

论 著

机器学习算法和能谱CT预测体内一水草酸钙结石*

黄丽丰¹ 刘宣含¹ 刘满荣²
丁 可^{1,*}

1.南宁市第二人民医院放射科

2.南宁市第二人民医院超声科

(广西 南宁 530000)

【摘要】目的 本研究的目的是基于能谱CT多定量参数, 并利用机器学习算法构建最佳预测模型, 用于分析体内一水草酸钙结石。**方法** 回顾性搜集123例尿路结石患者, 将数据按7: 3比例随机分为训练集(n=86)和测试集(n=37)。使用六种机器学习算法包括逻辑回归(logistic)、极端梯度增强(XGBoost)、随机森林(RandomForest)、高斯朴素贝叶斯(GNB)、支持向量机(SVM)、K近邻(KNN)构建体内一水草酸钙结石的预测模型, 用曲线下面积(AUC)值、校准曲线、DCA曲线、准确率-召回率曲线(PR曲线)、平均精确度(AP)进行综合分析并选出最佳预测模型。模型的可解释性采用SHapley Additive exPlanations(SHAP)算法。**结果** 最终发现3个独立预测影响因子, 分别是年龄、性别、有效原子序数(Zeff), 并用其建立的逻辑回归分类模型在5折交叉验证中的训练集中显示最佳识别能力。**结论** 基于机器学习算法的逻辑回归分类模型可以有效预测体内一水草酸钙结石。

【关键词】 人工智能; 机器学习;
能谱CT成像; 一水草酸钙结石

【中图分类号】R445.4

【文献标识码】A

【基金项目】广西卫生健康委员会科研课题
(Z-A20221157);广西住院医师规范化培训重点专业
基地项目;

全国职工创新补助资金项目

DOI:10.3969/j.issn.1672-5131.2025.03.051

Machine Learning Algorithm and Energy Spectrum CT to Predict Calcium Oxalate Monohydrate Stones in Vivo*

HUANG Li-feng¹, LIU Xuan-han¹, LIU Man-rong², DING Ke^{1,*}.1.Department of Radiology, The Second Nanning People's Hospital, Nanning 530000, the
Guangxi Zhuang Autonomous Region, China2.Department of Ultrasound, The Second Nanning People's Hospital, Nanning 530000, the
Guangxi Zhuang Autonomous Region, China

ABSTRACT

Objective The aim of this study was to construct an optimal prediction model based on energy-spectrum CT multi-quantitative parameters and using machine learning algorithms for analyzing calcium oxalate monohydrate stones in vivo. **Methods** 123 patients with urinary tract stones were retrospectively collected, and the data were randomly divided into a training set (n=86) and a test set (n=37) in a ratio of 7:3. Six machine learning algorithms including logistic, Extreme Gradient Boost (XGBoost), RandomForest, Gaussian Plain Bayes (GNB), Support Vector Machines (SVM), and K Nearest Neighbors (KNN) were used to construct predictive models for in vivo Calcium Oxalate Monohydrate stones using Area Under the Curve (AUC) values, calibration curves, DCA curves, precision-recall curves (PR curves), and average precision (AP) for comprehensive analysis and selection of the best prediction model. Model interpretability was performed using the SHapley Additive exPlanations (SHAP) algorithm. **Results** Three independent predictive influences, namely age, gender, and effective atomic number (Zeff), were finally found and the logistic regression classification model built with them showed the best recognition ability in the training set in 5-fold cross-validation. **Conclusion** A logistic regression classification model based on machine learning algorithms can effectively predict calcium oxalate monohydrate stones in vivo.

Keywords: Artificial Intelligence; Machine Learning; Energy Spectrum CT Imaging; Calcium Oxalate Monohydrate Stones

