

RESEARCH ARTICLE

A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions

Mei Ma^{1,2*}, Xiujuan Lei^{1*}

1 School of Computer Science, Shaanxi Normal University, Xi'an, China, **2** School of Mathematics and Statistics, Qinghai Normal University, Qinghai, China

* Current Address: School of Computer Science, Shaanxi Normal University, Xi'an, China.

* xjlei@snnu.edu.cn



Abstract

Expressive molecular representation plays critical roles in researching drug design, while effective methods are beneficial to learning molecular representations and solving related problems in drug discovery, especially for drug–drug interactions (DDIs) prediction. Recently, a lot of work has been put forward using graph neural networks (GNNs) to forecast DDIs and learn molecular representations. However, under the current GNNs structure, the majority of approaches learn drug molecular representation from one-dimensional string or two-dimensional molecular graph structure, while the interaction information between chemical substructure remains rarely explored, and it is neglected to identify key substructures that contribute significantly to the DDIs prediction. Therefore, we proposed a dual graph neural network named DGNN-DDI to learn drug molecular features by using molecular structure and interactions. Specifically, we first designed a directed message passing neural network with substructure attention mechanism (SA-DMPNN) to adaptively extract substructures. Second, in order to improve the final features, we separated the drug–drug interactions into pairwise interactions between each drug's unique substructures. Then, the features are adopted to predict interaction probability of a DDI tuple. We evaluated DGNN-DDI on real-world dataset. Compared to state-of-the-art methods, the model improved DDIs prediction performance. We also conducted case study on existing drugs aiming to predict drug combinations that may be effective for the novel coronavirus disease 2019 (COVID-19). Moreover, the visual interpretation results proved that the DGNN-DDI was sensitive to the structure information of drugs and able to detect the key substructures for DDIs. These advantages demonstrated that the proposed method enhanced the performance and interpretation capability of DDI prediction modeling.

OPEN ACCESS

Citation: Ma M, Lei X (2023) A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions. *PLoS Comput Biol* 19(1): e1010812. <https://doi.org/10.1371/journal.pcbi.1010812>

Editor: Nir Ben-Tal, Tel Aviv University, ISRAEL

Received: September 26, 2022

Accepted: December 12, 2022

Published: January 26, 2023

Copyright: © 2023 Ma, Lei. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data and code of DGNN-DDI is freely available at <https://github.com/mamei1016/DGNN-DDI>.

Funding: This work was supported by the National Natural Science Foundation of China (Grant numbers: 62272288 to XJL, 61972451 to XJL, 61902230 to XJL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Drug–drug interactions (DDIs) may cause adverse effects that damage the body. Therefore, it is critical to predict potential drug–drug interactions. The majority of the prediction techniques still rely on the similarity hypothesis for drugs, sometimes neglect the

molecular structure, and fail to include the interaction information between chemical substructure when predicting DDIs. We exploited this idea to develop and confirm the role that molecular structure and interaction information between chemical substructure play in DDIs prediction. The model includes a molecular substructure extraction framework to explain why substructures contribute differently to DDIs prediction, and a co-attention mechanism to explain why the interaction information between chemical substructure can improve DDIs prediction. Compared to state-of-the-art methods, the model improved the performance of DDIs prediction on real-world dataset. Furthermore, it could identify crucial components of treatment combinations that might be efficient against the emerging coronavirus disease 2019 (COVID-19).

This is a *PLOS Computational Biology Methods* paper.

Introduction

With the rapid development of machine learning techniques, many AI technologies have been successfully applied in a variety of tasks for drug discovery, such as drug–drug interactions (DDIs) [1]. One of the main issues in these studies is how to learn expressive representation from molecular structure [2]. Most of the conventional molecular representation are based on hand-crafted features and heavily rely on time-consuming biological experimentations [3].

The most common molecular representation method called simplified molecular input line entry specification (SMILES), is the molecular linear notation that encodes the molecular topology on the basis of chemical rules [4,5], while this line of work suffered from insufficient labeled data for specific molecular tasks. More recently, among the promising deep learning architectures, graph neural networks (GNNs) have gradually emerged as a powerful candidate for modeling molecular data [6–8]. A molecule is naturally treated as a graph based on its geometry, where an atom serves as the node and a chemical bond serves as the edge. Therefore, a molecule is normally mapped to an undirected graph and defined as $G = (V, E)$, where V and E are the sets of all atoms and chemical bonds in the molecule, respectively. Moreover, to better encode the interactions between atoms, a message passing neural network named MPNN was designed to utilize the attributed features of both atoms and edges [9]. It is a general framework for learning node embeddings or learning the entire graph representations. The MPNN used the basic molecular graph topology to obtain structural information through neighborhood feature aggregation and pooling methods [10,11].

DDIs prediction is one of the applications of molecular representation [12–14]. DDIs is referred to as a situation where the pleasant or adverse effects caused by the co-administration of two drugs, which may cause adverse drug events and side effects that damage the body [15,16]. In order to avoid such events, it's urgent to develop computational approaches to detect DDIs [17].

Various machine learning methods have been proposed and have greatly contributed to the DDIs prediction [18–20]. The vast majority of these techniques rely on the drug similarity assumptions, where it is believed that if drugs A and B interact to produce a specific biological effect, then drugs similar to drug A (or drug B) are likely to interact with drug B (or drug A) to produce the same effect. Drugs are, therefore, processed depending on their similarities in

chemical structures; as well as additionally, in other features such as their individual side effects, targets, pathways [21,22].

Recently, many deep learning methods have been developed and have shown encouraging performance in DDIs prediction tasks [23–26]. For instance, Deng et al. [24] proposed a multi-modal deep learning framework combined diverse drug features to predict DDIs. Feng et al. [25] applied deep graph auto-encoder to learn latent drugs representations fed to a deep feedforward neural network for DDIs prediction. Liu et al. [27] introduced a deep attention neural network framework for drug–drug interaction prediction, which can effectively integrate multiple drug features. For adverse drug–drug interaction (ADDI), Zhu et al. [28] employed eight attributes and developed a discriminative learning algorithm to learn attribute representations of each adverse drug pair for exploiting their consensus and complementary information in multi-attribute ADDI prediction. And then they designed three dependence guided terms among molecular structure, side effect and ADDI to guide feature selection and put forward a discriminative feature selection model DGDFS for ADDI prediction [29]. Because combination therapy can boost efficacy and reduce toxicity, recent approaches have used deep learning to identify synergistic drug combinations for the new coronavirus disease 2019 (COVID-19) [30–32]. Jin et al. [31] presented a new deep learning architecture Comb-oNet for predicting synergistic drug combinations for COVID-19. Howell et al. [33] developed a computational model of SARS-CoV-2-host interactions used to predict effective drug combinations.

Although these methods achieved inspiring results, there are still mostly unexplored in DDIs prediction tasks especially as far as feature extraction from the raw representations (i.e., chemical structures) of drugs are concerned. Most of the existing methods predict DDIs relying on the similarity assumption of drugs or on manually engineered features [34,35]. However, molecular structure-based methods regard drugs as independent entities, and predict DDIs only by relying on drug pairs. This is no need for external biomedical knowledge. It has been proven that DDIs usually depend only on a few substructures as a whole [36,37]. SSI-DDI [34] and GMPNN-CS [35], two recent methods that both leveraged the powerful feature extraction ability of deep learning, work directly on raw molecular chemical structures of drugs using GNNs. SSI-DDI used graph attention (GAT) layers [38] to learn a comprehensive feature representation of a drug from substructures, while GMPNN-CS introduced gated message passing neural network which learns chemical substructures with different sizes and shapes from the molecular graph representations of drugs for DDIs prediction. However, the gates are computed before the message passing, which means that they did not fully exploit the molecular structure information.

In this study, we proposed a DDIs prediction approach called DGNN-DDI that uses dual GNN to extract drug feature representation while taking into account drug substructure and the interaction information between chemical substructure. To extract the molecular substructures features, we first constructed a directed message passing neural network with substructure attention mechanism (SA-DMPNN) by fully considering the flexible-sized and irregular-shaped of drug molecules substructures. Second, we used co-attention mechanism [39] to determine the importance weight by learning the interaction scores between the substructures features of the two drugs. After that, we concatenated the weighted substructures features for each drug to obtain the final feature, which was used to predict the potential interaction probability of the existing drugs and drugs. We evaluated our model using real-world dataset, and the experiments demonstrated that our DGNN-DDI is superior in predicting the potential DDIs. The method was applied to predict anti-COVID-19 drug combinations. The main contributions of this work are summarized as the following:

1. DGNN-DDI takes into account the key molecular substructure feature of the drugs, which is conducive to learning high-quality features.
2. DGNN-DDI leverages the interaction information between chemical substructure to identify substructures with interactions, which can enhance the final feature of drug and also contribute to improving the prediction accuracy of DDIs.
3. The method is applied to predict anti-COVID-19 drug combinations using real-world dataset.

Results and discussion

Dataset

We evaluated the model performance in DrugBank [40], which is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information. The dataset contains 1706 drugs and 191808 DDIs tuples classified into 86 interaction types, which describes how one drug affects the metabolism of another one. Each drug is associated with their SMILES and we converted it into molecular graph using RDKit. In the dataset, each drug pair is only associated with a single type of interaction.

