

基于可解释性的深度学习模型在宫颈鳞状细胞癌预后预测中的研究

武舒琪¹, 刘毅蓉¹, 房瑞玲^{1,2}, 赵晋芳^{1,2}, 罗天娥^{1,2}

1. 山西医科大学公共卫生学院卫生统计学教研室, 山西 太原 030001;

2. 煤炭环境致病与防治教育部重点实验室(山西医科大学)

摘要:目的 探讨整合多组学数据构建不同深度学习模型对宫颈鳞状细胞癌患者的预后预测价值并比较预测性能。方法 使用 TCGA(The Cancer Genome Atlas)数据库中宫颈鳞状细胞癌(CESC)的多组学数据经预处理后分别构建 AutoSurv、DeepSurv 和 Cox-nnet 三种预后预测模型,采用一致性指数(C-index)、时间依赖性受试者工作特征曲线下面积(AUC)和对数秩检验P值评价模型的预测效能。选择预测能力更优的 AutoSurv 模型,应用 DeepSHAP 进行可解释性分析,确定有助于区分高危和低危患者的 mRNA 和 miRNA。结果 不同模型测试集的 C-index、1 年 AUC、3 年 AUC 和 5 年 AUC 值结果显示:AutoSurv 模型预测性能最佳(C-index:0.78(95%CI:0.75~0.81),AUC(1Year):0.84(95%CI:0.81~0.87)、AUC(3Year):0.78(95%CI:0.76~0.85)、AUC(5Year):0.77(95%CI:0.75~0.81)。Kaplan-Meier 生存曲线 Log-rank 检验结果表明,AutoSurv 模型划分的高、低风险组患者生存差异有统计学意义($P=0.013$)。AutoSurv 模型在预测宫颈鳞状细胞癌患者预后方面的表现优于 DeepSurv 和 Cox-nnet 模型,通过 DeepSHAP 量化对预测结果有重要影响的 mRNA 和 miRNA 后识别出多个与宫颈鳞状细胞癌预后相关的变量,包括 POLR2H、FN1、FYN、UBA52 和 RPS27A 等,并富集到 DNA 损伤修复和 NF- κ B 等分子信号通路。结论 AutoSurv 模型能准确预测 CESC 患者的生存风险(C-index:0.78(95%CI:0.75~0.81)),结合 DeepSHAP 方法提供有价值的可解释性分析,识别影响宫颈鳞状细胞癌预后的关键预测变量,揭示潜在的生物学机制,为临床工作者制定个性化治疗策略提供科学依据,为癌症的靶向治疗提供研究方向。

关键词: AutoSurv 模型、可解释性、深度学习、预后预测、宫颈鳞状细胞癌

中图分类号:R737.33 文献标志码:A 文章编号:1003-8507(2026)14-2676-07

DOI:10.20043/j.cnki.MPM.202509110

Interpretable deep learning models for prognosis prediction in cervical squamous cell carcinoma

WU Shu - qi*, LIU Yi - rong, FANG Rui - ling, ZHAO Jin - fang, LUO Tian - e

* Department of Health Statistics, School of Public Health, Shanxi Medical University, Taiyuan, Shanxi 030001, China

Abstract: Objective To investigate the prognostic value of different deep learning models constructed using integrated multi-omics data for patients with cervical squamous cell carcinoma and to compare their predictive performance. **Methods** Using preprocessed multi-omics data from the The Cancer Genome Atlas (TCGA) database for cervical squamous cell carcinoma (CESC), three prognostic prediction models—AutoSurv, DeepSurv, and Cox-nnet—were constructed. The predictive performance of the models was evaluated using the C-index, time-dependent area under the receiver operating characteristic curve (AUC), and the P-value of the log-rank test. The AutoSurv model, which demonstrated superior predictive capability, was selected for interpretability analysis using DeepSHAP to identify mRNAs and miRNAs that help distinguish high-risk from low-risk patients. **Results** The C-index, 1-year AUC, 3-year AUC, and 5-year AUC values for the different model test sets showed that the AutoSurv model had the best predictive performance (C-index: 0.78 (95% CI: 0.75-0.81), AUC (1-year): 0.84 (95% CI: 0.81-0.87), AUC (3-year): 0.78 (95% CI: 0.76-0.85), AUC (5-year): 0.77 (95% CI: 0.75-0.81). The Log-rank test of the Kaplan-Meier survival curves indicated that the survival difference between the high- and low-risk groups classified by the AutoSurv model was statistically significant ($P=0.013$). The AutoSurv model outperformed the DeepSurv and Cox-nnet models in predicting the prognosis of patients with

基金项目:山西省自然科学基金(201801D121210),国家青年科学基金(82204164)

作者简介:武舒琪(2000—),女,硕士在读,研究方向:深度学习方法在疾病预后预测中的应用

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

cervical squamous cell carcinoma. After quantifying the mRNAs and miRNAs that significantly influence the prediction results using DeepSHAP, multiple variables associated with the prognosis of cervical squamous cell carcinoma were identified, including POLR2H, FN1, FYN, UBA52, and RPS27A, among others, which were enriched in molecular signaling pathways such as DNA damage repair and NF- κ B. **Conclusion** The AutoSurv model can accurately predict survival risk in CESC patients (C-index: 0.78 (95% CI: 0.75 - 0.81)). Combined with the DeepSHAP method, it provides valuable interpretability analysis, identifies key predictive variables influencing the prognosis of cervical squamous cell carcinoma, and reveals underlying biological mechanisms. This provides a scientific basis for clinicians to develop personalized treatment strategies and offers a research direction for targeted cancer therapy.

