This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Multiview Feature Fusion and Contrastive Learning for Drug-Target Interaction Prediction

Xiaoting Zeng[1,3], Li Li[2,3], Yu Liang[1,3], Weilin Chen[4,⋆], Baiying Lei[2,3,⋆]

[1] School of Computer and Software, Shenzhen University, Shenzhen, China.
[2] School of Biomedical Engineering, Shenzhen University, Shenzhen, China.
[3] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Shenzhen, China.
[4] Marshall Laboratory of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen, China.

Abstract. Drug-target interaction (DTI) prediction is crucial for drug discovery, as it accelerates candidate screening and reduces development costs. However, existing computational methods are often limited to a single perspective and cannot simultaneously consider the biological information and complex associations of drugs and targets. Although multimodal data have been introduced, the complementarity and interaction of multi-source information remain underutilized, making efficient multi-view feature fusion a key challenge. In this paper, we propose a DTI prediction framework based on multi-view feature fusion and contrastive learning, named MFCL-DTI. It integrates sequence feature as well as structural and semantic information of heterogeneous graph. A multi-view adaptive fusion module facilitates cross-view feature fusion, while multi-view contrastive learning enhances feature representation. Experimental results demonstrate that MFCL-DTI outperforms existing methods, validating its effectiveness in DTI prediction.

Keywords: Drug-target interaction · Multi-view · Contrastive learning · Multi-source information · Feature fusion.

1 Introduction

Drug-target interaction (DTI) refers to drug molecules binding to biological targets to modulate biological functions. In drug development, DTI validation is crucial for confirming drug efficacy. Accurate DTI prediction can narrow candidate drug screening scope, reducing experimental costs and development cycles [3, 8]. Traditional experimental methods are time-consuming and labor-intensive, prompting a shift toward computational approaches [2]. In the big data era, computational DTI prediction aims to guide in vivo experiments and wet lab validation, accelerating drug discovery and repositioning [7, 19].

⋆ Corresponding authors: Weilin Chen (Email: cwl@szu.edu.cn), and Baiying Lei (Email: leiby@szu.edu.cn)

Recent advancements in deep learning, coupled with the accumulation of large-scale biomedical data, have significantly improved computational model-based DTI prediction methods. Current methods are categorized into sequence-based, network-based, and hybrid approaches [14]. Sequence-based methods predict DTIs using molecular sequences, such as drug structures or protein sequences, with deep learning for feature extraction. DeepConv-DTI [9] employs CNNs for proteins and ECFP [15] for drugs, while HyperAttentionDTI [20] uses attention mechanisms for interaction modeling. Transformer-based models like TransformerCPI [1] and MolTrans [6] also show strong performance. However, sequence-based methods rely solely on sequence information often fails to fully reveal the complex biological relationships.

Network-based methods build heterogeneous networks integrating drugs, targets, and biological entities, using graph techniques for DTI prediction. DTINet [12] employs inductive matrix completion to learn features from heterogeneous data. NeoDTI [17] and EEG-DTI [13] use graph convolution network for node representation, while IMCHGAN [10] applies graph attention network to relational networks. SGCL-DTI [11] leverages supervised graph contrastive learning. However, heterogeneous graphs often miss detailed molecular and sequence features, limiting interaction capture, and most graph methods focus on single views, neglecting complex drug-target relationships.

Hybrid methods for DTI prediction integrate multimodal data to address single-modal limitations, such as sequences and heterogeneous graphs. MultiDTI [21] maps heterogeneous and sequence data into a shared space, while HampDTI [18] extracts features from sequences, integrates them with heterogeneous graph learning to capture meta-paths, and combines meta-path graph embeddings for DTI prediction. MOVE [14] uses sequence and network features with contrastive learning for enhanced representation. MMA-DPI [4] based on multimodal attribute learning, fuses micro-level attributes and macro-level attributes for DTI prediction. However, these methods often rely on simplistic fusion strategies like feature concatenation, lacking deeper exploration of fusion mechanisms. Additionally, interactions between modalities are frequently overlooked, underutilizing the full potential of multimodal data.