泌尿系结石是泌尿外科的常见病, 术后复发率极高, 给患者造成一定的困扰^[1]。泌尿系结石的形成与年龄、性别、饮食习惯及地域分布等多种因素有关^[2-3]。泌尿系结石治疗效果与多种因素有关^[4]。体外冲击波碎石术(ESWL)具有损伤小、费用便宜等优点, 大部分泌尿系结石患者首选ESWL治疗^[5]。草酸钙结石最常见^[6], 其中二水草酸钙、磷酸钙、磷酸镁铵结石对ESWL治疗效果较好, 但一水草酸钙结石效果较差, 因此在治疗前加以鉴别显得尤为重要^[7-8]。随着计算机科学技术的发展, 人工智能越来越多的应用于医疗系统中, 可以辅助诊断、疾病转归预测、术后并发症评估等, 实现精准医疗^[9]。其中, 机器学习是人工智能的一个分支, 可识别复杂数据, 预测或分类潜在数据, 分析高级探索性数据^[10]。本研究的目的是基于能谱CT多定量参数, 并运用机器学习算法有效识别出体内一水草酸钙结石, 达到避免盲目碎石的目的。

1 资料与方法

1.1 研究对象 回顾性分析广西医科大学第三附属医院2020年12月至2021年12月经临床诊断为尿路结石患者共123例。

纳入标准: 结石术前均行能谱CT平扫、术后行红外光谱分析。排除标准: 临床资料失访者; 患者具有严重的心肝肾疾病等。

1.2 临床数据收集 收集泌尿系结石术后红外光谱分析结果, 单纯性结石共6种包括一水草酸钙(COM)、二水草酸钙(COD)、碳酸磷灰石(DAH)、六水磷酸镁铵、无水尿酸(UA)、胱氨酸结石, 混合性结石共11种包括COM+COD、COM+DAH、COM+UA、COD+DAH、COD+尿酸铵、COD+二水磷酸氢钙、COM+COD+DAH、COM+COD+UA、DAH+六水磷酸镁铵、DAH+六水磷酸镁铵+羟基磷酸钙、COM+DAH+羟基磷灰石。按结石成分 $\geq 70\%$ 定义该结石类型, 把上述结石分为COM结石组(70例)和非COM结石组(53例)。

1.3 扫描技术及能谱CT多定量参数的收集 采用我院GE能谱CT设备(Discovery CT750HD), 以能谱模式(GSI)进行扫描, 扫描参数为: 80/140kV快速切换的管电压(0.5ms), 自动管电流, 转速0.6s/r, FOV 50cm, 螺距0.984: 1, 探测器准直 $64 \times 0.625\text{mm}$, 扫描层厚和间隔5mm, 重建层厚1.25mm。利用ADW4.7后处理工作站中的GSI Viewer软件进行能谱分析。能谱曲线斜率=(HU40KeV-HU100KeV)/60; 测量结石ROI的有效原子序数; 选取基物质对(钙基图CD-水基图WD)并进行测量, 钙水比值(CWR)=CD/WD。

1.4 模型开发、验证和测试 将所有患者及其特定特征按7: 3比例被随机分为训练队列和测试队列。为了更好的泛化机器学习模型, 将训练队列按9: 1的比例随机分为训练集和验证集, 并使用5折交叉验证来评估训练队列中模型的准确性。

【第一作者】黄丽丰, 女, 住院医师, 主要研究方向: 腹部影像诊断。E-mail: 2535430554@qq.com

【通讯作者】丁 可, 男, 主任医师, 主要研究方向: 腹部影像诊断。E-mail: 272480365@qq.com

1.5 统计分析 采用用R版本3.6.3和python版本3.7对所得数据进行分析。计数资料以构成比(%)表示，计量资料服从正态分布以“均数±标准差”表示，不服从正态分布以中位数(P₂₅，P₇₅)表示。将所有的临床特征和能谱CT影像特征进行Lasso回归分析和多因素逻辑回归分析，筛选独立预测影响因子，并基于其特征建立6个模型。模型超参数的选择在训练集上使用5折交叉验证，并选出最佳模型用于数据集的分类建模任务，同时使用Python中的SHAP包来显示每个特征的重要性之间的关系。本研究中以P<0.05为差异具有统计学意义。

2 结 果

2.1 临床特征和能谱CT特征基本资料比较 见表1。

2.2 LASSO回归 采用LASSO回归分析选择特征因子，结果显示(最小均方误差的λ=0.048)32个变量减少到5个，分别是年龄+性别+结石长径+血尿酸+zeff。

2.3 多因素逻辑回归 见表2。

表2 多因素逻辑回归分析

Predictor	Estimate	SE	Z	P	Odds Ratio	Lower	Upper
(Intercept)	-14.084	3.393	-4.151	0.0	0.0	0.0	0.0
性别	1.838	0.619	2.972	0.003	6.287	1.975	22.798
年龄	0.078	0.021	3.662	0.0	1.082	1.04	1.132
结石长径	-0.434	0.247	-1.756	0.079	0.648	0.387	1.029
血尿酸	-0.005	0.002	-1.95	0.051	0.995	0.99	1.0
zeff	0.987	0.217	4.55	0.0	2.682	1.827	4.324

