

基于CT的放射组学对直肠癌T分期的多分类诊断价值

邱晨阳, 邹兵兵*

安徽医科大学第一附属医院普外科, 安徽 合肥

收稿日期: 2025年3月22日; 录用日期: 2025年4月15日; 发布日期: 2025年4月22日

摘要

目的: 研究基于增强CT的放射组学模型在直肠癌术前多分类T分期诊断中的可行性和有效性。方法: 本研究纳入了500例符合条件的直肠癌患者(T1期为48例, T2期为127例, T3期为259例, T4期为66例), 并随机分为训练组(n = 400例)和验证组(n = 100例)。直肠癌病变的感兴趣区由资深放射科医生进行像素化标注。我们使用pyradiomics从人工分割的感兴趣区域中提取特征, 然后采取10种不同机器学习算法构建了放射组学多分类模型, 并使用这10种算法构建临床基线数据多分类模型。使用ROC曲线、micro-average AUC、macro-average AUC和准确率评估多分类模型的性能。结果: 10种放射组学多分类模型中预测效果最好的Model SVM (micro-average AUC = 0.845, macro-average AUC = 0.777, 准确率 = 0.600)和10种临床基线数据多分类模型中预测效果最好的Model SVM (micro-average AUC = 0.841, macro-average AUC = 0.760, 准确率 = 0.610)相比, 在进行直肠癌T1、T2、T3和T4分期诊断方面表现更好。结论: 综上所述, 与临床基线数据多分类模型相比, 基于CT的放射组学多分类模型在进行直肠癌分期诊断方面具有更优异的预测性能。

关键词

直肠癌, 放射组学, 直肠癌分期多分类预测模型

Diagnostic Value of CT-Based Radiomics for Multi-Class T-Staging of Rectal Cancer

Chenyang Qiu, Bingbing Zou*

Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei Anhui

Received: Mar. 22nd, 2025; accepted: Apr. 15th, 2025; published: Apr. 22nd, 2025

*通讯作者。

文章引用: 邱晨阳, 邹兵兵. 基于 CT 的放射组学对直肠癌 T 分期的多分类诊断价值[J]. 临床医学进展, 2025, 15(4): 2513-2522. DOI: 10.12677/acm.2025.1541207

Abstract

Purpose: To investigate the feasibility and efficacy of a contrast-enhanced CT-based radiomics model in the preoperative multi-class T-staging diagnosis of rectal cancer. **Methods:** This study enrolled 500 eligible rectal cancer patients (48 T1, 127 T2, 259 T3, and 66 T4 cases) and randomly divided them into a training cohort (n = 400) and a validation cohort (n = 100). Regions of interest (ROIs) for rectal cancer lesions were manually segmented by senior radiologists. Using Pyradiomics, we extracted features from the annotated ROIs and constructed multi-class radiomics models with 10 machine learning algorithms. Additionally, clinical baseline models were developed using the same algorithms. Model performance was evaluated using ROC curves, micro-average AUC, macro-average AUC, and accuracy. **Results:** Among the 10 radiomics models, the Model SVM demonstrated the highest predictive performance (micro-average AUC = 0.845, macro-average AUC = 0.777, accuracy = 0.600). In comparison, the Model SVM from clinical baseline models achieved micro-average AUC = 0.841, macro-average AUC = 0.760, and accuracy = 0.610. The radiomics-based model exhibited superior diagnostic capability for distinguishing T1, T2, T3, and T4 stages. **Conclusion:** CT-based radiomics multi-class models outperform clinical baseline models in preoperative T-staging of rectal cancer, providing enhanced diagnostic accuracy and clinical utility.

Keywords

Rectal Cancer, Radiomics, Multi-Classification Prediction Model for Rectal Cancer Staging

Copyright © 2025 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

目前, 结直肠癌(Colorectal cancer, CRC)在全球肿瘤发病率和死亡率排名中分别位列于第3位和第2位[1][2]。其中直肠癌(Rectal cancer, RC)占结CRC的三分之一, 在东亚地区中具有最高的发病率[3]。如今对于RC患者的治疗多采用以手术为主, 联合放化疗的综合治疗方案, 使得RC患者的无病生存期得以增加, 让RC患者得到了更长的预期寿命和更好的生活质量, 同时基于术前分期的新辅助放化疗也使得肿瘤的R0切除率得以提高[4]-[6]。而选择最佳的治疗方案是建立在准确的术前分期的基础上的, 选择最佳治疗策略将对患者的生活质量和预后产生重要影响[7]。因此在术前进行RC的T分期的精确判断对于RC患者的精准治疗与选择最佳治疗策略有着较高的实用价值。如今, 基于CT技术的无创、实用、便捷和稳定等优势, 使其成为直肠癌术前分期的常规辅助评估方法。放射组学作为一种近期发展愈发兴盛的方法, 提出将CT、MRI等医学影像数据中的生物学和病理学特征转化为数值特征, 并将这些特征用于定量地进行肿瘤分期的评估[8]。相关研究证明, 通过使用放射组学对影像数据进行分析, 能够提取出肉眼难以分辨的形态学特征和纹理特征等, 从而帮助RC患者选择最佳治疗策略[9][10]。使用放射组学训练基于CT的模型以预测食管鳞状细胞癌T分期和长度方面的应用也得到证明[11]。因此, 我们旨在使用放射组学的方法进行RC的CT分期的术前预测, 在临床实践中为RC的预处理评估提供一种非侵入性、简单和经济的方法。

