A Conformation Enhanced Graph Attention Framework for Predicting Synergistic Drug Combinations

1st Huaze Long
School of Computer Science and Technology
Wuhan University of Science and Technology
Wuhan, China
loong@wust.edu.cn

2nd Shuting Jin*
School of Computer Science and Technology
Wuhan University of Science and Technology
Wuhan, China
shutingjin@wust.edu.cn

3rd Junlin Xu
School of Computer Science and Technology
Wuhan University of Science and Technology
Wuhan, China
xjl@wust.edu.cn

4th Zhiwei Xu
School of Computer Science and Technology
Wuhan University of Science and Technology
Wuhan, China
xuzhiwei@wust.edu.cn

2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3315-1557-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/BIBM66473.2025.11357183

***Abstract*—Identifying synergistic drug combinations is crucial for improving cancer therapies. However, the combinatorial explosion of candidate drug pairs and the complexity of drug–cell interactions make exhaustive experimental validation infeasible. Existing deep learning approaches often rely solely on two-dimensional molecular graphs or static cell line expression profiles, limiting their ability to capture spatial molecular conformations and dynamic cellular responses. To overcome these limitations, we propose ConGraSyn, a conformation enhanced graph attention framework for drug synergy prediction. For each drug, ConGraSyn integrates descriptor-based fingerprints with molecular graphs, where atom features encode interatomic distances through a scale attention mechanism. A gated fusion module adaptively combines fingerprint-level and graph-level embeddings into a unified drug representation. For cell lines, static protein–protein interaction–informed embeddings are augmented with drug-induced perturbations simulated from basal expression and drug embeddings. The resulting drug and cell features are concatenated and fed into a classifier to predict synergistic outcomes. Across benchmark datasets, ConGraSyn matches or exceeds representative deep learning baselines and consistently outperforms classical machine learning methods. Ablation and sensitivity analyses attribute the gains to its conformation-aware drug modeling and perturbation-informed cell representations. Leave-one-out validation, representation visualizations and case studies indicate robust generalization and biologically plausible novel synergies, supporting its utility for identifying effective drug combinations. The complete source code and datasets are publicly available at: https://github.com/HuazeLoong/ConGraSyn.**

***Index Terms*—Drug synergy, Graph neural networks, Molecular conformation, Multi-omics integration**

I. Introduction

Drug combination therapy has emerged as one of the fundamental strategies in modern oncology, aiming to enhance therapeutic efficacy, mitigate adverse effects, and delay drug resistance by leveraging the synergistic effects of multiple agents. Despite its promise, identifying clinically effective synergistic drug combinations remains a formidable challenge due to the combinatorial explosion of candidate space and the highly complex, often nonlinear interactions between drugs and cellular systems.

Early approaches grounded in systems biology, such as network propagation and pathway simulation, attempted to integrate drug–target information with transcriptomic profiles. While these methods offered valuable mechanistic insights, they depended heavily on predefined biological knowledge and often generalized poorly across diverse cell types or previously unseen compounds. These limitations have prompted a shift toward more flexible, data-driven strategies that learn directly from large-scale drug–cell interaction data. In this context, deep learning (DL) models have increasingly emerged as powerful alternatives for drug synergy prediction. Representative examples include DTSyn [1], which employs dual-transformer modules to capture hierarchical associations; and MMSyn [2], which integrates structural graphs, molecular fingerprints, and the Simplified Molecular Input Line Entry System (SMILES) [3] based encodings as drug features, while representing cancer cell lines with gene expression profiles, DNA copy number variations, and pathway activities.

Despite their progress, two key limitations persist. First, most models represent molecular structures as two-dimensional (2D) graphs or predefined fingerprints, overlooking three-dimensional (3D) geometries, which encode essential spatial information. Second, cellular context is often simplified as static expression profiles, which fail to capture the dynamic transcriptomic responses induced by drug treatment. These simplifications restrict the model's ability to represent spatial conformations of molecules and the dynamic nature of

*Corresponding author.

drug–cell interactions.

To address these challenges, we propose ConGraSyn, a conformation enhanced graph attention framework for synergistic drug combination prediction (Fig. 1). First, to enhance the spatial fidelity of molecular representations, we explicitly embed 3D geometric information—such as atomic coordinates and interatomic distances—into 2D molecular graphs. This integration preserves the chemical connectivity while enriching the graph with spatial context, thereby enabling the model to better capture molecular conformations. On top of this enriched representation, we introduce a multiscale conformational learning (MCL) [4] module within an improved Graph Transformer. By applying distance-aware attention masks across multiple spatial thresholds, the MCL module adaptively captures both local and global structural semantics. In addition, this spatially enriched representation is fused with molecular fingerprints to incorporate complementary global structural features. Second, to enrich the cellular context beyond static expression profiles, we leverage a pretrained TranSiGen model [5] to generate drug-induced transcriptomic features, which are then fused integrated with protein–protein interaction (PPI)-informed omics data to capture the dynamic cellular responses to drug treatment. This design enables a more faithful integration of molecular geometry and cellular dynamics for drug–cell interaction modeling. This study makes the following contributions:

- We integrate 3D geometry into 2D molecular graphs and use multiscale distance-aware attention to model local and global conformations, which are further gated with molecular fingerprints for complementary structural information.
- We integrate PPI-informed static gene expression profiles with drug-induced perturbation signatures to construct a more comprehensive cell line representation, enabling the model to better characterize drug–cell interactions.
- Extensive evaluations demonstrate that ConGraSyn achieves competitive performance, while ablation and case studies validate the contributions of conformation-aware drug modeling and perturbation-informed cell representations for identifying plausible synergistic therapies.

II. Related Works

Deep learning has been widely applied to synergistic drug combination prediction, leveraging the growing availability of chemical, genomic, and pharmacological data. Early models such as DeepSynergy [6] employed fully connected neural networks to integrate chemical descriptors of drugs with gene expression profiles of cell lines. AuDNNsynergy [7] extended this line of work by incorporating multi-omics features—including gene expression, mutations, and copy number variations—via autoencoders. However, it continued to rely on static, low-resolution drug descriptors.

