GNNSynergy: A multi-view graph neural network for predicting anti-cancer drug synergy

This report gives supplementary information to the manuscript "GNNSynergy: A multi-view graph neural network for predicting anti-cancer drug synergy". It provides more detailed information in the following three sections: data set, methods and results. The first section describes the cell line and true scores contained in the DrugComb database. The second section informs about the hyperparameter space for the different methods and parameter sensitivity analusis of GNNSynergy. The last section provides further experiment results for GNNSynergy.

1. DATASET

Table S1 shows the cancer cell lines included in the DrugComb database. The 81 cell lines originated from 11 different tissue types.

Table S1. The 81 cell lines tested in the DrugComb Database, covering 11 different tissue types. Amount indicated the amount of cell lines in the tissue.

Tissue (# cell lines)	Number of drug pairs	Cell Line
Large Intestine (10)	35297	SW-620; HT29; HCT116; HCT-15; KM12; COLO 205; HCC-2998; LOVO; SW837; RKO
Breast (8)	29421	T-47D; MDA-MB-231; MCF7; BT-549; MDA-MB-468; HS 578T; OCUBM; MDAMB436;
Lymphoid (8)	25234	K-562; CCRF-CEM; SR; MOLT-4; RPMI-8226; L-1236; HDLM-2; L-428;
Kidney (8)	37363	ACHN; SN12C; TK-10; A498; 786-0; CAKI-1; UO-31; RXF 393;
Lung (13)	39934	NCIH23; NCI-H226; A549; EKVX; NCI-H322M; NCI-H522; HOP-92; HOP-62; A427; SKMES1; NCIH2122; NCIH520; NCIH1650
Ovary (11)	31186	SK-OV-3; OVCAR3; OVCAR-5; IGROV1; OVCAR-8; OVCAR-4; A2780; PA1; ES2; UWB1289; OV90;
Skin (11)	33389	UACC62; UACC-257; SK-MEL-28; SK-MEL-5; M14; LOX IMVI; SK-MEL-2; A375; RPMI7951; A2058; HT144
Brain (6)	24883	SF-268; U251; SF-539; SF-295; SNB-75; T98G
Prostate (3)	9872	PC-3; DU-145; VCAP
Bone (2)	1878	A-673; TC-71
Soft (1)	80	RD

Fig. S1 shows the distribution of true synergy scores and predicted synergy scores in the processed DrugComb dataset. We can observe that the true score distribution presents a normal distribution, and there are few synergistic drug pairs and antagonistic drug pairs. In addition, the score distribution predicted by GNNSynergy is basically consistent with the real score distribution, but there are some errors at the same time.

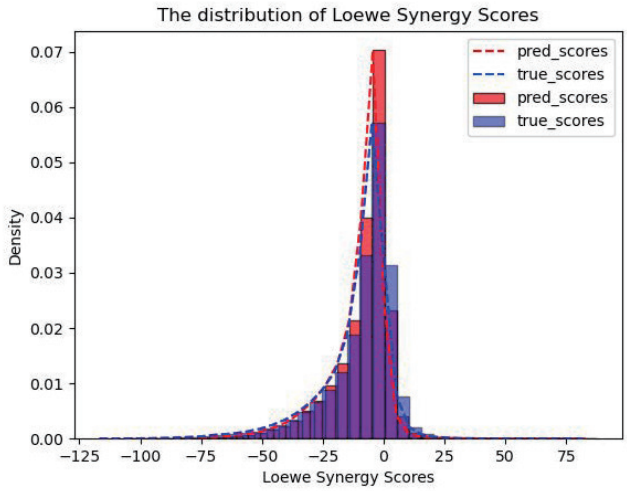


Fig. S1. The distribution of true synergy scores and predicted synergy scores in the processed DrugComb dataset.

