Supplementary Materials

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

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

July 29, 2023

Abstract

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 about dataset, parameter setting for baseline methods and GNNSynergy, as well as some additional experimental results.

1 Experimental Data

Table S1 shows the cancer cell lines included in the DrugComb database. There are 81 cell lines originated from 11 different tissues.
We performed the t-SNE analysis to visualize the cell lines in 2D space to reflect relationships among them in Fig. S1. Originally, each cell line has 972 features for its gene expression profile, and here we applied t-SNE to map the 972-dimensional vectors into the 2D space. It shows that cell lines within the same tissues are closer to each other. Given a specific cell line, the information of other cell lines in the same tissue would be useful for drug synergy prediction.

2 Parameter Setting and Sensitivity Analysis

2.1 Hyper-parameter Setting

The hyper-parameters of all the methods were optimized by the validation set, using grid search. Table S2 shows the search range for different hyper-parameters in the baseline methods, including Elastic Net (EN), Random Forest (RF), Gradient Boosting Machine (GBM), TreeCombo (XGBoost) and comboLTR. For the deep learning methods, such as DeepSynergy, TranSynergy and MatchMaker, we used the default parameters used in their original papers.

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

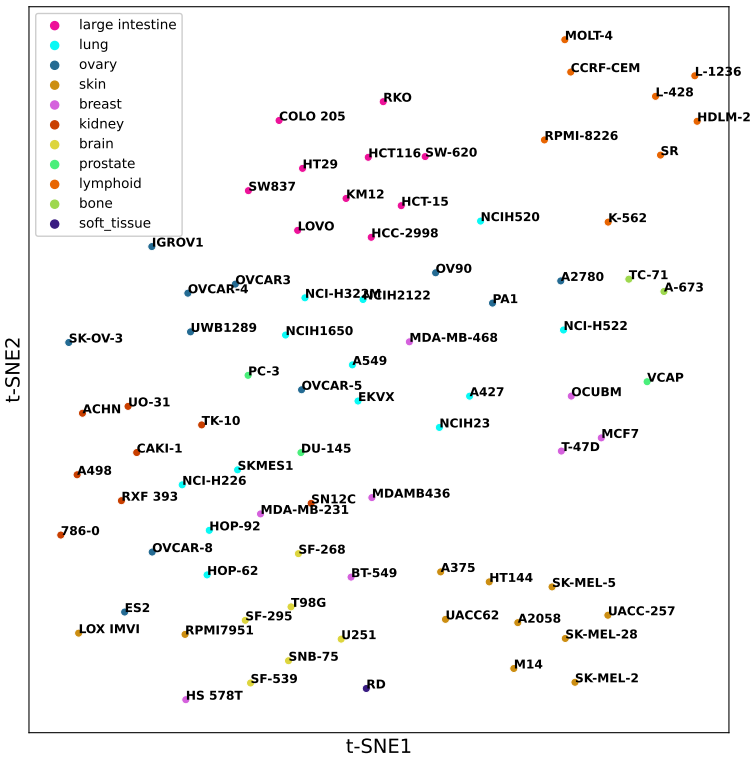


Figure S1: Visualization of different cell lines with t-SNE analysis. Different colors indicate different tissues. It shows that cell lines within the same tissues are closer to each other.

Table S2: **Hyper-parameters 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 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 $3d = 384$ neurons, and it has $3d \times k$ 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 0.0 and 0.6. 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-3. Lastly, we set the maximum number of training epochs to 2000, and we considered the early-stop mechanism that the optimization will stop if the validation loss does not decline within the recent 300 epochs. Table S3 summarizes the parameter settings in our GNNSynergy.

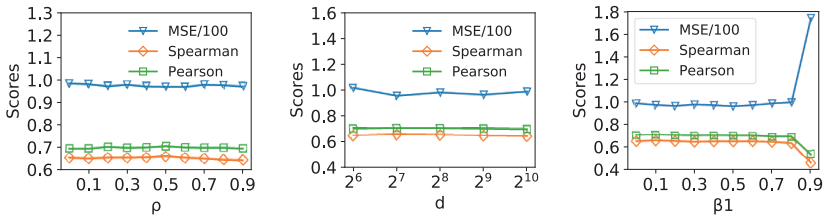
Table S3: **Hyper-parameter settings for GNNSynergy**

Module	Hyperparameter	Value
Global	learning rate in single-view (η_1)	1e-3
	learning rate in multi-view (η_2)	1e-3
	max epoch	2000
	early-stop epoch	300
GCN	dimensionality (d)	128
	dropout probability (ρ)	0.5
MLP in Single-View	dimensionality	[128, 256, 128]
	dropout probability (β)	[0.2, 0.6]
MLP in Multi-View	dimensionality	[640, 256, 128]
	dropout probability (γ)	[0.0, 0.6]

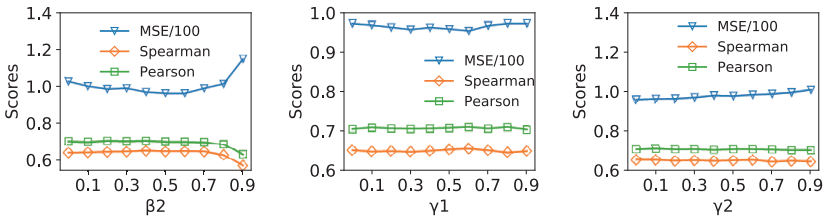
2.2 Parameter Sensitivity Analysis

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. S2.

