

Article

# MPHGCL-DDI: Meta-Path-Based Heterogeneous Graph Contrastive Learning for Drug-Drug Interaction Prediction

Baofang Hu <sup>1,\*</sup> , Zhenmei Yu <sup>1</sup> and Mingke Li <sup>2</sup>

<sup>1</sup> School of Data and Computer Science, Shandong Women's University, Jinan 250030, China; 34019@sdwu.edu.cn

<sup>2</sup> School of Information Science and Engineering, University of Jinan, Jinan 250024, China; 202321100467@stu.ujn.edu.cn

\* Correspondence: 34022@sdwu.edu.cn

**Abstract:** The combinatorial therapy with multiple drugs may lead to unexpected drug-drug interactions (DDIs) and result in adverse reactions to patients. Predicting DDI events can mitigate the potential risks of combinatorial therapy and enhance drug safety. In recent years, deep models based on heterogeneous graph representation learning have attracted widespread interest in DDI event prediction and have yielded satisfactory results, but there is still room for improvement in prediction performance. In this study, we proposed a meta-path-based heterogeneous graph contrastive learning model, MPHGCL-DDI, for DDI event prediction. The model constructs two contrastive views based on meta-paths: an average graph view and an augmented graph view. The former represents that there are connections between drugs, while the latter reveals how the drugs connect with each other. We defined three levels of data augmentation schemes in the augmented graph view and adopted a combination of three losses in the model training phase: multi-relation prediction loss, unsupervised contrastive loss and supervised contrastive loss. Furthermore, the model incorporates indirect drug information, protein–protein interactions (PPIs), to reveal latent relations of drugs. We evaluated MPHGCL-DDI on three different tasks of two datasets. Experimental results demonstrate that MPHGCL-DDI surpasses several state-of-the-art methods in performance.

**Keywords:** drug-drug interaction; heterogeneous graph contrastive learning; meta-path; data augmentation; protein–protein interaction



**Citation:** Hu, B.; Yu, Z.; Li, M. MPHGCL-DDI: Meta-Path-Based Heterogeneous Graph Contrastive Learning for Drug-Drug Interaction Prediction. *Molecules* **2024**, *29*, 2483. <https://doi.org/10.3390/molecules29112483>

Academic Editor: Shaoyong Lu

Received: 21 April 2024

Revised: 17 May 2024

Accepted: 22 May 2024

Published: 24 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The drug-drug interaction (DDI) refers to the change of a drug's expected effect when it is combined with another drug [1]. The combinatorial therapy with multiple drugs, often termed as polypharmacy, is a commonly adopted strategy for treating various complex diseases [2]. It is estimated that, during the period of 2010–2011, 36% of elderly individuals in the United States regularly used at least five medications or supplements simultaneously. And approximately 15% of individuals aged from 61 to 80 are at a potential risk of DDIs due to medication combinations [3]. The occurrence of DDIs often leads to unexpected side effects, which may be harmful to patients or weaken the effectiveness of drugs [4]. In severe cases, it even endangers patients' lives. Therefore, accurately predicting DDIs is essential for drug development and drug safety surveillance.

Although in vitro experiments and clinical trials can be performed to identify drug-drug interactions [5,6], experiments with multiple drugs are impractical due to the large number of possible drug combinations and various comorbidities. Furthermore, in most cases, it is not only necessary to determine whether drug-drug interactions exist, but also to understand what type of DDI event has occurred. This undoubtedly greatly increases the difficulty of experiments. With the advent of the big data era, more and more healthcare-relevant data are becoming more readily available in the pharmaceutical industry. There

are considerable research works trying to leverage these related data for DDI event prediction. For instance, Lin et al. [7] utilized multi-source feature fusion and a transformer self-attention mechanism to incorporate various biochemical attributes (chemical substructure, protein, enzyme and pathway transport) for DDI event prediction. Zhang et al. [8] recognized DDIs using knowledge graph convolutional networks (KGCNs) with neural factorization machines.

With the widespread adoption of graph neural networks (GNNs), an increasing number of studies are employing GNNs to predict drug-drug interactions (DDIs). The multi-source drug information is heterogeneous and can be represented using a heterogeneous graph (HG). Meta-paths, which are sequences of object types, can effectively capture the semantic relations between drugs. Recently, numerous studies have employed meta-path-based heterogeneous graph neural networks (HGNNs) for learning node embeddings, encompassing social networks [9,10], recommendation systems [11,12] and biological healthcare [13–15]. In these studies, meta-paths offer an interpretable way to reveal how entities connect through intermediary entities, for example, how users in recommendation systems connect through items and how drugs connect via target proteins or chemical substructures.

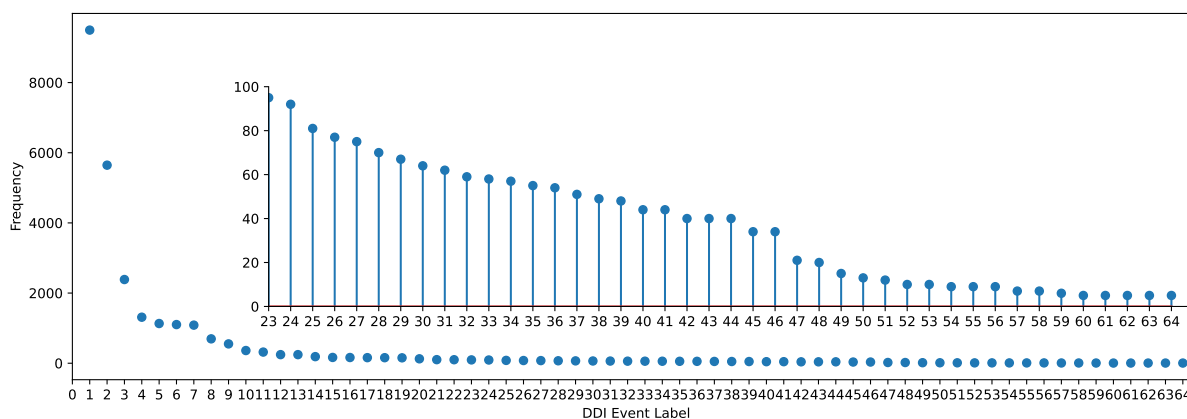
Despite the success of HGNNs, several unresolved issues remain. One notable challenge is the reliance of most existing models on substantial volumes of training data, which are difficult to obtain. To address this issue, unsupervised learning, which can extract supervision from the data themselves and learn high-quality representations, has been applied to HG. Particularly, contrastive learning [16,17], as a major type of unsupervised learning, has recently gained widespread attention. Some heterogeneous graph contrastive learning (HGCL) methods have already been proposed [18–21]. Chen et al. [20] proposed a heterogeneous graph contrastive learning model with meta-path-based augmentations (HGCMA), which is designed for downstream tasks in social network or recommendation systems with a small amount of labeled data. Yu et al. [21] proposed a novel framework, which considers both meta-path contexts and weighted negative samples, for learning node embeddings in social networks.

Several studies have also explored the potential of graph contrastive learning in DDI event prediction [22–26]. Wang et al. [22] presented a multi-view graph contrastive representation learning framework, MIRACLE, to predict DDIs by capturing inter-view molecule structure and intra-view interactions between molecules simultaneously. Zhang et al. [23] proposed a hierarchical triple-view contrastive learning framework (HTCL-DDI), leveraging a dual attention-aware network in the molecular view to aggregate the intramolecular compositional and structural information. Han et al. [25] proposed a supervised contrastive learning method, MDDI-SCL, implemented by three-level loss functions to predict multi-type DDIs.

However, most models seldom incorporate data augmentation strategies and contextual semantic information of meta-paths. Data augmentation, which is commonly utilized in contrastive learning, can further enhance the performance and robustness of models by expanding the training dataset [27,28]. Integrating data augmentation into graph contrastive learning is certainly not ineffective. Especially in DDI event prediction, the DDI datasets exhibit significant class imbalance, as depicted in Figure 1, where many DDI events in Dataset1 have very few instances. This imbalance severely impacts the predictive performance of rare events.

Based on the above discussion, the primary motivation of our work lies in data augmentation schemes and meta-path-based contrastive view construction in heterogeneous graph contrastive learning. We proposed a meta-path-based heterogeneous graph contrastive learning model for DDI event prediction, MPHGCL-DDI. The model utilized multi-source drug information and constructed two drug HGs: a drug HG based on the raw data and an extended drug HG constructed by masking certain features of drug biological attributes. Based on the two drug HGs, we constructed two meta-path-based contrastive views: an average graph view and a data augmentation graph view. In the contrastive

learning phase, we adopted unsupervised contrastive loss and supervised contrastive loss to learn the drug pair embeddings. For model evaluation, we adopted two datasets to assess the predictive performance of MPHGL-DDI for three different multi-type DDI prediction tasks, and compared the performance with several state-of-the-art models. Experimental results demonstrated that our model surpasses several state-of-the-art methods in performance.