Besides, we performed the t-SNE analysis to visualize the high dimensional vector representation of cell lines mapped to a 2D space to reflect relationships between cell lines (Fig. S2). It showed that the distance between cell lines in the same tissue would be more closer, that is, the information of cell lines in the same tissue is more useful.

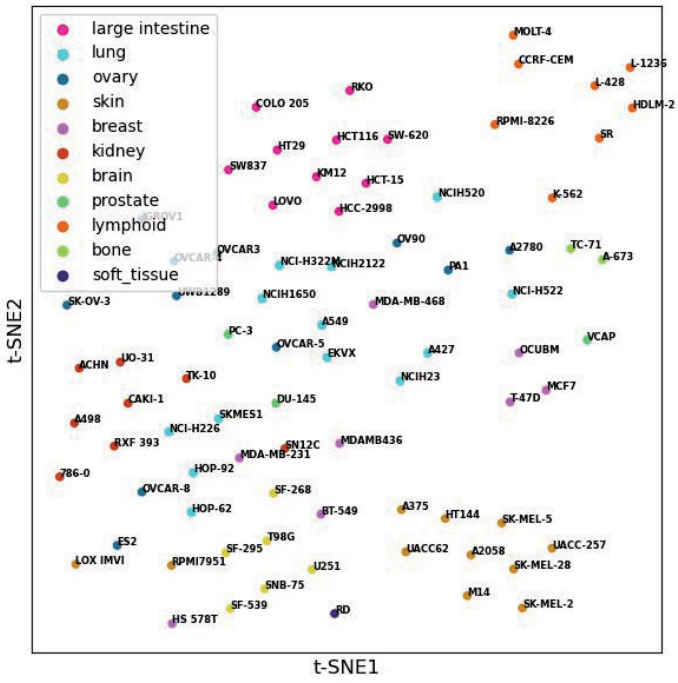


Fig. S2. Visualization of different cell lines with t-SNE analysis. Different colors indicate different tissues of each cell line. The distance between cell lines in the same tissue is closer.

2. HYPERPARAMETERS SETTING AND ANALYSIS

A. Hyperparameters Setting

The hyperparameters of all methods were optimized by valadation set, using grid search. Table S2 display the search range of hyperparameter space considered by Elastic Net (EN), Random Forest (RF), Gradient Boosting Machine (GBM), TreeCombo (XGBoost) and comboLTR, respectively. For the deep learning methods, such as DeepSynergy, TranSynergy and MatchMaker, we used the default parameters used in their original paper.

Table S2. Hyperparameter space considered for Baselines

Model	Hyperparament	Values consided
EN	Constant α	0.001; 0.01; 0.1; 1; 10; 100
	L1 ratio	0.2; 0.4; 0.6; 0.8
RF	numbers of estimators (decision trees)	10; 100; 500; 1000
	maximum tree depth	4; 6; 8; 10; 12
GBM	numbers of estimators (decision trees)	10; 100; 500; 1000
	maximum tree depth	4; 6; 8; 10; 12
	learning rate	0.05; 0.10; 0.15
XGBoost	maximum tree depth	4; 6; 8; 10; 12
	learning rate	0.05; 0.10; 0.15
comboLTR	rank_uv	20; 32; 64; 128; 256
	repeats	20; 70; 120; 170; 220; 270
	ranks	20; 40; 60; 80; 100

