DeepDrug: A general graph-based deep learning framework for drug relation prediction

Xusheng Cao[1⋆], Rui Fan[1⋆], and Wanwen Zeng[1,2⋆⋆]

[1] College of Software, Nankai University, Tianjin, 300350, China.
[2] Department of Statistics, Stanford University, Stanford, CA 94305, USA.

Abstract. Computational approaches for accurate predictions of drug-related interactions such as drug-drug interactions (DDIs) and drug-target interactions (DTIs) are highly demanding for biochemical researchers due to their efficiency and cost-effectiveness. Despite the fact that many methods have been proposed and developed to predict DDIs and DTIs respectively, their success is still limited due to a lack of systematic evaluation of the intrinsic properties embedded in their structure. In this paper, we develop a deep learning framework, named DeepDrug, to overcome these shortcomings by using graph convolutional networks to learn the graphical representations of drugs and proteins such as molecular fingerprints and residual structures in order to boost the prediction accuracy. We benchmark our methods in binary-class DDIs, multi-class DDIs and binary-class DTIs classification tasks using several datasets. We then demonstrate that DeepDrug outperforms other state-of-the-art published methods both in terms of accuracy and robustness in predicting DDIs and DTIs with varying ratios of positive to negative training data. Ultimately, we visualize the structural features learned by DeepDrug, which display compatible and accordant patterns in chemical properties, providing additional evidence to support the strong predictive power of DeepDrug. We believe that DeepDrug is an efficient tool in accurate prediction of DDIs and DTIs and provides a promising path in understanding the underlying mechanism of these biochemical relations. The source code of the DeepDrug can be downloaded from https://github.com/wanwenzeng/deepdrug.

Keywords: drug-drug interactions · drug-target interations · deep learning · graph convolutional neural network.

Introduction

The search for biomedical relations between chemical compounds (drugs, molecules) and protein targets is an important part of drug discovery [5]. At the fundamental level, drugs interact with biological systems by binding with protein targets and affecting their downstream activity. Prediction of Drug-Target Interactions (DTIs) is thus important for identification of therapeutic targets or characteristics of drug targets. Knowledge of DTIs also provide a key towards understanding and predicting higher-level information such as side effects, therapeutic mechanisms and even novel insights for drug repositioning or drug repurposing [47]. For instance, Sildenafil was initially developed to treat pulmonary hypertension, but identification of its side effects allowed it to be repositioned for treating erectile dysfunction [6]. In addition, since most human diseases are complex biological processes that are resistant to the activity of any single drug [20][15], polypharmacy has become a promising strategy among pharmacists. Prediction and validation of Drug-Drug Interactions (DDIs) can sometimes reveal potential synergies in drug combinations to improve therapeutic efficacy of individual drugs [37]. More importantly, negative DDIs are major causes of adverse drug reactions (ADRs) [24], especially among the elderly who are more likely to take multiple medications [13]. Critical DDIs have resulted in the withdrawal of drugs from market, such as withdrawal of mibefradil and cerivastatin from the US market [29][36]. Hence, early detection of negative DDIs or undesirable toxicity can ensure drug safety and prevent further investment of resources in non-viable entities.

Over the past decade, the emergence of various biochemical databases, such as DrugBank [42], TwoSides [38], RCSB Protein Data Bank [7], PubChem [21], has provided convenient reference for

⋆ These authors contributed equally to this work.
⋆⋆ Corresponding author: wwzeng@nankai.edu.cn

DTIs and DDIs for health professionals. However, prediction of novel biochemical interactions still remains a challenging task. *In vitro* experimental techniques are reliable but expensive and time-consuming. *In silico* approaches have received far more attention due to their cost-effectiveness and increasing accuracy in relation predictions. The state-of-the-art computational methods for interaction prediction rely on machine learning algorithms that incorporate large-scale biochemical data. Most of these efforts are based on the principle that similar drugs tend to share similar target proteins and vice versa [40]. Hence, the most popular framework formulates the prediction of DTIs and DDIs as a classification task and uses some forms of similarity function as inputs [43]. Deep learning based methods that make use of different feature extraction techniques in conjunction with various neural network architectures have also been explored by researchers, such as DeepDDI [33] for DDI predictions and DeepDTA for DTI predictions [31]. Another common approach is to construct a heterogeneous network in the chemogenomics space to predict potential interactions using random walks [46]. Efficient relation extraction from biomedical research articles has also been automated by applying natural language processing (NLP) techniques on large amount of relevant text corpus [18]. The rise of machine learning methods and their integration with biomedical science has promoted drug-related research tremendously in the last two decades. The readers are referred to [45][2] and references therein for detailed reviews on the topic.

In spite of these advances, there is still room for improvement in several aspects. First of all, the accurate prediction of unseen biomedical relations depends heavily on the feature extraction technique or similarity kernel used. Since different forms of feature extraction or similarity kernel introduce varying amount of human-engineered bias, they often give different levels of performance depending on the context and no single kernel outperforms others universally [3]. Similarity-based methods also have difficulty applying on large-scale dataset due to the significant computational complexity of measuring similarity matrices [30]. Network-based methods built upon topological properties of the multipartite graph suffer from the same problem to different extents depending on the complexity of the graph [26]. NLP approaches by emphasizing on semantics in textual data have achieved limited success in capturing the underlying biochemical reaction principles caused by intrinsic chemical or genomics structural properties though addition of structure-based information can improve the performance of the predictive model [1].