Graph-based methods introduced structural awareness into drug representation. DeepDDS [8] applied graph neural networks to molecular graphs to learn substructure-aware embeddings, which were then fused with gene expression profiles for synergy prediction. To improve biological relevance, GraphSynergy [9] embedded PPI networks to represent both drug targets and disease-associated proteins, using graph convolutional networks (GCNs) and attention modules to identify key regulatory components. Similarly, PRODeepSyn [10] integrated PPI networks with cell line omics features through GCNs to generate biologically informed cell representations.

Transformer-based models have further improved the capacity to model hierarchical drug–gene–cell interactions. TranSynergy [11] applied self-attention to drug–target and gene dependency data, and used Shapley-based pathway enrichment analysis to enhance interpretability. Multimodal fusion methods aim to capture complementary biological signals by integrating diverse data sources. DeepTraSynergy [12] incorporated drug–target interactions, PPI networks, and cell–target associations into a Transformer-based framework, trained with multitask objectives including synergy, toxicity, and binding affinity. SynergyX [13] jointly modeled molecular graphs, fingerprints, SMILES strings, and pathway-level omics features through attention-based integration.

III. Method

A. Pipeline of ConGraSyn

Each input sample is a triplet $(\text{drug}_1, \text{drug}_2, \text{cell})$ with a binary label indicating whether the drug pair exhibits synergistic activity. ConGraSyn predicts drug combinations synergy by jointly modeling chemical structures, spatial conformations, and the transcriptomic context of the target cell line. For each drug, we derive a unified representation by integrating descriptor-based molecular fingerprints (MACCS [14], PubChem [15], and ErG [16]) with graph-based embeddings. Fingerprints are projected through a multilayer perceptron (MLP) to obtain $\mathbf{h}_{\text{fp}}$, while spatially enriched molecular graphs are constructed from SMILES and encoded using a multiscale distance-aware attention mechanism to yield $\mathbf{h}_{\text{str}}$. A gated fusion module adaptively combines these two representations into the final drug embedding $\mathbf{h}_{\text{drug}}$. For cell lines, gene expression and mutation profiles are independently propagated using graph attention networks (GATs) [17] on the PPI network topology, and their outputs are subsequently fused via attention pooling to produce the static cell embedding $\mathbf{h}_{\text{sta}}$. Drug-specific perturbation responses $\mathbf{h}_{\text{per}}$ are simulated by fusing basal expression with drug embeddings and decoding the predicted transcriptomic changes. The compressed perturbation features and static embedding are concatenated to form the final cell representation $\mathbf{h}_{\text{cell}}$. Finally, $\mathbf{h}_{\text{drug}_1}$, $\mathbf{h}_{\text{drug}_2}$, and $\mathbf{h}_{\text{cell}}$ are concatenated and fed into a binary classifier to predict the probability of drug synergy, optimized with binary cross-entropy loss.

B. Drug Representation Learning

To comprehensively capture the chemical, structural, and spatial characteristics of drug molecules, we integrate both descriptor-based molecular fingerprints and conformation-aware graph representations. Each drug is first represented

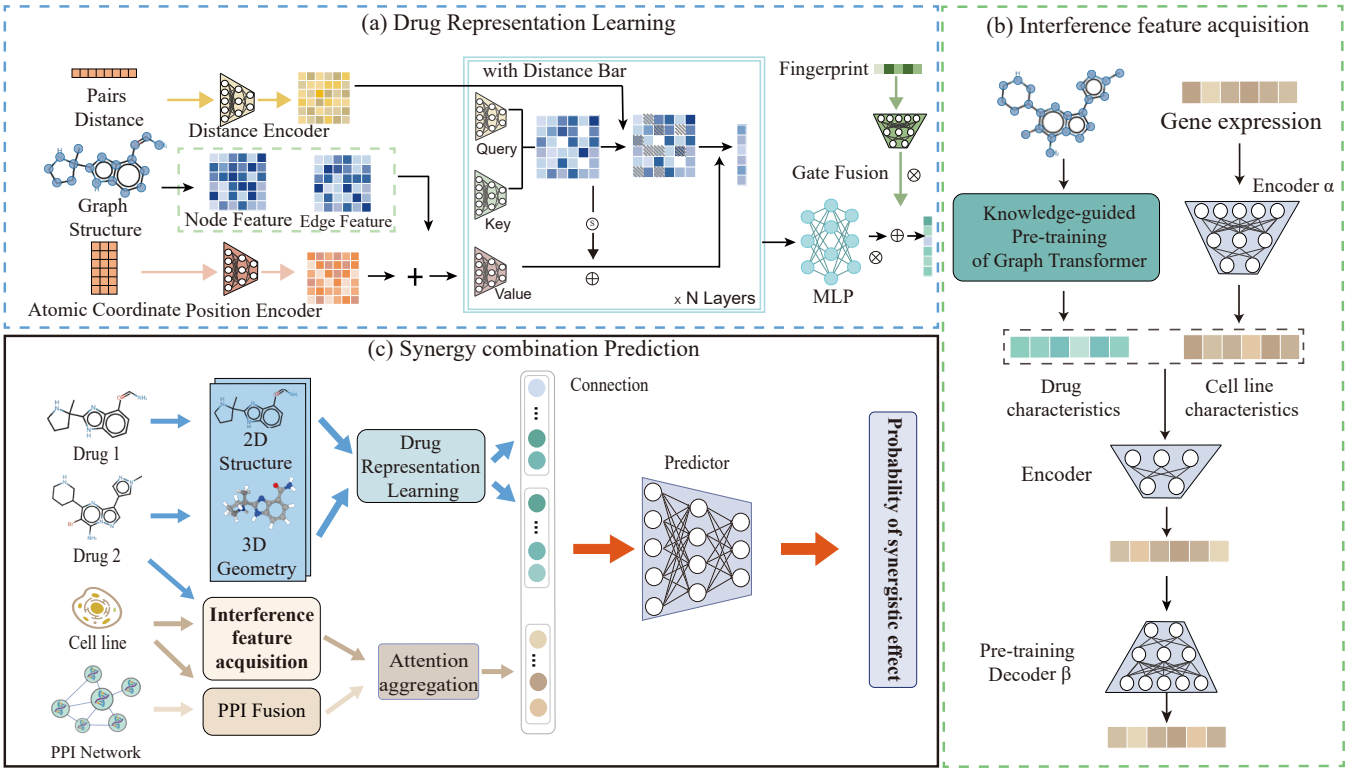

Fig. 1. The proposed ConGraSyn architecture. **a** ConGraSyn refines drug representations by embedding 3D molecular geometry, including atomic coordinates and interatomic distances, into 2D molecular graphs. A multiscale distance-aware attention mechanism is applied within a graph Transformer to capture both local and global conformational patterns. Drug fingerprints are further fused with graph-level embeddings through a gated fusion module to obtain unified drug representations. **b** For cell lines, PPI-informed static embeddings are combined with drug-induced perturbation features generated from a pretrained expression decoder. Attention-based pooling aggregates gene-level embeddings into cell-level representations. **c** The final drug and cell features are concatenated and fed into a predictor to estimate the probability of synergistic effects.