For GNNSynergy, we used only one layer and applied the dropout technique with dropout probability $p = 0.7$ in GCN layer. Besides, the dimension of trainable weight matrix $\mathbf{W}$ was set to 256, initialized by Kaiming initialization. In addition, the hidden layers of the feed forward neural network for aggregate three sub graphs embeddings had [256, 512, 256] neurons, respectively. Moreover, we adapted the Dropout layer with dropout probabilities β_1 and β_2 between each hidden layers, where $\beta_1 = 0.5$ and $\beta_2 = 0.2$. Similarly, the dimensions of the hidden layers in the multi-layer perceptron merging all sub views were set to [1280, 512, 256] and the dropout rate between layers were set 0.5 and 0, separately. The learning rate η in the optimization algorithm was set to 1e-3 in singel-view stage and was set to 1e-5 in multi-view stage, without learning rate decay. We adopted Adam optimizer with default parameters and set the maximum training epoch to 2000. We further considered the early-stopping mechanism that is the optimization will stop if the validation loss does not decline for 300 epochs and saved the model parameters when the validation loss is minimal. Follow the data normalization mentioned by Preuer *et al.*, we employed standardizing and hyperbolic tangent agian as input feature normalization, i.e. norm + tanh. All the best hyperparameter settings are filtered out by grid search and summarized in Table S3.

B. Parameter Sensitivity Analysis

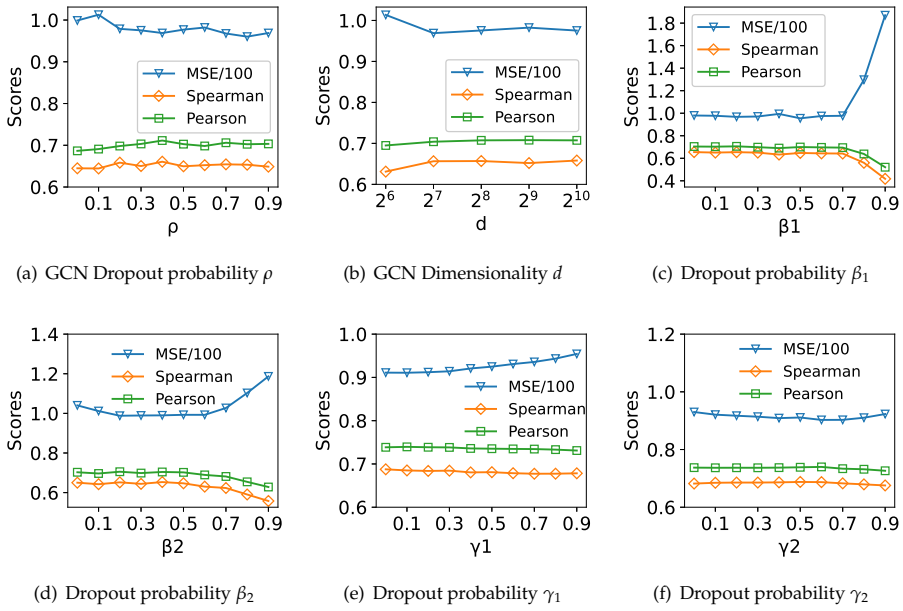
We present the sensitivity analysis for the parameters in our GNNSynergy, including the GCN dimensionality d, GCN dropout probability ρ, the dropout probability β_1, β_2 of MLP in Single-View, and the dropout probability γ_1, γ_2 of MLP in Multi-View, as shown in Fig. S3.

From the Fig. S3 (a), we can observe that the model performance will perform worse when GCN dropout probability $\rho > 0.7$. Fig. S3 (b) indicated that the performance become relatively stable in the medium values. Fig. S3 (c) and Fig. S3 (d) show the performance of different values, respectively. We change one of them and fix the other to its default value. As for β_1, we can observe that the MSE of GNNSynergy gradually decrease with larger dropout probability until before 0.2. Similarly, the performance get worse when $\beta_2 > 0.5$. Eventually, we set β_1 as 0.2 and

Table S3. Hyperparameter space considered for GNNSynergy

Module	Hyperparameter	Value
Global	learning rate in single-view (η_1)	1e-3
	learning rate in multi-view (η_2)	1e-5
	max epoch	2000
	early-stop epoch	300
GCN	dimensionality (d)	256
	dropout probability (ρ)	0.7
MLP in Single-View	dimensionality	[256, 512, 256]
	dropout probability (β)	[0.2,0.5]
MLP in Multi-View	dimensionality	[1280, 512, 256]
	dropout probability (γ)	[0.0, 0.5]

β_2 as 0.5 in our experiments. In Fig. S3 (e) and Fig. S3 (f), we can observe that the performance of GNNSynergy achieve the best when $\gamma_1 = 0, \gamma_2 = 0.5$.