To address these challenges, we propose a DTI prediction method based on multiview feature fusion and contrastive learning. Our approach uses multi-scale convolution to extract features from sequences while incorporating heterogeneous network to extract structural and semantic features from both neighborhood and meta-path views. To achieve efficient fusion of features from different views, we design a Multiview Adaptive Fusion Module (MVAF) to facilitate feature interaction and integration across views, fully leveraging multi-source information. Furthermore, multiview contrastive learning enhances feature representations by contrasting sequence, neighborhood, and meta-path views. The main contributions are: (1)We propose an innovative DTI prediction method named MFCL-DTI, which leverages multiview feature fusion and contrastive learning to efficiently extract features from multimodal data. (2) We design a multiview adaptive fusion module to enable cross-view feature interaction and integration,

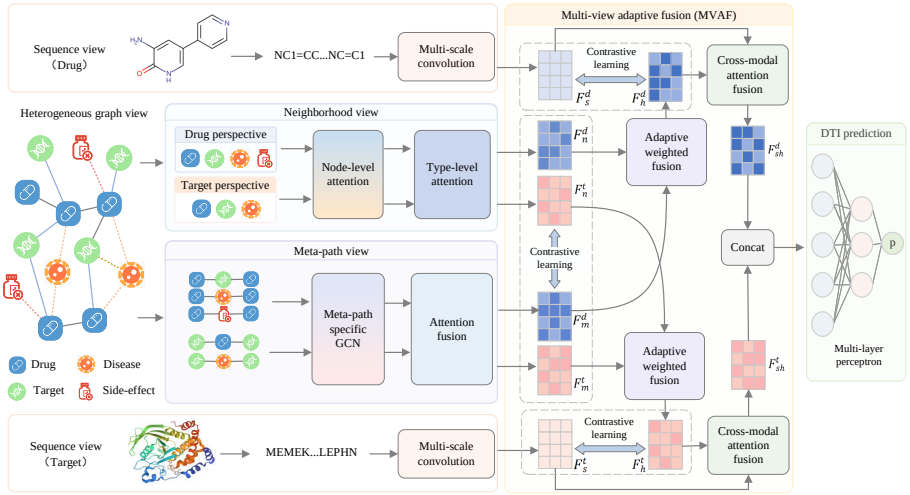

Fig. 1. Overview of MFCL-DTI framework, integrating heterogeneous graphs and sequence features for DTI prediction.

and introduce multiview contrastive learning to enhance feature representation. (3) Experiments show that MFCL-DTI outperforms state-of-the-art methods, validating its effectiveness in DTI prediction.

2 Methodology

2.1 Architecture

The overall workflow of MFCL-DTI is illustrated in Fig. 1, including three key steps: multiview feature extraction, multiview adaptive fusion, and DTI prediction. First, features of drugs and targets are extracted from the sequence view using a multi-scale convolution and from the heterogeneous graph view, which captures structural (neighborhood view) and semantic (meta-path view) information. MVAF Module integrates features from different views with contrastive learning for feature enhancement. Finally, the fused features are concatenated and processed by a multilayer perceptron to produce DTI prediction.

2.2 Sequence View Feature Extraction

Drug and target sequences, represented by SMILES and FASTA, contain rich biological information. Traditional single-scale convolution captures local patterns but struggles with global structures and long-range dependencies. Therefore, we employ multi-scale convolution to extract features at varying receptive fields, capturing richer multi-scale information. Sequences are encoded into low-dimensional vectors by embedding layer, yielding E_s for drug sequences D_s and

similarly for target sequences. Multi-scale convolution extracts local features by convolving E_s with varying kernel sizes and finally we obtain drug feature F_{cnn}:

$$F_{cnn} = \text{Concat}(\text{Conv1D}_k E_s),\ k \in \{4, 8, 16\} \tag{1}$$

where k represents the size of the convolution kernel, and each convolution kernel extracts features of different scales in the drug sequence.

