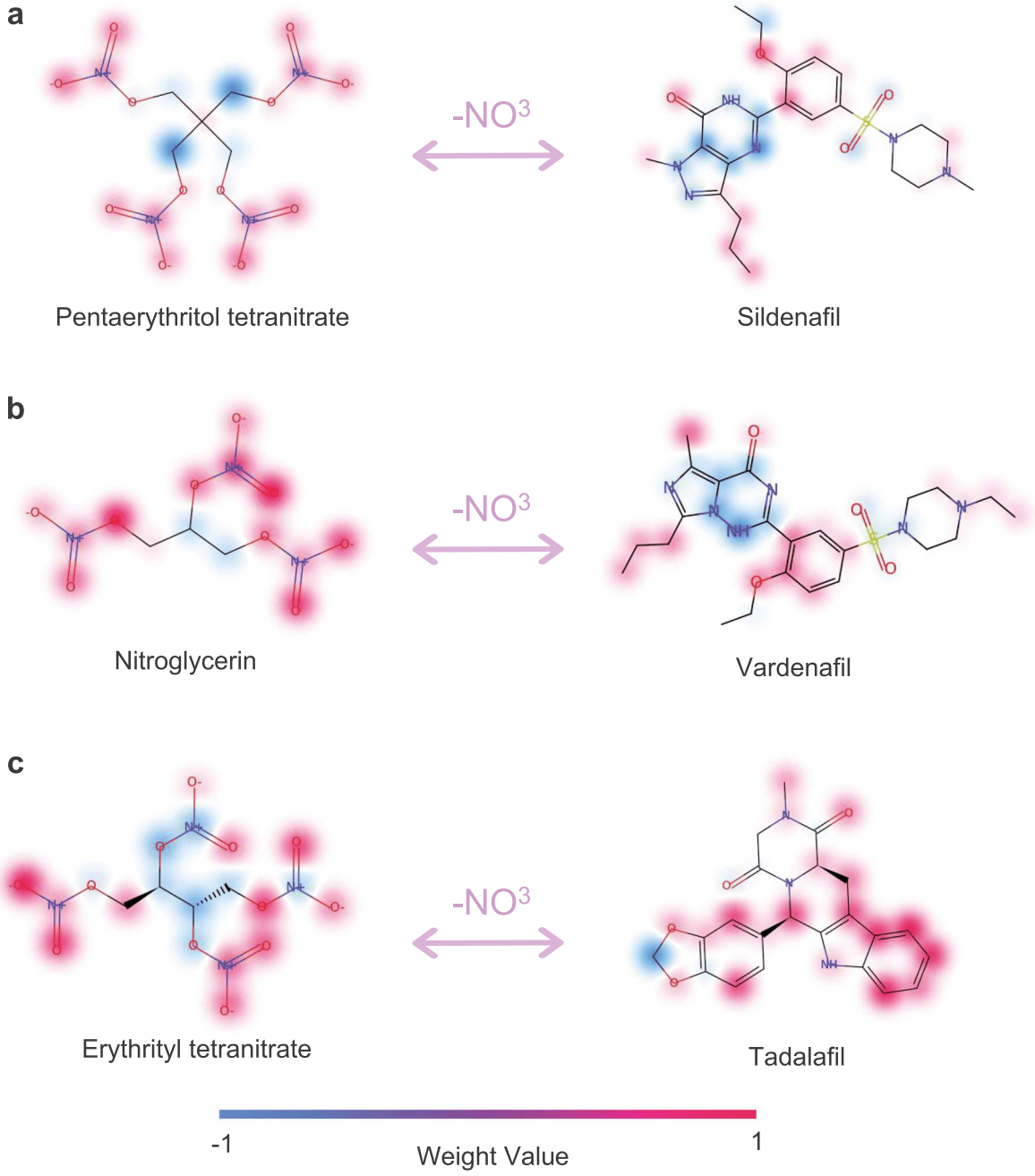

Figure 6. 3DGT-DDI provides explainability for DDI prediction. The left part of the figure illustrates nitrate-based drugs with each atom colored by the attention weight. The right part of the figure illustrates PDE5 inhibitors drugs. Note that 3DGT-DDI always pays more attention to the nitrate group.

analyze the role of nitrate groups. Through Figure 6, we can see that our 3D graph model gives a higher weight to the nitrate groups of drugs reacting with PDE5, which shows that our 3D graph model can use the reasonable information of molecular sparsity to predict DDI. In 'ethrytyl tetranitrate', we find that there are nitrate groups with low attention, which may be because the model pays more attention to oxygen, so it is unable to accurately capture the characteristics of complete nitrate groups. In general, our 3D graph model has been proved to be able to learn interpretable atomic-level characterization and capture some chemical knowledge, which may build a bridge between pre-training and downstream tasks, so as to improve the performance and explainability of the model.

