

METHODOLOGY

Open Access



# Drug-target interaction prediction with collaborative contrastive learning and adaptive self-paced sampling strategy

Zhen Tian<sup>1,2</sup>, Yue Yu<sup>1,2</sup>, Fengming Ni<sup>3\*</sup> and Quan Zou<sup>2</sup>

## Abstract

**Background** Drug-target interaction (DTI) prediction plays a pivotal role in drug discovery and drug repositioning, enabling the identification of potential drug candidates. However, most previous approaches often do not fully utilize the complementary relationships among multiple biological networks, which limits their ability to learn more consistent representations. Additionally, the selection strategy of negative samples significantly affects the performance of contrastive learning methods.

**Results** In this study, we propose CCL-ASPS, a novel deep learning model that incorporates Collaborative Contrastive Learning (CCL) and Adaptive Self-Paced Sampling strategy (ASPS) for drug-target interaction prediction. CCL-ASPS leverages multiple networks to learn the fused embeddings of drugs and targets, ensuring their consistent representations from individual networks. Furthermore, ASPS dynamically selects more informative negative sample pairs for contrastive learning. Experiment results on the established dataset demonstrate that CCL-ASPS achieves significant improvements compared to current state-of-the-art methods. Moreover, ablation experiments confirm the contributions of the proposed CCL and ASPS strategies.

**Conclusions** By integrating Collaborative Contrastive Learning and Adaptive Self-Paced Sampling, the proposed CCL-ASPS effectively addresses the limitations of previous methods. This study demonstrates that CCL-ASPS achieves notable improvements in DTI predictive performance compared to current state-of-the-art approaches. The case study and cold start experiments further illustrate the capability of CCL-ASPS to effectively predict previously unknown DTI, potentially facilitating the identification of new drug-target interactions.

**Keywords** Drug-Target Interaction, Contrastive Learning, Self-Paced Sampling

## Background

The discovery of novel drug-target interactions is at the heart of pharmaceutical research, driving the development of innovative therapies and the realization of personalized medicine [1]. Traditional wet-lab methods are expensive and time-consuming, whereas computational-based methods significantly improve drug candidate identification efficiency [2]. Consequently, computational-based methods have received increasing research interest. A key challenge lies in leveraging the complementary relationships encoded within multiple accessible

\*Correspondence:

Fengming Ni  
n\_fengming@jlu.edu.cn

<sup>1</sup> School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, Henan, China

<sup>2</sup> Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou, China

<sup>3</sup> Department of Gastroenterology, The First Hospital of Jilin University, Changchun 130021, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

interaction and association networks related to drugs and targets (proteins).

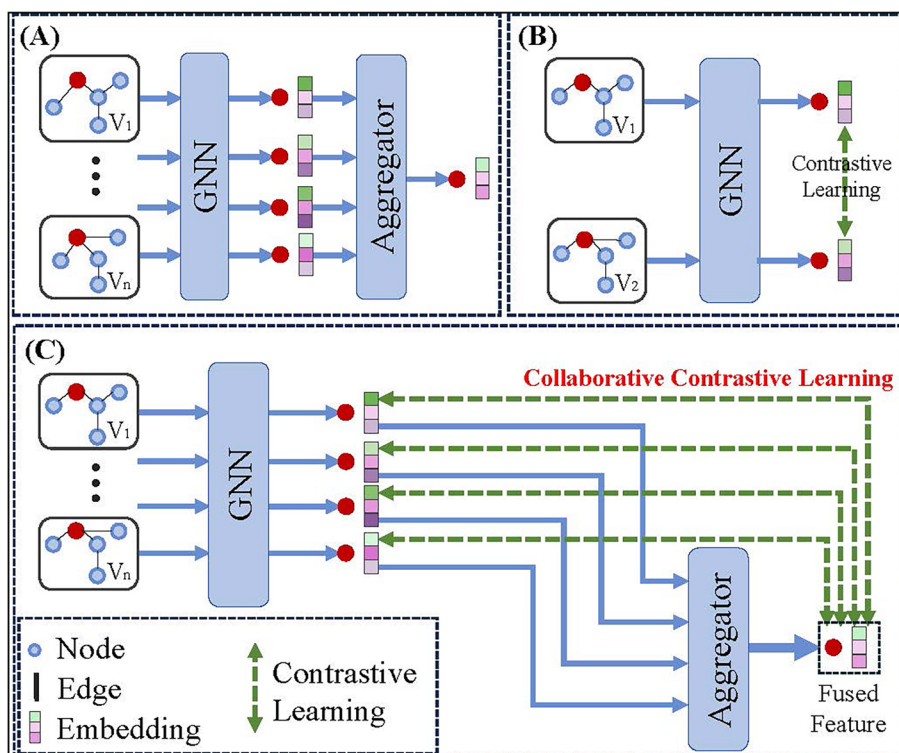
The embedding learning of drugs and targets is crucial for DTI prediction. Current methods typically learn representations from each network independently, followed by fusion via concatenation [3–6], average with attention mechanisms [7, 8], as displayed in Fig. 1A. Although these approaches utilize multiple networks, they learn embeddings separately, neglecting the complementary relationships among multiple networks. This hinders the learning of consistent drug and target representations. Recent work has demonstrated the effectiveness of contrastive learning (CL) in learning consistent representations across different networks [9]. However, traditional contrastive learning approaches are primarily designed for two networks (Fig. 1B), limiting their applicability in scenarios involving multiple networks. This study addresses the challenge by leveraging contrastive learning to collaboratively obtain embeddings across multiple networks, the brief illustration is presented in Fig. 1C. The proposed model enables the learning of fused embeddings that are consistent with their individual network-derived representations. Meanwhile, traditional contrastive learning methods typically treat identical nodes across views as positive samples, while other node

pairs as negative samples [10]. This strategy fails to fully exploit the structural information within individual networks and the complementary relationships between multiple networks. To overcome this drawback, we propose the adaptive self-paced sampling strategy.

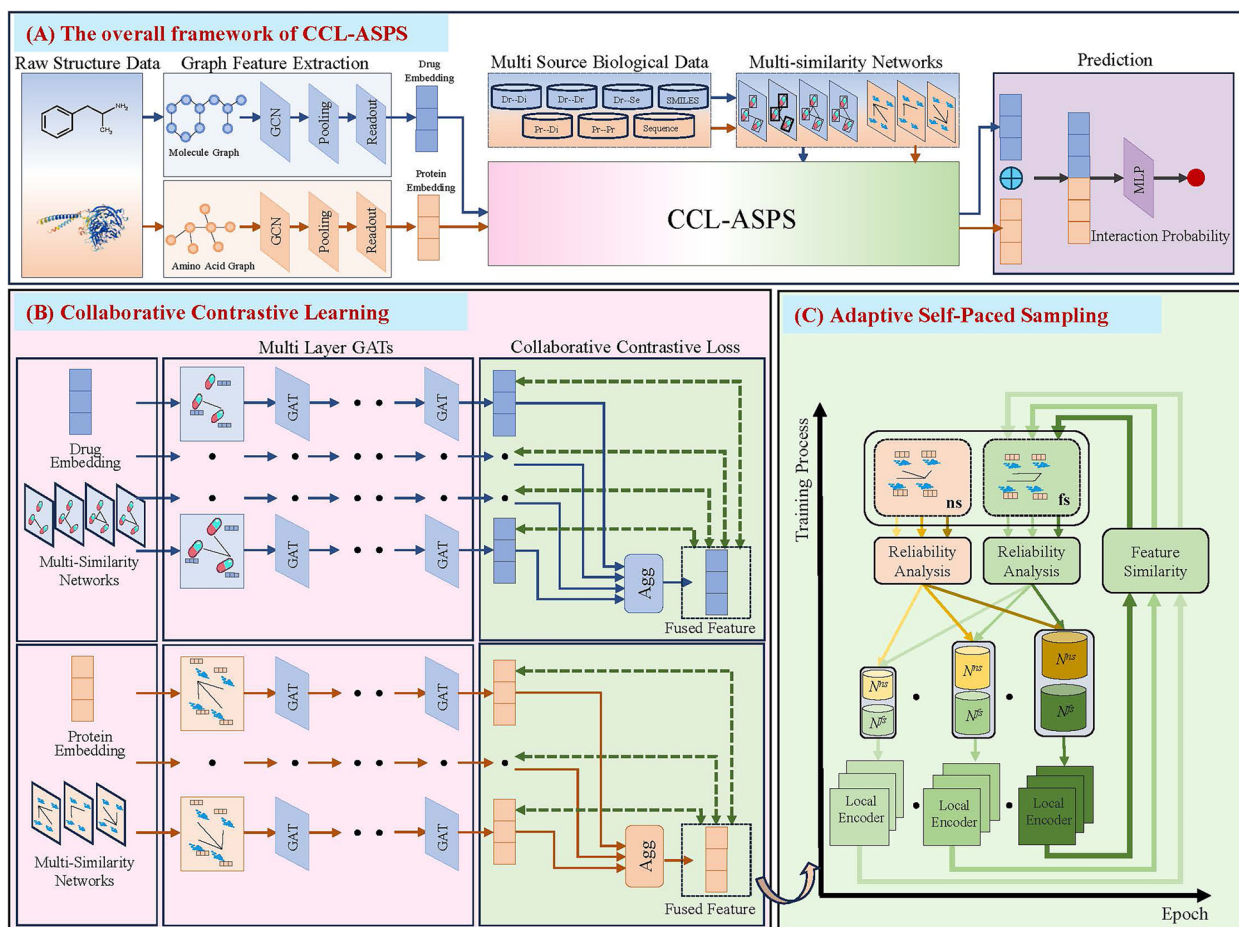
In this study, we present Collaborative Contrastive Learning with Adaptive Self-Paced Sampling strategy to predict drug-target interactions, as displayed in Fig. 2A. Initially, CCL-ASPS learns the embeddings of drugs and targets from their corresponding 2D graph structures. Subsequently, these learned embeddings are leveraged as inputs for the collaborative contrastive learning module. Furthermore, CCL-ASPS incorporates the adaptive self-paced sampling strategy to select more informative negative samples for contrastive learning. Finally, a multilayer perceptron (MLP) decoder is employed to predict potential DTIs. Extensive experiment results demonstrate that CCL-ASPS achieves optimal performance compared to established baselines.

In summary, the main contributions of CCL-ASPS are as follows:

- CCL-ASPS applies the collaborative contrastive learning strategy to learn more consistent representations of drugs and targets. To the best of our



**Fig. 1** **A** Representation learning by the traditional concatenation operation. **B** Representation learning by the typical contrastive learning strategy. **C** Representation learning by our proposed collaborative contrastive learning



**Fig. 2** The workflow of CCL-ASPS. **A** Overall framework. First, CCL-ASPS applies graph neural networks to extract drug and target representations from molecule and amino acid residue graphs, respectively. Afterward, these representations and similarity networks are fed into the core CCL-ASPS module. Finally, the learned representations from the CCL-ASPS module are utilized by an MLP decoder to predict potential drug-target interactions. **B** Collaborative Contrastive Learning. This module employs multi-layer GATs to learn drug and target embeddings from each similarity network. The learned embeddings are then aggregated to generate fused representations. To ensure consistency of the fused representations, the collaborative contrastive loss function is applied to guide the training process. **C** Adaptive Self-Paced Sampling. This strategy dynamically selects more informative negative samples for contrastive learning. Initially, drug (target) feature similarities are measured based on the fused representations. In the subsequent step, reliability scores are calculated for each negative sample pair based on individual network similarity (*ns*) and fused representation similarity (*fs*). Ultimately, samples with high reliabilities are selected in a self-paced manner

knowledge, this is the first attempt to learn consistent features from multiple networks collaboratively.

- CCL-ASPS employs the adaptive self-paced sampling strategy to select more informative negative samples for contrastive learning.
- Extensive experiment results demonstrate that CCL-ASPS achieves state-of-the-art performance compared to established baselines.

**Related work**

**DTI prediction**

Computational-based approaches for predicting drug-target(protein) interactions can be broadly classified

into three categories based on the utilized data type: structure type-based approaches, network type-based approaches and hybrid approaches. Structure type-based approaches typically extract drug representations from SMILES strings and protein representations from amino acid sequences [11–14]. For instance, DrugBAN [13] leverages a structure-based representation learning scheme, capturing individual atom and amino acid features to derive joint drug-protein representations. Ru [15] first extracts self-associated features (SAFs) of drugs and targets from similarity and sharing networks, then obtains adjacent-associated features (AAFs) based on their neighbors. Modality-DTA utilizes six modalities of

drug SMILES and protein sequences to capture informative drug and protein representations [16]. Network type-based approaches focus on learning representations by constructing similarity networks from interaction and association data related to drugs and proteins [3, 7, 17]. Luo proposes DTINet [17], which constructs multiple drug and protein similarity networks based on association and interaction networks. This method then utilizes a compact representation learning algorithm to extract low-dimensional embeddings. Hybrid approaches aim to combine the strengths of both structure and network information. Some studies learn representations from structure and network data in parallel, followed by integration operations to combine the learned features [4, 18, 19]. Alternatively, sequential learning strategies involve an initial step of extracting representations from structure data, followed by feeding them into a network data-based encoder for further refinement [20–22]. Although structure type-based approaches and network type-based approaches have made significant advancements, structure type-based approaches often overlook the extensive network data present in biological information, while network type-based approaches fail to fully exploit the detailed biological structure information. These limitations can restrict their predictive capabilities. While hybrid methods leverage both types of data, existing approaches often fail to fully consider the complementary relationships across multiple networks.

