Hans Journal of Biomedicine 生物医学, 2025, 15(2), 432-442
Published Online March 2025 in Hans. https://www.hanspub.org/journal/hjbm
https://doi.org/10.12677/hjbm.2025.152050

Hans汉斯

MOGCWMLP：基于图卷积网络和加权多层感知机的多组学数据整合模型用于改进肺癌分期

赵　宇，李　悦，康骏凯，张小轶*

北京工业大学化学与生命科学学院，北京

收稿日期：2025年2月14日；录用日期：2025年3月19日；发布日期：2025年3月28日

摘　要

癌症是全球范围内导致死亡的主要疾病之一，尤其是对晚期或发生转移的癌症治疗依然面临巨大的挑战。癌症的精准分期在临床上对治疗方案的选择和患者预后评估至关重要。传统的分期方法主要依赖影像学和临床检查数据，然而随着基因组学和分子生物学技术的飞速发展，利用多组学数据进行癌症的早期诊断和分期变得越来越重要。为了提高癌症分类和分期的准确性，本研究提出了一种新的多组学数据分析框架MOGCWMLP。该框架基于图卷积网络(GCN)对不同组学数据进行特征学习，结合加权多层感知机(MLP)网络进行分类决策。具体来说，MOGCWMLP框架集成了RNA-seq、miRNA和lncRNA等三种不同类型的组学数据，通过学习每种数据的特征并进行加权融合，最大化不同组学数据的互补信息。实验结果表明，MOGCWMLP模型在肺鳞癌(LUSC)数据集上的分类精度显著优于现有的单组学模型和多组学模型，尤其是在多组学数据整合的情况下，分类性能得到显著提升。此外，采用可学习的加权融合机制，能够动态调整各视图的贡献，从而进一步优化模型的分类效果。该研究为癌症精准诊断和个性化治疗提供了有效的工具，也为多组学数据的整合提供了新的思路。

关键词

癌症分期，图卷积网络(GCN)，加权多层感知机(MLP)，多组学数据

MOGCWMLP: A Multi-Omics Data Integration Model Based on Graph Convolutional Networks and Weighted Multilayer Perceptron for Improved Lung Cancer Staging

Yu Zhao, Yue Li, Junkai Kang, Xiaoyi Zhang*

*通讯作者。

文章引用: 赵宇, 李悦, 康骏凯, 张小轶. MOGCWMLP: 基于图卷积网络和加权多层感知机的多组学数据整合模型用于改进肺癌分期[J]. 生物医学, 2025, 15(2): 432-442. DOI: 10.12677/hjbm.2025.152050

College of Chemistry and Life Science, Beijing University of Technology, Beijing

Received: Feb. 14th, 2025; accepted: Mar. 19th, 2025; published: Mar. 28th, 2025

Abstract

Cancer remains one of the leading causes of mortality worldwide, particularly in advanced or metastatic cases, where treatment remains a significant challenge. Accurate cancer staging is critical in clinical practice for determining optimal treatment strategies and assessing patient prognosis. Traditional staging methods primarily rely on imaging and clinical examination data. However, with rapid advancements in genomics and molecular biology, lever aging multi-omics data for early cancer diagnosis and staging has become increasingly important. To enhance the accuracy of cancer classification and staging, this study proposes an ovel multi-omics data analysis framework, MOGCWMLP. This framework utilizes graph convolutional networks (GCN) for feature learning across different omics data types and incorporates a weighted multilayer perceptron (MLP) for classification decision-making. Specifically, MOGCWMLP integrates three distinct types of omics data—mRNA, miRNA, and lncRNA—by extracting and fusing their features through a weighted mechanism, there by maximizing the complementary information among different omics modalities. Experimental results demonstrate that the MOGCWMLP model achieves significantly higher classification accuracy on the lung squamous cell carcinoma (LUSC) dataset compared to existing single-omics and multi-omics models. Notably, the integration of multi-omics data leads to substantial improvements in classification performance. Furthermore, the incorporation of a learnable weighted fusion mechanism enables the dynamic adjustment of each modality's contribution, further optimizing the model's classification effectiveness. This study provides an effective tool for precise cancer diagnosis and personalized treatment, while also offering new insights into the integration of multi-omics data.

Keywords

Cancer Staging, GCN, MLP, Multi-Omics Data

1. 引言

癌症在全球范围内不仅是导致死亡的主要原因之一[1]，也是制约人类预期寿命进一步提高的重要障碍。过去几十年间，尽管在癌症的发病机制、诊断、治疗和预后等方面取得了显著进展，癌症的死亡率也逐渐下降，但其发病率仍持续攀升[2]。尤其是晚期或发生转移的癌症患者，治愈仍然面临巨大挑战[3]。

癌症分期是评估肿瘤严重程度、指导治疗方案制定以及预测预后的关键步骤[4]。目前，国际上广泛采用的分期系统是由国际抗癌联盟(UICC)和美国癌症联合委员会(AJCC)制定的 TNM 分期系统[5]。该系统根据原发肿瘤的大小和侵袭程度(T)、区域淋巴结受累情况(N)以及远处转移的存在与否(M)进行分类。不同分期的癌症往往需要采取不同的治疗手段，例如手术、化疗、放疗、免疫治疗等方法的综合应用。此外，明确癌症分期不仅有助于早期诊断和精准治疗，还能改善患者的预后并显著提高治愈率[6]。为此，将肿瘤分为不同的阶段被认为是肿瘤治疗领域的重要问题之一。随着测序技术的进步和“精准医学时代”

