

# 基于深度学习构建肺腺癌患者免疫相关基因生存预后模型的比较研究\*

任月<sup>1</sup> 秦阳<sup>1</sup> 兰宁<sup>1</sup> 张永超<sup>2</sup> 刘毅蓉<sup>1</sup> 赵晋芳<sup>1</sup> 罗天娥<sup>1,3,Δ</sup>

**【摘要】目的** 探讨使用多组学数据和免疫相关基因构建不同深度学习模型对肺腺癌患者的预后预测价值并比较其预测性能。**方法** 从 TCGA 数据库肺腺癌多组学数据提取免疫基因后分别用原始数据及单因素 Cox 回归过滤法、方差过滤法和弹性网络 (elastic net, EN) 方法降维后的数据构建 DeepOmix、Nnet-survival、Cox-nnet 和 Deepsurv 预后预测模型,采用一致性指数和时间依赖性 ROC 评价模型的预测效果。**结果** 比较多组学模型 C 指数、3 年 AUC 和 5 年 AUC 值结果显示,DeepOmix 模型结合生物信号通路信息采用不同降维方法构建的肺腺癌预后预测模型与 DeepOmix、Nnet-survival、Cox-nnet 和 Deepsurv 模型相比预测性能最好 (C 指数在 0.83 以上,3 年 AUC 在 0.89 以上,5 年 AUC 在 0.94 以上)。在筛选变量构建预后预测模型中 EN 降维在绝大多数情况下具有良好的预测准确度。DeepOmix 模型能显著区分高、低风险组患者,且高风险组预后较差 ( $P < 0.001$ )。**结论** DeepOmix 模型结合生物信号通路以及 EN 降维可以分析高维低样本量数据,构建的肺腺癌预后预测模型与其他模型相比预测性能较高。

**【关键词】** DeepOmix 深度学习 预后预测 生存分析 肺腺癌

**【中图分类号】** R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2025.01.010

## A Comparative Study on the Survival Prognosis Model of Immune-related Genes in Patients with Lung Adenocarcinoma based on Deep Learning

Ren Yue, Qin Yang, Lan Ning, et al (Department of Health Statistics, Public Health of School, Shanxi Medical University, Taiyuan 030001)

**【Abstract】Objective** To explore the prognostic value of different deep learning models constructed using multi-omics data and immune-related genes in patients with lung adenocarcinoma and compare their predictive performance. **Methods** DeepOmix, Nnet-survival, Cox-nnet and Deepsurv prognosis prediction models were constructed from the TCGA database lung adenocarcinoma multi-omics data after extracting immune genes using original data and the data after dimensionality reduction by single-factor Cox regression, variance and elastic net (EN) method, respectively. C-index and time-dependent ROC were used to evaluate the predictive effect of the models. **Results** The results of comparing the C-index, 3-year AUC and 5-year AUC values of the multi-omics models showed that the DeepOmix model combined with biological signaling pathway information using different dimensionality reduction methods to construct prognosis prediction models for lung adenocarcinoma had the best prediction performance compared with the DeepOmix, Nnet-survival, Cox-nnet and Deepsurv models (C-index was above 0.83, 3-year AUC was above 0.89 and 5-year AUC was above 0.94). In screening variables to construct prognosis prediction models, EN had good predictive accuracy in the majority of cases. The DeepOmix model significantly distinguished between patients in the high-risk and low-risk groups, and the prognosis was worse in the high-risk group ( $P < 0.001$ ). **Conclusion** The DeepOmix model combined with biological signaling pathway and EN dimension reduction can analyze high-dimensional low sample size data, and construct a prognosis prediction model for lung adenocarcinoma with high predictive performance compared with other models.

**【Key words】** DeepOmix; Deep learning; Prognosis prediction; Survival analysis; Lung adenocarcinoma

肺腺癌 (lung adenocarcinoma, LUAD) 是肺癌最常见的亚型,由于其高度异质性,需要临床与组学数据结合探索肿瘤患者预后情况。多组学数据为全面解释癌症的发生发展提供了可能<sup>[1]</sup>,但高维低样本组学数据分析迎来了新的挑战。深度学习 (deep learning, DL) 可以分析高维低样本量的组学数据分子特征间的非线性关系,并通过多个处理层响应传递,在生存预测方面得到广泛应用<sup>[2-5]</sup>。目前基于深度学习整合多组学数据构建预后预测模型常见的是不同 DL 模型方法之间的比较,尚未见与不同降维方法结合的预测性能比较。研究

