

doi: 10.3969/j.issn.1674-1242.2025.01.004

# 基于 T2WI 和 ADC 图像放射组学特征的机器学习模型 鉴别早期宫颈癌和慢性宫颈炎

杨守义, 吴卓翰, 韩朝钢, 侯悦, 柳佳丽, 马嘉敏, 邓义  
(韶关市妇幼保健院放射科, 广东韶关 512000)

**【摘要】目的** 探讨基于 T2WI 和 ADC 图像放射组学特征的机器学习模型在鉴别早期宫颈癌和慢性宫颈炎中的价值。**方法** 回顾性分析 2019 年 9 月至 2023 年 2 月韶关市妇幼保健院病理确诊的宫颈病变患者, 其中早期宫颈癌患者 34 例(阳性组), 慢性宫颈炎患者 46 例(阴性组)。患者被分为一个训练集(56 例, 其中 24 例阳性和 32 例阴性)和一个独立的测试集(24 例, 其中 10 例阳性和 14 例阴性)。收集每例患者的 T2WI 和 ADC 图像, 通过 3D Slicer 5.4.0 软件和 PyRadiomics 软件模块从宫颈的容积感兴趣区中提取 837 个放射组学特征。采用 3 种数据归一化方法、2 种数据降维方法、4 种特征选择方法和 10 种机器学习模型分类器。在建模过程中, 最优特征参数被设置为 1~12。对训练集进行 10 倍交叉验证, 以确定模型的超参数。采用独立测试集的 AUC 评价模型性能。**结果** 总共建立了 2880 个机器学习模型, 其中 Mean\_PCC\_RFE\_5\_SVM 模型的预测性能最佳。**结论** 基于 T2WI 和 ADC 图像放射组学特征的机器学习模型具有区分早期宫颈癌和慢性宫颈炎的应用价值, 与其他机器学习模型相比, 支持向量机具有更高的诊断效率。

**【关键词】** 宫颈癌; 慢性宫颈炎; 放射组学特征; 机器学习

**【中图分类号】** R445.1、R736.1

**【文献标志码】** A

文章编号: 1674-1242(2025)01-0025-06

## The Value of Machine Learning Models Based on T2WI and ADC Imaging Radiomics Features in Differentiating Early Cervical Cancer from Chronic Cervicitis

YANG Shouyi, WU Zhuohan, HAN Chaogang, HOU Yue, LIU Jiali, MA Jiamin, DENG Yi

(Department of Radiology, Shaoguan Maternal and Child Health Hospital, Shaoguan, Guangdong 512000, China)

**【Abstract】 Objective** To explore the value of machine learning models based on T2WI and ADC imaging radiomics features in differentiating early cervical cancer from chronic cervicitis. **Methods** A retrospective analysis was conducted on 80 patients with pathologically confirmed cervical lesions in Shaoguan Maternal and Child Health Hospital from September 2019 to February 2023, including 34 with early cervical cancer(positive group) and 46 with chronic cervicitis(negative group). The patients were split into a training set (56 cases: 24 positive and 32 negative), and an independent test set (24 cases: 10 positive and 14 negative). T2WI and ADC images of each patient were obtained, and a total of 837 radiomics features were extracted from cervical VOIs by 3D slicer software and the PyRadiomics software module.

收稿日期: 2024-01-06。

作者简介: 杨守义, 副主任医师; 研究方向: 腹部放射学及神经影像学。

通信作者: 邓义, 主任医师, 邮箱 (E-mail): jxdy@126.com。

Three data-normalization methods, two dimensionality-reduction methods, four feature selection methods and ten machine learning methods were used. The optimal number of features in the modeling process was set to 1 to 12. Ten-fold cross-validation was performed with the training data set to determine the model hyperparameters. Model performance was evaluated by AUC with the independent test set. **Results** A total of 2880 machine learning models were established; of these, Mean\_PCC\_RFE\_5\_SVM had the best performance. **Conclusions** Machine learning models based on T2WI and ADC imaging radiomics features have significant value in differentiating early cervical cancer from chronic cervicitis. Compared with other machine learning models, the support vector machine models had higher diagnostic efficiency.