**Figure 1.** The frequency of DDI events in Dataset1.

## 2. Results

### 2.1. Experiment Settings

We evaluated the performance of our model based on three different prediction tasks: (i) Task1: DDI event prediction between two known drugs; (ii) Task2: DDI event prediction between one known drug and one new drug; and (iii) Task3: DDI event prediction between two new drugs. New drugs are the drugs missing in the training set, but existing in the test set.

We adopted the hold-out method to divide each dataset into training, validation and test sets with a common partition ratio 7:1:2 [28]. In order to improve the stability and reliability of the experimental results, we randomly divided each dataset five times and took the average value as the final result. Specifically, in Task1, for every event type, we randomly split instances into training, validation and test sets to ensure that training/validation/test sets contained DDIs from all types. In Task2 and Task3, we split drugs instead of instances into training, validation and test sets.

As DDI events prediction is a multi-class classification task on highly imbalanced datasets, we evaluated the model using several commonly adopted evaluation metrics, including accuracy (ACC), area under the precision–recall curve (AUPR), macro-F1, macro-recall (macro-Rec), and macro-precision (macro-Pre).

All the experiments were run on a server with 32G memory and a 40G GPU from NVIDIA company. For our proposed model, MPHGL-DDI, we initialized model parameters using Xavier initialization [29] and trained the model using the Adam optimizer [30].  $K$  in neighbor filtering was set to the average number of connections of all the objects under each meta-path. The number of attention heads was set to 8. Other hyper-parameters of the model were fine-tuned with different step sizes.

### 2.2. Comparison with Baselines

We compared our method with the following state-of-the-art methods.

- MDDI-SCL [25]. It employs a multi-layer self-attention mechanism to learn the latent features of drugs and performs a multi-scale fusion to the outputs of different layers based on contrastive learning.

- MM-GANN-DDI [31]. It is a multi-modal graph-agnostic neural network for predicting DDI events. It fuses six drug modalities with the topological features of the DDI graph through a graph attention neural network.
- MCFF-MTDDI [32]. It extracts drug chemical structure features and drug pairs' extra label features, and integrates these features through a multi-channel feature fusion module, thereby predicting multiple types of DDIs.
- MP-DDI [33]. The model captures the complex semantics and learns high-quality representations of drugs using meta-paths. The original model does not consider what type of interaction event occurs. In experiments, we changed the original model for the DDI event prediction using a multi-layer perceptron.
- RaGSECo [26]. It is based on relation-aware graph structure embedding with co-contrastive learning. The model constructs two heterogeneous graphs: a multi-relational DDI graph and a multi-attribute drug-drug similarity graph, and learns representations of drug pairs using co-contrastive learning.

We categorized the aforementioned approaches based on three perspectives: whether to adopt deep fusion of diverse biological attributes, the attention mechanism to multi-modal information and the utilization of contrastive learning, as shown in Table 1.

**Table 1.** Summary of the experimental comparing algorithms.

Model	Deep Fusion	Attention Mechanism	Contrastive Learning
MDDI-SCL	No	Yes	Yes
MM-GANN-DDI	Yes	Yes	No
MCFF-MTDDI	Yes	Yes	No
MP-DDI	Yes	Yes	No
RaGSECo	Yes	No	Yes
MPHGCL-DDI	Yes	Yes	Yes

Table 2 presents the metric scores achieved by these methods on the two datasets. The comparison results demonstrate that our MPHGCL-DDI outperformed the competitors in both datasets. We also have the following observations. (1) The methods with deep fusion of diverse biological attributes perform better than the simply concatenate method. RaGSECo and MPHGCL-DDI perform better than MDDI-SCL. In MDDI-SCL, the initial feature of a drug is obtained by concatenate operation of biological attributes, which neglects the different importance of attributes. Deep fusion enables learning of higher-level representations, which is more conducive to predicting DDI events. (2) The methods involving contrastive learning perform better than the ones without contrastive learning. MPHGCL-DDI performs better than MCFF-MTDDI and MP-DDI. Contrastive learning can make node representations more discriminative and enhance the effectiveness of representation learning to a certain extent. The data augmentation can improve the model's generalization ability, so MDDI-SCL performs better than RaGSECo. (3) Our model, which adopts an attention mechanism to fuse multi-modal information, outperforms RaGSECo, in part because RaGSECo treats each type of biological attribute information equally. Multi-modal biological attributes contribute differently for each drug feature. MP-DDI performs poorly despite incorporating attention mechanisms to integrate meta-path information, because it only considers DDIs during representation learning, without taking into account the types of DDI events.

To further compare MPHGCL-DDI with baselines, we grouped events according to their occurrence frequency in two datasets and investigated the macro-F1 scores of events in different groups for Task1. Due to the different sizes of the two datasets, the grouping methods for the two datasets were also different. Each dataset was divided into five groups, as listed in Table 3.

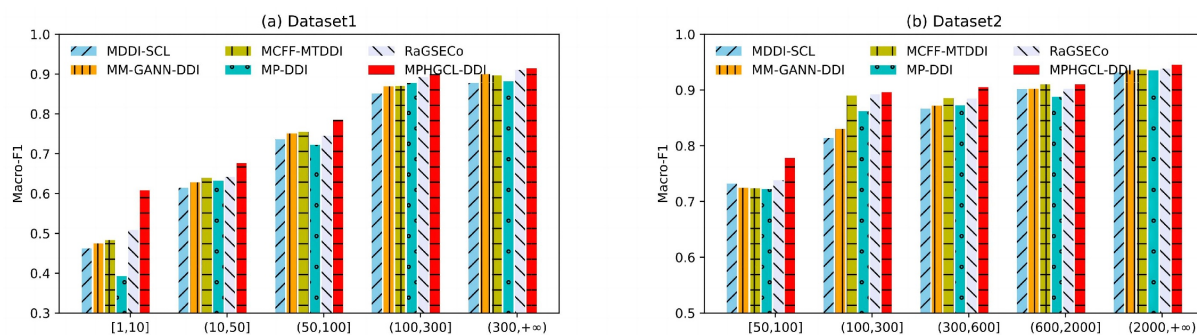
**Table 2.** Performance comparison with the state-of-the-art methods.

Dataset	Task	Methods	Acc	AUPR	Macro-F1	Macro-Rec	Macro-Pre
Dataset1	Task1	MDDI-SCL	0.9378	0.9782	0.8755	0.8767	0.8804
		MM-GANN-DDI	0.9386	0.9786	0.8980	0.895	0.9088
		MCFF-MTDDI	0.9350	0.9757	0.8918	0.8820	0.9100
		MP-DDI	0.9384	0.9621	0.8768	0.8838	0.9076
		RaGSECo	0.9461	0.9838	0.9050	0.9043	0.9121
		MPHGCL-DDI	0.9487	0.9897	0.9107	0.9087	<b>0.9140</b>
	Task2	MDDI-SCL	0.6767	0.6947	0.5304	0.4814	0.6254
		MM-GANN-DDI	0.6705	0.6855	0.5580	0.5156	0.6518
		MCFF-MTDDI	0.6650	0.6800	0.5574	0.5139	0.6507
		MP-DDI	0.6685	0.6723	0.5024	0.4934	0.6144
		RaGSECo	0.6855	0.7115	0.5860	0.5631	0.6514
		MPHGCL-DDI	0.6872	0.7208	0.5931	0.5682	0.6561
	Task3	MDDI-SCL	0.4589	0.3938	0.1919	0.1678	0.2585
		MM-GANN-DDI	0.4386	0.3786	0.2505	0.2480	0.2774
		MCFF-MTDDI	0.4400	0.387	0.2437	0.2351	0.2823
		MP-DDI	0.4398	0.3813	0.2183	0.2064	0.2431
		RaGSECo	0.4591	0.4114	0.2600	0.2513	0.3001
		MPHGCL-DDI	0.4634	0.4227	0.2682	<b>0.2591</b>	<b>0.3035</b>
Dataset2	Task1	MDDI-SCL	0.9516	0.9862	0.9321	0.9500	0.9162
		MM-GANN-DDI	0.9521	0.9868	0.9324	0.9623	0.9174
		MCFF-MTDDI	0.9517	0.9875	0.9348	0.9513	0.9171
		MP-DDI	0.9487	0.9743	0.933	0.9511	0.9146
		RaGSECo	0.9498	0.9890	0.9354	0.9633	0.9201
		MPHGCL-DDI	0.9541	0.9927	0.9407	0.9704	0.9213
	Task2	MDDI-SCL	0.6595	0.6794	0.5578	0.5712	0.5605
		MM-GANN-DDI	0.6530	0.6781	0.5627	0.5636	0.5727
		MCFF-MTDDI	0.6543	0.6821	0.5696	0.5801	0.5683
		MP-DDI	0.6478	0.6685	0.5603	0.5594	0.5691
		RaGSECo	0.6632	0.6874	0.5713	0.5937	0.5773
		MPHGCL-DDI	0.6685	0.6995	0.5889	0.6071	0.5808
	Task3	MDDI-SCL	0.4696	0.4261	0.2838	0.2773	0.3160
		MM-GANN-DDI	0.4731	0.4311	0.2875	0.2794	0.3323
		MCFF-MTDDI	0.4713	0.4374	0.2913	0.2839	0.3195
		MP-DDI	0.4704	0.4256	0.2864	0.2732	0.3069
		RaGSECo	0.4796	0.4403	0.2950	0.2886	0.3298
		MPHGCL-DDI	0.4847	0.4436	0.3021	0.2965	0.3346

**Table 3.** Proportions of events in five groups to all events.

Dataset1	[1, 10]	(10, 50]	(50, 100]	(100,300]	(300, +∞)
	20.00%	21.54%	24.62%	15.38%	18.46%
Dataset2	[50, 100]	(100, 300]	(300, 600]	(600, 2000]	(2000, +∞)
	26.00%	19.00%	19.00%	17.00%	19.00%

As shown in Figure 2, the performances of all models prominently declines with the decrease in the frequency of DDI event occurrences, and MPHGCL-DDI outperforms all baselines on each group of DDI events, especially on rare events with a significant improvement, which demonstrates that MPHGCL-DDI has considerable advantages in predicting rare DDI events. In addition, we found that MM-GANN-DDI, MCFF-MTDDI and MP-DDI, which do not use contrastive learning, achieve relatively unsatisfactory performances on the rare event group. This illustrates that contrastive learning can further help the prediction for rare DDI events.



