

机器学习算法挖掘整合临床实验室多参数检验大数据构建结直肠息肉风险预测模型及效能评估

王晶^a, 郝晓燕^a, 白丛霞^a, 邓娟^b, 周磊^a, 王娟^a, 刘家云^a
(空军军医大学西京医院 a. 检验科; b. 消化内科, 西安 710032)

摘要: **目的** 通过机器学习算法挖掘常规检验大数据构建结直肠息肉预测模型, 为结直肠癌前病变的早期诊断和预防性治疗提供新思路。**方法** 本研究收集空军军医大学西京医院2023年10月~2024年12月诊断为结直肠息肉的患者423例, 以及健康受试者305例。收集年龄、性别及实验室检测指标共66项, 采用拉索交叉验证(LassoCV)、极端梯度提升递归特征消减交叉验证(XGBoost-RFECV)、支持向量机递归特征消除交叉验证(SVM-RFECV)三种方法进行特征筛选。应用合成少数类过采样技术(SMOTE)对数据进行平衡处理, 处理后共846例, 使用随机森林(Random Forest)、高斯朴素贝叶斯(GNB)、决策树(Decision Tree)、逻辑回归(Logistic Regression)、梯度提升决策树(GBDT)、补集朴素贝叶斯(CNB)、自适应提升(AdaBoost)、极端梯度提升(XGBoost)和轻量级梯度提升(LightGBM)9种机器学习算法构建结直肠息肉早期预测模型, 其中760例为训练集, 86例为验证集用于验证模型的预测效能。**结果** 经筛选最终纳入活化部分凝血活酶时间(APTT)、凝血酶时间(TT)、凝血酶原时间(PT)、平均血小板体积(MPV)、血小板分布宽度(PDW)、直接胆红素(DBIL)、尿素(BU)、钠(Na⁺)共8个特征(均 $P < 0.05$)用于构建结直肠息肉预测模型。采用9种机器学习方法构建的模型中, XGBoost算法模型预测效果最佳, 诊断曲线下面积[AUC(95%CI)]为0.961(0.951~0.971), 灵敏度0.894, 特异度为0.889。**结论** XGBoost算法联合8种检验指标建立的结直肠息肉风险预测模型, 为肠镜检查的推广、结直肠癌的早筛早诊提供了新依据。

关键词: 结直肠息肉; 机器学习算法; 风险预测模型

中图分类号: R574; R446.1 文献标志码: A 文章编号: 1671-7414(2026)03-098-06

doi:10.3969/j.issn.1671-7414.2026.03.018

Machine Learning Algorithm-Based Mining and Integration of Clinical Laboratory Multi-Parameter Big Data for the Construction and Performance Evaluation of a Colorectal Polyps Risk Prediction Model

WANG Jing^a, HAO Xiaoyan^a, BAI Congxia^a, DENG Juan^b, ZHOU Lei^a, WANG Juan^a, LIU Jiayun^a (a. Department of Laboratory Medicine; b. Department of Gastroenterology, Xijing Hospital of Air Force Medical University, Xi'an 710032, China)

Abstract: **Objective** To establish a prediction model for colorectal polyps via mining big data derived from routine laboratory examinations through machine learning algorithms, and to provide novel insights for the early diagnosis and prophylactic treatment of precancerous lesions of colorectal cancer. **Method** A total of 423 patients diagnosed with colorectal polyps and 305 healthy controls were enrolled at Xijing Hospital of Air Force Medical University between October 2023 and December 2024. Sixty-six variables, including age, gender, and laboratory test indicators were collected. Feature selection was performed using three distinct methods: least absolute shrinkage and selection operator cross-validation (LassoCV), extreme gradient boosting-recursive feature elimination cross-validation (XGBoost-RFECV), and support vector machine-recursive feature elimination cross-validation (SVM-RFECV). The Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the dataset, resulting in a total of 846 eligible cases after processing. Subsequently, nine machine learning algorithms, namely Random Forest, Gaussian Naive Bayes (GNB), Decision Tree, Logistic Regression, gradient boosting decision tree (GBDT), complement naive Bayes (CNB), Adaptive Boosting (AdaBoost), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM), were utilized to establish an early prediction model for colorectal polyps. Among all cases, 760 were assigned to the training set and 86 to the validation set for evaluating the predictive performance of the model. **Results** A total of eight variables was finally incorporated into the colorectal polyp prediction model, including activated partial thromboplastin time (APTT), thrombin time (TT), prothrombin time (PT), mean platelet volume (MPV), platelet distribution width (PDW), direct bilirubin (DBIL), urea (BU) and sodium ion (Na⁺). Among the nine machine learning algorithms evaluated, the XGBoost model exhibited the optimal predictive