的到来，大量的生物数据已经被收集[7]，并可以在公共数据库中获得多组学数据[8] [9]，如基因组学、转录组学等多个层面的数据进行综合分析，充分利用多组学数据之间的互补性，能更好的提高预测准确性[10]-[14]。同时，随着深度学习技术方法的发展，其高效的学习能力以及灵活性，已经开始普遍应用于多组学数据的集成上[12] [15]-[20]。例如，Rong 等人[21]通过研究多组学数据预测癌症，表明神经网络在提升预测精度方面具有显著优势。Hu 等人[22]通过结合基因表达数据和拷贝数变异数据，采用随机森林和深度神经网络方法进行维度缩减，从而提升了多组学数据的预测精度。Paul 等人[23]使用多个遗传编程规则集进行分类，并提出了多数投票的遗传编程分类器。Broet 等人[24]提出了一种基于微阵列数据的统计评分方法，以确定将早期阶段与后期分开的基因表达特征。Huang 等人[25]提出了使用深层转移卷积神经网络和极端学习机器来诊断 CT 图像上的肺结节。Koike 等人[26]使用基于机器学习的模型进行分类以预测外周肺鳞状细胞癌的复发。Xu 等人[27]使用的堆积稀疏自动编码器(SSAE)用于乳腺癌组织病理学图像的核检测。Li 等人[28]提出了一种基于图卷积网络(GCN)的多组学整合方法 MoGCN，用于癌症亚型分类分析，该研究通过整合了多组学数据，利用自编码器进行降维，使用相似性网络融合(SNF)构建患者相似性网络(PSN)，并结合 GCN 在癌症亚型分类上表现出高精度和稳定性，具有广泛的临床应用潜力。然而，在多学数据的整合过程中很少有不同组学权重考虑其中，而考虑权重能够更好的利用不同组学数据间的交互关系，提高预测的准确率。

在本研究中我们设计了一个用于生物医学应用中分类任务的多组学数据分析框架 MOGCWMLP。MOGCWMLP 结合了不同组学数据的特点，采用多个视图来表达不同的数据特征，最后通过加权的感知机网络融合进行分类诊断。具体来说，MOGCWMLP 利用图卷积网络(GCN)进行组学特定的学习。与全连通神经网络相比，GCN 既利用了组学特征，又利用了相似网络描述的样本之间的相关性。同时在多组学数据整合的过程中，采用加权多层感知机网络，不同组学数据可能包含不同的重要信息。通过每个组学数据的权重来自动调整不同视图在最终决策中的贡献，避免某些视图模型的不合理影响，最大化利用多组学数据的互补信息，从而获得了更好的分类性能。

2. 材料与方法

2.1. 数据获取与预处理

本研究使用癌症基因组图谱数据库 TCGA (https://portal.gdc.cancer.gov/)中肺腺癌(LUSC)的多组学数据集，其中包括 RNA-seq、miRNA 和 lncRNA 三种类型的基因表达数据。由于噪声冗余特征可能会影响分类任务的性能，因此分别对每种组学数据类型进行了预处理和特征预选，详细信息见表 1。

Table 1. Data set information used in the study
表 1. 本研究使用的数据集信息

数据集	类别	原始特征数			训练特征数		
		mRNA	miRNA	lncRNA	mRNA	miRNA	lncRNA
LUSC	Normal: 35, Early-stage: 386, Late-stage: 85	19938	1881	16882	16036	352	4569

2.2. 模型构建

MOGCWMLP 模型利用多视图图卷积神经网络(GCN)架构，结合了不同组学数据的特点，采用多个视图来表达不同的数据特征，最后通过加权融合进行分类诊断。MOGCWMLP 的整体工作流程如图 1 所示。

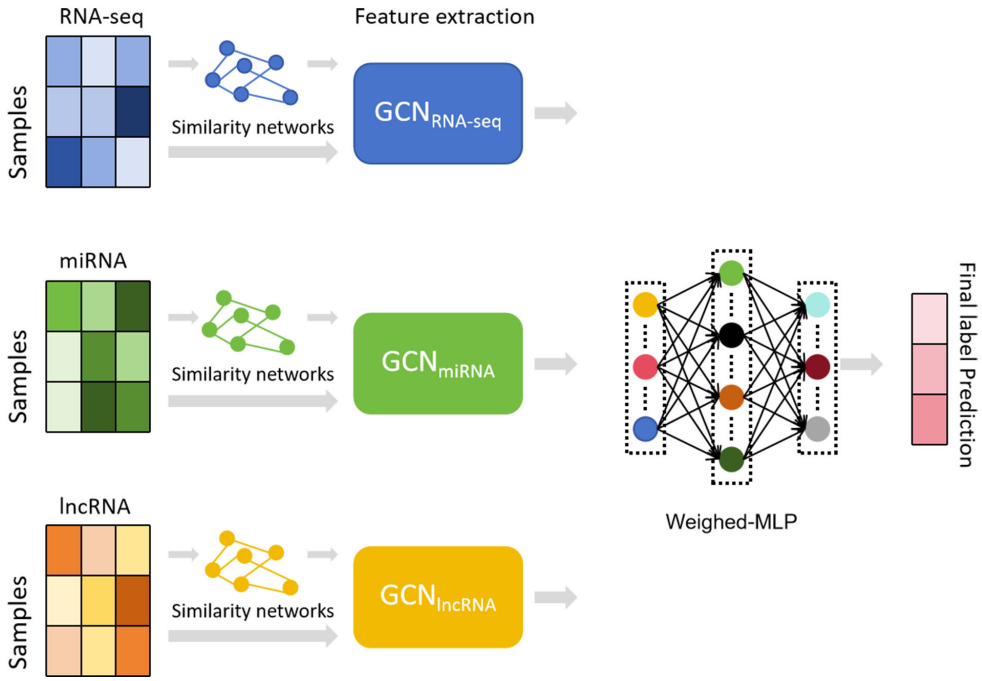

Figure 1. The workflow diagram of MOGCWMLP
图 1. MOGCWMLP 工作流程图

2.2.1. 图卷积神经网络(GCN)

