

MRGCDDI: Multi-Relation Graph Contrastive Learning without Data Augmentation for Drug-Drug Interaction Events Prediction

Yu Li, Lin-Xuan Hou, Zhu-Hong You, Yang Yuan, Cheng-Gang Mi, Yu-an Huang and Hai-Cheng Yi

Abstract— Predicting drug-drug interactions (DDIs) is a significant concern in the field of deep learning. It can effectively reduce potential adverse consequences and improve therapeutic safety. Graph neural network (GNN)-based models have made satisfactory progress in DDI event prediction. However, most existing models overlook crucial drug structure and interaction information, which is necessary for accurate DDI event prediction. To tackle this issue, we introduce a new method called MRGCDDI. This approach employs contrastive learning, but unlike conventional methods, it does not require data augmentation, thereby avoiding additional noise. MRGCDDI maintains the semantics of the graphical data during encoder perturbation through a simple yet effective contrastive learning approach, without the need for manual trial and error, tedious searching, or expensive domain knowledge to select enhancements. The approach presented in this study effectively integrates drug features extracted from drug molecular graphs and information from multi-relational drug-drug interaction (DDI) networks. Extensive experimental results demonstrate that MRGCDDI outperforms state-of-the-art methods on both datasets. Specifically, on Deng's dataset, MRGCDDI achieves an average increase of 4.33% in accuracy, 11.57% in Macro-F1, 10.97% in Macro-Recall, and 10.64% in Macro-Precision. Similarly, on Ryu's dataset, the model shows improvements with an average increase of 2.42% in accuracy, 3.86% in Macro-F1, 3.49% in Macro-Recall, and 2.75% in Macro-Precision. All the data and codes of this work are available at <https://github.com/Nokeli/MRGCDI>.

Index Terms— Graph Contrastive Learning, Drug-Drug

This work is supported in part by the Nature Science Fund for Distinguished Young Scholars of China under Grant. 62325308.

Yu Li is with School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: li.yu@nwpu.edu.cn).

Lin-Xuan Hou is with School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: houlinxuan@mail.nwpu.edu.cn).

Zhu-Hong You is with School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: zhuhongyou@nwpu.edu.cn).

Yang Yuan is with School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213164, China (e-mail: yuanyang@cczu.edu.cn).

Cheng-Gang Mi is with Foreign Language and Literature Institute, Xi'an International Studies University, Xi'an 710129, China (michenggang123@126.com).

Yu-An Huang is with School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China (yuanhuang@nwpu.edu.cn).

Cheng-Gang Mi is with School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China (yihaicheng@nwpu.edu.cn).

(Corresponding authors: Zhu-Hong You)

interaction, Heterogeneous Graph

I. INTRODUCTION

THIS Drug combination therapy, which involves using two or more drugs to treat a disease, is a well-established method in modern medicine. However, it may lead to drug-drug interactions (DDIs), which are classified into pharmaceutical interactions, pharmacokinetic interactions and pharmacodynamic interactions [1]. DDIs can impact the effectiveness of drug combinations and may result in unintended adverse effects, such as decreased drug efficacy or worsening of the condition [2]. Therefore, the study of DDIs holds significant medical value as it can help reduce morbidity and mortality associated with clinical treatments [3] and provide guidance for adjustments in drug development [4].

Due to the cost and complexity of clinical trials, several computational algorithms have been developed and applied to predict potential DDIs. Chemical structure-based methods can be divided into similarity-based methods and substructure-based methods. Based on assumption that similar drugs may interact with the same drugs, many similarity-based methods have emerged to predict DDIs [2, 5-9]. For instance, S. Vilar et al. [2, 8, 9] predicted DDIs by utilizing drug similarity information based on 2D and 3D molecular structures, interaction profiles, targets, and side effects. Given that a drug can be effectively dissected into various functional groups or chemical substructures, collectively influencing the overall pharmacological properties, several studies have leveraged the substructure of drugs for the prediction of DDIs [10-14]. For instance, A. K. Nyamabo et al. [11] proposed the substructure–substructure interaction–drug–drug interaction (SSI-DDI) method, which extracts drug molecular structure features directly from the original molecular map of the drug to predict DDIs. Despite the prevalence and demonstrated efficacy of chemical structure-based methods over time, a notable disadvantage lies in its substantial reliance on the manual curation of features and the application of specialized domain knowledge. To solve this challenge, a variety of network-based approaches have surfaced in recent years, including matrix factorization [15-20], random walk [21]. Common network-based methods acquired a limited of features, resulting in a lack of advantages in classification tasks. Consequently, deep learning-based models have emerged. Deep learning-

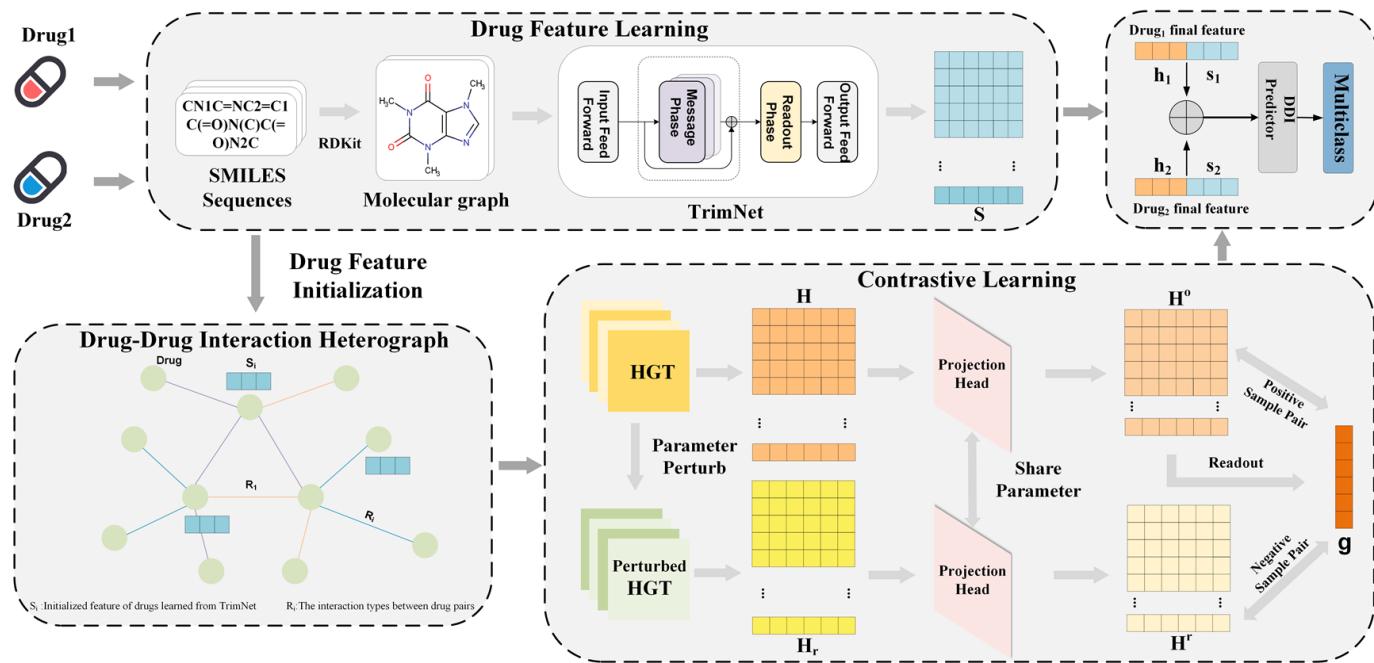


Fig. 1. The workflow of the proposed MRGCDDI method

based methods are widely used for predicting DDIs due to their ability to capture nonlinear relationships and perform end-to-end learning within the network. The models include deep neural network-based models, knowledge graph embedding-based models, graph contrastive learning-based models [22-35] and Hybrid method [46-47,52-53]. Su et al. [28] developed an attention-based DDIs prediction model that considers both drug properties and triple facts in the knowledge graph (KG). Yu et al. [52] develops a two-GAN architecture to capture the knowledge between drug attributes and the topological structure of the DDI network, thereby predicting DDI event types between drug pairs. Su et al. [53] employs the Transformer architecture with a relation-aware self-attention mechanism and integrates representation learning from both drug molecular graphs and knowledge graphs to predict DDIs. SumGNN [55] conduct a graph summarization module based on subgraph in knowledge graph to predict DDIs. Su et al [56] use capsule network to encode multi-relation DDI date from biomedical KGs. Wang et al. [30] proposed MIRACLE, a multi-view graph contrastive representation learning method that captures both inter-view molecular structures and interactions between molecules within views to predict DDIs. Zhang et al. [31] implement a multi-relational contrastive learning with a designed dual-view negative counterpart augmentation strategy to capture implicit information about rare DDI events. There are still have some limitations. (1) most methods can only predict whether drug pairs can interact or not as a binary classification task, rather than the more complex task of predicting specific DDI events, which represents a challenging and significant multi-class classification task [31]. Secondly, some DDI events lack sufficient labeled sample instances, and some existing graph neural network-based methods struggle to effectively learn features from a small number of labeled sample instances by comparing similarities and differences

between them. Finally, using the graph data augmentation method to capture implicit information of multi-relation DDI events may alter the graph semantics completely even if the perturbation is weak.

