

基于常规血液标志物的早期胰腺癌诊断模型

李晓云^{1*}, 齐贞光², 荆雪^{1#}

¹青岛大学附属医院消化内科, 山东 青岛

²青岛市第八人民医院消化内科, 山东 青岛

收稿日期: 2025年3月27日; 录用日期: 2025年4月22日; 发布日期: 2025年4月28日

摘要

目的: 胰腺导管腺癌(PDAC)早期诊断困难, 现有诊断方法耗时且侵入性强。本研究旨在构建一种基于常规血液标志物的非侵入性诊断模型, 以提高PDAC早期诊断效率。方法: 本研究为回顾性病例对照研究, 纳入2018年6月至2024年3月青岛大学附属医院677例胰胆疾病患者(PDAC组210例, 对照组467例)。采用倾向性评分匹配(PSM)平衡基线特征, 最终纳入410例(每组205例)。暴露变量为术前血液标志物(CA19-9、CEA、DBIL、CYS-C、COUNT评分), 结局变量为PDAC诊断, 协变量包括年龄、性别、身高、体重、肝肾功能及炎症营养指标等。结果: 最终模型纳入CA19-9、CEA、DBIL、CYS-C和COUNT评分五个变量, ROC曲线下面积(AUC)为0.86 (95%CI: 0.82~0.90), 准确率为77% (95%CI: 0.72~0.82)。多因素分析显示, CEA (OR = 1.16, 95%CI: 1.02~1.32)和DBIL (OR = 1.02, 95%CI: 1.01~1.02)为独立危险因素, CYS-C (OR = 0.06, 95%CI: 0.01~0.36)和COUNT评分 = 2 (OR = 0.21, 95%CI: 0.08~0.56)为保护因素。结论: 本研究构建的五维血液标志物模型可有效鉴别PDAC, 具有较高的诊断准确性和临床应用潜力, 有望成为PDAC早期诊断的非侵入性工具。

关键词

胰腺导管腺癌, 诊断/预测模型, 倾向性评分匹配

Diagnostic Model for Early Pancreatic Cancer Based on Routine Blood Biomarkers

Xiaoyun Li^{1*}, Zhenguang Qi², Xue Jing^{1#}

¹Department of Gastroenterology, The Affiliated Hospital of Qingdao University, Qingdao Shandong

²Department of Gastroenterology, Qingdao Eighth People's Hospital, Qingdao Shandong

Received: Mar. 27th, 2025; accepted: Apr. 22nd, 2025; published: Apr. 28th, 2025

*第一作者。

#通讯作者。

文章引用: 李晓云, 齐贞光, 荆雪. 基于常规血液标志物的早期胰腺癌诊断模型[J]. 临床医学进展, 2025, 15(4): 3181-3192. DOI: 10.12677/acm.2025.1541285

Abstract

Objective: Pancreatic ductal adenocarcinoma (PDAC) is challenging to diagnose early, and current methods are invasive and time-consuming. This study aimed to develop a non-invasive diagnostic model based on routine blood biomarkers to improve early PDAC detection. **Methods:** A retrospective case-control study was conducted at Qingdao University Hospital from June 2018 to March 2024, including 677 patients with pancreatobiliary diseases (210 PDAC, 467 controls). Propensity score matching (PSM) balanced baseline characteristics, resulting in 410 matched cases (205 per group). Exposure variables were preoperative blood biomarkers (CA19-9, CEA, DBIL, CYS-C, COUNT score), outcome variable was PDAC diagnosis, and covariates included age, sex, height, weight, liver/kidney function, inflammatory indicators and nutritional assessment indices. **Results:** The final model incorporated five variables: CA19-9, CEA, DBIL, CYS-C, and COUNT score, achieving an area under ROC curve (AUC) of 0.86 (95%CI: 0.82~0.90) and accuracy of 77% (95%CI: 0.72~0.82). Multivariate analysis identified CEA (OR = 1.16, 95%CI: 1.02~1.32) and DBIL (OR = 1.02, 95%CI: 1.01~1.02) as independent risk factors, while CYS-C (OR = 0.06, 95%CI: 0.01~0.36) and COUNT score = 2 (OR = 0.21, 95%CI: 0.08~0.56) were protective factors. **Conclusion:** The five-dimensional blood biomarker model constructed in this study effectively distinguishes PDAC with high diagnostic accuracy and clinical application potential, showing promise as a non-invasive tool for early PDAC diagnosis.

Keywords

Pancreatic Ductal Adenocarcinoma, Diagnostic/Prediction Model, Propensity Score Matching

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 背景

胰胆系统疾病在全球发病率逐年攀升,根据《Lancet Oncol》2024 年报告,全球胆道疾病年发病率达 58.3/10 万,其中中国高达 76.5/10 万[1]。胰腺导管腺癌(PDAC)作为胰胆系统恶性肿瘤代表,2022 年全球新增约 49.6 万例,中国新增 10.4 万例,5 年存活率仅为 11.5% [2]。尤其值得关注的是,胆道阻塞性疾病患者中约 15.3%最终确诊为 PDAC,而良性胰胆疾病与 PDAC 的临床表现高度重叠,导致 26.7%的患者初诊结果存在误差[3]。

本研究构建的五维预测模型整合了关键生物标志物:CA19-9 是一种唾液粘蛋白相关 lewis 血型抗原,在胰胆系统恶性肿瘤中高表达(敏感度 78.2%,特异性 82.5%) [4];CEA 对肝转移 PDAC 具有特异性预测价值[5];DBIL 是胆红素代谢中间产物,在胰头部肿瘤引起的梗阻性黄疸中升高显著[6];CONUT 评分基于血清白蛋白、总淋巴细胞计数和总胆固醇,客观评估 PDAC 患者的营养状态,研究显示营养不良与 PDAC 患者预后不良显著相关[7];CYS-C 作为肾小球滤过率标志物,在 PDAC 患者中水平异常提示肾小管功能受损,与生存期显著相关[8]。Meta 分析显示,联合标志物比单一指标能将诊断准确率提高 23.6% [3]。

目前 PDAC 的确诊路径依赖“临床症状 - 血清学检查 - 影像评估 - 组织学验证”的多步骤流程。虽然 CA19-9 是公认的血清学标志物,但其在胆道炎症和肝硬化患者中也会升高,特异性仅为 65.7%。影像学方法如增强 CT 和 MRI/MRCP 可提供形态学证据,但对 <2 cm 病灶的检出率有限(68.3%) [9]。最终确

