



基于多参数MRI的机器学习模型在肾脏实性小肿瘤定性诊断中的应用研究

崔梦秋¹, 王鹤², 蒋佳慧³, 陈思璐⁴, 曹阳⁵, 白旭⁶, 康欢欢¹, 郝雨薇¹, 刘百川¹, 杨大为³,
杨正汉³, 叶慧义¹, 王海屹¹

¹解放军总医院第一医学中心放射诊断科

²北京大学第一医院医学影像科

³首都医科大学附属友谊医院放射科

⁴解放军总医院第七医学中心放射诊断科

⁵黑龙江省鸡西市鸡矿医院放射科

⁶解放军总医院第二医学中心放射科

1. 研究背景

- 肾脏小肿瘤 (SRM) 为对比增强后直径 $\leq 4\text{cm}$ 的肾脏肿瘤
- SRM 占有所有已确诊肾脏肿瘤的 48% ~ 60%，占已切除肾脏肿瘤的 38%
- **良恶性诊断困难**
 - 在被切除的 SRM 中，经病理验证为 **良性比例可高达 30%**

难鉴别
易误切

TABLE 5. Histopathology results of laparoscopic vs open partial nephrectomy for renal tumor

	Laparoscopic	Open	p Value
% Pathological diagnosis:			
RCC	70.0	85.0	0.01 (Pearson chi-square test)
Benign	30.0	15.0	0.002 (Fisher exact test)
% RCC:	70.0	85.0	
Clear cell	51.4	57.6	
Papillary	35.7	12.9	
Other	12.8	9.4	
% Benign:	30.0	15.0	0.052 (Fisher exact test)
Oncocytoma	43.3	86.7	
Cyst	10.0	0.00	
Angiomyolipoma	23.3	6.7	
Other	0.0	6.7	

Wilcox RN et al. Curr Oncol Rep. 2020
 Finelli A et al. J Clin Oncol. 2019
 Gill, IS et al. N Engl J Med. 2010
 Suarez-Ibarrola, et al. Cancers .2020
 Silverman, Stuart G et al. Radiol. 2019
 Chang Emily H et al. Urol Oncol. 2020

1. 研究背景



经皮穿刺活检

- 诊断准确率高
- 有创检查
- 潜在并发症
(发生率为1% ~ 5%)
- 假阴性结果...



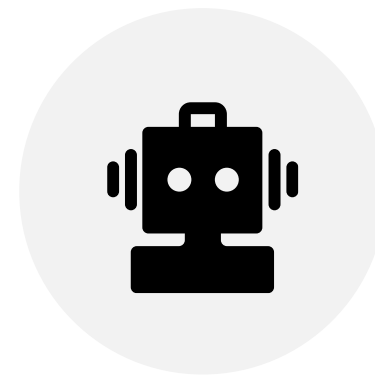
CT

- 成像速度快
- 软组织分辨率差
- 中等程度敏感性
- 电离辐射损伤...



MRI

- 软组织分辨率高
- 多平面、多序列、多参数成像评估
- 受影像科医生水平的影响...