表明,免疫细胞与肺腺癌的发生发展密切相关<sup>[6-7]</sup>。随着免疫疗法的发展,还需寻找免疫相关预后生物标志物改善 LUAD 患者预后。因此,构建免疫相关模型对肺腺癌预后具有重要意义。本次研究基于深度学习方法结合单因素 Cox 回归过滤法、方差过滤法和弹性网络 (elastic net, EN) 方法降维后的数据,分别构建 DeepOmix (Reactome)、DeepOmix (KEGG)、DeepOmix、Nnet-survival、Cox-nnet 和 Deepsurv 免疫相关预后模型,并比较其预测性能,为探索临床与多组学数据结合构建有效的预后预测模型改善肺腺癌患者预后提供建议。

\* 基金项目:山西省自然科学基金(201801D121210)

1.山西医科大学公共卫生学院卫生统计教研室(030001)

2.北京市丰台区疾病预防控制中心

3.重大疾病风险评估山西省重点实验室

Δ通信作者:罗天娥,E-mail: luotiane1977@163.com

## 方 法

### 1. 数据来源与处理

基于 UCSC Xena 网站 (<https://xenabrowser.net>)

下载肺腺癌的基因表达、DNA 甲基化、拷贝数变异、体细胞突变数据以及临床信息数据(生存时间、生存状态、年龄、性别、病理分期和吸烟量), count 数据已经过 log 转化。对于 DNA 甲基化数据,剔除一个探针对应多个基因以及未匹配到基因的探针后去除重复基因。

从免疫学数据库 ImmPort(<https://www.immport.org>)下载了 1793 个免疫相关基因,并在 LUAD 组学数据中筛选免疫相关基因,剔除缺失值超过 10% 的基因,用中位数对剩余缺失值进行插补,对 RNA 和 DNA 甲基化数据进行归一化处理。排除临床信息中缺失的样本信息后整合组学与临床数据。

从分子特征数据库(molecular signatures database, MSigDB)<sup>[8]</sup>下载 Reactome 和 KEGG 通路基因集,小通路可能被包含在大通路中,而较大通路可能与一般生物学通路有关,不具有疾病特异性。因此,排除超过 200 个基因和少于 20 个基因的通路分别获得 806 条 Reactome 通路,335 条 KEGG 通路,然后将组学数据与通路基因进行匹配。

## 2. 原理和方法

### (1) DeepOmix 模型

DeepOmix 模型由 Lianhe Zhao 等人<sup>[9]</sup>于 2021 年提出,用于整合多组学数据和进行生存预测。模型结构包括输入层、功能层、两个隐藏层和输出层。输入层由组学数据和临床信息组成,功能层为生物信号通路,输入层与功能层通过基因与信号通路的相关性连接,然后将通路特征传递到采用 tanh 激活函数的两个隐藏层,最后到达全连接的输出层,执行 Cox 比例风险回归(Cox-PH)。模型通过最小化平均负对数偏似然来优化参数  $L_2$  正则化,平均负对数偏似然函数公式如下:

$$l(\theta) = -\frac{1}{n_S} \sum_{i \in S} (h_i \delta - \log \sum_{j \in R(T_i)} e^{h_j \delta}) + \lambda (\|\theta\|_2^2)$$

其中  $h$  为第二个隐藏层的输出,  $R(T_i) = \{i \mid T_i \geq t\}$  是时间  $t$  的风险集,  $\|\theta\|_2^2$  是  $\{\delta, W\}$  的  $L_2$  范数,  $\delta$  是最后一层隐藏层与输出层的权重系数,而  $W$  是权重矩阵,  $\lambda$  是正则化的超参数。

DeepOmix(Reactome) 和 DeepOmix(KEGG) 分别结合 Reactome、KEGG 通路构建模型,其输入层与功能层通过基因与生物信号通路的相关性连接; DeepOmix 采用输入层与功能层全连接方式构建模型。

### (2) Nnet-survival 模型

Nnet-survival 是 Gensheimer 等人<sup>[10]</sup>于 2019 年提出的一种离散时间生存模型,模型将生存时间分为  $n$  个时间段,每个时间段的输出节点全连接到输出层,输出层结合时变基线风险率和非比例风险,输出  $n$  个时间段的生存概率。在每个时间段中应用损失函数,利用最小化批量随机梯度下降(stochastic gradient de-

scent, SGD)<sup>[11]</sup>训练模型,用 Sigmoid 激活函数传递非线性,总体损失函数是每个时间段的损失之和。

对于在时间间隔  $j$  未删失个体定义:

$$surv_s(j) = \begin{cases} 1, & t \geq t_j \\ 0, & \text{otherwise} \end{cases}; surv_f(j) = \begin{cases} 1, & t_{j-1} \leq t < t_j \\ 0, & \text{otherwise} \end{cases}$$

删失个体定义:

$$surv_s(j) = \begin{cases} 1, & t \geq \frac{1}{2}(t_{j-1} + t_j) \\ 0, & \text{otherwise} \end{cases}; surv_f(j) = 0$$