From the Fig. S2(a), the performance of GNNSynergy is relatively stable when ρ is set as different values. Thus, we recommend to set ρ in the range $[0.4, 0.6]$. In Fig. S2(b), we could observe that the medium values for d e.g., $d = 2^7$ is more favorable. In our study, we used $d = 2^7 = 128$. Fig. S2(c) and Fig. S2(d) show the performance of GNNSynergy using different values for β_1 and β_2 (the dropout probabilities in MLPs in single-view encoder), 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 starts to increase when β_1 is larger than 0.7. Similarly,



(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

Figure S2: Parameter sensitivity analysis for GNNSynergy.

the performance get worse when $\beta_2 > 0.6$. In this paper, we set β_1 as 0.2 and β_2 as 0.6 in our experiments. In Fig. S2(e) and Fig. S2(f), we can observe that GNNSynergy achieves the best performances when $\gamma_1 = 0.6, \gamma_2 = 0.0$.

2.3 Impact of Threshold t

As mentioned in the main manuscript, we construct 2 drug-drug synergy graphs for each cancer cell line based on a pre-defined threshold $t = 0$. Although this setting is following other previous studies, we still would like to show the impact of this parameter on the model performance. We evaluate the performance of GNNSynergy when t is set as different values, e.g., 0, 5 and 10. Table S4 shows that GNNSynergy achieves the best performance when $t = 0$.

Table S4: **The performance of GNNSynergy when t is set as different values.**

Values	MSE	RMSE	MAE	Sperman	Pearson
$t = 0$	95.462	9.613	6.422	0.667	0.714
$t = 5$	96.087	9.655	6.435	0.665	0.708
$t = 10$	96.487	9.648	6.429	0.662	0.709

3 Results

3.1 Performance for Individual Cell Lines

In our main manuscript, we showed the average performance of various methods over all the cell lines. Here, we further investigated the performance of GNNSynergy on each individual cell line. The spearman correlation coefficients and pearson correlation coefficients for GNNSynergy, MatchMaker and DeepSynergy are shown in Fig. S3. Obviously, GNNSynergy performs better than the MatchMaker in most of cell lines. Besides, Fig. S4 shows the cell-line-specific MSE for GNNSynergy.

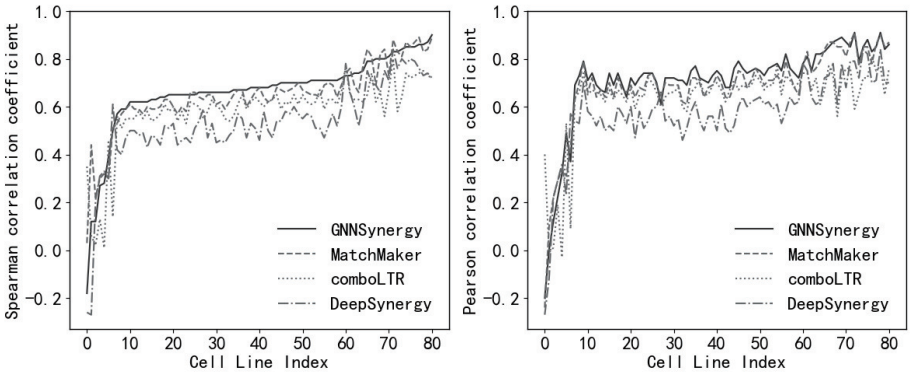


Figure S3: Spearman correlation coefficient and Pearson correlation coefficient comparison among GNNSynergy, DeepSynergy and MatchMaker. Here, MatchMaker is the best baseline as shown in our main manuscript. The gray horizontal dotted line represents the average score of GNNSynergy.

3.2 Attention Weights

As mentioned in the main manuscript, we adopted two weighting strategies to combine all sub-views, namely "Euqal Weights" and "Attention Weights". Here, we further explore the differences in the specific weights produced by these two strategies. Table S5 shows that the actual attention weights of 9 sub-views while the main-view cell is line SW-620.

3.3 Case studies for novel drug combination prediction

Table S6 shows 7 drug pairs predicted by our GNNSynergy and filtered by our literature search.