Experiment setting

In our study, we split the dataset randomly into training (60%), validation (20%), and test (20%) for the DDIs prediction task. The message passing steps T was searched from {1, 2, 3, 4, 5}, and the Multi-GNN layers L was searched from {1, 2, 3, 4, 5}. Because of the model was dual, T and L was determined to be 3 through subsequent parameter analysis. After parameter analysis, we considered the following hyper-parameter settings. Dimension of h_i in Eq 12 was searched from {32, 64, 128}. The model was trained on mini-batches DDI tuples tuned from {128, 256, 512} using the Adam optimizer [41] with a learning rate lr tuned from {1e-2, 1e-3, 1e-4}. Additionally, an exponentially decaying scheduler of $0.96e$ (where e is the current epoch) was set on the learning rate. We discovered that the best performance was obtained when the $T = L = 3$, $h_i \in R^{64}$, $lr = 1e - 4$ and batch size was 256. The number of epochs was 50. To avoid overfitting, we applied a weight decay of 5×10^{-4} for all methods. Like most of the literatures [42,43], we trained the comparison methods with the same parameter settings as DGNN-DDI, including learning rate, optimizer, batch size, weight decay, hidden dimension and number of epochs. But the message passing steps T or layers L was taken from original manuscript. The performance metrics included accuracy (ACC), area under the curve (AUC), F1-score (F1), precision (Prec), recall (Rec) and area under the precision and recall curve (AUPR).

Comparative analysis with other methods

On the DrugBank dataset, we compared the proposed model with cutting-edge approaches to verify its efficacy. Only chemical structural information is taken into account by these approaches as an input, and combined drug-drug information is integrated in some way during the learning process.

- SA-DDI [44]: a GNN that used a message-passing neural network and a substructure-substructure interaction module to learn thorough and useful features. SA-DDI extracted features with message passing step $T = 10$ for DDIs prediction.

Table 1. Comparison results in % of the proposed DGNN-DDI and baselines on the dataset.

	ACC	AUC	F1	Prec	Rec	AUPR
GAT-DDI	0.7894	0.8653	0.8045	0.7676	0.8682	0.8398
GMPNN-CS	0.9485	0.9834	0.9495	0.9346	0.9725	0.9785
SA-DDI	0.9565	0.9868	0.9573	0.9472	0.9746	0.9834
SSI-DDI	0.8965	0.9541	0.8993	0.8763	0.9321	0.9420
DGNN-DDI	0.9609	0.9894	0.9616	0.9472	0.9788	0.9863

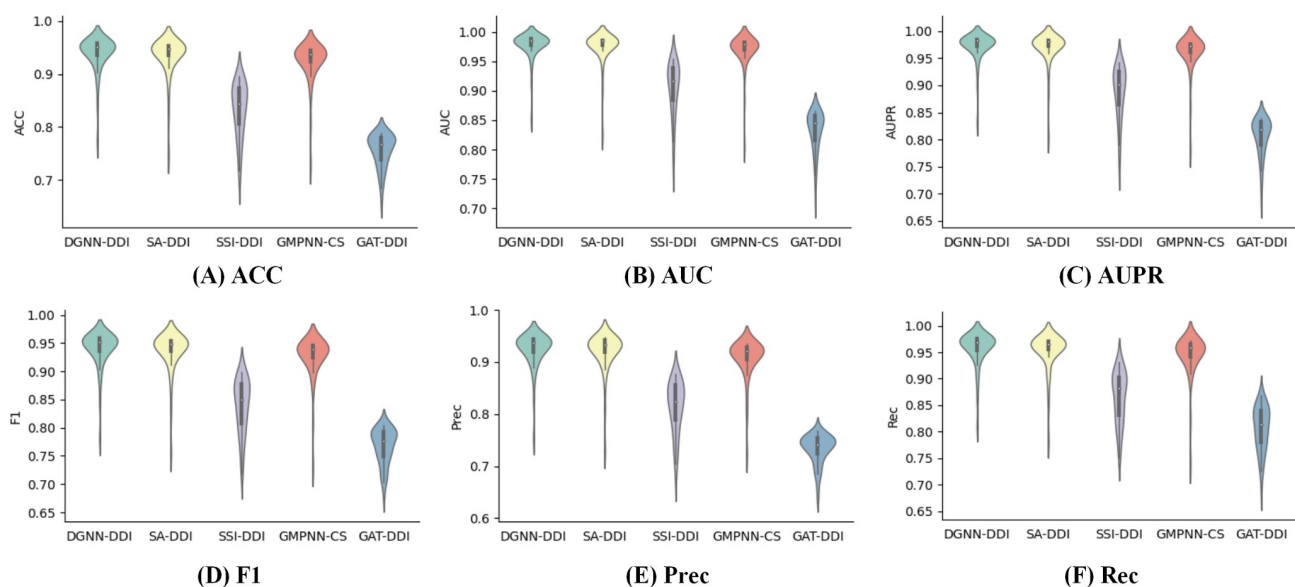
<https://doi.org/10.1371/journal.pcbi.1010812.t001>

- SSI-DDI [34]: considered each node hidden features as sub-structures and then computes interactions between these substructures to determine the final DDI prediction. SSI-DDI stacked $L = 4$ layers of graph attention (GAT) for DDIs prediction.
- GMPNN-CS [35]: a GNN architecture that introduced gates message passing mechanism to control the flow of message passing of GNN. GMPNN-CS learned chemical substructures with message passing step $T = 10$ for DDIs prediction.
- GAT-DDI [35]: replaced the GNN architecture in GMPNN-CS with GAT for drug representations, which are directly used for DDI prediction. GAT-DDI learned chemical substructures with message passing step $T = 10$ for DDIs prediction.

Table 1 summarizes metric scores of all prediction models, and results demonstrate that DGNN-DDI outperforms other methods on all metric scores for the DrugBank dataset, which show the effectiveness of the proposed DGNN-DDI for DDI prediction.

To further analyze the performances of prediction models, we used Fig 1A–1F to display all metric scores of different methods. These violin plots clearly show that DGNN-DDI produces better performances for these metrics than the competing methods.

Moreover, we conducted a statistical analysis to test the differences between DGNN-DDI and other state-of-the-art methods. We conducted statistical significance tests by using predicted scores, and paired t -test results are demonstrated in Fig 2. For the DDI prediction, the

**Fig 1. Violin plots displaying metric scores of all models.**

<https://doi.org/10.1371/journal.pcbi.1010812.g001>

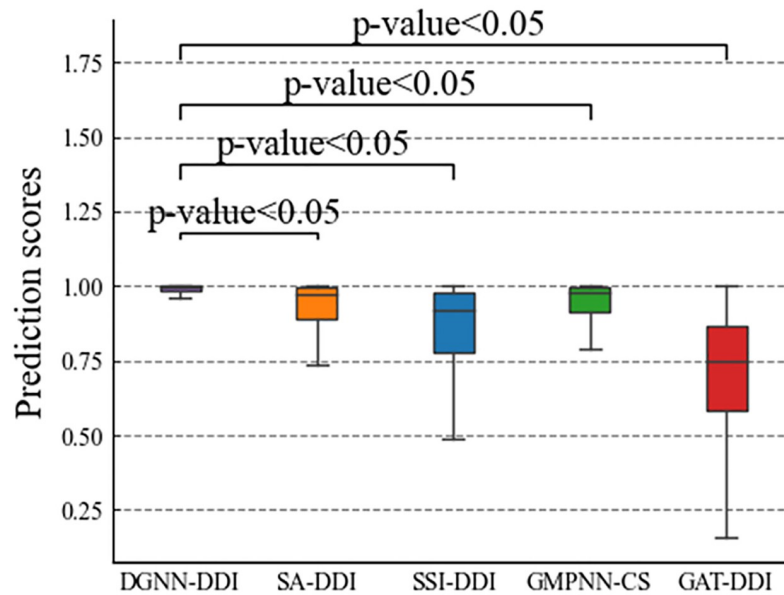


Fig 2. The significant difference between DGNN-DDI and other models in terms of predicted scores.

<https://doi.org/10.1371/journal.pcbi.1010812.g002>

proposed method DGNN-DDI significantly (p -value < 0.05) improves the performances compared to other state-of-the-art methods.

To further demonstrate the superiority of GNN-DDI, we set $T = 3$ or $L = 3$ for all comparison methods, which is consistent with DGNN-DDI. Fig 3 displays the ROC curves (receiver operating characteristic curves) and P-R curves (precision-recall curves) of all models. Clearly, DGNN-DDI performs best among all methods, demonstrating once more its strong potential for DDIs prediction.

Parameter analysis

The parameters T and L have a significant impact on the extraction of substructures with variable sizes and forms during the processing of molecular features learning. We tested all 25 combinations of steps T and layers L , as shown in Fig 4. The distribution of all metric scores under all 25 combinations are shown in Fig 4A–4E, respectively. It can be seen that when

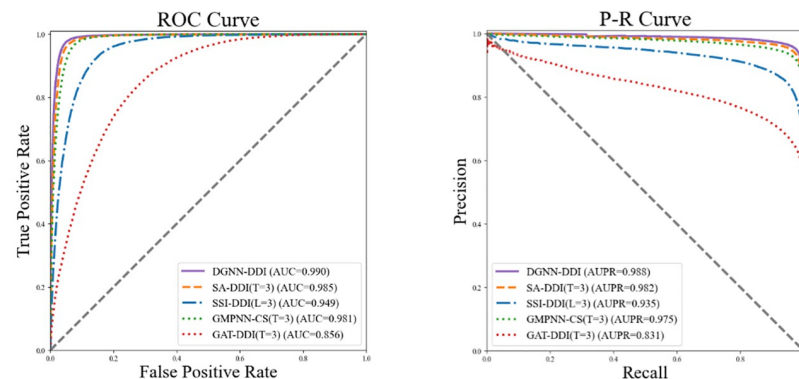


Fig 3. The ROC curves and P-R curves of all models, where $T = 3$ or $L = 3$ for all models.

