



doi:10.7659/j.issn.1005-6947.2024.09.013  
http://dx.doi.org/10.7659/j.issn.1005-6947.2024.09.013  
China Journal of General Surgery, 2024, 33(9):1459-1472.

· 专题研究 ·

## 基于机器学习算法的胰腺导管腺癌预后模型构建及其验证

张业光<sup>1</sup>, 赵攀<sup>2</sup>, 章慧<sup>2</sup>, 黄正红<sup>3</sup>, 黄坤<sup>2</sup>

(四川省绵阳市中医医院 1. 超声医学科 2. 普通外科, 四川 绵阳 621000; 3. 成都中医药大学医学技术学院, 四川 成都 611137)

### 摘要

**背景与目的:** 胰腺导管腺癌 (PDAC) 是胰腺癌中最常见的病理类型, 其远期预后差, 缺乏个体化的预后评价工具。本研究通过 SEER 数据库中的大样本真实世界数据, 基于机器学习算法, 构建 PDAC 患者预后列线图, 旨在精准化、个体化评价 PDAC 患者的预后, 为临床决策制定提供参考。

**方法:** 根据纳入和排除标准, 提取 SEER 数据库 2000—2018 年期间经病理学确诊为 PDAC 患者的临床病理及预后资料。按 7:3 随机分为训练集和验证集。在训练集中, 分别采用单 (多) 因素 Cox 比例风险模型、LASSO 回归模型和随机生存森林模型筛选影响 PDAC 预后的独立因素, 构建预测 6、12、36 个月肿瘤特异性生存期 (CSS) 和总生存期 (OS) 的列线图模型。随后, 分别在训练集和验证集中利用一致性指数 (C 指数)、受试者工作特征 (ROC) 曲线、校准曲线、生存曲线、决策曲线分析对模型进行验证和评估。

**结果:** 本研究共纳入 4 237 例患者, 其中训练集 2 965 例, 验证集 1 272 例, 两组基线资料均衡可比。训练集和验证集中患者的中位随访时间分别为 18 (9~36) 个月和 18 (9~37) 个月。多因素 Cox 比例风险模型显示, 年龄、T 分期、N 分期、M 分期、分化程度、手术、系统治疗和化疗是 OS 的独立影响因素 (均  $P < 0.05$ ); 年龄、T 分期、N 分期、M 分期、分化程度、手术和化疗是 CSS 的独立影响因素 (均  $P < 0.05$ )。LASSO 回归模型显示, 年龄、分化程度、T 分期、N 分期、M 分期、化疗、手术、淋巴清扫范围、放疗和系统治疗与 OS 相关; 分化程度、T 分期、N 分期、M 分期、化疗、手术、淋巴清扫范围、放疗和系统治疗与 CSS 相关。随机生存森林模型显示, 影响 OS 的重要性评分前五位变量分别为: 系统治疗、分化程度、N 分期、化疗和 T 分期; 而影响 CSS 的重要性评分前五位变量分别为: 系统治疗、分化程度、N 分期、化疗和 AJCC 分期。基于多因素 Cox 回归模型、LASSO 回归模型和随机生存森林模型的分析结果并结合临床重要性, 最终选择年龄、T 分期、N 分期、M 分期、分化程度、手术和化疗, 共七个临床特征成功构建预测 6、12、36 个月的 OS 和 CSS 的预测模型。模型验证结果表明, 对于 OS, 在训练集和验证集中的 C 指数分别为 0.692 (95% CI=0.681~0.704) 和 0.680 (95% CI=0.664~0.698); 对于 CSS, 在训练集和验证集中的 C 指数分别为 0.696 (95% CI=0.684~0.707) 和 0.680 (95% CI=0.662~0.698)。ROC 曲线表明模型具有良好的预测价值; 校准曲线均靠近理想的 45° 参考线。

**结论:** 年龄、TNM 分期、分化程度、手术和化疗是 PDAC 患者预后的独立影响因素; 基于这些变量构建的预测模型, 有较高的区分度和准确度。有助于临床医师为 PDAC 患者制定精准的、个体化的治疗和随访方案。

### 关键词

胰腺肿瘤; 预后; SEER 规划; 机器学习; 列线图

中图分类号: R735.9

**基金项目:** 四川省绵阳市卫健委基金资助项目 (202309); 四川省绵阳市中医医院基金资助项目 (MYSZYKYKT2023117)。

**收稿日期:** 2023-12-28; **修订日期:** 2024-03-14。

**作者简介:** 张业光, 四川省绵阳市中医医院主任医师, 主要从事超声医学 (普外方向) 方面的研究。

**通信作者:** 黄坤, Email: mshuangkun@163.com

# Construction and validation of a prognostic model for pancreatic ductal adenocarcinoma based on machine learning algorithm

ZHANG Yeguang<sup>1</sup>, ZHAO Pan<sup>2</sup>, ZHANG Hui<sup>2</sup>, HUANG Zhenghong<sup>3</sup>, HUANG Kun<sup>2</sup>

(1. Department of Ultrasound Medicine 2. Department of General Surgery, Mianyang Traditional Chinese Medicine Hospital, Mianyang, Sichuan 621000, China; 3. College of Medical Technology, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China)

## Abstract

**Background and Aims:** Pancreatic ductal adenocarcinoma (PDAC) is the most common pathological type of pancreatic cancer, with a poor long-term prognosis and a lack of individualized prognostic assessment tools. This study was conducted to construct a prognostic nomogram for PDAC patients based on large-sample real-world data from the SEER database using machine learning algorithms to provide precise and individualized prognostic evaluations to inform clinical decision-making.

**Methods:** The clinical and prognostic data of PDAC patients pathologically diagnosed from 2000 to 2018 were extracted from the SEER database based on inclusion and exclusion criteria. The data were randomly divided into training (70%) and validation (30%) sets. In the training set, independent prognostic factors were identified using univariate and multivariate Cox proportional hazards models, LASSO regression, and random survival forests. A nomogram was developed to predict 6, 12, and 36-month cancer-specific survival (CSS) and overall survival (OS). The model was then validated and assessed in both training and validation sets using the concordance index (C-index), receiver operating characteristic (ROC) curve, calibration curve, survival curves, and decision curve analysis.

**Results:** A total of 4 237 patients were included, with 2 965 in the training set and 1 272 in the validation set, showing comparable baseline characteristics. The median follow-up time was 18 (9–36) months for the training set and 18 (9–37) months for the validation set. The multivariate Cox model indicated that age, T stage, N stage, M stage, differentiation, surgery, systemic therapy, and chemotherapy were independent factors for OS (all  $P<0.05$ ). For CSS, age, T stage, N stage, M stage, differentiation, surgery, and chemotherapy were independent factors (all  $P<0.05$ ). The LASSO regression model found that age, differentiation, T stage, N stage, M stage, chemotherapy, surgery, lymph node dissection, radiotherapy, and systemic therapy were associated with OS, while T stage, N stage, M stage, chemotherapy, surgery, lymph node dissection, radiotherapy, and systemic therapy were linked to CSS. The random survival forest model identified the top five variables affecting OS as systemic therapy, differentiation, N stage, chemotherapy, and T stage; and for CSS, they were systemic therapy, differentiation, N stage, chemotherapy, and AJCC stage. Based on the analyses from the multivariate Cox, LASSO, and random survival forest model, along with clinical significance, a prediction model was successfully constructed using seven clinical features: age, T stage, N stage, M stage, differentiation, surgery, and chemotherapy to predict OS and CSS at 6, 12, and 36 months. The validation results showed C-indexes of 0.692 (95% CI=0.681–0.704) and 0.680 (95% CI=0.664–0.698) for OS in the training and validation sets, respectively, and 0.696 (95% CI=0.684–0.707) and 0.680 (95% CI=0.662–0.698) for CSS. ROC curves indicated good predictive value, and calibration curves closely matched the ideal 45° reference line.

**Conclusion:** Age, TNM stage, differentiation, surgery, and chemotherapy are independent prognostic factors for PDAC patients. The prognostic model based on these variables has high discrimination and

accuracy, assisting clinicians in developing precise and personalized treatment and follow-up plans for PDAC patients.