2. 材料和方法

2.1. 患者选择

本研究回顾性地分析了2017年1月至2023年6月在安徽医科大学第一附属医院接受RC根治手术

治疗的共计 500 例 RC 患者。(1) 纳入标准: ① 临床资料完整; ② 未接受新辅助化疗; ③ 具有完整的接受直肠高分辨 CT 平扫及盆腔 CT 动态增强扫描图像; ④ 手术过程顺利, 肿瘤得到根治性切除; ⑤ RC 为单发且为直肠原发恶性肿瘤; ⑥ 病理诊断明确为腺癌。(2) 排除标准: ① 临床资料不完整; ② 接受新辅助化疗; ③ 仅接受直肠高分辨 CT 平扫, 未行盆腔 CT 动态增强扫描, 或者 CT 图像质量差, 出现伪影; ④ 未能根治性切除或因肿瘤分期较晚无法施行手术治疗; ⑤ CRC 为多发或为直肠继发恶性肿瘤; ⑥ 病理诊断缺少或不为腺癌。最终, 500 名符合条件的患者按照 4: 1 的比例, 随机分为训练队列($n=400$ 例)和验证队列($n=100$ 例)。

2.2. 影像获取与标注

本研究使用门静脉期的增强 CT 图像来分割肿瘤病变[12]。在将每个患者的增强 CT 图像导出后使用 ITK-SNAP 软件(version 3.8.0, <http://www.itksnap.org/>)对每张 RC 增强 CT 切片进行手动分割。分割与标注工作是由两位经验丰富的放射科医师(工作时间均在 10 年以上)沿着肿瘤的轮廓像素级描绘每个切片上的肿瘤区域。增强 CT 的 T 分期评估是根据第 8 版 AJCC TNM 分类标准[13]。人工标注完成的增强 CT 如图 1 所示。

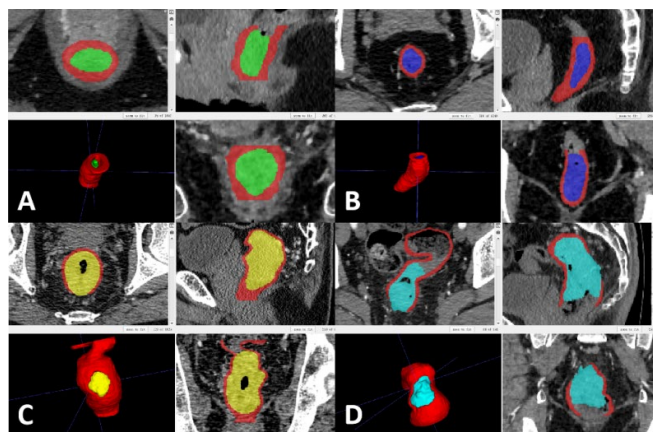


Figure 1. The results of pixel-level annotation for tumors with different stages (A)~(D). Red indicates the normal structure of the rectal wall, green represents the extent of involvement of T1 stage tumors, purple indicates the extent of involvement of T2 stage tumors, yellow represents the extent of involvement of T3 stage tumors, and blue indicates the extent of involvement of T4 stage tumors

图 1. 将不同分期肿瘤进行像素级标注的结果(A)~(D)。红色表示直肠肠壁正常结构, 绿色表示 T1 期肿瘤累及范围, 紫色表示 T2 期肿瘤累及范围, 黄色表示 T3 期肿瘤累及范围, 蓝色表示 T4 期肿瘤累及范围

2.3. 数据预处理

首先将数据标准化, 并对 CT 切片进行对比度调整、旋转以及裁剪和调整大小, 确保图像数据的一致性和可处理性(如图 2 所示)。



Figure 2. Preprocessing of CT slices (A)~(E). (A) shows the original image, (B) is the standardized version, (C) is the cropped and resized image, (D) is the contrast adjusted image, and (E) is the rotated image

