

文章编号: 1673-8640 (2025) 03-0253-06 中图分类号: R446.1 文献标志码: A DOI: 10.3969/j.issn.1673-8640.2025.03.009

基于多种血清肿瘤标志物的深度学习模型用于评估 早期结直肠肿瘤患病风险

余佳杰¹ 张智智¹ 罗清琼¹ 柯 星²

(1.上海市皮肤病医院检验科, 上海 200443;

2.上海交通大学医学院附属新华医院检验科, 上海200092)

摘要: **目的** 利用多种血清肿瘤标志物建立基于深度学习的随机森林(RF)算法模型,并探讨该模型在评估早期结直肠肿瘤患病风险中的作用。**方法** 选取2023年3月—2024年2月上海交通大学医学院附属新华医院结直肠癌(CRC)患者176例(CRC组)、癌前病变患者110例(癌前病变组)、健康体检者72名(正常对照组)。检测所有研究对象血清肿瘤标志物癌胚抗原(CEA)、糖类抗原(CA)50、CA125、CA15-3、CA19-9、CA242、CA72-4、细胞角蛋白19片段(CYFRA 21-1)和甲胎蛋白(AFP)水平。将各组建模对象按7:3的比例分别随机拆分为训练集和验证集。采用基于深度学习的RF算法构建模型。采用受试者工作特征(ROC)曲线评价模型筛查早期CRC和癌前病变的效能。**结果** 将CRC患者和癌前病变患者合并成结直肠肿瘤疾病组(286例)。结直肠肿瘤疾病组血清CA50、AFP、CEA、CA19-9、CA125、CA72-4、CYFRA21-1水平显著高于正常对照组($P<0.05$)。通过特征处理和RF算法建模,利用CEA、CA50、CA15-3、CA242、CYFRA 21-1、CA72-4和性别这7个特征联合构建了CRC和癌前病变的综合诊断模型CR7。在训练集中,以正常对照组为对照,CR7模型诊断结直肠肿瘤疾病的曲线下面积(AUC)为0.997;在验证集中,AUC为0.931。CR7模型区分早期CRC患者和正常对照者的AUC为0.983,区分癌前病变患者和正常对照者的AUC为0.991。**结论** 利用RF算法构建的结直肠肿瘤疾病筛查模型为实验室辅助诊断提供了一种新的有效方法,可用于评估早期CRC和癌前病变的患病风险。

关键词: 肿瘤标志物;深度学习;随机森林;早期诊断;结直肠癌

Predicting early colorectal tumor risk using a deep learning model based on multiple serum tumor markers YU Jiajie¹, ZHANG Zhizhi¹, LUO Qingqiong¹, KE Xing². (1. Department of Clinical Laboratory, Shanghai Dermatology Hospital, Shanghai 200443, China; 2. Department of Clinical Laboratory, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China)

Abstract: **Objective** A random forest (RF) algorithm model based on deep learning has been established using multiple serum tumor markers to investigate the role of the model in assessing the risk of early colorectal tumors. **Methods** A total of 176 patients with colorectal cancer (CRC) (CRC group), 110 patients with precancerous lesions (precancerous lesion group) and 72 healthy subjects (healthy control group) from Xinhua Hospital of Shanghai Jiao Tong University School of Medicine from March 2023 to February 2024 were enrolled. The levels of serum tumor markers, including carcinoembryonic antigen (CEA), carbohydrate antigen (CA) 50, CA125, CA15-3, CA19-9, CA 242, CA72-4, cytokeratin 19-fragment (CYFRA 21-1) and alpha-fetoprotein (AFP) were determined. The subjects were randomly divided into a training set and a validation set at a ratio of 7:3. A model was constructed using the RF algorithm based on deep learning. The performance of the model in screening CRC and precancerous lesions was evaluated by receiver operating characteristic (ROC) curve. **Results** A cohort of 286 patients, comprising patients diagnosed with CRC and those with precancerous lesions, were collectively categorized as colorectal tumor group. Serum levels of CA50, AFP, CEA, CA19-9, CA125, CA72-4 and CYFRA21-1 in

