

机器学习法在生存分析中的应用研究

刘玥^{1,2}, 刘启玲¹, 苏海霞^{2,3}, 杨鹏^{2,3}综述; 张玉海^{2,3} 审校

1. 陕西中医药大学公共卫生学院, 陕西 咸阳 712046; 2. 空军军医大学军事预防医学系, 陕西 西安 710032;
3. 特殊作业环境危害评估与防治教育部重点实验室, 陕西 西安 710032

摘要: 生存分析广泛应用于医学研究领域, Cox 比例风险模型是最常用的生存分析模型之一, 但实际应用受到限制。机器学习法在非线性数据处理和预测准确度方面能够弥补 Cox 比例风险模型的不足。本文对以神经网络为代表的机器学习法在生存分析领域的研究进展进行综述, 重点介绍 DeepSurv、Deep-Hit 和随机生存森林 3 种机器学习生存分析模型的原理和优势, 为复杂生存资料的分析提供方法学参考。

关键词: 生存分析; 机器学习; DeepSurv; Deep-Hit; 随机生存森林

中图分类号: R181.1 **文献标识码:** A **文章编号:** 2096-5087 (2024) 06-0496-06

Application of machine learning method for survival analysis

LIU Yue^{1,2}, LIU Qiling¹, SU Haixia^{2,3}, YANG Peng^{2,3}, ZHANG Yuhai^{2,3}

1. School of Public Health, Shaanxi University of Chinese Medicine, Xi'an, Shaanxi 712046, China; 2. Department of Military Preventive Medicine, Air Force Medical University, Xi'an, Shaanxi 710032, China; 3. Ministry of Education Key Laboratory of Environmental Hazard Assessment and Prevention in Special Operation, Xi'an, Shaanxi 710032, China

Abstract: Survival analysis has been widely used in the field of medical research. The Cox proportional hazard model is commonly used, but its practical application is limited. Machine learning method can compensate for the shortcomings of the Cox proportional hazard model in terms of nonlinear data processing and prediction accuracy. This article reviewed the advance of machine learning methods represented by neural networks, within the field of survival analysis, and highlighted the principles and benefits of three machine learning methods that DeepSurv, Deep-Hit and random survival forest, providing methodological insights for the analysis of complex survival data.

Keywords: survival analysis; machine learning; DeepSurv; Deep-Hit; random survival forest

生存分析一般用于研究个体或群体在一段时间内生存或死亡的概率并探索影响该结局的相关因素, 在医学研究领域得到广泛应用。最常用的统计模型为 Cox 比例风险模型, 该模型假设研究对象的结局对数风险函数是协变量的线性组合, 且符合等比例风险假设。但在实际工作中这些假设不一定成立, 导致模型应用受到限制。机器学习法是一种通过迭代训练学习、适应数据本身特征, 并用于预测、分类或决策的方法。针对非线性复杂数据, 通常利用深度机器学习构建神经网络, 在网络中添加激活函数完成非线性映射, 能更准确地拟合不满足相关假设生存资料的预测

模型。1995 年, Faraggi-Simon 模型首次利用神经网络对 Cox 比例风险模型进行非线性扩展^[1]。随着技术的进步, 逐渐在神经网络中添加丢弃 (dropout) 层或 Nesterov 动量技术, 从原理上避免过拟合、梯度消失和梯度爆炸等问题, 提高了模型的预测和泛化能力^[2-3]。但应用神经网络模型时也会出现数据特征与模型优势不匹配的问题, 导致模型拟合效果不佳。本文检索 PubMed、中国知网等数据库 1995—2023 年发表的相关研究文献, 对以神经网络为代表的机器学习法在生存分析领域的研究进展进行综述, 重点介绍 DeepSurv、Deep-Hit 和随机生存森林 (random survival for-

DOI: 10.19485/j.cnki.issn2096-5087.2024.06.009

基金项目: 国家自然科学基金项目 (82073662)