**Figure 2.** Results of MPHGL-DDI and baselines on events with different frequencies.

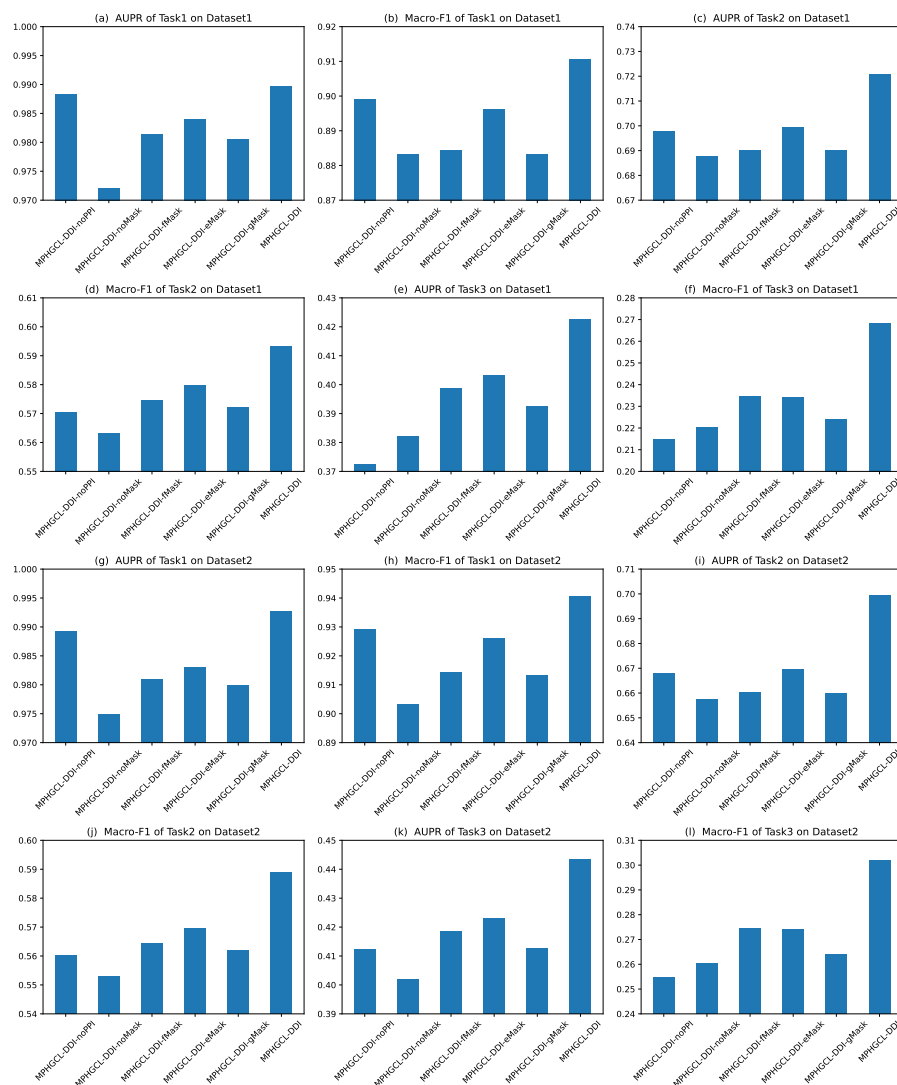
### 2.3. Ablation Study

To investigate the importance of various components of our model, we considered the following variants of MPHGL-DDI:

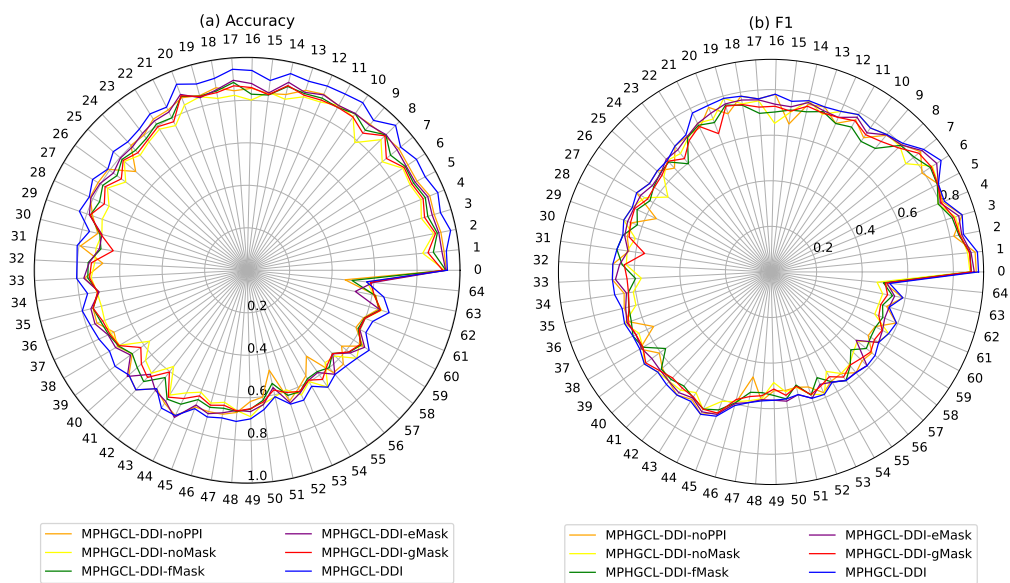
- MPHGL-DDI-noPPI: A variant without protein–protein interaction information, in which the meta-path DPPD is not considered.
- MPHGL-DDI-noMask: The model does not perform any augmentations in the training process.
- MPHGL-DDI-fMask: This model only performs masking of certain features of biological attributes in the training process.
- MPHGL-DDI-eMask: This model only performs masking of certain edges of the meta-path-based sub-graphs in the training process.
- MPHGL-DDI-gMask: This model only performs masking of certain meta-path-based sub-graphs in the training process.

Herein, we selected two representative metric scores (AUPR and macro-F1) to evaluate the prediction performance of MPHGL-DDI and its variants. Figure 3 illustrates the metric scores of six models on Task1, Task2 and Task3 of two datasets. The figure shows that MPHGL-DDI achieves higher metric scores than its variants, indicating the effectiveness of PPIs information and data augmentation. We also observed that MPHGL-DDI-noPPI performs better than other variants on Task1 but shows poorer performances on Task2 and Task3. MPHGL-DDI-noPPI incorporates all data augmentation schemes but lacks PPI information. On the contrary, other variants have PPI information, but lack certain data augmentation schemes. This observation confirms that the test DDIs include new drugs in Tasks 2 and 3, which may impact the model performance. The data augmentations are more effective when the drugs are known, while PPI information is more effective when the drugs are unknown.

To validate the stability of the model and its variants for different DDI events, we further evaluated accuracy and F1 of MPHGL-DDI and its variants for each DDI event on Task1 in Dataset1, which has 65 types of DDI events and fewer labeled instances on each DDI event than Dataset2. The experiment results are shown in Figure 4. The figure shows that MPHGL-DDI demonstrates more stable performance for different events compared with other models, further illustrating the effectiveness of PPIs information and data augmentation. In addition, in most DDI events, MPHGL-DDI-eMask achieves better results compared with MPHGL-DDI-fMask and MPHGL-DDI-gMask, indicating that masking some meta-path instances is more effective than masking features and nodes in data augmentation for heterogeneous graph contrastive learning.