In recent years, deep learning frameworks based on variants of graph neural networks such as graph convolutional network (GCNs) [23], graph attention networks (GATs) [39], gated graph neural networks (GGNNs) [25] have demonstrated ground-breaking performance on social science, natural science, knowledge graphs and many other research areas. In particular, GCNs have been applied to various biochemical problems such as molecular fingerprints learning with each node in the graphical model corresponding to an atom and each edge representing a chemical bond [10], as well as protein classification problems with each node describing a residual and each edge characterizing the distances between nodes [12]. As pharmacological similarities and genomics similarities arise mainly from structural properties, graphical representations of biochemical entities have shown capability of better capturing the structural features than Euclidean ones without requiring feature engineering [44][11].

Based on these observations, we propose DeepDrug, a graph-based deep learning framework, to learn drug relations such as pairwise DDIs or DTIs. The proposed model differs from previous drug relation prediction methods in the following aspects: 1) By taking advantage of the natural graph representation of drugs and proteins, DeepDrug requires only graph representations of drugs and proteins as input to learn structural features; 2) DeepDrug utilizes GCN modules to capture the intrinsic structure between atoms of a compound and residues of a protein. Comprehensive experiments on different benchmark datasets demonstrate that DeepDrug can successfully learn both DDIs and DTIs from graphical features in different tasks such as binary classification and multi-class classification and outperform other state-of-the-art models. In addition, we construct additional datasets with different ratios of positive and negative data to further validate the robustness of the model. By visualization techniques and computing Dice similarity scores among drugs in our study, we also demonstrate the effectiveness of the graphical model in learning structural information that are not explicitly introduced into the prediction framework.

In summary, the following contributions are made in this paper:

- DeepDrug provides a unified framework based on GCNs to extract structural features for both drugs and proteins for downstream DDIs and DTIs prediction. Compared to hand-crafted

features (e.g., molecular fingerprints) or string-based features (e.g., SMILES sequence [41]), the novel design of DeepDrug architecture can automatically capture structural features by considering the interactions among nodes and bonds in the input graphs.

– DeepDrug achieves state-of-the-art performance in DDIs and DTIs prediction tasks. Through comprehensive experiments, including binary-class classification of DDIs, multiple-class classification of DDIs and binary-class classification of DTIs, we demonstrate the superior performance of DeepDrug, highlighting the strong and robust predictive power of graph presentation strategy and GCN architecture.
– The visualization of structural features of DeepDrug proves the key insight that biomedical structure may determine their function and drugs with similar structures tend to have similar targets. These results suggest that DeepDrug can be a useful tool for effectively modeling DDIs and DTIs and greatly facilitate the drug discovery process.

Results

Overview of DeepDrug

We formulate the DDI and DTI prediction problem as a classification task. A key insight of our framework is that biochemical interactions are primarily determined by the structure of the participating entities and thence the performance of the predictive model depends ultimately on the accurate characterization of the structural information. Since chemical structure of drugs can be naturally represented as graphs (with nodes and edges denoting chemical atoms and bonds respectively) and protein structures also has a logical graph representation (with nodes and edges representing amino acids and biochemical interactions respectively), it is intuitive to employ graphical models for our predictive modeling.

The architecture of our proposed DeepDrug model is shown in Fig. 1. The model takes two inputs: 1) the drug's SMILES string and 2) target protein's PDB data or another drug's SMILES string. DeepDrug then partitions the prediction tasks into two stages. First, we extract graph representation of the biochemical entities. To achieve this, we use DeepChem [32] for drugs to convert their SMILES strings into graph representations in the form of feature matrices and adjacency matrices and ProteinGraph [27] for proteins to convert their PDB data into similar graph representations. Next, these graph representations are fed into GCNs for training. After GCN layers, we concatenate two embedding vectors and pass them into a dense layer with a sigmoid or softmax activation function to obtain the final prediction. More details of the DeepDrug framework as well as the datasets used in this study can be found in *Methods*.

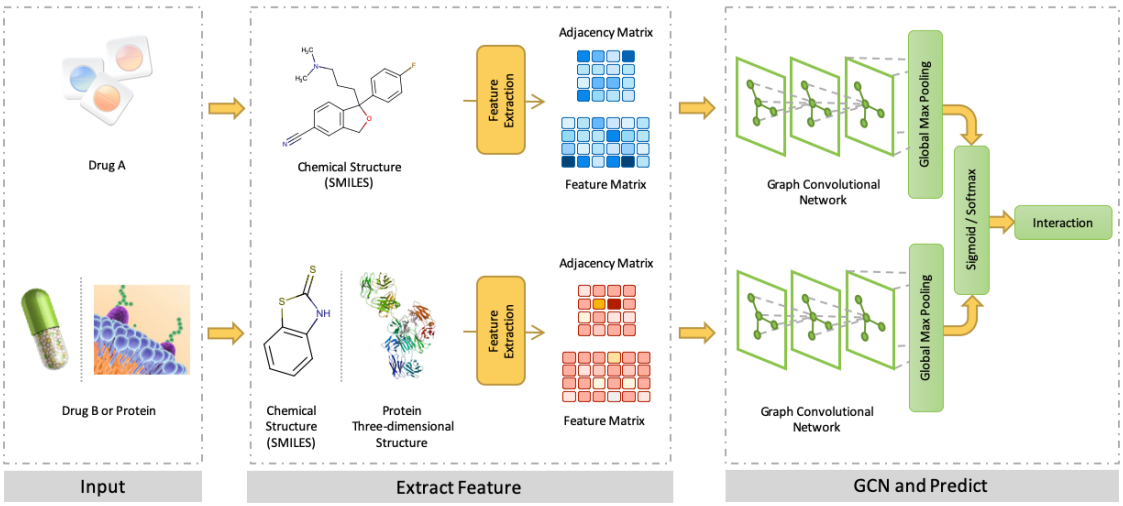

Fig. 1: Overview of DeepDrug.

