

比较基于机器学习的不同临床模型框架在非转移性胰头癌、胰体癌和胰尾癌患者总生存率和特异生存率方面的稳定性

吴冠宇¹, 李金昊^{2*}

¹四川大学计算机科学学院, 四川 成都

²重庆医科大学附属第二医院肝胆外科, 重庆

收稿日期: 2023年10月21日; 录用日期: 2023年11月15日; 发布日期: 2023年11月22日

摘要

近年来, 预测模型在临床实践中的应用越来越广泛, 针对不同患者的不同临床预测模型层出不穷。机器学习在医学中的应用正在逐渐增加。因此, 本文基于机器学习的视角, 研究非转移性胰头癌、胰体癌和胰尾癌患者的5年和全数据总生存期(OS)和癌症特异性生存期(CSS), 并尝试探讨整合模型、线性模型和生存树模型之间的差异性和稳定性。方法: 基于机器学习技术, 我们围绕患者的基本和临床信息构建了模型。选择了临床回顾性医学分析建模中目前最常用的七种模型进行比较, 评估了不同类型和相同类型之间连续和分类变量数据的区分能力和准确性。研究考虑了两个终点结果: 5年OS和全数据CSS。利用C-指数(一致性指数)、Brier分数、校准曲线以及净再分类指数(NRI)评估了模型的性能。结果: 从2000年至2018年, 共收集了6019例病理学确认的胰腺头、体和尾部癌症患者的数据。经过严格筛选, 最终纳入研究的病例为3675例。研究显示, 模型在预测CSS方面的准确性略优于OS。值得注意的是, 梯度提升生存分析(GBSA)在各种变量类型和生存期间的CSS预测中表现最佳, 无论是使用连续变量(C-指数: 0.753, 95% CI: 0.741~0.765)还是分类变量(0.743, 0.735~0.751)。NRI分析显示, 相较于分类变量, 对于与OS相关的连续变量, 应用Cox比例风险(CoxPH)生存分析提高了5年生存的预测能力30.5%, CSS模型提高了26.8%。NRI的散点图显示了模型之间在预测能力上的差异。结论: 在所研究的模型中, GBSA表现出最高的预测能力和区分度。此外, 随着临床指标的细化, 多变量模型的预测能力可能会进一步提升。基于机器学习的临床前模型的整合在未来可能为肿瘤患者提供更精确的个性化治疗方案。

关键词

机器学习, 胰腺癌, 生存框架, 跨学科应用, 预测模型

*通讯作者。

To Compare the Stability of Different Clinical Model Frameworks in the Overall Survival Rate and Specific Survival Rate of Patients with Non-Metastatic Pancreatic Head Cancer, Pancreatic Body Cancer and Pancreatic Tail Cancer Based on Machine Learning

Guanyu Wu¹, Jinhao Li^{2*}

¹College of Computer Science, Sichuan University, Chengdu Sichuan

²Department of Hepatobiliary Surgery, The Second Affiliated Hospital of Chongqing Medical University, Chongqing

Received: Oct. 21st, 2023; accepted: Nov. 15th, 2023; published: Nov. 22nd, 2023

Abstract

In recent years, the integration of predictive models into clinical practice has gained momentum, revolutionizing patient care in oncology. This study employs a machine learning perspective to investigate the 5-year and long-term overall survival (OS) as well as cancer-specific survival (CSS) in patients with non-metastatic pancreatic cancer. The aim is to discern disparities and reliability among integrated, linear, and survival tree models. **Methods:** Utilizing machine learning techniques, we constructed models using essential patient data. Seven commonly employed models in retrospective clinical analysis were selected for comparison, evaluating their discriminative power and accuracy for continuous and categorical variables within and between different cancer types. Two outcome measures were considered: 5-year OS and full-data CSS. Model performance was assessed using the Concordance index (C-index), Brier score, calibration curve, and Net Reclassification Index (NRI). **Results:** From 2000 to 2018, a total of 6019 pathologically confirmed pancreatic head, body, and tail cancer patients were collected. Following rigorous screening, 3675 patients were included in the study. The models exhibited slightly superior accuracy in predicting CSS compared to OS. Notably, Gradient Boosting Survival Analysis (GBSA) outperformed other models in predicting CSS for both continuous (C-index: 0.753, 95% CI: 0.741~0.765) and categorical variables (0.743, 0.735~0.751) across different variable types and survival periods. The NRI analysis revealed notable enhancements in predictive power when employing Cox proportional hazards (CoxPH) Survival Analysis for continuous variables in both OS (30.5% improvement) and CSS (26.8% improvement) compared to categorical variables. Scatter plots of NRI highlighted variations in prediction capability among models. **Conclusion:** Among the models scrutinized, GBSA exhibited the highest predictive power and discrimination. Additionally, the predictive capacity of multivariate models may be further enhanced with the refinement of clinical indicators. The integration of machine learning-based preclinical models holds promise for delivering more precise personalized treatments for cancer patients in the future.

Keywords

Machine Learning, Pancreatic Cancer, Survival Framework, Interdisciplinary Application,

Prediction Mode

Copyright © 2023 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. 研究背景

胰腺癌(PC)是一种常见的高度恶性肿瘤,通常发生在胰头,其次是胰体和胰尾。该病的早期症状隐匿。随着肿瘤的生长和扩散,患者将出现黄疸、体重减轻、食欲不振、上腹疼痛等不同的临床症状[1]。由于早期诊断和晚期治疗的难度,及时在疾病早期阶段进行检测和干预显得尤为重要。在肿瘤转移之前,及时检测和阻止病情进展将是未来临床工作的巨大挑战,而正确准确的预测工具的使用可能使这类患者能够获得最佳的治疗决策。我们在 Pubmed 上使用短语“胰腺癌人工智能”进行搜索,并发现自 2019 年以来,在这个领域发表的文章数量显著增加。目前,全球各国仍主要关注不同的模型和智能算法以提醒这类患者的预后。最初,该领域的发展目的主要基于各种实验室生化指标,但预后模型将患者的生存结果作为终点[2]。

