

基于机器学习的胰腺癌患者预后生存分析模型

付文斌, 易静*

重庆医科大学公共卫生学院, 重庆

收稿日期: 2026年1月24日; 录用日期: 2026年2月14日; 发布日期: 2026年2月25日

摘要

目的: 针对胰腺癌预后预测中传统回归模型处理高维非线性特征受限, 以及深度学习缺乏临床解释性的问题, 构建一种高精度且具有强解释性的生存分析模型。方法: 基于591例患者数据, 选取XGBoost、SVM、MLP与CNN作为基础模型。提出自适应加权与分层融合的双阶段策略进行集成优化, 并利用SHAP技术对24个临床特征进行量化解释分析。结果: 该模型准确率达90.0%, AUC为0.90, 召回率89.0%, 相比固定权重融合提升了2%。自适应加权与层次化融合分别贡献了1.5%和0.5%的效能增幅。年龄、医疗可及性及加工食品饮食是影响预后的关键因素。结论: 该方法有效解决了单一模型在复杂疾病预测中的不足, 在保证高准确率的同时实现了临床可解释性, 为制定个性化治疗方案提供了可靠依据。

关键词

胰腺癌, 机器学习, 自适应加权, SHAP

A Machine Learning-Based Model for Prognostic Survival Analysis in Pancreatic Cancer Patients

Wenbin Fu, Jing Yi*

College of Public Health, Chongqing Medical University, Chongqing

Received: January 24, 2026; accepted: February 14, 2026; published: February 25, 2026

Abstract

Objective: To address the limitations of traditional regression models in handling high-dimensional non-linear features and the lack of clinical interpretability in deep learning models, this study aims to construct a survival analysis model for pancreatic cancer that achieves both high precision and

*通讯作者。

strong interpretability. Methods: Based on data from 591 patients, XGBoost, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN) were selected as base models. A two-stage ensemble optimization strategy was proposed, incorporating adaptive weighting and hierarchical fusion. Furthermore, SHAP (SHapley Additive exPlanations) technology was employed to quantitatively interpret 24 clinical features. **Results:** The proposed model achieved an accuracy of 90.0%, an Area Under the Curve (AUC) of 0.90, and a recall rate of 89.0%, representing a 2% improvement over fixed-weight fusion methods. Specifically, adaptive weighting and hierarchical fusion contributed to performance gains of 1.5% and 0.5%, respectively. Feature analysis identified age, healthcare accessibility, and processed food intake as critical prognostic factors. **Conclusion:** This approach effectively overcomes the limitations of single models in complex disease prediction. By ensuring high accuracy while maintaining clinical interpretability, the model provides a reliable basis for developing personalized treatment plans.

Keywords

Pancreatic Cancer, Machine Learning, Adaptive Weighting, SHAP

Copyright © 2026 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. 引言

胰腺癌(Pancreatic Cancer)具有高度的侵袭性和早期隐匿性, 被公认为消化道恶性肿瘤中的“癌中之王”。据最新的全球癌症统计数据显示, 胰腺癌是癌症导致死亡的主要病种之一, 其五年总生存率(OS)长期滞留于 10% 以下[1]-[3]。由于缺乏特异性的早期筛查手段, 约 80% 的患者在确诊时已处于局部晚期或发生远处转移, 从而错失了最佳的手术切除窗口[4]。面对这一严峻的临床挑战, 构建具有早期精确风险评估功能的预后模型, 对于识别高危人群、推行个性化医疗及优化医疗资源配置具有重大的现实意义[5]。

在生存分析领域, 基于线性假设的传统统计学方法, 如 Cox 比例风险回归模型(Cox Proportional Hazards Model), 长期以来被视为“金标准”。然而, 随着精准医疗的发展, 基因组学、影像组学及电子健康记录(EHR)等高维异质性数据呈爆发式增长, 传统方法在处理这些复杂数据时往往捉襟见肘, 难以有效捕捉变量间复杂的非线性交互效应[6] [7]。近年来, 人工智能技术的飞速发展使得深度学习算法在肿瘤预后预测任务中展现了超越传统方法的卓越性能[8] [9]。尽管深度学习表现优异, 但其固有的“黑盒”性质导致决策过程缺乏透明度。临床医生难以理解模型“如何得出结论”以及具体是哪些特征在驱动预测结果, 这种可解释性的缺失严重阻碍了该技术在临床诊疗中的信任构建与实际应用[10] [11]。