基金项目: 国家重点研发项目(编号: 2022YFC3602302)。

作者简介: 王晶(1993-), 女, 硕士, 主管技师, 主要从事临床检验与分子生物学, E-mail: wjmed0429@163.com。

通讯作者: 刘家云(1971-), 男, 教授, 主任医师, 博士研究生导师, E-mail: jiayun@fmmu.edu.cn。

efficacy, with an area under the receiver operating characteristic curve, with an AUC(95%CI) of 0.961 (0.951 ~ 0.971), a sensitivity of 0.894, and a specificity of 0.889. **Conclusions** The risk prediction model for colorectal polyps, developed by integrating the XGBoost algorithm with eight laboratory indicators, provides a new evidence base for the popularization of colonoscopy and the early screening and diagnosis of colorectal cancer.

Keywords: colorectal polyps; machine learning algorithm; risk prediction model

结直肠癌是全球第三大常见癌症,死亡率居第二位^[1],在我国结直肠癌死亡率占比9.3%,位列第四^[2]。面临结直肠癌带来如此严重的医疗卫生和经济负担,早诊早治已然成为亟待解决的问题。大约70%~90%的结直肠癌是由结直肠息肉发展而来,息肉到肿瘤的发展需5~10年,这为筛查和早期干预提供了时间窗口^[3]。在此期间,有效的结直肠镜筛查和息肉切除可显著降低结直肠癌死亡率^[4]。

结直肠息肉在临床上因无特异症状,患者难以主动就诊,仅少部分体检者行肠镜检查被发现。肠镜检查作为结直肠癌筛查的“金标准”,因有创性限制其在健康体检者中普及,且常见的肿瘤标志物对结直肠息肉特异性也较低。因此我们尝试通过对体检者的常规检验大数据进行挖掘,对比结直肠息肉患者,基于机器学习方法建立多项目联合预测结直肠息肉模型,更精准地预测出健康体检者中的结直肠息肉患者,为这类高风险体检者进行肠镜检查提供更坚实的证据,为结直肠癌的早期诊断和一级预防提供新思路。

1 材料和方法

1.1 研究对象 本研究为单中心回顾性横断面研究,收集2023年10月~2024年12月空军军医大学西京医院消化内科内镜中心诊断为结直肠息肉的患者2 327例,根据纳排标准,最终入选结直肠息肉组423例;同期健康管理中心选取资料完整的健康受试者305例,共728例。采用人工少数类过采样法(synthetic minority over-sampling technique, SMOTE)技术对数据处理,平衡后的数据为结肠息肉组423例,对照组423例,共846例。纳入标准:①实验组:经结肠镜检查确诊的“结肠息肉”“结肠多发息肉”“下消化道息肉”者;对照组:各项实验室检测指标均在参考范围内,且超声、CT等其他检查无明显异常的体检者;②年龄符合35~75岁;③病历资料完整,血细胞检测、粪常规、电解质、肝肾功、血凝、肿瘤标志物均被检测。排除标准:①合并其他消化道疾病,如消化性溃疡、消化道穿孔、溃疡性结肠炎等;②合并包括结直肠癌在内的其他恶性肿瘤。本研究为回顾性研究,免去患者的知情同意书,研究经空军军医大学西京医院医学伦理委员会审查通过(批件号:KY20222284-F-1)。

1.2 研究方法

1.2.1 临床数据收集:通过检验科实验室信息系统(laboratory information system, LIS)筛选符合纳排标

准的受试者的年龄、性别及实验室检查结果,包括血细胞检测、粪常规、电解质、肝功十二项、肾功四项、血凝六项、血脂四项、肿瘤标志物四项(CEA+CA199+AFP+CA125)共66项常规指标,经医院管理信息系统(hospital information system, HIS)查阅患者肠镜结果及病历信息,对其他合并症进行排除。

