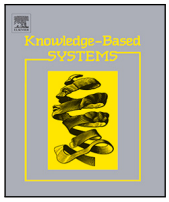
Knowledge-Based Systems 285 (2024) 111337

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



SRR-DDI: A drug–drug interaction prediction model with substructure refined representation learning based on self-attention mechanism

Dongjiang Niu, Lei Xu, Shourun Pan, Leiming Xia, Zhen Li *

College of Computer Science and Technology, Qingdao University, No. 308 Ningxia Road, Qingdao, 266071, Shandong, China

ARTICLE INFO

Keywords:
Drug–drug interaction
Substructure refinement
Self-attention
Drug similarity
Molecular graph

ABSTRACT

Drug–drug interaction (DDI) is an important safety issue during clinical treatment, where the mechanism of action of drugs may interfere with each other thereby causing adverse effects on the body leading to therapeutic failure. Since the deep learning method provide a powerful tool in DDI prediction, in this paper, we propose a DDI prediction model based on substructure Refined Representation Learning based on Self-Attention Mechanism, SRR-DDI, to improve the robustness of the substructure features determining the properties of the drug, thus improving the performance of DDI prediction. To improve the generalization of the model, the drug similarity feature is designed and introduced to help the model extract potential associations between drugs. The comparative experiments are set up on real-word data with two scenarios of warm start and cold start for performance evaluation, and the experimental results show that SRR-DDI outperforms the state-of-the-art methods. Finally, the visual interpretation experiment demonstrates that SRR-DDI can gradually refine the substructure features, and highlight the important substructures of drugs in DDI. In summary, SRR-DDI provides a powerful tool for predicting DDI and in-depth understanding of DDI.

1. Introduction

In the process of clinical treatment, some complex diseases [1] requires the concomitant administration of multiple drugs [2] for better therapeutic outcomes, but this also increases the risk of triggering drug–drug interactions (DDI) [3] which may lead to changes in drug efficacy or even put the patient's health at risk, so accurately predict and assess the risk of DDI is essential for the optimization of clinical treatment protocols [4]. DDI aims to identify potential drug–drug interaction problems to improve the drug design and development process and ensure the safety of drug therapy. Over the past few decades, researchers have conducted in-depth investigations into the mechanisms and factors that affect DDI, providing an important scientific basis for reducing adverse DDI and improving drug efficacy, but this process requires extensive domain knowledge and the experiments are expensive and time consuming.

With the development of artificial intelligence techniques, researchers can process and analyze large-scale DDI data more efficiently [5], which opens up more opportunities for AI Healthcare [6–8]. Artificial intelligence technologies can help researchers extract large amounts of data about DDI from the vast amount of literature [9] and clinical trial data [10], and construct prediction models to explore new DDIs accordingly, alleviating the problems in data acquirement and the need for extensive medical knowledge. DREAM [11] encoded the contextual information of each word in the sentence sequence from biomedical literature using BiLSTM [12], constructed dependency information in a long-distance word dependency graph by PageRank, and distinguished connectives using a graph attention mechanism for DDI extraction. On the other hand, Tiresias [13], a similarity-based linkage DDI prediction method, constructs interaction networks using the chemical structural similarity of drugs and literature-based interaction information for prediction.

However, the above work process only used information from the literature and clinical trials. It would be helpful if the artificial intelligence algorithms (AI) are used to explore deeper into the influence of molecular structure and composition on DDI prediction. The first problem that needs to be solved is how to represent the molecules that can be used as input to an AI algorithm. SMILES (Simplified Molecular Input Line Entry System) [14], which consists of a series of characters, each representing an atom, bond, and other specific chemical structural features in the drug molecule, can be used as the input for AI models. SMILES2Vec [15] uses SMILES as input, which is transformed into a low-dimensional feature representation of a drug through CNN (convolutional neural network)][16] and LSTM (long short-term memory network) [17] to predict the chemical properties of the drug. Hung et al. [18] extracted chemical similarity features from

* Corresponding author.
E-mail address: lizhen0130@gmail.com (Z. Li).

https://doi.org/10.1016/j.knosys.2023.111337
Received 4 August 2023; Received in revised form 17 October 2023; Accepted 21 December 2023
Available online 27 December 2023

the SMILES of a pair of drugs for pairwise interaction prediction based on the assumption that drugs with similar chemical properties have similar biological activities. In conclusion, the information contained in the 1D sequence representation of drugs can be applied to a variety of applications.

Relying on the sequence without considering chemical structure of the drug, the DDI prediction cannot fully take into account the chemical nature of the drug and cannot adequately capture the complex details of DDI. The graphical structure of a drug can be generated based on modeling the atoms and bonds in the drug molecule, in which the atoms are regarded as nodes and the bonds are regarded as edges in the graph. Meanwhile, with graph neural networks (GNNs) [19,20] being widely used in molecular science and chemistry for their excellent performance in processing graph data, the representation of drug molecules can also be learned through GNNs. GNNs progressively pass and integrate the features of nodes and edges by applying local information aggregation and updating mechanisms to obtain a global representation of molecular graph, it has played a notable role in various molecular tasks. To address the drawback of inadequate aggregation of neighborhood information in traditional GNNs, LAGNN [21] designs a local augmentation method, in which the network learns the features of a node in the graph by adding the feature distribution information of the neighboring nodes. Fang et al. [22] used Message Passing Neural Network (MPNN) to capture the local relationships between atoms in the graph structure by iteratively passing messages, and finally integrated the node information to obtain a representation of the molecular graph. The AttentionSiteDTI [23] proposed topological adaptive graph convolutional neural network to aggregate the molecular features of neighboring nodes through an adaptive topological weight matrix flexibly. To sum up, the GNN could extract features of graph structure effectively, providing additional scientific support to the field of computer-aided drug design.

Moreover, drug molecules are composed of several substructures, such as functional groups and ring structures, which own different biological activities and provide the chemical characteristics of the drug that is not available from the model based on single atomic information of the drug [24]. MR-GNN [25] represents the interactions between structured entities as multiple graphs, each corresponding to a different resolution. The graphs are aggregated through graph convolution layers to capture the contextual information of the substructures for each resolution. Moreover, a dual graph structure including semantic relationships and topological relationships is introduced exploit different levels of feature representation for DDI. SSI-DDI [26], a multilayer Graph Attention Network (GAT) [27] is used to extract the features of the drug, while feature of different substructures of the drug are obtained at each passing layer of the GAT, and interaction scores is calculated between different drug substructures for DDI prediction. Both methods have achieved good results in the field of DDI prediction using the substructure of drugs, demonstrating that the rational use of substructure of drugs can improve the performance of model.

GNNs learn the molecular representation by propagating and aggregating information between nodes in the graph, which limits the scope of message propagation, resulting in a limited understanding of the global information. On the other hand, the transformer model [28] achieves global information acquisition through a self-attention mechanism, thus overcoming the limitations of GNN on local message propagating to a certain extent. There are also some methods using transformer for DDI prediction. MDF-SA-DDI [29] combines two drugs in four different networks to obtain the potential features of the drug pairs, and a transformer is introduced for feature fusion, the output of which is used for the DDI prediction task. Moreover, the graphic structure could also be used in the transformer. MolGNet [30] updates the state of the nodes in the molecular graph by aggregating information through a transformer with a GRU network [31] added. In traditional graph neural networks, information transfer between nodes is usually achieved through a message-passing mechanism. The GTN [32] introduces the self- attention mechanism into the modeling of graph-structured data. Compared with traditional GNNs, GTN can dynamically adjust the importance between different nodes to better capture the feature in the graph structure.

At the same time, the deep learning model such as transformer contains a large number of parameters and suffers from the problem of long training time and slow convergence speed. To address this problem, there are many methods trying to improve the traditional normalization layer. The batch normalization [33] was used to solve the internal covariate transfer problem in deep learning model, which is added into each layer of the network to improving the training effect. Layer normalization [34] is different from batch normalization in that it applies the normalization to each feature dimension of each sample of the network. Similarly, in the field of natural language processing, Xiong et al. [35] pointed out that the most difficult stage of transformer parameter optimization is at the early stage of model training, therefore, by adjusting the position of layer normalization, Post-LN and Pre-LN were designed and have been proved to improve the training process of transformer. In this paper, we try to select appropriate normalization layers and find the optimal position for it to speed up the convergence process for molecular representation and DDI prediction.

The similarity between drugs is an important approach of molecular activity prediction, based on the principle that drug molecules with similar structures are likely to own similar properties [36]. Similarly, we would like to explore whether the similarity is helpful in DDI prediction. The DDI effects between commercially known drugs and the DDI effects with unknown drugs are two majority tasks in DDI prediction. The latter task is more difficult since the trained model cannot cover features of all possible drugs with limited known DDI data. So the similarity provides another view of molecular representation, and how to introduce similarity features into DDI prediction to increase the generalizability of the model is also a problem that needs to be addressed in this paper.