诊仍依赖 EUS-FNA 等侵入性操作,但受取材位置和技术限制,假阴性率达 12.8%~18.6%。这种多模态诊断流程不仅增加医疗负担,还可能延误治疗时机[10]。因此,建立简便高效的非侵入性诊断模型已成为临床亟需。

2. 资料与方法

2.1. 研究人群与分组

本研究为回顾性病例对照研究,于 2018 年 6 月至 2024 年 3 月在青岛大学附属医院开展。共纳入 677 例胰胆系统疾病患者,其中 PDAC 组 210 例,良性壶腹周围疾病组 467 例。研究数据来源于医院电子病历系统、临床实验室信息系统及放射科 PACS 系统。纳入标准:(1) 年龄 ≥ 18 岁;(2) 首次因胰胆系统疾病就诊;(3) 具备完整的临床资料及实验室检查结果;(4) 最终诊断明确。PDAC 诊断基于 WHO 消化系统肿瘤分类标准(2019 版),通过超声内镜引导下细针穿刺活检(EUS-FNA)或手术切除标本的组织病理学证实。良性壶腹周围疾病包括良性中央胆总管狭窄、胆管炎、慢性胰腺炎、胰腺囊肿及胰腺脂肪瘤,均基于组织病理学确诊。排除标准:(1) 既往接受胰腺或胆道手术;(2) 合并先天性胆囊炎、肝炎、肝硬化、神经内分泌肿瘤或硬化性胆囊炎;(3) 恶性肿瘤病史;(4) 免疫功能障碍或长期使用免疫抑制剂;(5) 临床资料不完整或关键数据缺失。为减少选择偏倚,良性疾病对照组来源与 PDAC 组相同时间段和医疗单位,且初始临床表现和影像学特征相似,均表现为胰胆系统症状。本研究经青岛大学附属医院伦理委员会(批准文号 QYFY WZLL 29856)批准。

2.2. 变量的收集

收集的变量均为术前变量,包括:(1) 人口学特征:年龄(连续变量)、性别(男/女)、身高、体重、BMI (体重 kg/身高 m^2);(2) 实验室检查:血细胞计数和分类(WBC、HB、PLT、NE、LYM、MONO)、肝功能(ALT、AST、ALB、GLB、PA、TBIL、DBIL、IBIL、ALP、r-GT)、肾功能(Cr、CYS-C)、纤维蛋白原(FIB)检测结果,肿瘤标志物(糖类抗原 19-9(CA19-9)、癌胚抗原(CEA)、甲胎蛋白(AFP))、血脂(TG、TC、apoA1、apoB、HDL、LDL);(3) 影像学特征:病灶位置(胰头/胰体/胰尾/壶腹);(4) 炎症相关指标:PLT 与 LYM 比值(PLR)、NE 与 LYM 比值(NLR)、FIB 与 ALB 比值(FAR)、FIB 与 PA 比值(FPR)、全身免疫炎症指数(SII;其中 $\text{SII} = \text{PLT} \times \text{NE} / \text{LYM}$);(5) 营养评估指标:预后营养指数 PNI ($\text{ALB} + 5 \times \text{LYM}$)、营养控制状况评分(COUNT,基于血清白蛋白、总淋巴细胞计数、总胆固醇的评分)、尼泊尔评分(NPS,基于血清白蛋白、总胆固醇、NE/LYM、LYM/MONO 评分)。选择这些变量基于:年龄和性别为 PDAC 已知危险因素;肝肾功能指标反映胆道梗阻及全身状况;影像学特征协助鉴别诊断;营养评估指标反映患者整体状态。

2.3. 统计学方法

采用 SPSS 22.0、R 4.1.2 和 Zstats 软件进行统计分析。连续变量采用均数 $\pm$ 标准差(正态分布)或中位数(四分位数范围)表示,分类变量以频数或百分比表示。组间差异比较采用卡方检验(分类变量)、Student t 检验(正态分布连续变量)或 Mann-Whitney U 检验(偏态分布连续变量)。采用 1:1 倾向性评分匹配(propensity score matching, PSM)平衡组间基线特征的差异,卡钳值为两组间倾向性评分绝对差值的 0.05 倍。采用单因素分析 + Lasso 回归筛选变量,构建胰腺癌的列线图预测模型。用受试者工作特征(receiver operating characteristic, ROC)曲线的曲线下面积(area under curve, AUC)评价列线图模型预测准确性,用一致性曲线图评估模型的预测概率与实际观测概率之间的拟合度。用决策曲线(DCA)验证模型的临床有效性。以 $p < 0.05$ 为差异有统计学意义。

3. 结果

3.1. PSM 前后基线特征

本研究共纳入 677 名研究对象(胰腺癌组 210 名, 良性组 467 名)。匹配前两组在年龄(62.42 ± 14.23 岁 vs 61.96 ± 9.12 岁, $p = 0.609$)和体重(65.48 ± 12.03 kg vs 64.41 ± 9.86 kg, $p = 0.224$)无显著差异, 但在身高(165.13 ± 8.48 cm vs 166.67 ± 7.93 cm, $p = 0.027$)和性别分布(男性比例: 53.75% vs 64.76%, $p = 0.007$)存在统计学差异。

为减少基线差异带来的偏倚, 我们采用 1:1 倾向性评分匹配(PSM)方法, 匹配后获得 410 名研究对象(每组 205 名)。匹配后两组人口统计学特征达到理想平衡, 所有变量的标准化均数差(SMD)均小于 0.2, 证实匹配效果良好, 为后续生物标志物与胰腺癌关联分析提供了可靠基础(见表 1)。

Table 1. Baseline characteristics of patients before and after propensity score matching

表 1. 倾向性评分匹配前后患者基线特征

变量	PSM 前						PSM 后					
	Total (n = 677)	0 (n = 467)	1 (n = 210)	Statistic	P	SMD	Total (n = 410)	0 (n = 205)	1 (n = 205)	Statistic	p	SMD
年龄, [岁 Mean $\pm$ SD]	62.28 $\pm$ 12.86	62.42 $\pm$ 14.23	61.96 $\pm$ 9.12	t = 0.512	0.609	-0.051	62.69 $\pm$ 11.59	63.47 $\pm$ 13.62	61.92 $\pm$ 9.10	t = 1.356	0.176	-0.170
身高, [cm, Mean $\pm$ SD]	165.61 $\pm$ 8.33	165.13 $\pm$ 8.48	166.67 $\pm$ 7.93	t = -2.221	0.027	0.194	166.31 $\pm$ 7.77	166.09 $\pm$ 7.64	166.54 $\pm$ 7.91	t = -0.578	0.563	0.056
体重, [Kg, Mean $\pm$ SD]	65.15 $\pm$ 11.41	65.48 $\pm$ 12.03	64.41 $\pm$ 9.86	t = 1.217	0.224	-0.108	64.16 $\pm$ 10.51	63.83 $\pm$ 11.12	64.48 $\pm$ 9.87	t = -0.629	0.529	0.066
性别, [例(%)]				$\chi^2 = 7.177$	0.007					$\chi^2 = 0.000$	1.000	
1	387 (57.16)	251 (53.75)	136 (64.76)			0.231	262 (63.9)	131 (63.90)	131 (63.90)			0.000
2	290 (42.84)	216 (46.25)	74 (35.24)			-0.231	148 (36.1)	74 (36.10)	74 (36.10)			0.000