Subsequently, The feature F_{cnn} is processed with the ReLU activation function, followed by a convolution layer to extract high-order features. Mean pooling is then applied to obtain the drug sequence feature F_s^d:

$$F_s^d = \text{MeanPool}(\text{Conv1D}(\text{ReLU}(F_{cnn}))) \tag{2}$$

The same processing procedure is applied to the target sequence T_s, ultimately yielding the feature representation F_s^t of the target sequence.

2.3 Heterogeneous Graph Feature Extraction

To leverage diverse association information in heterogeneous networks, the neighborhood view captures structural features from direct node relationships, while the meta-path view extracts semantic features to uncover potential associations. For feature initialization, we use one-hot encoding to ensure node-type-specific features reside in independent spaces, then project them into a shared vector space using type-specific mapping matrix W_t. For node v, its initial feature o_v is mapped as follows to obtain f_v:

$$f_v = \text{Elu}(o_v \cdot W_t + b_o) \tag{3}$$

where $\text{Elu}(\cdot)$ is a nonlinear activation function, and b_o is a bias vector.

Neighborhood view feature extraction. For the neighborhood view, node features are learned by aggregating information from the node and its one-hop neighbors, capturing structural features in the heterogeneous network. Given a node type $p \in P$, N_v denote the one-hop neighbors of node v, and N_v^p represent neighbors of type p. During aggregation, node-level attention is used to weight contributions from different neighbors, as different types and nodes contribute unevenly to others. For type p, its attention coefficient $\tilde{\alpha}_v$ is defined as:

$$\alpha_v = \text{LeakyReLU}(\alpha_p^{\mathrm{T}} \cdot f_v + (\alpha_p^{\mathrm{T}} \cdot f_{v'})^{\mathrm{T}}) \tag{4}$$

$$\widetilde{\alpha}_v = \frac{\exp(\alpha_v)}{\Sigma_{v' \in N_v^p} \exp(\alpha_{v'})} \tag{5}$$

where α_p denotes the node-level attention vector for node type p, and $\alpha_{v'}$ represents the attention value of the neighboring node v' with respect to node v. Subsequently, by applying the node-level attention mechanism, the neighbors of each type are aggregated into a unified representation as follows:

$$f_v^p = \text{Elu}\left(\sum_{v\prime \in N_v^p} \widetilde{\alpha}_v \cdot f_{v'}\right) \tag{6}$$

where $f_{v'}$ represents the projected feature of node v'.

To further integrate features of all types for node v, we introduce a type-level attention mechanism, which dynamically evaluates the relative importance of each type's embedding by assigning weight coefficients. Specifically, the importance of each node type is measured as follows:

$$\beta_p = \text{softmax}\left(\frac{1}{\left|V_{\psi(v)}\right|}\sum_{v\in V_{\psi(v)}} \alpha_n^{\mathrm{T}} \cdot \tanh\left(f_v^p \cdot W_n + b_n\right)\right) \tag{7}$$

where $V_{\psi(v)}$ denotes the set of all nodes of type $\psi(v)$, and α_n is the type-level attention vector. β_p represents the importance of neighbors of type p to node v. W_n and b_n are learnable parameter matrices and bias vectors, respectively.

Finally, the feature F_n^v of node v, incorporating neighbor features, is obtained:

$$F_n^v = \sum_{p\in P} \beta_p \cdot f_v^p \tag{8}$$

Meta-path view feature extraction. To fully leverage the heterogeneous network information, we also extract implicit high-order semantic relationships from the meta-path view, enriching feature information for node interactions and enhancing network representation. Let M denote the set of predefined meta-paths. For each meta-path $m \in M$, given the neighbor set N_v^m (nodes connected to v via m), we construct a meta-path-specific GCN to aggregate information from meta-path neighbors as follows:

$$f_v^m = f_v \cdot [d(v)+1]^{-1} + f_{v'} \cdot [\sqrt{(d(v)+1)(d(v')+1)}]^{-1} \tag{9}$$

where f_v and $f_{v'}$ represent the projected features of node v and node v', and $d(\cdot)$ denotes the degree of the corresponding node.

