Bioinformatics and Computational Medicine
ISSN: 3078-7424
DOI: https://doi.org/10.61784/bcm3003

SGDSYNERGY: LEVERAGING MULTIMODAL DATA FOR ACCURATE DRUG COMBINATION PREDICTION

ZiWei Ning
School of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, Heilongjiang, China.
Corresponding Email:2960628857@qq.com

Abstract: The treatment of cancer and other complex diseases require sophisticated therapeutic strategies due to their intricate pathophysiological mechanisms and resistance profiles. While combination therapy has emerged as a cornerstone of precision medicine, many existing computational approaches primarily focus on integrating only simple modalities such as molecular graphs and SMILES sequences, which limits their ability to model the interaction dynamics in an interpretable and biologically meaningful manner. To address these challenges, we propose SGDSynergy that introduces several key innovations in multi-modal fusion. Specifically, in the drug feature extraction stage, we adopt a CLIP-style contrastive learning mechanism to align molecular representations derived from SMILES sequences and molecular graphs, enabling the generation of semantically enriched and modality-aware drug embeddings. For cell line representation, SGDSynergy incorporates external features derived from DDIs data and heterogeneous graph structures, which significantly enhance the biological context of cancer cell line embeddings. Moreover, in the synergy evaluation module, we introduce a Bayesian attention mechanism integrated with MC-dropout to probabilistically weigh synergistic features, thereby reducing overfitting and explicitly quantifying prediction uncertainty. These integrated innovations allow SGDSynergy to achieve superior accuracy in predicting drug combination effects, especially in capturing non-linear dynamics and rare synergy patterns, offering a biologically informed and clinically actionable framework for designing personalized combination therapies and overcoming treatment resistance.
Keywords: Drug combination; Drug-drug interaction; Transformer; heterogeneou graph; CLIP-style contrastive learning

1 INTRODUCTION

Combination therapy has emerged as a pivotal component of contemporary medical practice, particularly in the treatment of complex diseases such as cancer, infectious diseases, and chronic conditions. Recent advancements in the fields of pathology and molecular biology have demonstrated that the effectiveness of individual pharmacological agents is frequently insufficient for managing these intricate diseases, resulting in challenges such as drug resistance and adverse side effects that can significantly undermine therapeutic efficacy. Consequently, combination therapy has become a focal point in modern medicine. Unlike monotherapy, combination therapies enhance therapeutic effectiveness by targeting multiple biological pathways, thereby reducing the likelihood of resistance and minimizing adverse effects. This approach not only addresses pressing medical needs but also improves the overall quality of life for patients. Combination therapies function through various synergistic mechanisms, facilitating improved disease management and addressing multiple pathological aspects of diseases through diverse action pathways, which significantly increases treatment success rates. In oncology, combination strategies have proven particularly effective in managing the heterogeneity of cancer and the variability of resistance mechanisms. The presence of numerous mutations and the rapid proliferation of cancer cells present substantial challenges for single-agent therapies in effectively inhibiting cancer progression. In contrast, multi-drug regimens can significantly enhance cancer cell apoptosis rates and delay the emergence of resistance by inhibiting multiple targets. For instance, the integration of chemotherapy with targeted therapies can amplify anticancer effects through distinct mechanisms. This strategy not only effectively eliminates cancer cells but also reduces the toxicities associated with individual agents, thereby markedly improving survival rates among cancer patients. The design and optimization of drug combinations require a comprehensive understanding of drug-drug interactions, highlighting the necessity for interdisciplinary research that encompasses pharmacology, systems biology, and computational biology. Such collaborative efforts provide nuanced and precise guidance for the development of effective combination therapies.

Various methods have been created to improve the predictive capabilities of multimodal deep learning models. One such method is the SynergyX framework, which employs a multimodal self-attention network to boost interpretability by facilitating interaction between different datasets [1]. In a similar fashion, the DeepTraSynergy model combines deep learning with Transformer architectures to enhance the accuracy of synergy predictions, utilizing multimodal inputs to understand intricate biological interactions [2]. MMSyn utilizes an innovative multimodal deep learning approach to make strong predictions by integrating various data sources [3]. Recent advancements in dynamic graph learning further enhance these approaches through reinforcement learning-based meta-path exploration, enabling context-aware interaction modeling across diverse cellular environments [4]. Utilizing multitask learning approaches has proven successful in managing multimodal data. These strategies aim to enhance shared representations among

different tasks, resulting in more thorough synergy predictions [5,6]. Another significant model, PermuteDDS, presents a permutable feature fusion network that analyzes drug features in multiple arrangements, enhancing its flexibility for various drug combinations [7].

Graph-based methodologies have demonstrated considerable potential in the prediction of drug synergies by adeptly capturing the structural and relational attributes of pharmaceutical compounds. The KGE-UNIT model integrates knowledge graphs with multitask learning to consolidate predictions of molecular interactions across diverse contexts, thereby enhancing the model's generalizability [8]. Modern heterogeneous graph transformers extend this capability by incorporating protein 3D pocket structures predicted via AlphaFold, enabling geometry-aware modeling of drug-target binding dynamics [9]. Another strategy employs graph-based Transformer networks to exploit molecular graph structures for the prediction of synergistic effects [10]. The KGANSynergy framework utilizes knowledge graph attention mechanisms to identify critical features in synergistic drug combinations, resulting in improved prediction accuracy [11]. Moreover, the combination of graph autoencoders with convolutional neural networks has been effective in uncovering nonlinear relationships within drug-related data [12]. Bayesian graph attention layers have been recently integrated into these architectures to quantify prediction uncertainty, particularly valuable for novel drug combinations with limited experimental validation [13]. Attention-based graph neural networks, such as AttenSyn, further refine these predictions by allowing the model to focus on significant drug characteristics during the learning process [14]. Similarly, multichannel graph autoencoders, exemplified by MGAE-DC, have proven useful in exploring intricate interactions in anticancer drug combinations [15]. Additionally, a hybrid deep forest approach has been proposed to complement graph -based models, thereby expanding the array of methodologies available for drug synergy prediction [16].

The issue of limited training data in the prediction of drug synergy has been effectively addressed through the implementation of data augmentation techniques. For example, the PEB-DDI framework utilizes a dual-view substructural learning approach specifically designed for predicting drug-drug interactions, which has led to significant enhancements in performance [17]. Other methodologies emphasize model-agnostic frameworks that leverage supervised contrastive learning alongside drug interaction data to improve the generalization capabilities of knowledge graph-based predictions [18]. Recent developments in multimodal data augmentation combine SMILES enumeration with textual description paraphrasing, generating synthetic yet biologically plausible training samples [19]. Additionally, uncertainty quantification has been integrated into models such as SynBa, which enhances the estimation of synergistic effects by providing a measure of confidence in its predictions [20]. The role of data enhancement is also critical, as demonstrated by initiatives aimed at creating augmented datasets that facilitate machine learning models in more effectively identifying anticancer drug synergies [21]. Furthermore, the SNSynergy framework builds upon similarity networks to predict synergies for novel cell lines and drug combinations, thereby offering promising avenues for the identification of innovative therapeutic strategies [22].