3.2. 训练集和验证集的均衡性检验

本研究采用分层抽样法将 410 例患者随机划分为训练集($n = 287$, 70%)和测试集($n = 123$, 30%)。基线特征分析显示两组在所有关键变量上均达到统计学均衡。所有 42 项指标的组间比较 p 值均 >0.05 (t 检验用于正态分布变量, Mann-Whitney U 检验用于偏态变量, χ^2 /Fisher 检验用于分类变量), 证实数据分层的合理性。

3.3. 训练集的基本特征与差异性分析

训练集中 287 名研究对象被分为对照组($n = 142$)和胰腺癌组($n = 145$)。两组人口统计学特征比较显示, 年龄(62.94 ± 13.55 岁 vs 61.74 ± 9.70 岁, $t = 0.86$, $p = 0.390$)、BMI (23.11 ± 3.20 vs 23.26 ± 2.72 , $t = -0.42$, $p = 0.674$)、身高(165.70 ± 6.88 cm vs 166.79 ± 8.09 cm, $t = -1.22$, $p = 0.223$)和体重(63.61 ± 10.55 kg vs 64.83 ± 9.52 kg, $t = -1.03$, $p = 0.304$)均无统计学差异。性别分布两组相似, 男性比例分别为 67.61%和 64.14% ($\chi^2 = 0.38$, $p = 0.536$)。

肿瘤标志物分析显示, 胰腺癌组 CA199 [167.50 ($48.70 \sim 384.00$) vs 24.22 ($10.51 \sim 68.37$) U/mL, $Z = -7.45$, $p < 0.001$]、CEA [3.26 ($2.12 \sim 5.17$) vs 2.09 ($1.23 \sim 2.96$) ng/mL, $Z = -5.74$, $p < 0.001$]和 AFP [3.18 ($2.22 \sim 4.15$) vs 2.62 ($1.95 \sim 3.56$) ng/mL, $Z = -2.95$, $p = 0.003$]水平均显著高于对照组。

胰腺癌组肝胆功能指标明显异常, 表现为 TBIL [75.50 ($14.40 \sim 192.36$) vs 22.85 ($15.06 \sim 42.81$) μ mol/L, $Z = -4.04$, $p < 0.001$]、DBIL [54.51 ($5.01 \sim 144.03$) vs 9.06 ($5.44 \sim 24.97$) μ mol/L, $Z = -3.94$, $p < 0.001$]和 IBIL

[19.05 (9.50~38.30) vs 12.15 (7.61~19.98) $\mu\text{mol/L}$, $Z = -3.54$, $p < 0.001$]均显著高于对照组。ALP [182.00 (79.00~333.00) vs 123.60 (81.00~210.38) U/L, $Z = -2.35$, $p = 0.019$]和 ALT (158.01 $\pm$ 158.27 vs 122.31 $\pm$ 141.13 U/L, $t = -2.02$, $p = 0.045$)也显著升高。

炎症标志物分析表明,胰腺癌组系统性炎症反应更为显著,PLR [164.80 (121.23~218.38) vs 141.32 (107.28~193.79), $Z = -2.09$, $p = 0.037$]、NLR [2.55 (1.83~3.40) vs 2.18 (1.39~3.75), $Z = -2.06$, $p = 0.040$]和 SII [588.55 (400.47~801.84) vs 475.30 (284.09~783.53), $Z = -2.15$, $p = 0.032$]均显著升高,纤维蛋白原水平也相应增高[3.46 (3.03~3.94) vs 3.18 (2.68~3.83) g/L, $Z = -2.86$, $p = 0.004$]。

胰腺癌组血红蛋白水平略低[129.00 (116.00~140.00) vs 132.50 (122.00~141.75) g/L, $Z = -1.98$, $p = 0.047$],而肌酐水平显著降低[57.00 (45.00~69.80) vs 71.25 (57.33~84.00) $\mu\text{mol/L}$, $Z = -5.29$, $p < 0.001$]。

临床营养评分系统分析显示,COUNT 评分(Fisher 精确检验, $p = 0.038$)和 NPS 评分($\chi^2 = 13.77$, $p = 0.008$)在两组间分布差异显著。胰腺癌组 COUNT 评分 2~3 级比例为 10.35%(对照组 17.78%),NPS 评分 3~4 级比例为 46.21%(对照组 45.19%),其中高分级(3 分)患者比例(37.24%)显著高于对照组(25.19%),提示营养代谢状态可能较差(见表 2)。

Table 2. Comparison of clinical data between benign group and pancreatic cancer group