图 2. CT 切片的预处理(A)~(E)。(A)为原图, (B)为标准化, (C)为裁剪和调整大小, (D)为对比度调整, (E)为旋转

2.4. 放射组学影像特征提取及模型构建

完成肿瘤分割后, 将带有标记的 CT 图像和通信(DICOM)图像导出到 onekey 平台中, 通过 onekey 平台中的影像组学模块调用 pyradiomics 算法从每个直肠癌患者的直肠区域中提取 107 个放射组学特征。onekey 平台中的影像组学模块筛选特征及构建模型的步骤如下: 首先将所有放射组学特征使用 Z-score 对数据进行正则化, 将数据变化到服从 $N(0, 1)$, 然后通过 t-test 进行影像组学特征的初步筛选($P < 0.05$ 为显著)。然后, 对于其中具有高重复性的特征, 使用皮尔逊相关系数计算特征之间的相关性, 并保留任意两个相关性高的特征(皮尔逊相关系数大于 0.9)中的一个特征。最后, 使用最小绝对值收敛和选择算法 (Least Absolute Shrinkage and Selection Operator, LASSO)进行进一步的降维和优化特征选择, 选取其中的非 0 项作为后续模型的特征。经过这些步骤, 最终保留了 12 个特征, 包括了 1 个一阶特征、5 个形状特征和 6 个高阶纹理特征, 进行放射组学模型构建。使用 10 种不同的机器学习算法进行任务学习, 构建不同的模型, 采用 ROC、micro-average AUC、macro-average AUC 和准确率对模型的预测性能进行评估, 具体流程见图 3。

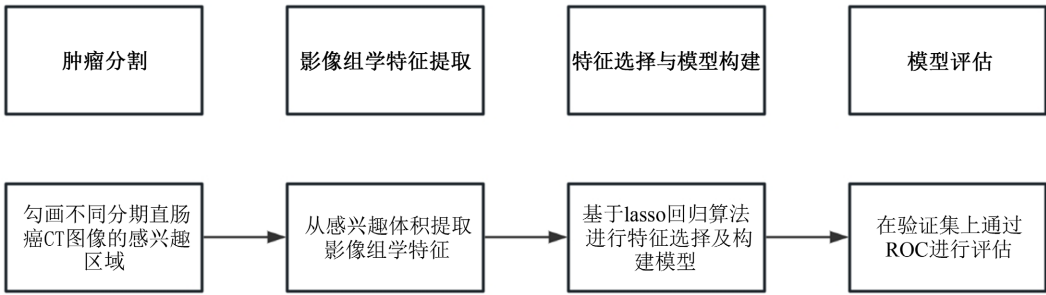


Figure 3. Workflow for constructing the radiomics model
图 3. 放射组学模型构建的流程

2.5. 临床基线数据模型的构建

为排除不必要因素的影响, 临床基线数据模型的构建过程与放射组学模型的构建几乎相同。首先, 使用皮尔逊相关系数计算临床基线数据之间的相关性进行临床基线数据的筛选, 这一过程筛选出了 6 个临床基线数据(包括 N 分期、肿瘤长度、脉管内癌栓、神经侵犯、CA199 水平以及性别), 并使用 LASSO 回归算法优化。同样, 采用相同的训练集和验证集, 并使用同放射组学模型相同的 10 种机器学习算法进行任务学习, 然后选择效果最优的算法用于临床基线数据模型的构建。采用同放射组学模型相同的指标对临床基线数据模型的预测性能进行评估。

