Inferring Drug-Related Diseases Based on Convolutional Neural Network and Gated Recurrent Unit

Ping Xuan [1], Lianfeng Zhao [1,*], Tiangang Zhang [2,*], Yilin Ye [1] and Yan Zhang [1]

[1] School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China
[2] School of Mathematical Science, Heilongjiang University, Harbin 150080, China
* Correspondence: zhaolianfenghlju@163.com (L.Z.); zhang@hlju.edu.cn (T.Z.);
Tel.: +86-136-8944-6615 (L.Z.); +86-188-4503-0636 (T.Z.)

check for updates

Abstract: Predicting novel uses for drugs using their chemical, pharmacological, and indication information contributes to minimizing costs and development periods. Most previous prediction methods focused on integrating the similarity and association information of drugs and diseases. However, they tended to construct shallow prediction models to predict drug-associated diseases, which make deeply integrating the information difficult. Further, path information between drugs and diseases is important auxiliary information for association prediction, while it is not deeply integrated. We present a deep learning-based method, CGARDP, for predicting drug-related candidate disease indications. CGARDP establishes a feature matrix by exploiting a variety of biological premises related to drugs and diseases. A novel model based on convolutional neural network (CNN) and gated recurrent unit (GRU) is constructed to learn the local and path representations for a drug-disease pair. The CNN-based framework on the left of the model learns the local representation of the drug-disease pair from their feature matrix. As the different paths have discriminative contributions to the drug-disease association prediction, we construct an attention mechanism at the path level to learn the informative paths. In the right part, a GRU-based framework learns the path representation based on path information between the drug and the disease. Cross-validation results indicate that CGARDP performs better than several state-of-the-art methods. Further, CGARDP retrieves more real drug-disease associations in the top part of the prediction result that are of concern to biologists. Case studies on five drugs demonstrate that CGARDP can discover potential drug-related disease indications.

Keywords: drug-disease association prediction; convolutional neural network; gated recurrent unit; attention mechanism at path level; drug repositioning

1. Introduction

In the past decades, there has been a gradual increase in new molecular entity research and development, but the number of new molecular entities approved by the Food and Drug Administration (FDA) has been decreasing [1–3]. Traditional drug development often requires 10–15 years and an investment of $1.5 billion [4–6]. Because FDA-approved drugs undergo biological experiments, clinical trials, and are evaluated for safety, drugs are often repositioned. Repositioning existing drugs for new indications or uses requires only 6.5 years, and the cost is $300 million, which is far less than the cost of developing a new drug [7–9].

Based on different biological premises and assumptions, researchers use different data types and biological preconditions to study drug repositioning. Research methods include retargeting based on

drug targets [10,11], relocation based on drug side effects [12–14], and heterogeneity based on drug diseases [15–18]. Most drug targets are directly linked to the pathogenesis of the diseases. Li et al. constructed a drug-target heterogeneous network using similarities between the targets and the drugs to integrate information between the target and the drug for drug repositioning. Zhao et al. [19] used target gene information and disease-causing gene information to calculate drug similarities and disease similarities, and they finally identified a gene-disease relationship through the Bayesian method. Wang et al. [20] proposed a three-layer heterogeneous network that integrates drug similarities, disease similarities, drug-disease associations, and drug-target interactions to disseminate information for predicting the relationship between the drugs and the diseases. However, drugs can cause off-target phenomena in the living environment and produce unexpected side effects; therefore, drug side effects are also one of the essential factors for repositioning drugs. Campillos et al. [16] proposed a drug side effect similarity to determine whether two drugs are involved in the same target. Gottlieb et al. [21] and Zhang et al. [22] used drug chemical substructure, side effects, etc. to calculate drug similarities using logistic regression and collaborative filtering algorithms to predict potential drug-diseases relationship. However, these methods are not suitable for drugs and diseases that do not have a common gene or target.

Most advanced methods are predictive for drug–disease networks. Liang et al. [23] used drug chemical substructure information, drug target domain information, and drug target annotation information to calculate drug similarities; the drug-disease associations were predicted through Laplacian regularized sparse subspace learning (LRSSL). Luo et al. [24] used drug chemical substructure information to calculate drug similarities, and they used disease semantics to calculate disease similarities. Then, they constructed a drug-disease two-layer heterogeneous network using a bi-random walk with a restart algorithm to reposition drugs. Zhang et al. [25] also used drug similarity and disease similarity to design drug–disease heterogeneous networks for repositioning drugs based on matrix decomposition with similarity constraints. Xuan et al. [26] proposed a matrix-based decomposition method for integrating drug similarity and disease similarity to predict drug–disease associations. However, these methods are shallow learning methods that cannot accommodate complex and non-linear information on drug similarity, disease similarity, and drug–disease associations. In addition, the paths of drugs and diseases as important auxiliary information were not deeply integrated in these previous methods. Therefore, a deep-learning-based prediction method must be developed to integrate the similarity, association information, and path information of drug–disease pairs. We propose a prediction method based on a convolutional neural network (CNN) and gated recurrent unit (GRU) called CGARDP for predicting drug-disease associations. The left part of CGARDP's prediction model focuses on local information related to a drug-disease pair, and the right part of the model learns the path information between drug-disease pairs. Experimental cross-validation results clearly show that CGARDP performs better than several of the most advanced prediction methods. Case studies involving five drugs show that CGARDP can detect potential candidate disease indications.

2. Materials and Methods

2.1. Dataset

We obtained drug-disease association data from study [26], which involved 763 drugs and 681 diseases. The chemical fingerprints extracted from the PubChem database [27] were used for representing the chemical substructures of drugs. Disease information can be obtained from the MeSH database [28]. We obtained drug similarity and disease similarity data from a work published on LRSSL [23].
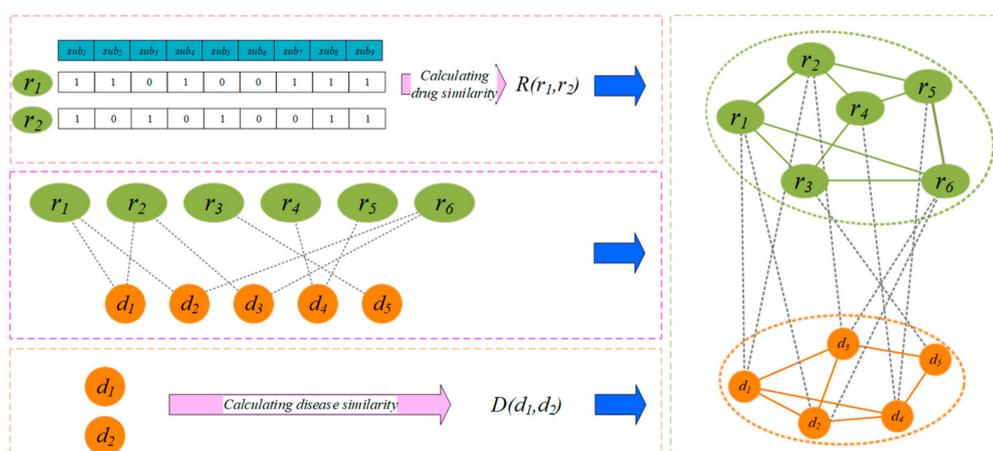
2.2. Construction of Drug-Disease Network