表 2. 良性组和胰腺癌组的临床数据比较

变量	训练集总共(n = 287)	良性组(n = 142)	胰腺癌(n = 145)	统计值	p 值
年龄, [岁, Mean $\pm$ SD]	62.34 $\pm$ 11.76	62.94 $\pm$ 13.55	61.74 $\pm$ 9.70	$t = 0.86$	0.390
性别, [例(%)]				$\chi^2 = 0.38$	0.536
男	189 (65.85)	96 (67.61)	93 (64.14)		
女	98 (34.15)	46 (32.39)	52 (35.86)		
身高, [cm, Mean $\pm$ SD]	166.25 $\pm$ 7.52	165.70 $\pm$ 6.88	166.79 $\pm$ 8.09	$t = -1.22$	0.223
体重, [kg, Mean $\pm$ SD]	64.22 $\pm$ 10.04	63.61 $\pm$ 10.55	64.83 $\pm$ 9.52	$t = -1.03$	0.304
体质指数, [kg/m ² , Mean $\pm$ SD]	23.19 $\pm$ 2.97	23.11 $\pm$ 3.20	23.26 $\pm$ 2.72	$t = -0.42$	0.674
WBC, [$\times 10^9/\text{L}$, M (Q ₁ , Q ₃)]	5.57 (4.70, 6.73)	5.69 (4.70, 6.89)	5.54 (4.70, 6.46)	$Z = -0.47$	0.640
NE, [$\times 10^9/\text{L}$, M (Q ₁ , Q ₃)]	3.39 (2.73, 4.29)	3.37 (2.62, 4.38)	3.44 (2.88, 4.23)	$Z = -1.22$	0.221
L, [$\times 10^9/\text{L}$, M (Q ₁ , Q ₃)]	1.41 (1.06, 1.85)	1.46 (1.05, 2.00)	1.36 (1.08, 1.75)	$Z = -1.19$	0.234
M, [$\times 10^9/\text{L}$, M (Q ₁ , Q ₃)]	0.42 (0.33, 0.53)	0.41 (0.32, 0.54)	0.44 (0.34, 0.52)	$Z = -0.89$	0.375
HB, [g/L, M (Q ₁ , Q ₃)]	131.00 (118.00, 141.00)	132.50 (122.00, 141.75)	129.00 (116.00, 140.00)	$Z = -1.98$	0.047
PLT, [$\times 10^9/\text{L}$, M (Q ₁ , Q ₃)]	213.00 (175.00, 268.00)	213.00 (166.00, 270.75)	215.00 (188.00, 266.00)	$Z = -1.14$	0.253
CRP, [mg/L, M (Q ₁ , Q ₃)]	4.12 (1.18, 8.45)	3.23 (0.98, 13.94)	4.20 (1.63, 6.37)	$Z = -0.12$	0.907
CA199, [U/mL, M (Q ₁ , Q ₃)]	59.52 (16.87, 211.40)	24.22 (10.51, 68.37)	167.50 (48.70, 384.00)	$Z = -7.45$	<0.001
CEA, [ng/mL, M (Q ₁ , Q ₃)]	2.60 (1.59, 4.15)	2.09 (1.23, 2.96)	3.26 (2.12, 5.17)	$Z = -5.74$	<0.001
AFP, [ng/mL, M (Q ₁ , Q ₃)]	2.87 (2.11, 3.87)	2.62 (1.95, 3.56)	3.18 (2.22, 4.15)	$Z = -2.95$	0.003
ALB, [g/L, Mean $\pm$ SD]	38.10 $\pm$ 4.94	37.94 $\pm$ 4.78	38.27 $\pm$ 5.11	$t = -0.56$	0.574
PA, [mg/L, Mean $\pm$ SD]	205.27 $\pm$ 67.90	209.62 $\pm$ 71.30	201.08 $\pm$ 64.42	$t = 1.06$	0.289
GLB, [g/L, M (Q ₁ , Q ₃)]	24.43 (21.40, 28.75)	24.77 (21.28, 28.80)	24.33 (21.40, 28.20)	$Z = -0.69$	0.487
AGR, [M (Q ₁ , Q ₃)]	1.55 (1.33, 1.79)	1.55 (1.30, 1.79)	1.56 (1.34, 1.79)	$Z = -0.61$	0.544
ALT, [U/L, Mean $\pm$ SD]	140.34 $\pm$ 150.84	122.31 $\pm$ 141.13	158.01 $\pm$ 158.27	$t = -2.02$	0.045
AST, [U/L, M (Q ₁ , Q ₃)]	45.00 (19.65, 117.35)	34.15 (20.25, 74.97)	60.00 (19.00, 143.00)	$Z = -2.40$	0.016
γ -GT, [U/L, Mean $\pm$ SD]	351.93 $\pm$ 465.54	275.94 $\pm$ 327.19	426.36 $\pm$ 560.66	$t = -2.78$	0.006
ALP, [U/L, M (Q ₁ , Q ₃)]	137.20 (80.90, 258.85)	123.60 (81.00, 210.38)	182.00 (79.00, 333.00)	$Z = -2.35$	0.019
TBIL, [$\mu\text{mol/L}$, M (Q ₁ , Q ₃)]	28.30 (14.66, 113.48)	22.85 (15.06, 42.81)	75.50 (14.40, 192.36)	$Z = -4.04$	<0.001

续表

DBIL, [umol/L, M (Q ₁ , Q ₃)]	13.29 (5.11, 83.05)	9.06 (5.44, 24.97)	54.51 (5.01, 144.03)	Z = -3.94	<0.001
IBIL, [umol/L, M (Q ₁ , Q ₃)]	14.84 (8.60, 28.99)	12.15 (7.61, 19.98)	19.05 (9.50, 38.30)	Z = -3.54	<0.001
TG, [mmol/L, M (Q ₁ , Q ₃)]	1.24 (0.83, 1.88)	1.07 (0.79, 1.55)	1.49 (0.93, 2.11)	Z = -3.77	<0.001
TC, [mmol/L, M (Q ₁ , Q ₃)]	4.62 (3.87, 5.52)	4.34 (3.77, 5.15)	4.86 (4.23, 5.81)	Z = -3.57	<0.001
Apoa1, [g/L, Mean ± SD]	1.06 ± 0.42	1.11 ± 0.41	1.02 ± 0.42	t = 1.63	0.103
Apob, [g/L, M (Q ₁ , Q ₃)]	0.88 (0.69, 1.10)	0.85 (0.65, 0.96)	0.95 (0.71, 1.19)	Z = -2.99	0.003
HDL, [mmol/L, M (Q ₁ , Q ₃)]	1.06 (0.77, 1.43)	1.22 (0.93, 1.54)	0.94 (0.58, 1.29)	Z = -4.05	<0.001
LDL, [mmol/L, M (Q ₁ , Q ₃)]	2.96 (2.28, 3.61)	2.90 (2.26, 3.49)	3.00 (2.34, 3.76)	Z = -1.32	0.186
Cr, [mol/L, M (Q ₁ , Q ₃)]	64.12 (51.00, 79.00)	71.25 (57.33, 84.00)	57.00 (45.00, 69.80)	Z = -5.29	<0.001
CYSC, [mg/L, M (Q ₁ , Q ₃)]	0.92 (0.80, 1.02)	0.94 (0.84, 1.08)	0.88 (0.79, 1.01)	Z = -2.72	0.007
FIB, [g/L, M (Q ₁ , Q ₃)]	3.36 (2.86, 3.90)	3.18 (2.68, 3.83)	3.46 (3.03, 3.94)	Z = -2.86	0.004
LCR, M (Q ₁ , Q ₃)	0.38 (0.15, 1.19)	0.45 (0.08, 1.68)	0.34 (0.21, 0.83)	Z = -0.68	0.494
PLR, M (Q ₁ , Q ₃)	154.23 (113.07, 207.18)	141.32 (107.28, 193.79)	164.80 (121.23, 218.38)	Z = -2.09	0.037
NLR, M (Q ₁ , Q ₃)	2.42 (1.63, 3.51)	2.18 (1.39, 3.75)	2.55 (1.83, 3.40)	Z = -2.06	0.040
FAR, M (Q ₁ , Q ₃)	0.09 (0.07, 0.11)	0.08 (0.07, 0.11)	0.09 (0.08, 0.11)	Z = -2.15	0.031
FPR, M (Q ₁ , Q ₃)	0.02 (0.01, 0.02)	0.01 (0.01, 0.02)	0.02 (0.01, 0.02)	Z = -2.78	0.006
SII, M (Q ₁ , Q ₃)	539.13 (342.65, 794.77)	475.30 (284.09, 783.53)	588.55 (400.47, 801.84)	Z = -2.15	0.032
PNI, M (Q ₁ , Q ₃)	45.50 (42.11, 50.04)	46.00 (41.88, 50.14)	45.20 (42.20, 49.96)	Z = -0.49	0.627
COUNT, 例(%)				-	0.038
0	104 (37.14)	49 (36.30)	55 (37.93)		
1	137 (48.93)	62 (45.93)	75 (51.72)		
2	36 (12.86)	24 (17.78)	12 (8.28)		
3	3 (1.07)	0 (0.00)	3 (2.07)		
NPS, 例(%)				$\chi^2 = 13.77$	0.008
0	22 (7.86)	13 (9.63)	9 (6.21)		
1	57 (20.36)	32 (23.70)	25 (17.24)		
2	73 (26.07)	29 (21.48)	44 (30.34)		
3	88 (31.43)	34 (25.19)	54 (37.24)		
4	40 (14.29)	27 (20.00)	13 (8.97)		

