GNNSynergy: A Multi-view Graph Neural Network for Predicting Anti-cancer Drug Synergy

Zhifeng Hao, Jianming Zhan, Yuan Fang, Min Wu, and Ruichu Cai*

Abstract—Drug combinations play very important roles in cancer therapy, as they can enhance curative efficacy and overcome drug resistance. Due to the increasing size of combinatorial space, experimental screening for all the drug combinations becomes infeasible in practice. Therefore, there is a great need to develop accurate computational approaches that can predict potential drug combinations to direct the experimental screening. In this paper, we propose a novel method called GNNSynergy to learn drug embeddings for drug synergy prediction. Given a specific cancer cell line, we propose a multi-view graph neural network framework which considers the current cell line as main view while other cell lines from the same tissue as sub-views. In each view, we first construct different graphs to describe drug synergistic and antagonistic interactions, and adopt graph neural network as encoder to learn drug embeddings. We further combine both the main view and sub-views via an attention mechanism to derive the final drug embeddings for drug synergy prediction. We perform extensive experiments on DrugComb database and the experimental results demonstrate that our proposed GNNSynergy significantly outperforms state-of-the-art methods for novel synergistic drug combination prediction.

Index Terms—Drug combination, Cancer cell lines, Multi-view, Attention mechanism, Graph neural network.

1 Introduction

DRUG combination, as an important concept for improving efficacy of complex disease treatment [1, 2], has received a lot of attention in the cancer therapeutics. Compared to drug monotherapy, combination therapy can help to reduce or even eliminate drug resistance. In addition to cancers, drug combinations have also been actively studied for the treatment of other diseases, such as AIDS [3], fungal and bacterial infections [4, 5, 6, 7].

Drug combinations can be identified by clinical experiments or high-throughput screening methods (HTS). As the number of drugs approved by FDA has been continuously growing, the number of possible pairwise drug combinations has become too huge. Even though HTS can produce a large number of measurements in reasonable time [8, 9], it is still impracticable to exhaust all these drug combinations. Nevertheless, HTS experiments have generated quite a number of known drug combinations across a couple of cell lines for various cancer types [8, 10]. Such data generated by HTS methods enables computational pre-screening for synergistic drug pairs, which would be a useful complement to wet-lab experiments.

Various machine learning methods have been proposed for synergistic drug combination prediction [11, 12, 13], including traditional machine learning methods and deep learning models. Traditional machine learning methods manually extract features from drug chemical structures, cell-line gene expression profiles and dose response curves, and then feed the extracted features to Random Forest [14, 15] and XGBoost [16] to predict synergy scores across different cell lines. An ensemble method called EPSDC [17] was proposed to predict drug combinations from multiple data sources by integrating various base prediction models (e.g., SVM, Naive Bayes, etc). Deep learning methods also take the drug chemical descriptors and cell-line gene expression profiles as inputs, and directly feed them into different network structures, e.g., deep feed-forward network [18, 19] and Transformer [20], to learn latent drug features automatically for drug synergy prediction. Kim *et al* [21] used drug encoder and cell line encoder to exact the features, and fed these features into a feed forward network for drug synergy prediction. Hu *et al* [22] proposed a deep neural network model termed DTSyn based on a multi-head attention mechanism to identify novel drug combinations. Both traditional machine learning and deep learning methods above can predict the synergy score between two drugs in a specific cell line by using cell line profiles as features. However, most of the existing works did not take into account the data for other similar cell lines from the same tissue to improve their predictions.

Moreover, known drug combinations can be modeled as a network/graph, where nodes are drugs and edges show their synergistic effects. It is denoted as a drug-drug synergy (DDS) graph, which enables modeling the long-range dependencies between drugs. However, above machine learning methods fail to consider such graph-structured DDS data for drug synergy prediction. Meanwhile, some network-based methods [23, 24, 25] have been proposed for drug synergy prediction. The authors in [23] quantified the relationship between drug targets and disease proteins in human protein-protein interaction (PPI) networks to predict novel drug combination pairs. A method called NEWMIN was proposed in [24] to predict drug

- *Zhifeng Hao, Jianming Zhan and Ruichu Cai are currently with School of Computer Science, Guangdong University of Technology, Guangdong, China 510006. E-mail: zfhao@fosu.edu.cn, {zhanjimmy520, cairuichu}@gmail.com.*
- *Yuan Fang is currently with School of Computing and Information Systems, Singapore Management University, Singapore 178902. E-mail: yfang@smu.edu.sg.*
- *Min Wu is currently with the Institute for Infocomm Research (I2R), A*Star, Singapore 138632. Email: wumin@i2r.a-star.edu.sg.*
- ** Corresponding author*

Manuscript received April 19, 2005; revised August 26, 2015.

combinations via SkipGram-based network embedding in multiplex network (i.e., 6 drug-drug similarity networks). In [25], the authors developed a method called ISDCSMP to prioritize synergistic drug combinations via integrating multi-source information in a heterogeneous network. These three recent studies also did not exploit DDS graph data for drug combination prediction.

Recently, graph neural networks (GNN) [26] have been successfully applied to tackle various bioinformatics tasks, including disease gene prediction [27], drug discovery [28], synthetic lethality prediction [29], etc. GNN methods have also been proposed for cell-line-specific drug synergy prediction [30, 31, 32]. For example, Jiang *et al* learned drug embeddings via graph neural network from a heterogeneous network with protein-protein interactions and drug-target interactions. Wang *et al* proposed a method called DeepDDS using graph neural network to obtain the drug embeddings from drug molecular graphs [31]. Different from [30, 31], the aforementioned DDS graph can also be fed into GNN to learn drug embedding for drug synergy prediction. However, there are two main challenges to apply GNN on DDS graph for this important task. First, drug pairs have different synergistic effects [19], i.e., synergism (positive Loewe scores) and antagonism (negative Loewe scores). If a single DDS graph contains different drug relationships, it would be very challenging to apply GNN methods on this DDS graph as most of them tend to work in a homogeneous graph where nodes and edges belong to the same relationship. Second, there are multiple cell lines from the same tissue and it may help to boost the model performance by integrating the data from other cell lines. How to effectively integrate multiple cell lines for drug synergy prediction remains a challenge.

To address the above challenges, we propose a novel multi-view graph neural network method named GNNSynergy to predict drug synergy. We build individual model to predict drug synergy scores for each cancer cell line (i.e., cell-line-specific model). Given a specific cell line, we propose a multi-view framework which considers the current cell line as main view while other cell lines from the same tissue as sub-views. In each view (i.e. cell line), we first construct two DDS graphs to describe different types of drug synergistic effects, and design graph neural networks as encoders to learn drug embeddings. We further combine both the main view and sub-views to derive the final drug embeddings for drug synergy prediction. Experimental results on DrugComb dataset show that our proposed GNNSynergy model significantly outperforms state-of-the-art methods for drug synergy prediction.

Note that Kim *et al* [21] also used other cell lines to build their prediction model for the target cell line. They applied transfer learning for the understudied target cell line where the labelled data is insufficient. Our setting for GNNSynergy is clearly different from [21]. We perform supervised learning for each cell line and then conduct multi-view (multiple cell line) integration for drug feature learning. Overall, our main contributions in this paper are summarized as follows.