2.4 分类多模型的建立和评估 使用曲线下面积(AUC)值、校准曲线、DCA曲线、PR曲线和平均精度(AP)值进行综合分析，Logistic回归模型稳定性相对较好。见图1。

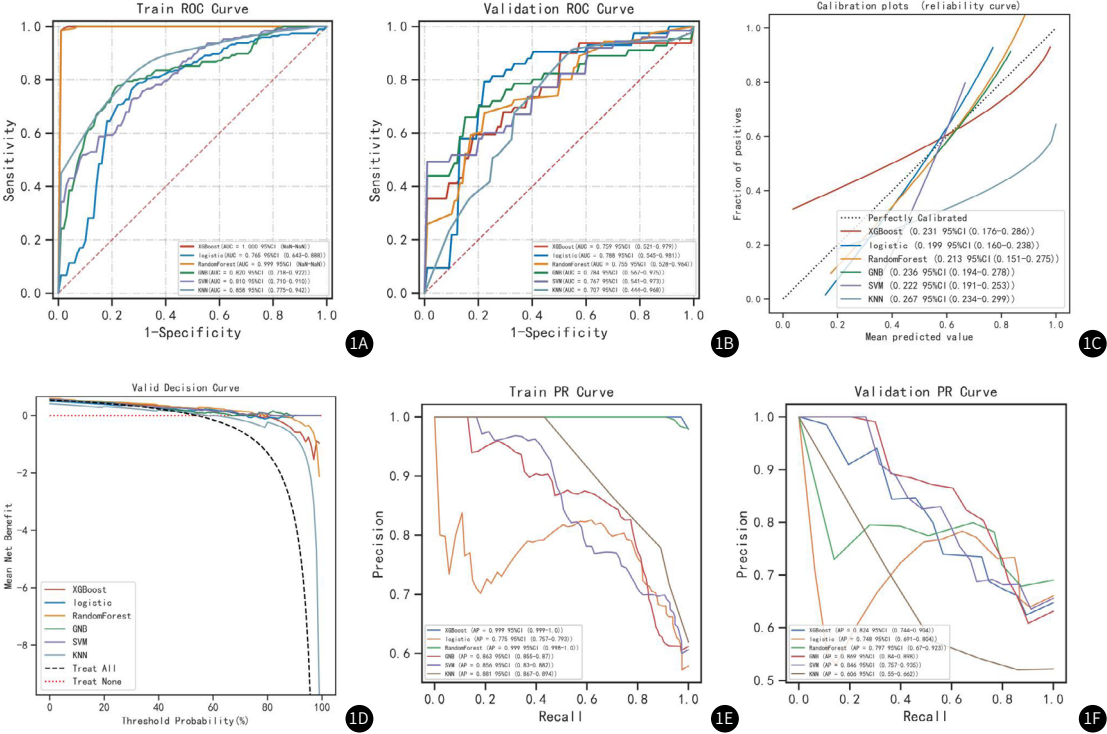


图1A-图1F ML模型综合分析(Fig.3 ML model synthesis analysis.)。图1A，训练集ROC和AUC；图1B，验证集ROC和AUC；图1C，对于验证集的校准曲线；图1D，验证集DCA曲线；图1E-1F，训练集和验证集PR曲线、AP。

2.5 最佳模型的建立和评估 结果表明，训练集的AUC值为0.771，测试集的AUC值为0.882，验证集的AUC为0.759。AUC指标下验证集未超出测试集或超出小于10%，可认为模型拟合成功。决策曲线分析显示该预测模型阈值概率约在0.0-0.95的范围内，说明具有较高的临床实用价值。综合分析，逻辑回归分类模型可以用于数据集的分类建模任务。见图2。

