

RESEARCH

Open Access



CCL-DTI: contributing the contrastive loss in drug–target interaction prediction

Alireza Dehghan¹ , Karim Abbasi² , Parvin Razzaghi^{3*} , Hossein Banadkuki⁴ and Sajjad Gharaghani^{4*}

*Correspondence:
p.razzaghi@iasbs.ac.ir;
s.gharaghani@ut.ac.ir

¹ Department of Bioinformatics,
Kish International Campus,
University of Tehran,
Kish 1417614411, Iran

² Laboratory of System Biology,
Bioinformatics and Artificial
Intelligence in Medicine
(LBB&AI), Faculty of Mathematics
and Computer Science, Kharazmi
University, Tehran 1417614411,
Iran

³ Department of Computer
Science and Information
Technology, Institute
for Advanced Studies
in Basic Sciences (IASBS),
Zanjan 4513766731, Iran

⁴ Laboratory of Bioinformatics
and Drug Design (LBD), Institute
of Biochemistry and Biophysics,
University of Tehran,
Tehran 1417614411, Iran

Abstract

Background: The Drug–Target Interaction (DTI) prediction uses a drug molecule and a protein sequence as inputs to predict the binding affinity value. In recent years, deep learning-based models have gotten more attention. These methods have two modules: the feature extraction module and the task prediction module. In most deep learning-based approaches, a simple task prediction loss (i.e., categorical cross entropy for the classification task and mean squared error for the regression task) is used to learn the model. In machine learning, contrastive-based loss functions are developed to learn more discriminative feature space. In a deep learning-based model, extracting more discriminative feature space leads to performance improvement for the task prediction module.

Results: In this paper, we have used multimodal knowledge as input and proposed an attention-based fusion technique to combine this knowledge. Also, we investigate how utilizing contrastive loss function along the task prediction loss could help the approach to learn a more powerful model. Four contrastive loss functions are considered: (1) max-margin contrastive loss function, (2) triplet loss function, (3) Multi-class N-pair Loss Objective, and (4) NT-Xent loss function. The proposed model is evaluated using four well-known datasets: Wang et al. dataset, Luo’s dataset, Davis, and KIBA datasets.

Conclusions: Accordingly, after reviewing the state-of-the-art methods, we developed a multimodal feature extraction network by combining protein sequences and drug molecules, along with protein–protein interaction networks and drug–drug interaction networks. The results show it performs significantly better than the comparable state-of-the-art approaches.

Keywords: Drug discovery, Multimodal deep learning, Drug–target interaction, Contrastive loss function

Introduction

Drug–target interactions (DTI) prediction is vital to drug discovery, as it helps to identify potential interactions between drugs and targets [1–4]. In particular, DTI prediction focuses on identifying whether the specific proteins interact with a drug compound or not [5]. Additionally, it offers guidance on drug repurposing, multi-drug pharmacology,



drug resistance, and side effect prediction [6, 7]. The traditional biomedical measurement of DTI through in vitro experiments is considered reliable, but it is costly, time-consuming, and inefficient, particularly when dealing with large-scale datasets [8–11]. However, computational methods for DTI prediction have been receiving increased attention [12–14]. The current techniques for predicting DTI can be categorized into three distinct groups: ligand-based [15], docking-based [16], and machine learning-based approaches [10].

In recent years, DTI prediction has gotten more attention [17–19]. The introduced methods could be divided into two categories: feature-based methods and similarity-based methods. Zhang and Xie [20] introduced a DTI model based on non-negative matrix factorization. They introduced a new $L_{2,1}$ regularization term to guarantee the sparsity of the feature matrices derived through non-negative matrix factorization. They have proved that the obtained solution converges to the KKT point. Feature-based methods include two main modules: the feature extraction module and the task prediction module. In the feature extraction module, raw sequences of protein and drug molecules should be mapped to discriminative feature spaces. Ozturk et al. [21] introduced a DeepDTA model, which utilizes two 1D convolution networks to learn feature space for drugs and proteins. Then, the drug and protein feature vectors are concatenated to be fed into the task prediction model. Karimi et al. [22] introduced a semi-supervised method that first learns two sequence-to-sequence models to learn an initial representation of a drug-target pair. Then, it is used as an initializer for the RNN-CNN network as a feature extractor of the pair. Li et al. [13] introduced a co-contrastive learning-based method for DTI prediction to learn more discriminative representation for drug target pairs. To do so, they have utilized inhomogeneous graph representation. Qian et al. [23] introduced an approach using the drug chemical text information and the drug 2D structure image as input. Moreover, they have utilized a bi-directional multi-head cross-attentional module to encode drug and target interaction features. Zhang et al. [24] have used a transformer based model containing graph-based layers to extract features from drug molecules and a convolutional network to extract features from protein sequences. Yazdani-Jahromi et al. [25] introduced a method called AttentionSiteDTI. They treat the drug-target complex as a sentence to identify the effective protein binding sites that contribute to the drug-target interaction. In the task prediction module, the goal is to take the feature descriptor of the drug-target pair to produce the task label as output. Many approaches use a simple multilayer perceptron as a task prediction network. Tayebi et al. [26] introduced UnbiasedDTI, which focuses on the imbalance issue of the active/inactive classes in DTI. They have introduced an ensemble of deep-learning models to cope with this issue. He et al. [27] extract cross-view knowledge, including the sequence and network views for drugs and targets. They have utilized contrastive loss to learn better feature vectors for drugs and targets. To do so, they have defined auxiliary contrastive losses, including (1) contrasting similar and dissimilar drug feature vectors in sequence view, (2) contrasting similar and dissimilar drug feature vectors in network view, (3) contrasting similar and dissimilar target feature vectors in sequence view and (4) contrasting similar and dissimilar target feature vector

in network views. Li et al. [13] introduced a new Supervised Graph Co-contrastive Learning for Drug–Target Interaction Prediction called SGCL-DTI. They have defined two graphs: topological graph and semantic graph where in these graphs, nodes are the drug–target pairs. Then, supervised contrastive loss is defined over these feature representations. Zhnag et al. [28] introduced a new method in DTI called MRB-DTA. They have introduced a modified version of the transformer encoder with skip connections. Also, they have introduced an effective approach to better encode the knowledge of the interaction site between drug and protein. In [29], a graph convolutional network (GCN) extracts features from proteins and drugs. To do so, they have extracted protein 2D graph by using protein contact matrix and its physicochemical properties of residues. To extract the intra-molecular interactions, they have utilized cross-attention layers. Then, inter- and intra-molecular features are fused to feed into the MLP network.

