

机器学习在结直肠癌相关风险预测中的研究进展^{*}马善源 综述, 彭 亮[△] 审校

上海交通大学医学院附属松江医院检验科, 上海 201600

摘 要: 结直肠癌(CRC)是全球第 3 大常见恶性肿瘤及第 2 大癌症相关死因, 新发 CRC 患者中有 20% 在就诊时即有转移, 另外 25% 的患者会发生转移, 因此, CRC 的早期筛查与精准风险预测是降低其病死率的关键。手术和放化疗仍然是 CRC 治疗的主要手段, 但治疗效果往往与发病部位、肿瘤分期等息息相关。目前常用于 CRC 早期筛查、转移和治疗监测的手段有肠镜、直肠指检、粪便隐血、肿瘤标志物和影像学检查等, 但单一手段往往灵敏度、准确性不高。机器学习是一种通过算法和模型使计算机从多源数据中自动学习并进行预测或决策的技术, 属于人工智能的一个分支。机器学习风险预测模型正越来越多地应用于疾病的筛查、诊断、治疗及预后等各环节。该文就机器学习在 CRC 相关风险预测中的应用进行综述。

关键词: 结直肠癌; 机器学习; 风险预测; 筛查; 诊断

中图法分类号: R735. 3; R319

文献标志码: A

文章编号: 1672-9455(2026)16-2181-06

Research progress of machine learning in risk prediction related to colorectal cancer^{*}Ma Shanyuan, Peng Liang[△]

Department of Laboratory Medicine, Songjiang Hospital Affiliated to Shanghai

Jiao Tong University School of Medicine, Shanghai 201600, China

Abstract: Colorectal cancer (CRC) is the third most common malignant tumor and the second leading cause of cancer-related death in the world, approximately 20% of newly diagnosed CRC patients present with metastasis at the time of diagnosis, and another 25% will develop metastasis subsequently. Therefore, early screening and precise risk prediction are crucial for reducing CRC mortality. Surgery, radiotherapy and chemotherapy remain the main treatment methods for CRC, but the treatment effects are often closely related to the site of onset, tumor stage, etc. Currently, colonoscopy, digital rectal examination, fecal occult blood, tumor biomarkers and imaging examination are commonly used for early screening, metastasis detection and treatment monitoring of CRC, however, a single method often has low sensitivity and accuracy. Machine learning, a branch of artificial intelligence, is a technology that enables computers to automatically learn from multi-source data through algorithms and models and make predictions or decisions. Machine learning-based risk prediction models are increasingly applied in various aspects of disease screening, diagnosis, treatment and prognosis. This article reviews the application of machine learning in CRC-related risk prediction.

Key words: colorectal cancer; machine learning; risk prediction; screening; diagnosis

国际癌症研究中心(IARC)2024 年发布的癌症负担报告显示, 2022 年全球新发结直肠癌(CRC)190 万例, 占全球所有新发肿瘤的 9. 6%, CRC 相关死亡 90 万例, 占所有癌症相关死亡的 9. 3%, 是全球第 3 大常见恶性肿瘤及第 2 大癌症相关死因^[1]。我国国家癌症中心 2024 年发布的最新数据显示, 2022 年我国新发恶性肿瘤中 CRC 发病率居第 2 位、病死率居第 4 位^[2], 早期筛查与精准风险预测是降低其病死率的关键。目前 CRC 筛查手段包括结肠镜、粪便免疫化学

检测、CT 结肠成像及分子标志物检测等, 但存在侵入性强、依从性低、假阳性率高、经济负担重等局限。机器学习按学习范式可分为监督学习、无监督学习和强化学习等不同类型。监督学习是一种基于标注数据训练模型的学习方式, 通过学习输入与标签之间的映射关系, 实现对新数据的预测。训练目标是预测输出与真实标签之间的误差最小化。常见算法包括线性回归、正则化、Logistic 回归、支持向量机、朴素贝叶斯、随机森林、神经网络、K 邻近(KNN)。无监督学习

^{*} 基金项目: 上海市卫生健康系统重点学科(2024ZDXK0065); 上海市松江区科技攻关项目(2023SJKJGG058)。

[△] 通信作者, E-mail: jysjyyjyk@126. com。

中文引用格式: 马善源, 彭亮. 机器学习在结直肠癌相关风险预测中的研究进展[J]. 检验医学与临床, 2026, 23(16): 2181-2186.

英文引用格式: Ma S Y, Peng L. Research progress of machine learning in risk prediction related to colorectal cancer[J]. Lab Med Clin, 2026, 23(16): 2181-2186.