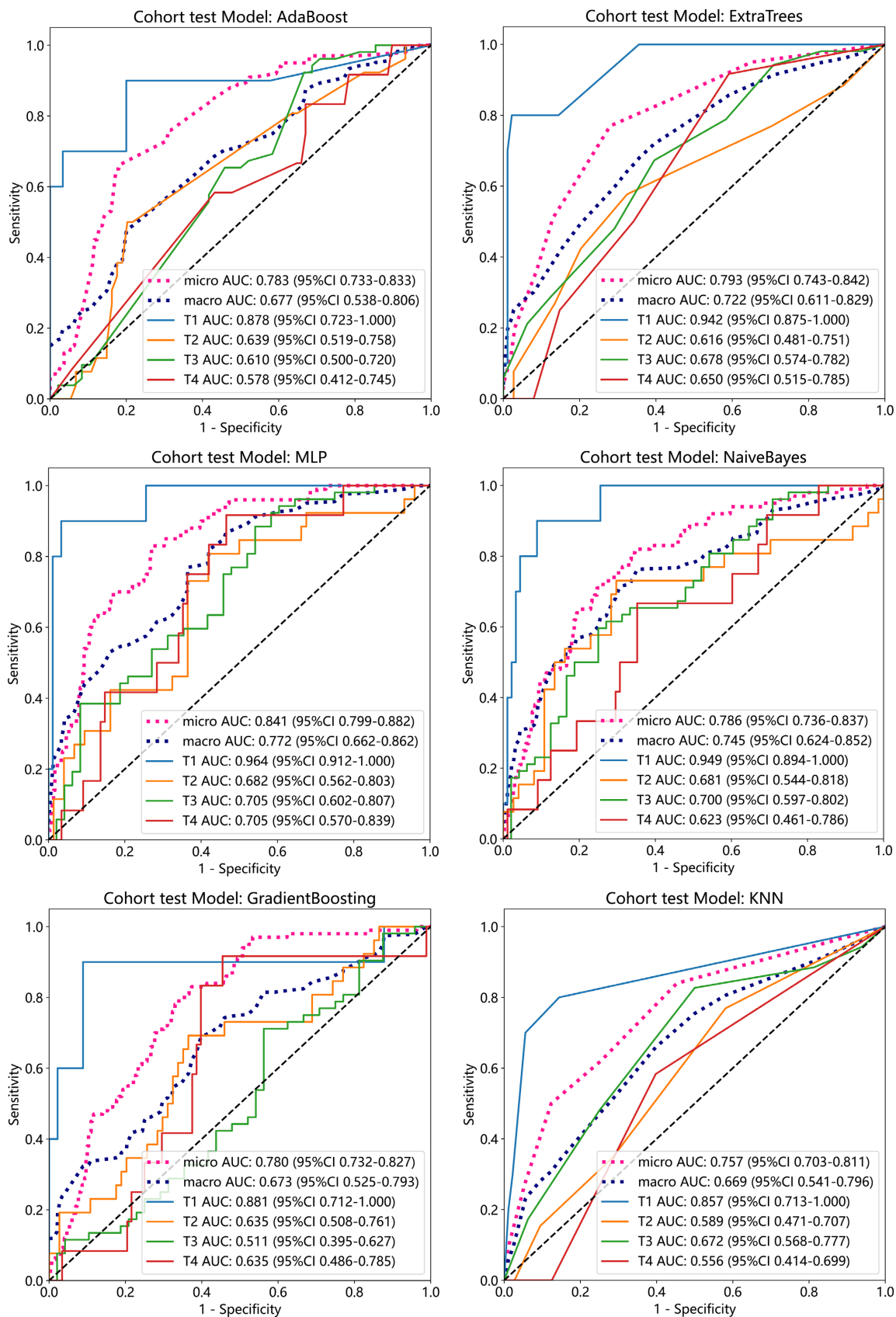
3. 结果

3.1. 样本详细信息

本研究共纳入 500 名患者, 其中 324 名为男性, 176 名为女性。年龄范围在 31 岁至 90 岁之间。直肠上部、中部和下部肿瘤的患者分别为 133、300 和 67 例。其中, 48 名患者为 TNM 分期 T1 期, 127 名患者为 T2 期, 259 名患者为 T3 期, 66 名患者为 T4 期。

3.2. 放射组学模型的预测性能

放射组学多分类模型在测试集中的 micro-average AUC、macro-average AUC 及 ROC 见图 4。在对比分析中, Model SVM 展现出显著优于其他放射组学模型的预测性能。首先, 其 micro-average AUC 达到