t: t 检验(t-test); Z: Mann-Whitney 检验(Mann-Whitney test); χ^2 : 卡方检验(Chi-square test); -: Fisher 精确检验 (Fisher exact test); SD: 标准差(standard deviation); M: 中位数(Median); Q₁: 第一四分位数(1st Quartile); Q₃: 第三四分位数 (3rd Quartile)。

3.4. LASSO 回归筛选变量

经过单因素 logistic 回归分析(表 3), 将 $p < 0.05$ 的变量纳入 Lasso 回归进行筛选, 在 19 个变量中共筛选出 5 个变量, 分别为 CA19-9、直接胆红素(DBIL)、癌胚抗原(CEA)、胱抑素-C (CYS-C)、COUNT 评分。Lasso 模型的最佳参数(λ)选择, 根据最小标准的 10 重交叉验证, 取 $\lambda = 0.051610$, 筛选得到最终变量, Lasso 交叉验证图及路径图(见图 1)。

3.5. 列线图模型的建立和验证

以筛选的 5 个变量构建胰腺癌的列线图模型(图 2)。模型在训练集上展现出良好的判别能力, 受试者工作特征曲线(ROC)下面积(AUC)达到 0.86 (95%CI: 0.82~0.90)。采用优化的截断阈值 0.434, 模型实现了 77% (95%CI: 0.72~0.82)的总体准确率, 为临床应用提供了可靠基础(图 3、表 4)。

Table 3. Univariate logistic regression analysis
表 3. 单因素 logistic 回归分析

变量	β	S.E	Z	p	OR (95%CI)
CA199	0.01	0.00	2.00	0.046	1.01 (1.00~1.00)
CEA	0.21	0.06	3.52	<0.001	1.24 (1.10~1.39)
AFP	0.22	0.08	2.83	0.005	1.25 (1.07~1.45)
HB	-0.02	0.01	-2.18	0.029	0.98 (0.97~0.99)
TBIL	0.01	0.00	5.07	<0.001	1.01 (1.01~1.01)
DBIL	0.01	0.00	5.50	<0.001	1.01 (1.01~1.02)
IBIL	0.02	0.01	3.05	0.002	1.02 (1.01~1.03)
ALT	0.01	0.00	1.99	0.047	1.01 (1.01~1.01)
AST	0.01	0.00	2.59	0.010	1.01 (1.01~1.01)
ALP	0.01	0.00	3.21	0.001	1.01 (1.01~1.01)
γ -GT	0.01	0.00	2.63	0.009	1.01 (1.01~1.01)
TG	0.49	0.16	3.12	0.002	1.63 (1.20~2.21)
TC	0.31	0.09	3.41	<0.001	1.37 (1.14~1.64)
apoB	0.97	0.39	2.46	0.014	2.65 (1.22~5.74)
HDL	-0.73	0.23	-3.15	0.002	0.48 (0.31~0.76)
Cr	-0.04	0.01	-5.06	<0.001	0.96 (0.95~0.98)
CYSC	-1.98	0.74	-2.68	0.007	0.14 (0.03~0.59)
FIB	0.35	0.15	2.32	0.020	1.42 (1.06~1.91)
PLR	0.00	0.00	0.84	0.399	1.00 (1.00~1.00)
NLR	0.00	0.03	0.01	0.990	1.00 (0.95~1.05)
FAR	8.16	4.64	1.76	0.079	3491.56 (0.39~31114247.14)
FPR	2.58	8.22	0.31	0.754	13.16 (0.00~130482925.17)
SII	0.00	0.00	0.30	0.761	1.00 (1.00~1.00)
COUNT					
0					1.00 (Reference)
1	0.07	0.26	0.29	0.774	1.08 (0.65~1.80)
2	-0.81	0.40	-2.00	0.046	0.45 (0.20~0.98)
3	15.45	840.27	0.02	0.985	5129687.82 (0.00~Inf)

续表

NPS					
0					1.00 (Reference)
1	0.12	0.51	0.24	0.812	1.13 (0.42~3.06)
2	0.78	0.50	1.58	0.113	2.19 (0.83~5.78)
3	0.83	0.49	1.71	0.087	2.29 (0.89~5.94)
4	-0.36	0.55	-0.66	0.509	0.70 (0.24~2.04)

