

RESEARCH

Open Access



Accurate prediction of synergistic drug combination using a multi-source information fusion framework

Shuting Jin^{1,2}, Huaze Long¹, Anqi Huang¹, Jianming Wang³, Xuan Yu⁴, Zhiwei Xu¹ and Junlin Xu^{1*}

Abstract

Background Accurately predicting synergistic drug combinations is critical for complex disease therapy. However, the vast search space of potential drug combinations poses significant challenges for identification through biological experiments alone. Nowadays, deep learning is widely applied in this field. However, most methods overlook the important role of protein–protein interaction networks formed by gene expression products and the pharmacophore information of drugs in predicting drug synergy.

Results We propose MultiSyn, a multi-source information integration method for the accurate prediction of synergistic drug combinations. Specifically, we design a semi-supervised learning framework using an attributed graph neural network to integrate protein–protein interaction networks of gene expression products with multi-omics data, constructing initial cell line representations that incorporate multi-source information. Furthermore, we refine the initial cell line representation by adaptively integrating it with normalized gene expression profiles, enabling the extraction of cell line features that encapsulate global information. In addition, we decompose drugs into fragments containing pharmacophore information based on chemical reaction rules and construct a heterogeneous graph comprising atomic and fragment nodes. To enhance the capture of molecular structural information, we introduce a heterogeneous graph transformer to learn multi-view representations of heterogeneous molecular graphs. Extensive experiments show that MultiSyn outperforms several classical and state-of-the-art baselines in synergistic drug combination prediction tasks.

Conclusions This study provides a powerful tool for inferring promising synergistic drug combinations. By leveraging attention mechanisms and pharmacophore information, MultiSyn identifies key substructures that are critical for synergy. Further visualization and case studies validate its effectiveness in capturing biologically meaningful features and identifying potential drug combinations.

Keywords Drug combination, Graph neural network, Multi-source information, Omics data

*Correspondence:

Junlin Xu

xjl@wust.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Drug combination therapy enhances treatment efficacy against complex diseases by leveraging agents with distinct mechanisms of action. Compared to monotherapies, such combinations can reduce toxicity and delay the onset of drug resistance [1, 2]. Traditional discovery methods primarily depend on clinical observations and *in vitro* experiments; however, the exponential growth in candidate drug pairs makes exhaustive experimental validation infeasible. As a result, increasing attention has been directed toward computational approaches for predicting drug synergy. Early computational strategies were grounded in systems biology models, which integrate prior biological knowledge to simulate protein interactions and signaling pathways for guiding combination strategies. For instance, Zhao et al. [3] proposed a model that predicts effective drug combinations by analyzing correlations between candidate and confirmed pairs. Yin et al. [4] utilized three-node enzymatic motifs to characterize synergistic and antagonistic interactions. However, these methods often suffer from limited scalability due to complex modeling assumptions and dependence on incomplete biological networks.

To address these limitations, researchers have begun exploring machine learning (ML) methods to automatically learn drug properties and predict the synergistic effects of various drug combinations, thereby reducing the costs and time associated with drug trials. One of the early ML-based drug synergistic combination classification methods proposed a strategy that integrates molecular and pharmacological data as side information and performs drug combination prediction and classification by maximizing the F1 score [3]. Iwata et al. developed a more complex model using target proteins and ATC drug codes to predict beneficial drug combinations [5]. SyDRa [6] applied a random forest algorithm to identify synergistic anticancer drug combinations based on three features: drug-chemical structure, drug-target network, and pharmacogenomics.

With the continuous advancement of ML technologies, deep learning (DL), a subfield of ML, has become an essential tool for drug combination prediction due to its ability to process complex data and automatically extract high-level features. By leveraging deep neural networks, DL can learn intricate patterns from

raw data, demonstrating significant potential, particularly with high-dimensional and large-scale datasets. Early DL based drug synergy prediction models, such as DeepSynergy [7], made predictions by integrating molecular and genomic data. More recent DL methods combine multiple drug and cell line features, including gene expression, copy number variations, and drug targets, to enhance prediction accuracy [8–10]. Although these methods have advanced the field, they mostly focus on capturing complex information from a single perspective, lacking a comprehensive consideration of biological networks related to specific biological functions and the pharmacophore structural features of drugs. These limitations hinder current models from fully capturing the underlying biological and chemical mechanisms of drug synergy. In other domains of drug discovery, the integration of multi-omics data to enhance predictive performance has been extensively validated. For example, DeepDRA [11] is a DL framework that integrates multi-omics profiles with drug descriptors and molecular fingerprints to enhance drug response prediction. This integration improves both predictive accuracy and generalization across datasets. HGTDRA [12] builds a large-scale, heterogeneous biomedical knowledge graph and employs a heterogeneous graph transformer to extract features, enabling flexible input handling and comprehensive integration of information from diverse biomedical entities.

In light of this, we propose a multi-source information integration method, MultiSyn, as illustrated in Fig. 1, which enhances the precise prediction of anticancer drug combinations by comprehensively integrating biological networks, multi-omics data, and drug structural features. MultiSyn is designed to address two key limitations in prior models: the underuse of pharmacophoric substructures—functional groups essential for drug activity—and the features used to represent cell lines lack corresponding biological network context. Specifically, we design a semi-supervised attributed graph neural network that employs graph attention network (GAT) to integrate cell line-associated protein–protein interaction (PPI) networks and multi-omics data, obtaining more accurate initial feature embeddings for cell lines. Furthermore, we refined the final cell line features by combining these initial representations with normalized gene expression

(See figure on next page.)

Fig. 1 The framework of the proposed MultiSyn model. **a** MultiSyn integrates PPI networks and cell line-related omics data using an attributed GAT to construct initial cell line features containing multi-source information. **b** The initial multi-source cell line features are combined with gene expression data to obtain the final cell line features, which are then connected with drug structural features containing pharmacophore information for the final drug synergy prediction. **c** The improved heterogeneous graph transformer learns drug structural features containing pharmacophore information

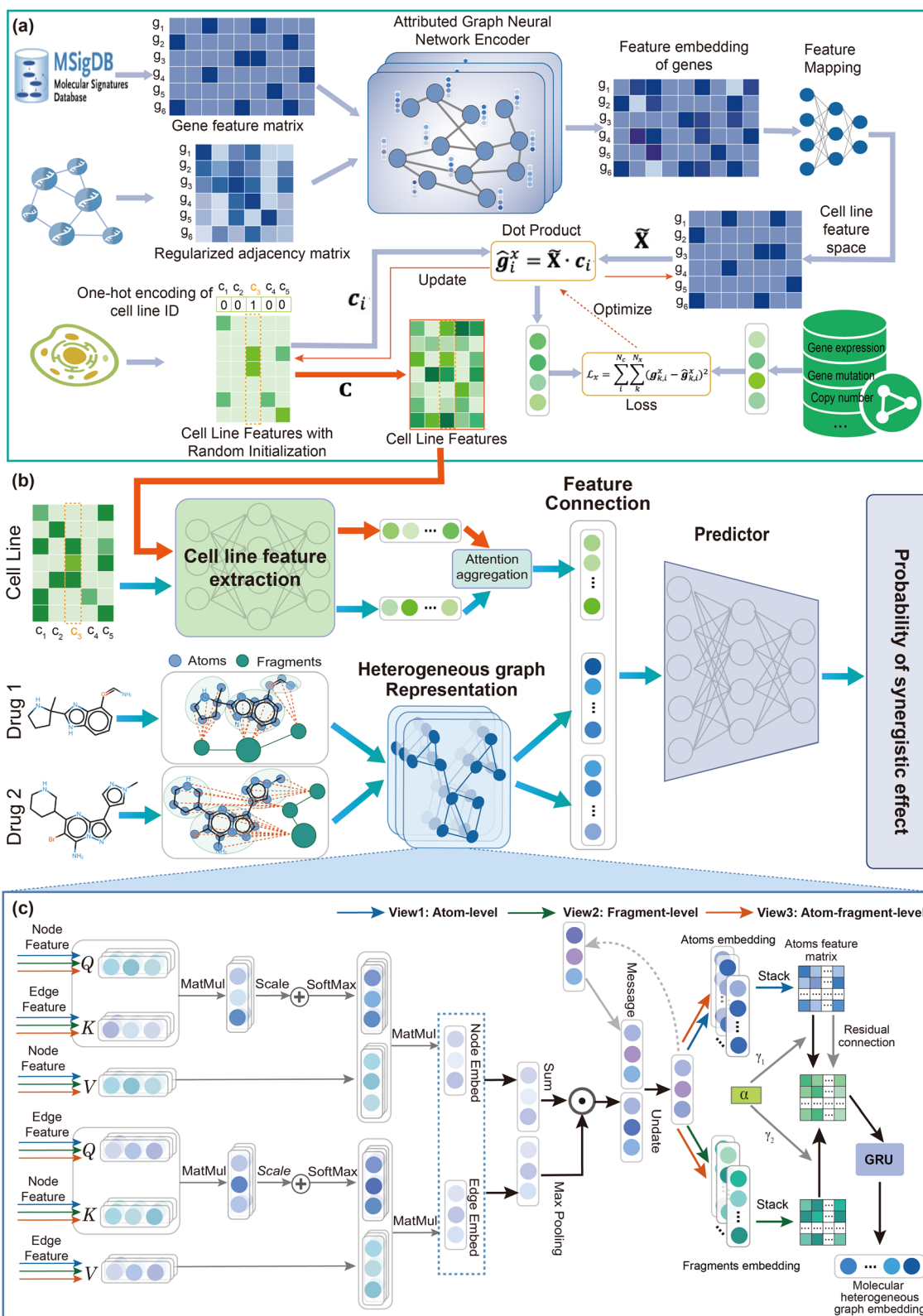


Fig. 1 (See legend on previous page.)

data. Additionally, to capture chemical structural information related to specific biological functions, we leverage domain-specific chemical knowledge to represent each drug molecule as a heterogeneous graph consisting of atomic nodes and fragment nodes that carry pharmacophore information. An improved heterogeneous graph transformer is then used to extract and process the structural information of drugs. Finally, drug features are combined with cell line representations and fed into a predictor for accurate drug synergy prediction. We evaluate the performance of MultiSyn by comparing it with several classical and state-of-the-art methods on benchmark datasets, demonstrating its superior predictive capability. Moreover, additional experiments are conducted to assess the effectiveness of pharmacophore substructure recognition and to explore the practical efficiency of the model in identifying potential synergistic drug combinations. The main contributions of this study are as follows:

- We propose MultiSyn, a multi-source information fusion method for drug synergy prediction that integrates multi-omics data, biological networks, and drug molecular features containing pharmacophore information to identify synergistic combinations.
- We provide a fresh perspective for addressing the problem of drug synergy prediction. The incorporation of heterogeneous molecular graphs containing pharmacophore information enhances the interpretability of predictions, particularly in elucidating pharmacodynamic mechanisms.
- Experimental results on benchmark datasets demonstrate that the MultiSyn method outperforms existing approaches, and further case studies validate its effectiveness in practical applications.