对数似然函数为:

$$loglik = \sum_{i=1}^n \left( \ln(1 + surv_s(i) \cdot (surv_{pred}(i) - 1)) + \ln(1 - surv_f(i) \cdot surv_{pred}(i)) \right)$$

其中  $surv_{pred}(i)$  为个体  $i$  的条件生存概率。

### (3) Cox-nnet 模型

Cox-nnet 模型是 Cox 回归模型的扩展,由 Travers Ching 等人<sup>[12]</sup>于 2018 年提出。模型结构包括输入层、单个隐藏层和输出层,隐藏层节点数为输入层平方根取整数,输出层根据隐藏层的 tanh 激活函数执行 Cox-PH,模型采用 dropout<sup>[13]</sup>正则化防止过拟合,输出结果表达式为:

$$\alpha_i = G(Wx_i + b)^T \lambda$$

其中  $\alpha_i$  和  $x_i$  为个体  $i$  的预后指数和协变量,  $W$  和  $b$  分别表示输入层与隐藏层的权重系数矩阵和每个隐藏节点的偏置项,为 Cox-PH 层的回归系数,  $G(k)$  为 tanh 激活函数:

$$G(k) = \frac{\exp(k) - \exp(-k)}{\exp(k) + \exp(-k)}; k = Wx_i + b$$

### (4) Deepsurv 模型

Deepsurv 模型是 Katzman 等人<sup>[14]</sup>于 2018 年提出的一种深度前馈网络,该模型结合 Cox-PH 并增加了多层感知机(multilayer perceptron, MLP)额外引入了非线性激活函数,运用 Adam 优化<sup>[15]</sup>、修正线性单元(rectified linear unit, ReLu)、批量归一化(batch normalization)、Dropout 等方法,以及执行随机超参数优化搜索<sup>[16]</sup>。模型隐藏层由全连接层和 dropout 层交替连接,输出为一个节点  $\hat{h}_\alpha(x)$ 。损失函数为平均负对数偏似然加正则化项:

$$l(\alpha) = -\frac{1}{N_{E=1}} \sum_{i: E_i=1} (\hat{h}_\alpha(x_i) - \log \sum_{j \in R(T_i)} e^{\hat{h}_\alpha(x_j)}) + \beta \cdot \|\alpha\|_2^2$$

其中  $N_{E=1}$  是发生观察事件的患者数,  $\beta$  是  $L_2$  正则化参数,通过梯度下降法使损失函数最小化。

### 3. 评价指标

本研究中使用一致性指数(C 指数)和时间依赖性 ROC(时依 ROC)来评价模型的预测性能。

### 4. 软件实现

本研究数据预处理以及变量筛选均在 R 软件中完成,单因素 Cox 回归过滤波法采用 Survival 包,弹性网

络降维在 glmnet 和 Survival 包中实现;在 Python 软件中构建深度学习模型,DeepOmix 模型在 Python 软件 PyTorch 1.0 包中实现,Nnet-survival 模型使用 Tensorflow 2.1.0 为后端的 Keras 2.3.1 库构建,Cox-nnet 模型使用 Theano 1.0.5 包在 Python 中实现,Deepsurv 模型基于 Tensorflow 1.15.0 平台实现。

### 结 果

本研究基于 TCGA 组学数据筛选免疫相关基因后分别用原始数据及单因素 Cox 回归过滤法 ( $P < 0.1$ )、方差过滤法和 EN 方法降维后的数据构建 LUAD 预后预测模型。数据划分为互不重合的训练集和测试集,比值为 1:1,训练集确定模型参数拟合模型,测试集评估模型预测性能,重复 50 次,采用 C 指数和时依 ROC 的均数±标准差( $\bar{x} \pm s$ )评估模型性能。

#### 1. 多组学构建模型比较

整合 RNA、DNA 甲基化、CNV、SNP 数据和临床信息得到 5942 个协变量,经单因素 Cox 回归过滤法、方差过滤法、EN 降维整合数据分别为 1038、404 和 76 个协变量,共纳入 432 个 LUAD 患者(表 1)。多组学模型预测性能结果差异见表 2。DeepOmix 模型结合生物信号通路信息与其他模型相比预测性能增强,具有良好的鲁棒性。绝大多数情况下,EN 降维在多组

学数据中预测性能表现出显著优势,其纳入模型维度最小。DeepOmix(Reactome)的 C 指数(0.86)与 DeepOmix、Nnet-survival、Cox-nnet 和 Deepsurv 模型相比分别提高了 87%、80%、84% 和 81%(图 1a)。单因素 Cox 回归与方差过滤法相比预测性能较好,能够提取与预后显著相关的因素。多组学模型 3、5 年 AUC 值比较如图 1b、c 所示,DeepOmix(Reactome)和 DeepOmix(KEGG)3 年 AUC 值为 0.89 以上,5 年 AUC 值为 0.94 以上。