DeepDrug yields accurate DDI prediction

Since chemical drugs are small compounds that are generally easier to be processed into graphs with little ambiguity, we first evaluate the performance of DeepDrug for DDI prediction in a binary classification setting. We benchmark DeepDrug against a baseline method using random forest classification (RFC) as well as another deep learning methods, DeepDDI [33]. The baseline method takes in graph representations as input whereas the original framework of DeepDDI only takes in SMILES string. Three different sets of data are employed (see *Methods*). Our analysis shows that DeepDrug consistently outperforms other existing methods, with 13.2% higher AUROC (Area Under Receiver Operating Characteristic) and 15.1% higher AUPRC (Area Under Precision-Recall Curve) than the second best method (Fig. 2 and Supplementary Fig. 1). Comparing to DeepDDI, DeepDrug achieves 31.0% higher AUROC and 17.0% higher AUPRC, presumably because DeepDDI only uses the SMILES sequence information as input. DeepDrug, on the other hand, takes advantage of a novel graph representation and is potentially capable of learning the underlying structural properties to attain better performance.

Due to the rarity of occurrence of drug-drug interactions [4], the number of known DDIs among a typical drug database is usually very low. Hence, to be more realistic and practical, we also evaluate robustness of DeepDrug with imbalanced datasets by altering the ratio between positive samples and negative samples to 1:2, 1:4 and 1:8. As studied in previous works, AUROC is likely to be an overoptimistic metric to evaluate the performance of a prediction algorithm, especially on highly skewed data, while AUPRC can provide a better assessment in this scenario [8]. In our case, although the AUPRC scores of all tested methods drop as compared to those using balanced samples, DeepDrug still maintains much higher AUPRC scores than other methods (Fig. 2). Thus, the noticeable performance level of DeepDrug in terms of AUPRC over other prediction methods demonstrates its superior ability in predicting new DDIs with sparsely labeled samples.

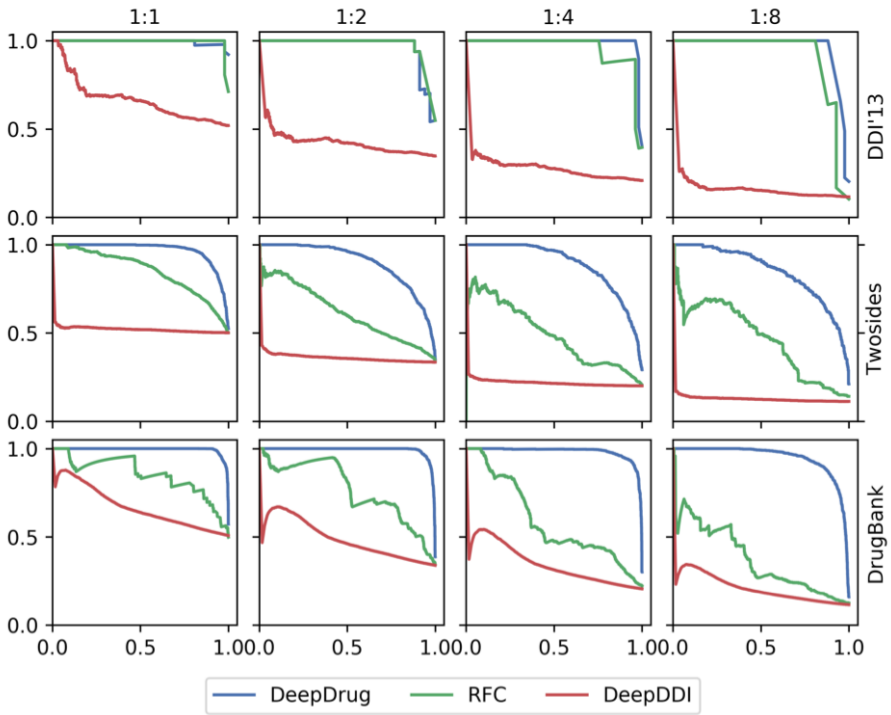

Fig. 2: AUPRC scores of DeepDrug, Random Forest Classification (RFC) and DeepDDI for binary DDI classification task using three datasets with various positive to negative sample ratios. DeepDrug outperforms the other two in all cases.

To further showcase the predictive capability of our model, we compare DeepDrug with other methods in two multi-class classification tasks. We pursue the classification of DDIs in DDI'13

dataset [16] based on the five interaction types defined by SemEval 2013 DDIExtraction Challenge [34]. In addition, we also classify the DrugBank DDI data into three classes: increase/decrease/negative. All of the DDI methods are evaluated using standard metrics including precision, recall, macro and micro F1 scores. Again, comparing to other methods, DeepDrug achieves the best performance across all evaluation standards (Table 1). Specifically, even at the highest imbalance ratio of 1:8, DeepDrug yields an micro F1 score of 0.9596, which is 10.74% better than the second best. The outperformance by DeepDrug indicates the merit of using graph representation of drug structure in DDI predictions. Similar to binary classification tasks presented earlier, we conduct numerical experiments on imbalanced datasets as well using the same positive to negative sampling ratios. The same trend is observed in multi-class classification results where the introduction of imbalance in dataset lowers F1 scores of all methods, DeepDrug manages to deliver much higher F1 scores than its counterparts. Therefore, by exploiting useful structural information from graph representation of drugs via GCNs, DeepDrug is shown to be robust in both binary and multi-class classification of DDIs.