是在无标注数据条件下,通过识别数据中的结构与模式进行学习的方法。无监督学习的核心思想是通过数据的统计特征、相似度等进行分析和挖掘,利用密度估计、聚类和降维等技术来捕获和发现数据隐藏的内在结构和模式。常见算法包括主成分分析(PCA)、t-分布随机邻域嵌入(t-SNE)、K 均值聚类算法(K-means)、混合高斯分布、潜在语义分析(LSA)、非负矩阵分解(NMF)等。强化学习是一种以环境反馈作为输入的、适应环境的机器学习算法,该方法通过试错来发现最优行为策略,进行从环境状态到动作映射的学习,以使动作从环境中获得的累积奖赏值最大。常见算法有 Value-Based、Policy-Based、Actor-Critic 等。随着人工智能技术的突破,机器学习通过整合不同来源如基因组学、影像组学、临床体征、实验室指标和肠道微生物学等数据显著提升了对 CRC 的早期识别及预后评估能力。当前研究聚焦于开发非侵入性生物标志物检测系统、优化内镜辅助诊断精度、预测转移与治疗反应,以及构建个体化筛查策略。本综述系统分析近 5 年机器学习在 CRC 领域的创新应用,重点介绍机器学习在 CRC 早期筛查中的应用,还涉及机器学习在 CRC 转移风险、手术并发症风险及预后和治疗反应预测模型中的应用。

1 机器学习在 CRC 早期筛查方面的应用

1.1 非侵入性生物标志物与液体活检相关模型 基于机器学习并结合非侵入性生物标志物及液体活检的预测模型显著提升了早期诊断 CRC 的灵敏度与特异性。Cao 等^[3]通过分析数据库中粪便 DNA 甲基化数据筛选关键甲基化位点并经实验验证,利用筛选出的位点甲基化值通过机器学习建模,cg13096260 和 cg12587766 双位点结合随机森林模型,显著提高了 CRC 及进展性腺瘤的检出率,经受试者工作特征(ROC)曲线分析,曲线下面积(AUC)为 0.928 5,外部数据评估新模型的诊断效能优于传统方法。Cao 等^[4]从 360 例参与者的训练队列中提取血浆游离 DNA(cfDNA)并进行测序,基于 cfDNA 片段组特征,采用包含 5 种机器学习算法的集成堆叠模型构建分类模型,以区分 CRC 患者与健康对照者。该模型在前瞻性队列中验证的 AUC 为 0.986,灵敏度为 94.88%,特异度为 98.00%,且灵敏度随肿瘤分期进展而升高。该模型充分挖掘了 cfDNA 片段组学的多维特征,兼具高灵敏度与高特异度,在 CRC 早期检测及患者广泛获益方面展现出良好的应用前景。Li 等^[5]从电子病历中提取癌胚抗原(CEA)、血红蛋白(Hb)、脂蛋白(a)[Lp(a)]和高密度脂蛋白(HDL)等指标,采用 Logistic 回归、随机森林、KNN、支持向量机(SVM)和朴素贝叶斯 5 种机器学习算法构建 CRC 诊断模型。其中,Logistic 回归模型表现最优(AUC=0.865),且 CEA、Hb、Lp(a)和 HDL 为其权重最高的 4 个特征。进而基于这 4 项指标建立简化诊断模型,其识别所有 CRC 患者的 AUC 为 0.849;该

模型在区分晚期结肠癌患者与健康个体时表现更优,AUC 达 0.905。此外,Li 等^[6]开发了一种用于 CRC 早期检测的微创方法,提出构建循环肿瘤细胞(CTC)RNA 数字评分系统,通过抗上皮细胞黏附分子(Ep-CAM)/表皮生长因子受体(EGFR)纳米颗粒捕获 CTC,再通过液滴数字 PCR(ddPCR)对纯化的 CRC 患者 CTC 中的 6 个 RNA 转录物进行定量,使用梯度提升算法(XGBoost)模型计算 CRC 评分,诊断准确率达 91%,显著优于 CEA、糖类抗原 125(CA125)和糖类抗原 199(CA199)。