在预处理和特征筛选后以去除噪声和荣誉特征后，使用 GCN 来学习每个组学数据类型的特征数据，其中使用皮尔逊相关性为每种类型数据构建样本相似性网络。

三种不同组学数据视图，分别表示为 X_1、X_2、X_3，每个视图的矩阵维度为 $N*D$，其中 N 是样本数，D 是每种组学数据的特征数，每种视图通过一个独立的图卷积网络进行处理，分别输入节点特征矩阵 X 和样本相似性邻接矩阵 A，每一层图卷积通过特征矩阵 X 和邻接矩阵 A 进行信息传播，捕获结点间的关系。具体计算过程如公式(1)、(2)所示：

$$H^{(l+1)} = \sigma\left(\hat{A}H^{(l)}W^{(l)} + b^{(l)}\right) \tag{1}$$

$$\hat{A} = D^{-1/2}AD^{-1/2} \tag{2}$$

其中，$H^{(l)}$ 是第 l 层的节点，$W^{(l)}$ 是该层的学习权重矩阵，$\hat{A}$ 是归一化的邻接矩阵如公式(2)所示，$b^{(l)}$ 是偏置项，σ 是激活函数(ReLU)，$H^{(l+1)}$ 是 GCN 通过层间信息传播，输出每种视图的图卷积后的特征表示。

图卷积的目的是通过邻接矩阵 A 捕捉样本间的关系，并通过权重矩阵 W 学习每个视图中的特征。每一层图卷积都会生成新的特征表示，这些表示将会传递到下一层，直到得到最终的图卷积输出。通过图卷积网络有效利用样本之间的相关性和特征之间的关系，充分发挥了图卷积网络(GCN)在多组学数据表征中的优势，使模型在多视图框架下学习到更丰富的特征表示，从而提升分类性能。

2.2.2. 特征加权感知机网络

在多组学数据的处理过程中，不同视图的特征重要性可能存在差异，因此需要采用加权融合机制来有效整合多个视图的信息。为了解决这一问题，我们对每个视图的输出特征赋予可学习的权重，然后对加权后的特征进行拼接，形成最终的特征向量 X_{final}，并将其输入到多层感知机(MLP)网络进行分类。该加权过程确保不同视图的特征在分类任务中按照其学习到的重要性进行贡献，以提升模型的决策能力。

在具体实现过程中，对每种视图最终输出的 $H^{(l)}$，通过学习的权重 w_l 进行加权。这些权重初始化为均匀分布，并在训练过程中不断优化。通过优化，模型能够学习到各视图的相对重要性，并动态调整它们的贡献，以确保最具信息量的组学特征得到更高的关注度。该加权学习策略能够有效捕捉不同组学层之间的互补信息，从而增强模型学习到的特征表示的鲁棒性，具体计算过程如公式(3)所示。

$$\text{Weighted\_features} = \sum_{v=1}^{V} w_l H^{(l)} \tag{3}$$

在加权特征拼接后，所得特征向量被传递到多层感知机(MLP)网络进行最终分类。MLP 结构由多个全连接层组成，能够对拼接后的组学数据进行高阶特征抽象，从而更全面地理解癌症分期分类模式。其中每层的输出都会经过 ReLU 激活函数，以引入非线性并提升特征表达能力。最终输出层采用 softmax 激活函数，以计算每个类别的预测概率，从而保证模型能够以高置信度将每个样本分配到最可能的癌症分期类别。具体的 MLP 计算过程见公式(4)。

$$\hat{y} = \text{MLP}\left(\text{X}_{\text{final}}\right) \tag{4}$$

加权融合策略最大化了最相关组学特征的贡献，确保从多个生物学视角中提取的信息能够得到最优整合。显著提高了癌症分期诊断的准确性，为个性化肿瘤治疗和临床决策提供了更加精确和数据驱动的框架。

3. 结果

3.1. 与现有模型进行对比

为了验证所提出 MOGCWMLP 模型的优势，我们将 MOGCWMLP 模型与传统机器学习和深度学习模型进行比较，包括 K 近邻(KNN) [29]、支持向量机(SVM) [30]、LightGBM [31]、MoGCN [28]。所有模型采用默认参数，在肺鳞癌(LUSC)数据集上与现有模型的比较结果如图 2 所示。

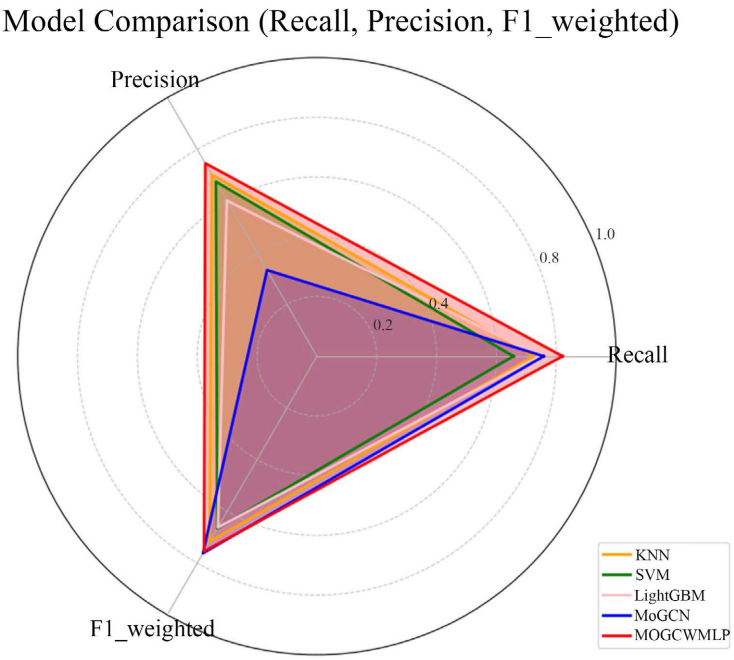

Figure 2. Performance comparison of different models based on accuracy, precision, and F1_weighted score
图 2. 不同模型在分类准确性、精确性和 F1 分数值上的性能对比

