

# 基于图神经网络的抗癌药物协同预测

杨培生

上海理工大学理学院, 上海

收稿日期: 2024年3月8日; 录用日期: 2024年5月27日; 发布日期: 2024年6月6日

## 摘要

药物组合作为个性化癌症治疗中的重要手段, 因其低毒性和较小不良反应而受到广泛关注。然而, 由于药物组合的指数性增长使得体外筛选药物协同组合变得耗时且费力。基于计算的方法逐渐被引入用于预测药物协同组合, 以提高效率。然而, 这些方法未能充分利用药物与细胞系之间复杂的关系。因此, 本研究提出了一种名为GNNSynergy的基于图神经网络的抗癌药物协同预测方法。GNNSynergy利用Loewe评分构建了药物组合与细胞系之间的相互作用超图和单个药物与细胞系之间的相互作用图, 其中药物和细胞系均作为图中的实体。然后, 我们采用超图神经网络和图卷积神经网络分别学习两种图中的节点特征。最后, 通过多层感知机和线性相关系数计算并融合两类网络的药物协同评分。GNNSynergy模型在基线实验中表现优于现有的先进模型, 在新药与新细胞系实验中展现出了强大的泛化能力。

## 关键词

图神经网络, 线性相关系数, 药物协同预测

# Anti-Cancer Drug Synergy Prediction Based on Graph Neural Network

Peisheng Yang

College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 8<sup>th</sup>, 2024; accepted: May 27<sup>th</sup>, 2024; published: Jun. 6<sup>th</sup>, 2024

## Abstract

Drug combination has gained wide attention in personalized cancer treatment due to its low toxicity and minimal adverse effects. However, the exponential growth of possible drug combinations makes it time-consuming and labor-intensive to screen for synergistic drug combinations *in vitro*. Computational approaches have been gradually introduced to predict drug synergy, aiming to im-

prove efficiency. However, these methods have not fully exploited the complex relationship between drugs and cell lines. Therefore, this study proposes a graph neural network-based anti-cancer drug synergy prediction method called GNNSynergy. GNNSynergy utilizes Loewe scores to construct interaction hypergraphs between drug combinations and cell lines, as well as interaction graphs between individual drugs and cell lines, where drugs and cell lines are treated as entities in the graphs. Subsequently, we employ a hypergraph neural network and a graph convolutional neural network to learn node features in the two types of graphs, respectively. Finally, drug synergy scores from the two networks are calculated and fused using a multi-layer perceptron and linear correlation coefficient. The GNNSynergy model outperforms existing advanced models in baseline experiments and exhibits strong generalization ability in experiments involving new drugs and cell lines.

## Keywords

Graph Neural Network, Linear Correlation Coefficient, Drug Synergy Prediction

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在过去几十年的时间里，科学家们一直致力于与癌症进行斗争，并提出了各种单药治疗方法，以提高患者的生存期。然而，由于单一用药的限制，癌症对抗癌药物的耐药性不断增强，导致单一用药在治疗癌症患者方面效果有限[1]。因此，近年来，科学家们逐渐开始关注联合用药对于癌症治疗效果的重要性。大量体外实验研究表明，与传统的单一用药相比，药物的联合使用可以有效延长癌症患者的生存期并提高其生存率[2] [3] [4] [5]，成为癌症研究的重要方向[6] [7] [8] [9]。随着高通量技术的发展，一些大型抗癌药物组合数据库如 O'Neil [10]、NCI-ALMANAC [11]、DREAM [12] 逐渐被建立，这些数据库提供了数百万种不同药物组合和癌细胞系的协同评分实验测量数据。因此，科学家们提出了许多基于高通量技术的抗癌药物协同预测模型，为发现临床上常见的药物组合的协同效应做出了巨大贡献。

抗癌药物协同预测通常被视为分类问题或回归任务。在分类任务中，药物组合根据其对癌细胞系的协同得分被分为协同或非协同两类；而回归任务则是预测药物组合的协同得分。近年来，研究人员提出了许多基于机器学习和深度学习的模型，利用药物组合的化学特征和癌细胞系的多组学特征作为输入，根据不同任务类型输出预测值：药物协同或非协同分类或药物组合对癌细胞系的协同得分。例如，文献[13]提出一种名为 CDRscan 的方法，通过一维卷积神经网络从药物的指纹信息和细胞系的基因表达数据中捕获药物与细胞系的深层表达以进行抗癌药物反应预测。文献[14]建立了一个综合的药物组合数据库，并提出了一种基于机器学习的模型。他们利用药物的化学特征、药物靶点信息和基因表达数据等多种特征，借助随机森林算法预测药物的协同效应。另一方面，王等人[15]构建了一个基于多项式回归的抗癌药物协同预测框架，利用响应张量、药物化学特征和细胞系的多组学特征作为输入来模拟药物协同效应。然而，由于传统机器学习方法对非线性关系建模的局限性，以及处理大规模高维数据的困难，越来越多的研究开始采用深度学习模型来取代传统机器学习方法。例如，Preuer 等人[16]根据药物的理化性质和癌细胞系的基因组学特征，构建了基于深度神经网络(DNN)的深度学习模型，用于抗癌药物的协同分类和回归任务。Bazgir 等人[17]将输入的高维特征向量表示为具有空间相关性的