Keywords: AUTOSurv model; Interpretability; Deep learning; Prognosis prediction; Cervical squamous cell carcinoma

宫颈鳞状细胞癌 (Cervical squamous cell carcinoma, CESC) 是全球女性第四大最常见的恶性肿瘤^[1]。癌症的多组学数据为全面解释癌症的发生发展提供了可能,但是其高维度、低样本量的特点使得数据分析存在很大挑战^[2]。深度学习能够有效捕捉组学数据分子特征间的非线性关系^[3],在预后预测中表现出色,但在实际研究中,样本量不足往往导致模型过拟合,进而影响预后预测的准确性。这类模型大多被视为“黑箱”,其内部决策机制难以理解,缺乏对已知生物学知识的有效利用。如何在保持模型预测性能的同时,提供对模型决策的解释,是一个亟待解决的关键问题。针对上述挑战,由 Jiang^[4] 等提出的 AutoSurv 模型将通路信息作为先验知识嵌入网络架构,利用 mRNA 和通路的生物学关联引导特征提取和降维,同时处理 mRNA、miRNA 和临床数据,并结合 DeepSHAP 进行可解释性分析,将预测结果转化为可理解的生物学知识。本研究首次将通路引导的可解释性深度学习模型 AutoSurv^[4] 应用于 CESC 的预后预测,并与传统的 DeepSurv^[5]、Cox-nnet^[6] 模型比较,为探索整合临床和多组学数据构建有效的预后预测模型、改善宫颈鳞状细胞癌患者预后提供数智化支持,为临床决策提供透明化的生物学依据。

1 数据和方法

1.1 数据来源与处理 从 UCSC Xena 网站 (<https://xenabrowser.net>) 获取 TCGA 数据库中 CESC 患者的 mRNA、miRNA、临床信息数据 (年龄、种族、癌症分期) 和生存数据 (生存时间和生存状态)。其中 mRNA 数据以 $\log_2(\text{FPKM} + 1)$ 的标准化计数提供,miRNA 表达数据以 $\log_2(\text{RPM} + 1)$ 的标准化计数表示。

从 DAVID 数据库 (<https://david.ncifcrf.gov/content.jsp?file=update.html>) 获取 Reactome 数据库版本 78 的通路信息。基于 Reactome 数据库注释,筛选出有通路注释的 mRNA,过滤关联较少的通路 (少于 15 个基因或超过 300 个基因)。使用 biomaRt 包从 Ensemble 数据库 (<https://www.ensembl.org/index.html?redirect=no>) 中获取相关基因的染色体

信息,移除位于 Y 染色体和线粒体 (MT) 上的基因,仅保留常染色体和 X 染色体上的基因。为了减少噪声,过滤方差小于 0.02 的 mRNA 和 miRNA,通过最小-最大归一化 (Min-Max Normalization) 处理将表达值缩放到 [0,1] 区间。对临床变量中的年龄,采用 Z-score 标准化;其余临床分类变量 (种族、分期) 进行二值化处理。癌症分期按照国际妇产科联合会 (International Federation of Gynecology and Obstetric, FIGO) 分期分为癌症早期 stage_i (I - IIA)、癌症晚期 stage_ii (IIB - IV)。mRNA、miRNA 与临床数据进行合并,形成多组学数据集。

为了确保模型评估的独立性,数据预处理时将数据集按 6:4 的比例划分为互不重合的训练集和测试集,训练集用于模型训练和超参数调优,而测试集在整个实验过程中未参与任何模型训练或超参数调优。

1.2 模型选择 本研究选用 AutoSurv^[4]、DeepSurv^[5] 和 Coxnnet^[6] 三种模型进行 CESC 患者的生存分析。对于 AutoSurv 模型,用通路信息引导的多组学变分自编码器^[7] (KLPMVAE) 提取 mRNA 和 miRNA 的低维潜在特征 z,将该潜在特征与临床变量和生存数据结合后输入潜在特征驱动的生存网络 (LFSurv),计算患者的预后指数 (Prognostic Index, PI);模型训练采用两阶段超参数搜索策略:KL-PMVAE 部分的学习率设为 0.01, L_2 正则化系数为 0.005;LFSurv 部分的学习率设为 0.05, L_2 正则化系数为 0.000 25;内部验证使用 Bootstrap 方法,最后通过 DeepSHAP 解释方法识别对预测结果贡献最大的 mRNA 和 miRNA。对于 DeepSurv 模型,输入患者的组学数据和生存数据,经过多个隐藏层的非线性变换后,输出对数风险函数;采用贝叶斯优化结合 5 折交叉验证进行超参数调优,确定学习率为 0.000 135 8,学习率衰减为 0.000 091, L_2 正则化系数为 0.001,衰减率 (momentum) 为 0.88,并用 Bootstrap 法进行内部验证。对于 Cox-nnet 模型,输入患者的组学数据和生存数据,经单隐藏层输出预后指数;使用分层策略对 L_2 正则化参数进行搜索,确定学习率为 0.001,衰减率为 0.9, L_2 正则化系数为 0.2865,配合早停机制