<https://doi.org/10.1371/journal.pcbi.1010812.g003>

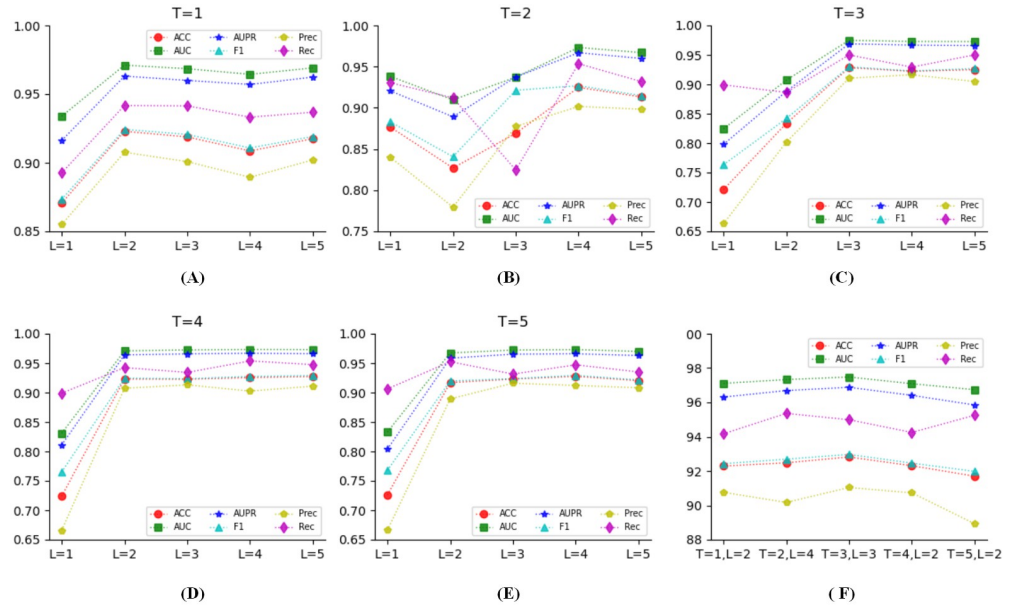


Fig 4. The effects of steps T and layers L .

<https://doi.org/10.1371/journal.pcbi.1010812.g004>

$T = 1, L = 2; T = 2, L = 4; T = L = 3; T = 4, L = 2; T = 5, L = 2$ all metric scores are superior to other combinations. We further compared these five combinations, as shown in Fig 4F, when $T = L = 3$ shows a better performance than other combinations.

The size of the batch is particularly significant since the DGNN-DDI is sampled and trained in batches. It will be challenging to converge if the batch size is too small. While if the batch is too large, a large amount of computation is required. As shown in Fig 5A, we investigated the

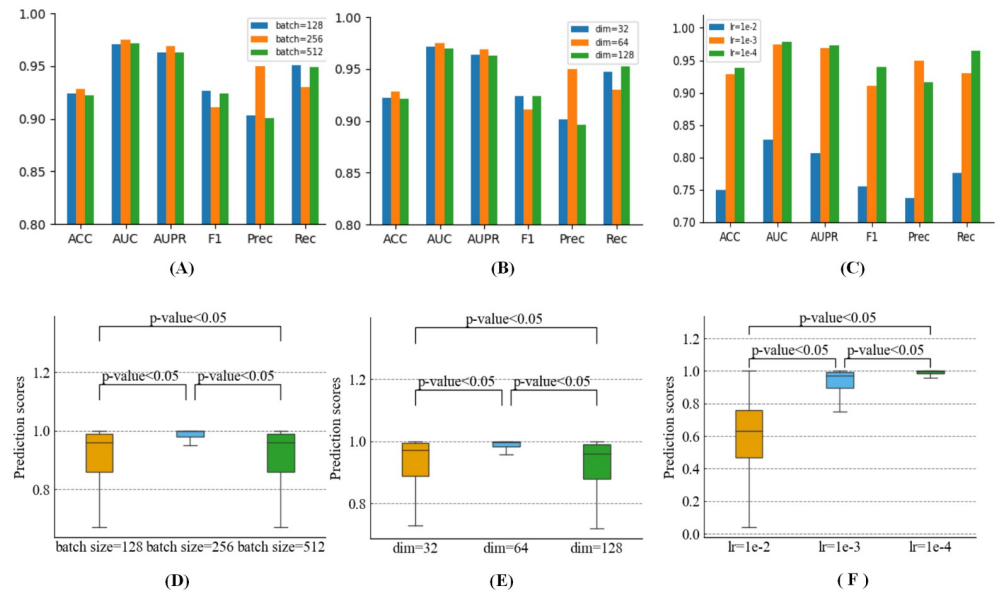


Fig 5. Parametric analysis. (A) Effects of batch size. (B) Effect of hidden dimension. (C) Effects of learning rate. (D) The significance analysis of batch size in terms of predicted scores. (E) The significance analysis of hidden dimension in terms of predicted scores. (F) The significance analysis of learning rate in terms of predicted scores.

<https://doi.org/10.1371/journal.pcbi.1010812.g005>

Table 2. Investigating the contributions of substructure-attention mechanism and co-attention layer.

	ACC	AUC	F1	Prec	Rec	AUPR
DGNN-DDI_no_SA	0.9072	0.9482	0.8979	0.8833	0.9155	0.9248
DGNN-DDI_no_CA	0.8882	0.9461	0.8915	0.8755	0.9413	0.9313
DGNNDDI_no_SA_CA	0.8858	0.9445	0.8898	0.8785	0.9308	0.9273
DGNN-DDI	0.9609	0.9894	0.9616	0.9472	0.9788	0.9863

<https://doi.org/10.1371/journal.pcbi.1010812.t002>

impact of various batch sizes on the methodology. The method has the best performance when the batch size is equal to 256. As demonstrated in Fig 5B and 5C, we also looked into how hidden dimensions and learning rates affected the performance of the model. Moreover, we conducted a significance analysis to test the differences on different values of batch size, learning rate and hidden dimension, respectively. Using predicted score, we conducted statistical significance tests, the paired *t*-test results are demonstrated in Fig 5D, 5E and 5F. For the three parameters, when the $h_i \in R^{64}$, $lr = 1e - 4$ and batch size is equal to 256, DGNN-DDI significantly (p -value < 0.05) improves the performances compared to other values.

Ablation study

In our designs, the successful construction of the DGNN-DDI highly relies on D-MPNN with substructure attention mechanism (SA-DMPNN) and interaction-aware substructure extraction (Multi-GNN). The substructure attention is used to extract substructures with arbitrary size and shape. The relevance of substructure interactions with co-attention is expected to enhance the model performance to the final DDI prediction. We conducted experiments by removing the substructure-attention mechanism (SA) and/or co-attention layer (CA). Table 2 summarizes the contributions of SA and CA. The results show that both SA and CA are necessary for DGNN-DDI.

Fig 6A–6C also shows that the model performs poorly without SA, CA, or SA and CA, showing the necessity of SA and CA. Furthermore, we observed that the results decrease

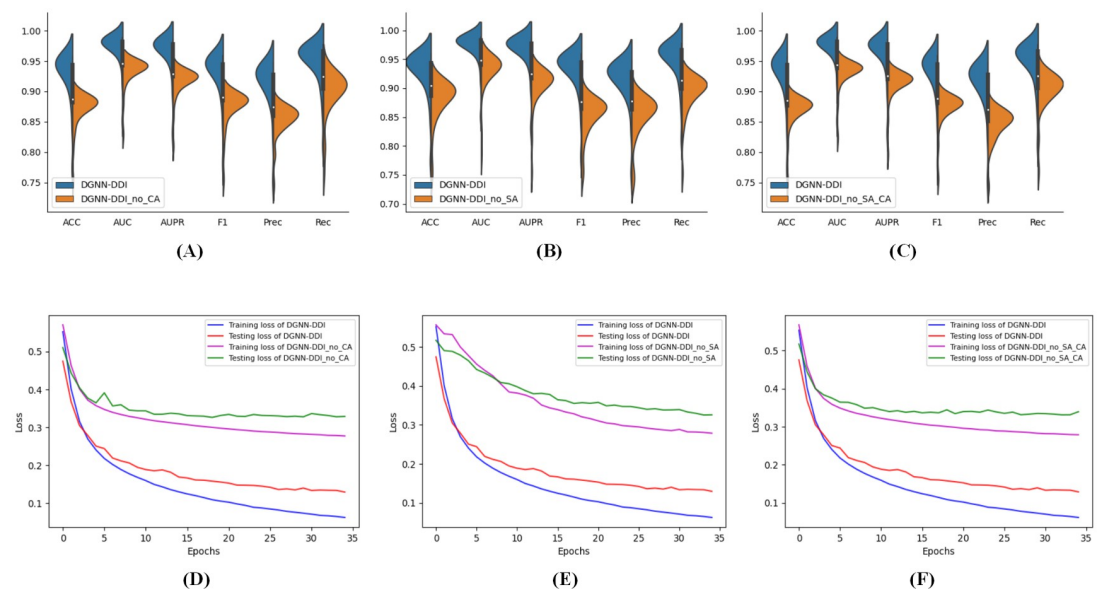


Fig 6. Analysis of the substructure attention mechanism (SA) and co-attention layer (CA). (A)-(C) The metric scores of DGNN-DDI and without SA and/or CA. (D)-(F) The training and testing losses for DGNN-DDI and without SA and/or CA.