In this paper, the research question is, "How do the different contrastive loss functions impact the drug target interaction prediction model's performance?". To investigate this research question, we present a new approach with two stages: (1) the first stage considers architecture to extract appropriate features for proteins and drugs, and (2) the second stage, a combinational loss function that includes task prediction loss and contrastive loss. For the feature extractor network, the first stage, we have utilized multimodal knowledge as input, including the drug molecule, protein sequences, protein–protein interaction networks, and drug–drug interaction networks. To extract features from the protein–protein interaction graph and drug–drug interaction graph, we have used the Node2vec network. To extract features from protein sequences and drug molecules, the 1D-convolution neural networks are used. We have used the two-sided attention mechanism to fuse the knowledge of these different modalities. Finally, the outputs of these networks are concatenated and fed into a multi-layered perceptron (MLP) to predict the affinity value. To recap, this comprehensive approach allows for a more complete understanding of the complex relationship between drugs and their targets, potentially leading to more accurate predictions. To investigate the effect of different contrastive loss functions, we have considered four important contrastive loss functions: (1) triplet loss function, (2) max-margin contrastive loss function, (3) Multi-class N-pair Loss Objective, and (4) NT-Xent loss function. The overall architecture of the proposed model is shown in Fig. 1. In the proposed approach, we have two loss functions to train the model: (1) the task prediction loss and (2) the contrastive loss function. In the training step, the model is first trained by the contrastive loss function, and then we train the model based on the prediction loss function. Next, this procedure is repeated until convergence is happened. It should be noted that providing data for contrastive loss functions is important. Each input data includes two drug–target pairs in the max-margin contrastive loss function. In the triplet loss function, we need three drug–target pairs, including anchor, positive, and negative. In Multi-class N-pair Loss Objective and NT-Xent loss functions, each input sample contains N drug–target pairs.

We evaluated the proposed approach on four well-known datasets: Wang et al. [30], Luo's dataset [31], KIBA [32], and Davis [33]. The results show significant improvements compared to state-of-the-art approaches and the base approach. It confirms that

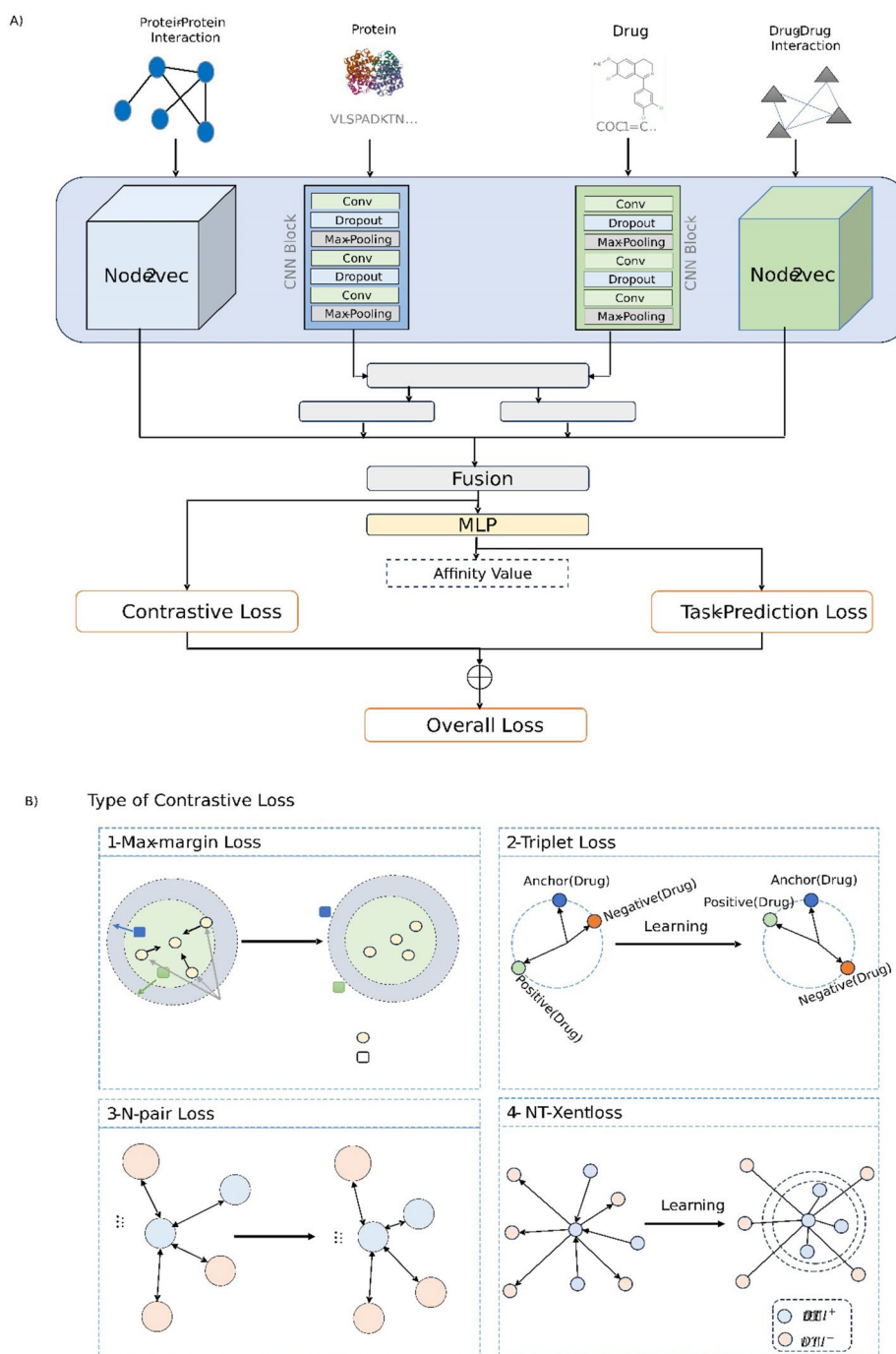


Fig. 1 An overview of the proposed approach

learning the discriminative feature space of the drug-target pair helps the task prediction model to predict the affinity value accurately.

