

CT影像组学鉴别脾脏高代谢与正常代谢的价值

荣昱辰¹, 黄 山², 武江芬³, 赵 红^{1*}

¹安徽医科大学第二附属医院放射科, 安徽 合肥

²安徽医科大学第二附属医院核医学科, 安徽 合肥

³推想医疗科技股份有限公司, 北京

收稿日期: 2023年11月27日; 录用日期: 2023年12月21日; 发布日期: 2023年12月27日

摘 要

目的: 评估CT影像组学对于鉴别脾脏高代谢与正常代谢的价值。方法: 回顾性分析正常代谢组274例及高代谢组93例的CT平扫资料, 按8:2比例将所有样本分为训练集和测试集, 提取脾脏纹理特征共1247个, 对其进行特征选择, 获得两组之间存在显著差异的特征参数共336, 采用5倍交叉验证, 并以逻辑回归(LR)、随机梯度下降(SGD)及线性判别分析(LDA) 3种机器学习方法建立预测模型, 通过训练集及测试集数据进行评估及验证, 绘制相应ROC曲线, 使用曲线下面积(AUC)评估模型的鉴别诊断效能, 以决策曲线分析预测模型的临床应用价值。结果: 分别建立预测模型, 分别获得3种机器学习方法的ROC曲线及AUC值, 其中以SGD方法建立的预测模型在测试集中的AUC为0.8287, 优于LR、LDA模型。决策分析曲线显示SGD模型较优于LR、LDA模型。结论: 基于CT影像组学可以有效鉴别脾脏高代谢与正常代谢。

关键词

脾, 体层摄影术, 正电子发射计算机断层显像, 影像组学

The Value of CT Radiomics Models in Differentiating Hypermetabolism and Normal Metabolism of Spleen

Yuchen Rong¹, Shan Huang², Jiangfen Wu³, Hong Zhao^{1*}

¹Department of Radiology, The Second Affiliated Hospital of Anhui Medical University, Hefei Anhui

²Department of Nuclear Medicine, Second Affiliated Hospital of Anhui Medical University, Hefei Anhui

³Beijing Infervision Technology Co., Ltd., Beijing

*通讯作者。

文章引用: 荣昱辰, 黄山, 武江芬, 赵红. CT 影像组学鉴别脾脏高代谢与正常代谢的价值[J]. 临床医学进展, 2023, 13(12): 19899-19910. DOI: 10.12677/acm.2023.13122803

Abstract

Objective: To evaluate the value of CT radiomics models in differentiating hypermetabolism from normal metabolism in spleen. **Methods:** The CT data of 274 cases of normal metabolism group and 93 cases of high metabolism group were retrospectively analyzed. All the samples were divided into training set and test set according to the ratio of 8:2, the texture features of the spleen were extracted, and the feature parameters with significant differences between the two groups were obtained. The 5-fold cross-validation method was used, and the prediction model was established by Logistic Regression (LR), stochastic gradient descent (SGD) and linear discriminant analysis (LDA) machine learning methods. The training set data were evaluated, and then the test set data were validated. The corresponding ROC curve was drawn, and the area under the curve (AUC) was used to evaluate the differential diagnostic performance of the model. **Results:** A total of 1247 texture features were extracted from spleen, 336 of which were significantly different between groups. To establish a prediction model, evaluate and verify it, the ROC curves of the three machine learning methods were obtained respectively. The AUC of the prediction model established by SGD method in the training set and test set was 0.8782 and 0.8287, respectively, with the highest accuracy. **Conclusion:** CT radiomics models can effectively distinguish hypermetabolism from normal metabolism in spleen.

Keywords

Spleen, Tomography, Positron Emission Computed Tomography, Radiomics

Copyright © 2023 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

脾脏具有广泛的免疫功能及其在造血和红细胞清除中的重要作用[1]。因此脾脏会受到多种疾病的影响,其中一些局灶性病变在影像中具有特征性表现很容易进行鉴别[2] [3] [4] [5] [6]。而在部分血液系统疾病(淋巴瘤、白血病等)以及部分非血液系统病变(炎症、自身免疫性疾病等)中,脾脏受累可表现为不同程度的脾脏 PET 异常代谢,而解剖学特征存在重叠[7]。这导致这些疾病在 CT 平扫上不易鉴别,同时这些疾病多为系统性全身性疾病,病情较严重[7] [8] [9] [10],需要进一步检查并尽快制定相应治疗方案,因此早期识别出脾脏弥漫性高代谢病变对下一步检查以及治疗方案的制定有重要意义。脾脏 CT 影像组学可以挖掘出解剖学特征以外的反映病灶的特征[11],如张华妮等[12]所建立的脾脏影像纹理特征模型能较好的对肝硬化食管静脉曲张严重程度进行预测,沈宁喆等[13]联合临床检验指标及影像学特征所建立的术前识别胃癌浆膜浸润的模型能够较为准确地预测浆膜浸润风险。PET/CT 不仅能评估脾脏的解剖学特征,还能发现脾脏代谢及功能学情况。这两者间均能在观察解剖学特征的同时发掘出其中深层信息,有异曲同工之处。因此,笔者设想能否发现脾脏影像组学模型与脾脏代谢间的关系。综上所述,在本次研究中,我们研究尝试建立不同机器学习算法的脾脏 CT 影像组学模型,并对比不同模型鉴别脾脏高代谢与正常代谢的诊断效能。

2. 资料与方法

2.1. 一般资料

收集从 2020 年 1 月~2021 年 12 月该时间段内行包含脾脏的 CT 检查的患者资料,其中包括一般资料和影像资料,一般资料包括年龄、性别、病史等,影像资料为包含全部脾脏的 CT 厚层平扫图像。然后从中筛选出在一月内行 PET/CT 检查的患者,并根据 PET/CT 结果将其分组。该研究,我院医学伦理委员会已批准,并且入组对象均已签署知情同意书。