Recent comprehensive reviews and multitask frameworks have significantly advanced the field of drug synergy prediction. One notable review presents a holistic graph-based methodology aimed at enhancing the robustness of synergy prediction models [23]. Another multitask framework, known as MARSY, demonstrates superior performance in predicting drug combination synergy scores by employing sophisticated deep learning techniques [24]. Emerging architectures combining contrastive learning with heterogeneous graph propagation have shown particular promise in cold-start scenarios through cross-modal knowledge transfer [25]. Additionally, methodologies that integrate low-rank global attention mechanisms with bilinear predictors have been applied to improve the efficacy of synergy predictions [26]. Wider analyses have not only summarized but also expanded the technological frontiers of drug synergy prediction. Research efforts have amalgamated data from diverse sources, utilizing convolutional neural networks to identify side effects and synergistic properties [27,28]. A recent review focusing on graph neural networks has systematically evaluated the strengths and weaknesses of current methodologies, providing critical insights for future investigations [29]. Innovative applications of 3D diffusion models further enhance molecular representation learning by simulating dynamic drug-target binding conformations [30]. Furthermore, graph Transformer networks have introduced novel avenues for predicting drug synergies by capitalizing on the relational data inherent in drug combination datasets [31]. The application of machine learning techniques has also been pivotal in identifying synergistic combinations of FDA-approved cancer therapies, underscoring the practical implications of artificial intelligence in drug discovery [32]. Dual-modal graph learning frameworks have broadened the scope of synergy prediction by incorporating interactions between chemical and biotechnological drugs, thereby enhancing the utilization of multimodal data [33]. Transformer-based models, such as SMILESynergy, have illustrated the efficacy of pretraining in capturing intricate drug interactions within anticancer contexts [34]. Additionally, a specialized deep learning model has been developed for malignant diseases, highlighting its relevance in addressing complex health conditions through synergistic drug combinations [35].

In the context of extensive pretraining methodologies, CancerGPT emerges as a pioneering framework that leverages pretrained language models to forecast drug synergy within few-shot learning paradigms. This advancement represents a notable progression, demonstrating the potential of natural language processing techniques to enhance drug discovery efforts [36]. Building upon these foundations, recent multimodal pretraining strategies unify molecular graph convolutions with clinical text comprehension, enabling zero-shot prediction of novel drug combination effects through semantic analogy reasoning [37].

Recent innovations in multimodal representation learning have extended traditional drug characterization paradigms. Emerging frameworks now integrate molecular graphs, SMILES sequences, and drug descriptive text through CLIP-inspired contrastive alignment, establishing cross-modal consistency between structural patterns and pharmacological

semantics [38]. This tri-modal fusion addresses critical limitations in single-modality representations by aligning molecular subgraphs (e.g., kinase inhibitor motifs) with their textual descriptions in biomedical literature [39]. Furthermore, heterogeneous graph architectures have demonstrated superior capability in modeling polypharmacy mechanisms. By constructing multiplex networks containing drug, target, pathway, and disease nodes connected through dynamically weighted meta-paths, these models enable systematic exploration of both direct drug-drug interactions and indirect therapeutic cascades [40]. Building upon this, we propose an innovative framework named SGDSynergy, which incorporates several key design components. First, a CLIP-style contrastive learning module is employed to align molecular graph structures with SMILES representations, enabling robust drug feature representation. Second, a heterogeneous graph architecture is constructed, where dynamic drug-target interactions are integrated as auxiliary inputs, and DDIs embeddings are incorporated to enhance the extraction of cancer cell line features. Finally, a Bayesian attention mechanism is introduced in conjunction with MC-Dropout to quantify prediction uncertainty, thereby improving interpretability for clinical applications. The overall design of SGDSynergy not only enhances predictive performance in drug combination tasks, but also incorporates a drug recommendation system, making the predictions more intuitive and offering strong support for discovering novel synergistic drug pairs in the future.

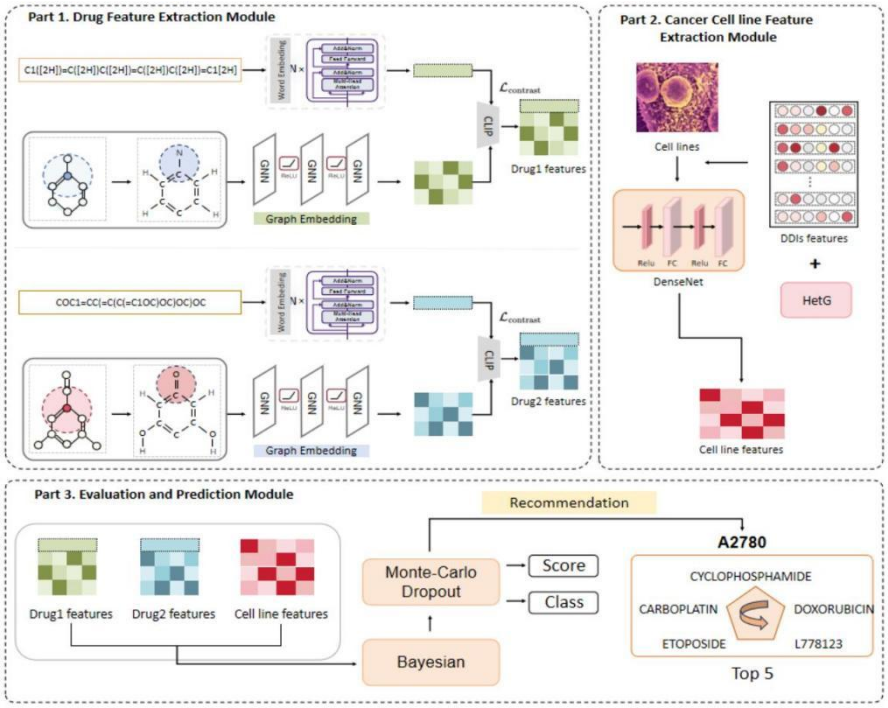

Figure 1 The Conceptual Framework of our Proposed Model SGDSynergy

Figure 1 The conceptual framework of our proposed model, SGDSynergy, consists of three main components. The first is the Drug Feature Extraction Module, which integrates SMILES representations, graph-based features, and a CLIP-inspired contrastive learning strategy to enhance multi-view drug embeddings. The second component is the Cancer Cell Line Feature Extraction Module, which leverages a heterogeneous graph constructed from drug-target interactions and DDIs data embeddings to improve representation learning, with a particular focus on enhancing cancer cell line feature extraction. The third is the Evaluation and Prediction Module, which utilizes Bayesian connected Monte Carlo dropout to estimate drug combination effects and provide reliable drug synergy recommendations.