作者简介: 刘玥, 硕士研究生在读, 公共卫生专业

通信作者: 张玉海, E-mail: zhyh@fmmu.edu.cn

ests, RSF) 3种机器学习生存模型的原理和特点,为复杂生存资料的分析提供方法学参考。

1 机器学习法扩展生存分析模型进展

机器学习法扩展生存分析模型的思路大致分为两类,一类是利用机器学习的数据特征提取及映射能力直接拟合复杂模型。1995年提出的Faraggi-Simon模型是最早将神经网络与传统Cox比例风险模型相结合的回归模型。它运用多层感知器内置函数取代Cox比例风险模型中的线性预测函数^[4],利用神经网络的非线性映射优势改善Cox比例风险模型的线性假定限制,但实践发现其总体预测性能并不优于传统Cox比例风险模型^[5-6]。2018年KATZMAN等^[7]在深度学习框架中改进Faraggi-Simon模型,开发了DeepSurv模型,发现其一致性指数优于传统Cox比例风险模型。同年,CHING等^[8]提出神经网络与Cox比例风险模型相结合的架构Cox-nnet,针对高通量组学数据表现出较高的预测准确度。ZHU等^[9]以卷积神经网络替换DeepSurv中的前馈神经网络,应用于肺癌的病理图像和整个组织切片的病理图像中,取得了较好的预测效果。类似的延展成果还有Survival-Net框架,它是基于神经网络和超参数贝叶斯优化的预测模型,主要针对基因组医学中的高维分子数据^[10],对临床生存结局的预测较为准确。

另一类分析思路是将持续时间离散化,在预计时间区间内分别计算风险或生存函数。2017年,LUCK等^[11]提出一种类似DeepSurv模型的算法,不同之处是遵循保序回归原理计算时间区间内的损失值,为生存分析增加了额外离散输出。FOTSO等^[12]运用神经网络估计多任务逻辑回归参数,可以直接计算时间区间内的生存概率。2018年LEE等^[13]提出一种将对数似然与排序损失相结合,用神经网络估计概率分布函数的方法,命名为Deep-Hit。

目前应用较多的生存分析分类回归模型是RSF模型^[14],它结合分类树理念和生存数据分析思想对生存时间进行预测,是比较灵活的连续型时间分类模型。

2 3种典型的机器学习生存分析模型

2.1 DeepSurv模型

DeepSurv模型是基于Cox比例风险模型部分似然理念,采用深度学习算法拟合风险函数的预测模型,通过多级融合自主特征学习,将多个线性、非线性因素转换为线性组合,并结合生存时间预测结局事件。该模型在调整网络深度的基础上添加了多种新技

术,如权重衰减、ReLU、批处理归一化、丢弃法、Nesterov动量、梯度修剪和学习率调度等。通过调整网络权重 θ 参数估计协变量对生存风险的影响。DeepSurv模型的网络结构如图1所示, x 为输入变量(即研究对象基线协变量数据),隐藏层使用全连接层和丢弃层,交替堆叠,通过对最后一层丢弃层的输出结果进行线性处理,得到最终风险率 $\hat{h}_\theta(x_i)$ 。DeepSurv有自动筛选特征变量的优势,能在不指定交互项的情况下进行预测。

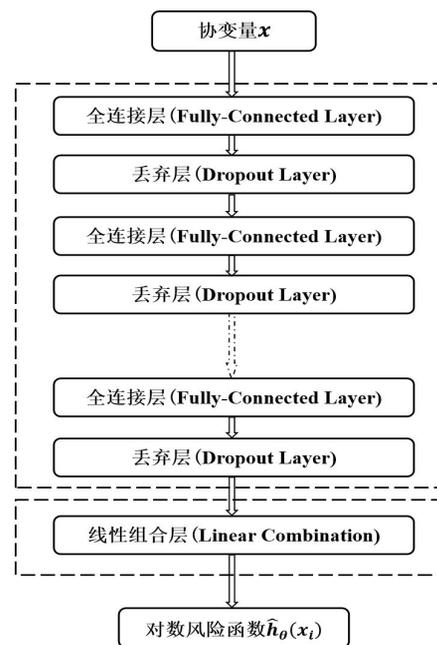


图1 DeepSurv网络结构图

Figure 1 DeepSurv network structure diagram

为使模型得出较为准确的预测结果,需不断调整参数以优化神经网络拟合度。损失函数是优化核心,与神经网络的结构、参数密切相关。DeepSurv模型在Cox对数似然基础上重设损失函数,通过最小化平均偏负对数似然函数来不断提升模型的预测和泛化能力,其损失函数如下:

$$l(\theta) = -\frac{1}{N_{E=1}} \sum_{i: E_i=1} \left(\hat{h}_\theta(x_i) - \log \sum_{j \in R(T_i)} e^{\hat{h}_\theta(x_j)} \right) + \lambda \|\theta\|_2^2$$