纳入标准:1) PET/CT 结果、影像及临床资料完整;2) 既往未接受过脾脏手术;排除标准:1) CT 扫描图像出现较大伪影,无法进行观察;2) 临床资料不完善;3) 脾脏出现较大病变区域或缺如。

2.2. 分组方法

根据 PET/CT 结果脾脏最大标准摄取值(SUV_{脾 max})与正常肝组织最大标准摄取值(SUV_{肝 max})的比值(SUV_{TLR}),将患者分为正常代谢组(SUV_{TLR} ≤ 1)及高代谢组(SUV_{TLR} > 1)。

$$SUV_{TLR} = \frac{SUV_{脾 max}}{SUV_{肝 max}}$$

2.3. 仪器与方法

2.3.1. CT 检查

本次扫描采用为西门子 64 层螺旋 CT 扫描仪,患者取仰卧位,双臂上举,于吸气末进行全脾脏扫描,所有 CT 扫描获得的图像以软组织窗显示。扫描条件为管电流 160~320 mA,管电压 120 kV,以 0.5 s 的机架旋转速度进行螺旋扫描,探测器线宽 192 × 0.6 mm,FOV416 mm,重建矩阵 512 × 512,图像重建采用迭代重建算法,扫描层厚 5 mm,重建层厚 1.0 mm,层间隔 1.0 mm。检查前去除患者身上所有影响成像的异物。

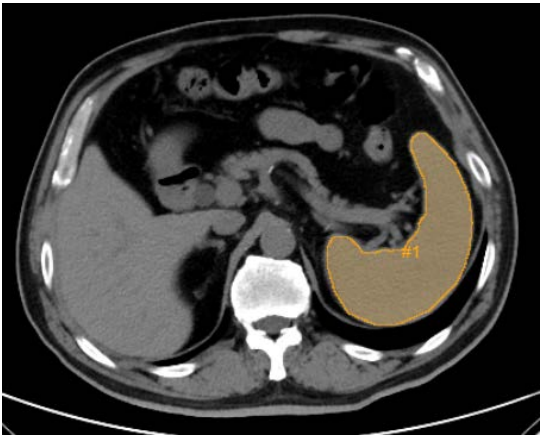
2.3.2. PET/CT 检查

进行 PET/CT 检查所使用的仪器为中国上海联影医疗科技有限公司生产的 UIH uMI510 PET/CT 扫描仪。PET/CT 显像示踪剂为由南京江原安迪科正有限公司购买的 18 氟标记的脱氧葡萄糖(18F-FDG),放化纯度为 95%~99%,该药剂符合注射放射性药物标准。患者禁食 6 小时左右,如监测末端指尖空腹血糖小于 7.0 mmol/l 时,则可以进行 18F-FDG PET/CT 检查。在整个检查过程中,患者需要保持放松状态并且平静呼吸。按照 0.10~0.12 mCi/Kg 计算注射量,然后于手臂肘静脉处注射显像示踪剂。平静休息约 50 min 并将排空尿液后,开始对患者进行全身断层显像,FOV 设为 512 mm × 512 mm,3D 模式下采集时间为 3 min/床位,全身扫描 7~8 个床位。PET 扫描前进行 CT 采集,从颅骨到大腿上部 1/3 处,扫描电压为 120 kV,上限为 180 mA,横切面厚度为 3.0 毫米。利用响应线 RAMLA 算法对 PET 图像进行重构,并利用 CT 数据进行衰减校正,分别获得冠状面、横状面及矢状面三个维度的融合图像。图像采集后,核医学医生检查图像质量,如果图像质量符合要求,患者方可离开,如不合格,则血药再次进行扫描。所有的 PET/CT 图像分析均由一名有 5 年以上工作经验的主治医师进行读片、诊断,然后由 10 年以上工作经验的高年资医师进行审核,意见出现不统一时,由第三位高年资医生介入,相互讨论后做出诊断。分别获取脾脏内病变的部位、大小、数目、SUV_{max} 及髓外病变等数据。获得 SUV_{max} 结果需要首先勾画病灶作为感兴趣区(ROI),然后通过电脑工作站自动算出 SUV_{max}。

2.4. CT 图像分割

将所以收集的脾脏轴面厚层 CT 平扫图像,从 PACS 系统中以 DICOM 格式导出,压缩文件并上传组

学分析软件，压缩文件为 ZIP 格式。在保证医师对临床信息和 PET/CT 结果未知的情况下，由一名有 10 年以上诊断经验的副主任医师独立于脾脏最大层面沿着脾脏的边缘手动勾画 ROI (图 1)，并且要尽量避免伪影及周围组织结构的影响。1 月后随机选取 30 例，该医师与另 1 名具有 5 年以上诊断经验的主治医师分别再次勾画肿瘤 ROI，提取病灶的影像组学特征。



#1 区域为于脾脏最大层面沿着脾脏的边缘手动勾画 ROI

Figure 1. Outline of ROI
图 1. 感兴趣区的勾画

2.5. 特征提取与选择

为了确保提取影像组学特征的可重复性，通过组间相关系数(ICC)检验评价 2 名医师之间及同名医师 2 次提取影像组学特征的一致性， $ICC > 0.75$ 为一致性较好。

提取所有感兴趣区的纹理特征，提取特征类型包括一阶特征、形态学特征、纹理特征，包括灰度共现矩阵(GLCM)、灰度大小区域矩阵(GLSZM)、灰度运行长度矩阵(GLRLM)、相邻灰色调差分矩阵(NGTDM)，其中形态学特征仅基于原始图像计算，一阶统计学特征和矩阵纹理特征分别在原始图像和滤波图像上计算，滤波器类型包括 Original、Wavelet、LoG、Square、Square Root、Logarithm、Exponential、Gradient；重采样大小设为 $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ ，插值方式设定为 BSpline，插值间距为 5 mm；具体特征提取的参数设置见表 1。针对 $ICC > 0.75$ 的影像组学特征以显著性检验观察影像组学特征差异，排除 $P > 0.05$ 者，保留组间差异明显者。