OR: 比值比, CI: 置信区间。

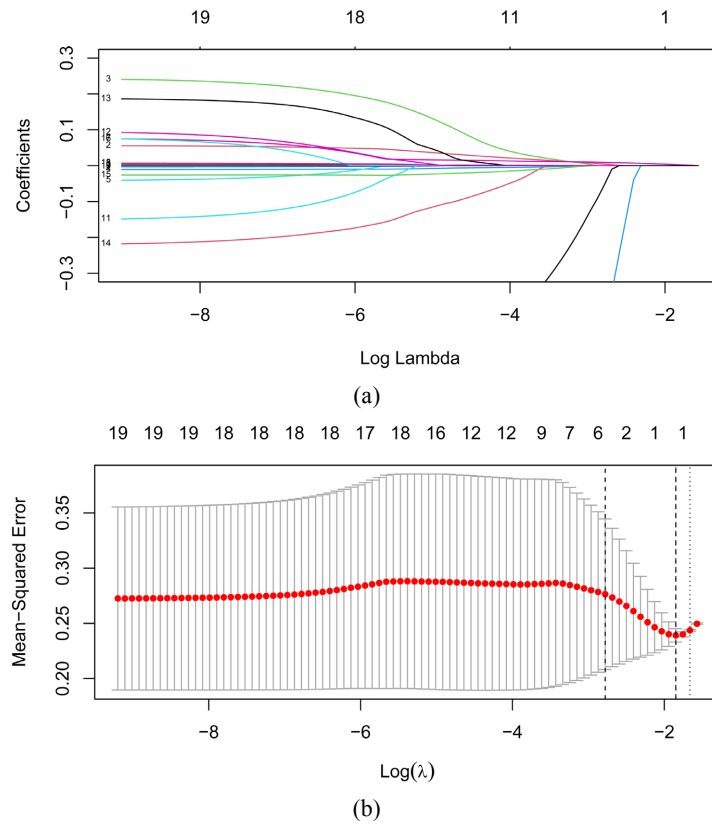


Figure 1. (a) Lasso regression coefficient path plot and (b) Cross-validation plot
图 1. (a) Lasso 回归系数路径图与(b) 交叉验证图

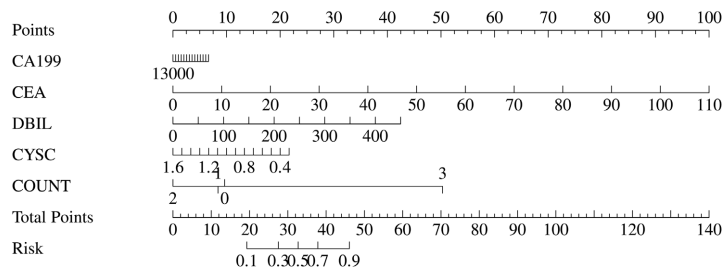
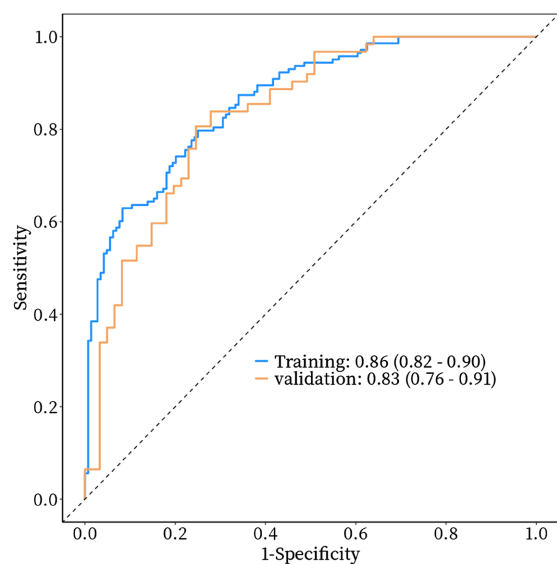


Figure 2. Nomogram model for predicting pancreatic cancer
图 2. 预测胰腺癌的列线图模型



注：Sensitivity 为敏感度；1-Specificity 为假阳性率；训练集 AUC 为 0.86；验证集 AUC 为 0.83。

Figure 3. ROC curves and AUC values for the training set and validation set
图 3. 训练集和验证集的 ROC 曲线及 AUC 面积

Table 4. Confusion matrices for the training set and validation set
表 4. 训练集和验证集混淆矩阵

Data	AUC (95%CI)	Accuracy (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	cut off
训练集	0.86 (0.82~0.90)	0.77 (0.72~0.82)	0.75 (0.68~0.82)	0.80 (0.73~0.86)	0.79 (0.72~0.86)	0.76 (0.69~0.83)	0.434
验证集	0.83 (0.75~0.90)	0.73 (0.64~0.81)	0.74 (0.63~0.85)	0.72 (0.61~0.84)	0.74 (0.63~0.85)	0.72 (0.61~0.84)	0.451

诊断性能分析显示，该模型在特异性方面略优于敏感性，分别为 80% (95%CI: 0.73~0.86)和 75% (95%CI: 0.68~0.82)。这种特性使其在确认诊断方面表现优异，可有效减少假阳性结果，从而降低不必要的医疗资源消耗和患者负担。同时，75%的敏感性也确保了模型在初筛环节具有足够的检出能力，仅有 25%的真实阳性病例可能被漏诊。

进一步分析预测价值指标，我们观察到模型具有相对平衡的阳性预测值(PPV)和阴性预测值(NPV)，分别为 79% (95%CI: 0.72~0.86)和 76% (95%CI: 0.69~0.83)。较高的 PPV 表明，当模型预测结果为阳性时，约 79%的病例确实为阳性，这为临床决策提供了较高的可信度。而 76%的 NPV 则提示在解释阴性结果时需保持适度谨慎，尤其是针对高风险人群。

在胰腺癌预测中，该预测模型具有良好的临床决策价值，特别是在中高风险阈值区间(0.4~0.9) (图 4)。

校准曲线图证明了该胰腺癌预测模型不仅具有良好的区分能力(之前提到的 AUC = 0.86)，还具有出色的校准性能(图 5)。

3.6. 确立独立危险因素

多因素 Logistic 回归分析显示(表 5)，CA19-9 (OR = 1.01, 95%CI: 1.01~1.01, p = 0.024)、CEA (OR = 1.16, 95%CI: 1.02~1.32, p = 0.025)和 DBIL (OR = 1.02, 95%CI: 1.01~1.02, p < 0.001)均为 PDAC 的独立危险因素。尤其值得注意的是，CA19-9 在多因素分析中达到了统计学显著性，表明其在控制其他因素后的独立预测价值。胱抑素 C 在多因素分析中保持强保护作用，且效应增强(OR = 0.06, 95%CI: 0.01~0.36, p = 0.002)，表明每单位增加与 PDAC 风险降低 94%相关。

