

RESEARCH

Open Access



MFSynDCP: multi-source feature collaborative interactive learning for drug combination synergy prediction

Yunyun Dong^{1*}, Yunqing Chang¹, Yuxiang Wang¹, Qixuan Han¹, Xiaoyuan Wen¹, Ziting Yang¹, Yan Zhang¹, Yan Qiang^{2*}, Kun Wu³, Xiaole Fan⁴ and Xiaoqiang Ren⁴

*Correspondence:
dongyunyun@tyut.edu.cn;
qiangyan@tyut.edu.cn

¹ School of Software, Taiyuan University of Technology, Taiyuan, Shanxi, China

² College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Taiyuan, Shanxi, China

³ School of Computing, University of Leeds, Leeds, West Yorkshire, UK

⁴ Information Management Department, Shanxi Provincial People's Hospital, Taiyuan, Shanxi, China

Abstract

Drug combination therapy is generally more effective than monotherapy in the field of cancer treatment. However, screening for effective synergistic combinations from a wide range of drug combinations is particularly important given the increase in the number of available drug classes and potential drug-drug interactions. Existing methods for predicting the synergistic effects of drug combinations primarily focus on extracting structural features of drug molecules and cell lines, but neglect the interaction mechanisms between cell lines and drug combinations. Consequently, there is a deficiency in comprehensive understanding of the synergistic effects of drug combinations. To address this issue, we propose a drug combination synergy prediction model based on multi-source feature interaction learning, named MFSynDCP, aiming to predict the synergistic effects of anti-tumor drug combinations. This model includes a graph aggregation module with an adaptive attention mechanism for learning drug interactions and a multi-source feature interaction learning controller for managing information transfer between different data sources, accommodating both drug and cell line features. Comparative studies with benchmark datasets demonstrate MFSynDCP's superiority over existing methods. Additionally, its adaptive attention mechanism graph aggregation module identifies drug chemical substructures crucial to the synergy mechanism. Overall, MFSynDCP is a robust tool for predicting synergistic drug combinations. The source code is available from GitHub at <https://github.com/kkioplkg/MFSynDCP>.

Keywords: Drug combination, Synergistic effect, Graph attention network, Anti-tumor, Deep learning

Introduction

In recent years, the field of malignant tumor biology has yielded a multitude of effective anticancer drugs. However, the inherent heterogeneity of tumors and the development of drug resistance often render single-drug therapies targeting individual markers ineffective [1]. In contrast, drug combination therapies [2–4] have shown great potential to improve efficacy. By acting on multiple targets and pathways, they can effectively



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

improve efficacy, reduce side effects, and overcome drug resistance [5, 6] However, there is a risk of antagonistic interactions or severe adverse reactions with some drug combinations, posing potential threats to patient health. Oncology, as one of the largest disease areas for global drug development, it has become particularly important to predict effective synergistic drug combinations from the huge number of anti-tumor drugs.

Traditional methods of predicting drug combinations primarily rely on numerous time-consuming and expensive clinical trials [7], which may cause patients to receive some unnecessary treatments and cause psychological or physiological harm. With the development of high-throughput drug screening technologies [8–10], researchers have accelerated the search for drug combinations with synergistic effects by using automated testing platforms and large-scale compound libraries to conduct extensive drug combination screening across hundreds of cancer cell lines. However, high-throughput drug screening methods are mainly based on *in vitro* cell models or animal models, and they ignore the complex interaction networks between drugs, biomolecules, and signaling pathways. These methods are unable to fully simulate the complexity of drug interactions in the human body [11]. Additionally, it is impractical to screen all possible drug combinations using this approach [12].

Recently, the advancement of artificial intelligence [13] and the availability of large-scale datasets have made it feasible to explore machine learning models [14–16] or deep neural networks [17] for drug combination predictions, which can reduce the cost of drug experiments while improving the prediction of synergistic effects of drug combinations. Mei [18] proposed an independent machine learning framework to simultaneously predict synergistic, antagonistic, and additive effects of drugs. This framework represents drug pairs through simple graphs of drug-targeted genes and cellular processes, thereby effectively explaining the molecular mechanisms behind drug interactions and reducing data complexity. Janizek et al. [19] utilized feature attribution methods, improving the quality of explanations by using a collection of interpretable machine learning models, and discovered hematopoietic differentiation characteristic drug combinations with therapeutic synergistic effects. Julkunen et al. [20] proposed a machine learning framework, *comboFM*, for predicting responses to drug combinations in preclinical studies. It models high-order tensor drug interactions specific to cellular contexts and uses powerful factorization machines to efficiently learn the latent factors of the tensor, predicting responses to new combinations in cells not yet tested. However, training a good machine model for the prediction of drug combination synergy tasks requires specialized domain knowledge and experience to manually select and construct relevant features [21].

The rapid development of deep learning has provided new possibilities for addressing these challenges. Deep learning models can automatically learn high-level abstract features from raw data [22], eliminating the reliance on manually designed features and better capturing complex relationships and nonlinear patterns in data. Various neural network models have been employed in the field of drug combination prediction. Preuer et al. [23] introduced a deep learning model named *DeepSynergy*, which uses chemical and genomic information as input and employs a normalization strategy to account for the heterogeneity of the input data. This was the first attempt to utilize deep learning in this field, and the performance of this model surpassed traditional machine learning

methods. Raffei et al. [24] utilized multimodal deep learning and transformer for multi-task predictions, including drug-target interactions, toxic effects, and synergistic effects of drug combinations. Yang et al. [25] proposed a deep learning model called Graph-Synergy, by adopting space-based graph convolutional network components and attention mechanisms, encodes high-order topological relationships in the Protein-Protein Interaction (PPI) network of protein modules. The approach focuses on identifying crucial proteins involved in biomolecular interactions within the PPI network, as well as interactions between drug combinations and cancer cell lines, with the aim of predicting synergistic drug combinations. However, these methods primarily focus on separately extracting features from drug molecules and cell line structures, without adequately considering the integration of cell line-drug combination pairs. This leads to a limitation in the model's ability to learn associated patterns from the data. Additionally, these methods have certain limitations in focusing on important substructures within drug molecules, which may pose challenges to the biological interpretation of the predictions.

Based on the considerations mentioned above, we propose a model for predicting the synergy of drug combinations based on multi-source feature interactive learning, named MFSynDCP. Specifically, we introduce a deep graph attention neural network to automatically learn and extract high-dimensional features of drugs. Then, we propose an adaptive attention mechanism graph aggregation module to capture the drug substructures that are most critical for predicting the synergistic effects of drug combinations. Additionally, we introduce a multi-source feature interactive learning controller, which incorporates a parameter self-learning gating structure within the controller to regulate information transfer between different data sources, thereby flexibly handling diverse features. Finally, we compared our method with recent deep learning prediction models, and the results show that our approach, MFSynDCP, has significant advantages in predicting drug combination synergy compared to recent deep learning models. Specifically, our model exhibits superior accuracy, enhanced predictive capabilities, and increased stability.

Materials and methods

Dataset

We used the large-scale tumor screening drug combination dataset published by O'Neil et al. [26] in 2016 as our benchmark dataset. This dataset involves screening of 583 different combinations of 22 experimental drugs and 16 approved drugs across 39 cancer cell lines, comprising 23,052 triplets, each consisting of two drugs and a cancer cell line. The Loewe Additivity [27, 28] scores for each pair of drugs were calculated using the Combenefit tool based on the 4×4 dose-response matrix in the dataset. The effect of each individual drug at the same dose served as a baseline, and the scores for the synergistic or antagonistic effects of drug combinations were quantitatively calculated by comparing the effect of the drug combination with the expected additive effect. According to the Loewe scores, combinations with scores greater than zero were considered synergistic, while those with scores less than zero were considered antagonistic. Considering the presence of noise, which results in synergy scores close to zero, we adopted a more stringent threshold for finer classification of these combinations. We chose 10 as the threshold to classify the drug pair-cell line triplets, considering drug combinations

with Loewe scores above 10 as synergistic and those with scores below zero as antagonistic. Ultimately, a balanced benchmark dataset was obtained, comprising 12,415 unique drug pair-cell line combinations, covering 36 anti-cancer drugs and 31 human cancer cell lines.

In this study, the SMILES (Simplified Molecular Input Line Entry System) sequence data of drugs [29] were sourced from the DrugBank database [30]. We obtained the SMILES expressions of the required drugs from the DrugBank database and used RDKit [31] to convert the SMILES sequences of the drugs into corresponding molecular graph representations. Drug compounds are viewed as graphical structures based on interactions between atoms. The transformed molecular graphs depict the overall structure of the molecules through a series of atoms and bonds, illustrating the connections and spatial arrangement of atoms. In these graphs, vertices represent the atoms in the drug structure, and edges indicate the chemical bonds connecting these atoms.

In this paper, relevant gene expression data for cancer cell lines were obtained from the Cancer Cell Line Encyclopedia (CCLE) [32]. Considering factors such as gene length and sequencing depth, the gene expression data were standardized using Transcripts Per Million (TPM) [33] to normalize expression levels. This normalization process ensures more accurate and reliable comparisons between different genes and cell lines. By standardizing gene expression data in this manner, it becomes feasible to conduct in-depth analyses across various cell lines, facilitating a better understanding of the biological characteristics and behaviors of different cancer types at the genetic level.