Related works

Multi-omics data-based methods

With the advancement of DL algorithms, an increasing number of studies have begun to combine various drug and cell line features, such as gene expression, copy number variation, and drug targets, to enhance drug synergy prediction. For example, one of the earliest DL models, DeepSynergy [7], predicts drug synergy by integrating molecular and genomic data. However, due to its architecture, DeepSynergy lacks interpretability, making it difficult to assess the contribution of specific drug features to the prediction results. Zhang et al. [8] have improved prediction accuracy by integrating multi-omics data. Although these methods have made progress in boosting prediction performance, they often fail to incorporate drug structural information and the complex biological networks between drugs, diseases,

and proteins, which limits their accuracy. To address this limitation, AuDNNsynergy [9] integrates drug structural data with genomic data from The Cancer Genome Atlas (TCGA) [13]. Although multi-omics data integration has improved predictive accuracy, these methods still face challenges in handling the heterogeneity of multi-omics data and enhancing model interpretability.

Graph-based methods

With the successful application of graph neural networks (GNNs) in biological networks and small molecule characterization [14, 15], increasing attention has been given to leveraging biological networks or graph-based models to extract drug features for drug synergy prediction. For instance, TranSynergy [16] employs restarted random walks on the PPI network to extract drug features, while integrating data such as a novel drug-target profile and gene expression. DTSyn [17] uses a multi-head attention mechanism to capture interactions among chemical substructures, gene-gene associations, and chemical-cell line interactions. DeepDDS [18] proposes a framework based on two types of GNNs—GATs and graph convolutional networks (GCNs)—to combine molecular structure features with gene expression profiles, performing drug synergy prediction. DualSyn [19] combines graph attention mechanisms with high-order relations and global information modules to capture complex drug-cell line interactions. SynergyGTN [20] leverages a graph transformer network to capture hierarchical graph representations of drug combinations while integrating PPI network information to improve biological interpretability. SDDSynergy [21] introduces an attention mechanism to capture substructure-level interactions, highlighting the significance of molecular substructures and identifying the key drivers of drug synergy. These methods further confirm the crucial role of biological networks and molecular structural features in predicting synergistic drug interactions.

Multimodal-based methods

Recently, multimodal learning frameworks have been developed to capture a more comprehensive understanding of drug synergy by combining diverse data sources, such as drug molecular graphs, SMILES strings, and cell line gene expression data. These frameworks often use contrastive learning to map the features of different modalities to a unified representation space, addressing data sparsity and expanding drug combination datasets. For example, Pisces [22] employs contrastive learning to align the multimodal features of drugs and cell lines, thereby improving both data representation and drug combination prediction accuracy. Similarly, MMSyn [23] introduces a multimodal framework that combines

molecular graphs, molecular fingerprints, SMILES strings, and cell line features to further boost prediction accuracy. Deeptrasynergy [24] uses multimodal inputs, including drug–target interaction, protein–protein interaction, and cell–target interaction, and employs a Transformer to predict drug combination synergy. While these methods have led to significant improvements in accuracy, they also come with increased computational complexity.

Results

Experimental settings

Datasets

To ensure a fair comparison with state-of-the-art methods, we adopt the experimental design of DeepDDS [18] and utilize its preprocessed O'Neil drug combination dataset as the benchmark [25]. This dataset comprises 36 drugs and 31 cancer cell lines, forming a total of 12,415 triplets, each consisting of two drugs and one cancer cell line. Similarly, we obtain gene expression data for the cell lines from the Cancer Cell Line Encyclopedia (CCLE) [26] and SMILES [27] sequences for the drugs from DrugBank [28]. Furthermore, we refer to the PRODeepSyn approach [29] by collecting gene expression data for cell lines from the ArrayExpress database [30], gene mutation data from the COSMIC database [31], and PPI data from the STRING database [32] to capture multi-source information for cell lines.

Evaluation protocol and metrics

To benchmark the predictive performance of MultiSyn against baselines, we performed 5-fold cross-validation (CV) on the benchmark dataset. To further validate the robust performance of MultiSyn, we adopted multiple leave-one-out strategies to assess its generalization ability. In each setting, all samples associated with a specific drug, drug pair, or tissue type were entirely excluded from the training set, ensuring that the model had no access to any related information during training. We evaluated performance using seven metrics: the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPR), accuracy (ACC), balanced accuracy (BACC), precision (PREC), true positive rate (TPR), and Cohen's kappa (KAPPA). The definitions and formulas of these metrics are provided in Additional file 1: Evaluation Indicators and Calculation Formulas.

Baselines

To comprehensively evaluate the performance of MultiSyn, we compared it against a range of methods. These

include six classical ML methods, namely SVM, AdaBoost, XGBoost, Gradient Boosting Machine (GBM), MLP, and Random Forest (RF), as well as seven state-of-the-art DL approaches. Part of the performance results for baseline models are directly referenced from the original publication of DeepDDS [18], where these methods were comprehensively evaluated on the O'Neil dataset using a 5-fold cross-validation protocol. To ensure a fair, reproducible, and consistent comparison with the literature, we adopted the same data preprocessing procedures and evaluation settings as described in DeepDDS. Additionally, for models without reported results in the DeepDDS literature, we reproduced them under identical experimental conditions using the default parameters provided in its open-source code for comparison (see Additional file 1: Implementation Details for further information). A brief summary of the baseline DL approaches is provided below:

- DeepSynergy [7] is a DL model for predicting drug combination synergy, integrating chemical and genomic data through feature extraction using a multilayer perceptron.
- Transynergy [16] uses a self-attention mechanism to process knowledge from cell line gene dependencies, gene-gene interactions, and genome-wide drug-target interactions to predict synergistic drug combinations.
- GraphSynergy [33] integrates PPI networks and graph convolutional networks to predict synergistic drug combinations by capturing molecular interactions and key contributing proteins.
- DeepDDS [18] leverages graph neural networks and attention mechanisms to analyze molecular structures and genomic data, enabling precise identification of synergistic drug combinations by capturing intricate interactions between drugs and cell lines.
- DFFNDDS [34] combines a fine-tuned pretrained language model with a dual feature fusion mechanism that integrates drug and cell line features at both the bit-wise and vector-wise levels to predict synergistic drug combinations.
- AttenSyn [35] employs an attention-based deep graph neural network, leveraging graph neural networks and attention pooling to extract molecular features and interaction information between drug pairs, ultimately predicting synergistic effects of anti-cancer drug combinations.
- MFSynDCP [36] extracts drug substructures through graph aggregation and employs a multi-source feature interaction controller to predict synergistic effects of drug combinations.

Table 1 Performance comparison of various methods on benchmark dataset for drug synergy prediction

Methods		AUROC	AUPR	ACC	BACC	PREC	TPR	KAPPA
ML	SVM	0.58 ± 0.01	0.56 ± 0.02	0.54 ± 0.01	0.54 ± 0.01	0.54 ± 0.01	0.51 ± 0.12	0.08 ± 0.04
	Adaboost	0.83 ± 0.01	0.83 ± 0.03	0.74 ± 0.01	0.74 ± 0.02	0.74 ± 0.02	0.72 ± 0.01	0.48 ± 0.03
	XGBoost	0.92 ± 0.01	0.92 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.68 ± 0.01
	GBM	0.85 ± 0.02	0.85 ± 0.01	0.76 ± 0.02	0.76 ± 0.02	0.77 ± 0.01	0.74 ± 0.01	0.53 ± 0.04
	MLP	0.65 ± 0.02	0.63 ± 0.05	0.56 ± 0.06	0.56 ± 0.05	0.54 ± 0.04	0.53 ± 0.22	0.12 ± 0.04
	RF	0.86 ± 0.02	0.85 ± 0.02	0.77 ± 0.01	0.77 ± 0.01	0.78 ± 0.02	0.74 ± 0.01	0.55 ± 0.04
DL	DeepSynergy	0.88 ± 0.01	0.87 ± 0.01	0.80 ± 0.01	0.80 ± 0.01	0.81 ± 0.01	0.75 ± 0.01	0.59 ± 0.05
	TranSynergy	0.90 ± 0.01	0.89 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.80 ± 0.01	0.64 ± 0.01
	GraphSynergy	0.91 ± 0.01	0.90 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.84 ± 0.01	0.80 ± 0.01	0.64 ± 0.01
	DeepDDS*	<i>0.93 ± 0.01</i>	<i>0.93 ± 0.01</i>	0.85 ± 0.07	0.85 ± 0.07	0.85 ± 0.07	0.85 ± 0.07	0.71 ± 0.21
	DFFNDDS	<i>0.93 ± 0.01</i>	0.92 ± 0.01	<i>0.86 ± 0.01</i>	<i>0.86 ± 0.01</i>	<i>0.85 ± 0.01</i>	<i>0.86 ± 0.03</i>	<i>0.72 ± 0.03</i>
	AttenSyn	0.92 ± 0.01	0.91 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.83 ± 0.03	0.82 ± 0.03	0.67 ± 0.01
MFSynDCP	0.92 ± 0.01	0.92 ± 0.01	0.85 ± 0.01	0.85 ± 0.01	0.86 ± 0.01	0.86 ± 0.01	0.70 ± 0.01	
MultiSyn		0.95 ± 0.01	0.94 ± 0.01	0.88 ± 0.01	0.88 ± 0.01	0.87 ± 0.01	0.88 ± 0.01	0.75 ± 0.02

The entries in bold denote the best results of all methods and those in italic denote the second-best result

DeepDDS* displays the best results from DeepDDS-GAT and DeepDDS-GCN

Performance comparison on CV

We performed 5-fold CV by randomly partitioning the dataset at the sample level, where each sample is represented as a triplet consisting of a pair of drugs and a cancer cell line. Specifically, the dataset was divided into five mutually exclusive folds, each comprising 20% of the data. In each iteration, one part is used as the test set, and the remaining four parts are used for training. This process is repeated five times, with each part serving as the test set once. The final performance is averaged across all iterations to provide a more reliable evaluation of the model.