Although some methods focus on molecular representation using graph structure by transformer at the atomic level [29,30]. The information contained in a single atom is limited, the structure of a compound could be regarded as a combination of several substructures, and the properties of these substructures will play an important role in certain reactions, so by extracting the most accurate substructural features that determine the drug properties not only improves the predictive performance of the model but also improves the interpretability of the model. Despite the substructure-based GNN models have been developed, the existing graph transformer [37] cannot deal with drug substructures directly, how to integrate the substructure information of drugs into the graph transformer, to utilize the advantage of transformer and improve the accuracy of molecular representation is still open to be solved. In this paper, we design a DDI prediction model using substructure refined representation learning based on self-attention mechanism, called SRR-DDI. Two experimental settings, warm start, and cold start, on two real datasets DrugBank [38] and Twosides [39] were implemented for evaluation of the model, which show that our model outperformed other state of art methods.

To sum up, the contributions of the proposed methods are listed as below:

- To integrate the substructure information of drugs into the DDI prediction, a substructure refinement module using graph transformer is proposed to extract the substructural features and improved the interpretability of the model.
- To speed up the convergence process for DDI prediction, we test different normalization layers and find the optimal position for it.
- The similarity feature is introduced into the DDI prediction to increase the generalizability of the model.

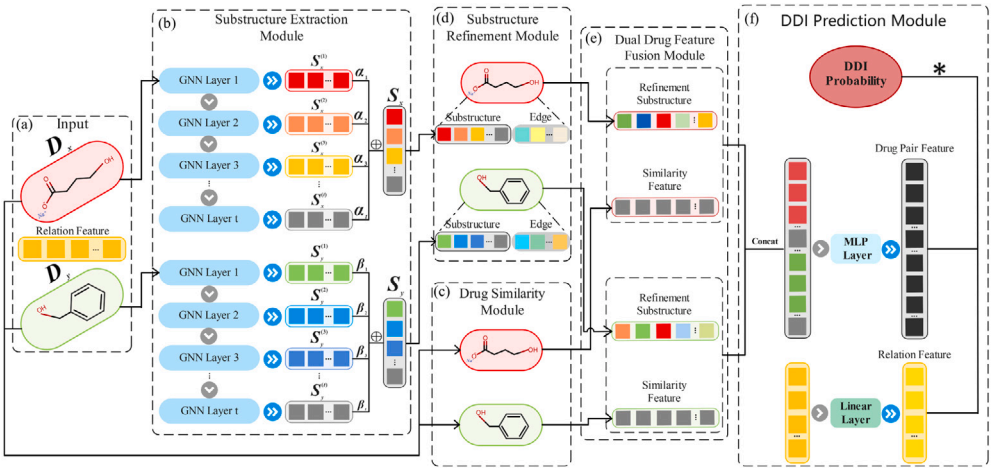


Fig. 1. The overview of SRR-DDI. (a) The input module receives the DDI triplet (D_x, D_y, r) as input to the model, including the two drugs and the types of interactions. (b) The substructure extraction module extracts the substructure feature of the drugs with different radius by multi-layer GNN. (c) The drug similarity module is used to calculate the similarity feature of drugs. (d) The substructure features are then passed to the substructure refinement module to yield a richer substructure feature of the drug. (e) The obtained substructure feature and drug similarity feature are fed into the dual drug feature fusion module to obtain the graph-level features. (f) The drug pair features and relation feature encoded by the model are fed into the DDI prediction module to obtain the final predicted probability scores.

2. Method

2.1. The overview of SRR-DDI

The overall framework of the SRR-DDI model is described in detail in this section, which is shown in Fig. 1. After a drug pair enters the SRR-DDI model, the drug pair is encoded as a graph structural feature by RDKit [40], and the data enters the drug similarity module to iteratively compute with all the drugs in the dataset to obtain the similarity feature. In the substructure extraction module, substructures of different scales of the drug can be obtained by multi-layer GNN networks, and different importance scores are assigned to them to obtain the substructure features of the drug. The extracted substructure features are fed into the substructure refinement module, where more accurate substructure features of the drug will be obtained through the graph transformer. Drug similarity feature and substructure features are aggregated in the dual drug feature fusion module to obtain the final representation of the drug. Finally, the probability score of a certain reaction between the two drugs is calculated by the DDI prediction module. The DDI prediction problem can be viewed as a binary classification task, specifically, given a database of drugs D, a database of response types R, and a triplet of DDIs $\{(D_x, D_y, r)_i\}_{i=1}^{N}$, where D_x, $D_y \in D$ and $r \in R$ represent a pair of drugs with interaction type r. The DDI prediction task of the model can be expressed as a function $F : D_x \times D_y \times r \rightarrow [0, 1]$, the output of the model determines the probability of interaction of type r of two drugs. Generally, 0.5 is used as the classification threshold.

2.2. Substructure extraction module

The drug molecules D_x, D_y are represented by the graph=(V, E) according to their structures, each atom in the drug molecule forms the node set $V = (v_i)_{i=1}^{n}$, n denotes the number of atoms in the drug, and each edge in the drug forms the bond set $E = \{(v_s, v_t)_i\}_{i=1}^{m}$, m denotes the number of bonds in the drug, and (v_s, v_t) denotes the bond between atoms v_s and v_t.

In general, drug molecules consist of many substructures that play an important role in the molecular property, therefore, a substructure feature extraction module is adopted to better capture the key structure features of molecule by aggregating substructure feature.

The traditional GNN model obtain the representation of the entire graph by iteratively updating the features of the nodes from their neighboring nodes. Inspired by SA-DDI [41], the message-passing process of our model is implemented through bonds. First, the model processes the input features to form the bond-level features of all edges in the graph as Eq. (1):

$$h_{ij} = \frac{1}{3}(h_i + h_j + e_{ij}) \tag{1}$$

where h_i and h_j denote the node features of the start and end points of the edges, respectively, e_{ij} is the feature of the bond between nodes i, j. h_{ij} is the representation of the bond between nodes i and j after model processing.

After getting the bond-level features, the softmax function is used to find the importance score of each bond-level feature in the whole graph representation, which are fed into the pooling layer to obtain the pooled features of the graph as Eq. (2):

$$g^{(t)} = AddPooling(softmax(GNN(H^{(t)}, A))H^{(t)}) \tag{2}$$

where $H^{(t)}$ is the matrix consisting of all bond-level features obtained from the GNN at layer t, and A is the adjacency matrix of the graph.

An attention score $s^{(t)}$ is calculated using $g^{(t)}$ as follows:

$$s^{(t)} = \alpha^{(t)}\sigma(W g^{(t)} + b) \tag{3}$$

where $\alpha^{(t)}$ is the weight vector of the layer t and σ is the tanh activation function. The importance weights of the substructures of each layer are then obtained by softmax normalizing the attention score $s^{(t)}$ of the substructure features of all layers, The final bond-level features h_{ij}^{final} which will be used to extract substructure feature are derived according to the Eq. (4):

$$h_{ij}^{final} = \sum_{t=1}^{T} softmax(s^{(t)})h_{ij}^{(t)} \tag{4}$$

where T is the number of network layers of the substructure extraction module. Finally, the $i-th$ substructure representation h_i^s is obtained through the hidden features of each node and its neighboring bond-level hidden features as Eq. (5):

$$h_i^s = F_n(x_i^{(0)} + \sum_{j \in \mathcal{N}_i} h_{ij}^{final}) \tag{5}$$

where F_n is a nonlinear function containing a multilayer perceptron, $\mathcal{N}_i$ is a neighbor nodes group of atom i and $x_i^{(0)}$ is the initial representation of node i.

