Mathematical Biosciences and Engineering

MBE, 18 (6): 7419–7439.
DOI: 10.3934/mbe.2021367
Received: 11 June 2021
Accepted: 11 August 2021
Published: 30 August 2021

http://www.aimspress.com/journal/MBE

Research article

A model for predicting drug-disease associations based on dense convolutional attention network

Huiqing Wang, Sen Zhao*, Jing Zhao and Zhipeng Feng

College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

* **Correspondence:** Email: 1209570351@qq.com.

Abstract: The development of new drugs is a time-consuming and labor-intensive process. Therefore, researchers use computational methods to explore other therapeutic effects of existing drugs, and drug-disease association prediction is an important branch of it. The existing drug-disease association prediction method ignored the prior knowledge contained in the drug-disease association data, which provided a strong basis for the research. Moreover, the previous methods only paid attention to the high-level features in the network when extracting features, and directly fused or connected them in series, resulting in the loss of information. Therefore, we propose a novel deep learning model for drug-disease association prediction, called DCNN. The model introduces the Gaussian interaction profile kernel similarity for drugs and diseases, and combines them with the structural similarity of drugs and the semantic similarity of diseases to construct the feature space jointly. Then dense convolutional neural network (DenseCNN) is used to capture the feature information of drugs and diseases, and introduces a convolutional block attention module (CBAM) to weight features from the channel and space levels to achieve adaptive optimization of features. The ten-fold cross-validation results of the model DCNN and the experimental results of the case study show that it is superior to the existing drug-disease association predictors and effectively predicts the drug-disease associations.

Keywords: drug-disease association prediction; Gaussian interaction profile kernel similarity; dense convolutional neural network; convolutional block attention module; random forest classifier

1. Introduction

The development of new drugs often goes through a long process including drug discovery, clinical trials, and drug marketing. It takes a lot of time and money to design complex biological

experiments. Newly-discovered drugs have low utilization rates in practice [1]. It is important to find suitable treatment drugs for diseases more efficiently, so researchers are adopting the research model of "new use of old drugs" to realize drug repositioning. They explore the therapeutic effects of marketed drugs on other diseases [2]. Drug-disease association prediction is an important branch in the direction of drug repositioning. It combines drug data and disease data and uses computational methods to find new indications for existing drugs, thereby providing certainty theoretical support for the treatment of diseases and the development of related drugs. In view of this, it is of great research significance to find an effective calculation method to realize drug-disease association prediction.

Drug-disease association prediction has been studied by many researchers. Based on the assumption that similar drugs tended to treat similar diseases [3], the researchers used similarity data of drugs and similarity data of diseases as raw information to predict the drug-disease associations. Wang et al. [4] and Gottlieb et al. [5] used molecular data of drugs and diseases to build a drug's similarity network and a disease's similarity network, and they input this information into the classifier to predict the drug-disease relationship. Zeng et al. [6] fused 10 heterogeneous networks containing information of drugs and diseases, and developed a method based on deep learning to realize drug repositioning. Yang et al. [7] used the structural similarity data of drugs and the semantic similarity data of diseases to reconstruct the drug-disease association matrix and found the new indications for existing drugs. Dai et al. [8] introduced disease-related genetic information to further improve the accuracy of drug-disease association prediction. However, these methods only considered the information on the chemical level of the drugs and the information on the medical level of the diseases, and did not make full use of the existing drug-disease association data.

In the field of silico prediction of interaction, Gaussian interaction profile kernel similarity has been widely used. You et al. [9] calculated the Gaussian interaction profile kernel similarity of diseases and the Gaussian interaction profile kernel similarity of miRNAs based on the miRNA-disease association data, and used them as input data, which effectively improved the prediction results of the model. Twan van Laarhoven et al. [10] predicted the drug-target interaction based on the Gaussian interaction profile kernel similarity of the drugs and the targets. Yan et al. [11] also introduced the Gaussian interaction profile kernel similarity of drugs in the study of drug-drug interactions and achieved better prediction results. Lan et al. [12] used the similarity of lncRNAs and diseases as the input of the model in lncRNA-disease association prediction, which included the Gaussian interaction profile kernel similarity. These studies show that the topology of interaction as a source of information for predicting interactions is important, and the use of Gaussian interaction profile kernel similarity to capture topological information in association data helps to improve the predictive ability of the model.

At present, most drug-disease association prediction methods are based on traditional machine learning, network propagation, and matrix factorization or completion methods. Wang et al. [4] and Gottlieb et al. [5] used support vector machine and logistic regression methods to predict drug-disease associations respectively. Liu et al. [13] analyzed the relationship between entities in the drug-disease heterogeneous network, and performed a two-step restart random walk with drugs and diseases as the center to determine the drug-disease associations. Under the assumption of a low-rank matrix, Yang et al. [7,14] proposed a regularization method with unclear boundaries and an overlap matrix completion method, which complemented the missing values in the drug-disease association matrix. Dai et al. [8] proposed a matrix factorization method to predict drug-disease associations. These methods have achieved certain results in the research of drug-disease association prediction. But they researched directly on the original similarity data of drugs and diseases, and it was difficult to mine the deep feature representations of the data.

The deep learning methods can learn the distribution of the original datasets by training a deep neural network with multiple hidden layers to form abstract high-level features [15], and then achieve accurate prediction and classification. It has been successfully applied to object detection [16,17], protein sites prediction [18], drug repositioning [6,19] and other fields [20,21]. In the study of drug-disease association prediction, Liu et al. [22] proposed the Hnet-DNN model, which used a deep neural network to extract features on the drug-disease heterogeneous network, and then a DNN classifier was trained to predict new drug-disease associations. Wang et al. [23] proposed the HNRD model, which used a deep neural network method to aggregate neighborhood information to learn the node embedding representations of drugs and diseases, and used it for drug-disease association prediction. Han et al. [24,25] proposed the calculation models of SAEROF and GIPAE, respectively using sparse auto-encoders and fully connected network to extract high-level feature representations of drug similarity data and disease similarity data, and input them into the classifier to predict drug-disease associations. These studies used deep learning technology to extract the deep abstract information of drug and disease data. They have achieved good predictive performance in the research of drug-disease association prediction. However, these methods only focused on the high-level network in the process of feature extraction. The interaction between high-level information and low-level information was neglected. They may have lost some information related to the prediction of drug-disease associations.

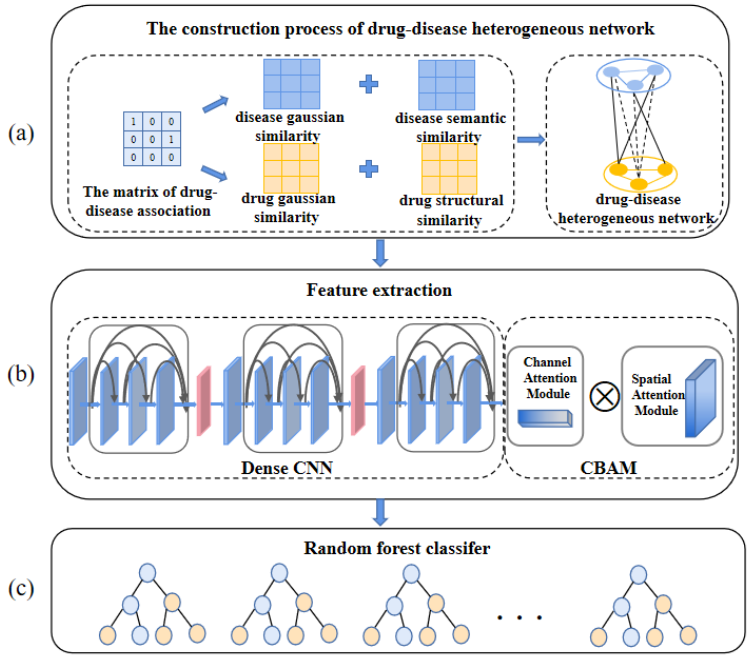

Figure 1. The Flowchart of our work: (a) constructs the drug-disease heterogeneous network; (b) dense convolutional attention network extracts high-level features of drugs and diseases; (c) random forest classifier predicts drug-disease associations.