作者简介: 余佳杰,男,1989年生,技师,学士,主要从事临床检验工作。

通信作者: 柯 星, E-mail: satoshiohno@sjtu.edu.cn。

引用本文: 余佳杰,张智智,罗清琼,等.基于多种血清肿瘤标志物的深度学习模型用于评估早期结直肠肿瘤患病风险[J].检验医学,2025,40(3):253-258.

colorectal tumor group were higher than those in healthy control group ($P<0.05$). Utilizing feature processing and RF algorithm model, a comprehensive diagnostic model named CR7 was developed, incorporating 7 key features: CEA, CA50, CA15-3, CA242, CYFRA21-1, CA72-4 and gender. In the training set, when using healthy control group as control, the CR7 model achieved an area under curve (AUC) of 0.997 for diagnosing colorectal tumor diseases. In the validation set, the AUC was 0.931. The CR7 model demonstrated an AUC of 0.983 for differentiating early CRC patients from healthy subjects and an AUC of 0.991 for differentiating patients with precancerous lesions from healthy subjects. **Conclusions** The colorectal tumor disease screening model constructed by the RF algorithm provides a new and effective method for laboratory-assisted diagnosis and can be used to assess the risk of early CRC and precancerous lesions.

Keywords: Tumor marker; Deep learning; Random forest; Early diagnosis; Colorectal cancer

结直肠癌 (colorectal cancer, CRC) 是临床常见的恶性肿瘤之一, 早期症状不明显, 患者确诊时多处于晚期, 发生转移的CRC患者5年生存率 $<20\%$ ^[1]。若能早期诊断CRC, 并通过规范治疗可使其5年生存率提高至 $75\%\sim 90\%$ ^[2]。目前, 临床上广泛使用的CRC筛查方法包括肠镜检查、影像学检查、粪便检测和血液学检测。肠镜检查虽然对CRC和进展期腺瘤有极高的敏感性, 但与粪便免疫化学试验 (fecal immunochemical test, FIT) 和愈创木粪便隐血试验 (guaiac-based fecal occult blood test, gFOBT) 等筛查方法相比, 具有侵入性、并发症多、成本较高等不足^[3]。此外, 多靶点粪便DNA (multi-target stool DNA, MT-sDNA) 检测、血浆SEPT9基因甲基化检测、粪便SDC2基因甲基化检测等非侵入性方法已被用于CRC筛查, 但这些方法对早期CRC和良性腺瘤的敏感性和特异性均较低^[4]。血清标志物如癌胚抗原 (carcinoembryonic antigen, CEA)、糖类抗原 (carbohydrate antigen, CA) 19-9、CA242、细胞角蛋白19片段 (cytokeratin 19-fragment, CYFRA 21-1) 等在肿瘤的诊断和进展监测中均有一定作用, 但这些肿瘤标志物在早期CRC和良性腺瘤筛查中的作用很少被研究^[5-7], 因此有必要充分挖掘这些肿瘤标志物的意义, 提高对早期CRC的筛查效能, 使患者能得到早期干预和治疗。机器学习通过随机选取大量样本, 利用相关因素之间的交互作用和非线性关系, 平衡样本误差的影响, 用大量数据训练得到的模型比只拟合简单测试样本的逻辑回归模型更可靠^[8-10]。本研究拟利用多种肿瘤标志物检测结果, 通过基于深度学习的随机森林 (random forest, RF) 算法建立结直肠肿瘤的诊断模型, 评估该模型对早期CRC和癌前病变的筛查效能, 以期能对