- We construct multiple drug-drug synergy (DDS) graphs to describe different types of synergistic effects. Drug embeddings are then learnt from these DDS graphs for drug synergy prediction via graph neural networks.
- We design a multi-view framework for drug synergy prediction, which can effectively integrate the data from other relevant cell lines within the same tissue. To the best of our knowledge, this is the first attempt to tackle this problem with multi-view graph neural networks.
- Our experimental results demonstrate that our proposed GNNSynergy significantly outperforms eight state-of-the-art methods.

2 Related Work

In this section, we first introduce existing machine learning methods for drug synergy prediction. We also introduce the recent development of graph neural network techniques for various bioinformatics tasks.

2.1 Machine Learning for Drug Synergy Prediction

Traditional machine learning methods have been shown to be very effective for drug synergy prediction. In [33], Li *et al*. proposed a probability ensemble approach (PEA) for drug combination prediction. In particular, they calculated the synergy score of a drug pair by integrating 6 molecular and pharmacological features with a Bayesian network. TreeCombo predicted the synergy scores of drug pairs with extreme gradient boosted trees (i.e., XGBoost) by taking chemical and physical descriptors of drugs and gene expression levels of cell lines as features [16]. Li *et al*. proposed to integrate cross-cell and cross-drug information and feed them into random forest for drug synergy prediction [14]. Similarly, TAIJI also utilized random forest, as well as the data from drug pharmacological and molecular properties, to predict drug synergy [15].

Recently, deep learning methods have been proposed for synergistic drug combination prediction. DeepSynergy [18], as the first deep learning method for drug synergy prediction, implemented a feed-forward network taking drug chemical features and cell line gene expression profiles as inputs. MatchMaker [19] further improved DeepSynergy, and trained two parallel feed-forward networks to learn latent drug representations for drug synergy prediction in a specific cell line. TranSynergy [20] proposed a self-attention Transformer for drug synergy prediction. It focused on incorporating the information from the gene-gene interactions, cell-line gene dependency, and drug-target interactions to predict the synergistic drug combinations.

2.2 Graph Neural Networks

Graph neural networks are deep neural networks modeling the graph data [26]. Various GNN models including graph convolutional network (GCN) [34] and graph attention network (GAT) [35] have been used for drug discovery [28], disease prediction [27], microbe-drug association prediction [36], synthetic lethality prediction [29], etc. Multi-view graph neural networks have also been proposed to fuse multiple data sources in various tasks. Zhang *et al*. proposed

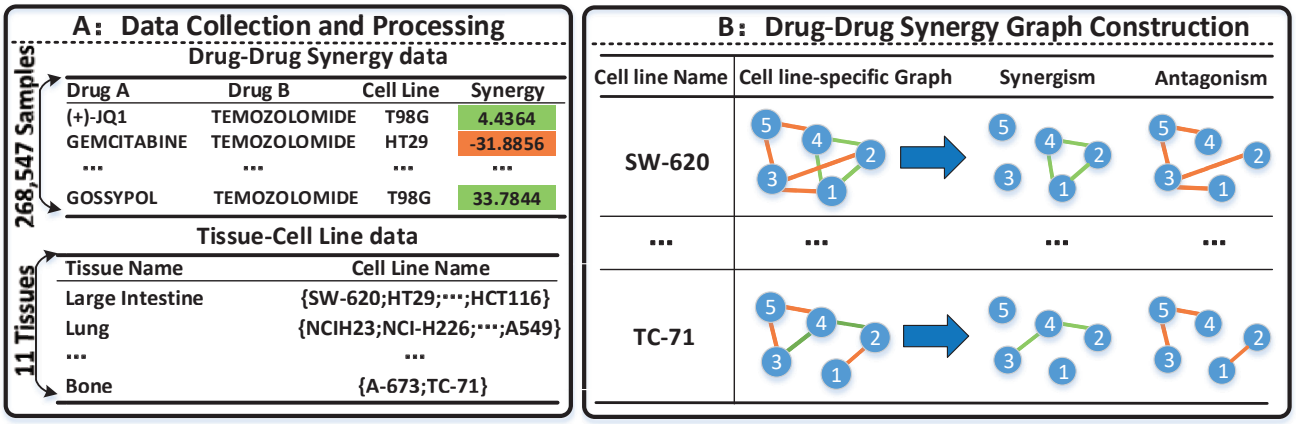


Fig. 1: Data preparation and DDS Graph Construction. (A) Drug synergy data and cell line data are collected from DrugComb database. (B) Two DDS graphs are constructed for each cell line to describe synergistic and antagonistic effects of drug pairs. Nodes represent drugs, and an edge represents the interaction (synergism or antagonism) between two nodes/drugs.

a multi-view graph convolutional network to integrate multiple modalities of brain images for Parkinson's disease prediction [37]. Fu *et al*. proposed a generic multi-view graph convolution network framework termed as MVGCN for link prediction in biomedical bipartite networks [38]. MVGCN shows very good generalization capacity on six benchmark datasets involving three different link prediction tasks in bipartite networks. Multi-view graph auto-encoder was proposed for drug-drug interaction prediction [39] and synthetic lethality prediction [40]. So far, there are no existing studies working on multi-view GNN to integrate multiple cell lines for drug synergy prediction.

3 METHODS

In this section, we first introduce data preparation and DDS graph construction, and then present the details of our GNNSynergy model.

3.1 Data Preparation and DDS Graph Construction

In this study, we use the data from DrugComb database [41], which originally consists of 4,015 drugs and 466,033 drug combinations in 112 cell lines from 11 different tissues. After removing redundant drug pairs, we collect 268,547 drug pairs in 81 cell lines from 11 tissues as shown in Fig. 1(A). We first select all the drug combination pairs in a specific cell line. For these selected drug combination pairs, they are classified as synergism, additivity or antagonism in [41]. Since additivity means no significant interactions, we construct 2 drug-drug synergy (DDS) graphs for synergism and antagonism only, based on their synergy scores as shown in Fig. 1(B). The synergism graph contains drug pairs with Loewe scores greater than a pre-defined threshold t. The antagonism graph contains drug pairs with scores less than $-t$. t is set as 0 in our experiments.

Above DDS graphs are the main inputs of our GNNSynergy model. We also take the chemical features of drugs as inputs. In particular, we follow MatchMaker [19] and use 541 chemical descriptors for drugs, which are calculated by *ChemoPy* Python library [42]. We further perform a feature normalization process and remove those features with zero values. Finally, we use 424 chemical features for drugs, and these selected chemical descriptors are listed in our supplementary materials. In this work, we focus on predicting the drug synergy scores in each cell line, which is a typical regression task. Next, we will introduce the detailed structure of our GNNSynergy model.

3.2 Overview of GNNSynergy

Fig. 2 shows the overall architecture of our proposed GNNSynergy model for cell-line-specific drug synergy prediction. Given a specific cell line C, we consider it as our main view, while all the other cell lines $C_1, \cdots, C_k$ within the same tissue are considered as sub-views. First, we leverage the single-view encoder to learn the drug embeddings for each cell line, including the main view C and sub-views $C_1, \cdots, C_k$. Second, we conduct a weighted concatenation to integrate the drug embeddings from all the sub-views. In particular, the weight α_j here is a trainable attention score to show the similarity between the sub-view C_j and main view C. Third, we combine the drug embeddings from both main view and sub-views as the final embedding. Lastly, we design a decoder for the final drug embeddings to reconstruct the drug synergy score matrix. Next, we will introduce each step in details.