Table 1. The parameter settings of feature extraction
表 1. 特征提取的参数设置

重采样大小(mm ³)	插值方式	插值间距(mm)	特征类型	滤波器
1 × 1 × 1	BSpline	5	形态学特征	Original ^a
			一阶特征	Original ^a
			GLCM	Wavelet
			GLDM	LoG
			GLRLM	Square
			GLSZM	Square Root
			NGTDM	Logarithm
				Exponential
				Gradient

^aOriginal: 从原始图像上提取相应的特征。

2.6. 模型建立和评估

按 8:2 比例将所有样本分为训练集和测试集,在训练集中进行 5 倍交叉验证,分别采用逻辑回归(LR)、随机梯度下降(SGD)及线性判别分析(LDA) 3 种机器学习方法建立基于影像组学特征的预测模型,并于测试集中进行检验,绘制相应的 ROC 曲线,使用曲线下面积(AUC)评估模型的鉴别诊断效能,用 Delong 检验判断各模型间 ROC 曲线是否有统计显著性差异,并计算取最大约登指数时相应预测准确率、敏感度、特异度、阴性预测值、阳性预测值,最后以决策曲线分析(DCA)评价模型的临床应用价值。

2.7. 统计

本次实验中,使用 SPSS 25.0 统计学软件进行相关统计学计算,用 n 表示计数资料,用均数 $\pm$ 标准差($\bar{x} \pm s$)表示计量资料。组学分析特征筛选采用显著性检验,以决策曲线分析(DCA)评价模型的临床应用价值。如果 $P < 0.05$ 则认为两组间差异具有统计学意义。

3. 结果

3.1. 临床资料

本次实验共纳入研究对象 367 名,其中正常代谢组 274 例及高代谢组 93 例。正常代谢组正常代谢组平均年龄为 60.9 ± 14.2 ;高代谢组平均年龄为 60.0 ± 16.0 。正常代谢组女性 138 例,男性 136 例;高代谢组女性 48 例,男性 45 例。正常代谢组与高代谢组两组间患者的年龄、性别无统计学差异($P > 0.05$),见表 2。临床诊断方面,正常代谢组中非血液系统恶性肿瘤患者 142 例,占正常代谢组 51.82%;淋巴瘤患者 79 例,占正常代谢组 28.83%;自身免疫性疾病患者 26 例,占正常代谢组 9.49%;多发性骨髓瘤患者 18 例,占正常代谢组 6.57%;发热待查患者 9 例,占正常代谢组 3.28%。高代谢组中非血液系统恶性肿瘤患者 19 例,占高代谢组 20.43%;淋巴瘤患者 29 例,占高代谢组 31.18%;自身免疫性疾病患者 15 例,占高代谢组 16.13%;多发性骨髓瘤患者 5 例,占高代谢组 5.38%;发热待查患者 25 例,占高代谢组 26.89%。

Table 2. General data of two groups of patients (n, $\bar{x} \pm s$)
表 2. 两组患者一般资料(n, $\bar{x} \pm s$)

一般资料		代谢水平		χ^2/t	P
		正常代谢组(n = 274)	高代谢组(n = 93)		
性别	女	138	48	0.043	0.835
	男	136	45		
年龄		60.9 ± 14.2	60.0 ± 16.0	0.547	0.215

3.2. 影像组学特征筛选

对所有感兴趣区进行影像组学特征提取,共提取了 1247 个影像组学特征,并以显著性检验进行筛选,最终筛选出 336 个影像组学特征,包括一阶统计特征、形状特征、纹理特征和高阶统计特征共 7 类:(1) 一阶特征(first order) n = 61;(2) 形状特征(shape) n = 14;(3)灰度共现矩阵(GLCM) n = 23;(4)灰度依赖矩阵(GLDM) n = 64;(5) 灰度运行长度矩阵(GLRLM) n = 78;(6) 灰度大小区域矩阵(GLSZM) n = 72;(7) 领域灰度差矩阵(NGTDM) n = 24。见图 2。

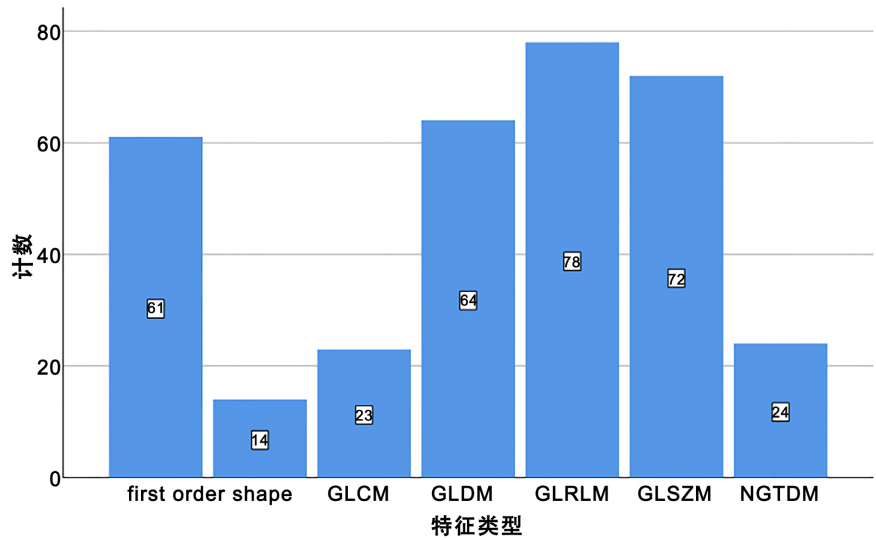


Figure 2. The number of various radiomics features after screening
图 2. 筛选后各类影像组学特征的数量

