DOI：10.20271/j.cnki.1673−9833.2026.4004

基于迁移学习的肿瘤耐药细胞识别方法

黄竣峰[1]，彭利红[2]

（1. 湖南工业大学 生物与医学工程学院，湖南 株洲 412007；
2. 湖南工业大学 计算机与人工智能学院，湖南 株洲 412007）

摘　要：针对肿瘤异质性导致患者对相同药物治疗反应存在显著差异，进而引发肿瘤复发或转移的问题，提出一种基于图结构与无监督域适应的肿瘤耐药细胞识别方法 GCDrug。该方法首先基于 k 最近邻分别构建源域（批量转录组测序）与目标域（单细胞转录组测序）样本的相似性图，利用图卷积网络提取特征；随后引入 Deep CORAL 损失，实现跨域分布自适应。在 10 个单细胞药物注释数据集上的实验结果表明，GCDrug 在 3 项分类指标上均优于现有主流方法，其中，在 Etoposide 数据集上的 F_1 分数达 0.956，显著优于现有方法。消融实验进一步验证：图结构模块与域适应模块的协同作用，能提升模型的判别性能与泛化能力。实验结果表明，所提方法能够准确识别肿瘤中的耐药个体细胞。

关键词：肿瘤异质性；无监督域适应；图结构；耐药细胞识别

中图分类号：TP18；TP317.4　　**文献标志码**：A　　**文章编号**：1673−9833(2026)04−0026−06

引文格式：黄竣峰，彭利红 . 基于迁移学习的肿瘤耐药细胞识别方法 [J]. 湖南工业大学学报，2026，40(4)：26−31.

A Tumor Drug-Resistant Cell Identification Method Based on Transfer Learning

HUANG Junfeng[1]，PENG Lihong[2]

（1. School of Biological Science and Medical Engineering，Hunan University of Technology，Zhuzhou Hunan 412007，China；
2. School of Computer Science and Artificial Intelligence，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：In view of the flaw that tumor heterogeneity causes patients to respond differently to the same drug therapy, thus leading to tumor relapse or metastasis, a tumor drug-resistant cell recognition method GCDrug has been proposed based on graph structure and unsupervised domain adaptation. Firstly, similarity maps are to be constructed between source domain (batch transcriptome sequencing) and target domain (single-cell transcriptome sequencing) samples based on k-nearest neighbors, thus extracting features using graph convolutional networks. Subsequently, Deep CORAL loss is introduced to achieve cross-domain distribution adaptation. Experimental results on 10 single-cell drug-annotated datasets demonstrate that GCDrug outperforms current mainstream methods across three classification metrics. Notably, on the Etoposide dataset, GCDrug achieves an F_1-score of 0.956, significantly outperforming existing methods. Ablation experiments further verify that the synergistic effect of the graph structure module and the domain adaptation module helps to improve the discriminative performance and generalization ability of the model. The experimental results indicate that the proposed method can accurately identify drug-resistant individual cells in tumors.

Keywords：tumor heterogeneity；unsupervised domain adaptation；graph structure；drug-resistant cell identification

收稿日期：2025−10−16

作者简介：黄竣峰，男，湖南工业大学硕士生，主要研究方向为生物医学工程，E-mail：m23083100009@stu.hut.edu.cn
通信作者：彭利红，女，湖南工业大学教授，主要研究方向为医学人工智能，E-mail：plhhnu@163.com

1　研究背景

癌症是当前最严峻的健康挑战之一，由异常细胞失控性增殖驱动导致 [1]。这种异常增殖可侵袭并破坏正常的组织结构，进而引发多种疾病 [2]。其中，肿瘤的异质性通常会让患同种癌症的不同患者对相同药物治疗的反应有着较大差异 [3]。尽管近年来抗癌治疗取得了显著进展，但肿瘤内耐药细胞的存活与增殖仍是导致肿瘤复发或转移的关键因素 [4]。在众多癌症治疗临床失败的案例中，耐药性往往是其中的主要原因 [5]。现如今，多个大型药物筛选项目，如 GDSC（genomics of drugs sensitivity in cancer）[6−7]、CCLE（cancer cell line encyclopedia）[8] 等，系统性地测试了数百种药物在大量肿瘤细胞系上的反应，并配套提供了多组学数据，这些资源为计算模型提供了训练数据，为大规模识别肿瘤内的耐药细胞提供了数据基础。