防止过拟合,最终通过 Bootstrap 法进行内部验证。三种模型均以平均负对数偏似然(含 L_2 正则化)作为损失函数进行优化。

1.3 可解释性分析 采用 DeepSHAP 方法对模型的预测结果进行解释,识别对预后预测具有重要影响的变量。该方法通过计算贡献分数(SHAP 值),将低风险组作为参考组,量化高风险组患者中每个输入变量对预测风险的贡献,确定对低风险组和高风险组之间 PI 差异贡献最大的变量,识别对预测结果最具影响力的 mRNA 和 miRNA。

1.4 预测性能指标 本研究中使用一致性指数(concordance index, C-index)、时间依赖性受试者工作特征曲线下面积(Time-dependent AUC)、对数秩检验 P 值(Log-rank test P -value)来评价模型的预

测性能。

1.5 软件实现 本研究数据预处理以及变量筛选均在 R 4.5.3 软件中完成;在 Python 软件中构建深度学习模型,AutoSurv 模型使用 PyTorch 2.4.1 包在 Python 中实现,Deepsurv 模型基于 Tensorflow 1.15.0 平台实现,Cox-nnet 模型使用 Theano 1.0.5 包在 Python 中实现。

2 结 果

2.1 CESC 患者多组学数据的描述 从 UCSC Xena 网站下载 TCGA 数据库 CESC 患者的 mRNA、miRNA、临床信息数据和生存数据,经预处理后整合得到 261 个患者,mRNA、miRNA 去除批次效应后进行归一化。所用数据集关键特征如表 1 所示。

表 1 数据集关键特征的描述

Table 1 Description of key features in the dataset

类别	特征描述	特征数	样本量
mRNA	原始数量	60 660	310
	Reactome 通路注释基因数	8 809	310
	通路过滤后(移除基因数 < 15 或 > 300 的通路)	6 753	310
	染色体过滤后	6 719	310
	低方差过滤后	3 154	310
	最终 mRNA 特征数	3 154	310
	Reactome 通路		
Reactome 通路	初始通路数	1 525	
	过滤后(移除基因数 < 15 或 > 300 的通路)	581	
miRNA	原始数量	1 881	312
	低方差过滤后	708	312
	最终 miRNA 特征数	708	312
最终样本量	有完整组学数据		261

为了实现可靠的模型性能评估,本研究在划分数据集时权衡了训练充分性与验证可靠性,将数据集按 6:4 划分为训练集和测试集,经检验训练集和测试集的患者基本资料和结局分布均衡,见表 2。

表 2 TCGA-CESC 患者基本资料

Table 2 Basic information on TCGA-CESC patient

项目	Train (n = 166)	Test (n = 95)	χ^2/z	P 值
年龄(岁)	48.3 ± 14.0	46.9 ± 12.9	0.916	0.648
种族[n(%)]			2.289	0.130
White	134(80.7)	69(72.6)		
其他	32(19.3)	26(27.4)		
疾病分期[n(%)]			0.500	0.480
stage_i	112(67.5)	60(63.2)		
stage_ii	54(32.5)	35(36.8)		
生存状态[n(%)]			0.592	0.442
存活	126(75.9)	68(71.6)		
死亡	40(24.1)	27(28.4)		

2.2 模型性能对比 使用 TCGA 的 CESC 多组学数据分别构建 AutoSurv、DeepSurv 和 Cox-nnet 预后预

测模型,模型性能通过 C-index 和 1 年、3 年、5 年 AUC 进行评估,具体结果汇总于表 3。三个模型测试集上的 KM 生存曲线以及时间依赖性 ROC 曲线分别见图 1(AutoSurv)、图 2(DeepSurv)、图 3(Cox-nnet)。时间依赖性 ROC 曲线用于评估模型在不同随访时间点(诊断后 1、3、5 年)的预测准确性,其曲线下面积(AUC)可解释为在特定时间点,模型随机赋予一名最终发生事件患者的风险评分高于未发生事件患者的概率。AutoSurv 模型在 CESC 测试集上的 C-index 为 0.78(95% CI:0.75~0.81),表现出良好的判别能力,在 1 年、3 年和 5 年的 AUC 值分别为 0.84(95% CI:0.81~0.87)、0.78(95% CI:0.76~0.85)、0.77(95% CI:0.75~0.81)。Log-rank 检验的差异具有统计学意义($P=0.013$),表明模型能够有效区分高、低风险组患者,且高风险组预后较差。AutoSurv 在 CESC 数据集上的表现优于 DeepSurv 和 Cox-nnet 模型,具体体现如下:在判别准确性方面,AutoSurv 具有更高的 C-index 和各时间点 AUC 值;在风险分层