3.3. 模型诊断性能评估

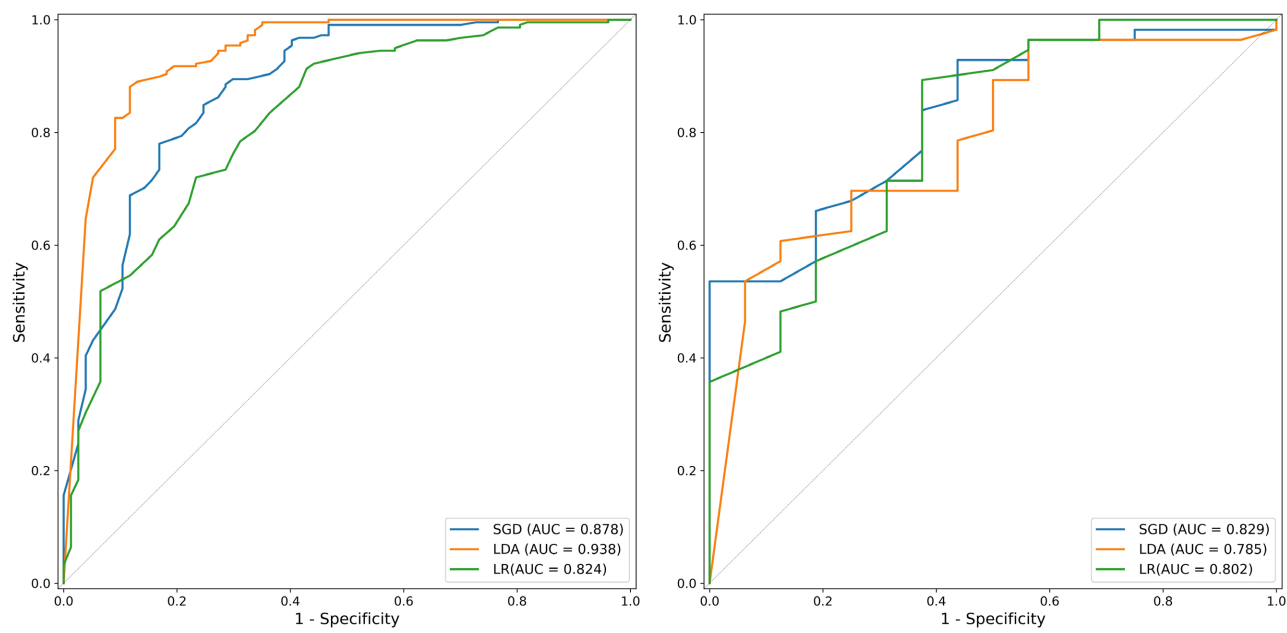
采用 5 倍交叉验证以 LR、SGD 及 LDA 3 种机器学习方法建立预测模型，并于测试集中进行检验，分别获得 3 种机器学习方法的 ROC 曲线，在训练集中 ROC 曲线下面积(AUC)分别为 0.8241 [95% CI (0.7649, 0.8723)]、0.8782 [95% CI (0.8276, 0.9230)]、0.9376 [95% CI (0.8978, 0.9666)]，在测试集中分别为 0.8019 [95% CI (0.6745, 0.9189)]、0.8287 [95% CI (0.7197, 0.9239)]、0.7846 [95% CI (0.6540, 0.8986)]；其中，测试集中 SGD 模型的 AUC 优于 LR 模型及 LDA 模型。三种预测模型的 AUC 值结果见表 3；三种预测模型的 ROC 曲线结果见图 3。在测试集用 Delong 检验判断各模型间 ROC 曲线是否有统计显著性差异，结果示各模型之间无统计学差异($P > 0.05$)；见表 4。

Table 3. The AUC of the three prediction models in the training set and the test set
表 3. 三种预测模型在训练集与测试集的 AUC 值结果对照

组别		AUC	95% CI
训练集	LR	0.8241	0.7649~0.8723
	SGD	0.8782	0.8276~0.9230
	LDA	0.9376	0.8978~0.9666
测试集	LR	0.8019	0.6745~0.9189
	SGD	0.8287	0.7197~0.9239
	LDA	0.7846	0.6540~0.8986

Table 4. The Delong test of the ROC curves among the three models in the test set
表 4. 测试集中三个模型间 ROC 曲线 Delong test 结果

模型	模型	P 值
SGD	LDA	0.10
SGD	LR	0.36
LDA	LR	0.67



训练集 AUC: LR = 0.8241, SGD = 0.8782, LDA = 0.9376; 测试集 AUC: LR = 0.8019, SGD = 0.8287, LDA = 0.7846。

Figure 3. ROC for the training and test sets of the three prediction model
图 3. 三种预测模型训练集及测试集的 ROC 曲线

取最大约登指数为最佳临界值, LR 模型训练集及测试集的诊断准确率、敏感度、特异度、阴性预测值及阳性预测值分别为 0.73 [95% CI (0.68, 0.78)]、0.72、0.77、0.90、0.49 和 0.83 [95% CI (0.74, 0.92)]、0.89、0.63、0.89、0.63; SGD 模型训练集及测试集的诊断准确率、敏感度、特异度、阴性预测值及阳性预测值分别为 0.79 [95% CI (0.75, 0.84)]、0.78、0.83、0.93、0.57)和 0.64 [95% CI (0.53, 0.75)]、0.54、1.00、1.00、0.38; LDA 模型训练集及测试集的诊断准确率、敏感度、特异度、阴性预测值及阳性预测值分别为 0.88 [95% CI (0.84, 0.92)]、0.88、0.88、0.88、0.88)和 0.67 [95% CI (0.56, 0.76)]、0.88、0.61、0.94、0.39; 见表 5。其中测试集中 LR 模型准确率优于 SGD 模型及 LDA 模型。三个预测模型的预测值和真实值的混淆矩阵, 见图 4~6。决策分析曲线表明三个预测模型均有良好的临床使用性, 并 SGD 模型较优于 LR、LDA 模型, 见图 7。