Based on the above problems, we introduce dense convolutional neural network (DenseCNN) [17] and convolutional block attention module (CBAM) [26], and propose a deep learning model DCNN based on dense convolutional attention network to predict drug-disease associations. The flowchart is shown in Figure 1. First, we introduce the Gaussian kernel function to calculate the Gaussian interaction profile kernel similarity of the drugs and the Gaussian interaction profile kernel similarity

of the diseases based on the drug-disease association data. We merge the Gaussian interaction profile kernel similarity and the structural similarity of drugs, and combine the Gaussian interaction profile kernel similarity and semantic similarity of diseases to construct the feature space of drugs and diseases together. Next, the dense convolutional neural network is introduced to be a feature extractor to focus on different levels of drug information and disease information in the network at the same time, improving the effectiveness of feature representations. Then, the convolutional block attention module is added to the weight feature maps and score the importance of drug information and disease information in the feature extraction process. Finally, a random forest classifier (RF) is used to predict drug-disease associations. The experimental results of ten-fold cross-validation show that the DCNN model is superior to existing methods, effectively learning the information representations of drugs and diseases, and improving the predictive performance of drug-disease associations.

2. Materials and methods

2.1. Dataset

At present, most drug-disease association predictions are studied on datasets F, C and DN [7,14,22–25], the specific information of these datasets is shown in Table 1. Datasets F, C and DN all contain structural similarity data of drugs, semantic similarity data of diseases and drug-disease association data. Among them, the structural information of drugs comes from the DrugBank database, a comprehensive database containing extensive information about drugs (https://go.drugbank.com) [28]. The semantic information of diseases is from the Online Mendelian Inheritance in Man (OMIM) database, which focuses on human genes and diseases (https://www.ncbi.nlm.nih.gov/omim/) [29]. The drug-disease association data can be verified in the Comparative Toxicogenomics Database (CTD) (http://ctdbase.org/) [30].

For the structural similarity of drugs which ranges in [0, 1], we first download the chemical structure information of the drugs in Canonical Simplified Molecular-Input Line-Entry System (SMILES) format from the Drug Bank database [32]. Then the binary fingerprint of the chemical structure of each drug is obtained by the tool of Chemical Development Kit [31]. Finally, the similarity of the drug structure is calculated based on the obtained binary fingerprint.

Table 1. The information of datasets。

Dataset	F	C	DN
Number of drugs	593	663	1490
Number of diseases	313	409	4516
The structural similarity of drugs	593×593	663×663	1490×1490
The semantic similarity of diseases	313×313	409×409	4516×4516
Drug-disease associations data	593×313	663×409	1490×4516
Number of positive samples	1933	2532	1008
Number of negative samples	183676	268635	6727832
Total number of samples	185609	271167	6728840
Number of train samples	3479	4557	1814
Number of validation samples	387	507	202

For the semantic similarity of diseases, we download the medical description of each disease from the Online Mendelian Inheritance in Man (OMIM) database. Then according to the number of occurrences of the Medical Subject Headings vocabulary (MeSH) in the medical description, the semantic similarity of each pair of diseases is calculated, which is in the range of [0, 1] [33,34].

2.2. Construct the drug-disease heterogeneous network

2.2.1. Calculation of gaussian interaction profile kernel similarity of drugs and gaussian interaction profile kernel similarity of diseases

For making full use of the drug-disease association data to improve the accuracy of drug-disease association prediction, we use Gaussian kernel function to calculate the Gaussian interaction profile kernel similarity between any two drugs and any two diseases, and capture topological information in drug-disease association data. Gaussian interaction profile kernel similarity measures the distance of the binary vector of two drugs (diseases) and its ranges in [0, 1]. The greater the similarity value between the two drugs (diseases), the more similar the two drugs (diseases) are. The calculation process of Gaussian interaction profile kernel similarity between drugs and diseases is shown in formulas (2.1) and (2.2).

$$G_{drug} = exp\left(\frac{\|X_i - X_j\|^2}{-2\theta^2}\right) \tag{2.1}$$

$$G_{disease} = exp\left(\frac{\|Y_i - Y_j\|^2}{-2\theta^2}\right) \tag{2.2}$$

Among them, G_{drug} and $G_{disease}$ represent the Gaussian interaction profile kernel similarity values of drugs and diseases, X_i and X_j represent the binary vectors corresponding to drugs i and j, and Y_i and Y_j represent the binary vectors corresponding to diseases i and j. The parameter θ is used to control the local scope of the Gaussian kernel function. In formula (2.1), we set θ to be $\sqrt{\frac{\sum_{i=1}^{r\_n}\|X_i\|^2}{r\_n}}$, r_n is the number of drugs. In formula (2.2), we set θ to be $\sqrt{\frac{\sum_{i=1}^{d\_n}\|Y_i\|^2}{d\_n}}$, d_n is the number of diseases.

2.2.2. Fusion of similarity data

In order to simultaneously consider the information in drug structural similarity, disease semantic similarity, and drug-disease association data, and improve the predictive ability of the model, we have merged drug similarity data and disease similarity data from different perspectives to construct the feature space of drugs and diseases jointly. In the drug-disease association datasets, when the association of drug-disease is unknown, the corresponding Gaussian interaction profile kernel is 0 [24]. We fill the Gaussian interaction profile kernel similarity matrix of the drugs with the structural information of the drugs, and fill the Gaussian interaction profile kernel similarity matrix of diseases with the semantic information of the diseases [24,25]. The fusion process of the drug mixture similarity matrix and the disease mixture similarity matrix is shown in formulas (2.3) and (2.4).

$$Drug_{sim} = \begin{cases} G_{drug}(i,j), & if \ \exists \ G_{drug}(i,j) \\ S_{drug}(i,j), & otherwise \end{cases} \tag{2.3}$$

$$Disease_{sim} = \begin{cases} G_{disease}(i,j), & if \ \exists \ G_{disease}(i,j) \\ S_{disease}(i,j), & otherwise \end{cases} \tag{2.4}$$

Among them, $Drug_{sim}$ refers to the drug similarity after mixing. $G_{drug}(i,j)$ represents the Gaussian interaction profile kernel similarity of the drugs. $S_{drug}(i,j)$ represents the structural similarity of the drugs. $Disease_{sim}$ refers to the similarity of the diseases after mixing. $G_{disease}(i,j)$ represents the Gaussian interaction profile kernel similarity of the diseases. $S_{disease}(i,j)$ represents the semantic similarity of the diseases.

2.2.3. Construction of drug-disease heterogeneous network

The drug-disease heterogeneous network is composed of three parts: the mixed similarity data of drugs and the mixed similarity data of diseases, the drug-disease association matrix, and the construction process is shown in Figure 2. We define drug-disease association data as a $M \times N$ matrix A. When A_{ij} is equals to 1, it means that there is a known association between the drug and the disease, and when A_{ij} is equals to 0, it means that the association between the drug and the disease is unknown. The mixed similarity of drugs is the $M \times M$ matrix $Drug_{sim}$, the mixed similarity of diseases is the $N \times N$ matrix $Disease_{sim}$, the more similar of two drugs or diseases are, the more likely the drugs or the diseases are to act on similar functions. In this paper, the known associations in the drug-disease association data are regarded as positive samples, which are represented by solid lines in the drug-disease heterogeneous network, and the same number of unknown associations are randomly selected as negative samples, which are represented by dashed lines in the heterogeneous network.