从图 2 可以看出，MOGCWMLP 模型在肺鳞癌(LUSC)数据集上的准确性(ACC)、精确性(Precision)、F1 分数(F1_weighted)共三个评价指标最优，显示出模型具有有效诊断肺鳞癌(LUSC)样本的分期能力。其中准确性(ACC)达到了最高水平的 0.8214，显著超越了其他模型，表明其在精准分类方面的优越性。精确性(Precision)达到了 0.7458，说明该模型相较于其他方法产生的误判较少，分类决策更为可靠。MOGCWMLP 在 F1 分数(F1_weighted)上同样表现最佳达到了 0.7546，表明其在分类任务中实现了精确率与召回率的良好平衡，这对于癌症分期等任务尤为关键。值得注意的是，虽然 MoGCN 模型[28]也表现出较强的性能，但在所有指标上略逊于 MOGCWMLP。而 KNN、SVM 和 LightGBM 等传统机器学习模型在分类准确率等方面的表现相对较低，进一步验证了图卷积网络(GCN)结合加权多层感知机(WMLP)融合策略的有效性。

实验结果表明，传统的机器学习方法依赖于人工特征工程，而 MOGCWMLP 通过 GCN 进行端到端的特征许欸小，能够自动挖掘多组学数据中的复杂模式，捕获疾病分期相关的关键特征。MOGCWMLP 结合 GCN 和可学习的 WMLP 结构，使其能够更好地适应复杂的多组学数据分布，提高模型的泛化能力，从而在多个评估指标上超越 MoGCN。MOGCWMLP 充分整合了多组学数据，在癌症分期任务中优于传统机器学习方法及其他深度学习模型。模型利用 GCN 进行特征学习，并结合可学习的加权融合机制，能够有效提取多组学数据之间的复杂关系，从而提升分类性能。

3.2. 不同组学数据间的表现

为了进一步验证多组学数据整合能够提升分类性能，我们设计了一系列对比实验，评估了 MOGCWMLP 模型在不同组学数据组合下的表现。具体而言，我们比较了以下几种训练方式：1) 三种组学数据联合训练(mRNA + miRNA + lncRNA)；2) 两种组学数据联合训练(mRNA + miRNA, mRNA + lncRNA, miRNA + lncRNA)；3) 单一组学数据训练(mRNA, miRNA, lncRNA)。图 3 展示了不同组合下模型在分类准确性(ACC)、精确性(Precision)和 F1 分数(F1_weighted)三项指标上的表现。

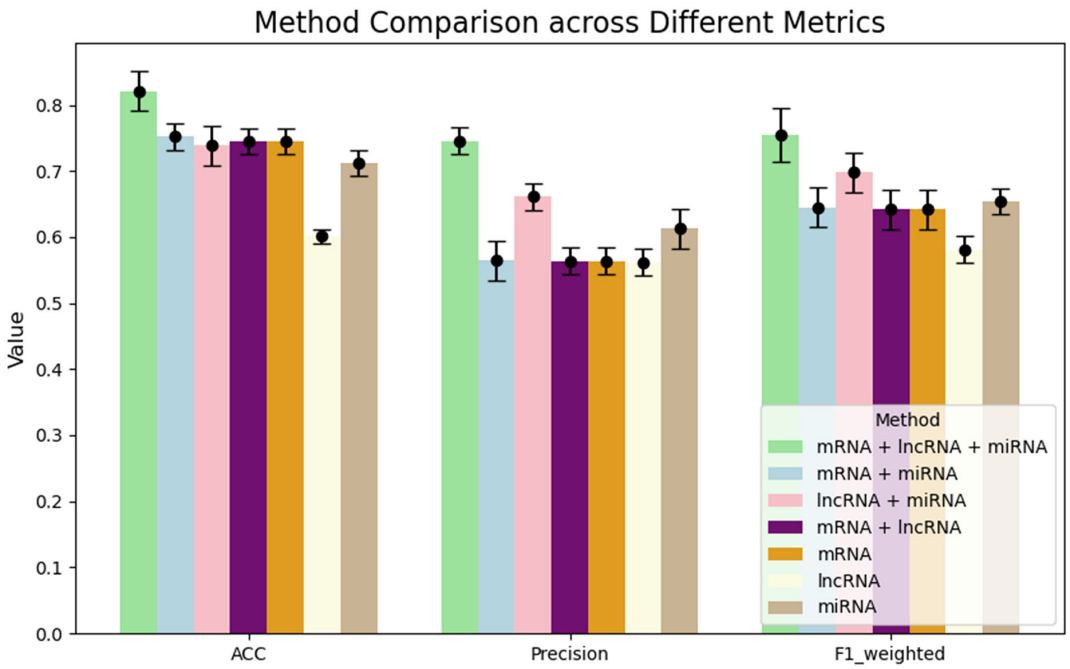

Figure 3. Comparison of classification performance across different omics-data combinations
图 3. 不同组学数据组合下分类性能的对比