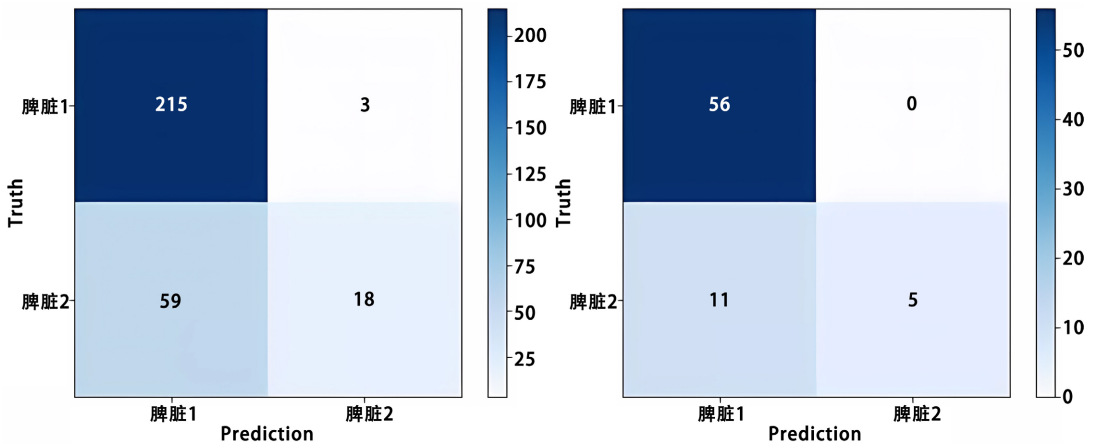


Figure 4. Confusion matrix of the training set and test sets of the LR model
图 4. LR 模型训练集及测试集的混淆矩阵

Table 5. Clinical efficacy of the three prediction models
表 5. 三个预测模型临床效能

组别		准确率	敏感度	特异度	阳性预测性	阴性预测值
训练集	LR	0.73 [95% CI (0.68, 0.78)]	0.72	0.77	0.90	0.49
	SGD	0.79 [95% CI (0.75, 0.84)]	0.78	0.83	0.93	0.57
	LDA	0.88 [95% CI (0.84, 0.92)]	0.88	0.88	0.88	0.88
测试集	LR	0.83 [95% CI (0.74, 0.92)]	0.89	0.63	0.89	0.63
	SGD	0.64 [95% CI (0.53, 0.75)]	1.00	0.54	1.00	0.38
	LDA	0.67 [95% CI (0.56, 0.76)]	0.88	0.61	0.94	0.39

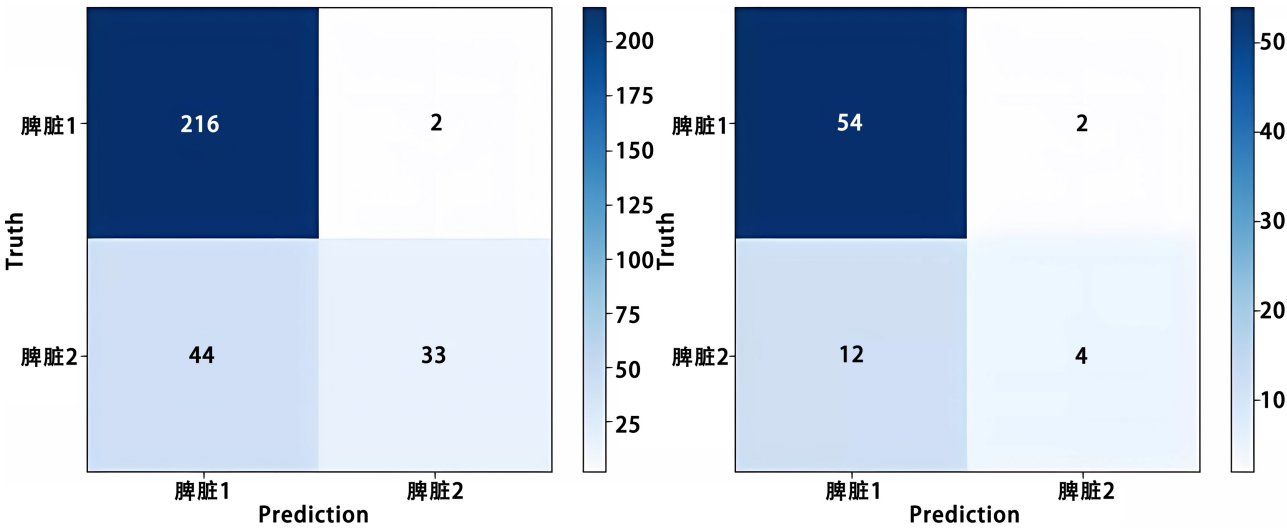


Figure 5. Confusion matrix of the training set and test sets of SGD models
图 5. SGD 模型训练集及测试集的混淆矩阵