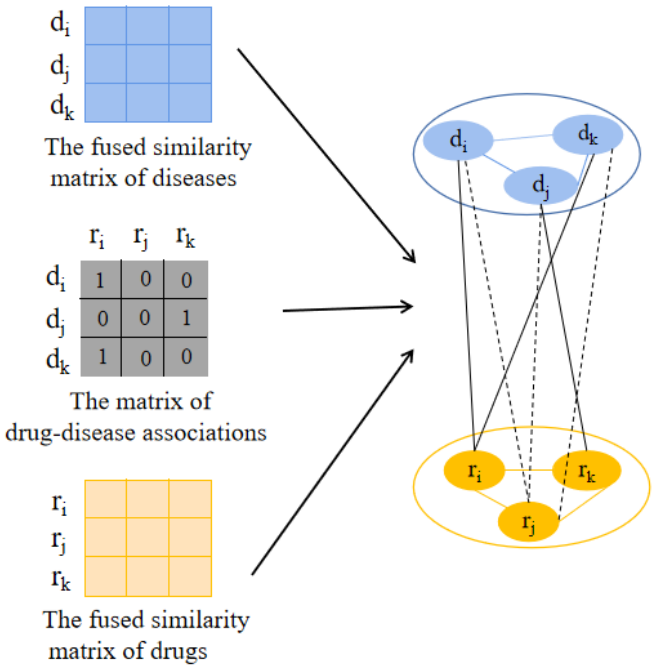

Figure 2. Construction process of drug-disease heterogeneous network.

2.3. Feature learning of drugs and diseases

2.3.1. Dense convolutional neural network

In order to extract high-quality feature representations of drugs and diseases and reduce the loss of information in the feature extraction process, we use dense convolutional neural network to learn in-depth information about drugs and diseases automatically. Each layer of DenseCNN obtains additional input from all preceding layers and passes on its own feature-maps to all subsequent layers [17]. At the same time, it pays attention to the low-level and high-level information of the network, and realizes the information complementarity between different levels. The structure of the dense convolutional neural network is shown in Figure 3.

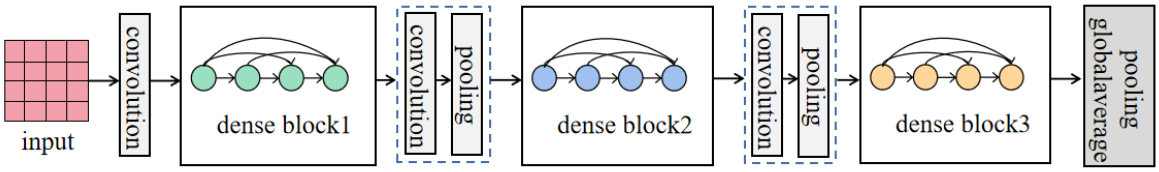

Figure 3. Diagram of dense convolutional neural network structure。

The implementation process of dense convolutional neural network is as follows.

First, we input the data containing drug and disease information into one-dimensional convolution to generate the low-level feature maps of drug and disease information, as shown in formula (2.5).

$$X^0 = \sigma(I \times W + b) \tag{2.5}$$

Among them, I refer to the similarity data after stitching. The data length is L ($L = M + N$, M and N are the number of drugs and diseases respectively). W refers to the weight matrix and b is the bias term. σ is ReLU (Rectified Linear Unit) [35] activation function. X^0 is the output of the one-dimensional convolutional layer.

Next, the low-dimensional feature maps obtained by one-dimensional convolution are used as the input of the dense convolution block to further extract the high-level feature representations of drug and disease information. The dense convolution process is shown in formula (2.6).

$$X^S = H([X^0, X^1, \ldots, X^{S-1}]) \tag{2.6}$$

We define $H(.)$ as a composite function [36] of three consecutive operations: batch normalization (BN) [37], followed by a rectified linear unit (ReLU) [35] and a convolution (Conv). X^S represent the feature maps generated by the Sth convolutional layer in the dense convolution block, and $[.]$ represents the concatenation along the feature dimension. The output of the dense convolution block is the concatenation of the feature dimension of the low-level feature maps X^0 and the feature maps $[X^1, X^2, \ldots, X^S]$ generated by each convolutional layer in the dense convolution block, namely $[X^0, X^1, X^2, \ldots, X^S]$. The structure of a dense convolution block is shown in Figure 4. X^1, X^2, X^3, X^4, are the 4 convolutional layers in a dense convolution block.

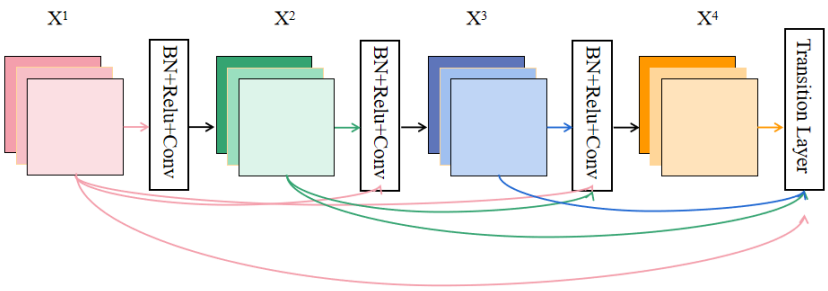

Figure 4. Diagram of the structure of a dense convolution block.

Then the transition layer is used to replace the down-sampling layer in the traditional convolutional neural network to complete the convolution and activation operations. The purpose is to reduce the dimensionality of the feature maps and reduce the risk of model overfitting. The transition layer consists of a convolutional layer and an average pooling layer.

Finally, 3 identical dense convolution blocks are connected in series to form stacked dense convolution blocks to extract high-level features of drug similarity information and disease similarity information. The parameters selection experiment can be seen in Table S1 of the supplementary materials.

2.3.2. Convolutional block attention module

Considering the importance of information contained in different channels and different spaces of dense convolutional neural network, we introduce a convolutional block attention module (CBAM) to weight the proposed features to achieve the importance of drug and disease information, thereby improving the network's ability to predict drug-disease associations. The convolutional block attention module is a lightweight attention module that can be integrated into any convolutional neural network without increasing memory and time overhead. Its structure is shown in Figure 5. Given the feature maps of the drugs and diseases, CBAM will infer the attention map in turn along the two independent dimensions of the channel and space, and then multiply the attention map with the feature maps of drugs and diseases to achieve adaptive optimization of the features [26].

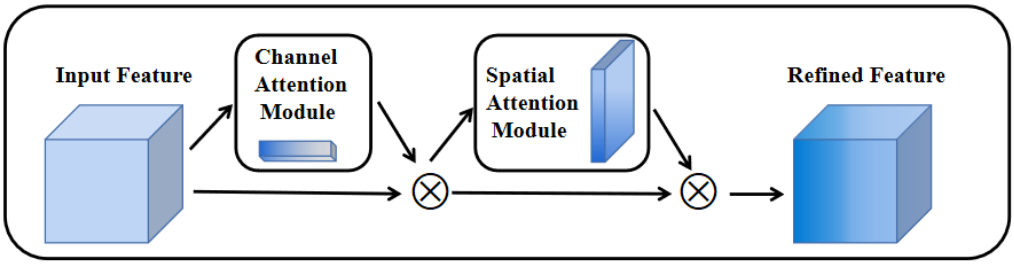

Figure 5. Diagram of the structure of a dense convolution block.

The implementation process of the convolutional block attention module is as follows.