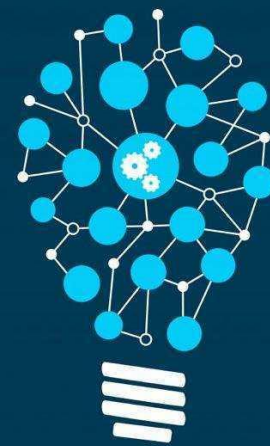
CT

- 无创、可重复、超前
- 性能稳定、减少主观性
- 减少低年资临床医生的负担

2. 研究目的及意义

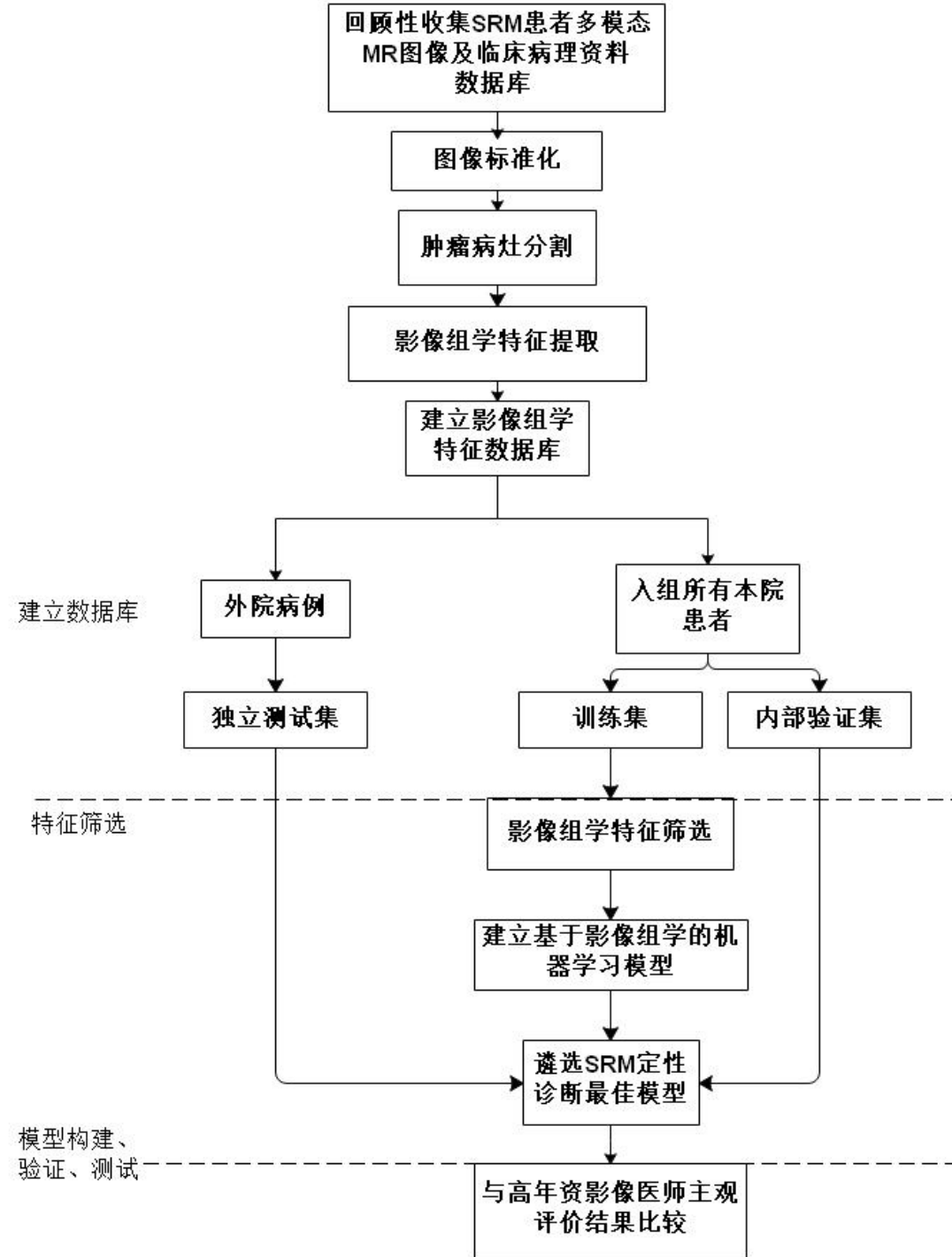
- 目的：评估基于多参数MRI建立的机器学习模型对SRM良恶性进行分类的效能
- 意义：减轻临床工作量，为SRM患者提供个性化的诊疗方案

MACHINE
LEARNING



3.研究方法——研究流程

- ①病例的收集及分组
- ②影像标准化处理
- ③病灶分割
- ④影像组学特征提取及筛选
- ⑤构建机器学习模型
- ⑥模型诊断能力评估及比较



3. 研究流程—— 1. 病例的收集及分组

• 病例的收集及分组

- 收集本院2014年1月至2021年3月符合纳排标准的病例798例作为训练集及验证集
- 收集友谊医院、北京大学第一医院于2017年3月至2021年5月符合纳排标准的81例SRM病例作为独立测试集

入组标准 (需符合下列所有标准)