To address the limitations of the above methods, we propose a novel DDI events prediction method named MRGCDDI. The workflow of MRGCDDI is shown in Figure 1. Firstly, TrimNet [36] is utilized to obtain the structural features of the drug molecular graph, serving as the features for the drug nodes in the DDI events graph. The various relationships between drug pairs are employed as edges in the DDI events graph. Subsequently, leveraging the relationships between drug pairs and the structural features of drug nodes, Heterogeneous Graph Transformer (HGT) [37] is employed to learn drug representations. To further enhance the prediction accuracy of DDI events with limited labeled, we use SimGRACE [38], a graph contrastive learning method without data augmentation, where use HGT model with its perturbed version as two encoders to obtain two different correlated views and employ contrastive learning on these views to captures semantic features of drug nodes from the DDI events graph. Finally, the prediction of interaction events between drug pairs is performed using a Multi-layer Perceptron (MLP). To evaluate the validity of MRGCDDI more effectively, five-fold cross-validation is applied and the result demonstrate that MRGCDDI outperforms the baseline method.

In summary, the main contributions of this paper are as follows:

- 1) The effectiveness of the multi-class classification prediction for DDIs event is improved by combining the structural information of drugs and the interaction information between drug pairs contained in the DDI event graph.
- 2) A graph contrastive learning without data augmentation is employed, which reduce influence of semantics alter

by the general augmentations method and enhances the preservation of semantic features inherent in both drug molecule graphs and DDI event graph.

- 3) We compare MRGCDDI with baseline methods and variants of our model for ablation study on two datasets. In addition, we perform hyper-parameter sensitivity analysis experiments. Extensive experimental results demonstrate the superior performance of MRGCDDI.

II. MATERIAL AND METHODS

A. Problem formulation

In this study, the drug set is denoted as $D = d_1, d_2, \dots, d_N$, and its associated molecular structure graph is denoted as $G_{drug} = \{g_1, g_2, \dots, g_N\}$, where each graph g consists of atoms as nodes and bonds as edges, and N represents the total number of drug entities. The drug pairs and the interaction events relationships between them are represented by G_{event} , which consists of a set of triple facts denoted as $G_{event} = \{(d_i, r, d_j) | d_i, d_j \in D, r \in R\}$, where R is the set of relationships of interaction events between different drug pairs. For multiclass classification of predicting the DDI events, the classification task is modeled as $d_i \times d_j \rightarrow \hat{y}_{(i,j)} \in R$. Our goal is to construct a model to implement this task. The main notations of this paper and their corresponding descriptions are listed in TABLE 1. We evaluate our method MRGCDDI on two datasets: (1) Deng's dataset [39] contains a total of 37,264 DDIs between 570 drugs with 65 types of DDI events. (2) Ryu's dataset [3] contains a total of 191,570 DDIs between 1,700 drugs with 86 types of DDI events. In the above two datasets, there exists only one type of interaction between drug pairs, specifically describing how one drug affects the metabolism of another drug. We conduct a five-fold cross-validation on these datasets, dividing the data into training, validation, and test sets in a 7:2:1 ratio.

B. Drug representation learning

To ensure a clear representation for downstream prediction tasks and to emphasize the critical atoms in a drug molecule graph for target characterization, we use TrimNet [36]. This method employs a triple information mechanism to efficiently learn molecular representations and extract representations of drug molecules. TrimNet is a variant of the message-passing neural network (MPNN) [40], which is a generalized graph neural network capable of learning features directly from molecular graphs. It contains two common phases: a message passing phase and a readout phase.

Specifically, in the message passing phase, the message vector m_v^{t+1} of an atom node v in the drug molecule graph is generated by propagating atom or bond features from neighboring nodes. The features of the embedded atom v are subsequently updated based on the message vector. The message passing phase runs for T time steps and contains the message function M_t and the node update function U_t . In each time step, the hidden state h_v^t of the node is updated to the new hidden state h_v^{t+1} based on the message vector m_v^{t+1} ,

which is expressed as:

$$m_v^{t+1} = \sum_{\omega \in N(v)} M_t(h_v^t, h_\omega^t, e_{v\omega}^t) \quad (1)$$

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1}) \quad (2)$$

where $N(v)$ denotes all the neighbors of an atom node v in the drug molecular graph, $e_{v\omega}^t$ denotes the hidden state of the connecting edges between atom node u and v , and the message function M_v is defined by the multi-head attention mechanism:

$$M_t = \sum_{k=1}^K \alpha_{v,\omega}^{t+1,k} \odot \mathbf{W}_h^k h_\omega^t \odot \mathbf{W}_e^k e_{v\omega}^t \quad (3)$$

$$p_{v,\omega}^{t+1,k} = \text{LeakyReLU}(\mathbf{u}^T [\mathbf{W}_h h_v^t || \mathbf{W}_e e_{v\omega}^t || \mathbf{W}_h h_\omega^t]) \quad (4)$$

$$\alpha_{v,\omega}^{t+1} = \text{softmax}(p_{v,\omega}^{t+1,k}) \quad (5)$$

where $||$ denotes the vector join operation, \odot is the element-wise product, softmax and LeakyReLU are both nonlinear functions. The scalar $\alpha_{v,\omega}^{t+1}$ represents the attention of the different nodes, the vector u is the learnable weights, and W_h and W_e are the shared learnable weight matrices. The node update function U_t is a gated recurrent unit [41], which is subsequently layer-normalized for aggregating information about neighbors. The representation is as follows:

$$h_v^{t+1} = LN(GRU(h_v^t, m_v^{t+1})) \quad (6)$$

The above steps are iterated T times during the message passing phase to produce the final representation h_v^t for each node v .

In the readout phase, TrimNet employs Set2Set [42] as the readout function to generate a graph-level representation. Specifically, Set2Set aggregates node features using various attention weights and concatenates them with historical information:

$$d_t = LSTM(d_{t-1}^*) \quad (7)$$

$$\lambda_{i,t} = \text{softmax}(h_v^T, d_t) \quad (8)$$

$$r_t = \sum_{i=1}^N \lambda_{i,t} h_v^T \quad (9)$$

$$d_t^* = d_t || r_t \quad (10)$$

where d_t is a query vector that can read r_t from memory, and r_t denotes the attention readout result. The Set2Set process is iteratively executed T times to derive the ultimate representation of the drug molecule map d_T^* , which is redefined as s_V . The features of all the drugs are represented as S .

C. Drug-drug Interaction Event Graph Learning

Heterogeneous Graph Transformer (HGT) is specifically developed to address highly multi-relational data found in practical knowledge bases [37]. The workflow of HGT is shown in Figure 2. We adopt the drug features S learned through TrimNet as node attributes to construct the DDI event graph G_{event} . Subsequently, HGT encoder is utilized to learn the representations of different drug nodes from G_{event} .

