Peng *et al. BMC Bioinformatics* 2020, **21**(Suppl 13):394
https://doi.org/10.1186/s12859-020-03677-1

BMC Bioinformatics

RESEARCH Open Access

A learning-based method for drug-target interaction prediction based on feature representation learning and deep neural network

Jiajie Peng[1,2], Jingyi Li[1,2] and Xuequn Shang[1,2*]

From The 18th Asia Pacific Bioinformatics Conference
Seoul, Korea. 18-20 August 2020

*Correspondence:
shang@nwpu.edu.cn
[1]The School of Computer Science, Northwestern Polytechnical University, 710072 Xian, China
[2]The Key Laboratory of Big Data Storage an Management, Northwestern Polytechnical Universitythe, Ministry of Industry and Information Technology, 710072 Xian, China

Abstract

Background: Drug-target interaction prediction is of great significance for narrowing down the scope of candidate medications, and thus is a vital step in drug discovery. Because of the particularity of biochemical experiments, the development of new drugs is not only costly, but also time-consuming. Therefore, the computational prediction of drug target interactions has become an essential way in the process of drug discovery, aiming to greatly reducing the experimental cost and time.

Results: We propose a learning-based method based on feature representation learning and deep neural network named DTI-CNN to predict the drug-target interactions. We first extract the relevant features of drugs and proteins from heterogeneous networks by using the Jaccard similarity coefficient and restart random walk model. Then, we adopt a denoising autoencoder model to reduce the dimension and identify the essential features. Third, based on the features obtained from last step, we constructed a convolutional neural network model to predict the interaction between drugs and proteins . The evaluation results show that the average AUROC score and AUPR score of DTI-CNN were 0.9416 and 0.9499, which obtains better performance than the other three existing state-of-the-art methods.

Conclusions: All the experimental results show that the performance of DTI-CNN is better than that of the three existing methods and the proposed method is appropriately designed.

Keywords: DTIs prediction, Convolutional neural network, Feature representation learning

© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Drug targets are special molecules that can bind to drugs and produce effects in cells, the main molecular targets for drugs are proteins [1]. Drug-target interactions (DTIs) prediction is of great significance for drug repositioning [2], drug discovery [3], side-effect prediction [4] and drug resistance [5]. However, identifying the drug-target interactions via biochemical and chemical biological experiments is costly and time-consuming [6]. Recently, as genomic, chemical, and pharmacological data become more and more complete, new opportunities for identifying drug target interactions have been emerged [2]. Therefore, many researchers have attempted to predict DTIs by using silico or computational approaches to guide in vivo validation in recent years, and thus significantly reduce the cost and time for identifying the drug-target interactions [2].

The traditional computational DTIs prediction approaches are mainly categorized into docking-based approaches [7] [8] and ligand-based approaches [9]. However, the docking is difficult to play a good performance when the three-dimensional structures of the target protein are unknown [10]. The ligand-based approaches are very effective in DTIs prediction, but it often requires a large number of known binding data and thus the prediction results are poor with only a small amount of known data [11].

In recent years, network-based approaches have demonstrated great advantages compared to docking-based and ligand-based methods [12] [13]. First, network-based approaches have a good prediction performance even without the three-dimensional structure of the target. Secondly, they are simple and fast for only by performing simple physical processes [13]. In the past decade, DTIs prediction approaches based on machine learning have also been widely studied. A key idea of these approaches is the assumption that similar drugs may share similar targets [14].Thus predicting DTIs is often regarded as a binary classification issue by using chemical structures of drugs and targets as input features and considering known DTIs as labels [2]. However, most existing prediction methods are limited to homogeneous networks, which ignore a rich variety of topological information and the complex interaction relationship of heterogeneous data [3].

In recent years, a variety of computing methods based on heterogeneous data sources have been developed to predict DTI. *Wang et.al* used a heterogeneous network data to obtain the diffusion feature and directly use the obtained diffusion distributions to derive the prediction scores of DTIs [3]. However, the direct use of diffusion state as a feature or prediction score is vulnerable to the deviation caused by noise and high dimension of heterogeneous network data, resulting in inaccurate DTI prediction [2]. *Luo et.al* provided the DTINet as a new prediction method, which extracted the low-dimensional characteristic information from heterogeneous data sources and used the inductive matrix completion (IMC) approach to predict the drug-target interaction fraction [2]. The inductive matrix completion approach predicts the new DTI by using a small amount of known drug-protein interaction information, combined with the extracted drug and protein characteristic information. This method has been proved to be superior to the commonly used Laplacian regularization least square approach [15], heterogeneous network model [16], cooperation matrix factorization (CMF) [17] and Bipartite local model by learning from local information and neighbors [18].