Table 1: Evaluation metrics for DeepDrug, Random Forest Classfication (RFC) and DeepDDI in two multi-class classification tasks.

Dataset		DrugBank (3-class)				DDI'13 (5-class)			
Sample Ratio		1:1	1:2	1:4	1:8	1:1	1:2	1:4	1:8
DeepDrug	F1(micro)	0.910166	0.940608	0.949001	0.959612	0.704545	0.820896	0.874477	0.928058
	F1(macro)	0.863203	0.861269	0.855011	0.81421	0.475786	0.362041	0.358907	0.374305
RFC	F1(micro)	0.656716	0.71673	0.769221	0.866794	0.625	0.753731	0.811715	0.913669
	F1(macro)	0.450059	0.404906	0.289853	0.309548	0.341737	0.263676	0.241108	0.234832
DeepDDI*	F1(micro)	0.428136	0.267981	0.092342	0.092342	–	–	–	–
	F1(macro)	0.313026	0.225724	0.092732	0.092732	–	–	–	–

*Results on DDI'13 dataset are not applicable due to DeepDDI's unique labeling strategy.

DeepDrug accurately predicts DTIs

Although proteins generally have more intricate structures than chemical drugs due to their three-dimensional arrangement of sequences of amino acids, they can still be effectively represented by 3D graphs and used for predictive modeling. In our DeepDrug framework, we classify the DrugBank DTI dataset with binary labels. The performance is evaluated using standard metrics against a baseline method, random forest classification, and an established DTI prediction model, DeepDTA [31]. Overall, DeepDrug achieves the highest performance in most cases (Fig. 3 and Supplementary Fig. 2). Specifically, at the highest imbalance ratio, DeepDrug yields an AUPRC of 0.72, which is 9.09% better than the second best. For evaluating robustness, we use the same sampling ratios of positive to negative samples as those in DDI predictions. As shown in Fig. 3, DeepDrug attains a higher level of AUPRC score than other methods under various imbalance ratios, confirming its remarkable capability in predicting DTIs in sparsely labeled datasets. We note that DeepDrug is able to homogenize both drug and protein inputs into similar graph representations so that DDI and DTI predictions can be implemented using the same framework. This competitive edge is speculated to be one of the reasons that give rise to DeepDrug's accurate prediction of DTIs.

DeepDrug captures structural features

To understand the latent features captured by GCNs, we investigate the embedding for each drug in one of our datasets, DDI'13, by collecting their outputs after the pooling layer. We then visualize the overall embedding using t-SNE [28]. As shown in Fig. 4, we observe that when we project high-dimensional embedding of drug entities non-linearly onto a low-dimensional space, there is clustering of groups of drugs, especially the one on top of the figure, that implies the presence of certain form of similarity or close relationship. To verify this, we compute the pairwise Dice

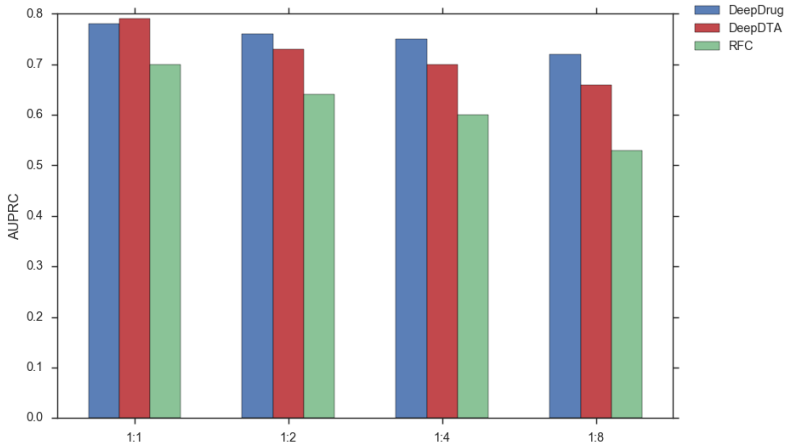

Fig. 3: AUPRC scores of DeepDrug, RFC and DeepDTA for binary DTI classification task with various positive to negative sample ratios. DeepDrug outperforms the other two in most cases.

coefficient [9] using Hetionet knowledge base [17] for each pair of drugs and compare the Dice coefficient of the cluster with the average of the overall database. The Dice similarity scores of the cluster (0.4102) is found to be much higher than the global average (0.1978). Notice that no similarity score in any form is explicitly used as inputs but DeepDrug is able to encode structurally similar entities with similar latent embeddings. For further understanding, we isolate all 12 drugs in the cluster and compare their chemical structures as well as their functionalities with other randomly sampled drugs in the dataset that are far away from the cluster. A subset of our sampled drugs are presented in Fig 5. The striking observation is that drugs in the cluster share very similar structural compositions. All of the 12 drugs contain a dimenthyamino group and at least one phenyl group. In terms of functionality, the cluster of drugs identified by DeepDrug embeddings is highly similar among themselves as well. Out of the 12 drugs in the cluster, 5 of them are meant for depression treatment (e.g. *Tramadol, Citalopram, Amitriptyline, Imipramine* in Fig. 5) and another 5 are related to migraines or pain relief (e.g. *Sumatriptan, Almotriptan* in Fig. 5). This provides another strong evidence that DeepDrug is capable to learn structural information which ultimately determines the functionality of the input entities. Such embedding capability is considered to be the main driving force to the outstanding performance of DeepDrug.