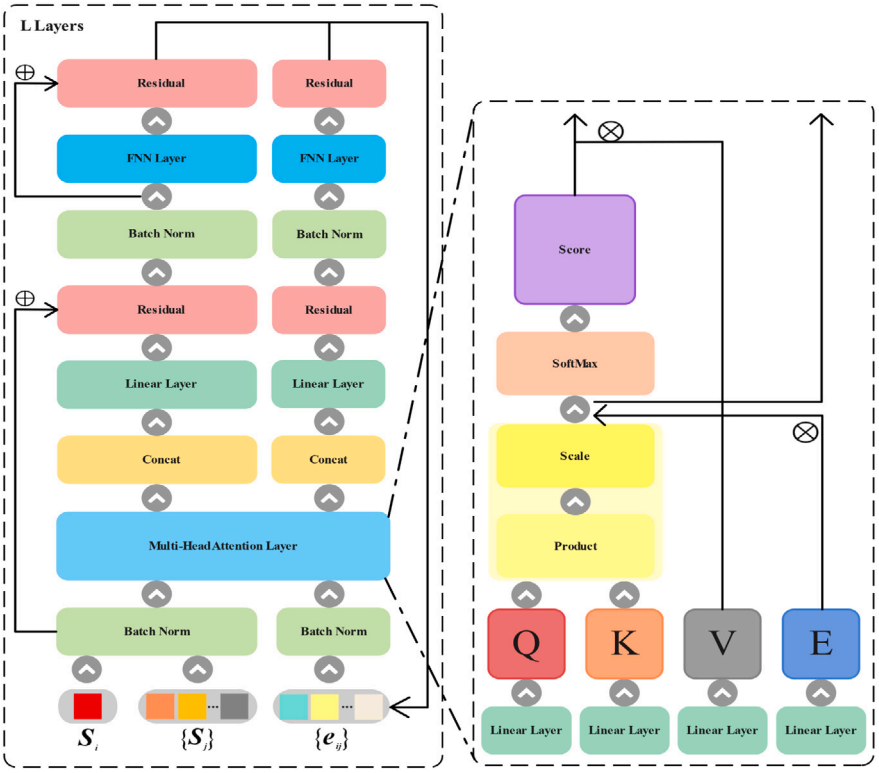


Fig. 2. Introduction of Substructure Refinement Module.

2.3. Substructure refinement module

Many DDI prediction methods use the extracted drug substructure feature to perform the DDI prediction task directly through GNN or use it to construct the global features of the drug for prediction. Considering improving the comprehensiveness of the substructure feature, we designed a substructure refinement module through graph transformer that can directly encode the substructure feature of a drug with the interactions between the various substructures in the drug. Moreover, the feature of molecular bond is also introduced in the module. Meanwhile, to reduce the phenomenon of gradient vanishing and gradient explosion, so as to speed up the training process of the model and improve the accuracy of the model prediction, we appropriately adjusted the positional order of some layers in the graph transformer, and the details of the module are shown in Fig. 2. Specifically, the obtained substructure feature and edge feature are fed into the substructure refinement module, the refined substructure feature will pay more attention to the important part of the drug in the chemical properties and DDI.

For the self-attention mechanism in the graph transformer, query (Q), key (K), and value (V) all come from the input substructure, Unlike traditional attentions mechanisms, the edge feature E is introduced in drug graphs to improve the comprehensive of the feature. Compared to traditional transformers, the feature of each substructure and edge are normalized, to reduce the overfitting situation of the model and improve the generalization ability, SRR-DDI uses batch normalization and adjusts its position to ensures the model is robust to changes in the scale and distribution of the input data, and also alleviates problems such as gradient disappearance and gradient explosion, which is helpful in improving the training stability and convergence of the model.

Then the multi-headed attention operation is performed as Eqs. (6), (7), and (8):

$$score_{ij}^{h} = softmax\{(\frac{Q^{l}BN(h_i^{s(l)}) \cdot K^{l}BN(h_j^{s(l)})}{\sqrt{d_k}}) \cdot E^{l}BN(e_{ij}^{(l)})\} \tag{6}$$

$$\hat{h}_i^{s(l)} = Concat \parallel_{h=1}^{H} \{\sum_{j \in \mathcal{N}_i} score_{ij}^{h} \cdot V^{l}BN(h_j^{s(l)})\} \tag{7}$$

$$\hat{e}_{ij}^{(l)} = Concat \parallel_{h=1}^{H} \{(\frac{Q^{l}BN(h_i^{s(l)}) \cdot K^{l}BN(h_j^{s(l)})}{\sqrt{d_k}}) \cdot E^{l}BN(e_{ij}^{(l)})\} \tag{8}$$

where $h_i^{s(l)}$ and $e_{ij}^{(l)}$ are the substructure and edge features obtained at the layer l, $Concat \parallel_{h=1}^{H}$ denotes multi-head concatenating, H denotes the number of attention heads, Q^l, K^l, V^l are learnable weight matrices, BN denotes batch normalization operation, the dot product result is divided by $\sqrt{d_k}$ to prevent the result from being too large, and d_k is the dimension of K^l, $\mathcal{N}_i$ is the set of substructures centered at node i and its neighboring substructures. The residual concatenation operation is then performed as Eq. (9):

$$\hat{\hat{h}}_i^{s(l)} = BN(h_i^{s(l)}) + \omega_h^{RE}\hat{h}_i^{s(l)} \tag{9}$$

where ω_h^{RE} is the learnable weight matrices. The same residual concatenation operation is performed on the bond of the drug to obtain $\hat{\hat{e}}_{ij}^{(l)}$. Both features $\hat{\hat{h}}_i^{s(l)}$ and $\hat{\hat{e}}_{ij}^{(l)}$ are fed into the normalization layer and the feed forward neural network layer:

$$\hat{\hat{\hat{h}}}_i^{s(l)} = FN2\{\sigma(FN1(BN(\hat{\hat{h}}_i^{s(l)})))\} \tag{10}$$

where both $FN1$ and $FN2$ are the feed forward operation and σ is the activation function. Similarly, the bond feature of the drug goes through a similar operation as Eq. (10) to obtain the new bond-level feature $\hat{\hat{\hat{e}}}_{ij}^{(l)}$. Then $\hat{\hat{\hat{h}}}_i^{s(l)}$ and $\hat{\hat{\hat{e}}}_{ij}^{(l)}$ are fed into the last residual connected layer as follow to obtain the refinement feature as Eq. (11):

$$h_i^{s(l+1)} = BN(\hat{\hat{\hat{h}}}_i^{s(l)}) + \hat{\hat{\hat{h}}}_i^{s(l)} \quad e_{ij}^{(l+1)} = BN(\hat{\hat{\hat{e}}}_{ij}^{(l)}) + \hat{\hat{\hat{e}}}_{ij}^{(l)} \tag{11}$$

where $h_i^{s(l+1)}$ is the feature of the substructure after the layer L, and $e_{ij}^{(l+1)}$ is the feature of the bond between node i and node j after the layer L.

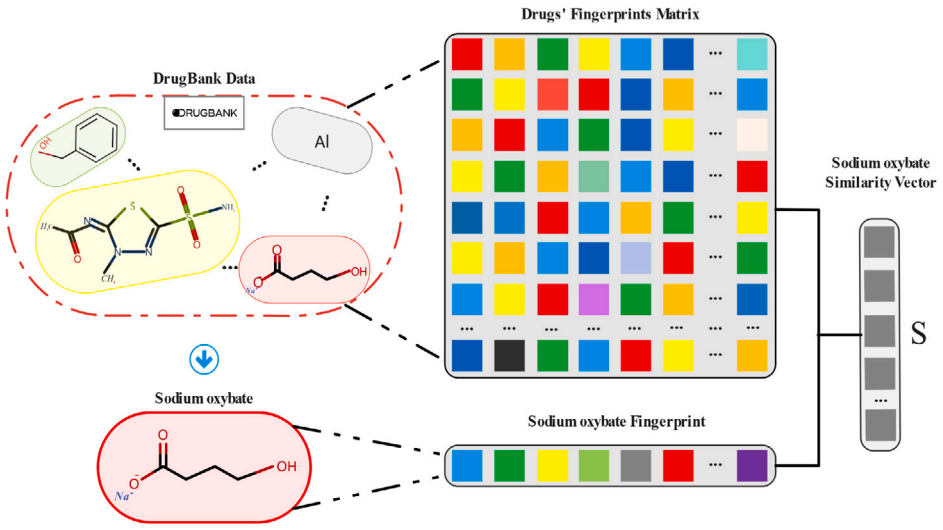


Fig. 3. Introduction of Drug Similarity Module. All drugs are obtained from the DrugBank, and the similarity score of any drug to all drugs in the database is calculated as the similarity features of the drug.

2.4. Drug similarity module

The drug similarity feature is introduced in the SRR-DDI by considering the effect of structural, chemical properties and functional similarities between drugs on their interactions, thus helping the model to better generalize to drug co-response relationships that have not been learned during the training process. By grouping similar drugs into the same or close clusters, the model can learn relationships between drugs from existing interaction information and apply these relationships to the prediction of new drugs. In the SRR-DDI model, we calculated the similarity scores between drugs using the Tanimoto coefficient [42], the process of which is shown in Fig. 3. For each molecule, the Morgan fingerprint [43] is extracted, which encodes molecular structure feature based on the chemical environment. Then, the similarity scores of the two molecular fingerprints are obtained through Eq. (12):

$$s_{xy} = \frac{\sum_i a_i \cdot b_i}{\sum_i a_i^2 + \sum_i b_i^2 - \sum_i a_i \cdot b_i} \tag{12}$$

where a_i and b_i denote the value of the ith element in the fingerprint vector of two molecules, respectively. Following the above method, all the drugs in the database are we traversed to calculate the similarity scores of a drug to form the similarity feature S for the drug.