In this paper, we improve the prediction method by learning low-dimensional vector representations of features from heterogeneous networks, and adopting convolution neural networks (CNN) as classification model. An efficient DTI prediction method

is presented, DTI-CNN, which can be used to identify the drug-target interactions, guide biochemical experiments and reduce the cost of research. Here are four major contributions:

- We propose a learning-based method for drug-target interaction prediction that contains three components, named as heterogeneous-network-based feature extractor, denoising-autoencoder-based feature selector and CNN-based interaction predictor.
- Based on random walk with restart (RWR) and denoising autoencoder (DAE) model, DTI-CNN can cope with the noisy, incomplete and high-dimensional features from heterogeneous data sources, including drug, proteins, side-effects and diseases information.
- Based on a deep CNN model, DTI-CNN can handle the low dimensional feature vectors and predict the probability of interaction between each pair of drugs and proteins.
- Based on our DTI prediction task, the results indicate that DTI-CNN is better than the other three state-of-the-art methods and is appropriately designed.

Methods

We propose an learning-based method called DTI-CNN to predict drug-target interactions. The workflow of DTI-CNN is shown in Fig. 1. First, the heterogeneous network was constructed by integrating a variety of drug and protein related information sources, and the initial drug feature vector and protein feature vector were obtained by RWR model.

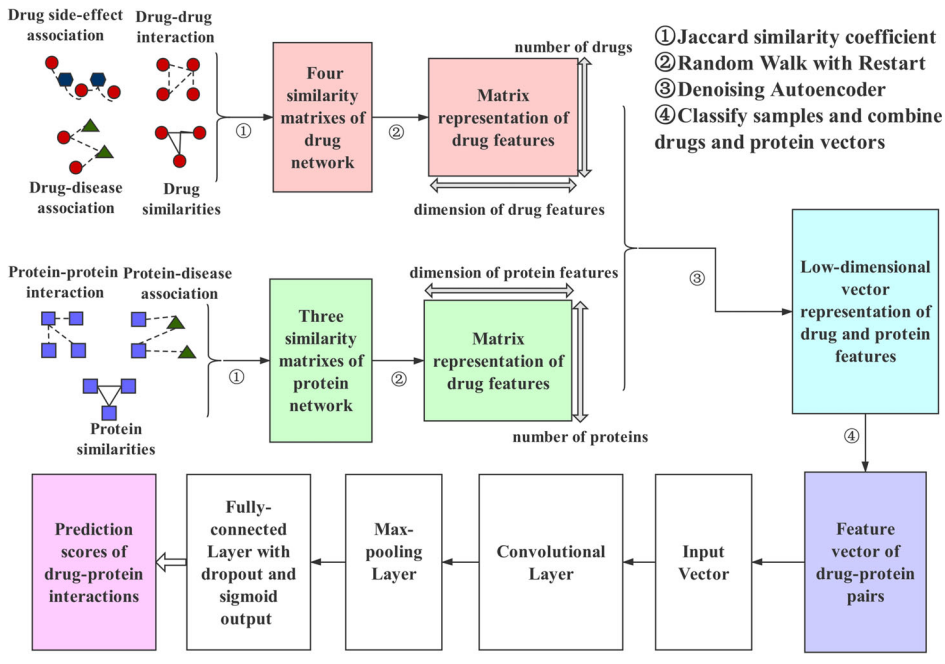

Fig. 1 The flowchart of DTI-CNN pipeline. The DTI-CNN contains heterogeneous-network-based feature extractor, denoising-autoencoder-based feature selector and CNN-based interaction predictor. First, the features are extracted from seven networks of drug and protein by the Jaccard similarity coefficient and RWR algorithm, then we get the low-dimensional representation of drug and protein features by adopting the DAE model. Third, a deep CNN model is constructed to predict the interaction of each pair of drugs and proteins

In the second step, the high-dimensional features of drugs and proteins are reduced by adopting the DAE model, and the low-dimensional representations of them are obtained respectively. Finally, according to the known drug-protein interactions, the samples are divided into positive samples and negative samples. Combining the feature vector of drug-protein pairs, CNN was adopted to predict the association between each pair of drugs and proteins.

Heterogeneous-network-based feature extractor

The heterogeneous networks is constructed based on two types of networks as follows. One is the drug-related networks, including drug-drug interactions, drug-disease associations, drug-side-effect associations, drug similarities (based on chemical structures of drugs) [2]. The other is protein-related networks, including protein-disease associations, protein-protein interactions, protein similarities (based on primary sequences of proteins) [2]. Firstly, the Jaccard similarity algorithm [19] is executed on each association and interaction matrix respectively, so we can generate a similarity matrix for each network.

Taking the drug-disease interactions matrix as an example, two sets A and B are given as two rows in the adjacent matrix, which represent the interactions between two different drugs and all diseases. The Jaccard similarity coefficient[20] is an indicator of the similarity between two sets, defined as follows:

$$Sim(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{1}$$

The original adjacent matrix is a description of the relationship between a single row and column node, and the Jaccard similarity coefficient calculation is based on the adjacent two row vectors of the original adjacent matrix. Thus the similarity matrices S represent the similarity between each drug or protein node and all features of the column nodes. The element $S_{i,j}$ represents similarity of row i and row j in the original adjacenct matrix.

After all the original data is transformed into similarity matrices, the RWR algorithm [21] is applied to each similarity matrix, which represents a weighted network. The diffusion state of each drug or protein is obtained on each network, which includes the topological structure relation of each drug or protein with all the other nodes in the network.