- 肿瘤最大径 $\leq 4\text{cm}$
- 接受手术治疗或穿刺活检且病理结果完整有效
- 术前在GE3.0T MR扫描仪上完成检查
- MR图像序列完整, 包括T1WI、T2WI、DWI、ADC图、同反相位及动态增强三期图像
- 患者临床及病理信息完整: 年龄、性别、BMI指数、症状、病理亚型等

排除标准 (符合下列标准之一)

- MR图像质量欠佳 (图像清晰度不足以用来诊断; 伪影严重等)
- 患者病理信息及临床信息有无法补充或忽略的重大缺失

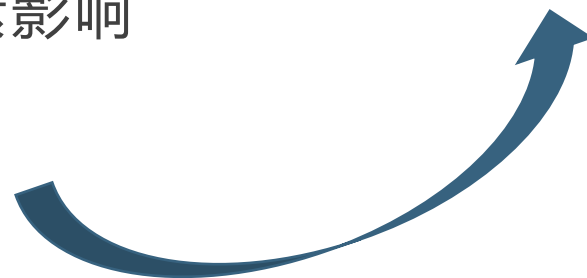
3. 研究流程—— 2. 影像标准化处理

图像异质性严重

- 多个中心：301、友谊医院、北京大学第一医院
- 多种病理亚型：透明细胞癌、非透明细胞癌、血管平滑肌脂肪瘤...
- 多个序列：T1、T2、增强三期（CP、NP、DP）
- 多种扫描参数：具体扫描参数受多种因素影响

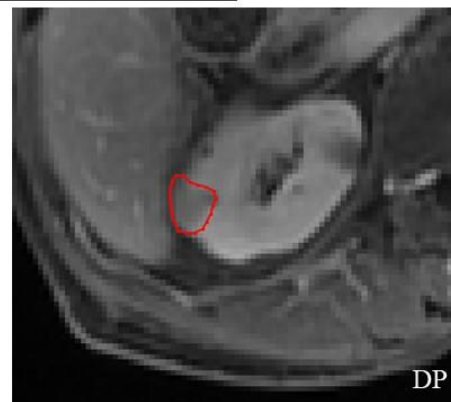
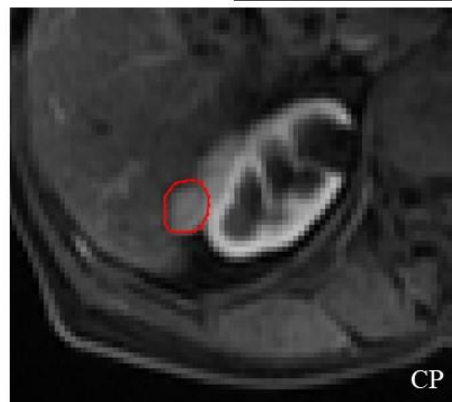
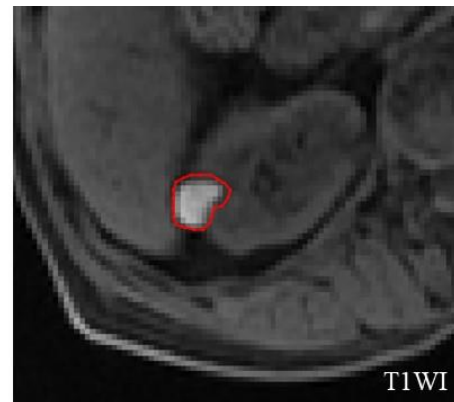
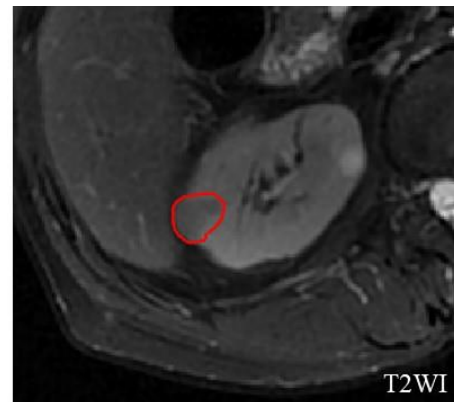
解决方法