式中, $N_{E=1}$ 为观察到结局的病例数; E_i 、 T_i 和 x_i 分别为观测对象 i 的生存结局、生存时间和特征变量;风险集 $R(t) = \{i: T_i \geq t\}$ 是在时间 t 时仍有发生结局风险的病例集合, j 为尚在风险集内的观测对象; λ 为L2正则化参数; θ 为待调整的网络参数。同时,神经网络需要引入非线性激活函数拟合非线性函数关系,提高模型表达能力。目前,激活函数已超

过 20 种，常用的有 Sigmoid、Tanh、ReLU 和 Leaky-ReLU。

DeepSurv 模型利用神经网络优势解决现实中非线性、复杂交互等问题，还具有灵活的建模框架。尤其在大样本及高维数据中表现更为突出 [15-18]，研究者可以根据数据类型和领域知识，构建不同类型的神经网络架构，例如替换为处理具有时间或序列信息特征数据的循环神经网络或处理图片及影像资料的卷积神经网络等。这种柔性建模框架使 DeepSurv 模型可以适应各种数据类型和分析需求。多项研究表明，DeepSurv 模型的预测效果优于传统 Cox 比例风险模型和 RSF 模型 [19-23]。DeepSurv 模型的灵活性在复杂的图像预测中也表现出同样的优势 [24]。

2.2 Deep-Hit 模型

当研究对象可能发生多个结局事件且结局事件的

发生互斥时，称这些结局事件之间存在竞争风险，导致所有研究对象都随访至出现感兴趣结局的难度增加。此外，在生存分析中，如果将其他原因导致的死亡等同看作无信息删失，就意味着如果随访时间足够长，这部分研究对象仍有可能发生所研究疾病的死亡结局，这不仅与事实相悖，在一定程度上还会高估相应结局的发生风险。因此，考虑到竞争风险的存在，LEE 等 [13] 构建了 Deep-Hit 模型。

Deep-Hit 模型以学习真实死亡概率估计值 \hat{p} 为目的训练神经网络，真实死亡概率是结合了结局发生时间和竞争风险联合概率分布。为拟合联合概率分布，Deep-Hit 模型采用 Softmax 作为输出层，采用多任务网络构建，其中包括共享子网络和多个特定因果子网络。为更好地应对神经网络优化过程中梯度消失和网络退化等问题，还采用了保持残差连接的方法。见图 2。