The reason for using RWR is that the similarity matrix obtained in the previous step only calculates the similarity of the two nodes in isolation. RWR can be used to consider global structure information in the network. If the distribution state of the two nodes is close, they can be considered to be in a similar position relative to other nodes in the network. According to the RWR principle, the greater the similarity between the two nodes, the higher the transition probability of them [22].

Taking the drug-disease similarity matrix $A_{i,j}$ as an example, we can get the drug-disease transition probability matrix B according to the $A_{i,j}$, whose elements $B_{i,j}$ describe the transition probability from the drug node i to the disease node j [23], that is defined as follows:

$$B_{i,j} = \frac{A_{i,j}}{\sum_j A_{i,j}}. \tag{2}$$

Next, the final drug-disease diffusion state matrix can be obtained by iterative convergence as follows:

$$S_i^{t+1} = (1 - p_r) S_i^t B + p_r e_i. \tag{3}$$

Where S_i^t is the result after t iterations, and each element stores the probability of accessing a disease node from the drug node i after iteration in the process of random walk, p_r is the restart probability, and e_i is an n-dimensional unit matrix.

After all the similarity matrices is transformed into diffusion state matrices, we splice the single diffusion state matrix of drug and protein networks, so that we can get two diffusion state matrices about drug and protein. The row of the drug diffusion matrix represents different drugs, and the column represents the four nodes of drugs, diseases, side effects and drugs, in which the element $D_(i,j)$ represents the transition probability between drug i and node j. The row of the protein diffusion state matrix represents different proteins, and the column represents the three nodes of protein, disease and protein, in which the element $P_(i,j)$ represents the transfer probability of protein i and node j.

Denoising-autoencoder-based feature selector

The vector of diffusion state matrix obtained in the previous step is high-dimensional, noisy and incomplete. In order to obtain the essential features, we apply a DAE model which carry on the data operation on the basis of the autoencoder. The main idea of DAE is shown in Fig. 2. Taking the diffusion state matrix about drug as an input example, by adding noise to the input training data and making the self-encoder learn to remove this noise, the real input which has not been polluted by noise can be obtained [24]. Therefore, the encoder can obtain the most essential features from the original input to get more robust representation. This is why its generalization ability is better than that of the general encoder [25]. Autoencoders use automatic Encoders to obtain low-dimensional data through neural networks based on the input data. Similarly, the Decoders to recover the original input from low-dimensional data [26].

In the model, we reduce the dimension of drug features to 100 dimensions and protein features to 400 dimensions. We set the noise figure to 0.2, and use the softplus [27] and RMSProp fuction [28] to optimize the mean-square error (MSE) [29]. At last, the backpropagation (BP) algorithm is used to train our DAE [30].

CNN-based interaction predictor

Convolutional Neural Networks is a classical and widely used structure since 1980s [31] and can greatly reduces the complexity of convolution neural network [32]. The parameters of each layer network are shared and the number of parameters to be trained

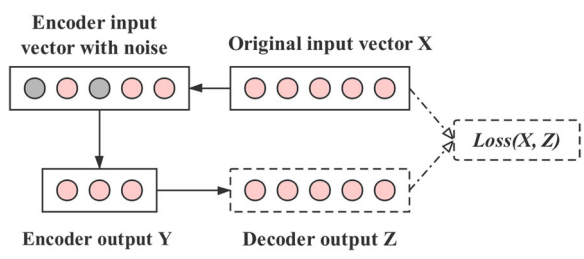

Fig. 2 The schematic illustration of denoising autoencoder. The original input data is high-dimensional, noisy and incomplete, the DAE adds noise to it and makes the self-encoder learn to remove the noise, which makes the encoder learn more robust and low-dimensional representation in the input data. Then the decoder is used to recover the original input from low-dimensional data, the loss between the original input and the decoder output is optimized by the RMSProp algorithm

is reduced during training. Compared with the standard fully connected neural network, it has better performance in image classification, sentence classification and other classification tasks [33].

Inspired by the success of CNN in classification tasks [34], we use CNN as the supervised learning model and the structure of the prediction model is shown in Fig. 3. The prediction model contains the convolution, max-pooling, fully-connected and output layer. A convolution layer with a rectified linear unit (ReLU) activation fuction [34] is used as a feature extractor [35]. A max-pooling layer is employed to reduce the dimension of features, and the fully-connection layer and the output sigmoid layer are used to classify the tasks.