为了解决上述在预测精度与临床可解释性之间难以兼顾的难题, 本文提出了一种高精度且具有强解释性的胰腺癌预后生存分析框架。该框架采用了多模型集成策略, 通过引入自适应权重分配机制与分层融合结构, 在确保模型泛化能力与系统稳定性的同时, 显著提升了预测性能[12]。此外, 本研究引入了 SHAP (SHapley Additive Explanations) 等前沿博弈论解释方法, 对模型输出进行量化归因, 直观展示了关键医学特征对预后的影响机理及其内在关联, 为临床辅助决策提供了可靠的依据[13]。

2. 对象与方法

2.1. 研究对象

在 SEER (The Surveillance, Epidemiology, and End Results) 数据库中筛选符合资料的胰腺癌患者数据,

探讨胰腺癌患者的预后生存模型。本次研究一共筛选了 591 例样本, 包括性别, 年龄, 生存时间, 癌症分期和其他一些相关指标。

2.2. 研究方法

首先初步筛选所有符合条件的数据, 并对收集到的数据按照纳排标准进行进一步, 删除缺失值比例严重的数据并对所有的异常值进行处理, 最终一共得到符合条件的样本 591 例。采用 R 中的 `mice` 包进行多重插补填补缺失数据, 随后, 对所有分类特征实施独热编码, 将其转换为机器学习算法可识别的欧几里得空间向量。为了消除不同量纲特征对模型收敛速度和权重分配的负面影响, 本研究采用 Z-score 标准化对所有连续变量进行无量纲化处理。

为全面探索潜在关联模式, 选取四种典型机器学习方法作为基础模型进行部署。包括极端梯度提升树(XGBoost)、支持向量机(SVM)、前馈神经网络(MLP), 针对非结构化的表格数据, 本研究采用了一维卷积神经网络(CNN)架构。具体而言, 我们将经过预处理的 24 个临床特征映射为特征向量序列 $X = [X_1, X_2, \dots, X_{24}]$ 。通过设定卷积核大小为 3, 步长为 1, 在特征序列上进行滑动卷积操作, 以捕获不同临床指标之间潜在的局部依赖关系与非线性组合特征。随后接入最大池化层以提取显著特征并降低维度, 最终通过全连接层输出预测概率。为了突破单一模型的性能瓶颈, 本研究提出了一种创新性的双阶段融合算法框架。第一阶段: 设基础模型集合为 $M = \{m_1, m_2, m_3, m_4\}$, 对应 XGBoost、SVM、MLP 与 CNN。对于验证集中的样本 x , 定义模型 m_k 的性能评分 S_k (基于 AUC 值)。我们引入 Softmax 函数动态计算各模型的置信度权重 α_k , 计算公式如下:

$$\alpha_k = \frac{e^{\lambda S_k}}{\sum_{j=1}^4 e^{\lambda S_j}} \quad (1)$$

其中 λ 为温度系数, 用于调节权重分布的平滑度。自适应加权集成基于验证集上的性能表现(如 AUC 值), 动态计算各基础模型的置信度权重。通过加权平均策略生成初始集成预测结果。该步骤旨在利用不同模型的优势互补, 降低整体预测方差。第二阶段: 非线性元学习融合。为进一步挖掘高阶特征交互效应, 本研究引入了一个轻量级多层感知机作为元学习器, 元学习器包含两层隐藏层(节点数分别为 32 与 16, 激活函数为 ReLU), 输出层采用 Sigmoid 函数。模型训练采用带有 L_2 正则化的二元交叉熵损失函数(Binary Cross-Entropy Loss), 以防止过拟合:

$$L = -\frac{1}{N} \sum_{i=1}^N \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right] + \beta \|W\|^2 \quad (2)$$

其中 N 为批量样本数, β 为正则化系数。不同于传统的 Stacking 策略, 我们将第一阶段的初始预测概率与原始临床特征向量进行拼接, 共同输入元学习器。这种设计不仅保留了原始数据的物理意义, 还允许模型对初步预测结果进行非线性校准, 从而在保持高泛化能力的同时显著提升分类精度。