<https://doi.org/10.1371/journal.pcbi.1010812.g006>

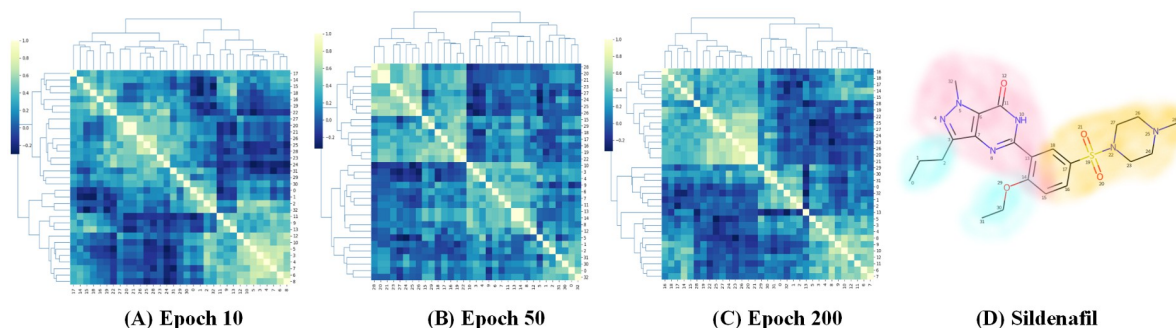


Fig 7. Heat maps of the atom similarity matrix for drug sildenafil.

<https://doi.org/10.1371/journal.pcbi.1010812.g007>

greatly when applying either SA or CA. However, the improvement of using both is larger than the only one, highlighting the effectiveness of DGNN-DDI. Additionally, as demonstrated in Fig 6D–6F, SA and CA can expedite training while also enhancing generalization ability.

Visual explanations for DGNN-DDI

We conducted visual explanation-related experiments to rationalize the DGNN-DDI. To investigate how the atom hidden vectors evolved during the learning process, we obtained the similarity coefficient between atom pairs by measuring the Pearson correlation coefficient for those hidden vectors. We chose the hidden vectors after the last iteration (i.e., h_i in Eq 12). Figs 7 and 8 give two drugs with their atom similarity matrices during the learning process. The cluster heat maps show some degree of chaos at the beginning and then clearly group into clusters during the learning process where the corresponding substructures for clusters are highlighted in the drugs. Taking Fig 7 as an example, we found that the atoms in sildenafil at epoch 50 approximately separate into three clusters. This finding is in accordance with our intuition regarding the sildenafil structure.

These results suggest that the DGNN-DDI can capture the structure information of a molecule.

Furthermore, we investigated the performances of DGNN-DDI for each DDI type and calculated the metric scores for interaction types independently by using predicted scores and real labels. The performance metrics for each DDI type are shown in Fig 9. Among 86 DDI types, DGNN-DDI achieves the highest AUC scores and the highest AUPR scores in 80 DDI

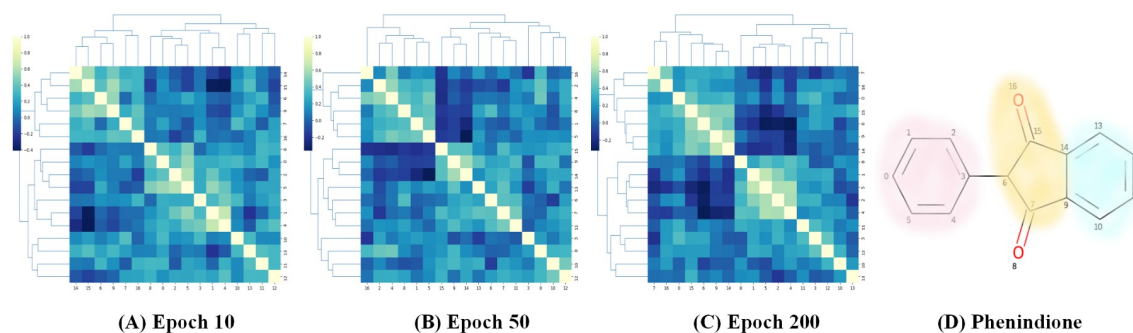


Fig 8. Heat maps of the atom similarity matrix for drug phenindione.

<https://doi.org/10.1371/journal.pcbi.1010812.g008>

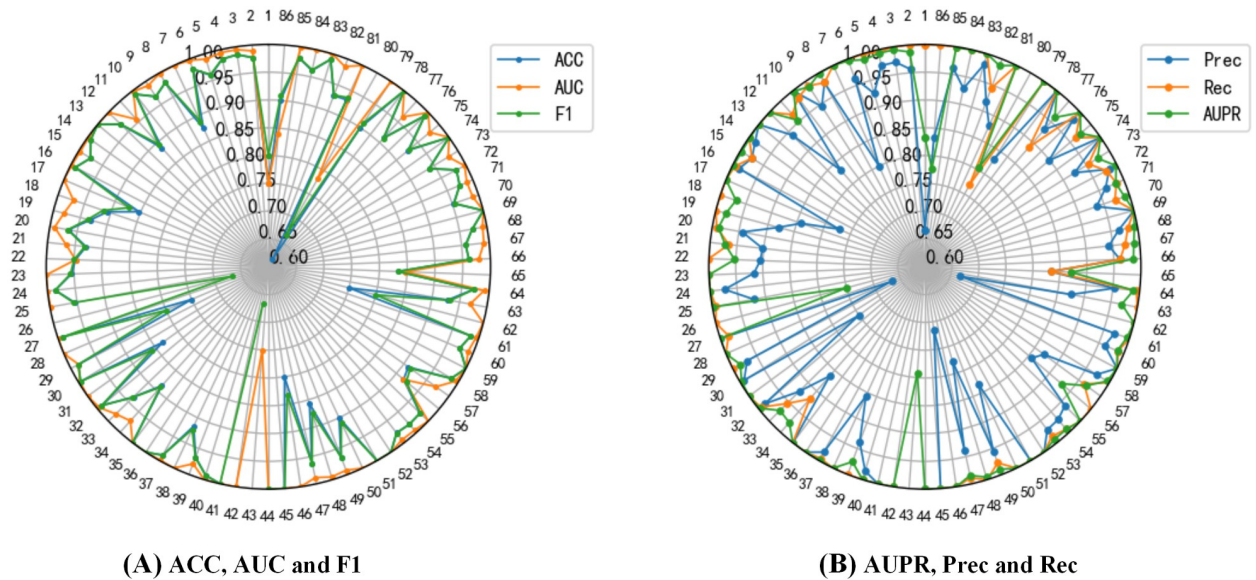


Fig 9. Performance for each DDI type.

<https://doi.org/10.1371/journal.pcbi.1010812.g009>

types (more than 85%). In general, Fig 9 demonstrates that DGNN-DDI produces good performance in most of DDI types.

Case study

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 2019 has triggered an ongoing global pandemic of the severe pneumonia-like disease coronavirus disease 2019 (COVID-19) [30]. However, the research and development of traditional medicines for the new coronavirus are very expensive in terms of time, manpower, and funds. We hypothesized that combining drugs with independent mechanisms of action could result in synergy against SARS-CoV-2, thus generating better antiviral efficacy [45]. We prioritized 73 combinations of 32 drugs with potential activity against SARS-CoV-2 and then tested them [46]. Twelve synergistic combination drugs were identified. To further investigate which substructure among the 12 synergistic instances contributes most significantly to medication synergistic combos, we visualized the most crucial substructures for combination drugs. Specifically, we first determine the indices (h, t) of the highest pairwise interaction score from Eq 1:

$$(h, t) = \operatorname{argmax}_{i,j} (\hat{g}_x^{(i)} M_r \hat{g}_y^{(j)}) \quad i, j = 1, \dots, L \quad (1)$$

This can be extended to top k pairwise interaction scores for further analysis. To keep the study simple, we used only the highest one ($k = 1$). (h, t) tells us that the substructures of concern are from the h -th Multi-GNN layer for d_x and t -th Multi-GNN layer for d_y , which were primarily responsible for the DDI outcome. We chose three atoms with the largest substructures attention, which were described by Eq 13, as the center of the most vital substructures.