MFSynDCP

Figure 1 illustrates the end-to-end learning framework MFSynDCP proposed for predicting the synergistic effects of drug combinations. The model comprises five parts: a feature extraction module for drugs (Fig. 1a), a feature extraction module for cell lines (Fig. 1b), a graph aggregation module based adaptive attention mechanism (Fig. 1c), a multi-source feature interactive learning controller (Fig. 1d), and a synergy prediction module (Fig. 1e). The process begins with the transformation of drug SMILES strings into molecular structure graphs using RDKit. The input layer receives the molecular structure graphs of two drugs, as well as the gene expression profiles of the cell lines affected by these drugs. A Graph Attention Network (GAT) extracts features from the nodes and edges of the drug molecular graphs. The design includes an adaptive attention mechanism graph aggregation module, which dynamically focuses on key information within the drug pair and comprehensively captures important interaction features between the drugs. A Multi-Layer Perceptron (MLP) encodes the genomic features of cancer cells, utilizing nonlinear transformations and mappings to capture and extract potential gene expression information. To fully consider the intrinsic correlation and interaction between them, their feature vectors are concatenated and processed through a multi-source feature interactive learning controller. This controller efficiently handles the concatenated feature vectors, delving into and conveying deeper-level features, ensuring the smooth integration of multi-source heterogeneous data. Finally, the processed integrated features are passed through a linear layer to output the predicted synergy scores of the drug pairs. Based on predetermined thresholds, the model determines whether the drug combination has a synergistic or antagonistic effect.

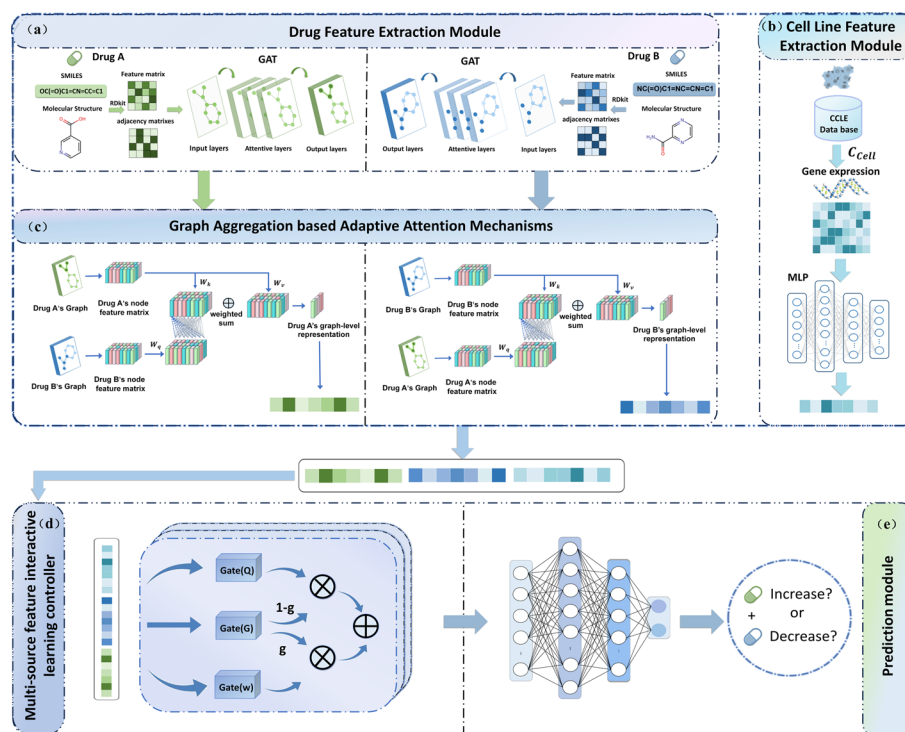


Fig. 1 The workflows of the MFSynDCP model framework process. Drug molecular graphs are generated based on the SMILES sequences of drugs, and their feature embeddings are obtained using a GAT. Additionally, feature embeddings of cancer cell line gene expression profiles are acquired using a MLP. The embedding vectors of the drugs and cell lines are then concatenated and input into a multi-source feature interactive learning controller for the fusion of multi-source features. Finally, the fused features are fed into the prediction module for predicting the synergistic effects of drug combinations

Drug feature extraction based on GAT

We use the software RDkit to convert the drug's SMILES string into a molecular graph, where nodes are atoms and edges are chemical bonds between atoms. The drug graph is defined as $G = (V, E)$, with V being a set of N nodes each represented by a d -dimensional vector, and E as a set of edges represented by the adjacency matrix A of the drug molecule's topological graph. $x_i \in V$ represents an atom, and $e_{ij} \in E$ represents a chemical bond between atoms. DeepChem [34], a cheminformatics software package that provides tools and algorithms for processing and analyzing chemical molecule data, is used to calculate atomic properties in each node of the drug molecular graph as initial features. Each atom x_i is represented as a vector $[x_{i1}, x_{i2}, \dots, x_{i5}]$, where the elements of the vector correspond respectively to the atomic symbol, the number of adjacent atoms, the atom's implicit valence, the count of adjacent hydrogen atoms, and the atom's inclusion in a benzene ring structure.

When dealing with graph structures, traditional CNN models experience a significant decrease in performance in Euclidean space. In the task of predicting drug combinations, extracting feature representations of drugs from chemical molecular graphs is necessary, but traditional convolutional networks are not up to the task. Given that Graph Convolutional Networks (GCN), typically used for processing graph structures, treat each neighboring node as equally important and fail to capture

the varying significance between nodes, we have adopted the GAT as the primary model for extracting drug molecular structure. By introducing graph attention layers in its architecture and utilizing a multi-head self-attention mechanism, the GAT learns advanced features of nodes in the graph. It dynamically allocates attention weights based on the relationships between nodes and their neighbors, thereby more accurately capturing the differences in importance between nodes and enhancing the representation capability of drug features.

For each vertex atom i in the transformed drug molecular graph, the correlation coefficient between each neighboring atom $j \in N_i$ and the atom i itself is calculated individually.

The process of calculating the correlation coefficient as follows:

$$e_{ij} = a([Wh_i || Wh_j]), j \in N_i \quad (1)$$

where $W \in \mathbb{R}^{K' \times K}$ represents a weight matrix, and the attention mechanism a is a single-layer feedforward neural network, parameterized by the weight vector $\vec{a}^T \in \mathbb{R}^{2K'}$. The feature vector h_i corresponds to the features of node i . The function $||$ concatenates the transformed features of atoms i and j . Normalize the attention coefficients using the softmax function:

$$\alpha_{i,j} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{z \in N_i} \exp(e_{iz})} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [W \vec{h}_i || W \vec{h}_j]))}{\sum_{z \in N_i} \exp(\text{LeakyReLU}(\vec{a}^T [W \vec{h}_i || W \vec{h}_z]))} \quad (2)$$

Based on the calculated attention coefficients $\alpha_{i,j}$, the features are weighted and summed up, and then processed through the activation function σ_d . This process results in new features for each atom node i after integrating the information from its neighboring atoms, using the multi-head attention mechanism, and the formula is as follows:

$$h'_i = ||_{m=1}^M \sigma_d \left(\sum_{j \in N_i} \alpha_{ij}^m W^m h_j \right) \quad (3)$$

The proposed method for processing drug features allows for the simultaneous consideration of various aspects of feature information from neighboring atomic nodes. It dynamically allocates weights based on the relationships between nodes, focusing more on drug structures that play a key role in the synergistic interaction of drug-drug combinations in specific cell lines. This approach captures the interactions between different atomic nodes in the drug molecular graph, enhancing the representation capability of drug features. This improved representation is crucial during the feature fusion process, as it allows for a more accurate consideration of the contributions of different neighboring nodes in the drug molecular graph. Consequently, this enhances the accuracy and reliability of drug synergy prediction tasks.

Adaptive attention mechanism graph aggregation module (AAGAM)

To better learn the interaction information between drug pairs and gain deeper insight into the impact mechanisms of drug structures on cancer cell genomes, we propose a graph aggregation module based on an adaptive attention mechanism. This module is designed

to identify which drug substructures are more crucial for predicting the synergistic effects of drug pairs. By assigning attention scores to each substructure and performing a weighted summation of embedding vectors, the module reveals the molecular mechanisms underlying drug combinations' resistance or sensitivity responses in cancer treatment. These insights deepen our understanding of drug interactions and enable more accurate predictions of synergistic effects, we are able to not only extract interactive information between drug pairs but also identify the significant chemical substructures within drugs. As illustrated in Fig. 2, an attention score is assigned to each substructure of a drug. The module performs a weighted summation of the embedding vectors of all nodes, thereby obtaining an aggregated representation of the drug in graph form. This process allows for the extraction of interaction information between drug pairs and reveals the molecular mechanisms behind drug combinations' resistance or sensitivity responses in cancer treatment. The attention scores for the drug pairs are calculated using specific formulas as follows:

$$S_A = \text{softmax} \left(\sum_{j=1}^M \tanh(E_{Ai}^1 W_k (E_B^1 W_q)^T), j \right) \tag{4}$$

$$S_B = \text{softmax} \left(\sum_{j=1}^N \tanh(E_{Bi}^1 W_k (E_A^1 W_q)^T), j \right) \tag{5}$$