紧凑图像, 并利用卷积神经网络(Graph Convolutional Neural Network, GCN) [18]构建回归模型。而文献 [19]构造了超图, 以药物组合和癌细胞系作为节点, 以协同得分作为超图边, 在输入方面使用了癌细胞的基因表达和药物的 SMILES (Simplified Molecular-Input Line-Entry System)特征。基于超图神经网络(Hypergraph Neural Network, HGNN) [20]构建了名为 HypergraphSynergy 的深度学习模型, 用于抗癌药物协同分类和回归任务。

尽管在抗癌药物协同预测领域, 基于高通量的模型不断取得进展, 但仍存在一些问题。例如, 基于神经网络的深度学习方法未考虑到药物组合和癌细胞系之间的拓扑结构。尽管 HypergraphSynergy 模型采用超图神经网络来学习药物组合和癌细胞系之间的拓扑信息, 但忽略了单一药物和癌细胞系之间的关联。为解决上述问题, 本文第四章提出了一种新的抗癌药物协同预测模型。该模型不仅考虑到药物组合和癌细胞系内在的联系, 还构建了单个药物与细胞系之间的关联, 以获取药物与细胞系之间的交互信息, 从而取得了显著的实验效果。我们的研究做出了如下贡献:

1) 构造了两种类型的网络, 并用不同的图神经网络进行训练, 不仅获取了药物组合与细胞系之间的内在联系, 同时捕获了单药物与细胞系之间的内在关系。

2) 在开放数据库上与现有先进模型对比实验表明, 我们的模型在性能上有明显改进, 具有突出的泛化能力和鲁棒性。

## 2. 材料与方法

### 2.1. 数据集

我们从多个公开的数据集收集模型所需的数据: 抗癌药物协同数据, 药物分子信息和癌细胞系的基因表达特征。这些数据库包括: O'Neil 抗癌药物协同预测数据集, PubChem [21]药物分子数据集, COSMIC [22]细胞系基因表达数据集。

为了获得药物组合与癌细胞系的相互作用信息, 我们从大型肿瘤筛查数据集 O'Neil 获取药物组合与癌细胞系的 Loewe 协同评分。O'Neil 数据集包含 38 种药物和 39 个人类癌细胞系的 23,062 个数据样本。每个样本为包含两种药物和一种细胞系的 Loewe 协同作用评分。然后使用公式(1)对初始相互作用

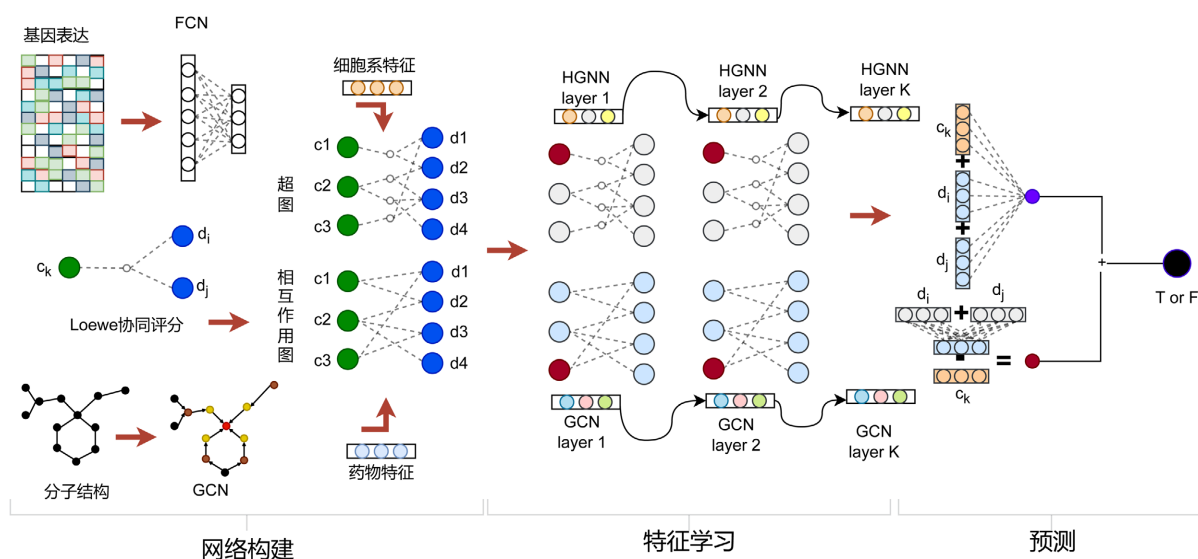


Figure 1. GNNSynergy flowchart

图 1. GNNSynergy 流程图

数据进行二值化：

$$P_{kij} = \begin{cases} 1 & \text{Loewe}_{kij} \geq 30 \\ 0 & \text{其他} \end{cases} \quad (1)$$