### Contrastive learning view generation

Inspired by the successful application in computer vision [23–25], contrastive learning has attracted widespread interest in graph-level representation learning and biological entity interaction prediction [9, 10, 26]. Typically, contrastive learning constructs two augmented views and aims to bring positive sample representations closer while pushing negative sample representations further away. Many studies employ techs such as node dropping, edge perturbation, attribute masking, or subgraph on the singular data source to create CL views [27–29]. Several approaches attempt to obtain CL views directly from structure or interaction networks. For instance, SMGCL [9] utilizes drug structure to construct a drug similarity network as one CL view and applies the drug-disease interaction network as another. MCHNLDA [30] builds one representation structure graph and one lncRNA-gene-disease interaction network as two views. MOVE [31] takes sequence information as one view and then treats multiple interaction networks as another view. However, these methods either rely on one biological data source or combine multiple networks into a single CL view. As a result, they fail to fully exploit the complementary relationships across multiple networks.

### Contrastive learning sampling

Traditional contrastive learning approaches typically rely on a simple strategy where identical nodes across the two augmented views are considered positive pairs, while all other node pairs are treated as negative samples [10, 26, 31–33]. However, this strategy overlooks potentially informative similar node pairs, leading to suboptimal performance. To address this limitation, several studies have proposed more sophisticated sampling strategies. For instance, HeCo [34] constructs meta-paths through interaction networks and selects the top-k nodes with the highest number of meta-paths as positive samples. SGCL-DTI [35] constructs a semantic network based on node class information and selects first-order neighbors in this network as positive samples. SMGCL [9] calculates scores between nodes based on their learned representations and selects the top-k nodes with the highest scores as positive samples. Recent advancements incorporate curriculum learning into the sampling process. Curriculum learning imitates human learning by progressively introducing the model to more complex samples [36]. For instance, CuCo [29] introduces a score function that ranks negative samples from easiest to hardest, and a pace function that gradually increases the number of negative samples presented to the model during training.

In contrast to these existing methods, the proposed CCL-ASPS method adopts a two-stage approach to sample pair selection. Initially, it adaptively selects sample pairs based on node similarities within each individual network. Subsequently, it obtains additional informative samples by calculating similarities between the fused representations.

### Results

This section presents a comprehensive evaluation of the proposed CCL-ASPS method. All the experiments in this study were conducted with 10 replicates, and the figures present the averaged results from these replicates. The experimental setup is first established, detailing the parameter settings and evaluation metrics employed. Subsequently, CCL-ASPS is benchmarked against baseline methods. To further validate its effectiveness, an ablation study is conducted, and the impact of different feature combination strategies is evaluated. Additionally, the learned representations are visualized at various training stages. Furthermore, this study performs case study and cold start experiments, demonstrating the ability of CCL-ASPS to discover new drug-target interactions. Finally, the time complexity of CCL-ASPS is analyzed.

### Parameter setting and metrics

Key parameters of CCL-ASPS are set for optimal performance. The feature dimensions for drugs and proteins are set to 64, and the learning rate is set to 0.001. For the drug and protein graph structure feature extraction module, the number of GCN layers are both set to 1, the training epoch numbers are both set to 2000. For the collaborative contrastive learning module, the negative sample rate  $\beta$  is set to 0.8, the contrastive loss weight  $\gamma$  is set to 0.3, the dropout value is set to 0.2, the number of GAT layers is set to 2, the number of training epochs is set to 5000.

The performance of CCL-ASPS and baseline methods is evaluated with five metrics: Accuracy (ACC), Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPR), Matthews Correlation Coefficient (MCC), and F1-Score (F1).

Specifically, our implementation utilizes Python 3.10, with PyTorch 2.0 as the deep learning framework and PyG (PyTorch Geometric) for graph processing. Data handling and analysis were performed using NumPy and Pandas. Visualization was conducted using Matplotlib and Seaborn. The experiments were conducted on an NVIDIA GeForce RTX 2070 GPU within a Windows operating system environment. More specific implementation details with code and dataset can be obtained from [10.5281/zenodo.13329691](https://doi.org/10.5281/zenodo.13329691).

### Comparison with baseline methods

To demonstrate the effectiveness of CCL-ASPS, this study conducts comparative analyses with a set of established baseline methods.

GCN [37] leverages the inherent relationships within drug-target interaction networks, which makes it a valuable baseline for our task.

GAT [38] is a graph-based method that excels in capturing important relationships in graph data.

SVM [39] is a classic machine learning algorithm widely used in binary classification tasks.

RF [40] is one ensemble learning method that forms the final prediction by aggregating the predictions of multiple decision trees.

DTI-CNN [3] utilizes Random Walk with Restart and a denoising autoencoder to obtain low-dimensional vector representations of drugs and proteins.

IMCHGAN [41] leverages GAT to learn embeddings of drugs and targets from DTI heterogeneous networks, then fuses the embeddings through an attention mechanism.

DrugBAN [13] learns the joint representations of drug-protein pairs by calculating the weights between drug atoms and amino acid subsequences.

GraphDTA [42] learns drug representations based on drug molecular graphs and protein representations from protein sequences, respectively.

HyperAttentionDTI [14] learns the drug and protein subsequence embeddings from drug SMILES strings and protein sequences, respectively, and updates these representations by calculating the attention weights between them.

GraphCDR [28] applies Graph Neural Network (GNN) to extract biochemical features of drugs and cancers, and then utilizes contrastive learning to improve the generalization ability.

MSGCL [43] constructs an anchor view and a learner view based on multiple interactive networks, then applies contrastive loss to increase the consistency of features learned by the two views.

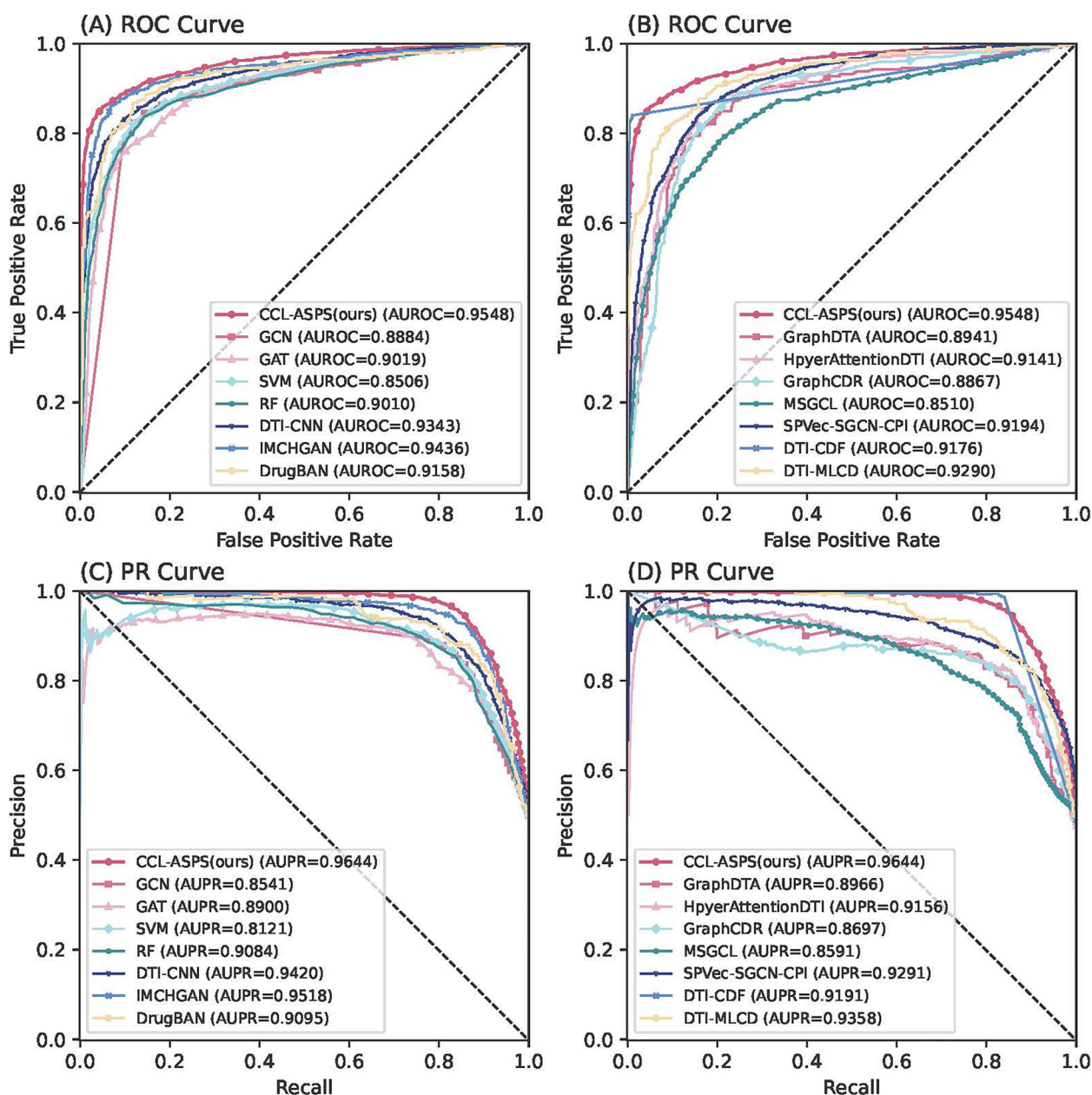
SPVec-SGCN-CPI [44] applies SPVec to learn compound and protein features from compound SMILESs and protein sequences, respectively.

DTI-CDF [45] constructs one fused similarity matrix to build a heterogeneous DTIs graph, then extracts path-category-based multi-similarities features of drug-target pairs.

DTI-MLCD [46] extracts drug features based on molecular descriptors, molecular fingerprints and Word2vec. Meanwhile, DTI-MLCD combines three sequence-derived features to obtain protein features.

The performance of CCL-ASPS is evaluated using a five-fold cross-validation strategy. Negative samples are randomly selected from the unlabeled sample set under the positive-negative sample ratio of 1:1, while both positive and negative samples are divided into training and testing sets. All compared baseline methods are evaluated on the same dataset using their optimized hyperparameters. Figure 3 presents the ROC curve and PR curve for CCL-ASPS and the baseline methods. We split the figure of AUROC results into 2 subfigures. Each subfigure contains the ROC Curves of CCL-ASPS and seven baselines. The figure of AUPR results are also divided into 2 subgraphs according to the same rules.

As presented in Fig. 3, CCL-ASPS consistently outperforms all comparison baselines. Specifically, Fig. 3A and B display the AUROC metrics for CCL-ASPS and other baselines. CCL-ASPS achieves the highest AUROC value of 0.9548, followed by IMCHGAN with an AUROC of 0.9436, and DTI-CNN with an AUROC of 0.9400. Figure 3C and D illustrate the AUPR metrics for CCL-ASPS and other baselines. Among all models, CCL-ASPS achieves the best AUPR result of 0.9644,



**Fig. 3** The ROC and PR Curves of the proposed model and all baselines

while IMCHGAN attains the second-best result with a value of 0.9518.

**Comparison with baseline methods under different ratios**

This experiment investigates the influence of positive-negative sample ratios on CCL-ASPS performance. Three ratios are evaluated: 1:1, 1:5, and 1:10. Negative samples are randomly selected from the unlabeled sample set at these predefined ratios.