最后, 尽管深度集成模型提升了预测性能, 但其决策过程往往缺乏透明度。为解决这一“黑盒”问题, 本研究引入了 SHAP 框架。基于合作博弈论, SHAP 值能够量化每个特征对最终预测结果的边际贡献。本研究计算了所有样本的 SHAP 值, 通过全局解释识别影响患者生存预后的关键风险因子(如关键生物标志物或临床特征), 并通过依赖图(Dependence Plots)分析特定生理变量在“生存”与“死亡”判定过程中的作用阈值及其与其他变量的交互规律。这为验证模型的临床合理性及辅助医生制定个性化诊疗方案提供了直观依据。

本研究所有分析均依托于 R 4.4.0 来完成。

3. 结果

3.1. 患者生存情况

图 1 展示了胰腺癌不同阶段患者的平均生存时间(单位: 月)。从图中可以看出, 早期的第一阶段(Stage I)具有较长的生存时间, 约为 35 个月, 而随着癌症进展到第二阶段(Stage II), 生存时间显著下降至约 20 个月。进入第三阶段(Stage III)和第四阶段(Stage IV)后, 生存时间进一步缩短, 第四阶段的患者生存时间大约为 5 个月。这一分析表明, 胰腺癌的生存期与其诊断阶段密切相关, 早期发现和治疗对于提高生存率至关重要。