**Key words** Pancreatic Neoplasms; Prognosis; SEER Program; Machine Learning; Nomograms

**CLC number** R735.9

胰腺癌是预后极差的恶性肿瘤,其5年总生存率约为11.5%,是人类健康的严重威胁<sup>[1-4]</sup>。而胰腺导管腺癌(pancreatic ductal adenocarcinoma, PDAC)作为胰腺癌中最常见的病理学类型,其预后更差,5年总生存期(overall survival, OS)率约为6%<sup>[1]</sup>。虽然,手术治疗被认为是该肿瘤最佳的治疗手段<sup>[5]</sup>。然而,有近50%的PDAC患者在初诊时已经出现包括肝脏在内的远处器官转移<sup>[6-8]</sup>。因而,这部分患者在初诊时,往往失去了手术治疗的机会。另一方面,同样初诊IV期肿瘤,其预后也可能存在差异<sup>[9]</sup>。因此,构建准确预测PDAC患者生存率的预测模型,并对其进行危险分层,从而制定个体化的治疗方案,这对PDAC患者的治疗和预后评价将大有裨益。列线图可以纳入多种与预后相关因素,将不同因素对生存预后影响的大小进行量化并将结果可视化地展现出来,以对患者的生存率进行个体化的预测,近年来在临床医学领域被广泛应用<sup>[10-13]</sup>。基于美国SEER(Surveillance, Epidemiology, and End Results, SEER)数据库多中心、大样本病例数的特点,本研究回顾分析SEER数据库中经过病理学确诊的PDAC患者的临床病理及预后资料,构建PDAC患者生存率的预测模型,旨在探讨PDAC患者预后的独立影响因素及其列线图预测模型的临床应用价值,以期指导临床实践。

## 1 资料与方法

### 1.1 资料来源

通过SEER\*Stat v8.3.9软件提取美国SEER数据库中2000—2018年期间,经过病理学确诊为PDAC患者的临床病理及预后资料。随访截止日期为2018年12月31日。

### 1.2 纳入和排除标准

纳入标准:(1)患者初诊时为PDAC;(2)经病理学确诊;(3)病理学类型符合ICD-O-3编码为:“8500/3”。排除标准:(1)多源性肿瘤;(2)经过尸

检或死亡证明确诊的病例;(3)研究指标无法获取;(4)临床或随访信息记录不完整。

### 1.3 提取数据

参照笔者<sup>[14-18]</sup>先前的研究,提取患者的诊断年龄、种族、性别、婚姻状况、肿瘤大小、分期、治疗信息、生存时间及生存结局等资料。根据数据特征并结合文献<sup>[16,19-22]</sup>报道,本研究中,年龄界定标准设置:<60岁和≥60岁。TNM分期采用美国癌症联合委员会(American Joint Committee on Cancer, AJCC)第7版。

### 1.4 统计学处理

本研究的研究终点分别为:肿瘤特异性生存期(cancer-specific survival, CSS)和OS。CSS定义:随访期间因PDAC死亡的时长(月),非PDAC死亡,随访截止,或失访数据作为删失数据;OS定义:随访期间因任何病因死亡的时长(月),随访截止或失访数据作为删失数据<sup>[23]</sup>。应用Stata/MP 16.0软件和R(version 4.2.3)软件进行统计分析。正态分布计量资料以均数±标准差( $\bar{x} \pm s$ )表示,非正态分布计量资料采用中位数(四分位间距)[ $M(IQR)$ ]进行描述。计数资料采用率表示,组间对比采用 $\chi^2$ 检验。按照7:3将数据集随机划分为训练集和验证集,训练集用于列线图模型的构建及其内部验证,验证集被用于外部验证。在训练集中,先构建单因素Cox比例风险模型,根据单因素Cox分析结果并结合临床重要性<sup>[24]</sup>,进入多因素Cox比例风险模型分析,最终选出独立预后因素。根据多因素Cox比例风险模型计算风险比(HR)及对应95%置信区间(CI)<sup>[24]</sup>。在多因素Cox风险比例模型分析中,用方差扩展因子(variance inflation factors, VIFs)评估模型多重共线性,当取值大于10,表明变量间存在严重的多重共线性问题,需排除后方可进行后续分析。同步运用LASSO(least absolute shrinkage and selection operator)算法和随机生存森林模型进行独立的预后分析。LASSO方法的原理<sup>[15]</sup>是在最小二乘法的基

础上增加一个惩罚项来压缩估计参数，进而对预后影响较大的自变量进行自动筛选并计算出相应的回归系数。模型筛选的主要参数是  $\lambda_{\min}$  和  $\lambda_{1se}$ 。前者是指，获得最小目标参量均值的  $\lambda$  值，而后者是指在  $\lambda_{\min}$  一个方差  $se$  范围内得到最简单模型的  $\lambda$  值。由于  $\lambda$  值到达一定值之后，再继续增加模型自变量个数并不能很显著地提高模型性能。因此， $\lambda_{1se}$  可以给出一个性能优良，同时自变量数最少的模型。随机生存森林是一种机器学习算法，具有高鲁棒性，它不受比例风险假定、对数线性假定等条件的约束，并且能够通过两个随机采样的过程来防止算法的过度拟合问题<sup>[25]</sup>。建模变量筛选原则：基于多因素 Cox 回归模型、LASSO 回归模型和随机生存森林模型的分析结果并结合临床重要性最终决定入选变量用于模型构建。通过 Bootstrap 1 000 次重抽样的方法，绘制校准曲线和受试者工作特征（ROC）曲线对模型准确度进行验证；计算一致性指数（C 指数）对模型区分度进行验证。评价模型在临床实践中的应用价值，根据构建的模型分别在训练集和验证集中计算出每例患者的总分值，将训练集中总分值的中位数设置为危险分层阈值<sup>[23]</sup>。采用 Kaplan-Meier 法计算生存率，生存率比较采用 Log-rank 法。所有检验均为双侧尾。检验水准  $\alpha=0.05$ 。

2 结 果

2.1 纳入病例特征

经严格筛选，最终 4 237 患者符合纳入标准，按 7:3 随机分为训练集（ $n=2\,965$ ）和验证集（ $n=1\,272$ ）（图 1）。两组间的基线特征均衡（均  $P<0.05$ ）（表 1）。2 965 例训练集患者中，死亡 2 496 例（84.18%），其中肿瘤相关性死亡 2 268 例（76.49%）。1 272 例验证集患者中，死亡 1 065 例（83.73%），其中肿瘤相关性死亡 966 例（75.94%）。

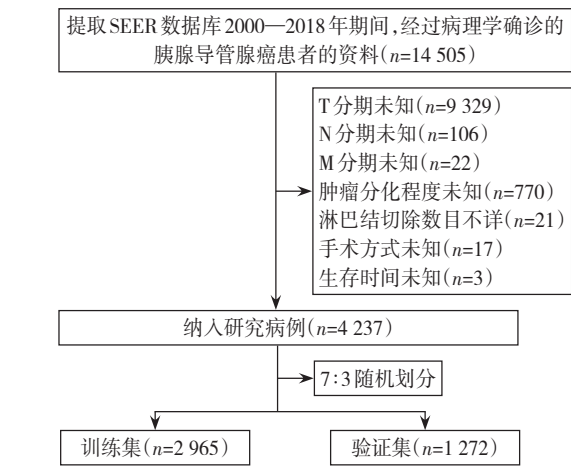


图 1 病例筛选流程图  
Figure 1 Case screening process

表 1 训练集和验证集 PDAC 患者的基线特征表[n (%) ]  
Table 1 Baseline characteristics of patients with PDAC in training set and validation set [n (%)]

变量	训练集( <i>n</i> =2 965)	验证集( <i>n</i> =1 272)	<i>P</i>
年龄(岁)			
<60	664(22.39)	293(23.03)	0.677
≥60	2 301(77.61)	979(76.97)	
性别			
女性	1 473(49.68)	649(51.02)	0.443
男性	1 492(50.32)	623(48.98)	
种族			
白种人	2 385(80.44)	1 036(81.45)	0.711
黑种人	305(10.29)	127(9.98)	
其他种族	275(9.27)	109(8.57)	
婚姻			
已婚	1 855(62.56)	789(62.03)	0.082
未婚	338(11.40)	160(12.58)	
未知	107(3.61)	63(4.95)	
离婚/丧偶	665(22.43)	260(20.44)	
肿瘤部位			
胰头	2 074(69.95)	892(70.12)	0.110
胰体	248(8.36)	131(10.30)	
胰尾	354(11.94)	131(10.30)	
其他	289(9.75)	118(9.28)	
分化程度			
高分化	332(11.20)	145(11.40)	0.284
中分化	1 582(53.35)	708(55.66)	
低/未分化	1 051(35.45)	419(32.94)	
T分期			
T1	138(4.66)	70(5.50)	0.289
T2	325(10.96)	156(12.27)	
T3	2 286(77.10)	947(74.45)	
T4	216(7.28)	99(7.78)	
N分期			
N0	1 023(34.5)	479(37.66)	0.053
N1	1 942(65.5)	793(62.34)	



表 1 训练集和验证集PDAC患者的基线特征表[n (%) ] (续)  
Table 1 Baseline characteristics of patients with PDAC in training set and validation set [n (%) ] (continued)

变量	训练集(n=2 965)	验证集(n=1 272)	P
M分期			
M0	2 719(91.70)	1 157(90.96)	0.462
M1	246(8.30)	115(9.04)	
AJCC分期			
I	255(8.60)	129(10.14)	0.207
II	2 317(78.14)	957(75.24)	
III	147(4.96)	71(5.58)	
IV	246(8.30)	115(9.04)	
化疗			
无	871(29.38)	394(30.97)	0.315
有	2 094(70.62)	878(69.03)	
手术			
无	266(8.97)	129(10.14)	0.253
有	2 699(91.03)	1 143(89.86)	
手术方式			
非手术治疗	266(8.97)	129(10.14)	0.486
肿瘤局部切除或部分胰腺切除	2 287(77.13)	969(76.18)	
全胰腺切除或扩大胰十二指肠切除术	412(13.9)	174(13.68)	
淋巴清扫			
1~3枚淋巴结	110(3.71)	58(4.56)	0.299
≥4枚淋巴结	2 554(86.14)	1 075(84.51)	
未清扫	301(10.15)	139(10.93)	
放疗			
无	2 184(73.66)	921(72.41)	0.419
有	781(26.34)	351(27.59)	
系统治疗			
无	1 019(34.37)	464(36.48)	0.199
有	1 946(65.63)	808(63.52)	