3.3 Single-View Encoder for Drug Embedding Learning

As shown in Fig. 2(B), we construct 2 DDS graphs for each cell line to represent synergism and antagonism, respectively, and 2 GCNs are employed as feature encoders. Here, we use standard graph convolution [34] as follows.

$$\mathbf{H}^{(l)} = \sigma(\hat{\mathbf{A}}\mathbf{H}^{(l-1)}\mathbf{W}^{(l)}), \tag{1}$$

$$\hat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}}\tilde{\mathbf{A}}\mathbf{D}^{-\frac{1}{2}}, \tag{2}$$

where $\tilde{\mathbf{A}} = \mathbf{I} + \mathbf{A}$ is the adjacency matrix of the input graph with self-loop and $\mathbf{I}$ is the identity matrix. $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a degree matrix with $\mathbf{D}_{ii} = \sum_{j=1}^{n} \tilde{\mathbf{A}}_{ij}$ and $\mathbf{W}^{(l)} \in \mathbb{R}^{n\times d}$ is a layer-specific trainable weight matrix. $\sigma(\cdot)$ is the activation

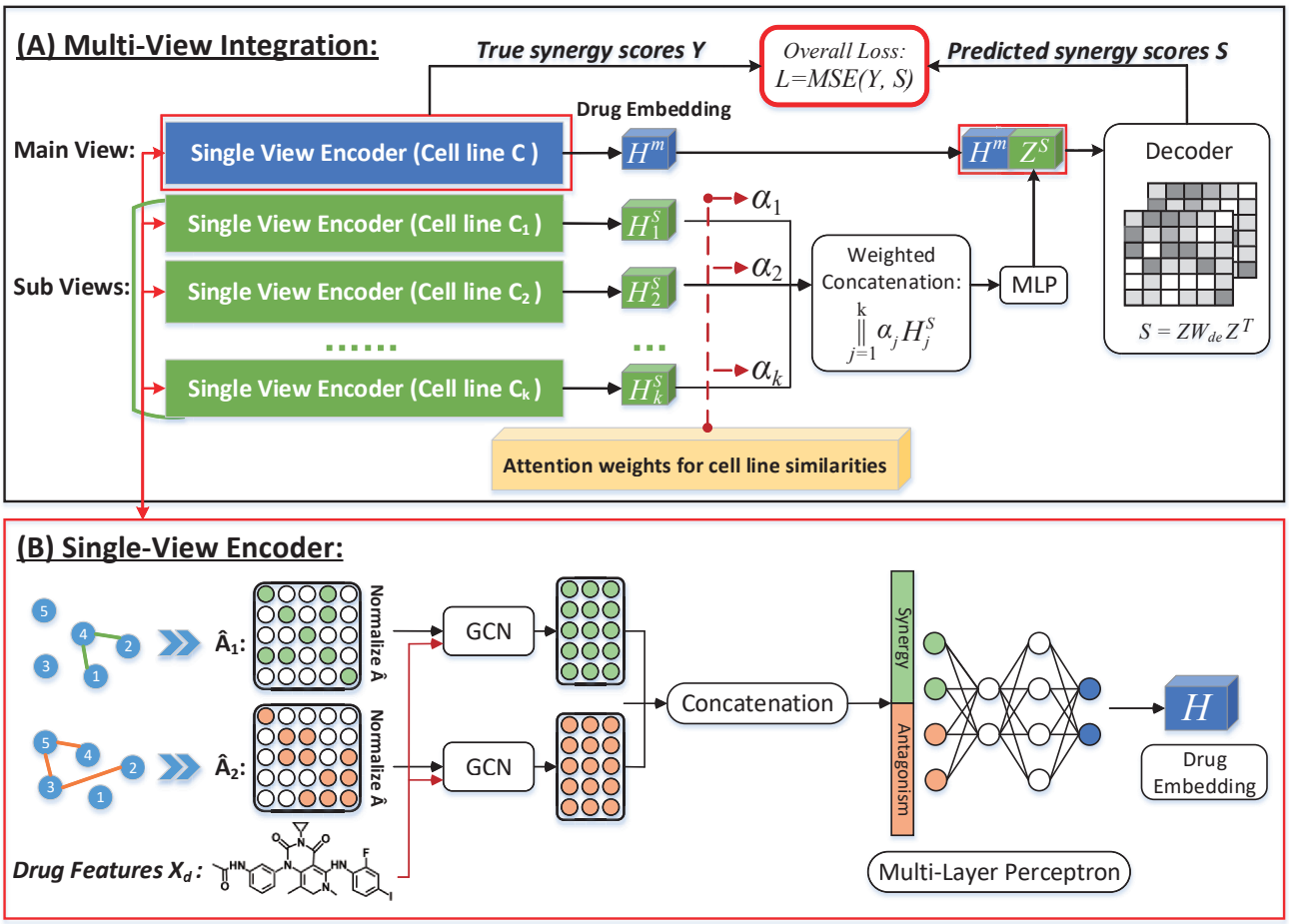


Fig. 2: The architecture of GNNSynergy for cell-line-specific drug synergy prediction. (A) Multi-view encoder integrates multiple cell lines to learn the final drug embeddings, and (B) single-view encoder learns drug embeddings for each cell line.

function (e.g., sigmoid or ReLU). $\mathbf{H}^{(l-1)}$ and $\mathbf{H}^{(l)}$ are the input and the output for the l^{th} layer of the GCN. When $l = 1$, the input $\mathbf{H}^{(0)}$ is the initial node features.

In this paper, the initial features for drugs are derived from their chemical structures, denoted as $\mathbf{X}_d$. We use 1-layer GCN with ReLU as activation function for feature encoding. For example, the GCN output for the synergism graph A_1 is computed in Equation 3.

$$\mathbf{H}_1 = ReLU(\hat{\mathbf{A}}_1\mathbf{X}_d\mathbf{W}_1), \tag{3}$$

where $\mathbf{H}_1 \in \mathbb{R}^{n\times d}$, n is the number of drugs, d is the dimension of drug embedding, and $\mathbf{W}_1$ is a learnable weight matrix. Similarly, the GCN outputs for antagonism graph A_2 are denoted as $\mathbf{H}_2$, respectively.

To learn the drug embeddings for a specific cell line, we further concatenate the embeddings from the two DDS graphs, namely, $\mathbf{H}_1$ and $\mathbf{H}_2$. Then, the concatenated embeddings are fed into a multi-layer perceptron (MLP) for dimension reduction as shown in Equation 4.

$$\mathbf{H} = MLP(\mathbf{H}_1||\mathbf{H}_2), \tag{4}$$

where $MLP(\mathbf{X})$ is the output of a 4-layer MLP with input $\mathbf{X}$ as shown in Fig. 2, and $||$ is the concatenation operation.

As aforementioned, we have a specific cell line as main view, and others in the same tissue as sub-views. Hence, we denote the drug embeddings learned from main view as $\mathbf{H}^m$, and those from sub-views as $\mathbf{H}^s$. For example, the drug embeddings from the sub-view C_j are denoted as $\mathbf{H}^s_j$ $(1 \le j \le k)$ as shown in Fig. 2(A).