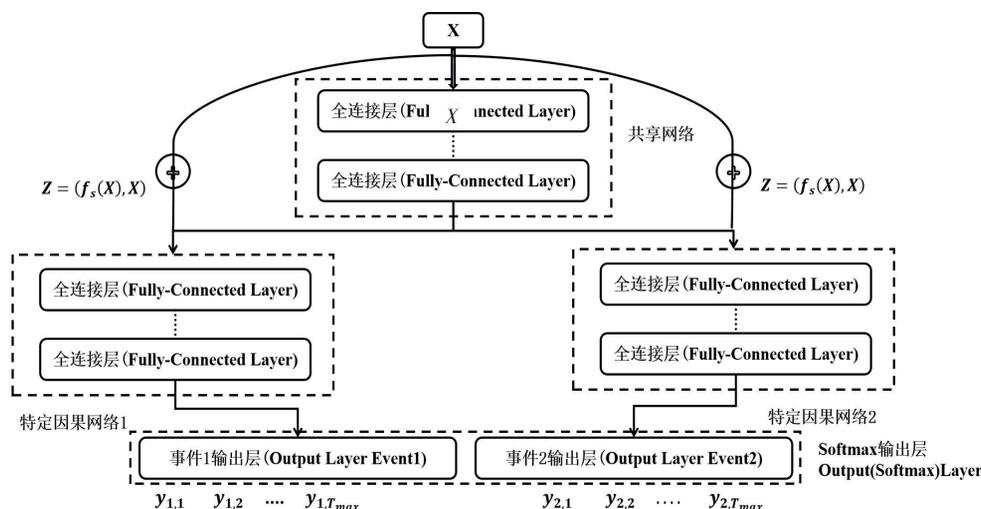


图 2 Deep-Hit 网络结构图

Figure 2 Deep-Hit network structure diagram

为提升模型的预测与泛化能力，Deep-Hit 模型通过 Adam 优化器及反向传播方法进行模型训练优化，使用可对删失数据进行特定处理的损失函数 L_{Total} ，公式如下：

$$L_{Total} = \alpha L_1 + (1 - \alpha) L_2$$

式中， α 为超参数，用于调节每个损失所占比例，可以采用超参数网格搜索进行优化； L_1 为生存时间及结局事件联合分布的对数似然函数； L_2 为特定病因的损失函数。

Deep-Hit 模型的优势在于既可处理单一潜在风险因素的影响，又可平稳应对有多个竞争风险因素的情况，针对非等比例风险生存数据有较好的表现。Deep-Hit 模型也是基于 Cox 比例风险模型基本假设

构建的神经网络模型，除去神经网络自主学习生存时间分布的优势，在理论上没有完全规避假设限制。2023 年，BAE 等 [25] 在 5 种恶性肿瘤生存分析预测中发现仅有 2 种恶性肿瘤的 Deep-Hit 模型表现优于传统 Cox 比例风险模型。ADEOYE 等 [19, 26] 对口腔癌预测及口腔疾病转化率预测研究显示，Deep-Hit 模型的预测准确度劣于 DeepSurv 模型和传统 Cox 风险比例模型。因此在模型选择时，需权衡模型的复杂程度、训练难度、耗费人力与物力、预测准确度和可靠性等，如果不优先考虑竞争风险，Deep-Hit 也许并不是最优选择。

2.3 RSF 模型

RSF 模型是综合随机森林与生存分析的基于集

成学习思想进行模型训练的预测方法。先采用 Bootstrap 法从原始数据中抽取 B 个样本集组成 IB (in-bag) 数据集, 每个样本集随机平均排除 37% 的数据, 称为袋外数据 (out-of-bag, OOB)。分别对每个样本集构建二叉生存树, 在树的每个节点上随机选取

p 个候选变量, 使用子节点之间生存差异最大化的候选变量对节点进行拆分, 计算每棵树的累积风险函数 (cumulative hazard function, CHF), 集成 CHF 再计算均值。最后采用 OOB 数据计算集成 CHF, 预测误差。见图 3。