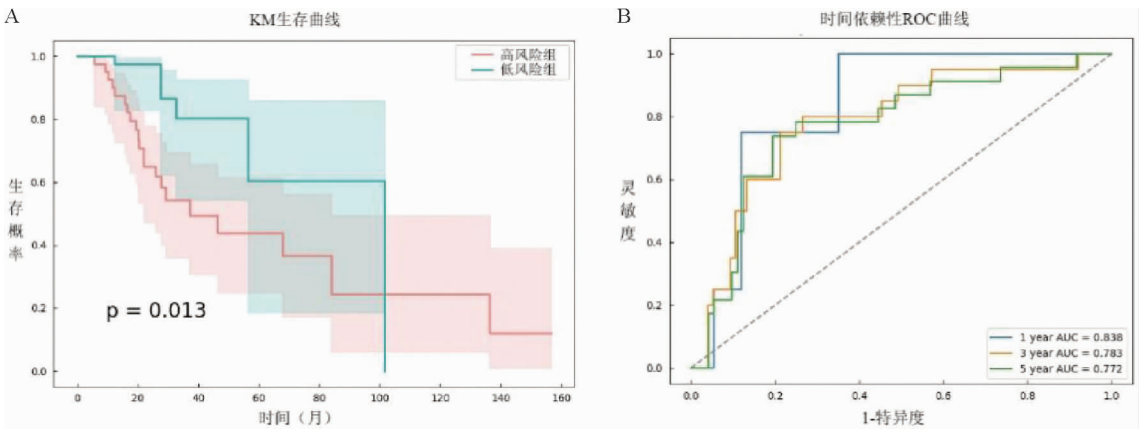
能力方面,其 KM 曲线高、低风险组分离更为显著, DeepSurv 模型的 KM 生存曲线(图 2A)未能有效区分不同风险组 ($P = 0.413$);在预测稳定性方面, AutoSurv 在时间依赖性 ROC 曲线(图 1B)中展现出卓越且持续的判别性能。该模型结合生物信号通路信息与其他模型相比预测性能增强,具有良好的鲁棒性。

表 3 模型预测性能评估结果

Table 3 The evaluation results of the models predictive performance

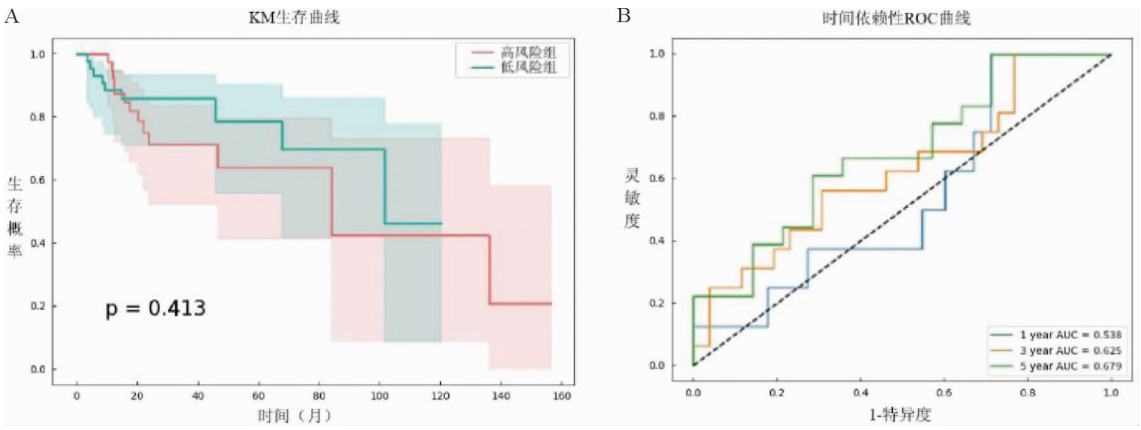
组别	Train			Test		
Model	AutoSurv	Deepsurv	Cox - nnet	AutoSurv	Deepsurv	Cox - nnet
C - index(95% CI)	0.85(0.82 ~ 0.87)	0.84(0.79 ~ 0.86)	0.84(0.69 ~ 0.88)	0.78(0.75 ~ 0.81)	0.59(0.55 ~ 0.62)	0.68(0.57 ~ 0.79)
AUC(1Year)(95% CI)	0.88(0.85 ~ 0.90)	0.87(0.83 ~ 0.89)	0.81(0.70 ~ 0.83)	0.84(0.81 ~ 0.87)	0.54(0.51 ~ 0.59)	0.75(0.56 ~ 0.71)
AUC(3Year)(95% CI)	0.85(0.82 ~ 0.88)	0.84(0.80 ~ 0.87)	0.85(0.74 ~ 0.87)	0.78(0.76 ~ 0.85)	0.63(0.59 ~ 0.67)	0.63(0.55 ~ 0.77)
AUC(5Year)(95% CI)	0.86(0.83 ~ 0.89)	0.81(0.75 ~ 0.83)	0.80(0.69 ~ 0.82)	0.77(0.75 ~ 0.81)	0.68(0.63 ~ 0.70)	0.70(0.51 ~ 0.76)
Log - rankP	<0.001	<0.001	<0.001	0.013	0.413	<0.01