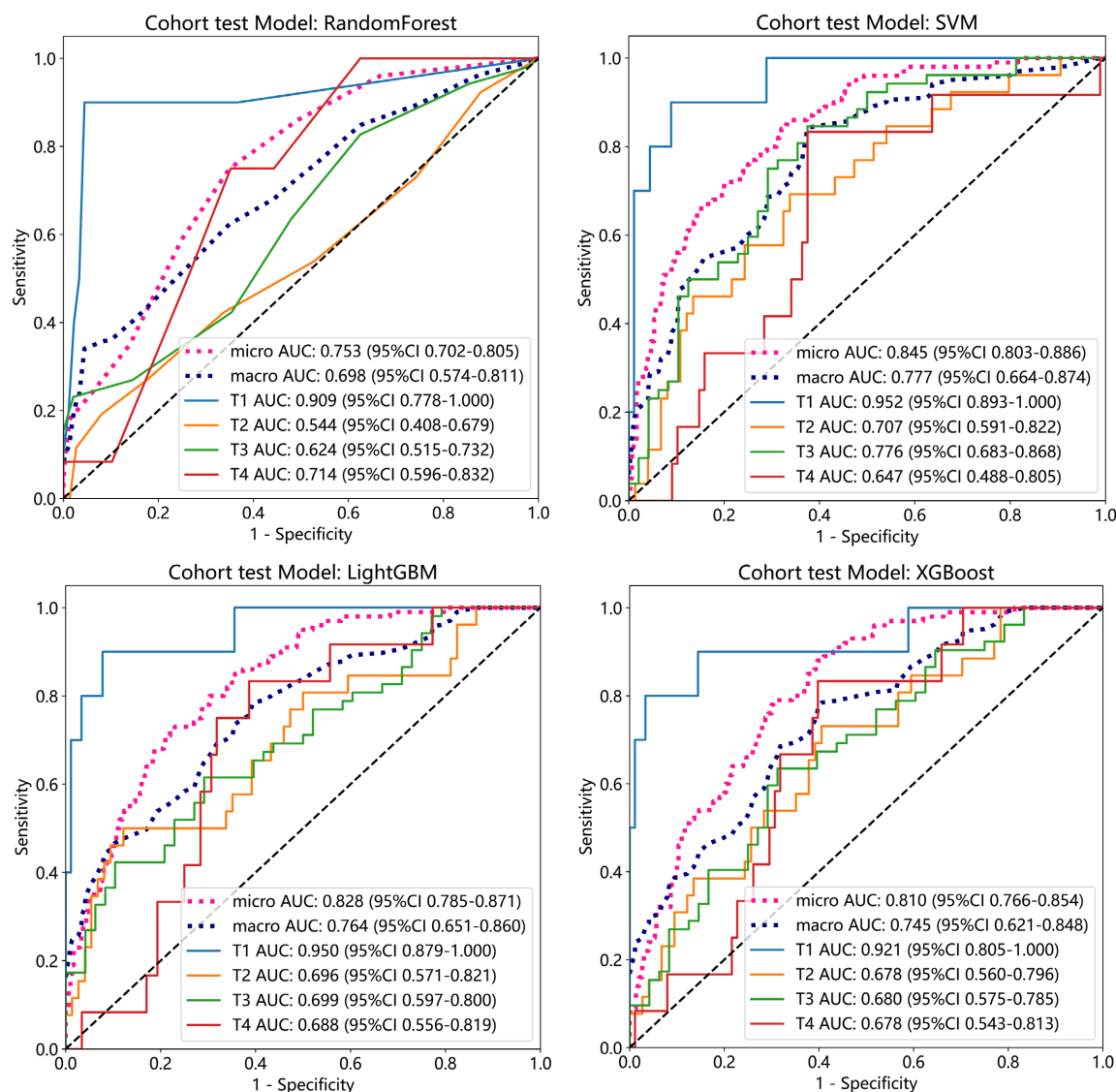


Figure 4. ROC of the radiomics multi-classification model

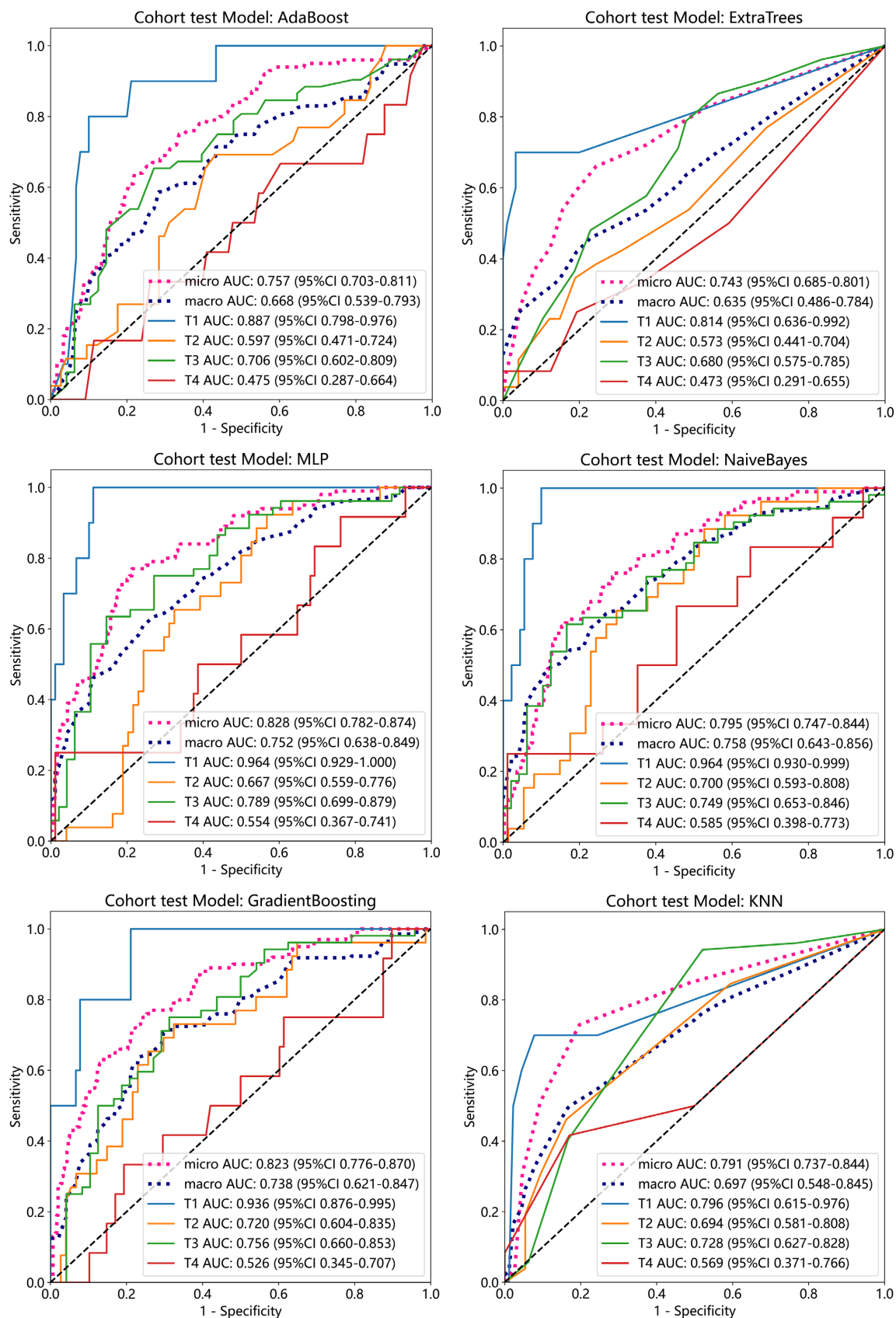
图 4. 放射组学多分类模型的 ROC

0.845 (95% CI 为 0.803~0.886), 为所有模型中最高值, 且在关键临床分类任务中表现尤为突出: 针对 T1 类别的预测, Model SVM 的 AUC 高达 0.952 (95% CI: 0.893~1.000)。此外, Model SVM 在 T3 (AUC 为 0.776)和 T4 (AUC 为 0.647)分类中均保持稳定性能, 未出现显著短板。综合全局性能、关键类别判别精度及跨类别稳定性, Model SVM 在多分类放射组学任务中展现出最优的泛化能力与临床适用性。这些数据表明放射组学模型中 Model SVM 的预测性能要优于其他放射组学模型。

3.3. 临床基线数据模型的预测性能

在临床基线数据模型训练时采用的测试集中, 各个模型的 micro-average AUC 和 macro-average AUC 及 ROC 见图 5。该图表明临床基线数据模型中 Model SVM 的预测性能要优于其他临床基线数据模型。