2.5. Dual drug feature fusion module

To make full use of the drug feature and drug similarity feature, a dual fusion module is designed as shown in Fig. 4. The refined substructure feature is input to the fusion module with two pooling operations, which will be concatenated to improve the comprehensiveness of drug features through Eq. (13).

$$\hat{G}_x = Concat\{MaxPooling(h^s), MeanPooling(h^s)\} \tag{13}$$

where $Concat$ is the concatenating operation, which can avoid the information loss of single pooling method and increase the robustness and stability of the model. $MaxPooling$ is a global maximum pooling operation that enables the graph level molecular representation to focus more on the most critical information in the drug, and $MeanPooling$ is a global average pooling operation that captures the global and average feature in the drug. Finally, the drug graph level feature is combined with the drug similarity feature to obtain the final feature representation G_x of the drug D_x as Eq. (14):

$$G_x = Concat(\hat{G}_x, \omega_x S) \tag{14}$$

where ω_x is the learnable weight matrix. Similarly, we can get the graph feature representation G_y of another drug D_y using the same method.

2.6. Drug–drug interaction prediction module

After obtaining the final feature G_x and G_y of two drugs D_x and D_y, given a DDI triplet (D_x, D_y, r) for two drugs and their interaction types, the joint probability score of the DDI prediction is calculated as Eq. (15):

$$P(D_x, D_y, r) = \sum\{L_p(Concat(G_x, G_y)) * E_r(r)\} \tag{15}$$

where L_p is a linear layer that uses the parameter-corrected linear rectification function (PReLU) as the activation function, E_r is an embedding layer that deals with relational features that the model can use for subsequent computation, $*$ is the operation of multiplying by elements between two matrices.

Finally, the cross-entropy function is selected as the loss function to minimize the difference between the predicted results and the true labels as Eq. (16):

$$\ell = -\frac{1}{ך}\sum_{(D_x, D_y, r)_{i=1}}^{ך}\{y_i \log p_i + (1 - y_i)\log(1 - p_i)\} \tag{16}$$

where ך is the total number of training samples, y_i is the true label of sample i, $y_i = 1$ means there is an interaction between D_x and D_y, and 0 otherwise, and p_i is the probability that the model predicts an interaction for drug pair i.

3. Experiment

3.1. Dataset

We evaluated the performance of SRR-DDI model on two real world datasets, DrugBank [38] and Twosides [39]. The DrugBank is integrated from multiple sources, including pharmaceutical literature, drug package inserts, drug registries, etc., which is organized and validated by a team of professionals to ensure the accuracy and reliability of the data. DrugBank contains 1,706 drugs covering 191,808 DDI triplets, a triplet containing a pair of drugs and one reaction type of those two drugs, and there are 86 DDI types in total. The Twosides is much larger relative to DrugBank, including 645 drugs and 4576287 DDI triples, containing a total of 963 DDI types.

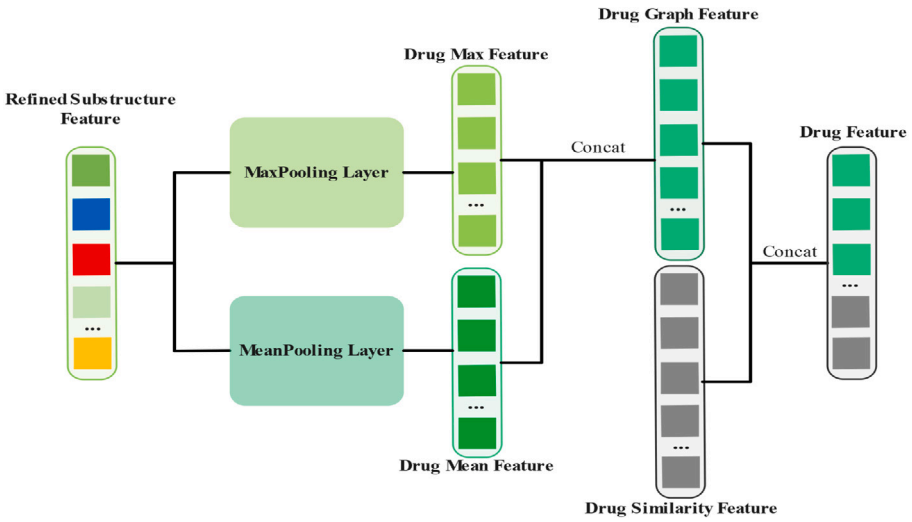


Fig. 4. Introduction of Dual Drug Feature Fusion Module.

3.2. Baseline

This study aimed to address the problem of inadequate mining of substructure feature and drug global representation in the field of DDI prediction. We compared SRR-DDI in the following state-of-the-art baseline models dealing with drug substructures:

- DeepDDI [44]: A DDI prediction model using names of drug–drug and their structural information.
- MR-GNN [25]: A DDI prediction method through multi-resolution graph representation and bi-graph structure was designed to extract and utilize the feature representation of the substructure.
- SSI-DDI [26]: A DDI prediction method based on substructure interactions.
- GAT-DDI [45]: A DDI prediction model based on graph attention mechanism. Through a multilayer GAT network, the model could learn the complex relationships between nodes in a drug molecule.
- SA-DDI [41]: A DDI prediction model with learnable size-adaptive substructure.
- DGNN-DDI [46]: A DDI prediction model using substructure co-attention mechanism.

3.3. Experimental settings

In our experiments, each triplet in the dataset is a positive sample, and then we use the method proposed by Wang et al. [47] to generate a negative sample. Because the dataset mainly contains known positive DDIs, in order to perform the binary classification task, we need to construct some negative samples from other pairs with unknown interactions randomly. To more comprehensively assess the model and determine its potential for real-world applications, we used the mean and variance of the following six metrics to evaluate the performance of SRR-DDI[1], which are also used in other DDI prediction method:

- Accuracy (ACC): Accuracy is an indicator of the proportion of samples correctly predicted by the model and is used to measure the model's classification accuracy in the overall sample.
- Area Under the Curve (AUC): The area under the ROC curve, which measures the model's classification accuracy and the higher the AUC value, the better the model's ability of classification.
- Precision (Prec), recall (Rec) and F1 Score (F1):

$$Precision = \frac{TP}{(TP + FP)} \tag{17}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{18}$$

$$Precision = 2 \times \frac{(Precision \times Recall)}{(TPrecision + Recall)} \tag{19}$$

where TP, FP and FN represent, respectively, true positive, false positive and false negative. The $F1$ Score is a combination of precision and recall metrics to measure the accuracy and completeness of the model in sample classification.

- Average Precision (AP): The Precision–Recall (P–R) curve can be plotted by taking the Recall value as the horizontal coordinate and Precision as the vertical coordinate. The value of AP ranges from 0 to 1, and the higher the value, the better the balance between accuracy and recall, and the better the performance of the model. The value of AP can be obtained by integrating the P–R curve:

$$AP = \int_0^1 P(r)dr \tag{20}$$

3.4. Performance evaluation on warm start scenario

In the warm start scenario, to ensure the fairness comparisons with other baseline models, we followed the same experiment setup used in DeepDDI [44], SA-DDI [41] and SSI-DDI [26]. The dataset is partitioned into training, validation, and test sets according to the ratio of 6:2:2. To more accurately assess the performance and stability of the model and compare with other methods under the same setting, we also performed three-fold cross-validation experiments. For warm start setting, each drug in the validation and test sets will also exist in the training set individually, but the drugs pairs in the training set will not appear in the other two sets. For example, if DDI between drug A and drug B are in the validation set, the interaction between A and C, or B and D will appear in the training set. But the interaction between A and B will not be in the training sets, i.e., no unknown drugs will appear in the validation and test sets in the warm start scenario. In DrugBank, the batch is set to 512, the Adam optimizer is selected to update the model parameters, the exponential decay learning rate is 1e-3 and the weight decay is set to 5e-4 to prevent overfitting. In Twosides, the batch is set to 1024, all dropout settings are set to 0.3 and the model depth is appropriately deepened, and the rest of the settings are the same as that in the DrugBank to maintain consistency.

[1] The source codes of SRR-DDI are available at https://github.com/NiuDongjiang/SRR-DDI

Table 1
Performance evaluation of SRR-DDI and baseline models in warm start scenario on DrugBank.

