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. Baseline methods

To demonstrate the efficacy of our model, we compare CSCL-DTI with six state-of-the-art methods.

- **DeepDTA** [1] adopts two separate CNN modules to predict the binding affinity between a drug-target pair by extracting features from SMILES and amino acid sequences, respectively.
- **DeepConv-DTI** [2] utilizes the drug molecular fingerprints and protein amino acid sequences as input. It also employs CNN to capture local residual patterns.
- **MolTrans** [3] employs an augmented Transformer to extract the contextual features from both drug SMILES sequences and amino acid sequences of protein.
- **GraphDTA** [4] adopts GCN, Graph Attention Networks (GAT) [5], and Graph Isomorphism Networks (GIN) [6] to extract molecular graph features of drugs, while utilizing CNNs to extract features of amino acid sequences for predicting the binding affinity.
- **TransformerCPI** [7] extracts the useful features from drug SMILES and protein amino acid sequences by employing an attention-based Transformer for predicting compound-protein interaction.
- **IIFDTI** [8] fuses the interactive and independent features between drug-target pairs to predict DTI. On one hand, for drug-target pairs, it extracts interactive features from substructures using a bidirectional encoder-decoder extractor. On the other hand, it separately models the independent features of drugs and proteins by employing GAT and CNN, respectively.

III. 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 [9] 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 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 two benchmarking datasets. In general, our proposed CSCL-DTI outperforms the state-of-the-art baseline methods on 14 out of 18 situations (3 metrics × 2 ratios × 3 datasets). 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**
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**

IV. 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. For these parameters, we run the sensitivity analysis. Note that here all the experiments are conducted on GPCR dataset.

1) Impact of the number of GCN layers: We evaluate our model by varying n from 1 to 4 with a step value of 1. Fig. 1a displays how the performance gradually rises and finally falls as n varies, with $n = 2$ achieving its best performance.

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

3) 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. 1c shows that our model

exhibits an increase followed by a decrease in performance, and $\lambda = 0.1$ yields the optimal model performance.

4) 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. 1d.

5) Impact of weight factor: Weight factor α and β represent the proportions of contrastive loss and cross-entropy loss in the total loss. To evaluate their impact, we choose their values from {0.01, 0.1, 1, 1.5, 50, 100}. Fig. 1e and Fig. 1f demonstrate that the best performance is achieved when α and β are simultaneously set to 0.1 and 1.5, respectively.

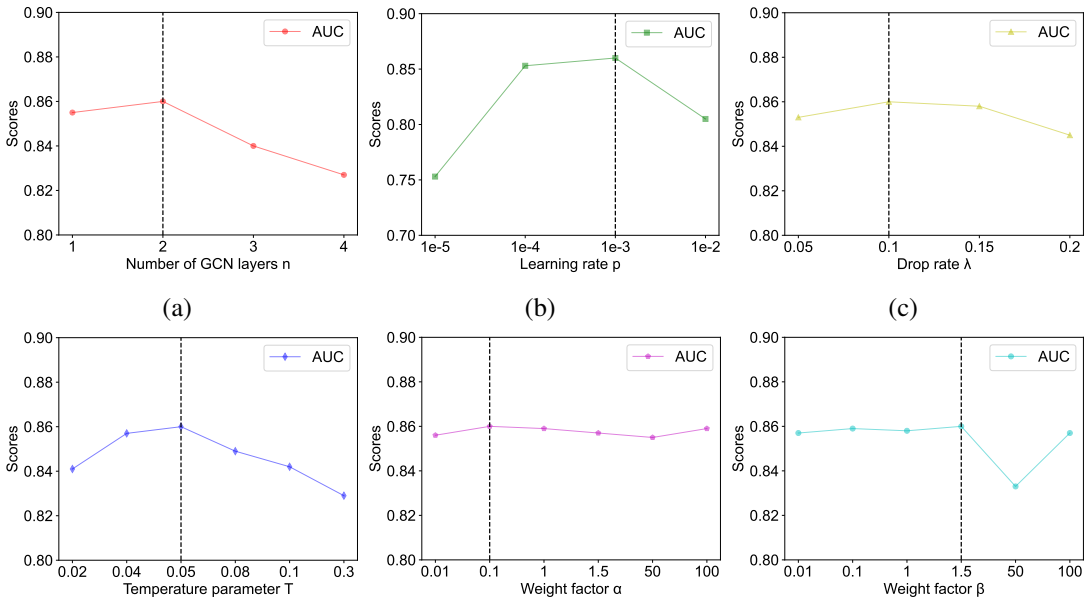

Fig. 1. Parameter sensitivity analysis for CSCL-DTI on GPCR dataset.

V. Case Study

We conducted a case study on the DrugBank dataset to further validate the effectiveness of our proposed CSCL-DTI. 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 [8], 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 presents the top 10 predicted candidate targets for the new drug *Diacerein* predicted by CSCL-DTI among a total of 4,254 targets. From the result we can observe that 5 out of 10 targets were successfully predicted (marked in bold). For example, the interaction between drug *Diacerein* and target *Nitric oxide synthase* (Uniprot ID: P35228) are confirmed by the literature (PubMed ID: 12747270) [10]. Specifically, in experimental

canine osteoarthritic cartilage, *Diacerein* decreases the amount of cartilage chondrocyte DNA fragmentation and death while also inhibiting caspase-3 and inducible *Nitric oxide synthase*. This offers further new information regarding the mechanisms of action of *Diacerein* on osteoarthritis development [10]. Similarly, Table IV 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). All these results suggested that CSCL-DTI is a beneficial method for accurately predicting interacting candidates for unknown drugs and targets. This can help biologists and pharmacologists identify possible protein targets for future drug development.

Table. III. The predicted candidate targets for new drug *Diacerein*.

Rank	Target name	Target Uniprot ID	Evidence
1	**Nitric oxide synthase**	**P35228**	**PMID: 12747270**
2	Retinoic acid receptor gamma	P13631	Unconfirmed
3	Muscarinic acetylcholine receptor	P11229	Unconfirmed
4	**Glutamate receptor**	**Q05586**	**PMID: 24121043**
5	**TGF-beta receptor type-2**	**P37173**	**PMID:10329300**
6	**Cytochrome P450**	**P11510**	**PMID: 34821124**
7	**Bile salt export pump**	**O95342**	**PMID: 29355060**
8	Cholinesterase	P06276	Unconfirmed
9	Estrogen receptor	P03372	Unconfirmed
10	Prostaglandin G/H synthase 1	P23219	Unconfirmed

Table. IV. 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] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018.

[2] I. Lee, J. Keum, and H. Nam, "DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences," *PLoS computational biology*, vol. 15, no. 6, p. e1007129, 2019.

[3] K. Huang, C. Xiao, L. M. Glass, and J. Sun, "MolTrans: molecular interaction transformer for drug–target interaction prediction," *Bioinformatics*, vol. 37, no. 6, pp. 830–836, 2021.

[4] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: Predicting drug–target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp. 1140–1147, 2021.

[5] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[6] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[7] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, and M. Zheng, "TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments," *Bioinformatics*, vol. 36, no. 16, pp. 4406–4414, 2020.

[8] 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.

[9] 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.

[10] J. Pelletier, F. Mineau, C. Boileau, and J. Martel-Pelletier, "Diacerein reduces the level of cartilage chondrocyte dna fragmentation and death in experimental dog osteoarthritic cartilage at the same time that it inhibits caspase-3 and inducible nitric oxide synthase," *Clinical and experimental rheumatology*, vol. 21, no. 2, pp. 171–178, 2003.