从图 3 中我们可以分析得到使用三种组学数据(mRNA + miRNA + lncRNA)训练的 MOGCWMLP 模型在所有指标上均取得了最佳性能，验证了多组学数据的互补性，即不同组学数据能够提供不同层面的生物信息，有助于提高癌症分期的分类能力。两种组学数据联合训练的模型略优于单组学模型，进一步验证了多组学数据的互补性和信息融合的有效性。例如，miRNA + lncRNA 组合的模型在分类任务中取得了较好的平衡性，而 mRNA + miRNA 组合在精确性以及 F1 分数上表现相对较弱。单组学模型的分类性能整体低于三种组学整合模型和两种组学整合模型接近，从生物学角度来看，癌症的发展涉及多个层面的基因调控，而 mRNA、miRNA 和 lncRNA 共同作用于基因表达调控过程，单独使用某一类型的组学数据可能无法全面刻画疾病的分子机制，限制了分类能力。因此多组学数据的联合分析可以更全面地揭示癌症分期的分子机制。

实验结果表明，MOGCWMLP 通过加权融合多种组学数据，在分类任务中均取得了显著的性能提升。相比单组学数据，融合多种组学数据能够提供更加全面的信息支持，使得模型具备更强的泛化能力。进一步验证了 MOGCWMLP 在多组学数据整合和精准医学领域的应用潜力，为未来的癌症诊断和个性化治疗提供了重要的计算工具。

3.3. 加权融合参数对模型性能的影响

为了验证加权融合参数的有效性，我们进行了两种情况下的对比实验：一种是权重不可学习时，设置为固定均值；另一种是权重可学习时，模型能够动态调整不同视图的权重，随着训练的进行模型能够动态适应不同视图的重要性自动学习不同视图的重要性。这两种方式的比较结果如图 4 所示。

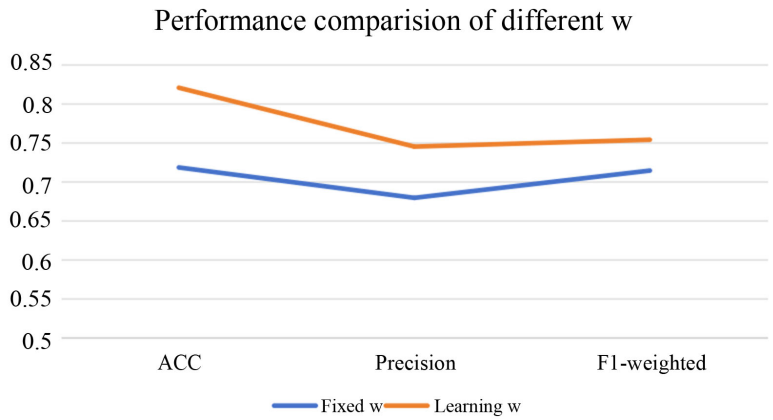

Figure 4. Performance comparison of fixed vs. learnable weights for fusion
图 4. 固定权重与可学习权重的融合性能对比

从图 4 中我们可以观察到，在分类准确性(ACC)、精确性(Precision)和 F1 分数(F1_weighted)三个指标上，可学习权重的模型均表现出优于固定权重的性能。这表明通过动态调整各视图的权重，模型能够更合理地融合多组学数据，提高最终的分类精度和诊断准确性。在固定权重的情况下，所有组学数据的贡献是相等的，但在实际应用中，不同组学数据对癌症分期的影响可能存在差异。例如，某些特定癌症可能更多地受到 miRNA 变化的影响，而在另一种癌症类型中，mRNA 可能更具诊断价值。通过引入可学习权重，MOGCWMLP 能够根据数据分布自动调整不同组学数据的贡献，使得模型适应性更强。在训练过程中，模型能够根据每个视图的贡献自动适应和学习不同视图的重要性，通过动态调整各视图的权重，模型能够避免某些视图对最终决策的不合理影响，使得多视图信息能够更加高效地融合，最大化利用多组学数据的互补信息，从而进一步提升了最终的分类精度和诊断准确性。

实验结果表明，可学习的加权机制显著提升了 MOGCWMLP 模型在多组学数据整合中的表现。相较于传统的固定权重方法，动态调整权重能够更好地捕捉不同组学数据的特征，避免信息丢失或过度偏倚。这一特性使得 MOGCWMLP 在癌症分期任务中具备了更高的稳定性和准确性，为精准医学中的多组学数据整合提供了有效的解决方案。