For the input drug and disease feature matrix X, the channel attention map M_C is first generated through the channel attention module, and then the M_C and the original feature matrix X are dotted to obtain X'. The calculation process is shown in formula (2.7).

$$X' = M_C(X) \otimes X \tag{2.7}$$

The channel attention module aggregates the information of the drug and disease feature matrix through average pooling and maximum pooling, and generates the average pooling feature X_{Avg}^C and the maximum pooling feature X_{Max}^C respectively, and then generates the channel attention map $M_C \in R^{C\times 1\times 1}$ through the shared network. The schematic diagram of channel attention is shown in Figure 6a, and the specific calculation process is shown in formula (2.8).

$$\begin{aligned} M_C((X)) &= \sigma\left(MLP\left(AvgPool(X)\right) + MLP\left(MaxPool(X)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(X_{Avg}^C\right)\right) + W_1\left(W_0(X_{Max}^C)\right)\right) \end{aligned} \tag{2.8}$$

The output X' of the channel attention module generates a spatial attention map M_S through the spatial attention module. The spatial attention map M_S performs a dot multiplication with X' to obtain the weighted feature X'' of drugs and diseases. The calculation process is shown in formula (2.9).

$$X'' = M_S(X') \otimes X' \tag{2.9}$$

The spatial attention module generates the average pooling feature $X_{Avg}^S \in R^{1\times H\times W}$ and the maximum pooling feature $X_{Max}^S \in R^{1\times H\times W}$ of drug and disease information through the average pooling layer and the maximum pooling layer respectively, and then a spatial attention map $M_S \in R^{H\times W}$ is generated through convolution operation. The schematic diagram of spatial attention is shown in Figure 6 (b), and the calculation process is shown in formula (2.10).

$$M_S(X') = \sigma\left(f^{7\times 7}([AvgPool(X'); MaxPool(X')])\right) = \sigma\left(f^{7\times 7}\left(\left[X_{Avg}^S; X_{Max}^S\right]\right)\right) \tag{2.10}$$

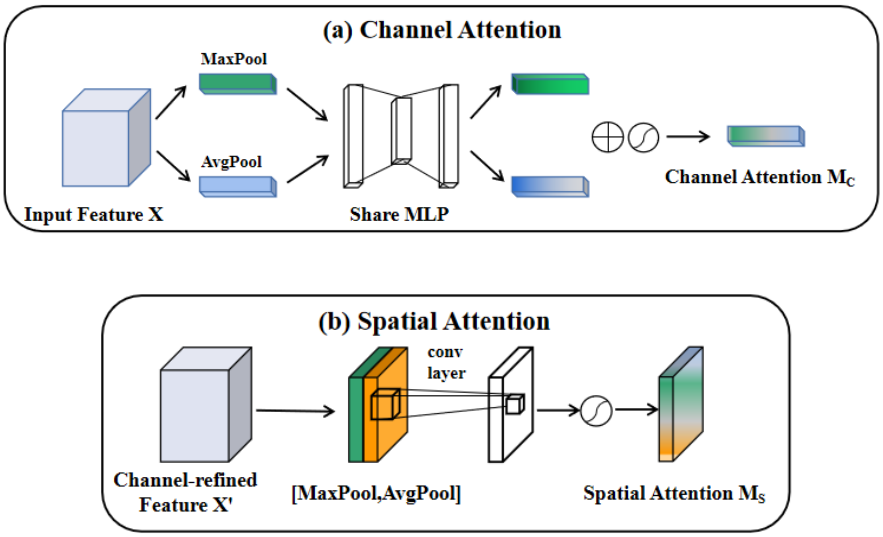

Figure 6. CBAM channel module and spatial module structure diagram (a) channel attention module (b) spatial attention module.

2.4. Random forest classifier predicts drug-disease associations

In the classification stage, we train a random forest classifier based on the feature information of drugs and diseases to realize the prediction of drug-disease associations. The random forest classifier uses a highly parallelized algorithm to detect the interaction between the high-level features of drugs and diseases during the training process, and efficiently calculates the importance of each feature to the output result, which is significant in the training and classification of samples. The random forest has strong generalization ability and high classification accuracy and there is no need to adjust too many parameters in the process of random forest training. In the training process of the random forest classifier, we focus on adjusting three parameters, namely: "n_estimators" = 100, "max_depth" = 50, "max_features" = "auto". The randomness of random forests is reflected in two aspects that the bootstrap technology is used to generate the sample of the decision tree randomly and when the tree is split, a feature subset is randomly selected from all the feature values of the samples to obtain the best classification method. These two random processes in the random forest avoid the occurrence of overfitting effectively.

3. Results

3.1. Evaluate prediction performance of DCNN

In this paper, the area under the receiver operating characteristic (ROC) curve (AUC), accuracy, recall, precision and F1-score are applied to evaluate the performance of DCNN model. AUC is the area under the ROC curve with False Positive Rate (FPR) as the abscissa and True Positive Rate (TPR) as the ordinate. Accuracy represents the proportion of the correct samples predicted by the model to the total samples, and recall represents the proportion of all positive samples predicted by the classifier. Precision is the proportion of correct predictions in the positive samples predicted by the classifier. F1-score is also known as the balanced F score which is the weighted harmonic mean of recall and precision. The higher the values of these indicators mean that the model realizes better performance. The calculation process of each indicator is shown in formulas (3.1–3.6).

$$TPR = \frac{TP}{TP + FN} \tag{3.1}$$

$$FPR = \frac{FP}{FP + TN} \tag{3.2}$$

$$recall = \frac{TP}{TP + FN} \tag{3.3}$$

$$precision = \frac{TP}{TP + FP} \tag{3.4}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.5}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{3.6}$$

To evaluate the predictive ability of the model DCNN in this paper on drug-disease associations, we performed ten-fold cross-validation on the three datasets of F, C, and DN. Ten-fold cross-validation is to divide all data into 10 equal parts randomly. Each fold verification experiment takes turns using 9 pieces of data as the training set to train the model, and 1 piece as the validation set to evaluate the model. Then the average of the results of each fold is the final result of this ten-fold cross-validation. In order to get a more stable result, we carried out 10-fold cross-validation for ten times, and took the average value as the final result. The results are shown in Table 2.

Table 2. Experimental results of the ten-fold cross-validation yielded by DCNN on datasets of F, C and DN.

Dataset	accuracy(%)	precision(%)	recall(%)	F1-score(%)
F	95.16	94.46	96.16	95.19
C	95.51	95.62	95.42	95.51
DN	94.59	94.52	94.64	94.88

It can be seen from Table 2 that the average accuracy of the ten-fold cross-validation of the model DCNN in this paper is 95.16%, the average accuracy is 94.46%, the average recall rate is 96.16%, and the average F1 score is 95.19% on the F dataset. Our model still obtains a high precision value even with good recall values and F1 scores. This shows that our model can not only predict the drug-disease associations better, but also can identify more real positive samples. The higher accuracy value indicates that the DCNN model can accurately identify the currently known and unknown drug-disease association pairs. Similarly, the model DCNN in this paper has achieved good performance on datasets C and DN. These results show that the deep learning model based on dense convolutional attention network can effectively mine the in-depth feature information of drug similarity data and disease similarity data to predict the drug-disease associations accurately.

In order to more intuitively describe the predictive ability of the dense convolutional attention network for drug-disease associations, we draw the ROC curve of the model DCNN on three datasets with ten-fold cross-validation, as shown in Figure 7 where the blue line represents the average AUC of the ten-fold cross-validation, and the curves in other colors are the result of each fold cross-validation.