As the key component of a CNN, convolution layer can help the model to learn local and global structures from the input vector [35]. In our model, the convolutional layer consists of 4 kernels. Given the input vector X of length S, and the length of weight vector is $4 * 1$. For each kernel, convolution operation is independent and thus we can obtain four particular feature with the length $(S - 4) + 1$, which was extracted from the input vector and named feature map M. The M is obtained as follows:

$$M_i = \sum_{j=1}^{4} W_j X_{i+j}. \tag{4}$$

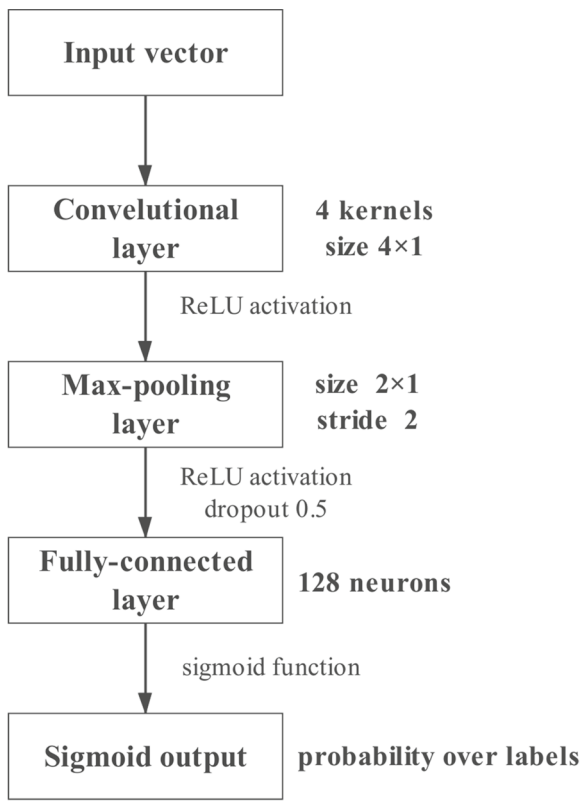

Fig. 3 The structure of the convolutional neural network model

where $(i \in 0, 1, \ldots, S-4)$, and W is initialized by a truncated normal distribution and used as a weight vector. Then a ReLU function is used to the feature map obtained last step:

$$f(x) = max(0, x). \tag{5}$$

The ReLU function is selected for the excitation function, which can effectively simplify the calculation process and avoid the gradient explosion and disappearance [36].

Next layer is the max-pooling layer, which can extract the maximum value in the pooled region and the pooled region continues to move forward at a certain step size in an input sequence, thus reducing the dimension in each feature map [36]. In our model, the pooled size is $2 * 1$, and the step size is 2. Given an input vector A_i $(i \in 0, 1, \ldots, S-4)$, the length of output of this layer is $\frac{(S-2)+1}{2}$.

After the first two layers, we use a one-dimensional vector to connect the important features extracted from all the kernels and then pass them to the fully connected layer. The number of hidden units in this layer is 180 and the output of this layer is calculated as follows:

$$f = W * y. \tag{6}$$

where $W \in R^{(n*180)}$ is the weight matrix, y is the output of pooling layer, and f is the ReLU function.

The final output sigmoid layer is constructed for the binary classification. The Sigmoid function maps output values between 0 and 1 for classification, which is extracted by the following equation:

$$S(x) = \frac{1}{1 + e^{-x}}. \tag{7}$$

Results

Data preparation

To make the performance evaluationn, DTI-CNN was tested on the drug-target interactions prediction task.

We obtained the heterogeneous network from *Luo*'s paper, which include 12,015 nodes and 1,895,445 edges in total [2]. The isolated nodes are excluded. The heterogeneous network integrates four types of nodes (drugs, proteins, diseases and side-effects) and six type of edges (drug-protein interactions, drug-drug interactions, drug-disease associations, drug-side-effect associations, protein-disease associations and protein-protein interactions) [2]. Based on chemical structures of drugs and primary sequences of proteins, we also built up the multiple similarity networks [16].

The drug nodes, known DTIs and drug-drug interactions were extracted from the DrugBank database [37]. The protein nodes and protein–protein interactions were obtained from the HPRD database [38]. The disease nodes, drug–disease and protein–disease associations were extracted from the Comparative Toxicogenomics Database [39]. The side-effect nodes and drug–side-effect associations were collected from the SIDER database [40].

In our model, we first constructed seven similarity matrices after the Jaccard similarity algorithm. We obtained the drug-related similarity matrices including drug-drug similarity matrix, drug-disease similarity matrix, drug-side-effect similarity matrix and drug similarities matrix. The protein-related similarity matrices include protein-disease similarity matrix, protein-protein similarity matrix and protein similarities matrix. Secondly, we perform RWR algorithm for the two kinds of matrices respectively and splice the single diffusion state matrix of drug and protein networks. After this step, we get two diffusion state matrices corresponding to drug and protein respectively. The rows of the drug diffusion matrix represent different drugs, and the columns represent proteins, diseases, side effects and drugs nodes. The values in the matrix represent the associations between drugs and the four biological entities. The rows of the protein diffusion state matrix represent different proteins. The columns represent proteins, diseases and drugs nodes. The values in the matrix represent the associations between proteins and the three biological entities. Then, we reduce the dimension of drug diffusion state matrix and protein diffusion state matrix respectively by using DAE model. Finally, we obtain the drugs feature vector matrix of 100 dimensions and a total of 708 samples. Similarly, the proteins feature vector matrix is 400 dimensions and 1512 samples.