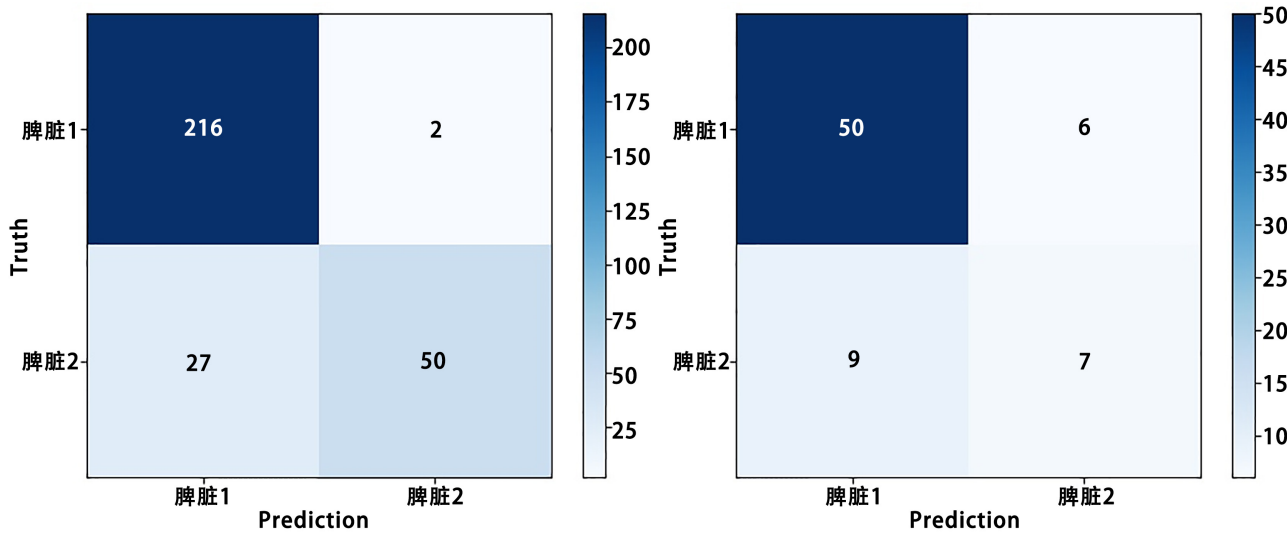
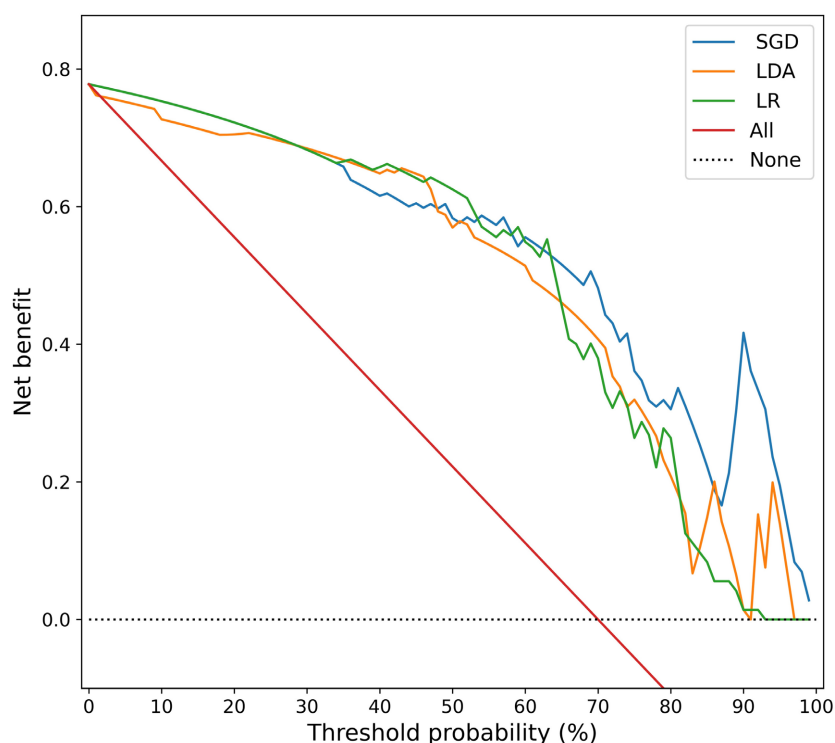


Figure 6. Confusion matrix of the training set and test sets of the LDA model
图 6. LDA 模型训练集及测试集的混淆矩阵



LR、SGD、LDA 三个预测模型均有良好的临床使用性，SGD 模型较优于 LR、LDA 模型。

Figure 7. The DCA curves of the three prediction models

图 7. 三种预测模型的 DCA 曲线

4. 讨论

本次研究中患者年龄普遍偏大(正常代谢组 60.9 ± 14.2 、高代谢组 60.0 ± 16.0)，部分伴有基础性疾病，例如糖尿病、高血压、心脏疾病等，同时部分恶性肿瘤患者有手术史。同时，本次研究高代谢组中，非血液系统性恶性肿瘤患者 19 例，占高代谢组 20.43%；淋巴瘤患者 29 例，占高代谢组 31.18%；自身免疫性疾病患者 15 例，占高代谢组 16.13%；多发性骨髓瘤患者 5 例，占高代谢组 5.38%；发热待查患者 25 例，占高代谢组 26.89%。脾脏是一个多功能器官，大部分血液系统疾病以及非血液系统病变可累及脾脏，包括淋巴瘤、白血病、炎症、自身免疫性疾病等。非血液系统性恶性肿瘤患者脾脏高代谢，部分考虑病灶转移至脾脏导致代谢异常，同时部分应考虑为恶性肿瘤治疗后导致代谢增高的假阳性。针对后者治疗完成后的 PET 应在化疗或化学免疫治疗后至少 3 周进行，最好在 6 至 8 周进行，在放疗或放化疗后 8 至 12 周进行[14]。而本研究可通过 CT 影像组学较早发现脾脏异常代谢，并对后续鉴别异常代谢是否为假阳性具有重要意义。淋巴瘤是常见的血液系统恶性肿瘤性疾病，其脾脏浸润可高达 40%，淋巴瘤的脾脏受累可以是原发性的或继发性的，后者更为常见。原发性脾淋巴瘤定义为脾淋巴瘤受累，伴或不伴脾门淋巴结肿大。继发性脾淋巴瘤定义为脾门和淋巴结以外的淋巴瘤受累[15] [16]。原发性脾淋巴瘤很少见，最常见的是弥漫性大 B 细胞淋巴瘤。影像学检查显示，原发性脾淋巴瘤常表现为单个大肿块或多个局灶性脾肿块。脾肿大伴邻近器官侵犯也有报道[17] [18]。在临床工作中，判断有无脾脏浸润对造血系统恶性肿瘤性病变非常重要，其对肿瘤分期、治疗方案及预后评估均有影响，如在李旭等[19]的研究中发现脾脏受累者复发/进展率更高。多发性骨髓瘤虽然是一种罕见疾病，但却是第二常见的血液系统恶性肿瘤，仅次于非霍奇金淋巴瘤[20]。研究表明近年来观察到多发性骨髓瘤髓外浸润的发病率较高，侵袭性较强，预后