**【Key words】** Cervical Cancer; Chronic Cervicitis; Radiomics Feature; Machine Learning

## 0 引言

宫颈癌是世界上最常见的女性恶性肿瘤之一。根据2020年全球癌症统计数据,宫颈癌是第四大最常见的癌症,也是妇女癌症死亡的第四大原因,2020年全球估计有60.4万例宫颈癌新病例和34.2万例死亡病例<sup>[1]</sup>。早期发现和诊断对宫颈癌的治疗具有重要意义。慢性宫颈炎非常常见,是指宫颈的慢性炎症,通常是由物理/化学刺激或急性炎症缺乏及时治疗引起的<sup>[2]</sup>。慢性宫颈炎通常被认为是宫颈癌的危险因素<sup>[3]</sup>。

磁共振成像(Magnetic Resonance Imaging, MRI)是子宫病变常用的检查方法。慢性宫颈炎和早期宫颈癌在MRI检查中常表现为子宫黏膜信号异常,特别是在T2加权成像(T2-Weighted Imaging, T2WI)和表观弥散系数(Apparent Diffusion Coefficient, ADC)图像上容易发现异常。然而,仅依赖放射科医生的主观视觉评估,要正确区分慢性宫颈炎和早期宫颈癌显然具有巨大的挑战性。随着放射组学和机器学习技术的快速发展,它们在宫颈病变中的应用逐渐增加,已有研究表明它们在宫颈癌<sup>[4-7]</sup>的诊断、病理分级和预后方面具有重要价值,但关于MRI在慢性宫颈炎和早期宫颈癌的鉴别诊断中的应用研究较少。T2WI因能够提供高信噪比、高空间分辨率和良好的软组织对比度的子宫图像,已被应用于宫颈癌的放射组学研究。ADC值反映了水分子在组织中的扩散运动,消除了T2信号对弥散加权成像(Diffusion Weighted Imaging, DWI)的穿透效应。既往研究发现,与ADC值相比,ADC图像随机森林模型在组织学分级、宫旁浸润、分期、淋巴结转移和复发的无创评估方面具有更大

的优势<sup>[8]</sup>。在肿瘤研究领域,已有很多基于T2WI和ADC图像<sup>[9-12]</sup>的放射组学研究。本研究基于宫颈T2WI和ADC图像的放射组学特征,采用多种类型的机器学习模型进行样本训练,探讨这种影像学方法在区分慢性宫颈炎和早期宫颈癌方面的临床应用价值。

## 1 资料与方法

### 1.1 临床资料

所有患者接受核磁共振(Magnetic Resonance, MR)检查前均签署知情同意书,由于这是一项回顾性研究,对患者没有任何不良影响,因此无须获得患者进一步的知情同意。本研究经我院伦理委员会批准。

患者纳入标准:①早期宫颈癌[原位癌和国际妇产科联盟(International Federation of Gynecology and Obstetrics, FIGO) IA-IB2分期]经活检或手术病理证实,并根据《FIGO 2018分期》进行分期<sup>[13]</sup>。②术前15天内采集MRI图像。③图像清晰。

患者排除标准:患有宫颈平滑肌瘤、息肉及2个大于10mm的纳氏囊肿。

根据纳入和排除标准,我们收集了我院2019年9月至2023年2月的80例患者,其中慢性宫颈炎患者46例,早期宫颈癌患者34例。

采用上海华东师范大学磁共振重点实验室提供的软件FAE v0.5.5<sup>[14]</sup>,将患者随机分为训练集( $n=56$ )和测试集( $n=24$ )。