早期结直肠肿瘤的诊断提供准确、便捷的决策支持。

1 材料和方法

1.1 研究对象

选取2023年3月—2024年2月上海交通大学医学院附属新华医院CRC患者176例 (CRC组), 其中男100例、女76例, 年龄 (65 ± 10) 岁; 癌前病变患者110例 (癌前病变组), 其中男64例、女46例, 年龄 (63 ± 10) 岁; 包括管状腺瘤62例、增生性息肉30例、炎性息肉18例。所有患者均根据临床表现和病理检查结果确诊。依据中国临床肿瘤学会2024年发布的《结直肠癌诊疗指南》^[11], 将早期CRC定义为局限于黏膜固有层或穿透结直肠黏膜肌层浸润至黏膜下层, 但未侵犯肌层的恶性肿瘤。176例CRC患者中早期35例、中晚期141例。排除标准: 1) 临床诊断不明确; 2) 反复检查或治疗; 3) 合并其他恶性肿瘤。对于多次入院的患者, 仅收集未治疗 (包括手术治疗和药物治疗) 前的首次数据。另选取同期上海交通大学医学院附属新华医院健康体检者72名作为正常对照组, 其中男36名、女36名, 年龄 (59 ± 11) 岁。3个组之间年龄和性别差异均无统计学意义 ($P>0.05$)。本研究经上海交通大学医学院附属新华医院伦理委员会批准 (XHEC-D-2023-093), 所有研究对象均知情同意。

1.2 肿瘤标志物检测

采集所有研究对象清晨空腹静脉血3 mL, $1\ 000\times g$ 离心10 min, 分离血清。采用cabas 8000 e801电化学发光免疫分析仪 (瑞士罗氏公司) 和配套试剂检测血清CEA、CA50、CA15-3、CA19-9、CA242、CYFRA21-1、CA72-4、CA125和甲胎蛋白 (alpha-fetoprotein, AFP) 水平。剩余的样本置于 $-80\ ^\circ\text{C}$ 保存。检测前评估

各项室内质控、准确度、精密度等性能指标，确保患者数据准确可靠。

1.3 特征筛选和模型选择

特征相关性筛选旨在减少特征之间的冗余，保留与因变量高度相关的特征。在RF算法中，特征相关分析的阈值设置为0.9，即与因变量有高度相关性的特征被优先保留。使用L1正则化（也称为LASSO回归）作为特征筛选方法来进一步优化特征，并将正则化系数设置为1，以控制模型的复杂性，防止过度拟合。通过特征系数和相对权重来衡量模型中每个特征的重

表1 正常对照者一般资料分析

组别	例数	年龄/岁	性别	
			男/例	女/例
训练集	50	59 ± 11	25	25
验证集	22	59 ± 11	11	11

表2 CRC患者一般资料分析

组别	例数	年龄/岁	性别		临床分期	
			男/例	女/例	早期/例	中晚期/例
训练集	123	65 ± 10	70	53	25	99
验证集	53	65 ± 10	30	23	10	42

表3 癌前病变患者一般资料分析

组别	例数	年龄/岁	性别		病理分类		
			男/例	女/例	管状腺瘤/例	增生性息肉/例	炎性息肉/例
训练集	77	63 ± 10	45	32	43	21	13
验证集	33	63 ± 10	19	14	19	9	5

1.4 统计学方法

采用SPSS 12.0软件进行统计分析。采用Spsspro在线平台（<https://www.spsspro.com/>）执行深度学习算法。呈非正态分布的计量资料以中位数（*M*）[四分位数（*P*₂₅，*P*₇₅）]表示，组间比较采用Mann-Whitney *U*检验。计数资料以例或率表示，组间比较采用 χ^2 检验。采用受试者工作特征（receiver operating characteristic, ROC）曲线评价模型筛查CRC和癌前病变的效能。以

要性。将CRC患者、癌前病变患者和正常对照者按7：3的比例分别随机拆分为训练集和验证集，随机种子设置为1，见表1~表3。对缺失值 $\geq 30\%$ 的数据实行特征清洗，对缺失值 $< 30\%$ 的数据实行中位数替补。采用Z分数标准化方法对定量数据进行归一化，以保证在各种条件下的可比性。使用基于scikit-learn库的RF模型并设置超参数进行建模。RF模型的输出值以0~1之间的风险评分表示，风险评分是直接代表模型预测能力的定量指标。风险评分越高，模型预测事件的风险越高。