表 2 PDAC 患者的单因素 Cox 分析  
Table 2 Univariate Cox regression analysis for CSS and OS in PDAC patients

变量	OS		CSS	
	HR(95% CI)	P	HR(95% CI)	P
年龄(岁)				
<60	1		1	
≥60	1.14(1.04~1.26)	0.007	1.10(1.00~1.22)	0.052
性别				
女性	1		1	
男性	1.08(0.99~1.16)	0.068	1.04(0.96~1.13)	0.386
种族				
白种人	1		1	
黑种人	1.10(0.97~1.25)	0.144	1.06(0.93~1.22)	0.374
其他种族	0.93(0.81~1.07)	0.322	0.90(0.78~1.05)	0.175
婚姻				
已婚	1		1	
未婚	1.18(1.04~1.34)	0.009	1.17(1.03~1.34)	0.019
未知	1.22(0.99~1.50)	0.060	1.16(0.93~1.45)	0.185

2.2 PDAC 患者的预后分析

2.2.1 单因素 Cox 比例风险模型 在训练集 2 965 例患者中, 中位随访 18 (9~36) 个月。在验证集 1 272 例患者中, 中位随访 18 (9~37) 个月。单因素 Cox 比例风险模型分析显示: 年龄、婚姻状况、分化程度、T 分期、N 分期、M 分期、AJCC 分期、化疗、手术、手术方式、淋巴清扫范围、放疗和系统治疗与患者的 OS 相关 (均  $P<0.05$ ); 婚姻状况、肿瘤部位、分化程度、T 分期、N 分期、M 分期、AJCC 分期、化疗、手术、手术方式、淋巴清扫范围、放疗和系统治疗与患者的 CSS 相关 (均  $P<0.05$ ) (表 2)。

2.2.2 多因素 Cox 比例风险模型 结合临床重要性和单因素 Cox 比例风险模型分析结果, 进一步筛选变量纳入多因素 Cox 比例风险模型分析, 结果显示: 年龄、T 分期、N 分期、M 分期、分化程度、手术、系统治疗和化疗是 OS 的独立影响因素 (均  $P<0.05$ ); 年龄、T 分期、N 分期、M 分期、分化程度、手术和化疗是 CSS 的独立影响因素 (均  $P<0.05$ ) (图 2)。

2.2.3 LASSO 回归模型 为了尽可能减少变量共线性对结果的影响, 采用了基于 10 折交叉验证方法 LASSO 回归分析, 以 lambda.1se 为模型参数, 结果显示: 年龄、分化程度、T 分期、N 分期、M 分期、化疗、手术、淋巴清扫范围、放疗和系统治疗与 OS 相关, 此时 lambda.1se=0.034; 分化程度、T 分期、N 分期、M 分期、化疗、手术、淋巴清扫范围、放疗和系统治疗与 CSS 相关, 此时 lambda.1se=0.037 (图 3)。

表2 PDAC 患者的单因素 Cox 分析 (续)  
Table 2 Univariate Cox regression analysis for CSS and OS in PDAC patients (continued)

变量	OS		CSS	
	HR(95% CI)	P	HR(95% CI)	P
肿瘤部位				
胰头	1		1	
胰体	0.96(0.83~1.11)	0.622	0.96(0.82~1.11)	0.580
胰尾	1.02(0.90~1.16)	0.731	0.95(0.83~1.08)	0.431
其他	1.11(0.97~1.27)	0.114	1.15(1.00~1.32)	0.045
分化程度				
高分化	1		1	
中分化	1.39(1.21~1.59)	<0.001	1.45(1.25~1.68)	<0.001
低/未分化	2.02(1.75~2.33)	<0.001	2.14(1.84~2.49)	<0.001
T 分期				
T1	1		1	
T2	2.36(1.82~3.05)	<0.001	2.35(1.79~3.10)	<0.001
T3	2.73(2.16~3.44)	<0.001	2.86(2.23~3.66)	<0.001
T4	4.58(3.51~5.98)	<0.001	4.72(3.56~6.27)	<0.001
N 分期				
N0	1		1	
N1	1.39(1.28~1.51)	<0.001	1.45(1.33~1.59)	<0.001
M 分期				
M0	1		1	
M1	2.63(2.30~3.01)	<0.001	2.72(2.36~3.12)	<0.001
AJCC 分期				
I	1		1	
II	1.94(1.65~2.28)	<0.001	2.15(1.80~2.57)	<0.001
III	3.02(2.40~3.80)	<0.001	3.34(2.62~4.27)	<0.001
IV	4.93(4.02, 6.03)	<0.001	5.60(4.50~6.96)	<0.001
化疗				
无	1		1	
有	0.52(0.48~0.57)	<0.001	0.54(0.50~0.59)	<0.001
手术				
无	1		1	
有	0.30(0.27~0.35)	<0.001	0.29(0.26~0.34)	<0.001
手术方式				
非手术治疗	1		1	
肿瘤局部切除或部分胰腺切除	0.30(0.26~0.34)	<0.001	0.29(0.25~0.33)	<0.001
全胰腺切除或扩大胰十二指肠切除术	0.34(0.29~0.40)	<0.001	0.33(0.28~0.39)	<0.001
淋巴清扫				
1~3 枚淋巴结	1		1	
≥4 枚淋巴结	0.90(0.73~1.11)	0.331	0.94(0.75~1.17)	0.579
未清扫	2.40(1.90~3.04)	<0.001	2.58(2.01~3.32)	<0.001
放疗				
无	1		1	
有	0.70(0.64~0.77)	<0.001	0.71(0.65~0.78)	<0.001
系统治疗				
无	1		1	
有	0.47(0.43~0.51)	<0.001	0.47(0.44~0.52)	<0.001

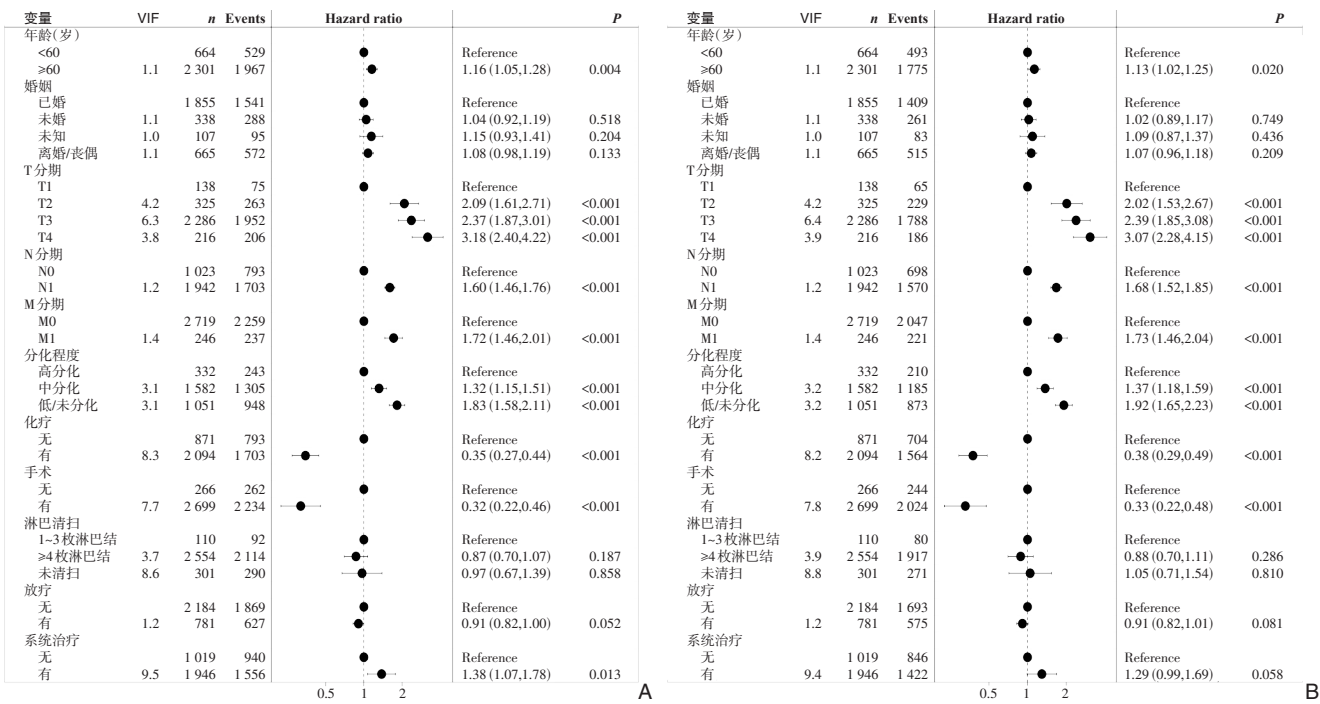


图2 多因素Cox回归分析结果森林图 A: OS; B: CSS

Figure 2 Forest plot of multivariate Cox regression analysis results A: OS; B: CSS