- 重采样
 - 将不同体素大小的图像重采样到相同的体素大小



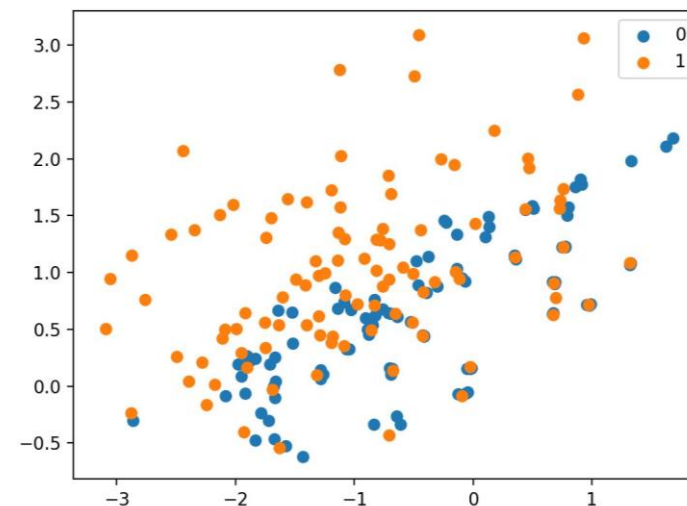
3. 研究流程—— 3. 病灶分割

- 由两位放射科医师对图像进行共同分割
 - 勾画肿瘤边界尽量贴合肿瘤轮廓，无须扩大肿瘤范围
 - 如第一层、最后一层肿瘤边界模糊可不勾画该层面，从而降低部分容积效应的干扰
 - 假包膜需作为肿瘤一部分勾画在内
 - 肿瘤位于肾窦时，需避免将肾窦脂肪勾画在内



3. 研究流程—— 4. 组学特征提取及筛选

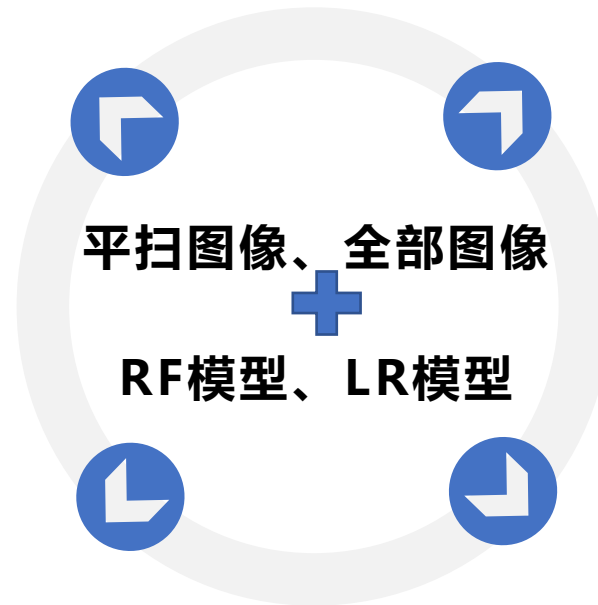
- 使用Pyradiomics 3.0.1软件包对五个序列进行特征提取，从每个患者的MRI中共提取了6231个特征
- 特征筛选
 - 单变量特征选择方法 (t-test)：可筛选相关性最高的变量，保留相关性良好的变量
 - 最小绝对收缩选择算子 (the least absolute shrinkage and selection operator, LASSO) 算法：采用十折交叉验证提升特征的鲁棒性
- 使用欠采样方法 (NearMiss-3技术) 解决类别不平衡的现象



3. 研究流程—— 5. 构建机器学习模型

- 基于平扫图像的random forest模型

- 基于全序列图像的random forest模型



- 基于全序列图像的logistic regression模型

- 基于平扫图像的logistic regression模型

3. 研究流程—— 6. 模型诊断能力评估及比较

诊断表现

- 分别计算各模型的灵敏度、特异度、准确度
- 绘制各模型的ROC曲线，分析ROC曲线并评估诊断效能

模型性能评估

- 使用Delong检验比较各模型的AUC值
- 使用Mcnermar检验比较各模型灵敏度和特异度

临床实际性能

与两位专家的诊断表现进行比较

3. 研究流程—— 7. 统计学方法

- 采用 χ^2 检验、fisher确切概率法分析定性数据
- 采用独立样本t检验分析符合正态分布的定量数据，采用Mann-Whitney U检验分析不符合正态分布的定量数据
- 使用准确度、灵敏度、特异度对高年资放射科医师和模型的诊断能力进行评估
- 使用ROC曲线及AUC值分析模型的性能
- 采用McNemar检验对各模型及医生诊断的灵敏度及特异度进行比较
- 采用DeLong检验对不同模型及两位医生的多个ROC曲线进行显著性检验
- $P < 0.05$ 视为具有统计学意义