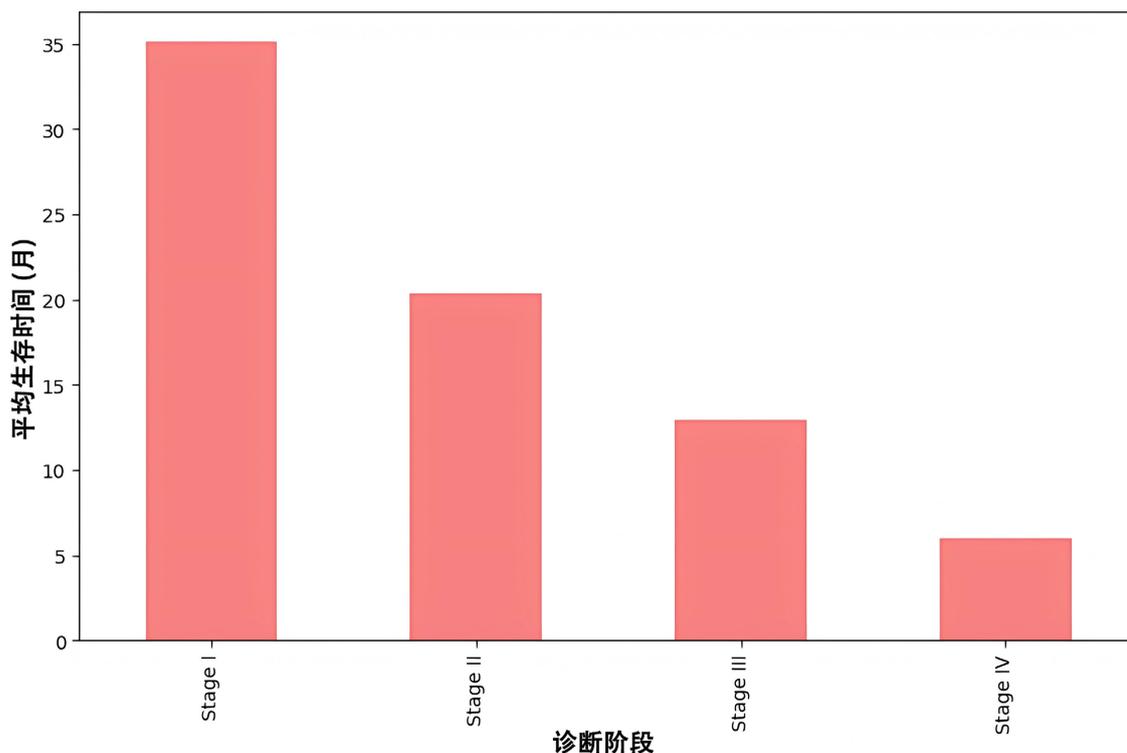


Figure 1. Survival time of pancreatic cancer patients at different stages

图 1. 不同阶段胰腺癌患者生存时间

3.2. 模型结果

表 1 展示了各个模型及其融合策略在准确率、精确度、召回率和 F1-score 上的实验结果。从表中可以看出, 自适应加权与层次化融合策略在所有评估指标上均优于其他单一模型和传统加权融合方法, 进一步验证了所提方法的有效性和优势。其中, XGBoost 模型在准确率(0.85)、精确度(0.83)和召回率(0.84)方面表现较好, 但由于其基于单一决策树, 面对复杂的非线性数据时, 可能无法充分挖掘数据中的深层次特征, 导致 F1-score 为 0.83, 略低于其他方法。支持向量机(SVM)模型在准确率(0.80)、精确度(0.76)和召回率(0.72)上均表现较为一般, 尤其是召回率较低, 表明在预测少数类样本(如生存)时存在困难。相比之下, 深度学习模型如多层感知机(MLP)和卷积神经网络(CNN)则展现了更强的预测能力, MLP 在准确率(0.86)、精确度(0.84)和 F1-score (0.84)上都表现优异, CNN 稍微提高, 准确率为 0.87, F1-score 为 0.85, 进一步证明了深度学习模型在处理复杂特征时的优势。当使用固定加权的融合方法时, 模型的表现有所提升, 准确率达到 0.88, F1-score 为 0.86。这一方法通过固定权重将多个模型的输出进行加权, 能够结

合各模型的优点,但由于权重无法根据模型的具体表现动态调整,仍然存在一定的局限性。自适应加权与层次化融合策略(本研究所提方法)在所有评估指标中表现最佳。该方法的准确率达到0.90, F1-score为0.88,召回率为0.89,精确度为0.88。自适应加权策略通过动态调整各个模型的权重,使得每个模型在不同任务中能够贡献最佳表现;而层次化融合机制通过多阶段的联合学习进一步优化模型,提升了整体预测性能。此外,融合策略在数据不平衡的情况下,特别是在少数类(如生存样本)的预测上表现更为突出,显著提高了召回率,解决了类别不平衡问题。从实验结果来看,自适应加权与层次化融合策略显著提高了预测的准确性和稳定性,特别是在处理复杂和高维数据时,能够更好地发挥不同模型的优势。该方法不仅提高了整体预测效果,还增强了模型的泛化能力,证明了其在胰腺癌生存状态预测中的有效性。

Table 1. Performance metrics of the machine learning models

表 1. 机器学习模型的性能指标

模型	准确率	精确度	召回率	F1-score
XGBoost	0.85	0.83	0.84	0.83
SVM	0.80	0.76	0.72	0.74
MLP	0.86	0.84	0.85	0.84
CNN	0.87	0.85	0.86	0.85
固定加权融合	0.88	0.86	0.87	0.86
自适应加权 + 层次化融合	0.90	0.88	0.89	0.88

3.3. 消融实验

通过三种消融实验验证框架组件贡献:(i) 比较去适应、层次化对四模型的影响;(ii) 评估传统模型的性能;(iii) 分析深度模型的效果。表 1 单模型结果显示,三类设置的计算过程无额外假设且具有可追溯性。

“固定加权融合”与“自适应加权 + 层次化融合”并列展示用于对比。表 2 揭示使用传统模型时,深度模型四项指标下滑,凸显深度特征在优化判别边界上的关键作用。虽然优于传统模型,但性能仍不及固定加权融合,表明集成互补假设空间,可增强性能。数据驱动估权在均匀与固定加权中展现稳健性优势,优于传统模型且次于固定加权,无需自适应调整。多模型叠加分析表明,采用自适应权重与二层非线性重组策略在四项指标上平均提升约 0.02 分,效果显著。

Table 2. Results of the ablation study

表 2. 消融实验结果

消融/子集设置	准确率	精确度	召回率	F1-score
均匀加权	0.85	0.82	0.82	0.82
仅传统模型	0.83	0.80	0.78	0.79
仅深度模型	0.87	0.85	0.86	0.85
固定加权融合	0.88	0.86	0.87	0.86
自适应加权 + 层次化融合	0.90	0.88	0.89	0.88