其中  $Loewe_{kij}$  表示药物  $k$  与药物  $i$  所组集合与细胞系  $j$  的 Loewe 协同评分。若  $p_{kij} = 1$  则表示该药物组合对该药物存在协同作用，若  $p_{kij} = 0$  反之。

药物的分子结构由药物的 SMILES 表示，从 PubChem 数据集获得。

癌细胞系的基因表达特征采用 COSMIC 数据集中的细胞系项目数据集，且只考虑与 COSMIC 癌症基因普查相关的 651 个基因数据(<https://cancer.sanger.ac.uk/census>)。并且对细胞系基因表达的每百万转录本值进行  $\log_2$  转化并进行 z-score 归一化。

获得上述样本后，我们去除没有基因表达的癌细胞系和没有 SMILES 特征的药物分子。从 O'Neil 数据集的 38 种药物和 39 个人类癌细胞系中获取 18,950 个有用数据样本的 Loewe 协同作用评分，其中具有协同作用的样本有 1694 个，没有协同作用的样本有 17,256 个。

## 2.2. 模型

在本研究中，我们的模型主要有三个部分组成：网络构建，特征学习和预测。模型的主体结构见图 1 所示。首先，我们癌细胞系的基因表达数据，药物的分子结构特征和药物协同数据 Loewe 评分作为输入，分别构建药物与细胞系协同作用超图和药物与细胞系相互作用图；其次通过超图神经网络和图神经网络分别对超图和相互作用图进行学习，提取药物与细胞系的深层特征。最后融合两种图的学习成果进行药物的协同预测。

### 2.2.1. 网络构建

本研究以药物的分子特征，细胞系的基因表达特征以及药物的协同 Loewe 评分为输入。

对于所有药物的初始特征矩阵  $X_d$ ，我们通过获取单个药物的 SMILES 字符串，利用 DeepChem 数据库的 ConvMolFeaturizer 方法，将药物转换成分子图  $G = (X, A)$  的形式，其中  $X \in R^{n \times d}$  为单个药物所有原子的特征矩阵， $A \in R^{n \times n}$  为邻接矩阵，若  $A_{ij} = 1$  则表示第  $i$  与第  $j$  个原子有化学键相连通，若  $A_{ij} = 0$  反之。其中  $n$  表示药物原子的个数， $d = 75$  表示每个原子的特征维度。我们通过图神经网络对药物分子图进行初步处理获得模型的初始特征  $X_d$ ，如公式(2)所示，经过  $k$  层图神经网络之后第  $d$  个药物的特征矩阵为  $H_d \in R^{n \times F}$ 。

$$H_d^k = \sigma(\hat{A}_d H_d^{k-1} W_d^k) \quad (2)$$

其中， $W_d^k$  为第  $k$  层 GCN 的可学习权重矩阵； $\hat{A}_d = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$  为归一化邻接矩阵， $D$  为度矩阵， $I$  为单位矩阵； $H_d^{k-1}$  为第  $k-1$  层的输出，当  $k=1$  时， $H_d^0$  表示药物分子的单个初始特征矩阵； $\sigma$  为 ReLU 激活函数。由于药物分子具有不同的原子个数，所以我们对最后一层神经网络的输出应用全局最大池化层，使得每个药物具有相同的特征维度。最后，将所有药物的特征输出串联得到模型药物的初始特征  $X_d$ 。

对于细胞系的初始特征矩阵  $X_c$ ，我们使用如公式(3)所示的  $k$  层全连接层捕获各个细胞系的潜在特征，经过  $k$  层全连接层后第  $c$  细胞系的特征可表示为  $f_c \in R^F$ ，其中特征维度与各药物的特征向量维度保持一致。

$$f_c^k = \sigma(f_c^{k-1} W_c^k) \quad (3)$$

其中， $W_c^k$  表示第  $k$  层可学习的权重矩阵； $f_c^{k-1}$  为第  $k-1$  层全连接层的输出，当  $k=1$  时， $f_c^0$  表示细胞系

的基因表达数据； $\sigma$  为  $\tanh$  激活函数。与药物处理相似，将各细胞系的输出串联，获得模型细胞系的初始特征  $X_c$ 。

药物协同超图构建我们用药物协同作用数据构建药物协同超图  $G_H = (V, E, W)$ 。  $V = D \cup C$  表示药物和细胞系两个集合组成的节点集 ( $D$  表示药物组成的集合,  $C$  表示细胞系组成的集合);  $E \in D \times D \times C$  表示超边组成的集合, 每一个超边由一个具有协同作用的药物 - 药物 - 细胞系三元组组合而成; 每一个超边具有各自的权重, 所有权重储存于对角矩阵  $W \in |E| \times |E|$ 。超图  $G_H$  可以进一步用关联矩阵  $Y \in R^{|V| \times |E|}$ , 若  $Y_{ij} = 1$ , 则表示节点  $v \in V$  被超边  $e \in E$  关联, 若  $Y_{ij} = 0$ , 反之。超图  $G_H$  的每一个节点具有一个初始的特征向量, 所有节点组成的特征矩阵为  $X = \begin{pmatrix} X_d \\ X_c \end{pmatrix} \in R^{|V| \times F}$ , 其中  $X_d$  和  $X_c$  分别表示药物和细胞系的初始特征矩阵,  $F$  表示特征维度。