4. 研究结果—— 1. 患者临床资料

- 良恶性患者间年龄、性别、肿瘤发生左右侧、吸烟史、高血压史、BMI指数间存在统计学差异

		良性	恶性	P值
年龄 (岁)		50.7±12.5	53.4±11.8	0.030
性别				< 0.001
	男	36(34.6%)	578(74.6%)	
	女	68(65.4%)	197(25.4%)	
左右侧				0.002
	左	65(62.5%)	358(46.2%)	
	右	39(37.5%)	417(53.8%)	
吸烟史				< 0.001
	有	21(20.2%)	294(37.9%)	
	无	83(79.8%)	481(62.1%)	
高血压史				0.032
	有	73(70.2%)	459(59.2%)	
	无	31(29.8%)	316(40.8%)	
BMI				0.009
	正常 (BMI < 24)	92(88.5%)	599(77.3%)	
	超重及肥胖 (BMI ≥ 24)	12(11.5%)	176(22.7%)	
合计		81	798	

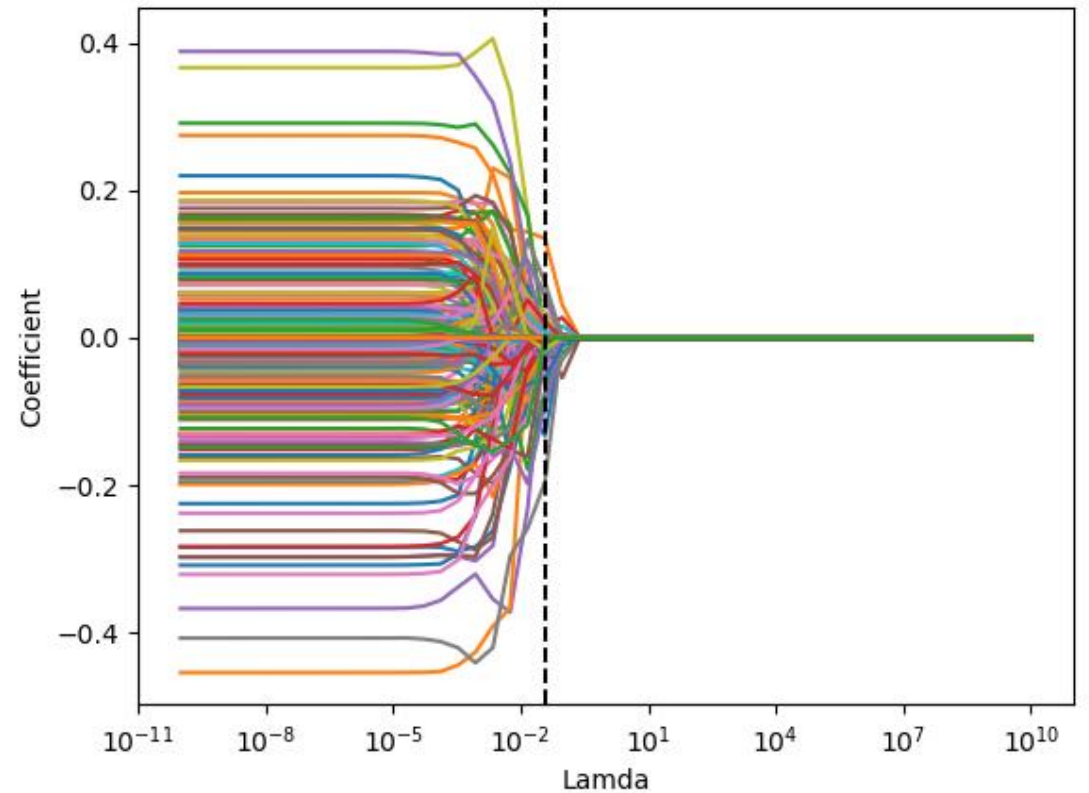