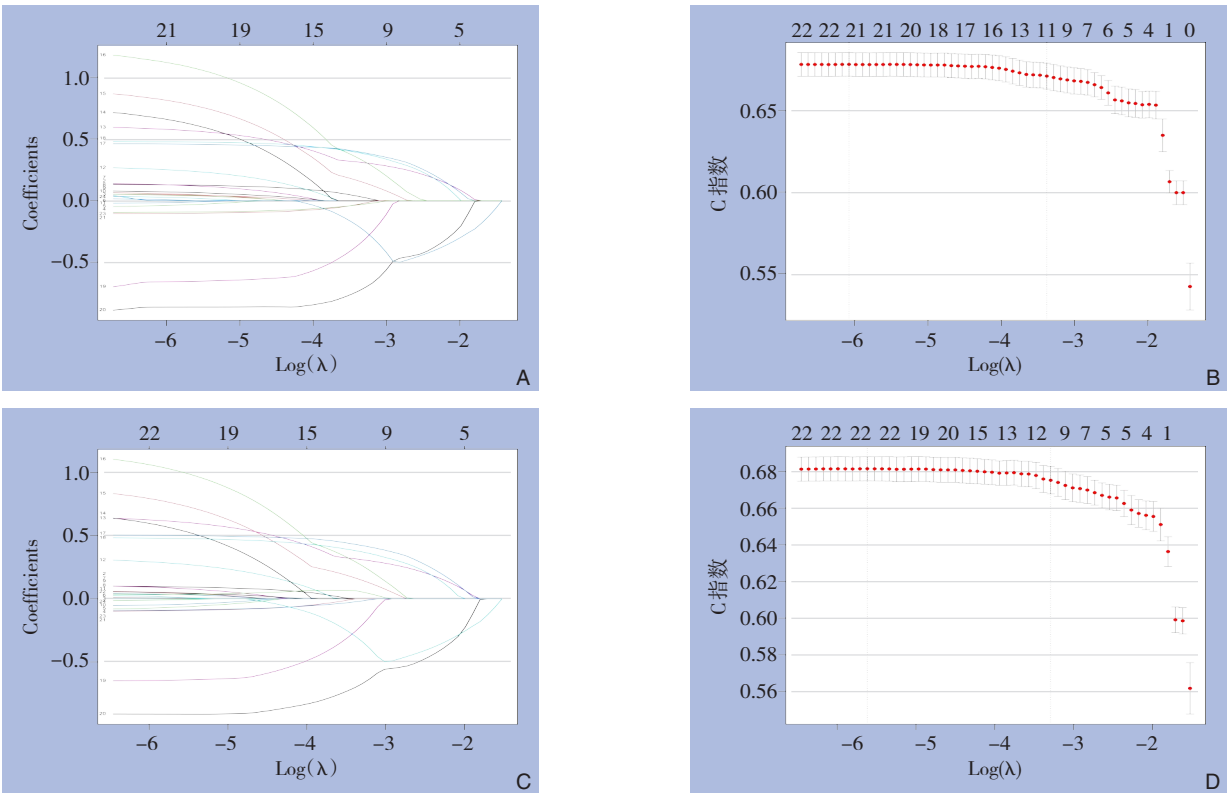


图3 基于LASSO回归的特征选择 A: LASSO回归系数随Log( $\lambda$ )的变化曲线(OS); B: 基于10折交叉验证C指数随Log( $\lambda$ )的变化曲线(OS); C: LASSO回归系数随Log( $\lambda$ )的变化曲线(CSS); D: 基于10折交叉验证C指数随Log( $\lambda$ )的变化曲线(CSS)

Figure 3 Feature selection based on LASSO regression A: LASSO regression coefficients vs. Log( $\lambda$ ) curve (OS); B: C-index from 10-fold cross-validation vs. Log( $\lambda$ ) curve (OS); C: LASSO regression coefficients vs. Log( $\lambda$ ) curve (CSS); D: C-index from 10-fold cross-validation vs. Log( $\lambda$ ) curve (CSS)

**2.2.4 随机生存森林模型** 随机生存森林具有高鲁棒性的特点，不受比例风险假定、对数线性假定的约束，在分析过程中，能够通过两个随机采样的过程来防止算法的过度拟合问题。本研究进一步通过该方法来筛选变量，结果显示，影响 OS 的

重要性评分前五位变量分别为：系统治疗、分化程度、N 分期、化疗和 T 分期（图 4A）；而影响 CSS 的重要性评分前五位变量分别为：系统治疗、分化程度、N 分期、化疗和 AJCC 分期（图 4B）。

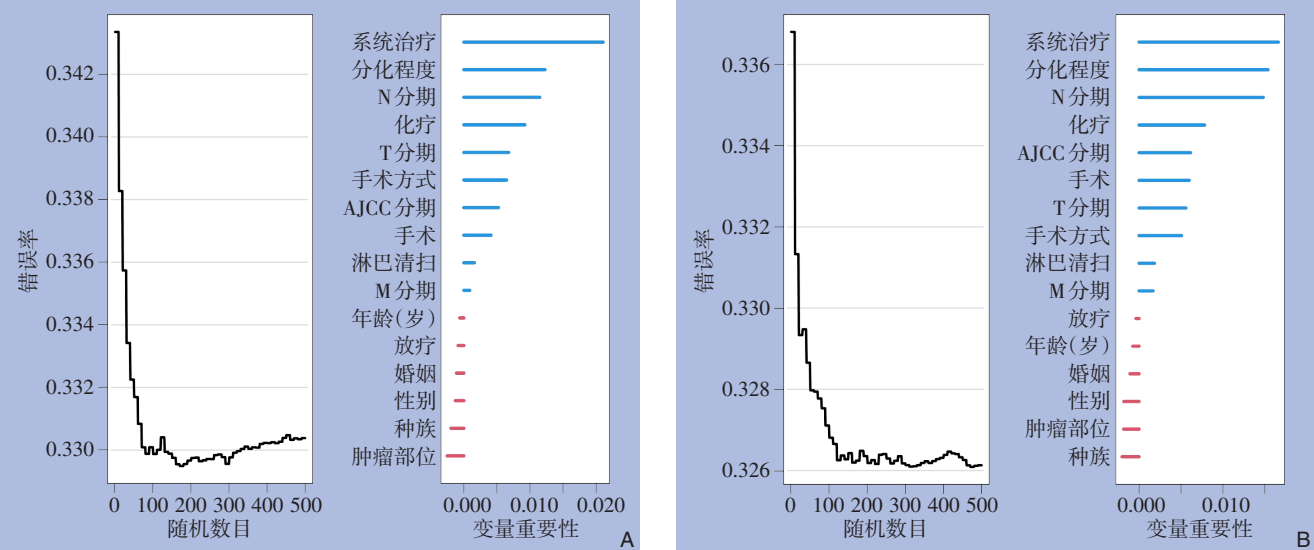


图4 变量重要性随机森林模型 A: OS; B: CSS  
Figure 4 Variable importance from the random forest model A: OS; B: CS

2.3 PDAC 患者生存预测模型的构建

基于多因素 Cox 回归模型、LASSO 回归模型和随机生存森林模型的分析结果并结合临床重要性，最终选择：年龄、T 分期、N 分期、M 分期、分化程度、手术和化疗，共七个临床特征参与构建列

线。预测 PDAC 患者在 6、12、36 个月的 OS（图 5A）和 CSS（图 5B）。图中每个临床特征都将被赋予相应的分值，所有特征的分值之和等于总分值，分值越低，预后越好。根据总分值即可对不同时间点 PDAC 患者的 OS 和 CSS 做出预测。

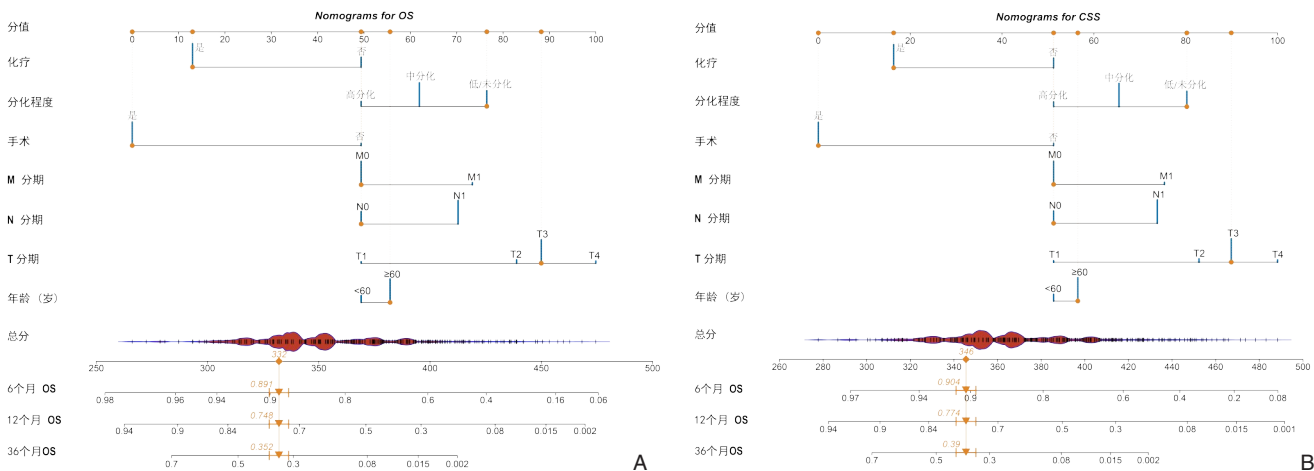


图5 预测PDAC患者6、12、36个月预后的列线图 A: OS; B: CSS  
Figure 5 Nomogram for predicting the prognosis of PDAC patients at 6, 12, and 36 months A: OS; B: CSS