2.6 模型的可解释性 SHAP算法用于获得每个预测变量对逻辑回归分类模型预测结果的重要性。见图3。

3 讨 论

泌尿系结石常常以混合性结石为主，其中草酸钙结石最常见^[11]。本研究中，草酸钙成分的结石约占76%，也就意味着这个百分比的患者很难通过ESWL取得良好的治疗效果。因此，如何在治疗前鉴别出含COM结石至关重要，将能为临床首选科学的治疗方案提供理论依据。

相关学者^[12-13]指出草酸钙结石和尿酸结石在男性更常见，女性则好发碳酸磷灰石和鸟粪石。本研究中，男性总体比例为65.85%，在COM结石分组中，男性比例为68.57%，与上述研究

结果一致。武桂帅^[14]指出草酸钙结石好发平均年龄约48岁，而尿酸结石好发于年龄较大的人群，平均年龄约54岁，这可能与老年人的高血压、糖尿病等多种代谢因素相关。本研究中平均年龄是46岁，含COM结石组中平均年龄是48岁，与上述结果类似。

GE能谱CT通过单一球管的高低双能(80KV和140KV)瞬时切换来实现成像，能为临床诊断提供更多信息。每种化合物和化学元素都有自己的X线吸收衰减系数，如果两者X线吸收衰减系数相同，那么两者的有效原子序数也是相同的^[15]。甘毅^[16]通过能谱CT与术后红外光谱分析结石成分，结果提示草酸钙结石、磷酸镁铵结石、尿酸结石、胱氨酸结石的有效原子序数分别是13.49±0.78、12.55±1.07、7.09±0.41、9.76±0.54，这与纪术晶^[17]的研究一致。刘红亮^[18]研究发现一水草酸钙、二水草酸钙、磷酸钙、磷酸氢钙、尿酸、磷酸铵镁、胱氨酸平均有效原子序数是14.29、14.50、16.12、14.11、6.91、8.48、10.20。上述实验说明平均有效原子序数可以进一步鉴别体内不同成分结石。本实验中，COM结石的有效原子序数高于非COM结石，与上述结果基本一致。