同时, 我们利用药物 Loewe 协同作用评分构造药物与细胞系相互作用网络  $G_R = (V, E)$ 。  $V = D \cup C$  表示药物和细胞系两个集合组成的节点集。  $E$  表示药物与细胞系的相互作用边, 对于图  $G_R$  的任意边  $e \in E$  表示药物  $d_i$  与细胞系  $c_j$  具有相互作用, 即存在一个包含  $d_i$  和  $c_j$  的三元组  $d_k - d_i - c_j$ , 使得 Loewe 协同评分大于阈值 30, 则称药物  $d_i$  与细胞系  $c_j$  具有相互作用。相互作用图  $G_R$  可用邻接矩阵  $A \in R^{|V| \times |V|}$  表示, 若  $A_{ij} = 1$  则表示药物  $i$  与细胞系  $j$  存在相互作用, 若  $A_{ij} = 0$  反之。图  $G_R$  节点的初始特征矩阵同超图  $G_H$  的初始特征矩阵  $X = \begin{pmatrix} X_d \\ X_c \end{pmatrix} \in R^{|V| \times F}$ 。

### 2.2.2. 特征学习

对于超图  $G_H = (V, E, W)$ , 我们使用超图神经网络学习节点的潜在特征, HGNN 是基于超图的谱卷积, 相对于传统图神经网络, HGNN 能更好的细化和学习超图结构的图网络。我们使用公式(4)得到超图神经网络的第  $k$  层输出:

$$Z^l = \sigma \left( O^{-1} Y W B^{-1} Y^T Z^{l-1} \epsilon^l \right) \quad (4)$$

其中,  $O \in R^{|V| \times |V|}$  ( $O_{vv} = \sum_{e=1}^{|E|} W_{ee} Y_{ve}$ ) 为节点的度矩阵;  $B \in R^{|E| \times |E|}$  ( $B_{ee} = \sum_{v=1}^{|V|} Y_{ve}$ ) 为超边的度矩阵;  $\epsilon^l$  是第  $l$

层的可学习权重;  $Z^l$  为第  $l$  层 HGNN 的输出, 当  $l=0$  时,  $Z^0 = X$ 。经过  $l$  层 HGNN 的迭代可聚合由共同超边连接的  $L$ -hop 邻居的信息, 从而更新超图节点的特征, 使得模型充分提取药物组合与细胞系之间的拓扑结构。令  $Z = Z^l$  表示最后一层 HGNN 的输出, 则通过矩阵分解我们可以得到药物节点  $d$  和细胞系节点的特征, 分别为  $Z_d = [d, :]$  和  $Z_c = [c, :]$ 。

针对  $G_R = (V, E)$ , 我们使用图卷积神经网络提取药物与细胞系之间的隐藏拓扑结构, 如公式(5), 第  $k$  层图卷积神经网络输出为  $H^k$ :

$$H^k = \sigma \left( \tilde{D}^{-\frac{1}{2}} \tilde{A}^{-\frac{1}{2}} \tilde{D}^{-\frac{1}{2}} H^{k-1} W^k \right) \quad (5)$$

其中,  $\sigma$  为 Relu 激活函数;  $H^k$  为第  $k$  层图卷积神经网络的输出, 当  $k=0$  时, 有  $H^0 = X$ ;  $\tilde{A} = A + I$ ;  $\tilde{D}$  是  $\tilde{A}$  的度对角矩阵;  $W^k$  表示 GCN 第  $k$  层的可学习权重矩阵。令  $H = H^k$  表示最后一层 GCN 的输出, 则通过矩阵分解我们可以得到药物节点  $d$  和细胞系节点的特征, 分别为  $H_d = [d, :]$  和  $H_c = [c, :]$ 。通过 GCN 模块, 我们可以捕获药物与细胞系的相互关联, 获得节点的深层特征。

基于多种图神经网络的学习模块, 我们不仅能提取药物组合与细胞系之间的联系, 还能捕获单个药物与细胞系之间的关联, 从而使得我们的模型能更好的提取不同实体之间的拓扑结构, 提高模型的稳定性和精度。

### 2.2.3. 预测

药物协同预测的目标是预测给定药物 - 药物 - 细胞系三元组合的 Loewe 协同作用评分(或概率), 基于不同类型的图, 我们构造了两种不同类型的预测模块预测药物组合 $[d_k, d_i]$ 与细胞系 $c_j$ 的协同评分。