注:加粗字体为最佳。



注:图 A:AutoSurv 模型 CESC 患者 KM 生存曲线;图 B:AutoSurv 模型 CESC 患者 1 年、3 年、5 年 ROC 曲线。

图 1 AutoSurv - CESC KM 曲线/ROC 曲线
Fig.1 AutoSurv - CESC KM/ROC curves

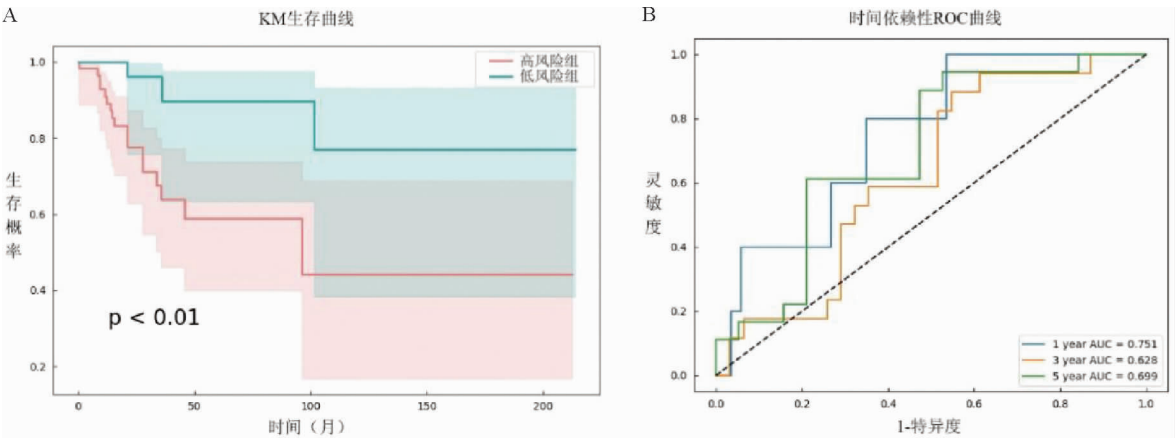


注:图 A:DeepSurv 模型 CESC 患者 KM 生存曲线;图 B:DeepSurv 模型 CESC 患者 1 年、3 年、5 年 ROC 曲线。

图 2 DeepSurv - CESC KM/ROC 曲线
Fig.2 DeepSurv - CESC KM/ROC curves

2.3 可解释性分析

2.3.1 特征重要性 为了揭示 AutoSurv 模型对 CESC 患者预后预测的决策机制,本研究采用 DeepSHAP进行可解释性分析,通过反向传播计算特



注:图 A:Cox - nnet 模型 CESC 患者 KM 生存曲线;图 B:Cox - nnet 模型 CESC 患者 1 年、3 年、5 年 ROC 曲线。

图 3 Cox - nnet - CESC KM/ROC 曲线
Fig. 3 Cox - nnet - CESC KM/ROC curves

征重要性,识别重要的潜在特征和临床变量。KL - PMVAE 提取的重要性排名前 10 的潜在特征和临床变量依次为:Z_5 (SHAP 值 = 0.540)、age (0.379)、Z_9 (0.255)、Z_10 (0.243)、Z_12 (0.208)、Z_3 (0.199)、Z_7 (0.166)、Z_15 (0.165)、Z_13 (0.157) 和 stage_i (0.156)。图 4 所示的 SHAP 蜂群图展示了每个特征对模型预后指数的影响方向和强度,每个点代表每个患者的 SHAP 值,红色表示高值,蓝色表示低值,SHAP 值为正表示增加死亡风险,为负表示降低死亡风险。分析结果表明,临床变量 age (年龄)重要性为第二 (0.379),随着年龄增长,患者的死亡风险上升;癌症分期为 stage_i (早期)降低模型预测的死亡风险。

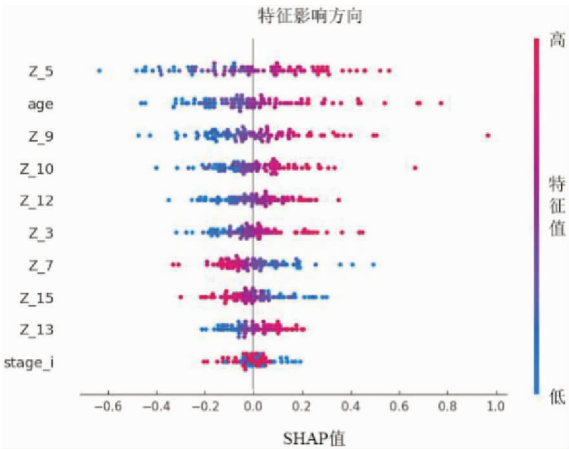


图 4 SHAP 蜂群图
Fig. 4 SHAP swarm plot

2.3.2 识别关键预测变量 计算每个输入变量 (mRNA、miRNA) 对潜在特征 z 的贡献分数,识别出对高低风险组区分贡献最大的 mRNA 和 miRNA。进一步评估除临床变量 age 和 stage_i 以外的前 8 个潜在

特征 (Z_5、Z_9、Z_10、Z_12、Z_3、Z_7、Z_15、Z_13) 的 KL - PMVAE 输入变量,识别每个潜在特征中重要性排名前 10 的关键变量,统计这些变量在 8 个潜在特征中出现的总频率。频率越高 (频率 > 1) 表明变量在模型预测中的稳健性越强。结果显示, CESC 中 POLR2H、FN1 在 5 个潜在特征中有较高的贡献分数, FYN、UBA52 和 RPS27A 在 4 个潜在特征中出现, PTPN11、UBC 和 PSMA5 在 3 个潜在特征中出现, PSMD12、PIK3R1、FGFR2 和 BTC 在 2 个潜在特征中出现。图 5 为重要性排名前 20 的关键预测变量及其出现频率,横轴表示 mRNA/miRNA 名称,纵轴表示该变量在潜在特征中出现的频率。