3.4 Multi-View Encoder for Cell Line Integration

Different cell lines from the same tissue may share similar patterns, and thus data fusion from multiple cell lines (multi-views) would improve the performance of drug synergy prediction. Given k sub-views (i.e., cell lines $C_1, \cdots, C_k$), we pretrain the single-view encoder for each cell line, and then derive their drug embeddings $\mathbf{H}^s_1, \cdots, \mathbf{H}^s_k$. We first combine these k sub-views via a weighted concatenation in Equation 5.

$$\mathbf{X}^s = \mathop{\Big\|}_{j=1}^{k} \alpha_j\mathbf{H}^s_j, \tag{5}$$

where α_j is the attention score that shows the similarity between the sub-view C_j and the main view C. In particular, the attention scores are initialized with Gaussian random weights, and then optimized during the model training.

As we may have several cell lines in a tissue, the dimension of $\mathbf{X}^s$ in Equation 5 is $k \times d$, and it would thus be much higher than that of $\mathbf{H}^m$ (i.e., the drug embedding from the main view). Hence, we implement another multi-layer

perceptron (MLP) to derive the drug embedding $\mathbf{Z}^s$ with reduced dimensionality in Equation 6. Here, $\mathbf{W}_1^s$, $\mathbf{W}_2^s$ and $\mathbf{W}_3^s$ are trainable weight matrices for the MLP in the multi-view encoder. Eventually, we concatenate the embedding $\mathbf{H}^m$ from the main view and the embedding $\mathbf{Z}^s$ from the sub-views as the final drug embedding $\mathbf{Z}$ through our multi-view encoder in Equation 7.

$$\begin{aligned}\mathbf{Z}^s &= MLP(\mathbf{X}^s) \\ &= ReLU(ReLU(\mathbf{X}^s\mathbf{W}_1^s)\mathbf{W}_2^s)\mathbf{W}_3^s, \quad (6)\\ \mathbf{Z} &= \mathbf{H}^m||\mathbf{Z}^s. \quad (7)\end{aligned}$$

3.5 Decoder and Overall Loss

After we obtain the final drug embedding matrix $\mathbf{Z}$, we can reconstruct a synergy score matrix S via a weighted inner-product decoder in Equation 8.

$$S = \mathbf{Z}\mathbf{W}_{de}\mathbf{Z}^T, \quad (8)$$

where $\mathbf{W}_{de}$ is a trainable matrix. Then, we leverage this score matrix S for drug synergy prediction. Meanwhile, we also calculate the mean square error between S and the ground-truth Y (i.e., the synergy score matrix for the main view C) as our model loss $\mathcal{L}$ in Equation 9.

$$\begin{aligned}\mathcal{L} &= MSE(Y, S) \\ &= \frac{1}{2N}\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}(Y_{ij} - S_{ij})^2, \quad (9)\end{aligned}$$

where N is total number of training samples and $M \in \mathbb{R}^{n\times n}$ is a mask matrix for training data, i.e., $M_{ij} = 1$ if the drug pair (d_i, d_j) is involved in the training set and 0 otherwise. Here, the MSE loss in Equation 9 is optimized to learn various model parameters in our GNNSynergy. Meanwhile, the single-view encoder for each cell line is pretrained by minimizing another MSE loss, where the predicted synergy score matrix S is calculated based on the single-view drug embedding $\mathbf{H}$ obtained by Equation 4 (instead of $\mathbf{Z}$). We aim to accelerate the training process and reduce the model training complexity through this pre-training step. Both the MSE loss of single-view encoder and the overall MSE loss are derivable and we adopt the Adam optimizer [43] to minimize them in this paper.

4 Results

In this section, we first introduce the experimental setup, and then demonstrate the empirical evaluation of the proposed GNNSynergy model.

4.1 Experimental Setup

4.1.1 Datasets

We used DrugComb database version v1.4 [41] for model training and evaluation. DrugComb is available at https://drugcomb.fimm.fi/. In particular, DrugComb integrates the data from four sources, namely, (i) the NCI ALMANAC dataset [44], (ii) the ONEIL dataset [8], (iii) the FORCINA dataset [45] and (iv) the CLOUD dataset [10]. Originally, DrugComb v1.4 consists of 466,033 drug combinations for 112 cancer cell lines in 11 different tissues. Given a drug pair within a specific cell line, it may have multiple entries in DrugComb with the same or different synergy scores. We first removed its duplicates with the same score, and then took the average score for the remaining entries as its final synergy score. Eventually, we worked on the processed DrugComb database with 268,547 drug pairs in 81 cell lines from 11 tissues as shown in Fig. 1(A).

4.1.2 Baselines

In our experiments, we compared our proposed GNNSynergy model with eight state-of-the-art methods. In particular, **Elastic Nets**, **Random Forest**, **Gradient Boosting Machines (GBM)** are common feature-based machine learning models, which were adopted as baselines for drug synergy prediction in [18]. **TreeCombo** [16] leveraged extreme gradient boosted trees (XGBoost) for drug synergy prediction. **comboLTR** [46] was recently proposed based on latent tensor reconstruction. **DeepSynergy** [18], **MatchMaker** [19] and **TranSynergy** [20] are deep learning based methods for drug synergy prediction. For a fair comparison, we used the same set of features (i.e., 424 features for each drug and 972 features for each cell line) for all the baseline methods.

4.1.3 Implementation details

For our GNNSynergy, we used 1-layer GCN with output feature dimensionality $d = 128$ and dropout rate $\rho = 0.5$. We used 4-layer MLPs (i.e., 1 input layer, 2 hidden layers and 1 output layer) in both single-view and multi-view encoders. The input layer of MLPs in single-view encoder has $2d = 256$ neurons, and it has $2d\times k$ neurons in multi-view encoder with k sub-view cell lines. For MLPs in single-view encoder, their 2 hidden layers and 1 output layer had 128, 256 and 128 neurons, respectively, and the dropout rates for 2 hidden layers were set as 0.2 and 0.6. For MLPs in multi-view encoder, their 2 hidden layers and 1 output layer had 640, 256 and 128 neurons, respectively, and the dropout rates for 2 hidden layers were set as 0.6 and 0. Note that we pre-trained the single-view encoder for each cell line with a learning rate of 1e-3, while the learning rate to optimize the overall GNNSynergy model was set as 1e-5. Lastly, we set the maximum number of training epochs to 2,000, and we considered the early-stop mechanism that the optimization will stop if the validation loss does not decline within the recent 300 epochs. In subsection 4.3, we will show the effects of some parameters above. We also empirically tuned the hyper-parameters for the baseline methods, and please find the details in our supplementary materials.

We conducted the 5-fold cross-validation (CV) to evaluate the performance of various methods. More importantly, the drug combination pairs were partitioned into five non-overlapping subgroups of equal size at random. We chose a subgroup as test set while the remaining four subgroups as training set. We further split the training data into two parts, i.e., training set and validation set, so that training set, validation set and test set have 60%, 20% and 20% of total samples, respectively. We calculated 5 metrics between predicted and ground-truth synergy scores, including Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pearson correlation coefficients and Spearman correlation coefficients. We implemented the deep learning methods using Pytorch

version 1.6, an open source python machine learning library, and traditional machine learning methods (Elastic Nets, Random Forest, TreeCombo and GBM) using scikit-learn version 0.23.2. We trained all the models on the NVIDIA Tesla K80 GPU. Note that the processed datasets, source codes and the supplementary materials are available at https://github.com/ZJMHub/GNNSynergy.