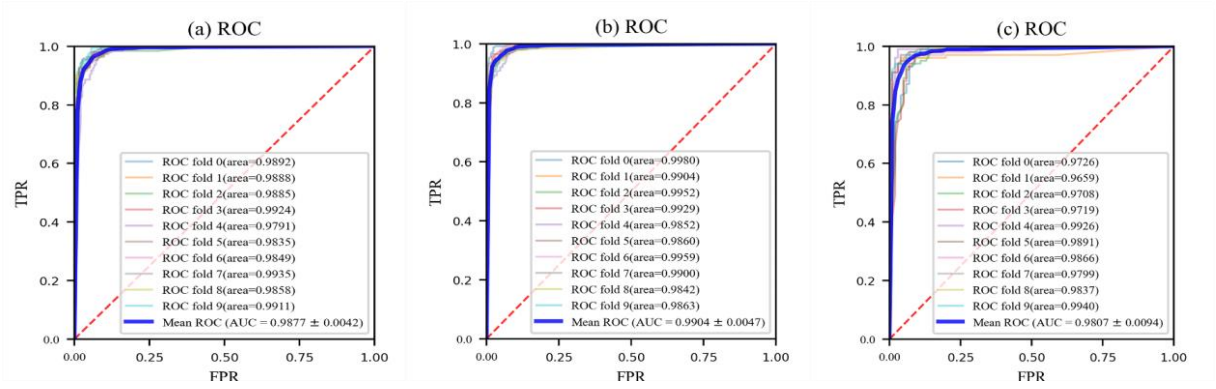

Figure 7. (a), (b), and (c) are the ROC curves of the DCNN model in datasets of F, C and DN, respectively.

As shown in Figure 7 that the average AUC values of the DCNN model on the three datasets of F, C, and DN are 0.9877, 0.9904, and 0.9807, respectively. The higher AUC value, the stronger model's ability to predict drug-disease associations. In the ten-fold cross-validation, the model in this paper achieved a higher AUC value, indicating that the drug-disease potential association scores predicted by the DCNN model had a high degree of credibility to provide a certain theoretical basis for the development of biological experiments. In addition, in order to verify the stability of the DCNN model, we calculated the standard deviation of the AUC on the three datasets to indicate the dispersion degree of verification results on each fold in the ten-fold cross-validation. The standard deviation of the AUC are 0.0042, 0.0047, and 0.0094, in the datasets F, C and DN respectively. The smaller standard deviation shows that the verification results of each fold of the model in this paper do not fluctuate greatly due to the inconsistency of the initial parameters, which further shows that our model has stable predictive performance. Besides, we performed ten times of ten-fold cross-validation on the three datasets to further verify the robustness of the DCNN performance. The relevant experimental results are shown in Table S2 and Figure S1 of supplementary materials. According to Figure S1, we can get that the AUC values of ten times of ten-fold cross-validation on the three datasets do not appear outliers, and the gap between the results of each validation is small. The small deviation among the results of these experiments indicates that the proposed model in this paper shows stable predictive performance in the prediction of drug-disease associations and that our computational model is robust.

3.2. Comparison with existing methods

In order to further assess the proposed DCNN model's prediction performance, on datasets F, C and DN, we have selected three types of comparative experiments including deep learning methods, traditional machine learning algorithms and matrix-based methods. Methods based on deep learning include: Gaussian interaction profile and Kernel-Based Autoencoder (GIPAE), Rotation Forest and Sparse Autoencoder deep neural network (SAEROF), Deep Neural Network Based on Heterogeneous Network Features (Hnet-DNN) and Neighborhood Information Aggregation in Neural Networks (HNRD). Methods based on traditional machine learning algorithms include: Random Forest (RF) and Support Vector Machine (SVM). Matrix-based methods include: Multi-view Multichannel Attention Graph Convolutional Network (MMGCN) and Additional Neural Matrix Factorization model (ANMF) [39,40].

In order to ensure the fairness of the experiment, the comparative experiments were all based on the F, C and DN datasets, and the ten-fold cross-validation method was used to evaluate all comparative experiments. We calculated the AUC value of the comparative methods and the DCNN model, and drew the ROC curve, as shown in Figure 8.

From figure 8, it can be concluded that the AUC values of our model, DCNN, are better than the results of the comparative experiment, in the ten-fold cross-validation on the datasets F, C and DN. The performance of two traditional machine learning methods namely RF and SVM are poorer than other methods in the comparative experiment, because traditional machine learning methods cannot extract deeper information representations when learning the characteristics of the input information. The matrix-based methods MMGCN and ANMF use the similarity information of drugs and the similarity information of diseases to complete or reconstruct the drug-disease associations matrix to supplement the missing values. Although this matrix-based method can directly obtain drug-disease associations scores, their prediction performance is generally lower than deep learning methods. Moreover, due to the sparsity of the DN dataset, we cannot evaluate the prediction performance of the

ANMF model on DN datasets. For the C dataset, compared with the four methods based on deep learning in the comparative experiment, the AUC values of the methods GIPAE, Hnet-DNN, SAEROF and HNRD are 0.9722, 0.9599, 0.9281, 0.9101, In particular, DCNN outperforms GIPAE by 1.82%, Hnet-DNN by 3.05%, SAEROF by 6.23% and HNRD by 8.03%, respectively. Similarly, on the F dataset and the DN dataset, the model DCNN also achieved good results, which showed that the proposed method can more effectively predict the drug-disease associations. In the feature extraction stage, the model DCNN in this paper used a combination of dense convolutional neural network and convolutional block attention module, fully considering the information interaction between different levels in the network, and scoring the importance of features. So the quality of drug information and disease information extracted by DCNN was higher than others, thereby it improved the prediction accuracy of drug-disease associations. The GIPAE, Hnet-DNN, SAEROF, and HNRD methods only paid attention to high-level information in the process of extracting features of drugs and diseases, and these four methods directly merged or connected the features, reducing the quality of the features and losing some information of drugs and diseases. The deep learning model based on the dense convolutional attention network proposed in this paper has achieved a greater improvement compared with the AUC value of the comparative experiment, indicating that the model DCNN has a better ability to rank drug candidates of some diseases. Therefore, the model predicts among all drug candidates with high rankings can be considered first in chemical and medical experiments. Thereby, the DCNN model can provide better theoretical guidance for the realization of drug repositioning.

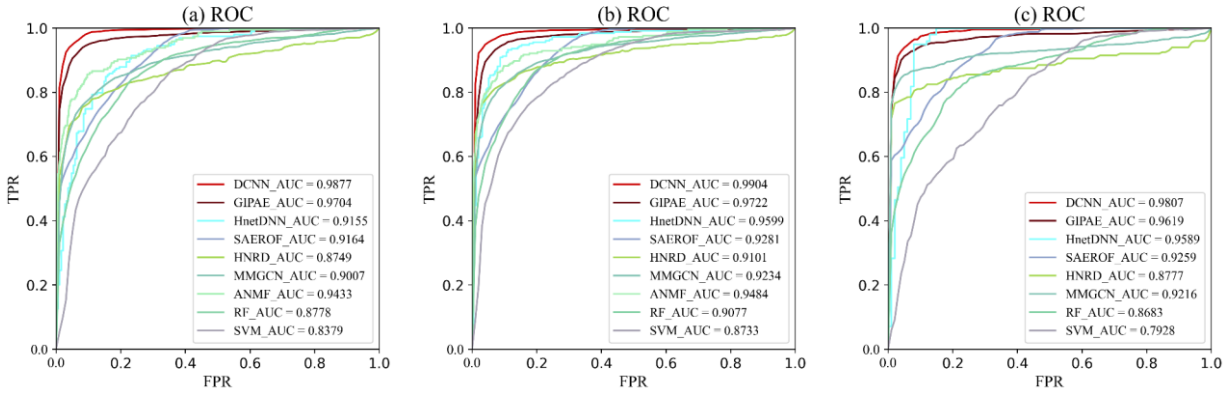

Figure 8. (a), (b), and (c) are the ROC curves of the proposed DCNN and four competitive methods on datasets F, C and DN.