Notations	Descriptions
D	Drug set
G_{drug}	Drug molecular structure graph
G_{event}	The drug pairs and the interaction events relationships between them
R	The interaction events relationships between drug pairs
(d_i, r, d_j)	Triple with drug pairs and the interaction events relationships between them
m_v^{t+1}	The message vector
M_t	The message function
U_t	The node update function
$N(v)$	All the neighbors of atom node
S	The initialized features of all drugs learned from TrimNet
s, t	Source node and target node
$\langle s, \phi(r), t \rangle$	The meta relation from source node to target node
$W_{\phi(r)}^{ATT}, W_{\phi(r)}^{MSG}$	The edge-based matrix
d_{scale}	The scaling factor
$K^i(d_i), Q^i(d_j)$	The linear projections
$\sim^{(l)}$	
$H_{HGT}^{[d_j]}$	The aggregated information from the various feature distributions of all the Neighbors (source nodes) of the target node
$H_{HGT}^{L[d_j]}$	The final representation of each source drug node learned from HGT
H, H_r	The Graph-level representations by HGT and its perturbed counterpart
δ_l, δ'_l	The weight tensor for the layer on the HGT encoder
μ	The coefficient measuring the magnitude of the perturbation
$\Delta\delta_l$	The perturbation term sampled from a Gaussian distribution
H^o, H^r	The representations and in the potential space
$K_{(i,j)}$	The combined representations of drugs
$y_{(i,j)}^r$	The true relationship type in which the drug pair interacts
$\hat{y}_{\{i,j\}}$	The predicted probability that the drug pair belongs to certain relationship type

TABLE I
NOTATIONS AND DESCRIPTIONS

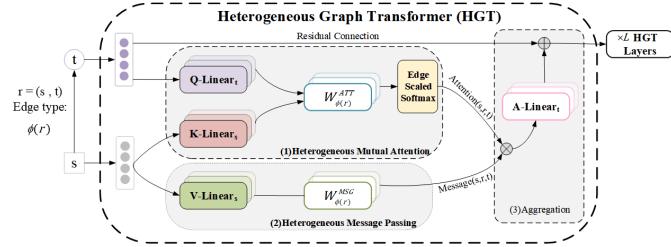


Fig. 2. In this model, "s" denotes source node d_i , while "t" represents target node d_j . The model takes both the edge $r = (s, t)$ and its corresponding meta relation $\langle s, \phi(r), t \rangle$ as input. HGT comprises three phases, (1) Heterogeneous Mutual Attention phase; (2) Heterogeneous Message Passing phase; and (3) Aggregation phase.

Specifically, HGT consists of three phases, (1) Heterogeneous mutual attention phase; (2) Heterogeneous message passing phase; and (3) Aggregation phase.

Heterogeneous mutual attention phase focuses on estimating the importance of each source node. For each node pair $r = (d_i, d_j)$, the corresponding meta relation is $\langle d_i, \phi(r), d_j \rangle$. When a source node d_i and all its neighbors $d_j \in N(d_i)$, their mutual attention is computed through their meta relationship. Specifically, the target node d_j is mapped to a Query vector while the source node d_i is mapped to a Key vector, and their dot product is computed as attention. The calculations are shown as follows:

$$\begin{aligned} & \text{Attention}(d_i, r, d_j) \\ &= \text{Softmax}_{\forall d_i \in N(d_j)} \left(\parallel_{i \in [1, h]} \text{ATT} - \text{head}^i(d_i, r, d_j) \right) \end{aligned} \quad (11)$$

$$\text{ATT-head}^i(d_i, r, d_j) = \left(K^i(d_i) W_{\phi(r)}^{\text{ATT}} Q^i(d_j) \right) \bullet \frac{1}{\sqrt{d_{\text{scale}}}} \quad (12)$$

$$K^i(d_i) = K - \text{Linear}^i \left(H_{HGT}^{(l-1)}[d_i] \right) \quad (13)$$

$$Q^i(d_j) = Q - \text{Linear}^i \left(H_{HGT}^{(l-1)}[d_j] \right) \quad (14)$$

where h denotes the number of attention heads in the multi-head attention mechanism, the edge-based matrix $W_{\phi(r)}^{\text{ATT}} \in \mathbb{R}^{\frac{d}{h} \times \frac{d}{h}}$ is specific to each different edge type $\phi(r)$. $H_{HGT}^{(0)} = S$ signifies the initial features of all drugs. d_{scale} is a scaling factor. $K^i(d_i)$ and $Q^i(d_j)$ represent linear projections, $K - \text{Linear}^i : \mathbb{R}^d \rightarrow \mathbb{R}^{\frac{d}{h}}$ and $Q - \text{Linear}^i : \mathbb{R}^d \rightarrow \mathbb{R}^{\frac{d}{h}}$, respectively, projecting source node d_i to the $i - th$ Key vector and target node d_j to the $i - th$ Query vector. \parallel denotes concatenating h attention heads to obtain the attention coefficients between the target node d_j and all the neighboring nodes $d_i \in N(d_j)$, followed by normalization using Softmax to obtain the $\text{Attention}(d_i, r, d_j)$ for each node pair.

Heterogeneous message passing phase relies solely on the neighborhood feature d_i of the target node d_j to extract the message. The formula for the multi-head message is as follows:

$$\text{Message}(d_i, r, d_j) = \parallel_{i \in [1, h]} \text{MSG} - \text{head}^i(d_i, r, d_j) \quad (15)$$

$$\text{SG-head}^i(d_i, r, d_j) = V - \text{Linear}^i \left(H_{HGT}^{(l-1)}[d_i] \right) W_{\phi(r)}^{\text{MSG}} \quad (16)$$

where $V - \text{Linear}^i : \mathbb{R}^d \rightarrow \mathbb{R}^{\frac{d}{h}}$ represents a linear projection that maps the source node d_j to the $i - th$ message vector. Following that, $W_{\phi(r)}^{\text{MSG}} \in \mathbb{R}^{\frac{d}{h} \times \frac{d}{h}}$ denotes a matrix that

incorporates edge dependencies. Finally, all h message heads are aggregated to obtain the $Message(d_i, r, d_j)$ for each node pair.

Aggregation phase is to aggregate the neighbor information based on the Attention weight. Following the computation of $Attentopn(d_i, r, d_j)$ and $Message(d_i, r, d_j)$ for each drug node pair, aggregating neighborhood information to the target node becomes essential. Specifically, $Attentopn(d_i, r, d_j)$ serves as the weight, and the corresponding $Message(d_i, r, d_j)$ from the source node is averaged to derive the updated vector as follows:

$$\tilde{H}_{HGT}^{(l)}[d_j] = \text{Mean}_{d_i \in N(d_j)} (\text{Attention}(d_i, r, d_j) \bullet \text{Message}(d_i, r, d_j)) \quad (17)$$

$$H_{HGT}^{(l)}[d_j] = A - \text{Linear} \left(\sigma \left(\tilde{H}_{HGT}^{(l)}[d_j] \right) \right) + H_{HGT}^{(l-1)}[d_j] \quad (18)$$

where $\tilde{H}_{HGT}^{(l)}[d_j]$ represents the aggregated information from the various feature distributions of all the neighbors (source nodes) of the target node d_j . Then a linear projection A - $\tilde{H}_{HGT}^{(l)}$ is applied to the updated vector $H_{HGT}^{(l)}[d_j]$ along with residual concatenation [43] yields the l -th layer HGT output of the target node d_j , denoted as $H_{HGT}^{(l)}[d_j]$. The entire HGT framework comprises a total of L layers, where the outputs of the previous layer serve as inputs for the subsequent layer, iteratively until the last layer. Finally, the final representation $H_{HGT}^L[d_j]$ of each source drug node d_j is obtained.

D. Multi-Relational Contrastive Learning

Contrastive learning is a self-supervised learning method that aims to learn valuable data representations by comparing similarities and differences between pairs of samples. Traditional graph contrastive learning methods generate positive and negative sample pairs by graph augmentation (dropping nodes, dropping edges, etc.) [44]. However, this approach corrupts the semantics of the original graph data, which reduces the quality of the representations. In light of these limitations, we adopt a graph contrastive learning framework SimGRACE [38] that does not use graph augmentation methods. SimGRACE employs the GNN model and its perturbed counterpart as dual encoders, thereby extracting two relevant views for comparison. Through the process of encoder perturbation, the semantic information within the graph data is effectively retained.