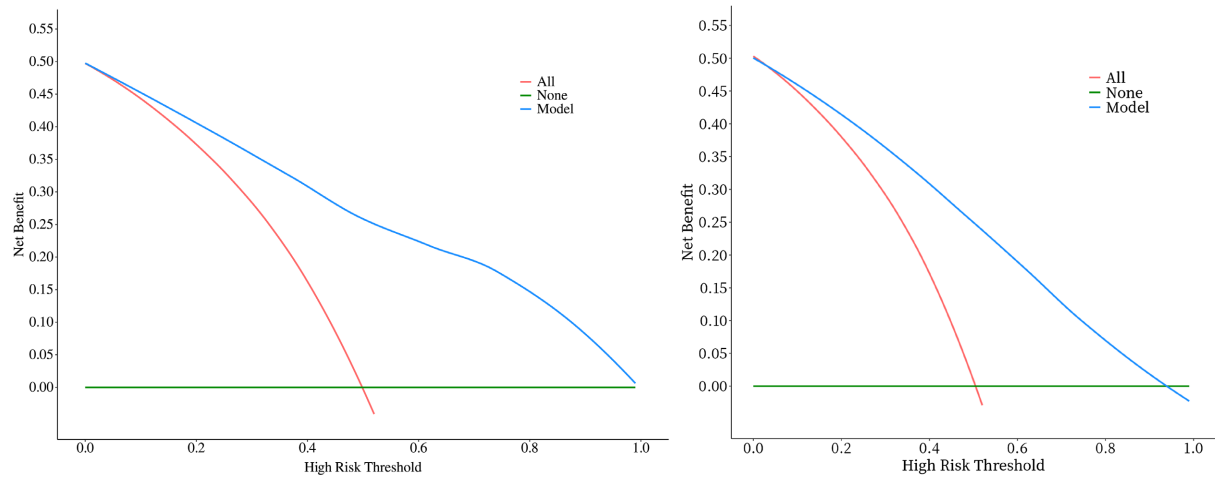


Figure 4. Decision curve analysis for the training set (left) and validation set (right)

图 4. 训练集(左)和验证集(右)的决策曲线分析

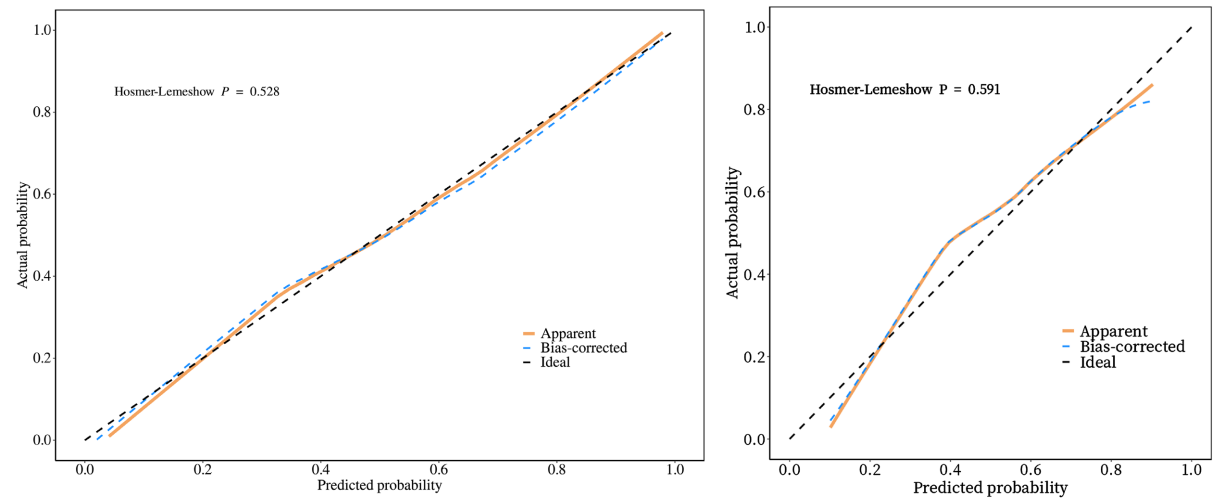


Figure 5. Calibration curves for the training set (left) and validation set (right)

图 5. 训练集(左)和验证集(右)的校准曲线

Table 5. Multivariate Logistic regression analysis

表 5. 多因素 Logistic 回归分析

变量	多因素 Logistic 回归分析				
	β	S.E	Z	p	OR (95%CI)
CA199	0.01	0.00	2.26	0.024	1.01 (1.01~1.01)
CEA	0.15	0.07	2.24	0.025	1.16 (1.02~1.32)
DBIL	0.02	0.00	5.30	<0.001	1.02 (1.01~1.02)
CYSC	-2.74	0.87	-3.14	0.002	0.06 (0.01~0.36)
COUNT					
0					1.00 (Reference)

续表

1	-0.20	0.30	-0.67	0.500	0.82 (0.45~1.47)
2	-1.58	0.51	-3.09	0.002	0.21 (0.08~0.56)
3	14.98	1264.15	0.01	0.991	3215656.77 (0.00~Inf)

4. 讨论

本研究检验了五种常规血液标志物(CA19-9、CEA、直接胆红素、胱抑素 C 和 COUNT 评分)对胰腺导管腺癌(PDAC)诊断的预测价值。这项纳入 677 例胰胆系统疾病患者的回顾性病例对照研究采用 1:1 倾向性评分匹配消除基线差异,并通过 LASSO 回归筛选出最具预测价值的变量组合。结果表明,基于这五项血液指标构建的列线图模型在鉴别 PDAC 方面表现优异,曲线下面积达 0.86 (95%CI: 0.82~0.90),在最佳阈值 0.434 处总体准确率为 77% (95%CI: 0.72~0.82)。

本研究结果与既往文献高度一致,同时呈现独特发现:CA19-9 方面,Yang 等对胰胆系统疾病患者的研究[4],与我们的模型(AUC = 0.86)相近。Mason 等通过队列研究证实 CEA 对胰腺癌具有显著预测价值[5],与本研究结果(OR = 1.16, 95%CI: 1.02~1.32)基本一致。Boyd 等人的研究证实直接胆红素是胰腺癌的独立危险因素[6],与我们的发现(OR = 1.02, 95%CI: 1.01~1.02)高度吻合。在胱抑素 C 作用上,本研究与既往报道存在差异。Kleeman 等在胰腺癌患者的研究中发现胱抑素 C 升高与不良预后相关[8],而我们的研究显示胱抑素 C 是胰腺癌诊断的保护因素(OR = 0.06, 95%CI: 0.01~0.36)。这种差异可能源于研究设计和目的的不同——Kleeman 等主要关注预后而非诊断,且未对肾功能等关键混杂因素进行充分调整。