as a fingerprint vector $\mathbf{f}_{\text{fp}} \in \mathbb{R}^{1489}$, constructed by concatenating MACCS keys (166D), PubChem substructure fingerprints (881D), and ErG pharmacophore descriptors (442D). To reduce redundancy and enhance nonlinear representation capacity, the fingerprint is transformed by a three-layer MLP with ReLU activation and dropout:

$$\mathbf{h}_{\text{fp}} = \text{MLP}_{\text{fp}}(\mathbf{f}_{\text{fp}}). \tag{1}$$

We parse the SMILES string of each drug into a molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ using RDKit, where nodes $i \in \mathcal{V}$ represent atoms and edges denote chemical bonds. For each atom i, its 2D feature vector $\mathbf{h}_i^{\text{2D}}$ is extracted using RDKit, including categorical features (e.g., atom type, hybridization state, aromaticity, and ring membership, encoded as one-hot vectors) and continuous properties (e.g., partial charge). To encode spatial information, 3D coordinates $\mathbf{p}_i \in \mathbb{R}^3$ are generated via RDKit's ETKDG conformer algorithm and embedded by a learnable network:

$$\mathbf{h}_i = \mathbf{h}_i^{\text{2D}} + \text{MLP}_{\text{coord}}(\mathbf{p}_i). \tag{2}$$

To model multi-scale spatial neighborhoods, we define a set of K distance thresholds $\{\delta_1, \delta_2, \ldots, \delta_K\}$ to build binary attention masks $\mathbf{M}^{(k)} \in \{0,1\}^{N\times N}$ for atoms within the same molecule, where $M_{ij}^{(k)} = 1$, if $d_{ij} = \|\mathbf{p}_i - \mathbf{p}_j\|_2 < \delta_k$, and i, j denote atom indices within the molecular graph.

Given atom features $\mathbf{h}_i$, we compute multi-head attention for each head $l \in \{1, \ldots, H\}$ and scale $k \in \{1, \ldots, K\}$:

$$\mathbf{q}_i^{(l)} = \mathbf{W}_Q^{(l)}\mathbf{h}_i, \quad \mathbf{k}_j^{(l)} = \mathbf{W}_K^{(l)}\mathbf{h}_j, \quad \mathbf{v}_j^{(l)} = \mathbf{W}_V^{(l)}\mathbf{h}_j, \tag{3}$$

$$e_{ij}^{(l,k)} = \frac{(\mathbf{q}_i^{(l)})^\top \mathbf{k}_j^{(l)}}{\sqrt{d_k}} + \text{MLP}_{\text{bias}}(d_{ij}), \tag{4}$$

where d_k denotes the dimensionality of the query or key vectors in each head. The function $\text{MLP}_{\text{bias}}(\cdot)$ is a two-layer ReLU network that generates a distance-aware bias.

Masked attention scores are defined as:

$$\tilde{\alpha}_{ij}^{(l,k)} = \begin{cases} e_{ij}^{(l,k)}, & \text{if } M_{ij}^{(k)} = 1 \\ -\infty, & \text{otherwise} \end{cases} \tag{5}$$

and normalized by:

$$A_{ij}^{(l,k)} = \text{Softmax}_j(\tilde{\alpha}_{ij}^{(l,k)}). \tag{6}$$

The head-wise message passing is then:

$$\mathbf{m}_i^{(l,k)} = \sum_j A_{ij}^{(l,k)} \cdot \mathbf{v}_j^{(l)}. \tag{7}$$

Messages from all heads and scales are concatenated in a multi-scale manner:

$$\mathbf{m}_i = \text{Concat}_{l=1}^{H}\left(\text{Concat}_{k=1}^{K}\mathbf{m}_i^{(l,k)}\right), \tag{8}$$

$$\mathbf{z}_i = \text{LayerNorm}(\mathbf{h}_i + \text{MLP}_{\text{out}}(\mathbf{m}_i)), \tag{9}$$

where $\mathrm{MLP}_{\text{out}}$ is a two-layer MLP with GELU activation.

To obtain a fixed-size molecular embedding, we apply global mean pooling over the atoms:

$$\mathbf{h}_{\text{str}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i. \tag{10}$$

Finally, we integrate fingerprint and structure embeddings via a learnable gate:

$$\alpha = \sigma\left(\mathrm{MLP}_{\text{gate}}\left([\mathbf{h}_{\text{fp}};\mathbf{h}_{\text{str}}]\right)\right), \tag{11}$$

$$\mathbf{h}_{\text{drug}} = \alpha\cdot\mathbf{h}_{\text{fp}} + (1-\alpha)\cdot\mathbf{h}_{\text{str}}, \tag{12}$$

where $\sigma(\cdot)$ is the sigmoid activation function and $[\cdot;\cdot]$ denotes concatenation. This fusion enables adaptive balancing between predefined molecular descriptors and spatially informed learned features.

C. Cell Line Representation Learning

To capture the intrinsic molecular state of each cell line, we integrate both its gene expression profile and mutation profile with the PPI network. Let $\mathbf{H}_{\text{exp}} \in \mathbb{R}^{N_p\times F_{\text{exp}}}$ and $\mathbf{H}_{\text{mut}} \in \mathbb{R}^{N_p\times F_{\text{mut}}}$ denote the initial feature matrices for N_p proteins, where F_{exp} and F_{mut} represent the feature dimensions of expression and mutation annotations, respectively.