4.2 Comparison with Baselines

In this section, we show the performance comparison between our GNNSynergy and various baseline methods on DrugComb dataset in terms of 5 evaluation metrics. Note that we calculated the performance for various methods on each cell line first and took the average over all the 81 cell lines as the final performance. In particular, MSE, RMSE and MAE measure the errors between predicted and real synergy scores, showing better performance if they have lower values. On the opposite, the Pearson and Spearman correlation coefficients measure the correlation between predicted and real synergy scores, showing better performance with higher values.

TABLE 1: Performance comparison among various methods under 5-fold CV.

Method	MSE	RMSE	MAE	Pearson	Spearman
Elastic Nets	170.617	12.887	9.220	0.441	0.425
Random Forests	167.111	12.738	9.018	0.443	0.430
GBM	168.604	12.784	8.969	0.448	0.429
TreeCombo	166.987	12.696	8.880	0.463	0.455
DeepSynergy	137.283	11.555	7.973	0.575	0.533
TranSynergy	132.172	11.211	7.529	0.586	0.561
comboLTR	121.962	10.876	7.659	0.636	0.581
MatchMaker	104.148	10.129	6.655	0.686	0.649
GNNSynergy	**95.462**	**9.613**	**6.422**	**0.714**	**0.667**

As shown in Table 1, GNNSynergy significantly outperforms all the baselines in terms of all the 5 metrics. The MSE, RMSE and MAE of the GNNSynergy model are 95.462, 9.613 and 6.422, significantly lower than those of the baselines. Among the baseline methods, MatchMaker [19] achieves the best results. Compared to MatchMaker, GNNSynergy demonstrates an improvement of 4.08% and 2.77% in terms of Pearson and Spearman correlations, respectively. Besides, we can observe that deep learning methods (DeepSynergy, MatchMaker and TranSynergy) perform better than those traditional machine learning methods. Moreover, our GNNSynergy leverages the graph structural information of known drug combination pairs and the data from multiple cell lines, which can further improve the performance for drug synergy prediction.

4.3 External Validation Experiment

We further conducted an external validation experiment for all the methods. In particular, we used the data from one study (e.g., NCI ALMANAC [44]) to train the model and and then tested on the data from another independent study (e.g., ONEIL [8]). We selected the cell lines that appear in both the NCI ALMANAC dataset and ONEIL dataset so that both training and test sets are available for each selected cell line. Eventually, we selected 8 cell lines from 5 tissues for independent test, including cell lines SW-620, HT29,HCT116,NCIH23,SK-OV-3,OVCAR3,UACC62, and T-47D.

TABLE 2: Performance comparison of various methods on an independent validation set.

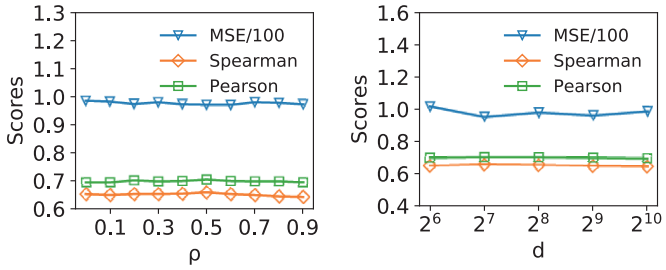
Method	MSE	RMSE	MAE	Pearson	Spearman
Elastic Nets	284.125	16.715	12.420	0.297	0.341
Random Forests	271.625	16.382	12.321	0.306	0.305
GBM	272.000	16.413	12.392	0.273	0.272
TreeCombo	273.375	16.446	12.310	0.292	0.287
DeepSynergy	263.750	16.086	11.228	0.368	0.370
TranSynergy	261.125	16.062	11.418	0.337	0.356
comboLTR	295.996	17.143	11.962	0.307	0.283
MatchMaker	240.000	15.555	10.791	0.442	0.438
GNNSynergy	**220.375**	**14.611**	**10.288**	**0.491**	**0.470**

As shown in Table 2, GNNSynergy still achieves the best performance in the independent validation experiment. The MSE, RMSE and MAE of the GNNSynergy are 220.375, 14.611 and 10.288, respectively, far below the error scores of MatchMaker (the second best performer). It can also be observed that traditional machine learning models do not perform as well as deep learning models, and some models perform poorly, such as comboLTR, which may be due to the overfitting issue.

4.4 Parameter Sensitivity Analysis

We present the sensitivity analysis for the parameters in our GNNSynergy, including the dropout probability ρ and the dimensionality of the latent space d in GCN.

In Fig. 3(a), the performance of GNNSynergy is relatively stable when ρ is set as different values in the stage of single-view training. Thus, we recommend to set ρ in the range $[0.4, 0.6]$. In Fig. 3(b), we could observe that the medium values for d e.g., $d = 2^7$ is more favorable. In our study, we used $d = 2^7 = 128$. Note that we have some other parameters, e.g., dropout probabilities in MLPs. Please refer to our supplementary materials (i.e., Fig. S3) and find their effects on the model performance.



(a) GCN dropout probability ρ (b) GCN dimensionality d

Fig. 3: Parameter sensitivity analysis for GNNSynergy.

4.5 Model Ablation Study

In this section, the ablation study shows the performance comparison between GNNSynergy and its variants.

First, we explore the impact of utilizing different values of pre-defined threshold t. As mentioned above, we constructed the DDS graphs according to the threshold t.

As shown in Fig. 4, 't = 0' refers to utilize synergy graph (Loewe scores > 0) and antagonism graph (Loewe scores < 0) as input. The variant 't = 5' refers to use synergy graph (Loewe scores > 5) and antagonism graph (Loewe scores < -5) and similarly for 't = 10'. As shown in Fig. 4, GNNSynergy achieves comparable performance when t is set as different values, i.e., the value of t does not affect the performance of GNNSynergy much. Meanwhile, we can observe that GNNSynergy achieves the slightly better performance when $t = 0$. The variant 't = 5' also slightly outperforms the variant 't = 10', indicating that our GNNSynergy utilizing more drug pairs for training tends to slightly improve its performance for drug synergy prediction.

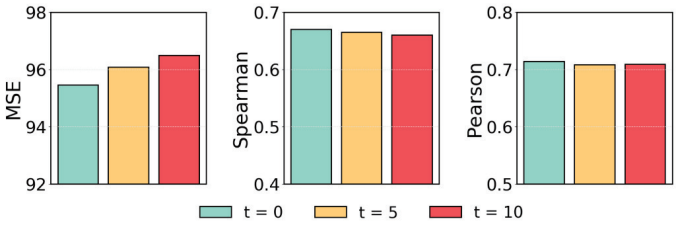


Fig. 4: Impact of different threshold t.

Second, we compare GNNSynergy with its variant using the single-view encoder only. As shown in Fig. 5, 'Multi-view' refers to GNNSynergy and 'Single-view' refers to this variant with the single-view encoder only. We can clearly observe that multi-view data integration improves the prediction performance, i.e., achieves lower error scores and higher correlation scores. The result in Fig. 5 demonstrates that data integration from multiple cell lines can help learn better drug embeddings for cell-line-specific drug synergy prediction.