For node v, the influence weights of different meta-path neighbors on its embedding vary. Therefore, a semantic-level attention mechanism is introduced to weight and integrate information from meta-path neighbors. The feature f_v^m undergoes an attention processing process similar to formulas (7) and (8), and finally generates the final meta-path feature F_m^v for node v.

2.4 Multi-view Adaptive Fusion

To integrate multiview features, we design MVAF module, combining adaptive weight fusion and cross-modal attention to merge meta-path, neighborhood, and sequence views. MVAF dynamically adjusts feature weights to capture complementary information, generating comprehensive representations for multi-source data utilization. In the heterogeneous network, the meta-path view models semantic associations, while the neighborhood view captures structural interactions. Specifically, we concatenate the two features to obtain F_{mn}:

$$F_{mn}^x = \text{Concat}(F_m^x, F_n^x), x \in \{d, t\} \tag{10}$$

where F_m^d and F_n^d represent the features of drug d in the meta-path view and

neighbor view, respectively. Similarly, F_m^t and F_n^t denote the features of target t in the meta-path view and neighbor view.

Since the contributions of different view features to the overall representation are interdependent, we design a dynamic weighting mechanism to assign adaptive fusion weights to the meta-path view F_m^x and neighborhood view F_n^x. The weight ω is learned in a data-driven manner and the fused network features for drugs F_h^d and targets F_h^t are computed as:

$$\omega = \sigma(F_{mn}^x \cdot W_c + b_c) \tag{11}$$

$$F_h^x = \omega \odot F_m^x + (1 - \omega) \odot F_n^x, \quad x \in \{d, t\} \tag{12}$$

where W_c and b_c are learnable parameters, and $\odot$ is element-wise multiplication.

To integrate multimodal feature information, we fuse the network view and sequence view through cross-modal attention. Learnable attention weights are assigned to each view, and the features are weighted and fused as follows:

$$F_{sh}^x = \sigma(w_s^x) \cdot F_s^x + \sigma(w_h^x) \cdot F_h^x, \quad x \in \{d, t\} \tag{13}$$

where w_s^x and w_h^x are learnable parameters, and F_{sh}^d and F_{sh}^t are the final fused features for drugs and targets, respectively.

2.5 Multi-view Contrastive Learning

To fully utilize multi-view information, we construct a contrastive learning task to optimize representations by maximizing mutual information of positive pairs and distinguishing negative pairs, enhancing prediction performance and cross-view fusion. Inspired by maximum mutual information [16] and the InfoNCE model [5], we use the InfoNCE loss to measure similarity between meta-path and neighbor views. For drugs, the loss between F_m^d and F_n^d is calculated as:

$$\mathcal{L}_d = -\log \frac{\exp\left(\text{sim}(F_m^d, F_n^d)/\tau\right)}{\Sigma_{d' \in V_{\psi(d)}} \exp\left(\text{sim}(F_m^d, F_n^{d'})/\tau\right)} \tag{14}$$

where $\text{sim}(\cdot)$ denotes cosine similarity, and τ is the temperature coefficient. The loss calculation for targets is analogous, yielding $\mathcal{L}_t$.

For the sequence and heterogeneous network view, contrastive learning is performed by computing the similarity of their respective node embeddings. The model is optimized by maximizing the similarity of positive pairs and minimizing differences between negative pairs. For drugs, the contrastive loss between the sequence and heterogeneous view is calculated as:

$$\mathcal{L}_d^s = -\log \frac{\exp\left(\text{sim}(F_s^d, F_h^d)/\tau\right)}{\Sigma_{d' \in V_{\psi(d)}} \exp\left(\text{sim}(F_s^d, F_h^{d'})/\tau\right)} \tag{15}$$

$$\mathcal{L}_d^h = -\log \frac{\exp\left(\text{sim}(F_h^d, F_s^d)/\tau\right)}{\Sigma_{d' \in V_{\psi(d)}} \exp\left(\text{sim}(F_h^d, F_s^{d'})/\tau\right)} \tag{16}$$

