OXFORD

Briefings in Bioinformatics, 2023, 1–12
https://doi.org/10.1093/bib/bbac597
Problem Solving Protocol

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

DSN-DDI: an accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning

Zimeng Li[†], Shichao Zhu[†], Bin Shao, Xiangxiang Zeng, Tong Wang and Tie-Yan Liu

Corresponding authors: Bin Shao. E-mail: binshao@microsoft.com; Xiangxiang Zeng. E-mail: xzeng@hnu.edu.cn; Tong Wang (Lead contact). E-mail: watong@microsoft.com

[†]Zimeng Li and Shichao Zhu contributed equally to this work.

Abstract

Drug–drug interaction (DDI) prediction identifies interactions of drug combinations in which the adverse side effects caused by the physicochemical incompatibility have attracted much attention. Previous studies usually model drug information from single or dual views of the whole drug molecules but ignore the detailed interactions among atoms, which leads to incomplete and noisy information and limits the accuracy of DDI prediction. In this work, we propose a novel dual-view drug representation learning network for DDI prediction ('DSN-DDI'), which employs local and global representation learning modules iteratively and learns drug substructures from the single drug ('intra-view') and the drug pair ('inter-view') simultaneously. Comprehensive evaluations demonstrate that DSN-DDI significantly improved performance on DDI prediction for the existing drugs by achieving a relatively improved accuracy of 13.01% and an over 99% accuracy under the transductive setting. More importantly, DSN-DDI achieves a relatively improved accuracy of 7.07% to unseen drugs and shows the usefulness for real-world DDI applications. Finally, DSN-DDI exhibits good transferability on synergistic drug combination prediction and thus can serve as a generalized framework in the drug discovery field.

Keywords: drug–drug interaction, drug side effect, dual-view learning, molecular graph

Introduction

Taking multiple drugs at the same time can provide therapeutic benefits but increase the risk of adverse side effects caused by the physicochemical incompatibility of the drug combination [1–3]. The identification of drug–drug interactions (DDIs) remains a challenging task because the huge number of drug–drug pairs leads to pharmaceutical research and clinical trials highly expensive and less efficient. A multitude of computational approaches to DDI prediction have been developed and proved to be an effective and an alternative choice to alleviate the challenge [4–8]. Most of these approaches follow the assumption that drugs with similar features are more likely to have similar interactions. To make full use of the raw features of drugs e.g. drug structures, chemical properties and molecular fingerprints, recent works mainly focus on utilizing the powerful feature extraction ability of deep neural network [9–12]. As a drug can be represented as a graph based on its molecular structure, graph neural networks (GNNs) have shown the impressive representation learning ability for drug molecules. Existing GNN-based approaches for DDI prediction [13–15] usually take the advantage of GNN's topological and semantic representation capabilities to model the drug itself, and then learn the representation of drug pairs based on the respective representation of each drug. Finally, the representations of drugs or drug pairs are used for final DDI prediction.

Considering that a drug can be simply divided into several functional groups or chemical substructures which jointly lead to the overall pharmacological properties [16], some studies were motivated to refine drugs into substructures for DDI prediction [17–20] .Existing works can be roughly classified into two categories: implicit and explicit ones, depending on how the substructures are used. The implicit one usually takes substructure features as inputs of the model, which does not explicitly learn substructure representations through the neural network [13, 17]. In contrast, the other kind of approaches, including SSI-DDI [18], GMPNN-CS [19], SA-DDI [20] and so on, extract the respective substructures of a pair of drugs in drug representation learning stage and predict DDI effect by identifying pairwise interactions between substructures of the two drugs in the final readout module, leading to an improvement in performance over previous methods. However, the substructures are only learned from the hidden representations of a single drug, whereas the interactive information between the two drugs that could provide more valuable information for substructure learning is often ignored.

Zimeng Li is a graduate student at Hunan University. He worked as an intern at Microsoft Research.
Shichao Zhu is a PhD candidate at Chinese Academy of Sciences. She worked as an intern at Microsoft Research.
Bin Shao is a principal researcher at Microsoft Research. He received his Ph.D. degree at Fudan University, Beijing, China. His research interest is molecular dynamics simulation.
Xiangxiang Zeng is a professor at Hunan University. His research interest is computer-aided drug design.
Tong Wang is a senior researcher of Microsoft Research. He received his Ph.D. degree at Tsinghua University, Beijing, China. His current research interests include computer-aided drug design, molecular dynamics simulation, bioinformatics and computational biology.
Tie-Yan Liu is ACM/IEEE fellow and a distinguished scientist of Microsoft. He received his Ph.D. degree at Tsinghua University, Beijing, China. His research interest is AI for science.
Received: October 3, 2022. **Revised:** November 18, 2022. **Accepted:** December 4, 2022
© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

In most of the DDI prediction algorithms, drug representation learning is a single-view process in the message passing module of the GNN for DDI prediction, which only encodes information from the drug itself and thus may hinder the accuracy improvement of DDI prediction. There are some works that try to adopt dual-view representation learning into DDI prediction [21, 22]. A state-of-the-art algorithm, MHCADDI [21] designs an external message passing mechanism between drugs' structures to integrate joint drug–drug information during the representation learning stage for individual drugs. However, it only uses the drug hidden representations derived from the last GNN block for DDI prediction, whereas hidden representations derived from all previous GNN blocks are ignored. Notably, during the message passing of the GNN, each GNN block gathers different orders of neighboring information, leading to hierarchical global representations of the whole drug molecule from different perspectives. Therefore, these hierarchical global representations can provide drug information at different substructure levels for final DDI prediction. Furthermore, dual-view learning in previous studies [23, 24] mostly focuses on the global representation learning of the whole drug molecule, whereas it can be further employed for local representation learning of each atom in the drug molecule. Specifically, dual-view learning could be adopted to update each node representation in building blocks of the GNN and integrate dual-view information on the atom level.

In this study, to make full use of drug hidden representations from different GNN blocks and leverage dual-view information on the atom level, we design DSN-DDI that employs local and global representation learning modules iteratively, learns drug substructures from intra-view and inter-view simultaneously and employs all hierarchical global representations for DDI prediction without depending on additional domain knowledge. This makes the model equally applicable to inductive settings where only the chemical structure of the drug is accessible. DSN-DDI can be divided into two parts: substructure learning module for learning substructures of a drug pair from dual view and DSN decoder module for substructure integration and final DDI prediction. Our comprehensive evaluations on DrugBank and Twosides datasets demonstrate that DSN-DDI has achieved 13.01% relative accuracy improvement on the transductive setting and 7.07% relative improvement to unseen drugs on the inductive setting. Furthermore, the significant improvement on known drugs makes the AUC reach 99.47% on DrugBank and 99.90% on Twosides for the transductive setting, respectively. In addition, DSN-DDI exhibits the usefulness of DDI prediction to new approved drugs and shows the good transferability for drug combination prediction. All these results suggest that DSN-DDI can not only act as a powerful tool for DDI prediction, but also serve as a generalized framework to greatly facilitate the research in the drug design and discovery field.

Method

Problem definition

We first summarize the commonly used notations in Table 1 for easy retrieval. We introduce some basic concepts and formalize the problem of DDI prediction as a binary classification problem, along with a new formulation of drug–drug graph.

Definition 1. DDI prediction. Given a set of triplets consisting of two drugs $\mathcal{D}$ and interaction types $\mathcal{R}$, the DDI prediction can be formulated as a function $f : \mathcal{D} \times \mathcal{R} \times \mathcal{D} \rightarrow [0, 1]$, where 1 indicates that a triplet (D_i, R, D_j) of two drugs D_i, D_j and an interaction type R exist and 0 means it does not exist. The goal of DDI prediction is to learn the function f and predict the probability of triplet existence that the given interaction type will occur between the two drugs. Notably, the DDI prediction task in this work is regarded as a binary classification task instead of directly predicting the interaction types. For different kinds of interaction types, the learnt models can make predictions when taking corresponding triplets, respectively.