(a) GCN Dropout probability ρ (b) GCN Dimensionality d (c) Dropout probability β_1

(d) Dropout probability β_2 (e) Dropout probability γ_1 (f) Dropout probability γ_2

Fig. S3. Parameter sensitivity analysis for GNNSynergy.

3. RESULTS

A. Investigation of prediction performance among all specific cell lines

We further investigated the performance of GNNSynergy among different cell lines. The spearman correlation coefficient and pearson correlation coefficient comparison among the GNNSynergy, MatchMaker and DeepSynergy are shown in Fig. S4. Obviously, GNNSynergy perform better than the MatchMaker in most of cell lines. Besides, Fig. S5 shows the cell line-specific MSE predicted by GNNSynergy.

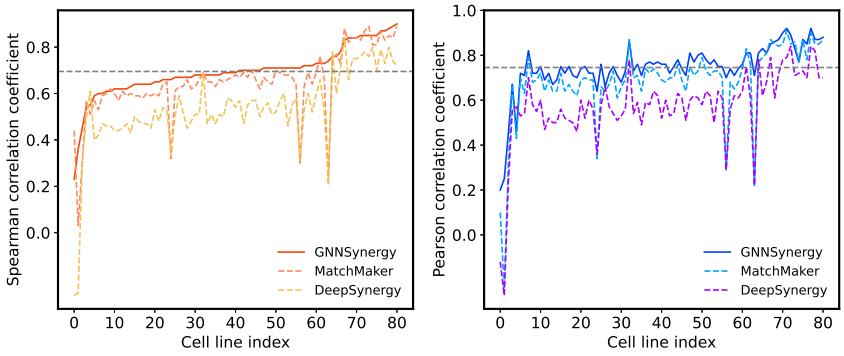


Fig. S4. The cell line-specific predicted Spearman correlation coefficient and Pearson correlation coefficient comparison among GNNSynergy, DeepSynergy and MatchMaker. Here, we choose to compare with MatchMaker because it is the best performance in the baselines. The gray horizontal dotted line represents the average score of GNNSynergy.

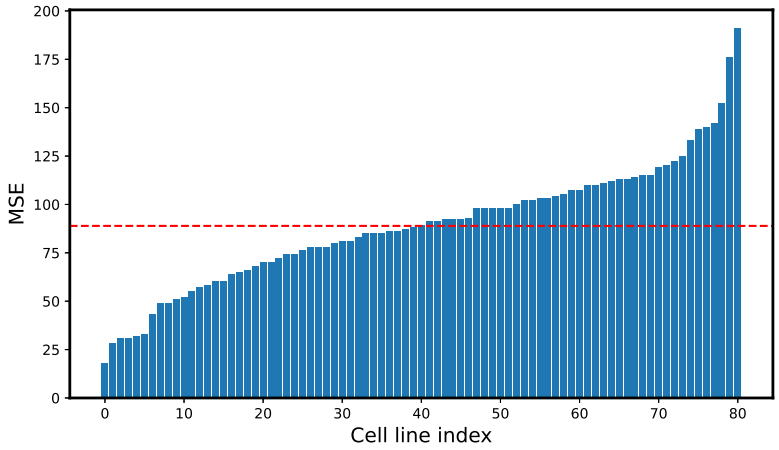


Fig. S5. The MSE scores predicted by GNNSynergy among all cell lines. The red horizontal dotted line represents the average score of GNNSynergy.

B. Model Ablation Study

We conducted four strategies to derive the weighted concatenation from different cell-lines. In order to further investigate, we list the weights for different cell-lines in SW-620 assigned by these strategies, as shown in Table S4.