1.2.2 数据准备:将年龄、性别及实验室检查共66个变量进行描述性分析,剔除缺失率超过30%的变量,数值变量使用随机森林(random forest, RF)^[5]进行填补,分类变量使用K临近算法(K-Nearest Neighbor algorithm, KNN)^[6]进行填补,过滤后进行归一化处理,得到处理后的数据供后续使用。

1.2.3 特征筛选:对结直肠息肉组和健康对照组共728例66个变量进行统计学差异分析,对有显著差异的指标($P<0.05$),采用拉索交叉验证(LassoCV)、极端梯度提升递归特征消减交叉验证(XGBoost-RFECV)、支持向量机递归特征消除交叉验证(support vector machines recursive feature elimination cross-validation, SVM-RFECV)三种机器学习算法进行特征因子筛选,三种方法筛选出的特征因子取交集,最终筛选出的指标作为构建机器学习模型的特征因子。

1.2.4 模型建立及验证:采用SMOTE过采样方法对数据进行平衡,使得疾病组和对照组的比例为1:1,最终匹配结果为结直肠息肉组423例,健康对照组423例,共846例。结合筛选出的特征因子,分别应用使用随机森林(Random Forest)、高斯朴素贝叶斯(Gaussian Naive Bayes, GNB)、决策树(Decision Tree)、逻辑回归(Logistic Regression)、梯度提升决策树(Gradient Boosting Decision Tree, GBDT)、补集朴素贝叶斯(complement naive Bayes, CNB)、自适应提升(Adaptive Boosting, AdaBoost)、极端梯度提升(XGBoost)和轻量级梯度提升(LightGBM)9种机器学习算法进行十折交叉验证^[7](将数据集均分为10份,1份作为验证集,9份为训练集,循环10次,确保每个等份都被用于验证集1次,最后取平均值)。对过拟合模型增加正则化策略以增加模型泛化能力^[8]。

1.3 统计学分析 采用GraphPad Prism 8、Excel及DxAITM智研分析平台进行数据分析(R版本4.2.3和python版本3.11.4)。首先对连续变量进行Shapiro-Wilks检验是否符合正态分布,符合正态分布计量数据采用均数 \pm 标准差($\bar{x}\pm s$)表示,组间比较使用 t -test检验;不符合正态分布计量数据则采用中位

数(四分位数间距)[$M(P_{25}, P_{75})$]表示,组间比较使用Mannwhitney- U 检验;分类变量用百分比(%)表示,组间比较使用 χ^2 检验;频数为0的分类变量组间使用Fisher精确检验是否显著; $P < 0.05$ 为差异具有统计学意义。

2 结果

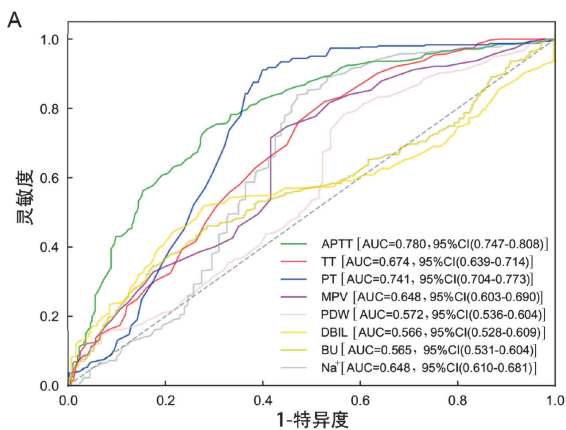
2.1 特征因子统计学差异 结直肠息肉患者及健康受试者共728例的66个特征因子中,性别、粪便隐血(FOB)、年龄、活化部分凝血活酶时间(APTT)、凝血酶时间(TT)、凝血酶原时间(PT)、PT国际标准化比值(INR)、纤维蛋白原降解产物(FDP)、D-二聚体(D-Dimer)、中性粒细胞百分比(N%)、淋巴细胞计数(L#)、淋巴细胞百分比(L%)、单核细胞计数(M#)、单核细胞百分比(M%)、嗜酸性粒细胞百分比(E%)、嗜酸性粒细胞计数(E#)、平均血红蛋白含量(MCH)、平均血红蛋白浓度(MCHC)、血小板计数(PLT)、平均血小板体积(MPV)、血小板分布宽度(PDW)、大血小板比率(P-LCR)、丙氨酸氨基转移酶(ALT)、AST/ALT、白蛋