Based on the results of the comparative experiments, we performed a one-way analysis of variance to evaluate whether the performance of the DCNN model proposed in this article significantly improved compared with the existing methods. The analysis results are shown in Table 3. As can be seen from Table 3, DCNN outperforms the other baseline methods and the statistical results indicate that DCNN yields significantly better performance under the p-value threshold of 0.05 in terms of AUCs on datasets of F, C and DN.

Table 3. The statistical results of one-way analysis of variance on the AUCs comparing DCNN and all of eight other methods.

p-Value between DCNN and Another Methods	F	C	DN
GIPAE	2.88×10^{-5}	2.93×10^{-5}	3.77×10^{-4}
SAEROF	4.65×10^{-13}	4.46×10^{-18}	3.07×10^{-12}
Hnet-DNN	5.99×10^{-15}	6.86×10^{-13}	2.93×10^{-7}
HNRD	6.85×10^{-20}	8.90×10^{-15}	5.06×10^{-16}
RF	3.36×10^{-19}	5.88×10^{-17}	5.07×10^{-19}
SVM	7.73×10^{-27}	2.91×10^{-35}	3.27×10^{-23}
MMGCN	2.94×10^{-22}	1.82×10^{-30}	2.51×10^{-14}
ANMF	5.46×10^{-15}	1.16×10^{-26}	-

3.3. Ablation experiments

To verify the effectiveness of Gaussian interaction profile kernel similarity, we conducted the ablation experiments. First, we directly input the structural similarity data of the drugs and the semantic similarity data of the diseases into the DCNN model to predict the drug-disease associations. The results are shown in the second row of Table 4. Then, we fused the drug structural similarity (disease semantic similarity) with the drug Gaussian interaction profile kernel similarity (disease Gaussian interaction profile kernel similarity), and input both the fused drug similarity and fused disease similarity into the DCNN model to predict the drug-disease associations. The results are shown in the third row of Table 4.

Table 4. The ten-fold cross-validation performance of ablation experiments of Gaussian interaction profile kernel similarity. Similarity R_ str means the structural similarity of drugs. Similarity R_ Gau means the Gaussian interaction profile kernel similarity of drugs. Similarity D_ sem means the semantic similarity of diseases. Similarity D_ Gau means the Gaussian interaction profile kernel similarity of diseases.

Similarity	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
R_ str;	F	87.79	87.84	87.80	87.81	92.69
D_ sem;	C	89.41	89.44	89.42	89.42	94.26
	DN	87.75	87.83	87.76	87.75	92.38
R_ str + R_Gau;	F	95.16	94.46	96.16	95.19	98.77
D_ sem + D_Gau;	C	95.51	95.62	95.42	95.51	99.04
	DN	94.59	94.52	94.64	94.88	98.07

It can be seen from Table 4 that when the input of the model only contains drug structural similarity and disease semantic similarity, the prediction results are poor in all indicators. After adding Gaussian interaction profile kernel similarity, all the indicators have been improved. Among them, the AUC values have increased by more than 5% on the F, C, and DN datasets. This shows that the Gaussian interaction profile kernel similarity matrix effectively extracts the topological information in the drug-disease association data. In this way, the topological similarity and biological similarity of

drugs (diseases) are considered meanwhile, which provide a wealth of input information for the model, thereby improving the prediction accuracy of drug-disease associations and providing a certain basis for drug development.

In addition, we have conducted comparative experiments between other kernel functions and Gaussian kernel function. We used Jaccard similarity coefficient and mutual information to extract the topological information of drugs and diseases, respectively [22,38]. We fused the Jaccard similarity (mutual information) of the drugs with the structural similarity of them and fused the Jaccard similarity (mutual information) of the diseases with the semantic similarity of them. Then the fused information is input into the model DCNN for ten-fold cross-validation, and the results are shown in Table S3 of the supplementary materials. Comparing rows 2, 3, and 4 in Table S3, we can see that when capturing the topological information in the drug-disease association data, the results of using Jaccard similarity and mutual information are lower than that of using Gaussian interaction profile kernel similarity. This indicates that Gaussian kernel function is more effective in capturing topological information in the drug-disease association data.

To verify the effectiveness of the data fusion operation, we input the single similarity information and the fused similarity information into the DCNN model, and performed ten-fold cross-validation separately. The results are shown in Table S4. It can be seen from Table S4 that when the fused similarity is used as the input of the DCNN model, the performance is higher than that when a single similarity information is used. This result shows that the fusion of topological similarity information and biological similarity information of drugs (diseases) from the information level is conducive to improving the accuracy of drug-disease association prediction. Therefore, the data fusion operation is meaningful.

In order to verify the important role of the use of various components in the model of this article in the prediction of drug-disease associations, we used ten-fold cross-validation method to conduct ablation experiments on the datasets of F, C and DN, including: the introduction of dense convolutional neural network and the use of different attention modules, the verification results are shown in Table5. The baseline model used the traditional convolutional neural network to extract the deep features of drug similarity information and disease similarity information, and input the features to random forest classifier to predict drug-disease associations.

Table 5. Ablation experiments of DCNN based on datasets of F, C and DN. The largest AUC value for each dataset is highlighted in bold.

CNN	√			
Dense CNN		√	√	√
+SE layer			√	
+CBAM layer				√
F	97.79	98.63	98.69	**98.77**
C	98.10	98.84	98.86	**99.04**
DN	95.84	97.85	97.82	**98.07**

From the second and third columns of Table 5, it can be seen that the introduction of dense convolutional neural network has made the AUC values improved to a certain extent, and the AUC values on the datasets F, C and DN have been increased by 0.84%, 0.74% and 2.01%, respectively. The improvement of these indicators shows that in the process of information extraction, compared with the traditional convolutional neural network, dense convolutional neural network pays attention

to both high-level and low-level features of drugs and diseases which can increase the drug and disease information flow in the network and realize the complementation of information between different levels, thereby obtaining higher-quality abstract representations. Our method solves the problem of the traditional convolutional networks that only pay attention to high-level information.

On the basis of the dense convolutional neural network, we respectively verified the effect of using the squeeze-excitation module (SE module) and the convolutional block attention module (CBAM module) to weight the features [27,27]. The verification results are shown in the fourth and fifth columns of Table 5. The SE module can score the importance of different channel features by weighting each feature map, and the CBAM module can learn the weights of different features from both spatial and channel levels. It can be seen from Table 5 that the introduction of the CBAM module has increased the AUC values of the model on the three datasets of F, C and DN by about 0.2%, while the SE module has a small impact on the model results. This is because dense convolutional neural network constructs channel information features and spatial information features on the local domain of each layer of the network, and merges them in the process of feature extraction. The addition of the convolutional block attention module is not only concerned with the importance of different features at the channel level, and the characteristics of drugs and diseases at the spatial level have been optimized, so that the model can better focus on important information about drugs and diseases to further improve the quality of the features. However, the squeeze-excitation module only distinguishes the information of different channels in the convolutional network, so it performs generally in the process of feature adaptive optimization.