We refine these two modalities separately using a graph attention network over the PPI adjacency matrix $\mathbf{A}_{\text{ppi}}$:

$$\mathbf{H}'_m = \mathrm{GAT}(\mathbf{H}_m, \mathbf{A}_{\text{ppi}}), \quad m\in\{\text{exp},\text{mut}\}, \tag{13}$$

where $\mathbf{H}'_m \in \mathbb{R}^{N_p\times d}$ denotes the interaction-aware gene embeddings for modality m, and d denotes the hidden dimension of the cell line representation space.

We introduce a trainable identifier matrix $\mathbf{C} \in \mathbb{R}^{N_c\times d}$, where each row $\mathbf{c}_i$ encodes the identity of cell line i. Attention-based pooling is applied over all genes to obtain a cell-level embedding:

$$\alpha_{ij}^{(m)} = \mathrm{softmax}_j\left(\mathbf{h}_{g_j}'^{(m)\top}\mathbf{W}_c^{(m)}\mathbf{c}_i\right), \tag{14}$$

$$\mathbf{h}_c^{(i,m)} = \sum_{j=1}^{N_p}\alpha_{ij}^{(m)}\,\mathbf{W}_h^{(m)}\mathbf{h}_{g_j}'^{(m)}, \tag{15}$$

where $\mathbf{W}_c^{(m)}, \mathbf{W}_h^{(m)} \in \mathbb{R}^{d\times d}$ are modality-specific learnable projection matrices. Here, $\mathbf{h}_{g_j}'^{(m)} \in \mathbb{R}^d$ denotes the j-th row of $\mathbf{H}'_m$, i.e., the interaction-aware embedding of gene j for modality m.

Finally, we concatenate the two modality-specific embeddings to form the static cell representation:

$$\mathbf{h}_{sta}^{(i)} = [\,\mathbf{h}_c^{(i,\text{exp})};\ \mathbf{h}_c^{(i,\text{mut})}\,]. \tag{16}$$

The resulting $\mathbf{h}_{sta}^{(i)}$ captures both expression and mutation characteristics of cell line i, informed by the global PPI network.

To model the dynamic transcriptional response of a cell line to chemical compounds, we simulate compound-specific perturbations based on molecular structure and basal expression. First, each compound is encoded into a latent vector $\mathbf{z}_{\text{drug}}$ using a pretrained molecular encoder KPGT [18]. Meanwhile, the control (untreated) gene expression profile $X_\alpha \in \mathbb{R}^{N_p}$ of the target cell line is transformed into a latent state via an expression encoder:

$$\mathbf{z}_{\text{base}} = \mathrm{Encoder}_{\text{expr}}(X_\alpha) \tag{17}$$

A predictive module TranSiGen, pretrained on large-scale drug–cell line perturbation data, is then used to estimate the latent perturbed state under compound influence:

$$\tilde{\mathbf{z}}_{\text{pert}} = \mathrm{MLP}_{\text{fusion}}([\mathbf{z}_{\text{base}};\mathbf{z}_{\text{drug}}])\,. \tag{18}$$

This latent state is subsequently fed into the pretrained decoder of TranSiGen, which has learned to reconstruct drug-induced gene expression profiles from latent perturbation embeddings:

$$X'_\beta = \mathrm{Decoder}_{\text{expr}}(\tilde{\mathbf{z}}_{\text{pert}}), \tag{19}$$

where X'_β denotes the predicted expression profile after drug treatment.

To avoid computational burden and overfitting, we compress the raw perturbation signature:

$$\mathbf{h}_{\text{per}} = \mathrm{MLP}_{\text{compress}}(X'_\beta - X_\alpha). \tag{20}$$

Given a drug pair (a,b), we compute two drug-specific perturbation vectors $\mathbf{h}_{\text{per}}^{(a)}$ and $\mathbf{h}_{\text{per}}^{(b)}$. These are concatenated with the static embedding to construct the final cell line representation:

$$\mathbf{h}_{\text{cell}} = \mathrm{MLP}_{\text{proj}}\left([\mathbf{h}_{sta};\ \mathbf{h}_{\text{per}}^{(a)};\ \mathbf{h}_{\text{per}}^{(b)}]\right). \tag{21}$$

D. Synergy Prediction and Optimization

Let $\mathbf{h}_{\text{drug}_1}$ and $\mathbf{h}_{\text{drug}_2}$ denote the final feature embeddings of the two drugs obtained via gated integration of molecular fingerprints and graph representations. Let $\mathbf{h}_{\text{cell}}$ represent the cell line embedding that combines static cell line expression features and drug-induced perturbation signals.

These three modality-specific representations are concatenated and passed through a MLP classifier MLP_{pred} to estimate the probability of synergy:

$$\hat{y} = MLP_{\text{pred}}([\mathbf{h}_{\text{drug}_1};\ \mathbf{h}_{\text{drug}_2};\ \mathbf{h}_{\text{cell}}]) \tag{22}$$

The model is trained using binary cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{23}$$

where N is the number of training samples, $y_i \in \{0,1\}$ denotes the ground-truth label indicating whether the i-th triplet exhibits synergy, and $\hat{y}_i$ is the predicted probability from the model.

IV. Experiment

A. Datasets

We evaluate our model on the O'Neil drug combination dataset [19], which contains 12,415 labeled triplets spanning 36 drugs and 31 cancer cell lines. Each triplet consists of a drug pair and a target cell line, annotated with a binary synergy label derived by thresholding the original synergy scores. Canonical molecular SMILES are obtained from DrugBank [20], while cell line representations integrate transcriptomic data from the Cancer Cell Line Encyclopedia (CCLE) [21] and ArrayExpress [22], gene mutations from COSMIC [23], and PPI structure from STRING [24], enabling multimodal cell context representation.

B. Baselines and Evaluation Metrics

To assess the predictive performance of our model, We compared our approach with a broad range of baselines, including three traditional machine learning algorithms—Support Vector Machine (SVM), XGBoost and Random Forest (RF), as well as six DL models selected as widely used synergy prediction baselines.