Table 1 reports the comparative experimental results under the same dataset partitioning setup, demonstrating that MultiSyn performs excellently on all key evaluation metrics. We observe that MultiSyn achieves an AUROC of 0.95, with ACC and BACC reaching 0.88, demonstrating improvements across nearly all metrics when compared to DeepDDS, which excels in AUROC and AUPR, and DFFNDDS, which offers balanced performance. Compared to the optimal ML method, XGBoost, MultiSyn shows significant improvements of 3.2% in AUROC and 10.2% in KAPPA.

To rigorously assess the statistical significance of MultiSyn's performance improvements, we conducted pairwise comparisons against a representative set of baseline DL models, including DeepSynergy, DeepDDS, AttenSyn, and MFSynDCP. To ensure a fair and statistically rigorous evaluation, we performed five independent runs of five-fold cross-validation, each using a distinct set of randomly generated fold indices. Within each run, all models were evaluated using the same

Table 2 Paired *t*-test *p*-values comparing MultiSyn with baseline models

Model	DeepSynergy	DeepDDS	AttenSyn	MFSynDCP
<i>p</i> -value	2.14×10^{-28}	1.61×10^{-14}	6.11×10^{-19}	3.00×10^{-24}

fold assignments to ensure consistency. This setup yielded 25 paired measurements per evaluation metric, which were used to assess statistical significance. Prior to hypothesis testing, we assessed the distribution of paired performance differences using the Shapiro–Wilk test and quantile–quantile (QQ) plots (see Additional file 1: Fig. S1, both of which supported the assumption of approximate normality and justified the use of parametric testing. We then applied paired *t*-tests to the results for key performance metrics, such as AUROC. As summarized in Table 2, the superiority of MultiSyn over all baseline models is statistically significant rather than incidental. These results reflect that the MultiSyn model achieves high accuracy and excels at identifying true sample labels, even in the presence of data imbalance, achieving superior predictive performance. Overall, the fair comparison of comprehensive performance highlights MultiSyn's great potential for accurately identifying synergistic drug combinations.

Performance evaluation by leave-one-out validation

To further evaluate the model's generalization ability in cold-start scenarios, we compared our method with thirteen drug synergy approaches across three

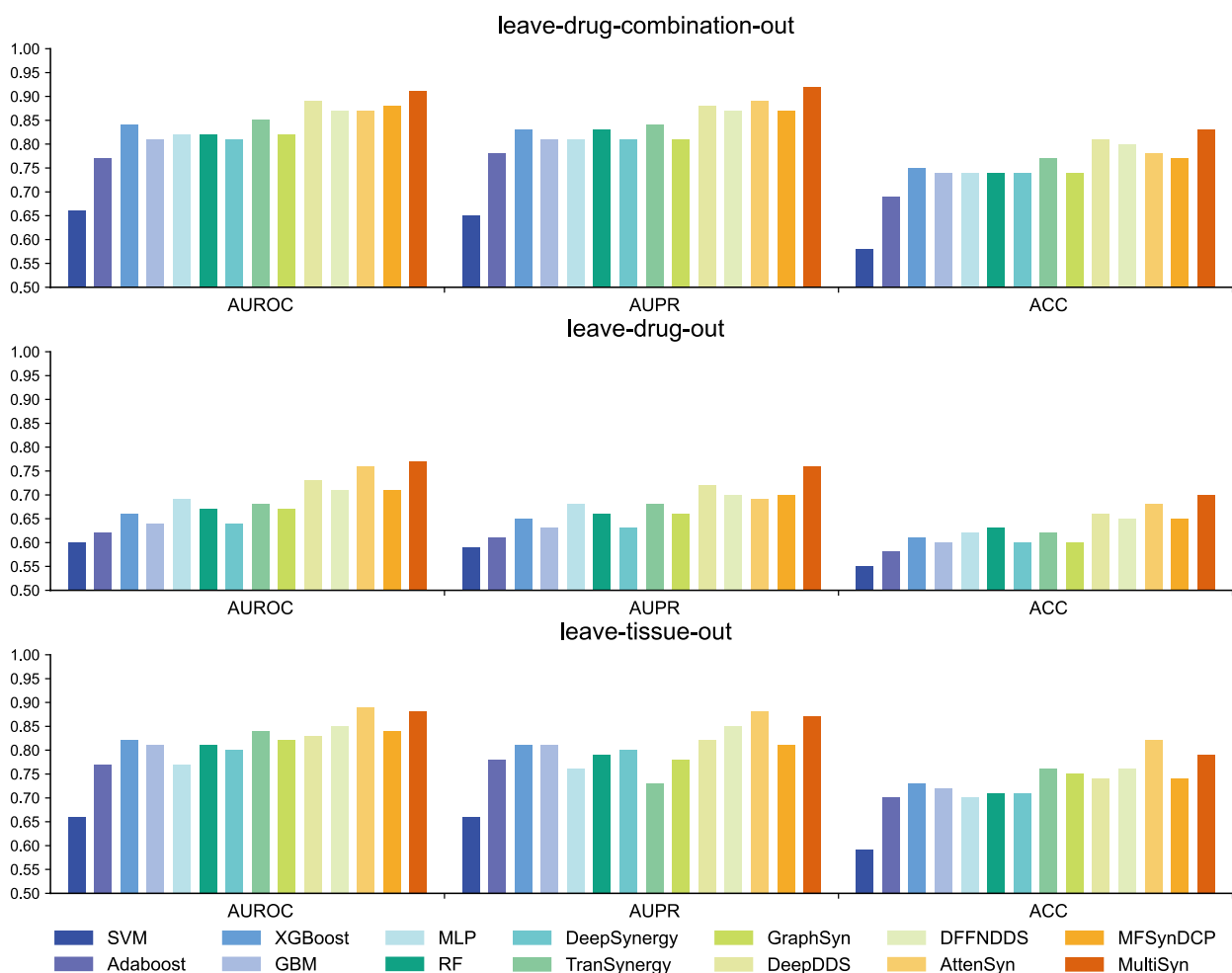


Fig. 2 Performance of various drug synergy prediction models under cold-start scenarios. Results are averaged over 5-fold cross-validation ($n = 5$), and all available values are provided in Additional file 2

cold-start scenarios: (i) leave-drug-combination-out: excludes specific drug combinations from the training data and tests the model on those combinations, (ii) leave-drug-out: excludes specific drugs from the training data and tests the model on those drugs, and (iii) leave-tissue-out: excludes data from certain tissue types and tests the model on those tissue types. As shown in Fig. 2, MultiSyn consistently achieves the highest scores across all tested cancer types, demonstrating its robustness and superior generalization ability. These results highlight the effectiveness of the MultiSyn framework in capturing complex molecular interactions and its adaptability to a wide range of cancer cell lines. Overall, these results validate the effectiveness of MultiSyn in learning robust and transferable feature representations, outperforming both classical and state-of-the-art methods in multiple cold-start scenarios.

Ablation study

To investigate the impact of the molecular graph feature extraction and cell line feature construction modules on the overall performance of the model, we considered the following variants of MultiSyn:

- MultiSyn-a: MultiSyn excludes features derived from heterogeneous graphs and relies solely on the commonly used molecular graph features containing atomic nodes.
- MultiSyn-g: The module that incorporates multi-source information for constructing cell line features (C_m) is removed, retaining only the gene expression-based cell line feature module (C_g).
- MultiSyn-m: The module that derives cell line features (C_g) from gene expression data is removed, retaining only the multi-source cell line feature module (C_m).

Table 3 Performance comparison of our proposed MultiSyn and its variants

Methods	AUROC	AUPR	ACC	BACC	PREC	TPR	KAPPA
MultiSyn	0.95	0.94	0.88	0.88	0.87	0.88	0.75
MultiSyn-a	0.92	0.91	0.84	0.84	0.82	0.83	0.69
MultiSyn-g	0.93	0.92	0.85	0.85	0.83	0.84	0.70
MultiSyn-m	0.93	0.92	0.85	0.85	0.83	0.85	0.71