Table 1. Notations and descriptions

Notations	Descriptions
$\mathcal{D}$	The set of drugs
$\mathcal{R}$	The set of DDI types
$\mathbf{G} = (\mathcal{V}, \mathcal{E})$	A drug graph
$\mathcal{V}$	The set of nodes in a drug graph
$\mathcal{E}$	The set of edges in a drug graph
$\mathbf{A} \in \mathbf{R}^{\mathbf{N}\times\mathbf{N}}$	The intra-adjacency matrix of a given drug
$\tilde{\mathbf{G}} := \mathcal{D}_\mathbf{i} \times \mathcal{D}_\mathbf{j}$	A bipartite graph of a drug pair
$\tilde{\mathbf{A}} \in \mathbf{R}^{\mathbf{N}_\mathbf{i}\times\mathbf{N}_\mathbf{j}}$	The inter-adjacency matrix of a drug pair
$\mathbf{e}_{\mathbf{ij}}$	The edge between atom $\mathbf{i}$ and $\mathbf{j}$, $\mathbf{e}_{\mathbf{ij}} \in \mathcal{E}$
$\mathbf{h}_\mathbf{i}$	The feature vector of the atom $\mathbf{i}$ from intra-view $\mathbf{A}$
$\tilde{h}_i$	*The feature vector of the atom i from inter-view* $\tilde{\mathbf{A}}$
$\mathbf{d}_\mathbf{i}$	The feature vector of the drug $\mathcal{D}_\mathbf{i}$
$\mathbf{f}$	The DDI prediction function $\mathbf{f} : \mathcal{D} \times \mathcal{R} \times \mathcal{D} \rightarrow [0, 1]$
$\mathbf{g}$	The graph mapping function $\mathbf{g} : \mathcal{E} \rightarrow \mathbf{R}^\mathbf{m}$

Definition 2. Drug–drug graph. A single drug's chemical structure can be naturally represented as an undirected graph of $G = (\mathcal{V}, \mathcal{E})$, in which $\mathcal{V}$ is the set of nodes with the input features of the corresponding atoms, $\mathcal{E}$ are edges between them. For the intra-view, we represent edges by an adjacency matrix $A \in R^{N\times N}, A_{ij} \in \{0, 1\}$, where $N := |\mathcal{V}|$ denotes the number of atoms, $A_{ij} = 0$ means that there is no edge between atom i and j. For the inter-view, we design a bipartite graph for the drug pair. Given a triplet of two drugs and an interaction type, the inter-connection across two drugs $\mathcal{D}_i, \mathcal{D}_j$ is defined as a bipartite graph $\tilde{G} := \mathcal{D}_i \times \mathcal{D}_j$, where the adjacency matrix $\tilde{A} \in R^{N_i\times N_j}, \tilde{A}_{ij} \in \{0, 1\}$ and N_i, N_j denote the number of nodes in the two drugs, respectively. $\tilde{A}_{ij} = 1$ denotes there is an edge between atom i in $drug_A$ and atom j in $drug_B$. In the bipartite graph, for each atom i in $drug_A$, we connect it with each atom j in $drug_B$. Finally, our dual-view substructure learning for DDI prediction is built upon the drug–drug graph with intra-connection A and inter-connection $\tilde{A}$ to learn the DDI prediction function f as a binary classification problem.

Overview of DSN-DDI

DSN-DDI is a GNN that learns drug substructures from dual-view iteratively and predicts DDI via interactions among all substructures of two drugs. A general sketch to illustrate substructure extraction and dual-view representation learning is shown in Figure 1. In each GNN block, a node is updated by aggregating the information from its neighboring nodes, followed by the substructure extraction process. A series of GNN blocks have receptive fields at different scales, leading to different extracted substructures (Figure 1**A**). For the neighboring nodes, in the intra-view, it is defined as the nodes that have a chemical bond with the central node in the drug molecule. In contrast, in the inter-view, we define inter edges by connecting a node in one drug with all nodes in the other drug to form a bipartite graph. Then the node in one drug is updated by aggregating all nodes' information in the other drug. Figure 2 shows the flowchart of DSN-DDI, which consists of the substructure learning module and DSN decoder.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

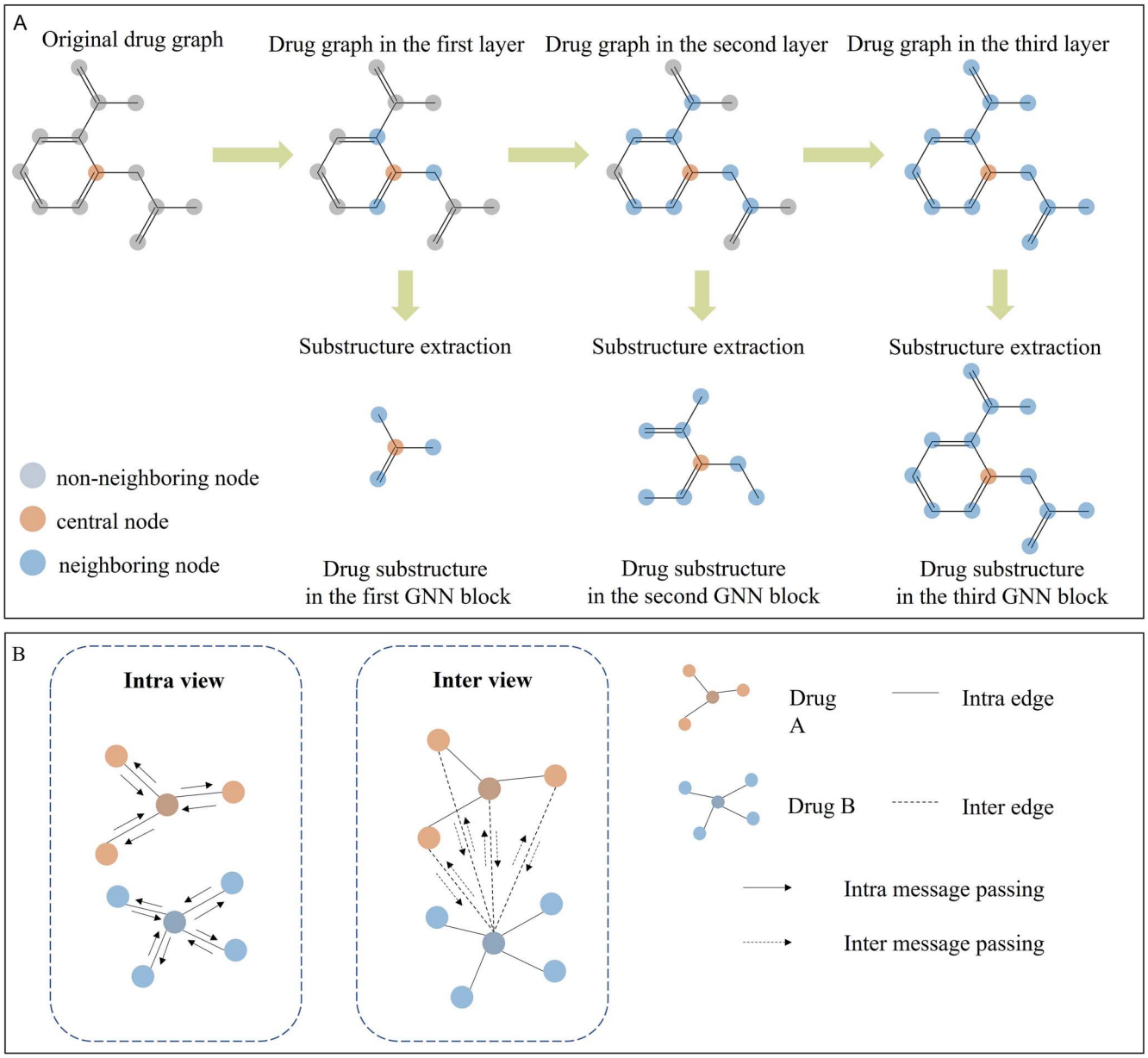

Figure 1. A general sketch to illustrate the substructure extraction and dual-view representation learning. (**A**) An illustration of drug substructure extraction in GNN blocks. The original drug (Aspirin) is represented by a molecular graph, in which the nodes represent the atoms, and the edges represent the chemical bonds. In each GNN block, a node (such as the central node colored orange) is updated by aggregating the information from its neighboring nodes. Then a substructure is extracted. A series of GNN blocks have receptive fields at different scales, leading to different extracted substructures. (**B**) The workflow of dual-view message passing. Drug A and drug B are represented by different graphs, respectively. In the intra-view, a node stands for an atom and an edge stands for a chemical bond. A node in drug A is updated by aggregating its neighboring nodes in drug A and vice versa. In the inter-view, we connect a node in one drug with all nodes in the other drug to form inter edges. Then the node in one drug is updated by aggregating all nodes' information in the other drug. In the dual-view, a node is updated by both intra-view message and inter-view message simultaneously.

Taking two single drug graphs for intra-view learning and a bipartite graph for the inter-view learning as input, the substructure learning module has both local (atom level) and global (drug level) representation learning processes. The DSN encoder is designed to learn the local representations and encode the interactions within a drug (intra-view) and between atoms of a drug pair (inter-view) simultaneously. The intra-view and inter-view information is then aggregated to update node representations for the next DSN encoder. Following a DSN encoder block, a self-attention graph (SAG) pooling layer is designed to learn global representations for the drug pair. Since a series of encoders capture different orders of neighboring information, the substructures are extracted from different perspectives.

DSN decoder is a co-attention scoring function to predict the probability of the triplet (D_i, R, D_j), where D_i and D_j stand for the drug pair and R stands for a type of drug pair interaction. In this module, each pair of substructures from respective drugs are integrated for the final DDI prediction.