- **DeepSynergy** [6]: integrates molecular descriptors and gene expression using fully connected layers.
- **TranSynergy** [11]: employs self-attention to model drug–target interactions and gene dependencies.
- **DeepDDS** [8]: utilizes a graph neural network with attention to identify synergistic drug combinations for cancer cells.
- **DFFNDDS** [25]: uses a fine-tuned pretrained language model with dual-feature fusion to predict drug synergy.
- **MFSynDCP** [26]: extracts adaptive graph attention and multi-source feature integration to predict synergistic anti-tumor drug effects.
- **MultiSyn** [27]: incorporates pharmacophore-level drug structures and PPI-guided multi-omics features for synergy prediction.

Baseline results were reproduced under the same preprocessing pipeline and five-fold cross-validation (CV) setting for a fair comparison. We evaluated all models using seven commonly adopted metrics: the area under the receiver operating characteristic curve (AUROC), area under the precision–recall curve (AUPR), true positive rate (TPR), precision (PREC), accuracy (ACC), balanced accuracy (BACC), and Cohen's kappa coefficient (KAPPA).

C. Performance Comparison

To evaluate the predictive performance of ConGraSyn, we conducted five-fold CV on the benchmark dataset, randomly partitioned into five approximately equal-sized folds. In each iteration, one fold was used for testing while the remaining four served for training. Table I summarizes the comparative results, which are compiled in our previous work MultiSyn [27]. Our previous framework MultiSyn already achieved competitive performance, ranking among the top models across all metrics. ConGraSyn maintains this high accuracy, achieving AUROC (0.95), AUPR (0.95), ACC (0.88), and BACC (0.88), while slightly improving TPR (0.89) and showing lower variance across folds. Notably, it attains a KAPPA of 0.75, indicating robust agreement with the ground truth. These results suggest that ConGraSyn preserves the strong predictive capability of MultiSyn while offering more stable generalization, benefiting from its conformation-aware drug modeling and multiscale attention mechanisms.

D. Ablation Study

To evaluate the contribution of each core component in the ConGraSyn framework, we conducted a series of ablation experiments. Specifically, we examined the impact of different architectural configurations on model performance across same seven evaluation metrics, with the results shown in Table II. We first assessed the role of drug feature fusion. When using only molecular fingerprints (ConGraSyn-FP) or only 2D graph-based structure embeddings (ConGraSyn-2D), performance dropped noticeably compared with the full model, with AUROC declining from 0.95 to 0.92 and KAPPA from 0.75 to 0.70. This indicates that either representation alone is insufficient to fully characterize drug properties. Combining fingerprints and 2D graph embeddings (ConGraSyn-FP2D) slightly improved AUROC to 0.93 but still underperformed the complete model, confirming the added value of 3D conformation-aware features for capturing spatial structural information. We then investigated the importance of drug-induced cell perturbation signatures. Excluding differential expression features and using only the static expression embedding derived from PPI-aware attention pooling (ConGraSyn-Static) led to a performance decrease, suggesting that dynamic, drug-specific cell responses provide complementary information beyond static cell features. Overall, the complete ConGraSyn model achieved the best performance across all metrics, underscoring the importance of both multimodal drug encoding and multi-source cell line representation.

E. Parameter Sensitivity Analysis

To assess how hyperparameters affect the performance of ConGraSyn, we conducted a sensitivity analysis using AUROC as the primary evaluation metric. Figure 2 illustrates the overall trends, with the upper panel showing the effect of embedding dimensionality and the lower panel showing the effect of learning rate. For the graph embedding dimensionality, performance improved steadily as the dimension increased from lower to moderate values, reaching its highest point at an intermediate setting. However, further enlarging the embedding dimension led to a noticeable decline, indicating that excessively large representations introduce redundancy and reduce generalization. This suggests that there is an optimal range where the embedding size balances structural expressiveness and model complexity. Similarly, for the learning rate, extremely high values caused divergence while excessively small values slowed convergence and slightly reduced performance. Moderate learning rates within an intermediate range achieved

TABLE I
FIVE-FOLD CROSS-VALIDATION RESULTS ON THE BENCHMARK DATASET. BEST VALUES ARE HIGHLIGHTED IN **BOLD**, AND SECOND-BEST ARE UNDERLINED.

Models	AUROC	AUPR	TPR	PREC	ACC	BACC	KAPPA
SVM	$0.58_{(0.01)}$	$0.56_{(0.02)}$	$0.51_{(0.12)}$	$0.54_{(0.01)}$	$0.54_{(0.01)}$	$0.54_{(0.01)}$	$0.08_{(0.04)}$
RF	$0.86_{(0.02)}$	$0.85_{(0.02)}$	$0.74_{(0.01)}$	$0.78_{(0.02)}$	$0.77_{(0.01)}$	$0.77_{(0.01)}$	$0.55_{(0.04)}$
DeepSynergy	$0.88_{(0.01)}$	$0.87_{(0.01)}$	$0.75_{(0.01)}$	$0.81_{(0.01)}$	$0.80_{(0.01)}$	$0.80_{(0.01)}$	$0.59_{(0.05)}$
TranSynergy	$0.90_{(0.01)}$	$0.89_{(0.01)}$	$0.80_{(0.01)}$	$0.84_{(0.01)}$	$0.83_{(0.01)}$	$0.83_{(0.01)}$	$0.64_{(0.01)}$
XGBoost	$0.92_{(0.01)}$	$0.92_{(0.01)}$	$0.84_{(0.01)}$	$0.84_{(0.01)}$	$0.83_{(0.01)}$	$0.83_{(0.01)}$	$0.68_{(0.01)}$
MFSynDCP	$0.92_{(0.01)}$	$0.92_{(0.01)}$	$0.86_{(0.01)}$	$0.86_{(0.01)}$	$0.85_{(0.01)}$	$0.85_{(0.01)}$	$0.70_{(0.01)}$
DeepDDS-GAT	$0.93_{(0.01)}$	$0.93_{(0.01)}$	$0.85_{(0.07)}$	$0.85_{(0.07)}$	$0.85_{(0.07)}$	$0.85_{(0.07)}$	$0.71_{(0.21)}$
DFFNDDS	$0.93_{(0.01)}$	$0.92_{(0.01)}$	$0.86_{(0.03)}$	$0.85_{(0.01)}$	$0.86_{(0.01)}$	$0.86_{(0.01)}$	$0.72_{(0.03)}$
MultiSyn	$0.95_{(0.01)}$	$0.94_{(0.01)}$	$0.88_{(0.01)}$	$0.87_{(0.01)}$	$0.88_{(0.01)}$	$0.88_{(0.01)}$	$0.75_{(0.02)}$
ConGraSyn	$\mathbf{0.95_{(0.01)}}$	$\mathbf{0.95_{(0.01)}}$	$\mathbf{0.89_{(0.01)}}$	$\mathbf{0.87_{(0.01)}}$	$\mathbf{0.88_{(0.01)}}$	$\mathbf{0.88_{(0.01)}}$	$\mathbf{0.75_{(0.01)}}$