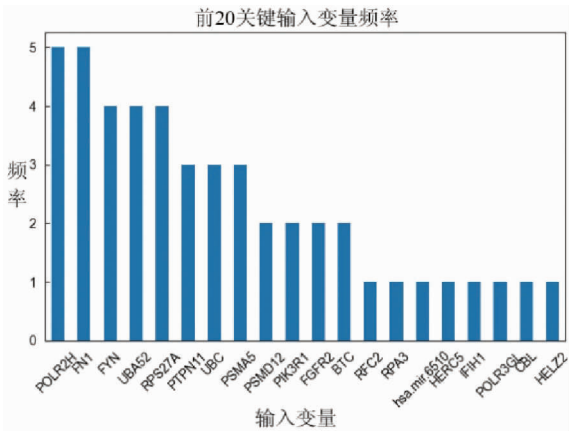


图 5 潜在特征关键输入变量频率分布
Fig. 5 Frequency distribution of key input factors for potential features

2.3.3 关键生物学通路 基于以上 8 个潜在特征识别出关键生物学通路,出现频率 > 1 的通路为稳健通路。结果显示 CESC 中 R - HSA - 5651801、R - HSA

- 1169091 和 R - HSA - 1474228 出现在 3 个潜在特征中, R - HSA - 166663、R - HSA - 354194 和 R - HSA - 1660661 等通路在 2 个潜在特征中出现。图 6 展示了重要性排名前 20 的生物学通路及其出现频率。经 Reactome 数据库查询, 高频通路主要涉及以下生物学过程: R - HSA - 5651801 (DNA 损伤修复)、R - HSA - 1169091 (非经典 NF - κ B 通路)、R - HSA - 1474228 (细胞外基质降解)、R - HSA - 166663 (补体初始激活)、R - HSA - 354194 (整合素信号通路)、R - HSA - 1660661 (鞘脂从头合成)、R - HSA - 202430 (ZAP - 70 转运)、R - HSA - 69231 (细胞周期检查点调控)、R - HSA - 2426168 (SREBF 激活基因表达)、R - HSA - 176412 (APC/C 磷酸化)。

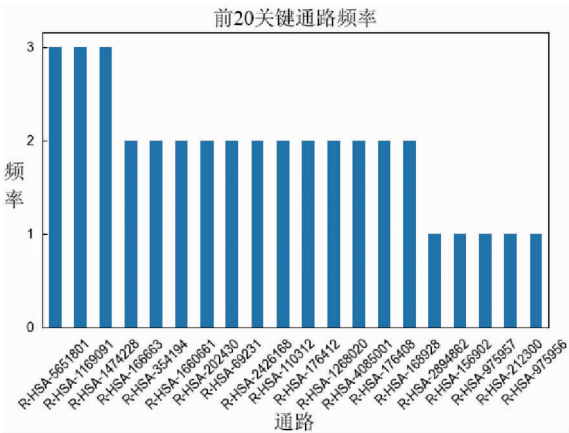


图 6 潜在特征关键输入通路频率分布

Fig. 6 Frequency distribution of key input pathway for potential features

2.4 蛋白质互作网络 通过 STRING 数据库 (<https://string-db.org/cgi/input.pl>) 构建了关键预测变量排名前 10 的 mRNA/miRNA (POLR2H、FN1、FYN、UBA52、RPS27A、PTPN11、UBC、PSMA5、PSMD12 和 PIK3R1) 的蛋白质互作网络 (PPI), 见图 7, 其中节点表示基因, 边表示蛋白质间的相互作用, 边的厚度和颜色强度反映了从 STRING 获得的相互作用置信度得分。结果表明: UBC (泛素 C) 和 RPS27A (核糖体蛋白 S27a) 为两大枢纽蛋白, 有广泛的连接性, 显示出泛素化修饰在宫颈鳞状细胞癌中的核心调控地位。网络可划分为三个主要的功能模块: 泛素 - 蛋白酶体系统模块 (UBC、UBA52、PSMA5、PSMD12), 细胞信号转导模块 (FN1、FYN、PTPN11、PIK3R1), 基因表达调控模块 (POLR2H、RPS27A)。

3 讨论

本研究首次将通路引导的可解释性深度学习模

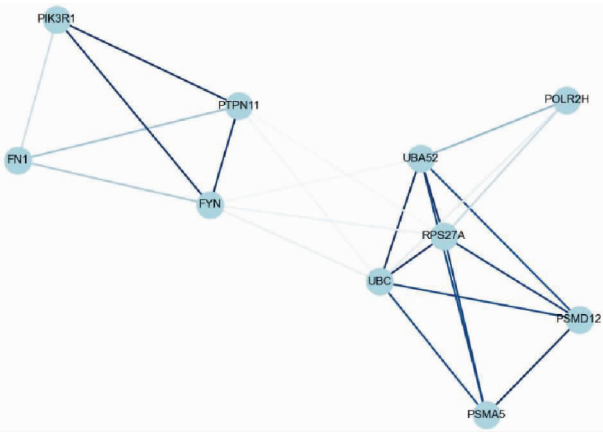


图 7 蛋白质互作网络图
Fig. 7 Protein interaction network graph

型 AutoSurv 应用于 CESC 的预后预测。结果表明, 与 DeepSurv 和 Cox - nnet 相比, AutoSurv 在预测准确性和可解释性方面均表现优越。原因体现在两个方面: 模型引入生物学先验知识、结合 DeepSHAP 解释框架。AutoSurv 成功识别出与 CESC 患者生存相关的 mRNA、miRNA 和信号通路, 为个体化预后评估提供了新工具, 为阐明 CESC 分子机制提供了潜在靶点。