1.2 挥发性有机化合物(VOCs)相关模型 VOCs 检测是另一新兴方向。Chen 等^[7]利用电子鼻分析呼气 VOCs,采用机器学习建模以早期诊断 CRC,模型的 AUC 为 0.87,早期患者灵敏度达 90%。而 Van 等^[8]利用场不对称离子迁移谱(FAIMS)和气相色谱-离子迁移谱(GC-IMS)技术检测粪便 VOCs 模式,并采用气相色谱-飞行时间质谱(GC-TOF-MS)分析单个 VOCs 的丰度,使用 10 折交叉验证进行二元分类预测,然后将区分性特征纳入 5 种独立的机器学习算法,其中,FAIMS 联合随机森林模型识别林奇综合征患者进展期肿瘤的灵敏度达 100%,为高危人群提供了无创监测手段。Bhattacharyya 等^[9]的荟萃分析显示,粪便 VOCs 诊断 CRC 的汇总灵敏度为 0.86,特异度为 0.90,而基于 VOCs 诊断 CRC 的机器学习算法提高了灵敏度和特异度,AUC 为 0.84~0.96。尽管 VOCs 用于 CRC 检测的准确性很高,但存在样本量较小、标准化不足的局限,需多中心验证提升可推广性。

1.3 肠道菌群相关模型 肠道菌群在 CRC 的发展中起着重要作用,作为早期发现和治疗的生物标志物,其机器学习相关研究日益成熟。Bakir-Gungor 等^[10]提出了基于分类的分组评分建模(G-S-M),在属、科、目 3 个分类水平研究与 CRC 相关的宏基因组数据。最终,基于属水平的 7 个 CRC 相关菌群特征构建的机器学习模型表现出最优的诊断效能,AUC 为 0.90。Han 等^[11]利用公开可用的病例对照数据集采用 sccomp 贝叶斯机器学习算法,通过概率建模精准验证 CRC 患者菌群组成差异,为微生物干预提供靶点。Lu 等^[12]从 38 例参与者的粪便标本中测序了 16S rRNA 基因,基于粪便微生物群操作分类单位(OTU),采用 8 种监督学习算法构建 CRC 诊断模型。其中,朴素贝叶斯模型诊断 CRC 的准确度为 0.917,AUC 为 0.926。研究结果表明,肠道微生物群失调与 CRC 之间存在关联,并证明了肠道微生物群诊断 CRC 的可行性^[12]。

1.4 组学数据相关模型 在多组学整合策略构建 CRC 风险预测模型方面,Qiu 等^[13]联合蛋白质组学与单细胞转录组筛选肿瘤相关抗原(TAA),使用 10 种机器学习算法开发诊断模型,基于抗细胞周期蛋白依赖性激酶调节亚基 1B(CKS1B)、抗 S100 钙结合蛋白 A11(S100A11)等 5 种自身抗体构建的随机森林模型

诊断 CRC 的 AUC 达 0.82。血清外泌体(EV)蛋白质组学作为非侵入性生物标志物用于诊断 CRC 的潜力尚不明确, Yin 等^[14]开发了一个与 EV 相关的随机森林模型,在多种机器学习算法中具有最高的诊断效率,在训练集和测试集的 AUC 分别达到 0.960、0.963。代谢组学作为一种潜在的非侵入性检测方法,可以使用已识别的生物标志物来检测患者体内潜在的癌症。Yang 等^[15]开发了一种基于化学描述符的机器学习模型,以识别与 CRC 相关的代谢物,模型在 CRC 各期都有较好的预测效能,充分证明了多组学机器学习模型在构建 CRC 风险预测模型中的潜在价值。

1.5 影像与内镜辅助诊断相关模型 在影像与内镜辅助诊断方面, Endo 等^[16]开发了基于快速区域卷积神经网络(Faster R-CNN)的人工智能系统,通过整合 CT 和结肠镜检查的双体位数据,将检出 ≥ 6 mm 肿瘤的灵敏度从 0.705 提升至 0.883。Kim 等^[17]纳入 233 例接受术前 PET/CT 检查的 CRC 患者,基于 PET 建立影像组学特征评分(Rad_Score), Logistic 回归分析显示, Rad_Score 是微卫星不稳定性(MSI)状态的独立预测因子,模型的 AUC 为 0.867。CT 结肠成像不能明确区分良性与癌前结直肠息肉, Grosu 等^[18]在结直肠息肉 CT 成像上进行手动分割,从分割中提取了表征形状、灰度直方图统计和图像纹理的定量图像特征,共得到 1 906 个特征-滤镜组合,基于这些特征构建机器学习诊断模型,随机森林模型的 AUC 为 0.91,显示出无创区分息肉类型的前景,这对于确定适当的治疗策略至关重要。