Drug representation learning

As shown in Figure 2**B**, the DSN encoder consists of a representation extraction layer for both two drugs with the shared weights,

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

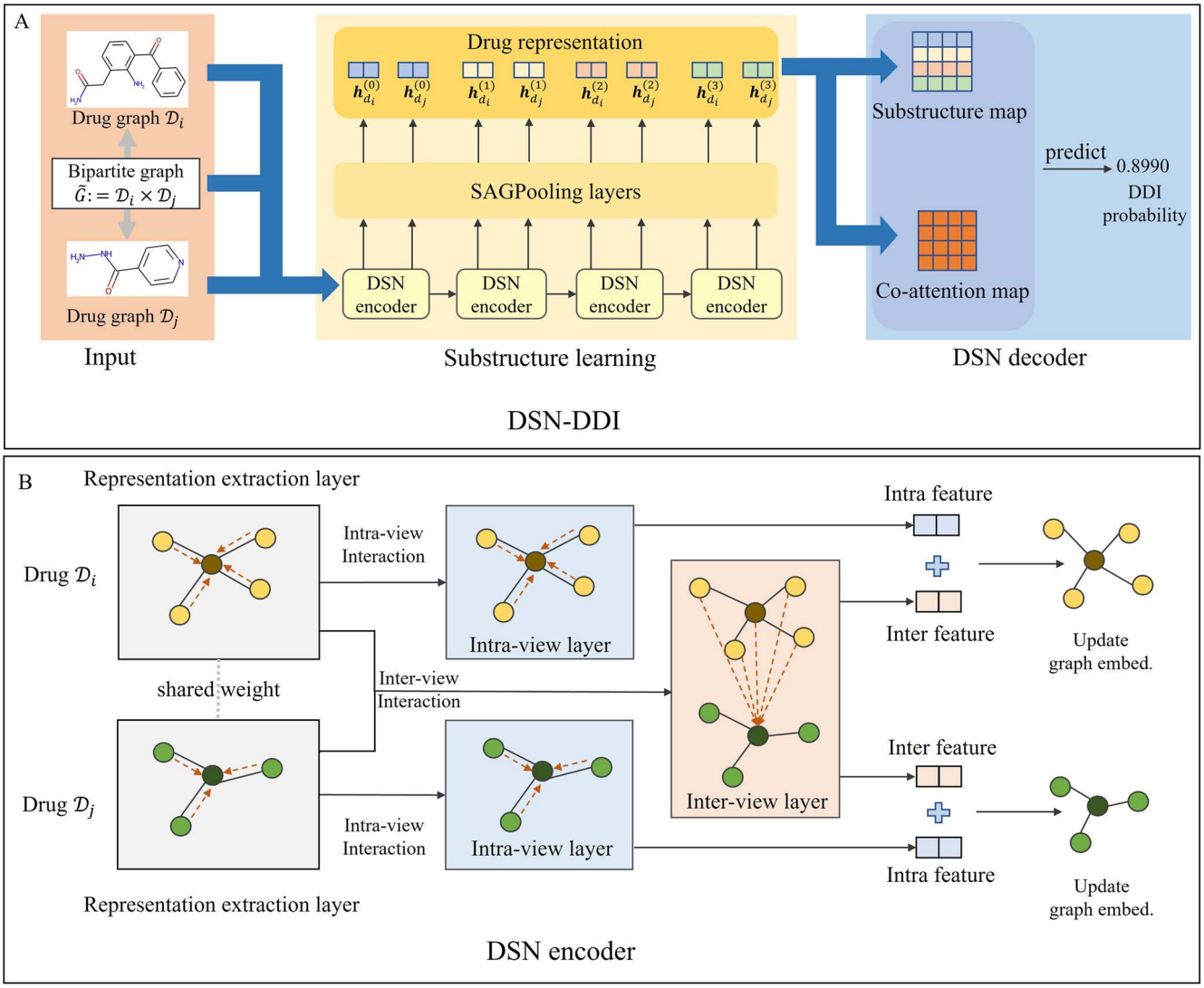

Figure 2. The general sketch of DSN-DDI. (**A**) An overall framework of DSN-DDI. Two single drug graphs (D_i and D_j) for intra-view learning and a bipartite graph ($\tilde{G}$) for inter-view learning are taken as model input. DSN-DDI consists of the substructure learning module and DSN decoder. For the substructure learning module, a series of repetitive DSN encoders learn the local atom representations for each atom, followed by the SAGPooling layers to learn the global drug representations and generate drug substructure embeddings $h_{d_i}^{(l)}, h_{d_j}^{(l)}$. Finally, all substructure embeddings are fed into the DSN decoder that is defined as a co-attention scoring function for a given triplet for DDI prediction. (**B**) DSN encoder. A pair of drugs are encoded by a representation extraction layer with the shared weights and then are simultaneously fed into the intra-view layers for respective drugs and the inter-view layer for the drug pair. The inter-view and intra-view information are then aggregated to update node representations for the next DSN encoder.

a respective intra-view layer for each drug and an inter-view layer between two drugs.

For the representation extraction and intra-view layers, our model employs the graph attention [25] to capture the intra-view interaction. Specifically, atoms are allowed to send attentional messages to each other along the edges in the drug graph A, and each atom then aggregates all the messages sent to it (Equation (1)).

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in\{i\}\cup\mathcal{N}_i} \frac{1}{c_i}\alpha_{ij} W^{(l)} h_j^{(l)} + b^{(l)}\right), \tag{1}$$

$$\alpha_{ij} = \mathrm{softmax}_j\left(a^{(l)T}\left[W_k^{(l)} h_i^{(l)} \parallel W_{k'}^{(l)} h_j^{(l)}\right]\right), \tag{2}$$

where $W^{(l)} \in R^{d^{(l+1)}\times d^{(l)}}, a^{(l)} \in R^{d^{(l+1)}}$, $b^{(l)} \in R^{d^{(l+1)}}$ are l-th layer's trainable matrices and bias in the intra-view layer, $d^{(l)}, d^{(l+1)}$ are the dimensions of l-th layer's input features and output features, respectively. $h_j^{(l)} \in R^{d^{(l)}}$ denotes the hidden representations of neighboring nodes in chemical structure A of the drug itself G, and $h_i^{(l+1)} \in R^{d^{(l+1)}}$ denotes the updated hidden representations of node i. The $\mathcal{N}_i$ denotes the set of neighbors of node i i.e. the atoms that have a bond with atom i. c_i is a normalization constant that is chosen as $|\mathcal{N}_i| + 1$. The attentional coefficient α_{ij} is computed as Equation (2) to account for the relevance of atom i and j.

Different from the intra-view interaction, the inter-view adjacency matrix is formulated as a bipartite graph $\tilde{G} := \mathcal{D}_i \times \mathcal{D}_j$ based on two drugs, to capture the interactions across drug boundaries via a co-attentional mechanism [21]. Specifically, atoms in drug $\mathcal{D}_i$ are allowed to send or receive attentional messages to the atoms in another drug $\mathcal{D}_j$, as follows.

$$\tilde{h}_i^{(l+1)} = \sigma\left(\sum_{j\in\tilde{\mathcal{N}}_i} \frac{1}{\tilde{c}_i}\beta_{ij}\tilde{W}^{(l)} h_j^{(l)} + \tilde{b}^{(l)}\right), \tag{3}$$

$$\beta_{ij} = \mathrm{softmax}_j\left(\tilde{a}^{(l)T}\left[W_{dx}^{(l)} h_i^{(l)} \parallel W_{dy}^{(l)} h_j^{(l)}\right]\right), \tag{4}$$

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

where $\tilde{W}^{(l)} \in R^{d^{(l+1)} \times d^{(l)}}, \tilde{a}^{(l)} \in R^{d^{(l+1)}}, \tilde{b}^{(l)} \in R^{d^{(l+1)}}$ are l-th layer's trainable matrix and bias in the inter-view layer, where $d^{(l)}, d^{(l+1)}$ are the dimensions of input features and out features, respectively. $\tilde{h}_j^{(l)} \in R^{d^{(l)}}$ denotes the hidden representations of neighboring nodes, and $\tilde{h}_i^{(l+1)} \in R^{d^{(l+1)}}$ denotes the updated hidden representations of node i. The $\tilde{\mathcal{N}}_i$ denotes the set of neighbors of node i in the bipartite graph $\tilde{G}$ i.e. all atoms in the other drug. c_i is a normalization constant that is chosen as $|\tilde{\mathcal{N}}_i|$. The attentional coefficient β_{ij} is computed as Equation (4) to account for the relevance of atom i and j in each drug.

The representations learned from dual-view interactions are then aggregated in both node and graph levels. For node level representation on the l-th layer, we aggregate the intra-view representations and inter-view representations as follows:

$$h_i^{(l)} = \mathrm{MLP}([h_i^{(l)} \| \tilde{h}_i^{(l)}]), \tag{5}$$

The integrated node representations are fed into the next layer to provide multi-perspective information for each node. Furthermore, the layer-wise integrated node representations are also used to produce the graph level representations via substructure pooling layers.