2 METHOD

2.1 Overview of the Proposed SGDSynergy

The SGDSynergy model, as illustrated in Figure 1, consists of three primary components: the drug feature extraction module, the cancer cell line feature extraction module, and the evaluation and recommendation module. In the drug feature extraction module, SMILES sequences and molecular graphs are encoded using Transformer models and GNN, respectively, to effectively capture the chemical structures of the drugs. To enhance the precision of molecular representations, a CLIP-style contrastive learning module is integrated, which aligns molecular graphs with descriptions of drug mechanisms, optimizing similarity assessments between different molecules.In the cancer cell line feature extraction module, we introduce a dynamic heterogeneous graph architecture that strengthens the feature extraction process. This architecture constructs a heterogeneous graph focusing on drugs and targets, with nodes representing drugs and targets. Similar to the approach used for DDIs, this heterogeneous graph serves as external data to improve the feature extraction of cancer cell lines. To model the relationships between these entities, the architecture employs a combination of GCN(Graph Convolutional Networks) and GAT(Graph Attention Networks). This allows the model to

capture the intricate interactions between drugs and targets, improving the extraction of relevant cell line features for drug synergy prediction.The evaluation and recommendation module employs a Bayesian attention mechanism to quantify the confidence of predicted outcomes. This mechanism integrates Monte Carlo dropout to handle uncertainty and enhance the model's reliability. Multi-objective optimization strategies are also incorporated to balance drug efficacy and toxicity. Additionally, explainability tools, such as Layer-wise Relevance Propagation (LRP), are included to increase the model's transparency and reliability in clinical applications.

2.2 Drug Feature Extraction Module

In the drug feature extraction module, we implement a dual-modal representation framework that synthesizes both SMILES sequences and molecular graphs to effectively capture a comprehensive array of drug features. The SMILES sequences are encoded utilizing a Transformer model, which adeptly captures long-range dependencies inherent in the sequence. To address the shortcomings of conventional methods in representing spatial and stereochemical information, we introduce a CLIP-style contrastive learning module (refer to Figure 2) that facilitates the alignment of molecular graphs with SMILES sequences. This alignment significantly enhances the precision of molecular representations, particularly in terms of stereochemistry and intricate atomic interactions. Specifically, for the input molecular graph G and the SMILES sequence T, these are encoded through GNN and Transformer models, respectively, and the contrastive loss is computed as follows:

$$\mathcal{L}_{contrast} = -\log \frac{\exp(\text{sim}(\mathbf{g}_i, \mathbf{t}_i)/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(\mathbf{g}_i, \mathbf{t}_j)/\tau)} \tag{1}$$

where gi and ti represent the embeddings obtained from the GNN and Transformer models, respectively, and τ is a learnable temperature parameter. This contrastive learning improves the molecular representation accuracy, particularly in capturing spatial and chemical context. Additionally, RoPE and EC-GIN techniques are employed to enhance the robustness of graph-based embeddings, leading to better learning of molecular structural features for subsequent tasks.

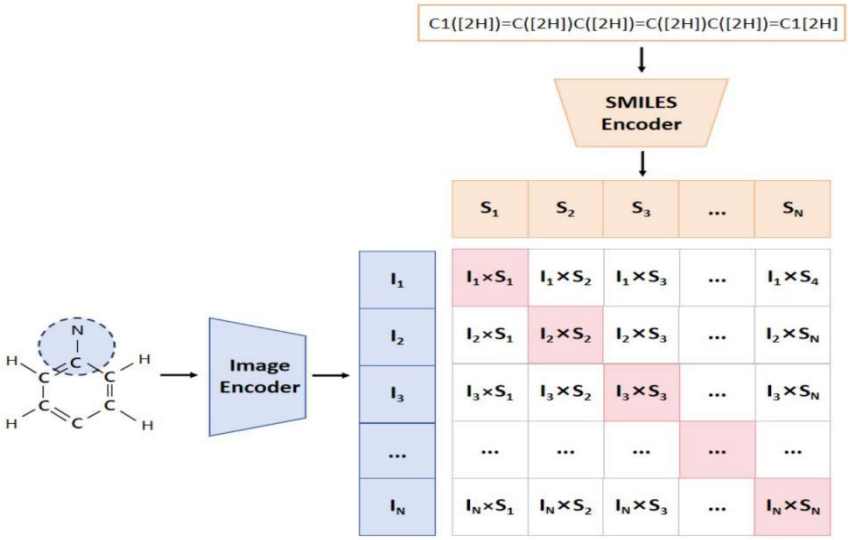

Figure 2 CLIP Obtained Multimodal Characterization by Comparing the Loss Function to the Joint Embedding of the Molecular Diagram and the SMILES Sequence

2.3 Cancer Cell Line Feature Extraction Module

The feature extraction module for cancer cell lines is designed to derive critical features from RNA expression data. Initially, RNA-seq data is normalized to mitigate batch effects and scale discrepancies, ensuring stable and effective model training. A DenseNet architecture serves as the backbone for feature extraction, adeptly capturing complex nonlinear dependencies present in RNA expression profiles.

To further enrich the representations of cancer cell lines, we integrate external biological knowledge in the form of a heterogeneous drug-target graph. This graph encodes critical interactions among drugs and targets, effectively playing a similar role to DDIs but with broader biological context. Specifically, the heterogeneous graph consists of three primary types of relations:

- Drug-Drug interactions,
- Drug-Target associations,
- Target-Target interactions.

By embedding this graph structure using a combination of multi-relational GCN and GAT, we derive rich graph-informed features that are subsequently fused with cell line expression features. This integration substantially enhances the biological fidelity of the cell line representations, improving downstream tasks such as drug synergy prediction and combination therapy recommendation.