Fig 10 show the results of this case study. Contributions of substructures are presented as a heat map (map with green fill) of the molecular graph. Each row contains two pair of drugs, for each pair of drugs, the indices h, t means the radius of a substructures. Therefore, in the heat map, each substructure contribution is shown mainly concentrated around its center. We

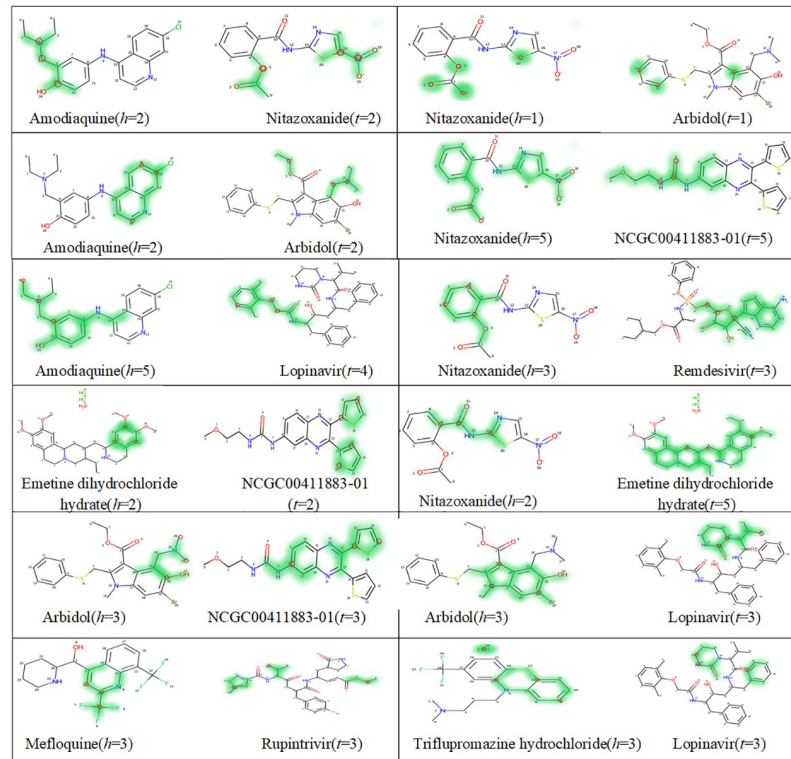


Fig 10. The key substructures contributing to the SARS-CoV-2 drug combinations. The center of the most important substructure and its receptive field are shown as red circle and green colors respectively.

<https://doi.org/10.1371/journal.pcbi.1010812.g010>

can see that the drug nitazoxanide with remdesivir, amodiaquine, emetine dihydrochloride hydrate, arbidol and NCGC00411883-01 exhibiting significant synergy against SARS-CoV-2, which is consistent with the result of a previous study [46]. When synergistic with different drugs, the key substructures of drug nitazoxanide are basically the unified, and cresyl acetate (“CC(=O)Oc1cccc1C”) or part of it can be found in all of them. However, the key substructures of drug arbidol are vary wildly. We hypothesized that this variety might be caused by various chemical substructures that function in various ways in the medication combinations used to treat SARS-CoV-2, which was consistent with the notion put out that substructures with various weights affect DDI prediction. Overall, these results highlight the utility of drug repurposing and preclinical testing of drug combinations for discovering potential therapies to treat SARS-CoV-2. Additionally, Fig 11 displays a map of pharmacological combinations with results of their synergism.

Conclusion

This paper presented a novel molecular structure-based deep learning model DGNN-DDI for predicting DDIs between a pair of drugs. The DGNN-DDI used the substructure attention and co-attention mechanism to obtain the substructure with irregular size and shape, and enhance the representation capability for the model. On DrugBank dataset, we contrasted the suggested model with cutting-edge approaches to confirm its superiority. Moreover, we visualized the atom similarity of certain molecules. Finally, we showed the key substructures for the SARS-CoV-2 drug combinations as a case study. The visual interpretation results showed that

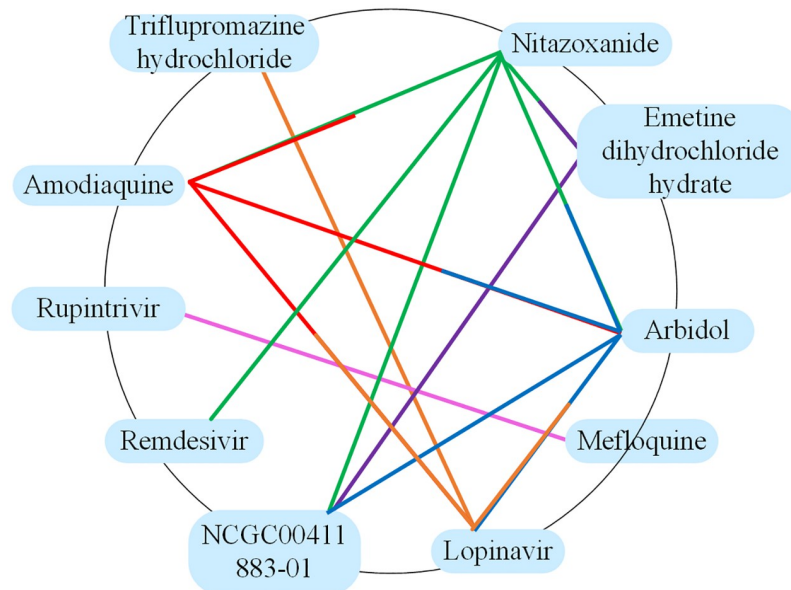


Fig 11. Drug combination to treat SARS-CoV-2.

<https://doi.org/10.1371/journal.pcbi.1010812.g011>

the DGNN-DDI was sensitive to the structure information of drugs and able to detect the key substructures for DDIs. These advantages demonstrated that the proposed method improved the performance and interpretation capability of DDIs prediction modeling.

Materials and method

This section gives the technical details of the DGNN-DDI. First, we defined the problem that has to be resolved. Then, we presented the input representation and all involved computational steps of our method. The overall framework is shown in Fig 12.

Problem formulation

The purpose of the DDIs prediction task is to develop an advanced model that takes two drugs and an interaction type as input and generates an output indicating whether there exists an interaction between them. Formally, given a dataset of DDIs $M = \{(d_x, d_y, r)\}_{i=1}^N$, where d_x, d_y is taken from the drugs set D , r denotes the interaction type between two drugs, taken from the interaction types set I . Our major objective is to find a model $f: D \times D \times I \rightarrow \{0, 1\}$, which predicts the probability that this type of interaction will occur between the two drugs.

Input representation

The input of the model is a DDI tuple (d_x, d_y, r) . Drugs d_x and d_y are both represented by SMILES strings. We preprocessed the SMILES into graph using RDKit [47] as shown in Fig 13A, where the nodes representing atoms, while edges representing the bonds between the atoms. Therefore, a drug is typically defined as a molecular graph $G = (V, E)$, where $V = \{v_i\}_{i=1}^n$ is the set of nodes, and $E = \{(v_i, v_j)_s\}_{s=1}^m$ is the set of edges. Each node v_i has a corresponding feature vector $x_i \in R^d$. Similarly, while each edge $e_{ij} = (v_i, v_j)$ has a feature vector $x_{ij} \in R^d$. The features used for atoms and bonds are given in the Table 3.

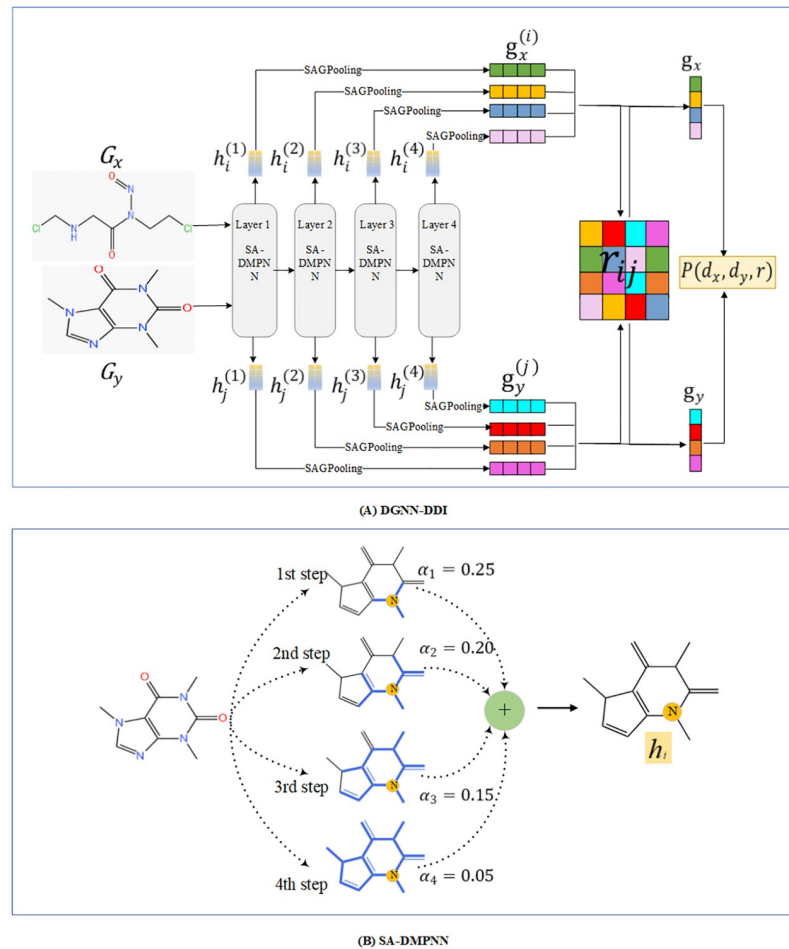


Fig 12. The overview of proposed DGNN-DDI for DDI prediction. (A) The workflow of DGNN-DDI. (B) The SA-DMPNN updates the node-level features with T steps where T is 4 in this example.