Similarly, the contrastive losses for the sequence and network view of targets are $\mathcal{L}_t^s$ and $\mathcal{L}_t^h$. The final cross-modal loss for the sequence and heterogeneous network view is computed as:

$$\mathcal{L}_{cr} = 0.25(\mathcal{L}_d^s + \mathcal{L}_d^h + \mathcal{L}_t^s + \mathcal{L}_t^h) \tag{17}$$

2.6 DTI Prediction

For DTI prediction, the final fused features F_{sh}^{d} and F_{sh}^{t} are concatenated and input into multi-layer perceptron (MLP). Finally, the model outputs a score $\hat{y}$, representing the interaction probability:

$$\hat{y} = \sigma(\text{MLP}(\text{Concat}(F_{sh}^{d}, F_{sh}^{t}))) \tag{18}$$

For DTI prediction task, we use the Binary Cross-Entropy (BCE) loss:

$$\mathcal{L}_{BCE} = -\left[y \cdot \log \hat{y} + (1-y) \cdot \log(1-\hat{y})\right] \tag{19}$$

where y is the true label.

For model optimization, the loss from the contrastive learning task is combined with the loss of DTI prediction to form the total loss function:

$$\mathcal{L} = \mathcal{L}_{BCE} + \lambda(\mathcal{L}_{cr} + \mathcal{L}_{d} + \mathcal{L}_{t}) \tag{20}$$

where λ is a hyperparameter that balances the contribution of the contrastive learning loss and the DTI prediction loss.

3 Experiments and Results

Dataset. The DTI prediction task employs a drug-related heterogeneous network dataset from DTInet [12], integrating multi-source data to establish a drug-target information network through shared biomedical entities. The network comprises 708 drugs, 1512 targets, 4192 side effects, and 5603 diseases, with interactions including 1923 drug-target, 10036 drug-drug, 199214 drug-disease, 80164 drug-side effect, 7363 target-target, and 1596745 target-disease relationships. Drug and target sequences are represented using SMILES sequences and amino acid sequences, respectively, sourced from DrugBank and UniProt.

Implementation Details and Evaluation Metrics. Due to the variable lengths of drug SMILES and target sequences, we set maximum lengths of 150 for SMILES and 1500 for target sequences, truncating longer sequences and zero-padding shorter ones. In experiments, the latent space dimensionality is 1024, batch size is 16, and the Adam optimizer is used with a learning rate of 1×10^{-5}. Training runs for a maximum of 40 epochs. Model effectiveness is evaluated using six metrics: Area Under the ROC Curve (AUC), Area Under the Precision-Recall Curve (AUPR), Accuracy (ACC), Precision, Recall, and F1 Score (F1).

3.1 Comparative experiments

To evaluate the performance of MFCL-DTI, we conduct comparative experiments with seven mainstream methods, including DTINet [12], NeoDTI [17], MultiDTI [21], IMCHGAN [10], HampDTI [18], SGCL-DTI [11] and MOVE [14].

Table 1. Comparative experimental results.

Methods	AUC	AUPR	ACC	Precision	Recall	F1
DTInet	0.8612	0.8861	0.8029	0.8263	0.7853	0.7996
NeoDTI	0.9150	0.9173	0.8550	0.8505	0.8628	0.8562
MultiDTI	0.8948	0.9030	0.8133	0.8230	0.8207	0.8101
IMCHGAN	0.9193	0.9240	0.8398	0.8619	0.8083	0.8342
HampDTI	0.9115	0.9079	0.8542	0.8400	0.8750	0.8571
SGCL-DTI	0.9220	0.9321	0.8362	0.8137	0.8828	0.8425
MOVE	0.9328	0.9280	0.8591	**0.8467**	0.8794	0.8599
MFCL-DTI	**0.9472**	**0.9512**	**0.8698**	0.8452	**0.9095**	**0.8665**

As shown in Table 1, MFCL-DTI outperforms all baseline models across evaluation metrics and achieves the best performance in AUC (0.9472) and AUPR (0.9512), with relative improvements of 1.54% and 2.50% over the second-best model. The model also demonstrates strong performance in ACC, Recall, and F1, indicating superior discriminative capability, overall accuracy, and balanced performance.