The more similar the chemical substructures of two drugs are, the more likely the drugs are to act on similar functions. The chemical substructure vector S_i of a drug r_i is an 869-dimensional

binary vector. We defined $S_i = \{sub_{i,1}, sub_{i,2}, \ldots, sub_{i,j}, \ldots, sub_{i,869}\}$, where $sub_{i,j}$ is the j-th chemical substructure of the i-th drug. LRSSL [23] measured the drug similarities by calculating the cosine similarities between the chemical substructures of drugs. We also use $\mathbf{R} = R_{[i,j]} \in \mathcal{R}^{N_r \times N_r}$, which represents drug similarity, where $R_{[i,j]}$ is in the range of [0, 1] and is the similarity of r_i and r_j, and N_r denotes the number of drugs.

To evaluate the similarity between diseases, we establish directed acyclic graphs (DAG) of semantic terms for corresponding diseases, which contain all semantic terms related to that disease. Wang et al. [28] successfully calculated the semantic similarity between diseases using their related terms in the DAG graph. LRSSL computed the similarities between diseases by using Wang's method, and we obtained the disease similarity from LRSSL. Let $\mathbf{D} = D_{[i,j]} \in \mathbf{R}^{N_d \times N_d}$ be a similarity matrix of diseases such that each element is between 0 and 1.

In light of the relationship between drugs and diseases, we add an edge between the corresponding drug and disease (Figure 1). Matrix $A \in R^{N_r \times N_d}$ denotes the edge set; if $A_{ij} = 1$, drug r_i is associated with the disease d_j, otherwise, $A_{ij} = 0$.

Figure 1. Construction of a drug-disease heterogeneous network based on the similarity calculation.

2.3. Prediction Model Based on CNN and GRU

To predict the potential representation of the association between a drug and a disease, we propose a novel prediction model based on a CNN and GRU. We apply the CNN module in the left part to learn the combinatorial representation of drug r_i and disease d_j; further, we apply GRU in the right part to capture the path representation between r_i and d_j. Finally, the two representations were integrated by a combined strategy to achieve the final correlation scores of r_i and d_j. We take drug r_1 and disease d_3 as an example to describe the learning framework for the left and right parts, and we use x, $\mathbf{x}$, $\mathbf{X}$ to represent the scalar, vector, and matrix, respectively.

The probability that a drug is associated with a disease is higher when there are more drugs similar to another drug associated with a disease, such as r_1 and d_3. As shown in Figure 2, drugs similar to r_1 are $\{r_2, r_3, r_6\}$, and the drugs associated with d_3 are $\{r_2, r_6\}$. The drugs associated with d_3 are similar to r_1, and therefore, the probability of d_3 being associated with r_1 is very high. The first row of matrix $\mathbf{R}$ denotes the similarity between r_1 and all drugs, and the third row of the matrix $\mathbf{A}^T$ denotes as the associations between d_3 and all drugs.

A drug is associated with more diseases that are similar to a disease, so the more likely the drug is associated with the disease, such as r_1 and d_3. As shown in Figure 2, diseases similar to d_3 are $\{d_1, d_2, d_5\}$ and the r_1 associated with $\{d_1, d_2\}$; therefore, r_1 and d_3 are more likely to be related. The third row of the matrix $\mathbf{D}$ denotes the similarity between d_3 and all diseases, and the first row of matrix $\mathbf{A}$ denotes the association between r_1 and all diseases.

Therefore, we combine the left and right feature representations into the feature matrix $\mathbf{X} = X_{[i,j]} \in R^{2 \times (N_r + N_d)}$ of r_1 and d_3, N_r is the number of drugs and N_d is the number of diseases. The first row of the matrix $\mathbf{X}$ denotes the eigenvector of drug r_1, and the second row denotes the eigenvector of disease d_3.

Figure 2. Construction of the feature matrix by integrating the similarities and associations.

### 2.3.1. Convolution Module on the Left

Convolutional Layer

As shown in Figure 3, to capture the boundary information of X, we first apply a padding operation obtain a new matrix named X'. Then, we use X' as an input to the left convolution module [29] to learn the potential representation of a drug-disease pair. We assume that the size of the filter is set as W_f and W_h for each layer of convolution. When there are n_{conv} filters, the convolution filter $W_{conv} \in R^{W_f \times W_h \times n_{conv}}$ is applied to X'. Then, we obtain the feature matrix $\mathbf{Z}_{conv} \in R^{(2-W_h+2p+1) \times (d-W_f+2p+1) \times n_{conv}}$, where p is the number of padding layer in the feature matrix of the CNN model, and d is the length of $\mathbf{X}'$. $X'_{conv}(i,j)$ is the element at the i-th row and the j-th column of $\mathbf{X}'$, and $X'_{conv}(k,i,j)$ represents a region within the filter when the k-th filter slides to the $X'_{conv}(i,j)$. The formal definitions of $X'_{conv}(k,i,\ j)$ and $\mathbf{Z}_{conv,k}(i,j)$ are as follows:

$$X'_{conv}(k,i,j) = X'_{conv}\big(i:i+w_f,\ j:j+\ w_h\big), X'_{conv} \in R^{W_f \times W_h}, \tag{1}$$

$$\mathbf{Z}_{conv,k}i,j = \ f\big(X'_{conv}(k,i,\ j) * \mathbf{W}_{conv}(k,:,:) + \mathbf{b}_{conv}(k)\big), \tag{2}$$

$$i \in [1, 2 - W_h + 2p + 1],\ j \in \big[1, d - W_f + 2p + 1\big],\ k \in [1, n_{conv}], \tag{3}$$