通过 R 语言 survminer 包确定 DeepOmix 模型测试集预后指数的最佳截断值,以截断值为界分为低风险组和高风险组,原始数据、单因素 Cox 回归过滤、方差过滤以及 EN 的 Kaplan-Meier 生存曲线[图 1d,以 DeepOmix(Reactome)模型 EN 数据为例]均能显著区分高、低风险组患者,且高风险组预后较差,差异具有统计学意义( $P < 0.001$ )。

表 1 模型维度比较

	原始数据	方差过滤	单因素 Cox 回归	EN
DeepOmix(Reactome)	4161	299	657	57
DeepOmix(KEGG)	4016	274	632	59
DeepOmix	5942	404	1038	76
Nnet-survival	5942	404	1038	76
Cox-nnet	5942	404	1038	76
Deepsurv	5942	404	1038	76

表 2 多组学数据构建模型预测性能比较( $\bar{x} \pm s$ )

	DeepOmix(Reactome)	DeepOmix(KEGG)	DeepOmix	Nnet-survival	Cox-nnet	Deepsurv
<b>C 指数</b>						
原始数据	0.83±0.02	0.84±0.02	0.59±0.03	0.55±0.04	0.59±0.03	0.56±0.04
方差过滤	0.83±0.02	0.84±0.02	0.65±0.03	0.62±0.02	0.61±0.03	0.58±0.04
单因素 Cox 回归	0.85±0.02	0.84±0.02	0.68±0.03	0.64±0.03	0.65±0.03	0.61±0.03
弹性网络	<b>0.86±0.02</b>	<b>0.85±0.02</b>	<b>0.75±0.02</b>	<b>0.69±0.01</b>	<b>0.72±0.03</b>	<b>0.70±0.03</b>
<b>3 年 AUC</b>						
原始数据	0.93±0.06	<b>0.93±0.06</b>	0.60±0.04	0.54±0.05	0.59±0.04	0.56±0.04
方差过滤	0.89±0.13	0.89±0.13	0.66±0.03	0.62±0.03	0.58±0.06	0.55±0.05
单因素 Cox 回归	0.92±0.09	0.89±0.13	0.72±0.03	0.66±0.03	0.66±0.03	0.61±0.04
弹性网络	<b>0.93±0.06</b>	0.89±0.13	<b>0.78±0.03</b>	<b>0.72±0.02</b>	<b>0.73±0.06</b>	<b>0.71±0.04</b>
<b>5 年 AUC</b>						
原始数据	0.98±0.07	<b>0.98±0.07</b>	0.58±0.04	0.49±0.05	0.58±0.04	0.56±0.05
方差过滤	0.94±0.15	0.94±0.15	0.64±0.03	0.56±0.03	0.57±0.05	0.50±0.06
单因素 Cox 回归	0.97±0.10	0.94±0.15	0.71±0.03	0.60±0.03	0.65±0.03	0.60±0.05
弹性网络	<b>0.98±0.07</b>	0.94±0.15	<b>0.76±0.03</b>	<b>0.68±0.02</b>	<b>0.71±0.05</b>	<b>0.64±0.05</b>