$$g_x = \text{multiply}(S_A, E_{Ai} W_v) \tag{6}$$

$$g_y = \text{multiply}(S_B, E_{Bi} W_v) \tag{7}$$

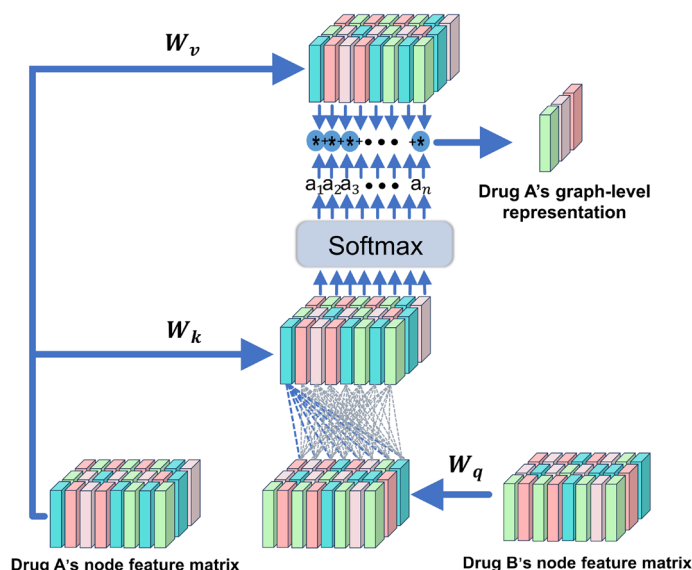


Fig. 2 The calculation steps for the adaptive attention mechanism in graph aggregation. Taking drug A as an example, it involves a weighted summation of the embeddings of all nodes to compute the final graph-level representation of drug A

$$G_x = \text{norm}(\text{mean}(E_{Ai}) + g_x) \quad (8)$$

$$G_y = \text{norm}(\text{mean}(E_{Bi}) + g_y) \quad (9)$$

where E_{Ai}^l and E_{Bi}^l represent the graph embedding matrices of drugs A and B from the last layer of the GAT, respectively. M and N are the number of nodes in drugs A and B, respectively. W_k and W_v are the feature matrices learned by drug A through two linear layers, while W_q is the feature matrix learned by drug B through a linear layer. Attention scores are used to assign weights to each node vector, with S_A and S_B being the calculated attention scores for drugs A and B, respectively. g_x represents the weighted attention scores for W_v for drug A. G_x is the normalization of the sum of the average of E_{Ai}^l along the first dimension and g_x , yielding the final graph-level representation. Similarly, G_y is the final graph-level representation for drug B, obtained by normalizing the sum of the average of E_{Bi}^l along the first dimension and g_y .

Employing a MLP for cell line feature extraction

The CCLE gene expression profile we selected contains a wealth of gene information, making it challenging to construct a model that predicts synergy due to the high dimensionality of the feature space. To address the dimensional disparity between drug and cell line feature vectors, we turned to the ‘‘Landmark gene set’’ provided by the LINCS project [35]. Subsequently, our focus shifted towards identifying genes that overlapped between the landmark gene set and the CCLE gene expression profile for further exploration. Gene annotation information from the CCLE and the GENCODE [36] annotation database was utilized to remove redundant data and transcripts of non-coding RNA. In the end, a total of 954 genes were chosen from the initial expression profile to serve as input for the model.

For the collected gene expression features of cell line X_C , redundant gene data is removed using gene annotation information from the CCLE and GENCODE databases, ensuring the accuracy and reliability of the gene data. This results in a cell line feature matrix $C \in \mathbb{R}^{S \times U}$, where S is the number of cell lines and U is the dimension of features for each cell line. The latent features $v \in \mathbb{R}^V$ of cell line C are captured through a q_c -layer fully connected neural network, and the formula is as follows:

$$v = \tanh\left(\dots \tanh\left(cW_c^0\right) \dots W_c^{q_c}\right) \quad (10)$$

where $W_c^{q_c}$ represents the learnable weight parameters of the q_c layer, c is the gene expression data of cell line features, and \tanh is the activation function. After processing through the MLP, a new feature matrix C' is obtained, which is dimensionally consistent with the extracted drug features, facilitating their subsequent fusion.

Multi-source feature interactive learning controller (MFIC)

Due to the involvement of multi-source data in the concatenated drug embedding vectors and cell line feature embedding vectors, including drug structure, biological activity, cell response, gene expression, etc., simply concatenating these vectors may

not fully utilize the information from different data sources. Additionally, with the increase in network layers, the multitude of parameters can make the network difficult to train. To fully integrate and deeply mine the features of drugs and cell lines while accelerating network training, we propose a multi-source feature interactive learning controller. By introducing a gating structure into the network, it can control the flow of information between different data sources, flexibly handling various features. This approach better facilitates the fusion of multi-source data and ensures smooth data transmission across multiple layers.

As shown in Fig. 3, the multi-source feature interactive learning controller divides the input data into two parts: one part undergoes nonlinear transformation, and the other part can pass through the layer without transformation. Based on this, the received information is selectively transmitted between layers, reducing the number of training parameters. For the vector F after concatenating drug features and cell line features, by setting the transformation gating $G(F)$, the model can effectively control two possible transformations, W and T , on the input feature vector F through self-learning. Specifically, $G(F)$ is set as a sigmoid transformation gate, converting its input values into probabilities between 0 and 1 to control the flow of input information. When the sigmoid gate output is close to 1, most of the input information undergoes transformation T ; when the sigmoid gate output is close to 0, it indicates that most of the input information undergoes transformation Q . To simplify the model, the final output formula is as follows:

$$y = G(F) \times W(F) + (1 - G(F)) \times Q(F) \quad (11)$$

where the transformation W is set as a relu function, and the transformation Q is set to perform a linear operation on the input. y is the final output of the module, and the dimensions of F , y , G , and Q are consistent. The addition of the sigmoid gating structure makes the form more flexible than the original, The simplified formula is as follows:

$$y = \begin{cases} \text{relu}(F), & \text{if sigmoid}(F) = 1 \\ \text{linear}(F), & \text{if sigmoid}(F) = 0 \end{cases} \quad (12)$$

The transformation gate G automatically learns whether to use the relu function for transformation or to apply a linear operation in the current state. The Stochastic Gradient Descent (SGD) algorithm is used to adjust network parameters, as shown below:

$$W_{t+1} = W_t + \eta \frac{1}{N} \sum_{n=1}^N \theta(-y_n W_t x_n) (y_n x_n) \quad (13)$$

where W represents the weight parameters, η is the learning rate, N is the number of samples per training session, θ are the network parameters, and x_n and y_n are the input and output, respectively.

The gating mechanism, as an alternative to simple vector concatenation, more effectively utilizes information from multiple data sources such as drug structure, biological activity, cell response, and gene expression. It flexibly handles different features to achieve the fusion of multi-source data. Furthermore, by introducing a gating structure, it addresses the issue of increased parameter quantity making the network difficult to train, thereby accelerating network training and enhancing model performance.

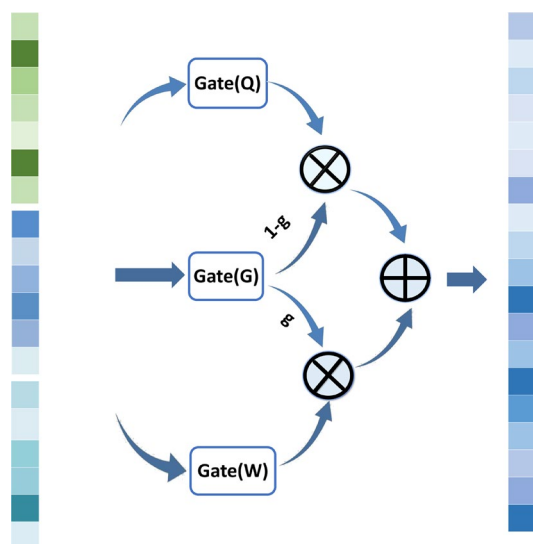


Fig. 3 The Data Processing Process of MFIC

Predicting the synergistic effects between drug combinations and cell lines

In our model, we utilize a GAT and an adaptive attention mechanism to process drug embedding vectors, and a MLP to process cell line embedding vectors. These processed vectors are then concatenated to form a new vector. This concatenated vector is further refined through a multi-source feature interactive learning controller, ensuring smooth data transmission across multiple layers. The fused vector is then passed through an MLP and a softmax layer to generate a classification for the synergistic effects of the drug combination. This process is depicted in Fig. 1d.

During the training process of our model, \hat{y} represents the predicted synergistic score of the drug combination by the model, and y represents the actual synergistic score. We use cross-entropy as the loss function to measure the difference between predicted and actual values, and optimize the model’s performance by minimizing this loss function. The specific loss function is as follows:

$$L_{i1} = -[y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)] \tag{14}$$

During the model training process, each sample is passed twice through the same network architecture, resulting in two different prediction outputs, \hat{y}_1^i and \hat{y}_2^i . The adoption of dropout mechanism leads to the random elimination of some neurons during the network propagation. Consequently, \hat{y}_1^i and \hat{y}_2^i represent distinct prediction probabilities generated by the two different sub-networks formed by the network’s two passes. This methodological approach of employing dual sub-networks introduces variability in the predictions, which substantially aids in enhancing the model’s generalization capacity and reducing the risk of overfitting.

To regularize the predictions from the two sub-networks, we minimize the Kullback–Leibler (KL) divergence between their respective output distributions. The KL divergence quantifies the difference between the probability distributions of \hat{y}_1^i and \hat{y}_2^i ,

measuring how much one distribution deviates from the other. This regularization term encourages the two sub-networks to generate similar output distributions, promoting consistency and reducing uncertainty in the model's predictions, as shown below:

$$\text{KL}(\hat{y}_1^i || \hat{y}_2^i) = \sum_{i=1}^N [(\hat{y}_1^i)(\log(\hat{y}_1^i) - \log(\hat{y}_2^i))] \quad (15)$$

Furthermore, the cross entropy loss takes into account both predictions \hat{y}_1^i and \hat{y}_2^i by averaging their combined values. The final loss function can be represented as follows:

$$L^i = 1/2[(L_{i1}\hat{y}_1^i + L_{i1}\hat{y}_2^i) + \alpha(\text{KL}(\hat{y}_1^i || \hat{y}_2^i) + \text{KL}(\hat{y}_2^i || \hat{y}_1^i))] \quad (16)$$

where α is a learnable parameter. We consider both the prediction error and the difference between the model's output distributions. By optimizing the final loss function, we encourage the model to better fit the training data, thereby improving the model's generalization ability and robustness.