3.4. 特征基因富集分析

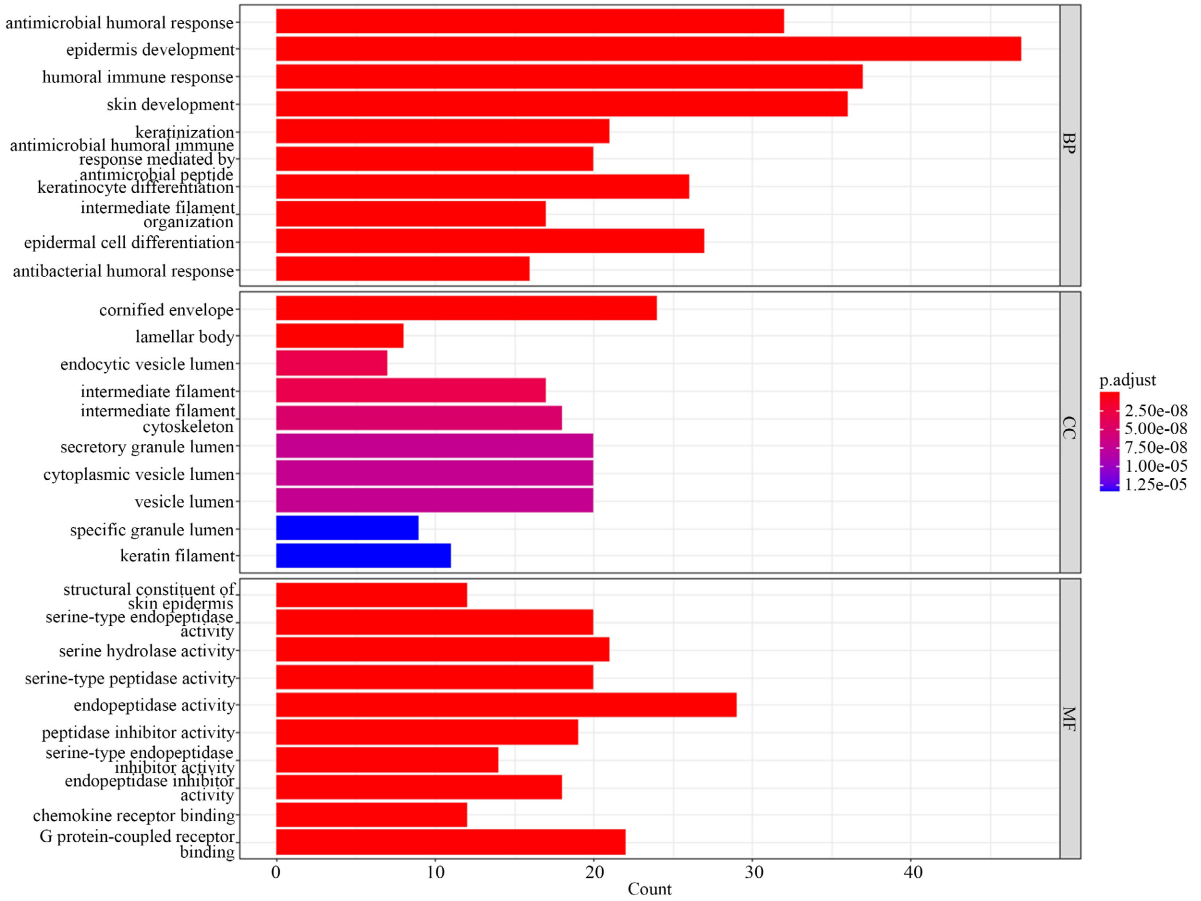

Figure 5. GO functional enrichment analysis of the top 300 feature genes
图 5. 前 300 个特征基因的 GO 功能富集分析

针对本研究使用的肺腺癌(LUSC)数据集，对模型提取出的前 300 个特征基因进行 GO 功能富集分析(图 5)，以确定其生物学功能。在生物学过程(Biological Process, BP)分主要富集于皮肤发育(Epidermis Development)、角化(Keratinization)以及免疫相关反应，表明肺鳞癌特征基因与皮肤屏障功能及免疫系统密切相关。在免疫应答相关通路中，具体富集在抗菌体液免疫反应(Antimicrobial Humoral Response)、体液免疫反应(Humoral Immune Response)、抗菌体液免疫反应(Antimicrobial Humoral Immune Response Mediated by Antimicrobial Peptide)、抗菌体液反应(Antibacterial Humoral Response)，表明 LUSC 相关的特征基因与免疫调节过程密切相关，尤其涉及体液免疫及抗菌肽介导的抗菌免疫，这可能与肿瘤微环境中的免疫激活或抑制状态有关。细胞组分(Cellular Component, CC)分析主要富集于分泌颗粒腔(Secretory Granule Lumen)、胞质囊泡腔(Cytoplasmic Vesicle Lumen)、特异性颗粒腔(Specific Granule Lumen)、囊泡腔(Vesicle Lumen)，表明肺鳞癌特征基因在细胞囊泡运输过程中发挥重要作用，可能涉及分泌、内吞及胞外基质

调控，这些过程可能影响肿瘤细胞的微环境适应性及免疫逃逸。分子功能(Molecular Function, MF)主要富集在丝氨酸肽酶活性(Serine-Type Peptidase Activity)、丝氨酸型内肽酶活性(Serine-Type Endopeptidase Activity)、内肽酶活性(Endopeptidase Activity)、肽酶抑制剂活性(Peptidase Inhibitor Activity)、丝氨酸型内肽酶抑制剂活性(Serine-Type Endopeptidase Inhibitor Activity)、内肽酶抑制剂活性(Endopeptidase Inhibitor Activity)，这些功能表明蛋白酶和蛋白酶抑制剂在肺鳞癌的发生发展中起关键作用。例如，丝氨酸蛋白酶参与肿瘤微环境重塑、炎症反应及细胞外基质降解，而蛋白酶抑制剂可能调控癌细胞的侵袭能力。

GO 富集结果表明，肺鳞癌特征基因在角化、皮肤发育、免疫应答及蛋白水解酶调控方面具有显著富集，进一步验证了肺鳞癌起源于鳞状上皮细胞的特性。角蛋白及中间纤维的富集反映了细胞结构稳定性的重要性，而蛋白酶及趋化因子受体的作用暗示肺鳞癌可能通过调节细胞外基质降解和免疫微环境促进肿瘤进展。丝氨酸蛋白酶和其抑制剂可能成为潜在的靶向治疗点，调节蛋白酶活性可能影响肺鳞癌细胞的侵袭性和耐药性。