### 1.2 方法

#### 1.2.1 MR扫描

本研究使用1.5T MR扫描仪(GE SIGNA Explorer, 美国)和8通道阵列线圈获得DWI和

T2WI。轴向 DWI 扫描参数：单次回波平面序列，重复时间 (Repetition Time, TR) 6285ms, 最小回波时间 (Echo Time, TE) 72.6ms, 激励次数 (Number of Excitations, NEX) 6, 加速因子 2, 回波平面成像 (Echo Planar Imaging, EPI) 因子 125, 层厚与层间隔分别为 5mm 和 1mm, 视野 (Field of View, FOV) 380mm, 矩阵  $128 \times 128$ ,  $b$  值为 0 和  $600\text{s}/\text{mm}^2$ ; ADC 图像由 MR 扫描仪内置软件自动生成。矢状面 T2WI 扫描参数：TR6648ms, TE106ms, NEX1.5, FOV260mm, 矩阵  $320 \times 256$ , 层厚与层间隔分别为 4mm 和 0.4mm, 脂肪饱和和抑制序列压脂。同时行骨盆冠状位、轴位 T2WI 和轴向 T1WI 检查。

### 1.2.2 图像分割与特征提取

将 T2WI 和 ADC 图像加载到 3D Slicer 5.4.0 软件 (哈佛医学院, 美国) 中进行图像后处理。

图像感兴趣区分割由有 5 年腹部 MRI 诊断经验的放射科医生进行。在每个矢状面脂肪抑制 T2 加权成像 (Fat-Suppression Weighted Imaging, FST2WI) 图像上逐层手工分割整个宫颈黏膜和纤维基质层, 得到容积感兴趣区 (Volume of Interest, VOI)。同样, 对轴向 DWI 序列的  $b=0$  图像进行分层手工分割, 包括整个宫颈黏膜和纤维基质层, 然后将该 VOI 应用到 ADC 图像中, 如图 1 所示。

放射组学特征通过 3D Slicer 5.4.0 软件的扩展软件包 PyRadiomics 提取, 重采样的体积大小为 (1,1,1), 参数 binWidth=25。每个病例均提取 1674 个放射组学特征。放射组学特征的类别包

括一阶统计特征、灰度共生矩阵 (Gray-Level Co-occurrence Matrix, GLCM) 特征、灰度依赖矩阵 (Gray-Level Dependence Matrix, GLDM) 特征、相邻灰度差分矩阵 (Neighboring Gray Tone Difference Matrix, NGTDM) 特征、灰度区域大小矩阵 (Gray-Level Size Zone Matrix, GLSZM) 特征及灰度运行长度矩阵 (Gray-Level Run Length Matrix, GLRLM) 特征。

### 1.2.3 数据清理、拆分和平衡

本研究使用 FAE v0.5.5 软件检查了数据的有效性, 并删除了所有无效的条目。将样本分为训练集 (56 例, 其中 24 例为阳性, 32 例为阴性) 和独立测试集 (24 例, 其中 10 例为阳性, 14 例为阴性)。为了解决训练集中的样本不平衡问题, 本研究通过随机重复的正/负样本进行采样, 直至得到一个平衡的数据集。

### 1.2.4 数据归一化、降维和特征选择

本研究使用 3 种方法对特征矩阵进行归一化。第一种方法是最小-最大归一化, 这是对原始数据的线性转换, 将值映射到  $0 \sim 1$ 。第二种方法是零均值归一化, 将数据转换为遵循标准正态分布, 均值为 0, 标准差为 1。第三种方法是均值归一化, 将数值范围缩放到  $[-0.5, 0.5]$  范围内, 再次得到平均值 0。为了减少特征空间中的高维问题, 本研究确定了每个特征对的相似性, 当特征对的皮尔逊相关系数 (Pearson Correlation Coefficient, PCC) 大于 0.90 时, 随机删除其中一个。本研究还对特