where $\mathbf{W}_{conv}(k,:,:)$ is the sliding window weight matrix of the k-th filter, $\mathbf{b}_{conv}$ is the bias vector, f is a ReLU function [30], $\mathbf{Z}_{conv,k}i,j$ is the element at the i-th row and j-th column of the k-th feature map $\mathbf{Z}_{conv,k}$.
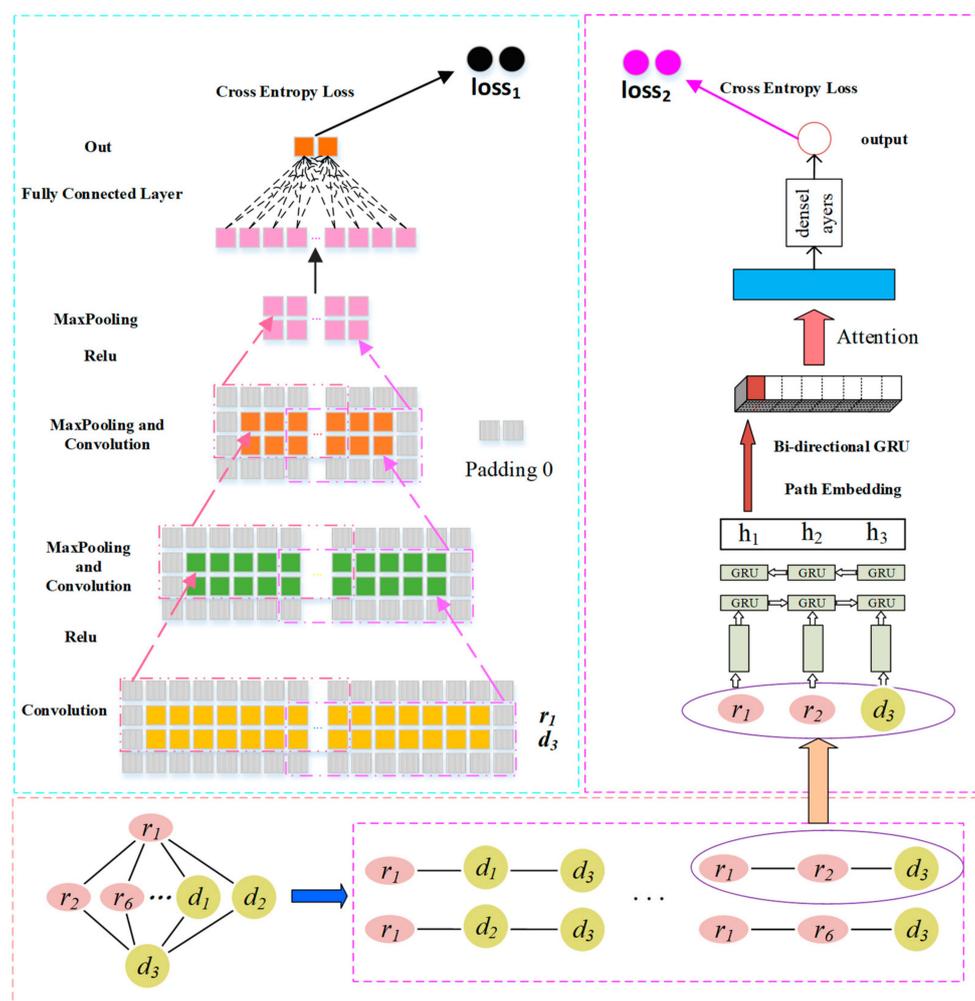
Figure 3. Drug-disease association prediction framework based on convolutional neural network (CNN) and gated recurrent unit (GRU).

Pooling Layer

The feature maps $Z_{conv,k}$ are pooling layers for downsampling to remove unimportant sample data, thus further reducing the number of parameters. We use max pooling to complete the pooling operation and set its sampling window size to $W_m \times W_p$. The pooling outputs of all the feature maps are $Z_{convpool,k}$:

$$Z_{convpool,k}(i, j) = \text{Max}\big(Z_{conv,k}\big(i : i + W_m, j : j + W_p\big)\big), \tag{4}$$

$$i \in [1, 2 - W_m + 2p + 1], \ j \in \left[1, d - W_p + 2p + 1\right], \ k \in [1, n_{conv}], \tag{5}$$

where $Z_{convpool,k}$ is the k-th feature map, and $Z_{convpool,k}(i, j)$ is the element at its' i-th row and j-th column, and p is the number of padding layer in the $Z_{conv,k}$. We obtain the feature representation of the node pair $Z_{convpool,k}(i, j)$, which is flattened and sent to the fully connected layer. The characteristic of the output represents the final result obtained by flattening the fully connected layer as a potential association for the final drug–disease pair c:

$$c = \sigma(Z_{convpool,k} \cdot W_l), \tag{6}$$

$$W_l \in R^{(\frac{2 - W_h + 2p}{s} + 1) \times (\frac{d - W_f + 2p}{s} + 1) \times 2}, \tag{7}$$

where σ is a sigmoid function [31], W_l is a fully connected layer feature matrix, and $\cdot$ is the dot product symbol.

2.3.2. GRU with Attention-Based Path Encoder on the Right

For the prediction of the novel association between drug r_i and disease d_j, the different paths between the two nodes contribute differently to their associations. Thus, a path-level attention mechanism is introduced to select more important paths for the association between r_i and d_j. This mechanism consists of two parts: a path encoder and a path attention layer, as shown in Figure 3.

GRU-Based Sequence Encoder

The GRU module [32] tracks the state of paths with a gating mechanism instead of using separate memory cells. There are two types of gates: the reset gate r_t and the update gate z_t. These gates jointly control the amount of information that is updated to the state. To illustrate the updated process of the state, we take r_1 and d_3 as an example. There are four paths between r_1 and d_3 to form a set $P_{13} = \{r_1 \rightarrow r_2 \rightarrow d_3, r_1 \rightarrow r_6 \rightarrow d_3, r_1 \rightarrow d_1 \rightarrow d_3, r_1 \rightarrow d_2 \rightarrow d_3\}$. The node in each path inputs its corresponding feature vector x_t. The i-th path in P_{13} is represented by P_{13}^i, and the new state h_t of the t-th node is calculated as:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \widetilde{h}_t, \tag{8}$$

where h_{t-1} is the state of the $t - 1$ state in the path, and $\widetilde{h}_t$ is the candidate state of the current node. This is a linear interpolation between the previous state h_{t-1} and the current new state $\widetilde{h}_t$ computed with new information. The update gate z_t controls the extent to which the previous node information is introduced into the current state. The closer the gate z_t is to 1, the more the state information of the previous node is brought in. z_t is updated as:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \tag{9}$$

where x_t is the vector at the t-th node, W_z is the weight matrix of the node vector, U_z is the weight matrix of the previous state, and b_z is a bias vector. The candidate state $\widetilde{h}_t$ is calculated as:

$$\widetilde{h}_t = \tan h(W_h x_t + r_t \cdot (U_h h_{t-1}) + b_h), \tag{10}$$

where r_t is the reset gate that controls how much the past state contributes to the candidate state. If r_t is zero, it will forget all previous states. W_h and U_h are matrices of the candidate state, b_h is the bias vector of the candidate state, and $\cdot$ is the Hadamard product symbol. The reset gate is updated as:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), \tag{11}$$

where σ is the sigmod function, W_r is the weight matrix of the node vector x_t in the reset gate, U_r is the weight matrix of the candidate state h_{t-1}, and b_r is the bias vector.