In addition to capturing the weight of drug molecules, 3DGT-DDI can also capture the weight of the text. We selected an example of a training set and test set in the DDI2013.

Figure 7 shows the weights of text and drug molecule pairs in the sample. From Figure 7a, we can notice that the model correctly judges the important contribution of the words 'additive', 'effect' and 'given' to the prediction category. From Figure 7b, we can see that the word most

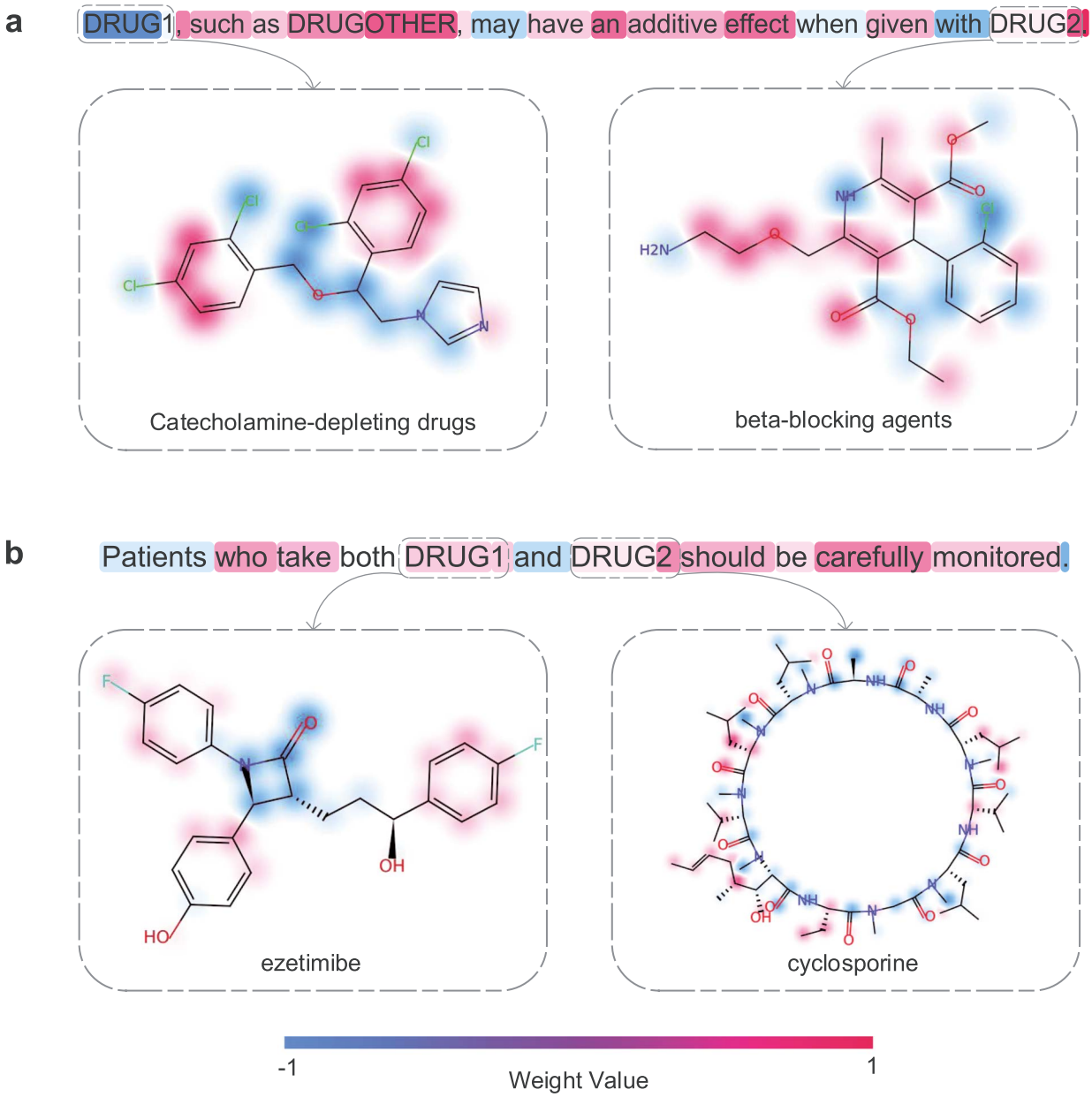

Figure 7. Weight visualization of 3DGT-DDI on DDI2013. The text and drug in the text were colored according to the attention weight. **(A)** is a sample of the 'effect' type in a training set of DDI2013 **(B)** is a sample of the 'advice' label in a testing set of DDI2013.