当前，传统机器学习方法已被广泛用于肿瘤内对耐药细胞的识别，并取得了一定的进展 [9]。结合现有药物筛选细胞系数据，已有一些研究从批量转录组测序（bulk RNA-seq）数据联合单细胞 RNA 测序（scRNA-seq）数据，从单细胞水平上揭示肿瘤耐药个体。如 Chen J. Y. 等 [10] 提出一个深度迁移学习框架 scDEAL，其通过整合大规模 bulk RNA-seq 数据，协调与肿瘤药物相关的 bulk RNA-seq 与 scRNA-seq 数据，并将在 bulk RNA-seq 数据上训练的模型迁移至 scRNA-seq 数据上，识别单细胞水平上的肿瘤耐药细胞。Zheng Z. T. 等 [11] 提出了一种整合对抗域适应的单细胞药物反应预测框架（SCAD）。该框架从大型公共数据集中的群体细胞系中学习批量转录组图谱和药物基因组学信息，并将这些知识转移到推断单个细胞的药物疗效。另外，Huang K. S. 等 [12] 提出一种推断耐药癌细胞的半监督少镜头迁移学习（SSDA4Drug）方法，这是一种半监督的小样本迁移学习方法，可用于推断耐药肿瘤细胞。SSDA4Drug 利用半监督的对抗性结构域适应从 bulk RNA-seq 和 scRNA-seq 数据中提取药物基因组特征，能将癌症药物敏感性知识从批量细胞系转移到单个细胞中。He D. 等 [13] 开发了一种新的上下文感知解压自动编码器（CODE-AE），其显著提高了从细胞系数据中预测患者特定临床药物反应的精确度和稳健性。尽管这些方法对于肿瘤内耐药细胞识别取得了一定成果，然而上述方法均未能充分考虑 scRNA-seq 数据样本数量稀缺、分布偏移显著等特性，且在对于 bulk RNA-seq 和 scRNA-seq 数据的处理方面，忽略了每个数据集中样本间的相似度关联，使得模型的预测准确性和泛化性存在一定问题。

为解决上述问题，本文提出一种用于识别耐药肿瘤细胞的基于图无监督域适应方法（GCDrug），即基于图结构与无监督域适应的肿瘤耐药细胞识别方法。实验结果表明该方法能将 bulk RNA-seq 数据构建的肿瘤耐药细胞识别模型高效、可靠地迁移至 scRNA-seq 数据上，从而完成对 scRNA-seq 数据进行肿瘤耐药细胞识别。本文主要贡献：1）分别根据源域（bulk RNA-seq 数据），目标域（scRNA-seq 数据）的域内基因表达谱相似性，构建图结构；2）迁移任务上采用 Deep CORAL（Deep CORrelation ALignment）[14] 方法，减小源域和目标域特征间的二阶统计量（协方差）域间差异，从而达到全局域适应，提高识别模型在目标域上的准确性和泛化性。

2　方法

2.1　数据来源及预处理

1）源域数据。批量细胞系药物注释数据来自癌症药物敏感性基因组学数据库（GDSC[6−7]），该数据库提供了 1 074 种癌细胞系对 226 种药物的反应信息。其中肿瘤药物耐药标签通过半数抑制浓度（IC50）和剂量－反应曲线下面积（AUC）进行评估。参考了 CODE-AE[13] 的方法，为每种细胞系分配了药物反应标签。将所有细胞系对该药物的 IC50 值进行排序，然后以所有细胞系对该药物的 IC50 平均值为阈值，将细胞系分为两类。敏感的细胞系标记 1，耐药的细胞系标记 0。同时，还从 GDSC 数据库中获取了这些细胞系对应的 bulk RNA-seq 数据。

2）目标域数据。总体目标域数据情况见表 1。

表 1　单细胞药物注释数据集的详细信息

Table 1　Detailed information of single-cell drug-response datasets

药物名	数据集	细胞系	基因数	癌症类型
Afatinib	CCLE	SCC47	2 876	口腔鳞状细胞癌
AR-42	CCLE	JUH006	2 933	黑色素瘤
Cetuximab	CCLE	SCC47	2 876	口腔鳞状细胞癌
Etoposide	GSE149215	PC9	2 668	肺癌
Gefitinib	CCLE	JUH006	2 933	口腔鳞状细胞癌
NVP-TAE684	CCLE	SCC47	2 876	口腔鳞状细胞癌
PLX4720[1]	CCLE	451Lu	2 856	口腔鳞状细胞癌
PLX4720[2]	GSE108383	A375	2 873	口腔鳞状细胞癌
Sorafenib	CCLE	SCC47	2 876	口腔鳞状细胞癌
Vorinostat	CCLE	JUH006	2 933	口腔鳞状细胞癌