GRU-Based Path Encoder

We assume that P_{ij}^t is the path set of drug r_i and disease d_j, and the t-th path contains nodes. We use a bidirectional GRU module to integrate the information in two directions of the path and combine the context information of the path nodes. A bidirectional GRU module contains a forward $\overrightarrow{GRU}$ module, which reads from the first node to the last node, and the backward $\overleftarrow{GRU}$ module, which reads from the last node to the first node as:

$$\overrightarrow{h_{ij}^t} = \overrightarrow{GRU}\left(P_{ij}^t\right), \tag{12}$$

$$\overleftarrow{h_{ij}^t} = \overleftarrow{GRU}\left(P_{ij}^t\right). \tag{13}$$

we concatenate h_{ij}^t and h_{ij}^t to obtain the representation $h_{ij}^t = [\overrightarrow{h_{ij}}, \overleftarrow{h_{ij}}]$ of the t-th path of r_i and d_j.

Path Attention

To distinguish the different contributions of multiple paths from r_i to d_j to their associated predictions, we introduce attention mechanisms to distinguish the importance of the path. The total path information g_{ij} is formulated as the weighted sum of all paths, and it is expressed as:

$$g_{ij} = \sum \alpha_{ij}^t h_{ij}^t, \tag{14}$$

where h_{ij}^t is the representation vector of the t-th path of r_i to d_j, and α_{ij}^t is the attention weight of h_{ij}^t to measure the importance of the t-th path. We introduce a path vector u_p to measure the importance of the path. The attention weight of each path can be defined as:

$$u_{ij}^t = \tan h\left(W_t h_{ij}^t + b_t\right), \tag{15}$$

$$\alpha_{ij}^t = \frac{\exp\left((u_p)^T u_{ij}^t\right)}{\sum_t \exp\left((u_p)^T u_{ij}^t\right)}, \tag{16}$$

where u_{ij}^t is the score function of the corresponding path, i.e., the score of the import of the path, W_t is the weight vector, b_t is the bias vector, α_{ij}^t is the attention weight of the t-th path, u_p is the weight vector, and $(u_p)^T$ indicated its transposition.

2.3.3. Combined Strategy

To fully combine the representation of the left-path node pair r_1 and d_3 and path information representation of the right path, we design a combined strategy for determining the association score of r_1 and d_3. We added a SoftMax classifier to ensure that left and right paths have certain predictive capabilities and to further improve the performance of predictive classification. The corresponding loss is defined as:

$$score_c = softmax\left(W_c c_{ij} + b_c\right), \tag{17}$$

$$loss_1 = y_{real} log score_c^0 + (1 - y_{real}) \log score_c^1, \tag{18}$$

$$score_g = softmax\left(W_v g_{ij} + b_v\right), \tag{19}$$

$$loss_2 = y_{real} log score_g^0 + (1 - y_{real}) \log score_c^1, \tag{20}$$

where c_{ij} is a representational learning method based on CNN learning drug r_i and disease d_j. g_{ij} is the representation obtained by learning on the right, W_c and W_v are the weight matrices of the left and right parts, respectively, b_c and b_v are the offset vectors, y_{real} is the actual correlation between the drug and the disease. Further, 1 means the drug is associated with the disease, and 0 is the unknown association, where $score_c^0$ indicates that there is no possibility of association between drug r_i and disease d_j, and $score_c^1$ indicates that there is no possibility of association between drug r_i and disease d_j. Finally, $loss_1$ and $loss_2$, are the cross entropy losses of the model in the probability of prediction and the true correlation value. The final loss function of our model is the weighted sum of $loss_1$ and $loss_2$:

$$loss = \alpha_1 loss_1 + (1 - \alpha_1) loss_2. \tag{21}$$

where α_1 is a super parameter, which is used to weigh the contribution of $loss_1$ and $loss_2$. Our final score is:

$$score = \alpha_1 score_c + (1 - \alpha_1) score_g. \tag{22}$$

2.3.4. Reducing Overfitting

Our neural network has nearly 50 million parameters, which turns out to too many parameters to learn without considerable overfitting. Thus, we introduce the following measures to prevent overfitting.

Dropout

Integrating the result from many different models is an excellent method to reduce test errors [33,34], but this method is too computationally expensive for large neural networks and takes several days to train. There is, however, a very efficient approach to model combination that only spends a factor of about two during training. The recently presented technique, called "dropout" [35], consists of setting the output of each hidden neuron to zero with probability 0.5. The neurons that are "dropped out" in this way do not participate in the forward pass and back-propagation. Thus, every time an input is presented, the neural network samples a different architecture, but all these architectures share weights. This technique reduces intricate co-adaptations of neurons, because a neuron cannot depend on the existence of other specific neurons. Therefore, it is forced to learn more robust, beneficial features in conjunction with many different random subsets of the other neurons. During the test, we multiply the output of all the neurons by 0.5, which reasonably approximates the geometric mean of the predictive distributions produced by the exponentially many dropout networks.

3. Results and Discussion

3.1. Evaluation Metrics

In this study, we applied five-fold cross-validation analysis to evaluate the performance of our method. All known drug-disease associations were treated as positive samples and divided randomly into five equal positive subsets. At the same time, unknown associations with a matching number were randomly selected and divided into five negative subsets. In each fold, four positive subsets and four negative subsets were selected for training and the remaining were used to testing. We trained the prediction model based on known associations in the training set and predicted associations in the testing set. Training and testing were repeated five times, and the average of the performance was adopted. In addition, we calculated the drug similarity each time we selected four positive samples. Then, the testing set for each drug was ranked; the higher the candidate disease ranked, the greater was the possibility of association between the drug and the disease.