TABLE II
FIVE-FOLD CROSS-VALIDATION RESULTS FOR THE ABLATION STUDY. BEST SCORES ARE **BOLD**.

Model Variant	AUROC	AUPR	TPR	PREC	ACC	BACC	KAPPA
ConGraSyn-FP	0.92	0.92	0.85	0.84	0.85	0.85	0.70
ConGraSyn-2D	0.92	0.92	0.85	0.84	0.85	0.85	0.70
ConGraSyn-FP2D	0.93	0.93	0.85	0.86	0.85	0.85	0.71
ConGraSyn-Static	0.93	0.93	0.85	0.85	0.86	0.86	0.73
ConGraSyn	**0.95**	**0.95**	**0.89**	**0.87**	**0.88**	**0.88**	**0.75**

the most stable and optimal results, confirming that balanced optimization dynamics are crucial for robust training. Based on these trends, we selected the embedding dimensionality and learning rate corresponding to the peak performance as the default configuration for all subsequent experiments, ensuring an optimal trade-off between accuracy and computational efficiency.

F. Leave-One-Out Validation

To test our model's generalization in cold-start conditions, we performed comparisons with ten drug synergy methods across three distinct experimental setups. The leave-drug-combination-out experiment removed certain drug combinations from the training set, assessing the model's ability to predict synergy for novel drug pairs. In the leave-drug-out experiment, a particular drug was excluded to evaluate performance on previously unseen drugs. Finally, the leave-tissue-out experiment excluded data from specific cell lines, examining how well the model generalizes across biological contexts. As shown in Figure 3, ConGraSyn consistently outperforms classical machine learning methods and achieves results comparable to or better than representative deep learning baselines across all cold-start scenarios. Notably, in both the drug combination and drug settings, the model maintains high accuracy for unseen drug pairs, demonstrating its robustness in handling novel data. Furthermore, in the tissue setting, ConGraSyn effectively generalizes to different cell lines, validating

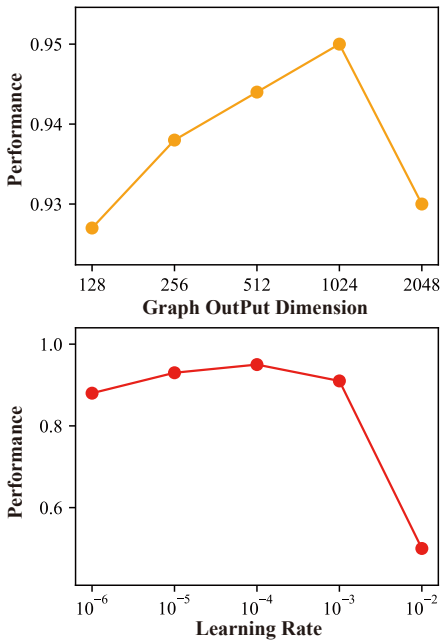

Fig. 2. Analysis of two parameters: graph output dimension and learning rate.

its adaptability to diverse biological contexts. These results highlight ConGraSyn's ability to capture complex drug–cell interactions and its strong generalization capacity.

G. Representation Visualization

To assess the representational power of the proposed ConGraSyn model, we applied the dimensionality reduction technique and visualized the feature embeddings both before and after training. We applied t-SNE to project high-dimensional feature data from both ConGraSyn and DeepDDS, a widely used baseline deep learning model that achieved second-best

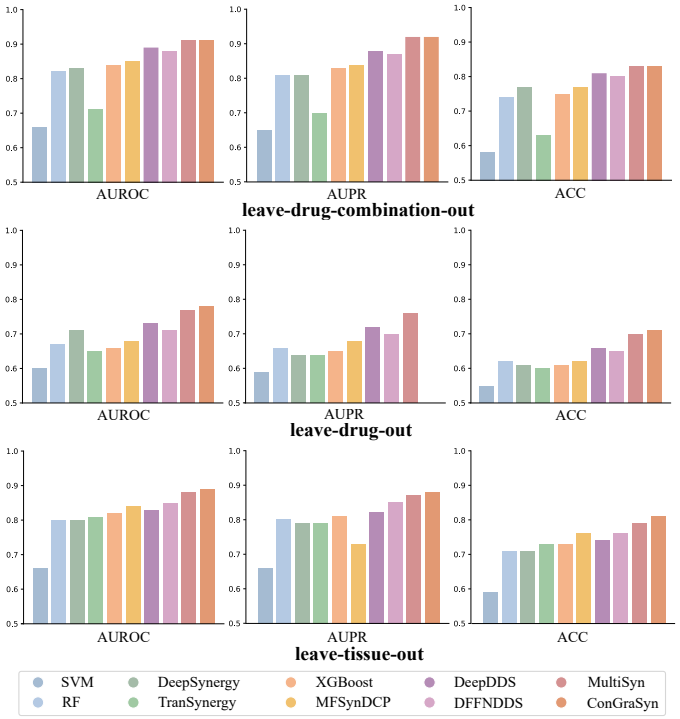

Fig. 3. Performance Evaluation by Leave-One-Out Validation.