To recap, the contributions of this paper are as follows:

1. We have utilized a multimodal feature extractor network. It means that the proposed method leverages various sources of information beyond considering the drug mole-

cule and protein sequences. Specifically, it takes into account the drug–drug interaction network and protein–protein interaction network, providing a broader perspective on the interplay between drugs and their targets.

2. We have proposed an attention-based fusion technique to combine the knowledge of the different modalities. To do so, we have utilized a two-sided attention mechanism.
3. We have used four powerful contrastive loss functions along the task prediction loss to learn more discriminative feature space.
4. We have done huge experiments to compare the contrastive loss functions in learning more discriminative feature space.
5. The results confirm the effectiveness of using contrastive loss functions along the task prediction loss function.

This paper is organized as follows: first, the problem formulation is given, and then, the proposed method is explained in detail. Next, the evaluations of the method's performance are presented. Finally, the paper highlights its effectiveness and suggests potential areas for further research.

Proposed method

In this section, we have given the details of the proposed method. The main contributions of this paper are to (1) fuse the multimodal knowledge using the attention-based module and (2) evaluate how different contrastive loss functions could impact drug–target interaction prediction. To do so, first, we have given the problem formulation. Next, the model's architecture is given, and finally, we have defined the different contrastive loss functions.

Problem formulation

Given $\{(d^{(i)}, p^{(i)}), l^{(i)}\}$ where $(d^{(i)}, p^{(i)})$ is a drug–target pair and $l^{(i)}$ is its corresponding affinity value or activity label (active or inactive). A drug, $d^{(i)}$, is shown by Simplified Molecular Input Line Entry System (SMILES) sequence, and i th protein is shown by amino-acid sequence. SMILES is a language to translate a three-dimensional chemical molecule into a string of symbols. The main goal is to design a system that takes the drug–target pair as input and predicts affinity value as output.

Model architecture

The architecture of the proposed approach is presented in this section. It consists of three subnetworks: protein feature encoder, drug feature encoder, and affinity value predictor (as task predictor). This paper uses the protein–protein interaction, drug–drug interaction networks, and protein sequence and drug molecule as input. PPI network is fed into the node2vec to extract feature vectors, and the same procedure is done for DDI. For extracting features from drug molecules and protein sequences, two 1D CNN networks are utilized. To combine the knowledge of the drugs and proteins, we have utilized the attention mechanism. In this case, we have utilized a two-sided attention mechanism. First, the drug features are considered as a query, and protein features are considered as key and value. Conceptually, it weights each local substructure of the protein sequence contributing to the drug features. Then, the protein features are

considered as a query, and drug features are considered as key and value. It determines the contribution of each local substructure of the drug molecule in updating protein features. Finally, the drug molecule features, drug–drug interaction graph features, protein sequence features, and protein–protein interaction graph features are concatenated and fed into the task prediction network. The task prediction network is a multilayer perceptron. A schematic view of the model architecture is shown in Fig. 1. In the following, the whole feature encoder is shown by N_E . To recap, the network N_E , takes drug SMILES, protein sequence, PPI, and DDI as input and returns the feature descriptor as output.

Contrastive loss function

In this section, the different types of loss functions are introduced and defined. In metric learning, metrics are learned to measure the similarity or dissimilarity between objects. Contrastive loss functions were introduced specifically for metric learning, aiming to optimize the parameters of these functions using deep neural networks. The resulting model can capture complex relationships between features and generate high-quality representations by embedding data points into a lower-dimensional space through deep neural networks. Ultimately, the objective is to create a model that renders a pair of examples with the same label more similar than a pair of examples with different labels. In this paper, four types of contrastive loss functions are used as auxiliary loss functions to learn a better model, and finally, in the experimental section, we evaluate these loss functions and explain how they perform.

Max-margin contrastive loss

The max-margin contrastive loss function was initially introduced by Hadsell et al. [34]. This loss function aims to maximize the distance between the pair of samples that belong to different classes. The max-margin contrastive loss function is defined as follows:

$$\mathcal{L}_{\text{max-margin}}(z_i, z_j) = 1_{l^{(i)}=l^{(j)}} \|z_i - z_j\|_2^2 + 1_{l^{(i)} \neq l^{(j)}} \max\left(0, m - \|z_i - z_j\|_2\right) \quad (1)$$

where z_i denotes the output of the feature encoder network for the i th sample $z_i = N_E(d^{(i)}, p^{(i)})$. This loss function for samples with similar labels minimizes the Euclidean distance between their corresponding feature vector. The Euclidean distance between the dissimilar samples (with different class labels) should be greater than the predefined margin threshold m .

Triplet loss function

The triplet loss function was first introduced by Weinberger [35], then it was used as a loss function by Facenet to train the deep neural network [36]. This loss function operates on triplets. Given $\{(d, p), (d, p)^+, (d, p)^-\}$ as a triplet include an anchor sample shown by (d, p) , positive sample shown by $(d, p)^+$ which has a same class label with anchor sample, and negative sample shown by $(d, p)^-$ which has a different class label with an anchor sample. This loss function is defined as follows:

$$\mathcal{L}_{\text{Triplet}}(z_i, z_j, z_k) = \max\left(0, \|z_i - z_j\|_2^2 - \|z_i - z_k\|_2^2 + m\right) \quad (2)$$

where m shows the margin, this loss function aims to minimize the distance between the feature embedding of the anchor and positive samples and maximize the distance between the anchor and negative samples.