在医学肿瘤学领域,机器学习和人工智能密切相关,涉及算法和数学模型的构建,无论是在诊断、新药开发还是生存概率预测方面,机器学习和人工智能都将占据更为重要的位置[3]。通过简单地结合患者的年龄、胰腺癌的部位、肿瘤的 TNM 分期、病理分级等指标,可以得到一个有效的模型。通过细化每个指标变量和分组,增加了模型包含的信息量,因此提高了模型的区分能力[4]。此外,在临床模型的分析中,是否将患者的临床指标作为连续变量或将临床数据分组作为分类变量进行研究也一直是需要研究的问题之一。众所周知,模型的区分性是评估模型的关键指标,而预测能力则决定了模型的准确性和效益[5]。同时,具有高区分能力的模型还可以在在一定程度上协助临床医生为患者做出个性化治疗决策,为癌症患者提供最大程度的治疗。

机器学习是一种基于数据和算法的人工智能技术,它通过让计算机从大量数据中自动学习和改进预测模型,可以对未知数据进行精确预测[6][7]。过去,与机器学习相关的一些胰腺癌研究也尝试探讨了不同临床模型的优缺点,个性化精准治疗将在未来医学肿瘤学领域占据更为重要的地位。机器学习将在未来临床工作中发挥更为重要的作用。在本文中,我们将机器学习与以前临床研究中常用的模型相结合,并深入研究了连续变量和分类变量对模型区分能力和预测能力的影响。基于胰腺癌患者整体生存(OS)和癌特异性生存(CSS)的两个终点结果,总结了过去常用的临床模型,为未来精准治疗奠定了部分基础。

2. 方法

2.1. 初步数据整理

我们使用前瞻性的维护监测、流行病学和结果(SEER)数据库提取和筛选患者数据,筛选了从 2000 年到 2018 年诊断为非转移性胰腺癌的 6019 名患者,并根据发病部位进行分类(部位编码 C25.0、C25.1、C25.2)。我们设置了以下纳入标准:1) 年龄在 20 到 84 岁之间;2) 病理诊断为原发性胰腺癌;3) 疾病确实在胰腺头、体和尾部发生;4) 仅患有胰腺癌,且为唯一的原发肿瘤;5) 未发现远处器官或淋巴结的转移。我们还设置了以下排除标准以进一步筛选数据:1) 未知病理等级;2) 患者 TNM 分期(第 7 版美国癌症联合委员会 TNM 分期)未知;3) 患者的生存状态或结果事件的原因缺失。由于 SEER 的一些数据

缺失或未知, 我们同时包括了患者的肿瘤分期和 T 分期, 以更方便、准确地筛选非转移性特征。除了上述条件, 我们还需要每位患者的具体年龄和性别。由于患者的治疗方式存在较大的个体差异, 因此未包括在内。由于 SEER 允许公开访问数据, 因此无需针对从 SEER 提取和整合数据进行道德批准。最终, 经过本研究的纳入和排除标准筛选, 共纳入 3675 名患者进行分析。其中, 患者的最长随访期为 107 个月。为了方便 NRI (净再分类指数) 的计算和分析, 我们将其称为全数据存活, 并分别分析了 5 年生存率(OS) 和 5 年特定生存率(CSS)的相似性和差异。

2.2. 模型选择和比较

我们选择了七种不同的临床模型, 分别是梯度提升生存分析(GBSA)、Cox 比例危险(CoxPH)生存分析、Cox 网络生存分析(CnSA)、分量智能梯度提升(CwGB)生存分析、随机生存森林(BSF)、生存树(ST)、额外生存树(EST), 并根据多变量模型和层次模型对其进行分类。在训练模型时, 我们添加了以下变量: 性别、肿瘤发生部位、病理分级, 并根据不同类型的年龄建立了模型。由于本文所包括的患者的肿瘤分期仅用于方便筛选数据, 并且依赖于 T 分期, 为防止多重共线性, 我们仅将 T 分期纳入模型。我们根据患者的临床参数准备了两组数据来训练模型。连续数值模型由患者的实际年龄和各种临床参数组成, 而分类数值模型由 60 岁的截断值和相同的临床参数组成。在保持其他临床指标不变的情况下, 通过将年龄作为连续变量和分类变量来训练模型, 探讨了不同类型的变量对模型拟合和预测能力的影响。

对于数据集, 我们将其随机划分为训练集和测试集(8:2), 在此基础上, 使用五折交叉验证(k-fold, k = 5)来训练和评估模型。对于所有使用的模型, 都使用相同的数据集和其划分方法, 参数更新的默认迭代次数为 100 次。不同的模型使用不同的算法来更新参数。从五折交叉验证的测试集的平均结果中计算出 C 指数(一致性指数)、Brier 评分, 并绘制校准图, 以解释不同模型的预测概率与实际观察值之间的差异。

2.3. 模型改进

为了进一步衡量连续值和分类值对模型的影响, 我们使用 NRI 来评估改进程度。在控制其他变量的情况下, 使用了临床研究中最常用的 CoxPH 生存分析来分别建立 5 年全数据风险回归模型, 采用 100 次自助抽样。探讨了年龄变量类型对模型预测能力的影响, 并绘制了散点图以提供可视化展示。本文中的所有统计分析均在 Python (版本 3.7.16)和 R (版本 4.2.3)中完成。

3. 结果

图 1 展示了我们的数据筛选过程, 其中有 2237 名患者的病理分级未知, 71 名患者的肿瘤分期未知, 36 名患者的生存信息缺失, 最终筛选出 3675 名患者并纳入研究。在队列中, 1197 名(32.6%)患者年龄在 60 岁以下, 2478 名(67.4%)患者年龄在 60 岁或以上。这些患者的平均年龄为 67.3 岁。男女比例大致相等, 大多数患者为白人。