P < 0.05 为差异有统计学意义。

2 结果

2.1 结直肠肿瘤疾病患者和正常对照者各项指标比较

将CRC患者和癌前病变患者合并成结直肠肿瘤疾病组（286例）。结直肠肿瘤疾病组血清CA50、AFP、CEA、CA19-9、CA125、CA72-4、CYFRA21-1水平高于正常对照组（*P* < 0.05 ）。见表4。

表4 结直肠肿瘤疾病组和正常对照组各项指标比较

组别	例数	CA50/（U·mL ⁻¹ ）	AFP/（ng·mL ⁻¹ ）	CEA/（ng·mL ⁻¹ ）	CA19-9/（U·mL ⁻¹ ）
结直肠肿瘤疾病组	286	6.71（3.66~10.58）	2.94（1.67~3.36）	2.56（1.31~5.65）	10.60（5.88~18.80）
正常对照组	72	1.15（0.50~4.00）	2.79（1.98~3.71）	1.41（1.06~2.16）	7.66（4.91~11.60）
Z值		-8.240	-2.094	-5.078	-3.489
P值		<0.001	0.036	<0.001	<0.001

组别	CA125/（U·mL ⁻¹ ）	CA15-3/（U·mL ⁻¹ ）	CA72-4/（U·mL ⁻¹ ）	CYFRA21-1/（ng·mL ⁻¹ ）	CA242/（U·mL ⁻¹ ）
结直肠肿瘤疾病组	13.25（7.70~15.63）	9.25（6.48~13.20）	2.00（1.50~5.39）	2.18（1.46~2.96）	6.96（2.45~10.33）
正常对照组	12.05（10.00~15.90）	8.00（5.72~12.00）	1.63（0.90~3.66）	1.73（1.47~2.14）	4.10（2.71~7.60）
Z值	-3.358	-1.677	-2.438	-4.053	-0.875
P值	0.001	0.093	0.015	<0.001	0.382

注：空白表示无此项。

2.2 模型特征

将包含286例CRC和癌前病变患者的集合设为正事件，72名正常对照者的集合设为负事件，分为训练集和验证集。经特征处理后，以超参数运行RF

算法模型。将该风险模型命名为CR7，输出风险评分。构建的CR7模型共纳入7个特征，分别为CEA、CA50、CA15-3、CA242、CYFRA21-1、CA72-4和性别。CR7模型的特征系数和相对权重见表5。

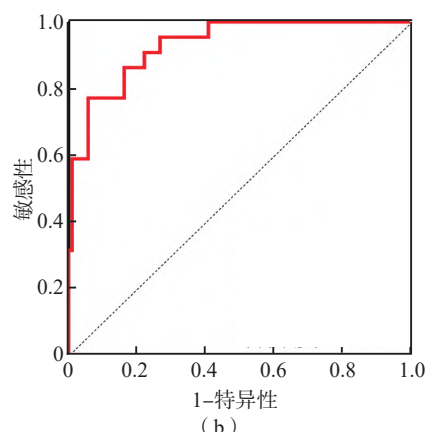
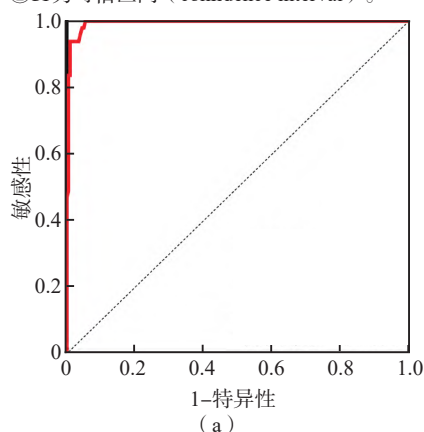
表5 CR7模型的特征系数和相对权重