<https://doi.org/10.1371/journal.pcbi.1010812.g012>

Graph neural network

When a graph is represented as $G = (V, E)$, a GNN maps a graph G to a vector $h_G \in R^d$ usually with a message passing phase and readout phase. As shown in Fig 13B and 13C, the message passing phase updates node-level features by aggregating messages from their neighbor nodes in G , and the readout phase generates a graph-level feature vector by aggregating all the node-level features, which is used to predict the label of the graph.

Message passing phase. Given a graph G , we denoted the feature of node v at step t as $x_v^{(t)} \in R^d$. We then updated $x_v^{(t)}$ into $x_v^{(t+1)} \in R^d$ using the following graph convolutional layer [9]:

$$m_v^{(t+1)} = \sum_{u \in N(v)} M_t(x_v^{(t)}, x_u^{(t)}, e_{uv}) \tag{2}$$

$$x_v^{(t+1)} = U_t(x_v^{(t)}, m_v^{(t+1)}) \tag{3}$$

where $N(v)$ denotes the neighbors of v in graph G . M_t and U_t are the message functions and node update functions, respectively.

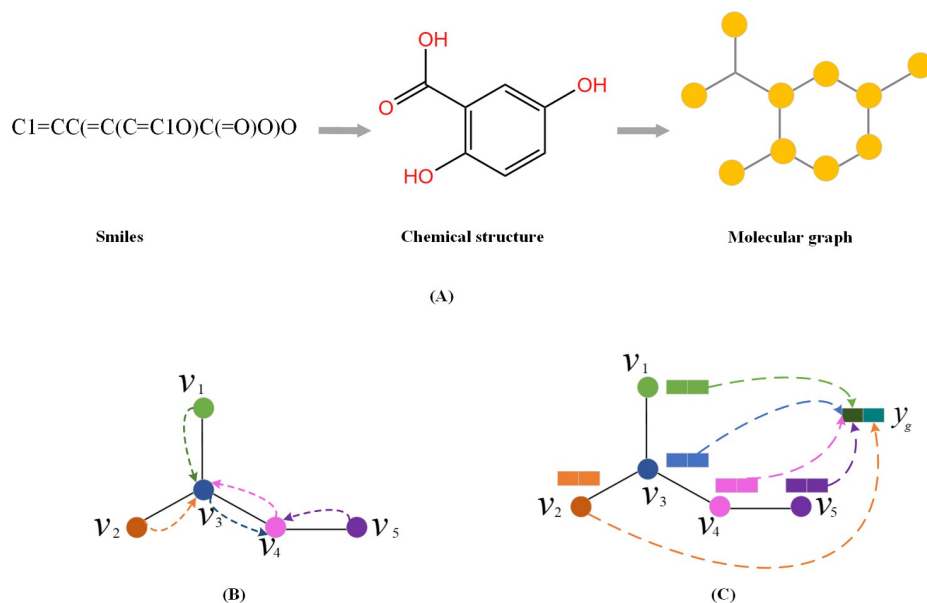


Fig 13. Molecule representation and graph embedding. (A) Preprocessed the smiles into graph. (B) Graph message passing phase. (C) Graph readout phase.

<https://doi.org/10.1371/journal.pcbi.1010812.g013>

Readout phase. To obtain a graph-level feature h_G , readout operation integrates all the node features among the graph G is given in Eq 4:

$$h_G = R(x_v^T : v \in G) \quad (4)$$

where R is readout function, and T is the final step.

So far, the GNN is learned in a standard manner, which has third shortcomings for DDIs prediction. First, the GNN extracts fixed-size substructures with a predetermined number of layers, it is insufficient to capture the global structure of the molecules. As shown in Fig 14A, a GNN with two layers is unable to know whether the ring exists in the molecule. Therefore, to capture the structures make up of k -hop neighbors, the k graph convolutional layers should be stacked. Second, a well-constructed GNN should be able to preserve the local structure of a compound. As shown in Fig 14B, the methyl carboxylate moiety is crucial for methyl

Table 3. Atom and bond features.

atom feature	Description	Size
atom type	[B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, I, At, meta]	16 (one-hot)
degree	number of covalent bonds [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	11 (one-hot)
hybridization	[sp, sp2, sp3, sp3d, sp3d2]	5 (one-hot)
implicit valence	implicit valence of the atom [0, 1, 2, 3, 4, 5, 6,]	7 (one-hot)
radical electrons	number of radical electrons	1(integer)
formal charge	formal charge of the atom	1 (integer)
aromatic	whether the atom is part of an aromatic system	1 (integer)
bond feature	Description	Size
bond type	[single, double, triple, aromatic]	4 (one-hot)
conjugated	whether the bond is part of a conjugated system	1 (integer)
ring	whether the bond is part of a ring	1 (integer)

<https://doi.org/10.1371/journal.pcbi.1010812.t003>

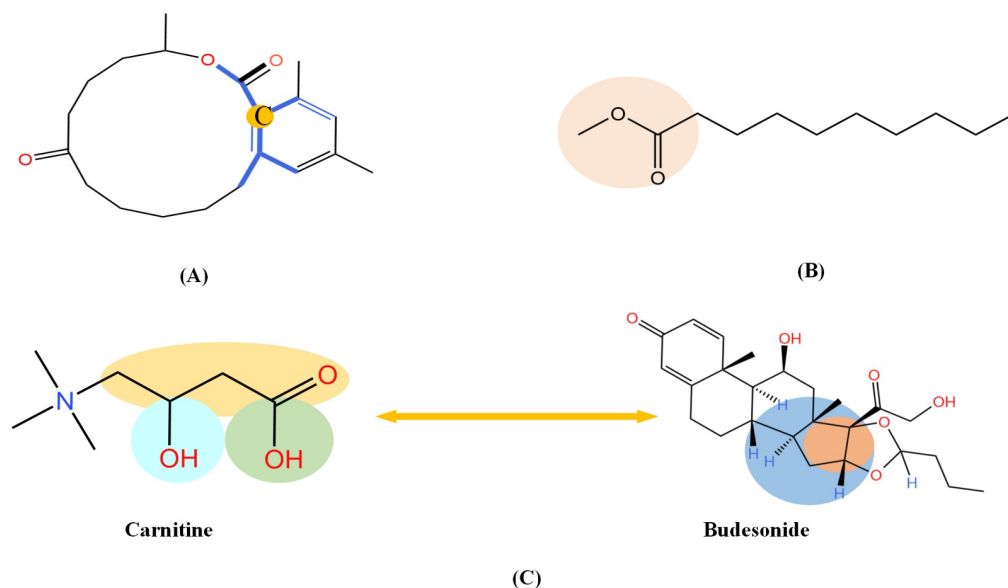


Fig 14. Both structure information and DDIs are important for GNN. (A) The sight of GNNs in the second layer is shown in blue as we take the carbon with orange as the center. In this example, a GNN with two layers fails to identify the ring structure of zearalenone. (B) The GNN should preserve local structure information (orange ellipse) (C) The interaction type of ‘blood calcium increased’ between drug pair ‘Carnitine’ and ‘Budesonide’ is caused by their partial significant substructures (elliptical parts).

<https://doi.org/10.1371/journal.pcbi.1010812.g014>

decanoate and the GNN should distinguish it from the less essential substituents in order to make a reasonable inference. Concretely, it is necessary to apply the attention mechanism to the key substructures. Third, DDIs usually depend only on a few substructures of the whole chemical structures. As depicted in Fig 14C, the interaction type of ‘blood calcium increased’ between drug pair ‘Carnitine’ and ‘Budesonide’ is caused by their partial important substructures [48]. It is feasible to break down DDIs into substructure–substructure interactions. The following, we adopted directed message passing neural network with substructure attention mechanism (SA-DMPNN) and interaction-aware substructure extraction to solve these three limitations.

Directed message passing neural network with substructure attention mechanism

The idea of substructure attention is to extract substructures with arbitrary sizes and shapes and assign each substructure a unique score. We used the D-MPNN [2] with substructure attention mechanism (SA-DMPNN) for molecule substructures extraction. The process is shown in Fig 12B. During the t -th step, the SA-DMPNN extracts substructures with a radius of t .

In the SA-DMPNN, each node will receive a message from the bond-level hidden feature. For each node v_i , the hidden feature at step t is $h_i^{(t)} \in R^d$, where $h_i^{(0)} = x_i$, we used $h_{ij}^{(t)}$ to represent a bond-level hidden feature with each bond $e_{i \rightarrow j}$. We initialized the bond-level hidden features as

$$h_{ij}^{(0)} = W_i x_i + W_j x_j + W_{ij} x_{ij} \quad (5)$$

where $W_i \in R^{h \times d}$, $W_j \in R^{h \times d}$, and $W_{ij} \in R^{h \times d}$ are learnable weight matrices.