根据肿瘤的发生部位, 胰头癌患者最多, 为 2298 例(62.5%), 胰体癌最少, 只有 553 例(15.1%)。在病理分期方面, I 期和 II 期患者较多, 分别为 1243 例(33.8%)和 1443 例(39.3%), 验证了该队列中大多数患者具有良好的病理分化程度。在 T 分期中, T3 阶段的患者比例最大, 达 1746 例(47.5%), T4 阶段最少, 仅 382 例(10.4%)(表 1)。

在不同模型比较中, 与 OS 相关的模型普遍优于与 CSS 相关的模型, 其中 GBSA 模型显示出最佳的拟合性能。在分类数据模型比较中, 与 OS 相关的 GBSA 模型的区分度为 0.733 (95%置信区间: 0.726~0.739), Brier 评分也显示出较低的值为 0.163 (0.045~0.281)(表 2)。在与 CSS 相关的模型中, GBSA 模型的 C 指数和 Brier 评分分别为 0.743 (0.735~0.751)和 0.162 (0.026~0.297)(表 3)。ST 模型在所有模型

中表现最差, OS 相关模型的 C 指数为 0.718 (0.709~0.728), CSS 相关模型为 0.731 (0.718~0.743), 其余模型的表现差异不大。在以连续数据构建的模型中, 与 CSS 相关的 GBSA 模型也显示出最佳的预测和拟合性能, C 指数为 0.753 (0.741~0.765), Brier 评分较低为 0.142 (0.013~0.271) (表 4)。在 OS 相关和 CSS 相关模型中, 由连续值组成的模型优于由分类值组成的模型, 但在 ST 模型和 CwGB 模型中观察到相反的结果(表 5)。图 2 显示了七种不同模型的校准曲线, 分别是 OS 的分类数值模型(A)、CSS 的分类数值模型(B)、OS 的连续数值模型(C)和 CSS 的连续数值模型(D)。

表 6 展示了基于 OS 和 CSS 的 CoxPH 模型在分类值和连续值方面的预测能力差异。在与 OS 相关的 NRI 计算中, 由连续值构建的模型在 5 年 OS 方面比由分类值构建的模型高出 0.305 (95% CI: 0.215~0.388) 并在全数据方面高出 0.313 (0.168~0.437)。在与 CSS 相关的 NRI 计算中, 5 年 CSS 高出 0.268 (0.185~0.333) 并在全数据方面高出 0.263 (0.122~0.339)。图 3 是具有不同类型变量的 CoxPH 模型中 NRI 相关的散点图, 其中带有分类值的 OS 模型和 CSS 模型分别为(A)和(B), 带有连续值的 OS 和 CSS 模型分别为(C)和(D)。