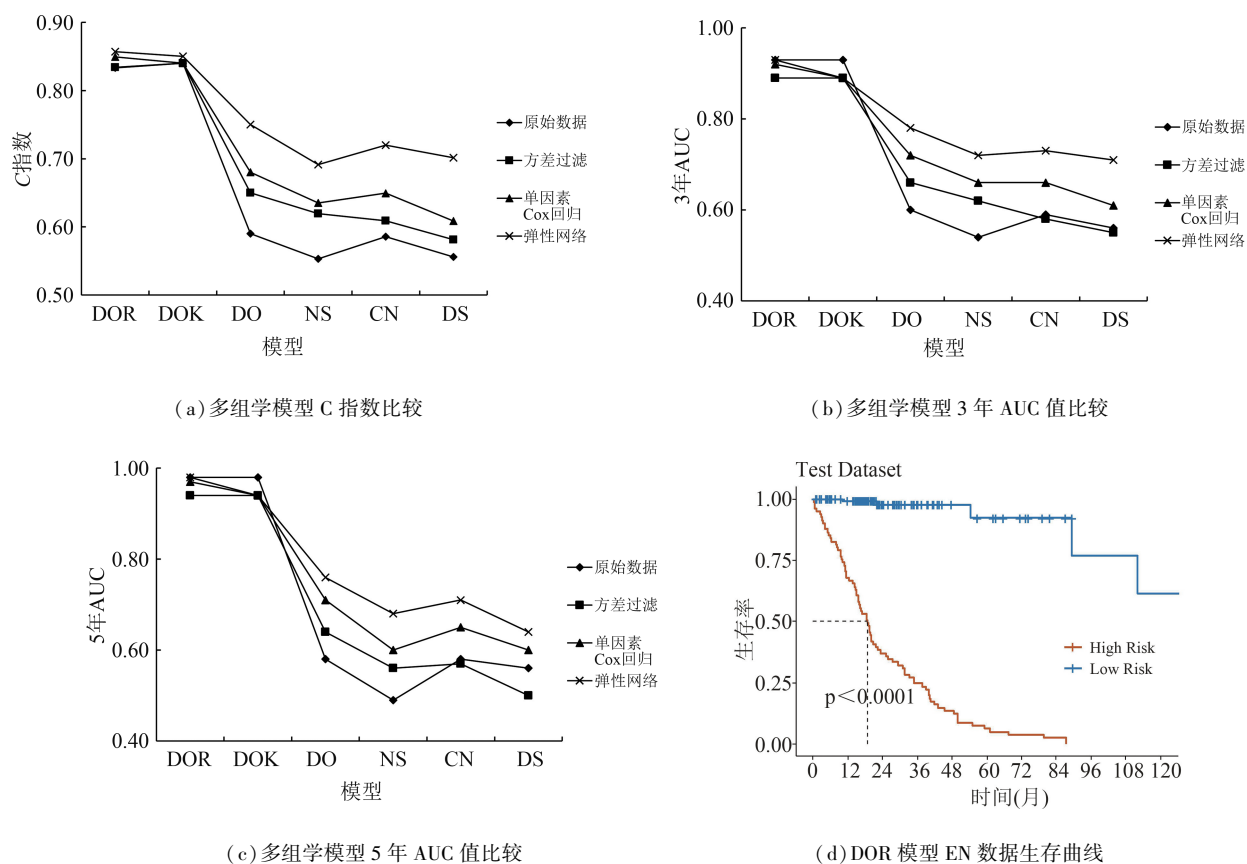
注:加粗字体为最佳

#### 2. 免疫相关 mRNA 的差异基因分析

为了进一步分析 DeepOmix 模型的生物信号通路,以低风险组为对照,对免疫相关 mRNA 进行差异基因分析。差异倍数取 1,Reactome 通路筛选 1075 个 mRNA,上调基因与下调基因均为 20 个(图 2a),KEGG 通路筛选 1041 个 mRNA,上调基因 24 个,下调基因 5 个(图 2b),Reactome 和 KEGG 通路差异基因有 9 个共有基因(图 3)。

根据差异基因分别绘制了 Reactome 和 KEGG 通

路气泡图,显示了前 15 个通路(图 4),Reactome 通路主要富集在白介素(interleukin, IL)家族信号通路、G 蛋白偶联受体(G protein-coupled receptors, GPCRs)通路、丝裂原活化蛋白激酶(mitogen-activated protein kinase, MAPK)信号通路、PI3K/AKT 通路负调控等。KEGG 通路结果显示了细胞因子-细胞因子受体相互作用、MAPK 信号通路、趋化因子信号通路、JAK-STAT 信号通路等。富集分析结果显示与炎症反应、细胞增殖与分化、趋化作用等方面有关。



注: DOR: DeepOmix ( Reactome ) 模型; DOK: DeepOmix ( KEGG ) 模型; DO: DeepOmix 模型; NS: Nnet-survival 模型; CN: Cox-nnet 模型; DS: Deepsurv 模型。