At last, we adopted the method of ten fold cross validation to divide the train set and test set, in which 90% of the positive and negative samples were used to train model and 10% of the positive and negative samples were used to test the model. According to the known drug-protein interactions matrix, we use the known drug-protein interaction pairs as positive samples. We randomly selected negative samples with the same number of positive samples. In total, we have 3846 samples. After splicing the corresponding protein vectors into drug vectors, we get drug-protein pair vectors of 500 dimensions.

Model parameters

For the RWR model, the restart probability is 0.5 and the number of iterations is 20.

The parameters used in the DAE model are as follows. For the drug features matrices, the original dimensions are 2832 and the DAE has one hidden layer with 100 units. For the protein features matrices, the original dimensions are 4536 and has one hidden layer with 400 units. For drug and protein features, there are 16 and 32 samples for each batch respectively. The number of epochs is 20. The noise scale value is 0.2 and we use Respro optimizer algorithm to train the model.

The parameters of the CNN model are as follows. We added a dropout layer before the fully-connected layer and the dropout percentage is 0.5. We run 35 epochs with 64 samples for each batch . We used Adam algorithm and set an initial learning rate as 0.001 to optimize the binary cross entropy loss [41].

Evaluation Metrics

The AUROC and AUPR scores were used to evaluate the model test and comparison. AUROC and AUPR scores are commonly used evaluation criteria for machine learning, which represent the area under the curve of ROC curve and PR curve respectively. The higher the score, the higher the

prediction accuracy of the model and the better the performance of the model.

ROC curve refers to the curve with false positive probability (FPR) as horizontal axis and true positive rate (TPR) as longitudinal axis, in which $FPR = \frac{FP}{TN+FP}$, $TPR = \frac{TP}{TP+FN}$. PR curve refers to the curve with Recall as horizontal axis and Precision as longitudinal axis, and $Recall = \frac{TP}{TP+FN}$, $Precision = \frac{TP}{TP+FP}$.

Performance evaluation on predicting drug-target interactions

We compared the performance of DTI-CNN with three existing state-of-art methods (DTINet [2], CMF [17] and NRLMF [42]) on the task of predicting drug-target interactions.All models are trained and tested with a 10-fold cross validation. Our comparative results are shown in Fig. 4 and Table 1. Comparing with other methods, the result shows that DTI-CNN can perform best on both scores at the same time and DTINet is the second best method. The AUROC of DTI-CNN is 0.9416, which is 0.03 higher than DTINet. The AUPR of DTI-CNN is 0.02 higher than DTINet. In summary, DTI-CNN performs better on drug-target interactions prediction task than the other three state-of-the-art DTI prediction methods.

Effects of DTI-CNN components

We choose two different approaches in feature selection module and interaction prediction module respectively to figure out the performance of each module of DTI-CNN. To test the effect of our feature selection model, we first use the singular value decomposition (SVD) method [43] to replace our DAE model and name this method as SVD-CNN. To test the effect of the CNN component, we substitute the CNN model with the support vector machine (SVM) model and name this method as DTI-SVM. All three methods are used on DTI prediction task to compare performance and the comparative results are shown in Fig. 5 and Table 2. Comparing with other two methods, the result shows that DTI-CNN can achieve higher AUROC and AUPR scores at the same time, indicating that the two modules of our model are appropriately designed.

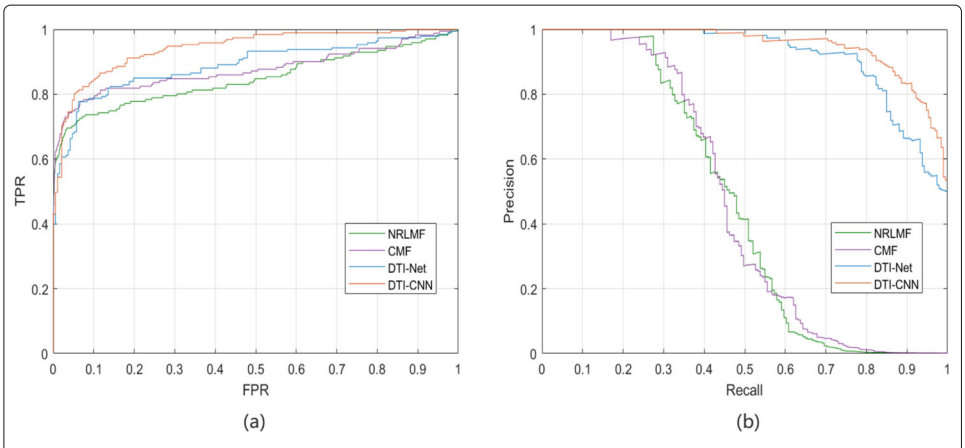

Fig. 4 The ROC **(a)** and P-R **(b)** curves of DTI-CNN, DTINet, CMF and NRLMF on drug-target interactions prediction task. The AUROC of DTI-CNN is 0.9416 and the AUPR of DTI-CNN is 0.9499, which performs better on the drug-target interactions prediction task than the other three state-of-the-art DTI prediction methods

Table 1 The AUROC, AUPR scores of DTI-CNN, DTINet, CMF and NRLMF on drug-target interactions prediction task