近年来，人工智能越来越多应用到医学领域中，在处理医疗

表1 临床特征和能谱CT特征基本资料比较

变量	总体(n=123)		P	总体(n=123)		P
	非COM组(n=53)	COM组(n=70)		LABEL0 70% (n=86)	LABEL1 30%(n=37)	
性别			0.465			0.793
F	20(37.736)	22(31.429)		30(34.884)	12(32.432)	
M	33(62.264)	48(68.571)		56(65.116)	25(67.568)	
年龄	40.000[30.000,60.000]	48.000[42.000,55.000]	0.047	46.000[34.000,56.000]	48.000[38.000,61.000]	0.343
结石部位			>0.05			0.379
肾脏	21(39.623)	19(27.143)		28(32.558)	12(32.432)	
肾脏+膀胱	0(0.000)	4(5.714)		1(1.163)	3(8.108)	
肾脏+输尿管	8(15.094)	14(20.000)		16(18.605)	6(16.216)	
膀胱	7(13.208)	2(2.857)		7(8.140)	2(5.405)	
输尿管	17(32.075)	31(44.286)		34(39.535)	14(37.838)	
结石长径	1.300[0.700,2.200]	1.000[0.700,1.700]	0.333	1.100[0.700,1.800]	0.900[0.600,2.000]	0.619
结石短径	0.800[0.600,1.400]	0.700[0.500,1.000]	0.169	0.700[0.600,1.100]	0.600[0.400,1.100]	0.376
血Ca	2.300[2.220,2.360]	2.270[2.200,2.350]	0.283	2.290[2.200,2.360]	2.280[2.200,2.350]	0.770
血P	1.120[1.020,1.230]	1.030[0.910,1.180]	0.034	1.080[0.950,1.200]	1.060[0.960,1.210]	0.633
血Na	140.613±2.336	141.024±2.187	0.322	140.899±2.284	140.727±2.204	0.702
血K	3.920[3.720,4.100]	3.770[3.560,4.070]	0.110	3.850[3.540,4.070]	3.840[3.750,4.120]	0.152
血cl	105.351±2.980	105.331±2.754	0.970	105.260±2.801	105.524±2.965	0.641
血Mg	0.879±0.098	0.866±0.086	0.459	0.877±0.095	0.859±0.081	0.310
血白蛋白	44.100[38.900,46.200]	42.200[38.700,44.800]	0.190	42.800[39.700,45.900]	42.300[37.000,45.000]	0.448
血肌酐	78.000[63.000,104.300]	89.500[73.000,111.000]	0.127	85.000[69.100,103.000]	87.000[65.000,121.100]	0.815
尿酸	408.253±124.445	371.310±102.206	0.076	386.036±111.315	390.000±119.362	0.861
尿PH值			0.943			0.661
5.0	3(5.660)	3(4.286)		5(5.814)	1(2.703)	
6.0	20(37.736)	31(44.286)		38(44.186)	13(35.135)	
6.5	17(32.075)	19(27.143)		23(26.744)	13(35.135)	
7.0	11(20.755)	15(21.429)		18(20.930)	8(21.622)	
7.5	2(3.774)	2(2.857)		2(2.326)	2(5.405)	
尿白细胞计数	44.000[6.000,103.000]	16.000[6.000,83.000]	0.316	20.000[6.000,83.000]	41.000[6.000,124.000]	0.379
CT40	1490.930[821.270,2378.110]	2072.020[1419.150,2384.670]	0.057	2030.330[1350.690,2403.260]	1580.820[1113.360,2249.050]	0.187
CT50	1089.320[588.330,1754.580]	1517.800[1072.530,1742.110]	0.085	1492.610[996.000,1765.260]	1148.220[808.620,1666.590]	0.148
CT60	881.580[562.000,1334.260]	1159.570[809.620,1328.170]	0.180	1115.180[754.740,1334.260]	881.580[609.570,1289.610]	0.188
CT70	762.790[507.720,1080.580]	938.120[660.720,1087.250]	0.225	898.440[629.060,1093.940]	756.280[516.130,1059.620]	0.230
CT80	688.970[483.230,925.280]	805.150[573.280,945.750]	0.219	764.410[561.880,962.620]	659.660[497.920,914.220]	0.243
CT90	611.180[477.230,863.800]	725.330[548.530,847.660]	0.400	694.602±217.552	641.077±235.864	0.229
CT100	604.725±200.547	648.912±194.565	0.484	638.146±198.687	610.641±196.307	0.485
CT110	556.650[430.440,733.540]	633.780[473.370,730.310]	0.398	626.490[463.370,735.830]	556.650[449.110,713.500]	0.422
CT120	549.531±173.268	576.445±173.504	0.015	572.510±172.395	547.038±176.112	0.460
CT130	533.455±165.241	555.090±169.938	0.005	553.189±167.663	528.517±168.418	0.460
CT140	517.250[398.210,642.740]	557.760[421.520,657.570]	0.294	555.930[421.470,657.060]	500.030[391.390,634.300]	0.407
斜率	17.279[7.850,25.459]	23.198[15.679,26.397]	0.015	23.166[14.758,27.099]	17.154[10.755,25.219]	0.125
zeff	12.310[9.040,13.120]	12.810[12.350,13.260]	0.005	12.750[12.000,13.290]	12.670[11.440,12.990]	0.110
水值	1312.920[1255.220,1428.530]	1312.000[1247.170,1359.400]	0.294	1303.720[1251.730,1373.550]	1327.840[1247.170,1435.690]	0.354
钙值	227.200[50.930,339.630]	309.490[240.480,344.120]	0.007	309.490[211.670,356.770]	268.470[173.920,331.040]	0.040
钙水比值	0.192[0.040,0.257]	0.227[0.193,0.269]	0.009	0.226[0.179,0.278]	0.201[0.142,0.243]	0.025

大数据等应用场景中优势明显。机器学习作为人工智能的核心，可以快速处理小样本数据和高质量数据集，在现有的训练集数据中探索规则和变量之间的相互关系，构建的模型即可预测未来数据^[19]。侯祺^[20]发现机器学习的RF和XGBoost，相比Logistic回归模型，在预测ESWL治疗输尿管结石患者的碎石成功率方面预测性能更好。国外学者^[21]通过体外单源双能CT构建5种机器学习预测模型，结果提示机器学习可以100%鉴别直径5cm以上的尿酸和非尿酸结石，75%鉴别纯水草酸钙、胱氨酸及鸟粪石。

Große Hokamp^[22]通过双能量CT和浅层神经网络学习可以预测体外不同成分结石，总体准确度约90%。本研究表明，通过可解释的机器学习模型，可以列出重要特征排序，显示每个特征对模型输出积极或消极的影响，可以有效鉴别出体内COM结石。