3.2 Ablation Study

To assess the contribution of different modules to model performance, we conducted ablation study with six variants: (i) w/o Seq: removing sequence information; (ii) w/o HN: removing the heterogeneous graph; (iii) w/o MVAF: replacing the MVAF with simple concatenation; (iv) w/o HN-C: removing contrastive learning between the neighbor and meta-path views; (v) w/o SH-C: removing contrastive learning between the sequence and graph views; and (vi) w/o Contrast: removing all contrastive learning strategies. As shown in Fig. 2, the complete MFCL-DTI model outperforms all variants in metrics such as AUC, AUPR, ACC, and F1 score, validating the effectiveness of multiview feature fusion and contrastive learning in enhancing DTI prediction performance.

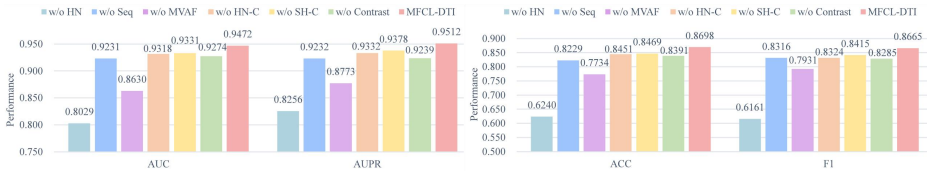

Fig. 2. Ablation experiments for verifying the effectiveness of model components.

3.3 Case Study

To validate MFCL-DTI's effectiveness, we conducted case study, presenting the top 10 ranked DTI pairs in Table 2 and largely confirmed in DrugBank database. Detailed analysis revealed strong alignment with known mechanisms, such as DB01142 inhibiting P23975 for antidepressant effects, DB00363 targeting P08908 for antipsychotic action, and DB0046 inhibiting P35354 for anti-inflammatory effects. These high-confidence predictions demonstrate MFCL-DTI's capability in identifying potential DTIs.

Table 2. Top 10 DTI predictions ranked by prediction scores.

DrugBank ID	Drug Name	Uniprot ID	Target name	Ground truth	Predicted label	Prediction score	Evidence
DB01142	Doxepin	P23975	Sodium-dependent norepinephrine transporter	1	1	0.9956	DrugBank
DB00363	Clozapine	P08908	5-hydroxytryptamine receptor 1A	1	1	0.9948	DrugBank
DB00465	Ketorolac	P35354	Prostaglandin G/H synthase 2	1	1	0.9927	DrugBank,TTD
DB00500	Tolmetin	P35354	Prostaglandin G/H synthase 2	1	1	0.9912	DrugBank,TTD
DB00734	Risperidone	P35348	Alpha-1A adrenergic receptor	1	1	0.9892	DrugBank
DB00573	Fenoprofen	P35354	Prostaglandin G/H synthase 2	1	1	0.9888	DrugBank
DB00193	Tramadol	P31645	Sodium-dependent serotonin transporter	1	1	0.9855	DrugBank
DB00248	Cabergoline	P28222	5-hydroxytryptamine receptor 1B	1	1	0.9826	DrugBank
DB01076	Atorvastatin	Q01959	Sodium-dependent dopamine transporter	0	1	0.9825	N/A
DB01224	Quetiapine	P08912	Muscarinic acetylcholine receptor M5	1	1	0.9823	DrugBank

4 Conclusion

In this work, we propose MFCL-DTI, a novel framework for DTI prediction leveraging multiview feature fusion and contrastive learning. It employs multi-scale convolution for sequence feature extraction and captures topological and semantic information through neighbor and meta-path views. A multiview adaptive fusion module dynamically integrates cross-view features, while contrastive learning enhances representation. Experimental results demonstrate MFCL-DTI's superior performance, validating its effectiveness in DTI prediction.