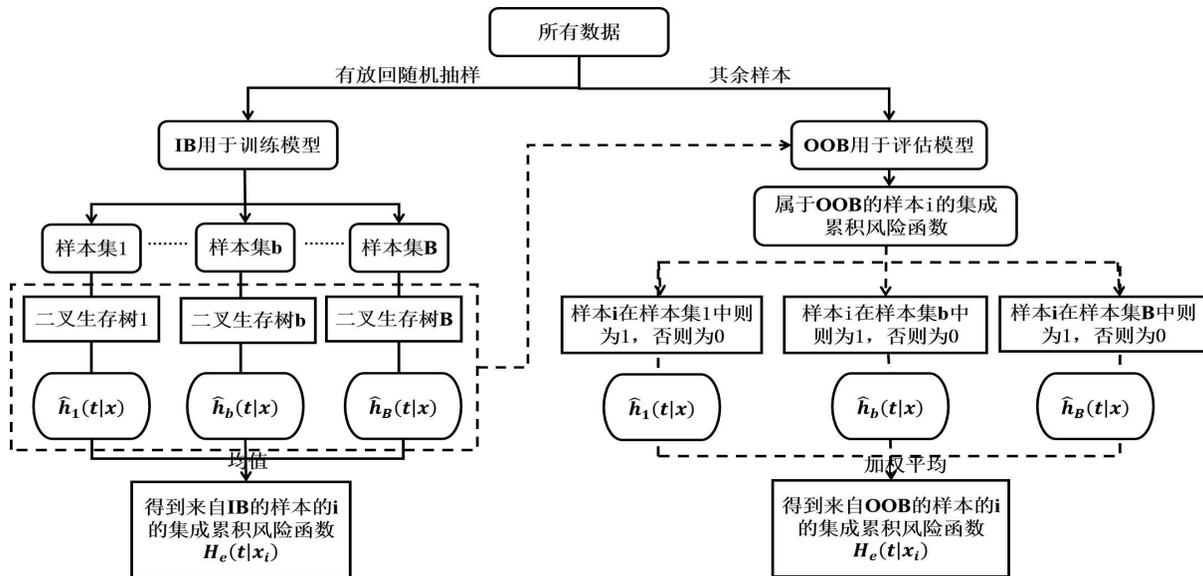


图 3 RSF 框架示意图

Figure 3 Schematic diagram of RSF framework

RSF 模型的核心是生成生存树和构建集成 CHF。与分类回归树算法类似, 生存树是通过递归分割技术搜索样本中所有的 x 变量和拆分值 c , 采用 log-rank 检验选择能使生存差异最大化的 x 和 c , 找到最佳拆分节点。层层筛检, 随着节点数量的增加, 异质样本被分离, 最终树上每个节点内都包括了同质且存活率相近的样本。叶节点 (饱和树最末端节点, h) 内所有样本的 CHF 相同。每棵生存树上叶节点处的 CHF 是利用该叶节点上的 n 个样本计算所得, 进一步进行累积发生函数 (cumulative incidence function, CIF) 分析并绘制 Nelson-Aalen 累积风险生存曲线。叶节点处的 CHF Nelson-Aalen 估计为:

$$\hat{H}_h(t) = \sum_{t_{i,h} \leq t} \frac{d_{i,h}}{Y_{i,h}} = \sum_{i=1}^l \frac{\text{在 } t_i \text{ 时刻死亡的人数}}{\text{存活到 } t_i \text{ 时刻的人数}}$$

式中, $d_{i,h}$ 表示在 $t_{i,h}$ 时刻死亡的人数 (不包括删失样本); $Y_{i,h}$ 表示在 $t_{i,h}$ 时刻仍存活的人数; i 表示第 i 个样本; h 表示结局事件发生的 $N(h)$ 个不同时刻。

由于集成学习思想, RSF 并不是通过最小化损失函数来优化模型, 而是通过计算每个个体的 CHF 或 CIF 得到最终集成模型的预测结果, 并利用预测误差衡量模型的泛化能力。

RSF 模型作为一种非参数预测模型, 针对分类、回归问题具有独特优势, 能有效回避传统 Cox 比例风险模型的等比例风险假设限制。但有研究显示, 当存在较多离群值时, RSF 模型的预测准确度会劣于传统 Cox 比例风险模型, 噪音比较大时也会出现过度拟合现象^[19, 27-30]。此外, RSF 模型对数据规范性的要求较高, 且只能计算死亡率的累加期望值也是该模型的不足之处。

3 小结

机器学习生存分析模型从不同层面弥补了传统 Cox 比例风险模型的局限。当存在非线性变量时, 传统 Cox 比例风险模型的难点在于需要筛选出具体的非线性变量, 而以神经网络为代表的机器学习生存分析模型可以自主学习无需单独识别。从预测效果来看, 机器学习生存分析模型通常优于传统 Cox 比例风险模型, 例如 DeepSurv 模型具有更好的区分度, RSF 在预测校准度上有一定优势^[31]。

传统 Cox 风险比例和机器学习生存分析模型都有各自的优缺点。前者以明确“风险因素”为主要目的, 结果较为直观易懂, 但可能会以牺牲预测准确性

为代价；后者以预测为主要目的，不仅需要估计更多参数，且需要更大的样本量支持参数估计结果，在追求预测准确性时可能会丧失一定的结果可解释性^[32-33]。研究者需根据数据特征和研究目的谨慎选择合适的分析方法。