Table 4 Performance comparison on independent dataset

Methods	AUROC	AUPR	ACC	BACC	PREC	TPR	KAPPA	
ML	SVM	0.47 ± 0.11	0.71 ± 0.13	0.54 ± 0.13	0.47 ± 0.15	0.70 ± 0.13	0.63 ± 0.11	-0.04 ± 0.15
	Adaboost	0.49 ± 0.09	0.69 ± 0.14	0.46 ± 0.17	0.47 ± 0.12	0.69 ± 0.14	0.46 ± 0.15	-0.05 ± 0.17
	XGBoost	0.52 ± 0.11	0.73 ± 0.12	0.45 ± 0.15	0.49 ± 0.11	0.71 ± 0.09	0.38 ± 0.17	-0.01 ± 0.14
	GBM	0.51 ± 0.10	0.71 ± 0.09	0.45 ± 0.12	0.47 ± 0.08	0.69 ± 0.14	0.43 ± 0.12	-0.03 ± 0.14
	MLP	0.53 ± 0.13	0.74 ± 0.12	0.53 ± 0.15	0.53 ± 0.15	0.74 ± 0.13	0.53 ± 0.13	0.05 ± 0.11
	RF	0.53 ± 0.14	0.76 ± 0.16	0.50 ± 0.14	0.54 ± 0.13	0.75 ± 0.14	0.49 ± 0.14	0.06 ± 0.11
DL	DeepSynergy	0.55 ± 0.15	0.71 ± 0.13	0.47 ± 0.41	0.53 ± 0.13	0.75 ± 0.14	0.39 ± 0.17	0.04 ± 0.15
	GraphSynergy	0.61 ± 0.12	0.80 ± 0.11	0.53 ± 0.11	0.55 ± 0.07	0.76 ± 0.10	0.45 ± 0.21	0.09 ± 0.13
	DeepDDS*	0.66 ± 0.12	0.82 ± 0.15	0.64 ± 0.15	0.62 ± 0.13	<i>0.80 ± 0.11</i>	0.67 ± 0.12	0.21 ± 0.29
	DFFNDDS	0.64 ± 0.11	0.82 ± 0.06	<i>0.66 ± 0.04</i>	0.56 ± 0.10	0.75 ± 0.08	0.66 ± 0.13	0.12 ± 0.13
	AttenSyn	0.65 ± 0.04	<i>0.82 ± 0.03</i>	0.65 ± 0.09	<i>0.61 ± 0.04</i>	0.79 ± 0.05	<i>0.73 ± 0.26</i>	0.19 ± 0.09
	MFSynDCP	0.63 ± 0.02	0.81 ± 0.01	0.61 ± 0.15	0.55 ± 0.03	0.78 ± 0.07	0.69 ± 0.34	0.11 ± 0.06
MultiSyn	0.67 ± 0.07	0.83 ± 0.05	0.67 ± 0.02	0.60 ± 0.03	0.80 ± 0.03	0.75 ± 0.04	<i>0.20 ± 0.08</i>	

The entries in bold denote the best results of all methods and those in italic denote the second-best result

DeepDDS* displays the best results from DeepDDS-GAT and DeepDDS-GCN

The comparison is based on a 5-fold CV test on the training dataset. The results are shown in Table 3. Among the variants, MultiSyn-a achieved the worst performance, indicating that incorporating molecular fragments containing pharmacophores and reaction information contributes to more reliable drug representations [37]. Such reliable representations are crucial for enhancing the model's ability to predict drug synergy. The variants MultiSyn-m and MultiSyn-g, which alter the source of cell line features, achieved similar results across most metrics. This suggests that cell line features derived from different sources can represent key characteristics of cell lines to some extent. However, both MultiSyn-m and MultiSyn-g showed decreased performance compared to MultiSyn across all metrics, further demonstrating that the single-source features in these variants are complementary. Their integration is essential for capturing more comprehensive and potentially synergistic interactions between drugs. By embedding genes within their functional and relational contexts, the integration of PPI networks offers a systems-level perspective that more comprehensively characterizes cell states and gene functions. Overall, the design of all modules in our model architecture focuses on capturing diverse and

representative key features from multiple sources, significantly enhancing the accuracy and efficiency of synergistic drug combination prediction.

Evaluation on independent dataset

To further assess the generalization capability of the proposed MultiSyn model, we conducted experiments on the independent AstraZeneca dataset [38], a widely recognized benchmark in drug combination prediction. This dataset comprises 668 unique drug pair–cell line combinations, involving 52 drugs and 24 cancer cell lines. Importantly, it differs substantially from the O'Neil dataset used for training, both in drug composition and cell line distribution. Some of the independent test results were directly cited from the original DeepDDS paper, while others were reproduced under the same data splits and preprocessing settings as DeepDDS, using official implementations and recommended hyperparameters provided by the respective authors. We performed five independent trials for our model, following the same protocol, to account for performance variance and support a robust evaluation of generalization.

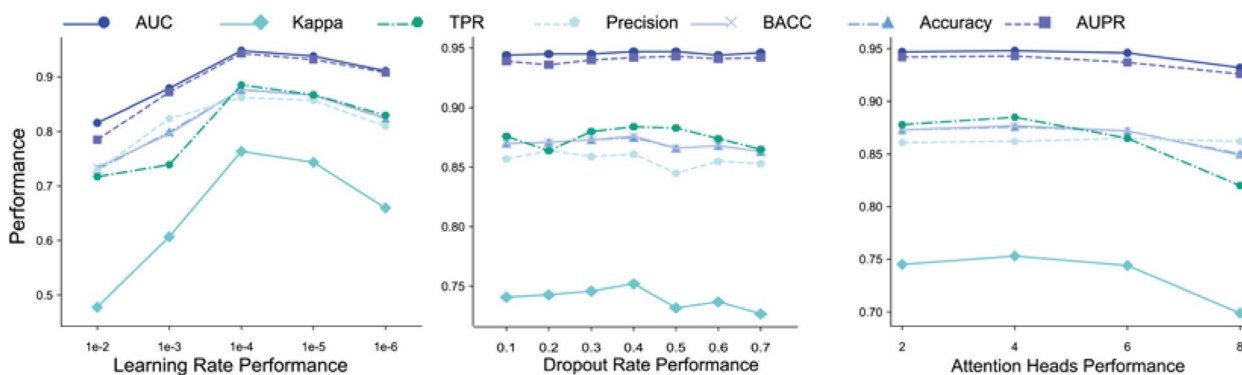


Fig. 3 Performance evaluation under different hyperparameter settings, each data point represents the average of 7 independent runs ($n = 7$)

As shown in Table 4, MultiSyn achieved the best overall performance across the majority of key metrics compared to both classical ML and DL-based methods. Compared to classical models such as XGBoost and GBM, MultiSyn achieved more than a 25% improvement in AUROC and over a 50% increase in TPR. Furthermore, relative to the best-performing DL baseline model, DeepDDS, MultiSyn not only achieved a higher AUROC but also outperformed it in AUPR, PREC, and TPR, demonstrating superior capability in addressing class imbalance and enhancing sensitivity to positive cases. Overall, these results demonstrate that MultiSyn not only fits the training data well but also generalizes effectively to novel drug–cell line combinations, highlighting its practical utility in drug discovery applications.

Parameter sensitivity analysis

To evaluate the impact of model parameters on the performance of the drug synergy prediction model, we conducted a parameter sensitivity analysis. Fig. 3 illustrates the model’s performance under various parameter configurations. We tested five different learning rates: $1e-2$, $1e-3$, $1e-4$, $1e-5$, and $1e-6$. The experimental results show that when the learning rate is set to $1e-4$, the performance of MultiSyn is optimal, with an AUROC of 0.95. Other metrics also significantly improve compared to the remaining learning rates. Performance fluctuates at other learning rates, especially when the learning rate is set to $1e-2$ or $1e-6$, where a notable decrease in model performance is observed. These results indicate that a moderate learning rate promotes stable convergence, while both excessively low and high learning rates can hinder effective model training. We also experimented with seven different dropout rates ranging from 0.1 to 0.7, in intervals of 0.1. The results revealed that a dropout rate of 0.3 achieved the best model performance. As the dropout rate increased beyond 0.3, particularly at 0.5 and 0.7, a noticeable decline in performance was observed.

These findings suggest that a moderate dropout rate helps mitigate overfitting and improves the model’s ability to generalize. In addition, we evaluated the impact of varying the number of attention heads in message passing, testing values of 2, 4, 6, and 8. The optimal performance was achieved with 4 attention heads, where the model reached an AUROC of 0.948, AUPR of 0.943, and accuracy of 0.876. Increasing the number of attention heads to 8 resulted in a slight decrease in performance, indicating that 4 attention heads are sufficient to capture the relevant features. Adding more heads beyond this point may increase computational overhead without offering significant performance gains. More parameter analysis experimental results are shown in Additional file 1: Fig. S2.

Representation visualization

To evaluate the representational capabilities of the model, we performed dimensionality reduction visualization on the feature representations before and after training with the MultiSyn model, selecting DeepDDS model for comparison. Specifically, we applied t-SNE (t-distributed stochastic neighbor embedding) to visualize the high-dimensional feature data of the A2780 and ZR751 cell lines into a two-dimensional space. As shown in Fig. 4, we compare the feature spaces learned by the proposed MultiSyn model and the baseline DeepDDS model, highlighting the distinction between positive (synergistic) and negative (non-synergistic) drug pairs. Before training, as illustrated in the left panels, both models exhibited significant overlap between the positive (red) and negative (blue) samples. This suggests that the boundary between the two classes is difficult to distinguish in the initial feature space, reflecting the inability of raw features to effectively differentiate between synergistic and non-synergistic drug combinations. After training, the t-SNE visualization of the MultiSyn model reveals tighter clustering within each category and clearer separation

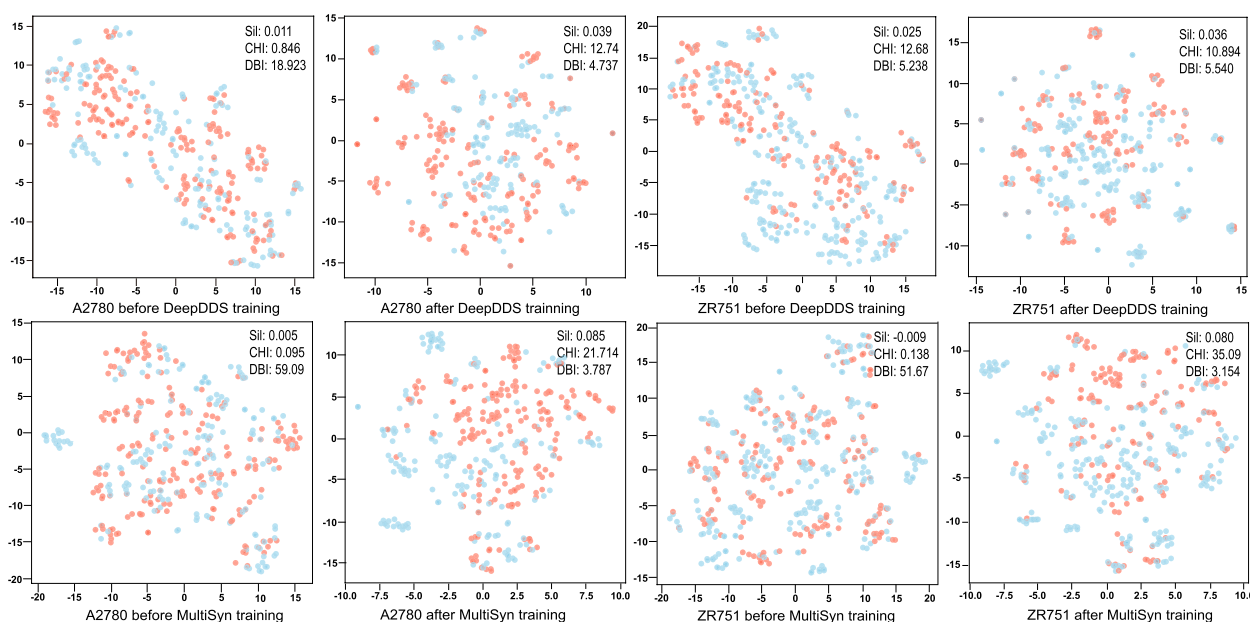


Fig. 4 t-SNE visualization of feature representations before and after training for A2780 and ZR751 cell lines. Red points indicate synergistic (positive) drug combinations, while blue points indicate non-synergistic (negative) combinations

between positive and negative samples for both A2780 and ZR751 cell lines.

