Appendix

I. Atom feature dimension

For drugs, each SMILES sequence is converted into a molecular graph where the features of atoms are defined by 75 dimensional physicochemical properties as shown in Table I.

TABLE I. The List of Predefined Atom Features.

Atom Feature	Size	Description
Atomic symbol	44	[C, N, O, S, F, Si, P, Cl, Br, Mg, Na, Ca, Fe, As, Al, I, B, V, K, Tl, Yb, Sb, Sn, Ag, Pd, Co, Se, Ti, Zn, H, Li, Ge, Cu, Au, Ni, Cd, In, Mn, Zr, Cr, Pt, Hg, Pb, Unknow] (One-hot)
Atomic degrees	11	Degree of atoms in a drug [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] (One-hot)
Implicit value	7	Implicit valence of atoms [0, 1, 2, 3, 4, 5, 6]
Formal charge	1	The formal charge of the atom, which usually ranges from -3 to +3
Radical electrons	1	The number of free radical electrons of an atom, which usually ranges from 0 to 2
Hybridization	5	The atomic hybridization mode [SP, SP2, SP3, SP3D, SP3D2] (One-hot)
Total hydrogen atoms	5	Total number of hydrogen atoms in the atom [0, 1, 2, 3, 4] (One-hot)

II. Robustness on imbalanced dataset

Since DTI datasets in practical applications often exhibit imbalance, we constructed imbalanced datasets to assess the robustness of CSCL-DTI. The ratio of positive to negative samples in the dataset was increased from 1:1 to 1:3. For GPCR dataset, we augmented its negative samples by selecting drug-target pairs from the GLASS database [1] with an affinity threshold below 6.0. The affinity threshold is defined as the negative logarithm of binding affinity values, including pIC50, pKi, and pEC50. This augmentation resulted in a positive-to-negative ratio of 1:3. For Human dataset, we followed the previous work [2] to maintain a positive-to-negative ratio of 1:3. For DrugBank dataset, negative samples were randomly selected from unknown drug-protein pairs, ensuring a 1:3 ratio with positive samples. Table II shows the comparison results between our proposed CSCL-DTI and baseline methods on three benchmarking datasets. In general, our proposed CSCL-DTI outperforms the state-of-the-art baseline methods on 14 out of 18 situations (3 metrics $\times$ 2 ratios $\times$ 3 datasets). From the observation we found that the AUC scores of most methods maintain stable (e.g., IIFDTI on Human dataset), while there is a corresponding decrease in AUPR scores. This phenomenon arises from the focus on recall and precision for the positive class (the minority class), resulting in AUPR penalizing false positives more severely compared to AUC score [3]. These findings reveal the reliability and robustness of our proposed CSCL-DTI, even when confronted with imbalanced datasets.

TABLE II. The AUC, AUPR and Recall results of our proposed CSCL-DTI and baseline methods on the imbalanced dataset.

Ratios		1:1			1:3		
Datasets	Methods	AUC	AUPR	Recall	AUC	AUPR	Recall
GPCR	DeepDTA	0.776±0.006	0.762±0.015	0.712±0.015	0.921±0.005	0.687±0.013	0.679±0.009
	DeepConv-DTI	0.752±0.011	0.685±0.010	0.713±0.021	0.908±0.007	0.611±0.008	0.668±0.011
	MolTrans	0.807±0.004	0.788±0.009	0.762±0.014	0.962±0.003	0.756±0.009	0.773±0.013
	GraphDTA	0.840±0.004	0.836±0.006	0.790±0.006	0.963±0.004	0.762±0.005	0.771±0.008
	TransformerCPI	0.842±0.007	0.837±0.010	0.796±0.015	0.971±0.006	0.761±0.007	0.789±0.014
	IIFDTI	0.845±0.008	0.842±0.007	0.783±0.017	**0.979±0.004**	0.777±0.007	0.764±0.016
	CSCL-DTI	**0.860±0.008**	**0.862±0.009**	**0.799±0.018**	0.976±0.005	**0.788±0.009**	**0.795±0.016**
Human	DeepDTA	0.972±0.001	0.973±0.002	0.935±0.017	0.969±0.006	0.968±0.005	0.891±0.012
	DeepConv-DTI	0.967±0.002	0.964±0.004	0.907±0.023	0.965±0.005	0.961±0.005	0.875±0.014
	MolTrans	0.974±0.002	0.976±0.003	0.933±0.022	0.980±0.004	0.961±0.008	0.823±0.012
	GraphDTA	0.972±0.005	0.973±0.005	0.946±0.006	0.967±0.007	0.927±0.005	0.642±0.004
	TransformerCPI	0.970±0.006	0.974±0.005	0.937±0.011	0.965±0.003	0.965±0.007	0.895±0.011
	IIFDTI	0.984±0.003	0.985±0.003	0.947±0.017	0.984±0.002	0.967±0.003	0.898±0.006
	CSCL-DTI	**0.987±0.001**	**0.988±0.001**	**0.951±0.005**	**0.986±0.002**	**0.969±0.003**	**0.899±0.007**
DrugBank	DeepDTA	0.784±0.004	0.519±0.007	0.635±0.010	0.756±0.005	0.346±0.006	0.551±0.004
	DeepConv-DTI	0.782±0.005	0.472±0.005	0.626±0.016	0.754±0.004	0.307±0.005	0.549±0.008
	MolTrans	0.501±0.010	0.203±0.006	0.417±0.015	0.728±0.006	0.208±0.005	0.543±0.007
	GraphDTA	0.786±0.006	0.517±0.008	0.638±0.008	0.759±0.004	0.344±0.006	0.564±0.004
	TransformerCPI	0.782±0.005	0.500±0.015	0.660±0.007	0.752±0.006	0.323±0.013	0.596±0.007
	IIFDTI	0.797±0.004	0.527±0.009	0.679±0.008	0.761±0.004	0.368±0.003	0.632±0.006
	CSCL-DTI	**0.808±0.002**	**0.557±0.007**	**0.689±0.010**	**0.769±0.003**	**0.376±0.002**	**0.634±0.006**