Specifically, SimGRACE consists of the following two main components:

- (1) **Encoder perturbation module.** First, the multi-relational DDI event graph G_{event} is constructed. Subsequently, the DDI event graph is encoded using HGT and a perturbed counterpart of HGT. This process generates two graph-level representations, H and H_r , respectively, which can be expressed as:

$$H = f(G_{event}; \delta), \quad H_r = f(G_{event}; \delta') \quad (19)$$

where f denotes the HGT, the relationship between the weight tensor δ_l and δ'_l for the l -th layer on the HGT encoder and the l -th layer on its perturbed version is expressed as follows:

$$\delta'_l = \delta_l + \mu \bullet \Delta\delta_l; \quad \Delta\delta_l \sim \mathcal{N}(0, \sigma_l^2) \quad (20)$$

where μ is a coefficient measuring the magnitude of the perturbation and $\Delta\delta_l$ is a perturbation term sampled from a Gaussian distribution with zero mean and variance σ_l^2 . In contrast to approaches that focus only with node-level representation learning, SimGRACE concentrates on graph-level representation learning. According to [45], mapping the representation to another potential space using a nonlinear transformation can improve the performance, referred to as the projection head. In this study, a two-layer nonlinear variation of the MLP is used to further acquire the representations H^o and H^r of H and H_r in the potential space, expressed as follows:

$$H^o = p(H), \quad H^r = p(H_r) \quad (21)$$

- (2) **Contrastive loss.** Here, the normalized temperature-scaled cross entropy loss (NT-Xent) [45] is used to make the representations closer between pairs of positive samples and more different between pairs of negative samples.

Specifically, during SimGRACE training, the original DDI event graph is input into both the HGT encoder and its perturbed counterpart, obtaining different representations of the DDI event graphs denoted as H^o and H^r . Utilizing a readout function, denoted as Λ , a global representation of $g \in \mathbb{R}^Q$ is learned from H^o , expressed as $g = \Lambda(H^o)$. The positive sample pairs are defined as g and H^o , while negative sample pairs are defined as g and H^r . The contrastive loss is formulated as follows:

$$L_{gcl} = -\frac{1}{|D| + |D|} \left(\sum_{a \in D} \mathbb{E}_{(D, R, HGT)} [\log F(h_a^o, g)] + \sum_{b \in D} \mathbb{E}_{(D, R, HGT')} [\log (1 - F(h_b^r, g))] \right) \quad (22)$$

where HGT denotes the HGT encoder, HGT' denotes the perturbed counterpart of the HGT encoder, $F(h_a^o, g) = \sigma(h_a^{oT} W g)$ and W is a trainable parameter matrix.

E. DDI Event Prediction

For each drug pair (d_i, d_j) in the DDI event graph, we get the drug features s_i and s_j using TrimNet, and their two final representations h_1 and h_2 . The concatenating method is often utilized to incorporate semantic information from feature maps of varying scales, leading to improve performance through the addition of dimensions. Thus, we employ the concatenation method to combine these representations, resulting in $K_{(i,j)} = concat(s_i, s_j, h_i, h_j) = (s_i, s_j, h_i, h_j)$. Finally, $K_{(i,j)}$ is fed into the multilayer perceptron, and the softmax function is applied to obtain the multi-class classification prediction results for different drug pairs:

$$\hat{y}_{(i,j)} = \text{Softmax} (MLP(K_{(i,j)})) \quad (23)$$

Datasets	Five groups				
	[1,10]	(10,50]	(50,100]	(100,300]	(300, +∞]
Deng's dataset	20.0%	21.5%	24.6%	15.4%	18.5%
Ryu's dataset	5.8%	21.0%	11.6%	14.0%	47.6%

TABLE II

THE RATIOS OF EVENT OCCURRENCES WITHIN FIVE GROUPS TO ALL EVENT OCCURRENCES.

where $\hat{y}_{(i,j)} \in \mathbb{R}^{|R|}$ and the training objective for predicting DDI events is to minimize the loss function:

$$L_{cf} = - \sum_{(i,j) \in Q} \sum_{r \in R} y_{(i,j)}^r \log \hat{y}_{(i,j)}^r \quad (24)$$

where Q is the training set, $\hat{y}_{(i,j)}^r$ denotes the predicted probability that the drug pair (d_i, d_j) belongs to relationship type r , and $y_{(i,j)}^r$ denotes the true relationship type in which the drug pair (d_i, d_j) interacts.

To train our proposed model, we optimized the combination of L_{gcl} and L_{cf} as follows:

$$L_{total} = \alpha L_{gcl} + \beta L_{cf} \quad (25)$$

where α and β are hyper-parameters utilized to balance contributions from different tasks, with β set to 1. They are set off from the text with rules above and below the title and after the last line.

III. EXPERIMENT AND RESULT

A. Experiment Setting

We conduct five-fold cross-validation to train MRGCDDI and the aforementioned baselines on two datasets. The Adam algorithm with a learning rate of 1e-3 is adopted to optimize all trainable parameters. The max epoch is set to 100, the dimensions of feature is 128 and the batch size is 256, the dropout is 0.2. The hyper-parameters α and μ is selected by grid search strategy. For Deng's dataset, α is 0.15 and μ is 0.3. For Ryu's dataset, α is 0.15 and μ is 0.2. For all baselines, In all baselines, the same hyperparameters were used as reported in their original work.

B. Baselines

The MRGCDDI is compared with a variety of baselines, including similar-based method (DeepDDI [3]), molecular graph based method (SSI-DDI [11], Trimnet-DDI [36],DGNN-DDI [51]), Network-based method (R-GCN [48], R-GAT [49],GNN-FILM [50], MRCGNN [31]), knowledge graph-based method (LaGAT [23]) and Hybrid methods (MUFFIN [46], GoGNN[47], DGANDDI[52], TIGER [53]).

DeepDDI [3]: This approach employs drug names and chemical structures as inputs, leveraging drug structural similarity and deep neural networks to predict the types of DDI events between drug pairs.

SSI-DDI [11]: Utilizing a shared attention mechanism, SSI-DDI assigns relevance scores to drug substructure embeddings obtained through multiple layers of Graph Attention Networks (GAT), thereby predicting different types of DDI events between drug pairs.

TrimNet-DDI [36]: Employing a novel triple information mechanism, TrimNet-DDI efficiently learns representations of drug molecules, subsequently inputting concatenated representations of drug pairs into MLP for DDI event prediction.

MUFFIN [46]: MUFFIN is a multi-scale feature fusion model that learns drug representations based on drug chemical structures and biomedical knowledge graphs to predict the type of DDI events between drug pairs.

GoGNN [47]: GoGNN uses a dual-attention mechanism to extract features from drug molecular graphs and DDI event graphs in a hierarchical manner, thereby predicting DDI event types between drug pairs.

LaGAT [23]: LaGAT extracts diverse semantic information about drugs from neighboring nodes in the knowledge graph, leveraging the drug pair links to predict the types of DDI events between drug pairs.

DGNN-DDI [51]: DGNN-DDI utilizes dual GNNs to capture drug feature representations by considering both drug substructures and the interaction information between these chemical substructures, aiming to predict the type of DDI events between drug pairs.

DGANDDI [52]: DGANDDI develops a two-GAN architecture to capture the knowledge between drug attributes and the topological structure of the DDI network, thereby predicting DDI event types between drug pairs.

TIGER [53]: TIGER employs the Transformer architecture with a relation-aware self-attention mechanism and integrates representation learning from both drug molecular graphs and knowledge graphs to predict DDIs.

R-GCN [48]: R-GCN is a variant of graph neural networks, which is specifically designed for handling graphs with multiple relations. We directly apply R-GCN on the DDI event graph to learn drug representations, followed by inputting concatenated representations of drug pairs into MLP for DDI events prediction.

R-GAT [49]: Extending the attention mechanism to the realm of relational graphs, R-GAT employs a masked self-attention mechanism that considers local relationship structures and node features. We directly utilize R-GAT on the DDI event graph to learn drug representations, and then input concatenated representations of drug pairs into MLP for DDI events prediction.

GNN-FILM [50]: GNN-FILM is a variant of graph neural networks that incorporates Feature-wise Linear Modulation (FiLM). We directly apply GNN-FILM on the DDI event graph to learn drug representations, subsequently input concatenated representations of drug pairs into MLP for DDI events prediction.