Model	ACC	AUC	F1	Prec	Rec	AP
DeepDDI	93.21 ± 0.27	97.03 ± 0.11	93.37 ± 0.22	91.26 ± 0.26	95.52 ± 0.43	95.95 ± 0.21
GAT-DDI	92.15 ± 0.11	96.19 ± 0.12	92.29 ± 0.16	90.28 ± 0.19	95.34 ± 0.33	95.02 ± 0.03
MR-GNN	93.26 ± 0.14	97.26 ± 0.04	93.35 ± 0.12	91.26 ± 0.21	95.69 ± 0.02	96.45 ± 0.07
SSI-DDI	94.24 ± 0.13	98.12 ± 0.03	94.29 ± 0.08	93.01 ± 0.02	96.11 ± 0.07	97.25 ± 0.02
SA-DDI	96.21 ± 0.13	98.78 ± 0.04	96.27 ± 0.11	95.01 ± 0.08	97.62 ± 0.02	98.36 ± 0.03
DGNN-DDI	96.12 ± 0.05	98.90 ± 0.02	95.98 ± 0.16	94.86 ± 0.07	97.89 ± 0.03	98.46 ± 0.05
SRR-DDI	**96.67 ± 0.06**	**99.05 ± 0.03**	**96.72 ± 0.05**	**95.28 ± 0.08**	**98.24 ± 0.03**	**98.74 ± 0.04**

Table 2
Performance evaluation of SRR-DDI and baseline models in warm start scenario on Twosides.

Model	ACC	AUC	F1	Prec	Rec	AP
DeepDDI	75.16 ± 0.23	82.42 ± 0.31	77.03 ± 0.05	71.65 ± 0.59	83.27 ± 0.84	79.47 ± 0.33
GAT-DDI	67.32 ± 2.04	75.16 ± 2.44	63.70 ± 3.11	71.54 ± 2.19	57.65 ± 5.09	72.48 ± 2.45
MR-GNN	85.39 ± 0.31	91.93 ± 0.21	86.46 ± 0.27	80.57 ± 0.37	93.28 ± 0.21	89.32 ± 0.22
SSI-DDI	82.21 ± 0.38	89.25 ± 0.45	83.32 ± 0.45	79.15 ± 0.32	87.55 ± 0.64	86.19 ± 0.41
SA-DDI	87.45 ± 0.06	93.15 ± 0.04	**88.35 ± 0.04**	82.43 ± 0.02	**95.18 ± 0.10**	**90.51 ± 0.08**
DGNN-DDI	85.33 ± 0.12	92.09 ± 0.12	84.87 ± 0.31	81.03 ± 0.17	89.35 ± 0.17	89.78 ± 0.24
SRR-DDI	**87.52 ± 0.05**	**93.20 ± 0.07**	88.32 ± 0.05	**83.06 ± 0.02**	94.38 ± 0.11	90.37 ± 0.06

Tables 1 and 2 summarized the warm start scenario's performance of SRR-DDI and other baseline models on two datasets, which shows that various evaluation metrics of our model was improved compared to previous methods. Specifically, the accuracy of SRR-DDI on the DrugBank and Twosides reached 96.67% and 87.52% respectively and the stability of it was higher according to the standard deviation obtained from the three-fold cross-validation, which was due to the fact that we targeted the substructures of the drugs and deeply mine the feature that better represents the drug properties through the refinement operation, thus reducing the redundant information in the molecular representation. Moreover, drug similarity features was used to obtain the potential linkage information between drugs to assist the DDI prediction. Similarly, the other evaluation metrics were also significantly improved, and these results fully reflected the superior performance of our proposed model.

In order to discovery the effect of number of layers on performance, the experiments of different graph transformer layers on DrugBank were implemented in the warm start scenario. Theoretically, the larger the depth of the model, the better the performance of the model, since deep architecture could combine multi-level features to increase the learning ability of the model. On the other hand, it also leads to an increase in the computation and the number parameters of model. The results of the experiments on three-fold are shown in Fig. 5, and the performance of the model reached the best when the number of layers of graph transformer was 3. It is concluded that model with less layers is difficult to extract sufficient information, and beyond a certain number of layers it will face the problems such as gradient disappearance and overfitting, so it is necessary to find a balance point between the performance and the efficiency.

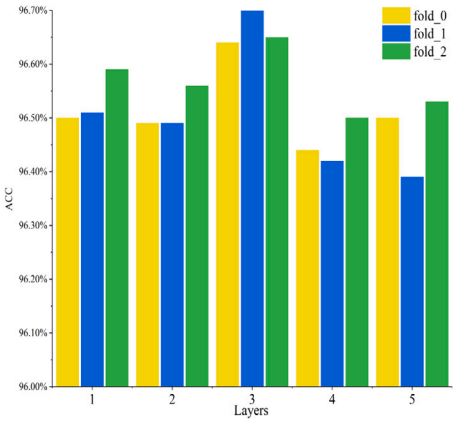


Fig. 5. Comparative results with respect to the different number of layers on three-fold.

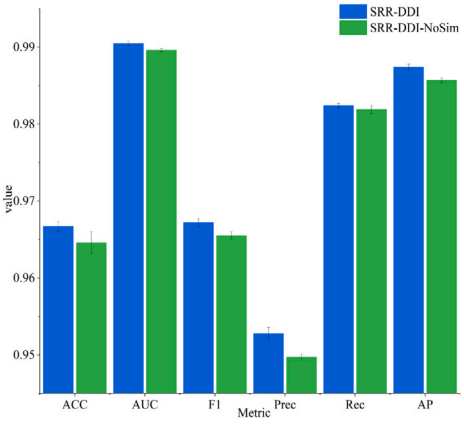


Fig. 6. Comparative results on feature of drug similarity.

3.5. Ablation experiments of SRR-DDI modules

To investigate and demonstrate the importance of each module of SRR-DDI, we have set up three ablation experiments for the substructure refinement module, drug similarity module and dual drug feature fusion module, respectively. All of ablation experiments are implemented on the DrugBank.

3.5.1. Ablation experiment for drug similarity module

Fig. 6 summarized the results of our ablation experiments for the drug similarity module. Although the drugs in both the validation and test sets were present in the training set in the warm start scenario, the paired DDI triplets in the validation and test sets were not visible in the training set, so we added drug similarity feature to compensate for this information gap. As shown by the experimental results, the performance of SRR-DDI was improved after adding feature of drug similarity, proving that the added information is effective for the DDI prediction task.

3.5.2. Ablation experiment for substructure refinement module

For the substructure refinement module, we performed three additional ablation experimental studies:

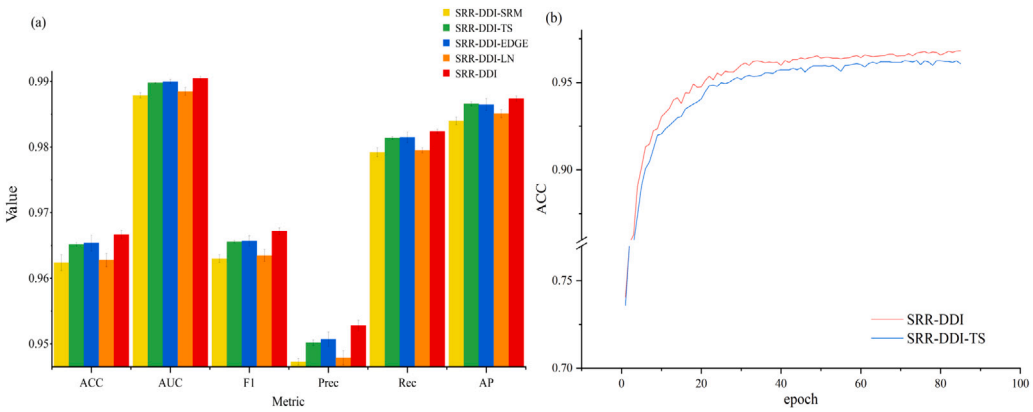


Fig. 7. Comparative results on the substructure refinement module with its variants. (a) Comparison of the performance metrics of the SRR-DDI and the variant model. (b) Accuracy curve with respect to different epochs in the validation set during training of SRR-DDI vs. SRR-DDI-TS.

- SRR-DDI-SRM: In this model, the substructure refinement module of SRR-DDI was removed and the other modules remained unchanged, which is used to demonstrate the validity of the substructure refinement module.
- SRR-DDI-TS: In this model, the substructure refinement module is retained, but a common graph transformer with different position of layer settings is used to replace our model, which illustrates the importance of the position of the normalization layer used in SRR-DDI.
- SRR-DDI-LN: In this model, the substructure refinement module was retained, and the normalization layer in the module was set to layer normalization to prove the validity of batch normalization in the module.
- SRR-DDI-EDGE: In this model, the edge feature in the self-attention mechanism of the substructure refinement module was removed to demonstrates the effectiveness of the introduction of edge feature to calculate the attention score.