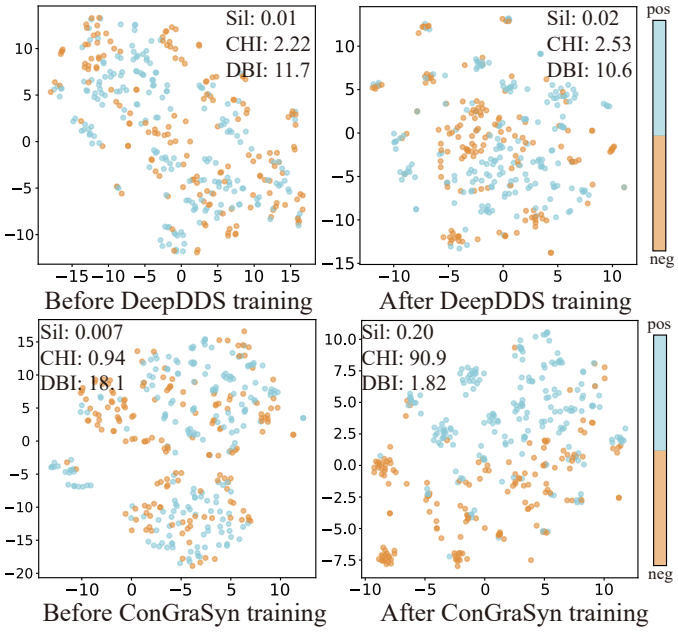

Fig. 4. Dimensionality Reduction and Representation Visualization.

performance in our experiments, into a two-dimensional space. DeepDDS serves as a common benchmark in drug synergy prediction tasks, providing a robust point of comparison. The visualizations for both models are presented in Figure 4. Before training, both models exhibited significant overlap between positive (synergistic) and negative (non-synergistic) drug pairs, indicating that the initial feature representations were not well separated. To quantify these improvements, we calculated three clustering metrics—Silhouette Score (Sil), Calinski-Harabasz Index (CHI), and Davies-Bouldin Index (DBI)—on the embeddings projected using t-SNE. After training, t-SNE visualizations revealed that ConGraSyn produced more distinct and compact clusters for both positive and negative drug pairs, demonstrating its superior ability to separate synergistic from non-synergistic combinations. Overall, the results highlight the effectiveness of the proposed ConGraSyn model in learning features that capture the latent relationships driving drug synergy.

H. Case Study

We applied ConGraSyn to identify novel drug combinations by enumerating all possible pairs in the dataset while strictly excluding those present in any training or validation folds. Based on predicted synergy scores, the top 10 combinations ranked by predicted synergy probability for the ZR751 cell line were selected, and a targeted literature survey was conducted via PubMed to assess supporting evidence. The predicted combinations and their PubMed identifiers (PMIDs) are summarized in Table III. Several predicted pairs have prior clinical support. For example, DOXORUBICIN and

TABLE III
TOP-RANKED NOVEL DRUG COMBINATIONS PREDICTED BY CONGRASYN.

Top	Drug1	Drug2	PMID
1	DOXORUBICIN	ETOPOSIDE	1984826,8435802
2	MITOMYCINE	PACLITAXEL	NA
3	DOXORUBICIN	PACLITAXEL	8905026,12586793
4	ETOPOSIDE	MK-8669	NA
5	GEMCITABINE	PACLITAXEL	25795409,16143163
6	CYCLOPHOSPHAMIDE	PACLITAXEL	10458222,9296218
7	ETOPOSIDE	PACLITAXEL	9554587,10673668
8	5-FU	PACLITAXEL	8629038
9	MK-4827	PACLITAXEL	NA
10	PACLITAXEL	VINBLASTINE	9426691

ETOPOSIDE have been evaluated in metastatic breast cancer, with response rates of 36%–40% reported in previously treated patients [28]. Similarly, carboplatin and etoposide showed an objective response rate of 42% in chemotherapy-naïve patients but limited efficacy in heavily pretreated cases [29]. ClinicalTrials.gov also lists combinations such as DOXORUBICIN and PACLITAXEL (NCT00006110), corroborating their potential clinical relevance. Overall, several of the top-ranked predicted combinations are supported by prior experimental or clinical evidence, while others may serve as hypotheses for future validation. These findings suggest that ConGraSyn can generate biologically plausible candidates that may assist in guiding experimental screening and inform potential clinical research.

V. CONCLUSION

We presented ConGraSyn, a conformation enhanced graph attention framework for predicting synergistic drug combinations by integrating 3D-enhanced drug structures and multi-source cell line features. Unlike approaches relying solely on 2D graphs or static expression profiles, ConGraSyn refines

drug representations with multi-scale distance-aware attention and models the cellular context by combining PPI-informed static embeddings with drug-induced perturbations. Evaluations on the benchmark with five-fold CV demonstrate that ConGraSyn outperforms classical baselines and achieves competitive results against state-of-the-art deep learning models. The leave-one-out validation and representation visualization demonstrate its strong generalization and effective feature learning in predicting drug synergy. Case studies further suggest the biological plausibility of several predicted novel combinations, highlighting its potential to guide experimental validation and inform rational design of combination therapies. While encouraging, ConGraSyn is currently limited by the size and oncology focus of available datasets and the underuse of other omics modalities. Future work will incorporate additional data sources, expand to broader disease contexts, and explore architectural refinements to enhance generalizability and scalability.

ACKNOWLEDGMENT

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 No. 2024AFB275), the Scientific Research Project of Education Department of Hubei Province (Grant No.Q20231109).

REFERENCES

[1] J. Hu, J. Gao, X. Fang, Z. Liu, F. Wang, W. Huang, H. Wu, and G. Zhao, "Dtsyn: A dual-transformer-based neural network to predict synergistic drug combinations," *Briefings in Bioinformatics*, vol. 23, no. 5, p. bbac302, 2022.

[2] Y. Pang, Y. Chen, M. Lin, Y. Zhang, J. Zhang, and L. Wang, "Mmsyn: A multimodal deep learning framework for enhanced prediction of synergistic drug combinations," *J. Chem. Inf. Model.*, vol. 64, no. 9, pp. 3689–3705, 2024.

[3] D. Weininger, "Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules," *J. Chem. Inf. Comput. Sci.*, vol. 28, no. 1, pp. 31–36, 1988.