MRCGNN [31]: Employing node feature and edge shuffling for data augmentation on the DDI event graph, subsequently MRCGNN utilizes graph contrastive learning to obtain features of different drug nodes, enabling the prediction of DDI event types between drug pairs.

C. Evaluation metrics

To assess the predictive capability of MRGCDDI, we conduct a five-fold cross-validation on a benchmark dataset, where

Method	Deng's DataSet				Ryu's DataSet			
	Acc.	Macro-F1	Macro-Rec.	Macro-Pre.	Acc.	Macro-F1	Macro-Rec.	Macro-Pre.
DeepDDI	0.7807	0.6055	0.5839	0.6611	0.9323	0.8643	0.8512	0.8928
SSI-DDI	0.7866	0.4216	0.3896	0.5139	0.9008	0.6663	0.6287	0.7507
Trimnet-DDI	0.8570	0.6548	0.6363	0.7046	0.9353	0.8288	0.8128	0.8627
R-GAT	0.8303	0.5983	0.5754	0.6553	0.9416	0.8617	0.8410	0.9039
GNN-FILM	0.8571	0.6318	0.6035	0.6985	0.9419	0.8638	0.8401	0.9077
MUFFIN	0.8269	0.5245	0.4844	0.6204	0.9510	0.8566	0.8339	0.8980
LaGAT	0.8199	0.6630	0.6851	0.6772	0.9123	0.8402	0.8433	0.9071
DGNN-DDI	N/A	N/A	N/A	N/A	0.9458	0.8709	0.8623	0.9292
DGANDDI	0.8571	0.6452	0.6452	0.69498	0.9480	0.8576	0.8563	0.9283
TIGER	0.8760	0.6921	0.6861	0.7366	0.9372	0.8797	0.8801	0.8893
R-GCN	0.8695	0.7026	0.6878	0.7500	0.9284	0.8487	0.8291	0.8881
GoGNN	0.8766	0.6938	0.6841	0.7316	0.9424	0.8589	0.8451	0.8949
MRCGNN	0.8979	0.7791	0.7688	0.8101	0.9567	0.8894	0.8727	0.9221
MRGCDDI	0.9001	0.7742	0.7554	0.8143	0.9643	0.8983	0.8819	0.9303

TABLE III

MULTICLASS CLASSIFICATION ABILITY COMPARISON OF MRGCDDI AND BASELINE METHODS ON FIVE-FOLD CROSS-VALIDATION

Methods	Deng's dataset				Ryu's dataset			
	Acc.	Macro-F1	Macro-Rec.	Macro-Pre.	Acc.	Macro-F1	Macro-Rec.	Macro-Pre.
w/o HGT&MCL	0.8570	0.6548	0.6363	0.7046	0.9353	0.8288	0.8128	0.8627
w/o Tri&MCL	0.8695	0.7026	0.6878	0.7500	0.9284	0.8487	0.8291	0.8881
MRGCDDI	0.9001	0.7742	0.7554	0.8143	0.9643	0.8983	0.8819	0.9303

TABLE IV

EXPERIMENTAL RESULTS OF ABLATION STUDY IN PREDICTING DDI EVENTS

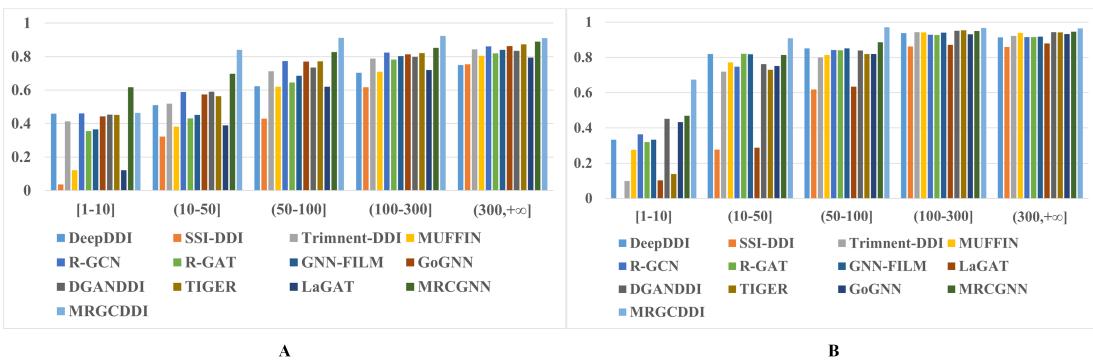


Fig. 3. The accuracy of five event groups on Deng's (A) and Ryu's (B) datasets

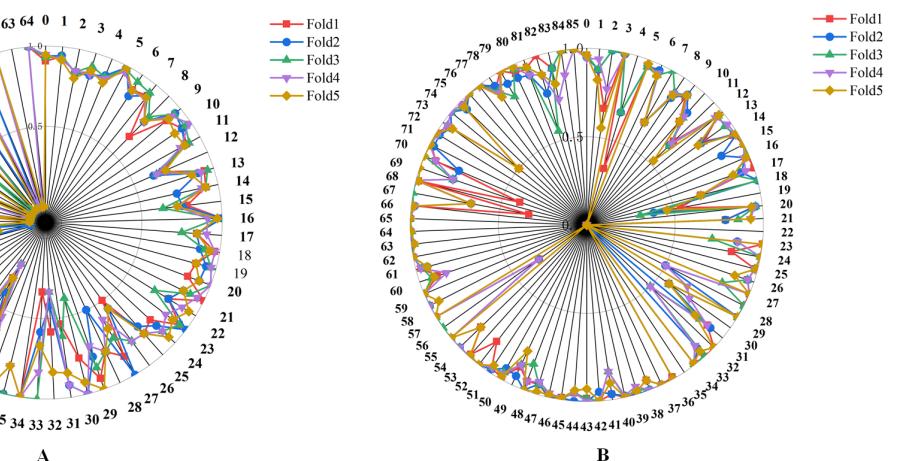


Fig. 4. The accuracy of five event groups on Deng's (A) and Ryu's (B) datasets.

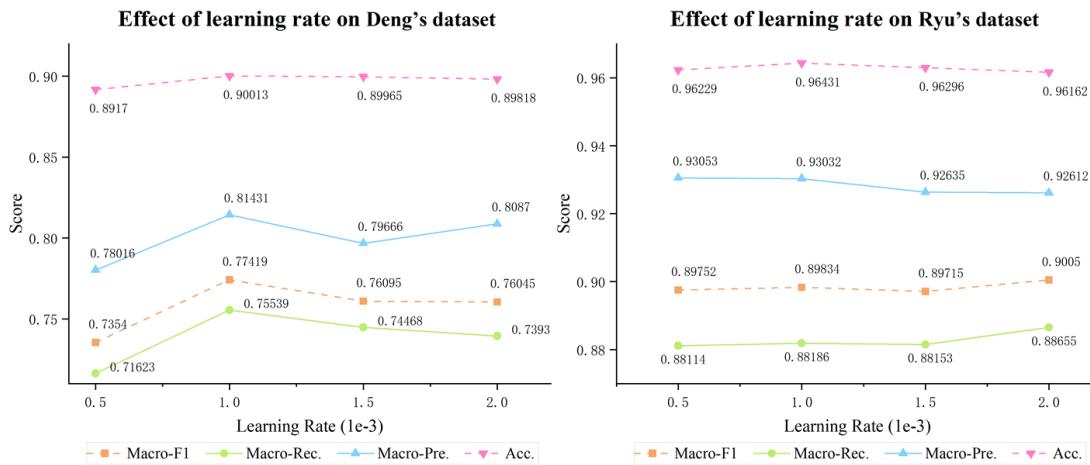


Fig. 5. Effect of learning rate on Deng's and Ryu's datasets

each dataset comprises pairs of drugs and their interaction events. Training samples were randomly divided into five subsets, with each subset sequentially used as the test set while the other four subsets were employed for model training. We evaluate the predictive accuracy of MRGCDDI on the test set and utilized the average prediction accuracy from the five-fold cross-validation as the final performance metric.