参考文献

- [1] FARAGGI D, SIMON R. A neural network model for survival data [J]. *Stat Med*, 1995, 14 (18): 73-82.
- [2] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. *J Mach Learn Res*, 2015, 15 (1): 1929-1958.
- [3] 邹军华, 段晔鑫, 任传伦, 等. 基于噪声初始化、Adam-Nesterov方法和准双曲动量方法的对抗样本生成方法 [J]. *电子学报*, 2022, 50 (1): 207-216.
- [4] 李文琦, 黄水平, 李海朋. BP神经网络和Cox比例风险模型在生存分析应用中的比较 [J]. *郑州大学学报*, 2014, 49 (6): 822-825.
- [5] XIANG A, LAPUERTA P, RYUTOV A, et al. Comparison of the performance of neural network methods and Cox regression for censored survival data [J]. *Comput Stat Data An*, 2000, 34 (2): 243-257.
- [6] SARGENT D J. Comparison of artificial neural networks with other statistical approaches: results from medical data sets [J]. *Cancer*, 2001, 91 (8): 1636-1642.
- [7] KATZMAN J L, SHAHAM U, CLONINGER A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network [J/OL]. *BMC Med Res Methodol*, 2018, 18 (1) [2024-05-10]. <https://doi.org/10.1186/s12874-018-0482-1>.
- [8] CHING T, ZHU X, GARMIRE L X. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data [J/OL]. *PLoS Comput Biol*, 2018, 14 (4) [2024-05-10]. <https://doi.org/10.1371/journal.pcbi.1006076>.
- [9] ZHU X, YAO J, ZHU F, et al. Wsisa: making survival prediction from whole slide histopathological images [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6855-6863.
- [10] YOUSEFI S, AMROLLAHI F, AMGAD M, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models [J/OL]. *Sci Rep*, 2017, 7 (1) [2024-05-10]. <https://doi.org/10.1038/s41598-017-11817-6>.
- [11] LUCK M, SYLVAIN T, CARDINAL H, et al. Deep learning for patient-specific kidney graft survival analysis [J/OL]. *arXiv*, 2017 [2024-05-10]. <https://doi.org/10.48550/arXiv.1705.10245>.
- [12] FOTSO S. Deep neural networks for survival analysis based on a multi-task framework [J/OL]. *arXiv*, 2018 [2024-05-10]. <https://doi.org/10.48550/arXiv.1801.05512>.
- [13] LEE C, ZAME W R, YOON J, et al. DeepHit: a deep learning approach to survival analysis with competing risks [C] // Thirty-Second AAAI Conference on Artificial Intelligence New Orleans, 2018: 2314-2321.
- [14] ISHWARAN H, KOGALUR U B, BLACKSTONE E H, et al. Random survival forests [J/OL]. *Ann Appl Stat*, 2008, 2 (3) [2024-05-10]. <https://doi.org/10.1214/08-AOAS169>.
- [15] BIGLARIAN A, BAKHSHI E, GOHARI M R, et al. Artificial neural network for prediction of distant metastasis in colorectal cancer [J]. *APJCP*, 2012, 13 (3): 927-930.
- [16] RAU H H, HSU C Y, LIN Y A, et al. Development of a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network [J]. *Comput Methods Programs Biomed*, 2016, 125 (4): 58-65.
- [17] YU H H, HUANG T, FENG B, et al. Deep-learning model for predicting the survival of rectal adenocarcinoma patients based on a surveillance, epidemiology, and end results analysis [J/OL]. *BMC Cancer*, 2022, 22 (13) [2024-05-10]. <https://doi.org/10.1186/s12885-022-09217-9>.
- [18] YU H, YANG W, WU S, et al. Deep-learning-based survival prediction of patients with cutaneous malignant melanoma [J/OL]. *Front Med (Lausanne)*, 2023, 10 [2024-05-10]. <https://doi.org/10.3389/fmed.2023.1165865>.
- [19] ADEOYE J, KOOHI-MOGHADAM M, LO A W I, et al. Deep learning predicts the malignant-transformation-free survival of oral potentially malignant disorders [J/OL]. *Cancers (Basel)*, 2021, 13 (23) [2024-05-10]. <https://doi.org/10.3390/cancers13236054>.
- [20] SKREDE O J, DE RAEDT S, KLEPPE A, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study [J]. *Lancet*, 2020, 395 (10221): 350-360.
- [21] KIM B, JANG Y J, CHO H R, et al. Predicting completion of clinical trials in pregnant women: Cox proportional hazard and neural network models [J]. *Clin Transl Sci*, 2022, 15 (3): 691-699.
- [22] WANG F F, CHEN L, LIU L H, et al. Deep learning model for predicting the survival of patients with primary gastrointestinal lymphoma based on the seer database and a multicentre external validation cohort [J]. *J Cancer Res Clin Oncol*, 2023, 149 (13): 12177-12189.
- [23] JIANG C, WANG K, YAN L Z, et al. Predicting the survival of patients with pancreatic neuroendocrine neoplasms using deep learning: a study based on surveillance, epidemiology, and end results database [J]. *Cancer Med*, 2023, 12 (11): 12413-12424.
- [24] KIM D Y, CHOI K H, KIM J H, et al. Deep learning-based personalised outcome prediction after acute ischaemic stroke [J]. *J Neurol Neurosurg Psychiatry*, 2023, 94 (5): 369-378.
- [25] BAE C Y, KIM B S, JEE S H, et al. A study on survival analysis methods using neural network to prevent cancers [J/OL]. *Cancers*, 2023, 15 (19) [2024-05-10]. <https://doi.org/10.3390/cancers15194757>.
- [26] ADEOYE J, HUI L L, KOOHI-MOGHADAM M, et al. Comparison of time-to-event machine learning models in predicting oral cavity cancer prognosis [J/OL]. *Int J Med Inform*, 2022, 157 [2024-05-10]. <https://doi.org/10.1016/j.ijmedinf.2021.104635>.