Tables 1, 2, and 3 present the drug-target interaction prediction performance of CCL-ASPS and baseline methods under different ratios. For clarity, the best results are shown in bold, and the second-best results are underlined. This formatting scheme is applied across all tables. CCL-ASPS consistently outperforms all baselines across all five metrics at a 1:1 ratio. The results are displayed in Table 1. Meanwhile, DTI-CDF achieves the second-best results on three metrics (ACC, MCC, and F1), and IMCHGAN achieves

**Table 1** The performance of CCL-ASPS and baselines under ratio 1:1

Methods	ACC	AUROC	AUPR	MCC	F1
GCN [37]	0.8422 ± 0.007	0.8884 ± 0.002	0.8541 ± 0.016	0.6685 ± 0.036	0.8410 ± 0.006
GAT [38]	0.8371 ± 0.003	0.9019 ± 0.002	0.8900 ± 0.005	0.6767 ± 0.007	0.8316 ± 0.004
SVM [39]	0.8506 ± 0.002	0.8506 ± 0.002	0.8121 ± 0.003	0.7049 ± 0.004	0.8452 ± 0.002
RF [40]	0.8398 ± 0.004	0.9010 ± 0.003	0.9084 ± 0.004	0.6810 ± 0.009	0.8358 ± 0.004
DTI-CNN [3]	0.8650 ± 0.003	0.9343 ± 0.002	0.9420 ± 0.002	0.7309 ± 0.006	0.8664 ± 0.002
IMCHGAN [41]	0.8917 ± 0.002	<u>0.9436 ± 0.001</u>	<u>0.9518 ± 0.002</u>	0.7851 ± 0.003	0.8854 ± 0.009
DrugBAN [13]	0.8619 ± 0.010	0.9158 ± 0.003	0.9095 ± 0.010	0.7251 ± 0.020	0.8657 ± 0.009
GraphDTA [42]	0.8409 ± 0.001	0.8941 ± 0.001	0.8966 ± 0.006	0.6827 ± 0.003	0.8447 ± 0.005
HpyerAttentionDTI [14]	0.8499 ± 0.013	0.9141 ± 0.011	0.9156 ± 0.018	0.7014 ± 0.026	0.8523 ± 0.013
GraphCDR [28]	0.7931 ± 0.022	0.8867 ± 0.018	0.8697 ± 0.022	0.6015 ± 0.032	0.8172 ± 0.017
MSGCL [43]	0.7874 ± 0.042	0.8510 ± 0.022	0.8591 ± 0.021	0.5751 ± 0.065	0.8021 ± 0.031
SPVec-SGCN-CPI [44]	0.8476 ± 0.011	0.9194 ± 0.010	0.9291 ± 0.009	0.6901 ± 0.022	0.8659 ± 0.009
DTI-CDF [45]	<u>0.8922 ± 0.010</u>	0.9176 ± 0.011	0.9191 ± 0.012	<u>0.7939 ± 0.018</u>	<u>0.8865 ± 0.013</u>
DTI-MLCD [46]	0.8645 ± 0.012	0.9290 ± 0.008	0.9358 ± 0.011	0.7299 ± 0.024	0.8619 ± 0.011
<b>CCL-ASPS(ours)</b>	<b>0.8955 ± 0.003</b>	<b>0.9548 ± 0.003</b>	<b>0.9644 ± 0.002</b>	<b>0.8013 ± 0.005</b>	<b>0.8932 ± 0.003</b>

**Table 2** The performance of CCL-ASPS and baselines under ratio 1:5

Methods	ACC	AUROC	AUPR	MCC	F1
GCN [37]	0.9045 ± 0.002	0.8910 ± 0.003	0.7185 ± 0.011	0.6360 ± 0.008	0.6939 ± 0.007
GAT [38]	0.9076 ± 0.001	0.9081 ± 0.002	0.7099 ± 0.001	0.66520 ± 0.005	0.7076 ± 0.004
SVM [39]	0.9041 ± 0.004	0.7672 ± 0.006	0.5246 ± 0.018	0.6207 ± 0.017	0.6827 ± 0.015
RF [40]	0.9012 ± 0.003	0.9180 ± 0.002	0.7261 ± 0.013	0.6224 ± 0.011	0.6823 ± 0.010
DTI-CNN [3]	0.9358 ± 0.002	0.9431 ± 0.001	0.8648 ± 0.001	0.7637 ± 0.007	0.8025 ± 0.005
IMCHGAN [41]	<u>0.9419 ± 0.002</u>	<u>0.9502 ± 0.002</u>	<u>0.8750 ± 0.006</u>	0.7855 ± 0.008	0.8166 ± 0.009
DrugBAN [13]	0.8900 ± 0.005	0.9416 ± 0.003	0.8575 ± 0.025	0.6812 ± 0.009	<b>0.8905 ± 0.001</b>
GraphDTA [42]	0.9158 ± 0.002	0.9185 ± 0.003	0.7962 ± 0.009	0.6955 ± 0.015	0.7461 ± 0.014
HpyerAttentionDTI [14]	0.9105 ± 0.005	0.9265 ± 0.009	0.8039 ± 0.021	0.6882 ± 0.017	0.7414 ± 0.014
GraphCDR [28]	0.7895 ± 0.011	0.8982 ± 0.003	0.6342 ± 0.024	0.5075 ± 0.013	0.5768 ± 0.008
MSGCL [43]	0.8611 ± 0.014	0.8457 ± 0.024	0.6058 ± 0.029	0.3412 ± 0.010	0.6036 ± 0.029
SPVec-SGCN-CPI [44]	0.8807 ± 0.029	0.9001 ± 0.044	0.7822 ± 0.032	0.6379 ± 0.020	0.6914 ± 0.017
DTI-CDF [45]	0.9378 ± 0.002	0.8896 ± 0.009	0.8087 ± 0.018	<b>0.8094 ± 0.011</b>	0.8379 ± 0.011
DTI-MLCD [46]	0.9162 ± 0.003	0.9299 ± 0.007	0.7918 ± 0.014	0.6782 ± 0.015	0.7299 ± 0.012
<b>CCL-ASPS(ours)</b>	<b>0.9490 ± 0.001</b>	<b>0.9516 ± 0.001</b>	<b>0.8953 ± 0.002</b>	<u>0.8084 ± 0.004</u>	<u>0.8403 ± 0.004</u>

the second-best results on the AUROC and AUPR metrics. Under a positive-negative sample ratio of 1:5 (Table 2), CCL-ASPS maintains superior performance in ACC, AUROC and AUPR metrics, while achieving the second-best results in the MCC and F1 metrics. The best MCC and F1 results are achieved by DTI-CDF and DrugBAN, respectively. For the 1:10 positive-negative sample ratio (Table 3), CCL-ASPS maintains the same trend as the 1:5 ratio, with the best results in the ACC, AUROC and AUPR metrics, and the second best results in the MCC and F1 metrics. Similarly, the best MCC and F1 results are still achieved by DTI-CDF and DrugBAN.

### Ablation experiments

This section exploits the contributions of each component within CCL-ASPS through ablation experiments. CCL-ASPS consists of three key components: (1) drug and protein graph structure-based representations extractions, (2) collaborative contrastive learning, and (3) adaptive self-paced sampling. To assess the impact of each component, three ablation experiments are conducted: CCL-ASPS w/o GF (disables graph structure-based representations extraction), CCL-ASPS w/o CCL (disables collaborative contrastive learning), CCL-ASPS w/o ASPS (disables adaptive self-paced sampling). The adaptive self-paced sampling is part of the collaborative

**Table 3** The performance of CCL-ASPS and baselines under ratio 1:10

Methods	ACC	AUROC	AUPR	MCC	F1
GCN [37]	0.9367 ± 0.001	0.8935 ± 0.003	0.6091 ± 0.011	0.5795 ± 0.008	0.6179 ± 0.008
GAT [38]	0.9363 ± 0.003	0.9085 ± 0.002	0.5918 ± 0.018	0.5841 ± 0.012	0.6213 ± 0.011
SVM [39]	0.9277 ± 0.001	0.7018 ± 0.002	0.3322 ± 0.004	0.4929 ± 0.004	0.5416 ± 0.004
RF [40]	0.9320 ± 0.001	0.9221 ± 0.003	0.6023 ± 0.013	0.5389 ± 0.009	0.5816 ± 0.008
DTI-CNN [3]	0.9588 ± 0.001	0.9433 ± 0.003	0.8212 ± 0.002	0.7415 ± 0.001	0.7652 ± 0.001
IMCHGAN [41]	0.9598 ± 0.005	0.9469 ± 0.009	0.8291 ± 0.074	0.7455 ± 0.044	0.7620 ± 0.048
DrugBAN [13]	0.9033 ± 0.006	0.9385 ± 0.003	0.7786 ± 0.022	0.6004 ± 0.020	<b>0.8829 ± 0.011</b>
GraphDTA [42]	0.9558 ± 0.003	0.9263 ± 0.001	0.7773 ± 0.021	0.7255 ± 0.021	0.7500 ± 0.019
HpyerAttentionDTI [14]	0.9405 ± 0.005	0.9273 ± 0.009	0.7563 ± 0.031	0.6645 ± 0.028	0.6947 ± 0.027
GraphCDR [28]	0.8989 ± 0.005	0.9175 ± 0.009	0.5828 ± 0.019	0.5514 ± 0.013	0.5765 ± 0.013
MSGCL [43]	0.9085 ± 0.008	0.8463 ± 0.025	0.4679 ± 0.037	0.2470 ± 0.008	0.5075 ± 0.033
SPVec-SGCN-CPI [44]	0.9060 ± 0.009	0.8379 ± 0.032	0.5250 ± 0.008	0.4245 ± 0.023	0.4307 ± 0.016
DTI-CDF [45]	0.9526 ± 0.001	0.8830 ± 0.009	0.7697 ± 0.016	<b>0.7924 ± 0.012</b>	0.7962 ± 0.012
DTI-MLCD [46]	0.9419 ± 0.001	0.9001 ± 0.007	0.6616 ± 0.012	0.6001 ± 0.013	0.6423 ± 0.011
<b>CCL-ASPS(ours)</b>	<b>0.9656 ± 0.001</b>	<b>0.9523 ± 0.002</b>	<b>0.8495 ± 0.005</b>	0.7787 ± 0.010	0.8003 ± 0.009

**Table 4** The ablation experiment results

Methods	AUROC	AUPR
CCL-ASPS w/o GF	0.9402 ± 0.002	0.9516 ± 0.002
CCL-ASPS w/o CCL	0.9503 ± 0.002	0.9615 ± 0.001
CCL-ASPS w/o ASPS	0.9506 ± 0.001	0.9614 ± 0.001
CCL-ASPS	<b>0.9548 ± 0.003</b>	<b>0.9644 ± 0.002</b>

contrastive learning framework. To better evaluate the effectiveness of the adaptive self-paced sampling and collaborative contrastive learning, we conduct an additional experiment (CCL-ASPS w/o CCL) on CCL-ASPS without using adaptive self-paced sampling.

As shown in Table 4, disabling any module leads to a decrease in performance compared to the full CCL-ASPS model. Specifically, CCL-ASPS w/o GF exhibits a 1.4% and 1.3% decrease in AUROC and AUPR, respectively. Similarly, CCL-ASPS w/o CCL shows a decrease of 0.4% and 0.3% in AUROC and AUPR, respectively. Finally, CCL-ASPS w/o ASPS led to a decrease of 0.4% and 0.3% in AUROC and AUPR, respectively. These ablation results demonstrate that each component of CCL-ASPS contributes to its superior drug-target prediction performance.

#### The effectiveness of collaborative contrastive learning under different feature combination strategies

A critical component of CCL-ASPS is collaborative contrastive learning. This strategy aims to learn more consistent features from the drug and protein similarity networks. The hypothesis is that CCL improves model

performance regardless of the chosen feature combination strategy.