3.4. 变量重要性

图 2 展示了不同特征在胰腺癌生存状态预测中的重要性。通过分析图中的条形图,可以看到,年龄是最重要的特征,其重要性值为 0.175,显著高于其他特征,表明年龄在胰腺癌生存状态的预测中占据了

核心地位, 可能与患者的预后密切相关。紧随其后的是医疗可及性(Access to Healthcare)和加工食品饮食(Diet Processed Food), 它们的重要性值分别为 0.12 和 0.11, 表明患者是否能方便地获得医疗服务以及饮食习惯, 特别是摄入加工食品的情况, 对胰腺癌的预测有着重要影响。此外, 体力活动水平(Physical Activity Level)和体重下降(Weight Loss)这两个特征的重要性也较为突出, 分别为 0.09 和 0.08, 表明这些生理和生活习惯因素与胰腺癌患者的生存状态存在一定关联。吸烟历史(Smoking History)和腹部不适(Abdominal Discomfort)等特征的贡献度也不可忽视, 分别为 0.07 和 0.06, 揭示了生活方式和临床症状对生存预测的重要性。虽然一些特征如酒精消费(Alcohol Consumption)、背痛(Back Pain)、肥胖(Obesity)、2 型糖尿病的发展(Development of Type2 Diabetes)、美国国籍(Country United States)、诊断阶段 IV (Stage at Diagnosis Stage IV)、家族病史(Family History)、以及糖尿病(Diabetes)的权重较低, 但它们在胰腺癌的生存状态预测中仍然起到了一定的作用。总体而言, 图中的分析有助于我们识别出哪些临床特征和生活方式因素对胰腺癌患者生存状态的预测至关重要, 为未来的临床决策和治疗策略提供了有价值的参考。