白(ALB)、球蛋白(G)、总蛋白(TP)、碱性磷酸酶(ALP)、 γ -谷氨酰转肽酶(GGT)、直接胆红素(DBIL)、总胆红素(TBIL)、线粒体天门冬氨酸氨基转移酶(mAST)、亮氨酸氨基转肽酶(LAP)、肌酐(CRE)、尿素(BU)、尿酸(UA)、钾(K^+)、钠(Na^+)、二氧化碳(CO_2)、钙(Ca^{2+})、总胆固醇(CHO)、高密度脂蛋白胆固醇(HDLC)、糖类抗原125(CA125)共43项组间差异具有统计学意义(均 $P < 0.05$)。

2.2 特征筛选 对上述43项存在统计学差异的指标进行特征筛选,通过LassoCV方法共选择出37个的因子;通过XGBoost-RFECV方法共选择出35个因子;通过SVM-RFECV方法共选择出8个因子;三种方法取交集,共同筛选出8个因子作为构建预测结直肠息肉风险预测模型的特征因子,包括APTT、TT、PT、MPV、PDW、DBIL、BU和 Na^+ ,其受试者操作特征(ROC)曲线下面积(AUC)值、灵敏度、特异度、约登指数、最佳阈值、准确度见表1。8个特征因子的ROC曲线AUC见图1A,各因子间相关性分析见图1B。

表1 各特征因子诊断结直肠息肉的效能分析结果

项目	AUC值(95%CI)	灵敏度	特异度	约登指数	最佳阈值	准确度
APTT	0.780 (0.747 ~ 0.808)	0.740	0.715	0.455	27.100	0.416
TT	0.674 (0.639 ~ 0.714)	0.820	0.473	0.292	17.600	0.580
PT	0.741 (0.704 ~ 0.773)	0.911	0.600	0.512	9.800	0.581
MPV	0.648 (0.603 ~ 0.690)	0.715	0.584	0.299	10.725	0.583
PDW	0.572 (0.536 ~ 0.604)	0.784	0.433	0.216	12.400	0.583
DBIL	0.566 (0.528 ~ 0.609)	0.518	0.721	0.239	4.500	0.424
BU	0.565 (0.531 ~ 0.604)	0.461	0.715	0.176	5.610	0.418
Na^+	0.648 (0.610 ~ 0.681)	0.839	0.530	0.369	142.228	0.585



A. 各特征因子ROC曲线; B. 各因子间相关性分析。



图1 特征筛选出的8个单因子诊断效能及因子间相关性分析

2.3 模型建立与验证 经SMOTE数据平衡后,结合特征因子APTT、TT、PT、MPV、PDW、DBIL、BU、 Na^+ 建立的结直肠息肉风险预测模型中,对XGBoost、Logistic Regression增加正则化策略,

LightGBM、DecisionTree调整树深度。依据AUC排序,训练集表现最佳者为RandomForest,其预测的结直肠息肉的AUC(95%CI)为0.998(0.997 ~ 0.999);验证集表现最佳者为XGBoost,其预测的结直肠息肉

的AUC(95%CI)为0.961(0.951 ~ 0.971)。二者不吻合, RandomForest极可能存在过拟合现象, XGBoost稳定性可能相对较好。经过采样技术优化后, 9种机器学习算法建立的结直肠息肉风险预测模型的10折交

叉验证训练集性能评价结果见表2; 验证集性能评价结果见表3; 9种机器学习算法预测模型对结肠息肉的预测效能评估及各特征因子贡献度可视化分析见图2。

表 2 9种机器学习算法验证结直肠息肉风险预测模型性能评价(训练集)