One of the most important disadvantages of the triplet loss function is that only one negative example in each sample is considered, and the relation of that negative example with other negative samples (especially from different negative classes) is not considered. This problem leads to slow convergence for the triplet loss function.

Multi-class N-pair loss objective

This loss function is introduced by Sohn [37] for the first time. Given $\{(d, p), (d, p)^+, (d, p)^{-1}, (d, p)^{-2}, \dots, (d, p)^{-N-1}\}$ as $(N+1)$ -tuple of the training samples where (d, p) is the anchor sample. Also, $(d, p)^+$ denotes the positive samples to (d, p) and $(d, p)^{-i}$ shows i th negative sample to (d, p) . Hence, the N-pair loss function is defined as follows:

$$\begin{aligned} \mathcal{L}_{NT-Xent} & \left(\{(d, p), (d, p)^+, (d, p)^{-1}, (d, p)^{-2}, \dots, (d, p)^{-N-1}\} \right) \\ & = \log \left(1 + \sum_{k=1}^{N-1} \exp \left(z^T z_k - z^T z^+ \right) \right) \end{aligned} \quad (3)$$

where z and z^+ denotes the output of the feature encoder network for anchor and positive sample. Also, z_k denotes the output of the feature encoder network for k th negative sample. As it is clear, it is the generalized version of the triplet loss function, which considers more than one negative example. It is shown that when N is set to two, it is identical to the triple loss function. One of the most important disadvantages of minimizing Eq. (3) loss function is that generating a batch is expensive. For each batch sample, we need $(N+1)$ -tuple. Sohn [37] considered this issue by introducing a new approach to generating batches.

NT-Xent loss function

NT-Xent was first introduced by Chen et al. [38] for normalized temperature-scaled cross-entropy loss. This loss function is similar to multi-class N-pair loss with the difference that a new variable called temperature is introduced to consider the scale of the similarity values. Chen et al. [38] introduced the NT-Xent loss function for semi-supervised learning. Khosla et al. [39] modified this loss function for a supervised setting, which is defined as follows:

$$\mathcal{L}_{NT-Xent-supervised}(z_i) = \frac{-1}{2N_{l(i)} - 1} \sum_{j \in P(i)} \log \frac{\exp(z_i z_j / \tau)}{\sum_{k \in A(i)} \exp(z_i z_k / \tau)} \quad (4)$$

where τ denotes the temperature parameter, one of the most important findings about the temperature is that it could help the approach to learn a better model from hard samples. Chen et al. [38] showed that the value of the temperature parameter is dependent on batch sizes and the number of training epochs. Also, $A(i)$ shows all samples in the batch distinct from i , and $P(i)$ is the set of all samples in the batch that they have the same label with i th sample.

The proposed approach uses these contrastive loss functions along the task-specific loss function to learn a better model. In other words, the overall loss function of the proposed model is defined as follows:

$$\mathcal{L}_{overall} = \mathcal{L}_{contrastive} + \mathcal{L}_{taskprediction} \quad (5)$$

where $\mathcal{L}_{contrastive}$ is one of four introduced contrastive loss functions and $\mathcal{L}_{taskprediction}$ is the task-specific loss function. If the affinity value is continuous, the task-specific loss function is the mean-squared error, and if it is discrete, the task-specific loss function is the categorical cross-entropy. It should be noted that all introduced contrastive loss functions are supervised, and they utilize the corresponding discrete class labels. Hence, we need to convert the continuous labels to discrete ones for the regression task to use in contrastive loss functions.

Experiments

In this section, the experimental results are given. Four well-known datasets are used to evaluate the proposed method: Wang et al. [30], Luo's dataset [31], KIBA [32], and Davis[33]. In the following, we first introduce datasets; next, the experimental setting is explained. After that, evaluation metrics are introduced, and finally, the obtained results are given and analyzed.

Datasets

Wang et al. dataset: there are six heterogeneous networks included in Wang et al. [30]: (1) drug–drug interactions network, (2) protein–protein interactions network, (3) drug–protein interaction network, (4) drug–disease associations, (5) protein–disease associations, and (6) drug side effects associations. The drug–target interaction network contains 1923 edges extracted from Drugbank Version 3.0 [40–43]. In this paper, we have used only the drug–drug interactions network, protein–protein interactions network, and drug–protein interaction network.

KIBA dataset: it is a well-known DTI dataset containing 117,657 interaction pairs. These pairs are from 2,068 unique drugs and 229 unique target proteins. The affinity value for each pair is measured by the KIBA score, which is an integration of IC_{50} , $K_{(i)}$, and $K_{(d)}$ scores [44]. KIBA is a large dataset, and there are many varieties in the unique number of drugs and proteins. For the KIBA dataset, similar to [44], the threshold value is set to 12.1 and it is used to convert the predicted continuous values into binary values.

Davis dataset: it is another well-known DTI dataset containing 25,772 interaction pairs. These pairs are from 68 unique drugs and 442 unique target proteins. In this dataset, the binding affinity is measured by k_d value. To have a more stable learned model, the k_d value should be transformed into the log space as follows:

$$pK_d = -\log_{10} \left(\frac{K_d}{10^9} \right) \quad (6)$$

This study also converts the predicted continuous values into binary values by applying thresholds. Similar to [44], the selected threshold for Davis is set to 7.

Luo Dataset: This dataset is a heterogeneous graph [31] in which there are four different types of nodes: proteins (1512 nodes), drugs (708 nodes), side-effects (4192 nodes), and diseases (5603 nodes). Also, there are eight types of edges (i.e., interaction), including drug–protein interaction (1923 edges), protein–protein interaction (7363 edges), drug–drug interaction (10,036 edges), drug–disease interaction (199,214 edges), drug–side effect interaction (80,164 nodes) and protein–disease interaction (1,596,745 edges).