对于超图神经网络得到的特征结果, 我们使用公式(6)获得药物组合 $[d_k, d_i]$ 与细胞系 $c_j$ 的协同概率 $\hat{p}_{kij}^H$ :

$$\hat{p}_{kij} = MLP(z_{d_k} \parallel z_{d_i} \parallel z_{c_j}) \quad (6)$$

其中  $MLP$  是一个多层感知机;  $\parallel$ 为拼接操作;  $z_{d_i}$  与  $z_{c_j}$  分别为药物  $i$  与细胞系  $j$  的超图神经网络输出结果。同时, 由于本研究为分类任务, 所以我们采用公式(7)二值交叉熵作为超图模块的损失函数:

$$\ell_H = -\frac{1}{|\tau|} \sum_{(d_k, d_i, c_j) \in \tau} (p_{kij} \log \hat{p}_{kij} + (1 - p_{kij}) \log (1 - \hat{p}_{kij})) \quad (7)$$

其中,  $\tau$  是训练集,  $p_{kij}$  表示真实协同标签。

针对相互作用图模型的 GCN 训练特征结果, 由于相互作用图提取的是单药物与细胞系之间的联系和相关性, 因此我们通过计算药物组合 $[d_k, d_i]$ 与细胞系 $c_j$ 的线性相关系数来预测药物是否有协同作用, 我们先通过公式(8)将药物组合的特征维度转换成与细胞系相同的特征维度:

$$h_{d_{ki}} = MLP(h_{d_k} \parallel h_{d_i}) \quad (8)$$

然后通过公式(9)计算药物组合 $[d_k, d_i]$ 与细胞系 $c_j$ 的相关系数:

$$Corr(d_{ki}, c_j) = \frac{(h_{c_j} - \mu_j)(h_{d_{ki}} - \mu_{ki})^T}{\sqrt{(h_{c_j} - \mu_j)(h_{c_j} - \mu_j)^T (h_{d_{ki}} - \mu_{ki})(h_{d_{ki}} - \mu_{ki})^T}} \quad (9)$$

其中,  $h_{c_j}$  为细胞系的图卷积神经网络输出结果,  $h_{d_{ki}}$  为药物组合 $[d_k, d_i]$ 的特征,  $\mu_j$  和  $\mu_{ki}$  为  $h_{c_j}$  和  $h_{d_{ki}}$  的平均值。为了凸显药物组合与细胞系之间的相似性, 我们引入超参  $\gamma$ 。由公式(10)可得药物组合与细胞系的线性相关矩阵  $M$ , 其中  $M_{kij}$  表示药物组合与 $[d_k, d_i]$ 细胞系  $c_j$  的敏感性。

$$M_{kij} = \gamma Corr(d_{ki}, c_j) \quad (10)$$

与超图模块的损失函数类似, 我们采用公式(11)二值交叉熵作为损失函数

$$\ell_R = -\frac{1}{|\tau|} \sum_{(d_k, d_i, c_j) \in \tau} (p_{kij} \log \hat{M}_{kij} + (1 - p_{kij}) \log (1 - \hat{M}_{kij})) \quad (11)$$

最后, 为了平衡两个图模块的贡献度, 我们将单药物与细胞系的相似性作为辅助任务设置为正则器, 通过设置超参数  $\alpha$  结合公式(7)与公式(11)对总损失函数公式(12)进行迭代优化:

$$\ell = \alpha \ell_H + (1 - \alpha) \ell_R \quad (12)$$

## 3. 实验

### 3.1. 基线

为了进行结果对比, 我们将我们的模型与当前最先进的药物协同作用预测模型进行对比。

Li 等人[23]提出了一种从特征集中提取特征的网络传播方法, 并利用该方法训练随机森林模型来预测药物协同作用。

ComboFM [24]模型结合剂量反应数据、药物化学特征和细胞系基因表达数据进行建模，并使用高阶因子分解机器学习输入数据的潜在特征以预测药物协同作用。

DTF [25]模型通过张量分解的方法从药物协同矩阵中提取潜在特征，并借助 DNN 模型进行药物协同作用的预测。

DeepSynergy 模型则采用前馈神经网络来预测药物协同评分。

HypergraphSynergy 模型构建了药物组合协同作用超图，并使用多层感知机训练超图神经网络以学习得到特征并预测药物协同作用。

## 3.2. 实验

### 3.2.1. 五倍交叉验证实验

为了确保实验结果的公平性和准确性，我们对模型进行了基线测试，并采用了五倍交叉验证方法。我们首先从数据集中随机选择 90%的样本作为交叉验证集，剩余的样本作为独立的测试集。在交叉验证集上，我们采用了五倍交叉验证的方法，将样本随机分成五等份，以验证模型在重新发现已知抗癌药物协同作用方面的能力。

见表 1 可以看出，我们的模型在指定指标上优于其他模型。这得益于我们的模型不仅能够提取药物组合与细胞系之间的关联，还能捕获单个药物与细胞系之间的联系。同时，我们的模型基于多层感知机和线性相关系数来进行药物协同预测，表明我们的模型具有良好的泛化性和强大的鲁棒性。通过与 Li 等人的 ComboFM 和 DTF 模型的对比，我们可以看出深度学习模型在药物协同作用预测任务上具有显著的优势。我们的模型相对于 DeepSynergy 模型的优势主要在于图神经网络模型能够更好地提取药物与细胞系之间的信息。同样，由于我们的模型不仅提取了药物组合与细胞系之间的信息，还提取了单个药物与细胞系之间的关联，因此相对于同样使用图神经网络的 HypergraphSynergy 模型，我们的模型也具有优势。