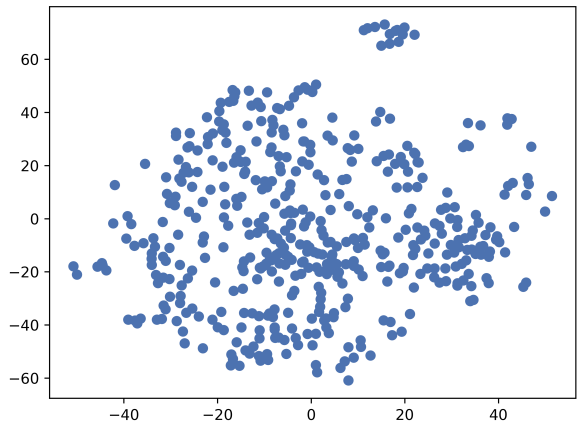

Fig. 4: The t-SNE visualization of drugs

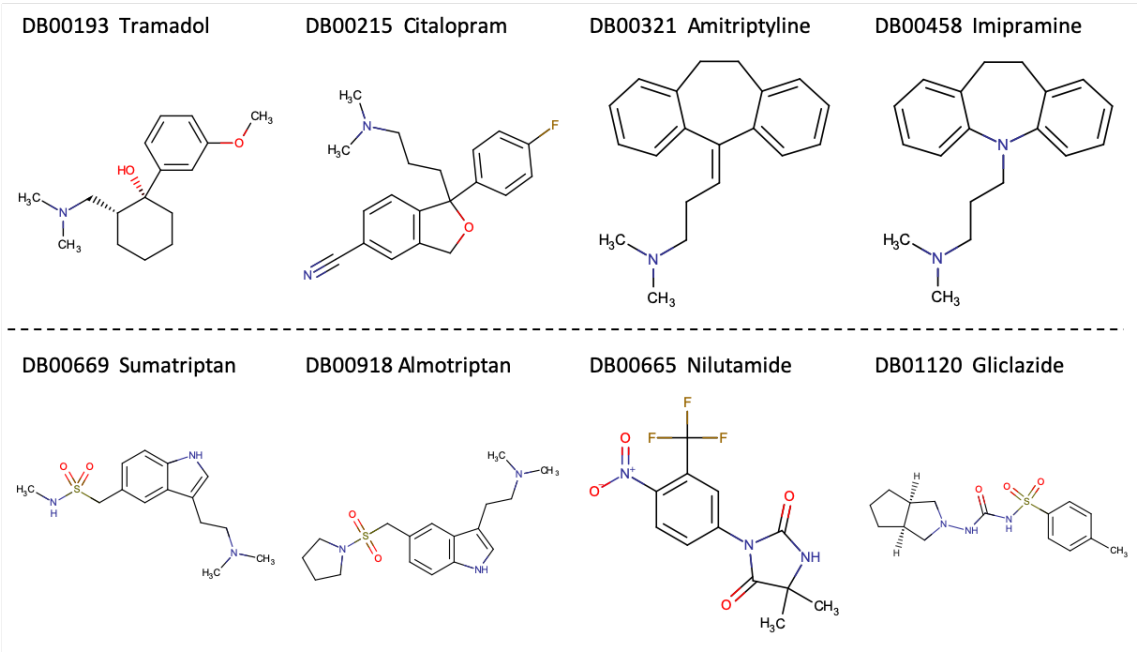

Fig. 5: Chemical structures of sampled drugs (with DrugBank ID). First 6 drugs are sampled from the cluster in the t-SNE plot, last two drugs are selected in regions far from the cluster and are added for comparison. Notice the first six drugs all contain a dimenthyamino group and at least one phenyl group whereas the other two do not. In terms of functionality, the first four drugs are for depressive disorder.Both *Sumatriptan* and *Almotriptan* treat headaches or pain. As references, *Nilutamide* is for prostate cancer and *Gliclazide* is for diabetes.

Sensitivity analysis

To further support the results shown earlier, we present a systematic sensitivity analysis of our model to provide justifications to the choice of hyperparamters used in our study (see *Methods*) and examine the robustness of our model. In particular, we analyze the sensitivity of DeepDrug with respect to the following parameters: presence of batch normalization [19], choice of global pooling operations, choice of activation functions, number of hidden units in each GCN layer and the total number of GCN layers. We use the binary classification task of DDIs as the testbed. From the results summarized in Fig. 6, we observe that using batch normalization, coupled with global max pooling (GMP) and the ReLU activation function, tends to produce better performance in terms of both AUROC and AUPRC scores though the amplification is not significant, indicating the stability of the GCN architecture. As the number of hidden units increases significantly (e.g. 32 and higher), both evaluation metrics start to saturate and the model becomes insensitive to the number of GCN layers as well. In general, DeepDrug is marginally responsive to most parameter choices, illustrating the robustness of the framework.

Discussion

In this work, we propose DeepDrug, a novel end-to-end deep learning framework for DDI and DTI predictions. DeepDrug takes in both drug SMILES strings and protein PDB inputs to characterize biochemical entities into graphical representations and utilizes GCNs to learn latent feature representations that give superior level of accuracy for predictive modeling. The competitive edge of graph-based architecture allows DeepDrug to incorporate both DDI and DTI predictions into a general framework. It also empowers DeepDrug to be applied to novel entities whose graphical representations can be extracted.