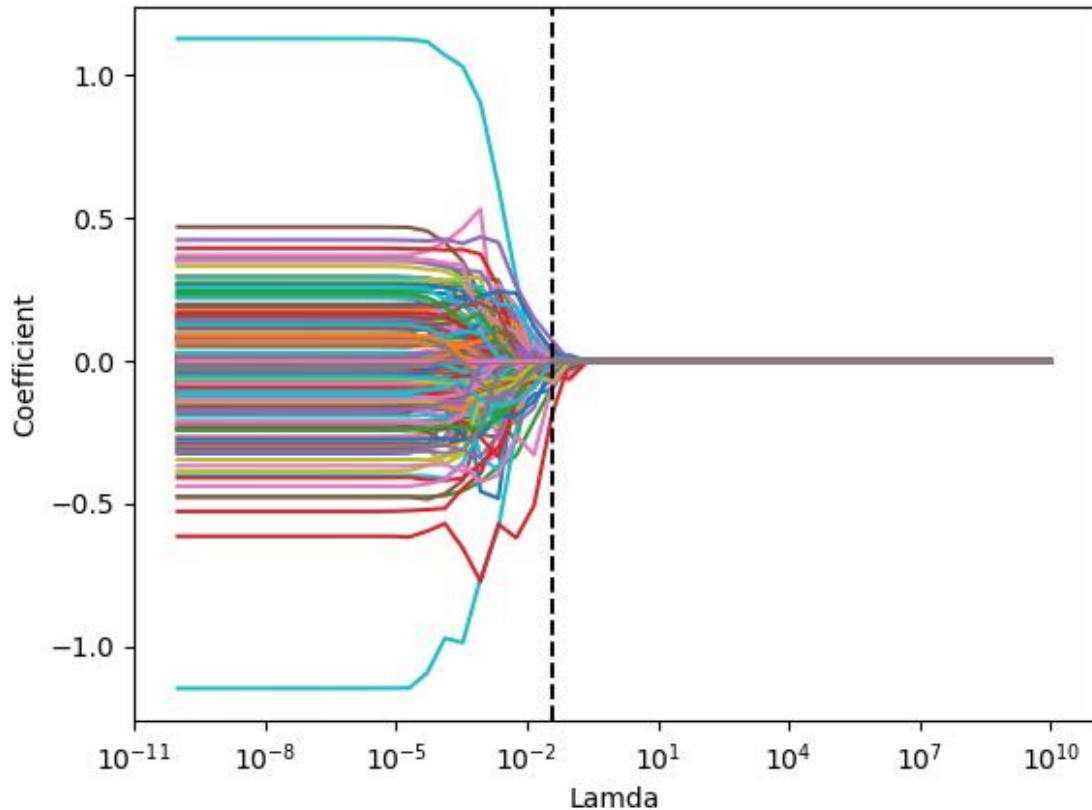
4. 研究结果—— 2. 患者病理资料

病理类型	本院n(%) ^a	外院n(%)
良性	94(11.8%)	10(12.3%)
PEComa ^b	76(9.5%)	9(11.1%)
嗜酸细胞腺瘤	12(1.5%)	1(1.2%)
其他 ^c	6(0.8%)	0
恶性	704(88.2%)	71(87.7%)
透明细胞癌	623(78.1%)	55(67.9%)
乳头状肾细胞癌	39(4.9%)	5(6.2%)
嫌色细胞癌	32(4.0%)	7(8.6%)
其他 ^d	10(1.3%)	4(4.9%)
合计	798	81

- a.表示数量（本组百分比）
- b. PEComa包括血管平滑肌脂肪瘤及上皮样血管平滑肌脂肪瘤
- c.包括后肾腺瘤，黄色肉芽肿性炎，混合性上皮和间质肿瘤，毛细血管瘤，平滑肌瘤，肾小球旁器细胞瘤
- d. 包括透明乳头状肾细胞癌、MiT家族性易位性肾细胞癌、嗜酸性实性和囊性肾细胞癌、弥漫大B细胞淋巴瘤、未分类的肾细胞癌、小圆细胞肿瘤