In particular, we searched the drug pairs together with the tissue information in PubMed abstracts, and we obtained 7 drug pairs as shown in Table S6. **TIOPRONIN** and **CAPTOPRIL** (Row 1) currently used in kindey disease

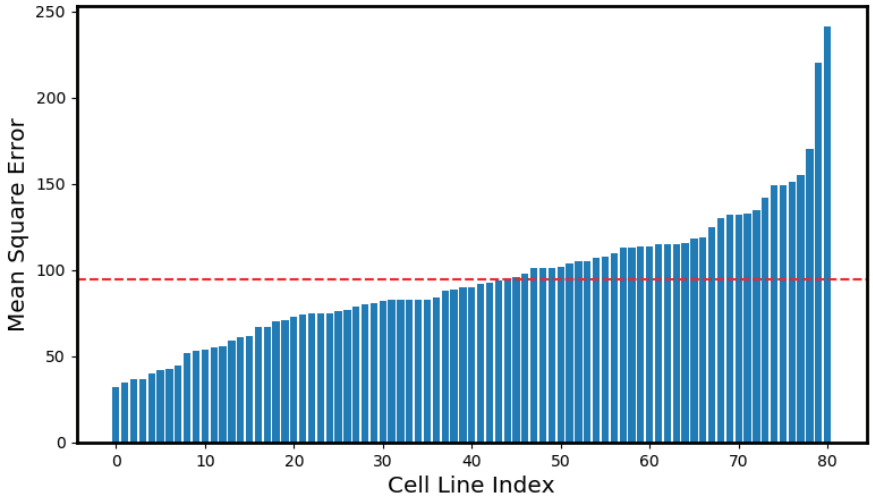


Figure S4: The MSE scores of GNNSynergy on each individual cell line. The red dotted line represents the average MSE score of GNNSynergy.

Table S5: **The weights for different sub-view cell lines in large intestine, when the main view cell line is SW-620.**

Sub-view Cell Lines	Equal Weights	Attention Weights
sub-view 1 (HT29)	0.1111	0.1101
sub-view 2 (HCT116)	0.1111	0.1432
sub-view 3 (HCT-15)	0.1111	0.0886
sub-view 4 (KM12)	0.1111	0.0985
sub-view 5 (COLO 205)	0.1111	0.0377
sub-view 6 (HCC-2998)	0.1111	0.1502
sub-view 7 (LOVO)	0.1111	0.0469
sub-view 8 (SW837)	0.1111	0.0571
sub-view 9 (RKO)	0.1111	0.1604

Table S6: **7 predicted drug combination pairs obtained by literature search. Note that this list was obtained by searching the PubMed abstracts with the tissue information instead of the cell line information.**

#	Rank	Drug 1	Drug 2	Cell line	Tissue	PMID
1	607	TIOPRONIN	CAPTOPRIL	SN12C, 786-0, CAKI-1	Kidney	15865542
2	1260	REFAMETINIB	TRAMETINIB	MDA-MB-468	Breast	35692083
3	1746	L-THYROXINE	Selenomethionine	K-562	Haematopoietic-and-lymphoid	22918596
4	2222	Selenomethionine	ACETYLCYSTEINE	A549, NCIH23	Lung	29799741
5	2415	BUTHIONINE SULFOXIMINE	D-Penicillamine	MCF7	Breast	28389407
6	2837	CAPTOPRIL	METFORMIN	CAKI-1	Kidney	34400938
7	3465	OMEPRAZOLE	CYSTEAMINE HCL	RKO	Large-intestine	35800723

regimens, has been demonstrated to lower the urinary supersaturation of cystine and significant improvement in urinary cystine capacity [1]. Tzeng *et al.* found that both **REFAMETINIB** and **TRAMETINIB** (Row 2) were significant in inhibiting the overexpression of LSM1 [2]. However, the combination of these two drugs was not tested in their paper. Also, Krysiak *et al.* investigated that **L-THYROXIN** and **Selenomethionine** (Row 3) administered together can produce a beneficial effect on haemostasis in euthyroid patients [3]. Besides, Hou *et al.*'s study showed that **Selenomethionine** and **ACETYLCYSTEINE** (also known as N-ACETYLCYSTEINE or NAC) in Row 4 in combination can significantly alleviate the downswing of cell viability, glutathione production, and phagorytosis induced by AFB1 and OTA in porcine alveolar macrophages [4]. **D-penicillamine** (DPEN) and **BUTHIONINE SULFOXIMINE** (Row 5) is an effective drug combination as DPEN-induced cancer cell toxicity was dependent on Cu and was significantly enhanced by depletion of glutathione using BUTHIONINE SULFOXIMINE (BSO) [5]. However, the experiments were conducted in breast cancer cell line MB231, which is different from our predicted MCF7 cell line. According to the research results of Paseban *et al.*, it was demonstrated that co-administration of ASA (aspirin), M (**METFORMIN**), C (**CAPTOPRIL**) and AT (atorvastatin), could improve glucose and lipid metabolism. Meanwhile, the administration of these drugs in diabetic rats resulted in a more pronounced increase in oxidative stress markers in both kidney and heart tissues compared to the administration of these drugs alone [6]. With regard to the last pair **OMEPRAZOLE** and **CYSTEAMINE HCL**, as reported in [7], OMEPRAZOLE treatment mitigates anxiety-related behaviours in the CYSTEAMINE HCL induced animal model of gastrointestinal (GI) disorder.