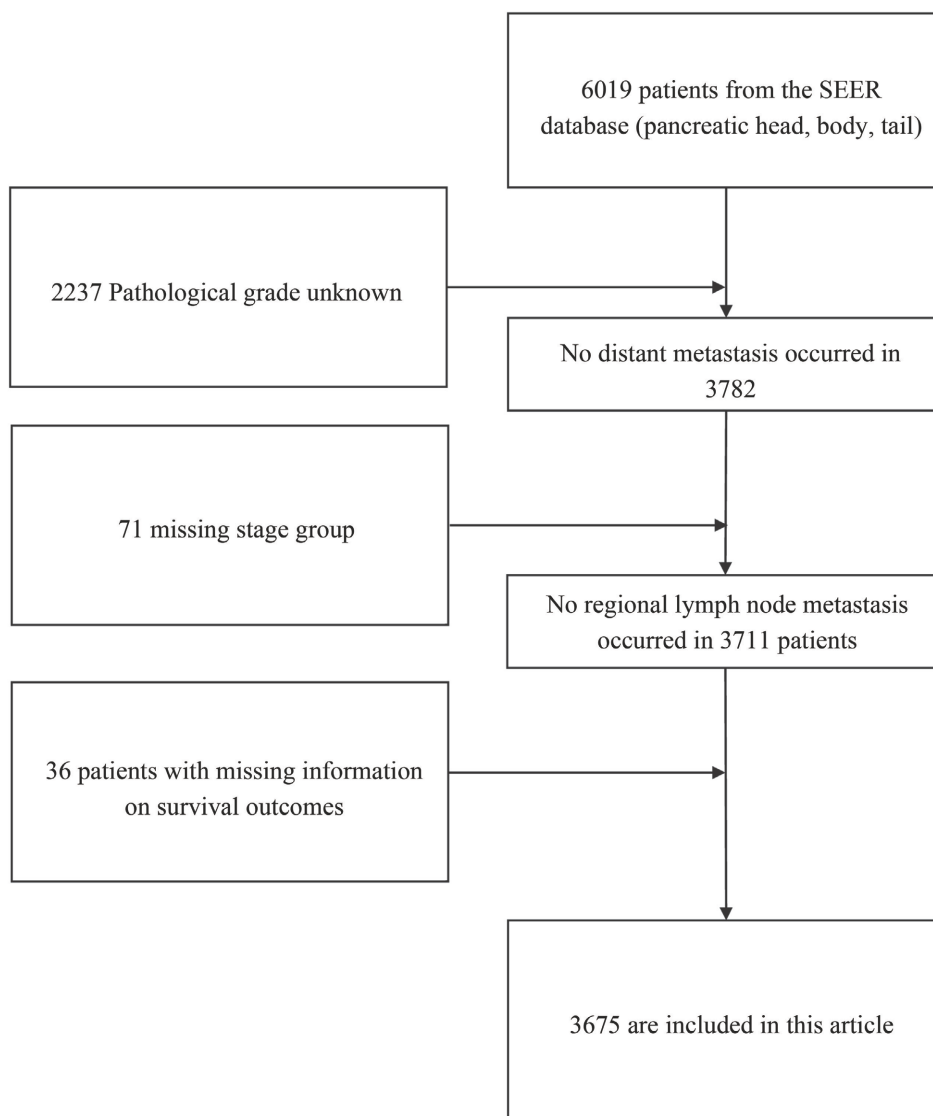


Figure 1. Data screening flow chart
图 1. 数据筛选流程图

Table 1. Basic information

表 1. 基本信息

N = 3675	
平均年龄	63.7 岁
年龄分组	
<60 岁	1197 (32.6%)
≥60 岁	2478 (67.4%)
性别	
女性	1855 (50.5%)
男性	1820 (49.5%)
种族	
白人	2844 (77.4%)
黑人	434 (11.8%)
美洲印第安人/阿拉斯加土著民族	16 (0.4%)
亚洲人或太平洋岛民	360 (9.8%)
未知	21 (0.6%)
原发部位	
腺头癌	2298 (62.5%)
胰体癌	553 (15.1%)
胰尾癌	824 (22.4%)
病理分级	
Grade I	1243 (33.8%)
Grade II	1443 (39.3%)
Grade III	928 (25.3%)
Grade IV	61 (16.6%)
临床分期	
IA	720 (19.6%)
IB	827 (22.5%)
IIA	1746 (47.5%)
III	382 (10.4%)
T 分期	
T1	720 (19.6%)
T2	827 (22.5%)
T3	1746 (47.5%)
T4	382 (10.4%)
胰腺癌生存期	
5 年生存期	1619 (44.1%)
总生存期	1539 (41.9%)
胰腺癌特异生存期	
5 年生存期	1801 (49.0%)
总生存期	1743 (47.4%)

Table 2. Comparison of discriminative power between survival models based on OS with full data from categorical variable cohorts**表 2.** 根据完整数据中分类变量队列基于总生存期的生存模型的区分能力比较

多变量模型	一致性指数(95% CI)	Brier 分数(95% CI)
Gradient Boosting Survival Analysis	0.733 (0.726~0.739)	0.163 (0.045~0.281)
Cox PH Survival Analysis	0.726 (0.719~0.732)	0.177 (0.038~0.316)
Cox net Survival Analysis	0.725 (0.719~0.731)	0.174 (0.036~0.312)
Component wise Gradient Boosting Survival Analysis	0.723 (0.716~0.729)	0.190 (0.101~0.280)
Tier-based models		
Random Survival Forest	0.723 (0.715~0.732)	0.161 (0.041~0.281)
Survival Trees	0.718 (0.709~0.728)	0.160 (0.042~0.279)
Extra Survival Trees	0.724 (0.717~0.731)	0.158 (0.040~0.276)

Table 3. Comparison of discriminative power between survival models based on CSS in the full data cohort of categorical variables**表 3.** 分类变量全数据队列中基于癌症特异性生存期的生存模型之间的区分能力比较

多变量模型	一致性指数(95% CI)	Brier 分数(95% CI)
Gradient Boosting Survival Analysis	0.743 (0.735~0.751)	0.162 (0.026~0.297)
Cox PH Survival Analysis	0.736 (0.727~0.745)	0.175 (0.012~0.338)
Cox net Survival Analysis	0.736 (0.727~0.746)	0.171 (0.005~0.338)
Component wise Gradient Boosting Survival Analysis	0.731 (0.721~0.741)	0.192 (0.090~0.293)
Tier-based models		
Random Survival Forest	0.735 (0.722~0.747)	0.156 (0.022~0.291)
Survival Trees	0.731 (0.718~0.743)	0.161 (0.028~0.293)
Extra Survival Trees	0.736 (0.725~0.747)	0.152 (0.020~0.285)

Table 4. Comparison of discriminative power between survival models based on CSS for full data from cohorts of continuous variables**表 4.** 根据连续变量队列的完整数据基于癌症特异性生存期的生存模型的区分能力比较

多变量模型	一致性指数(95% CI)	Brier 分数(95% CI)
Gradient Boosting Survival Analysis	0.753 (0.741~0.765)	0.142 (0.013~0.271)
Cox PH Survival Analysis	0.747 (0.738~0.757)	0.166 (-0.006~0.339)
Cox net Survival Analysis	0.747 (0.738~0.756)	0.160 (-0.016~0.337)
Component wise Gradient Boosting Survival Analysis	0.728 (0.719~0.736)	0.192 (0.091~0.294)
Tier-based models		
Random Survival Forest	0.737 (0.730~0.745)	0.168 (-0.023~0.361)
Survival Trees	0.705 (0.686~0.723)	0.208 (-0.008~0.426)
Extra Survival Trees	0.746 (0.737~0.754)	0.160 (-0.006~0.326)

Table 5. Comparison of discriminative power between survival models based on overall survival with full data from cohorts of continuous variables

表 5. 根据连续变量队列的完整数据基于总生存期的生存模型的区分能力比较

多变量模型	一致性指数(95% CI)	Brier 分数(95% CI)
Gradient Boosting Survival Analysis	0.743 (0.729~0.757)	0.146 (0.027~0.265)
Cox PH Survival Analysis	0.738 (0.730~0.746)	0.166 (0.017~0.315)
Cox net Survival Analysis	0.738 (0.730~0.746)	0.161 (0.012~0.310)
Component wise Gradient Boosting Survival Analysis	0.717 (0.711~0.723)	0.190 (0.099~0.281)
Tier-based models		
Random Survival Forest	0.727 (0.717~0.737)	0.164 (-0.005~0.334)
Survival Trees	0.697 (0.679~0.715)	0.189 (-0.004~0.383)
Extra Survival Trees	0.735 (0.726~0.744)	0.157 (0.008~0.306)