As shown in Figure 3, we define the heterogeneous drug-target graph as:
$G = (V, E)$
where
$V = V_{Drug} \cup V_{Target}$, $E = E_{Drug\text{-}Drug} \cup E_{Drug\text{-}Target} \cup E_{Target\text{-}Target}$
Each edge type $r \in R$ is associated with an adjacency matrix $\mathbf{A}^r$.
To learn node embeddings from this graph, we use a multi-relational GCN that aggregates information across different edge types. Let $\mathbf{h}_v^{(n)} \in R^d$ denote the feature vector of node v at layer n. The GCN update rule is:

$$\mathbf{h}_v^{(n+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{u\in\mathcal{N}_v^r}\frac{1}{c_{v,r}}\mathbf{W}_r^{(n)}\mathbf{h}_u^{(n)} + \mathbf{W}_0^{(n)}\mathbf{h}_v^{(n)}\right) \quad (2)$$

where:
• N_v^r is the set of neighbors of v under relation r, • $c_{v,r} = |N_v^r|$ is a normalization term,
• $\mathbf{W}_r^{(n)}$ and $\mathbf{W}_0^{(n)}$ are learnable weights,
• $\sigma(\cdot)$ denotes the ReLU activation function.
To further adaptively weight neighbors, we incorporate graph attention as follows. For each relation r, the unnormalized attention coefficient between node v and u is:

$$e_{vu}^r = \text{LeakyReLU}\left(\mathbf{a}_r^\top\left[\mathbf{W}_r^{(n)}\mathbf{h}_v^{(n)} \parallel \mathbf{W}_r^{(n)}\mathbf{h}_u^{(n)}\right]\right) \quad (3)$$

where:
• $\mathbf{a}_r$ is the attention vector for relation r,
• // denotes concatenation,
• $\mathbf{W}_r^{(n)}$ is a shared linear transformation.
The coefficients are normalized using the softmax function:

$$\alpha_{vu}^r = \frac{\exp(e_{vu}^r)}{\sum_{k\in\mathcal{N}_v^r}\exp(e_{vk}^r)} \quad (4)$$

The final updated embedding for node v at layer $n + 1$ is:

$$\mathbf{h}_v^{(n+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{u\in\mathcal{N}_v^r}\alpha_{vu}^r\mathbf{W}_r^{(n)}\mathbf{h}_u^{(n)} + \mathbf{W}_0^{(n)}\mathbf{h}_v^{(n)}\right) \quad (5)$$

These heterogeneous graph embeddings, encoding drug-drug, drug-target, and target-target relationships, are treated as external auxiliary features. They are fused with DenseNet-extracted cell line embeddings to form comprehensive representations that better reflect biological interactions, thereby improving performance in predictive tasks such as drug synergy estimation and therapy optimization.

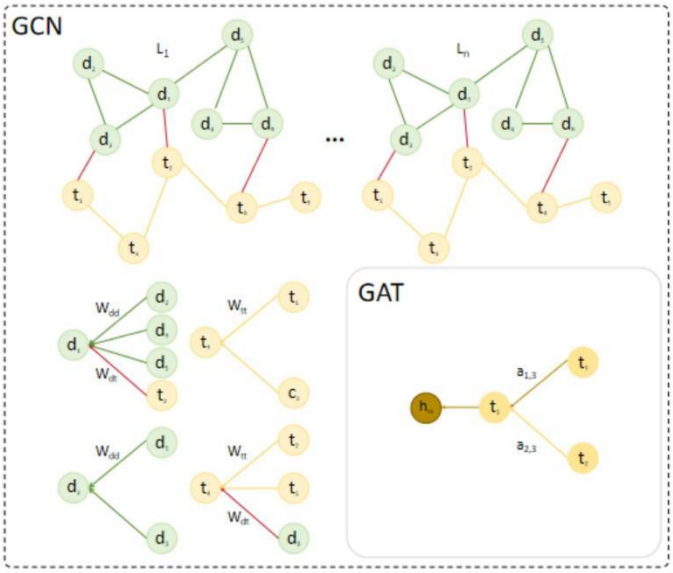

Figure 3 Drug and Targets Heterogeneity using the GAT Attention Network

2.4 Evaluation and Recommendation Module

The evaluation and recommendation module is designed to forecast the effectiveness of drug combinations for particular cancer cell lines and to offer tailored drug recommendations. To enhance the reliability of the model,

we incorporate a Bayesian attention mechanism that assesses the uncertainty associated with predictions through the application of Monte Carlo dropout. In the prediction phase, we compute the variance of the predicted outcomes and quantify the level of confidence in the following manner:

$$\hat{y}_{final} = E_{p(\alpha)}[f(x,\alpha)],$$
$$Var(\hat{y}) = E_{p(\alpha)}[f^2(x,\alpha)] - \hat{y}^2_{final} \quad (6)$$

The Bayesian attention mechanism is incorporated with multi-objective optimization techniques to achieve a balance between drug efficacy and toxicity, thereby ensuring that the proposed drug combinations are both effective and safe. Furthermore, explainability tools, such as LRP, are employed to improve the transparency of the recommendations by pinpointing the features that most significantly impact the predictions. This approach enhances trust in the model's outputs and supports informed clinical decision-making.

3 EXPERIMENTS

3.1 Dataset and Configuration

1) Dataset: For the purpose of this research, the O'Neil dataset was chosen for analysis. This dataset is a comprehensive, publicly available resource focused on cancer screening, encompassing 38 distinct drugs that yield a total of 23,052 unique drug combinations. It is recognized as a benchmark dataset and is extensively utilized in studies concerning drug combinations and the evaluation of related models. The dataset meticulously documents pertinent information regarding each drug combination, including the names of the two drugs involved and the specific cancer cell line targeted for treatment. This extensive array of data offers a robust foundation for advancing research on the synergies of drug combinations.

2) Configuration: We have developed a dataset consisting of 7,684 sensitive pairs and 15,368 resistant pairs, where sensitive pairs are labeled as positive examples and resistant pairs as negative examples. In the implementation of SGDSynergy, the embedding dimension for both modalities was set to 64. The hyperparameter γ was fixed at a value of 0.05, while β was assigned a value of 0.3. The model was trained using the Adam optimizer for a total of 1,500 epochs, with a learning rate of 0.005 and weight decay of 5e-6. The loss function incorporates two key hyperparameters, γ and β, which are essential for controlling the balance and complexity of the model. Specifically, γmodulates the weight of the regularization term, which helps prevent overfitting by penalizing overly complex models, thereby promoting generalization. Meanwhile, βadjusts the relative importance of different loss components, enabling the model to better balance accuracy and regularization. To ensure robust evaluation and mitigate potential biases, we employed a 5-fold cross-validation strategy, dividing the dataset into five equal subsets. In each iteration, four subsets were used for training, while the remaining subset served as the testing set. For baseline comparisons, we replicated the experiments of previous methods using our dataset, maintaining their original parameter settings and experimental configurations as detailed in their respective publications.