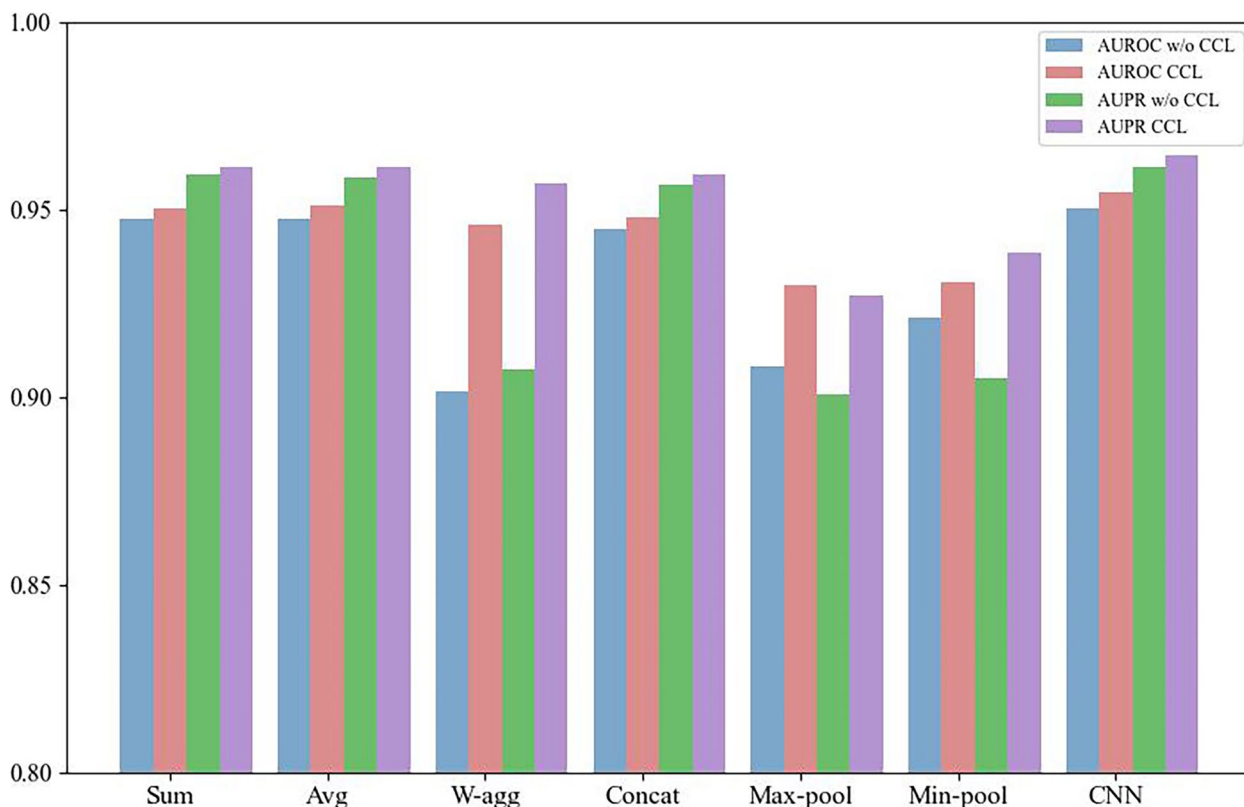
To validate the hypothesis, this study evaluates the impact of CCL across various feature combination strategies: summation (sum), averaging (avg), weighted aggregation (w-agg), concatenation (concat), max pooling (max-pool), min pooling (min-pool), and convolutional neural network (CNN). The AUROC and AUPR results, as presented in Fig. 4, depict the effectiveness of CCL across various feature combination strategies. The results consistently demonstrate that models employing CCL outperform those without it across all feature combination strategies, including summation, averaging, weighted aggregation, concatenation, max pooling, min pooling, and convolutional neural network. The superior AUROC and AUPR results with CCL are achieved by the CNN strategy. This strongly supports the hypothesis and highlights the robustness of CCL in enhancing the predictive capabilities of CCL-ASPS, irrespective of the chosen feature combination strategy.

#### Visualization and interpretation

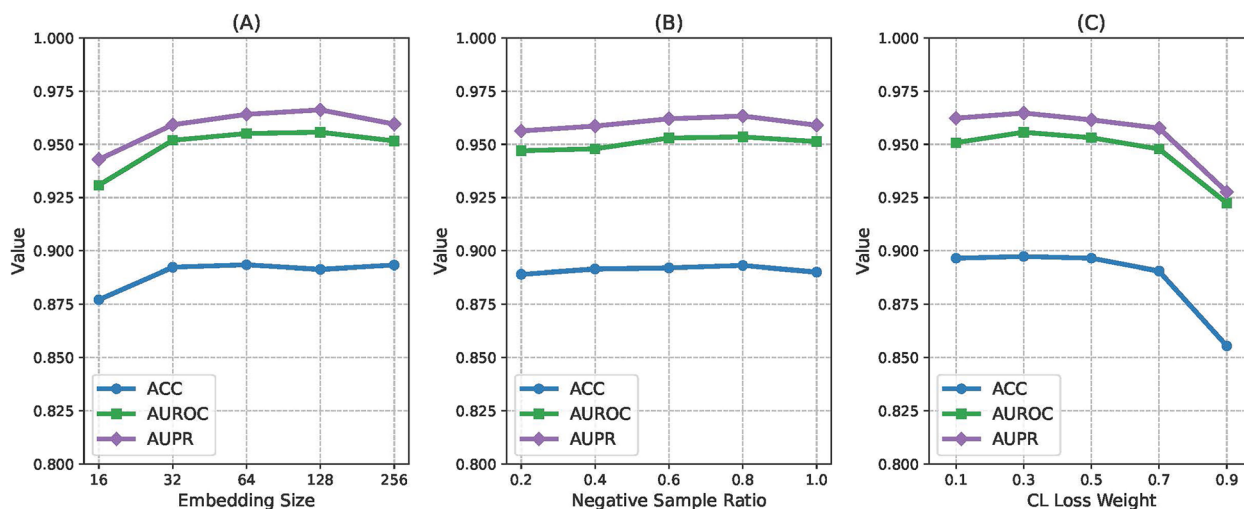
This section employs t-distributed Stochastic Neighbor Embedding (t-SNE) to visualize the learning process of CCL-ASPS. t-SNE projects high-dimensional feature representations into a 2D space for easier visualization. In this experiment, we showcase the ability of CCL-ASPS to distinguish features between positive and negative drug-target interaction pairs.

The learned drug and protein features are to form joint representations of drug-target pairs. Subsequently, these joint representations are compressed to two dimensions using t-SNE and visualized in a scatter plot, as displayed





**Fig. 4** The effectiveness of collaborative contrastive learning under different feature combination strategies



**Fig. 5** The visualization of positive and negative sample pairs under epoch 1, 50, 500

in Fig. 5. Blue dots represent positive drug-target pairs (labeled), while orange dots represent negative pairs (unlabeled). The figure depicts the distribution of joint representations at three training epochs: 1, 50, and 500. At epoch 1 (initial training stage), a substantial overlap between positive and negative samples is observed,

indicating indistinguishable representations during this phase. As training progresses to epoch 50, a clear separation begins to emerge in the 2D space. By epoch 500, although some overlap remains, the distinction between positive and negative sample pairs is more clear. The t-SNE visualization demonstrates the capability of

CCL-ASPS to progressively learn and distinguish positive and negative sample representations during training.

### Case study

The case study evaluates the ability of CCL-ASPS to predict novel drug-target interactions. Due to limitations in directly validating all predictions through wet-lab experiments, a two-step approach is employed for initial validation.

This case study focuses on two drugs: pravastatin and amphetamine, because the prediction probabilities for their associated proteins were notably high. Specifically, the prediction probabilities for the top 20 proteins related to the amphetamine all exceeded 0.96. For pravastatin, the prediction probabilities for the top 20 related proteins are all above 0.83, with the top 10 predictions exceeding 0.92. These high prediction probabilities indicate strong confidence in the model to accurately predict interactions between these drugs and their related proteins.

First, known positive samples are removed from the predictions. Subsequently, the top 20 predicted interacting proteins for pravastatin and amphetamine are selected. To validate these predictions, a comprehensive literature search is conducted with published research articles. The validation results are presented in Tables 5 and 6, respectively. Each table details the specific protein, its corresponding prediction rank, and the PMID number of the validating research paper. The case study

demonstrates the effectiveness of CCL-ASPS in identifying potential drug-target interactions. Notably, 19 out of the top 20 predicted interacting proteins for amphetamine are validated through the literature review. Similarly, 13 out of the top 20 predictions for pravastatin are supported by existing publications. These results highlight the capability of CCL-ASPS to screen and identify promising candidates for further investigation.

### Time complexity analysis

This section analyzes the time complexity of the proposed model, focusing on three key modules: graph feature extraction, collaborative contrastive learning with adaptive self-paced sampling, and predictor.

Both drug and protein graph structure feature extractions utilize a single layer GCN, resulting in a time complexity of  $O(|V_d * n_d + V_p * n_p|F)$ . Where  $V_d$  and  $V_p$  are the numbers of drugs and proteins respectively.  $n_d$  is the number of atoms in each drug and  $n_p$  is the number of residues in each protein.  $F$  represents the feature dimension. Since  $n_d$ ,  $n_p$  and  $F$  are within a fixed range, they can be treated as constants, simplifying the time complexity of this module to  $O(V_d + V_p)$ .

In the collaborative contrast learning module, for similarity network learning, the time complexity is  $O(|V_d + V_p| \times F \times F') + O(|E_d + E_p| \times F')$ , where  $F$  and  $F'$  are the input and output dimensions respectively.  $E_d$  and  $E_p$  are the drug and protein similarity

**Table 5** The top 20 predicted amphetamine-associated proteins by CCL-ASPS

Protein name	Rank	Probability	Evidence
Acetylcholinesterase	1	0.999	PMID:7155357
Norepinephrine transporter	2	0.999	PMID:31912366
Sodium-dependent serotonin transporter	3	0.999	PMID:16722230
Prostaglandin E2 receptor EP1 subtype	4	0.997	Unconfirmed
Solute carrier family 6 member 1	5	0.997	PMID:15615700
Monoamine oxidase type B	6	0.997	PMID:26841904
D(4) dopamine receptor	7	0.996	PMID:20118172
5-hydroxytryptamine receptor 2B	8	0.995	PMID:16996724
Monoamine oxidase type A	9	0.995	PMID:9393679
Muscarinic acetylcholine receptor M1	10	0.994	PMID:11752469
Alpha-2C adrenoreceptor	11	0.993	PMID:25522417
Alpha-1A adrenergic receptor	12	0.993	PMID:16794571
Cholinesterase	13	0.993	PMID:19423342
Serotonin receptor 2A	14	0.982	PMID:23879373
Alpha-2A adrenergic receptor	15	0.981	PMID:8945686
Glutamate receptor ionotropic	16	0.977	PMID:12850537
Alpha-2B adrenergic receptor	17	0.974	PMID:27751959
Beta-1 adrenergic receptor	18	0.972	PMID:15836801
Muscarinic acetylcholine receptor M5	19	0.970	PMID:18772318
Apoptosis regulator Bcl-2	20	0.964	PMID:32963743

**Table 6** The top 20 predicted pravastatin-associated proteins by CCL-ASPS

Protein name	Rank	Probability	Evidence
Dipeptidyl peptidase 4 (DPP-4)	1	0.998	PMID:37771303
Integrin beta-2	2	0.990	PMID:11385505
Ribonucleotide reductase large subunit	3	0.985	Unconfirmed
Prothrombin	4	0.984	PMID:27416543
Xanthine oxidase	5	0.978	Unconfirmed
Angiotensin-converting enzyme	6	0.966	PMID:17174637
Alanine-tRNA ligase, cytoplasmic	7	0.952	Unconfirmed
Tubulin alpha-1A chain	8	0.946	PMID:16678288
Proteasome subunit beta type-2	9	0.927	PMID:25042549
Thrombomodulin	10	0.923	PMID:32164317
Methionine synthase	11	0.917	PMID:29118764
PPase FKBP1A	12	0.912	PMID:12419552
Aminopeptidase N (AP-N)	13	0.901	PMID:16019130
Cytosolic purine 5'-nucleotidase	14	0.897	PMID:10588233
Purine nucleoside phosphorylase	15	0.895	Unconfirmed
Adenosine deaminase	16	0.890	PMID:16678288
Kinesin-like protein KIF1A	17	0.885	Unconfirmed
Integrin alpha-L gene	18	0.879	Unconfirmed
Thyroid peroxidase	19	0.851	Unconfirmed
C3/C5 convertase	20	0.837	PMID:32849586

edges in the similarity networks. The number of drug or protein similarity networks is constant and can be omitted. The time complexity of one layer convolution operation is  $O(|V_d + V_p| \times (F - k)^2 \times k^2 \times C_{in} \times C_{out})$ , where  $k$  is the kernel size,  $C_{in}$  and  $C_{out}$  are input and output channels respectively, which are all constants. Therefore, the time complexity of the convolution operation can be abbreviated as  $O(|V_d + V_p| \times F^2)$ . The drug and protein contrastive learning loss requires computing the feature similarity between any two nodes, hence the time complexity is  $O(|V_d^2 + V_p^2| \times F)$ . For the selection of contrastive learning sample pairs, we first calculate the feature similarity between all nodes, then sort them in ascending order, after which we select  $num_t$  negative samples. Thus the time

complexity of negative sample pair selection is  $O(|V_d^2 + V_p^2| \times F) + O(|V_d \log V_d|) + O(|V_p \log V_p|) = O(|V_d^2 + V_p^2| \times F)$ . By only considering the effect of the node number, the time complexity of the collaborative contrastive learning module is  $O(|V_d + V_p| \times F \times F) + O(|E_d + E_p| \times F) + O(|V_d + V_p| \times F^2) + O(|V_d^2 + V_p^2| \times F) + O(|V_d^2 + V_p^2| \times F) = O(|V_d^2 + V_p^2|)$ .

The time complexity of the predictor is  $O(|E_{dp}| \times F \times F)$ , where  $E_{dp}$  is the number of positive drug-target interaction pairs.

Combining the complexities of each module, the overall time complexity of CCL-ASPS is  $O(|V_d^2 + V_p^2|) + O(|E_{dp}|) = O(|V_d^2 + V_p^2|)$ .