注：上标 [1, 2] 代表源于不同数据集，下同。

目标域数据中，单细胞药物注释数据收集来自 5 个不同细胞系的 10 种单细胞药物反应数据集，这些

细胞系接受 9 种化疗药物和靶向药物治疗[11]。数据预处理时，利用 Scanpy[15] 的 Highly_Variable_gene 函数筛选高变基因，保留与药物敏感性相关信息最丰富的基因子集。

2.2 GCDrug 方法

2.2.1 相似性图结构构建

1）源域。使用 GDSC 数据库作为源域，表示为 $D_{\mathrm{s}}=\left\{\left(x_i^{\mathrm{s}},\ y_i^{\mathrm{s}}\right)\right\}_{i=1}^{n_{\mathrm{s}}}$，其中 s 代表源域 (基于细胞系药物反应前)，i 为细胞系；$x_i^{\mathrm{s}}\in\mathbf{R}^{d_{\mathrm{s}}\times n_{\mathrm{s}}}$ 表示细胞系的基因表达，d_{s} 为源域基因数，n_{s} 为细胞系数；y_i^{s} 为每个细胞系的二值化药物反应标签，即敏感（y_i^{s}=1）或耐药（y_i^{s}=0）。利用 K 最近邻[16]（K-nearest neighbors，KNN）算法构建细胞系相似性图。在图 A_{s} 中，每个节点 v_i 代表一个细胞系 i，如果两个节点 v_i 和 v_j 相连，则意味着对应的基因表达谱 x_i^{s} 和 x_j^{s} 相似。

2）目标域。scRNA-seq 数据依照源域的数据构建方法，表示为 $D_{\mathrm{t}}=\left\{\left(x_i^{\mathrm{t}},\ y_i^{\mathrm{t}}\right)\right\}_{i=1}^{n_{\mathrm{t}}}$，其中 t 代表目标域，$D_{\mathrm{t}}$ 为目标域基因数，n_{t} 为单细胞数量。在构建单细胞相似性图之后，将目标域数据划分为训练集和测试集（8:2）。在构建的图 A_{t} 中，每个节点 v_i 代表一个单细胞样本 i，如果两个节点 v_i 和 v_j 相连，则意味着对应的基因表达谱 x_i^{t} 和 x_j^{t} 相似。

样本间相似性图的 KNN 构建步骤具体如下。

1）特征标准化。对每个细胞系（或单细胞）的基因表达向量进行 Z-score 标准化，以消除量纲差异，标准化表达式为

$$z_{i,\,j}=\left(x_{i,\,j}-\mu_j\right)\big/\sigma_j\,, \tag{1}$$

式中：$x_{i,\,j}$ 为第 i 个样本在第 j 个基因上的表达值；μ_j 与 σ_j 分别为 j 基因在全体样本中的均值与标准差。

2）计算相似度矩阵。采用余弦相似度衡量样本间表达谱的相似性：

$$\boldsymbol{S}(i,\ j)=\left(\boldsymbol{z}_i\cdot\boldsymbol{z}_j\right)\big/\left(\|\,\boldsymbol{z}_i\,\|\|\,\boldsymbol{z}_j\,\|\right)。 \tag{2}$$

式中 $\boldsymbol{z}_i$、$\boldsymbol{z}_j\in\mathbf{R}^d$ 为两个样本的标准化表达向量。

3）构架 KNN 图。对每个样本，选取与其相似度最高的 K 个邻居（本文中参数设置 K=10），构建无向图邻接矩阵：

$$A_{i,\,j}=\begin{cases}1，\text{若样本}j\text{为样本}i\text{的}K\text{个最近邻之一；}\\0，\text{否则。}\end{cases} \tag{3}$$