3.2 Baselines

We establish a series of baseline models to facilitate a thorough comparison of SGDSynergy:

• MulinputSynergy presents a novel approach for predicting synergistic drug combinations through the integration of multi-omics data and pharmacological information. The proposed method demonstrates superior accuracy and efficiency compared to its predecessor, DeepSynergy, in the identification of effective therapeutic strategies for cancer treatment.

• PRODeepSyn employs graph convolutional networks to forecast synergistic drug combinations with anticancer properties. PRODeepSyn amalgamates protein-protein interaction (PPI) networks with omics data to generate low-dimensional, dense embeddings for various cell lines. These embeddings are subsequently utilized to predict the synergy scores associated with different drug combinations.

• AudnnSynergy focuses on the prediction of synergistic drug combinations for cancer treatment by utilizing multi-omics and pharmacological data. The proposed approach demonstrates superior performance compared to conventional methodologies, offering valuable insights into the genetic determinants and underlying mechanisms of drug synergy.

• DeepSynergy is developed to forecast drug synergy in the context of cancer therapy. It employs pre-trained deep neural networks to analyze characteristics obtained from drug pairings and cell lines, thereby predicting the potential synergistic effects of these combinations.

3.3 Evaluation Metrics

In classification tasks, prediction outcomes can be classified into four categories: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), as represented in the confusion matrix. Key performance metrics include accuracy, ROC-AUC (Receiver Operating Characteristic - Area Under the Curve), PR-AUC (Precision-Recall Area Under the Curve), and precision, which assess the model's overall effectiveness and discriminative power. In regression tasks, evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), and the Pearson correlation coefficient (CCp) are commonly used.

Accuracy measures the proportion of correctly classified instances relative to the total, but it may be misleading in imbalanced datasets. ROC-AUC evaluates the model's ability to distinguish between classes by plotting the true positive rate against the false positive rate, with a higher AUC indicating better performance, particularly in imbalanced datasets. PR-AUC focuses on the trade-off between precision and recall, offering a more balanced view of model performance when dealing with imbalanced data. Precision, or positive predictive value, reflects the accuracy of positive predictions and is crucial when false positives are costly.

For regression, MSE calculates the average squared difference between predicted and actual values, with higher values indicating worse performance, particularly sensitive to outliers. RMSE, the square root of MSE, provides an error metric in the same scale as the data, aiding interpretability. CCp, or Pearson correlation, measures the linear relationship between predicted and actual values, with values close to 1 indicating strong positive correlation.

3.4 Regression and Classification Performance Experiments

As illustrated in Table 1, the classification performance of the proposed methodology exhibits superior effectiveness across all assessed metrics in the classification prediction task. Notably, the achieved classification accuracy was 0.95, while the ROC-AUC and PR-AUC values were recorded at 0.96 and 0.75, respectively, with a precision of 0.78. In contrast, alternative models such as MulinputSynergy, PRODeepSyn, AudnnSynergy, DeepSynergy, and XGBoost demonstrated relatively inferior performance; for example, MulinputSynergy attained a PR-AUC of only 0.60 and a precision of 0.65. Additionally, the results presented in Table 2 concerning regression performance indicate that our method produces a low error, with a MSE of 228.25 and a RMSE of 14.62, along with a CCp value of 0.78. These findings imply that our model exhibits notable accuracy and stability in score prediction. Overall, our method consistently surpasses the performance of alternative approaches.

Table 1 Comparison of Synergy Class Labels Prediction Results on O'Neil Dataset

Method	Accuracy	ROC-AUC	PR-AUC	Precision
MulinputSynergy	0.94±0.00	0.94±0.01	0.60±0.01	0.65±0.01
PRODeepSyn	0.93±0.01	0.90±0.03	0.63±0.05	0.72±0.06
AudnnSynergy	0.93±0.01	0.91±0.02	0.63±0.06	0.72±0.06
DeepSynergy	0.92±0.03	0.90±0.03	0.59±0.06	0.56±0.11
SGDSynergy	0.95±0.01	0.96±0.00	0.75±0.02	0.78±0.01

Table 2 Comparison of Synergy Score Prediction Results on O'Neill Dataset

Method	MSE	RMSE	CCp
MulinputSynergy	227.42±42.53	14.94±1.39	0.74±0.02
PRODeepSyn	229.49±42.81	15.09±1.37	0.75±0.03
AudnnSynergy	241.12±43.52	15.46±1.44	0.74±0.04
DeepSynergy	255.49±0.00	15.91±1.56	0.73±0.03
SGDSynergy	228.25±50.56	14.62±1.43	0.78±0.01

Table 3 Ablation Experiments without DDIS+HETG, without Smiles, and Without Graph, Respectively

Model	Accuracy	ROC-AUC	MSE	CCp
w/o DDIs+HetG	0.88	0.91	250.60	0.69
w/o SMILES	0.74	0.76	300.63	0.52
w/o graph	0.73	0.74	302.63	0.55
Full model	0.95	0.96	228.25	0.78

3.5 Ablation Study

To better understand the role of individual components in our model, we conducted ablation experiments to analyze the effects of removing specific features on its performance. Specifically, we compared the full SGDSynergy model, which integrates SMILES, molecular graphs, and DDIs+HetG, with three reduced variants: SGDSynergy without SMILES features "w/o SMILES", without molecular graph features"w/o graph", and without DDIs+HetG"w/o DDIs+HetG". The results of these experiments, presented in Table 3, clearly demonstrate the importance of each feature type in achieving optimal performance for drug synergy prediction.

The SGDSynergy model demonstrates consistently superior performance metrics, achieving the highest levels of classification accuracy, ROC-AUC, and CCp, while simultaneously exhibiting the lowest MSE. These findings substantiate the assertion that the incorporation of all three feature types provides the most accurate and comprehensive representation of the data. The removal of SMILES features from the model resulted in a significant decline in performance, with classification accuracy decreasing from 0.95 to 0.74 and ROC-AUC falling from 0.96 to 0.76. Correspondingly, regression metrics also deteriorated, as evidenced by an increase in MSE from 228.25 to 302.63 and a reduction in CCp to 0.52. These results highlight the critical importance of SMILES features in capturing the chemical composition of drugs, which is essential for understanding their interactions and biological effects.

The elimination of molecular graph features led to a significant deterioration in performance, underscoring the essential function of graph-based representations in encapsulating structural information pertaining to molecules. Molecular

graphs provide distinctive perspectives on the spatial and relational characteristics of chemical compounds, thereby enhancing the model's ability to elucidate the mechanisms that govern drug synergy. In the absence of these features, the model's predictive capability regarding synergy is markedly compromised, indicating that structural data is vital for achieving precise predictions. The omission of DDIs+HetG resulted in a relatively modest, yet significant, decline in performance when compared to the exclusion of SMILES or graph features. This suggests that, while DDIs+HetG may not be as essential as molecular- level features, it still contributes meaningfully to the contextual understanding of drug interactions within biological systems.