Additionally, we have documented the running times for CCL-ASPS and all baseline methods, which are presented in Table 7. The running time of CCL-ASPS is 1293 seconds. The running times for GCN, SVM, RF, and DTI-CNN are all under 100 seconds. The running times for GAT, GraphCDR, SPVec-SGCN-CPI, DTI-CDF, and DTI-MLCD range between 100 and 1000 seconds. The running times for IMCHGAN and DrugBAN exceed 1000 seconds but are still shorter than that of CCL-ASPS. Models such as GraphDTA, HyperAttentionDTI, and MSGCL require running times longer than CCL-ASPS, with HyperAttentionDTI having the longest running time at 3217 seconds.

The running times reveal that while CCL-ASPS requires a relatively longer training time compared to some baseline models, it is still within a reasonable and manageable range. The majority of baseline models (e.g., GCN, SVM, RF, and DTI-CNN) with significantly shorter running times often exhibit limitations in scalability and accuracy.

Furthermore, the time complexity analysis demonstrate that the running time of CCL-ASPS increases quadratically as the dataset grows. Despite this quadratic growth, the running time remains within a predictable and manageable range, ensuring that CCL-ASPS can handle larger datasets efficiently.

### Discussion

Following the detailed experimental analysis presented in the Results section, this Discussion section further evaluates the robustness and practical utility of the proposed

**Table 7** The running times of CCL-ASPS and baselines

Model	Time(s)	Model	Time(s)	Model	Time(s)
GCN [37]	84	IMCHGAN [41]	1060	MSGCL [43]	2683
GAT [38]	108	DrugBAN [13]	1227	SPVec-SGCN-CPI [44]	844
SVM [39]	10	GraphDTA [42]	1717	DTI-CDF [45]	633
RF [40]	8	HpyerAttentionDTI [14]	3217	DTI-MLCD [46]	704
DTI-CNN [3]	43	GraphCDR [28]	419	<b>CCL-ASPS(ours)</b>	1293

CCL-ASPS, focusing on three key aspects: parameter sensitivity, cold start performance, and statistical significance.

### Parameter sensitivity analysis

The selection of hyperparameters is crucial for ensuring accurate and reliable prediction performance, and this section discusses the key parameters investigated along with their optimization process.

This experiment investigates the influence of three key hyperparameters: embedding size, negative sample ratio  $\beta$ , and contrastive loss weight  $\gamma$ . The results are presented in Fig. 6. To begin with, the proposed CCL-ASPS is evaluated with embedding sizes of 16, 32, 64, 128, and 256. The AUROC value increases from 0.931 at an embedding size of 16 to a maximum of 0.955 at 64. While the embedding size is further increased to 128 and 256, the AUROC value remains around 0.955. Similarly, the ACC and AUPR values also reach their highest at an embedding size of 64 (0.893 and 0.9641, respectively). Consequently, an embedding size of 64 is chosen for this study. Furthermore, the effectiveness of  $\beta$  in contrastive learning is analyzed. CCL-ASPS achieves the highest AUROC value when  $\beta = 0.8$ , as both excessively small and large values lead to a decrease in prediction performance. Additionally, the contrastive loss weight  $\gamma$  also impacts overall performance. Based on the experimental results, the optimal value for  $\gamma$  is set to 0.3.

The number of GCN layers in both the drug and protein graph structure extraction modules is set to 1. Grid search is employed to determine this value. Similarly, the GAT layers number for the collaborative contrastive learning

module is set to 2. A dropout rate of 0.2 is chosen for the GAT layer in the collaborative contrastive learning module. The training epochs are set to 2000, 2000, and 5000 for the drug graph structure feature extraction module, protein graph structure feature extraction module, and collaborative contrastive learning module, respectively. The learning rates of these three modules are all set to 0.001, which is selected from [0.1, 0.01, 0.001, 0.0001].

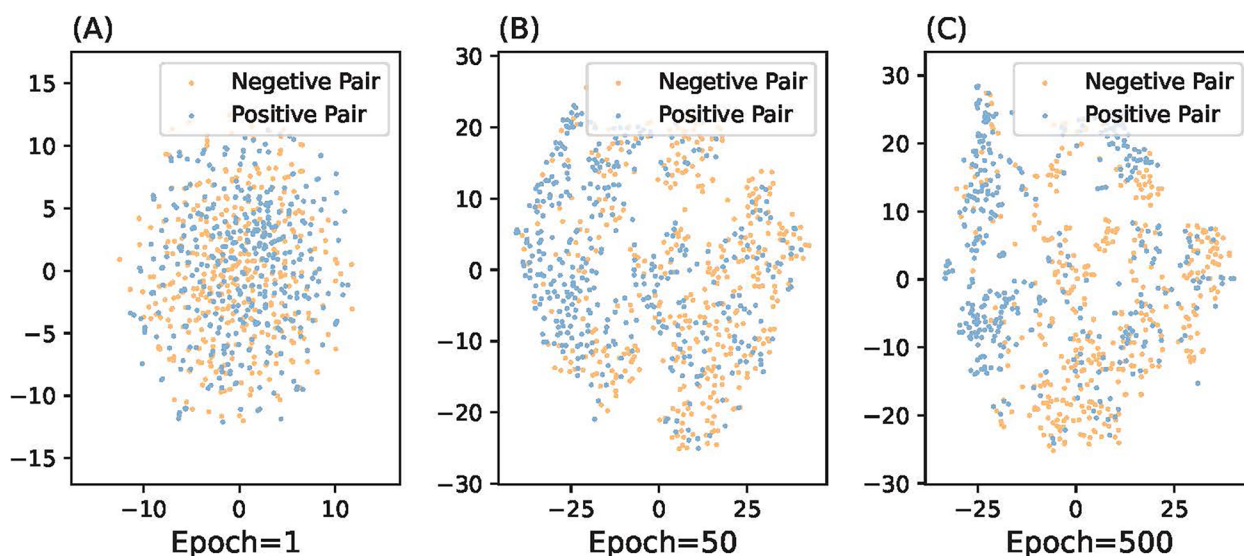
The parameter sensitivity analysis highlights that careful tuning of hyperparameters is essential to maximize the performance of CCL-ASPS. Future research could further enhance model effectiveness by employing adaptive or automated hyperparameter optimization techniques.

### Cold start

This section discusses the performance of CCL-ASPS under a cold start scenario, where the test set comprises drugs or proteins absent from the training data.

The K-fold cross-validation is employed to assess the model performance. In the context of drug cold start evaluation, drugs are first divided into k folds. During each iteration (fold), one subset of drugs with their associated interactions is designated as the test set. The remaining drugs in the k-1 subsets with their associated interactions constitute the positive samples for training. The negative samples are then randomly selected from all possible drug-target pairs between the drugs in the k-1 subsets and all targets.

This ensures the test set excludes drugs present in the training set, enabling a systematic evaluation of the model's response to unseen drugs. An identical k-fold



**Fig. 6** The parameter sensitive analysis

cross-validation strategy is employed for protein cold start evaluation, ensuring the generalizability of CCL-ASPS to unseen proteins.

Figure 7 presents the prediction performance of CCL-ASPS for drug and protein cold start with varying  $k$  values. As observed, both AUROC and AUPR scores exhibit a positive correlation with increasing  $k$ . For drug cold start, the AUROC score rises from 0.8939 ( $k=2$ ) to 0.9242 ( $k=20$ ), and the AUPR score increases from 0.9118 ( $k=2$ ) to 0.9261 ( $k=20$ ). A similar trend is observed in protein cold start. These observed trends are likely attributable to the increasing size and diversity of the training set with larger  $k$  values. A larger training dataset provides the model with a broader range of examples to learn from, consequently enhancing its capability for generalizing to unseen drugs and proteins.

The cold start analysis demonstrates the strong generalization capability of CCL-ASPS with unseen drugs or proteins. Future work could further improve the performance by incorporating additional biological knowledge, such as drug-target binding affinities or 3D structures of drug, to provide a more comprehensive understanding of the interactions.

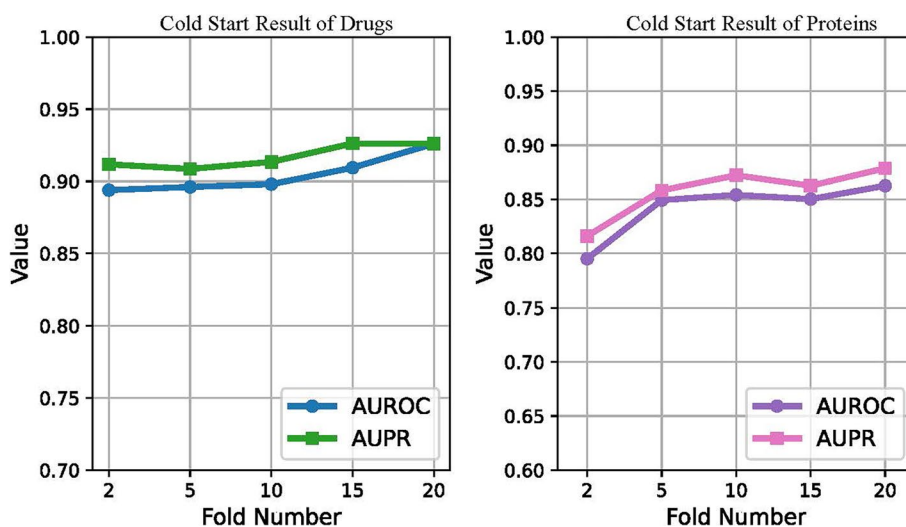
#### Statistic analysis

The experimental results demonstrate that our model, CCL-ASPS, significantly outperforms all baseline methods in predicting drug-target interactions. In this discussion, we conducted a comprehensive statistical analysis to compare the performance of CCL-ASPS against all baseline methods. The objective of this experiment is to analyze whether the observed improvements in performance are statistically significant.

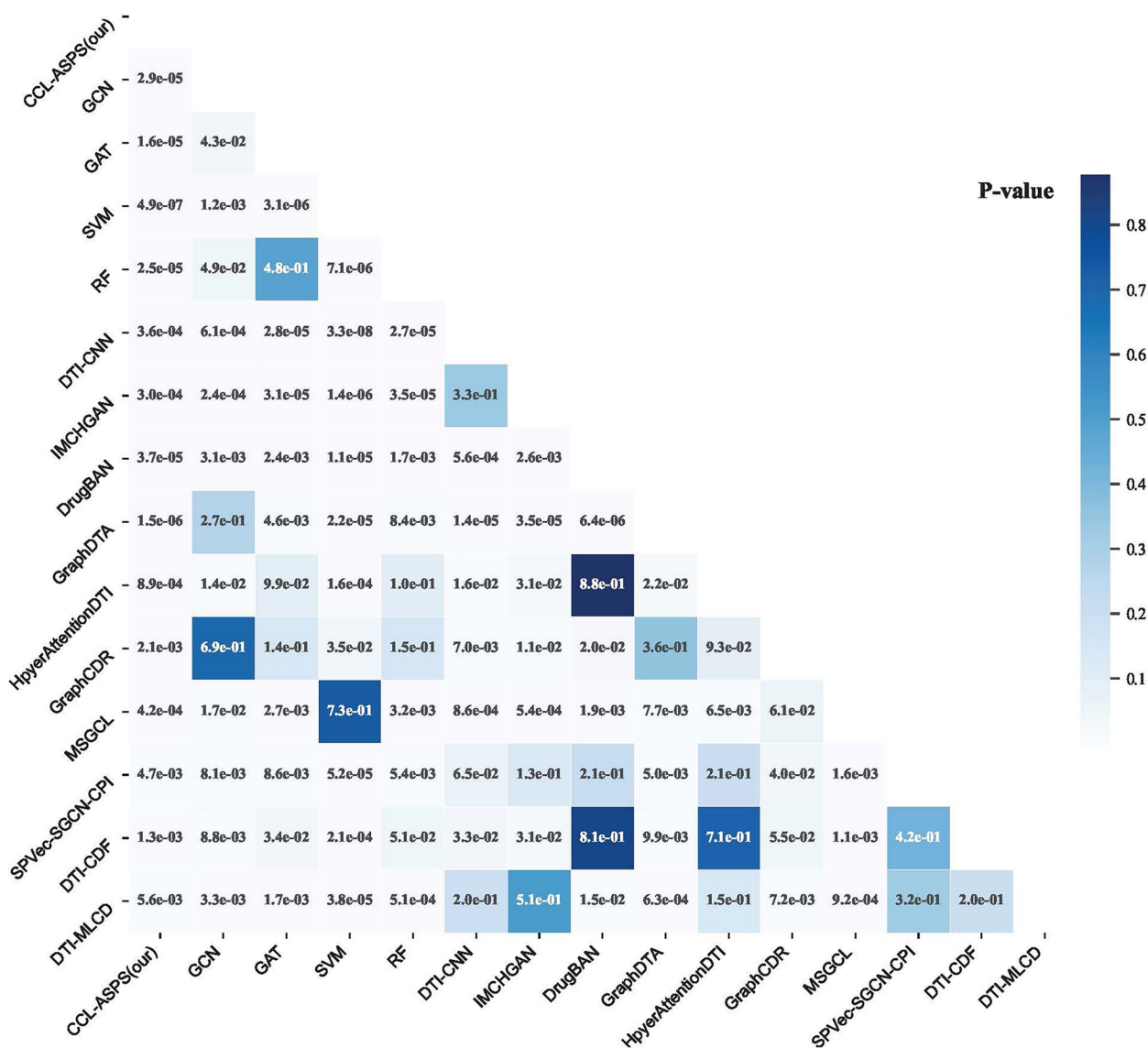
The performance of each baseline is assessed through ten independent runs. To evaluate the statistical significance of the performance improvements, we employ paired samples  $t$ -tests on the AUROC metric values. We use a significance level of 0.05 to determine the statistical significance. A  $p$ -value less than 0.05 indicates strong evidence against the null hypothesis, suggesting that the observed improvements in model performance are not due to random chance. The results of the paired  $t$ -tests are detailed in Fig. 8. CCL-ASPS outperforms all baseline methods with  $p$ -values less than 0.05, indicating that the improvements in model performance are statistically significant.