最终得到的图结构为无权无向图，节点为样本，边表示表达谱高度相似。

为避免孤立节点，构建过程中强制添加自环，即 $A_{i,\,i}$=1，确保每个节点至少包含自身信息。

综上所述，对于模型输入数据的相似性图结构构建流程，如图 1 所示。

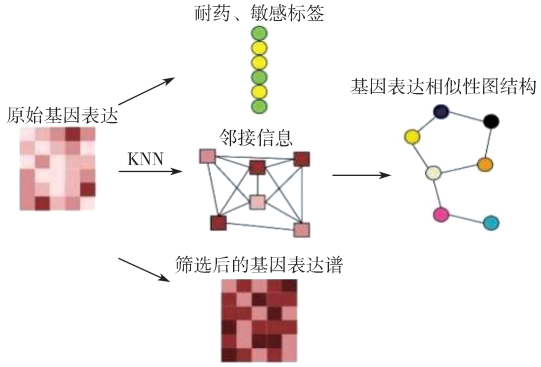


图 1 相似性图结构构建流程

Fig. 1 Flow chart of similar graph structure construction

2.2.2 特征提取器 E_{r}

为了充分挖掘空间转录组数据中样本间的拓扑结构信息，模型采用两层图卷积网络[17]（graph convolutional network，GCN）作为特征提取器。GCN 通过聚合邻居节点的特征信息，能够有效捕获节点的高阶邻域特征，从而增强模型对局部结构模式的表达能力。

给定图 $G=\left(v,\ \varepsilon\right)$，其中 v 为节点集合，ε 为边集合。设 $\boldsymbol{X}\in\mathbf{R}^{N\times d}$ 为节点特征矩阵（N 为节点数，d 为输入特征维度），$\boldsymbol{A}\in\mathbf{R}^{N\times N}$ 为邻接矩阵，则第 l 层 GCN 的前向传播公式为

$$\boldsymbol{H}^{(l+1)}=\sigma\left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{(l)}\boldsymbol{W}^{(l)}\right)。 \tag{4}$$

式中：$\tilde{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}_N$ 为添加自环的邻接矩阵；$\tilde{\boldsymbol{D}}$ 为 $\tilde{\boldsymbol{A}}$ 的度矩阵，且 $\tilde{\boldsymbol{D}}_{i,\,i}=\Sigma_j\tilde{\boldsymbol{A}}_{i,\,j}$；$\boldsymbol{W}^{(l)}\in\mathbf{R}^{d^{(l)}\times d^{(l+1)}}$ 为可学习的权重矩阵；σ (·) 为非线性激活函数（采用 ReLU[18]）；$\boldsymbol{H}^{(0)}=\boldsymbol{X}$，即初始输入为节点基因表达特征。

模块通过双层 GCN 结构实现从原始基因表达空间到 128 维嵌入空间的非线性映射，具备以下优势：1）结构感知，利用图卷积聚合邻居信息，增强细胞间的相似性表达；2）可微分，支持端到端训练，便于与后续 Deep CORAL 损失联合优化。

2.2.3 耐药细胞识别器 P_r

耐药细胞识别器是一种完全连接的分类器，旨在使用特征提取器中的潜在特征对耐药细胞进行识别。它是通过交叉熵损失最小化模型识别的源域标签和药物注释的源域标签（ground truth labels）之间的差异来训练的，其公式为

$$L_c=-\frac{1}{n_{\mathrm{s}}}\sum\nolimits_{x_i\in D_{\mathrm{s}}}\left(P_{x_i\in S}\log P_r\left(E_r\left(x_i,\ A\right)\right)+P_{x_i\in R}\log P_r\left(E_r\left(x_i,\ A\right)\right)\right), \tag{5}$$

式中：$P_{x_i\in S}$ 为 x_i 不属于耐药的概率；$P_{x_i\in R}$ 为 x_i 属于

耐药的概率；A 为源域中的图，耐药细胞识别器由源域和目标域共享。

2.2.4　无监督域适应

在跨域耐药细胞识别任务中，源域（bulk RNA-seq 数据）与目标域（scRNA-seq 数据）之间存在显著的分布差异。为了能更好地实现跨域迁移任务，本研究采用无监督域适应 Deep CORAL[14] 方法，通过对齐源域与目标域特征的二阶统计量（协方差矩阵），在特征空间中减小域间差异，从而提升模型在目标域上的泛化能力。相较于传统 CORAL[19] 方法，Deep CORAL 能够直接借助深度网络中的非线性映射，并且在整个训练流程实现端到端训练，从而实现对复杂域偏移更为鲁棒，且对于分类任务的无监督域适应有着更好的准确率。因此，设源域的特征矩阵为

$$\boldsymbol{X}_{\mathrm{s}} \in \mathbf{R}^{n_{\mathrm{s}} \times d}, \tag{6}$$