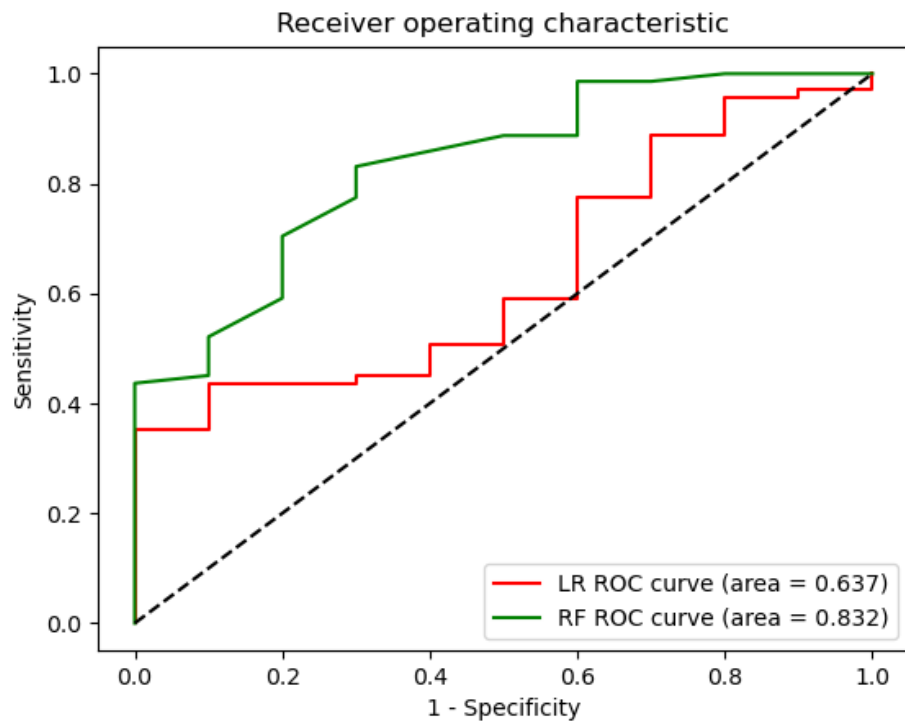
4. 研究结果—— 3. 组学特征筛选

- 通过单变量选择法、LASSO算法对特征进行降维，最终从平扫组及全序列组图像中分别筛选出27个、37个独立预测能力最优且鲁棒性最好的组学特征。

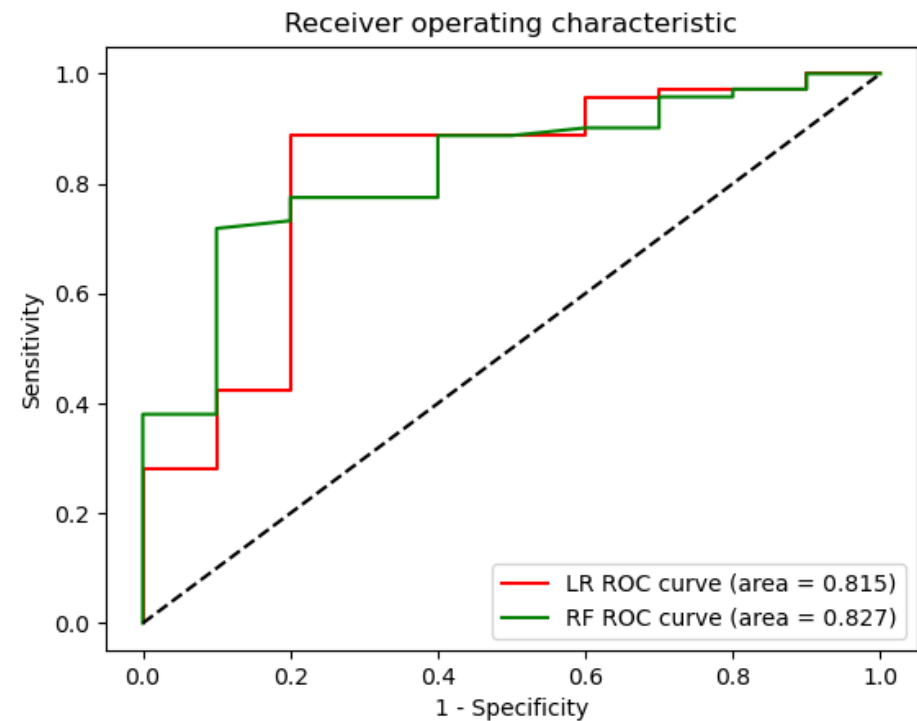


4. 研究结果—— 4. 模型及专家的诊断表现

- 在外院独立测试集中，基于平扫图像的RF模型和LR模型，基于全部图像的RF模型和LR模型的AUC值分别为0.832、0.637、0.827、0.815。



基于平扫图像的LR和RF模型ROC曲线分析



图基于全序列图像的LR和RF模型ROC曲线分析

4. 研究结果—— 4. 模型及专家的诊断表现

- 在外院独立测试集中，基于平扫图像的RF模型具有最佳诊断效能（AUC值），其灵敏度、特异度、准确度及AUC值分别为0.592、0.800、0.617、0.832
- 两位专家（专家A，工作年限20年；专家B,工作年限22年）使用盲法分别对外院病例进行评估

图像	模型	准确度	灵敏度	特异度	AUC (95%CI)
基于平扫图像 (T ₁ WI+ T ₂ WI)	RF	0.617	0.592	0.800	0.832 (0.716-0.933)
	LR	0.667	0.704	0.400	0.637 (0.501-0.780)
基于全序列图像 (T ₁ WI+ T ₂ WI +CP+NP+DP)	RF	0.728	0.704	0.900	0.827 (0.720-0.924)
	LR	0.753	0.746	0.800	0.815 (0.680-0.948)
专家A		0.840	0.859	0.700	0.780 (0.649-0.908)
专家B		0.914	0.972	0.500	0.736 (0.601-0.875)

4. 研究结果—— 5. 灵敏度及特异度比较

- 采用McNemar比较模型及专家间的灵敏度与特异度，结果表明
 - 灵敏度