卷和 FIT 阳性居民的结肠镜病变检出率较高, 与高以焕等^[18]的研究结果相一致, 提示问卷或 FIT 阳性居民应作为重点动员人群, 提高其结肠镜检查依从性。

本研究存在局限性。2021 年上城区区域优化, 部分结肠镜检查在非定点医疗机构完成, 无法获取全部筛查对象的结肠镜检查结果, 且技术水平存在差异, 可能导致结肠镜病变检出率被低估。建议增加定点医疗机构设置, 加强对定点医疗机构医生的结直肠癌筛查培训, 对筛查阳性居民做好健康教育, 落实早诊早治。

参考文献

- [1] XU L Y, ZHAO J H, LI Z H, et al. National and subnational incidence, mortality and associated factors of colorectal cancer in China: a systematic analysis and modelling study [J/OL]. *J Glob Health*, 2023, 13 [2024-05-14]. <https://doi.org/10.7189/jogh.13.04096>.
- [2] HAN B F, ZHENG R S, ZENG H M, et al. Cancer incidence and mortality in China, 2022 [J]. *J Natl*, 2024, 4 (1): 47-53.
- [3] 程向东, 杜灵彬. 2022 浙江省肿瘤登记年报 [M]. 北京: 清华大学出版社, 2023.
- [4] SIEGEL R L, WAGLE N S, CERCEK A, et al. Colorectal cancer statistics, 2023 [J]. *CA Cancer J Clin*, 2023, 73: 233-254.
- [5] 张芑, 蒋虹丽, 陈文, 等. 大肠癌人群筛查项目实施的国际进展 [J]. *上海预防医学*, 2021, 33 (4): 273-279.
- [6] 朱陈, 龚巍巍, 钟节鸣, 等. 浙江省重点人群结直肠癌筛查项目设计和实施方案 [J]. *中国肿瘤*, 2020, 29 (12): 899-903.
- [7] 柏愚, 杨帆, 马丹, 等. 中国早期结直肠癌筛查及内镜诊治指南 (2014 年, 北京) [J]. *胃肠病学*, 2015, 20 (6): 345-365.
- [8] 郑沛, 卢丽微, 朱福乾, 等. 永嘉县结直肠癌筛查结果分析 [J]. *预防医学*, 2022, 34 (7): 727-731.
- [9] 和金山, 朱陈, 潘婷婷, 等. 定性和定量粪便免疫化学试验筛查结直肠癌效果评价 [J]. *预防医学*, 2024, 36 (4): 317-321.
- [10] 国家癌症中心中国结直肠癌筛查与早诊早治指南制定专家组. 中国结直肠癌筛查与早诊早治指南 (2020, 北京) [J]. *中华肿瘤杂志*, 2021, 43 (1): 16-38.
- [11] 梅少林, 刘晓红, 杜冬明, 等. 丽水市居民结直肠癌发病与死亡趋势分析 [J]. *预防医学*, 2020, 32 (7): 666-669.
- [12] 杨金华, 沈飞琼, 薛峰, 等. 2007—2016 年浙江省嘉善县结直肠癌筛查方案应用评价 [J]. *中国肿瘤*, 2020, 29 (12): 919-924.
- [13] 杨会棉, 李超楠, 林恒娜. 2016—2019 年北京市门头沟区结直肠癌筛查结果分析 [J]. *社区医学杂志*, 2021, 19 (23): 1391-1394.
- [14] 洪尚游. 大肠癌结肠镜筛查的策略及依从性研究 [D]. 上海: 第二军医大学, 2012.