[4] J. Qiao, J. Jin, D. Wang, S. Teng, J. Zhang, X. Yang, Y. Liu, Y. Wang, L. Cui, and Q. Zou, "A self-conformation-aware pre-training framework for molecular property prediction with substructure interpretability," *Nat. Commun.*, vol. 16, no. 1, p. 4382, 2025.

[5] X. Tong, N. Qu, X. Kong, S. Ni, J. Zhou, K. Wang, L. Zhang, Y. Wen, J. Shi, and S. Zhang, "Deep representation learning of chemical-induced transcriptional profile for phenotype-based drug discovery," *Nat. Commun.*, vol. 15, no. 1, p. 5378, 2024.

[6] K. Preuer, R. P. I. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, and G. Klambauer, "Deepsynergy: Predicting anti-cancer drug synergy with deep learning," *Bioinformatics*, vol. 34, no. 9, pp. 1538–1546, 2018.

[7] T. Zhang, L. Zhang, P. R. O. Payne, and F. Li, "Synergistic drug combination prediction by integrating multiomics data in deep learning models," in *Translational Bioinformatics for Therapeutic Development*. Springer, 2020, pp. 223–238.

[8] J. Wang, X. Liu, S. Shen, L. Deng, and H. Liu, "Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations," *Briefings in Bioinformatics*, vol. 23, no. 1, 2022.

[9] J. Yang, Z. Xu, W. K. K. Wu, Q. Chu, and Q. Zhang, "Graphsynergy: A network-inspired deep learning model for anticancer drug combination prediction," *J. Am. Med. Inform. Assoc.*, vol. 28, no. 11, pp. 2336–2345, 2021.

[10] X. Wang, H. Zhu, Y. Jiang, Y. Li, C. Tang, X. Chen, Y. Li, Q. Liu, and Q. Liu, "Prodeepsyn: Predicting anticancer synergistic drug combinations by embedding cell lines with protein–protein interaction network," *Briefings in Bioinformatics*, vol. 23, no. 2, p. bbab587, 2022.

[11] Q. Liu and L. Xie, "Transynergy: Mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations," *PLoS Comput. Biol.*, vol. 17, no. 2, p. e1008653, 2021.

[12] F. Rafiei, H. Zeraati, K. Abbasi, J. B. Ghasemi, M. Parsaeian, and A. Masoudi-Nejad, "Deeptrasynergy: Drug combinations using multimodal deep learning with transformers," *Bioinformatics*, vol. 39, no. 8, p. btad438, 2023.

[13] Y. Guo, H. Hu, W. Chen, H. Yin, J. Wu, C.-Y. Hsieh, Q. He, and J. Cao, "Synergyx: a multi-modality mutual attention network for interpretable drug synergy prediction," *Briefings in Bioinformatics*, vol. 25, no. 2, p. bbae015, 2024.

[14] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, "Reoptimization of mdl keys for use in drug discovery," *J. Chem. Inf. Comput. Sci.*, vol. 42, no. 6, pp. 1273–1280, 2002.

[15] E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant, "Pubchem: Integrated platform of small molecules and biological activities," in *Annu. Rep. Comput. Chem.*, 2008, vol. 4, pp. 217–241.

[16] N. Stiefl, I. A. Watson, K. Baumann, and A. Zaliani, "Erg: 2d pharmacophore descriptions for scaffold hopping," *J. Chem. Inf. Model.*, vol. 46, no. 1, pp. 208–220, 2006.

[17] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Int. Conf. Learn. Represent. (ICLR)*, 2018.

[18] H. Li, R. Zhang, Y. Min, D. Ma, D. Zhao, and J. Zeng, "A knowledge-guided pre-training framework for improving molecular representation learning," *Nat. Commun.*, vol. 14, no. 1, p. 7568, 2023.

[19] J. O'Neil, Y. Benita, I. Feldman, M. Chenard, B. Roberts, Y. Liu, J. Li, A. Kral, S. Lejnine, and A. Loboda, "An unbiased oncology compound screen to identify novel combination strategies," *Mol. Cancer Ther.*, vol. 15, no. 6, pp. 1155–1162, 2016.

[20] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, and Z. Sayeeda, "Drugbank 5.0: A major update to the drugbank database for 2018," *Nucleic Acids Res*, vol. 46, no. D1, pp. D1074–D1082, 2018. [Online]. Available: https://doi.org/10.1093/nar/gkx1037

[21] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Y. Kim, A. Wilson, J. Lehár, G. Kryukov, and D. Sonkin, "The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, 2012. [Online]. Available: https://portals.broadinstitute.org/ccle

[22] F. Iorio, T. A. Knijnenburg, D. J. Vis, G. R. Bignell, M. P. Menden, M. Schubert, N. Aben, E. Gonçalves, S. Barthorpe, and H. Lightfoot, "A landscape of pharmacogenomic interactions in cancer," *Cell*, vol. 166, no. 3, pp. 740–754, 2016.

[23] J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, and E. Dawson, "Cosmic: The catalogue of somatic mutations in cancer," *Nucleic Acids Res*, vol. 47, no. D1, pp. D941–D947, 2019.

[24] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, and P. Bork, "String v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Res*, vol. 47, no. D1, pp. D607–D613, 2019.

[25] M. Xu, X. Zhao, J. Wang, W. Feng, N. Wen, C. Wang, J. Wang, Y. Liu, and L. Zhao, "Dffndds: Prediction of synergistic drug combinations with dual feature fusion networks," *J. Cheminformatics*, vol. 15, no. 1, p. 33, 2023.

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

[27] S. Jin, H. Long, A. Huang, J. Wang, X. Yu, Z. Xu, and J. Xu, "Accurate prediction of synergistic drug combination using a multi-source information fusion framework," *BMC Biol.*, vol. 23, p. 200, 2025.

[28] G. W. S. Jr., "Etoposide in the management of metastatic breast cancer," *Cancer*, vol. 67, no. S1, pp. 266–270, 1991.

[29] J. Crown, T. Hakes, B. Reichman, D. Lebwohl, T. Gilewski, A. Surbone, V. Currie, T.-J. Yao, C. Hudis, and A. Seidman, "Phase ii trial of carboplatin and etoposide in metastatic breast cancer," *Cancer*, vol. 71, no. 4, pp. 1254–1257, 1993.