Evaluation metrics

For the task of predicting drug combination synergy, the following metrics are used for evaluation: the area under the receiver operator characteristics curve (AUROC), the area under the precision – recall curve (AUPR), accuracy (ACC), balanced accuracy(BACC), precision (PREC), true positive rate (TPR), the Cohen's Kappa value (KAPPA). ACC is used to describe the model's ability to distinguish between synergistic and antagonistic drug combinations. BACC and KAPPA are two metrics that consider the model's predictive ability for both synergistic and antagonistic drug combinations and are suitable for handling imbalanced datasets. TPR and TNR respectively represent the model's predictive accuracy for positive and negative samples. PREC measures the accuracy of the model in predicting drug pairs as synergistic combinations. Generally, the higher these metrics, the stronger the predictive ability of the model. The calculation formulas for these metrics are as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (17)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (18)$$

$$\text{TNR} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (19)$$

$$\text{BACC} = \frac{\text{TPR} + \text{TNR}}{2} \quad (20)$$

$$\text{PREC} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (21)$$

$$\text{KAPPA} = \frac{p_o - p_e}{1 - p_e} \quad (22)$$

where TP , FP , TN , and FN respectively represent the number of correctly identified synergistic drug combinations, the number of antagonistic drug combinations incorrectly identified as synergistic, the number of correctly identified antagonistic drug combinations, and the number of synergistic drug combinations incorrectly identified as antagonistic. p_o is the ratio of the number of correctly classified samples to the total number of samples for each category, i.e., the overall classification accuracy. p_e is the ratio of the sum of the products of the actual and predicted quantities for each category to the square of the total number of samples, representing the rate of chance agreement. These metrics evaluate the model's ability to accurately recognize different types of samples and the consistency of the labeling task, reflecting the overall performance of the model and helping us judge the reliability of the model in predicting the synergy of drug combinations.

Experiment implementation

We use an RTX Nvidia 3090 GPU and is based on the PyTorch framework for training and testing. The Adam optimizer is used to update the model parameters. In the experiments, the batch size is set to 128; learning rate is set to 0.0001; dropout is set to 0.1; and cross-entropy is used as the loss function to measure the difference between the predicted results and the true labels.

Results and discussion

Performance comparison with other models

To evaluate the effectiveness of our model, it was compared with several existing methods on a benchmark dataset. These included methods based on machine learning for predicting drug combination synergy, such as Extreme Gradient Boosting(XGBoost), Random Forest(RF), Gradient Boosting Machines(GBM), Adaboost, Multilayer Perceptron(MLP), Support Vector Machines(SVM) and those based on deep learning, like DeepSynergy [23], TranSynergy [37], MGAE-DC [38], SDCNet [39], PRODeepSyn [40], DFFNDDS [41] and Deep Tensor Factorization(DTF) [42]. To delineate the distinctions between MFSynDCP and other deep learning-based approaches, we make the following summary for each deep learning model:

- *DeepSynergy*: DeepSynergy is a deep learning model that utilizes the chemical properties of two drugs and the gene expression of a cell line to forecast synergy scores. It utilizes a feedforward neural network to capture the potential pharmacological synergy between combinations of drugs.
- *TranSynergy*: TranSynergy integrates a Self-Attention Transformer to analyze drug synergy. It leverages input features such as drug-target interaction profiles, gene expression, and gene dependency profiles to compute a synergy score indicative of the effect of drug combinations on cell lines.
- *MGAE-DC*: The MGAE-DC framework utilizes a multi-channel graph autoencoder approach, with three distinct input channels designed to capture the effects of synergistic, additive, and antagonistic interactions among drugs. By employing

a graph convolutional network, it acquires drug embeddings and applies an attention mechanism to amalgamate these embeddings across various cell lines, aiming to forecast drug synergies.

- *SDCNet*: SDCNet, a novel encoder-decoder architecture, is crafted to forecast cell line-specific Synergistic Drug Combinations (SDCs). It employs a Relational Graph Convolutional Network (R-GCN) for the acquisition and amalgamation of deep drug representations from various cell lines, utilizing an attention mechanism to refine the learning of features.
- *PRODeepSyn*: The PRODeepSyn model integrates protein–protein interaction (PPI) networks with omics data via Graph Convolutional Networks to predict anticancer drug synergies.
- *DFNDDS*: DFNDDS is a deep learning model that predicts synergistic drug combinations by integrating a fine-tuned pretrained language model with a dual feature fusion mechanism, merging drug and cell line features at both bit-wise and vector-wise levels. This innovative approach ensures DFNDDS establishes itself as a dependable tool for identifying effective drug combinations.
- *DTF*: DTF integrates a tensor-based framework with deep learning techniques to forecast the synergistic effects of drug pairs, primarily utilizing tensor factorization and a deep neural network for its predictions.

We divided the dataset into a training set and a test set, accounting for 90% and 10% of the data, respectively. Five-fold cross-validation was used in experiments on the training set, where training samples were randomly divided into five roughly equal subsets. Four subsets were used as the training dataset, and the remaining one served as a validation set for assessing the model's performance and tuning the hyperparameters. To further ensure the model's generalization ability and prevent overfitting, early stopping was applied during the training process. Specific results are shown in Table 1.

Compared to other methods, our model achieved higher values in metrics such as AUROC, AUPR, ACC, BACC, TPR, KAPPA, indicating superior performance in the classification task of drug combination synergy prediction. Our model achieved an AUROC value of 0.930 ± 0.005 , demonstrating a stronger ability to distinguish between synergistic and non-synergistic drug combinations. The AUPR value also reached 0.929 ± 0.005 , indicating that the model maintains a high recall rate while achieving a high precision rate. In terms of the accuracy score (ACC), the model achieved a value of 0.855 ± 0.006 , exhibiting higher accuracy compared to other methods. In terms of precision (PREC) and recall rate (TPR), the model reached a value of 0.867 ± 0.012 , signifying its ability to correctly identify drug combinations with synergistic effects. To address the issue of class imbalance between synergistic and antagonistic drug combinations in the dataset, balanced accuracy (BACC) and KAPPA coefficient were used as evaluation metrics and reached values of 0.863 ± 0.004 and 0.709 ± 0.012 , respectively. These performance metrics provided a comprehensive evaluation of various aspects of the model. Here, it is noteworthy that MGAE-DC, SDCNet and the classical machine learning methods, such as XGB, also obtain competitive performance, but nevertheless still inferior to our method. Overall, our

Table 1 Performance comparison of MFSynDCP and competitive methods on fivefold cross validation

Methods	AUROC	AUPR	ACC	BACC	PREC	TPR	KAPPA
Ours	0.930 ± 0.005	0.929 ± 0.005	0.855 ± 0.006	0.855 ± 0.006	0.867 ± 0.012	0.863 ± 0.004	0.709 ± 0.012
MGAE-DC	0.922 ± 0.005	0.675 ± 0.010	0.851 ± 0.012	0.778 ± 0.005	0.723 ± 0.006	0.562 ± 0.003	0.603 ± 0.012
SDCNet	0.921 ± 0.007	0.920 ± 0.005	0.846 ± 0.007	0.832 ± 0.005	0.871 ± 0.005	0.795 ± 0.002	0.678 ± 0.008
DFFNDDS	0.912 ± 0.003	0.882 ± 0.005	0.821 ± 0.007	0.811 ± 0.003	0.821 ± 0.017	0.831 ± 0.013	0.661 ± 0.021
XGBoost	0.921 ± 0.005	0.922 ± 0.005	0.844 ± 0.005	0.844 ± 0.005	0.843 ± 0.007	0.840 ± 0.002	0.688 ± 0.009
PRODeep-Dyn	0.899 ± 0.005	0.922 ± 0.006	0.853 ± 0.007	0.853 ± 0.005	0.859 ± 0.007	0.856 ± 0.012	0.703 ± 0.003
TranSynergy	0.896 ± 0.007	0.892 ± 0.006	0.827 ± 0.013	0.827 ± 0.013	0.842 ± 0.006	0.801 ± 0.003	0.642 ± 0.013
DTF	0.892 ± 0.009	0.881 ± 0.008	0.814 ± 0.009	0.814 ± 0.009	0.822 ± 0.008	0.772 ± 0.031	0.633 ± 0.042
DeepSynergy	0.881 ± 0.005	0.874 ± 0.009	0.803 ± 0.007	0.803 ± 0.007	0.814 ± 0.011	0.752 ± 0.009	0.591 ± 0.048
GBM	0.852 ± 0.010	0.850 ± 0.007	0.772 ± 0.010	0.772 ± 0.010	0.773 ± 0.008	0.745 ± 0.014	0.544 ± 0.020
Random Forest	0.861 ± 0.010	0.850 ± 0.014	0.783 ± 0.012	0.783 ± 0.012	0.794 ± 0.016	0.751 ± 0.022	0.566 ± 0.024
Adaboost	0.828 ± 0.007	0.832 ± 0.010	0.743 ± 0.009	0.743 ± 0.009	0.746 ± 0.012	0.728 ± 0.006	0.486 ± 0.018
MLP	0.652 ± 0.024	0.640 ± 0.033	0.557 ± 0.045	0.560 ± 0.043	0.531 ± 0.042	0.924 ± 0.134	0.119 ± 0.085
SVM	0.586 ± 0.011	0.563 ± 0.011	0.542 ± 0.010	0.540 ± 0.010	0.534 ± 0.016	0.502 ± 0.067	0.081 ± 0.020