- [15] 许文敏. 广州市结直肠癌筛查策略评估及肠镜结果分析 [D]. 广州: 广州医科大学, 2020.
- [16] 杜倩, 邵春昕, 邹赤群, 等. 2015—2021 年北京市通州区结直肠癌早诊早治筛查结果分析 [J]. *中国肿瘤*, 2023, 32 (9): 698-702.
- [17] 蓝柳豪, 徐美慧, 吴胜智, 等. 2018—2019 年浙江省兰溪市结直肠癌筛查不同方案筛查效果的比较 [J]. *中国肿瘤*, 2023, 32 (3): 209-218.
- [18] 高以焕, 孔蕴馨, 董栋, 等. 风险评分模型、免疫法粪便隐血试验及二者联合筛查方案在徐州市人群结直肠癌筛查中的检出率比较 [J]. *现代预防医学*, 2023, 50 (5): 954-960.
- [27] LIN J, LI K, LUO S. Functional survival forests for multivariate longitudinal outcomes: dynamic prediction of Alzheimer's disease progression [J]. *Stat Methods Med Res*, 2021, 30 (1): 99-111.
- [28] KIM D W, LEE S, KWON S, et al. Deep learning-based survival prediction of oral cancer patients [J/OL]. *Sci Rep*, 2019, 9 (1) [2024-05-10]. <https://doi.org/10.1038/s41598-019-43372-7>.
- [29] BYUN S S, HEO T S, CHOI J M, et al. Deep learning based prediction of prognosis in nonmetastatic clear cell renal cell carcinoma [J/OL]. *Sci Rep*, 2021, 11 (1) [2024-05-10]. <https://doi.org/10.1038/s41598-020-80262-9>.
- [30] WANG D, PAN B, HUANG J C, et al. Development and validation of machine learning models for predicting prognosis and guiding individualized postoperative chemotherapy: a real-world study of distal cholangiocarcinoma [J/OL]. *Front Oncol*, 2023, 13 (2) [2024-05-10]. <https://doi.org/10.3389/fonc.2023.1106029>.
- [31] 陈雨轩, 韦红霞, 潘建红, 等. 扩展 Cox 模型在非线性生存资料分析中的预测能力比较 [J]. *南方医科大学学报*, 2023, 43 (1): 76-84.
- [32] 张梦, 张利, 刘梦雅, 等. 老年人自评健康状况与死亡风险的关联分析 [J]. *预防医学*, 2024, 36 (2): 105-108.
- [33] 李嵘, 张文丽, 李扬, 等. 基于深度学习的大规模肿瘤数据生存分析 [J]. *中国卫生统计*, 2022, 39 (1): 84-90.

收稿日期: 2024-02-23 修回日期: 2024-05-10 本文编辑: 徐文璐

(上接第 500 页)