目标域的特征矩阵为

$$\boldsymbol{X}_{\mathrm{t}} \in \mathbf{R}^{n_{\mathrm{t}} \times d}, \tag{7}$$

式（6）（7）中 d 为特征维度。

则 CORAL 损失定义为

$$\mathcal{L}_{\mathrm{CORAL}} = \frac{1}{4d^2}\left\|\boldsymbol{C}_{\mathrm{s}} - \boldsymbol{C}_{\mathrm{t}}\right\|_F^2 。 \tag{8}$$

式中 $\boldsymbol{C}_{\mathrm{s}}$、$\boldsymbol{C}_{\mathrm{t}} \in \mathbf{R}^{d\times d}$ 分别为源域和目标域的协方差矩阵；$\|\cdot\|_F$ 为 Frobenius 范数（矩阵元素平方和的平方根）；系数 $\frac{1}{4d^2}$ 用于归一化，防止维度敏感。

1）协方差矩阵计算方式。对于特征矩阵 $\boldsymbol{X} \in \mathbf{R}^{n\times d}$，其协方差矩阵计算式如下：

$$\boldsymbol{C} = \frac{1}{n-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} - \frac{1}{n}\left(\mathbf{1}^{\mathrm{T}}\boldsymbol{X}\right)^{\mathrm{T}}\left(\mathbf{1}^{\mathrm{T}}\boldsymbol{X}\right)\right), \tag{9}$$

式中：$\mathbf{1} \in \mathbf{R}^{n}$ 为全为 1 的向量。

式（9）等价于先对特征做中心化处理，再计算样本协方差。

2）梯度计算（可微性）。CORAL 损失是可微的，支持反向传播。对于源域特征 $\boldsymbol{X}_{\mathrm{s}}$，其梯度为

$$\frac{\partial \mathcal{L}_{\mathrm{CORAL}}}{\partial \boldsymbol{X}_{\mathrm{s}}} = \frac{1}{d^2\left(n_{\mathrm{s}}-1\right)}\left(\boldsymbol{X}_{\mathrm{s}} - \frac{1}{n_{\mathrm{s}}}\mathbf{1}\mathbf{1}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{s}}\right)\left(\boldsymbol{C}_{\mathrm{s}} - \boldsymbol{C}_{\mathrm{t}}\right)。 \tag{10}$$

同理可得目标域梯度。该梯度可通过链式法则传递到前端网络，实现端到端训练。

3）集成的损失函数设计。在训练阶段，将 CORAL 损失与分类损失联合优化：

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{C}} + \lambda \cdot \mathcal{L}_{\mathrm{CORAL}}, \tag{11}$$

式中：$\mathcal{L}_{\mathrm{C}}$ 为源域上的交叉熵损失；λ 为自适应权重，随着训练轮次动态调整，采用式（12）所示形式。

$$\lambda = 2\Big/\left[1+\exp\left(-10\cdot\frac{epoch}{max_epoch}\right)\right]-1, \tag{12}$$

式中：$epoch$ 为模型训练的当前轮次；max_epoch 为模型训练设定的最终轮次。

通过此设计，使得初期以分类为主，保证特征判别性，后期逐步增强对齐强度，提升迁移效果。

综上所述，GCDrug 模型的整体训练流程见图 2。

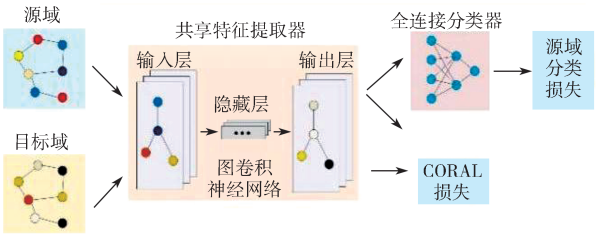


图 2　GCDrug 模型训练流程示意图

Fig. 2　GCDrug model training process diagram

3　实验结果与分析

3.1　实验环境

因为模型是轻量级，且数据集的单细胞样本数量较少，在普通的笔记本上也能够快速高效运行，故所有实验均在 Windows 11 专业版（版本为 24H2，内部版本为 26100.6584）平台上完成。其中 GPU 为 GeForce RTX 4060 laptop，内存为 16 GB。实验中，总共进行了 50 次的重复实验，且均设置不同的随机种子，以防止实验结果的偶然性，最终取 50 次实验结果的平均值作为模型性能评估依据。