Table 2 P-value comparison of MFSynDCP and comparative methods using t-test

Methods	p-value t-test of AUROC	p-value t-test of AUPR	p-value t-test of ACC
MGAE-DC	7.28×10^{-3}	7.73×10^{-27}	4.72×10^{-3}
SDCNet	4.42×10^{-2}	5.13×10^{-3}	3.56×10^{-5}
DFFNDDS	2.13×10^{-6}	1.01×10^{-11}	5.03×10^{-13}
XGBoost	1.36×10^{-3}	4.87×10^{-4}	1.22×10^{-5}
PRODeepDyn	7.22×10^{-9}	3.73×10^{-3}	6.71×10^{-3}
TranSynergy	9.63×10^{-9}	7.67×10^{-12}	1.15×10^{-3}
DTF	8.72×10^{-10}	3.25×10^{-12}	1.71×10^{-10}
DeepSynergy	5.34×10^{-13}	4.69×10^{-13}	3.28×10^{-15}
GBM	3.19×10^{-13}	7.21×10^{-16}	7.98×10^{-16}
Random Forest	6.01×10^{-13}	3.54×10^{-10}	1.66×10^{-13}
Adaboost	3.37×10^{-14}	9.82×10^{-16}	5.16×10^{-17}
MLP	9.69×10^{-18}	8.65×10^{-20}	1.77×10^{-13}
SVM	6.76×10^{-24}	1.49×10^{-24}	1.52×10^{-25}

method achieves superior performance on most evaluation metrics compared to the advanced deep learning methods and classical machine learning methods.

To further confirm the statistical significance of the superiority of our model, we conducted t-tests for a statistical analysis on three important metrics: AUROC, AUPR, and ACC, comparing the performance differences between our model and other benchmark models. As shown in Table 2, the obtained p-values consistently fall below the standard significance level of 0.05, indicating that our model significantly outperforms all compared models in a statistical sense. These results further showcase the exceptional performance of our model, MFSynDCP, on the aforementioned

performance metrics, validating its effectiveness in predicting synergistic cancer drug combinations.

In the experiments, selections were made for hyperparameter values, including the learning rate, dimension of the GAT layer, dropout ratio, and batch size. Notably, when the learning rate was set to 0.001, the dimension of the GAT layer to 64, the dropout ratio to 0.1, and the batch size to 128, the model was more effective in extracting drug features. This combination of parameters not only improved the model’s performance on the training set but also demonstrated good generalization ability on the validation set. Further, it was found that fine-tuning the dimension of the GAT layer significantly impacts the model’s sensitivity in handling complex drug molecular structures. An appropriate dropout ratio helps prevent overfitting, ensuring the stability of the model’s training.

Evaluation on independent test dataset

To further validate the generalization ability of our model on new datasets, the study also employed a large drug combination dataset released by AstraZeneca [43] in 2019 as an independent test set to evaluate the performance of MFSynDCP and other benchmark methods. This dataset is the result of the AstraZeneca-Sanger Drug Combination Prediction DREAM Challenge, a collaboration between AstraZeneca and the Sanger Institute, aimed at exploring fundamental characteristics of effective combination therapy and synergistic drug behavior. The dataset consists of 668 novel drug-drug-cell line triplets, comprising 57 drugs and 24 cell lines. By training on the benchmark dataset and testing on this independent test set, as shown in Fig. 4, our model demonstrated favorable performance, correctly predicting 492 drug combination pairs. Moreover, our model outperformed other comparison methods across all evaluation metrics.

To more intuitively demonstrate the effectiveness of our model, we selected three deep learning approaches and four machine learning methods for the plotting of ROC curves, as shown in Fig. 5. The graphical representation clearly illustrates that our model achieved a significant AUROC score of 0.701 ± 0.12 , surpassing other competing

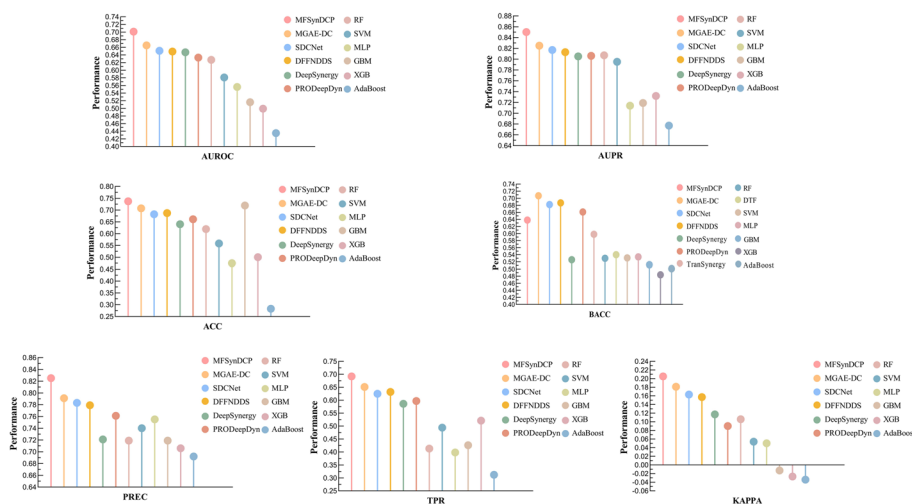


Fig. 4 Performance of MFSynDCP and its variants on the independent test dataset released by AstraZeneca

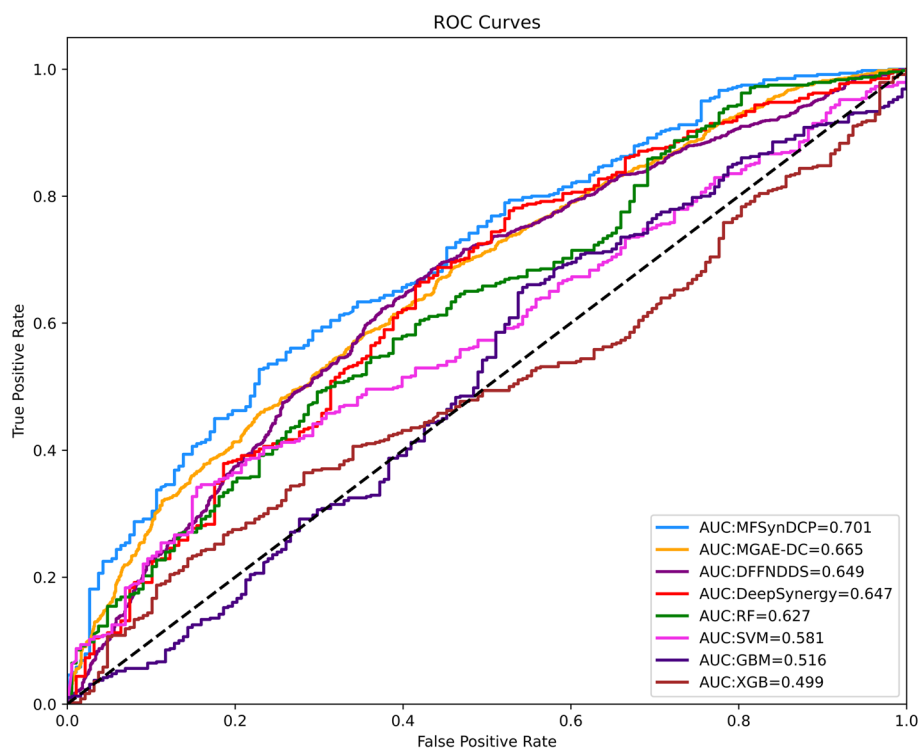


Fig. 5 ROC curves of MFSynDCP and competitive methods on independent test dataset

models. In contrast, the ROC curves of some benchmark models closely resembled random predictions, indicating their limited predictive accuracy for drug combinations. However, our model consistently exhibited exceptional performance on the test set of this challenge, thereby further validating its excellent generalization capability and practical applicability.

Ablation study

To investigate the importance and contribution of each component in the model, we conducted an ablation analysis by removing or replacing some model components. Specifically, we compared the results of MFSynDCP under the following conditions: (i) the proposed MFSynDCP, (ii) replace MLP with Variational Autoencoder (VAE) in MFSynDCP, (iii) replace GAT with GCN in MFSynDCP, (iv) replace GAT with Graph Isomorphism Network (GIN) in MFSynDCP, (v) replace GAT with GNN in MFSynDCP (vi) MFSynDCP without AAGAM, (vii) replace AAGAM with the global mean pooling (GMP) in MFSynDCP, (viii) MFSynDCP without MFIC. We conducted a fivefold cross-validation test based on the training dataset for comparison. The results on the benchmark dataset are summarized in the figure below. Figure 6 shows the experimental results of our model MFSynDCP compared with the other four variants.

The results indicate that the complete MFSynDCP framework achieved the best predictive performance in the 7 evaluation metrics, demonstrating its effectiveness. Specifically, we observed that GAT outperforms GCN, GIN, and GNN in extracting key chemical features. Through the adaptive attention mechanism of GAT, it can better capture important features and interaction information among drugs, thereby improving

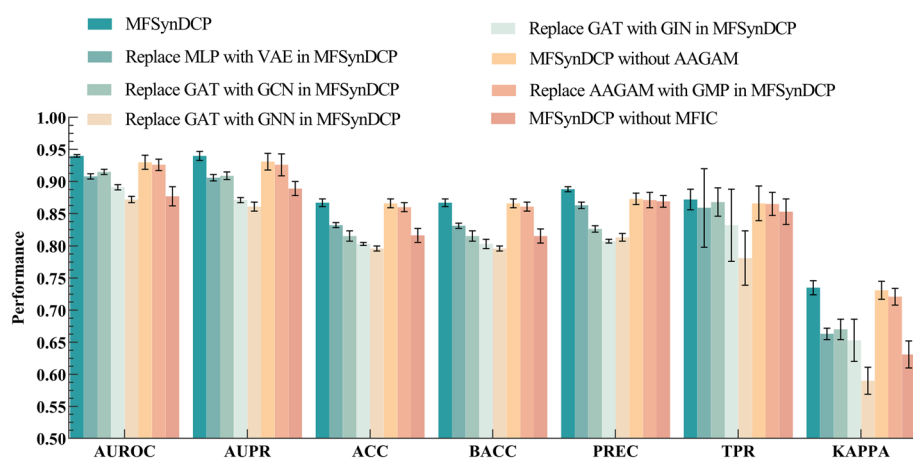


Fig. 6 The performance of our proposed MFSynDCP and its variants

the predictive performance of the model. In contrast, GCN, GIN, and GNN exhibit limitations in utilizing the complex relationships among drugs in certain aspects.