Methods	AUROC	AUPR
DTINet	0.9111	0.9290
NRLMF	0.8692	0.4411
CMF	0.9037	0.4173
DTI-CNN	0.9416	0.9499

Case study in three drugs

We extracted the known DTIs from the DrugBank database. We choose the three drugs with the largest number of interactions in known DTIs, which are Quetiapine, Olanzapine and Meprobamate. In the train set, we exclude all the features and interactions between the three drugs and their related proteins to avoid logic circle. In the test set, we input the features of the three drugs and its related proteins. In the "Quetiapine set", 24 of 24 known interactions are identified. In the "Meprobamate set", 23 of 24 known interactions are discovered. And in the "Olanzapine set", 23 of 23 known interactions are recognized. These results indicate that the DTI-CNN method has a good performance on drug-target interactions prediction.

Discussion

At present, there are three traditional kinds of DTI prediction approaches which contains molecular-based approach, ligand-based approach and network-based approach [10]. There are a variety of related models, and the demand for algorithm prediction accuracy is getting higher and higher. The feature learning approach based on deep learning is different from the traditional neural network [44]. Through layer-by-layer learning, we can learn the essential features of the data set without relying on a large scale of samples., so as to predict the unknown data more accurately [45].

The purpose of this research is to improve the prediction accuracy by using the CNN model based on depth learning on the basis of the developed method DTINet. The DTINet contains two modules which are feature extraction and classification. In the first

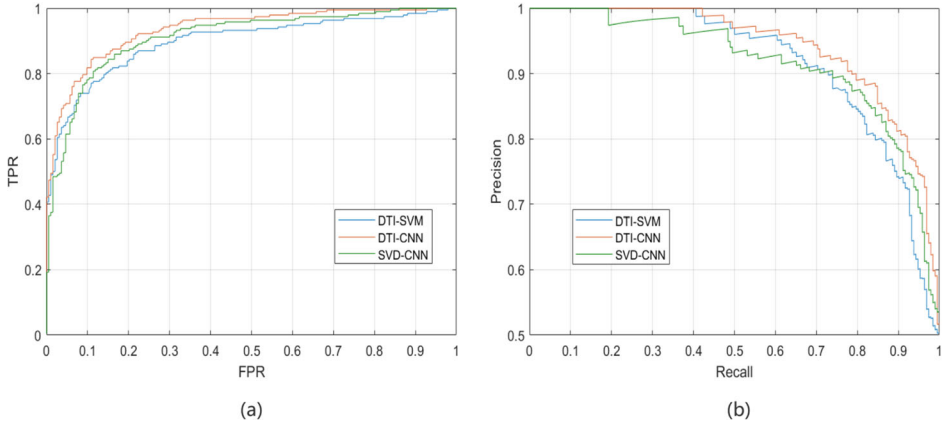

Fig. 5 The ROC **(a)** and P-R **(b)** curves of DTI-CNN, SVD-CNN and DTI-SVM on drug-target interactions prediction task. The DTI-CNN achieves both higher AUROC and AUPR scores than the SVD-CNN and the DTI-SVM method, indicating that the DTI-CNN has been appropriately designed

Table 2 The AUROC, AUPR scores of DTI-CNN, SVD-CNN and DTI-SVM on drug-target interactions prediction task

Methods	AUROC	AUPR
DTI-SVM	0.9068	0.9292
SVD-CNN	0.9306	0.9370
DTI-CNN	0.9416	0.9499

module, we replace the dimension reduction model based on SVD with DAE model to ensure that we can learn features that are more suitable for neural network learning. Then, we improve the classification prediction model of DTINet and use the CNN model as the new prediction model in the second module. Compared with the IMC approach used by DTINet, the CNN model can take into account the topological information and interrelation between the nodes in the network. In addition, we also choose three state-of-the-art DTI prediction methods as comparison. The experimental results show that the AUROC and AUPR scores of our model are both higher than DTINet, NRLMF and CMF.

In addition, we want to know which network contributes more to the DTI prediction. We sequentially remove a network from the original heterogeneous networks as new input data, and then use our DTI-CNN method to perform DTI prediction. The results are shown in Table 3. The result shows that the drug-drug interaction of drug networks and the protein similarities of protein networks contributed more to the DTI prediction. When the drug-drug interaction network was removed, the result achieves both the lowest AUROC and AUPR scores at the same time.

In the future, we will consider adding more relevant information to the heterogeneous network. For the CNN model, we can add the network structure appropriately to accommodate more complex input networks. In this work, although DTI-CNN is mainly designed to predict DTIs, it is an extendible method and can also be used to predict other related directions in the future, such as drug-drug, drug-side-effects and protein-disease.