For substructure pooling layers, we employ the self-attention graph pooling (SAGPooling) as the readout function. Specifically, the global representation of a drug graph is calculated by the sum of the node embeddings and their corresponding importance scores (Equation (6)).

$$h_d^{(l)} = \sigma\left(\sum_{i=1}^{N} \alpha_i h_i^{(l)}\right), \alpha_i = softmax\left(AX^{(l)}W_{SAG}^{(l)}\right), \tag{6}$$

where σ is a sigmoid non-linearity, A is the adjacency matrix, $X^{(l)}$ is the embeddings of all nodes in the l-th layer's, $W_{SAG}^{(l)} \in R^{d^{(l)} \times 1}$ are l-th layer's trainable matrix, d is the input dimension of node-level features and 1 is the output dimension of the important score, respectively.

DSN decoder

Following previous studies in DDI prediction [18, 19], the decoder is defined as a co-attention scoring function for a given triplet of two drugs and interaction types.

$$s\left(d_i, r, d_j\right) = \left(\sum_{l=1}^{L}\sum_{l'=1}^{L} \gamma_{ll'} d_i^{(l)} M_r d_j^{(l')}\right), \tag{7}$$

$$\gamma_{ll'} = a^T \tanh\left(W_0 d_i^{(l)} + W_1 d_j^{(l')}\right), \tag{8}$$

where $d_i^{(l)}$ and $d_j^{(l')}$ are the substructure embeddings of two drugs D_i and D_j in the layer l. M_r is the trainable representation matrix of the interaction type r, and σ is the sigmoid function. To obtain the different importance of each substructure pair in DDI prediction, we use the co-attention operation [26] to account for the importance $\gamma_{ll'}$ of each pairwise interaction between substructure embedding d_i and d_j in each layer. In Equation (8), $a \in R^d$ is a trainable weight vector, $W_0, W_1 \in R^{d \times d}$ are two trainable weight matrices.

With applying Equation (7) for multi-relational data, the number of parameters in M_r becomes too large. To alleviate this issue, we employ the diagonal matrices for regularizing the weights M_r of interaction types. With the diagonal parameterization, each M_r is defined through the diagonalization of the low-dimensional vector activated by a non-linear layer:

$$\mathbf{M_r} = diag\left(\sigma\left(\mathbf{W_m m_r}\right)\right), \tag{9}$$

where $m_r \in R^d$ denotes a trainable vector of the interaction type r. σ is an activation function i.e. elu and $W_m \in R^{d \times d}$ are a trainable matrix. The diagonalization can be seen as a sparsity constraint on the weight matrices for each interaction type. At the same time, we expect that diagonal parameterization can alleviate overfitting for datasets with rare interaction types.

Objective function

Here, DDI prediction is formulated as a binary classification problem, the model is trained end-to-end using the binary cross-entropy loss, given in Equation (10). The existing triplets (D_i, R, D_j) are considered positive samples, and the negative samples are generated by replacing either D_i or D_j. The sampling ratio of positive and negative samples is 1 : 1.

$$\mathcal{L}_{\mathcal{BCE}} = -\frac{1}{|\mathcal{T}|}\sum_{t=(d_i,r,d_j)\in\mathcal{T}}\left(\log p_t + \log\left(1 - p_t^{'}\right)\right), \tag{10}$$

where p_t denotes the probability of positive samples, $p_t^{'}$ denotes the probability of corresponding negative samples and $|\mathcal{T}|$ represents the number of triplets in the dataset. The probabilities of samples are computed by the scoring function defined in Equation (7).

Datasets

We evaluated DSN-DDI on two widely used dataset, DrugBank [27] and Twosides [1]. DrugBank contains 191 808 DDI triplets with 1706 drugs and 86 interaction types. There is only one interaction type for a drug pair, which describes how one drug affects the metabolism of another one. Following the same data split scheme in GMPNN-CS [19], the train, validation and test sets in DrugBank dataset were split with a ratio of 6:2:2 for the transductive setting. For the inductive setting, we also followed the same data split scheme in Nyamabo *et al.* [19], which can be found in the section of Results. Twosides contains 4 649 441 DDI triplets with 645 drugs and 1317 interaction types, proposed by Zitnik *et al.* [23] based on the original Twosides data. Different from DrugBank dataset, there are multiple interaction types for a drug pair in Twosides dataset. Following the same criterion in Zitnik *et al.* [23], the interaction types with <500 triplets were removed, resulting in 4 576 287 DDI triplets with 963 interaction types in the dataset. Then the training, validation and test sets were split with the same ratio of 6:2:2 for the transductive setting. Furthermore, to evaluate the generalization ability of DSN-DDI, we have built an FDA-approved drug dataset in which drugs in DrugBank approved before the year of 2017 were recruited in the training set, whereas those approved after the year of 2017 form the test set. More details can be found in the section of 'real-world' applications.

Baselines

We evaluated our algorithm DSN-DDI with the following state-of-the-art baselines, including both the substructure-based algorithms and algorithms with dual-view learning.

- MR-GNN [13]: utilizes the graph convolution networks to embed node-level features and then captures the representations of different substructures for each drug. These representations are all fed into a recurrent neural network for DDI prediction.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

- MHCADDI [21]: utilizes a co-attention mechanism to integrate inner and outer messages based on the structure information of a drug pair during the representation learning of individual drugs.
- SSI-DDI [18]: takes each node's hidden representations as substructures and then computes interactions between these substructures to predict DDI .
- GAT-DDI [25]: utilizes graph attention networks (GAT) for drug representations learning and DDI prediction.
- GMPNN-CS [19]: learns different chemical substructures and then models interactions between these substructures for final DDI prediction.
- SA-DDI [20]: employs a substructure-aware GNN and designs a novel substructure attention mechanism to learn drug for DDI prediction.

All these algorithms were evaluated on both transductive and inductive settings. In addition, we omitted comparisons with traditional machine learning baselines as prior work [17] has already found them to be significantly underperforming compared with deep learning approaches on this task. The results of baselines are reported from Nyamabo *et al*. and Yang *et al*. [19, 20].

Parameter settings

DSN-DDI contains four DSN encoder blocks. Each block has a representation layer with the shared weights for both two drugs, an intra-view layer for respective drugs and an inter-view layer for the drug pair. All layers adopt GAT mechanism equipped with two attention heads for message passing. Each DSN encoder outputs 128-dimensional hidden representations, consisting of 64-dimensional inter-view embedding and 64-dimensional intra-view embedding. Each node in drug graph has an original 55-dimentional chemical feature for input (Supplementary Table S1). We used grid-search for hyperparameters selection on validation set. The grid-search space is summarized in Supplementary Table S2. The model was initialized using Xavier initialization and trained on mini-batches with 1024 DDIs. We used the Adam SGD optimizer [28] and trained the models with 200 epochs on DrugBank and 120 epochs on Twosides. The learning rate was set to 0.01 for the transductive setting and 0.001 for the inductive setting, respectively.

Results

Performance evaluation on transductive setting for existing drugs

We conducted experiments on two standard benchmarks: DrugBank and Twosides, to evaluate the performance of our algorithm DSN-DDI. The statistics of the datasets are summarized in Supplementary Table S3. For both datasets, each drug is represented as a molecular graph converted from SMILES [29] by RDKit [30]. Each node in the drug graph has 55-dimensional initial chemical features, such as atomic symbols and degrees of the atoms (Supplementary Table S1).

Similar to previous studies, we evaluated performance for two scenarios: transductive and inductive. For the transductive setting, drugs in the test sets also exist in the training set, whereas the inductive setting contains drugs fully or partially not existing in the training set to examine the model generalization ability to new drugs. To make a fair comparison, following the same settings in the previous study [19], we adopted the same positive samples and negative samples with all baselines on DrugBank and Twosides and then performed 3-fold cross validation with the data split ratio of training:validation:test = 6:2:2. The data split on both datasets was performed on both positive samples and negative samples i.e. all positive and negative triplets are divided into training, validation and test sets with the same ratio, respectively. Experiment results are reported with the means and standard deviations of the following six metrics across the three folds: the accuracy (ACC), the area under the receiver operating characteristic (AUC) and the average precision (AP), the F1 score. The detailed definitions of each metric are given as follows.

- The accuracy (ACC): is defined as the number of correct predictions divided by the number of total predictions.
- The area under the receiver operating characteristic (AUC): is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.
- The average precision (AP): is calculated by taking the mean average precision over all classes.
- The F1 score: is the harmonic mean of precision and recall.