3.2　对比实验结果

在 10 个单细胞药物数据集上，对于本文提出的方法与 4 个基准方法（SSDA4Drug[12]、SCAD[11]、scDEAL[10]、CODEAE[13]）进行比较实验，其中对于分类指标 F_1 分数的结果如表 2 所示，其中加粗数值为最优值。

表 2　GCDrug 与 4 个基准方法的分类 F_1 分数结果

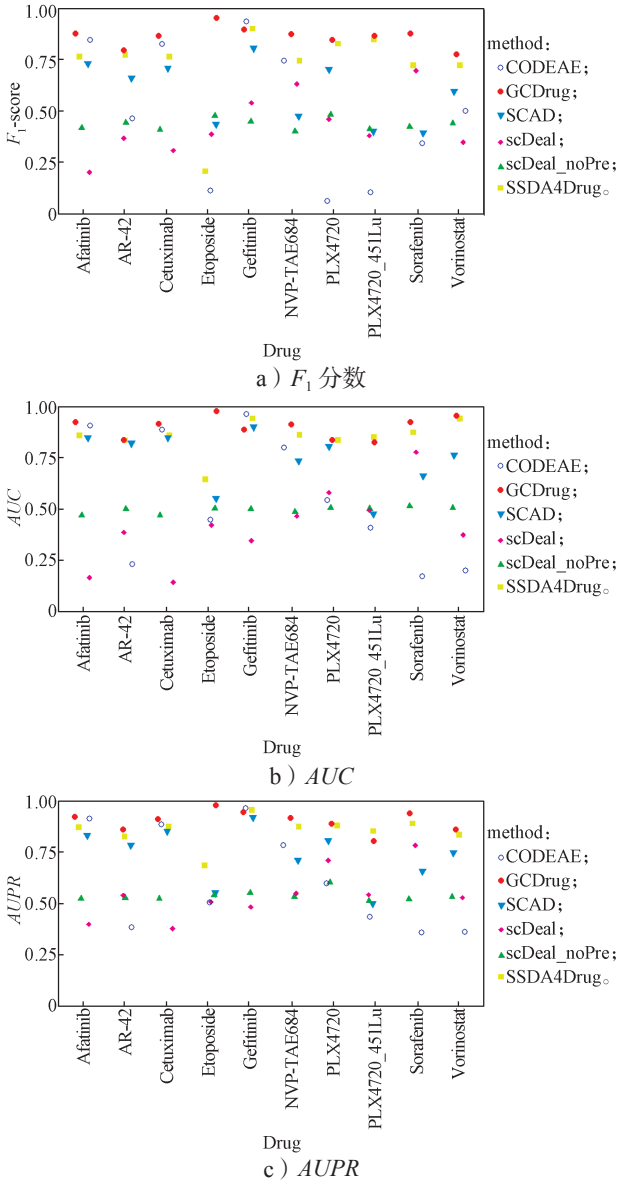
Table 2　Classification F_1-score results of GCDrug and four baseline methods

药物名	方　法				
	SSDA4Drug	SCAD	scDEAL	CODEAE	GCDrug
Afatinib	0.766	0.732	0.416	0.847	**0.878**
AR-42	0.775	0.661	0.441	0.464	**0.798**
Cetuximab	0.765	0.709	0.406	0.828	**0.867**
Etoposide	0.204	0.434	0.477	0.109	**0.956**
Gefitinib	0.903	0.807	0.540	**0.937**	0.899
NVP-TAE684	0.746	0.476	0.633	0.746	**0.876**
PLX4720[1]	0.829	0.705	0.482	0.059	**0.847**
PLX4720[2]	**0.848**	0.400	0.409	0.099	0.782
Sorafenib	0.723	0.393	0.698	0.341	**0.880**
Vorinostat	0.723	0.596	0.438	0.499	**0.778**