大样本量的研究和高维的数据库对于构建模型是很有必要的，这也是我们接下来需要进一步研究的方向。综上所述，通过人工智能结合临床特征、能谱CT多定量参数构建预测模型，可以提高体内COM结石的判断。

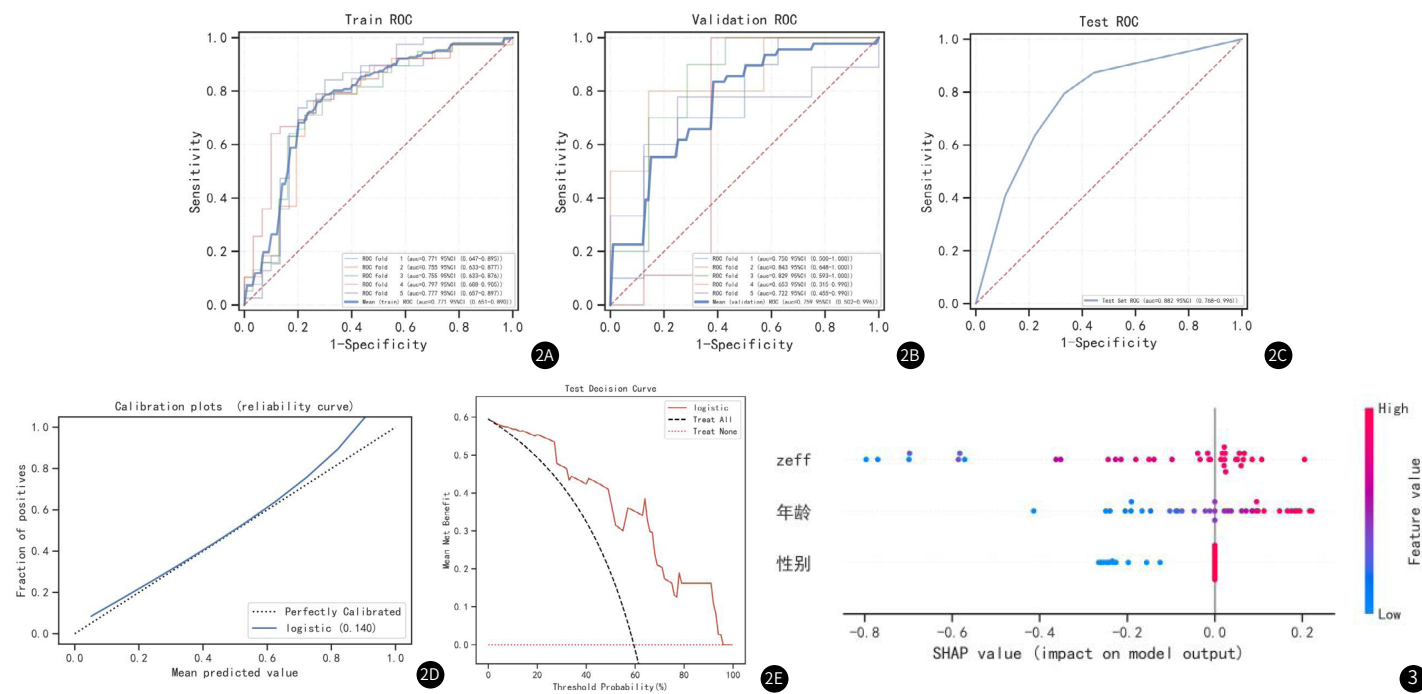


图2A-图2E 逻辑回归模型的训练、验证和测试。图2A，训练集ROC和AUC；图2B，验证集ROC和AUC；图2C，测试集ROC和AUC。30%患者的检测结果；图2D，校准曲线结果显示预测效果与实际效果较吻合；图2E，DCA曲线显示该模型预测具有较好的净利益。