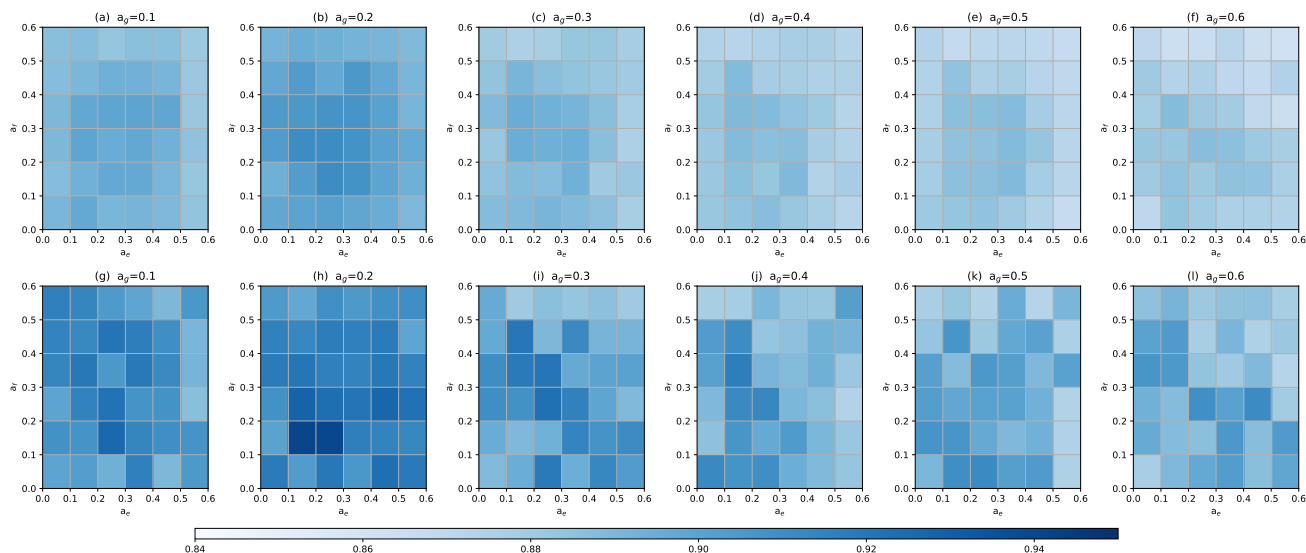
**Figure 3.** Experimental results of MPHGCL-DDI and its five variants in terms of AUPR and macro-F1 on three tasks.



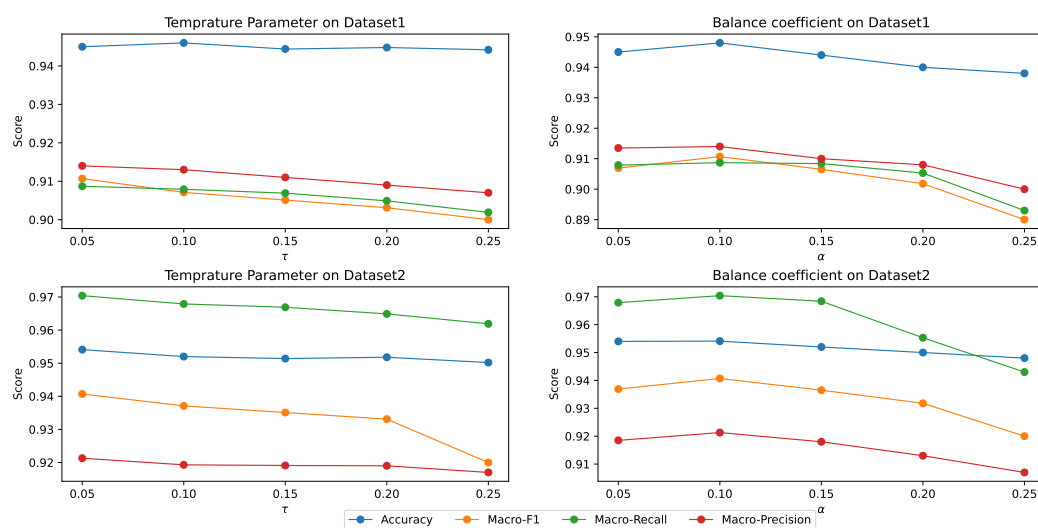
**Figure 4.** Performance comparison for each DDI event of Dataset1.

## 2.4. Hyper-Parameters Analysis

In this section, we performed sensitivity analysis on the main hyper-parameters of MPHGCL-DDI: three masking probabilities,  $a_f$ ,  $a_e$  and  $a_g$ , temperature parameter,  $\tau$ , and balance coefficient,  $\alpha$ . We evaluated the macro-F1 metrics for Task1 on two datasets. The results are shown in Figures 5 and 6.



**Figure 5.** Macro-F1 of MPHGCL-DDI with different masking probabilities. (a–f) Dataset1; (g–l) Dataset2.



**Figure 6.** Performance of MPHGCL-DDI with hyper-parameters  $\tau$  and  $\alpha$ .

With the help of the values in the color bar in Figure 5, we can observe that MPHGCL-DDI demonstrates relatively better performance across most combinations of masking probabilities. Performance only declines when the values of  $a_f$ ,  $a_e$  and  $a_g$  are at marginal levels. Therefore, we conclude that, overall, our augmentation scheme is insensitive to these hyper-parameters, demonstrating the robustness of our model. Additionally, we identified that the optimal hyper-parameter combinations on Dataset1 and Dataset2 are (0.2, 0.3, 0.2) and (0.2, 0.2, 0.2), respectively.

Figure 6 demonstrates the model's performance for different values of  $\tau$  and  $\alpha$ . From the results, we can observe that the performance of the model exhibits a smaller variation range with increasing  $\tau$  and  $\alpha$ , but there still exists a peak performance point. The optimal values on the two datasets are the same, which are  $\tau = 0.05$  and  $\alpha = 0.1$ . A higher

value of  $\alpha$  will lead to a decrease in model performance, because the high weight assigned to the contrastive learning task can cause the model to overly focus on the contrastive learning task during the training process, resulting in poorer performance on the DDI event prediction task.

### 2.5. Case Study

We performed case studies to assess the effectiveness of MPHGL-DDI. We utilized the DDI event instances of Dataset1 to train the model, and then predicted the drug pairs that do not exist on Dataset1. We focused on the top five most frequent DDI events, selected the top 10 prediction results for each event and checked them using the DDI Checker tool provided by DrugBank (<https://go.drugbank.com/drugs> accessed on 17 February 2024).

Out of the 50 selected drug pairs, 22 DDIs were confirmed in DrugBank and detailed in Table 4. For instance, the metabolism of dronedarone can be decreased when combined with ketoconazole. The serum concentration of isradipine can be increased when it is combined with cimetidine.

**Table 4.** The confirmed drug pairs of the selected 50 drug pairs.

Index	DDI Event	Drug1	Drug2
1	The metabolism decreases	Dronedarone	Ketoconazole
2	The metabolism decreases	Fluvoxamine	Isoniazid
3	The metabolism decreases	Ketoconazole	Erythromycin
4	The metabolism decreases	Imatinib	Isradipine
5	The metabolism decreases	Fluvastatin	Clemastine
6	The serum concentration increases	Isradipine	Cimetidine
7	The metabolism decreases	Cimetidine	Crizotinib
8	The risk or severity of adverse effects increases	Fluticasone propionate	Fluvoxamine
9	The metabolism decreases	Atomoxetine	Abiraterone
10	The serum concentration decreases	Fosphenytoin	Clozapine
11	The risk or severity of adverse effects increases	Bromocriptine	Citalopram
12	The serum concentration increases	Haloperidol	Cinacalcet
13	The serum concentration decreases	Eszopiclone	Mitotane
14	The risk or severity of adverse effects increases	Fentanyl	Dosulepin
15	The therapeutic efficacy decreases	Carbamazepine	Mianserin
16	The therapeutic efficacy decreases	Chlorpromazine	Cortisone acetate
17	The therapeutic efficacy decreases	Bosentan	Antipyrine
18	The serum concentration increases	Bendroflumethiazide	Cocaine
19	The serum concentration increases	Enzalutamide	Candesartancilexetil
20	The serum concentration increases	Conivaptan	Bisoprolol
21	The therapeutic efficacy decreases	Amoxapine	Donepezil
22	The risk or severity of adverse effects increases	Amoxapine	Alosetron

### 3. Discussion

In this study, we proposed a reliable computational model, MPHGL-DDI, for predicting DDI events. The model began with constructing a drug heterogeneous graph and meta-path-based contrastive views. Three levels of data augmentation schemes were designed within the augmented graph view. Subsequently, a graph encoder, comprising node feature transformation, inter-graph encoder and intra-graph encoder, was applied to obtain drug embeddings in each view. Finally, the representations of drug pairs were fed into a multi-layer perceptron (MLP) to predict DDI events. In experimental evaluation, MPHGL-DDI exhibited satisfactory performance across three tasks on two datasets. Furthermore, case studies demonstrated the model's reliable and accurate predictive performance. It is reasonable to conclude that MPHGL-DDI contributes to predicting DDI events.

The reliable performance of MPHGL-DDI benefited from the following factors:

- The model utilized heterogeneous graph contrastive learning. There are two contrastive views in the model: an average graph view and an augmented graph view. The former view is an average graph of all meta-path-based sub-graphs, representing

the connections between the drugs. The latter view integrates various meta-path-based sub-graphs using attention mechanisms, revealing the how the drugs connect with each other.