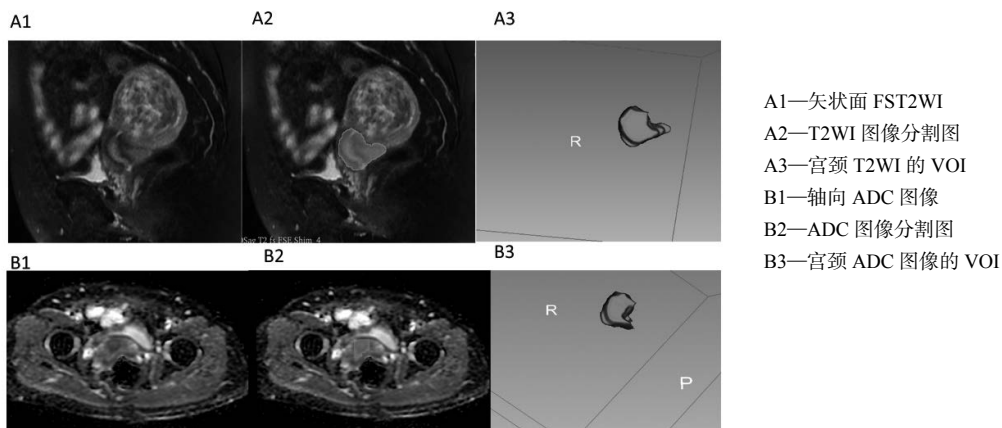


图 1 图像分割  
Fig.1 Image segmentation

征矩阵进行了主成分分析 (Principal Components Analysis, PCA), 这样变换后的特征矩阵的特征向量是相互独立的。

本研究使用了 Relief、递归特征消除 (Recursive Feature Elimination, RFE)、克鲁斯卡尔 - 瓦利斯 (Kruskal-Wallis, KW) H 测试、方差分析 (Analysis of Variance, ANOVA) 4 种特征选择方法。

### 1.2.5 机器学习模型构建

本研究使用 10 种类型的机器学习模型分类器, 包括自适应增强 (AdaBoost)、逻辑回归 (Logistic Regression, LR)、自动编码器 (Autoencoder, AE)、支持向量机 (Support Vector Machine, SVM)、线性判别分析 (Linear Discriminant Analysis, LDA)、随机森林 (Random Forest, RF)、朴素贝叶斯 (Naive Bayes, NB)、高斯过程 (Gaussian Process, GP)、最小绝对值收缩与选择算子 (Least Absolute Shrinkage and Selection Operator, LASSO)、决策树 (Decision Tree, DT)。

本研究对训练集进行了 10 倍交叉验证, 根据交叉验证集上的模型性能建立超参数。

### 1.3 统计分析

本研究采用受试者工作特征 (Receiver Operating Characteristic, ROC) 曲线分析并评价机器学习模型的性能。计算出 ROC 曲线下面积 (Area Under Curve, AUC)、准确性、敏感性、特异性、阳性预测值 (Positive Predictive Value, PPV) 和阴性预测值 (Negative Predictive Value, NPV)。上述所有过程都在 FAE v0.5.5 软件上完成。

## 2 结果

### 2.1 研究对象临床资料

早期宫颈癌患者 34 例, 年龄 25 ~ 68 岁, 年龄中位数为 52 岁, 标准差 (Standard Deviation, SD) 为 10.72。慢性宫颈炎患者 56 例, 年龄 27 ~ 66 岁, 年龄中位数为 45 岁, SD 为 7.57。在 34 例早期宫颈癌患者中, 原位癌 6 例, FIGO IA 1 期 6 例, IA 2 期 4 例, IB 1 期 11 例, IB 2 期 7 例。

### 2.2 模型的建立与比较

利用 3D Slicer 5.4.0 软件和 PyRadiomics 软件模块从每个 VOI 中提取 837 个放射组学特征。其中,

从原始图像中提取了 93 个特征, 包括 18 个一阶特征、24 个 GLCM 特征、16 个 GLSZM 特征、16 个 GLRLM 特征、5 个 NGTDM 特征和 14 个 GLDM 特征; 此外还有 744 个小波特征。每例患者均提取 1674 个放射组学特征, 其中包括 837 个 ADC 图像特征和 837 个 T2WI 特征。