The statistical significance analysis provides strong evidence that the performance improvements achieved by CCL-ASPS over baseline methods are substantial, demonstrating its superior predictive capabilities in drug-target interaction tasks.

In summary, this discussion focuses on three key aspects: parameter sensitivity, cold start performance, and statistical significance of CCL-ASPS. First, the parameter sensitivity analysis highlights the importance of fine-tuning hyperparameters such as embedding size, negative sample ratio, and contrastive loss weight to optimize the performance of CCL-ASPS. Second, the cold start analysis demonstrates the effectiveness of CCL-ASPS in handling unseen drugs or proteins. Third, the statistical significance analysis confirms that the performance gains of CCL-ASPS over other baseline models are statistically significant and unlikely due to random chance. Despite these strengths, certain limitations remain. For instance, integrating adaptive hyperparameter optimization techniques could further enhance model



**Fig. 7** The AUROC and AUPR results of drug and protein cold start experiments



**Fig. 8** The statistic results between CCL-ASPS and all baselines on metric AUROC

robustness and reduce manual effort. Furthermore, while the model performs well on the cold start experiment, its generalizability to other drug-target interaction datasets has yet to be explored.

**Conclusions**

This study presents CCL-ASPS, a novel approach that leverages collaborative contrastive learning and an adaptive self-paced sampling strategy to learn consistent representations from multiple networks.

Experimental results demonstrate the superior performance of CCL-ASPS in predicting drug-target interactions compared to baseline methods. Furthermore, the ablation experiment validates the effectiveness of each

model component, while the exploration of various feature combination strategies confirms the efficacy of CCL. Moreover, the case study and cold start evaluation showcase the capability of CCL-ASPS to predict potential drug-target interactions. Finally, the statistic analysis further confirms the superior performance of CCL-ASPS.

Despite its contributions, this work presents opportunities for further exploration. First, investigating more sophisticated learning strategies within the structure-based feature extraction component holds promise for further performance improvements. Secondly, the proposed CCL-ASPS could be applied to predict associations among other diverse biological entities, such as drug-disease interaction and protein-protein interaction.

## Methods

This section details the components of the proposed CCL-ASPS. The utilized dataset and data preprocessing are presented first. Subsequently, the components of CCL-ASPS are introduced: graph structure-based feature extraction, collaborative contrastive learning with adaptive self-paced sampling, and the final predictor. The source code and dataset for CCL-ASPS are publicly available at 10.5281/zenodo.13329691.

### Dataset and preprocessing

This study validates all innovations with one drug-target interaction dataset collected from multiple data sources. Interaction networks related to drugs and proteins are obtained from Luo [17]. SMILES strings of drugs are collected from DrugBank [47]. Protein amino acid sequences are retrieved from Uniport [48]. Protein 3D structures are acquired from the RCSB Protein Data Bank [49]. For proteins lacking PDB structures, predicted structures from AlphaFold [50] are utilized. As shown in Table 8, the final dataset contains 12015 nodes, including 708 drugs, 1512 proteins, 5603 diseases, and 4192 side effects. Furthermore, six types of interactions are collected, including 1923 drug-protein interactions (Dr-Pr), 199214 drug-disease associations (Dr-Di), 10036 drug-drug interactions (Dr-Dr), 80164 drug-side-effect associations (Dr-Se), 1923 protein-drug interactions (Pr-Dr), 1596745 protein-disease associations (Pr-Di), and 7363 protein-protein interactions (Pr-Pr).

Afterward, the drug SMILES strings are converted into atom interaction graphs. Each atom feature is initialized based on its chemical properties [51]. For proteins, the coordinates of amino acid residues are represented by their  $C_\alpha$  atoms. The distances between residues are calculated. Residues are considered in contact if their distance is below a specified threshold, which forms the amino acid residue interaction graphs.

**Table 8** Statistics of the collected data

Node	Number	Interaction	Number
Drug	708	Dr-Pr	1923
Protein	1512	Dr-Di	199214
Disease	5603	Dr-Dr	10036
Side-effect	4192	Dr-Se	80164
–	–	Pr-Dr	1923
–	–	Pr-Di	1596745
–	–	Pr-Pr	7363

Additionally, seven attributes of amino acid residues are utilized for initial representations [52].

Jaccard similarity [53] is employed to construct similarity networks based on each interaction and association network. Additionally, structure similarity networks are built based on SMILES strings and amino acid sequences using methods described in Luo [17].

### Drug graph structure-based feature

This section describes the extraction of drug representations from atom interaction graphs using Graph Convolutional Network (GCN) [37].

For each drug  $d_i$ , its initial atom representations are denoted as  $X_{d_i} \in \mathbb{R}^{n_{d_i} \times k_d}$ , where  $n_{d_i}$  and  $k_d$  represents the number of atoms in drug  $d_i$  and the feature dimension of atoms, respectively. Additionally, matrix  $A_{d_i} \in \mathbb{R}^{n_{d_i} \times n_{d_i}}$  denotes the adjacency matrix between atoms, where  $A_{d_i} = 1$  means there is an edge between two atoms and  $A_{d_i} = 0$  otherwise.

GCN updates the atom features by aggregating information from neighboring atoms, as defined by the following equation:

$$X'_{d_i} = \sigma \left( D_{d_i}^{-\frac{1}{2}} \check{A}_{d_i} D_{d_i}^{-\frac{1}{2}} X_{d_i} W_{d_1} \right), \quad (1)$$

where  $\check{A}_{d_i} = A_{d_i} + I$  is the adjacency matrix with self-loop,  $I \in \mathbb{R}^{n_{d_i} \times n_{d_i}}$  is the identity matrix,  $D_{d_i} \in \mathbb{R}^{n_{d_i} \times n_{d_i}}$  is the degree matrix with the value of each diagonal element equal to the degree of the corresponding atom node,  $W_{d_1} \in \mathbb{R}^{k_d \times f_d}$  is a learnable weight parameter and  $f_d$  is the output feature dimension of atoms,  $\sigma(\cdot)$  is the nonlinear activation function.

Following the update of features, a readout function is employed to combine all the atom features and generate the drug representation.

$$h_{d_i} = \text{Readout} \left( X'_{d_i} \right), \quad (2)$$

where  $h_{d_i} \in \mathbb{R}^{f_d}$  is the representation of drug  $d_i$ . In this study, all  $\text{Readout}(\cdot)$  functions are implemented using the global mean pooling operation.

To pre-train the GCN for drug feature extraction, a binary cross-entropy loss (BCELoss) is employed as the objective function, aiming to predict drug-drug interactions. This involves learning the joint representations of the drug-drug pairs with a CNN layer:

$$h_{(d_i, d_j)} = \text{Pooling} \left( \text{Conv} \left( h_{d_i} \| h_{d_j} \right) \right), \quad (3)$$

where  $h_{d_i}$  and  $h_{d_j}$  are the representations of drug  $d_i$  and  $d_j$ , respectively. The symbol  $\|$  denotes the concatenation operation.  $\text{Conv}(\cdot)$  represents the CNN layer with a kernel size of 3, an out channel of 4, padding and step size of

1.  $Pooling(\cdot)$  represents the max pooling operation. The resulting  $h_{(d_i, d_j)} \in \mathbb{R}^{4f_d}$  is the joint representation of drug  $d_i$  and  $d_j$ .

A multilayer perceptron is applied to predict the association probability of drug pairs:

$$P_{(d_i, d_j)} = \sigma \left( W_{d_3} \left( ReLu \left( W_{d_2} h_{(d_i, d_j)} \right) \right) \right), \quad (4)$$

where  $W_{d_2} \in \mathbb{R}^{4f_d \times f_d}$  and  $W_{d_3} \in \mathbb{R}^{f_d \times 1}$  are learnable parameters,  $\sigma(\cdot)$  and  $ReLu(\cdot)$  are nonlinear activation functions, and  $P_{(d_i, d_j)}$  is the predicted association probability between drug  $d_i$  and  $d_j$ .

The BCELoss serves as the objective function for drug feature extraction:

$$\mathcal{L}_d = - \sum_{(d_i, d_j) \in ddp} \hat{P}_{(d_i, d_j)} \log P_{(d_i, d_j)} + (1 - \hat{P}_{(d_i, d_j)}) \log(1 - P_{(d_i, d_j)}), \quad (5)$$

where  $ddp$  denotes a set of drug-drug pairs.  $(d_i, d_j)$  represents a single pair of drug  $d_i$  and  $d_j$ .  $P_{(d_i, d_j)}$  represents the predicted association probability of drug pair  $(d_i, d_j)$ .  $\hat{P}_{(d_i, d_j)}$  signifies the ground truth for drug-drug association, where  $\hat{P}_{(d_i, d_j)} = 1$  for true drug-drug association and  $\hat{P}_{(d_i, d_j)} = 0$  otherwise.

### Protein graph structure-based feature

Similar to drug feature extraction, a GCN-based approach is employed to extract protein features from the amino acid residue interaction graphs. The amino acid residue interaction graph for protein  $p_i$  is represented by an adjacency matrix  $A_{p_i} \in \mathbb{R}^{n_{p_i} \times n_{p_i}}$ , where  $n_{p_i}$  denotes the number of residues in the protein  $p_i$ . Each row in the initial feature matrix  $X_{p_i} \in \mathbb{R}^{n_{p_i} \times k_p}$  represents the initial features of an amino acid residue.  $k_p$  is the initial feature dimension of proteins.

A single-layer GCN is applied to update the residue representations, aggregating information from neighboring residues:

$$X'_{p_i} = ReLu \left( W_{p_2} \left( D_{p_i}^{-\frac{1}{2}} (A_{p_i} + I_{p_i}) D_{p_i}^{-\frac{1}{2}} X_{p_i} W_{p_1} \right) \right), \quad (6)$$

where  $I_{p_i} \in \mathbb{R}^{n_{p_i} \times n_{p_i}}$  is the identity matrix.  $D_{p_i} \in \mathbb{R}^{n_{p_i} \times n_{p_i}}$  is a diagonal matrix with the corresponding values are nodes degree.  $W_{p_1} \in \mathbb{R}^{k_p \times f_p}$  and  $W_{p_2} \in \mathbb{R}^{f_p \times f_p}$  are learnable parameters.  $f_p$  is the output feature dimension of

amino acid residue.  $ReLu(\cdot)$  is an activation function.  $X'_{p_i} \in \mathbb{R}^{n_{p_i} \times f_p}$  is the updated amino acid residue feature matrix.

Following the update, a self-attention pooling layer is employed to extract feature representations of key residues. Subsequently, a readout function is applied to obtain the protein representation for downstream tasks:

$$h_{p_i} = Readout \left( SAGPooling \left( X'_{p_i} \right) \right), \quad (7)$$

where  $SAGPooling(\cdot)$  is the self-attention pooling.  $h_{p_i} \in \mathbb{R}^{f_p}$  is the representation of protein  $p_i$ .

To pre-train the GCN for protein feature extraction, the binary cross-entropy loss is applied as the objective function, aiming to predict protein-protein interactions (PPIs). The pre-training process is analogous to drug feature extraction. The implementation is the same as in Eqs. (3, 4, 5):

$$h_{(p_i, p_j)} = Pooling \left( Conv \left( h_{p_i} \| h_{p_j} \right) \right), \quad (8)$$

$$P_{(p_i, p_j)} = \sigma \left( W_{p_3} ReLu \left( W_{p_2} h_{(p_i, p_j)} \right) \right), \quad (9)$$

$$\mathcal{L}_p = - \sum_{(p_i, p_j) \in ppi} \hat{P}_{(p_i, p_j)} \log P_{(p_i, p_j)} + (1 - \hat{P}_{(p_i, p_j)}) \log(1 - P_{(p_i, p_j)}), \quad (10)$$

where  $h_{p_i}$  and  $h_{p_j}$  are representations of protein  $p_i$  and  $p_j$ , respectively.  $h_{p_{ij}} \in \mathbb{R}^{4 \times f_p}$  is the joint representation of protein  $p_i$  and  $p_j$ .  $W_{p_2} \in \mathbb{R}^{4f_p \times f_p}$  and  $W_{p_3} \in \mathbb{R}^{f_p \times f_p}$  are learnable parameters.  $P_{(p_i, p_j)}$  is the predicted interaction probability between protein  $p_i$  and  $p_j$ .  $\hat{P}_{(p_i, p_j)}$  is the ground truth, with  $\hat{P}_{(p_i, p_j)} = 1$  indicating an interaction between two proteins and  $\hat{P}_{(p_i, p_j)} = 0$  otherwise.