As shown in Table 2, DSN-DDI achieved the best performance of all four metrics compared with state-of-the-art algorithms on both DrugBank and Twosides. Specifically, although the state-of-the-art algorithms have achieved high DDI prediction accuracy, our algorithm still achieved further improvement on all evaluation metrics. As a comparison, DSN-DDI achieved significant improvement on Twosides by making 13.01 and 11.86% relative improvements on ACC and F1 score over the second-best approach, SA-DDI. Furthermore, the AUC reaches 99.47% on DrugBank and 99.90% on Twosides, respectively, which indicates DSN-DDI made perfect DDI prediction on the transductive setting and nearly has solved the DDI prediction task for existing drugs. In addition, Supplementary Table S4 compared the prediction accuracy values on the DrugBank training and test sets, respectively, and the consistent performance indicates that we trained the model without over-fitting. These results have verified the powerful representational ability of DSN-DDI with several novel-designed components for dual-view substructure learning.

Performance evaluation on inductive setting for unseen drugs

The inductive setting is more challenging than the transductive setting since there exists unseen drugs in DDI triplets in the test sets. This cold-start scenario is an extremely difficult trial for the generalization ability of the model, without knowing any prior knowledge of the unseen drugs in the training process. In this setting, we split the dataset following the same scheme in literature [18, 19, 31]. Specifically, we randomly picked 20% of the drugs as unknown drugs and regarded the remaining drugs as existing ones. All positive and negative samples on the train dataset are all DDI triplets in which both drugs are existing drugs, whereas two partitioning schemes are used for splitting the test set:

- S1 Partition Scheme: the positive and negative samples on the test set have two unknown drugs. This task is to predict DDI for a pair of new drugs that no side effect is known in any combination with other drugs in the training set.
- S2 Partition Scheme: the positive and negative samples on the test set have one unknown drug and one known drug. This task is to predict DDI for a new drug that has no effect in any combination with another existing drug.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

Table 2. Performance evaluation of DSN-DDI and state-of-the-art algorithms for the transductive setting on DrugBank and Twosides datasets

Method	DrugBank				Twosides			
	ACC	AUC	AP	F1	ACC	AUC	AP	F1
MR-GNN	96.04 ± 0.05	98.87 ± 0.04	98.57 ± 0.06	96.10 ± 0.05	76.23 ± 0.23	85.00 ± 0.22	84.32 ± 0.35	77.88 ± 0.35
MHCADDI	83.80 ± 0.27	91.16 ± 0.31	89.26 ± 0.37	85.06 ± 0.31	-	88.20	-	-
SSI-DDI	96.33 ± 0.09	98.95 ± 0.08	98.57 ± 0.14	96.38 ± 0.09	78.20 ± 0.14	85.85 ± 0.13	82.71 ± 0.14	79.81 ± 0.16
GAT-DDI	89.81 ± 1.00	95.21 ± 0.70	93.56 ± 0.90	90.18 ± 0.74	50.00	50.00	50.00	-
GMPNN-CS	95.30 ± 0.05	98.46 ± 0.01	97.94 ± 0.02	95.39 ± 0.05	82.83 ± 0.14	90.07 ± 0.12	87.24 ± 0.12	84.08 ± 0.14
SA-DDI	96.23 ± 0.01	98.80 ± 0.02	98.36 ± 0.04	96.29 ± 0.09	87.45 ± 0.03	93.17 ± 0.04	90.51 ± 0.08	88.35 ± 0.04
DSN-DDI	**96.94 ± 0.02**	**99.47 ± 0.01**	**99.37 ± 0.02**	**96.93 ± 0.02**	**98.83 ± 0.04**	**99.90 ± 0.01**	**99.89 ± 0.01**	**98.83 ± 0.04**
Improvement	0.61(0.63)	0.52(0.53)	0.80(0.81)	0.55(0.57)	11.38(13.01)	6.73(7.22)	9.38(10.36)	10.48(11.86)

All values are shown in percentages. The highest value in each column is shown in bold. For performance improvement over the second-best approach, a relative improvement percentage is shown in the bracket.

Table 3. Performance evaluation of DSN-DDI and state-of-the-art algorithms for the inductive setting on DrugBank dataset

Method	S1 Partition (new drug, new drug)			S2 Partition (new drug, existing drug)				
	ACC	AUC	AP	F1	ACC	AUC	AP	F1(%)
MR-GNN	62.63 ± 0.77	70.92 ± 0.84	73.01 ± 1.23	45.81 ± 2.51	74.67 ± 0.33	83.15 ± 0.60	83.81 ± 0.69	69.88 ± 0.86
MHCADDI	66.50 ± 0.62	72.53 ± 0.92	71.06 ± 1.61	67.21 ± 0.59	70.58 ± 0.94	77.84 ± 1.08	76.16 ± 1.45	72.74 ± 0.65
SSI-DDI	65.40 ± 1.30	73.43 ± 1.81	75.03 ± 1.42	54.12 ± 3.46	76.38 ± 0.92	84.23 ± 1.05	84.94 ± 0.76	73.54 ± 1.50
GAT-DDI	66.31 ± 0.61	72.75 ± 0.78	71.61 ± 1.00	68.68 ± 0.60	69.83 ± 1.41	77.29 ± 1.63	75.79 ± 1.95	73.01 ± 0.85
GMPNN-CS	68.57 ± 0.30	74.96 ± 0.40	75.44 ± 0.50	65.32 ± 0.23	77.72 ± 0.30	84.84 ± 0.15	84.87 ± 0.40	78.29 ± 0.16
SA-DDI	67.15 ± 0.88	73.62 ± 1.25	73.39 ± 1.40	63.40 ± 1.53	75.55 ± 1.12	82.95 ± 1.05	84.11 ± 0.92	71.94 ± 1.57
DSN-DDI	**73.42 ± 1.29**	**81.79 ± 1.12**	**81.82 ± 1.48**	**70.34 ± 0.98**	**81.92 ± 1.20**	**91.01 ± 0.76**	**91.09 ± 0.93**	**80.18 ± 1.49**
Improvement	4.85(7.07)	6.83(9.11)	6.38(8.46)	1.66(2.42)	4.2(5.4)	6.17(7.27)	6.15(7.24)	1.89(2.41)

All values are shown in percentages. The highest value in each column is shown in bold. For performance improvement over the second-best approach, a relative improvement percentage is shown in the bracket.

Furthermore, to avoid potential bias in the unknown drug selection, we repeated this process three times in parallel and thus made 3-fold cross validation for the inductive setting.

Previous studies have proved that the chemical structures of new drugs in the test set are very different from existing drugs in the training set due to the large differences on scaffolds [18, 19]. As shown in Table 3, the scores of all evaluation metrics are obvious lower than those obtained in the transductive setting, which indicates accurate DDI prediction for unseen drugs is much more difficult. Furthermore, similar to the performance on the transductive setting, DSN-DDI also achieved the best performance on all evaluation metrics when compared with the state-of-the-art algorithms. Notably, our algorithm outperformed the second-best algorithm by a large margin e.g. relative improvements of 9.11 and 7.27% of AUC on S1 and S2 partitions, respectively. These results indicate the effective countermeasures of DSN-DDI that it does not only consider the structure of an individual drug, but also learns interactions from the drug pair, which greatly compensates for the lack of prior knowledge and interactive information of unseen drugs.

Ablation study for the effectiveness of model design

To study where the performance gains come from, we further conducted comprehensive ablation studies to investigate the importance of various components of DSN-DDI. Specifically, we conducted two sets of ablation studies with respect to the local- level and global-level representation learning designed in DSN-DDI.

First, we evaluated the key components in the local-level representation learning in DSN-DDI with its three variants. Evaluations were made on Twosides (Table 4) and DrugBank (Supplementary Tables S5 and S6) datasets, respectively. The following variants were evaluated and compared with DSN-DDI:

- wo_repre: a model where the representation extraction layer is removed. This serves to demonstrate the effectiveness of the shared representation learning for two drugs.
- wo_intra: a model in which the intra-view layer is removed. The substructures are derived only from the inter-view representations. This serves to demonstrate the importance of inter-view layer and the importance of simultaneously performing both inter-view and intra-view feature extraction.
- wo_inter: a model in which the inter-view layer is removed. The substructures are derived only from the intra-view representations. This serves to demonstrate the importance of inter-view layer for substructure extraction and DDI prediction.
- wo_inter_update: a model where the node representation is only updated by the information from the intra layer. It means that the hidden representations derived from the inter-view layer are only used for the substructure extraction while it has no direct influence on node feature update. This also serves to demonstrate the effectiveness of inter-view layer to node representation learning.