本研究一共开发了 2880 个分类模型, 通过比较它们在交叉验证集、训练集和测试集中的 AUC 值可知, Mean\_PCC\_RFE\_5\_SVM 模型 (以下简称 “SVM 模型”) 的预测性能最佳, 如表 1 所示。

表 1 SVM 模型的性能  
Tab.1 Performance of the SVM model

	测试集	训练集	交叉验证集
数量 / 例	24	56	56
A 组 / 例	10	24	24
B 组 / 例	14	32	32
截断值	0.559	0.420	0.551
准确性	0.875	0.892	0.767
敏感性	0.9	0.916	0.791
特异性	0.857	0.875	0.75
阳性预测值	0.818	0.846	0.703
阴性预测值	0.923	0.933	0.827
尤登指数	0.757	0.791	0.541
AUC	0.943	0.936	0.749
95% 可信区间	0.858-1.000	0.869-1.000	0.609-0.887

注: A 组为宫颈癌病例; B 组为慢性宫颈炎病例。

各数据集 SVM 模型的 ROC 曲线如图 2 所示。

构建模型时所选择的放射性组学特征及 SVM 系数如表 2 所示。

表 2 构建模型时所选择的放射性组学特征及 SVM 系数  
Tab.2 The radiomics features and SVM coefficients chosen for model construction

放射性组学特征	SVM 系数
ADC_SmallAreaEmphasis	-1.399
ADC_10Percentile	-2.253
T2WI_SmallDependenceLowGrayLevelEmphasis	1.491
T2WI_Skewness	-1.853
T2WI_GrayLevelNonUniformity	1.276

## 3 讨论

放射组学是一种医学计算机辅助技术, 它通过计算机算法将医学图像信息转换为一系列数据, 然后对其进行分析, 建立分类或预测模型, 以识别具有潜在临床价值的生物标志物并协助临床决策<sup>[15]</sup>。放射组学特征往往难以被直观地解释, 但它们可以反映生理方面的信息, 如组织结构特征、病理表型、

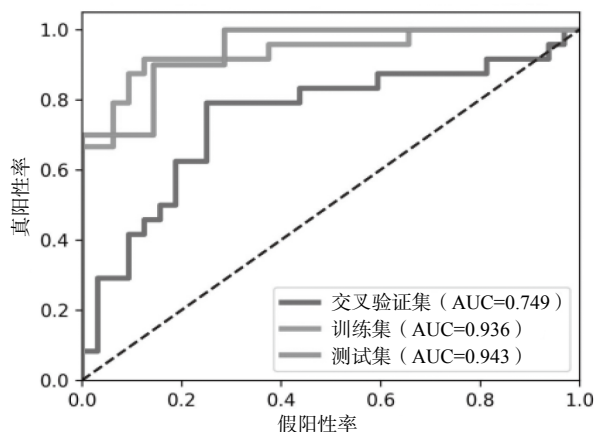


图2 各数据集 SVM 模型的 ROC 曲线  
Fig.2 The ROC curve of the data sets used in SVM model

肿瘤异质性等<sup>[16,17]</sup>。机器学习建模在医学领域得到了广泛的应用。许多研究表明，机器学习建模优于统计建模<sup>[18,19]</sup>。