The CGARDP model was used to obtain the test scores of the associations in the testing set. The scores were ranked in the descending order of the scores, given a threshold θ. If the scores were higher than θ, they were considered as positive samples, and those below θ were considered as negative samples. We calculate different true positive rates (TPRs), false positive rates (FPRs), accuracy (precisions), and recall (recall) in each θ as follows

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{TN + FP}, \tag{23}$$

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN} \tag{24}$$

where TP indicates the correct identification of the number of positive samples, TN indicates the correct identification of the number of negative samples, FP indicates the number of samples that will be predicted as a positive example, and FN indicates the number of samples identified as a negative sample. Thus, the receiver operating characteristic (ROC) curve [36] can be drawn using different TPRs and FPRs under different θ. The area under the curve (AUC) is called the drug-related AUC value. The average AUC of all drugs was used to assess the overall performance of our method. Because the ratio of positive and negative samples is 1:169, there is a large class imbalance. The class imbalance problem is concerned with positive cases, while the two indicators of the PR curve are focused on

positive samples; therefore, the PR curve has more credibility than the ROC curve [1]. Thus, we used the PR curve to measure the performance at the same time. Precision is defined as the percentage of real samples that are determined as positive samples, and recall as the percentage of true samples to the total number of actual positive samples.

In addition, biologists always choose to arrange higher-ranking candidate diseases for biological verifications, and therefore, the top of the ranking candidate list must have more positive samples. Therefore, we made another evaluation criterion a performance metric, i.e., we calculated the average recall rate of top-k (k = 30, 60, 90, 120 . . .). The higher the recall rate, the higher is the proportion of drug-related diseases that are correctly retrieved; further, the better the predictive performance, the higher is the positive sample that is successfully identified.

3.2. Comparison with Other Methods

To evaluate the performance of the CGARDP model, we compared it with several state-of-the-art methods including HGBI [37], MBIRW [24], LRSSL [23], and SCMFDD [25]. HGBI builds a three-layer heterogeneous network that uses a combination of drug, disease, and target for prediction. MBIRW builds a two-layer network of drugs and diseases to complete the drug reposition by walking among the drug-disease network. LRSSL, a Laplacian regularized sparse subspace learning method, combines the chemical substructure of the drug, the target domain, and the target annotation for prediction. SCMFDD calculates the Jaccard similarity of the chemical substructure of the drug and the semantic similarity of the disease to predict novel drug-disease association using matrix factorization.

For CGARDP and several other comparison methods, each method must adjust the parameters involved to optimize the prediction performance. In our method, the left convolutional neural network active windows W_f and W_h are 3 and 20, respectively. It has two convolutional layers; the first of contains 16 convolution kernels, and the second contains 32 convolution kernels, that is, n_{conv} is 16 and 32. The padding parameter P is (1,10). The size of the sampling window (W_m,W_p) is set to (2,2), and the super participation α_1 is 2. For fairness, the parameters of other methods are based on the parameters recommended in the corresponding literature (α = 0.4 for HGBI, α = 0.3 for MBIRW, $\mu = 0.01, \lambda = 0.01$ for LRSSL, $\mu = 2^0, \lambda = 2^2$ for SCMFDD).

As shown in Figure 4A and Table 1, CGARDP achieves the best average performance over all 763 drugs that we considered (AUC of ROC curve = 0.956). The AUC-ROC values of other methods, i.e., HGBI, MBIRW, LRSSL, and SCMFDD for 763 drugs are 0.683, 0.837, 0.838, and 0.726, respectively. In particular, CGARDP outperforms HGBI by 27.3%, MBIRW by 11.9%, LRSSL by 11.8%, and SCMFDD by 23%. Further, we list the AUCs of all five methods on 15 well characterized human drugs, each of which has more than 15 known related diseases. CGARDP yields the best average performance in terms of AUCs and achieves the best performances for 11 of the 15 common drugs. Among all methods, LRSSL performed second best, and LRSSL took full advantage of the multiple similarity of drugs. MBIRW achieved almost the same effect as LRSSL on AUC; however, it performance was less than LRSSL by 7% on AUPR. These differences in performance are possibly because MBIRW focuses on the topology information of the network. SCMFDD and HGBI perform considerably worse than LRSSL and MBIRW; however, SCMFDD performs 4.5% better than HGBI. This difference can be attributed to the fact that SCMFDD relies on the calculation of similarity, while HGBI constructs a three-layer network that introduces drug–protein information but does not make full use of this information. Compared with other methods, the superiority of CGARDP is due to its in-depth understanding of the node representation of the drug–disease association and the attentional representation of the path representation.
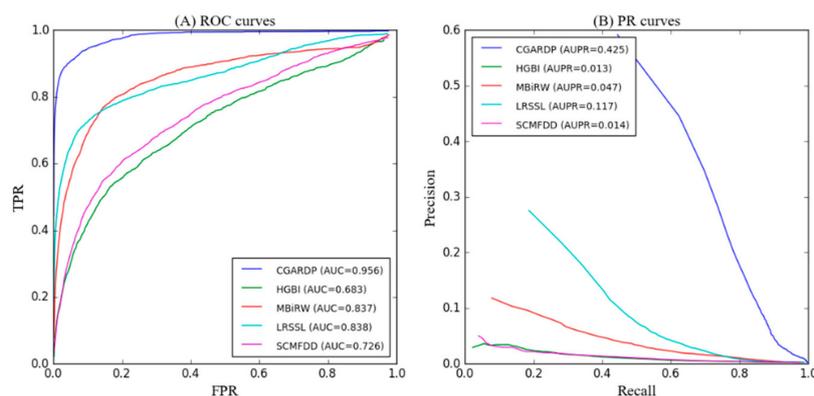
Figure 4. (**A**) Receiver operating characteristic (ROC) curves and (**B**) positive rate (PR) curves of CGARDP and other methods for all drugs.

Table 1. AUCs of CGARDP and other methods for all of the drugs and 15 well characterized drugs.