concerned by the model is 'carefully', which is a word used to express caution and is often used when advising others to take medicine. Moreover, the model also focuses on 'should' and 'monitored'. These words with high weight often appear in the suggestion statement. This further proves the explainability of the model. In addition, because our model constructs the 3D structure of molecules, in addition to the attention at the text level, we can see the weights of drug1 and drug2 in the model at the same time. Therefore, we can understand that the benzene ring and N of 'catecholamine depleting drugs' contribute more to the interaction, while the contribution of Cl is less. Similarly, other drug molecules can also use our model to get a visualization of atomic contribution. In a word, our model can capture deeper features at both text and molecular levels(words and atoms). Therefore, our model has strong explainability.

Conclusions

In this paper, we proposed 3DGT-DDI, an attention-based intelligent framework composed of pre-trained text models and 3D graph models for the prediction and in-depth interpretation of DDI relationships. Our model performs better than other baselines in SOTA on the DDI2013 dataset. In addition, we conducted verification and weight visualization on DrugBank, revealing the effectiveness of the 3D graph model and the interpretability of 3DGT-DDI. Experiments have shown that 3DGT-DDI is superior and more interpretable

than other baselines. Unlike traditional methods that used molecular 2D information, we introduced 3D information combining it with textual drug description information. The experimental results showed that the effective combination improves the performance of the model. The weight of 3D information and text showed better interpretability. At the same time, we used the hidden layer feature fusion method to further improve the ability to extract DDI types from the text training model, which significantly improves the performance of the model. More importantly, we combined the 3D structural features and position features of the drug entity. Visualizing the structure of text and drug molecules makes the training process of the model more transparent and operable. In general, 3DGT-DDI has learned the 3D structure of drug molecules and combined with text training of drug molecules. It has better performance and improves the effect of DDI prediction. Visualizing DDI can effectively help medical researchers to gain a deeper understanding of potential DDI and understand the mechanism behind combination medications or adverse reactions.

Key Points

- We proposed a DDI-type prediction model combining text description information and molecular 3D structure features. As a result, we obtained a better performance (a Macro F1-score of 84.48%) than other state-of-the-art methods on the DDI extraction 2013 dataset.
- To the best of our knowledge, we are the first to investigate the 3D graph model for Drug–Drug Interaction.
- We used the hidden layer feature fusion method to further improve the extraction of DDI types from the text model. At the same time, we combined the 3D structure features and position features of drug entities.
- We visualized the structure of text and drug molecules to help medical researchers have a deeper understanding of the relationship between DDI.

Acknowledgements

Thank the editors and reviewers for reviewing the paper.

Code and data availability

The source code and data are available at https://github.com/hehh77/3DGT-DDI.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 62176272), Guangzhou Science and Technology Fund (Grant No. 201803010072), Science, Technology and Innovation Commission of Shenzhen Municipality (JCYL 20170818165305521) and China Medical University Hospital (DMR-111-102, DMR-111-143, DMR-111-123). We also acknowledge the start-up funding from SYSU Hundred Talent Program.