序号	特征	特征系数	相对权重
1	CEA	0.279 8	1.000 0
2	CA50	0.253 2	0.905 2
3	CA15-3	0.145 4	0.519 9
4	CA242	0.137 8	0.492 5
5	CYFRA21-1	0.089 3	0.319 4
6	CA72-4	0.075 5	0.269 7
7	性别	0.019 0	0.067 8

表6 CR7模型诊断结直肠肿瘤疾病的效能

组别	AUC (95%CI) ^①	最佳临 界值	准确度/%	F1分数	敏感性/%	特异性/%	阳性预测值	阴性预测值	Youden指数
训练集	0.997 (0.994~1.000)	0.535	96.8	0.980	96.5	98.0	0.995	0.875	0.945
验证集	0.931 (0.878~0.983)	0.433	85.2	0.904	94.2	77.3	0.938	0.607	0.715

注：①CI为可信区间（confidence interval）。



注：（a）训练集；（b）验证集。

图1 CR7模型训练集和验证集的ROC曲线

表7 CR7模型内6个定量特征单项检测的筛查效能

项目	AUC (95%CI)	最佳临界值	敏感性/%	特异性/%	Youden指数
CA15-3	0.567 (0.490~0.644)	8.845 U·mL ⁻¹	53.8	62.5	0.163
CA72-4	0.592 (0.521~0.663)	5.225 U·mL ⁻¹	24.4	93.1	0.175
CYFRA21-1	0.660 (0.595~0.720)	2.000 ng·mL ⁻¹	62.1	72.2	0.343
CA242	0.535 (0.464~0.605)	3.070 U·mL ⁻¹	32.2	83.3	0.155
CA50	0.820 (0.767~0.873)	2.230 U·mL ⁻¹	89.5	59.7	0.492
CEA	0.697 (0.639~0.755)	2.840 ng·mL ⁻¹	44.3	91.7	0.360

2.4 CR7模型判断结果和实际结果比较

采用混淆矩阵将CR7输出的概率结果与实际诊断结果进行一致性比较。在训练集和验证集中，CR7模型判断结果与实际结果基本一致（ $P<0.001$ ）。见表8。

2.5 各组CR7模型风险评分比较

早期CRC组、癌前病变组CR7模型风险评分分别为 0.744 ± 0.154 、 0.869 ± 0.163 ，均高于正常对照组（ 0.202 ± 0.171 ）（ $P<0.001$ ）；且癌前病变组CR7模型风险评分高于早期CRC组（ $P<0.001$ ）。见图2。

2.3 CR7模型诊断结直肠肿瘤疾病的效能

ROC曲线分析结果显示，在训练集中，以正常对照组为对照，CR7模型诊断结直肠肿瘤疾病的曲线下面积（area under curve, AUC）为0.997；在验证集中，AUC为0.931。见表6、图1。CR7模型中单个特征的诊断效能见表7。CR7模型的筛查效能优于单项肿瘤标志物检测。

表8 CR7模型判断结果和实际结果比较

组别	正常对照组	结直肠肿瘤疾病组
训练集		
正常对照者	49	1
结直肠肿瘤疾病患者	7	193
χ^2 值		367.83
P值		<0.001
验证集		
正常对照者	17	5
结直肠肿瘤疾病患者	11	75
χ^2 值		355.42
P值		<0.001