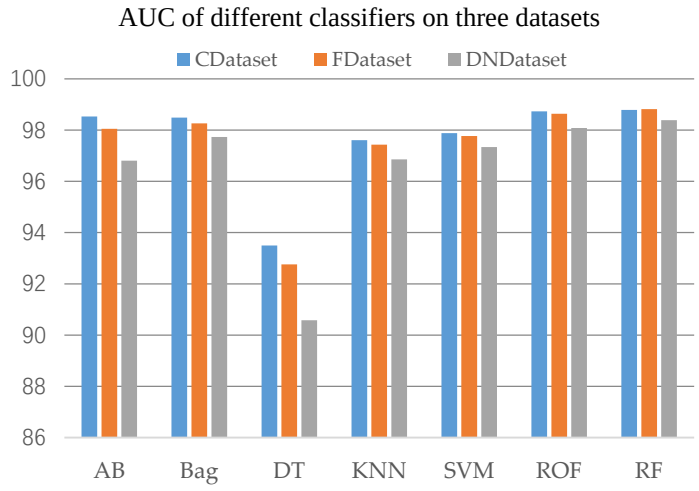

Figure 9. Performance comparison results of different classifiers.

3.4. Comparison among different classifiers

To evaluate the performance of the random forest classifier (RF) used by the DCNN model in the classification task, we used six other classifiers to conduct self-comparison experiments, based on the same feature extraction method, including: Adaboost classifier (AB), Bagging classifier (Bag), Decisiontree classifier (DT), Kneighbors classifier (KNN), Support vector machine (SVM) and Rotationforest classifier (ROF). The ten-fold cross-validation results of different classifiers are shown in Figure 9. It can be seen from Figure 9 that when DT, KNN and SVM are used as classifiers, the AUC values of the model are significantly lower than the AUC values obtained when RF is used as

the classifier. The reason for the worst performance of the decision tree is that the decision tree does not use an integrated algorithm and it is an extremely unstable model. A small deviation in the data will lead to a completely different decision tree. The classifiers AB, Bag and ROF show the same performance as RF basically, while the training time of these classifiers is much longer than the random forest classifier. In contrast, when the random forest classifier is trained, parallel computing is used so that the training speed is fast and the implementation process is simple. In addition, RF achieves high prediction accuracy, so we choose the random forest classifier to predict drug-disease associations.

3.5. Case study

In order to further verify the performance of the DCNN model in practical applications, we predicted and verified drugs that are potentially associated with obesity and stomach cancer. In the course of the case study, it is worth noting that when predicting the candidate drugs of a particular disease, all associations between the particular disease and all the drugs should be removed from the training set. First, we retrained the model using data that does not include obesity and stomach cancer, and then used the trained dense convolutional attention network to make predictions to obtain the correlation scores of candidate drugs for obesity and stomach cancer, respectively. The candidate drugs for these two diseases were ranked according to the correlation score, and the top 20 drugs were verified in the Comparative Toxicogenomics Database(CTD).

Obesity is an important factor that causes diabetes and cardiovascular disease in patients. At present, there are still difficulties in the treatment of obesity. This paper uses the DCNN model to predict the top 20 drug candidates that are potentially associated with obesity. The verification results in the CTD database are shown in Table 6. The first and third columns of Table 6 are the names of drugs that are potentially associated with obesity, and the second and fourth columns are the results of verification in the database. Therapeutic indicates that the drug is clinically used to treat the disease. Marker states that the drug is marked in the CTD database and proved to be related to the studied disease through genetic inference and other methods. Not confirmed means that there is no direct evidence in the CTD database to prove that the drug is related to a specific disease in our case study. From Table 6, we can see that in the case study of obesity, among the top 20 drug candidates predicted by the DCNN model, 17 drugs are considered for the treatment of obesity, among which one of the drugs is known in the drug-disease association data. This further illustrates that the results of our model are reliable to some extent.

Table 6. Top 20 drug candidates for the treatment of obesity.

Drug name	Evidence	Drug name	Evidence
Phenytoin	Marker	Diclofenac	Marker
Palmitic Acid	Marker	Triamcinolone	Marker
Esmolol	Marker	Warfarin	Marker
Pyridostigmine	Marker	Tretinoin	Marker
Phentermine	Therapeutic	Saquinavir	Marker
Pilocarpine	Marker	Methyclothiazide	Not confirmed
Metipranolol	Marker	Lidocaine	Marker
Gemfibrozil	Marker	Vinorelbine	Marker
Imipramine	Marker	Dipivefrin	Not confirmed
Pentostatin	Not confirmed	Cisplatin	Marker

Stomach cancer is a high-incidence cancer in the population, and the mortality rate is high. At present, the drug treatment of stomach cancer still needs further research. In this paper, the verification results of the top 20 candidate drugs for stomach cancer predicted by the DCNN model in the CTD database are shown in Table 7. It can be seen from Table 7 that 16 drugs were verified in the CTD database, which may have a certain effect on the treatment of stomach cancer. That means the model we proposed can predict the potentially related drug candidates for a specific disease, providing a theoretical guidance for the treatment of disease.

Table 7. Top 20 drug candidates for the treatment of stomach cancer.

Drug name	Evidence	Drug name	Evidence
Omeprazole	Marker	Ranitidine	Marker
Allopurinol	Marker	Dinoprostone	Marker
Benazepril	Marker	Phenoxybenzamine	Marker
Fondaparinux	Not confirmed	MethylAminolevulinate	Not confirmed
Nitroglycerin	Marker	Azathioprine	Marker
Mometasone	Marker	Docetaxel	Therapeutic
Tirofiban	Not confirmed	Busulfan	Marker
Meloxicam	Marker	Atazanavir	Marker
Trifluoperazine	Marker	Atorvastatin	Marker
Pseudoephedrine	Not confirmed	Fluvastatin	Marker

To further prove the generalization of the DCNN model, we added other two case studies (breast cancer and Alzheimer disease) with the same way of obesity and stomach cancer. The results are shown in Table S5 and Table S6 of the supplementary materials. It can be seen from Table S5 that among the top 20 drugs predicted by the DCNN model, 17 drugs are proven to be related to the treatment of breast cancer, and among these 17 kinds of drugs, 3 of them are used in clinical treatment. From Table S6 we know that 16 of the top 20 drugs are related to Alzheimer disease predicted by the DCNN model and among these 16 kinds of drugs, 2 of them are used for actual treatment. This further illustrates that the results of our model can provide theoretical support in practical applications to some extent.

4. Conclusions

This paper proposes a deep learning model DCNN for predicting drug-disease associations. The DCNN model introduces Gaussian interaction profile kernel similarity for diseases and drugs on the basis of drug structural similarity and disease semantic similarity, and jointly constructs the feature space of drugs and diseases. In the feature extraction stage, the dense convolutional neural network pays attention to the importance of information interaction between layers in the network, which increases the information flow of drugs and diseases, and solves the problem of information loss caused by only focusing on high-level features in existing methods. The use of the convolutional block attention module further enhances the abstraction ability of the model. It optimizes the features of drugs and diseases from the two levels of space and channel, which is conducive to the improvement of model prediction performance. In the ten-fold cross-validation experiment, the DCNN model achieved better AUC values on the three datasets than the comparison experiments, indicating that the DCNN model can more accurately predict the drug-disease associations. Furthermore, in the case

studies of obesity and stomach cancer, among the top 20 drug candidates predicted by the DCNN model, 17 and 16 drug candidates are verified in the CTD database respectively, proving that the method proposed in this article is in practical application reliability. In future work, we will explore more efficient calculation methods to further improve the model's ability to predict drug-disease associations.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, grant number 61976150; and the Key Research and Development Plan of Shanxi Province, grant number 201903D121151.

Conflict of interest

The authors declare no conflict of interest.