**Table 1.** Results of five-fold cross experiment

**表 1.** 五倍交叉实验结果

模型	AUC	AUPR	ACC	F1_score
Li 的模型	0.8750	0.5141	0.5181	0.9110
ComboFM	0.8945	0.5362	0.4655	0.9254
DTF	0.9138	0.5829	0.5828	0.9253
DeepSynergy	0.9060	0.5638	0.5445	0.9165
HypergraphSynergy	0.9230	0.6328	0.6025	0.9254
<b>ours</b>	<b>0.9258</b>	<b>0.6515</b>	<b>0.6348</b>	<b>0.9261</b>

### 3.2.2. 新药实验

为了验证模型在新药物组合和细胞系维度上的预测能力，我们选择将样本数据按照药物组合的维度进行划分，以确保测试集仅包含训练集中未出现的药物组合，从而评估模型对新药物组合的预测能力。在新药物组合实验对比中，我们选择在基线实验中表现最佳的 HypergraphSynergy 模型与我们的模型进行对比。结果见图 2 所示，我们的模型在所有指标上均优于 HypergraphSynergy 模型，这表明我们的模型具有更高的泛化能力和鲁棒性，使得在极端条件下我们的模型仍然能够保持较高的稳定性。

### 3.2.3. 新细胞系实验

为了验证我们的模型在极端条件下(如预测未知癌细胞)的高精度性能，我们采用了特定的数据集划分

方法。我们通过在细胞系维度上对数据集进行分割，确保训练集中的细胞系不会出现在测试集中，以模拟新细胞系预测的环境条件。

在与 HypergraphSynergy 模型进行对比的实验中，我们的模型展现出了更优秀的性能。见图 3 所示，我们的模型在列出的四个指标上均超过了对比模型。这可以归因于我们的模型捕获了更深层次的药物与细胞系之间的拓扑特征，从而使得我们的模型具备了更强的泛化能力，适用于更复杂的预测环境。