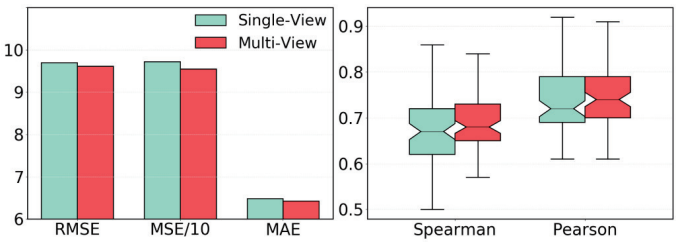


Fig. 5: Performance comparison between Single-view and Multi-view.

Third, we compare different variants/strategies to derive the weighted concatenation in Equation 5. The variant 'Equal Weights' assigns the same weights for all the sub-view cell lines (i.e., the weights are fixed). 'Attention Weights' refers to our GNNSynergy with the attention weights that are learned through model training. As shown in Fig. 6, the attention scores assigned for sub-view cell lines can help to improve the model performance. Note that the variant 'Equal Weights' still outperforms the best-performing baseline MatchMaker, demonstrating that data fusion from multiple cell lines indeed benefits.

Fourth, we compare different options to select sub-view cell lines. Note that we consider a specific cell line as main view and the other k cell lines within the same tissue as sub-views in our multi-view framework. We denote this strategy as 'Cell lines by tissue'. Here, we also explore two more strategies to select sub-view cell lines: (1) consider all the remaining 80 cell lines as sub-views (denoted as 'All cell lines'), and (2) randomly select k cell lines from 80 cell lines (denoted as 'Random cell lines'). Fig. 7 shows the performance comparison among these three strategies for sub-view selection. Obviously, 'Cell lines by tissue', i.e., selecting other cell lines within the same tissue as sub-views, achieves the best performance. This result demonstrates that integrating relevant cell lines (e.g., cell lines in the same tissue) would be helpful for learning drug embeddings and predicting the synergy scores.

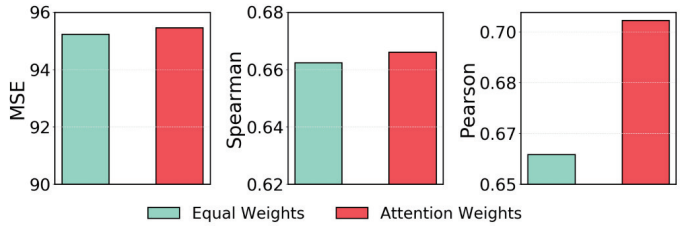


Fig. 6: Comparison between two weighting schemes for multi-view integration.

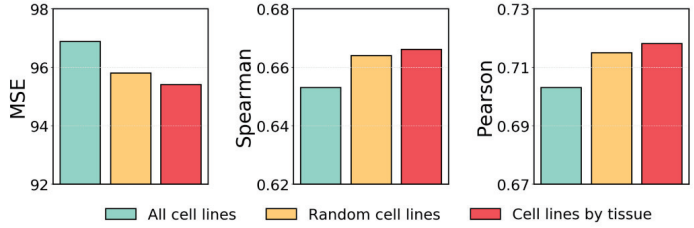


Fig. 7: Impact of different criteria for selecting cell lines.

4.6 Impact of Different Drug Features

The *ChemoPy* Python library [42] can derive 1135 descriptors for drugs, including 633 2D descriptors and 502 3D descriptors. After removing features with no values, we obtained the 541 2D and 472 3D descriptors for drugs as shown in Table 3. A detailed description for these descriptors can be found in our supplementary materials. In this subsection, we discuss the impact of 2D features and 3D features on the model performance, and the results are shown in Table 3.

TABLE 3: Performance comparison among different drug features.

Method	MSE	RMSE	MAE	Pearson	Spearman
w/ 2D features	95.462	9.613	6.422	0.714	0.667
w/ 3D features	96.431	9.640	6.418	0.709	0.662
combined features	94.234	9.604	6.415	0.711	0.672

From the results, we can observe that GNNSynergy achieves good performance using either 2D features or 3D features as input, which shows that both 2D and 3D features are important. After combining 2D and 3D features for drugs, we can observe that GNNSynergy can achieve the best performance.

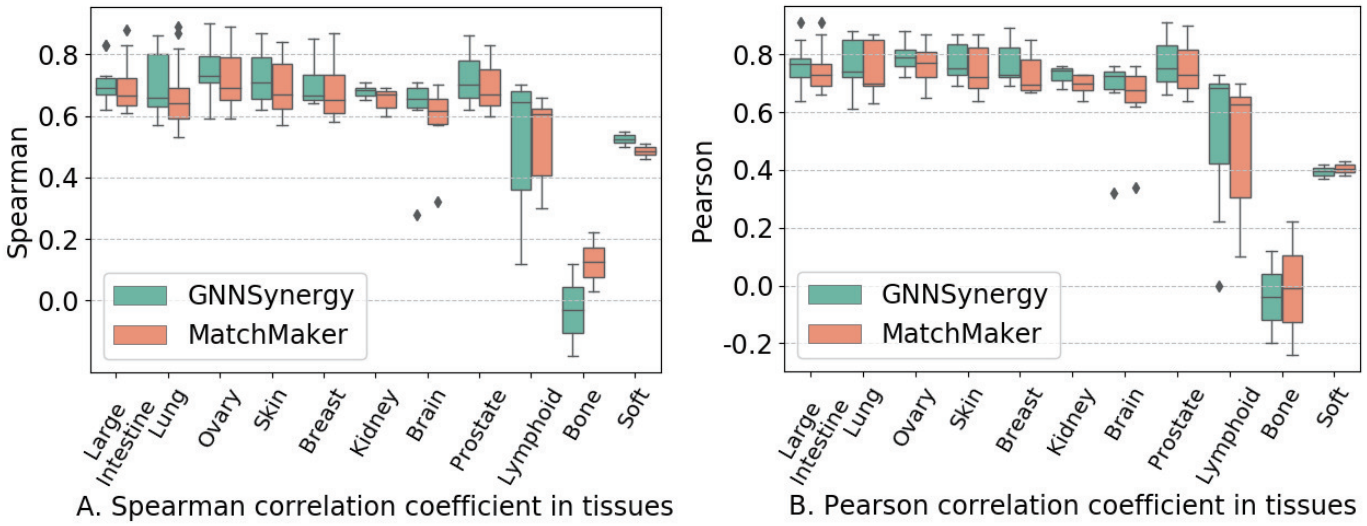


Fig. 8: Tissue-specific prediction performances of GNNSynergy and MatchMaker.

4.7 Investigation of Model Performance among Tissues

We further investigate the performance of our GNNSynergy model on individual tissues. In particular, we demonstrate the performance of our GNNSynergy and the best baseline MatchMaker on 11 different tissues, namely, large intestine, lung, ovary, skin, breast, kidney, brain, hamatopoietic and lymphoid, prostate, bone and soft tissue. The number of cell lines and drug combination pairs in each tissue can be found in Table S1 in our supplementary materials.

As shown in Fig. 8, our GNNSynergy outperforms MatchMaker in most tissues except bone. Moreover, GNNSynergy can achieve a Pearson correlation higher than 0.7 in 9 out of 11 tissues. For example, GNNSynergy achieves a Pearson correlation of 0.766 in large intestine cancer cells and 0.759 for breast cancer cells. Note that both GNNSynergy and MatchMaker achieve poor performances in bone and soft tissues. Bone has two cell lines (A-673 and TC-71) and soft tissue has only one cell line (RD), and thus our multi-view framework would not help much with only one sub-view cell line in bone and no sub-view in soft tissue. More importantly, we have very limited number of drug combination pairs in some cell lines, e.g., TC-71 in bone has 42 samples and RD in soft tissue has 80 samples. With such small number of samples, usually it is very challenging for us to train good models.