从表 2 所示 F_1 评分结果来看，本文提出的

GCDrug 模型在 10 个单细胞数据集上均表现出优越的性能，整体显著优于现有其他方法。与当前主流模型相比，GCDrug 在大多数药物（如 Afatinib、AR-42、Cetuximab、Etoposide、NVP-TAE684、Sorafenib 等）上均取得了最高的 F_1 分数，尤其在 Etoposide 数据集上，GCDrug 的 F_1 分数达 0.956，远超其他方法，表现出极强的预测准确性和鲁棒性。尽管在某些药物（如 Gefitinib 和 PLX4720）上，CODEAE 和 SSDA4Drug 方法表现略优，但 GCDrug 依然保持了较强的竞争力，且整体平均性能更稳定。

为了能更加直观地展示本文方法的优势，将本文方法与 4 个基准方法（scDEAL_noPe 是 scDEAL 方法未预训练的形式）的 F_1 分数、*ROC* 曲线下的面积（*AUC*）以及精准率和召回率曲线下的面积（*AUPR*）共 3 个分类指标的性能可视化为图 3。



a）F_1 分数

b）*AUC*

c）*AUPR*

图 3 对比实验结果

Fig. 3 Comparative experimental results

由图 3 所示对比实验结果可知，结合 3 个分类指标，本文方法不仅整体稳定性更强，而且在多数药物上全面领先，显示出其对正负样本不平衡问题的良好鲁棒性与判别能力。综合来看，本文方法在识别精准性、泛化能力与稳定性 3 个方面均表现出显著优势，验证了其在肿瘤内耐药细胞识别任务中的准确性和泛化性。

3.3 消融实验结果

为验证 GCDrug 各关键模块对整体性能的贡献，本文设计两套消融方案，3 个实验设计方案描述如下：

1）GCDrug（KNN）。本方案保留 KNN 图结构，但移除 CORAL 域适应（即令 λ=0），仅利用图卷积提取特征并进行分类。

2）GCDrug（Deep CORAL）。本方案保留 Deep CORAL 域适应，但是移除图结构（即 edge_index=None），将 GCN 替换为多层感知机（MLP），仅依赖基因表达特征。

3）GCDrug。本文完整模型，同时引入图结构与 Deep CORAL 对齐。

消融实验的 F_1 分数结果如表 3 所示。

表 3 GCDrug 消融实验的 F_1 分数结果

Table 3 F_1-score results of GCDrug ablation study

药物名	GCDrug（Deep CORAL）	GCDrug（KNN）	GCDrug
Afatinib	0.854	0.822	**0.878**
AR-42	0.783	0.775	**0.798**
Cetuximab	0.834	0.828	**0.867**
Etoposide	0.906	0.883	**0.956**
Gefitinib	0.856	0.837	**0.899**
NVP-TAE684	0.843	0.826	**0.876**
PLX4720[1]	0.821	0.815	**0.847**
PLX4720[2]	0.725	0.757	**0.782**
Sorafenib	0.835	0.852	**0.880**
Vorinostat	0.751	0.708	**0.778**

分析表 3 所示实验结果，对比 GCDrug（Deep CORAL）与 GCDrug 可知，在 Etoposide、Afatinib、Cetuximab 等数据集上，引入图结构后的 F_1 分数提升了 1.9%~7.9%，说明通过 KNN 构建的图结构能有效捕捉样本间的相似性，缓解单细胞样本稀缺问题。对比 GCDrug（KNN）与 GCDrug 可知，加入 Deep CORAL 损失后，所有数据集性能均得到提升，平均涨幅约为 2.1%，最高达 8.3%（Etoposide），这表明本文的域适应方法可显著增强模型在目标域的泛化能力。综上所述，仅保留单一模块时，其实验性能均显著低于完整模型的。可见，本文提出的方法在图结构和 Deep CORAL 域适应相结合的形式下，在肿瘤

内耐药细胞识别任务上具有优越的性能。

4　结语

本文针对肿瘤的异质性导致肿瘤的复发或转移这一问题。提出了一种基于图结构与无监督域适应的方法 GCDrug。通过双层图卷积网络挖掘样本间拓扑相似性，并引入 Deep CORAL 损失实现源域与目标域的二阶统计量对齐，在无需任何药物注释标签的条件下，能够准确识别肿瘤内耐药细胞，且在模型的泛化能力上有着优越的表现。基于 10 个公开单细胞药物注释数据集的系统性实验结果表明，GCDrug 的 F_1、*AUC*、*AUPR* 3 项指标均稳定优于现有主流方法的对应值。消融实验结果进一步证实，图结构模块与 Deep CORAL 域适应模块缺一不可，二者协同赋予模型对肿瘤内耐药细胞识别的鲁棒性与跨域知识迁移能力。

参考文献：

[1] UPADHYAY A. Cancer：An Unknown Territory; Rethinking Before Going Ahead[J]. Genes & Diseases，2021，8(5)：655−661.