The performance comparison of the four variants of the SRR-DDI and substructure refinement modules are shown in Fig. 7(a). It could be seen that SRR-DDI outperformed all variant architectures in all evaluation metrics. The accuracy of the model removing the substructure refinement module was much lower model than SRR-DDI because the lack of refinement information in the substructure prevents the substructure features from containing precise drug structure feature. Similarly, all the performance metrics of SSR-DDI-EDGE were similarly lower than those of the full SSR-DDI architecture, which was because edges carry chemical relationships and interactions between different nodes or groups in the molecular graph, and thus the reasonable introduction of edge features could effectively enhance the topological relationships and chemical property information of drugs. Fig. 7(b) summarized the accuracy curve with respect to different epochs of the validation set during the training process of SSR-DDI and SSR-DDI-TS, the accuracy of SSR-DDI in the validation set was higher than that of the SRR-DDI-TS. Moreover, and the number of epochs for finding the optimal value of SRR-DDI was also smaller than that of SRR-DDI-TS, which indicates that placing batch normalization before the residual connection was helpful. In Fig. 7(a), all the metrics of SRR-DDI were far superior to those of SRR-DDI-LN, which proves that batch normalization can better exploit the performance of SRR-DDI compared to layer normalization. In deep learning, the distribution of the input of each hidden layer often changes, which is regarded as internal covariate shift and will affect the performance of the model. The batch normalization is a differentiable transformation, which ensure that each layer could learn on input distributions with less internal covariate shift by introducing normalized activations. Moreover, high learning rate of model may result in the gradients vanish using other normalization method. However, the batch normalization could prevent small changes in layer parameters to keep high learning rate and speed up the convergence process [48].

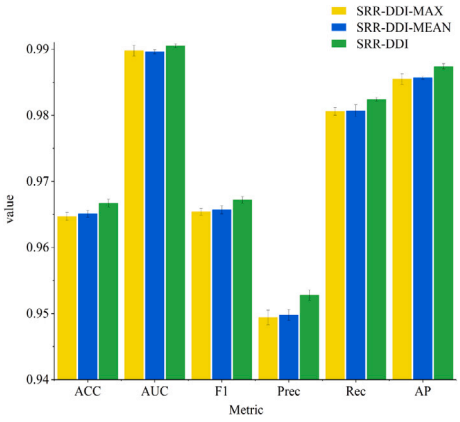


Fig. 8. Comparative results on the dual drug fusion modules with its variants.

3.5.3. Ablation experiment for dual drug feature fusion module

In the dual drug feature fusion module, two types of pooling method including MaxPooling and MeanPooling are both combined for fusion, to evaluate the impact of different pooling method, The SRR-DDI-MAX with a single MaxPooling layer and SRR-DDI-MEAN with a single MeanPooling layer.

The experimental results are shown in Fig. 8. Maximum pooling focuses only on the maximum value of the features in the drug graph, strengthens the most important substructure features in the drug, and ignores other potentially important drug features. Similarly, average pooling is an averaging operation of all the substructure features of the drug, which can capture the global feature and distribution in the molecular graph, but this leads to smoothing of the each substructure, i.e., the differences between drug nodes is missing. In summary, these two methods own advantages and disadvantages. In our fusion module, the two complementary pooling method are combined to improve the comprehensiveness of the molecular representation. Experimental results showed that our method outperformed single pooling methods(SRR-DDI-MAX and SRR-DDI-MEAN).

3.6. Performance evaluation on cold start scenario

To better demonstrate the learning ability of the proposed method, we also set up a more challenging cold start scenario for SRR-DDI on the DrugBank, as same as other DDI methods. In this scenario, the trained model is applied to unseen data. We segmented the dataset according to SSI-DDI [26], specifically, 20% of the drugs in the DrugBank were randomly selected as unknown drugs, the rest were known drugs, and

Table 3
Performance evaluation of SRR-DDI and baseline models in the cold start scenario on the DrugBank using two segmentation schemes.

	Model	ACC	AUC	F1	Prec	Rec	AP
S_1	DeepDDI	73.13 ± 1.09	81.52 ± 1.07	66.08 ± 0.89	84.09 ± 3.09	59.79 ± 4.24	81.75 ± 1.08
	GAT-DDI	70.13 ± 0.78	78.07 ± 0.91	72.53 ± 0.35	82.32 ± 1.47	56.27 ± 0.87	78.67 ± 0.46
	MR-GNN	74.32 ± 0.43	83.12 ± 0.27	70.01 ± 0.62	83.21 ± 1.06	66.09 ± 0.76	83.75 ± 0.54
	SSI-DDI	73.23 ± 0.78	80.92 ± 1.16	73.18 ± 1.17	81.54 ± 2.21	65.17 ± 1.87	**84.67 ± 0.67**
	SA-DDI	75.15 ± 0.64	82.76 ± 1.15	72.02 ± 0.89	83.87 ± 0.67	**67.09 ± 0.56**	84.07 ± 0.94
	DGNN-DDI	75.39 ± 0.64	83.14 ± 0.56	69.77 ± 0.67	84.34 ± 0.31	66.84 ± 0.97	83.15 ± 0.13
	SRR-DDI	**75.88 ± 0.32**	**83.19 ± 1.13**	**73.24 ± 1.34**	**85.72 ± 0.38**	66.87 ± 1.55	83.32 ± 0.22
S_2	DeepDDI	61.76 ± 3.43	67.84 ± 3.67	61.57 ± 1.15	72.07 ± 3.98	54.13 ± 2.96	71.62 ± 3.54
	GAT-DDI	66.09 ± 0.73	72.67 ± 0.57	**67.15 ± 0.21**	68.75 ± 1.08	59.12 ± 1.54	71.66 ± 0.68
	MR-GNN	63.41 ± 0.11	71.58 ± 0.32	52.97 ± 3.12	72.17 ± 2.86	55.13 ± 1.79	73.28 ± 0.79
	SSI-DDI	65.11 ± 0.56	72.36 ± 0.97	60.08 ± 1.79	76.89 ± 0.57	60.13 ± 0.98	71.06 ± 1.66
	SA-DDI	65.27 ± 0.68	73.43 ± 1.01	63.41 ± 0.84	78.98 ± 1.03	60.27 ± 1.43	73.26 ± 1.22
	DGNN-DDI	66.59 ± 0.34	73.34 ± 1.09	62.11 ± 0.77	78.56 ± 0.89	60.95 ± 0.64	72.32 ± 1.43
	SRR-DDI	**67.16 ± 0.76**	**73.72 ± 1.13**	65.56 ± 0.92	**80.52 ± 0.51**	**62.02 ± 0.67**	**73.35 ± 1.21**

all DDI triplets in the training set consisting of known drugs, and for the test set we adopted two segmentation schemes:

- S_1: The DDI triplet in the test set contains an unknown drug and an known drug for DDI prediction.
- S_2: All the drug pairs in the test set are two unknown drugs for DDI prediction.

Obviously, the cold start task is a much more challenging. Similarly, in order to assess model accuracy and stability, we still conducted three repetitions of SRR-DDI to provide a more comprehensive performance evaluation. In cold start setting, the exponentially decaying learning rate is 1e-2, and the batch is set to 256. In order to prevent overfitting and enable the model to make more accurate predictions on the test set, the number of iterations of the substructure refinement module is set to 2, and the rest of the parameters are consistent with the warm start scenario. The proposed methods are still compared to the above baselines, and the experimental results are summarized in Table 3. Although the performance of our model was not as good as that in the warm start scenario, SRR-DDI was still better than the other methods, especially in the ACC metrics, which was greatly improved in both groups of experiments.

3.7. Visual explanations for SRR-DDI

The internal mechanisms and decision-making processes of deep learning models are more complex and difficult to understand. In order to improve the interpretability of the prediction process of the model, the substructural features of drugs and attention mechanism are introduced to provide interpretable and visual results to increase understanding and trust in model predictions, and to provide interpretable and visual results to increase understanding and trust in model predictions, we have conducted visualization experiments, which highlights important parts that the model focuses on during the prediction process.

To further explore the interpretability of SRR-DDI in the DDI prediction process, we designed contribution assessment experiments for drugs in warm start scenario. Specifically, the input of the model is the DDI triplet (D_x, D_y, r), after obtaining the feature of the drug substructures, we calculated the interaction probability score of the ith substructure in D_x and D_y, which can also be regarded as the importance score of the $i-th$ substructure to the drug. The importance score is helpful in highlighting the key substructures that SRR-DDI focuses on in the process of prediction, thus validating the rationality of the proposed model.