## 2.4 PDAC患者生存预测模型的验证

通过C指数和时间依赖ROC曲线对模型的区分度进行评价。结果显示,对于OS,在训练集和验证集中的C指数分别为0.692(95% CI=0.681~0.704)和0.680(95% CI=0.664~0.698);对于CSS,在训练集和验证集中的C指数分别为0.696(95% CI=0.684~0.707)和0.680(95% CI=0.662~0.698)。ROC曲线表明模型具有良好的预测价

值(图6)。为了评价模型的准确度,采用Bootstrap法对模型进行内部验证和外部验证,自抽样次数B=1 000,并绘制校准曲线。验证结果显示,在训练集与验证集中,PDAC患者的在6、12、36个月的OS和CSS校正曲线均靠近理想的45°参考线,表明模型预测值与实际值之间具有良好的一致性(图7)。

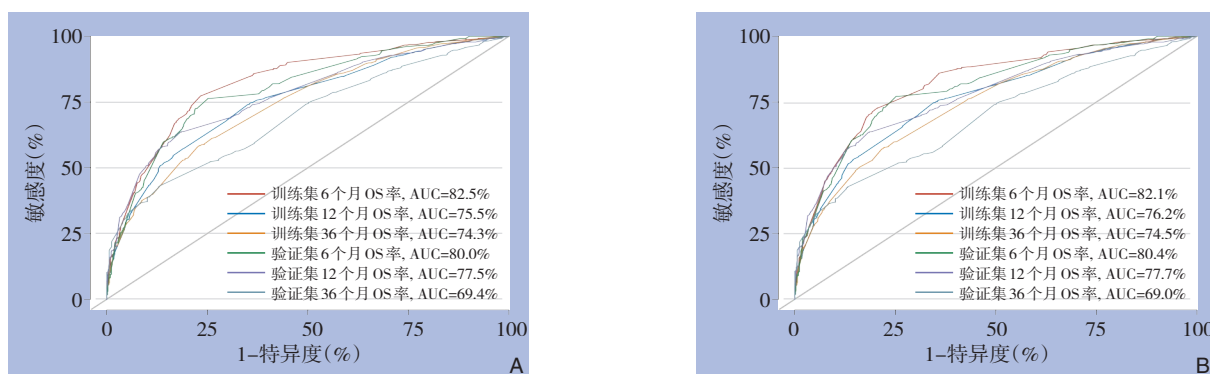


图6 模型在训练集和验证集中6、12、36个月预测能力验证的ROC曲线 A: OS; B: CSS

Figure 6 ROC curves validating the predictive ability of the model at 6, 12, and 36 months in the training and validation sets A: OS; B: CSS

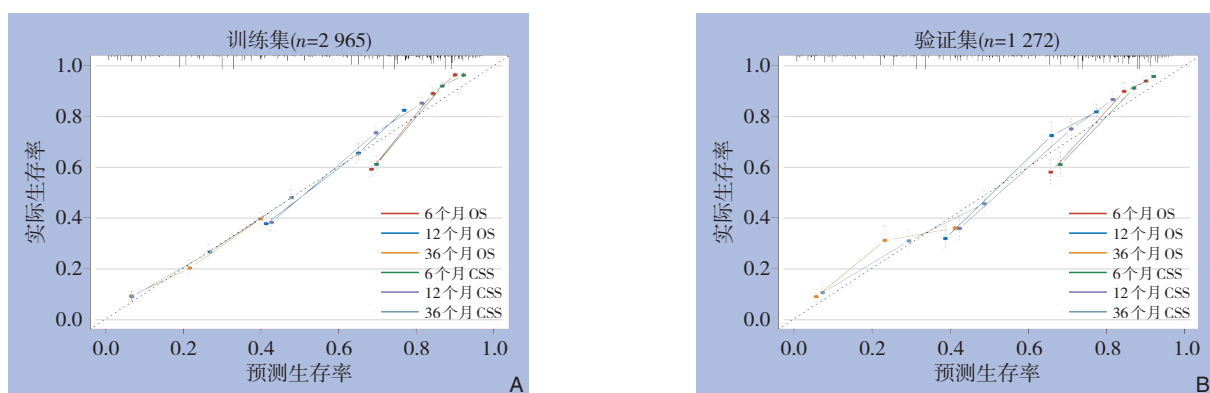


图7 PDAC患者6、12、36个月OS与CSS的校准曲线 A: 训练集; B: 验证集

Figure 7 Calibration curves for OS and CSS of PDAC patients at 6, 12, and 36 months A: Training set; B: Validation set

## 2.5 模型的决策曲线分析(decision curve analysis, DCA)

DCA是一种新的预测模型评价方法,它更注重临床获益评价。本研究通过DCA评估其在临床

应用中该模型是否有获益。结果显示,与TNM分期相比,列线图模型在预测6、12、36个月的OS和CSS均显示出更大的净收益(图8)。

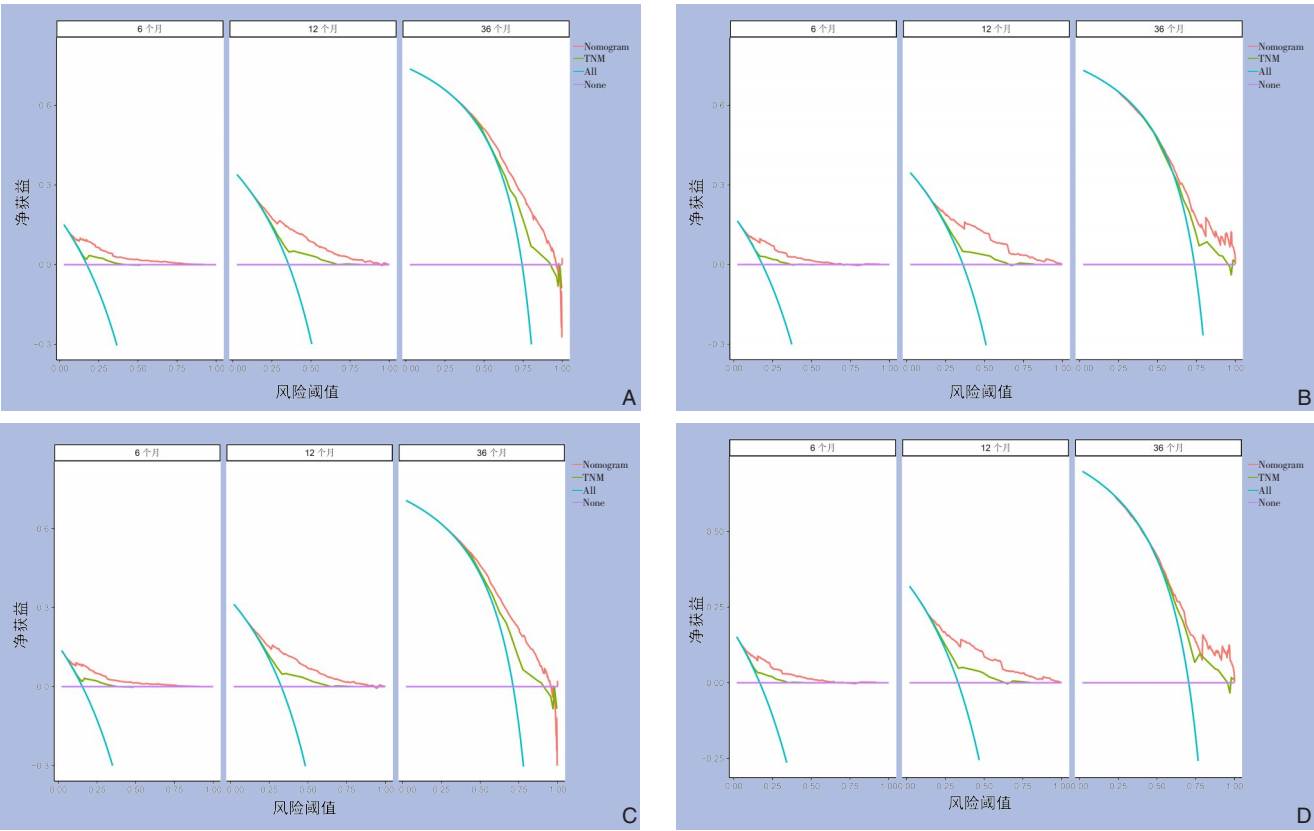


图8 训练集和验证集中列线图和TNM分期预测6、12、36个月OS和CSS的DCA比较 A: 训练集OS; B: 验证集OS; C: 训练集CSS; D: 验证集CSS

Figure 8 Comparison of DCA for the nomogram and TNM staging predicting OS and CSS at 6, 12, and 36 months in the training and validation sets A: OS for the training set; B: OS for the validation set; C: CSS for the training set; D: CSS for the validation set