2.6 CR7模型筛查早期CRC和癌前病变的效能

ROC曲线分析结果显示，CR7模型区分早期

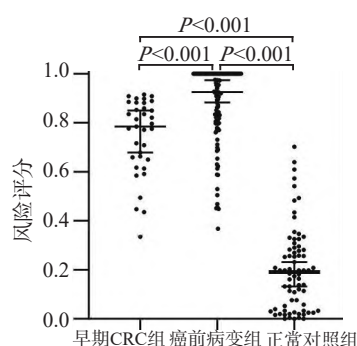
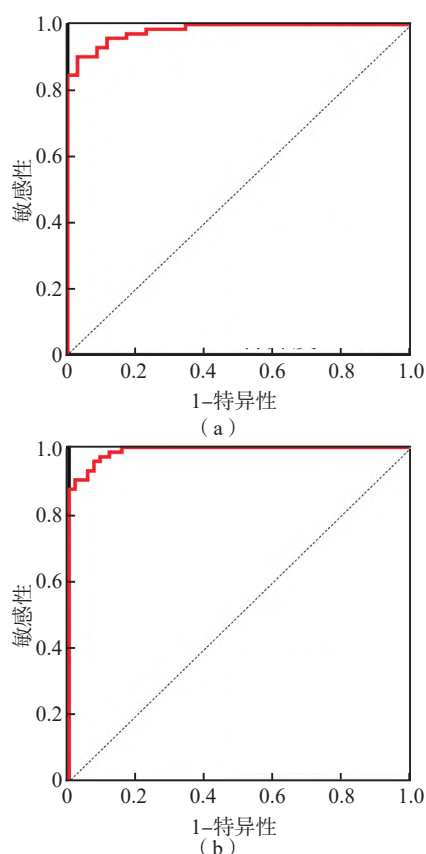


图2 各组CR7模型风险评分比较

CRC患者和正常对照者的AUC为0.983，区分癌前病变患者和正常对照者的AUC为0.991。见表9、图3。

表9 CR7模型的筛查效能

组别	AUC (95%CI)	最佳临界值	敏感性/%	特异性/%	Youden指数
早期CRC组	0.983 (0.965~1.000)	0.435	97.1	90.3	0.874
癌前病变组	0.991 (0.983~0.999)	0.580	92.7	95.8	0.885



注：(a) 早期CRC；(b) 癌前病变。

图3 CR7模型筛查早期CRC和癌前病变的ROC曲线

3 讨论

内镜检查和病理活检是诊断CRC的主要方法，但具有侵入性，不适合动态监测。非侵入性方法，如适用于大规模筛查的粪便隐血试验的敏感性和特异性较低，易出现假阳性结果^[12]。此

外，粪便或血浆DNA甲基化检测、ctDNA二代测序技术等都有其自身的限制。目前的诊断方法对早期结直肠肿瘤疾病，特别是早期CRC和癌前病变的筛查效果并不理想^[13-14]。因此，寻找安全、可靠的早期结直肠肿瘤疾病的诊断方法具有重要意义。有研究结果显示，结合深度学习技术分析多种标志物为肿瘤诊断提供了一个具有广阔前景的方法^[15-16]。因此，本研究检测了多种与胃肠道肿瘤诊断和预后相关的血清肿瘤标志物，利用RF算法建立了区分结直肠肿瘤疾病患者和健康人群的模型，并验证了该模型对早期CRC和癌前病变的筛查效能。

虽然血清肿瘤标志物已被广泛应用于各种肿瘤的诊断、病情和疗效监测，以及预后评估中，但单一标志物的敏感性和特异性均不足。本研究建立的基于深度学习的RF算法模型最终纳入了6种肿瘤标志物（CEA、CA50、CA15-3、CA242、CYFRA21-1、CA72-4）。CEA是《早期结直肠癌和癌前病变实验诊断技术中国专家共识》中推荐用于筛查CRC的诊断标志物^[17]。CA50是一种在大多数胃肠道腺癌中异常表达的细胞表面糖蛋白，可用于CRC的诊断和预后评估。CA72-4和CA242常被推荐作为胃肠道肿瘤筛查的补充指标^[18]。虽然CYFRA21-1和CA15-3通常被认为是肺癌和乳腺癌的生物标志物，但在最近的研究中，CYFRA21-1和CA15-3与其他标志物结合后亦可用于CRC的早期诊断^[19]。