差。尸检时, 大约三分之二的患者骨骼系统外存在浆细胞浸润, 而其中脾脏受累约占 30.8% [21]。因此运用 CT 影像组学尽早发现脾脏代谢增高对淋巴瘤、多发性骨髓瘤患者下一步检查以及后续治疗方案的制定十分重要。自身免疫性疾病除脾脏异常代谢外, 每种疾病都有相应的临床症状、影像特征以及实验室检查。因此, CT 影像组学发现脾脏异常代谢后并结合相应临床特点可提前对病情进行分析。发热待查患者往往以发热为表现, 常规影像学及实验室检查难以确定病因。不明原因的发热病因主要分为以下几种: 1) 感染性疾病、2) 恶性肿瘤性疾病、3) 非感染性炎性疾病(HIV 感染除外)、4) 其他原因(包括药物热、溶血、肉芽肿性疾病、栓属性静脉炎等) [22]。脾脏 CT 影像组学可对脾脏代谢异常进行早期筛查, 并帮助临床制定检查方案、尽早查明病因。

在本研究中共提取并筛选出: (1) 一阶统计特征 61 个; (2) 形状特征 14 个; (3) 灰度共现矩阵特征 23 个; (4) 灰度依赖矩阵特征 64 个; (5) 灰度运行长度矩阵特征 78 个; (6) 灰度大小区域矩阵 72 个; (7) 领域灰度差矩阵 24 个。在影像组学领域中使用的特征有形状特征、一阶统计特征和纹理特征。形状特征描述感兴趣区三维大小和形状的特征, 一阶统计特征表示肿瘤或病灶内灰度级方面的总体异质性, 纹理特征表示肿瘤内灰度的局部空间不均匀性。GLSZM 特征用于表征细胞核的不均匀性。本次研究提取的特征中一阶统计特征及二阶灰度特征所占比例较高分别为(61/336)和(261/336)。其中一阶统计特征表示病灶内灰度级方面的总体异质性, 二阶灰度特征表示病灶内灰度的局部空间不均匀性。这可能与脾脏的糖代谢增高, 导致感兴趣区即脾脏内异质性增高有关。

以 LR、SGD 及 LDA 3 种机器学习方法建立的预测模型在测试集中 AUC 值分别为 0.80 [95% CI (0.67, 0.92)]、0.83 [95% CI (0.72, 0.92)]、0.78 [95% CI (0.65, 0.90)]; 其中, SGD 模型的 AUC 优于 LR 模型及 LDA 模型。取最大约登指数为最佳临界值, 3 种机器学习方法建立的预测模型在测试集中准确率分别为 0.83 [95% CI (0.74, 0.92)]、0.64 [95% CI (0.53, 0.75)]、0.67 [95% CI (0.56, 0.76)], 其中 LR 模型准确率较 SGD 模型及 LAD 模型表现优秀。我们可以看到, 以 AUC 值评估预测模型效能时 SGD 模型最佳, 而取最大约登指数下的准确率评价预测模型效能时 LR 模型准确率最高。针对以上 AUC 与准确率矛盾的问题, 已有学者通过模拟研究证实矛盾的存在[23], 其中 AUC 与准确率结论的一致比例较低可达 93.5%。笔者指出 AUC 的隐含目标是处理样本分布偏斜的情况, 并且不想过度拟合单个类。针对该问题笔者认为, AUC 是比基于判别性和一致性的正式定义的准确度更好的度量, 并建议在评估和比较预测模型时, 使用 AUC 作为“单一数字”度量, 以提高准确性。在测试集中各模型间 ROC 曲线 Delong 检验判断是否有统计显著性差异时, 结果显示各模型之间 P 值均大于 0.05。虽然各模型间无统计学差异, 但 SGD 模型 AUC 值较 LR、LDA 模型高。

三个预测模型的 DCA 曲线表明三者均有良好的临床使用性, 并且在阈值在 0.5~1 之间, SGD 模型 DCA 曲线优于 LR、LDA 模型。因此综上, 本次研究中认为诊断效能最佳的为 SGD 模型, 测试集 AUC 值为 0.83, 有较好的诊断效能, 并且有良好的临床使用性。这标志着该模型对早期识别出脾脏弥漫性高代谢病变, 并对下一步检查以及治疗方案的制定有重要作用。

在本实验中存在一些不足之处, 本实验中仅在脾脏最大层面手动勾画 ROI, 这可能导致部分脾脏组学特征丢失, 同时手动勾画工作量较大, 在将来的研究中可以考虑逐层勾画脾脏, 来降低组学特征丢失的可能, 而针对手动勾画工作量较大的问题, 将来可以采用半自动勾画方法, 即人工进行管理、计算机辅助勾画。同时, 本研究仅对脾脏 CT 平扫影像进行了研究, 将来可以采取 CT 平扫和增强共同在影像组学中进行分析, 以期获得诊断效能更高的模型。

目前, 针对脾脏的 CT 影像组学研究仍在起步阶段, 而本研究通过脾脏 CT 影像组学鉴别脾脏高代谢与正常代谢的研究国内外还未见报道, 未来可采用多中心研究进一步扩大样本, 并且增加其他机器学习算法或采用深度学习等方法, 以找出最佳模型来鉴别脾脏高代谢与正常代谢。

总之, 本次研究表明, 基于脾脏 CT 影像组学所建立的模型用于鉴别脾脏高代谢与正常代谢具有较好的诊断性能, 并且 SGD 模型优于 LR 模型及 LAD 模型。这对早期识别出脾脏弥漫性高代谢病变, 并对下一步检查以及治疗方案的制定有重要意义。