At step t , we computed the bond-level neighborhood features for each node before utilizing a substructure-aware global pooling, then we obtained its bond-level graph representation $g^{(t)}$. The corresponding calculation equations are thus:

$$m_i^{(t)} = \sum_{v_j \in N(v_i)} \beta_{ij} h_{ji}^{(t)} \quad (6)$$

$$g^{(t)} = \sum_{i=1}^n m_i^{(t)} \quad (7)$$

where SAGPooling [49] can be used to calculate β_{ij} . Given a graph with bond-level feature matrix X and adjacency matrix A in which the nonzero position indicates that two bonds share a common node, SAGPooling computes the importance vector β_{ij} as follows:

$$\beta_{ij} = \text{softmax}(\text{GNN}(A, X)) \quad (8)$$

GNN is an arbitrary GNN layer for calculating projection scores. For each bond-level graph representation $g^{(t)}$, the substructure attention score can be computed as follows:

$$e^{(t)} = w^{(t)} \odot \tanh(Wg^{(t)} + b) \quad (9)$$

where \odot represents dot product, $w^{(t)}$ is a weight vector for step t . In order to make the coefficients of different steps easy to compare, we normalized $e^{(t)}$ by using the softmax function:

$$\alpha^{(t)} = \frac{\exp(e^{(t)})}{\sum_{k=1}^T \exp(e^{(k)})} \quad (10)$$

where each $\alpha^{(t)} \in R^1$ indicates the importance of the substructures with a radius of t . After updating bond-level hidden features T steps, we returned the final representation of h_{ij} by the weighted sum of bond-level hidden features across all steps according to the following:

$$h_{ji} = \sum_{t=1}^T \alpha^{(t)} h_{ji}^{(t)} \quad (11)$$

The substructure attention mechanism will make it possible that not all the nodes in a substructure are considered equally, refining even further the type of substructures being learned.

Finally, we returned to the node-level features by aggregating the incoming bond-level features as follows:

$$h_i = f(x_i + \sum_{v_j \in N(v_i)} h_{ji}) \quad (12)$$

where f is a multilayer perceptron, and h_i contains the substructure information from different receptive fields centering at i -th atom.

Interaction-aware substructure extraction

As mentioned above, shallow convolutional layers cannot capture global structure of the molecules, in order to overcome this limitation, we stacked multiple SA-DMPNN blocks to obtain substructure-level graph representation. The stacking structure is referred to as Multi-GNN for the sake of simplicity in descriptions. The process is shown in Fig 12A.

For a given drug d_x , suppose we have obtained the node-level features for each node in molecular graph G_x from the SA-DMPNN. At each Multi-GNN layer l , the features of each node are denoted as $h_i^{(l)}$, then the representation of the substructure $g_x^{(l)} \in G_x$ is therefore given by the Eq 13, which is represented by aggregating the node features $h_i^{(l)}$, each one weighted by a learnable coefficient β_i , which can be interpreted as its importance. The β_i can be obtained by

the SAGPooling.

$$g_x^{(l)} = \sum_{i=1}^n \beta_i h_i^{(l)} \tag{13}$$

After obtaining all the substructure information $g_x^{(l)}$ and $g_y^{(l)}$ of the input drugs d_x and d_y from all the Multi-GNN layers, we employed a co-attention mechanism to account for the importance γ_{ij} of each pairwise interaction between the substructures of G_x and G_y , which is given by:

$$\gamma_{ij} = b^T \tanh(W_x g_x^{(i)} + W_y g_y^{(j)}) \quad i, j = 1, \dots, L \tag{14}$$

where b is a learnable weight vector, W_x and W_y are learnable weight matrices. To prevent situations where similar substructures are given high ratings, we applied various weight matrices.

Furthermore, we updated the substructural features $g_x^{(i)}, g_y^{(j)}$ with γ_{ij} , respectively, which is formulated as follows:

$$\hat{g}_x^{(i)} = \sum_{j=1}^L \gamma_{ij} g_x^{(i)} \quad i = 1, \dots, L \tag{15}$$

$$\hat{g}_y^{(j)} = \sum_{i=1}^L \gamma_{ij} g_y^{(j)} \quad j = 1, \dots, L \tag{16}$$

Finally, the graph-level representation of d_x can be computed by the following:

$$g_x = \sum_{l=1}^L \hat{g}_x^{(l)} \tag{17}$$

The graph-level representation of d_y (i.e., g_y) can be calculated by using computational steps similar to that described in Eq 17. As opposed to the global pooling that considers every substructure equally important, we utilized the interaction information to enhance structure information of d_x and d_y by assigning cross-substructure interaction scores. The overall computational steps for graph-level representation of d_x and d_y are depicted in Fig 15.

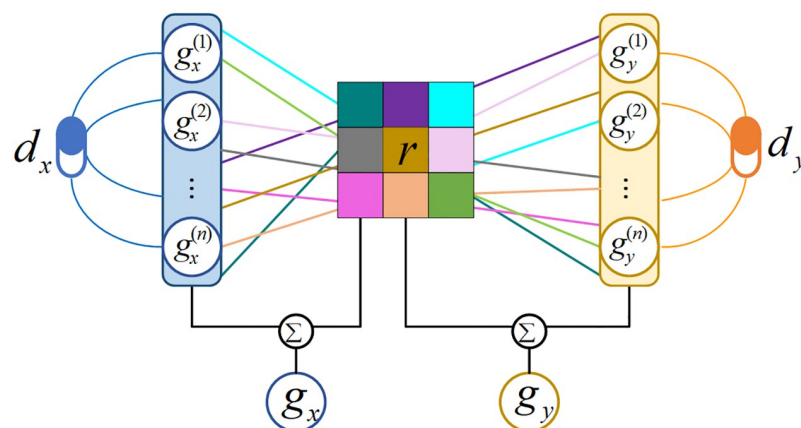


Fig 15. The overall computational steps for graph-level representation of d_x and d_y .

<https://doi.org/10.1371/journal.pcbi.1010812.g015>

Drug-drug interaction prediction

Given a DDI tuple (d_x, d_y, r) , the DDIs prediction can be expressed as the join probability of the tuple:

$$P(d_x, d_y, r) = \sigma(g_x^T M_r g_y) \quad (18)$$

where σ is the sigmoid function, and M_r is the learnable matrix representation of interaction type r . The learning process of the model can be achieved by minimizing the cross-entropy loss function [50], which is given as follows:

$$Loss = -\frac{1}{M} \sum_{i=1}^M y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (19)$$

where $y_i = 1$ indicates that an interaction exists between d_x and d_y , and vice versa; and p_i is the predictive interaction probability of a DDI tuple is computed by using Eq 18.

Author Contributions

Conceptualization: Mei Ma, Xiujuan Lei.

Data curation: Mei Ma.

Formal analysis: Mei Ma, Xiujuan Lei.

Funding acquisition: Xiujuan Lei.

Investigation: Mei Ma, Xiujuan Lei.

Methodology: Mei Ma, Xiujuan Lei.

Supervision: Xiujuan Lei.

Validation: Mei Ma.

Writing – original draft: Mei Ma.

Writing – review & editing: Mei Ma, Xiujuan Lei.

References

1. Zaikis D, Vlahavas I. TP-DDI: Transformer-based pipeline for the extraction of drug-drug interactions. *Artif Intell Med.* 2021; 119:102153. <https://doi.org/10.1016/j.artmed.2021.102153> PMID: 34531012
2. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. *J Chem Inf Model.* 2019; 59(8):3370–88. <https://doi.org/10.1021/acs.jcim.9b00237> PMID: 31361484
3. Lin X, Quan Z, Wang ZJ, Huang H, Zeng X. A novel molecular representation with BiGRU neural networks for learning atom. *Brief Bioinform.* 2020; 21(6):2099–111. <https://doi.org/10.1093/bib/bbz125> PMID: 31729524
4. Weininger D, Weininger A, Weininger JL. SMILES. 2. Algorithm for generation of unique SMILES notation. *J Chem Inf Comput Sci.* 1989; 29:97–101.
5. Xu Z, Wang S, Zhu F, Huang J. Seq2seq fingerprint: An unsupervised deep molecular embedding for drug discovery. *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics.* 2017:285–94.
6. Duvenaud DK, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel TD, Aspuru-Guzik A, et al. Convolutional networks on graphs for learning molecular fingerprints. arXiv:1509.09292. [Preprint]. 2015. [cited 2022.Nov 18]. <https://doi.org/10.48550/arXiv.1509.09292>.
7. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: Moving beyond fingerprints. *J Comput Aided Mol Des.* 2016; 30(8):595–608. <https://doi.org/10.1007/s10822-016-9938-8> PMID: 27558503

8. Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J Med Chem*. 2020; 63(16):8749–60. <https://doi.org/10.1021/acs.jmedchem.9b00959> PMID: 31408336
9. Justin G, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, George E. Dahl. Neural message passing for quantum chemistry. *Proceedings of the 34th International Conference on Machine Learning* 2017;70:1263–72.
10. Klicpera J, Groß J, Günnemann S. Directional message passing for molecular graphs. *arXiv:2003.03123*. [Preprint]. 2020. [cited 2022.Nov 18]. <https://doi.org/10.48550/arXiv.2003.03123>.
11. Wang Z, Liu M, Luo Y, Xu Z, Xie Y, Wang L, et al. Advanced graph and sequence neural networks for molecular property prediction and drug discovery. *Bioinformatics*. 2022: btac112. <https://doi.org/10.1093/bioinformatics/btac112> PMID: 35179547
12. Yu Y, Huang K, Zhang C, Glass LM, Sun J, Xiao C. SumGNN: Multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*. 2021:btab207. <https://doi.org/10.1093/bioinformatics/btab207> PMID: 33769494
13. Xuan Lin, Zhe Quan, Zhi-Jie Wang, Tengfei Ma, Xiangxiang Zeng. KGNN: Knowledge graph neural network for drug-drug interaction prediction. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. 2021:380.
14. Chen Y, Ma T, Yang X, Wang J, Song B, Zeng X. MUFFIN: Multi-scale feature fusion for drug-drug interaction prediction. *Bioinformatics*. 2021:btab169. <https://doi.org/10.1093/bioinformatics/btab169> PMID: 33720331
15. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med*. 2012; 4(125):125ra31. <https://doi.org/10.1126/scitranslmed.3003377> PMID: 22422992
16. Sun X, Vilar S, Tatonetti NP. High-throughput methods for combinatorial drug discovery. *Sci Transl Med*. 2013; 5(205):205rv1. <https://doi.org/10.1126/scitranslmed.3006667> PMID: 24089409
17. Qiu Y, Zhang Y, Deng Y, Liu S, Zhang W. A comprehensive review of computational methods for drug-drug interaction detection. *IEEE/ACM Trans Comput Biol Bioinform*. 2022; 19(4):1968–85. <https://doi.org/10.1109/TCBB.2021.3081268> PMID: 34003753
18. Yu H, Mao KT, Shi JY, Huang H, Chen Z, Dong K, et al. Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol*. 2018; 12(Suppl 1):14. <https://doi.org/10.1186/s12918-018-0532-7> PMID: 29671393
19. Shi JY, Mao KT, Yu H, Yiu SM. Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization. *J Cheminform*. 2019; 11(1):28. <https://doi.org/10.1186/s13321-019-0352-9> PMID: 30963300
20. Kastrin A, Ferk P, Leskošek B. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PLoS One*. 2018; 13(5):e0196865. <https://doi.org/10.1371/journal.pone.0196865> PMID: 29738537
21. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. *J Biomed Inform*. 2017; 70:54–64. <https://doi.org/10.1016/j.jbi.2017.04.021> PMID: 28465082
22. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinformatics*. 2017; 18(1):18. <https://doi.org/10.1186/s12859-016-1415-9> PMID: 28056782
23. Zhang Y, Qiu Y, Cui Y, Liu S, Zhang W. Predicting drug-drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning. *Methods*. 2020; 179:37–46. <https://doi.org/10.1016/j.jymeth.2020.05.007> PMID: 32497603
24. Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics*. 2020; 36(15):4316–22. <https://doi.org/10.1093/bioinformatics/btaa501> PMID: 32407508
25. Feng YH, Zhang SW, Shi JY. DPDDI: A deep predictor for drug-drug interactions. *BMC Bioinformatics*. 2020; 21(1):419. <https://doi.org/10.1186/s12859-020-03724-x> PMID: 32972364
26. Su X, You ZH, Huang Ds, Wang L, Wong L, Ji B, et al. Biomedical knowledge graph embedding with capsule network for multi-label drug-drug interaction prediction. *IEEE Trans Knowl Data Eng*. 2022: <https://doi.org/10.1109/TKDE.2022.3154792>
27. Liu S, Zhang Y, Cui Y, Qiu Y, Deng Y, Zhang ZM, et al. Enhancing drug-drug interaction prediction using deep attention neural networks. *IEEE/ACM Trans Comput Biol Bioinform*. 2022: <https://doi.org/10.1109/TCBB.2022.3172421> PMID: 35511833
28. Zhu J, Liu Y, Zhang Y, Chen Z, Wu X. Multi-attribute discriminative representation learning for prediction of adverse drug-drug interaction. *IEEE Trans Pattern Anal Mach Intell*. 2021: <https://doi.org/10.1109/tpami.2021.3135841> PMID: 34914581

29. Zhu J, Liu Y, Wen C, Wu X. DGDFS: Dependence guided discriminative feature selection for predicting adverse drug-drug interaction. *IEEE Trans on Knowl and Data Eng.* 2022; 34(1):271–85.
30. Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Pache L, et al. Discovery of sars-cov-2 antiviral drugs through large-scale compound repurposing. *Nature.* 2020; 586(7827):113–9. <https://doi.org/10.1038/s41586-020-2577-1> PMID: 32707573
31. Jin W, Stokes JM, Eastman RT, Itkin Z, Zakharov AV, Collins JJ, et al. Deep learning identifies synergistic drug combinations for treating covid-19. *Proc Natl Acad Sci U S A.* 2021; 118(39):e2105070118. <https://doi.org/10.1073/pnas.2105070118> PMID: 34526388
32. Bobrowski T, Chen L, Eastman RT, Itkin Z, Shinn P, Chen C, et al. Discovery of synergistic and antagonistic drug combinations against sars-cov-2 in vitro. *BioRxiv* 2020.06.29.178889. [Preprint]. 2020. [cited 2022.Nov 18]. Available from: <https://doi.org/10.1101/2020.06.29.178889> PMID: 32637956
33. Howell R, Clarke MA, Reuschl AK, Chen T, Abbott-Imboden S, Singer M, et al. Executable network of sars-cov-2-host interaction predicts drug combination treatments. *NPJ Digit Med.* 2022; 5(1):18. <https://doi.org/10.1038/s41746-022-00561-5> PMID: 35165389
34. Nyamabo AK, Yu H, Shi J. Ssi-ddi: Substructure–substructure interactions for drug–drug interaction prediction. *Brief Bioinformatics.* 2021; 22(6):bbab133. <https://doi.org/10.1093/bib/bbab133> PMID: 33951725
35. Nyamabo AK, Yu H, Liu Z, Shi J-Y. Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Brief Bioinformatics.* 2021; 23(1):bbab441.
36. Jia J, Zhu F, Ma X, Cao Z, Cao ZW, Li Y, et al. Mechanisms of drug combinations: Interaction and network perspectives. *Nat Rev Drug Discov.* 2009; 8(2):11–28. <https://doi.org/10.1038/nrd2683> PMID: 19180105
37. Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug-drug and drug-food interactions. *Proc Natl Acad Sci U S A.* 2018; 115(18):e4304–e11. <https://doi.org/10.1073/pnas.1803294115> PMID: 29666228
38. Veličković P, Casanova A, Liò P. Graph attention networks. *International Conference on Learning Representations ICLR.* 2018.
39. Lu JS, Yang JW, Batra D, Parikh D. Hierarchical question-image co-attention for visual question answering. *Proceedings of the 30th International Conference on Neural Information Processing Systems.* 2016:289–97.
40. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res.* 2014; 42(Database issue):D1091–7. <https://doi.org/10.1093/nar/gkt1068> PMID: 24203711
41. Kingma D, Ba J. Adam: A method for stochastic optimization. *International Conference on Learning Representations 2014.*
42. Zhang W, Jing K, Huang F, Chen Y, Li B, Li J, et al. SFLLN: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions. *Inf Sci.* 2019; 497:189–201.
43. Deng Y, Qiu Y, Xu X, Liu S, Zhang Z, Zhu S, et al. META-DDIE: Predicting drug-drug interaction events with few-shot learning. *Brief Bioinform.* 2022; 23(1):1–8. <https://doi.org/10.1093/bib/bbab514> PMID: 34893793
44. Yang Z, Zhong W, Lv Q, Yu-Chian Chen C. Learning size-adaptive molecular substructures for explainable drug-drug interaction prediction by substructure-aware graph neural network. *Chem Sci.* 2022; 13(29):8693–703. <https://doi.org/10.1039/d2sc02023h> PMID: 35974769
45. White JM, Schiffer JT, Bender Ignacio RA, Xu S, Kainov D, Ianevski A, et al. Drug combinations as a first line of defense against coronaviruses and other emerging viruses. *mBio.* 2021; 12(6):e0334721. <https://doi.org/10.1128/mbio.03347-21> PMID: 34933447
46. Jitobaom K, Boonarkart C, Manopwisedjaroen S, Punyadee N, Borwornpinyo S, Thitithanyanont A, et al. Synergistic anti-sars-cov-2 activity of repurposed anti-parasitic drug combinations. *BMC Pharmacol Toxicol.* 2022; 23(1):41. <https://doi.org/10.1186/s40360-022-00580-8> PMID: 35717393
47. Bento AP, Hersey A, Félix E, Landrum G, Gaulton A, Atkinson F, et al. An open source chemical structure curation pipeline using RDKit. *J Cheminform.* 2020; 12(1):51. <https://doi.org/10.1186/s13321-020-00456-1> PMID: 33431044
48. Yu H, Zhao S, Shi J. STNN-DDI: A substructure-aware tensor neural network to predict drug-drug interactions. *Brief Bioinform.* 2022; 23(4):bbac209. <https://doi.org/10.1093/bib/bbac209> PMID: 35667078
49. Lee Jh, Lee I, Kang J.W. Self-attention graph pooling. *Proceedings of the 36 th International Conference on Machine Learning.* 2019:6661–70.
50. Feng YH, Zhang SW. Prediction of drug-drug interaction using an attention-based graph neural network on drug molecular graphs. *Molecules.* 2022; 27(9):3004. <https://doi.org/10.3390/molecules27093004> PMID: 35566354