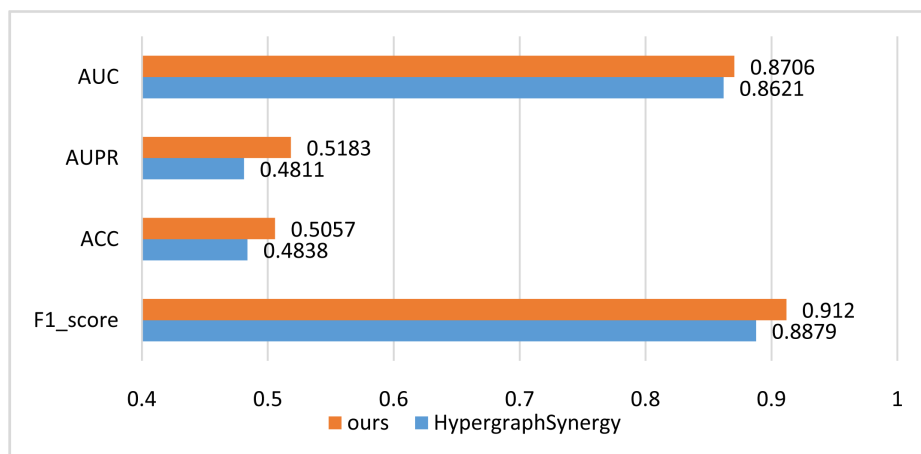


Figure 2. New drug experimental results

图 2. 新药实验结果

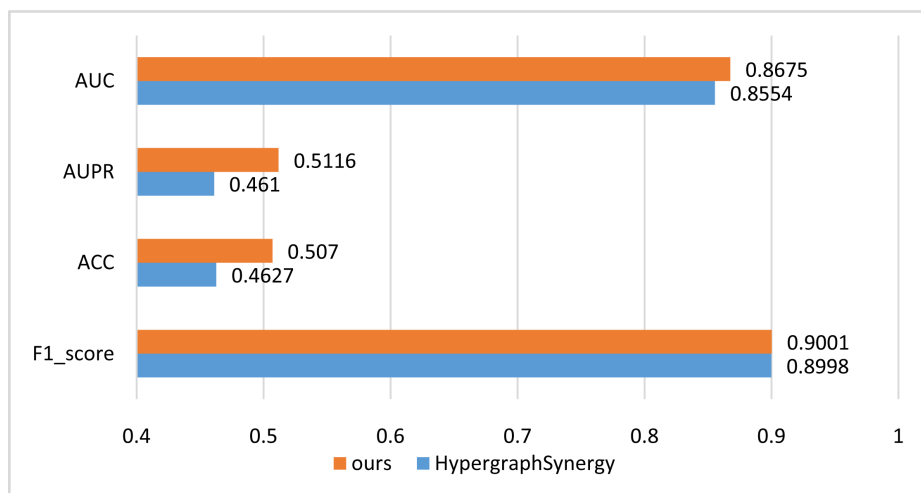


Figure 3. Results of new cell experiment

图 3. 新细胞实验结果

### 3.2.4. 消融实验

为了验证模型中各个模块的贡献度，我们对原始模型进行了两个变种实验：

1) 去超图：在原始模型的基础上，我们剔除了超图模块，并直接使用药物和细胞系的初始特征作为输入，通过多层感知机进行药物协同作用的预测。

2) 去相互作用图：与去超图类似，我们去除了相互作用图模块，而是直接使用药物和细胞系的初始特征来计算药物与细胞系之间的线性相关系数，从而预测协同作用。

通过使用相互作用五倍交叉验证方法，我们得到了实验结果见表 2 所示。可以观察到，在所有变种



模型中，精度都有所下降，这说明我们模型的各个模块都是必要的，它们是保证模型鲁棒性和泛化能力的关键因素。具体地说，原始模型相较于去超图模型的 AUC 值提高了 0.0271，这归功于我们构建的超图网络以及超图神经网络对其进行学习的能力，它们能够提取药物组合与细胞系之间潜在的拓扑特征和联系，从而提高了模型的预测精度。同时，原始模型相对于去相互作用图模型也有显著优势，这表明相互作用图能够有效地捕获到单个药物与细胞系之间的联系。两种不同类型的图结构极大地挖掘了不同实体之间的联系，使得我们的模型能够达到最佳性能。

**Table 2.** Results of ablation experiments

**表 2.** 消融实验结果

模型	AUC	AUPR	ACC	F1_score
去超图	0.8985	0.5934	0.5688	0.9270
去相互作用图	0.9249	0.6137	0.5913	0.9226
<b>ours</b>	<b>0.9258</b>	<b>0.6515</b>	<b>0.6348</b>	<b>0.9261</b>

#### 4. 结语

现有的药物协同作用预测模型利用机器学习或深度学习技术，通过提取实体特征进行预测。然而，这些模型在挖掘药物与细胞系特征方面存在不足，要么忽略了药物组合与细胞系之间的相互作用，要么忽略了单个药物与细胞系之间的联系。针对这些问题，本研究构建了药物组合与细胞系相互作用的超图和单个药物与细胞系相互作用的网络。我们分别采用超图神经网络和图卷积神经网络对这两种图进行训练，以提取药物与细胞系的深层特征。接着，我们通过多层感知机和线性相关系数为不同的图模型预测药物协同作用。实验证明了该模型的有效性，并证实在极端条件下能够成功预测新药物与新细胞系，突出了其具备良好的泛化性和鲁棒性。

#### 参考文献

- [1] Huang, M., Shen, A., Ding, J., *et al.* (2014) Molecularly Targeted Cancer Therapy: Some Lessons from the Past Decade. *Trends in Pharmacological Sciences*, **35**, 41-50. <https://doi.org/10.1016/j.tips.2013.11.004>
- [2] Grilli, R., Oxman, A.D. and Julian, J.A. (1993) Chemotherapy for Advanced Non-Small-Cell Lung Cancer: How Much Benefit Is Enough? *Journal of Clinical Oncology*, **11**, 1866-1872. <https://doi.org/10.1200/JCO.1993.11.10.1866>
- [3] Sandler, A., Gray, R., Perry, M.C., *et al.* (2006) Paclitaxel-Carboplatin Alone or with Bevacizumab for Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, **355**, 2542-2550. <https://doi.org/10.1056/NEJMoa061884>
- [4] Gandhi, L., Rodriguez-Abreu, D., Gadgeel, S., *et al.* (2018) Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, **378**, 2078-2092. <https://doi.org/10.1056/NEJMoa1801005>
- [5] Paz-Ares, L., Luft, A., Vicente, D., *et al.* (2018) Pembrolizumab plus Chemotherapy for Squamous Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, **379**, 2040-2051. <https://doi.org/10.1056/NEJMoa1810865>
- [6] Thai, A.A., Solomon, B.J., Sequist, L.V., *et al.* (2021) Lung Cancer. *The Lancet*, **398**, 535-554. [https://doi.org/10.1016/S0140-6736\(21\)00312-3](https://doi.org/10.1016/S0140-6736(21)00312-3)
- [7] Borghaei, H., Paz-Ares, L., Horn, L., *et al.* (2015) Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, **373**, 1627-1639. <https://doi.org/10.1056/NEJMoa1507643>
- [8] Brahmer, J., Reckamp, K.L., Baas, P., *et al.* (2015) Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, **373**, 123-135. <https://doi.org/10.1056/NEJMoa1504627>
- [9] Garon, E.B., Rizvi, N.A., Hui, R., *et al.* (2015) Pembrolizumab for the Treatment of Non-Small-Cell Lung Cancer. *The New England Journal of Medicine*, **372**, 2018-2028. <https://doi.org/10.1056/NEJMoa1501824>