SHAP 特征重要性结果提示临床应该重点关注高龄、处于癌症晚期 (IIB - IV) 的 CESC 患者。模型识别的 12 个关键输入变量, 已有研究证实与 CESC 预后相关: POLR2H 的表达水平反映肿瘤细胞的转录活性^[8]; FN1 参与细胞外基质重构, 调控肿瘤细胞的黏附、迁移和侵袭过程^[9]; FYN 调控整合素信号通路和细胞骨架重塑, 其失调与迁移增强和治疗抵抗相关^[9]; RPS27A 和 UBA52 维持蛋白质稳态和翻译调控^[10], 泛素化修饰异常在宫颈癌的发生发展中具有核心地位^[11]; PTPN11 是受体酪氨酸激酶信号通路的关键节点, 调控细胞增殖与存活^[12]。

在生物学通路层面, 识别的 Reactome 通路与 CESC 的发生发展机制高度吻合: DNA 损伤修复异常是宫颈癌诱导癌变的核心环节^[13]; NF - κ B 持续激活促进肿瘤细胞存活和化疗抵抗^[14]; 细胞外基质降解与肿瘤侵袭转移密切相关^[9]; 补体激活通过保护肿瘤细胞促进宫颈癌进展^[15]; 整合素信号过表达与富含 FN1 的细胞外基质相互作用促进肿瘤细胞迁移^[16]; 鞘脂生物合成产生的有毒代谢物对癌细胞的存活至关重要^[17]。

蛋白质互作网络揭示了基因和通路之间的关联, 将关键基因整合为三个功能模块: 泛素 - 蛋白酶体系统模块调控蛋白质稳态、DNA 损伤修复和细胞周期检查点等过程; 细胞信号转导模块介导细胞外基质信号向胞内传递, 驱动细胞迁移和侵袭; 基因表达调控模块维持转录和翻译的精确性。UBC 和 RPS27A 作为枢纽蛋白连接各模块。已有研究表明 FN1 通过激活 MAPK 通路影响细胞迁移, 在宫颈癌的发展和化疗

耐药性中有重要作用^[18];PTPN11 高表达水平与宫颈癌风险升高相关^[12];PIK3R1 通过 PIK3/AKT 通路调控肿瘤细胞的增殖和迁移^[19]。从蛋白质互作角度验证了基因和通路层面发现的内在一致性。

本研究存在一定局限性:第一,样本量有限,仅使用两种组学数据,未纳入甲基化数据、蛋白质组学数据等更多复杂数据;第二,只探讨了宫颈鳞状细胞癌数据,未扩展到其他癌症类型;第三,部分分析结果的统计效能有限,但所有评估指标均指向同一方向,Bootstrap 置信区间排除无效假设,生物学解释与已知通路相符。后续研究将纳入更多类型的组学数据,在更大样本量的独立队列中进行验证,深入探索模型识别的其他通路和靶点。

利益冲突声明 本研究不存在任何利益冲突

参考文献

[1] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2021 [J]. CA - A Cancer Journal for Clinicians, 2021, 71(1): 7 - 33.

[2] Ge S, Liu J, Cheng Y, et al. Multi - view spectral clustering with latent representation learning for applications on multi - omics cancer subtyping[J]. Brief Bioinform, 2023, 24(1): bbac500.

[3] Qi J, Lin W. Dynamic modeling, optimization, and deep learning for high - dimensional complex biological data [J]. Sci Sin Math, 2025, 55(7): 1413.

[4] Jiang LD, Xu C, Bai YT, et al. Autosurv: interpretable deep learning framework for cancer survival analysis incorporating clinical and multi - omics data[J]. NPJ Precis Oncol, 2024, 8(1): 4.

[5] 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.

[6] Ching T, Zhu X, Garmire LX. Cox - nnet: an artificial neural network method for prognosis prediction of high - throughput omics data[J]. PLOS Computational Biology, 2018, 14(4): e1006076.

[7] Cambridge Explorations in Arts and Sciences. Auto - Encoding Variational Bayes[EB/OL]. [2026 - 06 - 08]. <https://arxiv.org/abs/1312.6114>.

[8] Bhandari N, Acharya D, Chatterjee A, et al. Pan - cancer integrated bioinformatic analysis of RNA polymerase subunits reveal

RNA Pol I member CD3EAP regulates cell growth by modulating autophagy[J]. Cell Cycle, 2023, 22(18): 1986 - 2002.