- The data augmentation schemes introduce more variations and differences into the drug data and enhance the model's ability to generalize. There are three levels of data augmentation schemes: feature augmentation by masking features of biological attributes, edge augmentation by masking the edges of meta-path-based sub-graphs, and sub-graph augmentation by masking one meta-path. The three strategies progressively increase the perturbation intensity.
- The contrastive learning framework integrates both unsupervised contrastive loss and supervised contrastive loss, and improves the representation learning capacity of drug pairs.
- In addition, the model not only focuses on the direct biological attributes of drugs but also on their indirect information, including protein–protein interactions. This reveals the implicit relationships among drugs and is effective for DDI event prediction.

However, there are still two significant issues that need to be addressed in future work. On the one hand, our current results suggest that the trained model tends to assign higher scores to DDI events with more instances. The reason for this is the highly imbalanced distribution of instances across different DDI events. To address this issue, further research should focus on sampling methods and algorithmic models to make them applicable to the problem of data imbalance. During the model training phase, over-sampling methods can be employed to increase the number of samples in the minority class. In model improvement, a direction we can explore is combining multiple different models or different ablations of the same model based on ensemble techniques to achieve better overall performance. On the other hand, existing models demonstrate poor performance in predicting DDIs between two new drugs (Task3), which is a critical aspect in drug discovery. Thus, it is imperative for forthcoming studies to prioritize resolving these 'cold start' challenges.

## 4. Materials and Methods

### 4.1. Datasets

In this study, we adopted two datasets with a different scale of DDI events. The first dataset (Dataset1) was collected by Deng et al. [34] Dataset1 contains 572 drugs with 37,264 pairwise drug–drug interaction (DDI) instances associated with 65 DDI events. Each drug in Dataset1 has four biological attributes: chemical substructure, target protein, transport pathway and enzyme, all of which are extracted from the DrugBank database [35]. The second dataset (Dataset2) was from the research by Lin et al. [7]. It contains 1258 drugs with 323,539 pairwise DDI instances associated with 100 DDI events. Each drug in Dataset2 has three attributes: chemical substructure, target protein and enzyme.

However, these two datasets just include drugs' direct biological attributes and ignore possible implicit indirect information, such as interactions between target proteins (PPIs). When a drug acts on a known target protein, it may alter another potential target protein through the effect of protein–protein interactions, leading to potential adverse reactions [36,37]. To learn more accurate representations of drug pairs, we integrated protein–protein interactions (PPIs) into the aforementioned two datasets. The PPIs dataset was sourced from the reference [38], which contains 4603 protein–protein interactions. The integrated datasets used in this work are shown in Table 5.

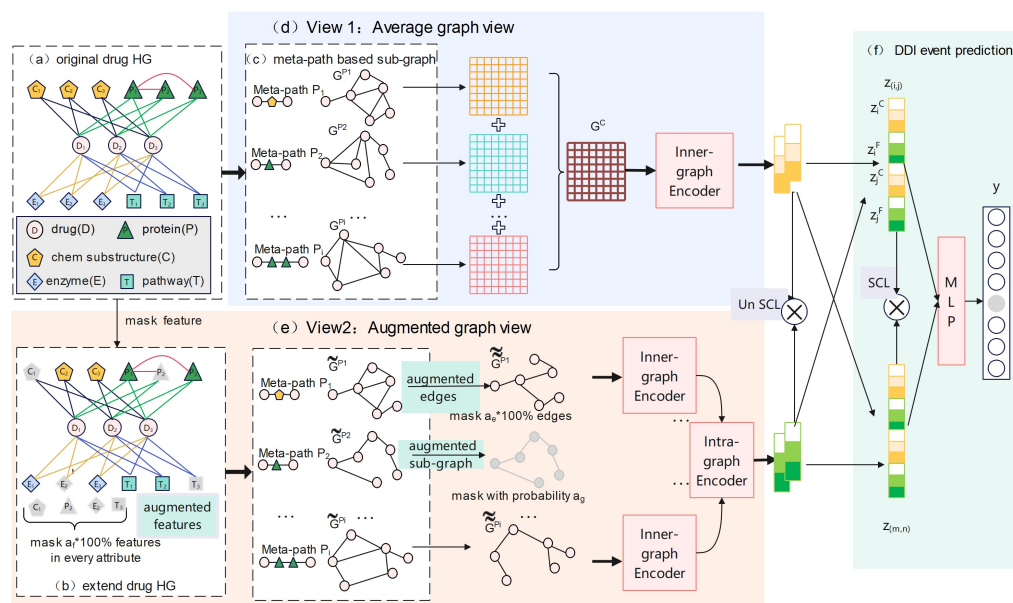
There are only a limited amount of labeled data available for each DDI event, with fewer than 100 labeled instances for 43 events in Dataset1. The frequency of each event in Dataset1 is depicted in Figure 1.

**Table 5.** Description of the two datasets.

Data Type	Dataset1	Dataset2
Drug number	572	1258
Drug-Drug interactions	37,264	323,539
Events	65	100
Drug-Chemical substructure relations	70,350	58,431
Drug-Target protein relations	3047	7386
Involved PPIs	2018	2346
Drug-Enzyme relations	2133	4479
Drug-Pathway relations	2778	Not included

#### 4.2. Methods

In this section, we introduce our proposed method, MPHGCL-DDI, as shown in Figure 7. The model consisted of six parts. Firstly, we constructed an original drug heterogeneous graph (drug HG, Figure 7a) based on each dataset shown in Table 5. Secondly, we masked certain features of biological attributes and constructed an extended drug heterogeneous graph (extended drug HG, Figure 7b). Thirdly, we constructed meta-path-based sub-graphs for the original drug HG and the extended drug HG, respectively (Figure 7c). Next, we designed three data augmentation schemes and constructed two contrastive views: an average graph view (Figure 7d) and an augmented graph view (Figure 7e). The former view is based on the original HG, while the latter view is based on augmentation schemes. Finally, we encoded the drugs of the two views and concatenated them for DDI event prediction (Figure 7f).

**Figure 7.** The overall framework of the MPHGCL-DDI model.

##### 4.2.1. Drug Heterogeneous Graph

Multi-source drug information describes different aspects of drugs and forms a typical heterogeneous graph. A heterogeneous graph (HG) is defined as a graph,  $G = (V, E)$ , where  $V$  and  $E$  denote the sets of nodes and edges, respectively. The number of nodes and edges are represented as  $|V|$  and  $|E|$ , respectively. Nodes and edges are associated with a node-type mapping function,  $\phi : V \rightarrow A$ , and an edge-type mapping function,  $\varphi : E \rightarrow R$ , respectively, where  $A$  and  $R$  denote the sets of node and edge types, respectively, and  $|A| > 1$  or  $|R| > 1$ .

We constructed a drug heterogeneous graph (drug HG), denoted as  $G = (V, E)$ , shown in Figure 7a, which contains five types of nodes: drug ( $D$ ), chemical substructure

(C), target protein (P), enzyme (E) and transport pathway (T). The edge set of the drug HG is  $R = \{D - C, D - P, D - E, D - T, P - P\}$ . Each type of edge represents a kind of relations between biological attributes.

#### 4.2.2. Meta-Path-Based Sub-Graph

After constructing the drug heterogeneous graph, we further defined multi-scale meta-paths between drugs to measure the similarity of different drugs.

A meta-path,  $P$ , is defined as a pattern of paths in the form of  $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_l} A_{l+1}$  (abbreviated as  $A_1 A_2 \dots A_{l+1}$ ), where  $A_i \in A, R_i \in R$ .

In our proposed model, we considered four initial 2-hop meta-paths  $\{P_1 : DCD, P_2 : DPD, P_3 : DED, P_4 : DTD\}$  and one 3-hop meta-path  $\{P_5 : DPPD\}$ . The 2-hop meta-paths pay attention to the direct relations between drugs through their biological attributes. However, learning short-chain information solely from 2-hop meta-paths may neglect the interactions between proteins. It may fail to capture the high-level complex semantic information of drugs. So we introduced a 3-hop meta-path,  $P_5 : DPPD$ , which integrated the PPIs information.

Here, we utilized the PathSim [38] to measure the similarity of nodes and constructed meta-path-based drug sub-graphs. For each pair of nodes,  $i$  and  $j$ , in an HG, if a meta-path,  $P$ , has start node and end node  $i$  and  $j$ , respectively, then the PathSim  $S(i, j)$  with respect to meta-path  $P$  is defined as Equation (1).

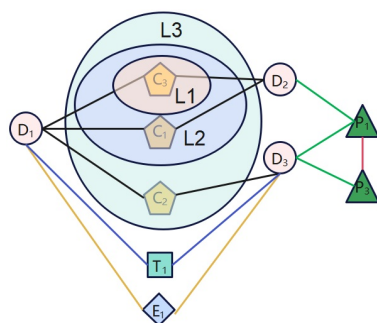
$$s(i, j) = \frac{2 \times |\{p_{i \rightarrow j} : p_{i \rightarrow j} \in P\}|}{|\{p_{i \rightarrow i} : p_{i \rightarrow i} \in P\}| + |\{p_{j \rightarrow j} : p_{j \rightarrow j} \in P\}|} \quad (1)$$

where  $p_{i \rightarrow j}$  is a path instance between  $i$  and  $j$ .