### Collaborative contrastive learning with adaptive self-paced sampling

This section describes the proposed collaborative contrastive learning and adaptive self-paced sampling strategy, illustrated in Fig. 2B and C. The strategy consists of three key components: multiple network learning, collaborative contrastive learning, and adaptive self-paced sampling. The overall workflow is outlined in Algorithm 1.



**Algorithm 1 CCL-ASPS****Input:**

drug and protein 2D structure features  $h_{d_i}$  and  $h_{p_i}$   
 multiple drug similarity networks  $S^{m_d}, m_d = 1, 2, \dots, n_{m_d}$   
 multiple protein similarity networks  $S^{m_p}, m_p = 1, 2, \dots, n_{m_p}$   
 positive sample set  $P$   
 number of training epochs  $T$

**Output:** trained model**Train:**

- 1: **for**  $t = 1$  to  $T$  **do**
- 2:   get negative sample number with equation (21);
- 3:   **for**  $S^m$  in  $S^{m_d}$  **do**
- 4:     learn feature  $h_{d_i}^{S^m}$  from network  $S^m$  with equation (13);
- 5:   **end for**
- 6:   get the fused features with equation (14);
- 7:   calculate feature similarity between any two drugs  $d_i$  and  $d_j$  with Jaccard similarity algorithm, denoted as  $sim(d_i, d_j)$ ;
- 8:   **for**  $S^m$  in  $S^{m_d}$  **do**
- 9:     select  $num_t$  smallest similarity pairs from  $S^m$  as negative sample set  $N^{ns}$ ;
- 10:    select  $num_t$  smallest pairs from  $sim(d_i, d_j)$  as negative sample set  $N^{fs}$ ;
- 11:    get final negative sample set  $N^m = N^{ns} \cup N^{fs}$ ;
- 12:    calculate contrastive learning loss  $\mathcal{L}_d^m$  using equation (15);
- 13:   **end for**
- 14:   similarly, calculate contrastive learning loss  $\mathcal{L}_p^m$  for proteins with equation (16);
- 15:   calculate overall contrastive loss  $\mathcal{L}^c$  using equation (17);
- 16:   predict drug-target interactions with equation (23);
- 17:   calculate predictive loss  $\mathcal{L}_{dp}$ ;
- 18:   overall loss  $\mathcal{L} = \mathcal{L}^c + \mathcal{L}_{dp}$ ;
- 19:   update model parameters via back propagation;
- 20: **end for**
- 21: return: trained model;

**Multiple network learning**

This section introduces the first component: multiple network learning. GAT [38] is employed to leverage information hidden within each drug and protein similarity network.

With each similarity network  $S^m$ , ( $m = 1, 2, \dots, n_{m_d}$ ), for each drug  $d_i$  and its pre-trained representation  $h_{d_i}$ , GAT computes normalized attention scores to its neighbors. The attention score  $\alpha_{ij}^l$  for neighbor  $d_j$  at layer  $l$  is calculated as:

$$\alpha_{ij}^l = \text{softmax} \left( \frac{e_{ij}^l}{\sum_{d_k \in \mathcal{N}(i)} e_{ik}^l} \right), \quad (11)$$

where  $e_{ij}^l$  is the unnormalized attention score between drug  $d_i$  and  $d_j$  at layer  $l$ .  $\mathcal{N}(i)$  denotes the set neighbors for drug  $d_i$ .

The unnormalized attention score  $e_{ij}^l$  is defined as:

$$e_{ij}^l = a^l \left( \text{LeakyReLU}(W_1^l h_{d_i}^{l-1} \| W_2^l h_{d_j}^{l-1}) \right), \quad (12)$$

where  $W_1^l \in \mathbb{R}^{f_a \times f_a}$  and  $W_2^l \in \mathbb{R}^{f_a \times f_a}$  are learnable weight matrices at layer  $l$ .  $a^l \in \mathbb{R}^{f_a}$  is a learnable parameter vector at layer  $l$ .

After computing the normalized attention scores  $\alpha_{ij}^l$  for all neighbors, the GAT layer aggregates the features of neighboring drugs in a weighted manner to update the representation of the target drug  $d_i$ :

$$h_{d_i}^l = \sigma \left( \sum_{d_j \in \mathcal{N}(i)} \alpha_{ij}^l W_3^l h_{d_j}^{l-1} \right), \quad (13)$$

where  $h_{d_i}^l$  denotes the output representation at layer  $l$ , with  $h_{d_i}^0 = h_{d_i}$ .  $W_3^l \in \mathbb{R}^{f_a \times f_a}$  is the weight matrix in layer  $l$ .

The output of the final layer is considered the final representation learned from the similarity network  $S^m$ , denoted as  $h_{d_i}^{S^m}$ .

Similar to drug representation learning, protein features can be extracted from each protein similarity network. The learned feature of protein  $p_i$  under the protein similarity network  $S^m$  is denoted as  $h_{p_i}^{S^m}$ .

#### Collaborative contrastive learning loss

After extracting features from all similarity networks, a convolutional neural network is employed to fuse the features of both drugs and proteins. The equation for drug feature fusion is as follows:

$$H_{d_i} = \text{Conv} \left( \prod_{m=1}^{n_{m_d}} h_{d_i}^{S^m} \right), \quad (14)$$

where  $\prod$  is the concatenate operation.  $\text{Conv}(\cdot)$  is a 1D convolution operation with an output channel of 16 and kernel size of 3.  $h_{d_i}^{S^m}$  is the feature of drug  $d_i$  learned from the  $m$ -th similarity network.  $H_{d_i}$  is the fused representation of drug  $d_i$ . An equivalent process is employed to obtain the fused representation of protein  $p_i$ , denoted as  $H_{p_i}$ .

Instead of calculating the contrastive loss for every pair of network outputs, this approach contrasts the features learned from individual networks with the fused features. This strategy encourages the representations learned from individual networks to be more consistent with the fused representations. For drugs, the contrastive learning loss aims to minimize the distance between each  $h_{d_i}^{S^m}$ , ( $m = 1, 2, \dots, n_{m_d}$ ) and the fused feature  $H_{d_i}$ . The loss function is defined as follows:

$$\mathcal{L}_d^m = \sum_{i=1}^{n_d} -\log \frac{\sum_{d_j \in P_{d_i}^m} \exp(\text{sim}(h_{d_i}^{S^m}, H_{d_j})/\tau)}{\sum_{d_j \in (P_{d_i}^m \cup N_{d_i}^m)} \exp(\text{sim}(h_{d_i}^{S^m}, H_{d_j})/\tau)}, \quad (15)$$

where  $n_d$  is the number of drugs.  $P_{d_i}^m$  and  $N_{d_i}^m$  are the positive and negative sample sets for drug  $d_i$  in drug similarity network  $S^m$ , respectively.  $h_{d_i}^{S^m}$  is the feature of drug  $d_i$  learned from  $S^m$ .  $H_{d_i}$  is the fused feature of drug  $d_i$ .  $\text{sim}(\cdot)$  is the cosine similarity and  $\tau$  is the temperature parameter.

An identical loss function is applied for protein  $p_i$  to minimize the distance between the feature  $h_{p_i}^{S^m}$  learned from  $m$ -th similarity network and the fused feature  $H_{p_i}$ . The equation for the protein contrastive loss is defined as follows:

$$\mathcal{L}_p^m = \sum_{i=1}^{n_p} -\log \frac{\sum_{p_j \in P_{p_i}^m} \exp(\text{sim}(h_{p_i}^{S^m}, H_{p_j})/\tau)}{\sum_{p_j \in (P_{p_i}^m \cup N_{p_i}^m)} \exp(\text{sim}(h_{p_i}^{S^m}, H_{p_j})/\tau)}, \quad (16)$$

where  $n_p$  is the proteins number,  $P_{p_i}^m$  and  $N_{p_i}^m$  are the positive and negative sample sets of protein  $p_i$ , respectively.

The contrastive losses from each individual similarity network are combined to compute the overall contrastive loss  $\mathcal{L}^c$ , as shown in the following equation:

$$\mathcal{L}^c = \sum_{m=1}^{n_{m_d}} (\mathcal{L}_d^m) + \sum_{m=1}^{n_{m_p}} (\mathcal{L}_p^m), \quad (17)$$

where  $n_{m_d}$  and  $n_{m_p}$  represent the drug and protein similarity network numbers, respectively.

#### Adaptive self-paced sampling

The selection of contrastive learning samples plays a pivotal role in the effectiveness of contrastive learning. This work follows a common approach where features of the same node across different views are considered positive samples. For instance, the positive sample set  $P_{p_i}^m$  for protein  $p_i$  in network  $S^m$  only includes  $p_i$  itself.

Unlike previous works that treat all non-positive node pairs as negative samples, this study introduces the adaptive self-paced sampling strategy to identify more informative negative samples. The details of this strategy are illustrated in Fig. 2C.

After removing positive samples, each protein  $p_i$  has  $n_p - 1$  candidate negative samples. To select more informative negative samples, this work employs two scoring functions to assess their reliabilities:

$$R_{ij}^m = 1 - S_{ij}^m, \quad (18)$$

$$R_{ij}^G = 1 - \text{sim}(p_i, p_j), \quad (19)$$

where  $R_{ij}^m$  is the network-specific reliability, reflecting the dissimilarity between proteins  $p_i$  and  $p_j$  within similarity network  $S^m$ .  $R_{ij}^G$  is the global feature reliability, calculated based on the cosine similarity between the fused feature representations of  $p_i$  and  $p_j$ .  $\text{sim}(p_i, p_j)$  is the cosine similarity between the fused features of protein  $p_i$  and  $p_j$ , which is defined as follows:

$$\text{sim}(p_i, p_j) = \frac{H_{p_i} \cdot H_{p_j}}{\|H_{p_i}\| \|H_{p_j}\|}, \quad (20)$$

where  $(\cdot)$  represents the dot product,  $\|\cdot\|$  is the euclidean norm operation.

A self-paced sampling strategy is employed to dynamically select informative negative samples during training. In each iteration (epoch)  $t$ , the number of negative samples  $num_t$  is determined using the following equation:

$$num_t = \lfloor (n_p - 1)\beta \frac{t}{T} \rfloor, \tag{21}$$

where  $T$  is the maximum number of training epochs.  $t$  is the current training epoch.  $\beta$  is the hyperparameter controlling the ratio of the negative sample size to the candidates.  $n_p$  is the number of proteins.  $\lfloor \cdot \rfloor$  represents the rounds down operation.

The self-paced sampling strategy prioritizes high-informative negative samples throughout training. Initially, the selection focuses on the most reliable negative samples within each similarity network  $S^m$ . The candidate samples are sorted based on their network-specific reliability scores  $R_{ij}^m$ . The  $num_t$  most reliable candidates are then chosen to form the network-specific negative sample set  $N^{ms}$ .

In parallel, the strategy feature-based negative sample set  $N^{fs}$  by leveraging global feature reliability scores  $R_{ij}^G$ . Finally, the negative sample set for each contrastive learning is formed by combining the two sets using the union operation, denoted as  $N^m = N^{ms} \cup N^{fs}$ .

The candidate negative samples at each similarity network are sorted by their network-specific reliability scores  $R_{ij}^m$  in descending order.

### Prediction

This section describes the prediction of drug-target interactions. First, CCL-ASPS removes the positive drug-target interactions from all possible drug-target pairs. Then, it generates negative samples by randomly sampling from the remaining unlabeled drug-target pairs. The selected negative samples and the removed positive samples are combined for training. Afterward, a convolutional neural network is employed to extract the joint representations of drug-target pairs:

$$H_{(d_i,p_j)} = Pooling \left( Conv \left( H_{d_i} \| H_{p_j} \right) \right), \tag{22}$$

where  $H_{d_i}$  and  $H_{p_j}$  are the fused representation of drug  $d_i$  and protein  $p_j$  respectively.  $H_{(d_i,p_j)} \in \mathbb{R}^{f_{dp}}$  is the joint representation of drug  $d_i$  and protein  $p_j$ .