分类模型	AUC(95%CI)	准确度	灵敏度	特异度	阳性预测值	阴性预测值
GBDT	0.921(0.918 ~ 0.924)	0.869	0.883	0.856	0.860	0.880
GNB	0.908(0.905 ~ 0.911)	0.844	0.861	0.828	0.834	0.857
CNB	0.847(0.844 ~ 0.850)	0.773	0.734	0.812	0.800	0.757
Logistic	0.916(0.911 ~ 0.921)	0.856	0.827	0.886	0.880	0.837
XGBoost	0.975(0.974 ~ 0.976)	0.921	0.924	0.919	0.920	0.924
LightGBM	0.961(0.960 ~ 0.962)	0.887	0.856	0.918	0.915	0.866
RandomForest	0.998(0.997 ~ 0.999)	0.986	0.991	0.981	0.981	0.991
AdaBoost	0.973(0.972 ~ 0.974)	0.919	0.924	0.914	0.915	0.924
DecisionTree	0.977(0.975 ~ 0.979)	0.921	0.922	0.919	0.920	0.923

表 3 9种机器学习算法验证结直肠息肉风险预测模型性能评价(验证集)

分类模型	AUC(95%CI)	准确度	灵敏度	特异度	阳性预测值	阴性预测值
GBDT	0.913(0.890 ~ 0.936)	0.868	0.886	0.849	0.859	0.885
GNB	0.902(0.878 ~ 0.926)	0.842	0.858	0.825	0.835	0.853
CNB	0.842(0.815 ~ 0.869)	0.751	0.712	0.790	0.776	0.740
Logistic	0.919(0.904 ~ 0.934)	0.850	0.815	0.884	0.877	0.830
XGBoost	0.961(0.951 ~ 0.971)	0.891	0.894	0.889	0.895	0.895
LightGBM	0.933(0.921 ~ 0.945)	0.853	0.820	0.887	0.886	0.833
RandomForest	0.961(0.952 ~ 0.970)	0.892	0.929	0.856	0.872	0.927
AdaBoost	0.945(0.929 ~ 0.961)	0.889	0.894	0.884	0.890	0.895
DecisionTree	0.896(0.882 ~ 0.910)	0.852	0.858	0.846	0.851	0.857

3 讨论

结直肠息肉是结直肠癌最重要的癌前病变, 在癌前病变发展为恶性肿瘤和转移前进行检测和切除, 可显著降低结直肠癌的发病率和死亡率。目前, 肠镜检查是提高结直肠息肉检出率的金标准, 并在临床实践中被证实能有效降低结直肠癌相关死亡风险。然而, 肠镜筛查参检率低是一个世界性难题^[9]。在现有的非侵入性筛查结直肠癌方法中, 粪便检测[如FOB和多靶点DNA检测(mtsDNA)^[10]]因其较高的特异度成为重要备选方案。然而, 研究显示体检人群的粪便样本弃检率高达35%~60%^[11], 极大限制了粪便检测的实际应用。相比之下, 血液检测具有无创、易获取、依从性高等优势。因此, 深入挖掘临床

实验室血液指标与结直肠病变的关联性, 有望为结直肠癌的早期筛查提供更可行地解决方案。

近年来, 人工智能在医疗领域的深入应用使得机器学习模型在疾病风险预测方面展现出显著潜力^[12]。本研究通过多阶段建模流程验证特征筛选的有效性: 首先基于原始数据进行特征筛选, 继而采用SMOTE过采样技术进行数据平衡, 通过多模型验证确保筛选特征的可靠性和泛化能力。在特征选择阶段, 本研究综合应用三种具有互补优势的机器学习方法: LassoCV具有计算效率高和自动特征选择的特性^[13], 但其对强相关性特征可能随机选择; XGBoost-RFECV擅长处理非线性关系和复杂特征交互, 但计算资源消耗较大; SVM-RFECV特别适用于小样本高维数据^[14],

具有坚实的理论基础,但对噪声敏感且计算效率较低。这种多方法融合的策略显著提升了特征选择的鲁棒性:既扩大了特征类型的覆盖范围,又通过方法间的相互验证降低了单一方法导致的选择偏差或过拟合风险。经比较筛选后,本研究确定XGBoost为最优预测模型。该集成学习算法具有以下核心优势^[15]:采

用迭代式决策树训练机制,通过残差优化逐步提升预测精度;整合L1/L2正则化约束,有效控制模型复杂度;利用二阶导数优化技术提高参数估计精度;特别适合结直肠息肉预测这类涉及多因素复杂交互的医学问题。这些特性共同保障了模型在保持较高预测性能的同时,具备良好的泛化能力。