1.6 病理诊断相关模型 机器学习辅助病理诊断同样取得进展。Bharath 等^[19]基于公共权威病理图像数据库,应用随机铰链指数分布耦合注意网络(RHED-CANet)构建 CRC 分类模型,该模型准确率为 99.9%,精密率为 99.7%。这些性能指标印证了该方法在降低假阳性和提高诊断准确性的优势。同样地, Ke 等^[20]开发了基于深度卷积神经网络(CNN)的集成模型,用于 CRC 组织病理学图像的分类,该集成模型在 CRC 组织病理学图像的 EBHI 数据集上取得了良好的分类性能。Steimetz 等^[21]则开发 Res-Net34 模型区分低/高级别上皮内瘤变,准确率达 95%,为临床治疗决策提供关键依据。机器学习的应用显著提高了病理诊断的准确性。

上述研究表明,机器学习通过整合 cfDNA 甲基化、VOCs、肠道菌群特征及影像组学等多源数据,显著提升了 CRC 早期筛查的灵敏度与特异度,展现了液体活检与多组学整合在无创筛查中的潜力。

2 机器学习在转移风险预测中的应用

Wu 等^[22]研究了肠道微生物群在 CRC 淋巴结转移(LNM)中的重要作用,利用与 LNM 相关的肠道微生物群特征开发了随机森林和多层感知器(MLP)机器学习模型来预测 CRC 患者的 LNM 状态,随机森林模型在训练集中取得了良好的分类效能, AUC 为

0.957,在验证集中 AUC 为 0.705。肝脏是 CRC 最常见的远处转移器官^[23], Ding 等^[24]通过美国监测、流行病学和最终结果数据库(SEER)提取 CRC 病例进行分析,通过多种机器学习算法构建预测 CRC 肝转移风险模型, CEA 水平、全身治疗、N 分期、T 分期及化疗均是预测肝转移最重要的指标,模型的 AUC 为 0.88。Feng 等^[25]则整合乳酸脱氢酶(LDH)、CA199 等 6 项指标构建 CRC 肝转移-Lab6 模型,该模型能有效预测 CRC 肝转移,模型的 AUC 为 0.94。在接受 CRC 肝转移切除治疗的患者中,肝外复发(EHR)与较差的总生存期(OS)独立相关^[26]。Kawashima 等^[27]基于 XGBoost 开发在线计算器,通过原发瘤 T/N 分期及肿瘤负荷评分构建 EHR 风险预测模型,模型一致性指数(C-index)为 0.77,有助于辅助临床医生制订 CRC 肝转移患者的治疗策略。肺是 CRC 转移的第 2 大部位^[28], Guo 等^[29]通过 SEER 数据库获取 CRC 肺转移患者的临床信息,基于年龄、组织学类型、分级、原发肿瘤部位、T 分期、N 分期、肿瘤大小、CEA 水平和肿瘤沉积等特征,采用随机森林、决策树、支持向量机、朴素贝叶斯、KNN、XGBoost 和梯度提升机 7 种机器学习算法构建肺转移风险预测模型,其中,随机森林模型在内部测试集的 AUC 达 0.980,利用国内医院数据对该模型进行外部验证,外部验证的 AUC 为 0.927。Zhang 等^[30]进行了一项多中心回顾性研究,基于 SEER 数据库 46 037 例 CRC 患者的临床病理特征,采用 Logistic 回归、KNN、支持向量机、决策树、随机森林和平衡随机森林 6 种机器学习模型预测肺转移风险并进行生存风险分层,其中,平衡随机森林作为不平衡学习策略表现最优, AUC 为 0.874。

上述研究表明,基于肠道菌群、临床指标的机器学习模型能有效预测 CRC 淋巴结、肝、肺转移风险,模型的 AUC 多处于 0.70~0.98,有助于实现个体化风险评估与治疗策略制订,提升转移监测的精准性。

3 机器学习在 CRC 手术并发症风险预测中的应用

机器学习模型在手术并发症预测中的应用日益广泛,并表现出较高的预测精度。下肢深静脉血栓形成(DVT)是一种常见的术后并发症,发生率高达 40%, Liu 等^[31]结合 Lasso 回归与 XGBoost 算法,利用术前前白蛋白、手术时长等 8 项指标构建预测 CRC 患者 DVT 风险的模型,该模型的 AUC 为 0.996,准确率为 0.964,特异度为 0.978, F1 评分为 0.958, SHAP 方法确定年龄和术前前白蛋白是影响机器学习模型预测准确率的主要决定因素。Fang 等^[32]针对老年手术患者,建立包含冠心病史、白蛋白水平的 XGBoost 模型,该模型预测术后并发症发生的 AUC 达 0.92,灵敏度为 86%,特异度为 89%, F1 评分为 0.85,优于 Logistic 回归模型。Ruan 等^[33]基于门控递归单元衰减(GRU-D)的深度学习架构,并采用各种特征采样方案来预测 CRC 手术后并发症的风险,随