To further clarify this point, we computed three widely used clustering metrics—Silhouette Score (Sil), Calinski–Harabasz Index (CHI), and Davies–Bouldin Index (DBI) on the t-SNE–projected embeddings of MultiSyn and DeepDDS to quantitatively assess the quality of the learned representations. We applied t-SNE with identical parameters and fixed random seeds to project all feature embeddings into a shared two-dimensional space, ensuring a consistent basis for evaluation. Under this setting, we focused on comparing the magnitude of performance improvement from pre-training to post-training within both models, as well as the clustering indices after training. After training, MultiSyn exhibited greater improvements across all three metrics. For example, on the A2780 cell line, MultiSyn’s Sil increased 15-fold, from 0.005 to 0.085; its CHI rose from 0.095 to 21.714; and its DBI decreased by over 90%, from 59.09 to 3.787, indicating substantially enhanced cluster compactness and separability. In comparison, DeepDDS exhibited a more modest 3.5-fold increase in Sil, from 0.011 to 0.039, and a reduction in DBI, from 18.923 to 4.737. In comparison, quantitative assessments revealed that while the baseline DeepDDS model also showed improvements in clustering and separation between classes after training, the MultiSyn model exhibited more compact intra-class clusters and sharper inter-class boundaries, indicating superior performance. This suggests that the MultiSyn

model excels in learning feature representations that capture subtle interactions driving drug synergy, particularly in challenging scenarios where the initial feature space exhibits substantial overlap. Collectively, these results demonstrate the effectiveness of the proposed MultiSyn model in learning features that capture latent relationships underlying drug synergy.

The effectiveness of identifying key substructures

The MultiSyn model incorporates pharmacophore-level substructural information, which aids in identifying key substructures. To validate this, we selected the A2780 and ZR751 cell lines and visualized four drug combinations known to exhibit synergy, with synergy scores of 1 predicted by MultiSyn. During the drug feature extraction process, we employed a heterogeneous graph neural network based on the Transformer architecture to obtain rankings of fragment and atomic node attention scores within the graph, and annotated the top-ranked fragment and the top 30% of nodes. As shown in the Fig. 5, many atomic nodes in the top 30% belong to the top 1 pharmacophore group. However, relying solely on the interpretability of top atoms is not effective in identifying specific and complete pharmacophore structures. In contrast, MultiSyn, which incorporates pharmacophore-level substructural information, can effectively identify key and complete pharmacophore information, utilizing these critical features for accurate synergy prediction. For example, the substructures of GEMCITABINE that the

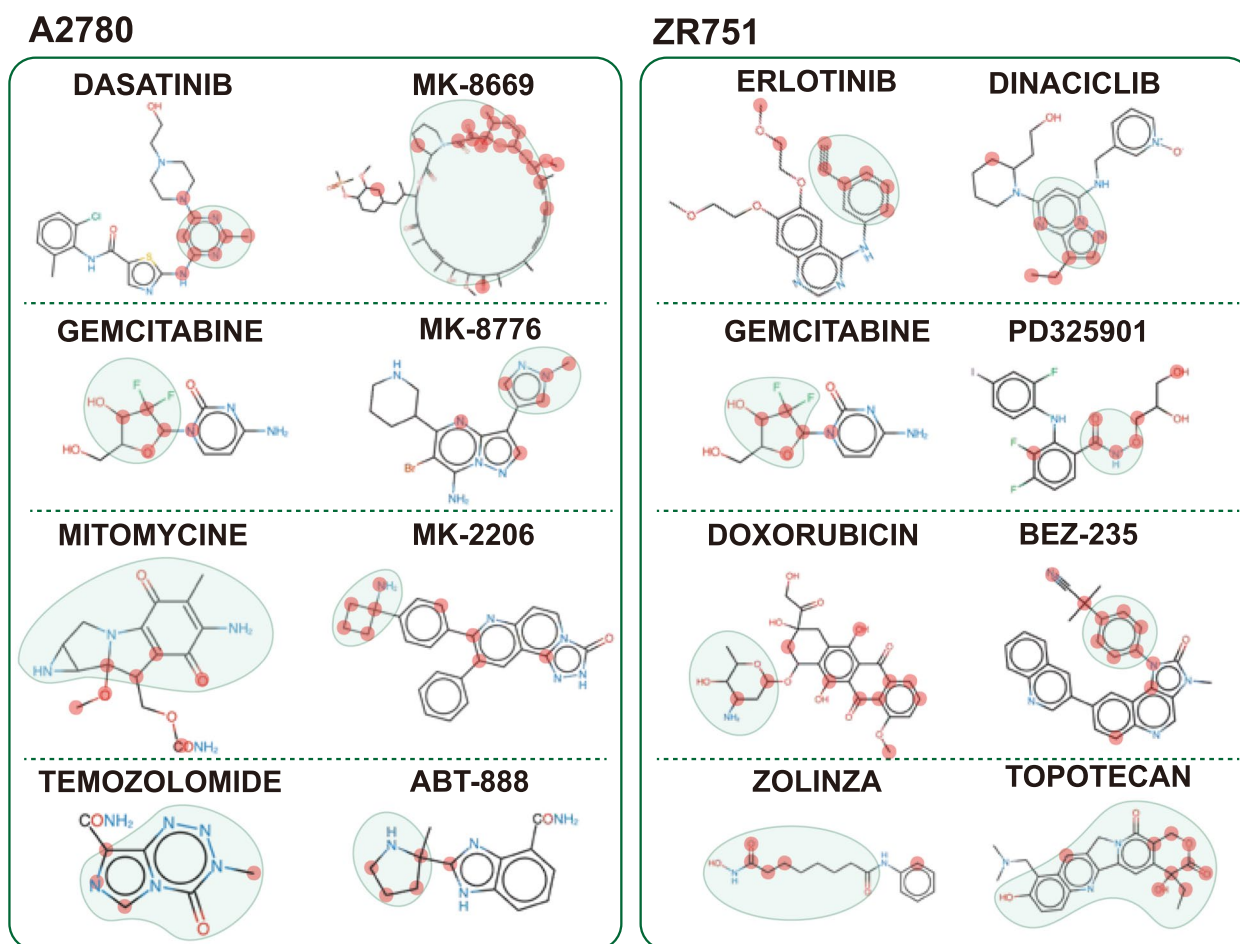


Fig. 5 Visualization of key drug substructure features and pharmacophoric level information in A2780 and ZR751 cell lines using the MultiSyn model

model focuses on include deoxyribose and the fluorine group. The sugar backbone is essential for its binding to DNA, where it disrupts DNA synthesis and induces cell death. The fluorine substitution in the sugar structure helps GEMCITABINE resist enzymatic degradation and prolongs its activity within the cell, making it more effective in inhibiting DNA synthesis. The nitrogen-containing heterocycles in MK-8776 can interact with various targets, such as kinases, and may be involved in inhibiting checkpoint kinases (e.g., CHK1). This makes tumor cells more sensitive to DNA-damaging agents like gemcitabine, and the bromine atom (Br) may alter the drug's binding properties, potentially enhancing its interaction with target proteins or DNA. Overall, the features of key substructures contribute to more accurate predictions of drug synergy, and the identification of key substructures containing pharmacophores is also crucial for the development of subsequent drugs and the discovery of lead compounds.

Prediction of novel drug combinations

In further experiments, we utilized the MultiSyn model to predict A2780 and ZR751 cell lines' potential novel drug combinations. To identify drug pairs that had not been experimentally validated, we systematically paired all drugs in the dataset while excluding combinations already present in the original data. This ensured that our predictions focused solely on previously untested drug combinations. Based on the synergy prediction scores generated by the model, we identified the top 10 drug combinations with the highest synergy potential for the A2780 and ZR751 cell lines. To further assess the validity of these predictions, we conducted a non-exhaustive literature search to explore existing experimental evidence supporting these combinations. Table 5 lists the predicted drug combinations for the A2780 and ZR751 cell lines, along with relevant references where they have been validated.