各放射组学模型及临床基线数据模型的准确率见表 1。



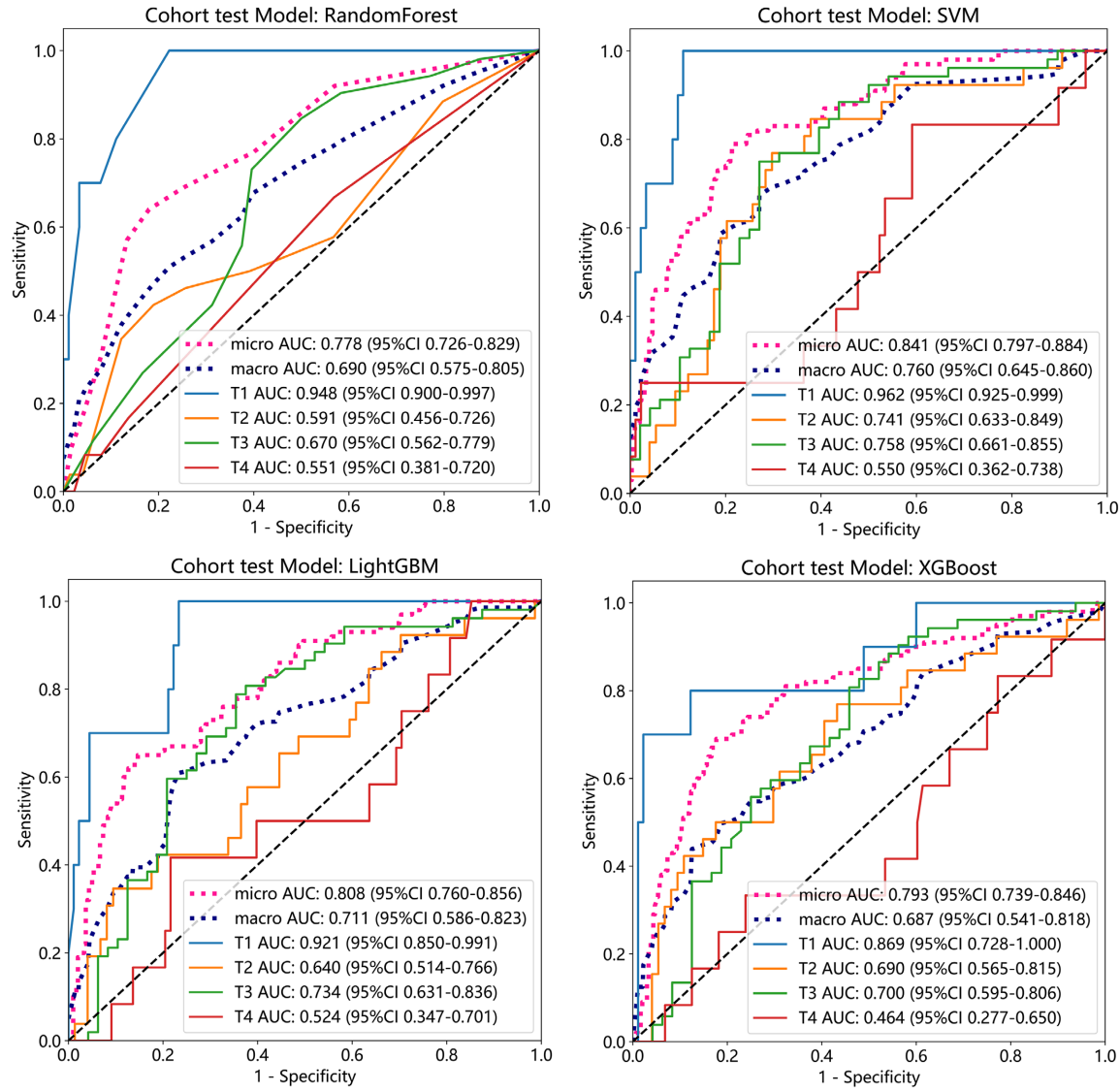


Figure 5. ROC of the clinical baseline data multi-classification model
图 5. 临床基线数据多分类模型的 ROC

Table 1. Accuracy of radiomics model and clinical baseline data model
表 1. 放射组学模型及临床基线数据模型的准确率

	放射组学模型	临床基线数据模型
Model AdaBoost	0.620	0.530
Model ExtraTrees	0.550	0.530
Model GradientBoosting	0.510	0.610
Model KNN	0.570	0.620
Model LightGBM	0.590	0.620
Model MLP	0.620	0.570
Model NaiveBayes	0.490	0.580

续表

Model RandomForest	0.470	0.580
Model SVM	0.600	0.610
Model XGBoost	0.530	0.590

4. 讨论

本研究的重要亮点是我们对 RC 的 T 分期进行了多分类预测。在先前的研究中, RC 的诊断模型多是二分类模型, 如 Hou M [14] 等人将 T1 和 T2 期 RC 划分为早期肿瘤, 将 T3 和 T4 期 RC 划分为进展期肿瘤, 使用 RC 高分辨率 MRI 进行训练, 得到 AUC 为 0.810、准确率为 0.773 的二分类诊断模型 Model HRT2, 这可能是由于二分类模型在实践中更易操作和解释, 而我们在进行 RC 的分期预测时, 不仅将 RC 预测结果分为早期肿瘤和进展期肿瘤, 更进一步将预测结果分为 T1、T2、T3 和 T4, 这有助于更精细地了解疾病的发展和临床特征。

另一突出的亮点是我们在标记样本时所遵循的原则。我们采用了一种精心设计的像素级标记方法, 确保每个样本都反映了肿瘤组织原有的特征, 同时也清晰地反映了肿瘤的分期情况。

5. 结论

综合来看, 在所有的 20 个多分类诊断模型中, 放射组学多分类模型中预测效果最好的 Model SVM, 其整体上优于其他模型, 对所有类别的分类性能都相对较好, 没有明显的短板, 在处理 RC T 分期多分类诊断任务上在 20 个放射组学模型中表现最为出色。

本研究仍有以下发展方向: 首先, 可结合自动分割和深度学习方法减少图像标注工作量, 提升直肠癌诊断模型性能。其次, 通过多中心数据扩充样本库。本研究为单中心, 需要进行外部验证。通过将模型应用于多个医疗中心的数据, 可验证其在不同数据集的性能和泛化能力。