图 1 多组学模型预测性能比较及生存曲线图

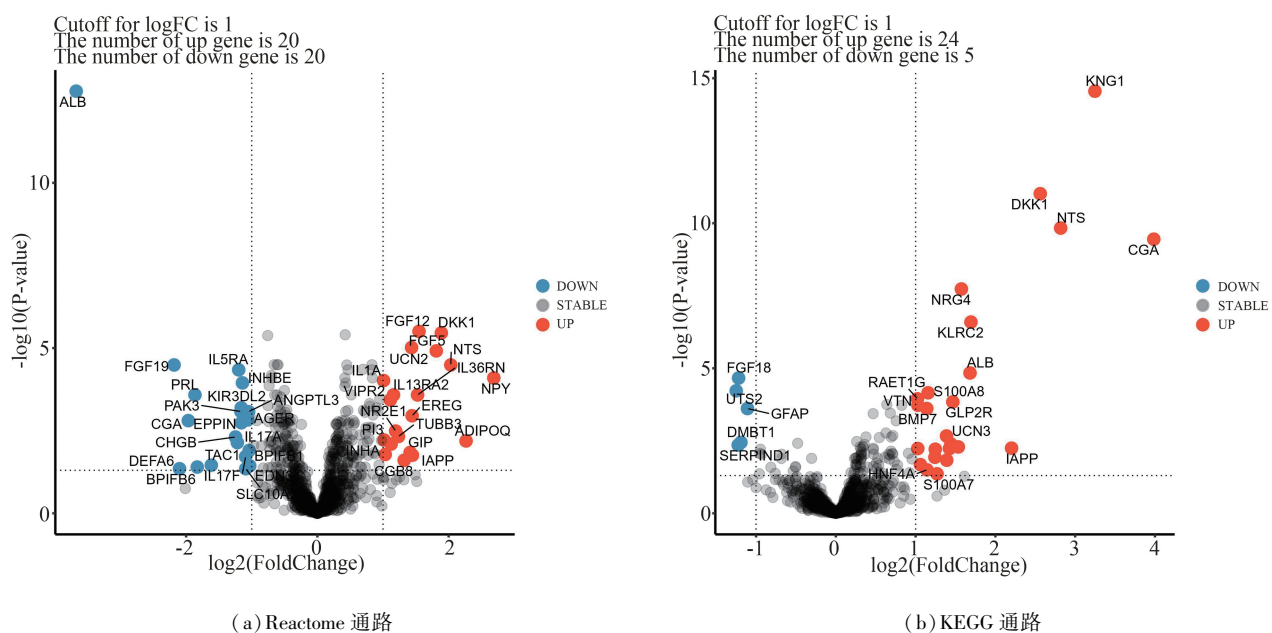


图 2 DeepOmix 模型差异基因分析火山图

### 讨论

基因表达、DNA 甲基化、CNV 以及 SNP 数据储存着重要的遗传信息,与肿瘤的发生发展密切相关,是预后预测疾病发生的生物标志物,也是肿瘤靶向治疗的

关键<sup>[17]</sup>。应用和整合不同分子分析技术为个性化医疗创造新机会的同时也带来了挑战。肿瘤样本的异质性和高维度是将多组学数据与临床结果预测(如生存时间)相结合的主要瓶颈。因此,需要探索更好的方法来构建预测模型并确定相关的潜在机制。

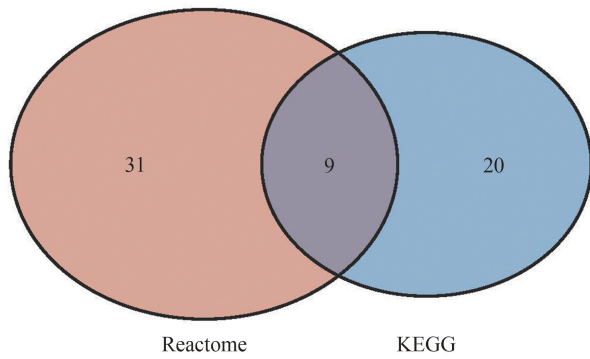
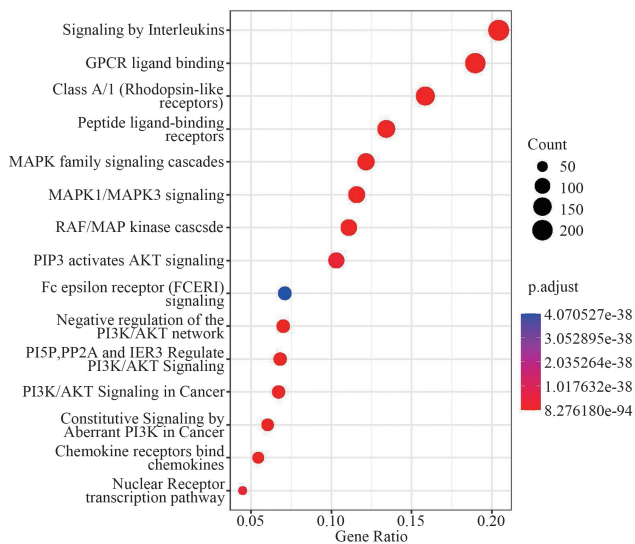
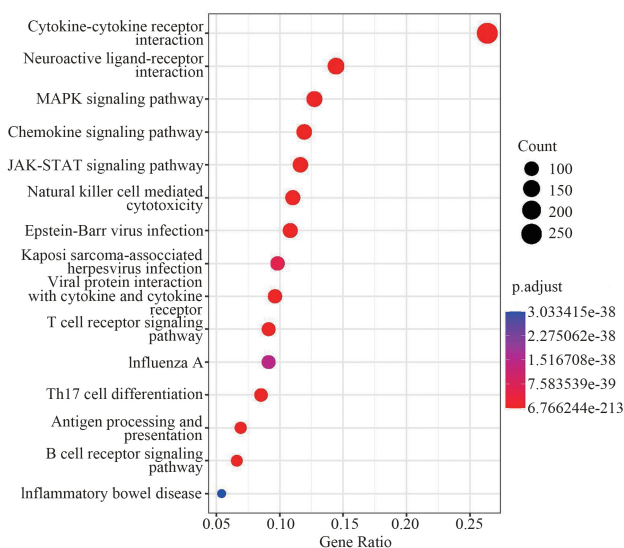