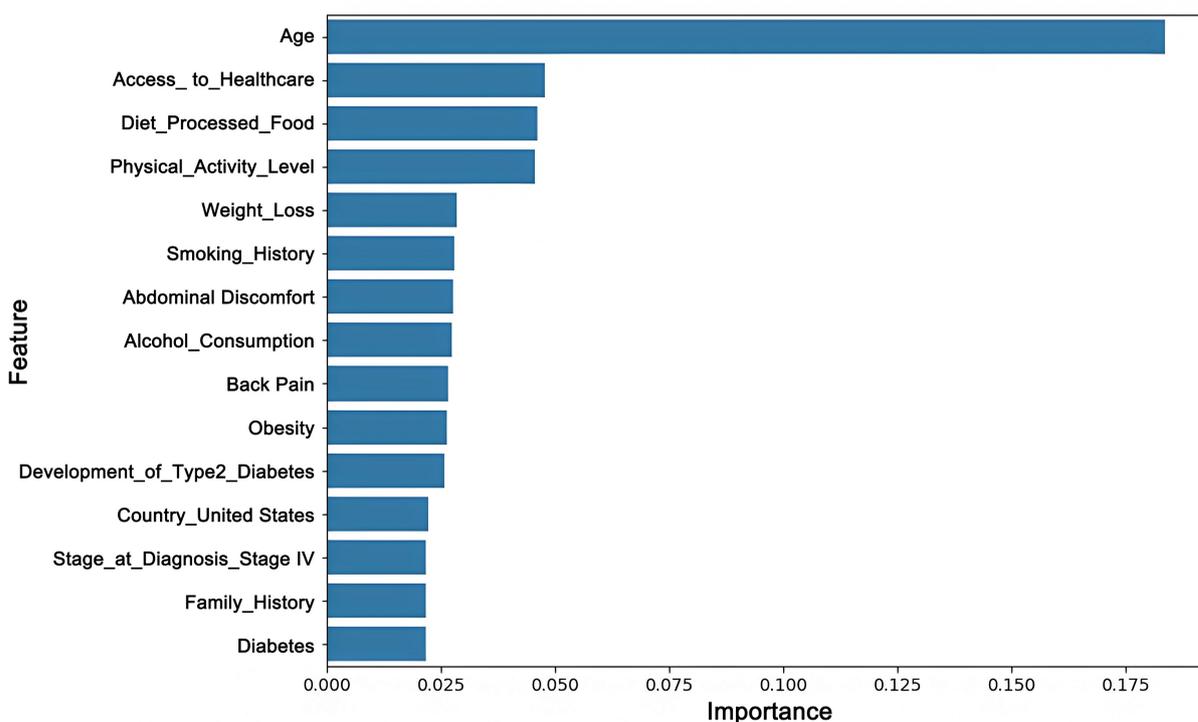


Figure 2. Feature importance ranking

图 2. 变量重要性排序

4. 讨论

本文提出一种新的用于预测胰腺癌患者生存状况的自适应加权、层次化融合多模型集成方法。综合使用多种机器学习模型来完成对胰腺癌患者预后风险的评估。其主要思想就是采用自适应加权的方法, 根据不同的模型对某一个数据集拟合程度来分配权重, 从而达到最好的预测效果。层次融合策略可以进一步提高预测的准确性, 在不同层次上做模型的集成, 既考虑单个模型的决策, 又考虑模型之间信息的交互, 从而得到更加全面、精确的预测结果。该方法目的在于改善单一模型面对复杂疾病预测时存在的不足, 给临床医生提供更加可靠的患者生存状况预估手段, 进而帮助制定个体化的治疗计划, 优化患者的生存率和生活品质。这个模型在实际运用中效果很好, 在准确率、F1 这些主要的评价指标上也得到了

非常好的成绩。在分类任务上表现出了很强的分类能力, 对于数据集的变化具有很强的稳定性、抗干扰性, 充分表现出很高的鲁棒性, 可以提供可靠的预测支持。研究发现自适应加权机制与多层融合策略结合后, 整体表现有很大程度的提升, 在消融实验中可以发现两者结合能够显著提升决策准确度, 在不同样本人群中的预测效果更加明显, 为解决实际问题提供了强有力的技术支持。使用动态权重调整的方式把多源数据融合成多层次的集成模型, 综合性能指标得到了极大的提升。该方法给后续的深度学习算法优化设计提供了重要的理论依据。SHAP 值分析表明, 在临床决策中, “年龄” 是重要的长期健康评价因素, 对疾病的危险性预测、治疗方案的选定、预后判断等具有较大影响。医疗资源是否容易获得会影响病人得到好的医疗服务, 也会对治疗效果和健康管理质量产生很大影响。高糖食品消费模式与多种慢性病发病率呈正相关, 食品摄入量的增减会直接影响这些疾病流行趋势的发展。

本文重点针对临床决策支持系统的核心要素, 即个体年龄、医疗资源获取能力、高加索饮食模式等方面展开讨论, 并对它们在改善诊疗途径和提升健康效果上的功能做了详尽的分析。从研究结果上看, 在跨族群、多中心等环境下使用该模型的时候, 容易出现泛化性能受限制的技术难题, 进而造成算法精度衰退, 这时要采用数据预处理的方法来克服这些问题。从实验结果可以看出, 现有的模型普遍存在着迁移学习效率低的问题, 这是因为训练样本集的地理分布不均衡、代表性不足, 不能很好地反映真实世界中复杂的情境特征。为了摆脱困境, 本文从理论上分析了相关的机制, 并给出了相应的改进措施, 以增强模型对于多元文化背景的适应能力, 提高其实践意义。本文主要研究了二分类预测模型在临床环境中的有效性, 判断患者生存状态, 使医生和医疗团队可以快速判断病人是否存活, 不必计算具体的存活时间, 给治疗决策、资源分配、病人关怀等提供重要的依据。尽管 SHAP 方法对提升对机器学习模型中特征作用重要性的理解有着很大的帮助, 它将局部解释与全局模型复杂性结合起来, 从而量化出每个特征在预测结果中所起到的边际作用。但是当应用于深度学习模型时, SHAP 方法在给出直观易懂的决策解析上仍有不足。深度学习模型因为结构复杂、非线性, 具有明显的黑箱性, 内部决策机理很难解释。虽然 SHAP 方法可以提供对某一个预测结果的局部解释能力, 但是要全面理解整套模型的行为模式和内在逻辑联系还存在许多困难。这一难题是由于 SHAP 值的计算基于 Shapley 值理论框架, 在高维数据以及多层网络的环境下存在明显的不足。尤其不能直观地反映出核心驱动因素对于全局输出的影响程度, 从而限制了它在深度学习领域中的实际效果。为提高医疗诊断系统的性能, 提高稳定性, 本文提出了一种新的解决方案, 即采用强化学习、深层神经网络和生存分析技术, 设计出一种专门用来精确推断患者预期寿命的新的算法架构。该集成算法不但能大幅度提高模型的泛化能力, 使模型在新数据集上也能有较高的预测精度, 还能显著降低计算开销, 在资源有限的情况下依然可以高效运行。本文使用可解释性技术手段来提高模型决策过程透明度和可信度, 使临床医生对推理过程有更深刻的理解, 从而提高模型的应用价值。这一创新方案或许会使医学领域出现重大革新, 精准医疗方向达成个性化治疗目标, 对于健康产业来说将会产生长久的影响。基于多模态异质数据的支撑, 可以构建出高度个性化的诊断与干预体系, 从而改善患者的预后情况, 推动个体化医疗的发展进程, 加快基础研究成果转化为临床应用的速度。

