Supplementary Meterial of CATSyn: Predicting Synergistic Drug Combinations through Context-Aware Heterogeneous Graph Convolution Model

I. Experiments

A. Datasets

We used two large-scale *in vitro* screening datasets for performance evaluation, i.e., the O'Neil dataset [1] [4] and the NCI-ALMANAC [5] dataset. The O'Neil dataset runs a high-throughput screening using a 4-by-4 dosing regimen to get an unbiased identification of synergistic and efficacious drug combinations, and covers 38 diverse anticancer drugs, 39 human cancer cell lines and in total 22,737 samples of combinations. The NCI-ALMANAC dataset comprises 304,549 samples distinguished by the ComboScores. The data are derived from pairwise combinations of 104 FDA-approved drugs across the 60 human cancer cell lines included in the NCI-60 panel. Synergistic effect is quantified by four synergy types, namely Loewe additivity (Loewe) [6], Bliss independence (Bliss), zero interaction potency (ZIP), and highest single agent (HSA), respectively. Following the work of [3], we choose the thresholds widely used 0, 30 for the Loewe score, -3.37, 3.68 for the Bliss score, -3.02, 3.87 for the HSA score, and -4.48, 2.64 for the ZIP score, to separate synergy effects into three categories: synergistic, additive and antagonistic.

B. Initial Features

Informative initial features for drugs and cell lines are required as input to the graph convolutional network for representation learning. For a drug, we choose drug molecular fingerprint, which is a numerical vector calculated on the drug's SMILES sequence. We choose 300 as the length of the vector in this study. For cell lines, we use the expression profiles of landmark genes obtained from [7] and processed by [8]. Specifically, for each cell line, its expression profiles are represented as a 972-dimensional vector.

C. Cross-Validation Strategies

We implemented a 10-fold cross-validation strategy to evaluate our model. The entire set of drug combinations and cell line triplets was randomly divided into ten parts of equal size. For each validation cycle, one part was selected as the test set, another as the validation set, and the remaining eight parts were used as the training set. This approach ensures a comprehensive and unbiased assessment of the model across different subsets of the data, allowing for a robust evaluation of its predictive performance.

Additionally, we conducted two unseen data cross-validation settings, namely 'leave drug out' and 'leave cell line out'. In the 'leave drug out' setting, drugs were divided into training, validation, and test sets. For any drug combination synergy data, if either drug in the pair was part of the validation or test drug sets, that data was excluded from training, ensuring no overlap of drugs in the combination training data with those in the validation or test drug sets. Similarly, the 'leave cell line out' setting was designed to ensure that cell lines in the validation and test cell line sets were not present in the training data. These setups were crucial for assessing the model's performance on unseen data and demonstrating its robustness in diverse scenarios.

D. Baseline Models for Comparison

- **DeepSynergy** [9] employs a three-layer feed forward neural network to predict synergy scores using drug chemical structure and cell line gene expression profiles.
- **MatchMaker** [8] conducts drug specific subnetworks to get more precise predictions of synergy scores.
- **TranSynergy** [10] integrates a self-attention transformer component with input dimension reduction and output fully connected components.
- **HypergraphSynergy** [11] represents drugs and cell lines as nodes and their synergistic combinations as hyperedges, leveraging a hypergraph neural network for learning and predicting drug synergy.
- **MGAE-DC** [3], a multi-channel graph autoencoder-based method, learns the drug embeddings by considering not only synergistic combinations but also additive and antagonistic ones as three input channels.
- **MFSynDCP** [12] a multi-source feature collaborative interactive learning meethod.

E. Evaluation Metrics

In the evaluation of our model, we used Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Pearson

[1]It should be noted that this dataset was also called Merck dataset in the literature, e.g., [1]. In this paper, we follow [2], [3] to denote it as O'Neil dataset.