In the multiclass classification prediction task, we use several commonly used metrics, including Accuracy (Acc.), Precision (Pre.), Recall (Rec.), F1 score (F1). Precision, Recall, and F1 score are calculated using the macro-average method, and their formulas are as follows:

$$Acc. = \frac{1}{n} \sum_{i=1}^n \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i} \quad (26)$$

$$Pre. = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i} \quad (27)$$

$$Rec. = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FN_i} \quad (28)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (29)$$

where TP , TN , FP and NP denote true positives, true negatives, false positives and false negatives, respectively. n denotes number of categories in DDI event.

D. Comparison of MRGCDDI and Baseline methods

Both MRGCDDI and the baseline methods are implemented by five-fold cross-validation. The comparison result between MRGCDDI and the baseline methods are presented in the Table 1, illustrating that MRGCDDI outperforms the baseline methods. Specifically, on Deng's dataset, MRGCDDI achieves an average increase of 4.33% on Acc., 11.57% on Macro-F1, 10.97% on Macro-Rec., 10.64% on Macro-Pre. Similarly, on Ryu's dataset, the model shows improvements with an average increase of 2.42% on Acc., 3.86% on Macro-F1, 3.49% on Macro-Rec., 2.75% on Macro-Pre., demonstrating the validity of MRGCDDI. By observing and analysing the

results of Table 3, we found that (1) DeepDDI show relatively poor performance because they only consider the structure similar of drug. (2) In comparison to R-GCN, R-GAT and GNN-FILM, the utilization of HGT in MRGCDDI has better performance, which can be attributed to its ability in ability in learning drug representations on multi-relational DDI event graphs, especially on Deng's dataset with a relatively small amount of data. (3) SSI-DDI, TrimNet-DDI and DGNN-DDI show slightly better performance but still inferior to MRGCDDI. The reason is they only consider the structural features of drugs, whereas MRGCDDI integrates both the structural features of drugs and the interaction information between drug pairs. (4) The LaGAT , while accounting for the heterogeneous information present in the knowledge graph, fails to consider the valid information that can be derived from the drug itself. Consequently, it perform worse performance than MRGCDDI. (5) The DGANDDI method is unable to consider multiple types of drug interactions, resulting in a lower level of accuracy than that achieved by the MRGCDDI method. (6) GoGNN and TIGER are inferior to MRGCDDI, primarily because multi-relational contrastive learning used in MRGCDDI captures the implicit information behind the multi-relational DDI event graph. (7) MRGCDDI is capable of retaining the semantics of the DDI event graph during encoder perturbation in the graph comparative learning phase. In contrast, the graph data enhancement methods used in MRCGNN (shuffle node features, shuffle edge relations) destroys the semantics of the original DDI event graph, thus degrading the quality of the drug representations learned by graph contrastive learning. Therefore, MRCGNN does not perform as effectively as MRGCDDI. Meanwhile, in Figure 4, it's evident that the model effectively learns to represent drug molecules. Despite the complexity and challenge of multi-classification, MRGCDDI demonstrates strong performance in accurately classifying the accuracy for each type of DDI event across five event groups.

To further compare MRGCDDI with baseline methods, we analyzed their Macro-F1 scores across five distinct groups of DDI events, paying particular attention to their performance on rare events, defined as occurrences between 1 and 10

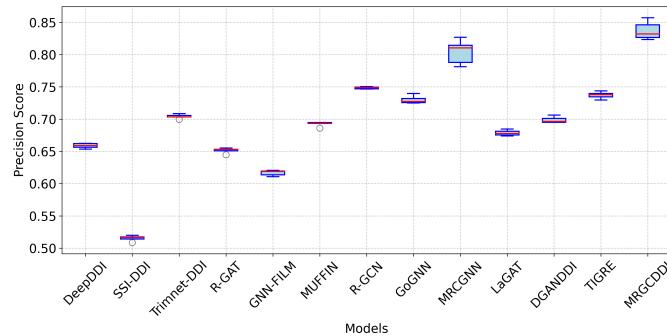


Fig. 6. The statistical test between MRGCDDI and baseline methods

times. This analysis is conducted for all groups of events as depicted in Table 2. As illustrated in Figure 3, the performance of all methods encounters a notable decline with decreasing event frequency. However, MRGCDDI outperforms all baseline models across every group of DDI events. Notably, there is a significant enhancement in predicting rare events in Ryu's dataset, indicating that MRGCDDI possesses a significant advantage in predicting rare DDI events. Furthermore, we observed the following: (1) SSI-DDI, TrimNet-DDI, and MUFFIN exhibit weaker performance in the rare events group. This might be attributed to their reliance solely on drug structure information, disregarding drug interaction information. (2) In comparison to R-GAT and GNN-FILM, as employed in MRGCDDI, HGT shows superior ability to learn drug representations within the multi-relation DDI event graph, resulting in enhanced prediction accuracy for rare event group. (3) The superiority of MRGCNN over GoGNN and MRGCNN suggests that multi-relation contrastive learning without data augmentation contributes to better prediction of rare DDI events. In summary, the optimal performance of MRGCDDI arises from the integration of diverse information aspects and the incorporation of contrastive learning.

Moreover, we conducted a statistical test about the differences between MRGCDDI and other state-of-the-art methods. As shown in Figure 6, compared to other baseline methods, MRGCDDI achieves a median precision score close to 0.81, the highest among all methods, demonstrating exceptional performance. The relatively small box size for MRGCDDI indicates that the distribution of the middle 50% of the data is concentrated, suggesting minimal fluctuations in accuracy across different runs, highlighting its high stability. The short length of the upper and lower whiskers further indicates that the data range remains narrow, with most accuracy scores close to the median and without extreme deviations. Additionally, the absence of outliers suggests that all data points fall within a reasonable range, further reinforcing the consistent and stable performance of MRGCDDI across multiple tests.

E. Ablation study

To investigate the importance of each component for our model, we consider the following variants of MRGCDDI: (1)MRGCDDI without TrimNet *w/oTri*: This variant removes multi-relational DDI event graph learning through TrimNet. Instead, it uses drug features learned directly from

drug molecule graphs for prediction. (2) MRGCDDI without HGT *w/oHGT*: In this variant, drug representation learning is removed, and drug representations are replaced with randomly generated features. (3) MRGCDDI without multi-relational contrastive learning (*w/o MCL*): This variant removes graph contrastive learning through SimGRACE and directly employs drug representations for prediction. Observing the results in Table 4, the following results can be obtained: (1) MRGCDDI *w/oHGT&MCL* has worse results than MRGCDDI *w/oTri&MCL*, indicating that drug interaction information holds greater significance drug structural features for predicting DDI events. (2) MRGCDDI outperforms MRGCDDI (*w/o Tri&MCL*) and MRGCDDI (*w/o HGT&MCL*), suggesting the effectiveness of the graph contrastive learning model we used without data augmentation on the DDI events prediction task.

F. Training Time

We compare the running time of our method with the existing baseline. As demonstrated in Figure 7, MRGCDDI consistently maintains a moderate running time. This balance ensures efficiency without sacrificing predictive performance. While several approaches prioritize speed over accuracy, our method achieves both. It offers a competitive runtime while delivering superior predictive results, surpassing the baseline in accuracy. By focusing on both aspects, MRGCDDI achieves an optimal trade-off between computational cost and prediction quality.

G. Hyperparameter sensitivity analysis

In our work, hyper-parameter sensitivity analysis is conducted on Deng's and Ryu's datasets to explore the influence of several critical hyper-parameters on model performance. The investigated hyper-parameters include the learning rate lr , the coefficient μ in Equation (13), and the coefficient α in Equation (19).

1) *Effect of learning rate lr*: We adjust the learning rate lr in MRGCDDI through a grid search strategy. As illustrated in Figure 5, the best performance across both datasets is achieved at a learning rate of 0.001.