 - 四个模型间比较
 - 基于全部图像的LR模型高于基于平扫图像的RF模型 ($P < 0.05$)
 - 与两位专家比较
 - 四个模型的灵敏度均低于专家B ($P < 0.05$)
 - 基于平扫图像的RF模型的灵敏度低于专家A ($P < 0.05$)
 - 特异度
 - 四个模型间均无差异
 - 四个模型的特异度与两位专家间均无差异

4. 研究结果—— 6. AUC值的比较

- 采用Delong检验比较模型间及专家间的灵敏度与特异度，结果表明 (P(Z))
- 基于平扫的LR模型和基于全序列的RF模型的AUC值存在差异

表 各模型及专家在外部独立测试集的AUC值Delong检验结果 (P (Z))

	基于全序列的RF模型	基于全序列的LR模型	基于平扫的RF模型	基于平扫的LR模型	专家A	专家B
基于全序列的RF模型	—	0.835 (0.208)	0.935 (0.081)	0.012 (0.524)	0.523 (0.639)	0.253 (1.144)
基于全序列的LR模型	0.835 (0.208)	—	0.836 (0.208)	0.067 (1.831)	0.541 (0.611)	0.393 (0.855)
基于平扫的RF模型	0.935 (0.081)	0.836 (0.208)	—	0.104 (1.627)	0.593 (0.534)	0.392 (0.856)
基于平扫的LR模型	0.012 (0.524)	0.067 (1.831)	0.104 (1.627)	—	0.093 (1.682)	0.310 (1.016)
专家A	0.523 (0.639)	0.541 (0.611)	0.593 (0.534)	0.093 (1.682)	—	0.669 (0.427)
专家B	0.253 (1.144)	0.393 (0.855)	0.392 (0.856)	0.310 (1.016)	0.669 (0.427)	—

5. 研究结论

基于多参数磁共振成像的机器学习模型表现

接近高年资医生

有助于肾脏实性小肿瘤的定性诊断