TABLE I
THE REGRESSION PERFORMANCES ON THE O'NEIL DATASET UNDER THE 10-FOLD RANDOM CV SETTING

Method	MSE	RMSE	PCC
	Loewe		
DeepSynergy [9]	226.04±11.12	15.03±0.37	0.71±0.02
TranSynergy [10]	208.73±12.26	14.44±0.32	0.77±0.02
Hypergraph [11]	184.14±18.21	13.57±0.72	0.79±0.02
MGAE-DC [3]	162.21±10.25	12.73±0.43	0.83±0.01
MFSynDCP	167.81±14.74	12.96±0.57	0.82±0.01
CATSyn (ours)	**149.25±15.44**	**12.20±0.63**	**0.84±0.01**
	Bliss		
DeepSynergy [9]	29.31±2.98	5.41±0.36	0.72±0.05
TranSynergy [10]	26.56±3.29	5.62±0.31	0.75±0.01
Hypergraph [11]	25.29±4.16	5.01±0.42	0.77±0.03
MGAE-DC [3]	17.36±3.17	4.15±0.39	**0.84±0.02**
MFSynDCP	19.20±3.45	4.37±0.43	0.83±0.01
CATSyn (ours)	**16.44±2.96**	**4.03±0.36**	**0.84±0.02**
	ZIP		
DeepSynergy [9]	16.11±0.86	4.01±0.11	0.77± 0.01
TranSynergy [10]	15.48±0.67	3.62±0.16	0.78±0.02
Hypergraph [11]	13.52±0.54	3.68±0.06	0.79±0.02
MGAE-DC [3]	10.68±0.41	3.27±0.06	0.85±0.01
MFSynDCP	11.67±0.66	3.40±0.08	0.85±0.02
CATSyn (ours)	**10.35±0.58**	**3.21±0.09**	**0.86±0.01**
	HSA		
DeepSynergy [9]	29.71±2.41	5.45 ± 0.22	0.71 ± 0.02
TranSynergy [10]	27.14±4.06	5.19±0.32	0.73±0.02
Hypergraph [11]	26.11±4.18	5.10±0.41	0.76±0.02
MGAE-DC [3]	17.89±2.17	4.22±0.26	0.83±0.01
MFSynDCP	18.16±3.07	4.26±0.33	0.82±0.01
CATSyn (ours)	**16.84±2.54**	**4.09±0.32**	**0.84±0.01**

TABLE II
THE REGRESSION PERFORMANCES ON THE NCI-ALMANAC DATASET UNDER THE 10-FOLD RANDOM CV SETTING

Method	MSE	RMSE	PCC
	Loewe		
DeepSynergy [9]	135.71±13.92	11.63±0.61	0.72±0.03
TranSynergy [10]	132.73±6.26	11.54±0.32	0.72±0.02
Hypergraph [11]	131.14±7.21	11.51±0.72	0.72±0.02
MGAE-DC [3]	121.18±1.92	11.01±0.09	0.75±0.01
MFSynDCP	125.76±7.56	11.20±0.32	0.74±0.01
CATSyn (ours)	**116.84±9.67**	**10.80±0.42**	**0.77±0.01**
	Bliss		
DeepSynergy [9]	18.08±0.28	4.25±0.03	0.68±0.01
TranSynergy [10]	16.96±0.61	4.13±0.31	0.70±0.01
Hypergraph [11]	16.89±1.16	4.10±0.42	0.71±0.03
MGAE-DC [3]	15.93±0.29	3.99±0.04	0.72±0.01
MFSynDCP	16.16±1.14	4.02±0.14	0.71±0.02
CATSyn (ours)	**14.82±0.75**	**3.85±0.09**	**0.73±0.01**
	ZIP		
DeepSynergy [9]	14.09±0.49	3.74±0.07	0.72±0.01
TranSynergy [10]	13.98±0.17	3.66±0.16	0.73±0.02
Hypergraph [11]	13.52±0.14	3.65±0.06	0.73±0.02
MGAE-DC [3]	12.88±0.12	3.59±0.02	0.75±0.01
MFSynDCP	12.56±0.21	3.54±0.03	0.75±0.01
CATSyn (ours)	**11.94±0.16**	**3.45±0.02**	**0.76±0.01**
	HSA		
DeepSynergy [9]	14.32±0.58	3.78±0.08	0.74±0.02
TranSynergy [10]	14.14±0.06	3.73±0.32	0.73±0.02
Hypergraph [11]	14.11±0.18	3.72±0.41	0.76±0.02
MGAE-DC [3]	13.63±0.15	3.69±0.02	0.76±0.01
MFSynDCP	13.70±0.26	3.70±0.02	**0.76±0.01**
CATSyn (ours)	**11.98±0.85**	**3.46±0.12**	**0.79±0.01**

Correlation Coefficient (PCC) as the primary metrics for assessing the accuracy of the regression predictions of the synergy scores.

F. Performance Comparisons

1) Random Cross-Validation (CV): Table I and II report the regression results of CATSyn and baseline models on the random CV setting. CATSyn demonstrates excellent performance in predicting drug synergy. In the regression task, CATSyn reduces the Loewe RMSE to 12.20. Similar trend has been observed in terms of PCC. The results illustrate CATSyn's superior predictive capabilities for the problem of drug synergy prediction.

Under various settings of synergy scores, our model consistently outperforms the existing state-of-the-art models by a significant margin in terms of MSE metric. Moreover, in all cases, our method also demonstrates superior performance in terms of PCC. Our model achieves similar results on the NCI-ALMANAC dataset, consistently outperforming the existing models by a significant margin. This results highlight the robustness and effectiveness of CATSyn in predicting drug synergy across different datasets and evaluation metrics.