Overall, through extensive experiments on existing DDI and DTI datasets and detailed comparison with other published methods, we demonstrate the promising performance of DeepDrug in drug-related interaction prediction tasks. The visualization of latent feature representations and

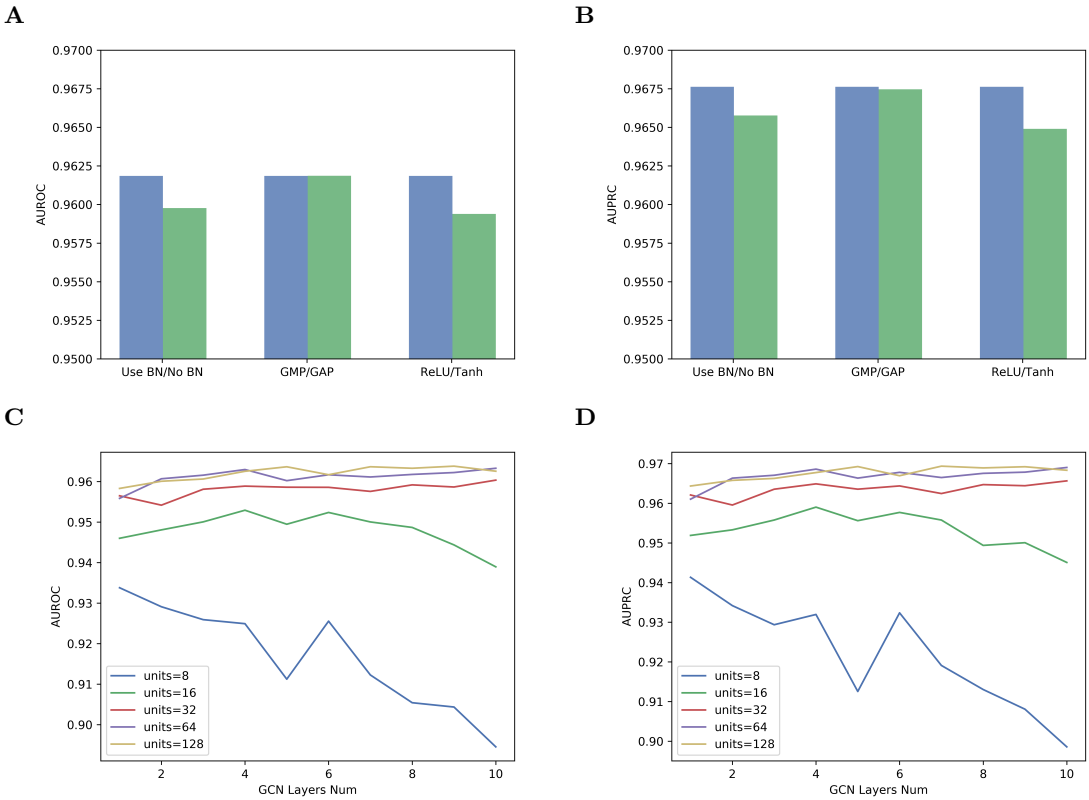

Fig. 6: Top: AUROC (A) and AUPRC (B) scores for use of Batch Normalization(BN), choice of Global Max Pooling or Global Average Pooling, choice of ReLU or Tanh activation function. Bottom: AUROC (C) and AUPRC (D) scores for different numbers of GCN layers and numbers of hidden units in each layer.

comparison of Dice similarity scores further support the ability of DeepDrug to learn structural properties of input entities. All these results suggest that DeepDrug can not only serve as a powerful tool in drug relation prediction, but also provide valuable insight in the discovery of interaction mechanisms of drugs.

We have shown the success of DeepDrug across a wide range of DDI and DTI prediction tasks, but there is still room for improvement. One possible future direction is to remove the limitation on the number of atoms in the drug/protein entities and identify DDIs or DTIs associated with excessively large molecules. A more in-depth systematic study of the embedded features learned by DeepDrug might also provide insight in interaction mechanisms or binding sites that could facilitate future biochemical research.

Methods

Datasets

The DDI data are collected from three different sources: DrugBank [42], Twosides [38] and DDI'13 [16]. Each sample contains a drug pair as well as a label representing an annotated DDI. For consistency, all drug pairs are identified in DrugBank and their SMILES sequence retrieved. One crucial observation during data collection is that drug entities across all datasets show extensive variations in their chemical structures. In particular, the number of atoms across all drug entities vary over a wide range. Although well-designed deep learning methods usually cope well with noisy inputs, we fix the number of atoms in drug entities to 50 by either discarding large molecules or padding small ones to standardize subsequent workflow. Negative samples are generated by randomly pairing drugs from DrugBank that do not have any known DDIs. We match the ratio of

positive samples to negative samples to 1:8 for robustness testing. The statistics of each dataset used in our study are summarized in Table 2.

The DTI data are collected from DrugBank databases as well. Each sample contains a drug-protein pair and an annotated DTI. Drug entities are again identified in DrugBank to collect their SMILES representations. Proteins are, on the other hand, identified in RCSB Protein Data Bank [7] to collect their standard structure files (e.g. PDB files). A similar normalization procedure is used here as well by limiting the number of atoms to 350. The statistics of proteins and DTIs used in this study are included in Table 2 as well. Similar to DDI dataset, negative samples are generated using random pairing of unlabeled drug-protein pairs.

Table 2: Statistics of datasets used in this study

DDI/DTI	Type	DrugBank	DDI'13	Twosides
DDI	Effect	–	244	–
	Mechanism	–	487	–
	Int	–	66	–
	Advise	–	191	–
	Increase	29598	–	–
	Decrease	9118	–	–
	Total	38716	988	41310
DTI	Positive	1268	–	–
	Total	1268	–	–
Drug		1599	488	495
Protein		559	–	–

Feature extraction

To extract graph representations of drugs, we use DeepChem [32] to convert the SMILES string of each drug into a feature matrix and an adjacency matrix. The feature matrix contains nodal information of each atom whereas the adjacency matrix represents the chemical bonds connecting the atoms as edges. Graph representation of proteins are extracted using ProteinGraph [27] from standard PDB structure files. Similarly, the outputs for each protein is a set of feature matrix representing nodal amino acids and an adjacency matrix describing the biochemical interactions as edges. The feature matrix and adjacency matrix of each entity is then fed into GCN for learning.