图 3 通路差异基因韦恩图



(a) Reactome 通路



(b) KEGG 通路

图 4 DeepOmix 模型差异基因富集分析气泡图

采用组学数据全部特征进行深度学习模型训练时,因存在大量与预测不相关的噪声和冗余特征可能导致模型出现过拟合<sup>[18]</sup>。对于这种小样本高维数据在建模前通常需要降维处理以剔除冗余特征简化模

型,提高模型的可解释性<sup>[19]</sup>。因此,本研究采用单因素 Cox 回归、方差过滤法以及 EN 降维的方法建模,通过 EN 数据构建 DeepOmix (Reactome)、DeepOmix (KEGG)、DeepOmix、Nnet - survival、Cox - nnet 和 Deepsurv 模型在绝大多数情况下具有良好的预测性能。EN 数据纳入的变量最少,剔除了组学数据中不重要的特征,有效提高了模型的预测准确性。单因素 Cox 回归与方差过滤法相比预测性能较好,能够提取与预后显著相关的因素。

DeepOmix 模型与 Cox 比例风险回归模型结合并通过组合基因功能模块拟合组学数据的非线性函数,充分利用基因与生物信号通路的先验知识提取与预后预测结果关联的显著模块,探索可能与预测结果相关的生物学机制。Nnet - survival 模型与 Cox 非比例风险回归模型结合将生存时间分段,对删失数据重新定义构建了离散时间生存模型,利用最小化批量随机梯度下降分段更新模型参数缩短了运行时间。Cox - nnet 模型较早地开发了 Cox 回归在深度学习中的应用,通过单个隐藏层 tanh 激活函数传递组学非线性关系,将深度学习与 Cox 回归结合在预测组学数据预后生存方面得到普遍应用。Deepsurv 模型增加了多层感知机额外引入非线性激活函数,应用随机超参数优化搜索估计了每个样本对网络参数化权重的危险率的影响,模型隐藏层由全连接层和 dropout 层交替连接,输出单个节点。

基于 LUAD 免疫相关基因构建的多组学预后预测模型中,DeepOmix 模型结合生物信号通路与其他模型相比有较高的预测准确度,KM 曲线能显著区分高、低风险组患者,且高风险组预后较差 ( $P < 0.001$ ),可用于分析多通路组学数据,以开发准确的预后预测模型,有利于识别有效的预后生物标志物、改善风险分层,为临床个性化治疗提供重要的理论依据。DeepOmix 模型生物信号通路的富集分析结果与肺腺癌的生长与转移息息相关。研究表明 IL-1 $\beta$  会通过 p38 信号传导促进 LUAD 细胞糖酵解,增强肿瘤细胞的迁移和侵袭<sup>[20-22]</sup>。而 LUAD 细胞产生的 IL-8 通过表皮生长因子受体反式激活刺激 LUAD 细胞增殖,参与 LUAD 的血管生成和转移。MAPK 信号通路在表皮生长因子的诱导下活化,促进了肺腺癌的进展和转移。趋化因子与肿瘤细胞结合,促进了肿瘤的生长、转移与侵袭。同时 GPCR 的一些成员在 LUAD 中具有预后因子的作用。因此,推荐 DeepOmix 模型与生物信号通路、EN 降维结合构建肺腺癌预后预测模型。

本研究目前仍存在一些不足,如使用了四种组学数据,蛋白质组学数据等更多的复杂数据还未扩展;只探讨了肺腺癌数据,其他肿瘤类型的数据还有待探讨。本研究下一步将分析 DeepOmix 模型的其他生物信号