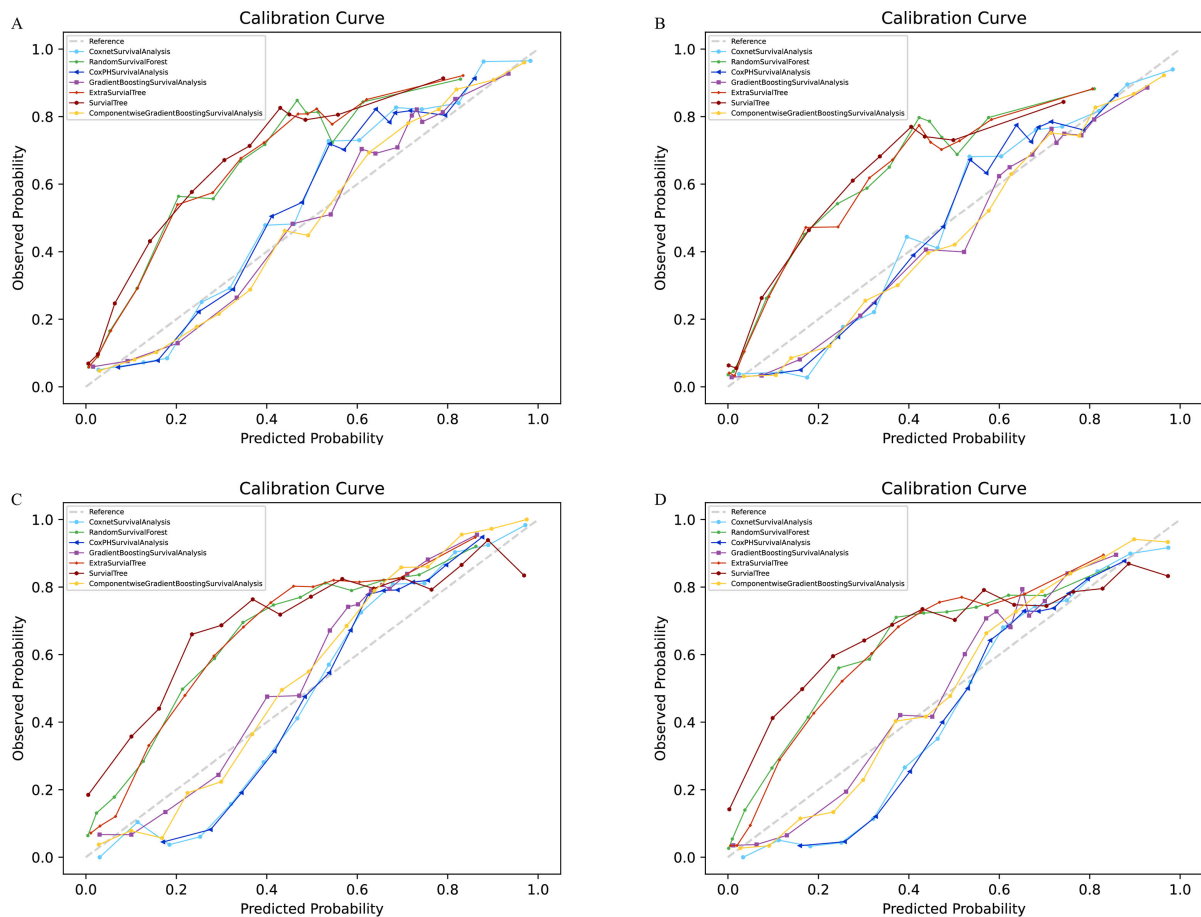


Figure 2. Calibration curves for the seven different models, namely the categorical numerical model for OS (A), the categorical numerical model for CSS (B), the continuous numerical model for OS (C), and the continuous numerical model for CSS (D)

图 2. 七种不同模型的校准曲线, 分别为 OS 的分类数值模型(A)、CSS 的分类数值模型(B)、OS 的连续数值模型(C)、以及 CSS 的连续数值模型(D)

Table 6. Based on the CoxPH model, the predictive ability of continuous variables is calculated compared with the model constructed by categorical variables

表 6. 基于 CoxPH 模型, 计算连续变量与基于分类变量构建的模型相比的预测能力

	总生存期-NRI (95 CI%)	癌症特异性生存期-NRI (95% CI)
5 年总生存期	0.305 (0.215~0.388)	0.268 (0.185~0.333)
全数据总生存期	0.313 (0.168~0.437)	0.263 (0.122~0.339)