Evaluation metrics

We must select important evaluation metrics for regression and classification tasks in the proposed approach. In the regression task, we choose two metrics to evaluate the performance: (1) The Concordance Index (CI) measures the degree of ranking agreement between the predicted and ground truth values. (2) The R^2 measure provides insight into the percentage of the dependent variable variance that the model can explain. For the classification task, we have considered five evaluation measures: (1) Recall, which measures the ratio of positive samples that are correctly classified from all positive samples; (2) Precision, which considers how good the classifier is at avoiding false alarms.; (3) Accuracy measures the ratio of correctly classified samples; (4) Area under the ROC curve (AUC-ROC), and (5) Area under the precision-recall curve (AUC-PR).

Results

This section presents the results obtained on four datasets. First, the results of the ablation study by Wang et al. are shown in Fig. 2. The ablation study evaluates six versions of the proposed method: (1) v1: the network is trained without attention-based fusion and contrastive loss functions. In this case, a simple concatenation is used to

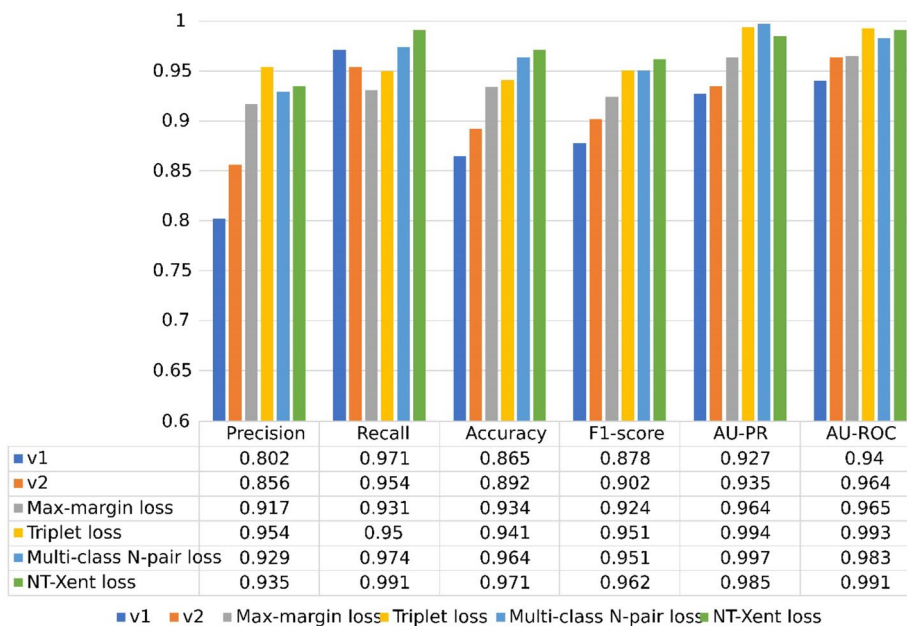


Fig. 2 The ablation study on the Wang et al. dataset

fuse the multimodal knowledge. (2) v2: in this case, the architecture is the same as the proposed model, and the contrastive loss is not used. In the following models, the architecture is the same as the proposed architecture, and the effect of the different contrastive loss functions is evaluated. (3) Triplet loss: The overall loss function is equal to the sum of the task prediction loss and the triplet loss; (4) Max-margin loss: The loss function for this case is the sum of task prediction loss and max-margin loss; (5) Multi-class N-pair loss: the overall loss function is the sum of task prediction loss and Multi-class N-pair loss, and (6) NT-Xent loss: the overall loss function sums the task prediction loss and the NT-Xent loss. As is shown in the proposed approach, the contrastive loss function is set to one of the four mentioned losses, and the obtained results are reported.

A comparison of the proposed method with state-of-the-art methods is shown in Fig. 3. Our approach is compared to five state-of-the-art approaches, including Multi-DTI [45], DTINet [31], NeoDTI [46], HNM [30], and TripletMultiDTI [3]. As shown in four metrics, the proposed method performs better than the other comparable approaches. It confirms that utilizing an appropriate contrastive loss function along the task prediction loss helps the model learn more discriminative feature space, leading to increased performance.

The obtained results on Luo’s dataset are given in Table 1. As shown in accuracy and AUROC, the proposed method performs better than the other approaches. Also, our approach achieves a comparable performance in other metrics compared to the best state-of-the-art approaches. It should be noted that MOVE utilizes a contrastive loss function [47], too, and our approach could improve three out of six measures over this approach.

Table 2 shows the results of the proposed method in the Davis dataset. For the Davis and KIBA datasets, we have compared the proposed method with the following

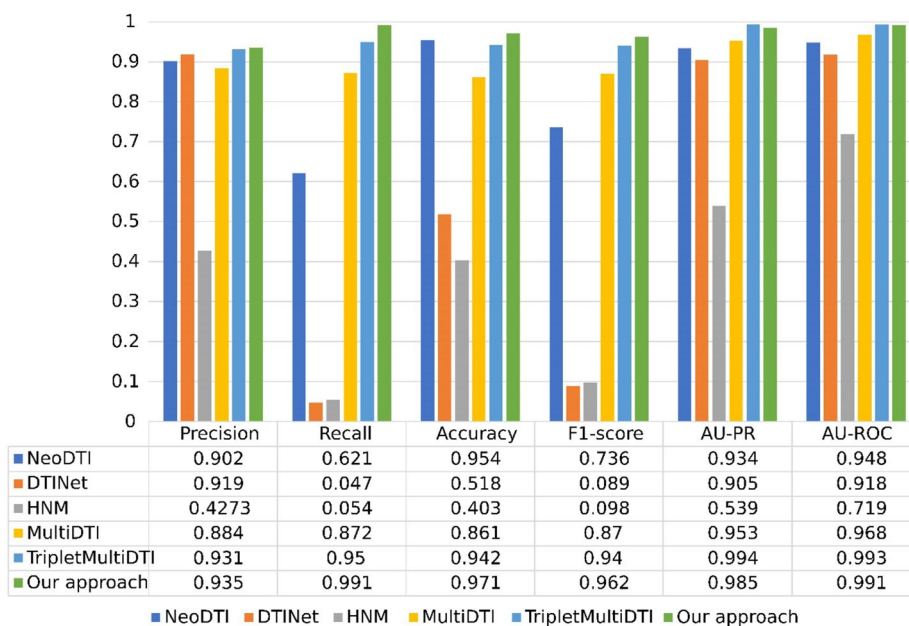


Fig. 3 Results obtained from Wang et al. dataset and comparisons with state-of-the-art approaches