We found that among the top 10 predicted drug combinations for the A2780 and ZR751 cell lines, 6 and 5

Table 5 Predicted novel synergistic combinations in the A2780 and ZR751 cancer cell lines

Top	Cell Line: A2780			Cell Line: ZR751		
	Drug1	Drug2	PMID	Drug1	Drug2	PMID
1	5-FU	ABT-888	29770165	ERLOTINIB	LAPATINIB	22431920
2	ZOLINZA	MK-2206	28393191	ETOPOSIDE	GEMCITABINE	NA
3	DOXORUBICIN	MK-2206	26698230,20571069	DEXAMETHASONE	TOPOTECAN	NA
4	BEZ-235	TOPOTECAN	NA	DOXORUBICIN	ETOPOSIDE	1984826,8435802
5	MRK-003	SUNITINIB	NA	ETOPOSIDE	TOPOTECAN	31383812
6	5-FU	BEZ-235	29180117,27777878	ETOPOSIDE	MK-8669	NA
7	ABT-888	MK-5108	24362082	DEXAMETHASONE	ETOPOSIDE	NA
8	DINACICLIB	MK-2206	27663592,32311593	GEMCITABINE	MK-2206	37521473
9	DEXAMETHASONE	MK-8669	NA	CYCLOPHOSPHAMIDE	TOPOTECAN	NA
10	L778123	MRK-003	NA	DOXORUBICIN	GEMCITABINE	12947059,10893285

pairs, respectively, were consistent with previous studies or clinical trial observations. This includes the confirmed synergistic effect between MK-2206 and DOXORUBICIN in ovarian cancer related to the A2780 cell line [39, 40], where research demonstrated that MK-2206 enhanced growth inhibition induced by chemotherapy drugs like DOXORUBICIN [41]. In the case of the ZR751 cell line related to breast cancer, studies showed that the combination of DOXORUBICIN and GEMCITABINE was highly effective in treating metastatic breast cancer with good tolerance [42, 43]. When GEMCITABINE was combined with DEXAMETHASONE, DEXAMETHASONE enhanced the antitumor effect of GEMCITABINE through the glucocorticoid receptor signaling pathway [44]. Additionally, another study [45] found that this combination exhibited synergistic effects in breast cancer, and in xenograft models of both ovarian and breast cancer, the combination of DEXAMETHASONE and GEMCITABINE effectively suppressed tumor growth [46]. In addition to literature support, we found that two pairs of the predicted combinations are mentioned in ClinicalTrials.gov, ERLOTINIB with LAPATINIB (NCT04591431) and DOXORUBICIN with GEMCITABINE (NCT00128856, NCT00191789). This further corroborates the potential clinical relevance of the drug combinations predicted by our method.

Discussion

In this study, we present MultiSyn, a novel framework that enhances the accuracy of synergistic drug combination prediction by integrating biological networks, multi-omics data, and chemical structural information associated with specific biological functions. Importantly, this work addresses two critical limitations in most existing DL models for drug synergy prediction—the lack of biological network context in modeling cell lines and

the underutilization of pharmacophoric features in drug representation. Our framework captures comprehensive cell line features by incorporating PPI networks and multi-omics data. Additionally, it constructs heterogeneous molecular graphs incorporating pharmacophore information through molecular decomposition, enabling a more functionally meaningful representation of drug structural information. Extensive comparative experiments on benchmark datasets and specialized scenarios demonstrate that MultiSyn provides more accurate and robust drug combination predictions. Further experimental analyses demonstrate the capability of MultiSyn to identify key drug substructures essential for synergy, offering new opportunities for medicinal chemists in rational drug design. Additionally, the precise identification of novel synergistic drug combinations aids clinical professionals in obtaining more reliable candidate combination therapies.

Although evaluations across diverse datasets and cold-start scenarios have demonstrated the feasibility of MultiSyn for accurate prediction of synergistic drug combinations, the method still has certain limitations. First, the scale and diversity of the datasets used are limited, which may introduce sampling bias and restrict the model's ability to generalize across broader chemical and biological spaces. The relatively small sample size also raises concerns regarding overfitting. Although we evaluated MultiSyn on an independent dataset, the performance improvements over existing DL models remain modest, suggesting that its generalizability across broader domains is still limited. Second, the focus on oncology-specific cell lines and treatment contexts may constrain the model's applicability to other therapeutic domains, such as neurology or infectious diseases, which involve distinct biological mechanisms and regulatory networks. Finally, the current evaluation does not include prospective clinical validation, which will be

essential for confirming the robustness and translational relevance of MultiSyn in real-world applications.

In future work, we plan to enrich the input space by incorporating more diverse and fine-grained data modalities—such as proteomics profiles, pathway activity scores, and drug–target interaction networks—to better capture the biological complexity underlying drug synergy. Furthermore, we aim to enhance the chemical and cellular diversity of our datasets by applying MultiSyn to larger-scale drug combination resources that encompass a wider range of compounds and cancer types. Finally, we will explore architectural simplifications and training strategies to improve computational efficiency while preserving the representational capacity of our heterogeneous graph encoder.

Conclusions

In summary, although MultiSyn still has room for improvement, the experimental results confirm its effectiveness. The proposed method further highlights the potential of multi-source information integration and underscores the importance of incorporating chemical structure features associated with specific biological functions for drug synergy prediction.

Methods

Pipeline of MultiSyn

Figure 1 illustrates the multi-source information fusion DL framework, MultiSyn, for predicting synergistic drug combinations. Given a pair of drug combinations and a cell line, we perform drug synergy prediction by extracting multi-source features of the cell line, which integrate biological network and omics data, alongside molecular features containing pharmacophore structural information of the drugs. Specifically, MultiSyn does not separately learn the biological network and omics data of the cell line but instead employs a semi-supervised learning method using an attributed GAT to integrate PPI networks and multi-omics data, thereby obtaining the initial multi-source features of the cell line. Furthermore, we combine the multi-source cell line features with cell line-related gene expression features to obtain the final cell line features. To extract drug structural features, we represent drugs from three perspectives: atom-level, fragment-level, and atom-fragment level, and learn features that carry pharmacophore information from these perspectives using a heterogeneous graph representation method, followed by multi-view feature fusion through attention mechanisms. Finally, we connect cell line features with drug pair features and utilize a predictor to make the final drug synergy prediction.

Cell line feature extraction

To effectively integrate cell line-related biological networks and omics data, we reference the previous work PRODeepSyn [29]. We first construct a semi-supervised learning model using an attributed graph neural network to merge PPI network information with cell line-related multi-omics data, obtaining the initial cell line features that incorporate multi-source information.

We first represent the PPI network of gene expression products as a graph $G^p = (V^p, E^p)$, where V^p denotes the set of nodes (genes) and E^p the set of edges (interactions between genes). The feature matrix for the gene nodes in the network is denoted as $\mathbf{X} \in \mathbb{R}^{N_p \times F_p}$, where N_p is the number of gene nodes in the PPI network and F_p is the feature dimension for the gene. \mathbf{X} represents the original gene features on the PPI graph, each row \mathbf{x}_i in \mathbf{X} represents the feature vector of the i th gene node. Using this setup, we employ an attribute GAT encoder to learn the embeddings of gene features \mathbf{X}' in the PPI network. These embeddings are then mapped to the feature space of the cell line through a fully connected layer, resulting in a new gene feature matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{N_p \times F_c}$.

In the attribute GAT encoder, attention is assigned to neighboring node embeddings based on the attribute characteristics of each node. The update rule for the node representations in the graph can be formalized as:

$$\mathbf{X}' = \text{Attribute GAT}(\mathbf{X}, \mathbf{P}) = \sigma \left(\sum_{u \in N(v)} \alpha_{vu} W^l \mathbf{h}_u^l \right) \quad (1)$$

where $N(v)$ represents the set of neighbors for node v , and $u \in N(v)$ denotes a neighboring node. σ is the activation function, W^l is a learnable weight matrix for the l th layer. \mathbf{h}_u^l represents the input feature of node u at the l th layer. The attention weight α_{vu} captures the influence of node u on node v and is computed based on the features \mathbf{h}_v^l and \mathbf{h}_u^l , as well as the adjacency matrix \mathbf{P} , where $P_{ij} = 1$ if gene i interacts with gene j . The output embedding $\mathbf{X}' \in \mathbb{R}^{N_p \times F_p}$ contains the updated node features for all nodes in the graph. The embeddings \mathbf{X}' are projected into the cell line feature space via a fully connected transformation, yielding the matrix $\tilde{\mathbf{X}}$. This projection ensures that the transformed gene features $\tilde{\mathbf{X}}$ and the cell line embedding \mathbf{c}_i share the same feature dimension, enabling the matrix multiplication.

Next, we use the cell line ID to represent one-hot encoding, and the feature matrix $\mathbf{C} \in \mathbb{R}^{N_c \times F_c}$ is randomly initialized, where N_c denotes the number of cell lines and F_c is the feature dimension for each cell line. Each row $\mathbf{c}_i \in \mathbb{R}^{F_c}$ corresponds to the embedding of the i th cell line. For the i th cell line, the relationship between gene features and the cell line status can be expressed as the product of the cell

line feature $\mathbf{c}_i \in \mathbf{C}$ and the gene feature matrix $\tilde{\mathbf{X}}$ derived from the PPI network:

$$\hat{\mathbf{g}}_i = \tilde{\mathbf{X}} \cdot \mathbf{c}_i \quad (2)$$

where $\tilde{\mathbf{X}} \in \mathbb{R}^{N_p \times F_c}$ is the transformed gene feature matrix, and the resulting $\hat{\mathbf{g}}_i \in \mathbb{R}^{N_p}$ approximates the molecular profile of cell line i across N_p genes.

The relationship between cell line and gene status is then optimized using the gene expression levels \mathbf{g}^e and gene mutation results \mathbf{g}^m corresponding to the cell line, updating both \mathbf{C} and $\tilde{\mathbf{X}}$. The loss function is defined as:

$$L_x = \sum_i^{N_c} \sum_k^{N_x} (g_{k,i}^x - \hat{g}_{k,i}^x)^2 \quad (3)$$

where N_x denotes the number of genes associated with omics modality $x \in \{e, m\}$, N_e and N_m represent the number of genes with valid expression and mutation data, respectively. These gene sets are typically subsets of the full PPI graph nodes V^p . $g_{k,i}^x$ is the ground-truth omics expression or mutation value of gene k in cell line i , and $\hat{g}_{k,i}^x$ is the corresponding predicted value obtained from the model. The loss L_x is used to optimize the parameters of both the gene transformation matrix $\tilde{\mathbf{X}}$ and the cell line embedding matrix \mathbf{C} via backpropagation.