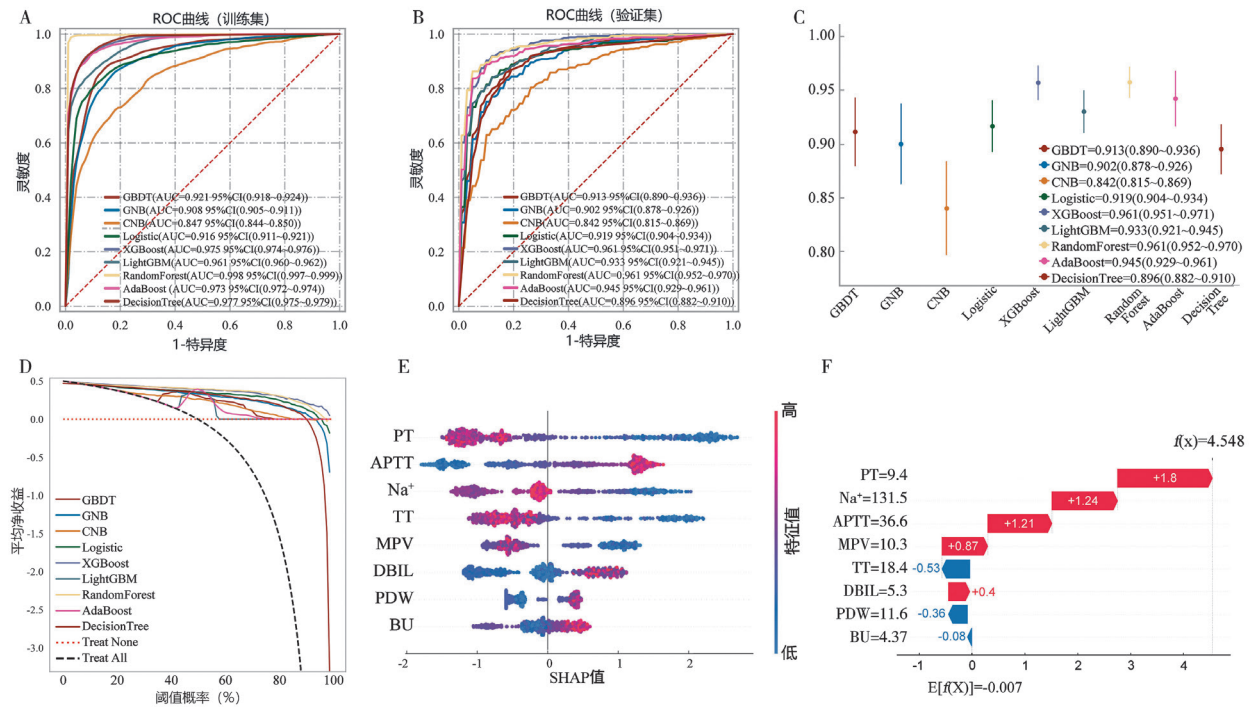


图2 9种机器学习模型训练集ROC曲线; B.9种机器学习模型验证集ROC曲线; C.验证集中各模型AUC评分森林图; D.各模型验证决策曲线图;

E.SHAP蜂群图:各特征在XGBoost模型中的贡献度; F.SHAP瀑布图:各特征对预测结果的累积贡献分解

图2 9种机器学习算法预测模型对结直肠息肉的预测效能评估及各特征因子贡献度分析

通过特征筛选出的8项关键指标,涉及凝血功能、血小板、肝肾功及电解质代谢,提示结直肠息肉的发生可能与系统微环境相关。多项研究表明,结直肠息肉患者常伴随凝血功能异常^[16],可能与局部炎症或血管内皮损伤相关。本模型纳入的凝血指标APTT、TT、PT的单因子诊断结直肠息肉效能较高且PT在模型中累积贡献最高,提示凝血功能检测或可作为息肉风险的间接标志物。MPV、PDW反映血小板活化^[17],而慢性炎症可能促进息肉形成,这一发现与既往关于炎症驱动结直肠肿瘤的研究相符。DBIL、BU、Na⁺则从内源性抗氧化^[18]、尿素循环紊乱^[19]、肿瘤微环境中钠离子蓄积增强免疫原性^[20]等方面支持对结直肠息肉的诊断作用。这些指标从凝血异常、血小板活化、氧化应激、氮代谢及离子平衡等多维度反映了结直肠息肉的特征,我们认为在结直肠息肉阶段,机体以自我保护、清除微小病变为调控方向,防止疾病进一步恶化。