Table S4. The weights for different cell-lines in SW-620 assigned by three strategies

Strategy	Equal Weights	Fixed Weights	Attention-FW	Attention-GR
sub-view 1	0.1111	0.1092	0.1192	0.1236
sub-view 2	0.1111	0.1138	0.1086	0.1558
sub-view 3	0.1111	0.1100	0.1191	0.0953
sub-view 4	0.1111	0.1109	0.1047	0.1083
sub-view 5	0.1111	0.1105	0.1045	0.0417
sub-view 6	0.1111	0.1086	0.1198	0.1699
sub-view 7	0.1111	0.1125	0.1198	0.0519
sub-view 8	0.1111	0.1095	0.0999	0.0643
sub-view 9	0.1111	0.1149	0.1043	0.1893

We also conducted a performance analysis about the question of divide the sub graphs from the DDS graph. We compare different variants to divide the sub graphs. The variant 'One Graph' refer to using the DDS graph including all drug pairs. The variant 'Two Graphs' divide the DDS graph into synergy graph and antagonism graph when threshold t was set to 0. At last, the variant 'Three Graphs' is our GNNSynergy. The comparison results were displayed in Fig. S6

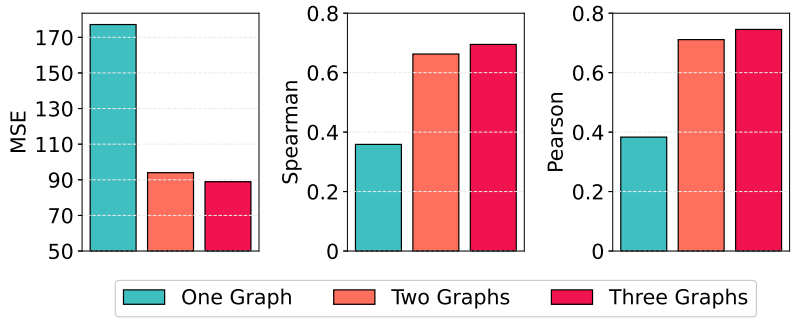


Fig. S6. The performance comparison of three variants to divide the DDS graph.

C. Case studies for novel drug combination prediction

After the training process of case studies, we total acquire 19 predicted drug combination synergistic pairs eventually. The all predicted pairs were list in Table S5.

Table S5. Top predicted drug combination pairs with literature support.

#	Rank	Drug 1	Drug 2	Cell line
1	302	AURANOFIN	TRAMETINIB	NCI-H226
2	497	AURANOFIN	PD325901	NCI-H226
3	834	AURANOFIN	CISATRACURIUM BESYLATE	NCIH23
4	992	CISATRACURIUM BESYLATE	AURANOFIN	UWB1289
5	1207	AURANOFIN	CARFILZOMIB (PR-171)	SF-268
6	2192	AURANOFIN	TRAMETINIB	HOP-92
7	2356	CISATRACURIUM BESYLATE	AURANOFIN	LOVO
8	3219	CARFILZOMIB (PR-171)	AURANOFIN	LOVO
9	3454	ZALCITABINE	AURANOFIN	SR
10	3700	NILOTINIB	SORAFENIB	T98G
11	3704	NILOTINIB	VEMURAFENIB	T98G
12	3714	IMATINIB	NILOTINIB	T98G
13	3716	NILOTINIB	GEFITINIB	T98G
14	3722	SORAFENIB	VEMURAFENIB	T98G
15	3729	IMATINIB	SORAFENIB	T98G
16	3742	ERLOTINIB HYDROCHLORIDE	NILOTINIB	T98G
17	3907	ISONIAZID	BORTEZOMIB	HDLM-2
18	3909	LEFLUNOMIDE	BORTEZOMIB	HDLM-2
19	3925	PROCARBAZINE	BORTEZOMIB	HDLM-2