后与 Logistic 回归模型进行比较,结果发现 GRU-D 在预测伤口和脏器间隙感染方面优于 Logistic 回归模型,并且随着更多数据点的可用,表现出更好的性能。Cai 等^[34]使用 3 种不同的机器学习方法来开发并验证一种用于预测 CRC 患者术后未计划再手术风险的机器学习模型,结果显示基于剖腹手术史、高血压或脑卒中、低蛋白血症、年龄、肿瘤淋巴结转移分期、手术时间和性别构建的模型具有良好的辨别力和临床实用性,可改善治疗决策和预后。

上述研究表明,机器学习在预测 CRC 术后并发症如 DVT、感染等表现优异,部分模型的 AUC 接近 0.996。通过整合术前生化指标、手术时长等特征,机器学习可高效识别高危患者,辅助临床进行干预决策,体现其在围术期风险管理中的重要价值。

4 机器学习在预后与治疗反应预测中的应用

用于 CRC 患者生存预后预测的模型已趋成熟。Liang 等^[35]通过机器学习结合炎症相关生物标志物和临床病理特征构建预测模型对 II/III 期 CRC 的预后进行评估,基于全身炎症评分(SIS)、中性粒细胞/淋巴细胞比值(NLR)和临床病理结果建立列线图,C-index 为 0.708,具有较高的预测效率。Liao 等^[36]从 GEO 数据库获取数据筛选肿瘤浸润性 CD8⁺ T 淋巴细胞的特异性标志物并通过机器学习构建预后评估模型,该模型的 AUC 为 0.75,提示该模型可以有效预测 CRC 患者的预后和药物反应。Luo 等^[37]纳入 180 例 CRC 肝转移患者的增强 CT 影像,提取形状、一阶、小波和纹理等影像组学特征;经惩罚 Cox 回归进行特征变量筛选后,分别采用弹性网(EN)和随机生存森林算法构建无病生存期(DFS)的影像组学预测模型,其中,随机生存森林模型显示出更好的 DFS 预后潜力。Hou 等^[38]对公共数据库进行综合分析,在 CRC 中鉴定出 11 个具有预后意义的免疫原性细胞死亡(ICD)相关基因,然后采用综合计算框架评估 101 种机器学习组合,最终通过整合正向逐步 Cox 回归(StepCox)和随机生存森林构建优化的 11 基因 ICD 相关签名(ICDRS),基于 ICDRS 的列线图在 1~3 年生存预测中取得了出色的时间依赖性(AUC > 0.90)。Yue 等^[39]则通过整合大量转录组和单细胞转录组数据,从多水平分析并鉴定 CRC 相关的氨基酸及其衍生物代谢(MAAD)关键基因;进而基于 10 种机器学习算法(包括 Lasso 回归、梯度提升、StepCox 和 Cox 比例风险提升模型等)构建 101 种算法组合,筛选出 Lasso 回归+梯度提升作为最优 MAAD 预后模型,生存分析显示,该模型对 T1~T2 期和 T3~T4 期 CRC 患者均具有良好的预后预测性能。

治疗反应预测方面,Bahrambanan 等^[40]比较了 7 种机器学习算法对放化疗应答的预测效能,发现随机森林与梯度提升模型准确率最高(93.8%),而互信息特征选择优于 χ^2 检验。Liao 等^[36]利用机器学习算法筛选与结肠癌预后及免疫治疗反应相关的肿瘤浸

润性 CD8⁺ T 淋巴细胞的特征,所构建模型的 AUC 为 0.75,该模型可以有效预测结肠癌患者的预后和药物反应。Zhu 等^[41]基于铁死亡、脂肪酸动力学和肿瘤微环境构建了预后及治疗反应的预测模型 FeFAM-score,该模型在验证队列中的 C-index 为 0.648~0.720。Wang 等^[42]通过机器学习算法构建的 CRC 线粒体相关基因预后及治疗效果预测模型发现,高危组患者总生存时间明显缩短,高风险组表现出升高的肿瘤免疫功能障碍和排除评分,表明免疫治疗效果降低。Weng 等^[43]构建了 mRNA 干细胞指数相关的代谢风险评估模型,该模型对验证集数据进行风险分层,高危评分组与 CRC 的恶性进展、预后较差、辅助化疗反应较差相关,并形成了免疫逃避的背景。