Acknowledgement. This work was supported partly by National Natural Science Foundation of China (Grant Nos. 62271328, 62171312 and U22A2024), National Natural Science Foundation of Guangdong Province (No. 2024A1515011950), Shenzhen Science and Technology Program (Grant Nos. JCYJ20241202124202004 and JCYJ20220818095809021), and Shenzhen Medical Research Funds (Nos. C2401023 and C2301005).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, L., Tan, X., Wang, D., Zhong, F., Liu, X., Yang, T., Luo, X., Chen, K., Jiang, H., Zheng, M.: Transformercpi: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. Bioinformatics **36**(16), 4406–4414 (2020)
2. Chen, R., Liu, X., Jin, S., Lin, J., Liu, J.: Machine learning for drug-target interaction prediction. Molecules **23**(9), 2208 (2018)
3. Chen, X., Yan, C.C., Zhang, X., Zhang, X., Dai, F., Yin, J., Zhang, Y.: Drug–target interaction prediction: databases, web servers and computational models. Briefings in bioinformatics **17**(4), 696–712 (2016)
4. Dong, W., Yang, Q., Wang, J., Xu, L., Li, X., Luo, G., Gao, X.: Multi-modality attribute learning-based method for drug–protein interaction prediction based on deep neural network. Briefings in bioinformatics **24**(3), bbad161 (2023)
5. Haresamudram, H., Essa, I., Plötz, T.: Contrastive predictive coding for human activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **5**(2), 1–26 (2021)

6. Huang, K., Xiao, C., Glass, L.M., Sun, J.: Moltrans: molecular interaction transformer for drug–target interaction prediction. Bioinformatics **37**(6), 830–836 (2021)
7. Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B., et al.: Predicting new molecular targets for known drugs. Nature **462**(7270), 175–181 (2009)
8. Langley, G.R., Adcock, I.M., Busquet, F., Crofton, K.M., Csernok, E., Giese, C., Heinonen, T., Herrmann, K., Hofmann-Apitius, M., Landesmann, B., et al.: Towards a 21st-century roadmap for biomedical research and drug discovery: consensus report and recommendations. Drug discovery today **22**(2), 327–339 (2017)
9. Lee, I., Keum, J., Nam, H.: Deepconv-dti: Prediction of drug-target interactions via deep learning with convolution on protein sequences. PLoS computational biology **15**(6), e1007129 (2019)
10. 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 Transactions on Computational Biology and Bioinformatics **19**(2), 655–665 (2021)
11. Li, Y., Qiao, G., Gao, X., Wang, G.: Supervised graph co-contrastive learning for drug–target interaction prediction. Bioinformatics **38**(10), 2847–2854 (2022)
12. Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, J., Chen, L., Zeng, J.: A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature communications **8**(1), 573 (2017)
13. Peng, J., Wang, Y., Guan, J., Li, J., Han, R., Hao, J., Wei, Z., Shang, X.: An end-to-end heterogeneous graph representation learning-based framework for drug–target interaction prediction. Briefings in bioinformatics **22**(5), bbaa430 (2021)
14. 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: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 535–540. IEEE (2022)
15. Rogers, D., Hahn, M.: Extended-connectivity fingerprints. Journal of chemical information and modeling **50**(5), 742–754 (2010)
16. Velickovic, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. ICLR (poster) **2**(3), 4 (2019)
17. Wan, F., Hong, L., Xiao, A., Jiang, T., Zeng, J.: Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. Bioinformatics **35**(1), 104–111 (2019)
18. Wang, H., Huang, F., Xiong, Z., Zhang, W.: A heterogeneous network-based method with attentive meta-path extraction for predicting drug–target interactions. Briefings in Bioinformatics **23**(4), bbac184 (2022)
19. Yu, Z., Huang, F., Zhao, X., Xiao, W., Zhang, W.: Predicting drug–disease associations through layer attention graph convolutional network. Briefings in bioinformatics **22**(4), bbaa243 (2021)
20. Zhao, Q., Zhao, H., Zheng, K., Wang, J.: Hyperattentiondti: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. Bioinformatics **38**(3), 655–662 (2022)
21. Zhou, D., Xu, Z., Li, W., Xie, X., Peng, S.: Multidti: drug–target interaction prediction based on multi-modal representation learning to bridge the gap between new chemical entities and known heterogeneous network. Bioinformatics **37**(23), 4485–4492 (2021)