Based on the similarities, for each drug, we selected its top-K neighbors with the largest similarity. The removal of loosely connected neighbors can significantly reduce the number of neighbors for each drug, which further improves the model's performance. After neighbor filtering, we constructed meta-path-based sub-graphs, in which all nodes are drugs. Given a set of meta-paths  $P_1, P_2, \dots, P_l$ , the set of sub-graphs is denoted as  $S = \{G^{P_1}, G^{P_2}, \dots, G^{P_l}\}$ .  $G^{P_i}$  represents the drug sub-graph based on meta-path  $P_i$ , and its induced adjacency matrix is denoted as  $M_{DD}^{P_i}$ . The elements of matrix  $M_{DD}^{P_i}$  are 0 or 1, depending on whether two drugs are neighbors after neighbor filtering.

#### 4.2.3. Augmentation Schemes

In this work, we adopted three levels of augmentation schemes: feature augmentation, edge augmentation and sub-graph augmentation. The three kinds of data augmentation scheme differ in the range of the information perturbation, as shown in Figure 8.



**Figure 8.** Illustration of three levels of augmentation schemes.

- **Level 1: Feature augmentation**  
At this level, we masked some features of every biological attribute. For each biological attribute of drugs, we randomly removed  $a_f \times 100\%$  features, where  $a_f$  is the masking

ratio, and obtained an extended corrupted drug heterogeneous graph,  $\tilde{G}$ . For example, we masked chemical substructure  $C_1$ , as shown in Figure 8. In this way, a path,  $D_1C_1D_2$ , between drug  $D_1$  and  $D_2$ , which belongs to meta-path  $DCD$ , is removed.

- **Level 2: Edge augmentation**  
At this level, we masked some edges in each meta-path-based drug sub-graph. As illustrated in Figure 8, the edge between drug  $D_1$  and drug  $D_2$  based on meta-path  $DCD$  consists of two paths:  $D_1C_1D_2$  and  $D_1C_2D_2$ . So, at level 2, all paths between two drug pairs based on a meta-path are masked. Specifically, for each meta-path-based sub-graph,  $\tilde{G}^P$ , in  $\tilde{S}$ , we randomly removed  $a_e \times 100\%$  edges in  $\tilde{G}^P$ , where  $a_e$  is the masking ratio, and we denoted the corrupted meta-path-based sub-graph as  $\tilde{\tilde{G}}^P$ .
- **Level 3: Sub-graph augmentation**  
In this augmentation scheme, we masked one meta-path-based sub-graph in  $\tilde{S}$ . As shown in Figure 8, we masked the meta-path  $DCD$  from the meta-path set. Consequently, the sub-graph based on meta-path  $DCD$  is removed. Considering this mask scheme as a coarse-grained graph perturbation, we only performed it with a certain probability at each training stage, and only masked one sub-graph every time. Specifically, in each training epoch, we sample a number,  $r$ , from a Bernoulli distribution,  $Bernoulli(a_g)$ , where  $a_g$  is the masking probability. If  $r$  equals 1, we randomly select a sub-graph and remove it from  $\tilde{S}$ ; otherwise, the augmentation is not performed.

#### 4.2.4. Contrastive Views

We constructed two views for contrastive learning: an average graph view and an augmented graph view. The average graph view is built on the raw drug HG without considering the importance of different meta-paths. Alternatively, the augmented graph view contains three levels of the aforementioned augmentation, and pays attention to the importance of meta-paths.

- **Average graph view**  
The average graph view aggregates all meta-paths from the original drug HG  $G$ , as shown in Figure 7d. After sub-graph construction, defined in Section 4.2.2, we obtained a set of sub-graphs,  $S = \{G^{P_1}, G^{P_2}, \dots, G^{P_l}\}$ , where  $G^{P_i}$  represents the drug sub-graph based on meta-path  $P_i$ . The average graph view focuses on the connectivity between drugs and is a kind of coarse view. We adopted the average pooling operation to the meta-path-induced adjacency matrices and obtained an aggregated average graph,  $G^C$ . The adjacency matrix of  $G^C$  is

$$M^C = \frac{1}{l} \sum_{i=1}^l M_{DD}^{P_i} \quad (2)$$

where  $l$  is the number of meta-paths and  $M_{DD}^{P_i}$  is the adjacency matrix of sub-graph  $G^{P_i}$ .

- **Augmented graph view**  
The augmented graph view fuses all meta-path-based augmented sub-graphs by the attention mechanism (as shown in Figure 7e). This view pays attention to each meta-path's contextual semantic information and is a kind of fine-grained view. In each training epoch, we firstly masked features on original drug HG and constructed extended drug HG  $\tilde{G}$ . Based on  $\tilde{G}$ , a new set of sub-graphs can be formed following the description in Section 4.2.2. The set of sub-graphs is denoted as  $\tilde{S} = \{\tilde{G}^{P_1}, \tilde{G}^{P_2}, \dots, \tilde{G}^{P_l}\}$ , and  $\tilde{G}^{P_i}$  represents the new drug sub-graph based on meta-path  $P_i$ . And then we further performed sub-graph augmentation and edge augmentation on  $\tilde{S}$  and denoted the set of corrupted sub-graphs as  $\tilde{\tilde{S}} = \{\tilde{\tilde{G}}^{P_1}, \tilde{\tilde{G}}^{P_2}, \dots, \tilde{\tilde{G}}^{P_l}\}$ , where  $\tilde{\tilde{G}}^{P_i}$  is the corrupted sub-graph after edge augmentation based on meta-path  $P_i$ . The size of  $\tilde{\tilde{S}}$  is  $|S|$  or  $|S| - 1$ , depending on whether the sub-graph is performed.

#### 4.2.5. Graph Encoder

After the construction of the two contrastive views, we further introduced a graph encoder to obtain drug embeddings in every view, and this consists of three components: node feature transformation, inter-graph encoder and intra-graph encoder.

- Node Feature Transformation

For the drugs in the two datasets, we collected their SMILES strings from DrugBank and adopted the ESPF algorithm [39] to extract features from the SMILES strings. The ESPF algorithm is an effective technique that decomposes the sequential structure into interpretable functional groups. It decomposes a SMILES string into a set of different sized frequent substructures, starting from the collection of all atoms and bonds. To obtain more enriched representations in the high dimension than the original vector space, for an arbitrary drug,  $i$ , we have

$$h_i = \sigma(W \cdot X_i + b) \quad (3)$$

where  $X_i \in R^F$  is the exacted feature of drug  $i$  using the ESPF algorithm, and  $h_i \in R^d$  is the projected feature of drug  $i$ .  $W \in R^{d \times F}$  is the learnable mapping matrix,  $b \in R^d$  is the learnable bias vector and  $\sigma(\cdot)$  is an activation function, respectively.

- Inter-Graph Encoder

For the drug graph in the average graph view,  $G^C$ , or each augmented sub-graph  $\tilde{G}^{P_i} \in \tilde{\mathcal{S}}$  in the augmented graph view, we utilized the node attention mechanism in a graph to embed each drug.

For each node,  $i$ , in the graph  $G^C$ , we collected its neighbors,  $N_i^C$ , and aggregated the embeddings of nodes in  $N_i^C$ . As different neighbors exhibit different degrees of importance to the target node in contrastive tasks due to their distinct features, it is appropriate to assign different weights to them. We adopted a graph attention layer [40] to aggregate the embeddings of nodes in  $N_i^C$ , which leverages a self-attention mechanism to assign different weights to neighbors, effectively and flexibly capturing the importance of each neighbor.

Specifically, for node  $i$ , the importance of its neighbor node,  $j$ , is calculated as:

$$e_{ij}^C = \text{LeakyReLU}((a^C)^T \cdot [h_i || h_j]) \quad (4)$$

where  $a^C \in R^{2d}$  is a learnable node attention vector, and  $||$  denotes the concatenate operation.

After obtaining the importance of all neighbors for node  $i$ , we normalized them to obtain the weight coefficient and computed the weighted combination of the representations for node  $i$ :

$$a_{ij}^C = \frac{\exp(e_{ij}^C)}{\sum_{k \in N_i^C} \exp(e_{ik}^C)} \quad (5)$$

$$z_i^C = \text{PReLU}(\sum_{j \in N_i^C} a_{ij}^C \cdot h_j + b^C) \quad (6)$$

To strength the representation of embedding, we adopted a multi-head attention mechanism. Specifically,  $H$  independent attention mechanisms are executed, and the outputs are concatenated as the final node representation:

$$z_i^C = \big||_{k=1}^H \sigma(\sum_{j \in N_i^C} [a_{ij}^C]_k \cdot W_k^C \cdot h_j + b^C) \quad (7)$$

where  $W_k^C \in R^{d \times d}$  is the a transformation matrix for each head to keep the dimension of  $z_i^C$  to be  $d$ , and  $[a_{ij}^C]_k$  is the normalized importance of node  $j$  to node  $i$  at the  $k$ -th attention head.