Drug Name	AUC				
	CGARDP	HGBI	MBiRW	LRSSL	SCMFDD
ampicillin	0.964	0.751	0.932	0.962	0.895
cefepime	0.990	0.910	0.970	0.971	0.914
cefotaxime	0.958	0.917	0.929	0.950	0.953
cefotetan	0.973	0.808	0.918	0.948	0.848
cefoxitin	0.880	0.890	0.912	0.979	0.894
ceftazidime	0.938	0.845	0.931	0.936	0.922
ceftizoxime	0.929	0.960	0.961	0.923	0.962
ceftriaxone	0.999	0.945	0.898	0.955	0.811
ciprofloxacin	0.905	0.811	0.813	0.928	0.820
doxorubicin	0.951	0.487	0.921	0.727	0.460
erythromycin	0.948	0.827	0.887	0.918	0.764
itraconazole	0.956	0.445	0.877	0.845	0.730
levofloxacin	0.898	0.943	0.975	0.964	0.872
moxifloxacin	0.992	0.812	0.948	0.957	0.932
ofloxacin	0.980	0.902	0.943	0.904	0.774
Average AUC	0.956	0.683	0.837	0.838	0.726

Because the number of unknown drug-disease associations far exceeds the known associations, there is a serious imbalance in data. The PR curve predicts performance metrics better than the ROC curve when there is a serious imbalance between the positive and negative samples. Figure 4B and Table 2 shows the AUPR for the average performance of all drugs, and CGARDP produces the best average performance on these drugs (AUC of PR curve = 0.425). Its average AUPR is 41.3%, 37.8%, 30.8%, and 41.1% higher than those of HGBI, MBIRW, LRSSL, and SCMFDD, respectively. For the 15 well-characterized drugs, CGARDP demonstrates the best performance for 11 of these drugs. In addition, 265 diseases were only association with one drug, and 116 diseases were associations with two drugs. Therefore, CGARDP can be used for diseases associated with only one or two drugs.

For all the prediction results on 763 drugs, we performed a Wilcoxon test to evaluate whether CGARDP's performance is significantly better than that of the other methods. The statistical results (Table 3) indicate that CGARDP yields the significantly better performance under the p-value threshold of 0.05 in terms of not only AUCs but also AUPRs.

Table 2. AUPRs of CGARDP and other methods for all of the drugs and 15 well characterized drugs.

Drug Name	AUPR				
	CGARDP	HGBI	MBIRW	LRSSL	SCMFDD
ampicillin	0.515	0.032	0.023	0.285	0.068
cefepime	0.766	0.163	0.315	0.625	0.054
cefotaxime	0.525	0.071	0.292	0.283	0.105
cefotetan	0.496	0.054	0.197	0.512	0.059
cefoxitin	0.420	0.151	0.394	0.286	0.065
ceftazidime	0.591	0.032	0.201	0.488	0.694
ceftizoxime	0.472	0.212	0.244	0.455	0.096
ceftriaxone	0.607	0.056	0.223	0.673	0.077
ciprofloxacin	0.429	0.082	0.118	0.280	0.064
doxorubicin	0.520	0.005	0.051	0.180	0.004
erythromycin	0.592	0.023	0.038	0.144	0.022
itraconazole	0.379	0.006	0.253	0.042	0.008
levofloxacin	0.212	0.136	0.071	0.539	0.098
moxifloxacin	0.735	0.049	0.650	0.384	0.088
ofloxacin	0.382	0.091	0.130	0.201	0.078
Average AUC	0.425	0.013	0.047	0.117	0.014

Table 3. The statistical result of the paired Wilcoxon test on the AUCs of 763 drugs comparing CGARDP and all of four other methods.

p-Value between CGARDP and Another Method	HGBI	MBiRW	LRSSL	SCMFDD
p-value of ROC curve	6.873×10^{-270}	6.302×10^{-72}	3.473×10^{-31}	9.326×10^{-180}
p-value of PR curve	4.365×10^{-40}	7.332×10^{-30}	2.321×10^{-12}	3.265×10^{-60}

A higher recall rate on top k ranked drugs means that real disease-related drugs are correctly identified. The average recall rates of the top k samples on all 763 drugs are shown in Figure 5. CGARDP consistently outperforms the other methods at various k values, and it ranked 89.9% in the top 30, 93.8% in the top 60, and 97.1% in the top 120. Before the top 90, LRSSL performed better than MBiRW, and then MBiRW surpassed LRSSL. The former ranks 63.4%, 71.3%, and 77.7% in the top 30, 60, and 120, respectively, and the latter is 53.1% and 66.3%. 79.3%. The possible reason for these different rankings is that MBiRW makes better use of global topology information, while LRSSL focuses more on neighbor node information. HGBI and SCMFDD have relatively close recall rates at different k values. HGBI ranks for k values of 30, 60, and 120 were 28.8%, 41.1%, and 54.9%, respectively, and those of SCMFDD are 30.6%, 45.0%, and 57.8%. Ultimately, we can conclude that CGARDP is indeed better than other methods in discovering the underlying disease of the drug.
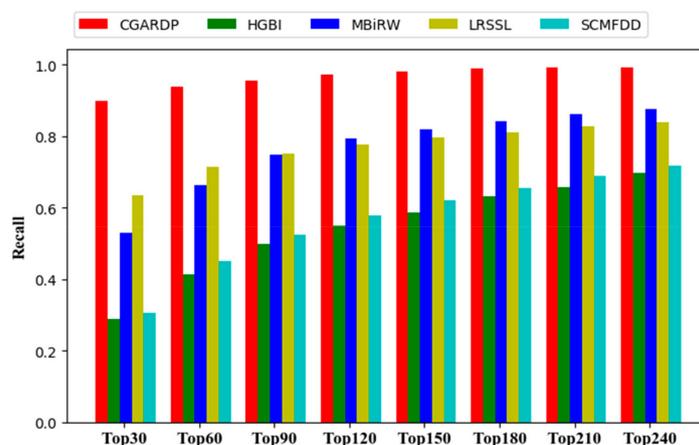
Figure 5. Recalls across all the tested drugs at different top k cutoffs.

3.3. *Case Studies on Ciprofloxacin, Ceftriaxone, Ofloxacin, Ampicillin, and Levofloxacin*

After the above five-fold cross-validation, we evaluated the performance of the method, and all known correlation data were used as training data to predict the unknown drug-disease association. Case studies of five drugs—Ciprofloxacin, Ceftriaxone, Ofloxacin, Ampicillin, and Levofloxacin—demonstrate the ability of CGARDP to detect high-quality candidate diseases for drugs. The analysis of each of the top ten candidates for each drug is presented in detail in Table 4.

Table 4. The top 10 candidates related to the drugs Ciprofloxacin, Ceftriaxone, Ofloxacin, Ampicillin, and Levofloxacin.