Graph Convolutional Network (GCN)

We present the key components of GCN relevant to DeepDrug and readers are referred to [23] for derivation and implementation details. The typical GCN model requires two inputs, a feature matrix $X \in \mathbb{R}^{N\times D}$ and an adjacency matrix $A \in \mathbb{R}^{N\times N}$, to give output $Z \in \mathbb{R}^{N\times D}$ with N being the number of nodes in the graph and D the number of features. The layer-wise forward propagation rule of a multi-layer GCN is defined as:

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}) \tag{1}$$

Here, $\tilde{A} = A + I_N$ is the adjacency matrix adjusted for self-connections by adding an identity matrix I_N. $\tilde{D} = \sum_j \tilde{A}_{ij}$ is the diagonal node degree matrix for normalization. $W^{(l)}$ is a layer-specific trainable weight matrix. $\sigma(\cdot)$ is the non-linear activation function. Lastly, $H^{(l)} \in \mathbb{R}^{N\times D}$ contains the activation values in the lth layer and $H^{(0)} = X$, $H^{(L)} = Z$. Hence, the overall GCN operations can be summarized as

$$f(X, A) = Z \tag{2}$$

As shown in Fig. 1, DeepDrug takes in pairs of inputs, such as a drug-drug pair or a drug-protein pair. The two entities go through separate processes of feature extractions to produce pairs of

feature matrices X_1, X_2 and adjacency matrices A_1, A_2. These are passed into two separate series of GCN layers to give two outputs Z_1, Z_2

$$\begin{aligned} f(X_1, A_1) &= Z_1 \\ f(X_2, A_2) &= Z_2 \end{aligned} \tag{3}$$

Based on a series of sensitivity studies (see *Sensitivity analysis*), we use 4 GCN layers each with 64 hidden units. The non-linear activation function in each layer uses the ReLU function.

$$\sigma = \mathrm{ReLU}(\cdot) = \max(0, \cdot) \tag{4}$$

The two outputs Z_1, Z_2 are then each passed into a Global Max Pooling (GMP) layer to summarize the features detected by GCN layers. The pooled feature maps for both entities are then concatenated and passed to a dense layer to compute the final prediction, y.

$$y = \sigma_d \left(W_d \left[\mathrm{GMP}(Z_1), \mathrm{GMP}(Z_2) \right] + b_d \right) \tag{5}$$

σ_d is the activation function of the dense layer which is either the sigmoid function or the softmax function depending on the classification task. W_d and b_d are the trainable weights and bias of the dense layer respectively.

Model training

Based on prediction results generated by DeepDrug, we define the mean square error (MSE) across a total of m samples as our objective function.

$$L_{MSE} = \frac{1}{m} \sum_{i}^{m} (y_i - \hat{y}_i)^2 \tag{6}$$

To minimize L_{MSE}, we first randomly split data in each dataset, then group 90% of total samples into the training set and 10% of total samples into the validation set. We then feed mini-batches of training data into DeepDrug to train our model for 100 epochs in order to reduce the loss function while constantly monitoring the model performance on the validation set. The loss function L_{MSE} is minimized using the Adam optimizer [22] with a learning rate of 0.001 and recommended values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. To prevent overfitting and other potential training issues, we implement several strategies in DeepDrug. First, all model weights are initialized using Xavier initialization techniques [14] to improve the starting behavior. Batch normalization layers are added to stabilize activation values of previous layers and speed-up the training process [19]. A dropout rate of 0.1 is applied to all hidden units for regularization effects [35]. For convergence, we also use an early stopping strategy with a window size of 5 to make sure the validation loss does not drop for 5 consecutive epochs. After training, we evaluate the performance of our model by plotting the Receiver Operating Characteristic curve, Precision-Recall Curve and computing the AUROC and AUPRC metrics.

Data Availability

The DDI and DTI datasets used in this work as well as the sourcecode for DeepDrug can be found at https://github.com/wanwenzeng/deepdrug.

References

1. Asada, M., Miwa, M., Sasaki, Y.: Enhancing drug-drug interaction extraction from texts by molecular structure information. arXiv preprint arXiv:1805.05593 (2018)
2. Bagherian, M., Sabeti, E., Wang, K., Sartor, M.A., Nikolovska-Coleska, Z., Najarian, K.: Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. Briefings in bioinformatics (2020)