本模型通过机器学习算法整合常规实验室检查数据(如血细胞检测、生化指标等),构建了一种具有

较高临床实用性的结直肠息肉预测工具。其核心优势在于仅需利用现有体检数据即可实现风险分层,无需额外检测,降低了筛查成本。通过将模型部署为检验数据自动化分析系统,可直接处理输入的检验结果并生成息肉风险评分,为临床医生提供客观的肠镜筛查依据,从而提高结直肠癌前病变的早期诊断率。该技术在医疗资源不足地区尤其具有应用价值,可结合基础检验项目实现初步风险评估,弥补基层筛查、检测能力的不足。然而,作为单中心回顾性研究且样本量有限,本模型可能存在选择偏倚,尚未纳入BMI、吸烟史和家族史等已知危险因素,未来需要通过扩大样本量进行多中心前瞻性研究进一步验证,并整合更全面的临床信息(如生活方式、营养状况等)以优化预测效能。此外,若能结合新兴的结直肠癌早期筛查分子标志物^[21-22]及多组学分析策略,模型的预测效能有望获得进一步提升。

参考文献:

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of inci-

- dence and mortality worldwide for 36 cancers in 185 countries[J]. *CA-A Cancer Journal for Clinicians*, 2021, 71(3): 209-249.
- [2] 郑荣寿,陈茹,韩冰峰,等.2022年中国恶性肿瘤流行情况分析[J]. *中华肿瘤杂志*, 2024, 46(3): 221-231.
ZHENG R S, CHEN R, HAN B F, et al. Cancer incidence and mortality in China, 2022[J]. *Chinese Journal of Oncology*, 2024, 46(3): 221-231.
- [3] ZAUBER A G, WINAWER S J, O' BRIEN M J, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths[J]. *New England Journal of Medicine*, 2012, 366(8): 687-696.
- [4] 杨婷,董泽华,李佳,等.结直肠黏膜癌前状态及癌前病变内镜下随访策略[J]. *中华消化杂志*, 2024, 44(7): 483-486.
YANG T, DONG Z H, LI J, et al. Endoscopic follow-up strategies for colorectal precancerous conditions and precancerous lesions[J]. *Chinese Journal of Digestion*, 2024, 44(7): 483-486.
- [5] JIANG T, GRADUS J L, LASH T L, et al. Addressing measurement error in random forests using quantitative bias analysis[J]. *American Journal of Epidemiology*, 2021, 190(9): 1830-1840.
- [6] EMMANUEL T, MAUPONG T, MPOLENG D, et al. A survey on missing data in machine learning [J]. *Journal of Big Data*, 2021, 8(1): 140.
- [7] 郝晓燕,周磊,白丛霞,等.基于机器学习方法建立血清多种肿瘤标志物联合区分胃炎和胃癌的三种模型及评价[J]. *现代检验医学杂志*, 2023, 38(2): 124-128.
HAO X Y, ZHOU L, BAI C X, et al. Establishment and evaluation of three models for the distinguish gastritis and gastric cancer with multiple serum tumor markers based on machine learning methods[J]. *Journal of Modern Laboratory Medicine*, 2023, 38(2): 124-128.
- [8] BARTLETT P L, LONG P M, LUGOSI G, et al. Benign overfitting in linear regression[J]. *Proceedings of the National Academy of Sciences of the United States of American*, 2020, 117(48): 30063-30070.
- [9] 李丹,汪嵘.结直肠癌肠镜筛查参检率影响因素研究[J]. *肿瘤研究与临床*, 2024, 36(7): 557-560.
LI D, WANG R. Factors influencing the participation rate in colorectal cancer screening by colonoscopy[J]. *Cancer Research and Clinic*, 2024, 36(7): 557-560.
- [10] 李怡玲,关旭,窦利州,等.粪便多靶点DNA与粪便潜血联合检测技术对结直肠癌早诊早筛的诊断价值分析[J]. *中华医学杂志*, 2022, 102(33): 2607-2613.
LI Y L, GUAN X, DOU L Z, et al. The clinical value of multi-target stool fecal immunochemical test-DNA in early screening and diagnosis for colorectal cancer[J]. *National Medical Journal of China*, 2022, 102(33): 2607-2613.
- [11] 朱小伶,贺雪,刘雪莲,等.品管圈在降低体检中心大便检查弃检率中的效果[J]. *当代护士*, 2018, 25(32): 176-177.