Table 1 Luo's dataset results and comparison with state-of-the-art methods

| | Acc | Precision | Recall | F1-measure | AUROC | AUPR |
|-------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| <i>State-of-the-art</i> | | | | | | |
| DTINet | 0.524 | 1.000 | 0.048 | 0.090 | 0.914 | 0.932 |
| MultiDTI | 0.536 | 0.649 | 0.538 | 0.417 | 0.822 | 0.842 |
| NeoDTI | 0.877 | 0.884 | 0.868 | 0.876 | 0.944 | 0.952 |
| SGCL-DTI | – | – | – | – | 0.977 | 0.977 |
| MOVE | 0.876 | 0.858 | 0.904 | 0.876 | 0.950 | 0.943 |
| <i>Our approach</i> | | | | | | |
| Task prediction loss | 0.728 | 0.994 | 0.597 | 0.746 | 0.854 | 0.861 |
| Triplet loss | 0.794 | 0.964 | 0.634 | 0.765 | 0.919 | 0.921 |
| Max-margin loss | 0.741 | 0.999 | 0.619 | 0.764 | 0.897 | 0.918 |
| Multi-class N-pair loss | 0.859 | 0.814 | 0.869 | 0.841 | 0.967 | 0.958 |
| NT-Xent loss | 0.878 | 0.847 | 0.873 | 0.860 | 0.978 | 0.964 |

The bold ones mean that it is the best performance of that measure among the comparable methods

Table 2 Davis dataset results and comparison with state-of-the-art methods

| | Approaches | CI measure | AU-PR | p value |
|------------------|-------------------------|----------------------|----------------------|-------------------------|
| State-of-the-art | KronRLS | 0.871 ± 0.0008 | 0.661 ± 0.010 | 2.91 × 10 ⁻¹ |
| | SimBoost | 0.872 ± 0.002 | 0.709 ± 0.008 | 2.41 × 10 ⁻¹ |
| | DeepDTA | 0.878 ± 0.004 | 0.714 ± 0.010 | 2.54 × 10 ⁻¹ |
| | DeepCDA | 0.891 ± 0.003 | 0.739 ± 0.006 | 2.65 × 10 ⁻¹ |
| | GraphDTA | 0.893 | – | – |
| | NerLTR-DTA | 0.936 | – | – |
| | SimCNN-DTA | 0.855 ± 0.0027 | 0.657 ± 0.0076 | 2.54 × 10 ⁻¹ |
| | TripletMultiDTI | 0.901 ± 0.001 | 0.902 ± 0.007 | 8.06 × 10 ⁻¹ |
| Our approach | Task prediction loss | 0.874 ± 0.003 | 0.821 ± 0.004 | 6.39 × 10 ⁻² |
| | Triplet loss | 0.901 ± 0.001 | 0.902 ± 0.007 | 8.06 × 10 ⁻¹ |
| | Max-margin loss | 0.901 ± 0.004 | 0.842 ± 0.007 | 5.48 × 10 ⁻² |
| | Multi-class N-pair loss | 0.928 ± 0.006 | 0.873 ± 0.004 | 2.83 × 10 ⁻¹ |
| | NT-Xent loss | 0.945 ± 0.005 | 0.879 ± 0.004 | |

The bold ones mean that it is the best performance of that measure among the comparable methods

approaches: KronRLS [48], SimBoost [44], DeepDTA [21], DeepCDA [1], SimCNN-DTA [49], GraphDTA [50], NerLTR-DTA [51], and TripletMultiDTI [3]. As shown, the obtained results are reported for the four different contrastive loss functions and a model with only task prediction loss. The obtained results are significantly better than TripletMultiDTI when the NT-Xent loss function is used as a contrastive loss function [3]. To statistically evaluate the proposed method, we have used the paired t-test. In this test, the null hypothesis states that there is no significant difference between the proposed approach and the comparing methods. Based on the reported p-values in Table 2, we reject the null hypothesis with a p-value lower than 30% for all state-of-the-art methods except the TripletMultiDTI approach.

Table 3 shows the results obtained by applying the proposed method to the KIBA dataset. As presented, the proposed method effectively increases the performance with respect to the comparable approaches. It should be noted that the task is a regression

Table 3 KIBA dataset results and comparison with state-of-the-art methods

| | Approaches | CI measure | AU-PR | p value |
|------------------|-------------------------|----------------------|----------------------|-------------------------|
| State-of-the-art | KronRLS | 0.782 ± 0.0009 | 0.635 ± 0.004 | 1.82 × 10 ⁻¹ |
| | SimBoost | 0.836 ± 0.001 | 0.760 ± 0.003 | 1.34 × 10 ⁻¹ |
| | DeepDTA | 0.863 ± 0.002 | 0.788 ± 0.004 | 1.72 × 10 ⁻¹ |
| | DeepCDA | 0.889 ± 0.002 | 0.812 ± 0.005 | 2.48 × 10 ⁻¹ |
| | GraphDTA | 0.891 | – | |
| | NerLTR-DTA | 0.893 | – | |
| | SimCNN-DTA | 0.821 ± 0.0011 | 0.721 ± 0.0018 | 1.61 × 10 ⁻¹ |
| Our approach | TripletMultiDTI | 0.895 ± 0.003 | 0.839 ± 0.007 | 2.00 × 10 ⁻¹ |
| | Task prediction loss | 0.882 ± 0.005 | 0.798 ± 0.002 | 2.42 × 10 ⁻¹ |
| | Triplet loss | 0.895 ± 0.003 | 0.839 ± 0.007 | 2.00 × 10 ⁻¹ |
| | Max-margin loss | 0.901 ± 0.010 | 0.870 ± 0.009 | 3.89 × 10 ⁻² |
| | Multi-class N-pair loss | 0.921 ± 0.006 | 0.881 ± 0.004 | 3.74 × 10 ⁻¹ |
| | NT-Xent loss | 0.924 ± 0.003 | 0.896 ± 0.001 | |