As can be observed in Fig. 6, in terms of encoding the genomic features of cancer cells, the use of MLP outperforms VAE. A possible reason for this could be the flexible network architecture and feature learning capability of MLP, which allows it to capture the key nonlinear relationships and complex features within the cancer cell genome more effectively than VAE.

Notably, the complete framework scored lower on the TPR metric compared to the model without the AAGAM, which might be due to the class imbalance between synergistic and antagonistic drug combinations in the dataset. To address this issue, BACC and KAPPA coefficients can be used for assessment, where the complete model achieved the highest performance scores.

The experimental results prove that the AAGAM in our model performs better than the versions without adaptive attention mechanisms and those using global mean pooling for graph aggregation. This is likely because, for molecular graphs, the global mean pooling method treats every substructure as equally important and simply averages the embeddings of all nodes. In contrast, our proposed AAGAM utilizes the interaction information of drug pairs, not just the molecular graph of a single drug, to obtain attention scores for each substructure, thereby achieving better performance compared to the other two comparison models.

Furthermore, it can be concluded that the models without the AAGAM and those replacing AAGAM with GMP did not show a significant difference in the AUROC, AUPR, ACC, BACC, TPR, and KAPPA metrics, and their performance was lower than the predictive indicators of our proposed model. This underscores the important role of our proposed adaptive attention mechanism in the graph aggregation module within this model. Simultaneously, the MFIC design makes a greater contribution to learning drug features compared to the AAGAM proposed in our study, possibly due to its effective handling and integration of feature information from different sources, including both drugs and cell lines. The MFIC led fusion module plays a key role in ensuring high-quality predictions of drug synergistic effects. The experimental results according to

AUROC, AUPR, ACC, BACC, TPR, and KAPPA indicate that the absence of MFIC led to a significant drop in model performance, suggesting that the MFIC-led fusion module effectively captures the synergistic effects between features. The higher scores in the PREC metric could also confirm the imbalance in the dataset; if there are fewer positive samples and more negative samples, the model may be biased towards predicting samples as negative to achieve a higher accuracy rate. Moreover, we observed that the complete MFSynDCP framework outperformed the other ablation scenarios in six metrics, further validating the importance of each component.

In summary, the experimental results clearly demonstrate that the adaptive attention mechanism and the MFIC-led fusion module play a crucial role in enhancing model performance. Their combination is capable of more comprehensively capturing the features of drug synergistic effects.

The impact of the input sequence of drug combination data on predictive performance

To mitigate the impact of drug order on the model's prediction results, during the training process, we treated [drug A, drug B, cell line] samples and [drug B, drug A, cell line] samples as two distinct input samples. This approach allowed us to examine the effect of different input feature orders on predicting synergy scores. As shown in Fig. 7, we observed that the prediction results under different input feature orders are concentrated near the diagonal line, with a Pearson correlation coefficient reaching 0.9. This indicates that our model is not sensitive to the order of drug combinations; accurate predictions are generated regardless of whether it's drug A-drug B or drug B-drug A. This further verifies the robustness and reliability of our model. Additionally, we observed that both the ROC AUC (Area Under the Receiver Operating Characteristic Curve) and PR AUC (Precision-Recall Area Under Curve) for [drug A, drug

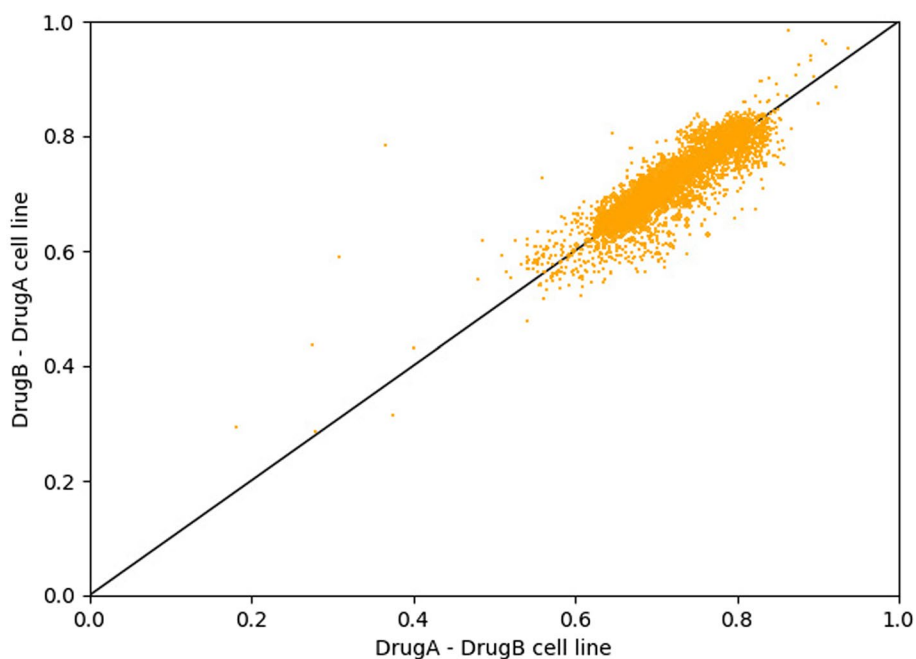


Fig. 7 Scatter plot of collaborative scores obtained based on different input orders of two drugs

B] and [drug B, drug A] reached or were close to 0.93, further proving the excellent performance of our model in this study.

The revelation of crucial chemical substructures in drugs

Deep learning models are often viewed as black boxes, and their lack of interpretability limits their further application in many fields, especially in practical scenarios of computational-aided drug discovery. To address this issue and explore the key substructures in drug combination prediction, we utilized the attention mechanism to visualize the critical substructures of drug pairs.

The MFSynDCP model proposed in this study employs a message-passing mechanism between nodes to update each node's information. This allows each node to capture information from its neighboring nodes and gradually accumulate and integrate information from surrounding nodes, enriching its feature representation. In this model, each neuron in the GAT network is connected to neighboring nodes from the previous layer through a set of learnable weights, enabling the neuron to acquire information from its neighbors and incorporate it into its feature expression. Furthermore, we introduced an adaptive attention mechanism-based graph aggregation module. This module assigns attention scores to each substructure of a drug and performs a weighted summation of all nodes' embedding vectors, resulting in a graph-aggregated representation of the drug. This process reveals the key chemical substructures that play a crucial role in synergy prediction. Therefore, the final drug feature representation actually contains information about the surrounding chemical substructures, including valency, solubility, and other physicochemical properties. This inspired our exploration of the attention mechanism in revealing important chemical substructures.

Specifically, the attention scores calculated using formulas 4 and 5 are used to represent the importance levels of corresponding substructures. These substructures' importance is visualized using different colors. Figure 8 displays the visualization results for three randomly selected drug pairs (ABT-888 and SORAFENIB, 5-FU and Erlotinib, L778123 and TEMOZOLOMIDE). In the initial stages of training, the attention scores show a more uniform distribution, indicating that the model has not yet focused on key structures with significant influence. However, as training progresses, the model gradually starts to assign higher importance to certain specific structures compared to others. Fig. 8a₁–c₁ presents the visualization results obtained after the model's training is complete, where deeper colors reflect more important substructures.

Taking Fig. 8b and b₁ as examples for explanation: A427 [44] is a human non-small cell lung cancer cell line, while 5-FU [45] and Erlotinib [46] are two drugs commonly used in lung cancer treatment. These drugs can be used in combination therapy. 5-FU and Erlotinib have been shown to have a more effective growth inhibitory effect on the A427 cell line. Our model successfully identified the amide group as an important chemical structure, which plays a key role in biomolecules, including many clinically approved drugs. Amides are widely present in drugs, not only because of their stability but also because their polarity allows drugs containing amide groups to interact with biological receptors and enzymes. This result demonstrates the good interpretability of our model.