上述研究表明,机器学习整合炎症标志物、影像组学、基因表达等多类数据,构建的预后与治疗反应预测模型表现稳健。这些模型有助于分层患者风险、预测生存及免疫治疗应答,为个体化治疗提供依据,推动精准医疗在 CRC 管理中的应用。

5 讨论

机器学习在 CRC 的研究与应用中已取得显著进展,尤其在早期筛查、转移预测、手术风险预警及治疗反应评估等方面表现出色。通过整合多组学数据,机器学习相关模型不仅提高了诊断的灵敏度与特异性,还为个体化医疗提供了有力工具。

5.1 面临的挑战 本文系统综述了近 5 年机器学习在 CRC 相关风险预测中的研究进展。尽管已经取得了显著进展,但其在实际应用中仍面临诸多挑战。

5.1.1 数据质量与规模限制 多数研究基于单中心、回顾性数据,样本量有限,数据缺失比较严重,缺乏前瞻性、多中心临床验证,影响模型的稳健性与泛化能力。

5.1.2 模型可解释性不足 多数机器学习模型,尤其是深度学习模型,常被视为“黑箱”,其决策过程难以理解,限制了在临床实践中的可信度和接受度。

5.1.3 标准化与泛化能力欠缺 不同研究使用的特征提取方法、算法架构和评估标准各异,导致模型之间难以比较,且跨人群、跨设备的泛化性能常不理想。

5.2 未来展望

5.2.1 多模态数据融合提升预测全面性 未来研究将更加注重整合临床、影像组学、基因组学、微生物组学等多维度数据,构建多层次、跨尺度的预测模型,以更全面地捕捉 CRC 发生与发展机制。

5.2.2 可解释性人工智能助力临床可信度 推动可解释性机器学习方法在医学模型中的应用,通过特征重要性分析、决策可视化等技术,增强模型透明度,促进临床医生的理解与采纳。

5.2.3 前瞻性临床验证与系统化部署 加强多中心、前瞻性临床试验,推动机器学习模型从研究向临床转化。开发轻量化、易集成的预测系统,实现与电子病历、影像系统的无缝对接。

5.2.4 标准化框架与开放共享平台建设 建立数据采集、特征工程、模型评估的标准化流程,推动公共数据集和基准平台的开放共享,促进算法比较与协同创新。

利益冲突 所有作者均声明不存在利益冲突。
作者贡献 马善源:负责文献检索、撰写初稿;彭亮:负责构思及终稿审定。

参考文献

[1] Global cancer burden growing, amidst mounting need for services[J]. Saudi Med J, 2024, 45(3):326-327.

[2] Han B F, Zheng R S, Zeng H M, et al. Cancer incidence and mortality in China, 2022 [J]. J Natl Cancer Cent, 2024, 4(1):47-53.

[3] Cao Q X, Dan Z J, Hou N Y, et al. Discovery and validation of colorectal cancer tissue-specific methylation markers; a dual-center retrospective cohort study[J]. Clin Epigenet, 2024, 16(1):122.

[4] Cao Y, Wang N, Wu X, et al. Multidimensional fragmentomics enables early and accurate detection of colorectal cancer[J]. Cancer Res, 2024, 84(19):3286-3295.

[5] Li H, Lin J, Xiao Y, et al. Colorectal cancer detected by machine learning models using conventional laboratory test data [J]. Technol Cancer Res Treat, 2021, 20: 15330338211058352.

[6] Li C, Wang Z L, Ding P, et al. A digital score based on circulating-tumor-cells-derived mRNA quantification and machine learning for early colorectal cancer detection[J]. ACS Nano, 2025, 19(19):18117-18128.

[7] Chen C D, Zheng Y X, Lin H F, et al. Development of electronic nose as a complementary screening Tool for breath testing in colorectal cancer[J]. Biosensors (Basel), 2025, 15(2):82.

[8] Van L E, Ramsoekh D, Daulton E, et al. Faecal volatile organic compounds to detect colorectal neoplasia in lynch Syndrome: a prospective longitudinal multicentre study [J]. Aliment Pharmacol Ther, 2025, 61(1):145-158.

[9] Bhattacharyya N, Sastimoglu Z, Armstrong D, et al. Fecal volatile organic compounds for colorectal cancer detection; a systematic review and Meta-analysis[J]. Comput Biol Med, 2025, 195:110561.

[10] Bakir-Gungor B, Temiz M, Canakcimaksutoglu B, et al. Prediction of colorectal cancer based on taxonomic levels of microorganisms and discovery of taxonomic biomarkers using the Grouping-Scoring-Modeling (G-S-M) approach[J]. Comput Biol Med, 2025, 187:109813.