2) Unseen Data Cross-Validation: Table III and IV shows the the regression result on unseen data CV settings (performance using other scoring metrics are shown in supplementary materials). In evaluating the robustness of CATSyn against unseen data, two distinct 10-fold cross-validation scenarios were employed: leave-drug-out and leave-cell-line-out. In the leave-drug-out CV, CATSyn significantly outperformed all baseline models. It achieved the lowest Mean Squared Error (MSE) of 357.23, a considerable improvement over the next best model. The leave-cell-line-out CV results further emphasized CATSyn's superior performance. It achieved an even more impressive MSE improvement.This result suggests that incorporating dynamic network related to cell line information can further strengthen the model's ability to generalize to unseen cell lines. These results, both in leave-drug-out and leave-cell-line-out settings, underscore CATSyn's exceptional ability to generalize and accurately predict drug synergy in scenarios involving novel drugs and cell lines, highlighting its potential as a robust tool in personalized medicine.

TABLE III
REGRESSION PERFORMANCE UNDER THE SETTING OF 10-FOLD LEAVE-DRUG-OUT CV USING LOEWE SCORE ON O'NEIL DATASET

Method	MSE	RMSE	PCC
DeepSynergy [9]	509.86±235.21	21.97±5.21	0.41±0.11
TranSynergy [10]	486.73±232.26	21.34±5.32	0.41±0.08
Hypergraph [11]	481.14±218.21	21.31±5.72	0.41±0.08
MGAE-DC [3]	474.02±236.71	21.15±5.17	0.45±0.08
MFSynDCP [12]	466.14±223.17	21.09±4.96	0.45±0.08
CATSyn (ours)	**435.05±97.14**	**20.73±2.30**	**0.47±0.06**

TABLE IV
REGRESSION PERFORMANCE UNDER THE SETTING OF 10-FOLD LEAVE-CELL-LINE OUT CV USING LOEWE SCORE ON O'NEIL DATASET

Method	MSE	RMSE	PCC
DeepSynergy [9]	443.21±282.41	20.21±5.94	**0.57±0.07**
TranSynergy [10]	408.73±322.26	19.28±5.32	0.55±0.02
Hypergraph [11]	398.14±298.21	18.93±5.72	0.55±0.02
MGAE-DC [3]	386.77±364.1	18.12±7.64	**0.57±0.14**
MFSynDCP [12]	369.50±317.24	18.06±6.64	0.27±0.11
CATSyn (ours)	**367.21±163.25**	18.11±3.96	**0.57±0.07**

TABLE V
ABLATION STUDY RESULTS ON 10-FOLD RANDOM CV SETTING

Method	MSE	RMSE	PCC
Loewe			
CATSyn	**149.25±15.44**	**12.20±0.63**	**0.84±0.01**
w/o CA Attention	165.47±15.66	12.84±0.60	0.81±0.03
Static Weight	162.25±13.45	12.76±0.58	0.80±0.02
w/o Universal	152.91±11.27	12.37±0.50	0.82±0.01
Bliss			
CATSyn	**16.44±2.96**	**4.03±0.36**	**0.84±0.02**
w/o CA Attention	19.03±2.24	4.32±0.30	0.80±0.02
Static Weight	18.52±2.11	4.29±0.29	0.80±0.01
w/o Universal	17.42±1.24	4.15±0.16	0.81±0.01
ZIP			
CATSyn	**10.35±0.58**	**3.21±0.09**	**0.86±0.01**
w/o CA Attention	11.17±0.27	3.33±0.05	0.84±0.01
Static Weight	10.78±0.39	3.27±0.06	0.85±0.01
w/o Universal	10.57±0.47	3.25±0.08	0.84±0.02
HSA			
CATSyn	**16.84±2.54**	**4.09±0.32**	**0.84±0.01**
w/o CA Attention	19.60±2.66	4.42±0.33	0.82±0.01
Static Weight	18.33±2.16	4.28±0.30	0.82±0.01
w/o Universal	18.25±1.99	4.27±0.25	0.82±0.01

G. Ablation Study

In our ablation study, we meticulously investigated the individual contributions of the context-aware attention module and the universal nodes. This analysis aimed to delineate the impact of each module on the overall performance of the model.