References

1. Ammar W, *et al.* Construction of the literature graph in semantic scholar. 2018; arXiv preprint arXiv:1805.02262.
2. Asada, M. *et al.* Enhancing drug-drug interaction extraction from texts by molecular structure information. *Proceedings of the 56th Annual Meeting of the Association For Computational Linguistics* 2018;**2**:680–685.
3. Asada M, Miwa M, Sasaki Y. Using drug descriptions and molecular structures for drug-drug interaction extraction from literature. *Bioinformatics* 2021;**37**(12):1739–46.
4. Cheng F, Zhao Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J Am Med Inform Assoc* 2014;**21**(e2):e278–86.
5. Chowdhury MFM, Lavelli A. Fbk-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information. In: *Second Joint Conference on Lexical and Computational Semantics (∗ SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Atlanta, Georgia: 2013 Association for Computational Linguistics, 2013, 351–5.
6. del Giglio A, Miranda V, Fede A, *et al.* Adverse drug reactions and drug interactions as causes of hospital admission in oncology. *J Clin Oncol* 2009;**27**(15_suppl):e20656–6.
7. Devlin, *J. et al.* Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 2018;**1**:4171–4186.
8. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *The lancet* 2000;**356**(9237):1255–9.
9. Fatehifar M, Karshenas H. Drug-drug interaction extraction using a position and similarity fusion-based attention mechanism. *J Biomed Inform* 2021;**115**:103707.
10. Fey, M. and Lenssen, J. E. (2019). Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*.
11. He Z, *et al.* See: Syntax-aware entity embedding for neural relation extraction. In: *Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA: Assoc Advancement Artificial Intelligence, 2018.
12. Herrero-Zazo M, Segura-Bedmar I, Martínez P, *et al.* The ddi corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *J Biomed Inform* 2013;**46**(5):914–20.
13. Hong L, Lin J, Li S, *et al.* A novel machine learning framework for automated biomedical relation extraction from large-scale literature repositories. *Nature Machine Intelligence* 2020;**2**(6):347–55.
14. Hu D, Yaxin Bi; Bhatia, R.; Kapoor, S. (eds) *An introductory survey on attention mechanisms in nlp problems*, *Proceedings of SAI Intelligent Systems Conference*. London, UK: Springer, 2019, 432–48.
15. Huang, L. *et al.* Egfi: Drug-drug interaction extraction and generation with fusion of enriched entity and sentence information. *Briefings in Bioinformatics* 2021;**23**:1–14.
16. Kim S, Chen J, Cheng T, *et al.* Pubchem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;**47**(D1):D1102–9.
17. Klicpera, *J. et al.* (2020). Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*.
18. Landrum G. *Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling*, 2013.
19. Leape LL, *et al.* Systems analysis of adverse drug events. *JAMA* 1995;**274**(1):35–43.
20. Lewis DD, *et al.* Training algorithms for linear text classifiers. In: *Proceedings of the 19th annual international ACM SIGIR conference on*

Research and development in information retrieval, Zurich, Switzerland: ACM USA, 1996, 298–306.
21. Li P, Wang J, Qiao Y, *et al.* An effective self-supervised framework for learning expressive molecular global representations to drug discovery. *Brief Bioinform* 2021;**22**(6):1–14.
22. Lin T-Y, *et al.* Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*, 2017, 2999–3007.
23. Liu, M. *et al.* Dig: A turnkey library for diving into graph deep learning research. *Journal of Machine Learning Research* 2021a;**22**: 1–9.
24. Liu, Y. *et al.* (2021b). Spherical message passing for 3d graph networks. *arXiv preprint arXiv:2102.05013*.
25. Mostafapour, V. and Dikenelli, O. (2019). Attention-wrapped hierarchical blstms for ddi extraction. *arXiv preprint arXiv:1907.13561*.
26. Nyamabo AK, Yu H, Shi JY. Ssi–ddi: substructure–substructure interactions for drug–drug interaction prediction. *Brief Bioinform* 2021;**22**(6):1–10.
27. Opitz, J. and Burst, S. (2019). Macro f1 and macro f1. *arXiv preprint arXiv:1911.03347*.
28. Quan C, *et al.* Multichannel convolutional neural network for biological relation extraction. *Biomed Res Int* 2016;**2016**: 1850404.
29. Schütt, K. T. *et al.* Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *31st Annual Conference on Neural Information Processing Systems (NIPS)* 2017;**30**.
30. Segura-Bedmar I, Martínez P, Herrero-Zazo M. Lessons learnt from the ddiextraction-2013 shared task. *J Biomed Inform* 2014;**51**:152–64.
31. Shi Y, *et al.* Word embedding representation with synthetic position and context information for relation extraction. In: Wu X; Soon, OY; Aggarwal, C; Chen, H (eds) *2018 IEEE International Conference on Big Knowledge (ICBK)*. Singapore: IEEE, 2018, 106–12.
32. Tosco P, Stiefl N, Landrum G. Bringing the mmff force field to the rdkit: implementation and validation. *J Chem* 2014;**6**(1): 1–4.
33. Unke OT, Meuwly M. Physnet: a neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation* 2019;**15**(6):3678–93.
34. Vaswani A, *et al.* Attention is all you need. *In Advances in neural information processing systems* 2017;**30**:5998–6008.
35. Wang W, Yang X, Yang C, *et al.* Dependency-based long short term memory network for drug-drug interaction extraction. *BMC bioinformatics* 2017;**18**(16):99–109.
36. Wienkers LC, Heath TG. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat Rev Drug Discov* 2005;**4**(10):825–33.
37. 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**(D1):D1074–82.
38. Wolf, T. *et al.* (2019). Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.
39. Wu, Y. *et al.* (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.
40. Zhang C, Zang T. Cnn-ddi: A novel deep learning method for predicting drug-drug interactions. In: Park, T; Cho, YR; Hu X; Yoo I ; Woo, HG; Wang, J; Facelli,j ; Nam, S; Kang, M (eds) *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. ELECTR NETWORK: IEEE, 2020, 1708–13.
41. 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.