The extracted joint representation is then fed into a multilayer perceptron to estimate the interaction probability between drug  $d_i$  and protein  $p_j$ :

$$p_{(d_i,p_j)} = \sigma \left( W_{dp_2} ReLU \left( W_{dp_1} H_{(d_i,p_j)} \right) \right), \tag{23}$$

where  $W_{dp_2} \in \mathbb{R}^{f_{dp} \times 1}$  and  $W_{dp_1} \in \mathbb{R}^{4f_{dp} \times f_{dp}}$  are the learnable weight matrices.  $\sigma(\cdot)$  is the sigmoid activation function.

The binary cross-entropy loss is applied as the objective function:

$$\mathcal{L}_{dp} = - \sum_{(d_i,p_j) \in dpp} \hat{p}_{(d_i,p_j)} \log p_{(d_i,p_j)} + (1 - \hat{p}_{(d_i,p_j)}) \log(1 - p_{(d_i,p_j)}), \tag{24}$$

where  $p_{(d_i,p_j)}$  is the predicted interaction probability between drug  $d_i$  and protein  $p_j$ .  $\hat{p}_{(d_i,p_j)}$  is the ground truth label.  $dpp$  is the training set of drug-protein pairs.  $\mathcal{L}_{dp}$  is the prediction loss.

The final loss function incorporates both the contrastive learning loss  $\mathcal{L}^c$  introduced earlier and the DTI prediction loss  $\mathcal{L}_{dp}$ :

$$\mathcal{L} = \gamma \mathcal{L}^c + (1 - \gamma) \mathcal{L}_{dp}, \tag{25}$$

where  $\gamma$  is a hyperparameter that controls the relative weight of each loss.

### Abbreviations

DTI	Drug-target interaction
CCL	Collaborative contrastive learning
ASPS	Adaptive self-paced sampling strategie
CL	Contrastive learning
MLP	Multilayer perceptron
SAFs	Self-associated features
AAFs	Adjacent-associated features
ACC	Accuracy
AUROC	Area under the receiver operating characteristic curve
AUPR	Area under the precision-recall curve
MCC	Matthews correlation coefficient
F1	F1-score
GNN	Graph neural network
CNN	Convolutional neural network
t-SNE	T-distributed stochastic neighbor embedding
BCELoss	Binary cross-entropy loss
PPIs	Protein-protein interactions

### Acknowledgements

Not applicable.

### Authors' contributions

Z.T. and Y.Y. conceived and designed the experiment. Y.Y. performed the experiment. Z.T. and Y.Y. wrote and revised the manuscript. Z.T., Q.Z. and F.N. supervised the whole process. All authors provided feedback on the manuscript. All authors read and approved the final manuscript.

### Funding

This work has been supported by the National Natural Science Foundation of China (Grant No. 62371423, 61802432), Key Scientific and Technological Project of Henan Province (Grant No. 232102211027).

### Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. In addition, all code and data can be obtained from <https://doi.org/10.5281/zenodo.13329691> [54].

### Data availability

Implementation details with code and dataset can be obtained from <https://doi.org/10.5281/zenodo.13329691>.

### Declarations

### Ethics approval and consent to participate

Not applicable.

**Consent for publication**

All authors read and approved the final manuscript.

**Competing interests**

The authors declare no competing interests.

Received: 18 June 2024 Accepted: 6 September 2024

Published online: 27 September 2024

**References**

- Feng Y, Wang Q, Wang T, et al. Drug target protein-protein interaction networks: a systematic perspective. *Biomed Res Int*. 2017;2017(1):1289259.
- Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, et al. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform*. 2016;17(4):696–712.
- Peng J, Li J, Shang X. A learning-based method for drug-target interaction prediction based on feature representation learning and deep neural network. *BMC Bioinformatics*. 2020;21(13):1–13.
- Xia X, Zhu C, Zhong F, Liu L. MDTips: a multimodal-data-based drug-target interaction prediction system fusing knowledge, gene expression profile, and structural data. *Bioinformatics*. 2023;39(7):btad411.
- Lin S, Wang Y, Zhang L, Chu Y, Liu Y, Fang 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(1):bbab421.
- Wei L, Ye X, Sakurai T. ToxinML: improving peptide toxicity prediction by fusing multimodal information based on mutual information. In: *Proc Conf Res Adapt Convergent Syst*. 2022. pp. 77–82.
- Zhang H, Cui H, Zhang T, Cao Y, Xuan P. Learning multi-scale heterogeneous network topologies and various pairwise attributes for drug-disease association prediction. *Brief Bioinform*. 2022;23(2):bbac009.
- Yang Y, Sun Y, Li F, Guan B, Liu JX, Shang J. MGCNRF: Prediction of Disease-Related miRNAs based on Multiple Graph Convolutional Networks and Random Forest. *IEEE Trans Neural Netw Learn Syst*. 2023.
- Gao Z, Ma H, Zhang X, Wang Y, Wu Z. Similarity measures-based graph co-contrastive learning for drug-disease association prediction. *Bioinformatics*. 2023;39(6):btad357.
- You Y, Chen T, Sui Y, Chen T, Wang Z, Shen Y. Graph contrastive learning with augmentations. *Adv Neural Inf Process Syst*. 2020;33:5812–23.
- Cheng Z, Zhao Q, Li Y, Wang J. IIFDTI: predicting drug-target interactions through interactive and independent features based on attention mechanism. *Bioinformatics*. 2022;38(17):4153–61.
- Song T, Zhang X, Ding M, Rodriguez-Paton A, Wang S, Wang G. DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions. *Methods*. 2022;204:269–77.
- Bai P, Miljković F, John B, Lu H. Interpretable bilinear attention network with domain adaptation improves drug-target prediction. *Nat Mach Intell*. 2023;5(2):126–36.
- Zhao Q, Zhao H, Zheng K, Wang J. HyperAttentionDTI: improving drug-protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*. 2022;38(3):655–62.
- Ru X, Zou Q, Lin C. Optimization of drug-target affinity prediction methods through feature processing schemes. *Bioinformatics*. 2023;39(1):btad615.
- Yang X, Niu Z, Liu Y, Song B, Lu W, Zeng L, et al. Modality-DTA: Multimodality Fusion Strategy for Drug-Target Affinity Prediction. *IEEE/ACM Trans Comput Biol Bioinform*. 2023;20(2):1200–10.
- Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun*. 2017;8(1):573.
- Zhang S, Yang K, Liu Z, Lai X, Yang Z, Zeng J, et al. DrugAI: a multi-view deep learning model for predicting drug-target activating/inhibiting mechanisms. *Brief Bioinform*. 2023;24(1):bbac526.
- Ye Q, Hsieh CY, Yang Z, Kang Y, Chen J, Cao D, et al. A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. *Nat Commun*. 2021;12(1):6775.
- Xiong Z, Liu S, Huang F, Wang Z, Liu X, Zhang Z, et al. Multi-relational contrastive learning graph neural network for drug-drug interaction event prediction. In: *Proc AAAI Conf Artif Intell*. vol. 37. 2023. pp. 5339–47.
- Long Y, Wu M, Liu Y, Fang Y, Kwok CK, Chen J, et al. Pre-training graph neural networks for link prediction in biomedical networks. *Bioinformatics*. 2022;38(8):2254–62.
- Wang H, Huang F, Xiong Z, Zhang W. A heterogeneous network-based method with attentive meta-path extraction for predicting drug-target interactions. *Brief Bioinform*. 2022;23(4):bbac184.
- Wu Z, Xiong Y, Yu SX, Lin D. Unsupervised feature learning via non-parametric instance discrimination. In: *Proc IEEE Conf Comput Vis Pattern Recognit*. 2018. pp. 3733–42.
- Ye M, Zhang X, Yuen PC, Chang SF. Unsupervised embedding learning via invariant and spreading instance feature. In: *Proc IEEE Conf Comput Vis Pattern Recognit*. 2019. pp. 6210–9.
- Ji X, Henriques JF, Vedaldi A. Invariant information clustering for unsupervised image classification and segmentation. In: *Proc IEEE/CVF Int Conf Comput Vis*. 2019. pp. 9865–74.
- Wei C, Liang J, Liu D, Wang F. Contrastive Graph Structure Learning via Information Bottleneck for Recommendation. *Adv Neural Inf Process Syst*. 2022;35:20407–20.
- Wang Y, Wang J, Cao Z, Barati Farimani A. Molecular contrastive learning of representations via graph neural networks. *Nat Mach Intell*. 2022;4(3):279–87.
- Liu X, Song C, Huang F, Fu H, Xiao W, Zhang W. GraphCDR: a graph neural network method with contrastive learning for cancer drug response prediction. *Brief Bioinform*. 2022;23(1):bbab457.
- Chu G, Wang X, Shi C, Jiang X. CuCo: Graph Representation with Curriculum Contrastive Learning. In: *IJCAI*. 2021. pp. 2300–6.
- Zhao X, Wu J, Zhao X, Yin M. Multi-view contrastive heterogeneous graph attention network for lncRNA-disease association prediction. *Brief Bioinform*. 2023;24(1):bbac548.
- Qu Y, He C, Yin J, Zhao Z, Chen J, Duan L. MOVE: Integrating Multi-source Information for Predicting DTI via Cross-view Contrastive Learning. In: *IEEE Int Conf Bioinformatics Biomed (BIBM)*. IEEE; 2022. pp. 535–40.
- Zhu Y, Xu Y, Yu F, Liu Q, Wu S, Wang L. Graph contrastive learning with adaptive augmentation. In: *Proc Web Conf 2021*. 2021. pp. 2069–80.
- Luo X, Ju W, Qu M, Chen C, Deng M, Hua XS, et al. Dualgraph: Improving semi-supervised graph classification via dual contrastive learning. In: *IEEE 38th Int Conf Data Eng (ICDE)*. IEEE; 2022. pp. 699–712.
- Wang X, Liu N, Han H, Shi C. Self-supervised heterogeneous graph neural network with co-contrastive learning. In: *Proc 27th ACM SIGKDD Conf Knowl Discov Data Min*. 2021. pp. 1726–36.
- Li Y, Qiao G, Gao X, Wang G. Supervised graph co-contrastive learning for drug-target interaction prediction. *Bioinformatics*. 2022;38(10):2847–54.
- Wang X, Chen Y, Zhu W. A survey on curriculum learning. *IEEE Trans Pattern Anal Mach Intell*. 2021;44(9):4555–76.
- Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:160902907*. 2016.
- Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *arXiv preprint arXiv:171010903*. 2017.
- Cao DS, Liu S, Xu QS, Lu HM, Huang JH, Hu QN, et al. Large-scale prediction of drug-target interactions using protein sequences and drug topological structures. *Anal Chim Acta*. 2012;752:1–10.
- Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
- Li J, Wang J, Lv H, Zhang Z, Wang Z. IMCHGAN: inductive matrix completion with heterogeneous graph attention networks for drug-target interactions prediction. *IEEE/ACM Trans Comput Biol Bioinform*. 2021;19(2):655–65.
- Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: Predicting drug-target binding affinity with graph neural networks. *Bioinformatics*. 2021;37(8):1140–7.
- Ruan X, Jiang C, Lin P, Lin Y, Liu J, Huang S, et al. MSGCL: inferring miRNA-disease associations based on multi-view self-supervised graph structure contrastive learning. *Brief Bioinform*. 2023;24(2):bbac623.
- Zhang Y, Li J, Lin S, Zhao J, Xiong Y, Wei DQ. An end-to-end method for predicting compound-protein interactions based on simplified homogeneous graph convolutional network and pre-trained language model. *J Cheminform*. 2024;16(1):67.

45. Chu Y, Kaushik AC, Wang X, Wang W, Zhang Y, Shan X, et al. DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. *Brief Bioinform.* 2019;22(1):451–62.
46. Chu Y, Shan X, Chen T, Jiang M, Wang Y, Wang Q, et al. DTI-MLCD: predicting drug-target interactions using multi-label learning with community detection method. *Brief Bioinform.* 2020;22(3):bbaa205.
47. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(D1):D1074–82.
48. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 2022;51(D1).
49. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000;28(1):235–42.
50. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9.
51. Li M, Zhou J, Hu J, Fan W, Zhang Y, Gu Y, et al. Dgl-lifesci: An open-source toolkit for deep learning on graphs in life science. *ACS Omega.* 2021;6(41):27233–8.
52. Gao Z, Jiang C, Zhang J, Jiang X, Li L, Zhao P, et al. Hierarchical graph learning for protein-protein interaction. *Nat Commun.* 2023;14(1):1093.
53. Niwattanakul S, Singthongchai J, Naenudorn E, Wanapu S. Using of Jaccard coefficient for keywords similarity. In: *Proc Int Multiconf Eng Comput Sci.* vol. 1. 2013. pp. 380–4.
54. Yu Y. CCL-ASPS. Zenodo. 2024. <https://doi.org/10.5281/zenodo.13329691>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.