1) Context-Aware Attention Module: To emphasize the significance of the model's dynamic nature with respect to cell lines, our ablation study of the context-aware attention module focused on evaluating this aspect. We implemented two alternative approaches to ascertain its overall impact. The first approach is replacing the module with a mechanism that simply averages the results from all heads. This approach was implemented to test the impact of assigning different weights to various channels and heads on the model's effectiveness. The second approach substituted the module with a cell line-independent learnable weighted average. This method was designed to test the impact of dynamically assigning weights based on cell line information. These modifications allowed us to comprehensively analyze how the context-aware attention module contributes to the model's performance. The performance is shown in Table V, the **w/o** CA Attention is the result of removing context-aware attention module, the Static Weight is the result of replacing the context-aware attention module with a static weight.

The results from our ablation study clearly indicate the critical importance of dynamic weighting in the context-aware attention module. The superiority of this module over the alternatives, which both yielded similar outcomes, highlights its ability to dynamically adjust weights based on cell line-specific information. This dynamic weighting is key in effectively capturing the unique features of different cell lines and the dynamic feature of drugs on different cell lines, significantly enhancing the model's predictive accuracy. These findings emphasize the essential role of context-specific dynamic weighting in accurately predicting drug synergy across varied cellular environments.

TABLE VI
ABLATION STUDY OF UNIVERSAL NODES ON 10-FOLD LEAVE CELL LINE OUT CV SETTING

Method	MSE	RMSE	PCC
Loewe			
CATSyn	**367.21±163.25**	18.11±3.84	**0.57±0.07**
w/o Universal	397.37±181.66	19.53±4.02	0.51±0.08
Bliss			
CATSyn	**47.69±19.25**	**6.73±1.41**	**0.43±0.10**
w/o Universal	50.37±18.08	6.98±1.24	0.36±0.07
ZIP			
CATSyn	**31.60±13.09**	**5.39±1.50**	**0.44±0.09**
w/o Universal	33.12±9.23	5.70±0.80	0.41±0.10
HSA			
CATSyn	**39.78±22.17**	**6.13±1.17**	**0.54±0.09**
w/o Universal	46.13±14.80	6.70±1.11	0.45±0.07

2) Universal Nodes: In our ablation study for the Universal Node, we chose to remove this component from the model and also excluded its input feature from the final synergy score prediction module. This approach allowed us to directly assess the impact of adding universal node on the model's performance, particularly in terms of its ability to generalize across different cell lines and enhance the robustness of the predictions. Therefore we conduct both random CV and leave cell line out CV test. The performance on random CV setting is shown in Table V, the **w/o** Universal line. And the performance on leave cell line out CV setting is shown in Tabel VI.

The observed decline in model performance in both random CV and leave-cell-line-out CV tests, upon removal of the universal nodes, underscores its critical role. This decline suggests that the module significantly contributes to the model's ability to handle diversity across different cell lines, enhancing both accuracy and generalizability. Without it, the model's capacity to integrate and utilize the unique characteristics of each cell line is diminished, leading to a notable reduction in predictive performance.

REFERENCES

[1] M. Besharatifard and F. Vafaee, "A review on graph neural networks for predicting synergistic drug combinations," *Artificial Intelligence Review*, vol. 57, no. 3, p. 49, 2024.

[2] 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*, vol. 23, no. 6, p. bbac403, 2022.

[3] P. Zhang and S. Tu, "MGAE-DC: Predicting the synergistic effects of drug combinations through multi-channel graph autoencoders," *PLoS computational biology*, vol. 19, no. 3, p. e1010951, 2023.

[4] 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.

[5] 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.

[6] S. Loewe, "The problem of synergism and antagonism of combined drugs," *Arzneimittel-forschung*, vol. 3, no. 6, pp. 285–290, 1953.

[7] F. Iorio, T. A. Knijnenburg, D. J. Vis, G. R. Bignell, M. P. Menden, M. Schubert, N. Aben, E. Gonçalves, S. Barthorpe, H. Lightfoot *et al.*,

"A landscape of pharmacogenomic interactions in cancer," *Cell*, vol. 166, no. 3, pp. 740–754, 2016.

[8] H. I. Kuru, O. Tastan, and A. E. Cicek, "MatchMaker: a deep learning framework for drug synergy prediction," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 19, no. 4, pp. 2334–2344, 2021.

[9] 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.

[10] 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.

[11] X. Liu, C. Song, S. Liu, M. Li, X. Zhou, and W. Zhang, "Multi-way relation-enhanced hypergraph representation learning for anti-cancer drug synergy prediction," *Bioinformatics*, vol. 38, no. 20, pp. 4782–4789, 2022.

[12] Y. Dong, Y. Chang, Y. Wang, Q. Han, X. Wen, Z. Yang, Y. Zhang, Y. Qiang, K. Wu, X. Fan *et al.*, "Mfsyndcp: multi-source feature collaborative interactive learning for drug combination synergy prediction," *BMC bioinformatics*, vol. 25, no. 1, p. 140, 2024.