Conclusion

In this paper, we propose a learning-based method named DTI-CNN to predict the drug-target interactions. Firstly, the Jaccard similarity coefficient and RWR model are used to obtain the relevant features of drugs and targets from heterogeneous networks. Then, we use DAE model to reduce dimensions and identify the essential features. Thirdly, based on the features obtained from the last step, a CNN model is constructed to make a prediction of DTIs.To demonstrate the advantages of DTI-CNN, we compare it with three advanced methods. In addition, we also evaluate the effect of each DTI-CNN module. All

Table 3 The AUROC, AUPR scores of sequentially strip out a network from the original heterogeneous network as new input data of DTI-CNN on drug-target interactions prediction task

Networks	AUROC	AUPR
Without drug-drug	0.9299	0.9370
Without drug-disease	0.9343	0.9416
Without drug side-effect	0.9344	0.9444
Without drug similarities	0.9345	0.9425
Without protein-protein	0.9418	0.9499
Without protein-disease	0.9364	0.9452
Without protein similarities	0.9327	0.9411

the experimental results show that the performance of DTI-CNN is better than that of the existing methods and the proposed method is appropriately designed. The case study also shows that DTI-CNN can be used to predict the drug-target interactions.

Abbreviations
DTIs: Drug-target interactions; IMC: Inductive matrix completion; CMF: Cooperation matrix factorization; CNN: Convolution neural networks; RWR: Random walk with restart; DAE: Denoising autoencoder; MSE: Mean-square error; BP: Backpropagation; ReLU: Rectified linear unit; FPR: False positive probability; TPR: True positive rate; SVD: Singular value decomposition; SVM: Support vector machine.

Acknowledgements
Not applicable.

About this supplement
This article has been published as part of BMC Bioinformatics Volume 21 Supplement 13, 2020: Selected articles from the 18th Asia Pacific Bioinformatics Conference (APBC 2020): bioinformatics. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-21-supplement-13 .

Authors' contributions
JP and XS designed the algorithm; JL implemented the algorithm; JP and JL wrote this manuscript. All authors read and approved the final manuscript.

Funding
Publication costs are funded by National Natural Science Foundation of China (No. 61702421, U1811262, 61772426), the international Postdoctoral Fellowship Program (no. 20180029), China Postdoctoral Science Foundation (No. 2017M610651), Top International University Visiting Program for Outstanding Young scholars of Northwestern Polytechnical University.

Availability of data and materials
The datasets generated during and analysed during the current study are available in the github repository, [https://github.com/MedicineBiology-AI/DTI-CNN.git].

Ethics approval and consent for publication
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Published: 17 September 2020

References
1. Feng Y, Wang Q, Wang T. Drug target protein-protein interaction networks: A systematic perspective. BioMed Res Int. 2017;2017:1289259–1289259.
2. 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. 2017;8(1):573.
3. Huang Y, Zhu L, Tan H, Tian F, Zheng F, Mısır M, Wang X, Liu W. Predicting drug-target on heterogeneous network with co-rank. In: The 8th International Conference on Computer Engineering and Networks (CENet2018). Cham: Springer International Publishing; 2020. p. 571–81. ISBN: 978-3-030-14680-1.
4. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, Zhou W, Huang J, Tang Y. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol. 2012;8(5):1002503.
5. Yamanishi Y, Kotera M, Kanehisa M, Goto S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. Bioinformatics. 2010;26(12):246–54.
6. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: in vitro safety pharmacology profiling: an essential tool for successful drug development. Drug Discov Today. 2005;10(21):1421–33.
7. Dominguez C, Boelens R, Bonvin A. Haddock: a protein- protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003;125(7):1731–7.
8. Morris G, Huey R, Lindstrom W, Sanner M, Belew R, Goodsell D, Olson A. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30(16):2785–91.
9. Keiser M, Roth B, Armbruster B, Ernsberger P, Irwin J, Shoichet B. Relating protein pharmacology by ligand chemistry. Nat Biotechnol. 2007;25(2):197.
10. 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. 2018;35(1):104–11. https://academic.oup.com/bioinformatics/article-pdf/35/1/104/27182907/bty543.pdf.
11. González-Díaz H, Prado-Prado F, García-Mera X, Alonso N, Abeijón P, Caamano O, Yanez M, Munteanu C, Pazos A, Dea-Ayuela MA, et al. Mind-best: Web server for drugs and target discovery; design, synthesis, and assay of mao-b inhibitors and theoretical- experimental study of g3pdh protein from trichomonas gallinae. J Proteome Res. 2011;10(4):1698–718.