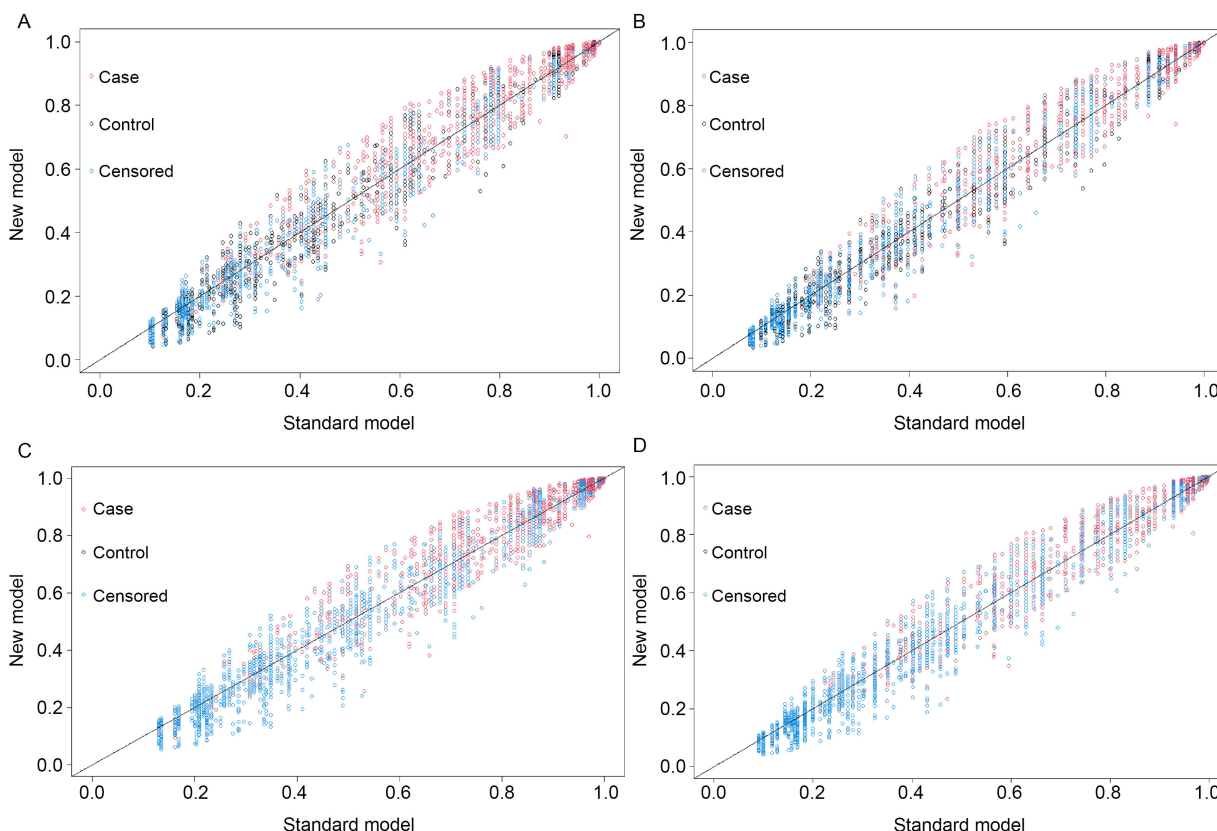


Figure 3. Scatter plot of NRI correlation in the CoxPH model with different types of variables, where the OS model and CSS model with categorical values are (A) and (B), respectively, and the OS and CSS models with continuous values are (C) and (D)

图 3. 在不同类型变量的 CoxPH 模型中, NRI 相关性的散点图, 其中具有分类值的 OS 模型和 CSS 模型分别为(A)和(B), 而具有连续值的 OS 模型和 CSS 模型分别为(C)和(D)

4. 讨论

在本分析中, 我们利用从 SEER 数据库筛选出的患者建立和训练模型, 以验证临床中最常用的模型, 并通过比较一系列模型对不同数据的区分和拟合能力来筛选出每个模型的不同特征。在开始这项研究之前, 我们意识到医学领域对于胰腺癌的研究文献中没有解释多元模型和基于层次的模型之间的相似性和差异, 而有些文章是基于某种机器学习或人工智能模型的[8]。因此, 我们希望探索基于机器学习的更多不同类型的模型特征。

过去, 机器学习更多地应用于基因组学、影像学或生物分子等微观领域[9] [10]。在临床患者的精确个体化治疗方面, 关于机器学习的应用研究较少。目前可以找到一份关于个体化治疗的报告。该报告评估了接受化疗的 50 名可手术切除的胰腺癌患者。通过收集患者的手术和病理信息, 并基于机器学习

logistic 回归方法预测个体复发的风险概率[11]。与此同时, 机器学习也在筛查胰腺癌患者的健康方面发挥了作用[5] [12]。

一般来说, 胰腺癌患者的生存时间受到疾病的病理情况、侵袭程度、接受的治疗和合并症的影响。当然, 这也包括患者的基本情况, 如年龄和营养状况[13]。为了尽量排除个体化治疗所引起的更多差异, 我们的研究未包括与治疗相关的数据, 只使用了患者的基本情况、病理分级、肿瘤分期等作为训练模型的指标, 并试图找出不同类型模型之间的差异。基于这个分析的结果, 我们的研究提供了以下结论的一些证据: 随着变量的细化和细粒度的增加, 大多数模型的区分度得到了改善; 多元模型优于基于层次的模型; 对于在临床实践中常用的 CoxPH 回归模型, 连续值可以提高模型的预测能力, 相对于分类值。

在本研究中, GBSA 展现了最佳的效果, 它是一个考虑多个特征对生存分析结果影响的多元模型。在这个模型中, 每个特征都被视为一个独立变量, 可以使用不同的权重来表示其对生存分析结果的贡献。通过结合所有特征, 该模型可以预测在给定时间内发生事件的概率。GBSA 是一个非参数模型, 不需要预先假设每个预测变量对风险比例的影响是恒定的[14]。与 CoxPH 模型等半参数模型相比, GBSA 对数据具有更强的拟合性能, 但同时也可能出现过拟合现象。因此, 在具体研究中应进行详细讨论[15]。

RSF 是一个基于层次结构的模型, 结合了随机森林和生存分析的优点, 同样不假设风险比例是恒定的[16]。在分析结果中, RSF 与多元模型在预测能力和区分度方面的整体差异并不显著。据报道, RSF 可以改善胰腺癌患者术后预后模型的预测能力[16]。在许多情况下, 多元模型比基于层次结构的模型更精确、准确[17]。然而, 在一些特殊情况下, 如数据集的分布不均匀、存在混杂因素或异常值, 基于层次结构的模型可能更具鲁棒性。考虑到术后患者的治疗效果和个体化治疗的差异, 本研究未包括患者的术后相关变量。在分析结果中, RSF 并未展现出优异的结果, 因为包含的数据都是非转移性胰腺癌患者, 没有涉及到个体化治疗。RSF 的特色在本文中可能没有用武之地, 但这不能掩盖其在构建预后模型中的地位。