4. 讨论与结论

本研究提出的 MOGCWMLP 框架在癌症分期分类任务中展示了卓越的性能，尤其是在多组学数据整合的背景下。我们首先使用三种不同类型的组学数据(RNA-seq, miRNA, lncRNA)，并采用图卷积网络(GCN)对每种组学数据进行特征学习。通过构建样本相似性网络并学习各组学特征，我们能够捕捉到不同组学数据之间的深层次关系。此外，在特征融合阶段，MOGCWMLP 采用了加权多层感知机(WMLP)进行加权融合，以便充分利用每个组学数据视图的独特信息，从而避免某些视图对最终决策的过度影响。研究结果表明，MOGCWMLP 作为一种先进的计算工具，在精准肿瘤学与多组学数据分析方面具有重要的应用价值。

与现有的传统癌症分期方法(如基于单一组学数据的 GCN 模型)相比，MOGCWMLP 框架通过有效的多组学数据融合显著提升了分类精度。在实验中，我们将 MOGCWMLP 与不同的组合模型进行了对比，结果表明，MOGCWMLP 在所有测试任务中均取得了最佳的分类性能。特别是，当使用三种组学数据的融合时，MOGCWMLP 模型的表现优于使用任何两种组学数据的模型。此外，使用多组学数据的模型无论在准确性、精确性还是 F1 分数方面，都优于单组学模型。

加权融合机制的有效性是本研究的另一大亮点。在实验中，我们比较了可学习和不可学习的加权机制。通过引入可学习的加权参数，MOGCWMLP 能够动态地调整不同视图的权重，使得模型能够根据每种组学数据的特征自动优化各视图的贡献。结果显示，采用可学习权重的模型在所有评估指标上均优于固定权重模型，进一步证明了多组学数据融合在癌症分类中的重要性。通过这一机制，MOGCWMLP 能够更好地利用不同组学数据之间的互补性，从而提高分类准确性。

然而，本研究也存在一些局限性。首先，虽然 MOGCWMLP 在肺鳞癌(LUSC)数据集上表现出色，但目前的研究主要集中在肺鳞癌这一特定癌症类型。未来的研究可以验证 MOGCWMLP 在其他癌症类型中的适用性和效果，例如乳腺癌、胃癌等。此外，模型的训练时间较长，尤其是在大规模数据集上进行训练时，计算复杂度较高。因此，未来可通过优化算法或采用更强的计算资源来加速模型的训练过程。

尽管如此，MOGCWMLP 框架仍为癌症精准分期和个性化治疗提供了一个有力的工具。随着深度学习和多组学数据分析技术的不断进步，我们预计该方法在临床癌症诊断中将发挥越来越重要的作用。未来，结合更多的组学数据(如表观遗传学、代谢组学等)以及更复杂的深度学习模型，MOGCWMLP 有望在癌症的早期诊断、精准分期和个性化治疗方案的制定方面取得更广泛的应用。

参考文献

[1] Cheever, M.A., Allison, J.P., Ferris, A.S., Finn, O.J., Hastings, B.M., Hecht, T.T., *et al.* (2009) The Prioritization of

Cancer Antigens: A National Cancer Institute Pilot Project for the Acceleration of Translational Research. *Clinical Cancer Research*, **15**, 5323-5337. https://doi.org/10.1158/1078-0432.ccr-09-0737

[2] Siegel, R.L., Miller, K.D., Fuchs, H.E. and Jemal, A. (2022) Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, **72**, 7-33. https://doi.org/10.3322/caac.21708

[3] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., *et al*. (2024) Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **74**, 229-263. https://doi.org/10.3322/caac.21834

[4] McPhail, S., Johnson, S., Greenberg, D., Peake, M. and Rous, B. (2015) Stage at Diagnosis and Early Mortality from Cancer in England. *British Journal of Cancer*, **112**, S108-S115. https://doi.org/10.1038/bjc.2015.49

[5] Amin, M.B., Greene, F.L., Edge, S.B., Compton, C.C., Gershenwald, J.E., Brookland, R.K., *et al*. (2017) The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge from a Population-Based to a More "Personalized" Approach to Cancer Staging. *CA: A Cancer Journal for Clinicians*, **67**, 93-99. https://doi.org/10.3322/caac.21388

[6] Teichgraeber, D.C., Guirguis, M.S. and Whitman, G.J. (2021) Breast Cancer Staging: Updates in the *AJCC Cancer Staging Manual*, 8th Edition, and Current Challenges for Radiologists, from the *AJR* Special Series on Cancer Staging. *American Journal of Roentgenology*, **217**, 278-290. https://doi.org/10.2214/ajr.20.25223

[7] Zhang, Z., Bajic, V.B., Yu, J., Cheung, K.-H. and Townsend, J.P. (2011) Data Integration in Bioinformatics: Current Efforts and Challenges. In: Mahdavi, M.A., Ed., *Bioinformatics—Trends and Methodologies*, InTech, 41-56. https://doi.org/10.5772/21654

[8] Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) Review the Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge. *Współczesna Onkologia*, **1**, 68-77. https://doi.org/10.5114/wo.2014.47136

[9] Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., *et al*. (2013) The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nature Genetics*, **45**, 1113-1120. https://doi.org/10.1038/ng.2764

[10] van de Wiel, M.A., Lien, T.G., Verlaat, W., van Wieringen, W.N. and Wilting, S.M. (2015) Better Prediction by Use of Co-Data: Adaptive Group-Regularized Ridge Regression. *Statistics in Medicine*, **35**, 368-381. https://doi.org/10.1002/sim.6732

[11] Singh, A., Shannon, C.P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S.J., *et al*. (2019) DIABLO: An Integrative Approach for Identifying Key Molecular Drivers from Multi-Omics Assays. *Bioinformatics*, **35**, 3055-3062. https://doi.org/10.1093/bioinformatics/bty1054