12. Lv Z, Ao C, Zou Q. Protein function prediction: From traditional classifier to deep learning. PROTEOMICS. 2019;19(14):1900119. https://onlinelibrary.wiley.com/doi/pdf/10.1002/pmic.201900119.
13. Wu Z, Li W, Liu G, Tang Y. Network-based methods for prediction of drug-target interactions. Frontiers Media SA. 9: 1134.
14. Lan W, Wang J, Li M, Wu F-X, Pan Y. Predicting drug-target interaction based on sequence and structure information. IFAC-PapersOnLine. 2015;48(28):12–6.
15. Xia Z, Wu L-Y, Zhou X, Wong S. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. In: BMC Systems Biology. BioMed Central; 2010. p. 6.
16. 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.
17. Zheng X, Ding H, Mamitsuka H, Zhu S. Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2013. p. 1025–33.
18. Mei J-P, Kwoh C-K, Yang P, Li X-L, Zheng J. Drug–target interaction prediction by learning from local information and neighbors. Bioinformatics. 2012;29(2):238–45.
19. Niwattanakul S, Singthongchai J, Naenudorn E, Wanapu S. Using of jaccard coefficient for keywords similarity. In: Proceedings of the International Multiconference of Engineers and Computer Scientists; 2013. p. 380–4.
20. Jaccard P. Étude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin del la Société Vaudoise des Sciences Naturelles. 1901;37:547–79.
21. Tong H, Faloutsos C, Pan J-Y. Random walk with restart: fast solutions and applications. Knowl Inf Syst. 2008;14(3): 327–46.
22. Lee S, Lee J, Lim J, Suh I. Robust stereo matching using adaptive random walk with restart algorithm. Image Vis Comput. 2015;37:1–11. https://doi.org/10.1016/j.imavis.2015.01.003.
23. Liu H, Song Y, Guan J, Luo L, Zhuang Z. Inferring new indications for approved drugs via random walk on drug-disease heterogenous networks. BMC Bioinformatics. 2016;17(17):539.
24. Peng J, Wang X, Shang X. Combining gene ontology with deep neural networks to enhance the clustering of single cell rna-seq data. BMC Bioinformatics. 2019;20(8):284. Springer.
25. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning, ICML '08. New York, NY, USA: ACM; 2008. p. 1096–103. https://doi.org/10.1145/1390156.1390294.
26. Peng J, Guan J, Shang X. Predicting parkinson's disease genes based on node2vec and autoencoder. Front Genet. 2019;10:226. https://doi.org/10.3389/fgene.2019.00226.
27. Ramachandran P, Zoph B, Le Q, Quoc V. Searching for activation functions. arXiv e-prints. 2017. https://ui.adsabs.harvard.edu/abs/2017arXiv171005941R. Provided by the SAO/NASA Astrophysics Data System.
28. Mukkamala M, Hein M. Variants of RMSProp and Adagrad with logarithmic regret bounds. In: Doina P, Yee Whye T, editors. Proceedings of the 34th International Conference on Machine Learning. vol. 70. Sydney: PMLR; 2017. p. 2545–53.
29. Allen D. Mean square error of prediction as a criterion for selecting variables. Technometrics. 1971;13(3):469–75.
30. LeCun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989;1(4):541–51.
31. Ming Liang, Xiaolin Hu. Recurrent convolutional neural network for object recognition. 20153367–75. https://doi.org/10.1109/CVPR.2015.7298958.
32. Spanhol F, Oliveira L, Petitjean C, Heutte L. Breast cancer histopathological image classification using convolutional neural networks. In: 2016 International Joint Conference on Neural Networks (IJCNN). IEEE; 2016. p. 2560–7.
33. Ren F, Yang C, Qiu Q, Zeng N, Cai C, Hou C, Zou Q. Exploiting discriminative regions of brain slices based on 2d cnns for alzheimer's disease classification; 2019. p. 1–1. https://doi.org/10.1109/ACCESS.2019.2920241.
34. Clevert D-A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). arXiv e-prints. 2015arXiv:1511̇07289. https://ui.adsabs.harvard.edu/abs/2015arXiv151107289C.
35. Luo P, Ding Y, Lei X, Wu F. deepdriver: predicting cancer driver genes by convolutional neural networks. Front Genet. 2019;10:13.
36. Peng J, Hui W, Li Q, Chen B, Hao J, Jiang Q, Shang X, Wei Z. A learning-based framework for miRNA-disease association identification using neural networks. Bioinformatics. 2019;35(21):4364–71. https://academic.oup.com/bioinformatics/article-pdf/35/21/4364/30330838/btz254.pdf.
37. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, et al. Drugbank 3.0: a comprehensive resource for omics research on drugs. Nucleic Acids Res. 2010;39((suppl_1)):1035–41.
38. Keshava Prasad T, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al. Human protein reference database 2009 update. Nucleic Acids Res. 2008;37((suppl_1)):767–72.
39. Davis A, Murphy C, Johnson R, Lay J, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King B, Rosenstein M, Wiegers T, et al. The comparative toxicogenomics database: update 2013. Nucleic Acids Res. 2012;41(D1):1104–14.
40. Kuhn M, Campillos M, Letunic I, Jensen L, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol. 2010;6(1):343.
41. Liu L, Qi H. Learning effective binary descriptors via cross entropy. IEEE. 20171251–8.
42. Liu Y, Min W, Chunyan M, Peilin Z, Xiao-Li L, M P. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction; 2016. p 1004760.
43. GOLUB G, REINSCH C. Handbook series linear algebra. singular value decomposition and least squares solutions. Numer Math. 1970;14:403–20.
44. Arun K, Govindan V. A Hybrid Deep Learning Architecture for Latent Topic-based Image Retrieval. Data Sci Eng. 2018;3:166–95. https://doi.org/10.1007/s41019-018-0063-7.
45. Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y-H, Lu H. Deep-learning-based drug-target interaction prediction. J Proteome Res. 201716. https://doi.org/10.1021/acs.jproteome.6b00618.

Publisher's Note

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