[9] Hałas - Wiśniewska M, Zawadka P, Arendt W, et al. From adhesion to invasion: Integrins, focal adhesion signaling, and actin binding proteins in cervical cancer progression - a scoping review [J]. Cells, 2025, 14(20): 1640.

[10] Yan ZC, Yu JW, Wang SY, et al. Identification of E3 ubiquitin ligase - associated prognostic genes and construction of a prediction model for uterine cervical cancer based on bioinformatics analysis [J]. Hormones & Cancer, 2024, 15(1): 395.

[11] Yao S, Chen S, Wang A, et al. BAG2 inhibits cervical cancer progression by modulating type I interferon signaling through stabilizing Sting[J]. Advanced Science, 2025, 12(29): e70005.

[12] Li S, Wang X, Li Q, et al. Role of SHP2/PTPN11 in the occurrence and prognosis of cancer: a systematic review and meta - analysis[J]. Oncology Letters, 2022, 25(1): 19.

[13] Elayappillai SP, Dogra S, Lausen J, et al. ATR inhibition increases reliance on PARP - mediated DNA repair revealing an improved therapeutic strategy for cervical cancer[J]. Gynecologic Oncology, 2024, 191(2): 182 - 193.

[14] Zhong L, Li J, Zhong J, et al. IFI16 induced by p53 activates the NF - κ B pathway to counteract cisplatin - induced apoptosis in cervical cancer cells [J]. Journal of Cellular and Molecular Medicine, 2025, 29(15): e70728.

[15] Khan A, Das BC, Abiha U, et al. Insights into the role of complement regulatory proteins in HPV mediated cervical carcinogenesis[J]. Seminars in Cancer Biology, 2022, 86(15): 583 - 589.

[16] Hazelbag S, Kenter G, Gorter A, et al. Overexpression of the α v β 6 integrin in cervical squamous cell carcinoma is a prognostic factor for decreased survival[J]. Journal of Pathology, 2007, 212(3): 316 - 324.

[17] Spears ME, Lee N, Hwang S, et al. De novo sphingolipid biosynthesis necessitates detoxification in cancer cells [J]. Cell Reports, 2022, 40(13): 111415.

[18] Jin S, Wang W, Xu X, et al. miR - 34b - 3p - mediated regulation of STC2 and FN1 enhances chemosensitivity and inhibits proliferation in cervical cancer: MiR - 34b - 3p targets STC2 and FN1 to enhance chemosensitivity [J]. Acta Biochimica et Biophysica Sinica, 2024, 56(5): 740.

[19] Wang Y, Chen A, Zheng C, et al. miR - 92a promotes cervical cancer cell proliferation, invasion, and migration by directly targeting PIK3R1 [J]. Journal of Clinical Laboratory Analysis, 2021, 35(8): e23893.

收稿日期:2025-09-06

(上接第 2655 页)

[34] Ali FEM, Badran KSA, El - Maksoud MSA, et al. The role of Wnt/ β - catenin signaling in lung cancer progression and therapy: a comprehensive review [J]. Medical Oncology, 2025, 42(6): 183.

[35] Lam AP, Gottardi CJ. β - catenin signaling: a novel mediator of fibrosis and potential therapeutic target[J]. Current Opinion in Rheumatology, 2011, 23(6): 562 - 567.

[36] Liu T, GonzalezDLSF, Hirsch M, et al. Noncanonical Wnt signaling promotes myofibroblast differentiation in pulmonary fibrosis [J]. American Journal of Respiratory Cell and Molecular Biology, 2021, 65(5): 489 - 499.

[37] Chai D, Du H, Shi Y. Advances in the biological functions of miR - 205 in various diseases (Review)[J]. Mol Med Rep, 2026, 33(5).

[38] Yang WJ, Tan SM, Yang LX, et al. Exosomal miR - 205 - 5p enhances angiogenesis and nasopharyngeal carcinoma metastasis by targeting desmocollin - 2 [J]. Molecular Therapy - Oncolytics, 2022, 24(2): 612 - 623.

[39] 朱梦杰,丁文俊,冯天照,等. 特发性肺纤维化实验模型构建现状与应用进展[J]. 中国实验动物学报,2025,33(10):1522 -

1530.

Zhu MJ, Ding WJ, Feng TZ, et al. Current status and application advances in experimental models of idiopathic pulmonary fibrosis [J]. Acta Laboratorium Animalis Scientia Sinica, 2025, 33(10): 1522 - 1530. (In Chinese)

[40] Ishida Y, Kuninaka Y, Mukaida NFI, et al. Immune mechanisms of pulmonary fibrosis with bleomycin [J]. International Journal of Molecular Sciences, 2023, 24(4): 3149.

[41] Della latta V, Cecchettiini A, Del Ry S, et al. Bleomycin in the setting of lung fibrosis induction: From biological mechanisms to counteractions[J]. Pharmacological Research, 2015, 97(1): 122 - 130.

[42] Li R, Jia YZ, Kong XH, et al. Novel drug delivery systems and disease models for pulmonary fibrosis [J]. Journal of Controlled Release, 2022, 348(2): 95 - 114.

[43] Diener C, Keller A, Meese E. Emerging concepts of miRNA therapeutics: from cells to clinic [J]. Trends in Genetics, 2022, 38(6): 613 - 626.

收稿日期:2026-04-04