Similarly, for each sub-graph,  $\tilde{G}^{P_i}$ , in the augmented graph view, we also applied the aforementioned multi-head node attention mechanism to obtain node representations for each sub-graph. The embedding of drug  $i$  in the sub-graph  $\tilde{G}^{P_i}$  based on meta-path  $P_i$ , denoted as  $z_i^{P_i}$ , can be calculated as:

$$z_i^{P_i} = \prod_{k=1}^H \sigma \left( \sum_{j \in N_i^{P_i}} [a_{ij}^{P_i}]_k \cdot W_k^{P_i} \cdot h_i + b^{P_i} \right) \quad (8)$$

- Intra-Graph Encoder

After encoding every drug in each meta-path-based drug sub-graph of the augmented graph view, we further integrated these embeddings to obtain the aggregated drug embedding for contrastive learning. Similar to the node attention mechanism in the aforementioned inner-graph encoder, different meta-paths represent different semantic information and have different importance. Therefore, we employed a graph-level attention mechanism to automatically learn the importance of different meta-paths. To be specific, for each meta-path-based drug sub-graph,  $\tilde{G}^{P_i} \in \tilde{\mathcal{S}}$ , we firstly calculated a summary vector by averaging the transformed node embeddings for all nodes in  $\tilde{G}^{P_i}$ :

$$s^{P_i} = \frac{1}{|D|} \sum_{i \in D} \tanh(W_1 \cdot z_i^{P_i} + b_1) \quad (9)$$

where  $D$  is the drug set,  $W_1 \in R^{d \times d}$  is a learnable weight matrix, and  $b_1 \in R^d$  is a learnable bias vector.

Secondly, we computed the importance of each meta-path utilizing a learnable vector,  $q_1 \in R^d$ , as follows:

$$e^{P_i} = q_1^T \cdot s^{P_i} \quad (10)$$

We normalized the importance score utilizing a Softmax function and obtained the weight co-efficient:

$$\beta^{P_i} = \frac{\exp(e^{P_i})}{\sum_{j=1}^{|\tilde{\mathcal{S}}|} \exp(e^{P_j})} \quad (11)$$

where  $|\tilde{\mathcal{S}}|$  means the number of sub-graphs.

Finally, we obtained the the aggregate final drug embedding of the augmented graph view:

$$z_i^F = \sum_{j=1}^{|\tilde{\mathcal{S}}|} \beta^{P_j} \cdot z_i^{P_j} \quad (12)$$

#### 4.2.6. Drug-Drug Interaction Event Prediction

For each drug pair  $(i, j)$ , we now have their embeddings in the average graph view,  $z_i^C$  and  $z_j^C$ , and their embeddings in the augmented graph view,  $z_i^F$  and  $z_j^F$ . We concatenated them to form a drug pair representation as following:

$$z_{(i,j)} = z_i^C || z_i^F || z_j^C || z_j^F \quad (13)$$

After that,  $z_{(i,j)}$  was fed into a multi-layer perceptron (MLP) followed by a Softmax function to obtain the multi-class prediction of the drug pair:

$$\hat{y}_{(i,j)} = \text{Softmax}(\text{MLP}(z_{(i,j)})) \quad (14)$$

where  $\hat{y}_{(i,j)} \in R^{|E|}$ , and  $|E|$  is the number of DDI events.

#### 4.2.7. Model Training

During model training, we optimized model parameters using a combined loss function, which consists of three parts: unsupervised contrastive loss, supervised contrastive loss and prediction loss.

- Unsupervised Contrastive Loss

After the graph encoder described in Section 4.2.5, we obtained the embeddings of the two contrastive views,  $z_i^C$  and  $z_i^F$ . In multi-view graph contrastive learning, a contrastive objective is adopted to distinguish the embeddings of the same node from other node embeddings. The InfoNCE loss function [41] is a commonly adopted contrastive objective and defined as:

$$L_i = -\log \frac{\exp(\text{sim}(z_i, z_j) / \tau)}{\sum_{k \in N} \exp(\text{sim}(z_i, z_k) / \tau)} \quad (15)$$

where  $\text{sim}(z_i, z_j)$  measures the similarity between node embeddings  $z_i$  and  $z_j$ ,  $\tau$  is a temperature hyper-parameter and  $N$  is the negative samples set.

The contrastive loss in this work is:

$$L_{uc} = \frac{1}{2|D|} \left( \sum_{i=1}^{|D|} -\log \frac{\exp(\text{sim}(z_i^C, z_i^F) / \tau)}{\sum_{k=1}^{|D|} \exp(\text{sim}(z_i^C, z_k^F) / \tau)} + \sum_{i=1}^{|D|} -\log \frac{\exp(\text{sim}(z_i^F, z_i^C) / \tau)}{\sum_{k=1}^{|D|} \exp(\text{sim}(z_i^F, z_k^C) / \tau)} \right) \quad (16)$$

where  $|D|$  is the size of the drug set. The first term in the parentheses represents the contrastive loss under the average graph view, while the second term is the loss under the augmented graph view.

- Supervised Contrastive Loss

Given that DDI event prediction is a multi-class classification task, supervised contrastive learning can learn more comprehensive drug embeddings. Herein, we further designed a supervised contrastive learning method to learn the embeddings of drug pairs. The latent features of drug pairs obtained by supervised contrastive learning have the following property: the embeddings of drug pairs belonging to same event are more similar, while the embeddings of different events are more different.

The loss function of supervised contrastive learning can be computed using the following equation:

$$L_{sc} = \frac{1}{N_{batchsize}} \sum_{(i,j) \in batch} \frac{1}{N_{e(i,j)}} \sum_{(m,n) \in e(i,j)} -\log \frac{\exp(\text{sim}(z_{(i,j)}, z_{(m,n)}) / \tau)}{\sum_{(u,v) \in batch} \exp(\text{sim}(z_{(i,j)}, z_{(u,v)}) / \tau)} \quad (17)$$

where,  $z(i, j)$  is the embedding of drug pair  $(i, j)$ , which is obtained by Equation (13),  $e(i, j)$  is the DDI event label of drug pair  $(i, j)$ , and  $N_{e(i,j)}$  is the number of instances in event  $e(i, j)$ .

- Prediction loss

The training objective of DDI event prediction is to minimize the loss function:

$$L_s = - \sum_{(i,j) \in \Omega} \sum_{e \in E} y_{(i,j)}^e \log \hat{y}_{(i,j)}^e \quad (18)$$

where  $\Omega$  is the training set,  $E$  is the events set,  $\hat{y}_{(i,j)}^e$  indicates the predicted probability that the drug pair  $(i, j)$  belongs to the event  $e$ , and  $y_{(i,j)}^e$  is the corresponding true label.

- Total Loss

For training our model MPHGL-DDI, we optimized the total loss that combines Equations (16)–(18):

$$L = L_s + \alpha(L_{uc} + L_{sc}) \quad (19)$$

where  $\alpha$  is a hyper-parameter that balances the contributions of contrastive loss and the prediction loss.

## 5. Conclusions

In this paper, we proposed a multi-relational DDI event prediction model based on heterogeneous graph contrastive learning (MPHGCL-DDI). This model constructed two meta-path-based contrastive views: an average graph view and an augmented graph view. The two views revealed the relations between drugs from different perspectives.

We defined three data augmentation schemes with different levels of data perturbation and adopted a combination loss, consisting of unsupervised contrastive loss, supervised contrastive loss and prediction loss, to train the model. Additionally, protein–protein interactions were integrated into the datasets to learn more effective representations of drug pairs. Experimental results demonstrated that our proposed model outperformed state-of-the-art models. We also conducted case studies to identify new DDIs not included in the current dataset. The actual case results further supported the effectiveness of the model in DDI event prediction.