参考文献

- [1] Lewis, S.M., Williams, A. and Eisenbarth, S.C. (2019) Structure and Function of the Immune System in the Spleen. *Science Immunology*, **4**, eaau6085. <https://doi.org/10.1126/sciimmunol.aau6085>
- [2] Shabtaie, S.A., Hogan, A.R. and Slidell, M.B. (2016) Splenic Cysts. *Pediatric Annals*, **45**, e251-256. <https://doi.org/10.3928/00904481-20160523-01>
- [3] Ozkale Yavuz, O., Ozcan, H.N., Oguz, B., et al. (2021) Imaging Findings of Benign and Malignant Pediatric Splenic Lesions. *Abdominal Radiology (New York)*, **46**, 3245-3252. <https://doi.org/10.1007/s00261-021-03004-z>
- [4] 程晓悦, 贺文. 血管源性急腹症 CT 诊断及应用[J]. CT 理论与应用研究, 2018, 27(5): 643-651.
- [5] Damouny, M., Mansour, S. and Khuri, S. (2022) Primary Angiosarcoma of the Spleen: An Aggressive Neoplasm. *World Journal of Oncology*, **13**, 337-342. <https://doi.org/10.14740/wjon1542>
- [6] Malgras, B., Najah, H., Dohan, A., Barat, M. and Soyer, P. (2022) Diagnosis and Treatment of Focal Splenic Lesions. *Journal of Visceral Surgery*, **159**, 121-135. <https://doi.org/10.1016/j.jvisc Surg.2021.11.010>
- [7] Saboo, S.S., Krajewski, K.M., O'Regan, K.N., et al. (2012) Spleen in Haematological Malignancies: Spectrum of Imaging Findings. *The British Journal of Radiology*, **85**, 81-92. <https://doi.org/10.1259/bjr/31542964>
- [8] Fianchi, L., Quattrone, M., Criscuolo, M., et al. (2021) Extramedullary Involvement in Acute Myeloid Leukemia. A Single Center Ten Years' Experience. *Mediterranean Journal of Hematology and Infectious Diseases*, **13**, e2021030. <https://doi.org/10.4084/MJHID.2021.030>
- [9] Stölzel, F., Lüer, T., Löck, S., et al. (2020) The Prevalence of Extramedullary Acute Myeloid Leukemia Detected by (18)FDG-PET/CT: Final Results from the Prospective PETAML Trial. *Haematologica*, **105**, 1552-1558. <https://doi.org/10.3324/haematol.2019.223032>
- [10] Liu, Y. (2009) Clinical Significance of Diffusely Increased Splenic Uptake on FDG-PET. *Nuclear Medicine Communications*, **30**, 763-769. <https://doi.org/10.1097/MNM.0b013e328323fa254>
- [11] Mayerhoefer, M.E., Materka, A., Langs, G., et al. (2020) Introduction to Radiomics. *Journal of Nuclear Medicine*, **61**, 488-495. <https://doi.org/10.2967/jnumed.118.222893>
- [12] 张华妮, 李嘉颖, 周敏辉, 陈海波, 孙志超. 脾脏影像纹理特征预测肝硬化患者食管静脉曲张严重程度的研究[J]. 浙江医学, 2023, 45(5): 470-474+562.
- [13] 沈宁喆, 郑经纬, 潘步建, 章巍腾, 陈孝冬. 基于术前脾脏 CT 影像组学列线图预测胃癌的浆膜浸润[J]. 温州医科大学学报, 2023, 53(1): 15-21+28.
- [14] Barrington, S.F., Mikhael, N.G., Kostakoglu, L., et al. (2014) Role of Imaging in the Staging and Response Assessment of Lymphoma: Consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *Journal of Clinical Oncology*, **32**, 3048-3058. <https://doi.org/10.1200/JCO.2013.53.5229>
- [15] Rabushka, L.S., Kawashima, A. and Fishman, E.K. (1994) Imaging of the Spleen: CT with Supplemental MR Examination. *Radiographics*, **14**, 307-332. <https://doi.org/10.1148/radiographics.14.2.8190956>
- [16] Leite, N.P., Kased, N., Hanna, R.F., et al. (2007) Cross-Sectional Imaging of Extranodal Involvement in Abdominopelvic Lymphoproliferative Malignancies. *Radiographics*, **27**, 1613-1634. <https://doi.org/10.1148/rg.276065170>
- [17] Bhatia, K., Sahdev, A. and Reznick, R.H. (2007) Lymphoma of the Spleen. *Seminars in Ultrasound, CT, and MR*, **28**, 12-20. <https://doi.org/10.1053/j.sult.2006.10.010>
- [18] Dachman, A.H., Buck, J.L., Krishnan, J., Aguilera, N.S. and Buetow, P.C. (1998) Primary Non-Hodgkin's Splenic Lymphoma. *Clinical Radiology*, **53**, 137-142. [https://doi.org/10.1016/S0009-9260\(98\)80061-5](https://doi.org/10.1016/S0009-9260(98)80061-5)
- [19] 李旭, 姚稚明. (18)F-FDG PET/CT 探查初诊弥漫大 B 细胞淋巴瘤脾脏受累的预后价值[J]. 医学研究杂志, 2020, 49(6): 119-124.
- [20] Firth, J. (2019) Haematology: Multiple Myeloma. *Clinical Medicine (London, England)*, **19**, 58-60. <https://doi.org/10.7861/clinmedicine.19-1-58>
- [21] Oshima, K., Kanda, Y., Nannya, Y., et al. (2001) Clinical and Pathologic Findings in 52 Consecutively Autopsied Cases with Multiple Myeloma. *American Journal of Hematology*, **67**, 1-5. <https://doi.org/10.1002/ajh.1067>

- [22] Casali, M., Lauri, C., Altini, C., *et al.* (2021) State of the Art of (18)F-FDG PET/CT Application in Inflammation and Infection: A Guide for Image Acquisition and Interpretation. *Clinical and Translational Imaging*, **9**, 299-339. <https://doi.org/10.1007/s40336-021-00445-w>
- [23] Ling, C.X., Jin, H. and Zhang, H. (2003) AUC: A Better Measure than Accuracy in Comparing Learning Algorithms. In: Xiang, Y. and Chaib-draa, B., Eds., *Advances in Artificial Intelligence. Canadian AI 2003. Lecture Notes in Computer Science*, Vol. 2671, Springer, Berlin, 329-341. https://doi.org/10.1007/3-540-44886-1_25