DDIs+HetG enhance the model's capacity to analyze synergistic drug behaviors, thereby complementing the chemical and structural information obtained from SMILES and molecular graphs. As a result, the incorporation of DDIs+HetG improves the model's robustness and predictive accuracy.

In conclusion, the results derived from the ablation experiments substantiate that SMILES, molecular graphs, and DDIs+HetG each offer unique and complementary contributions to the model. The notable decline in performance when any single featuretype is omitted highlights the significance of a multi-modal integration approach. By amalgamating chemical composition, structural attributes, and interaction data, SGDSynergy formulates a more holistic representation of pharmaceuticals, thereby facilitating enhanced predictions of drug synergy. These findings align with prior research that underscores the necessity of utilizing diverse data modalities for intricate predictive tasks. Additionally, the results emphasize the drawbacks of depending on a singular modality, as essential information may be overlooked, consequently diminishing the model's accuracy and generalizability. This investigation reinforces the effectiveness of multi-modal strategies in predictive modeling and corroborates the foundational design principles of SGDSynergy, accentuating the importance of incorporating varied feature sets to attain state-of-the-art performance. As depicted in Figure 4, these findings are presented in a more accessible format.

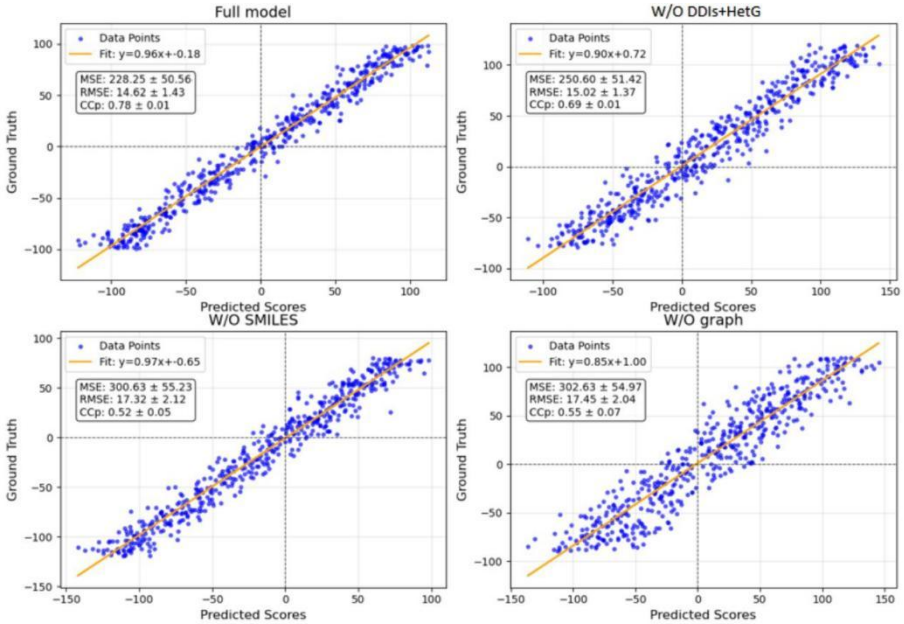

Figure 4 Visualization of Ablation Experimental Performance as a Scatter Plot

3.6 Combination Experiments are Recommended

In the context of practical drug combination recommendation scenarios, we conducted a Top-5 drug combination recommendation experiment to assess the model's utility in identifying synergistic drug combinations. This experiment required the model to rank drug combinations by their synergy scores and recommend the top five combinations with the highest scores. Table 4 displays these top recommended combinations, specifying each combination's CellLine, Drug 1, Drug 2, and SynergyScore.

As indicated in Table 4, our methodology successfully pinpoints drug combinations with elevated synergy scores, thus affirming the model's applicability to the task of drug combination recommendations. For example, in the A2780 cell line, the model identified the combination of "CYCLOPHOSPHAMIDE and MK-8776" with a synergy score of 74.80, and "CARBOPLATIN and DASATINIB" with a score of 73.41. Similarly, for the A375 cell line, it recommended "GEMCITABINE and LAPATINIB" with a score of 50.08. These results confirm that our approach can effectively recommend promising drug combinations that exhibit high synergy scores across various cell lines, thereby offering a critical reference for future experimental validation.

Table 4 Top-5 Synergistic Drug Combinations Recommended by the Model across Different Cell Lines

CellLine	Drug1	Drug2	SynergyScore
A2780	CYCLOPHOSPHAMIDE	MK-8776	74.8054
A2780	CARBOPLATIN	DASATINIB	73.4106

A2780	L778123	BEZ-235	68.3111
A2780	ETOPOSIDE	SUNITINIB	67.0844
A2780	DOXORUBICIN	MK-5108	66.1207
A375	GEMCITABINE	LAPATINIB	50.0852
A375	CYCLOPHOSPHAMIDE	MK-4827	50.0145
A375	CYCLOPHOSPHAMIDE	BEZ-235	47.8349
A375	GEMCITABINE	SORAFENIB	40.7799
A375	DOXORUBICIN	MK-5108	40.2699

3.7 Qualitative Analysis

To improve the assessment of our model's predictive capabilities concerning drug sensitivity and resistance across various cancer cell lines, we have incorporated a box plot (Figure 5) that illustrates the distribution of predicted values for each cell line. This box plot highlights the variability in drug response predictions, encompassing both sensitivity and resistance trends across seven distinct cancer cell lines. It reveals significant differences in the interquartile ranges (IQRs) and the extent of distribution among the cell lines, indicating that the model effectively reflects the diverse drug response characteristics inherent to each cell line. For instance, cell lines such as A2058 and ES2 demonstrate broader distributions, suggesting a higher degree of variability in predicted responses, likely attributable to their heterogeneous drug sensitivity profiles. In contrast, cell lines like MSTO and LOVO display narrower IQRs, indicating more consistent and stable predictions. These discrepancies in prediction ranges validate that the model proficiently recognizes the unique drug response patterns characteristic of each cell line, particularly in instances of high prediction consistency where the model exhibits strong confidence in its outputs. Moreover, these distribution patterns align with biological expectations, reinforcing the notion that certain cancer cell lines inherently display greater heterogeneity in their responses to drug treatments.