2.6 不同风险组PDAC患者的生存获益

为了进一步评价模型在临床中应用价值，在训练集和验证集中分别计算出每例患者的总分值，将训练集中总分值的中位数设置为危险分层阈值。

对于OS和CSS，低风险组和高风险组的阈值分别为155分和158分。结果显示，在验证集和训练集中该模型均能对PDAC患者的生存预后做出良好的区分（均 $P<0.0001$ ）（图9）。

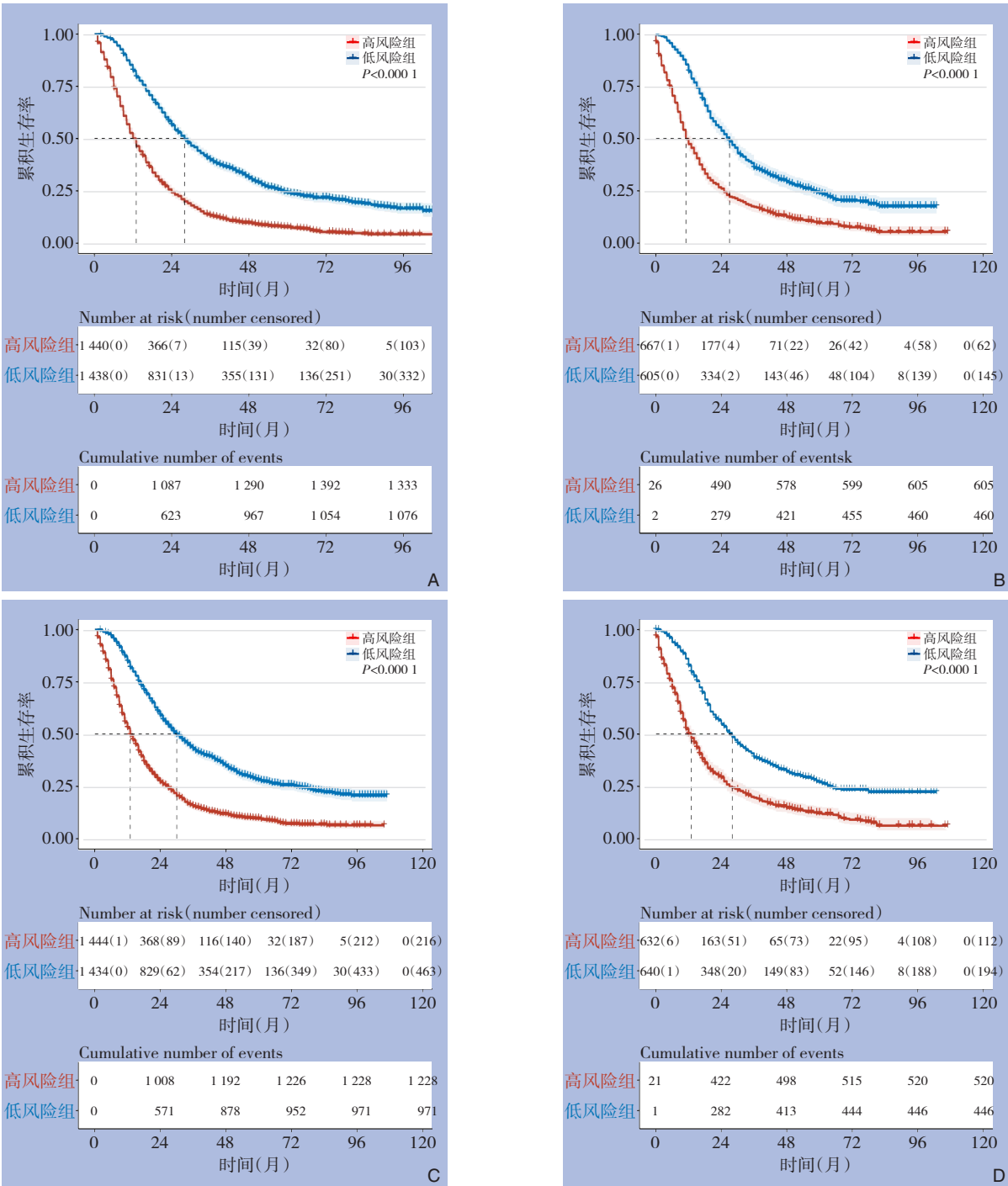


图9 不同风险患者的生存曲线 A: 训练集OS; B: 验证集OS; C: 训练集CSS; D: 验证集CSS

Figure 9 Survival curves for patients with different risk levels A: OS for the training set; B: OS for the validation set; C: CSS for the training set; D: CSS for the validation set

3 讨论

PDAC是最常见的胰腺恶性肿瘤，占胰腺肿瘤性疾病的90%以上<sup>[26]</sup>，占胰腺恶性肿瘤的85%<sup>[27]</sup>。据推测，到2030年PDAC可能将成为第二位肿瘤相关性致死原因<sup>[26,28]</sup>。同时，由于PDAC早期缺乏特异性临床症状，初次确诊时，患者常合并局部进

展或远处转移，因而约有80%~85%的患者初诊时已失去手术机会<sup>[29]</sup>。另一方面，即使接受根治性切除手术治疗的PDAC患者，仍有近80%的患者出现术后早期的复发<sup>[30]</sup>。因此，其远期预后仍然较差<sup>[1,31-32]</sup>，如何提高PDAC患者的OS和CSS，是当前研究的热点问题。列线图被广泛应用于肿瘤预后的评价，它可以个体化预测患者特定时间点的累

积生存概率，辅助临床医师做出针对性的临床决策<sup>[10-13,33]</sup>。

Deng等<sup>[34]</sup>研究表明，年龄（>60岁）、肿瘤直径大是PDAC患者1年OS的独立危险因素。此外，一系列研究均显示手术<sup>[2,35-37]</sup>、化疗<sup>[35,38]</sup>、放疗<sup>[37]</sup>、TNM分期<sup>[39]</sup>、分化程度<sup>[39]</sup>等均是PDAC患者预后的独立影响因素。本研究中，分别运用了单因素、多因素Cox回归、随机森林模型和LASSO回归的方法，并结合临床重要性<sup>[24]</sup>。对影响PDAC患者OS和CSS的变量进行了筛选。同上述研究一致，我们也最终确定了以下变量：年龄、TNM分期、分化程度、手术和化疗用于构建PDAC患者6、12、36个月OS和CSS的预测模型。该模型经过内部验证和外部验证结果均表现出良好的区分度和准确度。相比于单一TNM分期，该模型的临床净获益更大。且不同于TNM分期，列线图可以直观预测患者的生存概率<sup>[23,40]</sup>，帮助临床医师对患者的预后做出更准确的判断。在此基础上，还可进一步根据模型得分，对患者进行危险分层，从而区分患者是否属于高危人群，帮助临床医师做出最佳的临床决策。例如：1例70岁，低分化，T3N0M0期患者，该患者接受手术治疗，未接受化疗，OS总分为332分，6、12、36个月的累积生存率分别为89.1%、74.8%、35.2%；CSS总分为346分，6、12、36个月的累积生存率分别为90.4%、77.4%、39.0%。由于影响患者的OS的因素相对复杂，而相比之下CSS显得更加重要<sup>[10]</sup>。根据CSS对该患者进行危险分层。由于其总分346分高于阈值158分，因此该患者被认定为高危人群，需要更加密切观察患者，制定更为严格的治疗和随访计划。

据笔者所知，目前国内外专门针对PDAC的大样本真实世界研究较少。而本研究主要有以下优势：首先，在分析过程中，本研究使用了随机生存森林模型。该模型作为一种机器学习算法，属于集成学习范畴，具有高鲁棒性，且不受比例风险假定、对数线性假定等条件的限制，而且还能够通过两个随机采样的过程来防止算法的过度拟合问题<sup>[25]</sup>。因而，该算法得出的结论将更加可靠和稳定。其次，在分析过程中，本研究也同样独立地运用了LASSO、经典的单因素、多因素Cox比例风险模型，并且利用VIFs对模型的多重共线性问题进行了定量的评价与分析。最后，本研究总共纳入了4 237例患者，3 561例（84.05%）患者发

生死亡结局事件，其中3 234例（76.33%）患者出现肿瘤特异性死亡。因此，本研究的样本含量和事件结局数，可以为统计分析提供足够大的检验效能。

本研究中仍然存在一些局限性：首先本研究是回顾性研究，存在一定的选择偏倚，即使在本研究中我们制定了严格的纳入和排除标准，但这些仍然不能完全消除选择偏倚的影响。其次，由于SEER数据库获取权限的限制，一些重要的临床特征无法获得，比如：CA19-9的检测情况、肿瘤复发情况，手术、放疗、化疗的具体方案等。最后，SEER数据库是大型肿瘤数据库，容易受编码错误及删失值的影响<sup>[14]</sup>。

综上所述，本研究通过大样本量临床数据，构建的PDAC患者预后模型，有较高的区分度和准确度。在临床实践过程中，它可以对每例PDAC患者不同时间节点的生存率进行个体化的预测，并直观地展现出来。更重要的是，它可以帮助临床医师为PDAC患者制定精准的、个体化的治疗和随访方案。