5 Conclusion

Drug combination therapies play an important role in cancer treatment. In this paper, we presented a multi-view graph neural network model named GNNSynergy for drug combination prediction. We first built multiple drug-drug synergy (DDS) graphs to describe different types of drug interactions and then learned drug embeddings from these DDS graphs. We further designed a multi-view framework to integrate the drug embeddings learned from similar cell lines for drug synergy prediction. Experimental results demonstrated that our proposed GNNSynergy outperformed state-of-the-art methods. However, there are still some limitations in the current GNNSynergy model. First, we employed the drug chemical structures as features only. We plan to collect additional drug data (e.g., drug-target interactions and adverse drug-drug interactions) to learn more robust representations for drugs. Second, GNNSynergy, as well as other baselines, achieved poor performances for the understudied cell lines [21] with limited number of drug combination pairs. We will consider transfer learning and domain adaptation techniques [47] to address this issue in the future. Third, GNNSynergy cannot make predictions for new drugs/nodes, which are not tested or recorded in current database (e.g., DrugComb). We will apply inductive graph methods (e.g., GraphSAGE [48]) to improve our model.

Acknowledgments

This work has been supported by the Natural Science Foundation of China (61876043), Natural Science Foundation of Guangdong (2014A030306004, 2014A030308008), Guangdong High-level Personnel of Special Support Program (2015TQ01X140), Science and Technology Planning Project of Guangzhou (201902010058).

References

[1] J. Jia, F. Zhu, X. Ma, Z. W. Cao, Y. X. Li, and Y. Z. Chen, "Mechanisms of drug combinations: interaction and network perspectives," *Nature reviews Drug discovery*, vol. 8, no. 2, pp. 111–128, 2009.

[2] J. Lehár, A. S. Krueger, W. Avery, A. M. Heilbut, L. M. Johansen, E. R. Price, R. J. Rickles, G. F. Short Iii, J. E. Staunton, X. Jin *et al.*, "Synergistic drug combinations tend to improve therapeutically relevant selectivity," *Nature biotechnology*, vol. 27, no. 7, pp. 659–666, 2009.

[3] E. De Clercq, "The design of drugs for hiv and hcv," *Nature reviews Drug discovery*, vol. 6, no. 12, pp. 1001–1018, 2007.

[4] A. H. Groll and A. Tragiannidis, "Recent advances in antifungal prevention and treatment," in *Seminars in hematology*, vol. 46, no. 3. Elsevier, 2009, pp. 212–229.

[5] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "Nllss: predicting synergistic drug combinations based on semi-supervised learning," *PLoS computational biology*, vol. 12, no. 7, p. e1004975, 2016.

[6] X. Tan, L. Hu, L. J. Luquette, G. Gao, Y. Liu, H. Qu, R. Xi, Z. J. Lu, P. J. Park, and S. J. Elledge, "Systematic identification of synergistic drug pairs targeting hiv," *Nature biotechnology*, vol. 30, no. 11, pp. 1125–1130, 2012.

[7] R. J. Worthington and C. Melander, "Combination approaches to combat multidrug-resistant bacteria," *Trends in biotechnology*, vol. 31, no. 3, pp. 177–184, 2013.

[8] J. O'Neil, Y. Benita, I. Feldman, M. Chenard, B. Roberts, Y. Liu, J. Li, A. Kral, S. Lejnine, A. Loboda *et al.*, "An unbiased oncology compound screen to identify novel combination strategies," *Molecular cancer therapeutics*, vol. 15, no. 6, pp. 1155–1162, 2016.

[9] L. He, E. Kulesskiy, J. Saarela, L. Turunen, K. Wennerberg, T. Aittokallio, and J. Tang, "Methods for high-throughput drug combination screening and synergy scoring," *Methods in molecular biology (Clifton, NJ)*, vol. 1711, p. 351, 2018.

[10] M. P. Licciardello, A. Ringler, P. Markt, F. Klepsch, C.-H. Lardeau, S. Sdelci, E. Schirghuber, A. C. Müller, M. Caldera, A. Wagner *et al.*, "A combinatorial screen of the cloud uncovers a synergy targeting the androgen receptor," *Nature chemical biology*, vol. 13, no. 7, pp. 771–778, 2017.

[11] I. F. Tsigelny, "Artificial intelligence in drug combination therapy," *Briefings in bioinformatics*, vol. 20, no. 4, pp. 1434–1448, 2019.

[12] L. Wu, Y. Wen, D. Leng, Q. Zhang, C. Dai, Z. Wang, Z. Liu, B. Yan, Y. Zhang, J. Wang *et al.*, "Machine learning methods, databases and tools for drug combination prediction," *Briefings in Bioinformatics*, 2021.

[13] K. Fan, L. Cheng, and L. Li, "Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects," *Briefings in Bioinformatics*, 2021.

[14] H. Li, T. Li, D. Quang, and Y. Guan, "Network propagation predicts drug synergy in cancers," *Cancer research*, vol. 78, no. 18, pp. 5446–5457, 2018.

[15] H. Li, S. Hu, N. Neamati, and Y. Guan, "Taiji: approaching experimental replicates-level accuracy for drug synergy prediction," *Bioinformatics*, vol. 35, no. 13, pp. 2338–2339, 2019.

[16] J. D. Janizek, S. Celik, and S.-I. Lee, "Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine," *bioRxiv*, p. 331769, 2018.

[17] P. Ding, R. Yin, J. Luo, and C.-K. Kwoh, "Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological, and network knowledge," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1336–1345, 2018.

[18] K. Preuer, R. P. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, and G. Klambauer, "Deepsynergy: predicting anti-cancer drug synergy with deep learning," *Bioinformatics*, vol. 34, no. 9, pp. 1538–1546, 2018.

[19] H. Brahim Kuru, O. Tastan, and E. Cicek, "Matchmaker: A deep learning framework for drug synergy prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021.

[20] Q. Liu and L. Xie, "Transynergy: Mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations," *PLoS computational biology*, vol. 17, no. 2, p. e1008653, 2021.

[21] Y. Kim, S. Zheng, J. Tang, W. Jim Zheng, Z. Li, and X. Jiang, "Anticancer drug synergy prediction in understudied tissues using transfer learning," *Journal of the American Medical Informatics Association*, vol. 28, no. 1, pp. 42–51, 2021.

[22] J. Hu, J. Gao, X. Fang, Z. Liu, F. Wang, W. Huang, H. Wu, and G. Zhao, "DTSyn: a dual-transformer-based neural network to predict synergistic drug combinations," *Briefings in Bioinformatics*, vol. 23, no. 5, 08 2022, bbac302. [Online]. Available: https://doi.org/10.1093/bib/bbac302

[23] F. Cheng, I. A. Kovács, and A.-L. Barabási, "Network-based prediction of drug combinations," *Nature communications*, vol. 10, no. 1, pp. 1–11, 2019.

[24] L. Yu, M. Xia, and Q. An, "A network embedding framework based on integrating multiplex network for drug combination prediction," *Briefings in Bioinformatics*, 2021.