2) *Effect of the coefficients μ and α* : Similar to the analysis of the learning rate, in order to analyze the sensitivity of MRGCDDI to changes in μ and α , a series of experiments is conducted using a grid search strategy. The values of μ and α are varied from {0.05, 0.1, 0.15, 0.2} and {0.1, 0.2, 0.3}, respectively. The impact of these variations on the performance of the model was thoroughly investigated. The results presented in the Figure 7 show that the model achieves the best results when α is equal to 0.15 and μ to 0.3 in Deng's dataset, and when α is set to 0.15 and μ to 0.2 in Ryu's dataset. Additionally, the following conclusions are drawn: (1) Appropriate perturbation magnitude can enhance performance, while excessive perturbation may lead to degradation due to the incomplete preservation of the semantics of the graph data. (2) The contribution coefficient α of the contrastive learning task is closely related to the perturbation magnitude

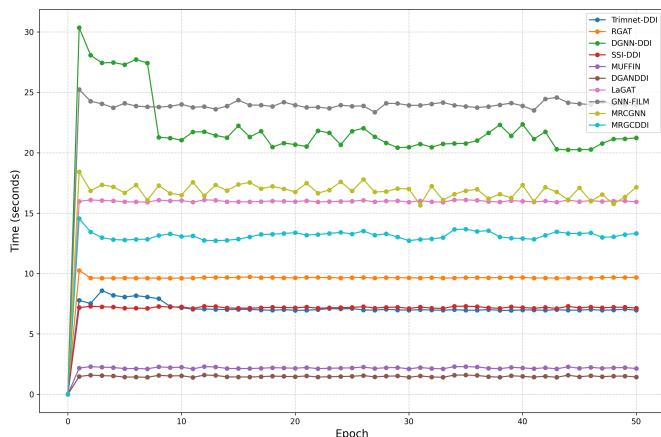


Fig. 7. The comparison of running time between MRGCDDI and baseline methods

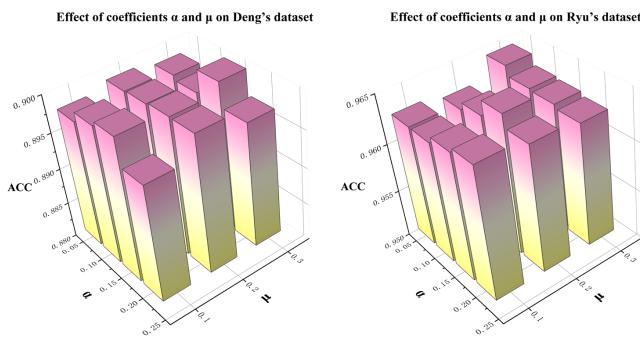


Fig. 8. Effect of coefficients α and μ on Deng's and Ryu's datasets

μ . Appropriate perturbation magnitude brings certain contribution, while excessive contribution coefficient may cause some irrelevant noise, ultimately leading to worse results.

IV. CONCLUSION

This paper proposes MRGCDDI, a multi-relational contrast learning graphical neural network approach that predicts DDI events without the need for data enhancement. MRGCDDI perturbs the encoder to preserve semantic information in the graphical data, integrating drug structure information in the drug molecule graph and drug interaction information in the DDI event graph. Experimental results on two benchmark datasets show that MRGCDDI outperforms the baseline and achieves satisfactory performance in predicting rare DDI events. In future work, there are several directions to improve DDI event prediction, such as incorporating more relevant entities like targets and genes, or including the 3D structural information of the drug in the DDI event graph.

REFERENCES

- [1] J. Aronson, "Classifying drug interactions," *British journal of clinical pharmacology*, vol. 58, no. 4, p. 343, 2004.
- [2] S. Vilar et al., "Similarity-based modeling in large-scale prediction of drug-drug interactions," *Nature protocols*, vol. 9, no. 9, pp. 2147-2163, 2014.
- [3] J. Y. Ryu, H. U. Kim, and S. Y. Lee, "Deep learning improves prediction of drug-drug and drug-food interactions," *Proceedings of the national academy of sciences*, vol. 115, no. 18, pp. E4304-E4311, 2018.
- [4] J. Jia, F. Zhu, X. Ma, Z. W. Cao, Y. X. Li, and Y. Z. Chen, "Mechanisms of drug combinations: interaction and network perspectives," *Nature reviews Drug discovery*, vol. 8, no. 2, pp. 111-128, 2009.
- [5] R. Ferdousi, R. Safdari, and Y. Omidi, "Computational prediction of drug-drug interactions based on drugs functional similarities," *Journal of biomedical informatics*, vol. 70, pp. 54-64, 2017.
- [6] A. Kastrin, P. Ferk, and B. Leskošek, "Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning," *PloS one*, vol. 13, no. 5, p. e0196865, 2018.
- [7] D. Sridhar, S. Fakhraei, and L. Getoor, "A probabilistic approach for collective similarity-based drug-drug interaction prediction," *Bioinformatics*, vol. 32, no. 20, pp. 3175-3182, 2016.
- [8] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabidan, and C. Friedman, "Drug—drug interaction through molecular structure similarity analysis," *Journal of the American Medical Informatics Association*, vol. 19, no. 6, pp. 1066-1074, 2012.
- [9] S. Vilar, E. Uriarte, L. Santana, N. P. Tatonetti, and C. Friedman, "Detection of drug-drug interactions by modeling interaction profile fingerprints," *PloS one*, vol. 8, no. 3, p. e58321, 2013.
- [10] A. K. Nyamabo, H. Yu, and J.-Y. Shi, "SSI-DDI: substructure–substructure interactions for drug-drug interaction prediction," *Briefings in Bioinformatics*, vol. 22, no. 6, p. bbab133, 2021.
- [11] A. K. Nyamabo, H. Yu, and J.-Y. Shi, "SSI-DDI: substructure–substructure interactions for drug-drug interaction prediction," *Briefings in Bioinformatics*, vol. 22, no. 6, p. bbab133, 2021.
- [12] X. Shen, Z. Li, Y. Liu, B. Song, and X. Zeng, "PEB-DDI: A Task-Specific Dual-View Substructural Learning Framework for Drug-Drug Interaction Prediction," *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [13] K. Huang, C. Xiao, T. Hoang, L. Glass, and J. Sun, "Caster: Predicting drug interactions with chemical substructure representation," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, no. 01, pp. 702-709.
- [14] Z. Yang, W. Zhong, Q. Lv, and C. Y.-C. Chen, "Learning size-adaptive molecular substructures for explainable drug-drug interaction prediction by substructure-aware graph neural network," *Chemical science*, vol. 13, no. 29, pp. 8693-8703, 2022.
- [15] S. Jain, E. Chouzenoux, K. Kumar, and A. Majumdar, "Graph regularized probabilistic matrix factorization for drug-drug interactions prediction," *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [16] J. Zhu, Y. Liu, Y. Zhang, and D. Li, "Attribute supervised probabilistic dependent matrix tri-factorization model for the prediction of adverse drug-drug interaction," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2820-2832, 2020.
- [17] N. Rohani, C. Eslahchi, and A. Katanforoush, "Iscmf: integrated similarity-constrained matrix factorization for drug-drug interaction prediction," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, pp. 1-8, 2020.
- [18] J.-Y. Shi, K.-T. Mao, H. Yu, and S.-M. Yiu, "Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization," *Journal of cheminformatics*, vol. 11, no. 1, pp. 1-16, 2019.
- [19] W. Zhang, Y. Chen, D. Li, and X. Yue, "Manifold regularized matrix factorization for drug-drug interaction prediction," *Journal of biomedical informatics*, vol. 88, pp. 90-97, 2018.
- [20] H. Yu et al., "Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization," *BMC systems biology*, vol. 12, no. 1, pp. 101-110, 2018.
- [21] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 701-710.
- [22] Y. Deng et al., "META-DDIE: predicting drug-drug interaction events with few-shot learning," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab514, 2022.
- [23] Y. Hong, P. Luo, S. Jin, and X. Liu, "LaGAT: link-aware graph attention network for drug-drug interaction prediction," *Bioinformatics*, vol. 38, no. 24, pp. 5406-5412, 2022.
- [24] M. R. Karim, M. Cochez, J. B. Jares, M. Uddin, O. Beyan, and S. Decker, "Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network," in *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*, 2019, pp. 113-123.