深度学习技术的发展提高了多种标志物筛查某些疾病的潜力。RF是利用多棵树对样本进行训练并预测的一种准确度较高的分类器，是一种特殊的bagging算法，其利用bootstrap方法生成每个训练集来构造决策树，在每个节点分裂时，随机抽取一部分特征来找到最优解。这相当于对样本和特征都进行了采样，可有效避免过度拟合。RF算法通过随机选择特征数目和训练数据，对同一个预测数据出现次数最多的预测标签为最终预测标签^[20]。近年来，大量研究证明了RF算法在肿瘤诊断和治疗指导中的作用。NOVIELLI等^[21]利用RF算法鉴定出与CRC诊断相关的肠道微生物群，揭示了微生物与肿瘤发展的相互作用。LUO等^[22]利用RF等算法得出风险评分，用于预测索拉非尼耐药性相关基因标志物，以优化肝癌患者化疗和免疫治疗的临床决策。LIN等^[23]

利用RF算法评价我国接受全乳腺切除术并有1~2个阳性前哨淋巴结的乳腺癌患者非前哨淋巴结的转移情况,结果显示构建的模型有助于个体化的临床治疗决策。

虽然影响CRC诊断的因素众多且关系复杂,但RF算法可以处理一些复杂、不确定、非线性问题,具有良好的容错能力和快速并行处理能力。本研究将多个血清肿瘤标志物基于RF算法组合建模,最终构建的CR7模型纳入了性别和6种肿瘤标志物。在训练集和验证集中,模型输出的风险评分的筛查效能均较好($AUC>0.9$),优于单项肿瘤标志物,能很好地区分结直肠肿瘤疾病患者和正常对照者。混淆矩阵分析结果显示,预测结果和实际诊断结果的一致性较高,提示建立的CR7模型显著提高了多种血清肿瘤标志物联合诊断的准确性,对早期CRC和癌前病变均具有良好的筛查潜力。本研究也存在样本数有限和未做外部验证的不足。在未来的大型前瞻性队列研究中,将确定被CR7模型识别为高危人群的个体在随后的几年中是否会被确诊;同时将纳入新的标志物,以建立可靠的、有实用价值的模型,用于评估CRC的发生风险。

综上所述,基于患者一般资料和多种血清肿瘤标志物检测结果,采用RF算法构建的CR7模型显著提高了对早期CRC和癌前病变的筛查能力,为CRC的实验室辅助诊断提供了一种新的准确、便捷的方法。