3. Bajusz, D., Rácz, A., Héberger, K.: Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? Journal of cheminformatics **7**(1), 20 (2015)
4. Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., et al.: A community computational challenge to predict the activity of pairs of compounds. Nature biotechnology **32**(12), 1213–1222 (2014)
5. Bleakley, K., Yamanishi, Y.: Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics **25**(18), 2397–2403 (2009)
6. Boolell, M., Allen, M.J., Ballard, S.A., Gepi-Attee, S., Muirhead, G.J., Naylor, A.M., Osterloh, I.H., Gingell, C.: Sildenafil: an orally active type 5 cyclic gmp-specific phosphodiesterase inhibitor for the treatment of penile erectile dysfunction. International journal of impotence research **8**(2), 47–52 (1996)
7. Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J.M., Dutta, S., et al.: Rcsb protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic acids research **47**(D1), D464–D474 (2019)
8. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proceedings of the 23rd international conference on Machine learning. pp. 233–240 (2006)
9. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
10. Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P.: Convolutional networks on graphs for learning molecular fingerprints. In: Advances in neural information processing systems. pp. 2224–2232 (2015)
11. Feng, Q., Dueva, E., Cherkasov, A., Ester, M.: Padme: A deep learning-based framework for drug-target interaction prediction. arXiv preprint arXiv:1807.09741 (2018)
12. Fout, A., Byrd, J., Shariat, B., Ben-Hur, A.: Protein interface prediction using graph convolutional networks. In: Advances in neural information processing systems. pp. 6530–6539 (2017)
13. Gallagher, P.F., Barry, P.J., Ryan, C., Hartigan, I., O'Mahony, D.: Inappropriate prescribing in an acutely ill population of elderly patients as determined by beers' criteria. Age and ageing **37**(1), 96–101 (2008)
14. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256 (2010)
15. Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., Bassik, M.C.: Synergistic drug combinations for cancer identified in a crispr screen for pairwise genetic interactions. Nature biotechnology **35**(5), 463 (2017)
16. Herrero-Zazo, M., Segura-Bedmar, I., Martínez, P., Declerck, T.: The ddi corpus: An annotated corpus with pharmacological substances and drug–drug interactions. Journal of biomedical informatics **46**(5), 914–920 (2013)
17. Himmelstein, D.S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S.L., Hadley, D., Green, A., Khankhanian, P., Baranzini, S.E.: Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Elife **6**, e26726 (2017)
18. Hong, L., Lin, J., Li, S., Wan, F., Yang, H., Jiang, T., Zhao, D., Zeng, J.: A novel machine learning framework for automated biomedical relation extraction from large-scale literature repositories. Nature Machine Intelligence pp. 1–9 (2020)
19. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
20. Jia, J., Zhu, F., Ma, X., Cao, Z.W., Li, Y.X., Chen, Y.Z.: Mechanisms of drug combinations: interaction and network perspectives. Nature reviews Drug discovery **8**(2), 111–128 (2009)
21. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., et al.: Pubchem 2019 update: improved access to chemical data. Nucleic acids research **47**(D1), D1102–D1109 (2019)
22. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
23. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
24. Lazarou, J., Pomeranz, B.H., Corey, P.N.: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. Jama **279**(15), 1200–1205 (1998)
25. Li, Y., Tarlow, D., Brockschmidt, M., Zemel, R.: Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493 (2015)
26. Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, J., Chen, L., Zeng, J.: A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature communications **8**(1), 1–13 (2017)
27. Ma, E.: Protein graph. https://github.com/ericmjl/protein-interaction-network (2020)

28. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov), 2579–2605 (2008)
29. Meinertz, T.: Mibefradil—a drug which may enhance the propensity for the development of abnormal qt prolongation. European heart journal supplements **3**(suppl_K), K89–K92 (2001)
30. Mousavian, Z., Masoudi-Nejad, A.: Drug–target interaction prediction via chemogenomic space: learning-based methods. Expert opinion on drug metabolism & toxicology **10**(9), 1273–1287 (2014)
31. Öztürk, H., Özgür, A., Ozkirimli, E.: Deepdta: deep drug–target binding affinity prediction. Bioinformatics **34**(17), i821–i829 (2018)
32. Ramsundar, B., Eastman, P., Walters, P., Pande, V., Leswing, K., Wu, Z.: Deep Learning for the Life Sciences. O'Reilly Media (2019), https://www.amazon.com/Deep-Learning-Life-Sciences-Microscopy/dp/1492039837
33. Ryu, J.Y., Kim, H.U., Lee, S.Y.: Deep learning improves prediction of drug–drug and drug–food interactions. Proceedings of the National Academy of Sciences **115**(18), E4304–E4311 (2018)
34. Segura Bedmar, I., Martínez, P., Herrero Zazo, M.: Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). Association for Computational Linguistics (2013)
35. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
36. Staffa, J.A., Chang, J., Green, L.: Cerivastatin and reports of fatal rhabdomyolysis. New England Journal of Medicine **346**(7), 539–540 (2002)
37. Sun, Y., Sheng, Z., Ma, C., Tang, K., Zhu, R., Wu, Z., Shen, R., Feng, J., Wu, D., Huang, D., et al.: Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. Nature communications **6**(1), 1–10 (2015)
38. Tatonetti, N.P., Patrick, P.Y., Daneshjou, R., Altman, R.B.: Data-driven prediction of drug effects and interactions. Science translational medicine **4**(125), 125ra31–125ra31 (2012)
39. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
40. Wang, C., Kurgan, L.: Survey of similarity-based prediction of drug-protein interactions. Current medicinal chemistry (2020)
41. Weininger, D.: Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. Journal of chemical information and computer sciences **28**(1), 31–36 (1988)
42. Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., Woolsey, J.: Drugbank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research **34**(suppl_1), D668–D672 (2006)
43. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M.: Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. Bioinformatics **24**(13), i232–i240 (2008)
44. Zamora-Resendiz, R., Crivelli, S.: Structural learning of proteins using graph convolutional neural networks. bioRxiv p. 610444 (2019)
45. Zhang, T., Leng, J., Liu, Y.: Deep learning for drug–drug interaction extraction from the literature: a review. Briefings in bioinformatics **21**(5), 1609–1627 (2020)
46. Zitnik, M., Agrawal, M., Leskovec, J.: Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics **34**(13), i457–i466 (2018)
47. Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., Hoffman, M.M.: Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. Information Fusion **50**, 71–91 (2019)