ZHU X L, HE X, LIU X L, et al. The effect of Quality Control Circle in reducing the abandonment rate of stool tests in the physical examination center[J]. *Modern Nurse*, 2018, 25(32): 176-177.
- [12] ZUO D, YANG L X, JIN Y, et al. Machine learning-based models for the prediction of breast cancer recurrence risk[J]. *BMC Medical Informatics and Decision Making*, 2023, 23(1): 276.
- [13] XIONG Y B, MA Y, RUAN L G, et al. Comparing different machine learning techniques for predicting COVID-19 severity[J]. *Infectious Diseases of Poverty*, 2022, 11(1): 19.
- [14] ZAREI GHOBADI M, AFSANEH E, EMAMZADEH R. Gene biomarkers and classifiers for various subtypes of HTLV-1-caused ATLL cancer identified by a combination of differential gene co-expression and support vector machine algorithms[J]. *Medical Microbiology and Immunology*, 2023, 212(4): 263-270.
- [15] LIANG D, WANG L, ZHONG P P, et al. Perspective: global burden of iodine deficiency: insights and projections to 2050 using XGBoost and SHAP [J]. *Advances in Nutrition*, 2025, 16(3): 100384.
- [16] ZHANG L, YE J, LUO Q Y, et al. Prediction of poor outcomes in patients with colorectal cancer: elevated preoperative prothrombin time (PT) and activated partial thromboplastin time (APTT) [J]. *Cancer Management and Research*, 2020, 12: 5373-5384.
- [17] LEBLANC R, PEYRUCHAUD O. Metastasis: new functional implications of platelets and megakaryocytes[J]. *Blood*, 2016, 128(1): 24-31.
- [18] CHEN Z J, VONG C T, GAO C F, et al. Bilirubin nanomedicines for the treatment of reactive Oxygen species (ROS)-Mediated diseases[J]. *Molecular Pharmaceutics*, 2020, 17(7): 2260-2274.
- [19] WU E, WEI G F, LI Y, et al. Serum urea concentration and risk of 16 site-specific cancers, overall cancer, and cancer mortality in individuals with metabolic syndrome: a cohort study[J]. *BMC Medicine*, 2024, 22(1): 536.
- [20] POKU L O, PHIL M, CHENG Y N, et al. 23Na-MRI as a noninvasive biomarker for cancer diagnosis and prognosis [J]. *Journal of Magnetic Resonance Imaging*, 2021, 53(4): 995-1014.
- [21] 张颖.结直肠癌患者血清SDC4,CXCL1水平检测在临床诊断及淋巴结转移评估中的应用价值[J]. *现代检验医学杂志*, 2025, 40(1): 116-121.
ZHANG Y. Application value of serum SDC4,CXCL1 level detection in clinical diagnosis and lymph node metastasis assessment in colorectal cancer patients[J]. *Journal of Modern Laboratory Medicine*, 2025, 40(1): 116-121.
- [22] 付超英,黄士月,张云霞.直肠癌组织中LncRNA TTN-AS1及角鲨烯环氧酶表达与临床病理特征及预后相关性研究[J]. *现代检验医学杂志*, 2024, 39(5): 12-16, 53.
FU C Y, HUANG S Y, ZHANG Y X. Correlation between the expression of LncRNA TTN-AS1 and squalene epoxidase in rectal cancer tissue and clinicopathological characteristics and prognosis[J]. *Journal of Modern Laboratory Medicine*, 2024, 39(5): 12-16, 53.

收稿日期: 2025-04-29

修回日期: 2025-07-11