参考文献

- [1] CHUANG J P, TSAI H L, CHEN P J, et al. Comprehensive review of biomarkers for the treatment of locally advanced colon cancer[J]. *Cells*, 2022, 11 (23): 3744.
- [2] MAES-CARBALLO M, GARCÍA-GARCÍA M, MARTÍN-DÍAZ M, et al. A comprehensive systematic review of colorectal cancer screening clinical practices guidelines and consensus statements[J]. *Br J Cancer*, 2023, 128 (6): 946-957.
- [3] DING Q, KONG X, ZHONG W, et al. Fecal biomarkers: Non-invasive diagnosis of colorectal cancer[J]. *Front Oncol*, 2022, 12: 971930.
- [4] YUAN R Q, ZHAO H, WANG Y, et al. SEPTIN9-SDC2-VIM methylation signature as a biomarker for the early diagnosis of colorectal cancer[J]. *Am J Cancer Res*, 2022, 12 (7): 3128-3140.
- [5] CAO H, ZHU L, LI L, et al. Serum CA724 has no diagnostic value for gastrointestinal tumors[J]. *Clin Exp Med*, 2023, 23 (6): 2433-2442.
- [6] PAKU M, UEMURA M, KITAKAZE M, et al. Clinical significance of preoperative and postoperative serum CEA and carbohydrate antigen 19-9 levels in patients undergoing curative resection of locally recurrent rectal cancer[J]. *Dis Colon Rectum*, 2023, 66 (3): 392-400.
- [7] LUO H, SHEN K, LI B, et al. Clinical significance and diagnostic value of serum NSE, CEA, CA19-9, CA125 and CA242 levels in colorectal cancer[J]. *Oncol Lett*, 2020, 20 (1): 742-750.
- [8] KOPPAD S, BASAVA A, NASH K, et al. Machine learning based identification of colon cancer candidate diagnostics genes[J]. *Biology*, 2022, 11 (3): 365.
- [9] 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.
- [10] KAVITHA M S, GANGADARAN P, JACKSON A, et al. Deep neural network models for colon cancer screening[J]. *Cancers*, 2022, 14 (15): 3707.
- [11] WAN F, CHEN G, ZHANG Z, et al. The Chinese Society of Clinical Oncology (CSCO): clinical guidelines for the diagnosis and treatment of colorectal cancer, 2024 update[J]. *Cancer Commun (Lond)*, 2024: 1-28.
- [12] LADABAUM U, MANNALITHARA A, WENG Y, et al. Comparative effectiveness and cost-effectiveness of colorectal cancer screening with blood-based biomarkers (liquid biopsy) vs fecal tests or colonoscopy[J]. *Gastroenterology*, 2024, 167 (2): 378-391.
- [13] HANNA M, DEY N, GRADY W M. Emerging tests for noninvasive colorectal cancer screening[J]. *Clin Gastroenterol Hepatol*, 2023, 21 (3): 604-616.
- [14] LI Y J, WANG X, WU Y J, et al. Access to colorectal cancer screening in populations in China, 2020: a coverage-focused synthesis analysis[J]. *Int J Cancer*, 2024, 155 (3): 558-568.
- [15] KE X, LIU W, SHEN L, et al. Early screening of colorectal precancerous lesions based on combined measurement of multiple serum tumor markers using artificial neural network analysis[J]. *Biosensors (Basel)*, 2023, 13 (7): 685.
- [16] KE X, CAI X, BIAN B, et al. Predicting early gastric cancer risk using machine learning: a population-based retrospective study[J]. *Digit health*, 2024, 10: 20552076241240905.
- [17] 中华医学会检验医学分会分子诊断学组. 早期结直肠癌和癌前病变实验诊断技术中国专家共识[J]. *中华检验医学杂志*, 2021, 44(5):372-380.
- [18] LIU S, XU M, QIAO X, et al. Prediction of serosal invasion in gastric cancer: development and validation of multivariate models integrating preoperative clinicopathological features and radiographic findings based on late arterial phase CT images[J]. *BMC Cancer*, 2021, 21 (1): 1038.
- [19] GAWEL S H, LUCHT M, GOMER H, et al. Evaluation of algorithm development approaches: development of biomarker panels for early detection of colorectal lesions[J]. *Clin Chim Acta*, 2019, 498: 108-115.
- [20] QIN Y, HUO M, LIU X, et al. Biomarkers and computational models for predicting efficacy to tumor ICI immunotherapy[J]. *Front Immunol*, 2024, 15: 1368749.
- [21] NOVELLI P, ROMANO D, MAGARELLI M, et al. Explainable artificial intelligence for microbiome data analysis in colorectal cancer biomarker identification[J]. *Front Microbiol*, 2024, 15: 1348974.
- [22] LUO T, CHEN X, PAN W, et al. The sorafenib resistance-related gene signature predicts prognosis and indicates immune activity in hepatocellular carcinoma[J]. *Cell Cycle*, 2024, 23 (2): 150-168.
- [23] LIN G, CHEN W, FAN Y, et al. Machine learning radiomics-based prediction of non-sentinel lymph node metastasis in chinese breast cancer patients with 1-2 positive sentinel lymph nodes: a multicenter study[J]. *Acad Radiol*, 2024, 31 (8): 3081-3095.

(收稿日期: 2024-04-02; 修回日期: 2024-12-26)

(本文编辑: 龚晓霖)