参考文献

- [1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **71**, 209-249. <https://doi.org/10.3322/caac.21660>
- [2] Siegel, R.L., Miller, K.D., Wagle, N.S. and Jemal, A. (2023) Cancer Statistics, 2023. *CA: A Cancer Journal for Clinicians*, **73**, 17-48. <https://doi.org/10.3322/caac.21763>
- [3] 张思维, 郑荣寿, 孙可欣, 等. 2016 年中国恶性肿瘤分地区发病和死亡估计: 基于人群的肿瘤登记数据分析[J].

- 中国肿瘤, 2023, 32(5): 321-332.
- [4] Mizrahi, J.D., Surana, R., Valle, J.W. and Shroff, R.T. (2020) Pancreatic cancer. *The Lancet*, **395**, 2008-2020. [https://doi.org/10.1016/s0140-6736\(20\)30974-0](https://doi.org/10.1016/s0140-6736(20)30974-0)
- [5] Topol, E.J. (2019) High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, **25**, 44-56. <https://doi.org/10.1038/s41591-018-0300-7>
- [6] Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T. and Kluger, Y. (2018) Deepsurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network. *BMC Medical Research Methodology*, **18**, Article No. 24. <https://doi.org/10.1186/s12874-018-0482-1>
- [7] Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V. and Fotiadis, D.I. (2015) Machine Learning Applications in Cancer Prognosis and Prediction. *Computational and Structural Biotechnology Journal*, **13**, 8-17. <https://doi.org/10.1016/j.csbj.2014.11.005>
- [8] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., *et al.* (2019) A Guide to Deep Learning in Healthcare. *Nature Medicine*, **25**, 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
- [9] Teng, B., Zhang, X., Ge, M., Miao, M., Li, W. and Ma, J. (2024) Personalized Three-Year Survival Prediction and Prognosis Forecast by Interpretable Machine Learning for Pancreatic Cancer Patients: A Population-Based Study and an External Validation. *Frontiers in Oncology*, **14**, Article ID: 1488118. <https://doi.org/10.3389/fonc.2024.1488118>
- [10] Tjoa, E. and Guan, C. (2021) A Survey on Explainable Artificial Intelligence (XAI): Toward Medical Xai. *IEEE Transactions on Neural Networks and Learning Systems*, **32**, 4793-4813. <https://doi.org/10.1109/tnnls.2020.3027314>
- [11] Rudin, C. (2019) Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, **1**, 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- [12] Sagi, O. and Rokach, L. (2018) Ensemble Learning: A Survey. *WIREs Data Mining and Knowledge Discovery*, **8**, e1249. <https://doi.org/10.1002/widm.1249>
- [13] Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., *et al.* (2018) Explainable Machine-Learning Predictions for the Prevention of Hypoxaemia during Surgery. *Nature Biomedical Engineering*, **2**, 749-760. <https://doi.org/10.1038/s41551-018-0304-0>