**Author Contributions:** B.H. guided the research and methodology; Z.Y. collected the data and wrote the manuscript; M.L. performed the experiments. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work is partially supported by the Shandong Natural Science Foundation (No. ZR2023MF037), China Ministry of Education Industry–Academia Collaboration Collaborative Education Project (No. 231107099010537) and Shandong Women’s University High-level Talent Introduction Research Project (No. 2020RCYJ18).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets that support the findings of this study are available from the corresponding authors upon reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Giacomini, K.M.; Krauss, R.M.; Roden, D.M.; Eichelbaum, M.; Hayden, M.R.; Nakamura, Y. When good drugs go bad. *Nature* **2007**, *446*, 975–977. [[CrossRef](#)]
2. Bansal, M.; Yang, J.; Karan, C.; Menden, M.P.; Costello, J.C.; Tang, H.; Xiao, G.; Li, Y.; Allen, J.; Zhong, R.; et al. A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* **2014**, *32*, 1213–1222. [[PubMed](#)]
3. Qato, D.M.; Wilder, J.; Schumm, L.P.; Gillet, V.; Alexander, G.C. Changes in prescription and over-the-counter medication and dietary supplement use among older adults in the United States, 2005 vs. 2011. *JAMA Intern. Med.* **2016**, *176*, 473–482. [[CrossRef](#)] [[PubMed](#)]
4. Qiu, Y.; Zhang, Y.; Deng, Y.; Liu, S.; Zhang, W. A comprehensive review of computational methods for drug–drug interaction detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *19*, 1968–1985. [[CrossRef](#)] [[PubMed](#)]
5. Ryall, K.A.; Tan, A.C. Systems biology approaches for advancing the discovery of effective drug combinations. *J. Cheminform.* **2015**, *7*, 7. [[CrossRef](#)]
6. Yap, T.A.; Smith, A.D.; Ferraldeschi, R.; Al-Lazikani, B.; Workman, P.; De Bono, J.S. Drug discovery in advanced prostate cancer: Translating biology into therapy. *Nat. Rev. Drug Discov.* **2016**, *15*, 699–718. [[CrossRef](#)] [[PubMed](#)]
7. Lin, S.; Wang, Y.; Zhang, L.; Chu, Y.; Liu, Y.; Fang, Y.; Jiang, M.; Wang, Q.; Zhao, B.; Xiong, Y.; et al. MDF-SA-DDI: Predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Brief. Bioinform.* **2022**, *23*, bbab421. [[CrossRef](#)] [[PubMed](#)]
8. Zhang, J.; Chen, M.; Liu, J.; Peng, D.; Dai, Z.; Zou, X.; Li, Z. A knowledge-graph-based multimodal deep learning framework for identifying drug–drug interactions. *Molecules* **2023**, *28*, 1490. [[CrossRef](#)] [[PubMed](#)]
9. Salamat, A.; Luo, X.; Jafari, A. HeteroGraphRec: A heterogeneous graph-based neural networks for social recommendations. *Knowl.-Based Syst.* **2021**, *217*, 106817. [[CrossRef](#)]
10. Hu, B.; Wang, H.; Wang, L. Wshe: User feedback-based weighted signed heterogeneous information network embedding. *Inf. Sci.* **2021**, *579*, 167–185. [[CrossRef](#)]
11. Wang, X.; Bo, D.; Shi, C.; Fan, S.; Ye, Y.; Philip, S.Y. A survey on heterogeneous graph embedding: Methods, techniques, applications and sources. *IEEE Trans. Big Data* **2022**, *9*, 415–436. [[CrossRef](#)]
12. Mukande, T. Heterogeneous Graph Representation Learning for multi-target Cross-Domain Recommendation. In Proceedings of the 16th ACM Conference on Recommender Systems, Seattle, WA, USA, 18–23 September 2022; pp. 730–734.
13. Tanvir, F.; Islam, M.I.K.; Akbas, E. Predicting drug–drug interactions using meta-path based similarities. In Proceedings of the 2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Melbourne, Australia, 13–15 October 2021; pp. 1–8.
14. Li, M.; Cai, X.; Li, L.; Xu, S.; Ji, H. Heterogeneous graph attention network for drug–target interaction prediction. In Proceedings of the 31st ACM International Conference on Information Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 1166–1176.
15. Liu, W.; Tang, T.; Lu, X.; Fu, X.; Yang, Y.; Peng, L. MPCLCDA: predicting circRNA–disease associations by using automatically selected meta-path and contrastive learning. *Briefings Bioinform.* **2023**, *24*, bbad227. [[CrossRef](#)] [[PubMed](#)]

16. You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; Shen, Y. Graph contrastive learning with augmentations. In Proceedings of the Advances in Neural Information Processing Systems 33 (NeurIPS 2020), Virtual, 6–12 December 2020; pp. 5812–5823.
17. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. In Proceedings of the Advances in Neural Information Processing Systems 33 (NeurIPS 2020), Virtual, 6–12 December 2020; pp. 18661–18673.
18. Wang, X.; Liu, N.; Han, H.; Shi, C. Self-supervised heterogeneous graph neural network with co-contrastive learning. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining, Virtual Event, Singapore, 14–18 August 2021; pp. 1726–1736.
19. Zhu, Y.; Xu, Y.; Cui, H.; Yang, C.; Liu, Q.; Wu, S. Structure-enhanced heterogeneous graph contrastive learning. In Proceedings of the 2022 SIAM International Conference on Data Mining (SDM), Alexandria, VA, USA, 28–30 April 2022; pp. 82–90.
20. Chen, X.; Wang, Y.; Fang, J.; Meng, Z.; Liang, S. Heterogeneous Graph Contrastive Learning with Metapath-Based Augmentations. *IEEE Trans. Emerg. Top. Comput. Intell.* **2023**, *8*, 1003–1014. [[CrossRef](#)]
21. Yu, J.; Ge, Q.; Li, X.; Zhou, A. Heterogeneous Graph Contrastive Learning with Meta-Path Contexts and Adaptively Weighted Negative Samples. *IEEE Trans. Knowl. Data Eng.* **2024**, 1–13. [[CrossRef](#)]
22. Wang, Y.; Min, Y.; Chen, X.; Wu, J. Multi-view graph contrastive representation learning for drug-drug interaction prediction. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 2921–2933.
23. Zhang, R.; Wang, X.; Wang, P.; Meng, Z.; Cui, W.; Zhou, Y. HTCL-DDI: A hierarchical triple-view contrastive learning framework for drug-drug interaction prediction. *Brief. Bioinform.* **2023**, *24*, bbad324. [[CrossRef](#)] [[PubMed](#)]
24. Zhuang, L.; Wang, H.; Hua, M.; Li, W.; Zhang, H. Predicting drug-drug adverse reactions via multi-view graph contrastive representation model. *Appl. Intell.* **2023**, *53*, 17411–17428. [[CrossRef](#)]
25. Lin, S.; Chen, W.; Chen, G.; Zhou, S.; Wei, D.Q.; Xiong, Y. MDDI-SCL: Predicting multi-type drug-drug interactions via supervised contrastive learning. *J. Cheminform.* **2023**, *14*, 81. [[CrossRef](#)] [[PubMed](#)]
26. Jiang, M.; Liu, G.; Zhao, B.; Su, Y.; Jin, W. Relation-aware graph structure embedding with co-contrastive learning for drug-drug interaction prediction. *Neurocomputing* **2024**, *572*, 127203. [[CrossRef](#)]
27. Suresh, S.; Li, P.; Hao, C.; Neville, J. Adversarial graph augmentation to improve graph contrastive learning. In Proceedings of the Advances in Neural Information Processing Systems 34 (NeurIPS 2021), Virtual, 6–14 December 2021; pp. 15920–15933.
28. Xiong, Z.; Liu, S.; Huang, F.; Wang, Z.; Liu, X.; Zhang, Z.; Zhang, W. Multi-relational contrastive learning graph neural network for drug-drug interaction event prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 5339–5347.
29. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010), Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
30. Kingma, P.D.; Ba, L.J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
31. Feng, J.; Liang, Y.; Yu, T. MM-GANN-DDI: Multimodal graph-agnostic neural networks for predicting drug-drug interaction events. *Comput. Biol. Med.* **2023**, *166*, 107492. [[CrossRef](#)]
32. Han, C.D.; Wang, C.C.; Huang, L.; Chen, X. MCFE-MTDDI: Multi-channel feature fusion for multi-typed drug-drug interaction prediction. *Brief. Bioinform.* **2023**, *24*, bbad215. [[CrossRef](#)] [[PubMed](#)]
33. Zhao, W.; Yuan, X.; Shen, X.; Jiang, X.; Shi, C.; He, T.; Hu, X. Improving drug-drug interactions prediction with interpretability via meta-path-based information fusion. *Brief. Bioinform.* **2023**, *24*, bbad041. [[CrossRef](#)] [[PubMed](#)]
34. Deng, Y.F.; Xu, X.R.; Qiu, Y.; Xia, J.B.; Zhang, W.; Liu, S.C. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics* **2020**, *36*, 4316–4322. [[CrossRef](#)] [[PubMed](#)]
35. Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A.C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; et al. DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res.* **2014**, *42*, 1091–1097. [[CrossRef](#)] [[PubMed](#)]
36. Zitnik, M.; Agrawal, M.; Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* **2018**, *34*, i457–i466. [[CrossRef](#)]
37. Park, K.; Kim, D.; Ha, S.; Lee, D. Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks. *PLoS ONE* **2015**, *10*, e0140816. [[CrossRef](#)] [[PubMed](#)]
38. Sun, Y.; Han, J.; Yan, X.; Yu, P.S.; Wu, T. PathSim: Meta path-based top-K similarity search in heterogeneous information networks. In Proceedings of the VLDB Endowment 2011, Istanbul, Turkey, 31 October–2 November 2011; pp. 992–1003.
39. Huang, K.; Xiao, C.; Glass, L.; Sun, J. Explainable substructure partition fingerprint for protein, drug, and more. In Proceedings of the NeurIPS Learning Meaningful Representation of Life Workshop, 2019. Available online: <https://static1.squarespace.com/static/58f7aae1e6f2e1a0f9a56616/t/5e370e2d12092f15876d5753/1580666413389/paper.pdf> (accessed on 1 May 2022).
40. Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
41. Oord, A.V.D.; Li, Y.; Vinyals, O. Representation learning with contrastive predictive coding. *arXiv* **2018**, arXiv:1807.03748.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.