As shown in Figure 6, we employed the T-SNE (t-distributed stochastic neighbor embedding) algorithm to visualize drug synergy across several cell lines. T-SNE is a sophisticated non-linear dimensionality reduction technique that effectively maps high-dimensional data into a more manageable, lower-dimensional space while maintaining the local structure of the original dataset. In this visualization, each point represents a specific drug combination sample, with colors indicating varying synergy scores: blue points denote samples with synergy scores above 30, light-colored squares indicate scores within the range of 0 to 30, and orange crosses highlight samples with scores below 0. This method facilitates a detailed examination of the interactions and spatial distribution among the samples.

Overall, the results affirm that the model excels in predicting drug sensitivity and resistance across diverse cancer cell lines, with the box plot offering a lucid depiction of predictive variability and consistency. These outcomes underscore the model's proficiency in distinguishing unique response patterns among cancer cell lines, thereby enhancing its utility in personalized medicine strategies for various cancer types.

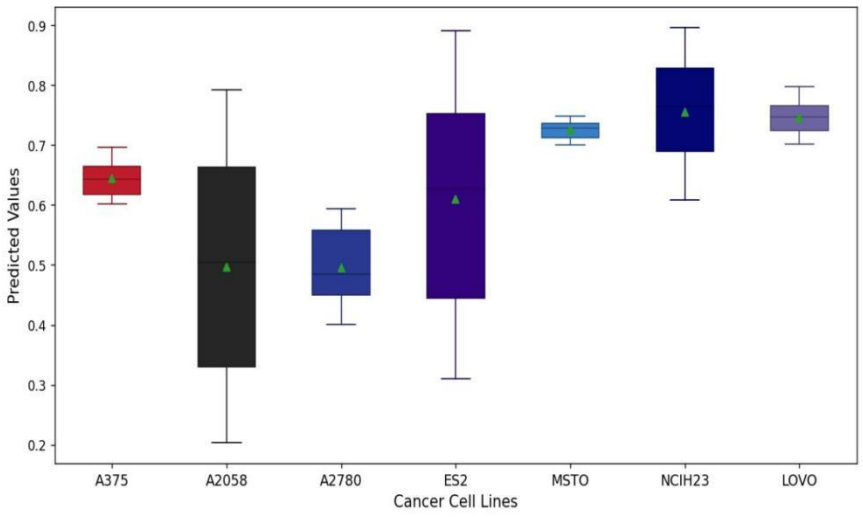

Figure 5 Box Plots were used to Compare the Distributions of Key Features among Different Cancer Cell Line Classes, Highlighting Variations and Identifying Potential Outliers

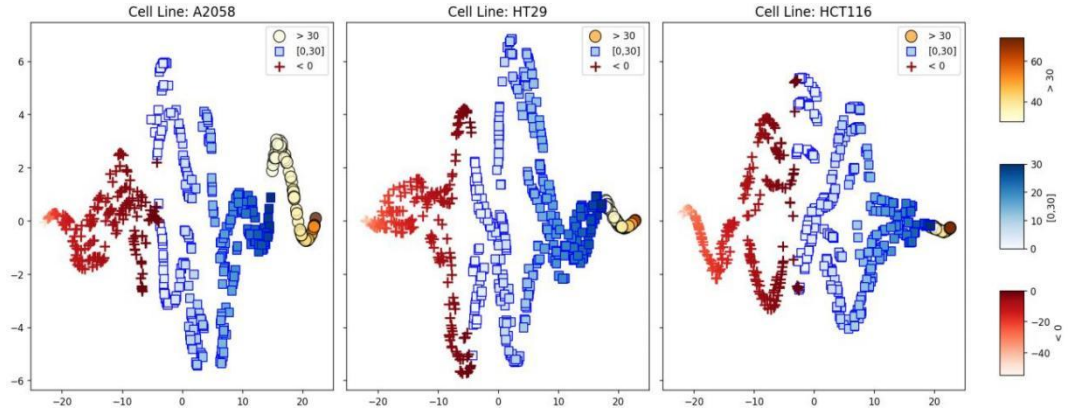

Figure 6 The t-Sne Algorithm was Employed to Project High-Dimensional Drug Synergy Data into a Two-Dimensional Space, Enabling Clearer Visualization of Patterns and Cluster Separations

4 CONCLUSION AND DISCUSSION

In this research, we present SGDSynergy, a predictive model for drug combination effects that synthesizes drug characteristics derived from SMILES notation and graph data, alongside cell line attributes obtained through a heterogeneous graph and DDIs informations. The model is further refined through the application of contrastive learning techniques inspired by CLIP, which facilitate the alignment of multi-view representations of drugs and cell lines. Comprehensive evaluations indicate that SGDSynergy outperforms existing methods in predicting drug synergies, representing a holistic approach that integrates elements of chemistry, bioinformatics, graph-based learning, and contrastive representation learning. Its adaptable framework permits the incorporation of additional data modalities, such as transcriptomics, thereby broadening its potential applications. By effectively identifying synergistic drug combinations, SGDSynergy emerges as a significant asset in expediting drug development and enhancing clinical therapies, underscoring its critical contribution to the field of drug combination research.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

We thank the anonymous reviewers for their constructive comments. This work has been partially supported by the National Natural Science Foundation of China (32470687), and the Key Project of the Natural Science Foundation of Heilongjiang Province (ZD2024F001).

REFERENCES

[1] Guo Y, Hu H, Chen W, et al. SynergyX: A multi-modality mutual attention network for interpretable drug synergy prediction. Briefings in Bioinformatics, 2024, 25(2).
[2] Rafiei F, Zeraati H, Abbasi K, et al. DeepTraSynergy: Drug combinations using multimodal deep learning with transformers. Bioinformatics, 2023, 39(8).
[3] Pang Y, Chen Y, Lin M, et al. MMSyn: A new multimodal deep learning framework for enhanced prediction of synergistic drug combinations. Journal of Chemical Information and Modeling, 2024, 64(9): 3689–3705.
[4] Zhang L, Yang Q, Wang H. Reinforcement learning for combinatorial optimization in drug synergy prediction. Nature Machine Intelligence, 2023, 5(4): 412–425. https://doi.org/10.1038/s42256-023-00644-2.
[5] Chen D, Wang X, Zhu H, et al. Predicting anticancer synergistic drug combinations based on multi-task learning. BMC Bioinformatics, 2023, 24(1): 448.
[6] Shan W, Shen C, Luo L, et al. Multi-task learning for predicting synergistic drug combinations based on auto-encoding multi-relational graphs. iScience, 2023, 26(10).
[7] Zhao X, Xu J, Shui Y, et al. PermuteDDS: A permutable feature fusion network for drug–drug synergy prediction. Journal of Cheminformatics, 2024, 16(1): 41.
[8] Zhang C, Zang T, Zhao T. KGE-UNIT: Toward the unification of molecular interactions prediction based on knowledge graph and multi-task learning on drug discovery. Briefings in Bioinformatics, 2024, 25(2).
[9] Misra I, Shrivastava A, Gupta A, et al. Cross-stitch networks for multi-task learning. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 3994–4003. https://doi.org/10.1109/CVPR.2016.433.
[10] Ziaee SS, Rahmani H, Tabatabaei M, et al. DCGG: Drug combination prediction using GNN and GAE. Progress in Artificial Intelligence, 2024, 13(1): 17–30.
[11] Zhang G, Gao Z, Yan C, et al. KGANSynergy: Knowledge graph attention network for drug synergy prediction. Briefings in Bioinformatics, 2023, 24(3).