[11] Han H Y, Li Y, Qi Y R, et al. Deciphering gut microbiome in colorectal cancer via robust learning methods[J]. Genes, 2025, 16(4):452.

[12] Lu F, Lei T, Zhou J, et al. Using gut microbiota as a diagnostic tool for colorectal cancer: machine learning techniques reveal promising results [J]. J Med Microbiol, 2023, 72(6):001699.

[13] Qiu Z, Cheng Y F, Liu H Y, et al. Screening colorectal cancer associated autoantigens through multi-omics analysis and diagnostic performance evaluation of corresponding autoantibodies[J]. BMC Cancer, 2025, 25(1):713.

[14] Yin H F, Xie J Y, Xing S, et al. Machine learning-based analysis identifies and validates serum exosomal proteomic signatures for the diagnosis of colorectal cancer[J]. Cell Rep Med, 2024, 5(8):101689.

[15] Yang R, Tsigelny I F, Kesari S, et al. Colorectal cancer detection via metabolites and machine learning[J]. Curr Issues Mol Biol, 2024, 46(5):4133-4146.

[16] Endo S, Nagata K, Utano K, et al. Development and validation of computer-aided detection for colorectal neoplasms using deep learning incorporated with computed tomography colonography[J]. BMC Gastroenterol, 2025, 25(1):149.

[17] Kim S, Lee J H, Park E J, et al. Prediction of microsatellite instability in colorectal cancer using a machine learning model based on PET/CT radiomics[J]. Yonsei Med J, 2023, 64(5):320-326.

[18] Grosu S, Wesp P, Graser A, et al. Machine learning-based differentiation of benign and premalignant colorectal polyps detected with CT colonography in an asymptomatic screening population; a proof-of-concept study[J]. Radiology, 2021, 299(2):326-335.

[19] Bharath E, Raja R V, Kalaivanan K, et al. Accurate colorectal cancer detection using a random hinge exponential distribution coupled attention network on pathological images[J]. Abdom Radiol, 2025, 50(7):2828-2857.

[20] Ke Q, Yap W S, Tee Y K, et al. Advanced deep learning for multi-class colorectal cancer histopathology: integrating transfer learning and ensemble methods[J]. Quant Imaging Med Surg, 2025, 15(3):2329-2346.

[21] Steimetz E, Simsek Z C, Saha A, et al. Deep learning model for detecting high-grade dysplasia in colorectal adenomas[J]. J Pathol Inform, 2025, 17:100441.

[22] Wu Y Z, Deng C G, Huang Z G, et al. Identification and predictive machine learning models construction of gut microbiota associated with lymph node metastasis in colorectal cancer[J]. mSystems, 2025, 10(8):e0033925.

[23] de Ridder J, de Wilt J H, Simmer F, et al. Incidence and origin of histologically confirmed liver metastases; an explorative case-study of 23 154 patients[J]. Oncotarget, 2016, 7(34):55368-55376.

[24] Ding Q Z, Li C Y, Wang C D, et al. Construction and interpretation of weight-balanced enhanced machine learning models for predicting liver metastasis risk in colorectal cancer patients[J]. Horm Cancer, 2025, 16(1):164.

[25] Feng S S, Zhou M L, Huang Z X, et al. A machine learning-based prediction model for colorectal liver metastasis [J]. Clin Exp Med, 2025, 25(1):156.

[26] Viganò L, Gentile D, Galvanin J, et al. Very early recurrence after liver resection for colorectal metastases: incidence, risk factors, and prognostic impact [J]. J Gastrointest Surg, 2022, 26(3):570-582.

[27] Kawashima J, Endo Y, Woldesenbet S, et al. Preoperative

identification of early extrahepatic recurrence after hepatectomy for colorectal liver metastases; a machine learning approach[J]. *World J Surg*, 2024, 48(11):2760-2771.

[28] Moorcraft S Y, Ladas G, Bowcock A, et al. Management of resectable colorectal lung metastases[J]. *Clin Exp Metastasis*, 2016, 33(3):285-296.

[29] Guo Z T, Zhang Z M, Liu L M, et al. Explainable machine learning for predicting lung metastasis of colorectal cancer[J]. *Sci Rep*, 2025, 15(1):13611.

[30] Zhang W Y, Guan X, Jiao S, et al. Development and validation of an artificial intelligence prediction model and a survival risk stratification for lung metastasis in colorectal cancer from highly imbalanced data; a multicenter retrospective study[J]. *Eur J Surg Oncol*, 2023, 49(12):107107.