As shown in Table 4, Supplementary Tables S5–S7, the full DSN-DDI architecture outperformed all variants, which indicates the effectiveness of all proposed modules. We summarize the conclusions as follows: (1) inter-view contributes most to DSN-DDI, especially for the Twosides dataset, since the performance of the variant wo_inter decreased significantly on both transductive

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

Table 4. Performance evaluation of DSN-DDI and its five variants on Twosides dataset for the transductive setting

Variant	ACC (%)	AUROC (%)	AP (%)	F1 (%)
wo_repre	96.88	99.50	99.43	96.89
wo_intra	97.89	99.77	99.75	97.90
wo_inter	80.69	88.04	84.96	81.81
wo_inter_update	85.37	92.66	91.15	85.83
wo_SAGPool	98.66	99.89	99.87	98.66
wo_Pool	94.34	98.56	98.35	94.41
wo_co-attention	97.92	99.71	99.67	97.92
DSN-DDI	**98.83**	**99.90**	**99.89**	**98.83**

The highest value in each column is shown in bold.

setting (Table 4) and inductive setting (Supplementary Table S7). Compared with DrugBank dataset, Twosides dataset has less drugs and more DDI triplets, which provides more inter information for the same drug pair during model training. Consequently, removing the inter-view layer greatly reduced the acquisition of DDI information and thus resulted in a poorer performance. Furthermore, the fact that the performance of two variants wo_inter and wo_intra declined to some extent implies that it is beneficial to learn drug–drug representations jointly from dual-view perspective rather than a respective view. (2) The performance dropped on wo_repre indicates that the leading sub-network with the shared parameters learns the common representations for all drugs and thus contributes to model performance gains. (3) When compared the variant wo_inter_update and DSN-DDI, the performance has also declined with a large margin, which indicates that besides for substructure extraction and final DDI prediction, the inter-view information is useful for drug representation learning when directly incorporated into node update process.

We further evaluated the usefulness of the components for global-level representation learning in DSN-DDI, including the following baselines:

- wo_SAGPool: a model where the SAGPooling layers is replaced by a simple sum function, without distinguishing the importance of nodes. In this setting, both the inter and intra drug embeddings are computed by the sum readout function. This serves to demonstrate the importance of performing SAGPooling for substructure extraction.
- wo_Pool: a model where the pooling layer is replaced by a maximum value operation, which means the global drug representation is computed as the largest value among all node representations.
- wo_co-attention: a model where the co-attention operation is removed from DSN decoder. Although the final DDI prediction is still computed from the pairwise interactions between substructures of the drug pair, all combinations of pairwise interactions have the same contribution to DDI prediction. As the co-attention mechanism were also used in previous studies [18], this serves to evaluate its contribution to our performance gains.

As shown in Table 4, the performance of the variant wo_Pool decreased significantly, which indicates the global pooling layer is important to learn the drug global representations from node local representations. In contrast, a minor performance drop was detected for the variant wo_SAGPool, which implies that the improvement brought from global pooling layer does not depend on any specific operations. The minor performance drop on wo_co-attention indicates that the co-attention mechanism only plays a minor role in model performance gains, which reflects substructures are distinct and robust and thus can be directly used for DDI prediction without such complicated attention mechanism. In summary, all components in DSN-DDI for the two-level representation learning contribute to DDI prediction, whereas the novel designed operations in the local-level representation learning are more useful.

Real-world DDI applications

To demonstrate the usefulness of DSN-DDI for real-world DDI applications, we evaluated the DDI prediction for the newly FDA-approved drugs by the model trained with existing information of old drugs. We collected the FDA drug approval information [32] for all the drugs in DrugBank dataset and divided them into two parts according to the drug approval date before or after the year of 2017. The DDI triplets containing two old drugs form the training set, whereas the remaining DDI triplets containing at least one new drug are recruited into the test set (Supplementary Table S8 for more details). We trained and evaluated DSN-DDI with the same hyperparameters used in the inductive setting. Furthermore, we picked the three state-of-the-art DDI prediction algorithms from the above performance evaluation, SSI-DDI, GMPNNN-CS and SA-DDI for comparison. These three algorithms were reproduced on the same dataset with their default hyperparameters. As shown in Supplementary Table S9, DSN-DDI outperformed SSI-DDI, GMPNN-CS and SA-DDI on all four metrics, ACC, AUROC, AP and F1 by a large margin. These results consolidate DSN-DDI has captured the generalized information of DDI among different drugs and thus it is applicable for new approved drugs. Figure 3 illustrates the detailed ROC curve and PR curve on the test set of four algorithms. DSN-DDI achieved significantly larger areas under both the ROC and PR curves than SSI-DDI, GMPNN-CS and SA-DDI, respectively, which also indicates DSN-DDI can distinguish the positive DDI effects from negative ones well.

Model transferability to synergistic drug combination prediction

In DSN-DDI, the dual-view message passing mechanism could also serve as a generalized framework for many tasks in drug discovery. It will be of interest to examine the usefulness and transferability of DSN-DDI for other tasks with drug pairs as input. Complex diseases, such as hypertension [33] and cancer [34] are often caused by multiple factors. Therefore, combination therapy with several drugs is a common and effective method in clinical treatment. Consequently, reliable and accurate drug combinations for complex diseases have significant meanings for clinical treatment.

Many machine learning-based and deep learning-based algorithms [35–39] have been used to predict synergistic drug

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

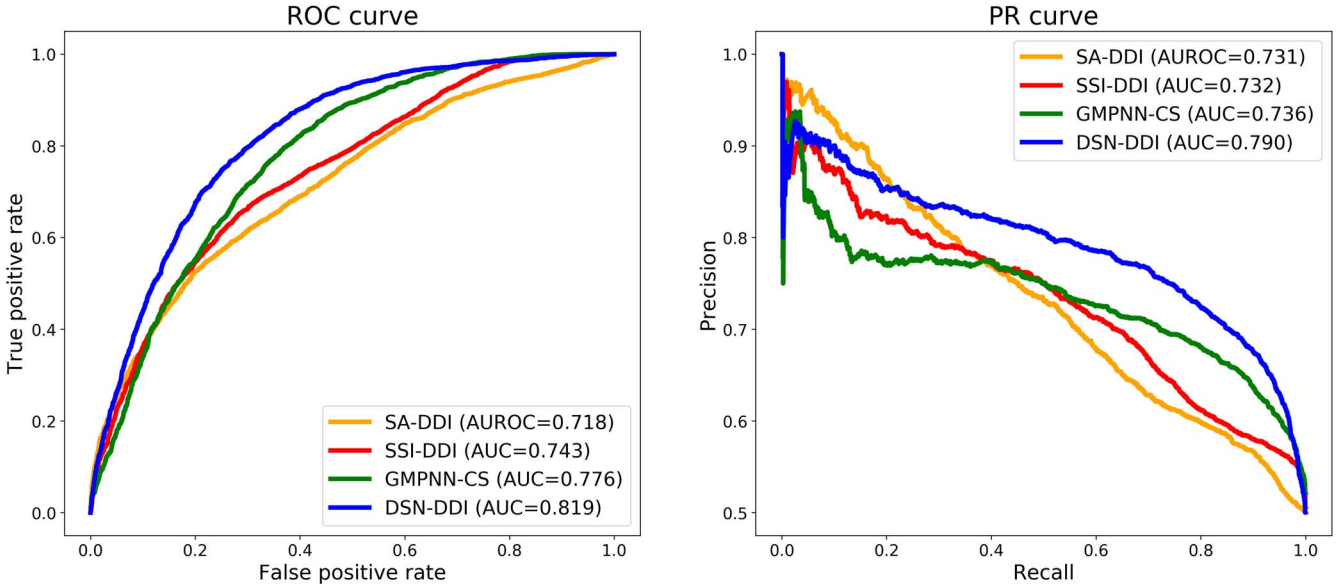

Figure 3. Performance evaluation of SSI-DDI, GMPNN-CS, SA-DDI and DSN-DDI for new FDA-approved drugs. (**A**) The receiver operating characteristic (ROC) curve of four algorithms evaluated on the test set. (**B**) The prevision versus recall (PR) curve of four algorithms evaluated on the test set.