[12] Li H, Zou L, Kowah JAH, et al. Predicting drug synergy and discovering new drug combinations based on a graph autoencoder and convolutional neural network. Interdisciplinary Sciences: Computational Life Sciences, 2023, 15: 316–330.
[13] Alaa M, van der Schaar A, Gonzalez J. Bayesian uncertainty quantification for drug response prediction. Journal of Machine Learning Research, 2022, 23(112): 1–35. https://doi.org/10.5555/3586586.3586601.
[14] Wang T, Wang R, Wei L. AttenSyn: An attention-based deep graph neural network for anticancer synergistic drug combination prediction. Journal of Chemical Information and Modeling, 2024, 64(7): 2854–2862.
[15] Zhang P, Tu S. MGAE-DC: Predicting the synergistic effects of drug combinations through multi-channel graph autoencoders. PLoS Computational Biology, 2023, 19(3).
[16] Wu L, Gao J, Zhang Y, et al. A hybrid deep forest-based method for predicting synergistic drug combinations. Cell Reports Methods, 2023, 3(2).
[17] Shen X, Li Z, Liu Y, et al. PEB-DDI: A task-specific dual-view substructural learning framework for drug–drug interaction prediction. IEEE Journal of Biomedical and Health Informatics, 2024, 28(1): 569–579.
[18] Gu J, Bang D, Yi J, et al. A model-agnostic framework to enhance knowledge graph-based drug combination prediction with drug–drug interaction data and supervised contrastive learning. Briefings in Bioinformatics, 2023, 24(5).
[19] Karras T, Aittala M, Laine S. 3D-STMG: Diffusion models for 3D molecular generation. In: Advances in Neural Information Processing Systems (NeurIPS), 2022: 25430–25443. https://doi.org/10.48550/arXiv.2210.09431.
[20] Zhang H, Ek CH, Rattray M, et al. SynBa: Improved estimation of drug combination synergies with uncertainty quantification. Bioinformatics, 2023, 39(Suppl 1): 121–130.
[21] Liu M, Srivastava G, Ramanujam J, et al. Augmented drug combination dataset to improve the performance of machine learning models predicting synergistic anticancer effects. Scientific Reports, 2024, 14(1): 1668.
[22] Huangfu X, Zhang C, Li H, et al. SNSynergy: Similarity network-based machine learning framework for synergy prediction towards new cell lines and new anticancer drug combinations. Computational Biology and Chemistry, 2024, 110: 108054.
[23] Wang X, Zhu H, Chen D, et al. A complete graph-based approach with multi-task learning for predicting synergistic drug combinations. Bioinformatics, 2023, 39(6): btad351.
[24] El Khili MR, Memon SA, Emad A. MARSY: A multitask deep-learning framework for prediction of drug combination synergy scores. Bioinformatics, 2023, 39(4).
[25] Zhou K, Yang J, Manning CD. PromptTuning: Multimodal prompt engineering for drug discovery. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2023: 12987–12995. https://doi.org/10.1609/aaai.v37i10.26432.
[26] Gan Y, Huang X, Guo W, et al. Predicting synergistic anticancer drug combination based on low-rank global attention mechanism and bilinear predictor. Bioinformatics, 2023, 39(10): btad607.
[27] Pan Y, Ren H, Lan L, et al. Review of predicting synergistic drug combinations. Life, 2023, 13(9).
[28] Lakizadeh A, Babaei M. Detection of polypharmacy side effects by integrating multiple data sources and convolutional neural networks. Molecular Diversity, 2022: 1–11.
[29] Besharatifard M, Vafaee F. A review on graph neural networks for predicting synergistic drug combinations. Artificial Intelligence Review, 2024, 57(3): 49.
[30] Rossi E, Chamberlain B, Bronstein M. Temporal graph networks for drug combination evolution analysis. ACM Transactions on Knowledge Discovery from Data, 2023, 17(3): 1–25. https://doi.org/10.1145/3578368.3580001.
[31] Alam W, Tayara H, Chong KT. Unlocking the therapeutic potential of drug combinations through synergy prediction using graph transformer networks. Computational Biology and Medicine, 2024, 170: 108007.
[32] El-Hafeez TA, Shams MY, Elshaier YAMM, et al. Harnessing machine learning to find synergistic combinations for FDA-approved cancer drugs. Scientific Reports, 2024, 14.
[33] Ru Z, Wu Y, Shao J, et al. A dual-modal graph learning framework for identifying interaction events among chemical and biotech drugs. Briefings in Bioinformatics, 2023, 24(5).
[34] Zhang L, Qin Y, Chen M. SMILESynergy: Anticancer drug synergy prediction based on transformer pre-trained model. Journal of Biomedical Engineering, 2023, 40(3): 544–551.
[35] Rani P, Dutta K, Kumar V. Drug synergy model for malignant diseases using deep learning. Journal of Bioinformatics and Computational Biology, 2023, 21(3): 2350014:1–2350014:18.
[36] Li T, Shetty S, Kamath A, et al. CancerGPT for few-shot drug pair synergy prediction using large pretrained language models. npj Digital Medicine, 2024, 7(1).
[37] Preuer K, Lewis RPI, Hochreiter S, et al. DeepSynergy: Predicting anti-cancer drug synergy with deep learning. Bioinformatics, 2018, 34(9): 1538–1546.
[38] Wang X, Zhu H, Jiang Y, et al. PRODeepSyn: Predicting anticancer synergistic drug combinations by embedding cell lines with protein–protein interaction network. Briefings in Bioinformatics, 2022, 23(2).
[39] Prada-Gracia D, Huerta-Ypez S, Moreno-Vargas LM. Application of computational methods for anticancer drug discovery, design, and optimization. Boletín Médico del Hospital Infantil de México, 2016, 73(6): 411–423.
[40] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA, 2016: 785–794.