[12] Kim, D., Li, R., Dudek, S.M. and Ritchie, M.D. (2013) ATHENA: Identifying Interactions between Different Levels of Genomic Data Associated with Cancer Clinical Outcomes Using Grammatical Evolution Neural Network. *BioData Mining*, **6**, Article No. 23. https://doi.org/10.1186/1756-0381-6-23

[13] Huang, Z., Zhan, X., Xiang, S., Johnson, T.S., Helm, B., Yu, C.Y., *et al*. (2019) SALMON: Survival Analysis Learning with Multi-Omics Neural Networks on Breast Cancer. *Frontiers in Genetics*, **10**, Article 166. https://doi.org/10.3389/fgene.2019.00166

[14] Günther, O.P., Chen, V., Freue, G.C., Balshaw, R.F., Tebbutt, S.J., Hollander, Z., *et al*. (2012) A Computational Pipeline for the Development of Multi-Marker Bio-Signature Panels and Ensemble Classifiers. *BMC Bioinformatics*, **13**, Article No. 326. https://doi.org/10.1186/1471-2105-13-326

[15] Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., *et al*. (2022) Multimodal Machine Learning in Precision Health: A Scoping Review. *npj Digital Medicine*, **5**, Article No. 171. https://doi.org/10.1038/s41746-022-00712-8

[16] Abdelaziz, E.H., Ismail, R., Mabrouk, M.S. and Amin, E. (2024) Multi-Omics Data Integration and Analysis Pipeline for Precision Medicine: Systematic Review. *Computational Biology and Chemistry*, **113**, Article 108254. https://doi.org/10.1016/j.compbiolchem.2024.108254

[17] Tian, J., Zhu, M., Ren, Z., Zhao, Q., Wang, P., He, C.K., *et al*. (2022) Deep Learning Algorithm Reveals Two Prognostic Subtypes in Patients with Gliomas. *BMC Bioinformatics*, **23**, Article No. 417. https://doi.org/10.1186/s12859-022-04970-x

[18] Lin, Y., Zhang, W., Cao, H., Li, G. and Du, W. (2020) Classifying Breast Cancer Subtypes Using Deep Neural Networks Based on Multi-Omics Data. *Genes*, **11**, Article 888. https://doi.org/10.3390/genes11080888

[19] Madhumita, and Paul, S. (2022) Capturing the Latent Space of an Autoencoder for Multi-Omics Integration and Cancer Subtyping. *Computers in Biology and Medicine*, **148**, Article 105832. https://doi.org/10.1016/j.compbiomed.2022.105832

[20] Rong, Z., Liu, Z., Song, J., Cao, L., Yu, Y., Qiu, M., *et al*. (2022) Mcluster-VAEs: An End-to-End Variational Deep Learning-Based Clustering Method for Subtype Discovery Using Multi-Omics Data. *Computers in Biology and Medicine*, **150**, Article 106085. https://doi.org/10.1016/j.compbiomed.2022.106085

[21] Rong, Z., Lingyun, D., Jinxing, L. and Ying, G. (2021) Diagnostic Classification of Lung Cancer Using Deep Transfer

Learning Technology and Multi-Omics Data. *Chinese Journal of Electronics*, **30**, 843-852. https://doi.org/10.1049/cje.2021.06.006

[22] Hu, Y., Zhao, L., Li, Z., Dong, X., Xu, T. and Zhao, Y. (2022) Classifying the Multi-Omics Data of Gastric Cancer Using a Deep Feature Selection Method. *Expert Systems with Applications*, **200**, Article 116813. https://doi.org/10.1016/j.eswa.2022.116813

[23] Paul, T.K. and Iba, H. (2009) Prediction of Cancer Class with Majority Voting Genetic Programming Classifier Using Gene Expression Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **6**, 353-367. https://doi.org/10.1109/tcbb.2007.70245

[24] Broët, P., Kuznetsov, V.A., Bergh, J., Liu, E.T. and Miller, L.D. (2006) Identifying Gene Expression Changes in Breast Cancer That Distinguish Early and Late Relapse among Uncured Patients. *Bioinformatics*, **22**, 1477-1485. https://doi.org/10.1093/bioinformatics/btl110

[25] Huang, X., Lei, Q., Xie, T., Zhang, Y., Hu, Z. and Zhou, Q. (2020) Deep Transfer Convolutional Neural Network and Extreme Learning Machine for Lung Nodule Diagnosis on CT Images. *Knowledge-Based Systems*, **204**, Article 106230. https://doi.org/10.1016/j.knosys.2020.106230

[26] Koike, Y., Aokage, K., Ikeda, K., Nakai, T., Tane, K., Miyoshi, T., *et al*. (2020) Machine Learning-Based Histological Classification That Predicts Recurrence of Peripheral Lung Squamous Cell Carcinoma. *Lung Cancer*, **147**, 252-258. https://doi.org/10.1016/j.lungcan.2020.07.011

[27] Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., *et al*. (2016) Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. *IEEE Transactions on Medical Imaging*, **35**, 119-130. https://doi.org/10.1109/tmi.2015.2458702

[28] Li, X., Ma, J., Leng, L., Han, M., Li, M., He, F., *et al*. (2022) MoGCN: A Multi-Omics Integration Method Based on Graph Convolutional Network for Cancer Subtype Analysis. *Frontiers in Genetics*, **13**, Article 806842. https://doi.org/10.3389/fgene.2022.806842

[29] Fix, E. and Hodges, J.L. (1989) Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review/Revue Internationale de Statistique*, **57**, 238-247. https://doi.org/10.2307/1403797

[30] Boser, B.E., Guyon, I.M. and Vapnik, V.N. (1992) A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, 27-29 July 1992, 144-152. https://doi.org/10.1145/130385.130401

[31] Meng, Q. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 3149-3157.