通路,进一步探索不同通路对模型预测性能的影响。

### 参 考 文 献

- [ 1 ] Golub TR, Slonim DK, Tamayo P, et al. Molecular Classification of Cancer; Class Discovery and Class Prediction by Gene Expression Monitoring [J]. *Science*, 1999, 286(5439): 531-537.
- [ 2 ] Lan K, Wang DT, Fong S, et al. A Survey of Data Mining and Deep Learning in Bioinformatics [J]. *J Med Syst*, 2018, 42(8): 139.
- [ 3 ] 王萌, 王策, 栗思思, 等. 深度学习模型融合正则化方法在高维数据特征筛选中的应用研究 [J]. *中国卫生统计*, 2021, 38(1): 73-75.
- [ 4 ] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507.
- [ 5 ] Chen S, Zhang E, Guo T, et al. Development and verification of a deep learning-based m6A modification model for clinical prognosis prediction of renal cell carcinoma [J]. *Journal of Cancer Research and Clinical Oncology*, 2023, 149(15): 14283-14296.
- [ 6 ] Qu Y, Cheng B, Shao N, et al. Prognostic value of immune-related genes in the tumor microenvironment of lung adenocarcinoma and lung squamous cell carcinoma [J]. *Aging*, 2020, 12(6): 4757-4777.
- [ 7 ] Ohshima K, Morii E. Metabolic Reprogramming of Cancer Cells during Tumor Progression and Metastasis [J]. *Metabolites*, 2021, 11(1): 28.
- [ 8 ] Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database(MSigDB)3.0 [J]. *Bioinformatics*, 2011, 27(12): 1739-1740.
- [ 9 ] Zhao L, Dong Q, Luo C, et al. DeepOmix: A scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis [J]. *Comput Struct Biotechnol J*, 2021, 19: 2719-2725.
- [ 10 ] Gensheimer MF, Narasimhan B. A scalable discrete-time survival model for neural networks [J]. *PeerJ*, 2019, 7: e6257.
- [ 11 ] Ketkar N. Stochastic Gradient Descent [M]// *Deep Learning with Python: A Hands-on Introduction*. Berkeley: Apress, 2017: 113-132.
- [ 12 ] Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data [J]. *PLoS Comput Biol*, 2018, 14(4): e1006076.
- [ 13 ] Srivastava N, Hinton GE, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. *J Mach Learn Res*, 2014, 15: 1929-1958.
- [ 14 ] Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network [J]. *BMC Medical Research Methodology*, 2018, 18(1): 24.
- [ 15 ] Kingma D, Ba J. Adam: A Method for Stochastic Optimization [EB/OL]. (2017-01-30) [2023-02-09]. <https://arxiv.org/pdf/1412.6980>.
- [ 16 ] Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization [J]. *Journal of Machine Learning Research*, 2012, 13(1): 281-305.
- [ 17 ] Gui CP, Wei JH, Chen YH, et al. A new thinking: extended application of genomic selection to screen multiomics data for development of novel hypoxia-immune biomarkers and target therapy of clear cell renal cell carcinoma [J]. *Briefings in Bioinformatics*, 2021, 22(6): 173.
- [ 18 ] 卢宇红, 宋佳丽, 王萌, 等. 基于神经网络融合稀疏分组 LASSO 的预测模型研究 [J]. *中国卫生统计*, 2021, 38(6): 821-827.
- [ 19 ] 王娜, 肖建明, 罗钰, 等. LASSO 与经典特征选择方法在高级别前列腺癌诊断中的价值 [J]. *中国卫生统计*, 2022, 39(4): 578-580.
- [ 20 ] Tan Q, Duan L, Huang Q, et al. Interleukin -1 $\beta$  Promotes Lung Adenocarcinoma Growth and Invasion Through Promoting Glycolysis via p38 Pathway [J]. *Journal of inflammation research*, 2021, 14: 6491-6509.
- [ 21 ] Luppi F, Longo AM, de Boer WI, et al. Interleukin-8 stimulates cell proliferation in non-small cell lung cancer through epidermal growth factor receptor transactivation [J]. *Lung Cancer(Amsterdam, Netherlands)*, 2007, 56(1): 25-33.
- [ 22 ] Wang Y, Qiu L, Chen Y, et al. Screening and Identification of Four Prognostic Genes Related to Immune Infiltration and G-Protein Coupled Receptors Pathway in Lung Adenocarcinoma [J]. *Frontiers in Oncology*, 2020, 10: 622251.
- (责任编辑:邓妍)
- (上接第 55 页)
- [ 23 ] Breiman L. Random forests, machine learning 45 [J]. *Journal of Clinical Microbiology*, 2001, 2: 199-228.
- [ 24 ] Muhammed NKP, Thiyagarajan P. Alzheimer's classification using dynamic ensemble of classifiers selection algorithms: A performance analysis [J]. *Biomedical Signal Processing and Control*, 2021, 68: 102729.
- [ 25 ] 何思宇, 汪颖懿, 左敏, 等. 基于动态选择技术的食品安全谣言不平衡数据集分类方法 [J]. *计算机应用与软件*, 2022, 39(11): 250-256+349.
- [ 26 ] 曲文龙, 李一漪, 陈笑屹, 等. 一种基于近邻样本评估的动态选择性集成预测算法 [J]. *西北大学学报(自然科学版)*, 2020, 50(5): 802-810.
- [ 27 ] Hou WH, Wang XK, Zhang HY, et al. A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment [J]. *Knowledge-Based Systems*, 2020, 208: 106462.
- [ 28 ] Khushi M, Dar KS, Alam TM, et al. A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data [J]. *IEEE Access*, 2021, 9: 109960-109975.
- [ 29 ] Lamari M, Azizi N, Hammami NE, et al. SMOTE-ENN-Based Data Sampling and Improved Dynamic Ensemble Selection for Imbalanced Medical Data Classification [J]. 2021, 1188: 37-49.
- (责任编辑:张悦)