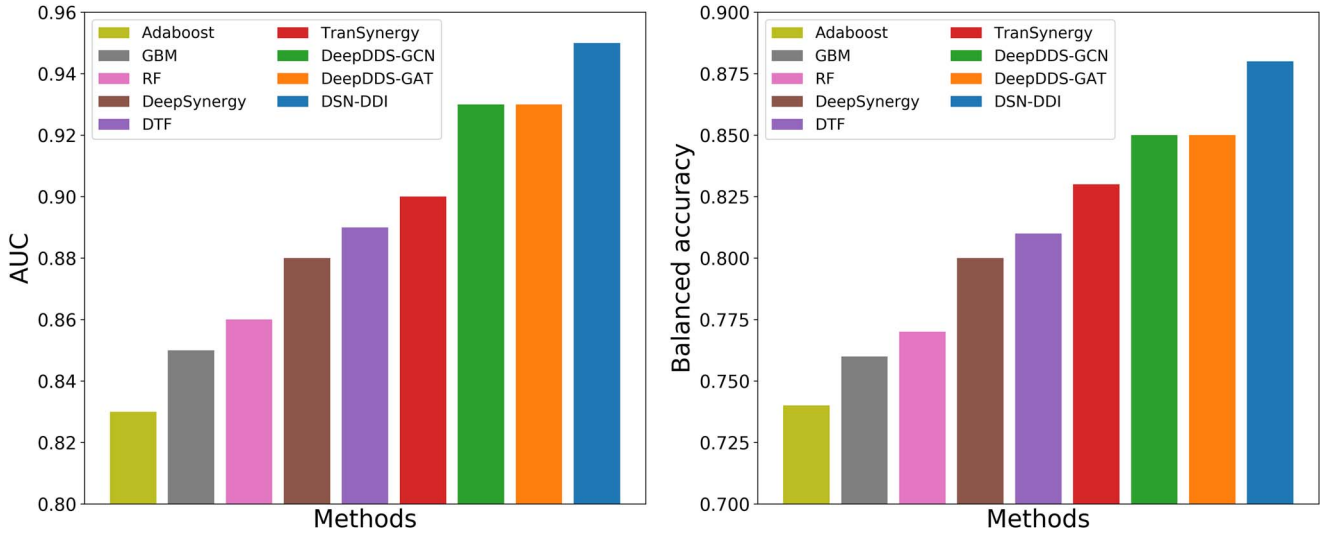

Figure 4. Performance evaluation of DSN-DDI and state-of-the-art algorithms for drug combination prediction. (**A**) Area under the receiver operator characteristics curve. (**B**) Balanced accuracy. All algorithms are evaluated in 5-fold cross validation.

combinations in recent years. Here, we chose the task of synergistic drug combination prediction to examine the transferability of DSN-DDI. Notably, DDI prediction is to predict the side effects between two drugs, whereas synergistic drug combination prediction is to predict the effects between the combination of two drugs and the specific diseases. Specifically, drug combination prediction is to predict whether a triplet $(drug_a, drug_b, cell)$ (i.e. a drug pair and a corresponding cell line) exists. Although this task is obviously different from DDI prediction we demonstrated in this study, we replaced the relation in DDI prediction with cell line and thus applied DSN-DDI for drug combination prediction. A balanced benchmark dataset with 12 415 unique drug pair-cell line combinations [40] were chosen for evaluation. It consists of 36 anticancer drugs and 31 human cancer cell lines. Following the same experiment settings in Wang *et al.* [40], we conducted 5-fold cross validation to evaluate the performance of our model. For model training, we adopted the same hyperparameters and configurations in DDI prediction to train DSN-DDI for drug combinations prediction. Three machine learning-based algorithms i.e. Random Forests, Gradient Boosting Machines and Adaboost, and four state-of-the-art deep learning-based algorithms i.e. TranSynergy [39], Deep Tensor Factorization [41], DeepSynergy [38] and DeepDDS [40] were chosen for comparison.

As shown in Figure 4, compared with other state-of-the-art algorithms, DSN-DDI achieved the best performance on AUC and balanced accuracy, indicating its outstanding ability for drug combinations prediction. Thanks to the dual-view message passing mechanism, DSN-DDI can fully exploit the information embedded in the single drug and the drug-pair simultaneously and establish a reliable connection to the corresponding cell line. Although DSN-DDI was originally designed for DDI prediction, it still performed well in drug combination prediction. The results have proved that our model has good transferability to tasks

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

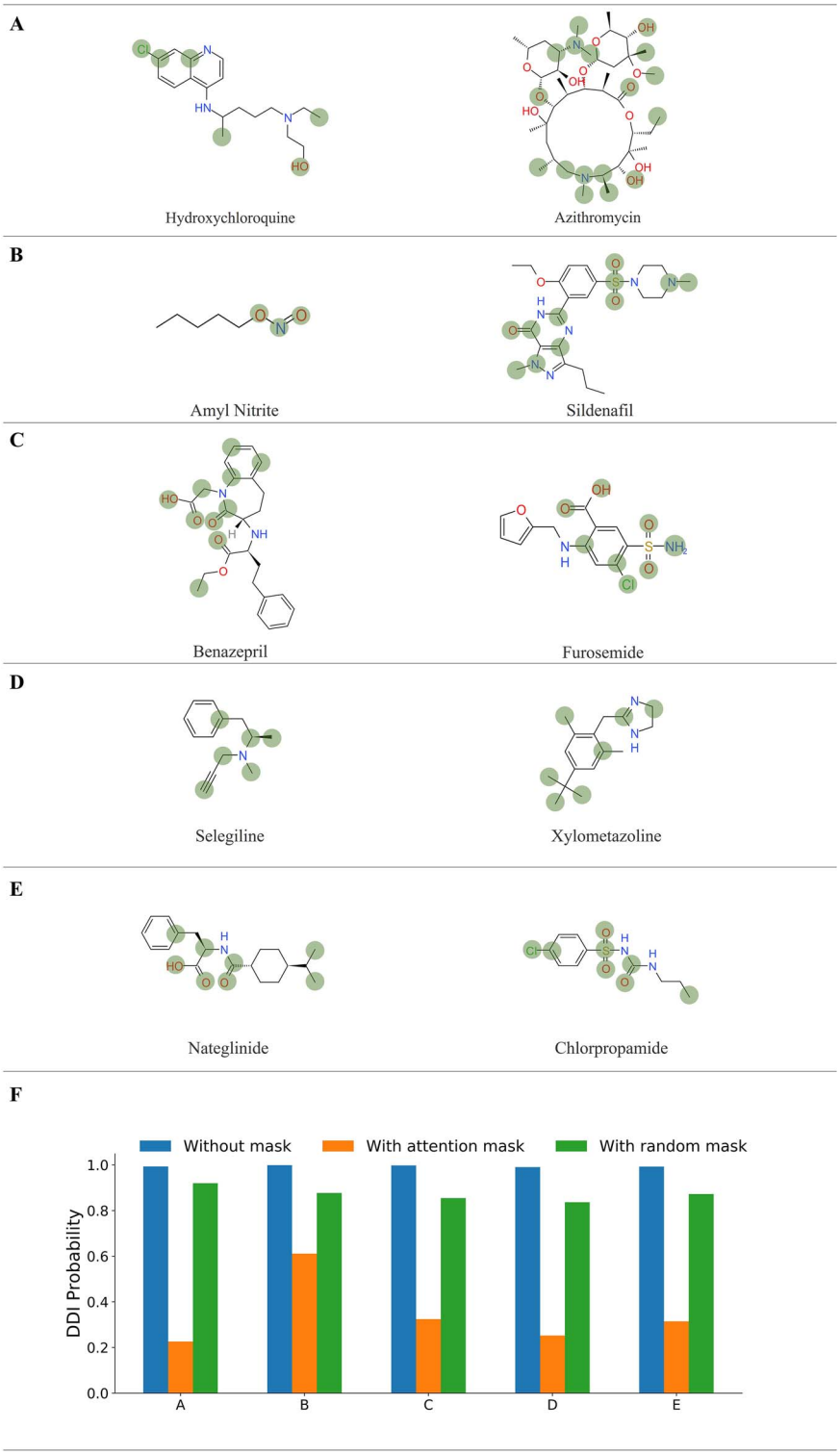

Figure 5. Visualization and evaluation of extracted substructures for DDI prediction. (**A**) to (**E**) are five DDI prediction cases. The learned substructures consist of the atoms highlighted with green shadows, which are selected according to high attention scores in drug global representation learning. (F) Comparison of DDI predictions with or without extracted substructures for all five cases in panel (**A**) to (**E**). 'without mask' refers to the prediction by DSN-DDI, 'with attention mask' refers to the prediction by the model after removing the highlighted nodes and 'with random mask' refers to the prediction by the model after removing the same number of nodes that were randomly selected.

with drug pairs as input. Furthermore, the insights and novel designed modules in DSN-DDI could be applied for more two-body interaction problems, such as drug–target interaction prediction, protein–protein interaction prediction, which will be explored in our future study.

Discussion

In this work, we demonstrate the superiority of DSN-DDI for DDI prediction. By both local representation learning and global representation learning, our algorithm can extract substructures

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

which are used for final DDI prediction. To further investigate the learned substructures, we extracted and illustrated the valid substructures with important atoms for five DDI cases and examined the performance drops after removing these atoms in Figure 5. We obtained the contribution scores of each atom of drugs in each block by the SAGpooling operation. Then we calculated the averaged contribution score in all blocks as the total attention score for each atom. The score represents the importance of atoms for DDI prediction. We extracted 30% of atoms for each drug with the highest attention scores to form substructures. As shown in Figure 5, the highlighted atoms are mostly concentrated on the functional groups and structural junctions of drugs. These atoms learned by DSN-DDI are gathered to some certain regions of the drug chemical structure to form steady substructures. As a result, DSN-DDI has made almost perfect predictions (i.e. a predicted score more than 0.99) for all five cases. When masking 10% of nodes with the highest attention scores in each drug, the DDI predicted probabilities scores have significantly decreased. As a comparison, masking the same number of nodes that were randomly selected only led to a moderate performance drop. These results imply that the incomplete drug structures disrupt the drug topology and affect the message passing in our model from dual-view. Furthermore, masking nodes with the largest attention scores has much larger impacts on DDI prediction. Therefore, these important atoms and the substructures learned by DSN-DDI have played a key role in DDI prediction.