在临床实践中，慢性宫颈炎和早期宫颈癌的 MRI 征象重叠，依靠放射科医生的视觉观察往往不易鉴别，因此本研究探索了机器学习模型在这方面的应用价值。

本研究结果显示，SVM 模型是 2880 个机器学习模型中性能最好的，其训练集 AUC=0.936，测试集 AUC=0.943，交叉验证集 AUC=0.749。SVM 模型选择了 5 个放射组学特征，具体如下：GLSZM 特征 ADC\_SmallAreaEmphasis 是对小面积分布的度量，值越大，表示面积越小，纹理越细；两个一阶特征，其中 ADC\_10Percentile 反映了 ADC 值的第 10 个百分位数，T2WI\_Skewness 指测量值相对于平均值分布的不对称性；两个 GLDM 特征，其中 T2WI\_SmallDependenceLowGrayLevelEmphasis 测量具有较低灰度值的小相关性的联合分布，T2WI\_GrayLevelNonUniformity 量化测量图像中灰度强度值的相似性<sup>[20,21]</sup>。

SVM 是一种有效和稳定的分类器。本研究使用线性核函数，因为它可以更容易地解释最终模型的特征系数。与非线性 SVM 相比，线性 SVM 的训练速度更快，需要优化的参数更少，在综合研究中具有显著优势。从少量样本中识别具有多个变量（交互权值）的复杂系统需要一个简单的预测器，

因此 SVM 也适用于小样本<sup>[22]</sup>。SVM 是应用最广泛的分类器之一，以往的研究也表明 SVM 模型在分类中表现良好<sup>[23-25]</sup>。

本研究表明，不同的数据归一化方法、数据降维方法、特征选择方法和机器学习模型分类器对模型的 AUC 有影响，因此本研究积极应用不同的方法来选择性能较好的模型。

本研究存在一定的局限性。首先，本研究不是一个多中心研究，基于 T2WI 和 ADC 图像的机器学习模型的诊断效率有待进一步讨论。其次，样本量相对较少，慢性宫颈炎组和早期宫颈癌组数据不平衡，但本研究采用了向上抽样的方法来保持数据平衡。

综上所述，基于 T2WI 和 ADC 图像放射组学特征的机器学习模型可以区分慢性宫颈炎和早期宫颈癌。与其他机器学习模型相比，支持向量机模型具有更高的诊断效率。

#### 参考文献

- [1] SUNG H, FERLAY J, SIEGEL R L, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA: A Cancer Journal for Clinicians*, 2021, 71(3): 209-249.
- [2] HESTER E E, MIDDLEMAN A B. A clinical conundrum: chronic cervicitis[J]. *Journal of Pediatric and Adolescent Gynecology*, 2019, 32(3): 342-344.
- [3] PANDEY N O, CHAUHAN A V, RAITHATHAN S, *et al.* Association