[31] Liu X, Shu X M, Zhou Y J, et al. Construction of a risk prediction model for postoperative deep vein thrombosis in colorectal cancer patients based on machine learning algorithms[J]. *Front Oncol*, 2024, 14:1499794.

[32] Fang C Z, Shi W B, Qiao Y, et al. Construction of a risk factor prediction model for postoperative complications in elderly patients with colorectal cancer using machine learning[J]. *J Robot Surg*, 2025, 19(1):441.

[33] Ruan X Y, Fu S Y, Storlie C B, et al. Real-time risk prediction of colorectal surgery-related post-surgical complications using GRU-D model[J]. *J Biomed Inform*, 2022, 135:104202.

[34] Cai L Q, Yang D Q, Wang R J, et al. Establishing and clinically validating a machine learning model for predicting unplanned reoperation risk in colorectal cancer[J]. *World J Gastroenterol*, 2024, 30(23):2991-3004.

[35] Liang L, Xu N, Ding L F, et al. Combined inflammation-related biomarkers and clinicopathological features for the prognosis of stage II/III colorectal cancer by machine learning[J]. *BMC Cancer*, 2024, 24(1):1548.

[36] Liao K L, Yang Q J, Xu Y H, et al. Identification of signature of tumor-infiltrating CD8 T lymphocytes in prognosis and immunotherapy of colon cancer by machine learning[J]. *Clin Immunol*, 2023, 257:109811.

[37] Luo X J, Deng H, Xie F, et al. Prognostication of colorectal cancer liver metastasis by CE-based radiomics and machine learning[J]. *Transl Oncol*, 2024, 47:101997.

[38] Hou S, Heng S, Xie S, et al. Integrated multi-omics and machine learning reveal an immunogenic cell death-related signature for prognostic stratification and therapeutic optimization in colorectal cancer [J]. *Front Immunol*, 2025, 16:1606874.

[39] Yue J, Fang H, Yang Q, et al. Integrating multi-omics and machine learning methods reveals the metabolism of amino acids and derivatives-related signature in colorectal cancer[J]. *Front Oncol*, 2025, 15:1565090.

[40] Bahrambanan F, Alizamir M, Moradveisi K, et al. The development of an efficient artificial intelligence-based classification approach for colorectal cancer response to radiochemotherapy: deep learning vs. machine learning[J]. *Sci Rep*, 2025, 15(1):62.

[41] Zhu J, Zhang J, Lou Y, et al. Developing a machine learning-based prognosis and immunotherapeutic response signature in colorectal cancer: insights from ferroptosis, fatty acid dynamics, and the tumor microenvironment [J]. *Front Immunol*, 2024, 15:1416443.

[42] Wang S K, Li Y E, Wang Z H, et al. Constructing a mitochondrial-related genes model based on machine learning for predicting the prognosis and therapeutic effect in colorectal cancer[J]. *Horm Cancer*, 2025, 16(1):661.

[43] Weng M, Li T, Zhao J, et al. mRNAsi-related metabolic risk score model identifies poor prognosis, immunoevasive contexture, and low chemotherapy response in colorectal cancer patients through machine learning[J]. *Front Immunol*, 2022, 13:950782.

(收稿日期:2026-01-08 修回日期:2026-06-15)
(编辑:陈秋莲 王明丰)

(上接第 2180 页)

[26] Delong E R, Delong D M, Clarke-Pearson D L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach[J]. *Biometrics*, 1988, 44(3):837-845.

[27] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[C]//Adv Neural Inf Process Syst 30 (NIPS 2017). Long Beach:Curran Associates Inc., 2017:4765-4774.

[28] Vimbi V, Shaffi N, Mahmud M. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection[J]. *Brain Inform*, 2024, 11(1):10.

[29] Lombardi A, Diacono D, Amoroso N, et al. A robust framework to investigate the reliability and stability of explainable artificial intelligence markers of mild cognitive impairment and Alzheimer's Disease[J]. *Brain Inform*, 2022, 9(1):17.

[30] Winchester L M, Powell J, Lovestone S, et al. Red blood cell indices and anaemia as causative factors for cognitive function deficits and for Alzheimer's disease[J]. *Genome Med*, 2018, 10(1):51.

[31] Stevenson A, Lopez D, Khoo P, et al. Exploring erythrocytes as blood biomarkers for Alzheimer's disease[J]. *J Alzheimers Dis*, 2017, 60(3):845-857.

(收稿日期:2025-10-28 修回日期:2026-05-13)
(编辑:陈秋莲 廖薇薇)