References

1. H. Luo, M. Li, M. Yang, F. Wu, Y. Li, J. Wang, Biomedical data and computational models for drug repositioning: a comprehensive review, *Brief Bioinformatics*, **22** (2020), 1604–1619.
2. L. S. Maryam, G. Nasser, M. S. Rasoul, V. Jaleh, J. R. Green, A review of network-based approaches to drug repositioning, *Brief Bioinformatics*, **19** (2017), 878–892.
3. P. Xuan, L. Zhao, T. Zhang, Y. Ye, Y. Zhang, Inferring Drug-Related Diseases Based on Convolutional Neural Network and Gated Recurrent Unit, *Molecules*, **24** (2019), 2712.
4. Y. Wang, S. Chen, N. Deng, W. Yong, Drug Repositioning by Kernel-Based Integration of Molecular Structure, Molecular Activity, and Phenotype Data, *Plos One*, **8** (2013), e78518.
5. A. Gottlieb, G. Y. Stein, E. Ruppin, R. Sharan, PREDICT: A method for inferring novel drug indications with application to personalized medicine, *Mol. Syst. Biol.*, **7** (2011), 496.
6. X. Zeng, S. Zhu, X. Liu, Y. Zhou, F. Cheng, deepDR: a network-based deep learning approach to in silico drug repositioning, *Bioinformatics*, **35** (2019), 5191–5198.
7. M. Yang, H. Luo, Y. Li, J. Wang, Drug repositioning based on bounded nuclear norm regularization, *Bioinformatics*, **35** (2019), i455–i463.
8. W. Dai, X. Liu, Y. Gao, L. Chen, J. Song, D. Chen, et al., Matrix Factorization-Based Prediction of Novel Drug Indications by Integrating Genomic Space, *Comput. Math. Method M.*, **2015** (2015), 275045.
9. Z. You, Z. Huang, Z. Zhu, G. Yan, X. Chen, PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction, *PLoS Comput. Biol.*, **13** (2017), e1005455.
10. V. L. Twan, S. B. Nabuurs, M. Elena, Gaussian Interaction Profile kernels for predicting drug–target interaction, *Bioinformatics*, **27** (2011), 3036–3043.
11. C.Yan, G. Duan, Y. Pan, F. Wu, J. Wang, DDIGIP: predicting drug-drug interactions based on Gaussian Interaction Profile kernels, *BMC Bioinform.*, **20** (2019), 538.
12. W. Lan, L. Min, K. Zhao, J. Liu, F. Wu, Y. Pan, et al., LDAP: a web server for lncRNA-disease association prediction, *Bioinformatics*, **33** (2017), 458–460.

13. L. Hui, Y. Song, J. Guan, L. Luo, Z. Zhuang, Inferring new indications for approved drugs via random walk on drug-disease heterogenous networks, *BMC Bioinform.*, **17** (2016), 539.
14. M. Yang, H. Luo, Y. Li, F. Wu, J. Wang, Overlap matrix completion for predicting drug-associated indications, *PLoS. Comput. Biol.*, **15** (2019), e1007541.
15. Y. Lecun, Y. Bengio, G. Hinton, Deep learning, *Nature*, **521** (2015), 436–444.
16. Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, et al., M2Det: A Single-Shot Object Detector based on Multi-Level Feature Pyramid Network, In *Proceedings of AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, 2019.
17. G. Huang, Z. Liu, V. Laurens, K. Q. Weinberger, Densely Connected Convolutional Networks, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017.
18. D. Wang, S. Zeng, C. Xu, W. Qiu, Y. Liang, T. Joshi, et al., MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction, *Bioinformatics*, **33** (2017), 3909–3916.
19. H. Wang, J. Wang, C. Dong, Y. Lian, Z. Yan, A Novel Approach for Drug-Target Interactions Prediction Based on Multimodal Deep Autoencoder, *Front. Pharmacol*., **10** (2020), 1592.
20. R. Hu, J. Andreas, M. Rohrbach, T. Darrell, K. Saenko, In Learning to Reason: End-to-End Module Networks for Visual Question Answering, In P*roceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017.
21. L. Yu, Z. Lin, X. Shen, J. Yang, X. Lu, M. Bansal, et al., MAttNet: Modular Attention Network for Referring Expression Comprehension, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018.
22. H. Liu, W. Zhang, Y. Song, L. Deng, S. Zhou, HNet-DNN: inferring new drug-disease associations with deep neural network based on heterogeneous network features, *J. Chem. Inf. Model.*, **60** (2020), 2367–2376.
23. Y. Wang, G. Deng, N. Zeng, X. Song, Y. Zhuang, Drug-Disease Association Prediction Based on Neighborhood Information Aggregation in Neural Networks, *IEEE Access*, **7** (2019), 50581–50587.
24. J. Han, Y. Huang, Z. You, SAEROF: an ensemble approach for large-scale drug-disease association prediction by incorporating rotation forest and sparse autoencoder deep neural network, *Sci. Rep.*, **10** (2020), 4972.
25. J. Han, Y. Huang, Z. You, Predicting Drug-Disease Associations via Using Gaussian Interaction Profile and Kernel-Based Autoencoder, *BioMed. Res. Int.*, **2019** (2019), 1–11.
26. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, In CBAM: Convolutional Block Attention Module, In *Proceedings of the European Conference on Computer Vision*, Munich, Germany, 2018.
27. H. Jie, S. Li, S. Gang, In Squeeze-and-Excitation Networks, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018.
28. D. S. Wishart, K. Craig, A. C. Guo, S. Savita, H. Murtaza, S. Paul, et al., DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Res.*, **34** (2006), D668–672.
29. H. Ada, A. F. Scott, A. Joanna, B. Carol, V. David, V. A. Mckusick, Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders, *Nucleic Acids Res*., **30** (2005), 514.

30. A. P. Davis, C. J. Grondin, R. J. Johnson, D. Sciaky, C. J. Mattingly, The Comparative Toxicogenomics Database: update 2019, *Nucleic Acids Res.*, **47** (2019), D948–D954.
31. C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E. L. Willighagen, The chemistry development kit (CDK): An open-source Java library for chemo- and bioinformatics, *J. Chem. Inf. Comput. Sci.*,**43** (2003), 493–500.
32. D. Weininger, SMILES, a chemical language and information system Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.*, **28**, (1988) 31–36.
33. M. A. Van Driel, J. Bruggeman, G. Vriend, H. G. Brunner, and J. A. M. Leunissen, A text-mining analysis of the human phenome, *Eur. J. Hum. Genet.*, **14** (2006), 535–542.
34. A. Hamosh, A. F. Scott, J. S. Amberger, D. Valle, and V. A. Mckusick, Online mendelian inheritance in man (OMIM), *Hum. Mutation*., **15** (2000), 57–61.
35. X. Glorot, A. Bordes, Y. Bengio, In Deep Sparse Rectifier Neural Networks, In *Proceedings of International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, USA, 2011.
36. K. He, X. Zhang, S. Ren, S. Jian, Identity Mappings in Deep Residual Networks, In *Proceedings of European Conference on Computer Vision*, Fort Lauderdale, Amsterdam, Netherlands, 2016.
37. S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, In *Proceedings of International Conference on Machine Learning*, Lille, France, 2015.
38. H. Wang, J. Wang, C. Dong, Y. Lian, D. Liu, Z. Yan, A Novel Approach for Drug-Target Interactions Prediction Based on Multimodal Deep Autoencoder, *Frontiers in Pharmacology*, **10** (2020).
39. X. R. Tang, J. W. Luo, C. Shen, Z. H. Lai, Multi-view Multichannel Attention Graph Convolutional Network for miRNA-disease association prediction, *Brief. Bioinform.*, **2021** (7006), 1–12.
40. X. X. Yang, L. Zamit, Y. Liu, J. Y. He, Additional Neural Matrix Factorization model for computational drug repositioning, *BMC Bioinform.*, **2019** (20), 423.