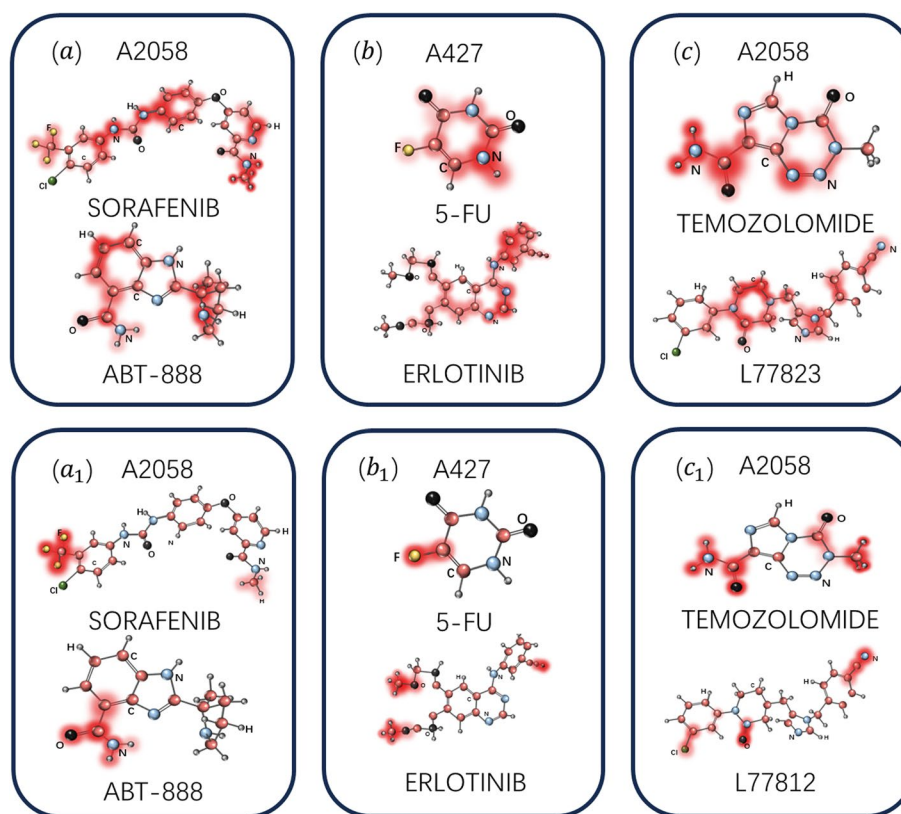


Fig. 8 The visualization results for three randomly selected drug pairs are presented. Figure 8a–8c display the visualization of attention scores for these three drug pairs before training. Figure 8a₁–c₁ shows the visualization of attention scores for these three drug pairs during the model training process, where deeper colors indicate more important substructures

The prediction of new synergistic drug combinations

The pursuit of innovative and effective drug combinations remains a cornerstone in the fight against cancer, presenting a complex yet crucial challenge in medical research. We introduce a refined methodological approach aimed at identifying synergistic drug combinations, effectively capitalizing on the sophisticated capabilities of the MFSynDCP model. Our approach integrates computational modeling with clinical predictive analysis, aiming to identify novel drug combinations that have the potential to alter current treatment modalities, thereby offering new research pathways and therapeutic strategies in the field of cancer treatment.

To assess the model's potential in discovering new synergistic drug combinations, we trained our MFSynDCP model using the O'Neil drug combination dataset. To generate candidate drug combinations, we selected 25 small-molecule anticancer chemical drugs approved by the U.S. Food and Drug Administration (FDA), removing drug combinations that duplicated those in the benchmark dataset. We then used our MFSynDCP model to predict the synergy of the final candidate drug pairs. Extensive literature searches were conducted to validate whether the model could identify new synergistic drug combinations.

In this study, new drug combinations were predicted using the widely studied A375 cancer cell line [47], forecasting unknown [drug, drug, cell line] triplets. We focused particularly on the top 7 ranked untested triplets in the prediction scores and conducted a non-exhaustive literature search. We found that 3 of the predicted drug combinations were consistent with previous research or clinical trial observations. For instance, the combination of Erlotinib and Regorafenib was used for the treatment of hepatocellular carcinoma, successfully overcoming the interference of epidermal growth factor [48]. These examples illustrate that MFSynDCP can successfully predict drug combinations consistent with previous research or clinical trial observations, further validating its potential in discovering new synergistic drug combinations.

The predicted new synergistic drug combinations each consist of two drugs, and each combination is assigned a predictive score reflecting its potential synergistic efficacy. Additionally, the predictive scores for all listed drug combinations are exceptionally high (close to 1), indicating these combinations show substantial potential for synergy in the model. At least three of these predicted drug combinations are consistent with existing research or clinical trial observations, enhancing the reliability of the predictions. All these predictions are made for the A375 cancer cell line, a melanoma (skin cancer) cell line. This specificity suggests that these combinations may not be equally effective against other types of cancer cells. Future research could focus on validating the actual efficacy and potential synergistic mechanisms of those combinations that are supported by literature but have not yet entered clinical trial stages, as well as those completely unsupported by existing literature. Overall, Fig. 9 demonstrates the capability of the MFSynDCP model in predicting potentially effective new synergistic drug combinations. It offers a promising beginning for further experimental and clinical research, highlighting the model's utility in guiding hypothesis generation and decision-making in drug development and personalized medicine.

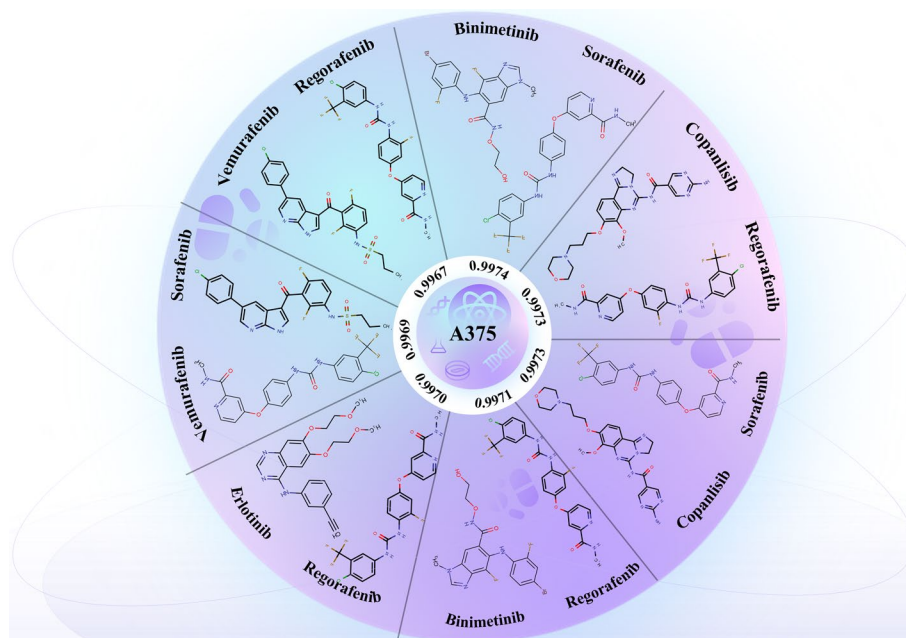


Fig. 9 The top 7 novel synergistic combinations predicted on A375 cancer cell line

Conclusions

In this paper, we proposed a deep graph neural network model named MFSynDCP, which is guided by a multi-source feature interactive learning controller and employs an adaptive attention mechanism for predicting the synergy of anticancer drug combinations. Specifically, the SMILES features of drugs are first transformed into drug molecular structure graphs, and a Graph Attention Network (GAT) is used to extract structural information from drug pairs. An adaptive attention mechanism-based graph aggregation module was designed to unearth the most critical chemical substructures for synergy prediction. Additionally, an innovative multi-source feature interactive learning controller was constructed to enhance the representation of drug pairs, enabling the fusion of multi-source data from drugs and cell lines and learning the interaction information between them. We also explored the learning process of MFSynDCP, uncovering the mechanisms of drug synergy among substructures, which provided a level of interpretability to the model and supported the explanation of drug synergy mechanisms. Our performance comparison experiments demonstrated that MFSynDCP outperforms other competitive methods.

However, the MFSynDCP model has certain limitations. The study focused solely on features of drugs and cell lines for synergy prediction, without considering the potential of biomedical knowledge graph methods in predicting effective combinations for diseases. In the future, we plan to integrate biomedical knowledge graphs to further enhance the overall performance of predicting synergistic anticancer drug combinations. Additionally, we recognize the importance of exploring gene contributions in synergy prediction and plan to incorporate this aspect into our future research endeavors. This will allow us to gain a more comprehensive understanding of the factors influencing synergy prediction and improve the predictive capabilities of our model.

Author contributions

YYD and YQC designed the research study. YXW, XYW, YZ, ZTY and QXH performed the research. XLF and XQR analyzed the data, KW and YQ discussed and critically revised the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

Funding

This work is supported by the National Natural Science Foundation of China (No.62306206);the Key Research and Development Program of Shanxi Province (No.202102020101007); and the Applied Basic Research Program of Shanxi Province (No.202203021212207);

Availability of data and materials

All data generated or analyzed during this study are included in this article.

Code availability

The source code is available from GitHub at <https://github.com/kkioplkg/MFSynDCP>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no conflict to disclose.