Table S7: **Drug chemical structure features collected from ChemoPy.**

Chemical Feature Type	**Feature Count**
Constitutional descriptors	30
Connectivity descriptors	44
Kappa descriptors	7
Basak descriptors	21
E-state descriptors	245
Burden descriptors	32
Moreau-Broto autocorrelation descriptors	32
Moran autocorrelation descriptors	32
Geary autocorrelation descriptors	32
Molecular property descriptors	6
MOE-type descriptors	60
Geometric descriptors (3D)	12
WHIM descriptors (3D)	70
MoRSE descriptors (3D)	210
RDF descriptors(3D)	180

3.4 Drug features

Drug chemical structure features are calculated by a Python library called ChemoPy. The ChemoPy Python library can compute 16 drug feature groups consisting of 19 de- scriptors, including 1135 descriptor values. We collected SMILE of all drugs from DrugBank and PubChem, then input them into Chemopy for calculation, and finally obtained the 541 2D and 472 3D descriptor values of the drugs Table S7 shows the chemical feature types and corresponding number of features.

4 References

[1] Dolin DJ, Asplin JR, Flagel L, Grasso M, Goldfarb DS. Effect of cystine-binding thiol drugs on urinary cystine capacity in patients with cystinuria. J Endourol. 2005 Apr;19(3):429-32. doi: 10.1089/end.2005.19.429. PMID: 15865542.

[2] Tzeng YT, Tsui KH, Tseng LM, Hou MF, Chu PY, Sheu JJ, Li CJ. Integrated analysis of pivotal biomarker of LSM1, immune cell infiltration and therapeutic drugs in breast cancer. J Cell Mol Med. 2022 Jul;26(14):4007-4020. doi: 10.1111/jcmm.17436. Epub 2022 Jun 12. PMID: 35692083; PMCID: PMC9279588.

[3] Krysiak R, Okopien B. Haemostatic effects of levothyroxine and selenomethionine in euthyroid patients with Hashimoto's thyroiditis. Thromb Haemost. 2012 Nov;108(5):973-80. doi: 10.1160/TH12-04-0275. Epub 2012 Aug 23. PMID: 22918596.

[4] Hou L, Zhou X, Gan F, Liu Z, Zhou Y, Qian G, Huang K. Combination of Selenomethionine and N-Acetylcysteine Alleviates the Joint Toxicities of Aflatoxin B1 and Ochratoxin A by ERK MAPK Signal Pathway in Porcine Alveolar Macrophages. J Agric Food Chem. 2018 Jun 13;66(23):5913-5923. doi: 10.1021/acs.jafc.8b01858. Epub 2018 Jun 1. PMID: 29799741.

[5] Sciegienka SJ, Solst SR, Falls KC, Schoenfeld JD, Klinger AR, Ross NL, Rodman SN, Spitz DR, Fath MA. D-penicillamine combined with inhibitors of hydroperoxide metabolism enhances lung and breast cancer cell responses to radiation and carboplatin via H2O2-mediated oxidative stress. Free Radic Biol Med. 2017 Jul;108:354-361. doi: 10.1016/j.freeradbiomed.2017.04.001. Epub 2017 Apr 5. PMID: 28389407; PMCID: PMC5495544.

[6] Paseban M, Niazmand S. The Comparison of Antioxidant Effect of Aspirin, Metformin, Atorvastatin and Captopril Co-administration in the Heart and Kidney Tissues of Diabetic Rats. Iran J Pharm Res. 2021 Winter;20(1):27-39. doi: 10.22037/ijpr.2019.112004.13481. PMID: 34400938; PMCID: PMC8170761.

[7] Sri Rethinavel H, Selvaraj DB, Balakrishnan SJ, Vergil Andrews JF, Joseph JHM, Kandasamy M. Omeprazole treatment manifests anxiolytic effects in a cysteamine hydrochloride induced mouse model of gastrointestinal disorder. Heliyon. 2022 Jun 24;8(6):e09787. doi: 10.1016/j.heliyon.2022.e09787. PMID: 35800723; PMCID: PMC9253648.