参考文献

- [1] Siegel, R.L., Miller, K.D., Fuchs, H.E. and Jemal, A. (2021) Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians*, **71**, 7-33. <https://doi.org/10.3322/caac.21654>
- [2] 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>
- [3] Lu, L., Mullins, C.S., Schafmayer, C., Zeißig, S. and Linnebacher, M. (2021) A Global Assessment of Recent Trends in Gastrointestinal Cancer and Lifestyle-Associated Risk Factors. *Cancer Communications*, **41**, 1137-1151. <https://doi.org/10.1002/cac2.12220>
- [4] Kasi, A., Abbasi, S., Handa, S., Al-Rajabi, R., Saeed, A., Baranda, J., et al. (2020) Total Neoadjuvant Therapy vs Standard Therapy in Locally Advanced Rectal Cancer: A Systematic Review and Meta-Analysis. *JAMA Network Open*, **3**, e2030097. <https://doi.org/10.1001/jamanetworkopen.2020.30097>
- [5] Sebag-Montefiore, D., Stephens, R.J., Steele, R., Monson, J., Grieve, R., Khanna, S., et al. (2009) Preoperative Radiotherapy versus Selective Postoperative Chemoradiotherapy in Patients with Rectal Cancer (MRC CR07 and NCIC-CTG C016): A Multicentre, Randomised Trial. *The Lancet*, **373**, 811-820. [https://doi.org/10.1016/s0140-6736\(09\)60484-0](https://doi.org/10.1016/s0140-6736(09)60484-0)
- [6] VanderWalde, N. and Grothey, A. (2020) Personalizing Treatment for Rectal Cancer: Total Neoadjuvant Therapy Is Leading the Way. *JAMA Network Open*, **3**, e2030508. <https://doi.org/10.1001/jamanetworkopen.2020.30508>
- [7] Linden, T., Hanses, F., Domingo-Fernández, D., DeLong, L.N., Kodamullil, A.T., Schneider, J., et al. (2021) Machine Learning Based Prediction of COVID-19 Mortality Suggests Repositioning of Anticancer Drug for Treating Severe Cases. *Artificial Intelligence in the Life Sciences*, **1**, Article 100020. <https://doi.org/10.1016/j.ailsci.2021.100020>
- [8] Gillies, R.J., Kinahan, P.E. and Hricak, H. (2016) Radiomics: Images Are More than Pictures, They Are Data. *Radiology*, **278**, 563-577. <https://doi.org/10.1148/radiol.2015151169>

-
- [9] Khorrami, M., Bera, K., Leo, P., Vaidya, P., Patil, P., Thawani, R., *et al.* (2020) Stable and Discriminating Radiomic Predictor of Recurrence in Early Stage Non-Small Cell Lung Cancer: Multi-Site Study. *Lung Cancer*, **142**, 90-97. <https://doi.org/10.1016/j.lungcan.2020.02.018>
 - [10] Zheng, J., Kong, J., Wu, S., Li, Y., Cai, J., Yu, H., *et al.* (2019) Development of a Noninvasive Tool to Preoperatively Evaluate the Muscular Invasiveness of Bladder Cancer Using a Radiomics Approach. *Cancer*, **125**, 4388-4398. <https://doi.org/10.1002/cncr.32490>
 - [11] Xu, C., Peng, Y., Zhu, W., Chen, Z., Li, J., Tan, W., *et al.* (2022) An Automated Approach for Predicting Glioma Grade and Survival of LGG Patients Using CNN and Radiomics. *Frontiers in Oncology*, **12**, Article 969907. <https://doi.org/10.3389/fonc.2022.969907>
 - [12] Sun, K., Hu, H., Chen, S., Ye, J., Li, G., Chen, L., *et al.* (2020) CT-Based Radiomics Scores Predict Response to Neoadjuvant Chemotherapy and Survival in Patients with Gastric Cancer. *BMC Cancer*, **20**, Article No. 468. <https://doi.org/10.1186/s12885-020-06970-7>
 - [13] Amin, M.B., Greene, F.L., Edge, S.B., Compton, C.C., Gershenwald, J.E., Brookland, R.K., *et al.* (2017) The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge from a Population-Based to a More “Personalized” Approach to Cancer Staging. *CA: A Cancer Journal for Clinicians*, **67**, 93-99. <https://doi.org/10.3322/caac.21388>
 - [14] Hou, M., Zhou, L. and Sun, J. (2022) Deep-Learning-Based 3D Super-Resolution MRI Radiomics Model: Superior Predictive Performance in Preoperative T-Staging of Rectal Cancer. *European Radiology*, **33**, 1-10. <https://doi.org/10.1007/s00330-022-08952-8>