[25] P. Ding, C. Liang, W. Ouyang, G. Li, Q. Xiao, and J. Luo, "Inferring synergistic drug combinations based on symmetric meta-path in a novel heterogeneous network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 04, pp. 1562–1571, jul 2021.

[26] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[27] S. K. Ata, M. Wu, Y. Fang, L. Ou-Yang, C. K. Kwoh, and X.-L. Li, "Recent advances in network-based methods for disease gene prediction," *Briefings in bioinformatics*, vol. 22, no. 4, p. bbaa303, 2021.

[28] M. Sun, S. Zhao, C. Gilvary, O. Elemento, J. Zhou, and F. Wang, "Graph convolutional networks for computational drug development and discovery," *Briefings in bioinformatics*, vol. 21, no. 3, pp. 919–935, 2020.

[29] R. Cai, X. Chen, Y. Fang, M. Wu, and Y. Hao, "Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers," *Bioinformatics*, vol. 36, no. 16, pp. 4458–4465, 2020.

[30] P. Jiang, S. Huang, Z. Fu, Z. Sun, T. M. Lakowski, and P. Hu, "Deep graph embedding for prioritizing synergistic anticancer drug combinations," *Computational and structural biotechnology journal*, vol. 18, pp. 427–438, 2020.

[31] J. Wang, X. Liu, S. Shen, L. Deng, and H. Liu, "DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations," *Briefings in Bioinformatics*, 09 2021, bbab390. [Online]. Available: https://doi.org/10.1093/bib/bbab390

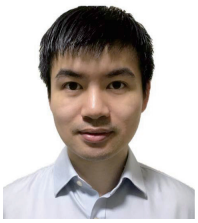
[32] P. Zhang, S. Tu, W. Zhang, and L. Xu, "Predicting cell line-specific synergistic drug combinations through a relational graph convolutional network with attention mechanism," *Briefings in Bioinformatics*, 2022.

[33] P. Li, C. Huang, Y. Fu, J. Wang, Z. Wu, J. Ru, C. Zheng,

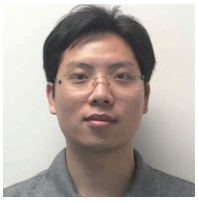
Z. Guo, X. Chen, W. Zhou *et al.*, "Large-scale exploration and analysis of drug combinations," *Bioinformatics*, vol. 31, no. 12, pp. 2007–2016, 2015.
[34] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *ICLR*, 2017.
[35] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.
[36] Y. Long, M. Wu, C. K. Kwoh, J. Luo, and X. Li, "Predicting human microbe–drug associations via graph convolutional network with conditional random field," *Bioinformatics*, vol. 36, no. 19, pp. 4918–4927, 2020.
[37] X. Zhang, L. He, K. Chen, Y. Luo, J. Zhou, and F. Wang, "Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson's disease," in *AMIA Annual Symposium Proceedings*, vol. 2018, 2018, p. 1147.
[38] H. Fu, F. Huang, X. Liu, Y. Qiu, and W. Zhang, "Mvgcn: data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks," *Bioinformatics*, 2021.
[39] T. Ma, C. Xiao, J. Zhou, and F. Wang, "Drug similarity integration through attentive multi-view graph auto-encoders," in *IJCAI*, 2018, pp. 3477–3483. [Online]. Available: https://doi.org/10.24963/ijcai.2018/483
[40] Z. Hao, D. Wu, Y. Fang, M. Wu, R. Cai, and X. Li, "Prediction of synthetic lethal interactions in human cancers using multi-view graph auto-encoder," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 4041–4051, 2021.
[41] B. Zagidullin, J. Aldahdooh, S. Zheng, W. Wang, Y. Wang, J. Saad, A. Malyutina, M. Jafari, Z. Tanoli, A. Pessia *et al.*, "Drugcomb: an integrative cancer drug combination data portal," *Nucleic acids research*, vol. 47, no. W1, pp. W43–W51, 2019.
[42] D.-S. Cao, Q.-S. Xu, Q.-N. Hu, and Y.-Z. Liang, "Chemopy: freely available python package for computational biology and chemoinformatics," *Bioinformatics*, vol. 29, no. 8, pp. 1092–1094, 2013.
[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[44] S. L. Holbeck, R. Camalier, J. A. Crowell, J. P. Govindharajulu, M. Hollingshead, L. W. Anderson, E. Polley, L. Rubinstein, A. Srivastava, D. Wilsker *et al.*, "The national cancer institute almanac: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity," *Cancer research*, vol. 77, no. 13, pp. 3564–3576, 2017.
[45] G. C. Forcina, M. Conlon, A. Wells, J. Y. Cao, and S. J. Dixon, "Systematic quantification of population cell death kinetics in mammalian cells," *Cell systems*, vol. 4, no. 6, pp. 600–610, 2017.
[46] T. Wang, S. Szedmak, H. Wang, T. Aittokallio, T. Pahikkala, A. Cichonska, and J. Rousu, "Modeling drug combination effects via latent tensor reconstruction," *bioRxiv*, 2021.
[47] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
[48] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

Zhifeng Hao is currently a Professor in the School of Computer, Guangdong University of Technology, and College of Science, Shantou University. He received his B.S. degree in Mathematics from the Sun Yat-Sen University in 1990, and his Ph.D. degree in Mathematics from Nanjing University in 1995. His research interests involve various aspects of Algebra, Machine Learning, Data Mining, Evolutionary Algorithms.

Jianming Zhan received a Master's degree in Computer Technology Engineering from Guangdong University of Technology in 2022, under the guidance of Professor Hao Zhifeng. He received his bachelor's degree in software engineering from Guangdong University of Technology in 2019. His research interests include deep learning, graph neural networks and bioinformatics.



Yuan Fang is currently an Assistant Professor at the School of Computing and Information Systems, Singapore Management University, Singapore. Previously, he was a scientist at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore and a data scientist at DBS Bank, Singapore. He received his Ph.D. degree in Computer Science from the University of Illinois at Urbana-Champaign, United States in 2014, and Bachelor's degree in Computer Science from National University of Singapore, Singapore in 2009. His work has been featured in the Best Papers collection of VLDB 2013. His current research focuses on graph-based machine learning, Web and social media mining, recommendation systems and bioinformatics.



Min Wu is currently a Senior Research Scientist in the Data Analytics Department at the Institute for Infocomm Research (I2R) under the Agency for Science, Technology and Research (A*STAR), Singapore. He received the B.Eng. from the University of Science and Technology of China (USTC), China in 2006 and his Ph.D. degree from Nanyang Technological University, Singapore in 2011. He received the best paper awards in the 15th International Conference on Bioinformatics (InCoB 2016) and the 20th International Conference on Database Systems for Advanced Applications (DASFAA 2015). He also won the IJCAI contest 2015 on repeated buyers prediction after sales promotion. His current research interests include machine learning, data mining and bioinformatics.

Ruichu Cai is currently with a Professor in the School of Computer and the director of the Data Mining and Information Retrieval (DMIR) group, Guangdong University of Technology. He received his B.S. degree in Applied Mathematics and Ph.D. degree in Computer Science from South China University of Technology in 2005 and 2010, respectively. He was a recipient of the National Science Fund for Excellent Young Scholars, Natural Science Award of Guangdong and so on awards. He has served as a senior PC/PC/Reviewer for AAAI, IJCAI, TPMAI, TNNLS, and other conferences and journals. His research interests cover a variety of different topics including causality, deep learning and their applications.