Drug Name	Rank	Disease Name	Description	Rank	Disease Name	Description
Ciprofloxacin	1	Conjunctivitis, Bacterial	Clinical Trials	6	Gram-Negative Bacterial Infections	Clinical Trials
	2	Campylobacter Infections	CDC	7	Chlamydia Infections	Clinical Trials
	3	Anthrax	CTD, Clinical Trials	8	Pneumonia, Pneumocystis	PubChem
	4	Klebsiella Infections	CTD, Clinical Trials	9	Eye Infections, Bacterial	Clinical Trials
	5	Soft Tissue Infections	Clinical Trials	10	Acanthamoeba Keratitis	PubChem
Ceftriaxone	1	Bone Diseases, Infectious	Clinical Trials	6	Tetanus	literature [38]
	2	Panic Disorder	Drug Bank	7	Legionnaires Disease	Drug Bank
	3	Hepatitis B	Clinical Trials	8	Cytomegalovirus Infections	Drug Bank
	4	Respiratory Syncytial Virus Infections	PubChem	9	Respiration Disorders	Clinical Trials
	5	Maxillary Sinusitis	Drug Bank	10	Respiratory Distress Syndrome, Adult	Clinical Trials
Ofloxacin	1	Corneal Ulcer	PubChem	6	Proteus Infections	CTD
	2	Epididymitis	CDC	7	Urinary Bladder Neck Obstruction	Inferred candidate by 1 literature
	3	Otitis Externa	Drug Bank	8	Glaucoma, Angle-Closure	PubChem
	4	Tuberculosis, Pulmonary	CTD, clinical Trials	9	Urinary Bladder Diseases	Inferred candidate by 1 literature
	5	Urethral Diseases	PubChem	10	Trichomonas Vaginitis	clinical Trials
Ampicillin	1	Burns	Inferred candidate by 3 literature	6	Candidiasis, Cutaneous	PubChem
	2	Meningitis, Bacterial	CTD	7	Otitis Media, Suppurative	Drug Bank
	3	Pseudomonas Infections	CTD	8	Pneumonia, Bacterial	CTD, Clinical Trials
	4	Skin Diseases, Infectious	Clinical Trials	9	Proteus Infections	CTD
	5	Radiation Injuries, Experimental	Inferred candidate by 1 literature	10	Sarcoma, Ewings	Drug Bank
Levofloxacin	1	Tuberculosis, Pulmonary	Clinical Trials	6	Listeriosis	Drug Bank
	2	Histoplasmosis	Drug Bank	7	Soft Tissue Infections	CTD, Clinical Trials
	3	Pneumonia, Mycoplasma	Clinical Trials	8	Respiratory Tract Fistula	Drug Bank
	4	Bronchitis	Clinical Trials	9	Rhinitis	Drug Bank
	5	AIDS-Related Opportunistic Infections	Clinical Trials	10	Mouth Diseases	Clinical Trials

First, A drug bank is a database of drugs pharmacology indication, drug interaction, and clinical trials for a disease. The Comparative Toxicogenomics Database (CTD) contains important information about the effects of drugs on the disease. The Centers for Disease Control and Prevention (CDC) records the trends and preventive treatments of common diseases. In Table 4, 12 candidate diseases are included from the drug bank, nine candidates are included in the CTD, and two candidates are included in the CDC; this table shows that these candidate diseases are indeed related to the corresponding

drugs. Second, ClinicalTrials.gov (https://clinicaltrials.gov/) is a database of clinical trials run by the National Institutes of Health (NIH), and it contains clinical trials of various drugs and related diseases. PubChem (https://pubchem.ncbi.nlm.nih.gov/) is a database of chemical modules supported by the NIH, and it stores biochemical experimental data and structural information on compounds, including drugs and their biological activities data. A total of 21 candidate diseases in Table 4 were included in ClinicalTrials.gov, and 7 candidates were included in PubChem, indicating that these candidates were supported by the experiment. In addition, a candidate for the "literature" marker was supported by the literature. The addition of ceftriaxone to metronidazole has a synergistic effect, which can reduce the production of toxins and promote wound healing; thus, the combination of metronidazole and ceftriaxone is preventive. Tetanus patients with sepsis and pneumonia have good efficacy, confirming that Ceftriaxone affects the candidate disease tetanus.

In addition, the CTD database also contains potential associations that the literature infers to exist, labelled as Inferred. Four candidate diseases in Table 4 were inferred from the CTD literature, indicating that the drug is more likely to be associated with the candidate disease. Case studies of candidate diseases for the five drugs confirmed that CGARDP was indeed able to detect potential candidate diseases for the drug.

3.4. Prediction of Novel Drug–Disease Associations

According to cross validation and case studies, we applied CGARDP to predict the novel drug–disease associations. All known drug–disease associations were utilized to train CGARDP's prediction model, the potential candidate associations were then obtained by using the model as listed in Supplementary Table S1.

4. Conclusions

A novel method based on CNN and GRU—CGARDP—was proposed to predict the potential drug–disease associations. The CRU based framework deeply integrates the similarity and association information of a drug–disease pair. The GGU based framework deeply learns the path information between the drug and the disease. CGARDP discriminates different contributions of the paths by constructing the attention mechanism and learns more informative representation of the drug-disease pair. The experimental results show that CGARDP outperforms other methods in terms of both AUCs and AUPRs. The case studies on five drugs confirm that CGARDP is able to retrieve potential candidate drug–disease associations.

References

1. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240.
2. Mullard, A. 2014 FDA drug approvals. *Nat. Rev. Drug Discov.* **2015**, *14*, 77–81. [CrossRef] [PubMed]

3. Fashoyin-Aje, L.; Donoghue, M.; Chen, H.; He, K.; Veeraraghavan, J.; Goldberg, K.B.; Keegan, P.; McKee, A.E.; Pazdur, R. FDA Approval Summary: Pembrolizumab for Recurrent Locally Advanced or Metastatic Gastric or Gastroesophageal Junction Adenocarcinoma Expressing PD-L1. *Oncologist* **2019**, *24*, 103–109. [CrossRef] [PubMed]

4. Dickson, M.; Gagnon, J.P. Key factors in the rising cost of new drug discovery and development. *Nat. Rev. Drug Discov.* **2004**, *3*, 417–429. [CrossRef] [PubMed]

5. Ellis, P.; Tamimi, N. Drug Development: From Concept to Marketing! *Nephron Clin. Pr.* **2009**, *113*, c125–c131.

6. Pushpakom, S.; Iorio, F.; Eyers, P.A.; Escott, K.J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; Mcnamee, C. Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **2019**, *18*, 41. [CrossRef] [PubMed]