To further explore the interpretability of the substructures generated by DSN-DDI, we extracted and illustrated the valid substructures with important atoms between dicoumarol and the other five drugs. By using the same visualization method, we identified the important nodes and effective substructures for the five DDIs in Supplementary Figure S1. It is worth noticing that DSN-DDI extracted the similar valid substructures i.e. barbituric acid in all five cases including pentobarbital, amobarbital, secobarbital, primidone and methylphenobarbital, respectively. The previous study has demonstrated that drugs with barbituric acid substructure can enhance the activity of human liver microsomes, thereby reducing the efficacy of dicoumarol [42]. Therefore, these important atoms and the substructures learned by DSN-DDI have good agreement with experimental and pharmacological results.

Conclusions

In this work, we have presented a dual-view substructure learning framework for predicting the possible polypharmacy side effects of drug pairs. Comprehensive experiments have confirmed the state-of-the-art performance of DSN-DDI for DDI prediction on both the transductive and the inductive settings. The DSN-DDI has achieved significant improvements with 13.01% accuracy on Twosides in the transductive setting compared with the state-of-the-art methods. More importantly, DSN-DDI has achieved significant performance improvement in the more challenging inductive scenario, with an average improvement of 7.07% on S1 Partition scheme and 5.40% on S2 Partition scheme in accuracy over the second-best algorithms. By performing intra-view message passing within each drug, as well as inter-view message passing between two drugs, we have demonstrated the power of integrating joint drug–drug information during the substructure representation learning for DDI prediction. Future research could further improve the generalization ability of the model for new drugs in the inductive learning setting, which approximates a real-world scenario where there is a new drug without any prior associated drug interactions.

Key Points

- How to make full use of both drug representation and drug pair representation is key to substructure extraction and DDI prediction
- We propose DSN-DDI which employs local and global representation learning modules iteratively and learns drug substructures from the intra-view and the inter-view simultaneously.
- Comprehensive evaluations demonstrate that DSN-DDI has achieved the state-of-the-art performance for both transductive setting and inductive setting and shows the usefulness for real-world DDI applications.
- DSN-DDI exhibits good transferability and can serve as a generalized framework in the drug discovery field.

Author contributions

T.W. led the study. T.W., X.Z. and B.S. conceived and designed the study. T.W. and S.Z. carried out model design. Z.L. carried out experiments on transductive setting and inductive setting. Z.L., S.Z. and T.W. carried out ablation study. Z.L. and T.W. carried out real-world DDI applications. T.W. and S.Z. wrote the manuscript. B.S., X.Z. and T.L. contributed to manuscript draft.

Data and materials availability

The code to reproduce the figures and results of this article can be found at https://github.com/microsoft/Drug-Interaction-Research/tree/DSN-DDI-for-DDI-Prediction. All raw data including DrugBank and Twosides datasets are from the original manuscript of GMPNN-CS.

Supplementary data

Supplementary data are available online at https://academic.oup.com/bib.

References

1. Tatonetti NP, Ye PP, Daneshjou R, *et al.* Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012;**4**:125ra131–1.
2. Jia J, Zhu F, Ma X, *et al.* Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 2009;**8**:111–28.
3. Han K, Jeng EE, Hess GT, *et al.* Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol* 2017;**35**:463–74.
4. Yu H, Mao K-T, Shi J-Y, *et al.* Predicting and understanding comprehensive drug-drug interactions via semi-nonnegative matrix factorization. *BMC Syst Biol* 2018;**12**:101–10.
5. Shi J-Y, Mao K-T, Yu H, *et al.* Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. *J Chem* 2019;**11**:1–16.
6. Li X, Xu Y, Cui H, *et al.* Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif Intell Med* 2017;**83**:35–43.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023

7. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. *J Biomed Inform* 2017;**70**:54–64.
8. Kastrin A, Ferk P, Leskošek B. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PLoS One* 2018;**13**: e0196865.
9. Huang K, Xiao C, Hoang T *et al.* Caster: predicting drug interactions with chemical substructure representation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA: AAAI Press. 2020;**34**:702–709.
10. Deng Y, Xu X, Qiu Y, *et al.* A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* 2020;**36**:4316–22.
11. Chen Y, Ma T, Yang X, *et al.* MUFFIN: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics* 2021;**37**: 2651–8.
12. Zhang Y, Qiu Y, Cui Y, *et al.* Predicting drug-drug interactions using multi-modal deep auto-encoders based network embedding and positive-unlabeled learning. *Methods* 2020;**179**: 37–46.
13. Xu N, Wang P, Chen L, *et al.* Mr-gnn: multi-resolution and dual graph neural network for predicting structured entity interactions. arXiv preprint arXiv:1905.09558. 2019.
14. Yu Y, Huang K, Zhang C, *et al.* SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics* 2021;**37**:2988–95.
15. Feeney A, Gupta R, Thost V, *et al.* Relation matters in sampling: a scalable multi-relational graph neural network for drug-drug interaction prediction. arXiv preprint arXiv:2105.13975. 2021.
16. M. W. Harrold, R. M. Zavod. (Taylor & Francis, 2014).
17. B. Jin, H. Yang, C. Xiao *et al.* Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. In: *Proceedings of the AAAI conference on artificial intelligence*. San Francisco, California USA: AAAI Press, 2017; **31**.
18. Nyamabo AK, Yu H, Shi J-Y. SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction. *Brief Bioinform* 2021;**22**:bbab133.
19. Nyamabo AK, Yu H, Liu Z, *et al.* Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Brief Bioinform* 2022;**23**:bbab441.
20. Yang Z, Zhong W, Lv Q, *et al.* Learning size-adaptive molecular substructures for explainable drug–drug interaction prediction by substructure-aware graph neural network. *Chem Sci* 2022;**13**: 8693–703.
21. Deac A, Huang Y-H, Veličković P, *et al.* Drug-drug adverse effect prediction with graph co-attention. arXiv preprint arXiv:1905.00534. 2019.
22. Wang Y, Min Y, Chen X *et al.* Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: *Proceedings of the Web Conference*. 2021;2921–33.
23. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018;**34**:i457–66.
24. Ma T, Xiao C, Zhou J, *et al.* Drug similarity integration through attentive multi-view graph auto-encoders. arXiv preprint arXiv:1804.10850. 2018.
25. Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. arXiv preprint arXiv:1710.10903. 2017.
26. Lu J, Yang J, Batra D, *et al.* Hierarchical question-image co-attention for visual question answering. *Adv Neural Inf Process Syst* 2016;**29**:4071–9.
27. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
28. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.
29. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;**28**:31–6.
30. Landrum G. RDKit: a software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* Academic Press. 2013.
31. Dewulf P, Stock M, De Baets B. Cold-start problems in data-driven prediction of drug–drug interaction effects. *Pharmaceuticals* 2021;**14**:429.
32. Mullard A. 2020 FDA drug approvals. *Nat Rev Drug Discov* 2021;**20**: 85–90.
33. Giles TD, Weber MA, Basile J, *et al.* Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study. *The Lancet* 2014;**383**:1889–98.
34. Kim Y, Zheng S, Tang J, *et al.* Anticancer drug synergy prediction in understudied tissues using transfer learning. *J Am Med Inform Assoc* 2021;**28**:42–51.
35. Sałat R, Sałat K. The application of support vector regression for prediction of the antiallodynic effect of drug combinations in the mouse model of streptozocin-induced diabetic neuropathy. *Comput Methods Programs Biomed* 2013;**111**:330–7.
36. Liu H, Zhang W, Zou B, *et al.* DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res* 2020;**48**:D871–81.
37. Barretina J, Caponigro G, Stransky N, *et al.* The cancer cell line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7.
38. Preuer K, Lewis RP, Hochreiter S, *et al.* DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018;**34**:1538–46.
39. Liu Q, Xie L. TranSynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput Biol* 2021;**17**:e1008653.
40. Wang J, Liu X, Shen S, *et al.* DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022;**23**:bbab390.
41. Chen Z, Gai S, Wang D. Deep tensor factorization for multi-criteria recommender systems. In: *2019 IEEE International Conference on Big Data (Big Data)*. New York, USA: IEEE 2019;1046–51.
42. Ioannides C, Parke DV. Mechanism of induction of hepatic microsomal drug metabolizing enzymes by a series of barbiturates. *J Pharm Pharmacol* 1975;**27**:739–46.

Downloaded from https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbac597/6966537 by Zhejiang University Library yuquan campus user on 03 January 2023