Received: 17 January 2024 Accepted: 28 March 2024

Published online: 01 April 2024

References

1. Humphrey RW, Brockway-Lunardi LM, Bonk DT, Dohoney KM, Doroshow JH, Meech SJ, et al. Opportunities and challenges in the development of experimental drug combinations for cancer. *J Natl Cancer Inst.* 2011;103:1222–6. <https://doi.org/10.1093/jnci/djr246>.
2. Wu L, Wen Y, Leng D, Zhang Q, Dai C, Wang Z, et al. Machine learning methods, databases and tools for drug combination prediction. *Brief Bioinform.* 2022;23:bbab355. <https://doi.org/10.1093/bib/bbab355>.
3. Jaaks P, Coker EA, Vis DJ, Edwards O, Carpenter EF, Leto SM, et al. Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature.* 2022;603:166–73. <https://doi.org/10.1038/s41586-022-04437-2>.
4. Yu J, Mu Q, Fung M, Xu X, Zhu L, Ho RJP, et al. Challenges and opportunities in metastatic breast cancer treatments: nano-drug combinations delivered preferentially to metastatic cells may enhance therapeutic response. *Pharmacol Ther.* 2022;236:108108. <https://doi.org/10.1016/j.pharmthera.2022.108108>.
5. Zhu C, Guan X, Zhang X, Luan X, Song Z, Cheng X, et al. Targeting KRAS mutant cancers: from druggable therapy to drug resistance. *Mol Cancer.* 2022;21:159. <https://doi.org/10.1186/s12943-022-01629-2>.
6. Yardley DA. Drug resistance and the role of combination chemotherapy in improving patient outcomes. *Int J Breast Cancer.* 2013. <https://doi.org/10.1155/2013/137414>.
7. Lin R, Yin G. Bayesian optimal interval design for dose finding in drug-combination trials. *Stat Methods Med Res.* 2017;26:2155–67. <https://doi.org/10.1177/0962280215594494>.
8. Lin GL, Wilson KM, Ceribelli M, Stanton BZ, Woo PJ, Kreimer S, et al. Therapeutic strategies for diffuse midline glioma from high-throughput combination drug screening. *Sci Transl Med.* 2019;11:0064. <https://doi.org/10.1126/scitranslmed.aaw0064>.
9. Kim H, Lee SJ, Lee IK, Min SC, Sung HH, Jeong BC, et al. Synergistic effects of combination therapy with AKT and mTOR inhibitors on bladder cancer cells. *Int J Mol Sci.* 2020;21:2825. <https://doi.org/10.3390/ijms21082825>.
10. Costa A, Campos D, Reis C, Gomes C. Targeting glycosylation: a new road for cancer drug discovery. *Trends Cancer.* 2020;6:757–66. <https://doi.org/10.1016/j.trecan.2020.04.002>.
11. Lu Y, Chan Y-T, Tan H-Y, Li S, Wang N, Feng Y. Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy. *Mol Cancer.* 2020;19:1–16. <https://doi.org/10.1186/s12943-020-01197-3>.
12. Zhao X-M, Iskar M, Zeller G, Kuhn M, Van Noort V, Bork P. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol.* 2011;7:e1002323. <https://doi.org/10.1371/journal.pcbi.1002323>.
13. Beam AL, Drazen JM, Kohane IS, Leong T-Y, Manrai AK, Rubin EJ. Artificial intelligence in medicine. *N Engl J Med.* 2023;388:1220–1. <https://doi.org/10.1056/NEJMe2206291>.
14. De Ville B. Decision trees. *Wiley Interdiscip Rev Comput Stat.* 2013;5:448–55. <https://doi.org/10.1002/wics.1278>.
15. Breiman L. Random forests. *Mach Learn.* 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
16. Noble WS. What is a support vector machine? *Nat Biotechnol.* 2006;24:1565–7. <https://doi.org/10.1038/nbt1206-1565>.
17. Šicho M, Luukkonen S, van den Maagdenberg HW, Schoenmaker L, Béguignon OJ, van Westen GJ, et al. DrugEx: deep learning models and tools for exploration of drug-like chemical space. *J Chem Inf Model.* 2023. <https://doi.org/10.1021/acs.jcim.3c00434>.
18. Mei S. A machine learning framework for predicting synergistic and antagonistic drug combinatorial efficacy. *J Math Chem.* 2022;60:752–69. <https://doi.org/10.1007/s10910-022-01331-0>.
19. Janizek JD, Dincer AB, Celik S, Chen H, Chen W, Naxerova K, et al. Uncovering expression signatures of synergistic drug responses via ensembles of explainable machine-learning models. *Nat Biomed Eng.* 2023. <https://doi.org/10.1038/s41551-023-01034-0>.
20. Julkunen H, Cichonska A, Gautam P, Szedmak S, Douat J, Pahikkala T, et al. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nat Commun.* 2020;11:6136.
21. Hall MA. Correlation-based feature selection for machine learning. *Hamilton: The University of Waikato*; 1999.
22. Shaheen F, Verma B, Asafuddoula M. Impact of automatic feature extraction in deep learning architecture. In: 2016 International conference on digital image computing: techniques and applications (DICTA). IEEE; 2016. pp. 1–8. <https://doi.org/10.1109/DICTA.2016.7797053>
23. Preuer K, Lewis RP, Hochreiter S, Bender A, Bulusu KC, Klambauer GJB. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics.* 2018;34:1538–46. <https://doi.org/10.1093/bioinformatics/btx806>.
24. Rafiei F, Zeraati H, Abbasi K, Ghasemi JB, Parsaeian M, Masoudi-Nejad AJB. DeepTraSynergy: drug combinations using multimodal deep learning with transformers. *Bioinformatics.* 2023;39:btad438. <https://doi.org/10.1093/bioinformatics/btad438>.
25. Yang J, Xu Z, Wu WK, Chu Q, Zhang Q. GraphSynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J Am Med Inform Assoc.* 2021;28:2336–45. <https://doi.org/10.1093/jamia/ocab162>.
26. O'Neil J, Benita Y, Feldman I, Chenard M, Roberts B, Liu Y, et al. An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther.* 2016;15:1155–62. <https://doi.org/10.1158/1535-7163.MCT-15-0843>.
27. Di Veroli GY, Fornari C, Wang D, Mollard S, Bramhall JL, Richards FM, et al. Combeneft: an interactive platform for the analysis and visualization of drug combinations. *Bioinformatics.* 2016;32:2866–8. <https://doi.org/10.1093/bioinformatics/btw230>.
28. Goldoni M, Johansson C. A mathematical approach to study combined effects of toxicants in vitro: evaluation of the bliss independence criterion and the Loewe additivity model. *Toxicol In Vitro.* 2007;21:759–69. <https://doi.org/10.1016/j.tiv.2007.03.003>.
29. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci.* 1988;28:31–6. <https://doi.org/10.1021/ci00057a005>.

30. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46:D1074–82. <https://doi.org/10.1093/nar/gkx1037>.
31. Bento AP, Hersey A, Félix E, Landrum G, Gaulton A, Atkinson F, et al. An open source chemical structure curation pipeline using RDKit. *J Cheminformatics.* 2020;12:1–16. <https://doi.org/10.1186/s13321-020-00456-1>.
32. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012;483:603–7. <https://doi.org/10.1038/nature11003>.
33. Velculescu VE, Madden SL, Zhang L, Lash AE, Yu J, Rago C, et al. Analysis of human transcriptomes. *Nat Genet.* 1999;23:387–8. <https://doi.org/10.1038/70487>.
34. Ramsundar B. *Molecular machine learning with DeepChem.* Stanford: Stanford University; 2018.
35. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 2012;41(D1):D955–61. <https://doi.org/10.1093/nar/gks1111>.
36. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long non-coding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;22:1775–89.
37. Liu Q, Xie L. TranSynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput Biol.* 2021;17(2):e1008653. <https://doi.org/10.1371/journal.pcbi.1008653>.
38. Zhang P, Tu S. MGAE-DC: predicting the synergistic effects of drug combinations through multi-channel graph autoencoders. *PLoS Comput Biol.* 2023;19(3):e1010951. <https://doi.org/10.1371/journal.pcbi.1010951>.
39. Zhang P, Tu S, Zhang W, Xu L. Predicting cell line-specific synergistic drug combinations through a relational graph convolutional network with attention mechanism. *Brief Bioinform.* 2022;23(6):bbac403. <https://doi.org/10.1093/bib/bbac403>.
40. Wang X, Zhu H, Jiang Y, Li Y, Tang C, Chen X, et al. PRODeepSyn: predicting anticancer synergistic drug combinations by embedding cell lines with protein–protein interaction network. *Brief Bioinform.* 2022;23(2):bbab587. <https://doi.org/10.1093/bib/bbab587>.
41. Xu M, Zhao X, Wang J, Feng W, Wen N, Wang C, et al. DFFNDDS: prediction of synergistic drug combinations with dual feature fusion networks. *J Cheminformatics.* 2023;15(1):1–12. <https://doi.org/10.1186/s13321-023-00690-3>.
42. Sun Z, Huang S, Jiang P, Hu P. DTF: deep tensor factorization for predicting anticancer drug synergy. *Bioinformatics.* 2020;36(16):4483–9. <https://doi.org/10.1093/bioinformatics/btaa287>.
43. Menden MP, Wang D, Mason MJ, Szalai B, Bulusu KC, Guan Y, et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat Commun.* 2019;10:2674. <https://doi.org/10.1038/s41467-019-09799-2>.
44. Parry JJ, Eiblmaier M, Andrews R, Meyer LA, Higashikubo R, Anderson CJ, et al. Characterization of somatostatin receptor subtype 2 expression in stably transfected A-427 human cancer cells. *Mol Imaging.* 2007;6:7290–2007. <https://doi.org/10.2310/7290.2007.0000>.
45. Wigmore PM, Mustafa S, El-Beltagy M, Lyons L, Umka J, Bennett G. Effects of 5-FU. In: Raffa RB, Tallarida RJ, editors. *Chemo fog.* New York: Springer; 2010. p. 157–64. https://doi.org/10.1007/978-1-4419-6306-2_20.
46. Bareschino MA, Schettino C, Troiani T, Martinelli E, Morgillo F, Ciardiello F. Erlotinib in cancer treatment. *Ann Oncol.* 2007;18:vi35–41. <https://doi.org/10.1093/annonc/mdm222>.
47. Zhang YP, Li YQ, Lv YT, Wang JM. Effect of curcumin on the proliferation, apoptosis, migration, and invasion of human melanoma A375 cells. *Genet Mol Res.* 2015;14:1056–67. <https://doi.org/10.4238/2015.February.6.9>.
48. D'Alessandro R, Refolo MG, Lippolis C, Carella N, Messa C, Cavallini A, et al. Modulation of Regorafenib effects on HCC cell lines by epidermal growth factor. *Cancer Chemother Pharmacol.* 2015;75:1237–45. <https://doi.org/10.1007/s00280-015-2751-6>.

Publisher's Note

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