Finally, the embedded features updated according to \mathbf{g}^e and \mathbf{g}^m are concatenated column by column to form the feature matrix \mathbf{C} , which contains multi-omics data for the cell lines, serves as the initial representation of cell lines in this module. To further enrich the cell line feature representations, we incorporate gene expression data from the CCLE dataset [26]. The gene expression data is pre-processed into a feature matrix, which is then passed through a multi-layer perceptron (MLP) to obtain \mathbf{C}_g . The final cell line feature representation is obtained by integrating the embeddings \mathbf{C}_g and \mathbf{C}_m through an attention mechanism, as expressed below:

$$\mathbf{C} = \text{Attention}(\mathbf{C}_g, \mathbf{C}_m) \quad (4)$$

Drug feature extraction

In this section, we represent drug molecules as multi-view molecular graphs containing structural information about drug pharmacophore and use the Graph Transformer framework for drug feature learning. We represent drug molecules as heterogeneous molecular graphs, focusing on the fragments containing pharmacophores. We first use the cheminformatics tool RDKit to build a molecular graph: $G = (V_x, E_x)$ by the drug's SMILES sequence, where V_x represents the set of atomic nodes in the molecule, and E_x represents the set of chemical

bonds in the molecule. Each atom v_i corresponds to an atom in the molecule, and each bond $(v_i, v_j) \in E_x$ indicates a chemical interaction between two atoms. Next, we use BRICS rules [47] to partition the drug molecular SMILES sequence into multiple fragments, each fragment is then further represented by graph G_i , where $i = 1, 2, \dots, N_f$, and N_f represents the total number of fragments obtained from the SMILES partition. Furthermore, we regard the fragments obtained by segmentation as new node types (fragments containing pharmacophores) as \mathcal{V}_y . At the same time, we define the interaction information between fragments as \mathcal{E}_y and the mapping relationship between fragments and atoms as \mathcal{E}_z , which links atoms in \mathcal{V}_x to their parent fragments in \mathcal{V}_y . These newly defined node and edge types allow us to construct a heterogeneous molecular graph based on the original molecular graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{V}_x \cup \mathcal{V}_y$ and $\mathcal{E} = \mathcal{E}_x \cup \mathcal{E}_y \cup \mathcal{E}_z$.

In this heterogeneous graph, the feature matrix of each node is represented as: $\mathbf{X}^0 \in \mathbb{R}^{N_{\mathcal{V}} \times F_{\mathcal{V}}}$, where $N_{\mathcal{V}}$ denotes the number of nodes \mathcal{V} , $F_{\mathcal{V}}$ represents the node feature dimension. The edge feature is expressed as: $\mathbf{E}^0 \in \mathbb{R}^{N_{\mathcal{E}} \times F_{\mathcal{E}}}$, where $N_{\mathcal{E}}$ is the number of edges in the heterogeneous graph and $F_{\mathcal{E}}$ is the dimension of each edge feature vector. Given the node feature matrix $\mathbf{X}^{(0)}$ and the edge feature matrix \mathbf{E} , we define the initial edge feature $\mathbf{E}_{ij}^{(0)}$ between node v_i and its neighbor v_j as \mathbf{X}_{v_i} , the node feature of v_i . To enhance the representation capability and stabilize training, we employ multi-head attention for node and edge feature updates as follows:

$$\tilde{\mathbf{X}}^l = \text{Concat} \left(\left[\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^h (\mathbf{X}_j^{l-1} \mathbf{W}^{V,h}) \right]_{h=1}^{N_h} \right) \quad (5)$$

$$\tilde{\mathbf{E}}_{ij}^l = \text{Linear}(\mathbf{E}_{ij}^{l-1} \oplus \mathbf{X}_i^{l-1} \oplus \mathbf{X}_j^{l-1}) \quad (6)$$

where N_h denotes the number of attention heads, and h indexes each individual attention head, \oplus represents vector concatenation, combining edge features and node features at both ends, and a linear layer is used to integrate information, \mathbf{V}_j^h represent of node v_j in the h -th attention head, α_{ij}^h is the attention coefficient between nodes v_i and v_j in the h th attention head. The attention coefficient is computed by aggregating information from all nodes:

$$\alpha_{ij}^l = \text{Softmax} \left(\frac{\mathbf{X}_i^{l-1} \mathbf{W}^Q (\mathbf{X}_j^{l-1} \mathbf{W}^K)^\top}{\sqrt{d_k}} \right) \quad (7)$$

where \mathbf{X}_i^{l-1} and \mathbf{X}_j^{l-1} are the hidden states of node i and node j at the $(l-1)$ -th layer, serving as inputs for the query (Q) and key (K). $\mathbf{W}^Q, \mathbf{W}^K$ are learnable projection

matrices, and d_k is the key dimension used for scaling. The output $\alpha_{ij}^l \in [0, 1]$ reflects the normalized attention weight of node j with respect to node i .

Finally, a Gated Recurrent Unit (GRU) is applied to aggregate node embeddings into a global molecular representation, atom nodes and fragment nodes are represented as feature matrices with shapes $[N_a, F_{dim}]$ and $[N_f, F_{dim}]$, here N_a and N_f denote the number of atom nodes and fragment nodes. In the heterogeneous graph representation of molecules, and F_{dim} is the feature dimension. Through the multi-head attention mechanism, the relationships between atom nodes and fragment nodes are dynamically calculated, and the fragment node features are weighted and fused to generate an updated atom node feature matrix. Meanwhile, the residual connection preserves the original atom node features, resulting in the fused feature matrix:

$$\mathbf{X}_{fusion} = \text{Multi-Attention}(\tilde{\mathbf{X}}_a, \tilde{\mathbf{X}}_p) + \tilde{\mathbf{X}}_a \quad (8)$$

where $\tilde{\mathbf{X}}_a$ and $\tilde{\mathbf{X}}_p$ represent the features of atom nodes and fragment nodes; the recursive update operation of the GRU is defined as:

$$\mathbf{X} = \text{GRU}(\mathbf{X}_{fusion}, \mathbf{X}_{init}) \quad (9)$$

where \mathbf{X}_{fusion} represents the fused node features obtained from the multi-head attention mechanism, and \mathbf{X}_{init} denotes the initial hidden states used to initialize the GRU. The GRU integrates both inputs to generate the final node representations \mathbf{X} . The global molecular representation is generated by applying mean pooling to all node features:

$$\mathbf{D}_{global} = \frac{1}{N_a} \sum_{i=1}^{N_a} \mathbf{X}_i \quad (10)$$

where \mathbf{X}_i represents the characteristics of node i . The resulting $\mathbf{D}_{global} \in \mathbb{R}^{F_{dim}}$ encodes the global structural feature of a single molecule and serves as its final representation.

Prediction of synergistic medication

In the prediction module, the features of the two drugs and the cell line are first integrated, followed by a multi-layer perceptron (MLP) to predict the drug synergy results. Concatenate the refined cell line embeddings \mathbf{C} with the drug pair features \mathbf{D}_A and \mathbf{D}_B obtained from the drug feature module. Finally, the classification task is performed using an MLP, as described by the following equations:

$$p_i = \text{softmax}(\text{MLP}(\mathbf{D}_A \parallel \mathbf{D}_B \parallel \mathbf{C})) \quad (11)$$

where \parallel denotes the concatenation operator and p_i denotes the predicted probability that the drug pair exhibits synergy. The model is trained by minimizing the cross-entropy loss function:

$$\text{CrossEntropyLoss} = \frac{1}{N} \sum_{i=1}^N -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (12)$$

where N represents the total number of samples in the training set, y_i is the label of the i th sample.

Abbreviations

HTS	High-throughput screening
ML	Machine learning
DL	Deep learning
GAT	Graph attention network
TCGA	The Cancer Genome Atlas
GNNs	Graph neural networks
GCNs	Graph convolutional networks
PPI	Protein-protein interaction
MLP	Multi-layer perceptron
GRU	Gated recurrent unit
CCLC	Cancer Cell Line Encyclopedia
CV	Cross-validation
AUROC	Area under the receiver operating characteristic curve
AUPR	Area under the precision-recall curve
ACC	Accuracy
BACC	Balanced accuracy
PREC	Precision
TPR	True positive rate
KAPPA	Cohen's kappa
GBM	Gradient boosting machine
RF	Random forest
t-SNE	t-Distributed stochastic neighbor embedding

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-025-02302-y>.

Additional file 1. Contains extended methodological details, evaluation analyses, and experimental results referenced in the main manuscript: Evaluation Indicators and Calculation Formulas: Provides definitions and formulas for key performance metrics including ACC, BACC, TPR, TNR, PREC, and KAPPA. Implementation Details: Describes model architectures and settings for MultiSyn and baseline methods, ensuring reproducibility. Normality Assessment for AUROC Differences: Includes Fig. S1, a quantile-quantile plot used to assess the validity of statistical tests for AUROC comparison. Parameter Analysis: Includes Fig. S2, which visualizes sensitivity to hyperparameters such as hidden dimensions, layer depth, and activation functions. Computational Complexity: Presents the theoretical analysis of MultiSyn's computational complexity across its modules. Additional file 1: Figures S1, S2. Fig. S1- [QQ plots], Fig. S2 -[Hyperparameter Settings].

Additional file 2. Shows the source data of MultiSyn and 14 baseline models under three cold-start validation scenarios.

Acknowledgements

Not applicable.

Authors' contributions

S.J. conceived and designed the study. H.L., A.H., and J.X. contributed to methodology development. H.L. and J.X. performed validation. H.L. drafted the original manuscript, while Z.X. and S.J. revised and edited it. S.J., J.X., and Z.X. supervised the study. J.W. carried out the investigation, and X.Y. curated the data. S.J. and J.X. oversaw project administration.

Funding

This work was partly supported by the National Natural Science Foundation of China (Grant Nos. 62402351 and 62302156); the Hubei Provincial Natural Science Foundation of China (Grant Nos. 2024AFB275 and 2024AFB307); the Scientific Research Project of Education Department of Hubei Province (Grant No. Q20231109).