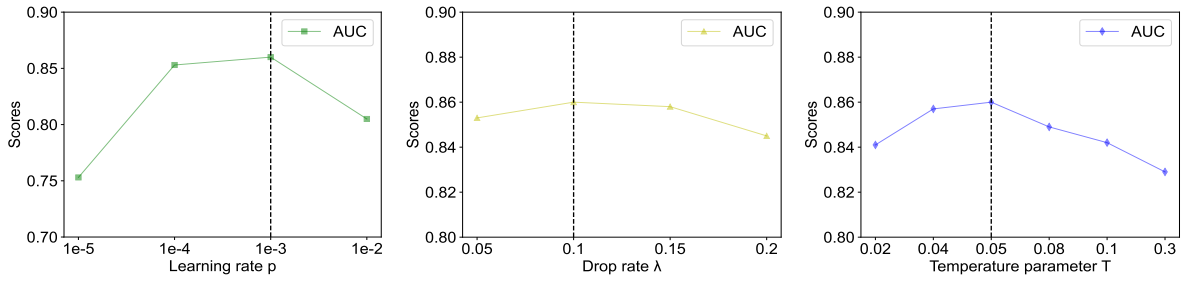
III. Parameter Analysis

The performance of our model is influenced by several significant parameters, such as the number of GCN layers n, learning rate p, drop rate λ, temperature parameter T, weight factor α and β. The following is the analysis of the parameters learning rate p, drop rate λ, temperature parameter T.

1) Impact of learning rate: We choose p values from $\{1e-5, 1e-4, 1e-3, 1e-2\}$. As depicted in Fig. 1a, our model achieves the best performance when p is set to $1e-3$.

2) Impact of drop rate: The drop rate λ also significantly influences model training. We evaluate our model by selecting λ from {0.05, 0.1, 0.15, 0.2}. Fig. 1b shows that our model exhibits an increase followed by a decrease in performance, and $\lambda = 0.1$ yields the optimal model performance.

3) Impact of temperature parameter: In contrastive learning, the temperature parameter (T) is a hyperparameter that adjusts the similarity measure. By varying T from {0.02, 0.04, 0.05, 0.08}, we observe that the optimal value for our model is 0.05, as shown in Fig. 1c.



(a) (b) (c)

Fig. 1. Comparison between CSCL-DTI and its variants on GPCR dataset.

IV. CASE STUDY

We conducted a case study on the DrugBank dataset to further validate the effectiveness of our proposed CSCLDTI. Specifically, we applied CSCL-DTI for *de novo* predictions on the important drug *Diacerein* (DrugBank ID: DB11994) and target *Aspartate aminotransferase* (Uniprot ID: Q2TU84), respectively. More specifically, following the previous work [4], we utilize all the known drug-target pairs in DrugBank as training samples to train the model, then we use the pre-trained model to predict the interactive probabilities between them and known drugs or targets. For the predicted results, we sort the candidate drugs (or targets) according to their predicted scores. Finally, the predicted interactions are verified by searching the previous literature on PubMed database.

Table III illustrates the top 10 predicted candidate drugs for the new target *Aspartate aminotransferase* among a total of 6,645 drugs. From the table we can find that 5 candidate drugs are successfully predicted in the top 10 predicted results (marked in bold).

TABLE III. The predicted candidate drugs for new target *Aspartate aminotransferase*.

Rank	Drug name	DrugBank ID	Evidence
1	**N-Acetylglucosamine**	**DB00141**	**PMID: 11327813**
2	Adenosine Phosphate	DB00131	Unconfirmed
3	Adenosine-5'-triphosphate	DB00171	Unconfirmed
4	Adenosine-5'-phosphosulfate	DB03708	Unconfirmed
5	**gamma carboxyl glutamic acid**	**DB03847**	**PMID: 6117008**
6	Hadacidin	DB02109	Unconfirmed
7	**Flavin adenine dinucleotide**	**DB03147**	**PMID: 11455601**
8	Adenosine 3',5'-diphosphate	DB01812	Unconfirmed
9	**2'-Deoxyadenosine 5'-triphosphate**	**DB03222**	**PMID: 7516581**
10	**Aspartic acid**	**DB00128**	**PMID: 26232224**

REFERENCES

[1] W. K. Chan, H. Zhang, J. Yang, J. R. Brender, J. Hur, A. Özgür, and Y. Zhang, "GLASS: a comprehensive database for experimentally validated gpcr-ligand associations," *Bioinformatics*, vol. 31, no. 18, pp. 3035–3042, 2015.

[2] M. Li, Z. Lu, Y. Wu, and Y. Li, "BACPI: a bi-directional attention neural network for compound–protein interaction and binding affinity prediction," *Bioinformatics*, vol. 38, no. 7, pp. 1995–2002, 2022.

[3] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 233–240.

[4] Z. Cheng, Q. Zhao, Y. Li, and J. Wang, "IIFDTI: predicting drug–target interactions through interactive and independent features based on attention mechanism," *Bioinformatics*, vol. 38, no. 17, pp. 4153–4161, 2022.