The experimental results are shown in Figs. 9 and 10(a). Each drug undergoes four iterations, layer 0 is the drug substructure features extracted by GNN, layer 1, 2 and 3 are the substructure features enhanced by the substructure refinement module iterations, the importance scores of the substructures in the drug are computed at each layer,

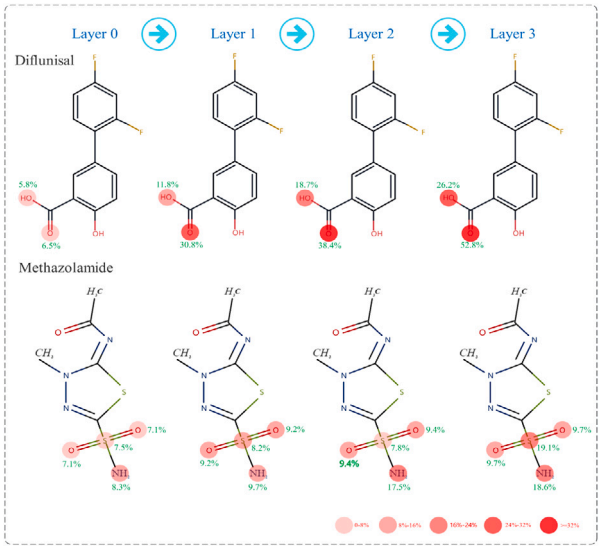


Fig. 9. The Visual Explanations Experiments on Diflunisal and Methazolamide, depicting the variation of substructure importance scores during training. (each red node in the drug represents the extracted substructure centered around that node, and the intensity of the color represents the important scores).

which are used as a visualization process for SRR-DDI work for each drug in the four iterations. Two pairs of drug from DrugBank examples: Diflunisal and Methazolamide, as well as Pentobarbital and Midazolam were used for visualization, as shown in Figs. 9 and 10(a). When DDI occurs, it typically involves chemical reactions between different functional groups within the two drugs. Therefore, co-administration of Diflunisal and Methazolamide may increase the risk of adverse effects. Diflunisal contains a carboxylic acid group (–COOH), which can participate in various chemical reactions such as esterification, acylation, amidation and so on. On the other hand, Methazolamide contains a sulfonamide group ($–SO_2NH_2$), these two compounds can undergo an amidation reaction, where the hydroxyl group (–OH) in the carboxylic acid group reacts with the amino group ($–NH_2$) in the sulfonamide group, forming an amide structure ($–CONH–SO_2NH_2$). From the results in Fig. 9, it could be observed that as the depth of the model increased, the importance scores of the carboxylic acid group and sulfonamide group also increased. It is indicated that our model prominently enhanced its attention towards these vital substructures during the prediction of DDI.

In the other pair of drugs, we demonstrated the effectiveness of our method from a single drug perspective. In Pentobarbital, the amide group (–CONH), composed of a carbonyl group (–C=O) and an amino group (–NH), plays an important role in the chemical reactions. In

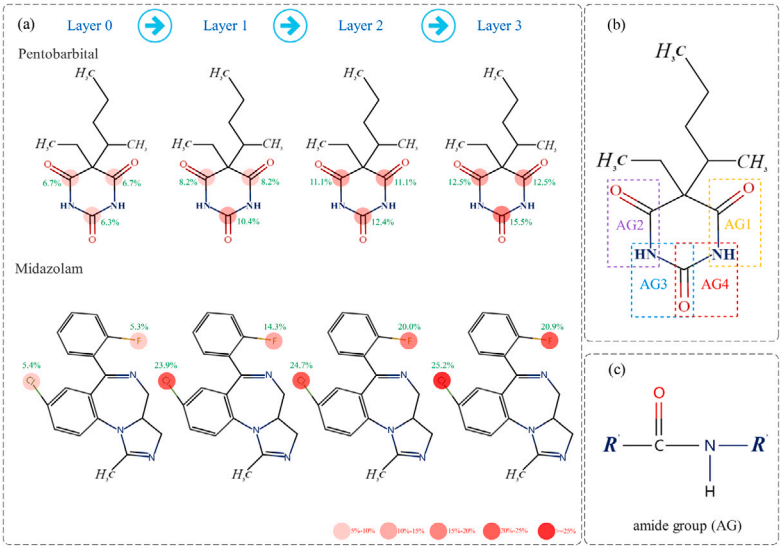


Fig. 10. The Visual Explanations Experiments on Pentobarbital and Midazolam. (each red node in the drug represents the extracted substructure centered around that node, and the intensity of the color represents the important scores). (a) Depicting the variation of substructure importance scores during training. (b) Index of AGs in Pentobarbital. (c) The chemical formular of the AG.

Fig. 10(a), it could be observed that there are four amide groups in the substructures consisting of Pentobarbital, and the importance scores and growth rates of these substructures have been marked. For the sake of description, we have labeled the amide group (we named it as AG) in Pentobarbital in Fig. 10(b) and shown the chemical formula of the amide group in Fig. 10(c). The importance scores of AG1 and AG2 are the same but lower than those of AG3 and AG4. This was because the carbonyl groups in the AG1 and AG2 only forms an amide group with one amino group respectively, while in the AG3 and AG4, the carbonyl group forms an amide group with two separate amino groups. Therefore, the importance score of the substructure with the shared carbonyl group in the AG3 and AG4 was higher than the formers, demonstrating the rationality of our designed substructure importance scores. In Midazolam, the presence of a fluorine atom (–F) and a chlorine atom (–Cl) can have significant effects on the molecular conformation and electronic properties, thereby influencing the interactions of the drug with receptors or other molecules. From the experimental results in Fig. 10(a), it is evident that the importance scores of the two functional groups centered at the fluorine atom and the chlorine atom are the highest, suggesting that SRR-DDI will pay more attention to the substructures that are decisive for the properties of the drug itself through refining the features of substructure. In conclusion, the two experiments we designed from two perspectives of DDI have fully demonstrated the interpretability of the SRR-DDI.

4. Conclusion

In this paper, we propose a model called SRR-DDI for predicting DDI (DDI). The reactions between drugs are usually determined by the nature and interactions of each functional group in the compounds, and we obtain the substructure feature of the drug through GNN, and then the substructure refinement module using graph transformer is proposed to enrich the substructure representation by incorporating the correlation information of other substructures. Compared to the traditional transformer model, we place the normalization layer in the residual layer before the attention and feed forward layers to improve the performance of the model. Moreover, the drug similarity feature is introduced to improve the generalization ability of the model.

We set up two experimental scenarios to test the performance of our model. Both in the warm start and cold start scenario, various evaluation metrics of SRR-DDI are improved compared with the existing stare of art methods. To verify the interpretability of SRR-DDI, we set up a visual experiment to demonstrate that the proposed model could pay more attention to the functional groups contribute more to the DDI, or allocate more attention to groups important to the chemical properties and biological activities. In future, we will pay more attention in cold start scenario to improve the accuracy of the DDI prediction for unknown drugs.

CRediT authorship contribution statement

Dongjiang Niu: Writing – original draft, Conceptualization, Methodology, Writing – review & editing. **Lei Xu:** Software, Data curation. **Shourun Pan:** Software, Visualization. **Leiming Xia:** Data curation, Validation. **Zhen Li:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was fund by National Natural Science Foundation of China (12371491)

References

[1] B. Al-Lazikani, U. Banerji, P. Workman, Combinatorial drug therapy for cancer in the post-genomic era, Nat. Biotechnol. 30 (7) (2012) 679–692, http://dx.doi.org/10.1038/nbt.2284.

[2] N.P. Tatonetti, P.P. Ye, R. Daneshjou, R.B. Altman, Data-driven prediction of drug effects and interactions, Sci. Transl. Med. 4 (125) (2012) 125ra31, http://dx.doi.org/10.1126/scitranslmed.3003377.

[3] Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, S. Liu, A multimodal deep learning framework for predicting drug–drug interaction events, Bioinformatics 36 (15) (2020) 4316–4322, http://dx.doi.org/10.1093/bioinformatics/btaa501.

[4] J. Zhu, Y. Liu, C. Wen, Mtma: Multi-task multi-attribute learning for the prediction of adverse drug–drug interaction, Knowl.-Based Syst. 199 (2020) 105978, http://dx.doi.org/10.1016/j.knosys.2020.105978.