图3 SHAP解释模型。SHAP中特征的属性。红点表示高特征值，蓝点表示低特征值。

参考文献

- [1] 李勤祥, 潘爱珍, 徐志峰, 等. CaSR986基因多态性与能谱CT泌尿系结石成分分析及临床特征的相关性[J]. 中国临床医学影像杂志, 2020, 31(2): 119-122.
- [2] 刘松林, 张刚, 董於坪. 山东烟台地区1146例泌尿系结石成分分析[J]. 国际泌尿系统杂志, 2022, 42(6): 1054-1057.
- [3] 刘范林, 蒋群, 杨小君, 等. 基于KUB、低剂量CT影像组学列线图预测攀花枝泌尿系统结石成分价值[J]. 中国CT和MRI杂志, 2023, 21(9): 133-136.
- [4] 崔增林, 刘高瑞, 姜言梦, 等. 输尿管结石患者应用术前全体积CT纹理联合输尿管软镜激光治疗的效果评价[J]. 中国CT和MRI杂志, 2023, 21(9): 140-142.
- [5] 王炫钦, 陈军. 上尿路结石CT值预测结石成分及对体外冲击波碎石术的指导意义[J]. 现代医学影像学, 2022, 31(4): 650-652.
- [6] 黄锦佩, 李锦棠, 陶勇, 等. 输尿管镜下钬激光碎石术(HLL)对治疗输尿管结石的结石清除率及对血清PCT、hs-CRP含量影响分析[J]. 罕少疾病杂志, 2021, 28(4): 75-76, 91.
- [7] 王典豹. 输尿管软镜手术治疗肾结石的效果分析[J]. 罕少疾病杂志, 2023, 30(6): 44-45.
- [8] 顾燕, 姜宁, 李健, 等. 体外冲击波碎石和输尿管镜钬激光碎石治疗输尿管上段结石疗效对比[J]. 实用医学杂志, 2018, 34(24): 4123-4126.
- [9] Mintz Y, Brodie R. Introduction to artificial intelligence in medicine[J]. Minim Invasive Ther Allied Technol, 2019, 28(2): 73-81.
- [10] MacEachern S J, Forkert N D. Machine learning for precision medicine[J]. Genome, 2021, 64(4): 416-425.
- [11] 覃天资, 梁洁, 吴计超, 等. 泌尿系结石患者的发病影响因素及结石成分特点[J]. 广西医学, 2023, 45(6): 656-660.
- [12] 涂熹, 庄稀亮, 黄朝友, 等. 泌尿系结石成分与性别、年龄等相关性研究[J]. 遵义医科大学学报, 2023, 46(2): 177-181.
- [13] Ye Z, Zeng G, Yang H, et al. The status and characteristics of urinary stone composition in China[J]. BJU Int, 2020, 125(6): 801-809.
- [14] 武佳帅. 代谢相关因素与泌尿系结石类型相关性研究及多因素分析[D]. 河北医科大学, 2022.
- [15] 王欣, 陈刘成, 沈龙山, 等. CT能谱成像基本原理及其临床应用进展[J]. 中国中西医结合影像学杂志, 2021, 19(2): 197-200.
- [16] 甘毅, 徐志峰, 潘爱珍, 等. 能谱CT有效原子序数对泌尿系结石成分的诊断价值[J]. 现代医学影像学, 2021, 30(11): 2070-2072.
- [17] 纪术晶, 卜令贤, 田德翠. 能谱CT对肾结石原子序数和CT值在体外震波碎石术的临床应用[J]. 医学影像学杂志, 2020, 30(5): 901-904.
- [18] 刘红亮. 能谱CT有效平均原子序数对离体尿路结石成分分析的前瞻性研究[J]. 影像研究与医学应用, 2018, 2(13): 46-47.
- [19] 吴林玉, 许茂盛. 重视人工智能在医学影像中的应用[J]. 中国中西医结合影像学杂志, 2022, 20(4): 307-309.
- [20] 侯祺, 相洋, 吴珊珊, 等. 机器学习算法模型预测体外冲击波碎石治疗输尿管结石的疗效[J]. 中华腔镜泌尿外科杂志(电子版), 2021, 15(4): 280-284.
- [21] Kriegshauser J S, Paden R G, He M, et al. Rapid kV-switching single-source dual-energy CT ex vivo renal calculi characterization using a multiparametric approach: refining parameters on an expanded dataset[J]. Abdom Radiol (NY), 2018, 43(6): 1439-1445.
- [22] Grosse H N, Lennartz S, Salem J, et al. Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an ex-vivo study[J]. Eur Radiol, 2020, 30(3): 1397-1404.

(收稿日期: 2023-04-25)

(校对编辑: 翁佳鸿)