[2] BHAT G R，SETHI I，SADIDA H Q，et al. Cancer Cell Plasticity：From Cellular, Molecular, and Genetic Mechanisms to Tumor Heterogeneity and Drug Resistance[J]. Cancer and Metastasis Reviews，2024，43(1)：197−228.

[3] CREE I A，CHARLTON P. Molecular Chess? Hallmarks of Anti-Cancer Drug Resistance[J]. BMC Cancer，2017，17(1)：10.

[4] ARAMINI B，MASCIALE V，GRISENDI G，et al. Dissecting Tumor Growth：The Role of Cancer Stem Cells in Drug Resistance and Recurrence[J]. Cancers，2022，14(4)：976.

[5] VASAN N，BASELGA J，HYMAN D M. A View on Drug Resistance in Cancer[J]. Nature，2019，575：299−309.

[6] YANG W J，SOARES J，GRENINGER P，et al. Genomics of Drug Sensitivity in Cancer (GDSC)：A Resource for Therapeutic Biomarker Discovery in Cancer Cells[J]. Nucleic Acids Research，2013，41：D955−D961.

[7] IORIO F，KNIJNENBURG T A，VIS D J，et al. A Landscape of Pharmacogenomic Interactions in Cancer[J]. Cell，2016，166(3)：740−754.

[8] BARRETINA J，CAPONIGRO G，STRANSKY N，et al. The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity[J]. Nature，2012，483：603−607.

[9] ADAM G，RAMPÁŠEK L，SAFIKHANI Z，et al. Machine Learning Approaches to Drug Response Prediction：Challenges and Recent Progress[J]. NPJ Precision Oncology，2020，4：19.

[10] CHEN J Y，WANG X Y，MA A J，et al. Deep Transfer Learning of Cancer Drug Responses by Integrating Bulk and Single-Cell RNA-Seq Data[J]. Nature Communications，2022，13：6494.

[11] ZHENG Z T，CHEN J Y，CHEN X J，et al. Enabling Single-Cell Drug Response Annotations from Bulk RNA-Seq Using SCAD[J]. Advanced Science，2023，10(11)：2204113.

[12] HUANG K S，LIU H. Identification of Drug-Resistant Individual Cells Within Tumors by Semi-Supervised Transfer Learning from Bulk to Single-Cell Transcriptome[J]. Communications Biology，2025，8：530.

[13] HE D，LIU Q，WU Y，et al. A Context-Aware Deconfounding Autoencoder for Robust Prediction of Personalized Clinical Drug Response from Cell-Line Compound Screening[J]. Nature Machine Intelligence，2022，4(10)，879−892.

[14] SUN B C，SAENKO K. Deep CORAL：Correlation Alignment for Deep Domain Adaptation[C]//Computer Vision-ECCV 2016 Workshops. Cham：Springer，2016：443−450.

[15] WOLF F A，ANGERER P，THEIS F J. SCANPY：Large-Scale Single-Cell Gene Expression Data Analysis[J]. Genome Biology，2018，19(1)：15.

[16] PETERSON L. K-Nearest Neighbor[J]. Scholarpedia，2009，4(2)：1883.

[17] KIPF T N，WELLING M. Semi-Supervised Classification with Graph Convolutional Networks[EB/OL]. 2016：arXiv: 1609.02907. [2025−10−06]. https://arxiv.org/abs/1609.02907.

[18] AGARAP A F. Deep Learning Using Rectified Linear Units (ReLU)[EB/OL]. 2018：arXiv: 1803.08375. [2025−10−06]. https://arxiv.org/abs/1803.08375.

[19] SUN B S，FENG J S，SAENKO K. Return of Frustratingly Easy Domain Adaptation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix Arizona：[s.n.]，2016：2058−2065.

（责任编辑：廖友媛）