[5] S. Scaboro, B. Portelli, E. Chersoni, E. Santus, G. Serra, Extensive evaluation of transformer-based architectures for adverse drug events extraction, Knowl.-Based Syst. 275 (2023) 110675, http://dx.doi.org/10.1016/j.knosys.2023.110675.
[6] G. Lee, C. Park, J. Ahn, Novel deep learning model for more accurate prediction of drug-drug interaction effects, BMC Bioinf. 20 (2019) 1–8, http://dx.doi.org/10.1186/s12859-019-3013-0.
[7] K.H. Hoang, T.B. Ho, Learning and recommending treatments using electronic medical records, Knowl.-Based Syst. 181 (2019) 104788, http://dx.doi.org/10.1016/j.knosys.2019.05.031.
[8] Z. Li, M. Jiang, S. Wang, S. Zhang, Deep learning methods for molecular representation and property prediction, Drug Discov. Today 27 (12) (2022) 103373, http://dx.doi.org/10.1016/j.drudis.2022.103373.
[9] C. Park, J. Park, S. Park, Agcn: Attention-based graph convolutional networks for drug-drug interaction extraction, Expert. Syst. Appl. 159 (2020) 113538, http://dx.doi.org/10.1016/j.eswa.2020.113538.
[10] J.D. Duke, X. Han, Z. Wang, A. Subhadarshini, S.D. Karnik, X. Li, S.D. Hall, Y. Jin, J.T. Callaghan, M.J. Overhage, et al., Literature based drug interaction prediction with clinical assessment using electronic medical records: Novel myopathy associated drug interactions, PLoS Comput. Biol. (2012) http://dx.doi.org/10.1371/journal.pcbi.1002614.
[11] Y. Shi, P. Quan, T. Zhang, L. Niu, Dream: Drug-drug interaction extraction with enhanced dependency graph and attention mechanism, Methods 203 (2022) 152–159, http://dx.doi.org/10.1016/j.ymeth.2022.02.002.
[12] M. Schuster, K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681, http://dx.doi.org/10.1109/78.650093.
[13] A. Fokoue, M. Sadoghi, O. Hassanzadeh, P. Zhang, Predicting drug-drug interactions through large-scale similarity-based link prediction, in: H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S.P. Ponzetto, C. Lange (Eds.), The Semantic Web, in: Latest Advances and New Domains, Springer International Publishing, Cham, 2016, pp. 774–789, http://dx.doi.org/10.1007/978-3-319-34129-3_47.
[14] D. Weininger, Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1) (1988) 31–36, http://dx.doi.org/10.1021/ci00057a005.
[15] G.B. Goh, N.O. Hodas, C. Siegel, A. Vishnu, Smiles2vec: An interpretable general-purpose deep neural network for predicting chemical properties, 2017, http://dx.doi.org/10.48550/arXiv.1712.02034, arXiv preprint arXiv:1712.02034.
[16] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, http://dx.doi.org/10.1109/5.726791.
[17] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J.B. Ghasemi, A. Masoudi-Nejad, DeepCDA: Deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks, Bioinformatics 36 (17) (2020) 4633–4642, http://dx.doi.org/10.1093/bioinformatics/btaa544.
[18] T.N.K. Hung, N.Q.K. Le, N.H. Le, L. Van Tuan, T.P. Nguyen, C. Thi, J.-H. Kang, An ai-based prediction model for drug-drug interactions in osteoporosis and paget's diseases from smiles, Mol. Inf. 41 (6) (2022) 2100264, http://dx.doi.org/10.1002/minf.202100264.
[19] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Trans. Neural Netw. 20 (1) (2009) 61–80, http://dx.doi.org/10.1109/TNN.2008.2005605.
[20] A. Rassil, H. Chougrad, H. Zouaki, Holistic graph neural networks based on a global-based attention mechanism, Knowl.-Based Syst. 240 (2022) 108105, http://dx.doi.org/10.1016/j.knosys.2021.108105.
[21] S. Liu, R. Ying, H. Dong, L. Li, T. Xu, Y. Rong, P. Zhao, J. Huang, D. Wu, Local augmentation for graph neural networks, in: K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, S. Sabato (Eds.), Proceedings of the 39th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 162, PMLR, 2022, pp. 14054–14072.
[22] X. Fang, L. Liu, J. Lei, D. He, S. Zhang, J. Zhou, F. Wang, H. Wu, H. Wang, Geometry-enhanced molecular representation learning for property prediction, Nat Mach Intell. 4 (2) (2022) 127–134, http://dx.doi.org/10.1038/s42256-021-00438-4.
[23] M. Yazdani-Jahromi, N. Yousefi, A. Tayebi, E. Kolanthai, C.J. Neal, S. Seal, O.O. Garibay, AttentionSiteDTI: An interpretable graph-based model for drug-target interaction prediction using NLP sentence-level relation classification, Briefings Bioinf. 23 (4) (2022) bbac272, http://dx.doi.org/10.1093/bib/bbac272.
[24] K. Huang, C. Xiao, T. Hoang, L. Glass, J. Sun, Caster: Predicting drug interactions with chemical substructure representation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 702–709, http://dx.doi.org/10.1609/aaai.v34i01.5412.
[25] N. Xu, P. Wang, L. Chen, J. Tao, J. Zhao, Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions, 2019, http://dx.doi.org/10.48550/arXiv.1905.09558, arXiv preprint arXiv:1905.09558.
[26] A.K. Nyamabo, H. Yu, J.-Y. Shi, SSI–DDI: Substructure–substructure interactions for drug–drug interaction prediction, Briefings Bioinf. 22 (6) (2021) bbab133, http://dx.doi.org/10.1093/bib/bbab133.
[27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, http://dx.doi.org/10.48550/arXiv.1710.10903, arXiv preprint arXiv:1710.10903.
[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017.
[29] S. Lin, Y. Wang, L. Zhang, Y. Chu, Y. Liu, Y. Fang, M. Jiang, Q. Wang, B. Zhao, Y. Xiong, D.-Q. Wei, MDF-SA-DDI: Predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism, Briefings Bioinf. 23 (1) (2021) bbab421, http://dx.doi.org/10.1093/bib/bbab421.
[30] P. Li, J. Wang, Y. Qiao, H. Chen, Y. Yu, X. Yao, P. Gao, G. Xie, S. Song, An effective self-supervised framework for learning expressive molecular global representations to drug discovery, Briefings Bioinf. 22 (6) (2021) bbab109, http://dx.doi.org/10.1093/bib/bbab109.
[31] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder–decoder for statistical machine translation, 2014, http://dx.doi.org/10.48550/arXiv.1406.1078, arXiv preprint arXiv:1406.1078.
[32] S. Yun, M. Jeong, R. Kim, J. Kang, H.J. Kim, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, R. Garnett (Eds.), Graph transformer networks, in: Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc., 2019.
[33] D. Arpit, Y. Zhou, B. Kota, V. Govindaraju, Normalization propagation: A parametric technique for removing internal covariate shift in deep networks, in: M.F. Balcan, K.Q. Weinberger (Eds.), Proceedings of the 33rd International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 48, PMLR, New York, New York, USA, 2016, pp. 1168–1176.
[34] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, 2016, http://dx.doi.org/10.48550/arXiv.1607.06450, arXiv preprint arXiv:1607.06450.
[35] R. Xiong, Y. Yang, D. He, K. Zheng, S. Zheng, C. Xing, H. Zhang, Y. Lan, L. Wang, T. Liu, On layer normalization in the transformer architecture, in: H. D. III, A. Singh (Ed.), Proceedings of the 37th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 119, PMLR, 2020, pp. 10524–10533.
[36] P. Willett, Similarity-based virtual screening using 2d fingerprints, Drug Discov. Today 11 (23) (2006) 1046–1053, http://dx.doi.org/10.1016/j.drudis.2006.10.005.
[37] V.P. Dwivedi, X. Bresson, A generalization of transformer networks to graphs, 2020, http://dx.doi.org/10.48550/arXiv.2012.09699, arXiv preprint arXiv:2012.09699.
[38] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: A comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res. 34 (suppl_1) (2006) D668–D672, http://dx.doi.org/10.1093/nar/gkj067.
[39] M. Zitnik, M. Agrawal, J. Leskovec, Modeling polypharmacy side effects with graph convolutional networks, Bioinformatics 34 (13) (2018) i457–i466, http://dx.doi.org/10.1093/bioinformatics/bty294.
[40] G. Landrum, et al., Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling, Greg Landrum 8 (2013) 31.
[41] Z. Yang, W. Zhong, Q. Lv, C.Y.-C. Chen, Learning size-adaptive molecular substructures for explainable drug–drug interaction prediction by substructure-aware graph neural network, Chem. Sci. 13 (29) (2022) 8693–8703, http://dx.doi.org/10.1039/D2SC02023H.
[42] J. Peng, J. Li, X. Shang, A learning-based method for drug-target interaction prediction based on feature representation learning and deep neural network, BMC Bioinf. 21 (13) (2020) 1–13, http://dx.doi.org/10.1186/s12859-020-03677-1.
[43] D. Rogers, M. Hahn, Extended-connectivity fingerprints, J. Chem. Inf. Model. 50 (5) (2010) 742–754, http://dx.doi.org/10.1021/ci100050t.
[44] J.Y. Ryu, H.U. Kim, S.Y. Lee, Deep learning improves prediction of drug–drug and drug–food interactions, Proc. Natl. Acad. Sci. 115 (18) (2018) E4304–E4311, http://dx.doi.org/10.1073/pnas.180329411.
[45] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, http://dx.doi.org/10.48550/arXiv.1710.10903, arXiv preprint arXiv:1710.10903.
[46] M Ma, X. Lei, A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions, PLoS Comput. Biol. 19 (1) (2023) e1010812, http://dx.doi.org/10.1371/journal.pcbi.1010812.
[47] Z. Wang, J. Zhang, J. Feng, Z. Chen, Knowledge graph embedding by translating on hyperplanes, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28, 2014, http://dx.doi.org/10.1609/aaai.v28i1.8870.
[48] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015, pp. 448–456.