除了 RSF, 还有其他基于层次的模型, 如 ST 和 EST, 它们是基于决策树的生存预测模型。EST 是 ST 的扩展和改进, 借鉴了随机森林算法中“随机特征选择”和“随机样本选择”的思想。为了减少过拟合并提高模型的泛化能力, 在构建每棵决策树时会随机选择部分特征和样本。ST 和 EST 的应用主要集中在流行病学领域[18]。数据根据特定规则进行分层, 每一层可以识别最重要的特征, 通过添加或删除一些层可以调整模型的复杂度, 因此这种模型更适用于分类数据集。在本文的分析结果中, ST 的表现比其他模型差, 特别是在构建连续数值相关模型方面, 区分度仅为 0.697 (0.679~0.715)。我们认为这可能是因为在使用 Kaplan-Meier 估计器评估每个分割点对目标变量(如生存时间)的影响时, 它只考虑将样本在每个分割点划分为不同子组后, 在某个时间点上对生存曲线的差异, 但不考虑在时间点之前和之后的整体趋势。因此, 在处理连续值时, 不恰当的分割点选择可能导致模型失去预测能力。这类问题可能通过修剪调整或设置更多停止标准, 并将连续值离散为几个类别来解决。EST 作为 ST 的扩展, 在本文数据集的区分度和预测能力方面有显著改善。自研究结束以来, 尚未有关于 EST 的研究报告, 现有的大多数文章极少提及 EST, 其潜力仍需要进一步探索[19]。

目前, CwGB 模型尚未被发现在临床实践中得到广泛应用。它与 GBSA 一样是一个基于梯度的算法, 但略有不同。CwGB 采用逐渐增强每个协变量对目标变量的影响的方法, 在每次迭代中只选择一个最优特征进行学习和调整系数, 因此其模型更容易收敛, 比 GBSA 更稳定。然而, 在结果中, CwGB 在性能方面表现并不理想, 可能考虑到在训练集训练模型时, CwGB 将不断选择最佳特征并增加它们的系数, 以使模型在训练集上表现良好, 但可能导致泛化能力差, 对验证集的结果不佳。因此, 如果数据集维度较高或需要考虑多个变量对目标变量的总体影响, 可以考虑使用这个模型, 但本文的数据集较小, 训练变量较少, 因此 GBSA 能更灵活地处理这些特征。

作为临床实践中最常用的生存分析方法, CoxPH 模型以其灵活性和兼容性而闻名。它对连续变量和分类变量都具有良好的拟合能力[20]。CnSA 模型是一种基于 LASSO (最小绝对值收缩和选择运算符) 惩罚项的生存分析方法, 它可以自动选择重要变量并建立预测模型。该模型在 CoxPH 上添加了一个正则化项, 以防止过拟合。在模型区分度和预测能力的比较中, 两者几乎没有太大区别。鉴于 CoxPH 的高兼容性和灵活性, 为了进一步研究连续值和分类值对模型拟合的影响, 我们选择了 CoxPH 模型来计算不同数值类型的 NRI。近年来, NRI 已广泛应用于临床研究中, 如糖尿病、血管疾病和肿瘤疾病等, 用于衡量模型之间的改善程度[21] [22]。在本文中, 基于连续值构建的模型在 5 年 OS 和 CSS, 以及完整数据的 OS 和 CSS 方面都显示出显著的改善。因此, 在未来相关研究中, 变量指标的细化和细粒度的增加可能会重新提升一些风险分层的效果。

本研究的数据全部来自现实世界, 由大型、开放且可靠的数据库收集。在研究分析中, 我们仍然存在一些不足之处: 1) 本研究中大量队列倾向于早期疾病; 2) 我们无法获取更多的大量人群; 3) 未包括患者的个体化治疗信息; 4) 包含的指标变量较少。尽管存在这些限制, 但我们决定完成这项研究, 以探索基于偏局部标量的临床常用生存分析方法在胰腺癌患者中的应用。因为我们的目的是比较基于不同类型数据的不同类型模型, 从而为临床治疗提供一些参考。机器学习代表了未来的生存分析框架, 通过模型的选择和迭代, 为肿瘤患者提供更精细、更准确、更合理的个体化治疗是未来治疗的发展方向。

参考文献

- [1] Halbrook, C., Lyssiotis, C., Pasca di Magliano, M. and Maitra, A. (2023) Pancreatic Cancer: Advances and Challenges. *Cell*, **186**, 1729-1754. <https://doi.org/10.1016/j.cell.2023.02.014>
- [2] Faur, A., Lazar, D. and Ghenciu, L. (2023) Artificial Intelligence as a Noninvasive Tool for Pancreatic Cancer Prediction and Diagnosis. *World Journal of Gastroenterology*, **29**, 1811-1823. <https://doi.org/10.3748/wjg.v29.i12.1811>
- [3] Jan, Z., El Assadi, F., Abd-Alrazaq, A. and Jithesh, P. (2023) Artificial Intelligence for the Prediction and Early Diagnosis of Pancreatic Cancer: Scoping Review. *Journal of Medical Internet Research*, **25**, e44248. <https://doi.org/10.2196/44248>
- [4] Wang, L., Liu, Z., Liang, R., Wang, W., Zhu, R., Li, J., et al. (2022) Comprehensive Machine-Learning Survival Framework Develops a Consensus Model in Large-Scale Multicenter Cohorts for Pancreatic Cancer. *eLife*, **11**, e80150. <https://doi.org/10.7554/eLife.80150>
- [5] Chen, W., Zhou, Y., Xie, F., Butler, R., Jeon, C., Luong, T., et al. (2023) Derivation and External Validation of Machine Learning-Based Model for Detection of Pancreatic Cancer. *The American Journal of Gastroenterology*, **118**, 157-167. <https://doi.org/10.14309/ajg.0000000000002050>
- [6] Kernbach, J. and Staartjes, V. (2020) Predicted Prognosis of Pancreatic Cancer Patients by Machine Learning-Letter. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, **26**, 3891. <https://doi.org/10.1158/1078-0432.CCR-20-0523>
- [7] Hayward, J., Alvarez, S., Ruiz, C., Sullivan, M., Tseng, J. and Whalen, G. (2010) Machine Learning of Clinical Performance in a Pancreatic Cancer Database. *Artificial Intelligence in Medicine*, **49**, 187-195. <https://doi.org/10.1016/j.artmed.2010.04.009>
- [8] Yokoyama, S., Hamada, T., Higashi, M., Matsuo, K., Maemura, K., Kurahara, H., et al. (2020) Predicted Prognosis of Patients with Pancreatic Cancer by Machine Learning. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, **26**, 2411-2421. <https://doi.org/10.1158/1078-0432.CCR-19-1247>
- [9] Toyama, Y., Hotta, M., Motoi, F., Takanami, K., Minamimoto, R. and Takase, K. (2020) Prognostic Value of FDG-PET Radiomics with Machine Learning in Pancreatic Cancer. *Scientific Reports*, **10**, Article No. 17024. <https://doi.org/10.1038/s41598-020-73237-3>
- [10] Iwatate, Y., Hoshino, I., Yokota, H., Ishige, F., Itami, M., Mori, Y., et al. (2020) Radiogenomics for Predicting p53 Status, PD-L1 Expression, and Prognosis with Machine Learning in Pancreatic Cancer. *British Journal of Cancer*, **123**, 1253-1261. <https://doi.org/10.1038/s41416-020-0997-1>
- [11] Sala Elarre, P., Oyaga-Iriarte, E., Yu, K., Baudin, V., Arbea Moreno, L., Carranza, O., et al. (2019) Use of Machine-Learning Algorithms in Intensified Preoperative Therapy of Pancreatic Cancer to Predict Individual Risk of Relapse. *Cancers*, **11**, Article No. 606. <https://doi.org/10.3390/cancers11050606>