Data availability

The full source code and datasets used in this study are available at the following GitHub repository: <https://github.com/HuazeLoong/MultiSyn>. A permanent archive of the same data and code has also been deposited in Zenodo: <https://doi.org/10.5281/zenodo.15194129>. Gene expression profiles from the CCLE: <https://portals.broadinstitute.org/ccle>. Drug molecular SMILES from DrugBank: <https://go.drugbank.com>. All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China. ²Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430065, Hubei, China. ³Department of Integrative Biotechnology, Yonsei University, Incheon 21983, Republic of Korea. ⁴Department of Radiology, Henan Provincial People's Hospital and the People's Hospital of Zhengzhou University, Zhengzhou 450000, Henan, China.

Received: 5 March 2025 Accepted: 19 June 2025

Published online: 03 July 2025

References

- Zhao S, Nishimura T, Chen Y, Azeloglu EU, Gottesman O, Giannarelli C, et al. Systems pharmacology of adverse event mitigation by drug combinations. *Sci Transl Med*. 2013;5(206):140206140.
- Hill JA, Ammar R, Torti D, Nislow C, Cowen LE. Genetic and genomic architecture of the evolution of resistance to antifungal drug combinations. *PLoS Genet*. 2013;9(4):1003390.
- Zhao XM, Iskar M, Zeller G, Kuhn M, Van Noort V, Bork P. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol*. 2011;7(12):1002323.
- Yin N, Ma W, Pei J, Ouyang Q, Tang C, Lai L. Synergistic and antagonistic drug combinations depend on network topology. *PLoS ONE*. 2014;9(4):93960.
- Iwata H, Sawada R, Mizutani S, Kotera M, Yamanishi Y. Large-scale prediction of beneficial drug combinations using drug efficacy and target profiles. *J Chem Inf Model*. 2015;55(12):2705–16.
- Li X, Xu Y, Cui H, Huang T, Wang D, Lian B, et al. Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif Intell Med*. 2017;83:35–43.
- Preuer K, Lewis RP, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*. 2018;34(9):1538–46.
- Zhang H, Feng J, Zeng A, Payne P, Li F. Predicting tumor cell response to synergistic drug combinations using a novel simplified deep learning model. In: *AMIA Annu Symp Proc*. vol. 2020. p. 1364. American Medical Informatics Association, Bethesda, MD, USA. 2021.
- Zhang T, Zhang L, Payne PRO, Li F. Synergistic Drug Combination Prediction by Integrating Multiomics Data in Deep Learning Models. In: Markowitz J, editor. *Translational Bioinformatics for Therapeutic Development*. Methods in Molecular Biology, vol. 2194. New York; 2020. pp. 223–238.
- Hosseini SR, Zhou X. CCSynergy: an integrative deep-learning framework enabling context-aware prediction of anti-cancer drug synergy. *Brief Bioinform*. 2023;24(1):bbac588.
- Mohammadzadeh-Vardin T, Ghareyazi A, Gharizadeh A, Abbasi K, Rabiee HR. DeepDRA: Drug repurposing using multi-omics data integration with autoencoders. *PLoS ONE*. 2024;19(7):0307649.
- Gharizadeh A, Abbasi K, Ghareyazi A, Mofrad MR, Rabiee HR. HGTD: Advancing drug repurposing with heterogeneous graph transformers. *Bioinformatics*. 2024;40(7):btae349.
- Giordano TJ. The cancer genome atlas research network: a sight to behold. *Endocr Pathol*. 2014;25:362–5.
- Jin S, Zeng X, Xia F, Huang W, Liu X. Application of deep learning methods in biological networks. *Brief Bioinform*. 2021;22(2):1902–17.
- Jiang Y, Jin S, Jin X, Xiao X, Wu W, Liu X, et al. Pharmacophoric-constrained heterogeneous graph transformer model for molecular property prediction. *Commun Chem*. 2023;6(1):60.
- Liu Q, Xie L. Transynergy: Mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput Biol*. 2021;17:1008653.
- Hu J, Gao J, Fang X, Liu Z, Wang F, Huang W, et al. Dtsyn: a dual-transformer-based neural network to predict synergistic drug combinations. *Brief Bioinform*. 2022;23:bbac302.
- Wang J, Liu X, Shen S, Deng L, Liu. Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform*. 2022;23(1):bbab390–bbab390.
- Chen Z, Li Z, Shen X, Liu Y, Lin X, Zeng D, et al. Dualsyn: A dual-level feature interaction method to predict synergistic drug combinations. *Expert Syst Appl*. 2024;257:125065.
- Alam W, Tayara H, Chong KT. Unlocking the therapeutic potential of drug combinations through synergy prediction using graph transformer networks. *Comput Biol Med*. 2024;170:108007.
- Liu Y, Zhang P, Che C, Wei Z. Sddsynergy: Learning important molecular substructures for explainable anticancer drug synergy prediction. *J Chem Inf Model*. 2024;64(24):9551–62.
- Lin J, Xu H, Woicik A, Ma J, Wang S. Pisces: A cross-modal contrastive learning approach to synergistic drug combination prediction. *bioRxiv*; 2022. p. 2022–11.
- Pang Y, Chen Y, Lin M, Zhang Y, Zhang J, Wang L. MMSyn: A New Multimodal Deep Learning Framework for Enhanced Prediction of Synergistic Drug Combinations. *J Chem Inf Model*. 2024;64(9):3689–705.
- Rafiei F, Zeraati H, Abbasi K, Ghasemi JB, Parsaeian M, Masoudi-Nejad A. DeepTraSynergy: drug combinations using multimodal deep learning with transformers. *Bioinformatics*. 2023;39(8):438.
- O'Neil J, Benita Y, Feldman I, Chenard M, Roberts B, Liu Y, et al. An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther*. 2016;15(6):1155–62.
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim SY, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603–7. <https://doi.org/10.1038/nature11003>.
- Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci*. 1988;28(1):31–36.
- Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):1074–82. <https://doi.org/10.1093/nar/gkx1037>.
- Wang X, Zhu H, Jiang Y, Li Y, Tang C, Chen X, Li Y, Liu Q, Liu Q. PRODeepSyn: predicting anticancer synergistic drug combinations by embedding cell lines with protein–protein interaction network. *Briefings in bioinformatics*. 2022;23(2):bbab587–bbab587.
- Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*. 2016;166(3):740–54.
- Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019;47(D1):D941–7.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13.

33. Yang J, Xu Z, Wu WKK, Chu Q, Zhang Q. Graphsynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J Am Med Inform Assoc*. 2021;28:2336–45.
34. Xu M, Zhao X, Wang J, Feng W, Wen N, Wang C, et al. DFFNDDS: prediction of synergistic drug combinations with dual feature fusion networks. *J Cheminform*. 2023;15(1):33.
35. Wang T, Wang R, Wei L. AttenSyn: an attention-based deep graph neural network for anticancer synergistic drug combination prediction. *J Chem Inf Model*. 2023;64(7):2854–62.
36. Dong Y, Chang Y, Wang Y, Han Q, Wen X, Yang Z, et al. MFSynDCP: multi-source feature collaborative interactive learning for drug combination synergy prediction. *BMC Bioinformatics*. 2024;25:140.
37. Zhang Z, Guan J, Zhou S. FraGAT: a fragment-oriented multi-scale graph attention model for molecular property prediction. *Bioinformatics*. 2021;37(18):2981–7.
38. Menden MP, Wang D, Mason MJ, Szalai B, Bulusu KC, Guan Y, et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat Commun*. 2019;10(1):2674.
39. De Vera AA, Reznik SE. Combining pi3k/akt/mtor inhibition with chemotherapy. In: *Protein Kinase Inhibitors as Sensitizing Agents for Chemotherapy*. Amsterdam: Elsevier; 2019;229–42.
40. Nitulescu GM, Margina D, Juzenas P, Peng Q, Olaru OT, Saloustros E, et al. Akt inhibitors in cancer treatment: The long journey from drug discovery to clinical use. *Int J Oncol*. 2015;48(3):869–85.
41. Hirai H, Sootome H, Nakatsuru Y, Miyama K, Taguchi S, Tsujioka K, et al. MK-2206, an allosteric Akt inhibitor, enhances antitumor efficacy by standard chemotherapeutic agents or molecular targeted drugs in vitro and in vivo. *Mol Cancer Ther*. 2010;9(7):1956–67.
42. Rivera E, Valero V, Arun B, Royce M, Adinin R, Hoelzer K, et al. Phase II study of pegylated liposomal doxorubicin in combination with gemcitabine in patients with metastatic breast cancer. *J Clin Oncol*. 2003;21(17):3249–54.
43. Perez-Manga G, Lluch A, Alba E, Moreno-Nogueira J, Palomero M, Garcia-Conde J, et al. Gemcitabine in combination with doxorubicin in advanced breast cancer: final results of a phase II pharmacokinetic trial. *J Clin Oncol*. 2000;18(13):2545–52.
44. Gong JH, Zheng YB, Zhang MR, Wang YX, Yang SQ, Wang RH, et al. Dexamethasone enhances the antitumor efficacy of Gemcitabine by glucocorticoid receptor signaling. *Cancer Biol Ther*. 2020;21(4):332–43.
45. Yuan Y, Zhou X, Ren Y, Zhou S, Wang L, Ji S, et al. Semi-mechanism-based pharmacokinetic/pharmacodynamic model for the combination use of dexamethasone and gemcitabine in breast cancer. *J Pharm Sci*. 2015;104(12):4399–408.
46. Stringer-Reasor EM, Baker GM, Skor MN, Kocherginsky M, Lengyel E, Fleming GF, et al. Glucocorticoid receptor activation inhibits chemotherapy-induced cell death in high-grade serous ovarian carcinoma. *Gynecol Oncol*. 2015;138(3):656–62.
47. Degen J, Wegscheid-Gerlach C, Zaliani A, Rarey M. On the art of compiling and using 'drug-like' chemical fragment spaces. *ChemMedChem*. 2008;3(10):1503.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.