从机制上看,CEA 作为细胞黏附分子可促进肿瘤细胞扩散和转移[5],胆红素升高反映了肿瘤导致的胆道梗阻及肝细胞损伤[6],而胱抑素 C 作为肾功能标志物,其保护作用可能反映了早期胰腺癌患者肾小管功能的代偿性改变。COUNT 评分是通过白蛋白、总淋巴细胞计数和总胆固醇计算得出的营养-炎症综合评分,本研究首次发现 COUNT = 2 评分对胰腺癌具有保护作用(OR = 0.21, 95%CI: 0.08~0.56)。轻度 COUNT 异常可能代表一种“适度应激状态”,在此状态下机体免疫系统处于警觉但未过度消耗的平衡点。适度的营养应激可激活细胞自噬机制,促进受损细胞清除,防止异常细胞累积和肿瘤发生[11]。在《Nat Rev Gastroenterol Hepatol》发表的研究表明,胰腺微环境中的适度代谢压力可诱导组织驻留巨噬细胞向 M1 型极化,增强对早期恶变细胞的监视和清除能力[12]。

本研究构建的五维整合预测模型展现出显著的临床转化价值。当前 PDAC 确诊路径需要多步骤流程,平均周期达 18.5 天,而本模型仅通过常规血液检查即可实现 77%的准确率,有望缩短诊断时间并降低医疗成本。与传统单一标志物相比,本模型特异性(80%)明显高于单一 CA19-9 (65.7%),有效解决了良恶性胰胆疾病鉴别诊断的关键挑战。该模型整合了肿瘤标志物(CA19-9、CEA)、肝功能(DBIL)、肾功能(CYS-C)及营养状态(COUNT 评分),全面反映胰腺癌患者的多维生理改变。在实际应用中,该模型可作为筛查高风险人群的辅助工具,尤其适用于影像学检查前的风险分层,帮助临床医生识别需优先进行侵入性确诊的患者。根据决策曲线分析,当风险阈值设定在 0.4~0.9 区间时,模型提供最大净获益,这正是临床怀疑 PDAC 患者的常见高风险区间。

本研究方法学优势确保了结果的可靠性:1:1 PSM 策略有效消除基线差异(SMD < 0.2),显著减少选择偏倚;严格入选标准和多中心数据采集确保研究对象代表性和数据质量;分层抽样(7:3)进行集划分,确保基线特征无统计学差异(p > 0.05);单因素分析联合 LASSO 回归控制多重共线性,同时保留最具预测价值的标志物;模型验证采用 ROC 曲线、校准曲线和决策曲线分析三重评估体系。尽管如此,本研究仍

存在局限性：首先，尽管采用了倾向性评分匹配，但仍存在潜在未测量混杂因素；其次，单中心研究限制了结果普适性，需多中心验证；第三，研究主要针对中国汉族人群，推广至其他种族时应谨慎；第四，回顾性设计仅能确立相关性而非因果；最后，模型缺乏前瞻性外部验证，临床实用性有待评估。

参考文献

- [1] Ward, Z.J., Gaba, Q. and Atun, R. (2024) Cancer Incidence and Survival for 11 Cancers in the Commonwealth: A Simulation-Based Modelling Study. *The Lancet Oncology*, **25**, 1127-1134. [https://doi.org/10.1016/s1470-2045\(24\)00336-x](https://doi.org/10.1016/s1470-2045(24)00336-x)
- [2] 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>
- [3] Wood, L.D., Canto, M.I., Jaffee, E.M. and Simeone, D.M. (2022) Pancreatic Cancer: Pathogenesis, Screening, Diagnosis, and Treatment. *Gastroenterology*, **163**, 386-402.e1. <https://doi.org/10.1053/j.gastro.2022.03.056>
- [4] Yang, J., Xu, R., Wang, C., Qiu, J., Ren, B. and You, L. (2021) Early Screening and Diagnosis Strategies of Pancreatic Cancer: A Comprehensive Review. *Cancer Communications (Lond)*, **41**, 1257-1274. <https://doi.org/10.1002/cac2.12204>
- [5] Mason, J., Lundberg, E., Jonsson, P., Nyström, H., Franklin, O., Lundin, C., Naredi, P., Antti, H., Sund, M. and Öhlund, D. (2022) A Cross-Sectional and Longitudinal Analysis of Pre-Diagnostic Blood Plasma Biomarkers for Early Detection of Pancreatic Cancer. *International Journal of Molecular Sciences*, **23**, Article 12969. <https://doi.org/10.3390/ijms232112969>
- [6] Boyd, L.N.C., Ali, M., Kam, L., Puik, J.R., Rodrigues, S.M.F., Zwart, E.S., et al. (2022) The Diagnostic Value of the CA19-9 and Bilirubin Ratio in Patients with Pancreatic Cancer, Distal Bile Duct Cancer and Benign Periapillary Diseases, a Novel Approach. *Cancers*, **14**, 344. <https://doi.org/10.3390/cancers14020344>
- [7] Ma, X., Zou, W. and Sun, Y. (2022) Prognostic Value of Pretreatment Controlling Nutritional Status Score for Patients with Pancreatic Cancer: A Meta-Analysis. *Frontiers in Oncology*, **11**. <https://doi.org/10.3389/fonc.2021.770894>
- [8] Kleeman, S.O., Thakir, T.M., Demestichas, B., Mourikis, N., Loiero, D., Ferrer, M., et al. (2023) Cystatin C Is Glucocorticoid Responsive, Directs Recruitment of Trem2+ Macrophages, and Predicts Failure of Cancer Immunotherapy. *Cell Genomics*, **3**, 100347. <https://doi.org/10.1016/j.xgen.2023.100347>
- [9] Chu, L.C. and Fishman, E.K. (2023) Pancreatic Ductal Adenocarcinoma Staging: A Narrative Review of Radiologic Techniques and Advances. *International Journal of Surgery*, **110**, 6052-6063. <https://doi.org/10.1097/js9.0000000000000899>
- [10] Conroy, T., Pfeiffer, P., Vilgrain, V., Lamarca, A., Seufferlein, T., O'Reilly, E.M., et al. (2023) Pancreatic Cancer: ESMO Clinical Practice Guideline for Diagnosis, Treatment and Follow-Up. *Annals of Oncology*, **34**, 987-1002. <https://doi.org/10.1016/j.annonc.2023.08.009>
- [11] König, T. and McBride, H.M. (2024) Mitochondrial-Derived Vesicles in Metabolism, Disease, and Aging. *Cell Metabolism*, **36**, 21-35. <https://doi.org/10.1016/j.cmet.2023.11.014>
- [12] Encarnación-Rosado, J. and Kimmelman, A.C. (2021) Harnessing Metabolic Dependencies in Pancreatic Cancers. *Nature Reviews Gastroenterology & Hepatology*, **18**, 482-492. <https://doi.org/10.1038/s41575-021-00431-7>