-
- [12] Malhotra, A., Rachet, B., Bonaventure, A., Pereira, S. and Woods, L. (2021) Can We Screen for Pancreatic Cancer? Identifying a Sub-Population of Patients at High Risk of Subsequent Diagnosis Using Machine Learning Techniques Applied to Primary Care Data. *PLOS ONE*, **16**, e0251876. <https://doi.org/10.1371/journal.pone.0251876>
- [13] Klatte, D., Boekestijn, B., Onnekink, A., Dekker, F., van der Geest, L., Wasser, M., *et al.* (2023) Surveillance for Pancreatic Cancer in High-Risk Individuals Leads to Improved Outcomes: A Propensity Score-Matched Analysis. *Gastroenterology*, **23**, e17-e18. <https://doi.org/10.1016/j.pan.2023.06.446>
- [14] Khajehpiri, B., Moghaddam, H., Forouzanfar, M., Lashgari, R., Ramos-Cejudo, J., Osorio, R., *et al.* (2022) Survival Analysis in Cognitively Normal Subjects and in Patients with Mild Cognitive Impairment Using a Proportional Hazards Model with Extreme Gradient Boosting Regression. *Journal of Alzheimer's Disease: JAD*, **85**, 837-850. <https://doi.org/10.3233/JAD-215266>
- [15] Chen, Y., Jia, Z., Mercola, D. and Xie, X. (2013) A Gradient Boosting Algorithm for Survival Analysis via Direct Optimization of Concordance Index. *Computational and Mathematical Methods in Medicine*, **2013**, Article ID: 873595. <https://doi.org/10.1155/2013/873595>
- [16] Lin, J., Yin, M., Liu, L., Gao, J., Yu, C., Liu, X., *et al.* (2022) The Development of a Prediction Model Based on Random Survival Forest for the Postoperative Prognosis of Pancreatic Cancer: A SEER-Based Study. *Cancers*, **14**, Article No. 4667. <https://doi.org/10.3390/cancers14194667>
- [17] Qiu, X., Gao, J., Yang, J., Hu, J., Hu, W., Kong, L., *et al.* (2020) A Comparison Study of Machine Learning (Random Survival Forest) and Classic Statistic (Cox Proportional Hazards) for Predicting Progression in High-Grade Glioma after Proton and Carbon Ion Radiotherapy. *Frontiers in Oncology*, **10**, Article ID: 551420. <https://doi.org/10.3389/fonc.2020.551420>
- [18] Tapak, L., Kosorok, M., Sadeghifar, M. and Hamidi, O. (2018) Multistate Recursively Imputed Survival Trees for Time-to-Event Data Analysis: An Application to AIDS and Mortality Post-HIV Infection Data. *BMC Medical Research Methodology*, **18**, Article No. 129. <https://doi.org/10.1186/s12874-018-0596-5>
- [19] Tran, T., Lee, J., Gunathilake, M., Kim, J., Kim, S., Cho, H., *et al.* (2023) A Comparison of Machine Learning Models and Cox Proportional Hazards Models Regarding Their Ability to Predict the Risk of Gastrointestinal Cancer Based on Metabolic Syndrome and Its Components. *Frontiers in Oncology*, **13**, Article ID: 1049787. <https://doi.org/10.3389/fonc.2023.1049787>
- [20] Lee, E. and Go, O. (1997) Survival Analysis in Public Health Research. *Annual Review of Public Health*, **18**, 105-134. <https://doi.org/10.1146/annurev.publhealth.18.1.105>
- [21] Foussard, N., Saulnier, P., Potier, L., Ragot, S., Schneider, F., Gand, E., *et al.* (2020) Relationship between Diabetic Retinopathy Stages and Risk of Major Lower-Extremity Arterial Disease in Patients with Type 2 Diabetes. *Diabetes Care*, **43**, 2751-2759. <https://doi.org/10.2337/dc20-1085>
- [22] Dong, D., Tang, L., Li, Z., Fang, M., Gao, J., Shan, X., *et al.* (2019) Development and Validation of an Individualized Nomogram to Identify Occult Peritoneal Metastasis in Patients with Advanced Gastric Cancer. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, **30**, 431-438. <https://doi.org/10.1093/annonc/mdz001>