The bold ones mean that it is the best performance of that measure among the comparable methods

task for the Davis and KIBA datasets. It means that the model predicts a continuous value. This leads us to utilize CI measures for both of these datasets. Also, we have converted the continuous affinity value to a binary label by thresholding like [1, 21, 44]. The CI measure and AUPR are increased by 2.9% and 5.6% over the best state-of-the-art method. In other words, it means the model learns a strong model by utilizing both the appropriate contrastive loss function and the prediction loss function. Based on the reported p-values in Table 3, we reject the null hypothesis with a p-value lower than 20% for most state-of-the-art methods.

Conclusion

This paper focuses on this research question: "How contrastive loss function along the task prediction loss could help the approach to learn a more discriminative model?". We have selected four important contrastive loss functions to investigate and used them as auxiliary loss functions. However, we believe that a feature extraction network may be beneficial in learning a strong model. Accordingly, after reviewing the state-of-the-art methods, we developed a multimodal feature extraction network by combining protein sequences and drug molecules, along with protein–protein interaction networks and drug–drug interaction networks. To fuse the multimodal knowledge, we have proposed to use an attention-based fusion technique.

One of the advantages of the proposed method, which leads to performance improvement, is that it utilizes a powerful loss function. The loss function guides the optimization process during the backpropagation. Hence, using powerful loss functions leads to an improvement in the performance and the generalization capabilities of trained models. The loss function in most DTI approaches is based on the error between the predicted outputs and the ground truth labels without considering the representation vector of the drug–target pair. As a result of this work, we introduce a novel loss function that combines the task prediction loss with a contrastive loss function.

To evaluate the proposed method, it is applied to four well-known datasets: Wang et al., Luo's dataset, Davis, and KIBA datasets. A huge experiment is done to show the effectiveness of the proposed method. Based on the results obtained, the proposed method could improve the performance.

One of the limitations of the proposed method is the computational complexity. In Multi-class N-pair Loss and NT-Xent loss functions, each batch sample needs $(N + 1)$ -tuple, which is practically intractable. Although, we have utilized an introduced approach by Sohn [32] to generate batches. Still, it needs more computing power. The other limitation is finding the best strategy to generate batches. In future work, providing more informative batches for DTI will be considered.

In recent years, ncRNAs are recognized as a new class of drug targets due to its effectiveness evidence in gene expression and disease progression [52, 53]. In future work, by providing protein-disease and ncRNA-disease graphs as additional inputs, we can modify the approach to predict small molecule-ncRNA associations.

Abbreviations

| | |
|---------|--|
| DTI | Drug–target interactions |
| PPI | Protein–protein interaction |
| DDI | Drug–drug interaction |
| CNN | Convolutional neural network |
| SMILES | Simplified molecular-input line-entry system |
| CI | Concordance Index |
| AUC-PR | Area under the precision-recall curve |
| AUC-ROC | Area under ROC curve |

Acknowledgements

We thank the anonymous reviewers for their constructive comments on the original manuscript.

Author contributions

AD: Result analysis, conceptualization, methodology, data curation, writing, review and editing. KA: Methodology, formal analysis, conceptualization, writing original draft and editing. HB: Result analysis, programming, visualization and writing the initial draft. PR and SGH: Supervision, conceptualization, review, editing and project administration.

Funding

None.

Availability of data and materials

The sample data and code for CCL-DTI: <https://github.com/dehghan1401/CCL-DTI>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 13 September 2023 Accepted: 22 January 2024

Published online: 30 January 2024

References

1. Abbasi K, Razzaghi P, Poso A, Amanlou M, Ghasemi JB, Masoudi-Nejad A. DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks. *Bioinformatics*. 2020;36(17):4633–42.
2. 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:btad411.
3. Dehghan A, Razzaghi P, Abbasi K, Gharaghani S. TripletMultiDTI: multimodal representation learning in drug–target interaction prediction with triplet loss function. *Expert Syst Appl*. 2023;232:120754.