- [25] X. Lin, Z. Quan, Z.-J. Wang, T. Ma, and X. Zeng, "KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction," in IJCAI, 2020, vol. 380, pp. 2739-2745.
- [26] S. Pang, Y. Zhang, T. Song, X. Zhang, X. Wang, and A. Rodriguez-Patón, "AMDE: A novel attention-mechanism-based multidimensional feature encoder for drug-drug interaction prediction," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab545, 2022.
- [27] Z.-H. Ren *et al.*, "A biomedical knowledge graph-based method for drug-drug interactions prediction through combining local and global features with deep neural networks," *Briefings in Bioinformatics*, vol. 23, no. 5, 2022.
- [28] X. Su, L. Hu, Z. You, P. Hu, and B. Zhao, "Attention-based Knowledge Graph Representation Learning for Predicting Drug-drug Interactions," *Briefings in Bioinformatics*, vol. 23, no. 3, p. bbac140, 2022.
- [29] J. Wang, X. Liu, S. Shen, L. Deng, and H. Liu, "DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab390, 2022.
- [30] Y. Wang, Y. Min, X. Chen, and J. Wu, "Multi-view graph contrastive representation learning for drug-drug interaction prediction," in Proceedings of the Web Conference 2021, 2021, pp. 2921-2933.
- [31] Z. Xiong *et al.*, "Multi-relational contrastive learning graph neural network for drug-drug interaction event prediction," in Proceedings of the AAAI Conference on Artificial Intelligence, 2023, vol. 37, no. 4, pp. 5339-5347.
- [32] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, and X. Li, "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data," *BMC bioinformatics*, vol. 18, pp. 1-12, 2017.
- [33] X. Zhang *et al.*, "Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug-drug interactions prediction," *Briefings in Bioinformatics*, vol. 23, no. 5, 2022.
- [34] X. Zhu, Y. Shen, and W. Lu, "Molecular substructure-aware network for drug-drug interaction prediction," in Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 4757-4761.
- [35] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457-i466, 2018.
- [36] P. Li *et al.*, "TrimNet: learning molecular representation from triplet messages for biomedicine," *Briefings in Bioinformatics*, vol. 22, no. 4, p. bbaa266, 2021.
- [37] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous graph transformer," in Proceedings of the web conference 2020, 2020, pp. 2704-2710.
- [38] J. Xia, L. Wu, J. Chen, B. Hu, and S. Z. Li, "Simgrace: A simple framework for graph contrastive learning without data augmentation," in Proceedings of the ACM Web Conference 2022, 2022, pp. 1070-1079.
- [39] Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, and S. Liu, "A multimodal deep learning framework for predicting drug-drug interaction events," *Bioinformatics*, vol. 36, no. 15, pp. 4316-4322, 2020.
- [40] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in International conference on machine learning, 2017: PMLR, pp. 1263-1272.
- [41] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [42] O. Vinyals, S. Bengio, and M. Kudlur, "Order matters: Sequence to sequence for sets," *arXiv preprint arXiv:1511.06391*, 2015.
- [43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
- [44] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," *Advances in neural information processing systems*, vol. 33, pp. 5812-5823, 2020.
- [45] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in International conference on machine learning, 2020: PMLR, pp. 1597-1607.
- [46] Y. Chen, T. Ma, X. Yang, J. Wang, B. Song, and X. Zeng, "MUFFIN: multi-scale feature fusion for drug-drug interaction prediction," *Bioinformatics*, vol. 37, no. 17, pp. 2651-2658, 2021.
- [47] H. Wang, D. Lian, Y. Zhang, L. Qin, and X. Lin, "Gognn: Graph of graphs neural network for predicting structured entity interactions," *arXiv preprint arXiv:2005.05537*, 2020.
- [48] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *The Semantic Web: 15th International Conference, ESWC 2018*, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings 15, 2018: Springer, pp. 593-607.
- [49] D. Busbridge, D. Sherburn, P. Cavallo, and N. Y. Hammerla, "Relational graph attention networks," *arXiv preprint arXiv:1904.05811*, 2019.
- [50] M. Brockschmidt, "Gnn-film: Graph neural networks with feature-wise linear modulation," in *International Conference on Machine Learning*, 2020: PMLR, pp. 1144-1152.
- [51] M. Ma, X. Lei, A dual graph neural network for drug-drug interactions prediction based on molecular structure and interactions, *PLOS Computational Biology*, 19 (2023) e1010812.
- [52] H. Yu, K. Li, J. Shi, DGANDDI: double generative adversarial networks for drug-drug interaction prediction, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20 (2022) 1854-1863.
- [53] X. Su, P. Hu, Z.-H. You, S.Y. Philip, L. Hu, Dual-Channel Learning Framework for Drug-Drug Interaction Prediction via Relation-Aware Heterogeneous Graph Transformer, *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, pp. 249-256.
- [54] Ren Z-H, Chang-Qing Y, Li L-P, et al. Biodkg-ddi: predicting drug-drug interactions based on drug knowledge graph fusing biochemical information. *Brief Funct Genomics* 2022;21(3):216-29.
- [55] Yue Y, Huang K, Zhang C, et al. Sumgnn: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics* 2021;37(18):2988-95.
- [56] Xiaorui S, You Z-H, Huang D-s, et al. Biomedical knowledge graph embedding with capsule network for multi-label drug-drug interaction prediction. *IEEE Trans Knowl Data Eng* 2022.
- [57] Cao, Xiyue, et al. "scPriorGraph: constructing biosemantic cell-cell graphs with prior gene set selection for cell type identification from scRNA-seq data." *Genome Biology* 25.1 (2024): 207.
- [58] Wei Song, Lewen Xu, Chenguang Han, Zhen Tian*, Quan Zou*. Drug-target interaction predictions with multi-view similarity network fusion strategy and deep interactive attention mechanism. *Bioinformatics*. 2024, 40(6): btac346.
- [59] Hongjie Wu, Junkai Liu, Tengsheng Jiang, Quan Zou, Shujie Qi, Zhiming Cui, Prayag Tiwari*, Yijie Ding*. AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. *Neural Networks*. 2024, 169: 623-636



Yu Li received the PhD degrees from the University of Chinese Academy of Sciences, Beijing, China. He is currently working as postdoctoral fellow with the Northwestern Polytechnical University. His research interests include bioinformatics, machine learning, and data mining.



Lin-Xuan Hou received her B.E. degree in Software Engineering from Northwest University, Xi'an, China, in 2022. She is pursuing her Master degree in computer science at Northwestern Polytechnical University, Xi'an, China. Her current research interests include data mining, network analysis, artificial intelligence, and its applications in bioinformatics.



Zhu-Hong You received the BE degree in electronic information science and engineering from Hunan Normal University, Changsha, China, in 2005, and the PhD degree in control science and engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2010. From June 2008 to November 2009, he was a visiting research fellow with the Center of Biotechnology and Information, Cornell University, Ithaca, New York. He is currently a professor with the Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Science, China. He has published more than 170 research papers in refereed journals and conferences in the areas of pattern recognition, bioinformatics, and complex-network analysis. He holds more than 10 patents. His current research interests include neural networks, intelligent information processing, sparse representation, and its applications in bioinformatics.



Hai-Cheng Yi received his Ph.D. degree in Computer Applied Technology at University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2022. From 2021 to 2022, He was a visiting scholar at Nanyang Technological University (NTU) in Singapore. He is currently an Associate Professor with the School of Computer Science, Northwestern Polytechnical University (NPU), Xi'an, China. His current research interests include artificial intelligence for science, graph deep learning, complex networks analysis, and bioinformatics.



Yang Yuan received the Ph.D in Computer Application Technology from The Xinjiang Institute of Physics & Chemistry, University of Chinese Academy of Sciences and is currently a lecturer at the School of Computer Science and Artificial Intelligence, Changzhou University, China. His current research interests include natural language processing, knowledge graph and Bioinformatics.



Cheng-Gang Mi received the B.S. degree from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2010 and the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2015. He is currently an associate professor with the Foreign Language and Literature Institute, Xi'an International Studies University, China. He has authored more than 30 papers in his research fields. His research works focus on statistical approaches to natural language processing, especially machine translation from and into morphologically rich languages. His research interests also include morphological analysis, automating knowledge acquisition and machine learning for NLP tasks.



Yu-An Huang is an Associate Professor in the School of Computer Science at Northwestern Polytechnical University. His research interests include computational biology and data mining.