利益冲突：所有作者均声明不存在利益冲突。

作者贡献声明：黄坤、张业光共同设计研究；黄正红和赵攀完成数据提取、整理和分析；章慧制作表格和图片；张业光撰写初稿；黄坤校正图表并最后完成全文校正。

## 参考文献

- [1] Timmer FEF, Geboers B, Nieuwenhuizen S, et al. Locoregional treatment of metastatic pancreatic cancer utilizing resection, ablation and embolization: a systematic review[J]. *Cancers (Basel)*, 2021, 13(7):1608. doi:10.3390/cancers13071608.
- [2] Zhou W, Wang D, Lou W. Current role of surgery in pancreatic cancer with synchronous liver metastasis[J]. *Cancer Control*, 2020, 27(1):1073274820976593. doi:10.1177/1073274820976593.
- [3] Siegel R, Miller K, Jemal A. Cancer statistics, 2019[J]. *CA Cancer J Clin*, 2019, 69(1):7-34. doi: 10.3322/caac.21551.
- [4] 黄坤, 何运胜, 李建波, 等. 胰腺癌肝转移核心基因的筛选与验证[J]. *中国普通外科杂志*, 2023, 32(3):390-399. doi: 10.7659/j.issn.1005-6947.2023.03.008.
- [5] Huang K, He YS, Li JB, et al. Screening and identification of hub gene involved in hepatic metastasis of carcinoma of pancreas[J]. *China Journal of General Surgery*, 2023, 32(3): 390-399. doi: 10.7659/j.issn.1005-6947.2023.03.008.
- [5] Serafini S, Sperti C, Friziero A, et al. Systematic Review and Meta-Analysis of Surgical Treatment for Isolated Local Recurrence of

- Pancreatic Cancer[J]. *Cancers*, 2021, 13(6): 1277. doi: [10.3390/cancers13061277](https://doi.org/10.3390/cancers13061277).
- [6] Maire F, Cibot JO, Compagne C, et al. Epidemiology of pancreatic cancer in France: descriptive study from the French national hospital database[J]. *Eur J Gastroenterol Hepatol*, 2017, 29(8):904–908. doi:[10.1097/MEG.0000000000000901](https://doi.org/10.1097/MEG.0000000000000901).
- [7] van der Geest LGM, Lemmens VEPP, de Hingh IHJT, et al. Nationwide outcomes in patients undergoing surgical exploration without resection for pancreatic cancer[J]. *Br J Surg*, 2017, 104(11): 1568–1577. doi:[10.1002/bjs.10602](https://doi.org/10.1002/bjs.10602).
- [8] Wei M, Shi S, Hua J, et al. Simultaneous resection of the primary tumour and liver metastases after conversion chemotherapy versus standard therapy in pancreatic cancer with liver oligometastasis: protocol of a multicentre, prospective, randomised phase III control trial (CSPAC-1)[J]. *BMJ Open*, 2019, 9(12):e033452. doi:[10.1136/bmjopen-2019-033452](https://doi.org/10.1136/bmjopen-2019-033452).
- [9] 张红英, 张超, 陈茂山, 等. 基于SEER数据库对初诊IV期乳腺癌患者的转移部位及其预后分析[J]. *中国普外基础与临床杂志*, 2023, 30(9):1079–1086. doi:[10.7507/1007-9424.202306009](https://doi.org/10.7507/1007-9424.202306009).
- Zhang HY, Zhang C, Chen MS, et al. Analysis of metastatic sites and prognosis of newly diagnosed stage IV breast cancer patients based on SEER database[J]. *Chinese Journal of Bases and Clinics in General Surgery*, 2023, 30(9): 1079–1086. doi: [10.7507/1007-9424.202306009](https://doi.org/10.7507/1007-9424.202306009).
- [10] Xu L, Wen N, Qiu JJ, et al. Predicting survival benefit of sparing sentinel lymph node biopsy in low-risk elderly patients with early breast cancer: a population-based analysis[J]. *Front Oncol*, 2020, 10:1718. doi:[10.3389/fonc.2020.01718](https://doi.org/10.3389/fonc.2020.01718).
- [11] Xiang JX, Zhang XF, Weiss M, et al. Multi-institutional development and external validation of a nomogram predicting recurrence after curative liver resection for neuroendocrine liver metastasis[J]. *Ann Surg Oncol*, 2020, 27(10): 3717–3726. doi: [10.1245/s10434-020-08620-5](https://doi.org/10.1245/s10434-020-08620-5).
- [12] 程晨, 吴云桦, 徐正水, 等. II–III期结肠癌根治术后复发危险因素分析及其列线图预测模型的应用价值[J]. *中华消化外科杂志*, 2021, 20(3): 331–338. doi: [10.3760/cma.j.cn115610-20210120-00034](https://doi.org/10.3760/cma.j.cn115610-20210120-00034).
- Cheng C, Wu YH, Xu ZS, et al. Risk factors for tumor recurrence after radical resection of stage II–III colon cancer and application value of its nomogram prediction model[J]. *Chinese Journal of Digestive Surgery*, 2021, 20(3): 331–338. doi: [10.3760/cma.j.cn115610-20210120-00034](https://doi.org/10.3760/cma.j.cn115610-20210120-00034).
- [13] 邓家仲, 英卫东. 肝细胞癌微血管侵犯危险因素分析及术前预测列线图模型构建[J]. *中国普通外科杂志*, 2021, 30(7):772–779. doi:[10.7659/j.issn.1005-6947.2021.07.003](https://doi.org/10.7659/j.issn.1005-6947.2021.07.003).
- Deng JZ, Jia WD. Analysis of risk factors for microvascular invasion in hepatocellular carcinoma and construction of preoperative predictive nomogram[J]. *China Journal of General Surgery*, 2021, 30(7): 772–779. doi: [10.7659/j.issn.1005-6947.2021.07.003](https://doi.org/10.7659/j.issn.1005-6947.2021.07.003).
- [14] 黄坤, 赵平武, 白斗, 等. 不同病理学类型的腹膜后脂肪肉瘤患者的预后分析及临床评价[J]. *中国普通外科杂志*, 2021, 30(10): 1212–1228. doi:[10.7659/j.issn.1005-6947.2021.10.011](https://doi.org/10.7659/j.issn.1005-6947.2021.10.011).
- Huang K, Zhao PW, Bai D, et al. Prognostic analysis and clinical evaluation of patients with different pathological types of retroperitoneal liposarcoma[J]. *China Journal of General Surgery*, 2021, 30(10): 1212–1228. doi: [10.7659/j.issn.1005-6947.2021.10.011](https://doi.org/10.7659/j.issn.1005-6947.2021.10.011).
- [15] 黄坤, 黄正红, 赵攀, 等. 胆囊鳞状细胞癌预后模型的构建与验证[J]. *中国普通外科杂志*, 2023, 32(8):1187–1198. doi:[10.7659/j.issn.1005-6947.2023.08.007](https://doi.org/10.7659/j.issn.1005-6947.2023.08.007).
- Huang K, Huang ZH, Zhao P, et al. Construction and validation of the prognosis model for gallbladder squamous cell carcinoma[J]. *China Journal of General Surgery*, 2023, 32(8): 1187–1198. doi: [10.7659/j.issn.1005-6947.2023.08.007](https://doi.org/10.7659/j.issn.1005-6947.2023.08.007).
- [16] 黄坤, 何运胜, 张红梅, 等. 手术治疗对胆囊小细胞癌患者预后价值的SEER数据库分析[J]. *中国普通外科杂志*, 2022, 31(8):997–1005. doi:[10.7659/j.issn.1005-6947.2022.08.002](https://doi.org/10.7659/j.issn.1005-6947.2022.08.002).
- Huang K, He YS, Zhang HM, et al. Analysis of prognostic value of surgical treatment in patients with small cell carcinoma of gallbladder based on SEER database[J]. *China Journal of General Surgery*, 2022, 31(8): 997–1005. doi: [10.7659/j.issn.1005-6947.2022.08.002](https://doi.org/10.7659/j.issn.1005-6947.2022.08.002).
- [17] 黄坤, 何运胜, 赵攀, 等. 手术治疗对直肠小细胞神经内分泌癌患者预后影响的SEER数据库分析[J]. *中国普外基础与临床杂志*, 2023, 30(2):179–184. doi:[10.7507/1007-9424.202208041](https://doi.org/10.7507/1007-9424.202208041).
- Huang K, He YS, Zhao P, et al. Effect of surgical treatment on prognosis in patients with rectal small cell neuroendocrine carcinoma based on SEER database[J]. *Chinese Journal of Bases and Clinics in General Surgery*, 2023, 30(2):179–184. doi:[10.7507/1007-9424.202208041](https://doi.org/10.7507/1007-9424.202208041).
- [18] 黄坤, 赵攀, 赵平武, 等. 手术治疗对胆囊鳞状细胞癌预后的影响[J]. *中国普外基础与临床杂志*, 2023, 30(8):964–969. doi: [10.7507/1007-9424.202304039](https://doi.org/10.7507/1007-9424.202304039).
- Huang K, Zhao P, Zhao PW, et al. Effect of surgical treatment on prognosis in patients with gallbladder squamous cell carcinoma[J]. *Chinese Journal of Bases and Clinics in General Surgery*, 2023, 30(8):964–969. doi:[10.7507/1007-9424.202304039](https://doi.org/10.7507/1007-9424.202304039).
- [19] 程杨, 陈小彬, 魏志鸿, 等. 术前淋巴细胞与C反应蛋白比值在预测胆囊癌术后复发中的应用及其临床价值[J]. *中国普通外科杂志*, 2023, 32(2): 190–199. doi: [10.7659/j.issn.1005-6947.2023.02.004](https://doi.org/10.7659/j.issn.1005-6947.2023.02.004).
- Cheng Y, Chen XB, Wei ZH, et al. Application of preoperative lymphocyte to C-reactive protein ratio in predicting postoperative recurrence of gallbladder cancer and its clinical value[J]. *China Journal of General Surgery*, 2023, 32(2): 190–199. doi: [10.7659/j.issn.1005-6947.2023.02.004](https://doi.org/10.7659/j.issn.1005-6947.2023.02.004).
- [20] 王晓晓, 陶立元, 裴敏玥, 等. 连续变量转换为分类变量的几种方法[J]. *中华儿科杂志*, 2022, 60(5): 420. doi: [10.3760/cma.j.cn112140-20220317-00212](https://doi.org/10.3760/cma.j.cn112140-20220317-00212).
- Wang XX, Tao LY, Pei MY, et al. Several methods of converting continuous variables into classified variables[J]. *Chinese Journal of*