4. Zhang Y, Hu Y, Han N, Yang A, Liu X, Cai H. A survey of drug–target interaction and affinity prediction methods via graph neural networks. *Comput Biol Med.* 2023;163:107136.
5. Palhamkhani F, Alipour M, Dehnad A, Abbasi K, Razzaghi P, Ghasemi JB. DeepCompoundNet: enhancing compound–protein interaction prediction with multimodal convolutional neural networks. *J Biomol Struct Dyn.* 2023. <https://doi.org/10.1080/07391102.2023.2291829>.
6. Xue H, Li J, Xie H, Wang Y. Review of drug repositioning approaches and resources. *Int J Biol Sci.* 2018;14(10):1232.
7. Mongia A, Majumdar A. Drug–target interaction prediction using multi graph regularized nuclear norm minimization. *PLoS ONE.* 2020;15(1):e0226484.
8. Li F, Zhang Z, Guan J, Zhou S. Effective drug–target interaction prediction with mutual interaction neural network. *Bioinformatics.* 2022;38(14):3582–9.
9. Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, Zhang Y. Drug–target interaction prediction: databases, web servers and computational models. *Brief Bioinform.* 2016;17(4):696–712.
10. Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, Lu H. Deep-learning-based drug–target interaction prediction. *J Proteome Res.* 2017;16(4):1401–9.
11. Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics.* 2020;36(22–23):5545–7.
12. Hu L, Fu C, Ren Z, Cai Y, Yang J, Xu S, Xu W, Tang D. SSELM-neg: spherical search-based extreme learning machine for drug–target interaction prediction. *BMC Bioinform.* 2023;24(1):38.
13. Li Y, Qiao G, Gao X, Wang G. Supervised graph co-contrastive learning for drug–target interaction prediction. *Bioinformatics.* 2022;38(10):2847–54.
14. Tanoori B, Zolghadri Jahromi M, Mansoori EG. Binding affinity prediction for binary drug–target interactions using semi-supervised transfer learning. *J Comput Aided Mol Des.* 2021;35:883–900.
15. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol.* 2007;25(2):197–206.
16. Shaikh N, Sharma M, Garg P. An improved approach for predicting drug–target interaction: proteochemometrics to molecular docking. *Mol BioSyst.* 2016;12(3):1006–14.
17. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Brief Bioinform.* 2021;22(1):247–69.
18. Ezzat A, Wu M, Li XL, Kwok CK. Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinform.* 2019;20(4):1337–57.
19. Tanoori B, Jahromi MZ, Mansoori EG. Drug-target continuous binding affinity prediction using multiple sources of information. *Expert Syst Appl.* 2021;186:115810.
20. Zhang J, Xie M. Graph regularized non-negative matrix factorization with L_{2,1} norm regularization terms for drug–target interactions prediction. *BMC Bioinform.* 2023;24(1):375.
21. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics.* 2018;34(17):i821–9.
22. Karimi M, Wu D, Wang Z, Shen Y. DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics.* 2019;35(18):3329–38.
23. Qian Y, Li X, Wu J, Zhang Q. MCL-DTI: using drug multimodal information and bi-directional cross-attention learning method for predicting drug–target interaction. *BMC Bioinform.* 2023;24(1):323.
24. Zhang P, Wei Z, Che C, Jin B. DeepMGT-DTI: transformer network incorporating multilayer graph information for drug–target interaction prediction. *Comput Biol Med.* 2022;142:105214.
25. Yazdani-Jahromi M, Yousefi N, Tayebi A, Kolanthai E, Neal CJ, Seal S, Garibay OO. AttentionSiteDTI: an interpretable graph-based model for drug–target interaction prediction using NLP sentence-level relation classification. *Brief Bioinform.* 2022;23(4):bba272.
26. Tayebi A, Yousefi N, Yazdani-Jahromi M, Kolanthai E, Neal CJ, Seal S, Garibay OO. UnbiasedDTI: mitigating real-world bias of drug–target interaction prediction by using deep ensemble-balanced learning. *Molecules.* 2022;27(9):2980.
27. He C, Qu Y, Yin J, Zhao Z, Ma R, Duan L. Cross-view contrastive representation learning approach to predicting DTIs via integrating multi-source information. *Methods.* 2023;218:176–88.
28. Zhang L, Wang CC, Chen X. Predicting drug–target binding affinity through molecule representation block based on multi-head attention and skip connection. *Brief Bioinform.* 2022;23(6):bbac468.
29. Zhang L, Wang CC, Zhang Y, Chen X. GPCNDTA: prediction of drug-target binding affinity through cross-attention networks augmented with graph features and pharmacophores. *Comput Biol Med.* 2023;166:107512.
30. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics.* 2014;30(20):2923–30.
31. 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. *Nat Commun.* 2017;8(1):573.
32. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, Aittokallio T. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J Chem Inf Model.* 2014;54(3):735–43.
33. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol.* 2011;29(11):1046–51.
34. Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. *Comput Vis Pattern Recognit.* 2006;2:1735–42.
35. Weinberger KQ, Blitzer J, Saul LK. Distance metric learning for large margin nearest neighbor classification. In: *Advances in neural information processing systems*; 2006. p. 1473–80.
36. Schroff F, Kalenichenko D, Philbin J. Facenet: a unified embedding for face recognition and clustering. In: *IEEE conference on computer vision and pattern recognition*; 2015. p. 815–23.
37. Sohn K. Improved deep metric learning with multi-class n-pair loss objective. *Adv Neural Inf Process Syst.* 2016;29:1857–65.

38. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: International conference on machine learning; 2020. p. 1597–607.
39. Khosla P, Teterwak P, Wang C, Sarna A, Tian Y, Isola P, Maschinot A, Liu C, Krishnan D. Supervised contrastive learning. *Adv Neural Inf Process Syst*. 2020;33:18661–73.
40. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2010;39:D1035–41.
41. Keshava Prasad T, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A. Human protein reference database—2009 update. *Nucleic Acids Res*. 2009;37(suppl_1):D767–72.
42. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Rosenstein MC, Wiegiers TC. The comparative toxicogenomics database: update 2013. *Nucleic Acids Res*. 2013;41(D1):D1104–14.
43. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol*. 2010;6(1):343.
44. He T, Heidemeyer M, Ban F, Cherkasov A, Ester M. SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. *J Cheminformatics*. 2017;9(1):1–14.
45. 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*. 2021;37(23):4485–92.
46. 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*. 2019;35(1):104–11.
47. 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). IEEE; 2022. p. 535–40.
48. Pahikkala T, Airolo A, Pietila S, Shakyawar S, Szwajda A, Tang J, Aittokallio T. Toward more realistic drug–target interaction predictions. *Brief Bioinf*. 2014;16(2):325–37.
49. Shim J, Hong ZY, Sohn I, Hwang C. Prediction of drug–target binding affinity using similarity-based convolutional neural network. *Sci Rep*. 2021;11(1):1–9.
50. 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.
51. Ru X, Ye X, Sakurai T, Zou Q. NerLTR-DTA: drug-target binding affinity prediction based on neighbor relationship and learning to rank. *Bioinformatics*. 2022;38(7):1964–71.
52. Chen X, Guan NN, Sun YZ, Li JQ, Qu J. MicroRNA-small molecule association identification: from experimental results to computational models. *Brief Bioinform*. 2020;21(1):47–61.
53. Chen X, Zhou C, Wang CC, Zhao Y. Predicting potential small molecule–miRNA associations based on bounded nuclear norm regularization. *Brief Bioinform*. 2021;22(6):bbab328.

Publisher's Note

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