- of TLR4 and TLR9 gene polymorphisms and haplotypes with cervicitis susceptibility[J]. **Scientific Reports**, 2019(9): 1-11.
- [4] JIANG X, LI J, KAN Y, *et al.* MRI based radiomics approach with deep learning for prediction of vessel invasion in early-stage cervical cancer[J]. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, 2021, 18(3): 995-1002.
- [5] MENG J, ZHU L, ZHU L, *et al.* Whole-lesion ADC histogram and texture analysis in predicting recurrence of cervical cancer treated with CCRT[J]. **Oncotarget**, 2017, 8(54): 92442-92453.
- [6] JAJODIA A, GUPTA A, PROSCH H, *et al.* Combination of radiomics and machine learning with diffusion-weighted mr imaging for clinical outcome prognostication in cervical cancer[J]. **Tomography**, 2021, 7(3): 344-357.
- [7] MAYERHOEFER M E, MATERKA A, LANGS G, *et al.* Introduction to radiomics[J]. **Journal of Nuclear Medicine**, 2020, 61(4): 488-495.
- [8] YAMADA I, OSHIMA N, MIYASAKA N, *et al.* Texture analysis of apparent diffusion coefficient maps in cervical carcinoma: correlation with histopathologic findings and prognosis[J]. **Radiology: Imaging Cancer**, 2020, 2(3): 1-11.
- [9] HU P, CHEN L, ZHOU Z. Machine learning in the differentiation of soft tissue neoplasms: comparison of fat-suppressed T2WI and apparent diffusion coefficient (ADC) features-based models[J]. **Journal of Digital Imaging**, 2021, 34(5): 1146-1155.
- [10] WANG X Y, LIU Y, WANG W, *et al.* The applied research of simultaneous image acquisition of T2-weighted imaging (T2WI) and diffusion-weighted imaging (DWI) in the assessment of patients with prostate cancer[J]. **Asian Journal of Andrology**, 2019, 21(2): 177-182.
- [11] ZHANG W, ZHANG W, LI X, *et al.* Predicting tumor perineural invasion status in high-grade prostate cancer based on a clinical-radiomics model incorporating T2-weighted and diffusion-weighted magnetic resonance images[J]. **Cancers**, 2022, 15(1): 1-15.
- [12] CHU C, WANG F, ZHANG H, *et al.* Whole-volume ADC histogram and texture analyses of parotid glands as an image biomarker in evaluating disease activity of primary sjögrens syndrome[J]. **Scientific Reports**, 2018, 8(1):1-9.
- [13] BHATLA N, AOKI D, SHARMA D N, *et al.* Cancer of the cervix uteri[J]. **International Journal of Gynecology & Obstetrics**, 2018,143: 22-36.
- [14] SONG Y, ZHANG J, ZHANG Y, *et al.* FeAture explorer (FAE): a tool for developing and comparing radiomics models[J]. **PLoS One**, 2020, 15(8): 1-12.
- [15] RIZZO S, BOTTA F, RAIMONDI S, *et al.* Radiomics: the facts and the challenges of image analysis[J]. **European Radiology Experimental**, 2018, 2(1):1-8.
- [16] TOMASZEWSKI M R, GILLIES R J. The biological meaning of radiomic features[J]. **Radiology**, 2021, 298(3): 505-516.
- [17] KOCAK B, Department of Radiology Istanbul Training and Research Hospital I, DURMAZ E S, *et al.* Radiomics with artificial intelligence: a practical guide for Beginners[J]. **Diagnostic and Interventional Radiology**, 2019, 25(6): 485-495.
- [18] CHAO X, WANG S, LANG J, *et al.* The application of risk models based on machine learning to predict endometriosis—associated ovarian cancer in patients with Endometriosis[J]. **Gynecologica Scandinavica**, 2022, 101(12): 1440-1449.
- [19] SUN L, ZHU W, CHEN X, *et al.* Machine learning to predict contrast-induced acute kidney injury in patients with acute myocardial infarction[J]. **Frontiers in Medicine**, 2020, 7: 1-11.
- [20] ZWANENBURG A, VALLIÈRES M, ABDALAH M A, *et al.* The Image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping[J]. **Radiology**, 2020, 295(2): 328-338.
- [21] KIM J K, PARK H W. Statistical textural features for detection of microcalcifications in digitized mammograms[J]. **IEEE Transactions on Medical Imaging**, 1999, 18(3): 231-238.
- [22] MAETSCHKE S R, MADHAMSHETTIWAR P B, DAVIS M J, *et al.* Supervised, semi-supervised and unsupervised inference of gene regulatory networks[J]. **Briefings in Bioinformatics**, 2013, 15(2): 195-211.
- [23] HAO P, DENG B Y, HUANG C T, *et al.* Predicting anaplastic lymphoma kinase rearrangement status in patients with non-small cell lung cancer using a machine learning algorithm that combines clinical features and CT Images[J]. **Frontiers in Oncology**, 2022,12:1-11.
- [24] CHEN X, HUANG Y, HE L, *et al.* CT-based radiomics to differentiate pelvic rhabdomyosarcoma from yolk sac tumors in children[J]. **Frontiers in Oncology**, 2020,10:1-8.
- [25] AREFAN D, CHAI R, SUN M, *et al.* Machine learning prediction of axillary lymph node metastasis in breast cancer: 2D versus 3D radiomic features[J]. **Medical Physics**, 2020, 47(12): 6334-6342.