7. Alfedi, G.; Luffarelli, R.; Condò, I.; Pedini, G.; Mannucci, L.; Massaro, D.S.; Benini, M.; Toschi, N.; Alaimo, G.; Panarello, L.; et al. Drug repositioning screening identifies etravirine as a potential therapeutic for friedreich's ataxia. *Mov. Disord.* **2019**, *34*, 323–334. [CrossRef] [PubMed]

8. Tobinick, E. The value of drug repositioning in the current pharmaceutical market. *Drug News Perspect.* **2009**, *22*, 119. [CrossRef]

9. Ashburn, T.T.; Thor, K.B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673–683. [CrossRef]

10. Suthram, S.; Dudley, J.T.; Chiang, A.P.; Chen, R.; Hastie, T.J.; Butte, A.J. Network-Based Elucidation of Human Disease Similarities Reveals Common Functional Modules Enriched for Pluripotent Drug Targets. *PLoS Comput. Boil.* **2010**, *6*, e1000662. [CrossRef]

11. Chiang, A.P.; Butte, A.J.; Chiang, A.P.; Butte, A.J.; Chiang, A.; Butte, A. Systematic Evaluation of Drug–Disease Relationships to Identify Leads for Novel Drug Uses. *Clin. Pharmacol. Ther.* **2009**, *86*, 507–510. [CrossRef]

12. Bamshad, M.J.; Ng, S.B.; Bigham, A.W.; Tabor, H.K.; Emond, M.J.; Nickerson, D.A.; Shendure, J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **2011**, *12*, 745–755. [CrossRef] [PubMed]

13. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141. [CrossRef] [PubMed]

14. Yang, D.; Zhang, Y.; Nguyen, H.G.; Koupenova, M.; Chauhan, A.K.; Makitalo, M.; Jones, M.R.; Hilaire, C.S.; Seldin, D.C.; Toselli, P.; et al. The A 2B adenosine receptor protects against inflammation and excessive vascular adhesion. *J. Clin. Investig.* **2006**, *116*, 1913–1923. [CrossRef] [PubMed]

15. Ghofrani, H.A.; Osterloh, I.H.; Grimminger, F. Sildenafil: From angina to erectile dysfunction to pulmonary hypertension and beyond. *Nat. Rev. Drug Discov.* **2006**, *5*, 689–702. [CrossRef] [PubMed]

16. Campillos, M.; Kuhn, M.; Gavin, A.-C.; Jensen, L.J.; Bork, P. Drug Target Identification Using Side-Effect Similarity. *Science* **2008**, *321*, 263–266. [CrossRef] [PubMed]

17. Sardana, D.; Zhu, C.; Zhang, M.; Gudivada, R.C.; Yang, L.; Jegga, A.G. Drug repositioning for orphan diseases. *Briefings Bioinform.* **2011**, *12*, 346–356. [CrossRef] [PubMed]

18. Cheng, F.; Liu, C.; Jiang, J.; Lu, W.; Li, W.; Liu, G.; Zhou, W.-X.; Huang, J.; Tang, Y. Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference. *PLoS Comput. Boil.* **2012**, *8*, e1002503. [CrossRef] [PubMed]

19. Zhao, S.; Li, S. A co-module approach for elucidating drug-disease associations and revealing their molecular basis. *Bioinformatics* **2012**, *28*, 955–961. [CrossRef] [PubMed]

20. Wang, F.; Zhang, P.; Cao, N.; Hu, J.; Sorrentino, R. Exploring the associations between drug side-effects and therapeutic indications. *J. Biomed. Inform.* **2014**, *51*, 15–23. [CrossRef] [PubMed]

21. Gottlieb, A.; Stein, G.Y.; Ruppin, E.; Sharan, R. PREDICT: A method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **2011**, *7*, 496. [CrossRef] [PubMed]

22. Zhang, P.; Wang, F.; Hu, J. Towards drug repositioning: A unified computational framework for integrating multiple aspects of drug similarity and disease similarity. In Proceedings of the AMIA Annual Symposium Proceedings, Washington, DC, USA, 15–19 November 2014; pp. 1258–1267.

23. Liang, X.; Zhang, P.; Yan, L.; Fu, Y.; Peng, F.; Qu, L.; Shao, M.; Chen, Y.; Chen, Z. LRSSL: Predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics* **2017**, *33*, 1187–1196. [CrossRef] [PubMed]

24. Luo, H.; Wang, J.; Li, M.; Peng, X.; Wu, F.-X.; Pan, Y. Drug repositioning based on comprehensive similarity measures and Bi-Random Walk algorithm. *Bioinformatics* **2016**, *32*, 2664–2671. [CrossRef] [PubMed]

25. Zhang, W.; Yue, X.; Lin, W.; Wu, W.; Liu, R.; Huang, F.; Liu, F. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinform.* **2018**, *19*, 233. [CrossRef] [PubMed]

26. Xuan, P.; Cao, Y.; Zhang, T.; Wang, X.; Pan, S.; Shen, T. Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics* **2019**, *13*. [CrossRef] [PubMed]

27. Wang, Y.; Xiao, J.; O Suzek, T.; Zhang, J.; Wang, J.; Bryant, S.H. PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, *37*, W623–W633. [CrossRef] [PubMed]

28. Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644–1650. [CrossRef] [PubMed]

29. Cheng, D.; Gong, Y.; Zhou, S.; Wang, J.; Zheng, N. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In Proceedings of the IEEE Conference on Computer Vision and Pattern recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1335–1344.

30. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010.

31. Elfwing, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **2018**, *107*, 3–11. [CrossRef] [PubMed]

32. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.

33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

34. Bell, R.M.; Koren, Y. Lessons from the Netflix prize challenge. *ACM SIGKDD Explor. Newsl.* **2007**, *9*, 75. [CrossRef]

35. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. In Proceedings of the International Conference on Learning Representations, Tsukuba, Japan, 11–15 November 2012.

36. Xuan, P.; Sun, C.; Zhang, T.; Ye, Y.; Shen, T.; Dong, Y. Gradient Boosting Decision Tree-Based Method for Predicting Interactions Between Target Genes and Drugs. *Front. Genet.* **2019**, *10*, 10. [CrossRef] [PubMed]

37. Wang, W.; Yang, S.; Zhang, X.; Li, J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* **2014**, *30*, 2923–2930. [CrossRef] [PubMed]

38. Brock, H.; Moosbauer, W.; Gabriel, C.; Necek, S.; Bidal, D. Treatment of severe tetanus by continuous intrathecal infusion of baclofen. *J. Neurol. Neurosurg. Psychiatry* **1995**, *59*, 193–194. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds are not available from the authors.