- [10] O'Neil, J., Benita, Y., *et al.* (2016) An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. *Molecular Cancer Therapeutics*, **15**, 1155-1162.
- [11] Holbeck, S.L., Camalier, R., Crowell, J.A., Govindharajulu, J.P., Hollingshead, M., Anderson, L.W., *et al.* (2017) The National Cancer Institute Almanac: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs with Enhanced Therapeutic Activity. *Cancer Research*, **77**, 3564-3576.  
<https://doi.org/10.1158/0008-5472.CAN-17-0489>
- [12] Menden, M.P., Bulusu, K.C., Guan, Y., *et al.* (2019) Community Assessment to Advance Computational Prediction of Cancer Drug Combinations in a Pharmacogenomic Screen. *Nature Communications*, **10**, Article No. 2674.
- [13] Chang, Y., Park, H., Yang, H., *et al.* (2018) Cancer Drug Response Profile Scan (CDRscan): A Deep Learning Model that Predicts Drug Effectiveness from Cancer Genomic Signature. *Scientific Reports*, **8**, Article No. 8857.
- [14] Liu, H., Zhang, W., Zou, B., Wang, J., Deng, Y. and Deng, L. (2020) DrugCombDB: A Comprehensive Database of Drug Combinations toward the Discovery of Combinatorial Therapy. *Nucleic Acids Research*, **48**, D871-D881.  
<https://doi.org/10.1093/nar/gkz1007>
- [15] Wang, T., Szedmak, S., Wang, H., Aittokallio, T., Pahikkala, T., Cichonska, A. and Rousu, J. (2021) Modeling Drug Combination Effects via Latent Tensor Reconstruction. *Bioinformatics*, **37**, i93-i101.  
<https://doi.org/10.1093/bioinformatics/btab308>
- [16] Preuer, K., Lewis, R.P.I., Hochreiter, S., Bender, A., Bulusu, K.C. and Klambauer, G. (2018) DeepSynergy: Predicting Anti-Cancer Drug Synergy with Deep Learning. *Bioinformatics*, **34**, 1538-1546.  
<https://doi.org/10.1093/bioinformatics/btx806>
- [17] Bazgir, O., Ghosh, S. and Pal, R. (2021) Investigation of REFINED CNN Ensemble Learning for Anti-Cancer Drug Sensitivity Prediction. *Bioinformatics*, **37**, i42-i50. <https://doi.org/10.1093/bioinformatics/btab336>
- [18] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks.  
<https://arxiv.org/abs/1609.02907>
- [19] Liu, X., Song, C.Z., Liu, S.C., Li, M.L., Zhou, X.H. and Hang, W. (2022) Multi-Way Relation-Enhanced Hypergraph Representation Learning for Anti-Cancer Drug Synergy Prediction. *Bioinformatics*, **38**, 4782-4789.  
<https://doi.org/10.1093/bioinformatics/btac579>
- [20] Bai, S., Zhang, C., Liu, L., Yao, Y., Zhang, X. and Cheng, J. (2021) Hypergraph Convolution and Hypergraph Attention. *Pattern Recognition*, **110**, Article ID: 107637. <https://doi.org/10.1016/j.patcog.2020.107637>
- [21] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., Zaslavsky, L., Zhang, J. and Bolton, E.E. (2019) PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Research*, **47**, D1102-D1109. <https://doi.org/10.1093/nar/gky1033>
- [22] Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C.G., Ward, S., Kok, C.Y., Jia, M., De, T., Teague, J.W., Stratton, M.R., McDermott, U. and Campbell, P.J. (2015) COSMIC: Exploring the World's Knowledge of Somatic Mutations in Human Cancer. *Nucleic Acids Research*, **43**, D805-D811.  
<https://doi.org/10.1093/nar/gku1075>
- [23] Li, H., Zeng, X., Liu, B., Luo, X., Yu, Q. and Xu, X. (2018) Network Propagation Predicts Drug Synergy in Cancers. *Cancer Research*, **78**, 5446-5457. <https://doi.org/10.1158/0008-5472.CAN-18-0740>
- [24] Julkunen, H., Rantanen, A.J., Barrett, A., Gomez, M., Gopalakrishnan, B., Harju, M., Hellman, T., Huhtala, S., Karlsson, A., Kallio, A., Laasonen, L., Lehtonen, R., Merisaari, J., Pahikkala, T., Parvinen, P., Ruusuvuori, P., Salo, H., Szedmak, S., Talmane, L. and Aittokallio, T. (2020) Leveraging Multi-Way Interactions for Systematic Prediction of Pre-Clinical Drug Combination Effects. *Nature Communications*, **11**, Article No. 6136.  
<https://doi.org/10.1038/s41467-020-19950-z>
- [25] Sun, Z., Cheng, F., Xu, J., Song, X., Su, R., Chen, Y., Tu, J., Yao, L., Liu, B., Wang, Y., Pan, P., Li, Y. and Cao, Z. (2020) DTF: Deep Tensor Factorization for Predicting Anticancer Drug Synergy. *Bioinformatics*, **36**, 4483-4489.  
<https://doi.org/10.1093/bioinformatics/btaa287>