- Pediatrics, 2022, 60(5): 420. doi: [10.3760/cma.j.cn112140-20220317-00212](https://doi.org/10.3760/cma.j.cn112140-20220317-00212).
- [21] Su BB, Bai DS, Yu JQ, et al. Can patients with pancreatic cancer and liver metastases obtain survival benefit from surgery? A population-based study[J]. J Cancer, 2021, 12(2): 539–552. doi: [10.7150/jca.51218](https://doi.org/10.7150/jca.51218).
- [22] Pu N, Chen Q, Gan W, et al. Lymph node metastatic patterns and survival predictors based on tumor size in pancreatic ductal adenocarcinoma[J]. Adv Ther, 2021, 38(8): 4258–4270. doi: [10.1007/s12325-021-01819-2](https://doi.org/10.1007/s12325-021-01819-2).
- [23] Chen MS, Liu PC, Yi JZ, et al. Development and validation of nomograms for predicting survival in patients with de novo metastatic triple-negative breast cancer[J]. Sci Rep, 2022, 12(1): 14659. doi: [10.1038/s41598-022-18727-2](https://doi.org/10.1038/s41598-022-18727-2).
- [24] Kim Y, Margonis GA, Prescott JD, et al. Nomograms to predict recurrence-free and overall survival after curative resection of adrenocortical carcinoma[J]. JAMA Surg, 2016, 151(4): 365–373. doi: [10.1001/jamasurg.2015.4516](https://doi.org/10.1001/jamasurg.2015.4516).
- [25] 陈哲, 许恒敏, 李哲轩, 等. 随机生存森林: 基于机器学习算法的生存分析模型[J]. 中华预防医学杂志, 2021, 55(1): 104–109. doi: [10.3760/cma.j.cn112150-20200911-01197](https://doi.org/10.3760/cma.j.cn112150-20200911-01197).
- Chen Z, Xu HM, Li ZX, et al. Random survival forest: applying machine learning algorithm in survival analysis of biomedical data[J]. Chinese Journal of Preventive Medicine, 2021, 55(1): 104–109. doi: [10.3760/cma.j.cn112150-20200911-01197](https://doi.org/10.3760/cma.j.cn112150-20200911-01197).
- [26] Huang XY, Qin WT, Su QS, et al. A new stemness-related prognostic model for predicting the prognosis in pancreatic ductal adenocarcinoma[J]. Biomed Res Int, 2021, 2021: 6669570. doi: [10.1155/2021/6669570](https://doi.org/10.1155/2021/6669570).
- [27] Cen C, Liu L, Li X, et al. Pancreatic ductal adenocarcinoma at CT: a combined nomogram model to preoperatively predict cancer stage and survival outcome[J]. Front Oncol, 2021, 11: 594510. doi: [10.3389/fonc.2021.594510](https://doi.org/10.3389/fonc.2021.594510).
- [28] Rahib L, Smith BD, Aizenberg R, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States[J]. Cancer Res, 2014, 74(11): 2913–2921. doi: [10.1158/0008-5472.CAN-14-0155](https://doi.org/10.1158/0008-5472.CAN-14-0155).
- [29] McGuigan A, Kelly P, Turkington RC, et al. Pancreatic cancer: a review of clinical diagnosis, epidemiology, treatment and outcomes[J]. World J Gastroenterol, 2018, 24(43): 4846–4861. doi: [10.3748/wjg.v24.i43.4846](https://doi.org/10.3748/wjg.v24.i43.4846).
- [30] Groot VP, Rezaee N, Wu WC, et al. Patterns, timing, and predictors of recurrence following pancreatectomy for pancreatic ductal adenocarcinoma[J]. Ann Surg, 2018, 267(5): 936–945. doi: [10.1097/SLA.0000000000002234](https://doi.org/10.1097/SLA.0000000000002234).
- [31] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2018, 68(6): 394–424. doi: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492).
- [32] He C, Sun S, Zhang Y, et al. A novel nomogram to predict survival in patients with recurrence of pancreatic ductal adenocarcinoma after radical resection[J]. Front Oncol, 2020, 10: 1564. doi: [10.3389/fonc.2020.01564](https://doi.org/10.3389/fonc.2020.01564).
- [33] 李吴寒, 张营, 潘晶晶, 等. 胃癌患者预后相关影响因素的列线图模型构建及验证[J]. 中国普通外科杂志, 2022, 31(10): 1381–1388. doi: [10.7659/j.issn.1005-6947.2022.10.014](https://doi.org/10.7659/j.issn.1005-6947.2022.10.014).
- Li WH, Zhang Y, Pan JJ, et al. Construction and validation of a nomogram for prognostic value of NLR and PLR in patients with gastric cancer[J]. China Journal of General Surgery, 2022, 31(10): 1381–1388. doi: [10.7659/j.issn.1005-6947.2022.10.014](https://doi.org/10.7659/j.issn.1005-6947.2022.10.014).
- [34] Deng QL, Dong S, Wang L, et al. Development and validation of a nomogram for predicting survival in patients with advanced pancreatic ductal adenocarcinoma[J]. Sci Rep, 2017, 7(1): 11524. doi: [10.1038/s41598-017-11227-8](https://doi.org/10.1038/s41598-017-11227-8).
- [35] Safi SA, Fluegen G, Rehders A, et al. Surgical margin clearance and extended chemotherapy defines survival for synchronous oligometastatic liver lesions of the ductal adenocarcinoma of the pancreas[J]. Int J Clin Oncol, 2021, 26(10): 1911–1921. doi: [10.1007/s10147-021-01961-5](https://doi.org/10.1007/s10147-021-01961-5).
- [36] De Dosso S, Siebenhüner AR, Winder T, et al. Treatment landscape of metastatic pancreatic cancer[J]. Cancer Treat Rev, 2021, 96: 102180. doi: [10.1016/j.ctrv.2021.102180](https://doi.org/10.1016/j.ctrv.2021.102180).
- [37] Fu N, Jiang Y, Weng Y, et al. Worth it or not? Primary tumor resection for stage IV pancreatic cancer patients: a SEER-based analysis of 15, 836 cases[J]. Cancer Med, 2021, 10(17): 5948–5963. doi: [10.1002/cam4.4147](https://doi.org/10.1002/cam4.4147).
- [38] Napoli N, Kauffmann E, Cacace C, et al. Factors predicting survival in patients with locally advanced pancreatic cancer undergoing pancreatectomy with arterial resection[J]. Updates Surg, 2021, 73(1): 233–249. doi: [10.1007/s13304-020-00883-7](https://doi.org/10.1007/s13304-020-00883-7).
- [39] Li G, Liao CY, Chen JZ, et al. Construction and validation of novel nomograms for predicting prognosis of pancreatic ductal adenocarcinoma after surgery according to different primary cancer locations[J]. Front Oncol, 2021, 11: 646082. doi: [10.3389/fonc.2021.646082](https://doi.org/10.3389/fonc.2021.646082).
- [40] Peng F, Qin T, Wang M, et al. Development and validation of a nomogram to predict survival in pancreatic head ductal adenocarcinoma after pancreaticoduodenectomy[J]. Front Oncol, 2021, 11: 734673. doi: [10.3389/fonc.2021.734673](https://doi.org/10.3389/fonc.2021.734673).

( 本文编辑 姜晖 )

本文引用格式: 张业光, 赵攀, 章慧, 等. 基于机器学习算法的胰腺导管腺癌预后模型构建及其验证[J]. 中国普通外科杂志, 2024, 33(9): 1459–1472. doi: [10.7659/j.issn.1005-6947.2024.09.013](https://doi.org/10.7659/j.issn.1005-6947.2024.09.013)

Cite this article as: Zhang YG, Zhao P, Zhang H, et al. Construction and validation of a prognostic model for pancreatic ductal adenocarcinoma based on machine learning algorithm[J]. Chin J Gen Surg, 2024, 33(9): 1459–1472. doi: [10.7659/j.issn.1005-6947.2024.09.013](https://doi.org/10.7659/j.issn.1005-6947.2024.09.013)